Dissertation
submitted to the
Combined Faculty for the Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M. Sc. Omar Hammouda
born in: Cairo, Egypt
Oral examination: December 3rd 2021

Novel genetic predictors of human cardiovascular diseases validated by CRISPR in medaka

Referees: Prof. Dr. Joachim Wittbrodt
Prof. Dr. med. Johannes Backs

Abstract

Accounting for nearly a third of global deaths, cardiovascular diseases (CVDs) represent a major health burden to our society. Most CVDs can be prevented with early diagnosis and intervention. Several of the major risk factors, which contribute in a cumulative manner to the onset and development of CVDs, can be regulated. Among the major risk factors, the resting heart rate is a vital indicator that reflects underlying physiology and is a recognised predictor for cardiovascular and overall mortality. Early life-style changes or medical interventions can foster control of modifiable risk factors and in turn, by reducing or preventing CVDs, positively impact on the patient's longevity and quality of life.

The crucial contribution of genetics to the onset and progression of CVDs has led the discovery of its diagnostic and prognostic genetic determinants utilising human genome-wide association studies (GWAS). These association studies correlate hundreds of genes and CVDs after diagnosis and therefore constitute a valuable resource for potential candidate discovery. However, to date many of these genes remain uncharacterized or without link to heart development or function. Moreover, genes correlated to CVDs, still lack causality and require functional validation in animal models. Due to the lack of rapid gene validation workflows and the pressured demand for mechanistic insights, further focus is mainly directed towards genes with known and established functions rather than novel uncharacterized genes. Consequently, in the realm of functional validation, a negative loop of re-discovery is created.

In this thesis, I addressed this negative loop by targeting and validating an understudied domain of human GWAS candidates with no known connection to the heart. I built on the high conservation of cardiac function and underlying genetics between teleost fish and humans, and developed a sensitive and robust functional gene validation assay using an isogenic strain of the Japanese rice fish medaka. By combining highly efficient CRISPR/Cas9-mediated mutagenesis with high-throughput heart rate phenotyping, I identified novel candidate genes with potential predictive power for human CVDs, and in turn redirected the spotlight onto these genes as putative diagnostic and prognostic CVD genetic markers.

Zusammenfassung

Herz-Kreislauf-Erkrankungen (HKE) sind für fast ein Drittel der weltweiten Todesfälle verantwortlich und stellen dadurch eine enorme gesundheitliche Belastung für unsere Gesellschaft dar. Die meisten HKE können durch eine frühzeitige Diagnose und medizinische Eingriffe verhindert werden. Mehrere der wichtigsten Risikofaktoren, die das Auftreten und die Entwicklung von HKE gegenseitig begünstigen, können positiv beeinflusst werden. Der Ruhepuls ist dabei, als Risikofaktor, einer der wichtigsten Indikatoren der die zugrunde liegende Physiologie widerspiegelt und als anerkannter Prädiktor für die Kardiovaskuläre- sowie die Gesamtmortalität dienen kann. Dabei können HKE mit Hilfe von frühzeitiger Veränderung des Lebensstils sowie durch medizinische Eingriffe, welche die Kontrolle veränderbarer Risikofaktoren fördern, reduziert oder gar verhindert werden was sich wiederum positiv auf die Langlebigkeit und Lebensqualität des Patienten auswirken kann.

Der wesentliche Beitrag genetischer Faktoren zum Auftreten und Fortschreiten von HKE hat zur Entdeckung ihrer diagnostischen und prognostischen genetischen Determinanten unter Verwendung von humangenomweiten Assoziationsstudien (GWAS) geführt. Diese Assoziationsstudien haben Hunderte von Genen und HKE nach der Diagnose miteinander korreliert und stellen daher eine wertvolle Ressource für die Entdeckung potenzieller genetischer Determinanten dar. Bis heute sind jedoch viele dieser Gene nicht charakterisiert oder es besteht keinerlei ursächlicher Zusammenhang mit Bezug auf Herzentwicklung oder -funktion. Darüber hinaus fehlt es jenen mit HKE korrelierten Genen noch an Kausalzusammenhang wodurch eine funktionelle Validierung in Tiermodellen unerlässlich ist. Mangels schneller Genvalidierungsabläufe und durch die stetige Nachfrage nach mechanistischen Erkenntnissen werden im weiteren Verlauf tendenziell eher Gene mit bekannten und etablierten Funktionen anstatt neuartiger, nicht charakterisierter Gene untersucht. Es entsteht somit aus logischer Konsequenz eine negative Wiederentdeckungsschleife im Bereich der Funktionsvalidierung von HKE Genen.

In dieser Dissertation habe ich das grundsätzliche Problem dieser negativen Wiederentdeckungsschleife ausgearbeitet, indem ich wenig erforschte, menschliche GWAS Kandidatengene ohne bekannte Verbindung zum Herzen untersucht und validiert habe. Ich baute dabei auf die physiologische Ähnlichkeit der Herzfunktion zwischen Echten Knochenfischen und Menschen, welche auf einem hohen Grad der Konservierung der zugrunde liegenden Gene basiert, und entwickelte unter Verwendung eines isogenen Stammes des japanischen Reisfischs Medaka einen empfindlichen und robusten funktionellen Genvalidierungsablauf. Durch die Kombination hocheffizienter CRISPR/Cas9 gesteuerter Mutagenese mit einer Hochdurchsatz-Herzfrequenz-Phänotypisierung identifizierte ich dabei neue Kandidatengene mit potenzieller Vorhersagekraft für menschliche HKE wodurch ich diese Gene als mutmaßliche diagnostische und prognostische HKE Genmarker hervorheben konnte.

Contents

Abbreviations

A	Atrium
ABCB1	ATP binding cassette subfamily B member 1
ABCB4	ATP binding cassette subfamily B member 4
Actb	Beta-actin
AP	Action potential
ASD	Atrial septal defect
ATP8B4	ATPase phospholipid transporting 8B4
AV	Atrioventricular
BAG3	BCL2 associated athanogene 3
BF	Bright-field
bp	Base pair(s)
bpm	Beats per minute
CABP2	Calcium binding protein 2
CABP4	Calcium binding protein 4
Cas	CRISPR-associated protein
CASQ2	Calsequestrin 2
CCDC141	Coiled-coil domain containing 141
CCTop	CRISPR/Cas9 target online predictor
CDC42	Cell division control protein 42 homolog
CEL I	Celery nuclease 1
CEP85L	Centrosomal protein 85 like
CLCNKA	Chloride voltage-gated channel K a
cmlc2	Cardiac myosin light chain 2
CMYA5	Cardiomyopathy associated 5
CNOT1	CCR4-NOT transcription complex subunit 1
COL9A1	Collagen type IX alpha 1 chain
COS	Centre for Organismal Studies
CR	CRISPR-Cas9
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CVD	Cardiovascular disease
ddPCR	Digital droplet PCR
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dpf	Days post fertilisation
DSB	Double-strand break
DUOX	Dual oxidase
E. coli	*Escherichia coli*
ECG	Electrocardiogram
EDN1	Endothelin 1
eDNA	Environmental DNA
EDTA	Ethylendiamine tetraacetic acid
eGFP	Enhanced green fluorescent protein
EML6	Echinoderm microtubule-associated protein-like 6
ERM	Embryonic rearing medium
ESC	Embryonic stem cells
EtBr	Ethidium bromide
EtOH	Ethanol
FB	Fin-clip lysis buffer
Fok1	*Flavobacterium okeanokoites* nuclease 1

fps	Frames per second
fwd	Forward
gDNA	Genomic DNA
GFP	Green fluorescent protein
GIGYF1	GRB10 interacting GYF protein 1
GIT2	GIT ArfGAP2
GRASP	Genome-wide repository of associations between SNPs and phenotypes
GRID2	Glutamate ionotropic receptor delta type subunit 2
GWAS	Genome-wide association study
H	Heart
HBIGS	Heidelberg Biosciences International Graduate School
HCN4	Hyperpolarization activated cyclic nucleotide gated potassium channel 4
HDR	Homology-directed repair
heiCas9	High efficiency tagged-Cas9
hGWAS	Human heart-GWAS
HOMEZ	Homeobox and leucine zipper encoding
hpf	Hours post fertilisation
HR	Heart rate
HRMA	High-resolution melt analysis
ICE	Inference of CRISPR edits
IDAA	Indel detection by amplicon analysis
InDel	Insertion or deletion
ISH	*In situ* hybridization
KCNH2	Potassium voltage-gated channel subfamily H member 2
KIT	Karlsruhe Institute of Technology
LB	Lysogeny broth
LDL	Low-density lipoprotein
LDLR	LDL receptor
MAML3	Mastermind like transcriptional coactivator 3
MINAR1	Membrane integral NOTCH2 associated receptor 1
mRNA	Messenger RNA
MUS81	MUS81 structure-specific endonuclease subunit
MYRF	Myelin regulatory factor
NACA	Nascent polypeptide associated complex subunit alpha
NGS	Next generation sequencing
NHEJ	Nonhomologous end-joining
NKX2-5	NK2 homeobox 5
NUBP2	Nucleotide binding protein 2
o/n	Overnight
OCA2	Oculocutaneous albinism 2
OGDH	2-oxoglutarate dehydrogenase
oNLS	Optimized nuclear localization signal
OR124-2	Odorant receptor, family E 124, member 2
OR5AU1	Olfactory receptor family 5 subfamily AU member 1
PADI2	Peptidyl arginine deiminase 2
PADI4	Peptidyl arginine deiminase 4
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
PFA	Paraformaldehyde
PIEZO1	Piezo type mechanosensitive ion channel component 1
PLEKHA8	Pleckstrin homology domain containing A8
PLG	Plasminogen
PPP1R9A	Protein phosphatase 1 regulatory subunit 9A

rev	Reverse
RGS3	Regulator of G protein signalling 3
RNA	Ribonucleic acid
RNF207	Ring finger protein 207
RNP	Ribonucleoprotein
ROI	Region of interest
RT	Room temperature
rx2	Retinal homeobox protein 2
SCMH1	SCM polycomb group protein homolog 1
SCN4AB	Sodium voltage-gated channel alpha subunit 4 B
SCN5A	Sodium voltage-gated channel alpha subunit 5
SD	Standard deviation
sgRNA	Single guide RNA
SH2B3	SH2B adaptor protein 3
SLC17A3	Solute carrier family 17 member 3
SLEDGE	Swift large-scale examination of directed genome editing
SMG6	SMG6, nonsense mediated mRNA decay factor
SNP	Single nucleotide polymorphism
SpCas9	*Streptococcus pyogenes* Cas9
SSPO	SCO-spondin
T7EI	T7 endonuclease 1
TAE	Tris-acetate-EDTA-buffer
TALEN	Transcription activator-like effector nucleases
TB	Terrific broth
TIDE	Tracking of indels by decomposition
tracrRNA	Trans-activating CRISPR RNA
TRAPPC12	Trafficking protein particle complex 12
TRAPPIII	Transport protein particle III
TTL	Tubulin tyrosine ligase
TTN	Titin
U	Unit
UFSP1	UFM1 specific peptidase 1
UV	Ultraviolet radiation
V	Ventricle
VEPH1	Ventricular zone expressed PH domain containing 1
WT	Wild-type
XYLB	Xylulokinase
ZFHX3	Zinc finger homeobox 3
ZFN	Zinc-finger nucleases

Contributions

In the following, people that contributed to the work presented in this thesis are listed:

Dr. Thomas Thumberger helped with sgRNA and primer design as well as CRISPR-related troubleshooting. He contributed to the quantification analysis of eye pigmentation in *oca2* crispants and its illustration in Figure 2.10. He also aided in the illustration of Figure 2.1-2.3 and Figure 2.6-2.7.

Frank Böttger and **Gero Hofmann** manufactured the 96-well plate mortar ('the Hammer') and its respective design and schematics illustration in Figure 2.5.

Dr. Jakob Gierten generated the transgenic cardiac dual-reporter line *cmlc2::eGFP cmlc2::H2A-mCherry*.

Tanja Kellner helped with synthesis and quality check of sgRNAs.

Alex Cornean designed and synthesized the *gfp_T1* sgRNA and helped with the live confocal *in vivo* imaging setup.

Meng Yue Wu under my supervision helped with synthesis and testing of sgRNAs, and genotyping. She also helped with heart rate analysis using the *HeartBeat* software and with the phenotyping of embryos for the data illustrated in Figure 2.20.

Verena Kaul under my supervision helped with synthesis and testing of sgRNAs, and genotyping. She also helped with heart rate analysis using the *HeartBeat* software.

Gasser ElMissiery under my supervision helped with phenotyping and genotyping of *scn4ab* F1 adult fish and F2 embryos for the data in Figure 2.22.

1

Introduction

1.1 Cardiovascular diseases: a global burden and universal enemy

Cardiovascular diseases (CVDs) are the leading cause of death globally, accounting for nearly 18 million deaths per year (~32% of all deaths worldwide) according to the World Health Organisation [1]. CVDs encompass a wide range of cardiac diseases involving the heart (e.g., cardiomyopathies, congenital heart diseases, arrhythmias, heart failure and myocardial infarctions), as well as vascular diseases involving the blood vessels (e.g., stroke, coronary artery disease and aortic aneurysm). CVDs are typically of multifactorial origin, with genetics crucially co-contributing to the complexity of CVDs in terms of onset and progression. The risk of CVD onset increases approximately threefold per decade of life, with a tenfold increased risk between ages 50 and 80 [2]. Among other risk factors, hypertension and hypercholesteremia represent major CVD risk factors that constantly cumulate throughout one's lifetime. Hypertension leads to the dysfunction of blood vessels, heart vasculature and conduction system through adaptive cardiac remodelling against extended hemodynamic changes [3]. Hypertension accounted for an estimated 54% of all strokes and 47% of ischemic heart disease events globally [4]. Hypercholesterolemia, or elevated blood cholesterol (especially low-density lipoprotein (LDL)) levels, results in atherosclerotic plaques and higher risk of thrombosis-mediated stroke or myocardial infarctions [5]. The resting heart rate is another highly important risk indicator reflecting underlying conditions such as elevated blood pressure, and is associated with overall mortality in patients with hypertension [6] and ischemic heart diseases [7–11]. The resting heart rate is additionally a standalone risk predictor for cardiovascular and all-cause mortality independent of other risk factors in CVD patients [9,12]. An elevated resting heart rate corresponded to an increased risk of CVD-related hospitalization or death by 3% per increased beat and 16% per 5-beat increase from baseline heartbeat [10].

1.2 Early diagnosis and prediction are key for CVDs prevention

Most CVDs can be prevented with early diagnosis and early intervention. An estimated 85% of CVD-related deaths are caused by heart attack or strokes. Both can be effectively prevented or managed by maintaining a healthy lifestyle or medical interventions. Indeed, the majority of CVD risk factors (hypertension, hypercholesterolemia, obesity, smoking, physical inactivity, alcohol consumption and diet) are similarly 'modifiable' through lifestyle or medical interventions resulting in improved CVD prognosis. Anti-hypertensive drug treatments have been reported to reduce stroke risk by up to 40% [13–16] and myocardial infarctions risk by 21% [16]. Cholesterol management with statins have been reported to decrease LDL levels by up to 70% [17–19] and accordingly reduce risk of developing CVDs. Similarly, inhibition of *Proprotein Convertase Subtilisin/Kexin Type 9* (*PCSK9*), a gene coding for a crucial enzyme in cholesterol homeostasis that by default leads to higher plasma LDL levels [20], significantly lowers LDL levels up to 60% [21,22]. Similarly, lowering the resting heart rate through surgery or drug treatment has been reported to improve clinical outcomes of CVDs [6,10,23,24]. Therefore, early diagnosis or genetic risk prediction of CVDs is of paramount importance to change the intervention's focus from treatment to prevention, substantially impacting CVD prognosis and patient longevity.

1.3 The quest for new diagnostic and predictive genetic markers for CVDs

The importance of early CVD diagnosis on the onset and progression of CVDs and the impact on the patient's longevity and quality of life has led to the search of CVD-related determinants to use for diagnostic or even predictive purposes. Genetics play a crucial role in CVD onset and progression, either directly by affecting heart development and function (e.g., congenital heart diseases, cardiomyopathies and cardiac arrhythmias), or by affecting the abovementioned risk factors such as hypertension, hypercholesterolemia and the resting heart rate [25–28]. Although CVD-research has been extensively performed for decades, yet neither all causative genetic factors have been identified, nor are the mechanisms of CVDs development fully elucidated. Many of the essential genes in heart development and function have been identified through forward genetics by linkage mapping and tracing of familial cases. For example, the LDL receptor gene (*LDLR*) was identified in a case of familial hypercholesterolemia and was the first demonstration of traceable CVD-causing mutation [29]. Furthermore, mutagenesis screens have identified even more cardiac key players [30,31]. However, the deleterious mutations in single genes driving large causal effects and diseases, also known as mendelian diseases, are relatively rare (less than 5%) and have low impact at the population level [26,27]. On the other hand,

CVDs appear to have a rather complex polygenic background, influenced by a multitude of polymorphisms yielding weak effects individually or in combination (additive versus epistatic interactions). These polymorphisms occur more frequently in a population and hence might have a more significant impact at the population level despite a weak, negligible impact at the individual level [26,27]. These polymorphisms therefore represent as attractive genetic markers with great potential in CVD diagnosis and prevention, as well as in the pharmacogenetic field. To achieve high diagnostic and prognostic capabilities, a deeper understanding of the genetic key players involved in cardiovascular functions is required to better assess the genetic predispositions of CVDs. Identifying and understanding the role of these CVD-relevant polymorphisms and their underlying gene functions is the key to generate a genetic blueprint of CVDs enabling accurate risk assessment and prediction.

1.4 The need of GWAS associations filtering by functional validation

The demand for novel CVD genetic markers has initiated the quest to identify such markers via human genome-wide association studies (GWAS) performed extensively over the last two decades. GWAS revealed thousands of polymorphisms with associations to various CVDs like congenital heart diseases [32] and cardiomyopathy [33]. These investigations also revealed polymorphisms associated to CVD-related phenotypes and major risk factors such as elevated blood pressure [34], hyperlipidaemia [35,36], elevated resting heart rate [12,37] and cardiac conduction and rhythm disorders [38–40]. However, despite the richness of these human CVD-GWAS associations as a resource for the discovery of novel genetic diagnostic and predictive markers, such bioinformatically driven associations still lack causality and require functional validation in animal models. Nowadays with increasing affordability of next generation sequencing (NGS), the number of GWAS studies is on the rise, and the absence of available rapid and cost-effective high-throughput gene validation pipelines lead to an imbalance between the output of associations and their consequent functional validation.

1.5 The negative loop of re-discovery

The vast majority of the disease-associated polymorphisms (more than 90%) lie in non-coding regions [41], the black box of our genetic code. These regions contain yet to be identified key gene regulatory elements (e.g., promoters, insulators and enhancers), which only in the past decade have started to get their deserved attention [42–44]. Nevertheless, although polymorphisms in coding regions (i.e., in exons) represent the minority of total disease-associated polymorphisms (less than 10%) [41], they still relate to hundreds of genes. Many of these genes have been not previously connected to heart development or

function, and some are even uncharacterised. The lack of quick gene validation pipelines and the constant demand for rapid mechanistic insights has, however, led to a negative loop of re-discovery. It resulted in the convenient shift of focus towards genes with a known connection to heart development or function rather than characterization of new genes or newly identified cardiac relations. Thus, before delving into the more complicated validation of polymorphisms, it is necessary to have at least some insight into the biological functions of the identified genes and their actual relevance to the associated-phenotype. Consequently, hundreds of GWAS candidate gene associations can be filtered down via experimental validation to the most disease-relevant candidates for further in-depth characterization. While mice are widely represented as the laboratory standard with highest relevance to humans, they are not really compatible with high-throughput gene validations due to their high-cost and space maintenance as well as long generation times. Additionally, their inaccessible hearts represent another obstacle for cardiac investigations. Invertebrate models such as *Drosophila melanogaster* have been used for concurrent identification and validation of CVD-associated genes from human GWAS using RNAi approaches [40,45,46]. Although, drosophila provide rapid insights into gene functionalities, investigations in higher vertebrate model are still a necessity. Alternatively, vertebrate teleost models such as the Japanese rice fish medaka (*Oryzias latipes*) offer a reasonable compromise between throughput and translational relevance.

1.6 Medaka as a vertebrate model to study cardiovascular diseases

Teleost fish species such as medaka and zebrafish *(Danio rerio*) have been exclusive vertebrate model organisms for decades [47,48]. Their extrauterine development within transparent chorions allows direct non-invasive visual access into the various organs throughout development. The additional availability of various genetic tools renders them excellent vertebrate model organisms for developmental, genetic and molecular studies of human pathologies [48–51]. Furthermore, their small size and economic maintenance, in addition to their short generation times and high fecundity makes them perfect vertebrate models for high-throughput applications such as genetic and compound screens [52–59]. One of the major limitations of human genetics is the low resolution of phenotype-associations in GWAS due to large background genetic noise, making it hard to discriminate between disease-causing and individual variants. Meanwhile, one distinctive property of medaka is its high tolerance to inbreeding, which has led to the establishment of highly isogenic laboratory strains [60], as well as the first wild vertebrate inbred panel as a valuable genomic resource [61–63]. The isogenic medaka inbred panel may therefore hold the solution towards the accurate identification of CVD-causing polymorphisms.

The aforementioned advantages, especially the exposed heart throughout embryonic and larval stages, have primed teleost fish for cardiovascular research [51]. Compared to a four-chambered heart in mammals, teleost fish have a simple, two-chambered heart with one

atrium and one ventricle (Figure 1.1). Nevertheless, teleost hearts are highly comparable to mammals with respect to development, function and physiology [51,64–69]. Many of specialised cell types and structures (e.g., trabeculae, atrioventricular valve and pacemaker) [69], as well as signalling pathways involved in early development are conserved between the species [65]. Orthologues of main human cardiac ion channel genes have been also found in teleost fish. Furthermore, teleost heart function parameters, such as electrocardiograms (ECG) and components thereof (e.g., P wave, QRS complex, T wave and QT interval) as well as atrial and ventricular action potentials (AP) were shown to be highly similar to those in humans [64,66–68] (Figure 1.1). Teleost heart rates (~100-160 beats per min (bpm)) are notably closer to human heart rates (60-100 bpm) than murine heart rates are (~ 500-700 bpm), and medaka hearts are closer to human hearts than zebrafish with respect to the lack of cardiac regeneration [70]. Interestingly, circadian heart rate rhythms closely resembling nocturnal heart rate dipping in healthy humans [71] were also observed in medaka embryos [72].

However, extensive studies have been performed in zebrafish [30,51,65,73–78], being the most popular teleost, and accordingly a wealth of information is available regarding cardiac development in zebrafish. On the other hand, cardiac development in medaka has hardly been investigated. A detailed description of the various stages in medaka development has been reported by Takashi Iwamatsu, which has and still serves as the ultimate medaka reference [79]. However, details surrounding heart development are only superficial, only morphological milestones such as the appearance of the heart anlage at around 38 hours post fertilization (hpf), the onset of heartbeat between 40-44 hpf, looping of the heart from 2.5 to ~6 days post fertilization (dpf) were described [79].

1.7 The demand for high-throughput applications to study heart development and function

Transparency and small size of teleosts offer an exceptional advantage over other higher vertebrate models, that is an exclusive non-invasive view on the developing heart. This allowed researchers to obtain unique insights into cardiac development and function via *in vivo* imaging as well as in the context of high-throughput mutagenesis and drug screens [56,57,80–82]. The heart rate is a clinically important indicator for overall cardiac performance, aberrant heart rates may provoke as well as reflect various underlying phenotypes that directly or indirectly affect cardiac function. Moreover, the heart rate is also widely regarded as a vital risk factor with predictive potential of CVD onset and overall mortality [7,10,12,40]. A 5-beat increase in resting heart rate corresponded to a 16% increased risk of CVD-related hospitalization or death [10], and a 20% increased risk of overall mortality [12].

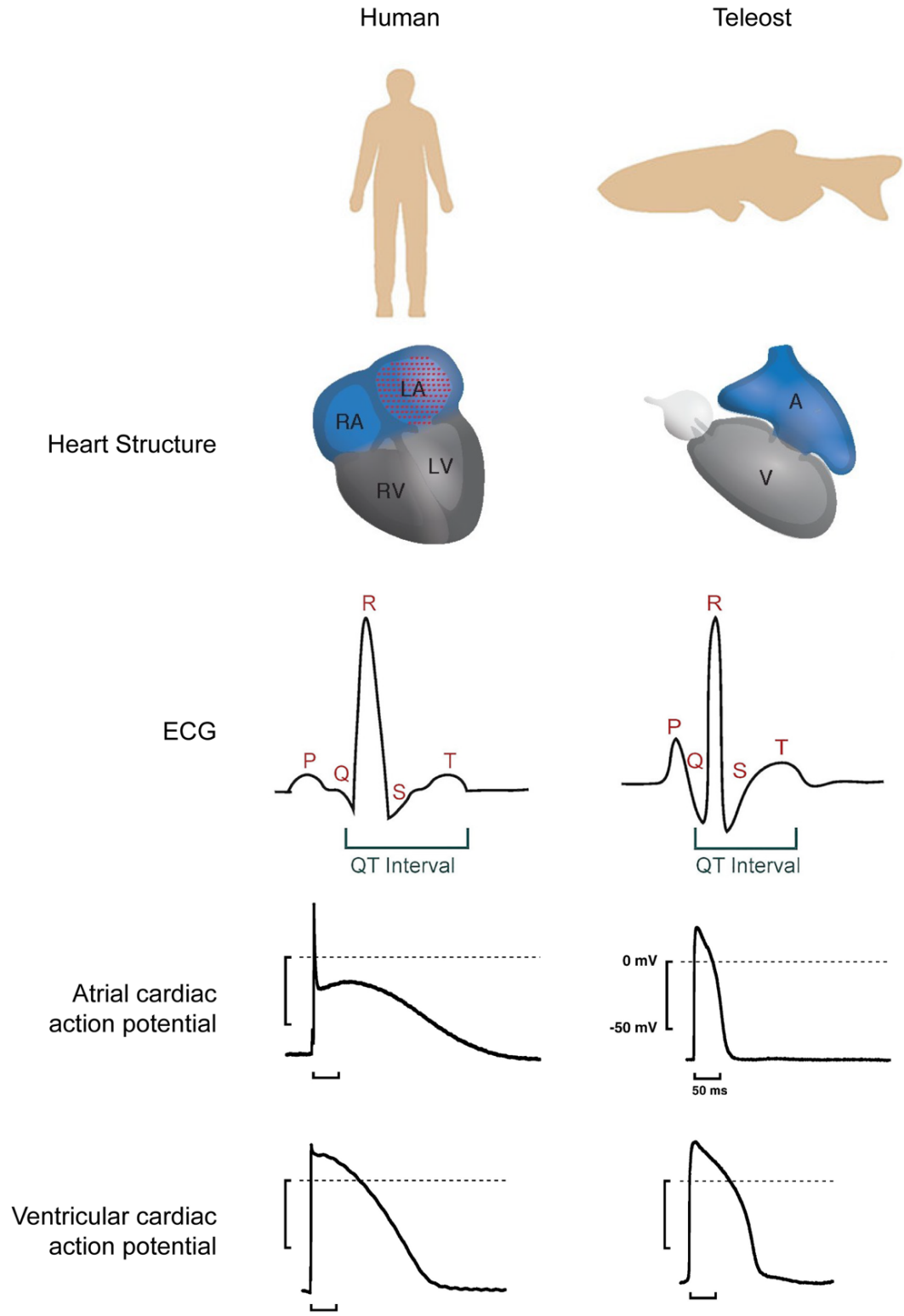


Figure 1.1 Basic cardiac functions are comparable between Humans and Teleost fish
Humans have a four-chambered heart compared to teleost's two-chambered heart. "ECG recordings of embryonic and adult zebrafish hearts closely resemble those obtained from humans, with distinguishable P wave, QRS complex and T wave, allowing quantification of comparative parameters such as QT interval". Comparable shapes of atrial and ventricular cardiac action potentials between humans and adult zebrafish. Heart illustrations adapted and modified from [83], Human and fish icons as well as ECGs with respective caption reprinted and modified from [69]. Atrial and Ventricular potentials were reprinted and modified from [64], with permission from Elsevier.

Although many clinically relevant aspects of cardiac development and function can be addressed and investigated, the lack of automated feature detection algorithms still represents a bottle neck for high-throughput phenotyping. The heart rate is a relatively simple feature to detect and quantify, accordingly, numerous assays have been developed for the automated imaging and quantification of heart rates in teleost fish [55,56,82,84–86]. Although the assays were constantly improving the feature detection algorithms, they still had shortcomings, limiting them in the context of high-throughput applications. A major drawback is the heavy reliance on immobilization techniques such as highly-viscous media, low-melting agar or sedation as the assays were mainly developed for dechorionated embryos or hatched larvae. That is most probably due to the preferential choice of zebrafish for such assays, which develop and hatch relatively fast. These time-consuming mounting procedures require extensive manual handling for specimen orientation, which is not a desirable feature when considering a high-throughput approach. Alternatively, many assays commonly employ anaesthetics such as tricaine to achieve embryo immobility, which has been reported to cause severe cardiac side effects upon long term exposure [87]. Therefore, an assay, able to automatically image, detect and quantify teleost heart rates in a high-throughput manner without the need of special sedation or mounting, was required in order to investigate teleost cardiac development and function in their true native form.

1.8 High-throughput imaging and heart rate analysis in teleost fish embryos

Focusing on the heart rate as a valuable phenotypic readout, we have established the heart rate assay: an assay for scalable, automated high-throughput imaging and heart rate analysis in fish embryos (Figure 1.2A) [72]. The heart rate assay overcomes the abovementioned drawbacks of the previous workflows by employing teleost embryos, encapsulated by a transparent chorion and inherently immobilised, and simultaneously permits a live non-invasive view on the developing heart. Thus, the need for anaesthesia and immobilization was completely bypassed enabling investigation under native physiological conditions. Embryos, loaded on microtiter plates, are subjected to automated imaging in a temperature-controlled imaging machine with movable optics. The acquired image sequences of each embryo can then be processed by the *HeartBeat* software, which detects the beating heart and subsequently extracts the heart rate as a quantitative readout.

Using the heart rate assay, we generated a timeline following the actual dynamics of the heart rate throughout embryonic development of medaka (Figure 1.2B). From onset of heart beat at ~40 hpf (stage 24) [79], medaka heart rates exhibited a steep increase, reaching a plateau of ~160 bpm at ~72 hpf (stage 29), followed by circadian oscillations similar to those reported in healthy humans [71]. Complementing Iwamatsu's annotations, this timeline provided valuable information regarding cardiac functional development in

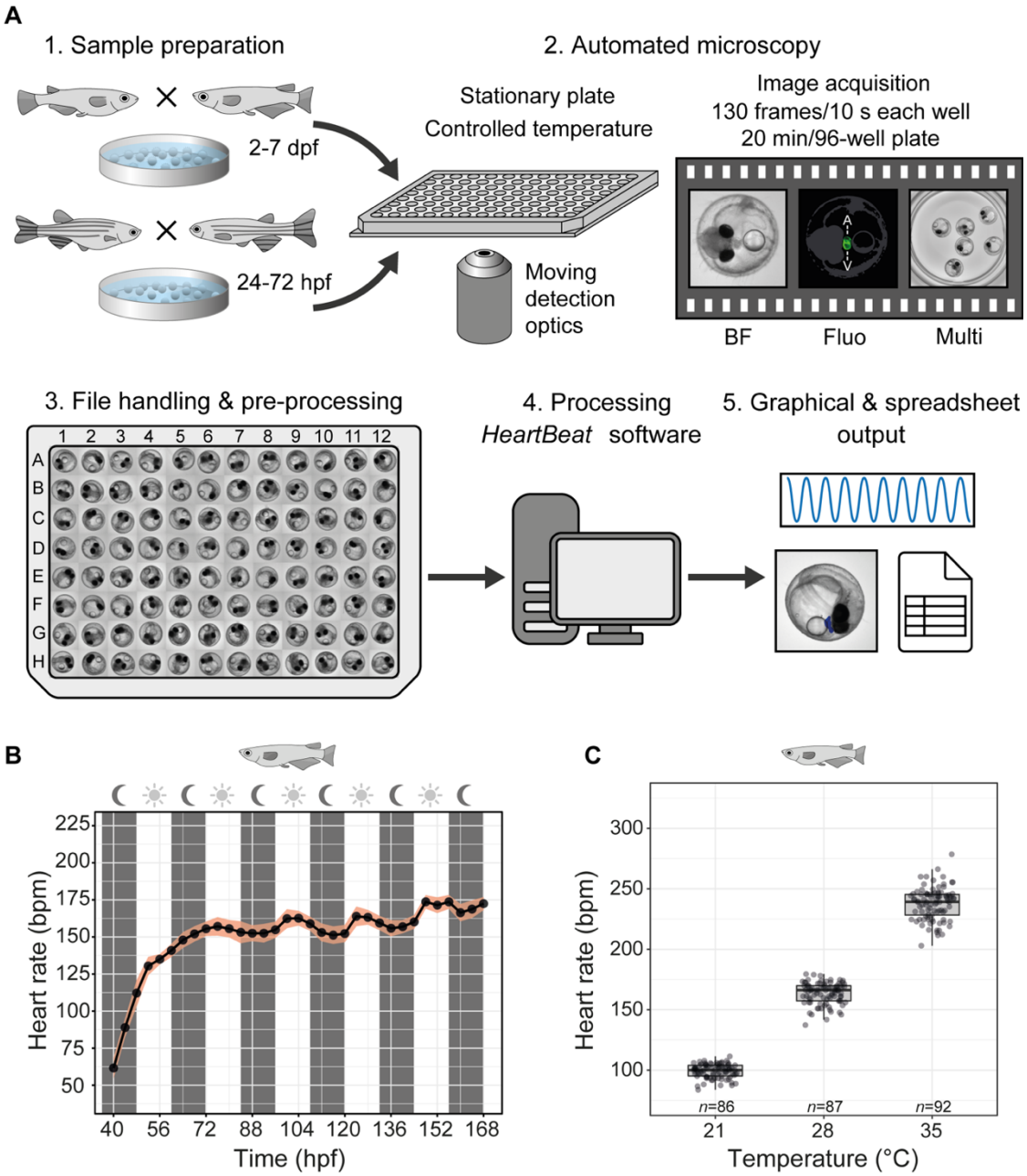


Figure 1.2 Automated high-throughput imaging and heart rate analysis in medaka embryos
A) Schematic overview of the heart rate assay workflow using teleost fish medaka and zebrafish embryos. Medaka and zebrafish embryo heart rates can be quantified in their respective time windows as indicated (hpf = hours post fertilization; dpf = days post fertilization), reflecting the time from onset of heart beat until hatching. Briefly, teleost embryos are loaded on microtiter plates (96-well plates depicted) without special mounting or sedation requirements. Image sequences (130 images) of each individual well are automatically recorded (at 13 frames per second) in a temperature-controlled environment inside an imaging machine with movable optics. The resulting videos (~10 seconds) from the image sequences are then subjected to analysis using the *HeartBeat* software, which detects and extracts the heart rate as a quantitative readout. **B)** Heart rate profiles of developing medaka at 28°C from onset of heartbeat until hatching, a heart rate plateau of ~160 beats per minute is reached at 72 hpf, from which 4 dpf was chosen as the default stage for heart rate experiments in medaka. **C)** Medaka (4 dpf) heart rate adapts in a linear fashion to external temperatures of 21, 28 and 35°C. Figure adapted and modified from [72].

medaka embryos, which was essential for the choice of the 4 dpf (approximately stage 32) time point as the most relevant stage for accurate heart rate phenotyping. Moreover, the heart rate assay was additionally applied on medaka embryos to assess the heart rate plasticity to environmental stimuli such as external temperature (Figure 1.2C). Medaka heart rates clearly responded to the shifts in temperatures, exhibiting a linear, steady and robust elevation in heart rates from ~100 to ~160 and ~230 bpm with increasing temperatures from 21 to 28 and 35°C, respectively. Taken together, the heart rate assay demonstrated its robust capability for reliable quantitative extraction of heart rates despite the random orientation of embryos within the chorion. It has also revealed relevant novel insights into medaka heart development and function. Medaka revealed particularly low inter-individual differences in their heart rates [72], supporting the isogenic-driven phenotypic robustness of medaka compared to zebrafish. Thus, indicating a foreseeable advantage of using medaka to detect subtle but clinically relevant heart rate phenotypes in a functional gene validation setting. Furthermore, the heart rate assay is highly versatile, with increased scalability achievable by concurrent imaging and analysis of multiple embryos, and is compatible with compound screening applications [72].

1.9 CRISPR/Cas9-mediated mutagenesis

Before the CRISPR era, targeted mutagenesis approaches such as zinc-finger nucleases (ZFNs) [88,89] and transcription activator-like effector nucleases (TALENs) [89,90] have been commonly employed to investigate and characterize gene functions. This was achieved through gene knockouts, or alternatively through knock-ins either by inserting a mutation or corrective DNA sequence of interest, or by tagging endogenous genes with a fluorescent protein. Both knockout and knock-in approaches rely on engineered site-specific DNA-binding protein domains attached to a non-sequence-specific cleavage domain (Fok1 nuclease), the latter introduces a double-strand break (DSB). DSBs in turn trigger DNA-repair either through nonhomologous end-joining (NHEJ) or homology-directed repair (HDR) [91]. Although HDR is the mechanism of choice for precise editing (knock-in of a specific mutation, correction or tag), NHEJ is the predominant repair mechanism employed by the cell. NHEJ is functional throughout the cell cycle and highly efficient at rapidly re-joining DSBs, so rapid that it causes random insertion or deletion (InDel) mutations. These InDel mutations usually result in deleterious frameshift mutations, exactly how it is exploited by the scientific community for rough gene knockouts [91].

The CRISPR/Cas system, short for clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas), was originally discovered as an adaptive RNA-mediated immune response in bacteria and archaea [92,93]. However, it has been reprogrammed for targeted genome editing [94] and since then revolutionised the biological and medical fields. The most commonly employed CRISPR/Cas system is the type II system, which relies on a single Cas protein, *Streptococcus pyogenes* Cas9

(SpCas9) for DNA targeting [94–96]. The CRISPR/Cas9 system is composed of the Cas9 endonuclease and a single guide RNA (sgRNA), which guides the Cas9 to its target site. The sgRNA in itself comprises of a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). The crRNA contains the variable stretch of DNA, called 'protospacer', that binds to its target sequence by base pairing. Upon hybridization with the tracrRNA, the crRNA-tracrRNA duplex then binds with the Cas9 protein to form the CRISPR/Cas9 complex [94]. The CRISPR/Cas9 complex is then directed towards its target site by the crRNA protospacer and binds to it as long as the recognition sequence is accompanied by a recognisable protospacer-adjacent motif (PAM) [94]. For SpCas9, an NGG PAM (N represents any nucleotide and G represents Guanine) requirement has been reported [94]. Later studies have shown the ability of the SpCas9 to also bind in presence of NAG PAM, albeit at lower efficiencies [97–99]. Cas9-mediated cleavage then follows, introducing a DSB roughly three bases from the PAM [94].

Since its emergence as a genome editing tool, the CRISPR/Cas9 system has undergone over the years various improvements, significantly boosting its efficiency [100]. Our group has impressively enhanced Cas9 *in vivo* editing efficiency by tagging the protein with an optimized artificial nuclear localization signal (oNLS). This oNLS tag facilitates the immediate nuclear localization of the new Cas9 variant, named as high efficiency tagged-Cas9 (heiCas9), and results in rapid gene editing [101]. Previously available ZFNs and TALENs mediated targeting approaches suffered from substantial protein engineering or complicated molecular cloning procedures. Each approach required continuous enzyme reengineering and separate synthesis for each desired target sequence, which is not only labour intensive but also expensive [89]. Alternatively, the only changing component in the CRISPR/Cas9 complex is the protospacer sequence of the crRNA, which can be easily redesigned to match any target site in a sequenced organism using available online target prediction tools [102,103]. Furthermore, the CRISPR/Cas9 complex can be potentially employed to concurrently edit multiple loci of interest by using several sgRNAs at the same time [104]. Consequently, the CRISPR/Cas9 system became an indispensable tool in biological research. Its versatility is clearly demonstrated through the vast diversity of applications utilizing this technology from gene knockouts [105–107], knock-ins for endogenous protein tagging [108,109], modelling of clinically relevant human mutations in animal models [110,111] and therapeutic correction of disease-causing gene variants [112]. Taken together, the CRISPR/Cas system represents a highly versatile gene editing approach that is relatively simple and inexpensive while offering unlimited potential in biological and medical applications. It comes of no surprise how it quickly became an integral part of the scientific community.

1.10 The need for high-throughput genotyping for rapid assessment of CRISPR-mediated targeting

Highly efficient genome editing tools such as CRISPR/Cas9, in addition to the continuous technological innovations in imaging, phenotyping and sequencing, have led to the emergence of high-throughput genetic screening applications [113–115]. Consequently, this has triggered the demand for rapid subsequent cost-effective genetic analyses for target validation. Almost all genotyping methods rely on prior amplification of the genomic locus of interest via polymerase chain reaction (PCR), using primers binding outside of the targeted locus. Following PCR amplification, a variety of techniques at all levels of costs, effort, time and resolution are available for downstream genetic analyses of the amplified locus (amplicon) [116].

Most common and convenient are the enzyme mismatch assays, employing restriction endonucleases such as CEL I (Surveyor Assay) [117–119], or T7 Endonuclease I (T7EI assay) [120] that recognise and cut only heteroduplex, not homoduplex, DNA. Their convenience stems from the fast delivery of targeting confirmation without the need for specific equipment. However, they can only reliably detect InDel mutations of at least 2-5 base pairs (bp) and do not provide detailed information regarding the nature or frequency of the InDel mutations. Other cost-effective InDel detection approaches with higher resolution (1 bp) such as high-resolution melt analysis (HRMA) [121,122], indel detection by amplicon analysis (IDAA) [123], and digital droplet PCR (ddPCR) [124] are also commonly employed, however they require specific instrumentation. NGS platforms offer the highest resolution and scalability, as well as the most information regarding frequency and nature of the InDels [125], but are also the most expensive and time/labour intensive. Alternative to NGS, Sanger sequencing of the PCR amplicon followed by trace decomposition using online tools such as Tracking of Indels by Decomposition (TIDE) [126] or Inference of CRISPR Edits (ICE) [127] offer an acceptable compromise of cost, time/labour and InDel information compared to NGS.

Nevertheless, despite the continuous advancements in efficiencies and throughputs of current genotyping workflows, the initial genomic DNA (gDNA) extraction step still remains a substantial bottleneck. Current gDNA extraction workflows although greatly improved over the years, were generally laborious (samples individually handled) and time consuming with limited applicability for high-throughput approaches [128]. Especially for teleost embryos, which require disintegration of the chorion for efficient lysis, usually carried out by physical deterioration (i.e., grinding with pestle or using magnetic beads). Efficient extraction protocols mainly rely on solid-phase extraction kits [129] or are chemically driven, as with phenol-chloroform extraction [130] providing the highest yield. Alkaline lysis represents another straightforward chemical disintegration approach [131]. However, it requires incubation steps followed by neutralization and does not reliably

work with medaka embryos. Current standard gDNA extraction protocol for teleost embryos is the Blin-Stafford lysis approach [132] followed by isopropanol precipitation. This protocol can take at least two hours and requires substantial manual handling of individual samples. Alternatively, the reported binding affinity of nucleic acids on membranes of different materials has therefore allowed the circumvention of incubation steps, by the amplification of nucleic acids directly off the membrane [133–136]. Taking advantage of nucleic acids binding affinity to cellulose paper, a research group has manufactured cellulose paper-based "dip-sticks" [137]. These dip-sticks not only were proven useful for sample collections in remote areas but have also drastically cut nucleic acid extraction time down to 30 seconds. However, the dip-sticks were incompatible with handling large sample numbers. Therefore, developing a rapid and scalable genotyping workflow was a valuable asset to complement high-throughput genetic screenings.

Aims & Approaches

The aim of this thesis was to identify novel genes with potential diagnostic and predictive power for cardiovascular diseases (CVDs) in humans. Most CVDs have a strong genetic component, making their underlying genetic risk factors a strong driver of rising rates of death worldwide. Therefore, genetic markers enabling early risk assessment and diagnosis of CVDs are highly demanded. Human genome-wide association studies (GWAS) have been extensively performed in search of genetic determinants of CVDs, however, these associated genes and polymorphisms still require functional validation in animal models. Although GWAS constitute a rich resource for potential candidate discovery, the pressure for rapid mechanistic insights and the absence of high-throughput validation tools to achieve them has shifted the focus onto genes with known pre-existing cardiac function rather than novel genes with no previous link to the heart. Consequently, uncharacterized genes that may collectively hold strong predictive power remain largely unaddressed. I therefore aimed to functionally validate these uncharacterized candidate genes, using a reverse genetic screening approach. Due to the conservation of its cardiac features, as well as its high isogeneity I decided to employ medaka as my model of choice for the genetic screen.

First, I wanted to evaluate the feasibility of performing functional gene validations using CRISPR in an F0 setting. Therefore, I examined the editing efficiency of the enhanced heiCas9 variant by targeting a series of genes yielding known and visible phenotypes.

Second, I wanted to investigate whether the low interindividual variability in highly isogenic medaka would allow me to observe statistically significant heart phenotypes in genetically mosaic mutant embryos. The heart rate is widely recognized as a vital predictor for CVDs, thus based on the functional conservation of basic cardiac function between teleost fish and humans, I chose the heart rate as the main phenotypic readout. Here, I designed a pipeline that combined the previously established heart rate assay with CRISPR-based genome editing for the rapid validation of candidate gene function in injected generation medaka (F0). As a proof of concept, I targeted a known cardiac gene to determine the relevance of the phenotypes I obtain in F0.

Third, I wanted to determine the baseline probability of an occurring heart rate phenotype. Thus, I applied my validation workflow onto a small set of randomly selected genes.

Fourth, I wanted to identify novel genes with potential diagnostic power for CVDs using my established pipeline. Therefore, I performed a CRISPR-screen to validate CVD-associated genes from human GWAS, with an emphasis on validating uncharacterized genes or genes with no prior evidence on heart function.

2 Results

2.1 Swift Large-scale Examination of Directed Genome Editing

The majority of experimental applications including sequencing, phenotyping and compound screening, have undergone a tremendous shift towards higher throughput, enabled by a multitude of technological advancements. We are currently witnessing the boom of CRISPR gene editing and high-throughput genetic screens, which require subsequent genetic analysis. Despite continuous improvements in current genotyping workflows, gDNA extraction still represents a significant bottleneck with respect to sample handling and processing times. Blin-Stafford lysis [132] followed by isopropanol precipitation is the current standard gDNA extraction protocol employed for medaka and zebrafish. However, it is time consuming with several incubation times and requires extensive manual handling of individual embryos before gDNA is extracted for PCR amplification and downstream processing (e.g. sequencing or mismatch cleavage assays), taking almost a full day of work. With the aim to validate a large number of candidate genes *in vivo* in medaka using CRISPR/Cas9, it was necessary to establish a genotyping pipeline with scalable, quick and efficient gDNA extraction for the rapid confirmation of CRISPR-mediated targeting. Zou et al have reported the use of cellulose paper-based "dip-sticks" for rapid nucleic acid extraction, significantly cutting sample processing time down to 30 seconds [137]. However, the dip-sticks still require manufacturing, and their inherent design makes them unsuitable for high-throughput applications. To exploit the characteristics of cellulose paper to bind nucleic acids, I have developed an alternative gDNA extraction workflow, compatible with high-throughput applications, which has significantly sped up genotyping and the evaluation of CRISPR-mediated gene editing.

2.1.1 Rapid verification of CRISPR-mediated gene targeting using filter-in-tips

To cut down processing and incubation times in genotyping protocols currently used, I took advantage of nucleic acid binding properties of cellulose to establish the filter-in-tip

as a gDNA extraction tool. Consequently, cellulose-based Whatman paper discs were cut and inserted into standard pipette tips (Figure 2.1A). To test the filter-in-tip approach, adult wild-type medaka fish were genotyped to determine their genetic sex by PCR. Fins of three male and three female medaka adults were clipped and ground in individual tubes containing SDS-based Fin-Clip lysis buffer (FB) using a pestle (Figure 2.1B, Step 1). Ground lysate was pipetted up into the filter-in-tips to soak the filter in the lysate, allowing the uptake of gDNA (Figure 2.1B, Step 2). A quick washing step by brief pipetting of water in and out of the filter-in-tip is essential to remove SDS remnants, which can impair PCR amplification by damaging the polymerase enzyme (Figure 2.1B, Step 3). Elution then follows by pipetting the PCR mix into the filter-in-tip, thus releasing the bound gDNA, and finally pipetted back into the PCR tube for direct amplification (Figure 2.1B, Step 4). All male medaka fish were successfully genotyped by amplification of the male-specific *DnmtY* locus [138], while the beta actin (*Actb*) locus was successfully amplified in all medaka specimens (Figure 2.2A), reflecting the applicability of the filter-in-tip approach. Processing time using the filter-in-tip approach was reduced down to 30 seconds, thus providing a rapid alternative to the standard gDNA extraction approach, which takes at least 2 hours.

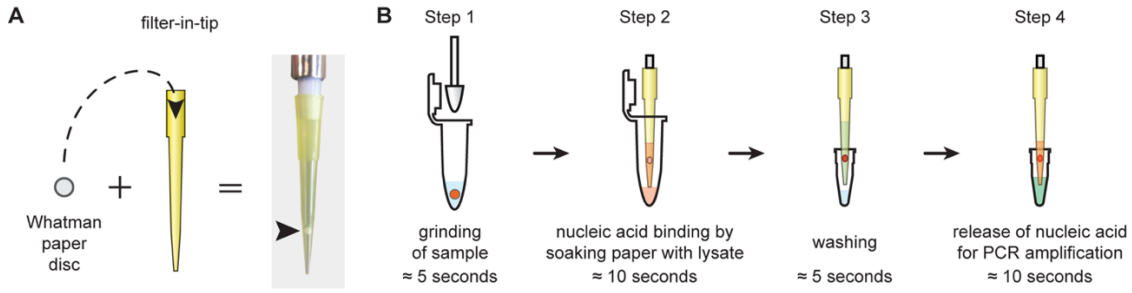


Figure 2.1 Design and workflow of genotyping using filter-in-tips
A) Filter-in-tip design: Whatman paper disc punched and inserted into standard yellow pipette tips (P200) and pushed towards the sample-proximal end. **B)** An overview of the filter-in-tip-based purification method. Step 1 lysis: sample (red) is ground in FB using a pestle. Step 2 binding: lysate is aspired using a filter-in-tip, soaked for ~10 seconds, allowing nucleic acids (red) to bind to the cellulose filter, before releasing lysate back to sample tube for storage. Step 3 washing: filter disc containing nucleic acids (red) is washed from lysis buffer remnants by pipetting nuclease free water in and out. Step 4 elution: release bound nucleic acid by pipetting up pre-mixed PCR mixture, soak for ~10 seconds, and pipet back into PCR tube for amplification. Figure adapted and modified from [139].

The comparability of the filter-in-tips to the standard gDNA extraction approach was assessed by genotyping genetically mosaic medaka embryos at the *oculocutaneous albinism 2* locus (*oca2* [101,107]) as a result of CRISPR/Cas9-mediated mutagenesis, commonly referred to as crispants [140]). A single *oca2* crispant embryo was first subjected to genotyping using the filter-in-tip-approach and the remainder of the lysate was then processed following the standard gDNA extraction protocol. Both approaches successfully yielded clear PCR bands (Figure 2.2B). Crispants, being genetically mosaic at the targeted locus, usually have a unique signature of mutant alleles resulting from various NHEJ-driven InDel mutations. Processing the amplified PCR bands for Sanger sequencing and downstream TIDE analysis revealed matching allele distribution profiles between both approaches (Figure 2.2B). Thus, confirming the high quality of the gDNA

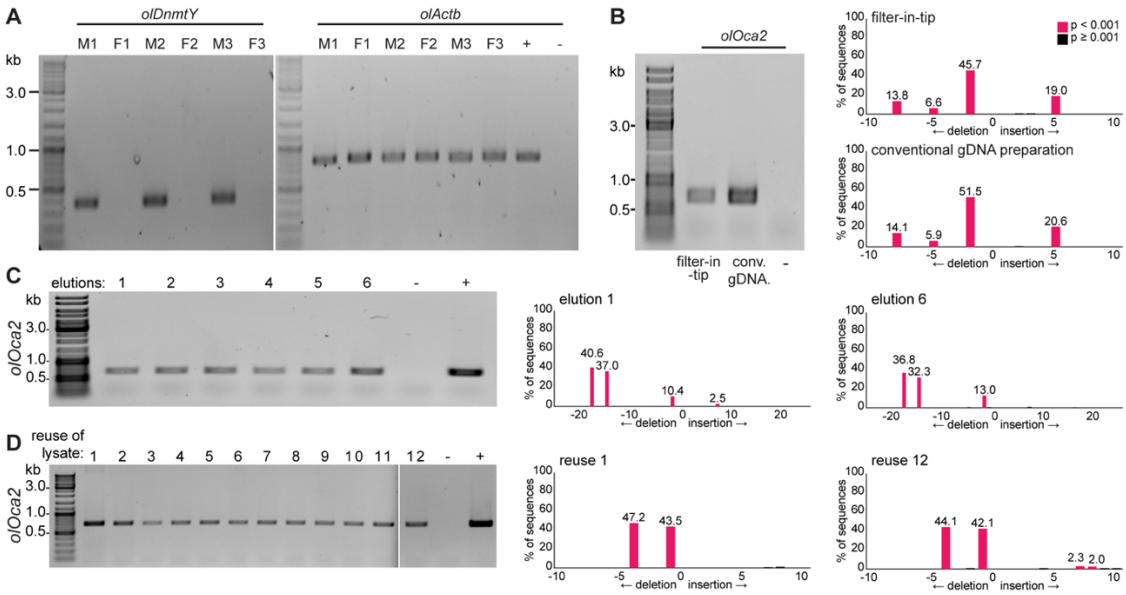


Figure 2.2 Rapid, sensitive and accurate genotyping using filter-in-tips
A) Genetic male sex determination by PCR amplification of *olDnmtY* locus in 3 male (M1-3) and 3 female (F1-3) medaka fin clips. **B)** Amplified gDNA extracted using filter-in-tips (left) compared to conventional Blin-Stafford gDNA preparation (right) of a single medaka *oca2* crispant, with matching allele distribution profiles of both amplicons (TIDE analysis). **C)** DNA retention capacity of a filter-in-tip assessed by soaking it only once (10 seconds) in a medaka *oca2* crispant lysate, washed once and eluted in 6 separate PCR reactions (10 seconds each), with matching allele distribution profiles between the first and sixth eluate (TIDE analysis). **D)** Reproducibility of filter-in-tip approach assessed by repeated extraction of gDNA from a single medaka *oca2* crispant lysate using 12 filter-in-tips, with matching allele distribution profiles between the first and twelfth amplicons (TIDE analysis). Conventional Blin-Stafford gDNA extraction (wt gDNA) was used as a positive control (+), water was used as a negative control (-). Figure adapted and modified from [139].

extracted using filter-in-tips compatible with downstream Sanger sequencing, useful for the evaluation of CRISPR-mediated mutations.

Next, the DNA retention capacity of the filter-in-tip was examined by loading a filter-in-tip once with an *oca2* crispant embryo lysate, washed once, and sequentially eluted in six separate PCR reactions. Amplification of the *oca2* locus was successful in all six reactions, with matching allele distribution profiles of the first and sixth eluate (Figure 2.2C). The results have shown a high binding capacity of the filter-in-tip, but also a limited release quota of bound nucleic acids. The reproducibility of the filter-in-tips was then addressed by sequential gDNA extraction from the same *oca2* crispant embryo lysate using 12 filter-in-tips. The successful amplification across all 12 attempts, with matching allele distribution profiles between the first and twelfth attempt clearly demonstrate the robustness and reproducibility of the filter-in-tip approach (Figure 2.2D).

The broad applicability of the filter-in-tips was further evaluated by extracting gDNA from varying sources of tissues derived from model organisms across kingdoms. To that extent, filter-in-tips were used to extract gDNA from *Arabidopsis thaliana* (single seedling; *At*), *Drosophila melanogaster* (individual adult fly; *Dm*), *Chironomous riparius* (individual adult mosquito; *Cr*), as well as *Mus musculus* ear punches (individual mouse

ear punch; *Mm* EP) and *Mus musculus* embryonic stem cells (2.5 x 10^5 ESC suspension; *mM* ESCs) (Figure 2.3). In all cases, successful amplification of desired loci was observed, reflecting the broad applicability of this fast, simple and reliable approach.

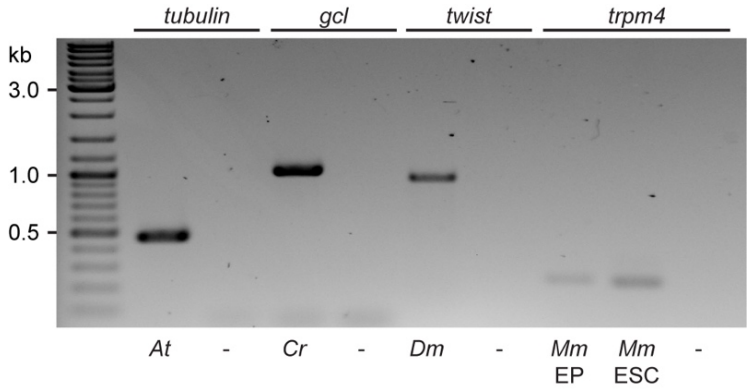


Figure 2.3 Filter-in-tip approach compatible with various organisms
Successful PCR amplification of candidate genes (*tubulin*; *germ cell-less*, *gcl*; *twist*; and *trpm4*) from lysates of *Arabidopsis thaliana* (*At*), *Chironomous riparius* (*Cr*), *Drosophila melanogaster* (*Dm*), mouse (*Mus musculus*, *Mm*) ear punches (EP) and mouse embryonic stem cells (ESCs) respectively using the filter-in-tips. Water was used as a negative control (-). Figure adapted from [139].

2.1.2 Standard pipette tips inherently bind sufficient nucleic acids for genotyping purposes

The filter-in-tips were mainly driven by the previous findings of the ability of nucleic acids to bind to cellulose paper [137]. However, the inherent ability of the traditional pipette tips to bind to and retain nucleic acids has not been previously tested due to the advertised hydrophobic nature of the tips. Thus, a direct comparison of nucleic acid binding was performed using pipette tips with inserted cellulose disc filter (filter-in-tips) and without cellulose disc (standard-plain), followed by PCR amplification. Two different samples of single embryo lysates and a lysate from a pool of five embryos were used (stored lysate, Figure 2.4A; fresh lysate, Figure 2.4B). Overall, no significant differences between traditional tips and our filter-in-tips were observed, except in few cases where adding the cellulose filter resulted in a slightly improved yield. This experiment has revealed that the pipette tip alone, although hydrophobic in nature, has enough nucleic acid binding properties sufficient for purification and amplification purposes.

To address whether the nucleic acid binding properties exhibited by the pipette tip is brand-specific, pipette tips from various companies were tested, and their respective nucleic acid binding efficiency with and without the cellulose filter disc was compared (Figure 2.4C). Overall, standard tips from various companies showed similar nucleic acid binding properties, with no significant advantage of adding the cellulose filter disc to the tip. An exception, however, is observed with low retention tips, where a significant increase in the yield of nucleic acid extraction was observed by adding the cellulose filter disc to the tip. Thus, demonstrating that the retention of nucleic acids in the tip mainly stems from the innate tip coating and structure resulting from the manufacturing process.

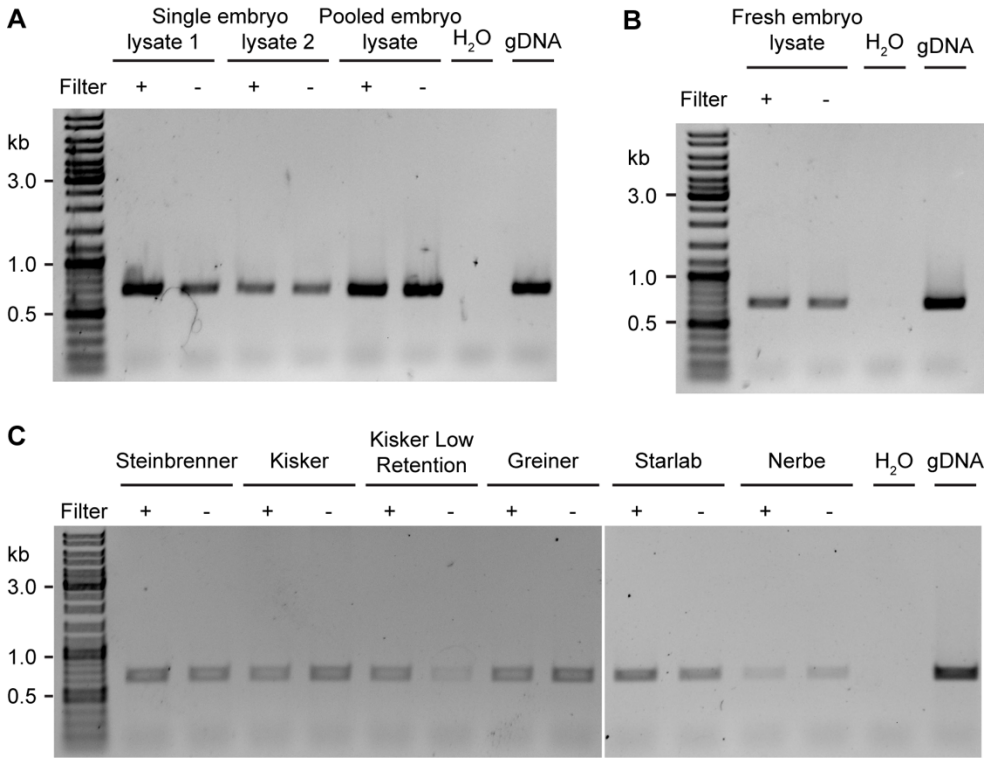
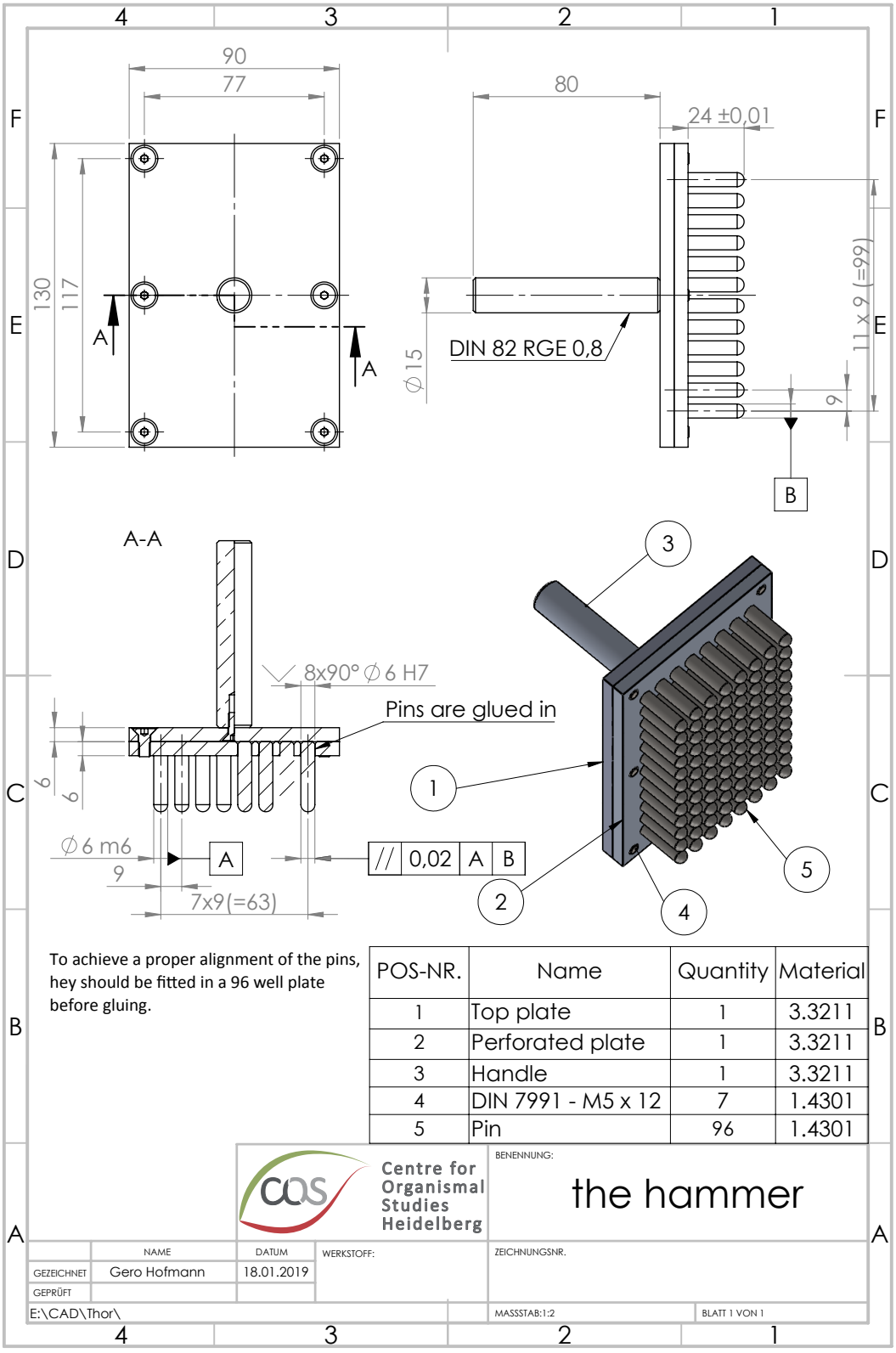


Figure 2.4 Standard pipette tips retain nucleic acids sufficient for genotyping
A-B) Direct comparison of gDNA extraction and amplification when using tips with inserted cellulose disc filter (filter-in-tips; +) and without cellulose disc filter (standard tip; -) with **(A)** two non-fresh independent single embryo lysates and a non-fresh pooled embryo lysate (5 embryos) as well as **(B)** with a fresh single embryo lysate. **C)** Pipette tips from various companies were compared with respect to their nucleic acid binding efficiency with (+) and without (-) the cellulose filter disc. Conventional Blin-Stafford gDNA extraction (gDNA) was used as a positive control (gDNA), water was used as a negative control (H_2O).

2.1.3 Eliminating the bottleneck: Genotyping in high throughput

The main limitation of the cellulose dip-sticks, similar to most gDNA extraction approaches, was their incompatibility with high-throughput applications [137]. Although processing time per sample was drastically cut down to 30 seconds, individual processing of multiple samples in a multi-well plate format would still require exhaustive manual handling. Moreover, efficient parallel mechanical sample lysis so far achieved only by use of magnetic beads [141] poses an additional bottleneck. Sample transfer, incubation on a shaking platform and subsequent removal from the plates are all time-consuming. Since mechanical lysis, which tears the chorion, is essential for gDNA extraction from medaka embryos, a mortar with stainless-steel pins specifically designed for 96-well plates was devised to enable simultaneous lysis (the Hammer; Figure 2.5). The hammer enabled simultaneous mechanical lysis of all specimens in a 96-well plate in a single step. Combining the quick mechanical lysis attained using the hammer and the subsequent parallel processing using the tips mounted on a multi-channel pipette unleashes the full



POS-NR.	Name	Quantity	Material
1	Top plate	1	3.3211
2	Perforated plate	1	3.3211
3	Handle	1	3.3211
4	DIN 7991 - M5 x 12	7	1.4301
5	Pin	96	1.4301

Figure 2.5 Custom-made 96 U-well plate mortar (the Hammer) design and schematics
Construction plan and schematics of the 96 U-well plate mortar. 96 stainless steel pins (6 mm diameter) were fitted on a 96 U-well plate before gluing to achieve proper alignment of pins. Pins were pushed and glued in between two perforated metal plates, the mortar is equipped with a textured aluminum handle for a firm grip. Materials and dimensions are indicated. Figure adapted from [139].

potential of the swift large-scale examination of directed genome editing approach (the SLEDGE-hammer approach; Figure 2.6A).

To test the SLEDGE-hammer approach for direct phenotype-genotype correlations, two phenotyping paradigms were used: i) the loss of ocular pigmentation through the CRISPR/Cas9-mediated knockout of *oca2* in medaka (*olOca2*) and zebrafish (*drOca2*) embryos, and ii) appearance of green fluorescent protein (GFP)-positive retinal cells by homology-directed repair (HDR) knock-in of GFP into the *retinal homeobox protein 2* locus (*gfp-rx2*) in medaka embryos (Figure 2.6B). For both paradigms, an initial double-stranded break was achieved through the extraordinary efficiency of heiCas9, triggering NHEJ-mediated knockout at the *oca2* locus [101,107], and repaired by HDR pathways to knock-in GFP into the *rx2* locus using a biotinylated dsDNA donor [102,142]. For each condition, injected embryos (4 dpf for medaka *oca2*; 2 dpf for zebrafish *oca2* and medaka *gfp-rx2*) were randomly selected and transferred into individual wells of a 96-well plate, with one empty well as negative control (empty well containing embryo rearing medium), and two un-injected control embryos (Figure 2.7A). As expected, medaka and zebrafish *oca2* crispants showed various degrees of loss of eye pigmentation, while tagged *rx2* was visible via GFP expression in medaka retinae (Figure 2.6C). Following phenotyping and scoring of embryos, direct phenotype-genotype correlations were determined using our SLEDGE-hammer approach to individually genotype the embryos (Figure 2.6A). Processing the entire plate took approximately 6 mins, demonstrating the outstanding gDNA sample preparation performance of the SLEDGE-hammer approach.

For the *oca2* crispants, locus amplification by PCR was successful in all specimens, indicating successful gDNA transfer from lysate to PCR mix (Figure 2.6D and Figure 2.7B). For the rapid confirmation of successful targeting of the *oca2* locus (i.e. CRISPR/Cas9-mediated InDel mutations), the T7EI mismatch cleavage assay was employed [120]. T7EI downstream analysis of the amplified PCR bands, revealed a positive qualitative correlation between the phenotypes observed and genotypes. All embryos in which *oca2* was targeted produced an additional T7EI cleavage band, while for the un-injected (wild-type) embryos, no T7EI digestion was observed (Figure 2.6D and Figure 2.7C).

For *gfp-rx2* knock-in attempts, genotyping was performed via PCR amplification of the full *rx2* locus (Lf/Lr). To test for the single-copy integration of the *gfp* into the locus, *gfp-rx2* specific PCR amplification was performed (Lf/*gfp* r). For all samples the full locus was successfully amplified with a correlation between GFP-expressing embryos and their respective genotypes due to a band shift resulting from the single-copy fusion of *gfp* sequence to the *rx2* locus (Figure 2.6D and Figure 2.7D). Genotypical validation of GFP-expressing embryos was obtained after specific amplification of the *gfp-rx2* insert (Figure 2.7E). Amplification of *gfp-rx2* occurred in some embryos despite their lack of GFP expression (yellow asterisks; Figure 2.7E), a likely consequence from exclusive integration of the transgene in non-retinal cells or from NHEJ-mediated mutations or concatenation events [142]. Taken together, the SLEDGE-hammer protocol enables rapid

verification of CRISPR-mediated targeting in a high-throughput manner with the possibility for individual phenotype-genotype correlations.

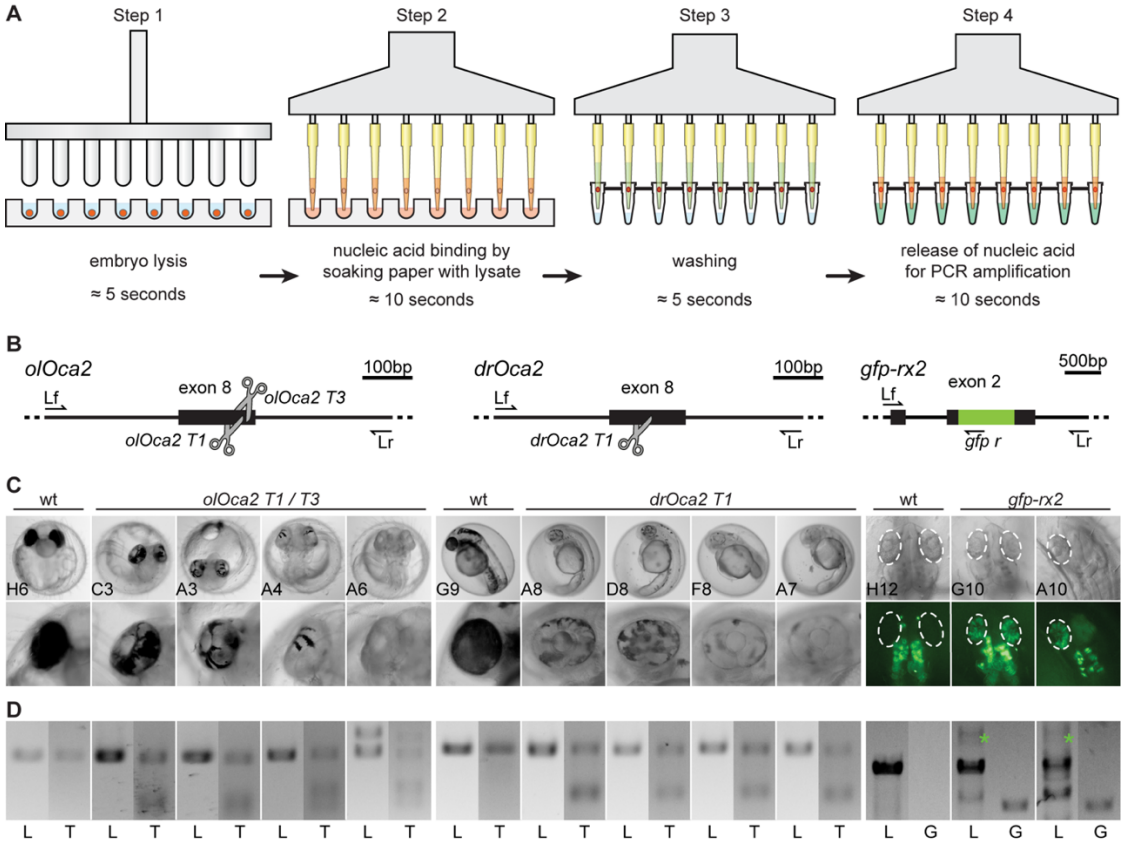


Figure 2.6 Genotyping in high throughput, the SLEDGE-hammer protocol
A) An overview of the high-throughput SLEDGE-hammer genotyping protocol. Step 1 lysis: sample consisting of 96 individual embryos in a 96-well plate (red) are simultaneously ground in FB using our custom-made 96-pinned mortar. Step 2 binding: lysates are aspired in parallel using a multi-channel pipette for ~10 seconds before releasing lysates back to the plate for storage. Step 3 washing: tips are washed from lysis buffer remnants by pipetting nuclease free water in and out. Step 4 elution: release bound nucleic acid by pipetting up pre-mixed PCR mixtures, soak for ~10 seconds, and pipet back into PCR tubes for amplification. **B)** Illustrative depiction of CRISPR/Cas9-mediated NHEJ-based knock-out in medaka and zebrafish *oca2* loci (*olOca2* and *drOca2* respectively), as well as HDR-mediated knock-in of *gfp* sequence in frame with medaka *rx2* [108]. sgRNA target sites (two sgRNAs for medaka *oca2*, one for zebrafish *oca2*, and one for medaka *rx2*) portrayed as scissors. *rx2* locus shown after single-copy HDR-mediated integration of *gfp* in frame. Primers used for PCR amplification of the respective genes indicated; locus forward (Lf), locus reverse (Lr), and *gfp* reverse (*gfp r*). **C)** Select representative embryos from 96-well plate (for plate coordinates of embryos indicated refer to Figure 2.7). Wild-type (wt) medaka and zebrafish display fully pigmented eyes, while *oca2* crispants show varying degrees of pigment loss in body and eyes (zoom-in). Successful HDR-mediated integration of *gfp* into *rx2* open reading frame results in GFP-expressing cells in retinae of injected embryos (dashed ellipses). Note: non-specific autofluorescence of body pigment. **D)** Genotyping of embryos in **(C)**. Locus PCR of respective genes (L; primers Lf/Lr) followed by T7EI mismatch assay (T) for qualitative confirmation of InDel formation in *oca2* loci of *oca2* crispants but not wild-type (wt) siblings. In *gfp-rx2* tagging, other than the non-*gfp*-integrated locus band amplified (L; primers Lf/Lr), single *gfp* integration can also be observed in embryos expressing GFP in the retinae (green asterisk), further confirmed by locus-gfp band amplification (G; primers Lf/*gfp* r). Figure adapted from [139].

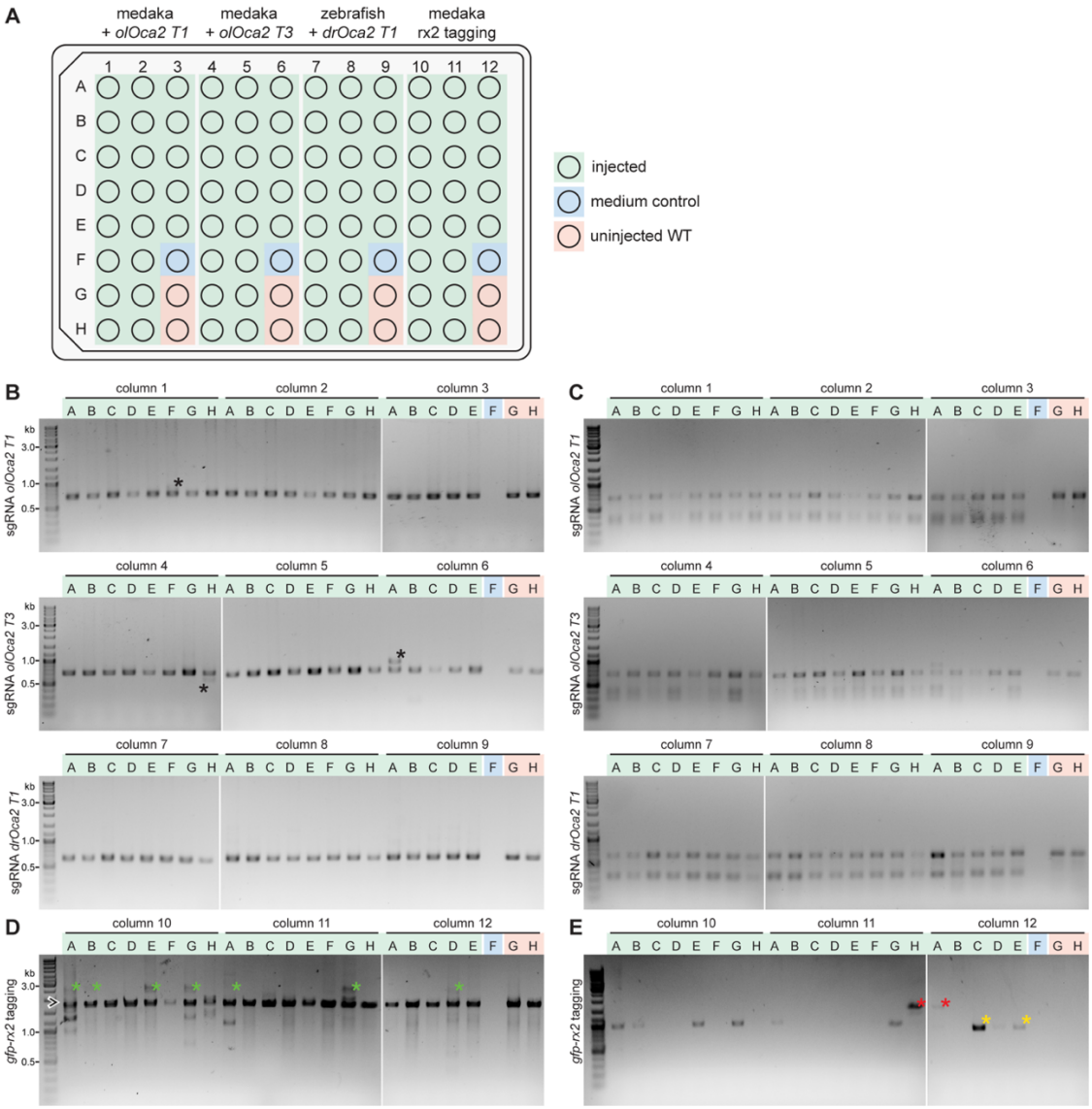


Figure 2.7 96-well plate format and high-throughput SLEDGE-hammer analysis
A) High-throughput genotyping of a 96-well microtiter plate. CRISPR/Cas9-mediated knock-out (green wells) of *oca2* using two individual sgRNAs targeting the locus in medaka: *olOca2 T1* (columns 1-3), *olOca2 T3* (columns 4-6) as well as one sgRNA targeting the locus in zebrafish: *drOca2 T1* (columns 7-9). In addition to HDR-mediated knock-in of *gfp* in frame with the *rx2* locus in medaka (columns 10-12). For each condition, a medium control (blue well) and two uninjected wild-type siblings (red wells) were included as control. **B)** Successful rapid extraction and PCR amplification of *oca2* loci across all *oca2* experimental group (columns 1-9) using the SLEDGE-hammer protocol. Note: large InDel alleles can give rise to additional bands (black asterisks). **C)** T7EI assay run on samples from **(B)** for qualitative validation of InDel formation in *oca2* crispants (digested bands) but not in uninjected wild-type embryos. **D)** *rx2* locus PCR amplification of *gfp-rx2* experimental group (columns 10-12). Note the possible discrimination between non-*gfp*-integrated *rx2* locus band (black arrowhead, 1719 bp) and single copy *gfp* integration band (green asterisks, 2547 bp), additional bands may arise from NHEJ-events. **E)** Further confirmation of *gfp* integration by amplification of *gfp-rx2* specific bands, matching embryos expressing GFP in the retinae. Single copy HDR-mediated *gfp* knock-in events in **(D)** could be verified by band size (953 bp). In some embryos NHEJ-mediated integrations (red asterisk, ~1400 bp) or concatenation events (yellow asterisks) may have occurred. Figure adapted from [139].

2.2 Functional validation of human CVD-associated genes in medaka

The key for cardiovascular diseases (CVDs) prevention is early diagnosis and treatment. Resting heart rate has been widely acclaimed as a modifiable, vital risk factor for both prediction and prevention of CVDs [10,12,23,143]. Since genetics plays a major role in CVDs onset and progression, attempts at understanding the genetic determinants of CVDs are constantly ongoing, making use of comprehensive human GWAS. These studies typically yield thousands of CVD-associated polymorphisms across hundreds of different genes, many of which are not yet characterised or connected to heart functions. Despite the wealth of the GWAS output, the lack of a rapid experimental validation pipeline, typically results in a negative loop of re-discovery. Only those genes with pre-existing indication of cardiac function are usually addressed further, while the remainder uncharacterised genes are neglected. In this section, I addressed this blind spot of discovery by coupling the two introduced high-throughput pipelines for heart-rate analysis [72] and genotyping using the SLEDGE-hammer [139] with highly efficient CRISPR/Cas9 targeted genome editing [101]. In consequence, this allowed me to develop an F0 CRISPR/Cas9 screen of heart-phenotype associated genes from human GWAS in highly isogenic medaka.

2.2.1 heiCRISPR/Cas9 system delivers highly efficient genomic editing already in injected generation

Before functional gene validations could be considered, an initial *in vivo* assessment of the efficiency of the hei-tagged CRISPR/Cas9 system [101] was required. Therefore, for quick and easy phenotyping, *gfp* was targeted using heiCas9 in a transgenic medaka double reporter line, expressing cytosolic GFP and nuclear red mCherry fluorescent proteins under the heart-specific *cardiac myosin light chain 2* (*cmlc2*) promoter. Injections of *heiCas9* mRNA with an sgRNA targeting *gfp* (*gfp_T1*) were performed in medaka embryos at the 1- and 4-cell stage. Targeting of *gfp* at the 1-cell stage resulted in a complete loss of GFP expression in the heart (n= 8/8) (Figure 2.8A and B), and only when targeting *gfp* at the 4-cell stage was a mosaic GFP pattern observed (Figure 2.8A and C), highlighting the extraordinary delivery and cleavage efficiency of heiCas9. To evaluate the extent of InDel formations in the 1- and 4-cell injected *gfp* crispant embryos, genotyping was performed by PCR amplification of the *gfp* locus, followed by downstream TIDE analysis to reveal the allele distribution profiles of each of the *gfp* crispants. TIDE analysis confirmed the complete loss of the wild-type allele in the 1-cell injected *gfp* crispants, but not in the 4-cell injected *gfp* crispants, complementary with the phenotypes observed (Figure 2.8B and C).

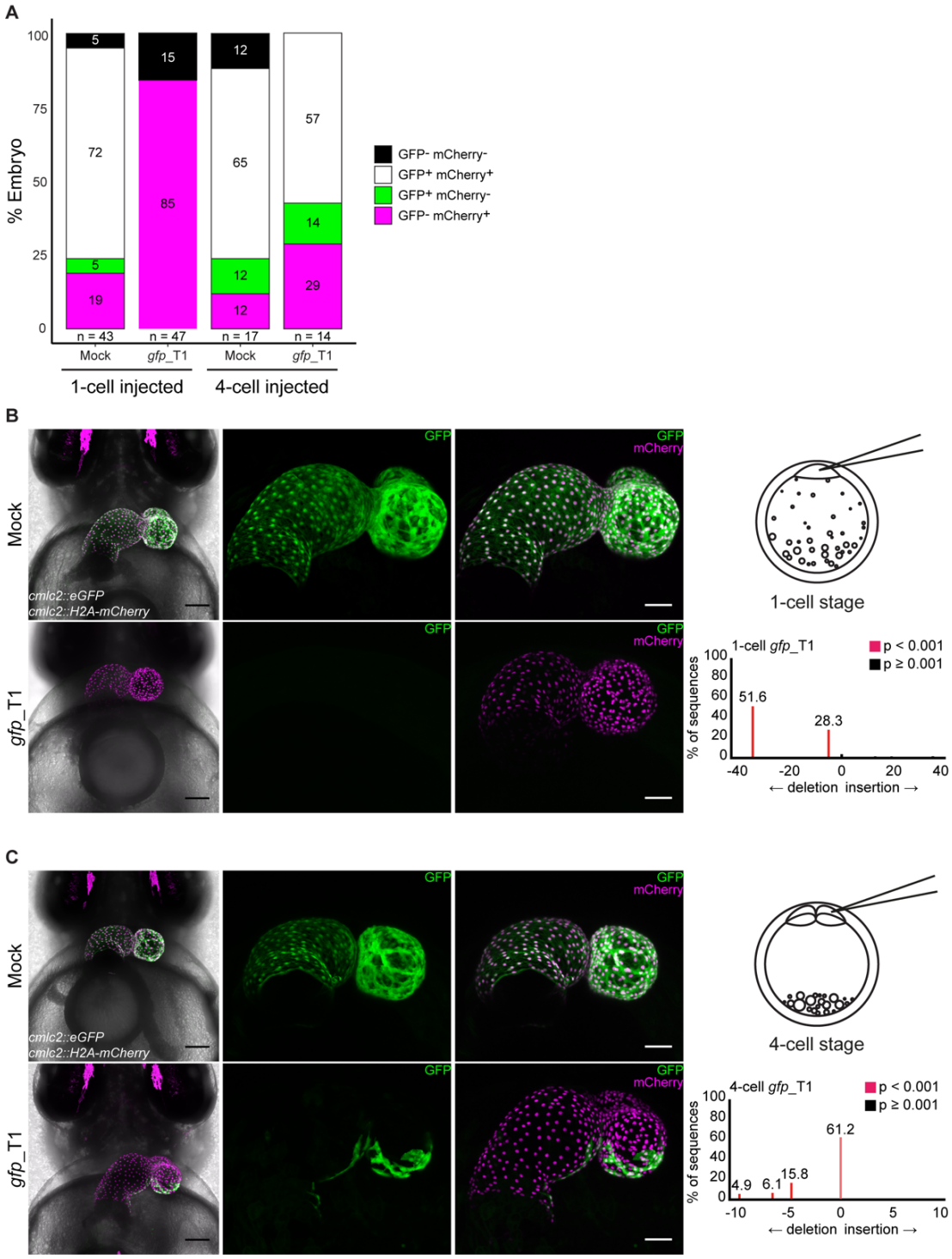


Figure 2.8 heiCas9 gene targeting reveals high editing efficiency in injected (F0) generation
A) Expression of GFP and mCherry reporters under the cardiac-specific *cmlc2* promoter in mock injected and *gfp_T1* crispants (4 dpf). Embryos were injected either at the 1-cell or 4-cell stage. Note: when injected at the 1-cell stage, there are no GFP-expressing embryos. Number of embryos successfully scored for each group is denoted as (n). **B-C)** Confocal images of representative hearts of dual reporter *cmlc2::eGFP cmlc2::H2A-mCherry* mock injected and *gfp_T1* crispant embryos (7 dpf), injected at **(B)** 1-cell stage and **(C)** 4-cell stage. Note: no GFP expression in 1-cell stage-injected embryos (**B**; n = 8/8), and mosaic expression upon injection at 4-cell stage (**C**; n = 4/4). Mutant allele distribution profiles (TIDE analysis) of genotyped *gfp_T1* crispants display the genetic mosaicism resulting from heiCas9-based targeting. Note: complete loss of wild-type allele in 1-cell stage-injected embryos (**B**; n = 8/8). Scale bars: 100 µm (first panel on left) and 50 µm (blow-up images). Figure adapted from [144].

Having determined the high efficacy of the heiCas9 system in medaka embryos in targeting an artificially inserted gene (i.e. *gfp*), it was next sought to assess the efficacy when targeting an endogenous gene. The *oca2* locus was chosen due to its observable loss of eye pigmentation phenotype [145], which also serves as an indicator of the bi-allelic knockout efficacy of the heiCas9 employed [107,139]. Two guide RNAs (NGG PAM) were previously designed to target the same *oca2* locus (exon 8) within 18 base pairs apart (Figure 2.9A) and were co-injected to ensure the efficient bi-allelic knockout. However, upon individual testing of the sgRNAs for the SLEDGE-Hammer genotyping experiments (Figure 2.6 and Figure 2.7), the difference in bi-allelic knockout efficiencies using each sgRNA separately became apparent (Figure 2.9B). *oca2_T3* alone yielded more of the almost pigment-less crispant embryos compared to *oca2_T1* alone, which rather results in mosaic pigment patterns. While using both sgRNAs together also yielded more of near pigment-less embryos than when using *oca2_T1* alone (data not shown), still *oca2_T3* alone produced the strongest pigmentation knockout phenotypes, indicating possible competitive inhibition when concomitantly using both sgRNAs together. Accordingly, *oca2_T3* sgRNA became the default guide RNA used to target *oca2*. These results highlight possible influences of other factors outside of the PAM sequence and genomic accessibility that play significant roles in efficient Cas9-mediated gene targeting [146–149].

Having CRISPR screens in perspective, ordering commercially available synthetic crRNA and tracrRNA duplexes (crRNA:tracrRNA duplex) was considered over in-house synthesis of sgRNAs for time-saving purposes. Therefore, an initial comparison of heiCas9-mediated *oca2* knockout efficacies between in-house sgRNAs and synthetic crRNA:tracrRNA duplexes was required and consequently performed. Medaka 1-cell stage embryos were injected with *heiCas9* mRNA in addition to either in-house *oca2* sgRNA or using synthetic *oca2* crRNA:tracrRNA. The latter was injected either in duplex or singletons forms (i.e. direct addition of crRNA/tracrRNA components to the injection mixture without duplex formation by pre-incubation). As information on the use of synthetic guides in medaka embryos was lacking, two different concentrations were tested, one matching the standard sgRNA concentration used (0.46 pmol/µl) and a higher concentration (1.4 pmol/µl) suggested by the supplier for zebrafish targeting.

Overall quantification of eye pigment loss in *oca2* crispants (4 dpf) revealed a substantially higher bi-allelic knockout efficacy when using in-house sgRNA compared to synthetic cr/tracr RNAs, both in duplex or singleton forms (Figure 2.10). In-house sgRNA led to the majority of embryos showing higher than 80% loss of eye pigmentation, with complete absence of ocular pigments in some embryos. In contrast, duplexed cr/tracr RNAs at either tested concentration showed reduced efficacies, with less than 20 % loss of eye pigmentation. While the use of singletons, unexpectedly, at a concentration of 1.4 pmol/µl did improve the expected loss of ocular pigment outcome, accompanied by increased aberrant phenotypes and lethality rates, the in-house sgRNA remained clearly superior. On a side note, incubating embryos in ice-cold medium after injection with *heiCas9* mRNA and *oca2* sgRNA compared to direct incubation in room temperature

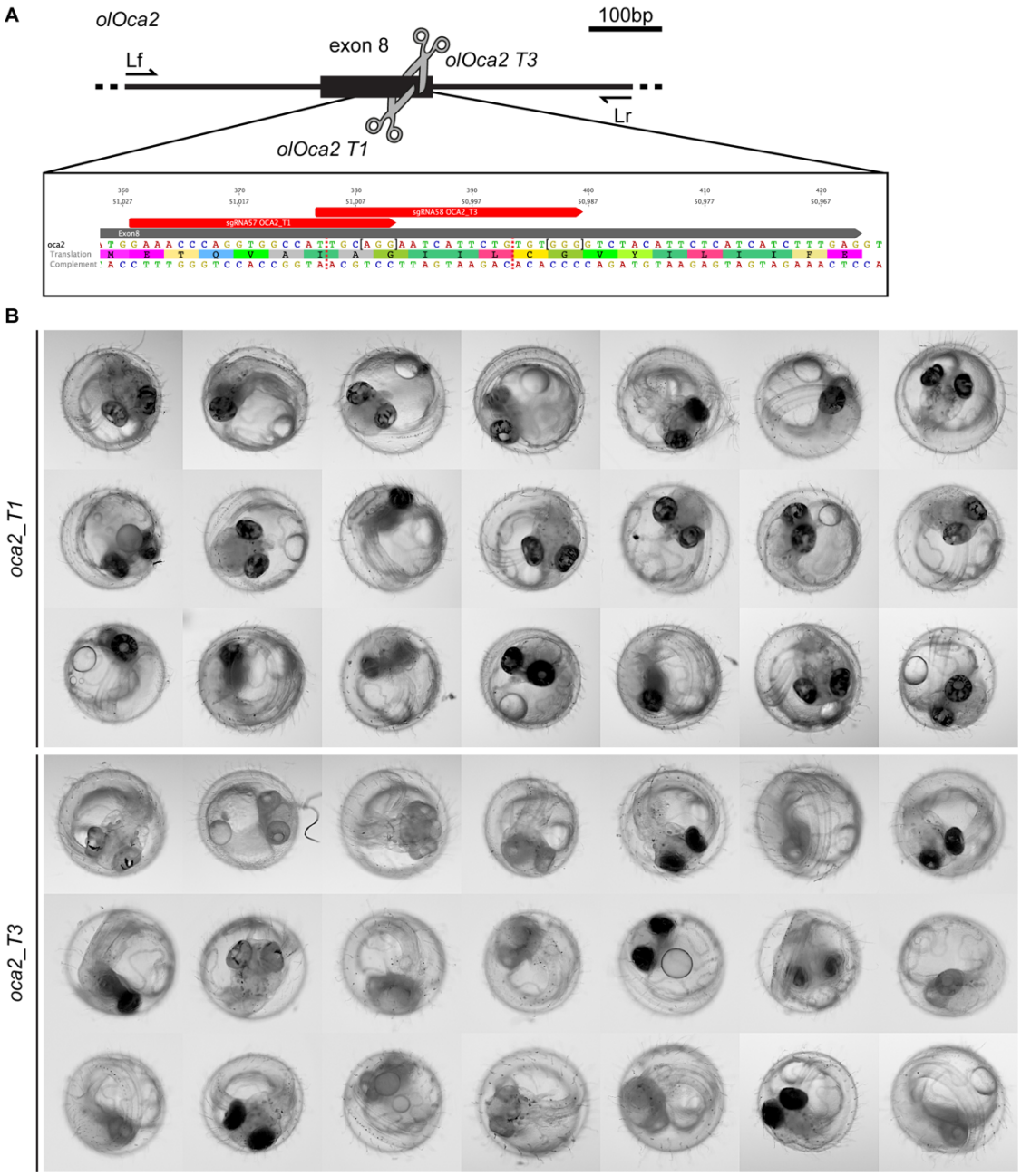


Figure 2.9 Two NGG-PAMed *oca2* sgRNAs with obvious differences in F0 knock-out efficacies
A) Schematic representation of the *oca2* locus in medaka (*olOca2*) to be targeted via CRISPR/Cas9. Two sgRNAs (*oca2_sg57_T1* and *oca2_sg58_T3*) designed to target the same region of interest (exon 8) within 18 bases from each other. Square brackets correspond to sgRNA's respective PAMs and dotted red line corresponds to respective Cas9 cut site. **B)** Medaka *oca2* crispant embryos (4 dpf) show difference in bi-allelic knock-out mediated loss of eye pigmentation phenotypes when injected at the one cell stage either with *oca2_sg57_T1* (*oca2_T1*) or with *oca2_sg58_T3* (*oca2_T3*).

medium did not affect the heiCas9 targeting efficiency. In sum, the results clearly demonstrated that our in-house sgRNAs are considerably more efficient and have less variability than the crRNA/tracrRNA counterparts, and that the heiCas9-mediated mutagenesis is extremely efficient already in the injected generation.

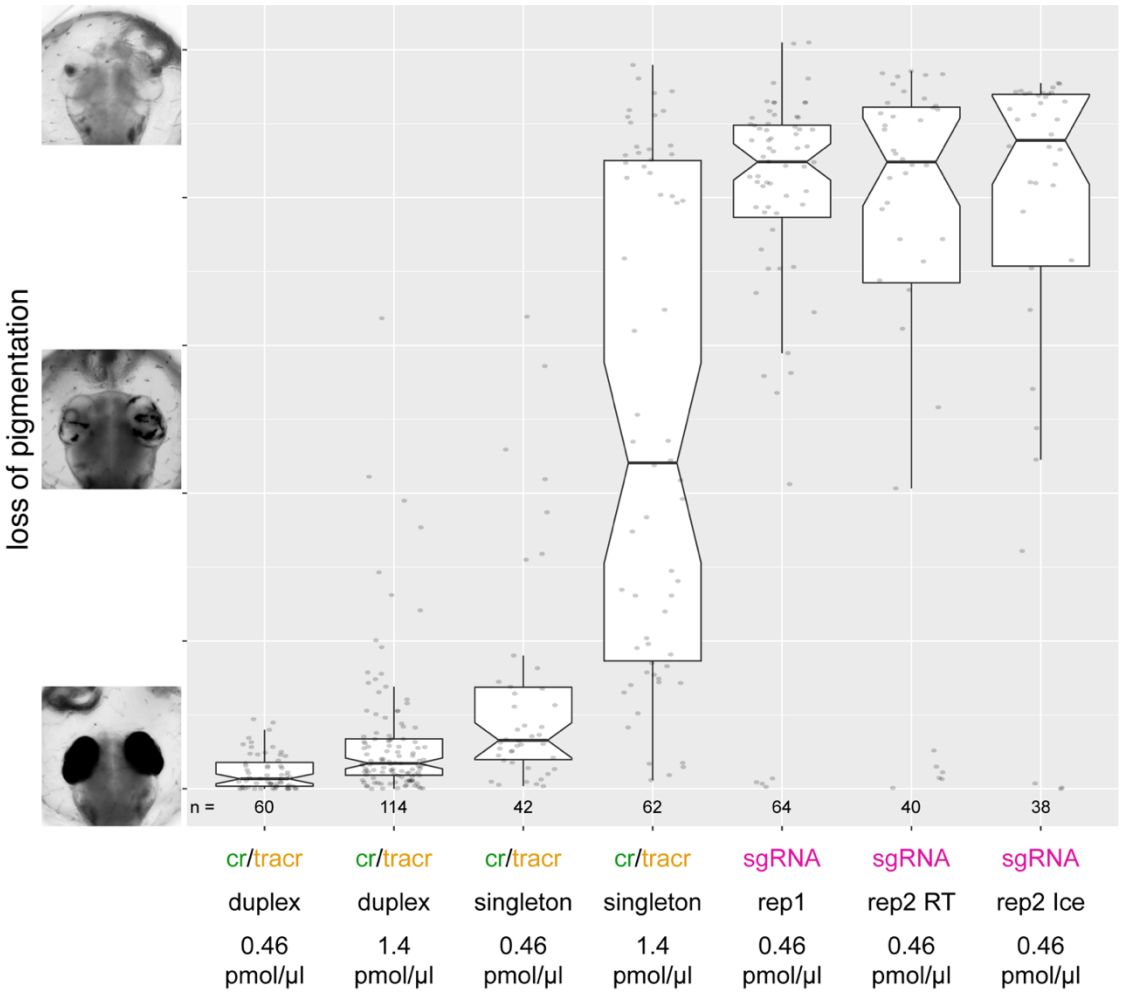


Figure 2.10 in-house sgRNAs result in superior heiCas9-mediated editing efficacy compared to synthetic cr/tracrRNA
Comparison of heiCas9-mediated, bi-allelic knock-out efficacies when targeting *oca2* using in-house synthesized sgRNA and commercial synthetic crRNA/tracrRNA (IDT) reveals drastic differences in proportions of near pigment-less embryos. sgRNA and crRNA were designed to target the same exact *oca2* sequence. Medaka embryos were injected at the 1-cell stage with *heiCas9* mRNA (150 ng/µl) in addition to *oca2_T3* sgRNA (15 ng/µl, 0.46 pmol/µl), or *oca2* synthetic crRNA/tracrRNA either by direct addition of each of crRNA and tracrRNA components individually in the injection mix (singletons) or by pre-incubating them together prior to their addition to the injected mix (duplex). Two concentrations (0.46 and 1.4 pmol/µl) of the crRNA/tracrRNA were tested. Medaka embryos (4 dpf) were imaged and mean grey values of individual eyes from *oca2* crispants were quantified for broad phenotypic classification between full pigmentation (0) and complete loss of pigmentation (255) (embryo representative examples displayed on the left). Note: no significant difference in knock-out efficiency observed when incubating injected embryos for prolonged period in ice-cold medium. Data is presented as notched box plots (median ± interquartile range) with original values overlaid as scatter plots. Number of individual eyes is denoted (n).

2.2.2 Functional gene validation pipeline confirms heart phenotypes in *nkx2-5* crispants

After the proven high efficiency of heiCas9-mediated knockout of *gfp* and *oca2* in F0, the feasibility of this approach to validate heart-relevant genes was next tested. I therefore combined the powerful CRISPR/Cas9-mediated targeted gene inactivation strategy to create crispants which I subjected to high-throughput imaging and heart rate analysis.

Using this pipeline allowed me to validate the impact of the gene-specific loss-of-function on the heart rate (Figure 2.11). A region of interest (ROI) surrounding a candidate CVD-associated coding SNP on the human gene was defined from which the medaka ortholog was identified. sgRNAs designed to target this orthologous ROI on the medaka gene were used for targeted CRISPR/Cas9-mediated gene editing, resulting in genetically mosaic crispant embryos for which heart rates of individual embryos are the ultimate quantified phenotyping readout. Crispant heart rates can then be compared to control siblings to measure the effect of loss-of-function of the targeted gene on the heart rate.

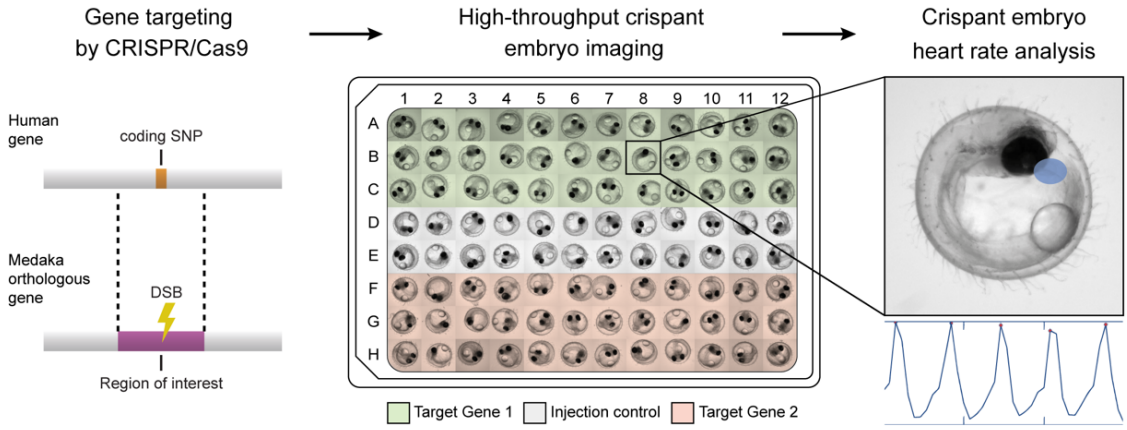


Figure 2.11 Schematic overview of the functional gene validation pipeline
Functional validation of human GWAS genes by defining a region of interest in the medaka gene orthologous to the potential disease-associated SNP in the corresponding human gene of interest, followed by CRISPR/Cas9-mediated targeting, resulting in double strand breaks (DSB) and subsequent NHEJ-mediated InDel mutations in a genetically mosaic fashion. Resulting crispant embryos (4 dpf) from two targeted genes (Target Gene 1 and 2; n = 36 embryos per condition), as well as siblings mock-injected with *GFP* mRNA as plate controls (Injection control; n = 24 embryos) are then individually loaded into 96-well plates and subjected to the established high-throughput heart rate assay. Figure adapted from [144].

A control test-run of the pipeline (Figure 2.12A) was performed to assess the robustness of the embryos by measuring their heart rate in response to extensive manipulation (e.g., handling, injecting, medium change and screening). Medaka wild-type embryos of the Cab strain were injected at the 1-cell stage with *gfp* mRNA as an injection tracer. Embryos are incubated until heart function has fully developed and the heart rate has reached a plateau at 4 dpf [72]. Embryos are then individually loaded on a 96-well plate and subjected to automated imaging (i.e. the heart rate assay) at two different temperatures (21 and 28°C) to measure heart function under different environmental conditions. The plate containing embryos is then be lysed and subjected to genotyping using the SLEDGE-Hammer protocol (whole plate or select embryos) followed by downstream targeting confirmation using T7EI mismatch assay [139]. Importantly, while embryos experience a variety of inevitable environmental and physical manipulations during the procedure, the heart rate did not seem significantly different between *gfp* mRNA (Mock) injected embryos and their un-injected wild-type (WT) siblings (Figure 2.12B).

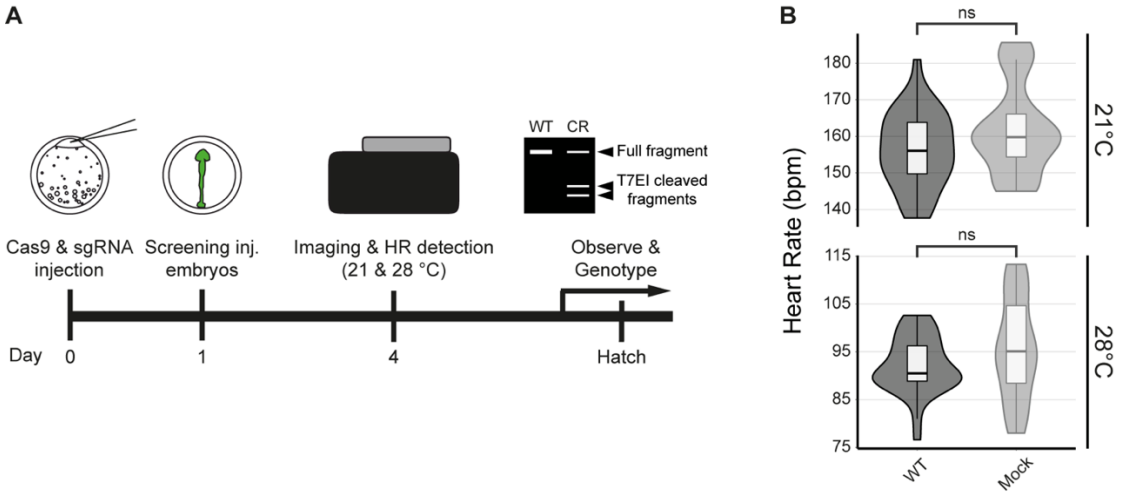


Figure 2.12 Embryo manipulation does not significantly impact on the heart rate
A) Schematic overview of the workflow of the validation assay. Medaka embryos are injected at the 1-cell stage either with *heiCas9* mRNA, sgRNA designed against target candidate gene and *GFP* mRNA as injection tracer, or only with *GFP* mRNA as injection control. At 1 dpf, positively injected embryos are screened and selected following GFP tracer expression. At 4 dpf, embryos are individually loaded into 96-well plates and subjected to the heart rate assay at two temperatures (21 and 28°C). After imaging, embryos are lysed and randomly selected embryos per condition are genotyped to confirm CRISPR-mediated editing following the SLEDGE-Hammer protocol. Alternatively, embryos can be kept longer for further observations until hatching. Note: Embryos are always maintained at 28°C from injection until imaging at 4 dpf. **B)** Comparative heart rate analysis of non-injected wild-type medaka embryos (WT; dark grey) and *GFP* mRNA mock-injected siblings (Mock) shows no significant heart rate phenotypes at both measured temperatures. Significant differences in (**B**) were determined by Student's *t*-test; ns (not significant; light grey). Data in (**B**) is presented as violin box plots (median ± interquartile range). Number of embryos successfully scored in (**B**) can be found in Appendix Table 1.

Next, I chose the cardiac-specific homeobox-containing transcription factor NKX2-5 as a positive control for an initial assessment of the F0 gene validation pipeline. A single amino acid mutation (R141C) in the NKX2-5 homeodomain was associated with atrial septal defect (ASD) in human patients and reported to cause delayed heart morphogenesis in adult mice [150]. Therefore, the region orthologous to R141C in medaka was chosen for CRISPR/Cas9 targeting using the highly efficient heiCas9. To address the specificity of the CRISPR/Cas9 system, a heart-unrelated gene, *oca2*, with no previously reported cardiac effects was chosen as a putative negative control [101,107,139]. Moreover, despite the medaka Cab strain being highly isogenic with low inter-individual variations, some minor differences in heart rate baselines were observed in wild-type embryos collected and imaged at different days. To account for day and plate specific fluctuations, mock-injected siblings were always included as internal plate controls.

In contrast to mock-injected embryos, *nkx2-5* crispants displayed cardiac-specific phenotypes, including enlarged heart chambers (Figure 2.13A), resembling previously reported phenotypes in zebrafish *nkx2-5* mutants [151]. Other cardiac phenotypes observed were heart-looping defects resulting in tubular-shaped hearts and retrograde blood flow presumably resulting from valve defects. Some crispants, however, showed rather unspecific, global developmental phenotypes (Figure 2.13B). Notably, *oca2* crispants did not show cardiac or global phenotypes (Figure 2.13A), verifying the lack of an influence

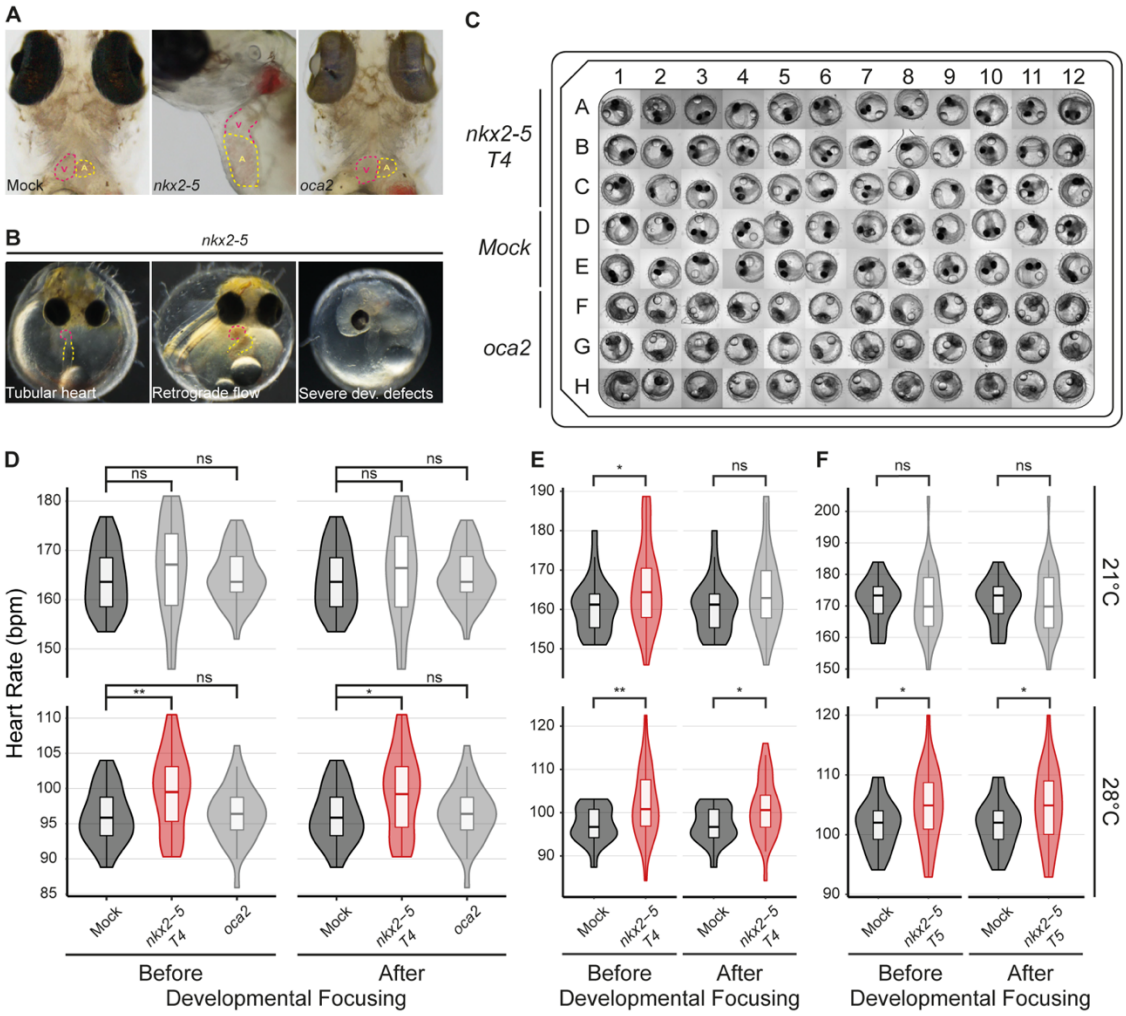


Figure 2.13 Functional gene validation pipeline consistently recapitulates relevant heart phenotypes in *nkx2-5* crispants

A) Morphological assessment of embryo (9 dpf) hearts (A: atrium, dotted red line; V: ventricle, dotted yellow line) reveal dilated heart chambers in *nkx2-5* crispants not observed in *oca2* crispants or mock-injected siblings (Mock), the latter two are indistinguishable. Note: the high efficiency of heiCas9-mediated bi-allelic knock out in the injected generation is reflected by the prominent loss of eye pigmentation in *oca2* crispants. **B)** Additional heart-specific (tubular heart/looping defects and retrograde blood flow) but also global (overall developmental defects) morphological phenotypes observed in *nkx2-5* crispants (4 dpf). **C)** 96-well plate layout containing medaka (4 dpf) *nkx2-5* and *oca2* crispant embryos as well as GFP-injected siblings (Mock). Note: prominent loss of eye pigmentation in *oca2* crispant embryos. **D)** Comparative heart rate analysis of embryos from **(C)** at 21 and 28°C, before (left) and after (right) developmental focusing reveal increased mean heart rates of *nkx2-5* crispant embryos compared to GFP-injected siblings (Mock; dark grey), significant at 21°C (red). No significant difference in heart rates was observed between *oca2* crispants and mock control. **E-F)** Consistent significant heart rate phenotypes of *nkx2-5* crispants at 21°C in **(E)** a second replicate using the same sgRNA (*nkx2-5_T4*) as well as **(F)** when using a different sgRNA targeting the same region of interest (*nkx2-5_T5*). Significant differences in (**D-F**) are depicted in red and were determined by Student's *t*-test; $*p < 0.05$, $**p < 0.01$, ns (not significant; light grey). Data in (**D-F**) is presented as violin box plots (median ± interquartile range). Number of embryos successfully scored in (**D-F**) can be found in Appendix Table 1. Panels **A** and **C-F** are adapted from [144].

of *oca2* targeting and the injection procedure on the heart and general development per se. Crispant embryos (n=36 per target gene) and mock-injected control siblings (n=24) were loaded onto a 96-well plate to assess changes in mean heart rates with statistical significance (Figure 2.13C). A general increase in mean heart rates of *nkx2-5* crispants

compared to mock control siblings is observed with a significant difference at 21°C (Figure 2.13D; left panel). Consistently, independent experimental replicates using the same sgRNA (*nkx2-5_T4*; Figure 2.13E) or a different sgRNA targeting the same exon (*nkx2-5_T5*; Figure 2.13F) yielded a significant heart rate phenotype at 21°C. Meanwhile, *oca2* crispants did not display any significant heart rate deviations compared to mock controls at both measured temperatures, confirming *oca2* as a bona fide negative control gene.

To address a concern that the occurrence of *nkx2-5* crispants displaying severe global developmental phenotypes (Figure 2.13B) might potentially skew heart rate effects, a developmental focusing filter was applied. Therefore, only those embryos having developed beyond stage 28 [79], when the heart rate had already reached a plateau [72], are considered for statistical analysis. Developmental focusing was therefore applied, and only three embryos needed to be excluded from the *nkx2-5* group and none from the mock or *oca2* groups. Nevertheless, it had no major impact on the outcome of the analysis (Figure 2.13D-F; right panel). The results demonstrate the robustness of the heart rate as a phenotypic readout as well as the overall validation pipeline employed. The results also showcase the assay's sensitivity at detecting mild heart rate phenotypes reflecting cardiac function already at embryonic stages of medaka development. This was mainly possible due to medaka's low inter-variability as a result of successive inbreeding over multiple generations.

2.2.3 Unbiased gene targeting verifies positive correlation between heart-related genes and heart rate phenotype

To determine the baseline occurrence of heart rate phenotypes in crispants from a randomized selection of genes, I used a random number generator to pick ten genes out of a total of 23622 annotated medaka coding genes in Ensembl [152] (Table 2.1). A random exon was chosen for each gene for CRISPR/Cas9 targeting, and heart rates of crispants were then measured as previously described.

Table 2.1 List of random gene selection
Medaka Ensembl gene names and codes, as well as orthologous human genes as annotated in the 95th Ensembl release.

Ensembl ID	Medaka Gene	Orthologous Human Gene
ENSORLG00000006335	novel gene - *cdc42*	na (*CDC42* by name)
ENSORLG00000000979	novel gene - *ogdh*	*OGDH*
ENSORLG00000005268	*duox*	*DUOX1*
ENSORLG00000007310	*git2*	*GIT2*
ENSORLG00000003492	*mus81*	*MUS81*
ENSORLG00000020766	*or124-2*	na

ENSORLG00000007400	*plekha8*	*PLEKHA8*
ENSORLG00000022757	*ttl*	*TTL*
ENSORLG00000005922	*cabp4*	*CABP2*
ENSORLG00000023106	novel gene - *eml6*	*EML6*

Heart rate phenotypes were observed in two out of ten randomly selected genes at both temperatures measured (Figure 2.14 and Figure 2.15). Remarkably, both genes, the *oxoglutarate dehydrogenase* (*ogdh*) and the *cell division control protein 42 homolog* (*cdc42)* were the only genes which have been previously connected to heart function. *cdc42* was previously reported to play a significant role in heart physiology, including cardiomyocyte proliferation, sarcomere organization and cell-cell adhesion during heart development [153,154]. On the other hand, *ogdh* had only been associated with heart phenotypes in human GWAS without experimental validation [155]. In addition to establishing the baseline of heart rate phenotypes upon the loss-of-function of any random medaka gene to 20%, I validated the role of *ogdh* in cardiac physiology, marking it as a potentially relevant cardiovascular disease marker.

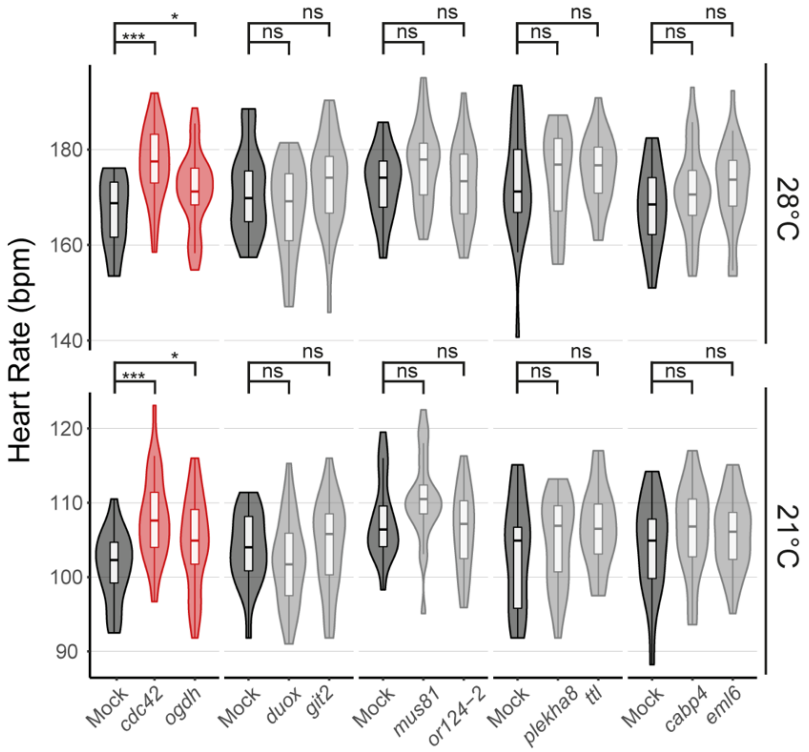


Figure 2.14 Only heart-related genes showed heart rate phenotypes in a random set of genes
Comparative heart rate analysis of GFP-injected (Mock; dark grey) and respective sibling crispant embryos (4 dpf) at 21 and 28°C after developmental focusing revealed heart rate phenotypes only in genes with previous cardiac-relevant findings (*cdc42*) or GWAS heart phenotype association (*ogdh*). Heatmap quantitative representation of the data available in (Figure 2.15). Breaks on the x-axis signify different experimental plates. Significant differences in mean heart rates between each crispant group and its plate-matching sibling mock group are depicted in red and were determined by Student's *t*-test; *$p < 0.05$, ***$p < 0.001$, ns (not significant; light grey). Data is presented as violin box plots (median ± interquartile range). Number of embryos successfully scored can be found in Appendix Table 2. Figure adapted and modified from [144].

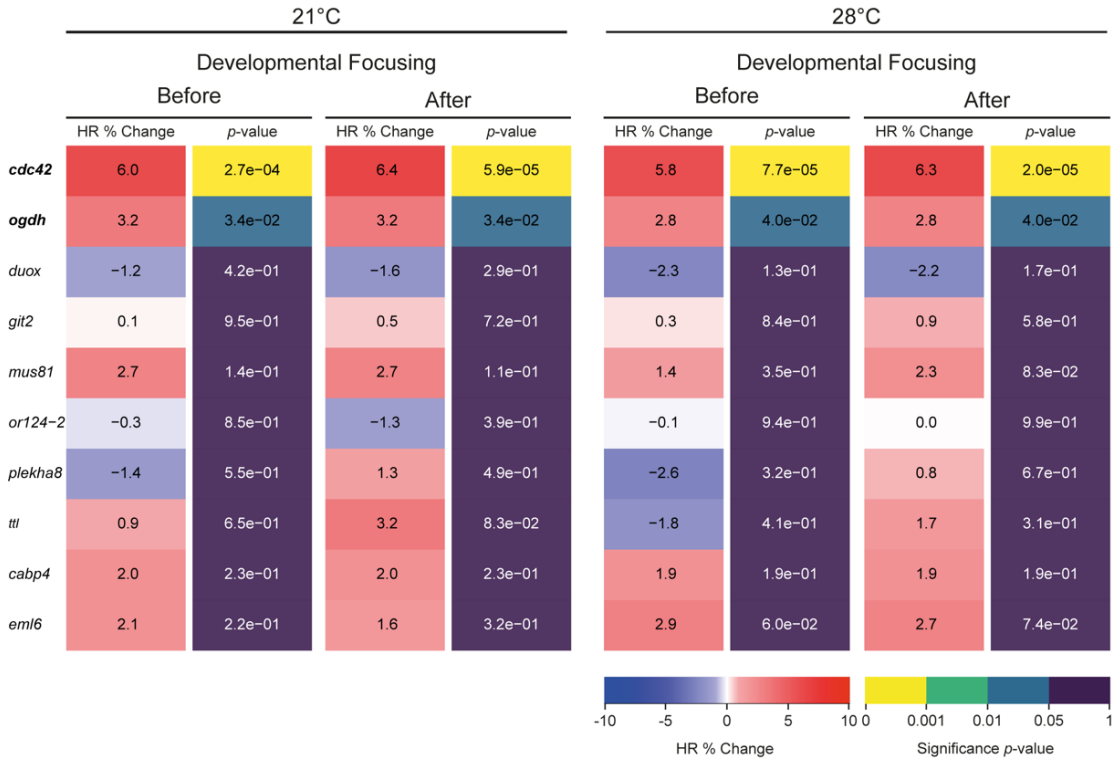


Figure 2.15 The heart rate is a robust phenotype unaffected by developmental focusing
Heatmap representation of the data shown in (Figure 2.14) before and after developmental focusing. Displayed are the percent change in mean heart rate (HR % Change) between each crispant group and its plate-matching sibling mock group, and the corresponding statistical significance (*p*-value) calculated by Student's *t*-test on the full distribution. Genes showing significant heart rate phenotypes are indicated in bold. Number of embryos successfully scored can be found in Appendix Table 2. Figure adapted and modified from [144].

The observed baseline probability of 20% of heart rate phenotypes resulting from targeting a randomly selected set of genes was unexpectedly high. To further investigate this, I performed repeated simulation rounds of random gene selection (data not shown). After four repeated rounds of random selection of ten genes using the random number generator, a similar mean baseline of 10-20% of randomly selected genes (per repeat) was observed to be either associated to heart phenotypes in human GWAS or were reported to have direct involvement in heart development or function. These findings thus further reflect on the vast polygenic nature of CVDs and how a prominent proportion of total genes (up to 20%) may contribute to CVDs onset and progression.

2.2.4 CRISPR/Cas9 screen of human heart-GWAS genes in medaka identifies new genes affecting heart rate

To identify novel players in cardiac function and diagnostic markers for CVDs, I performed a targeted selection of genes associated with heart phenotypes in human GWAS. The genome-wide repository of associations between SNPs and phenotypes (GRASP) [156] was used as a resource to compose a list of 40 candidate genes from human GWAS with a coding association (i.e. SNPs in exons) to heart phenotypes (hGWAS

genes; Table 2.2). Uncharacterized genes or those with no prior experimental link to heart functions were primarily chosen with few known heart genes included as positive controls. The selected hGWAS genes were segregated according to their associated phenotypic category into general heart-related phenotypes (n=17) or specifically heart-rate phenotypes (n=23) to assess the correlation between the associated phenotypes in human GWAS and the experimentally observed phenotypes in medaka crispant embryos.

Due to the proven superiority of the in-house synthesized sgRNAs compared to commercial synthetic crRNA/tracrRNA components (Figure 2.10), sgRNAs targeting the medaka gene region orthologous to the candidate heart GWAS-derived human SNP region were designed and synthesized in the lab. It is noteworthy to mention that NGG PAMs were mainly considered when designing sgRNAs. Contrary to the general belief that SpCas9 can also recognize NAG PAMs [98,99], none (0 out of 9) of the designed sgRNAs harbouring an NAG PAM resulted in CRISPR/Cas9-mediated InDel mutations *in vivo* in medaka embryos, while InDel mutations occurred when using NGG PAM-containing sgRNAs targeting the same region (data not shown).

Heart rates of candidate hGWAS genes crispants were scored and compared to mock-injected siblings before and after developmental focusing (Figure 2.16 and Figure 2.17). Across the hGWAS set of genes, a total of 16 genes exhibited significant heart rate phenotypes when targeted (Figure 2.18A). Notably, the majority of crispants (14 out of 16) yielded a significantly elevated mean heart rate compared to control siblings, while for two genes (*scn4ab* and *sspo*) a significantly decreased mean heart rate was observed. Furthermore, seven genes displayed significant heart rate phenotypes at both measured temperatures when targeted, while five and four genes displayed significant heart rate phenotypes only at 21°C or 28°C, respectively. The five positive controls (*TTN, NACA, CASQ2, KCNH2,* and *SCN5A*) clearly responded in the assay. However, *HCN4*, another positive control, surprisingly did not yield a significant heart rate phenotype, possibly due to compensation by its paralog *hcn4l*. *A priori* reports of involvement in heart function included cardiac contractility for *TTN* [157] and *NACA* [158]), and regulation of heart rate for *CASQ2* [159], *KCNH2* [160], and *SCN5A* [161,162]). By identifying 11 genes through the established assay (*CCDC141, GIGYF1, HOMEZ, MYRF, SMG6, CMYA5, CNOT1, SLC17A3, TRAPPC12, SSPO and PADI4*) I could strengthen our understanding of the genes' involvement in cardiac function [163–166], and in some cases demonstrate the very first experimental evidence linking these genes to cardiac function.

The overall proportion of genes yielding heart rate phenotypes in the targeted hGWAS (16/40; 40%) and the randomly selected (2/10; 20%) set of genes confirmed the positive correlation between phenotypes observed in embryonic medaka crispants and the associated phenotype in adult human GWAS. With special consideration that both heart rate yielding genes in the random set were actually associated to heart phenotypes (Figure 2.18B). Remarkably, this positive correlation became even stronger upon segregation of candidates according to their GWAS association group ("heart rate" and "non-heart rate"

Table 2.2 List of hGWAS candidate gene selection

Human heart phenotype-associated genes, extracted from human GWAS using GRASP 2.0 database, categorized according to their association into "heart rate" (**bold**) and "non-heart rate" (non-bold) related phenotypes in human GWAS.

Human Gene	Coding SNP ID	Associated heart phenotype	Association Reference	Medaka orthologue gene (Ensembl release)	Ensembl Gene Code
ATP8B4	**rs2452524**	**Pulse rate**	**(Hiura et al. 2010)**	***atp8b4*** **(98)**	**ENSORLG00000005106**
CASQ2	**rs4074536**	**QRS interval**	**(Sotoodehnia et al. 2010)**	***casq2*** **(89)**	**ENSORLG00000017885**
CCDC141	**rs17362588**	**Heart rate**	**(Hoed et al. 2013)**	**na (89) (TBLASTN=*ccdc141*)**	**TBLASTN = ENSORLG00000030409**
CEP85L	**rs3734381**	**QRS interval**	**(Sotoodehnia et al. 2010)**	***cep85l*** **(91)**	**ENSORLG00000015455**
CMYA5	**rs10942901**	**Heart rate**	**(Hoed et al. 2013)**	***cmya5*** **(91)**	**ENSORLG00000008983**
COL9A1	**rs592121**	**Pulse rate**	**(Hiura et al. 2010)**	***col9a1b*** **(98)**	**ENSORLG00000010431**
GIGYF1	**rs221794**	**Heart rate**	**(Hoed et al. 2013)**	***gigyf1*** **(89)**	**ENSORLG00000003655**
GRID2	**rs1385405**	**Pulse rate**	**(Hiura et al. 2010)**	***grid2*** **(98)**	**ENSORLG00000024663**
HOMEZ	**rs1055061**	**Sick sinus syndrome**	**(Holm et al. 2011)**	***homeza*** **(89)**	**ENSORLG00000012220**
KCNH2	**rs1805123**	**QT interval**	**(Pfeufer et al. 2009)**	***kcnh2*** **(98)**	**ENSORLG00000004137**
MINAR1	**rs2297773**	**Pulse rate**	**(Hiura et al. 2010)**	***minar1*** **(98)**	**ENSORLG00000016707**
MYRF	**rs174535**	**RR interval**	**(Eijgelsheim et al. 2010)**	***myrf*** **(91)**	**ENSORLG00000006459**
NACA	**rs2926743**	**Heart rate**	**(Hoed et al. 2013)**	***naca*** **(98)**	**ENSORLG00000012246**

Human Gene	**Coding SNP ID**	**Associated heart phenotype**	**Association Reference**	**Medaka orthologue gene (Ensembl release)**	**Ensembl Gene Code**
OR5AU1	**rs4982419**	**Pulse rate**	**(Hiura et al. 2010)**	**na (98) (TBLASTN=no name)**	**TBLASTN = ENSORLG00000024679**
PADI4	**rs2240335**	**Pulse rate**	**(Hiura et al. 2010)**	**na (98) (TBLASTN=*padi2*)**	**TBLASTN = ENSORLG00000007539**
PPP1R9A	**rs854524**	**Pulse rate**	**(Hiura et al. 2010)**	***ppp1r9a* (98)**	**ENSORLG00000004418**
RNF207	**rs846111**	**QT interval**	**(Newton-Cheh et al. 2009)**	***rnf207b* (91)**	**ENSORLG00000017207**
SCN5A	**rs1805126**	**QRS interval**	**(Ritchie et al. 2013)**	**na (89) (TBLASTN=*scn4ab*)**	**TBLASTN = ENSORLG00000003273**
SSPO	**rs10261977**	**Pulse rate**	**(Hiura et al. 2010)**	***sspo* (98)**	**ENSORLG00000004121**
TRAPPC12	**rs6767**	**Pulse rate**	**(Hiura et al. 2010)**	***trappc12* (98)**	**ENSORLG00000017859**
TTN	**rs12476289**	**QT interval**	**(Marroni et al. 2009)**	***ttn.2* (91)**	**ENSORLG00000018144**
UFSP1	**rs12666989**	**RR interval**	**(Eijgelsheim et al. 2010)**	**na (89) (TBLASTN=*ufsp1*)**	**TBLASTN = ENSORLG00000022928**
XYLB	**rs17118**	**PR interval**	**(Smith et al. 2009)**	***xylb* (91)**	**ENSORLG00000003755**
ABCB1	rs1128503	Drug response CVD	(Paré et al. 2013)	*abcb4* (91)	ENSORLG00000009269
BAG3	rs3858340	Sporadic dilated cardiomyopathy	(Villard et al. 2011)	*bag3* (89)	ENSORLG00000013813
CLCNKA	rs1805152	Sporadic dilated cardiomyopathy	(Villard et al. 2011)	*clcnk* (89)	ENSORLG00000018693
CNOT1	rs11866002	Aortic valve calcium	(Thanassoulis et al. 2013)	*cnot1* (91)	ENSORLG00000013734
EDN1	rs150035515	Aortic valve calcium	(Matsa et al. 2014)	*edn1* (89)	ENSORLG00000009276

Human Gene	**Coding SNP ID**	**Associated heart phenotype**	**Association Reference**	**Medaka orthologue gene (Ensembl release)**	**Ensembl Gene Code**
HCN4	rs529004	Aortic valve calcium	(Thanassoulis et al. 2013)	*hcn4* (91)	ENSORLG00000013180
MAML3	rs11729794	Congenital heart malformations	(Hu et al. 2013)	na (91) (TBLASTN=*maml3*)	TBLASTN = NCBI Ref Seq: XM_023954746.1
NUBP2	rs344359	LV systolic dysfunction	(Vasan et al. 2009)	*nubp2* (91)	ENSORLG00000007228
PIEZO1	rs2290902	Bicuspid aortic valve	(Wooten et al. 2010)	*piezo1* (91)	ENSORLG00000000402
PLG	rs13231	Aortic valve calcium	(Thanassoulis et al. 2013)	*plg* (91)	ENSORLG00000020532
RGS3	rs12341266	Hypertrophic cardiomyopathy	(Wooten et al. 2013)	*rgs3a* (91)	ENSORLG00000006823
SCMH1	rs10489520	Ischemic stroke	(Ikram et al. 2009)	*scmh1* (91)	ENSORLG00000014207
SH2B3	rs3184504	Tetrology of fallot	(Cordell et al. 2013)	*sh2b3* (91)	ENSORLG00000003569
SLC17A3	rs942379	Bicuspid aortic valve	(Wooten et al. 2010)	*si:ch1073-513e17.1* (91)	ENSORLG00000007671
SMG6	rs216193	Aortic root size	(Vasan et al. 2009)	*smg6* (91)	ENSORLG00000003317
VEPH1	rs1378796	Sporadic dilated cardiomyopathy	(Villard et al. 2011)	*veph1* (89)	ENSORLG00000012452
ZFHX3	rs2228200	Aortic valve calcium	(Thanassoulis et al. 2013)	*zfhx3* (91)	ENSORLG00000007874

phenotypes). A substantial portion (13/23; ~57%) of the heart rate-associated genes in hGWAS indeed resulted in heart rate phenotypes in medaka embryos. Conversely, a much lower proportion (3/17; ~18%) of non-heart rate-associated genes resulted in heart rate phenotypes. Taken together, I could uncover clinically relevant heart rate phenotypes in previously uncharacterized genes, expanding the list of putative risk factors in human adults.

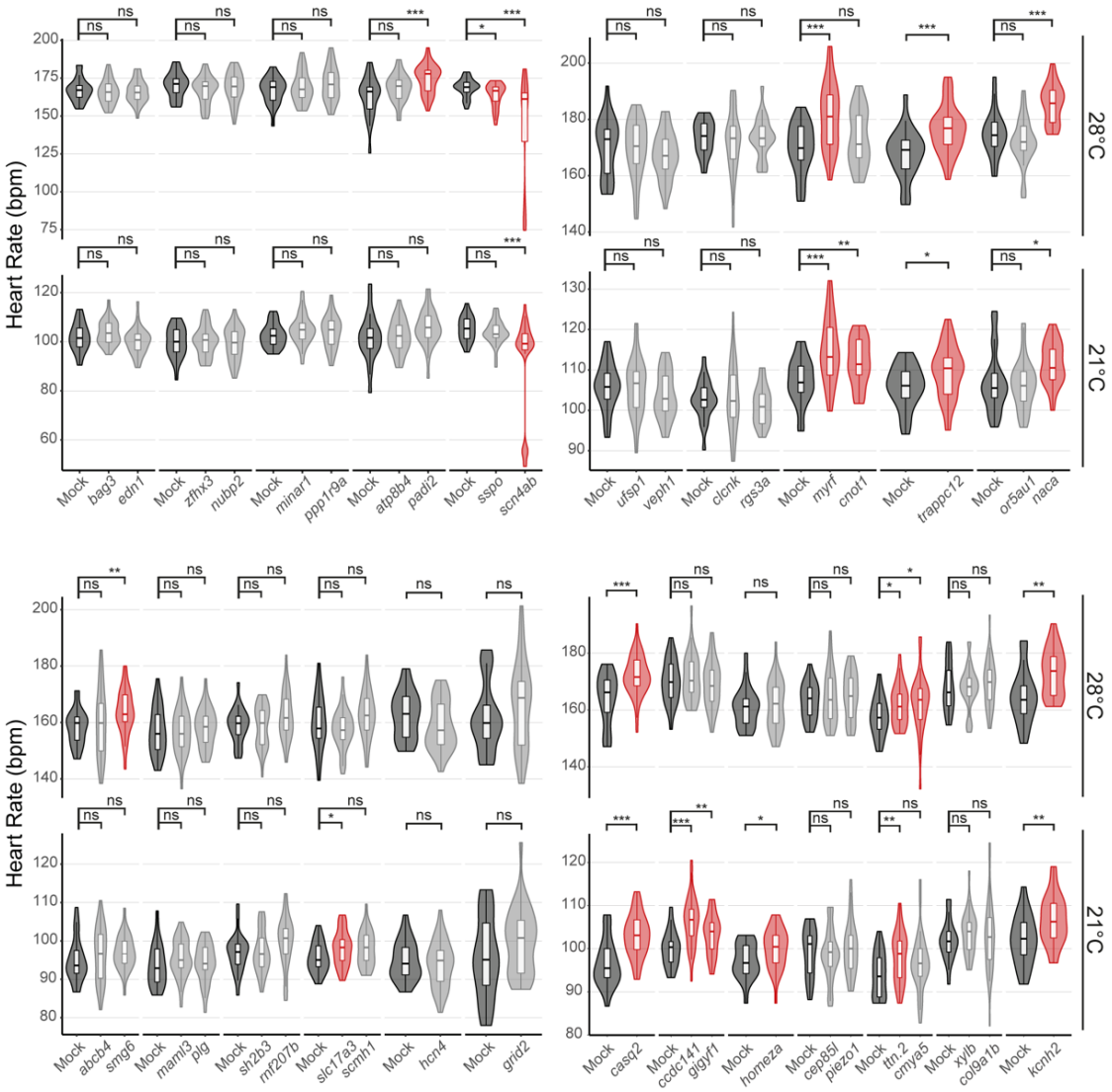


Figure 2.16 Comparative Heart rate analysis in targeted hGWAS set of genes
Comparative analysis of mean heart rates of GFP-injected (Mock; dark grey) and respective sibling crispant embryos (4 dpf) at 21 and 28°C after developmental focusing revealed heart rate phenotypes in 16 out of 40 genes. Heatmap quantitative representation of the data available in (Figure 2.17). Breaks on the x-axis signify different experimental plates. Significant differences in mean heart rates between each crispant group and its plate-matching sibling mock group are depicted in red and were determined by Student's *t*-test; **p* < 0.05, ***p* < 0.01, ****p* < 0.001, ns (not significant; light grey). Data is presented as violin box plots (median ± interquartile range). Number of embryos successfully scored can be found in Appendix Table 3. Figure adapted from [144].

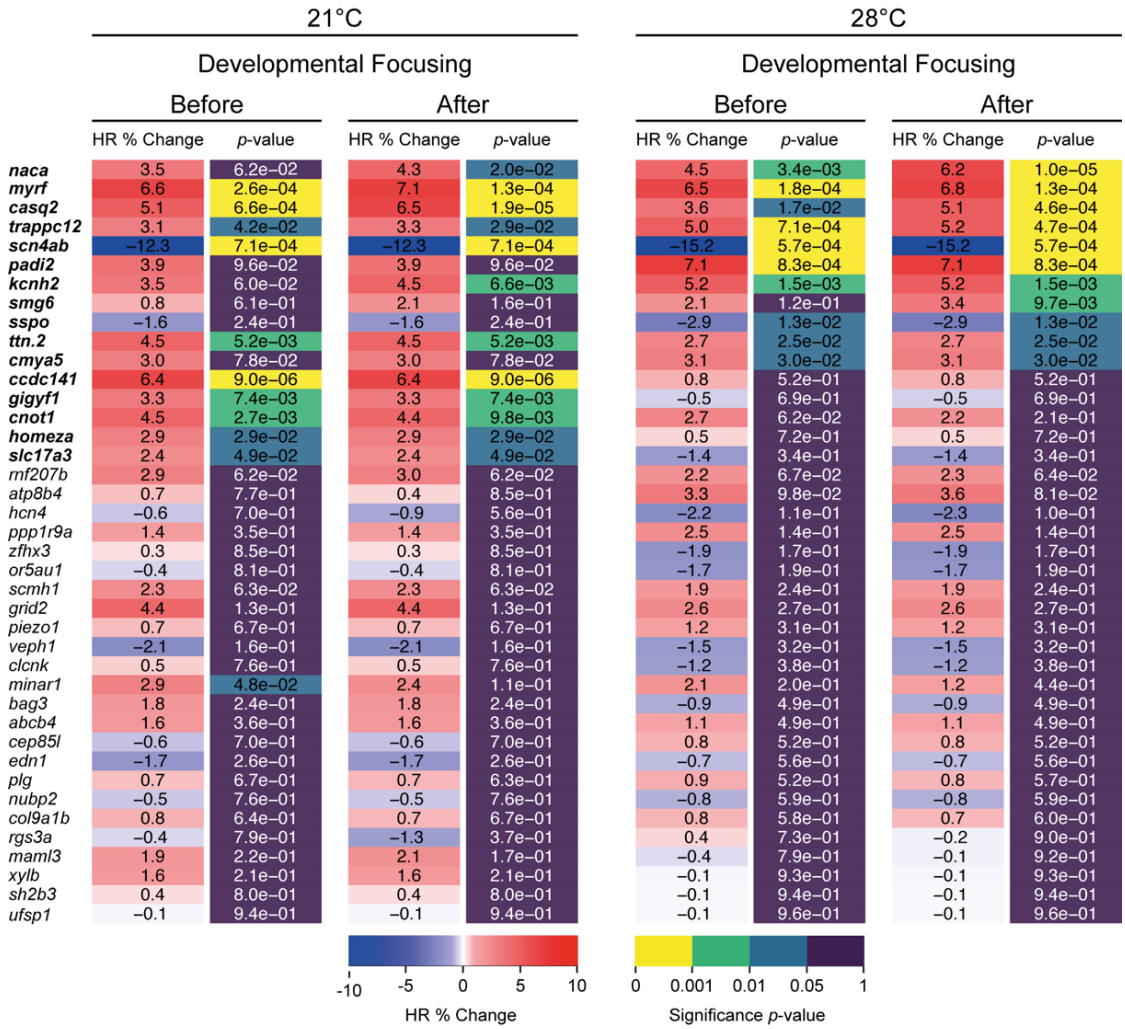


Figure 2.17 Developmental focusing did not significantly alter the outcome of targeted hGWAS analysis

Heatmap representation of the data shown in (Figure 2.16) before and after developmental focusing. Displayed are the percent change in mean heart rate (HR % Change) between each crispant group and its plate-matching sibling mock group, and the corresponding statistical significance (*p*-value) calculated by Student's *t*-test on the full distribution. Genes showing significant heart rate phenotypes are indicated in bold. Genes are ordered by their significance after developmental focusing at 28°C then 21°C. Number of embryos successfully scored can be found in Appendix Table 3. Figure adapted from [144].

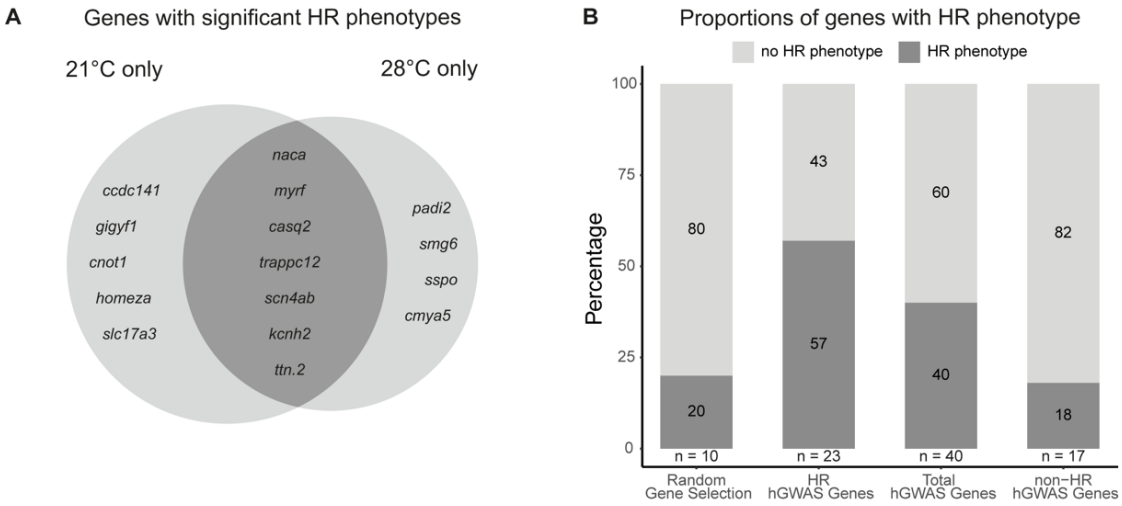


Figure 2.18 Targeted F0 validation of hGWAS genes yields 16 genes with significant heart rate phenotypes

A) Venn diagram recapitulating genes eliciting significant heart rate phenotypes (HR) at 21°C only, 28°C only or at both recorded temperatures (dark grey segment). **B)** Stacked plots depicting proportion of genes in each targeting group yielding significant heart rate phenotypes (dark grey). hGWAS corresponds to heart phenotype-associated genes selected from human GWAS (Table 2.2), of which some were associated to heart rate related phenotypes (HR hGWAS Genes) and the rest were associated to non-heart rate related phenotypes (non-HR hGWAS Genes). Number of genes in each group is denoted (n). Note the positive correlation between the observed phenotype in medaka crispant embryos and the associated phenotype in adult human GWAS. Figure adapted from [144].

2.2.5 Beyond the heart rate: CVD-relevant phenotypes observed in hGWAS CRISPR/Cas9 screen

Despite my focus of the F0 CRISPR/Cas9 screen on the heart rate phenotype, I still identified and scored other cardiovascular phenotypes for some of the candidate genes. In addition to the significantly elevated heart rate phenotype, medaka *trappc12* crispants displayed morphological heart aberrations, such as heart looping defects. In wild-type medaka, heart looping usually occurs at around stage 27, as the atrium shifts upwards to the right and lies adjacent to the ventricle [79]. Almost 26% of *trappc12* crispant embryos (12/46) showed heart looping defects not observed in *oca2* crispants (0/22) or mock-injected control siblings (0/26) (Figure 2.19). TRAPPC12 is part of the multi-subunit Transport Protein Particle III (TRAPPIII) complex involved in membrane trafficking and autophagy [167]. An attempt to visualize its expression domains at various embryonic stages corresponding to check points in medaka heart development was performed via *in situ* hybridization (ISH) using a *trappc12* probe. The ubiquitous and unspecific signal, however, makes any stage or expression domain-based speculation impossible (data not shown). This could result from *trappc12's* very low and ubiquitous expression, also evident from its low expression profile at the transcriptomic level across various developmental stages in medaka [168].

Furthermore, the occurrence of phenotypes beyond heart rate was assessed across a subset of the hGWAS genes validated using our assay. Incidences of severe global developmental phenotypes (Figure 2.20A), as well as cardiovascular-relevant phenotypes

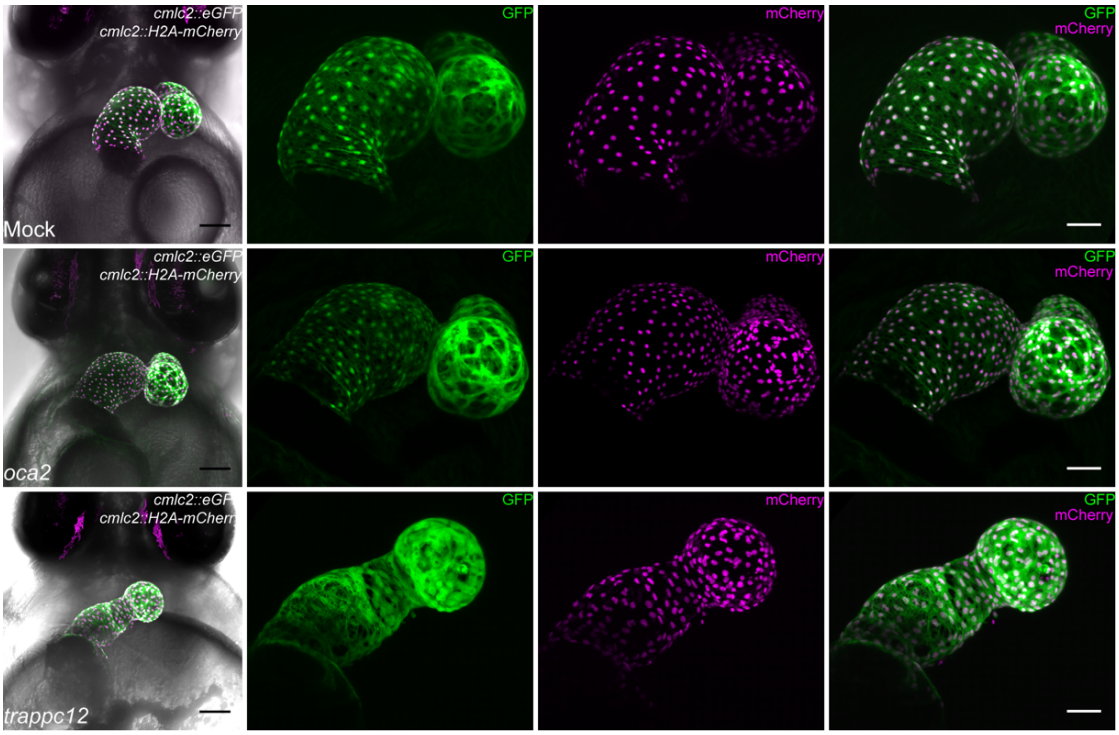


Figure 2.19 Heart looping defects revealed in medaka *trappc12* crispant embryos
Confocal images of representative hearts of the *cmlc2::eGFP cmlc2::H2A-mCherry* dual reporter line (7 dpf) reveal heart looping defects in *trappc12* crispant embryos but not in *oca2* crispants or mock-injected siblings. Scale bars: 100 µm (first panel on left) and 50 µm (blow-up images). Figure adapted from [144].

such as blood clot formation (Figure 2.20B) and morphological heart phenotypes (e.g., looping defects and clearly enlarged or diminished chambers) (Figure 2.20C) were scored for each target gene of this subset. Although it is reasonable to expect a heart rate phenotype from a morphologically defective heart, for a number of genes the percent incidence of morphological heart phenotypes or blood clots is interestingly high despite not yielding a significant heart rate phenotype. Therefore, the results demonstrate that although the heart rate is an important risk factor, other cardiovascular features are also as important to understand the functional role of the genes in heart regulation. Ideally, a screen scoring for a matrix of phenotypes would maximise the relevance in the context of cardiovascular-gene validations.

For *scn4ab* crispants, an unusual bimodal distribution of heart rates was observed, where some crispants displayed half the average heart rate at both measure temperatures (Figure 2.21A). An arrhythmia phenotype, known as an atrio-ventricular block (AV-block) previously reported in zebrafish *scn5a* mutants [161], was revealed upon in-depth investigation of these crispants. Characterized by an impairment of the electrical signal transmitted from atrium to ventricle, AV-block arrhythmias represent a clinically relevant phenotype in humans, and we were able to detect and score it using our assay. Measuring the beat frequency in atrium and ventricular chambers separately further exposed the impaired atrial to ventricular contractions in *scn4ab* crispants but not in control siblings (Figure 2.21B). Remarkably, the penetrance of the arrhythmia phenotype in the injected F0 generation was impressively high, exceeding 90% of *scn4ab* crispants. *scn4ab*

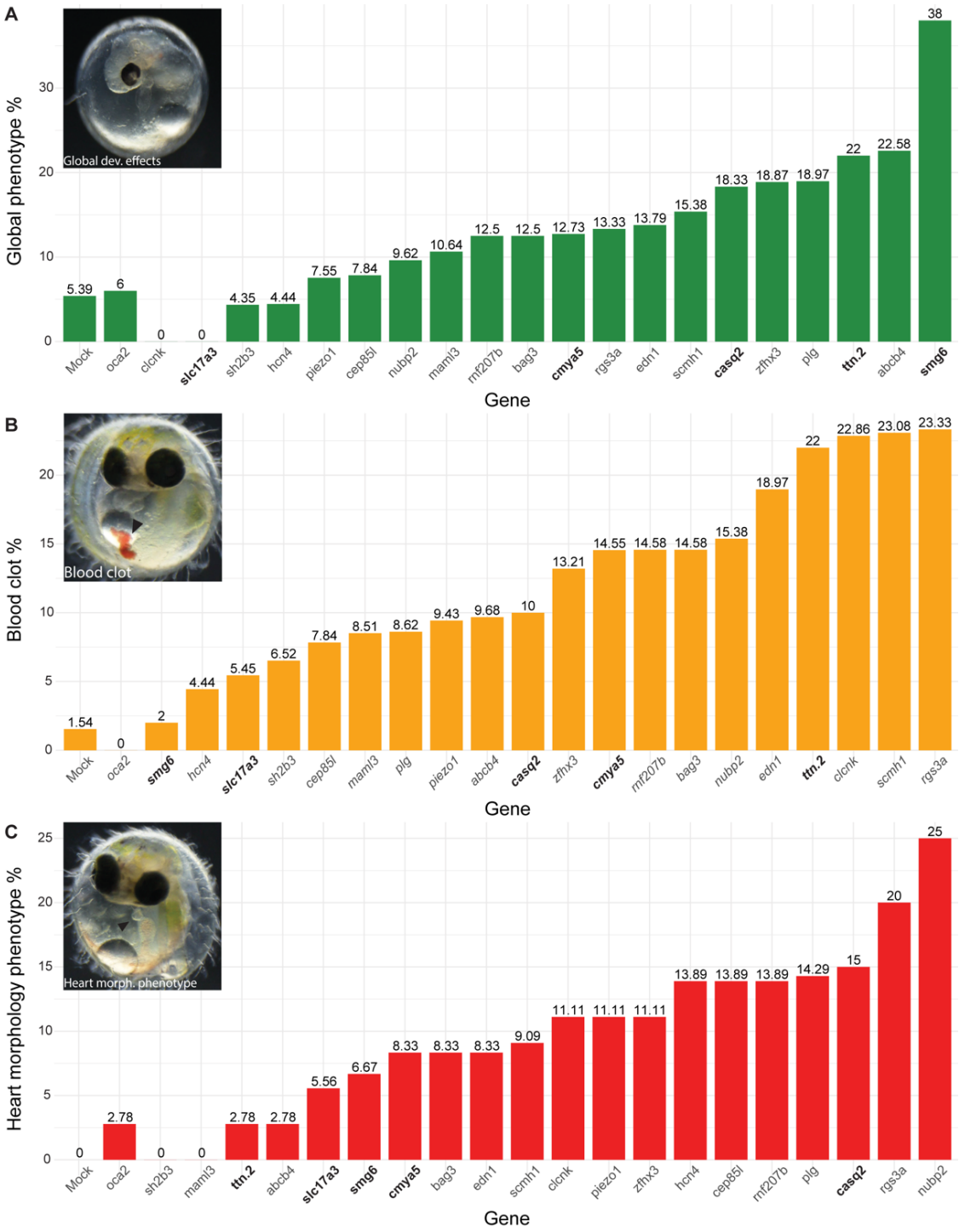


Figure 2.20 CVD-relevant phenotypes beyond heart rate observed in a subset of hGWAS crispants

Crispant embryos from a subset of the hGWAS genes were further examined for potential cardiovascular-relevant morphological phenotypes. Accordingly, displayed are the percentages of embryo crispants showing **(A)** severe developmental defects, **(B)** blood clots (black arrowhead) and **(C)** morphological cardiac defects for each gene targeted in the subset of hGWAS genes, as well as *oca2* crispants and mock-injected controls. Examples for each condition are displayed on the respective top left corner. Images were performed and observations were manually counted on a stereomicroscope under bright-field illumination. Genes yielding significant heart rate phenotypes are indicated in bold. Note the lack of an obvious correlation between heart rate, blood clot and morphological heart phenotypes.

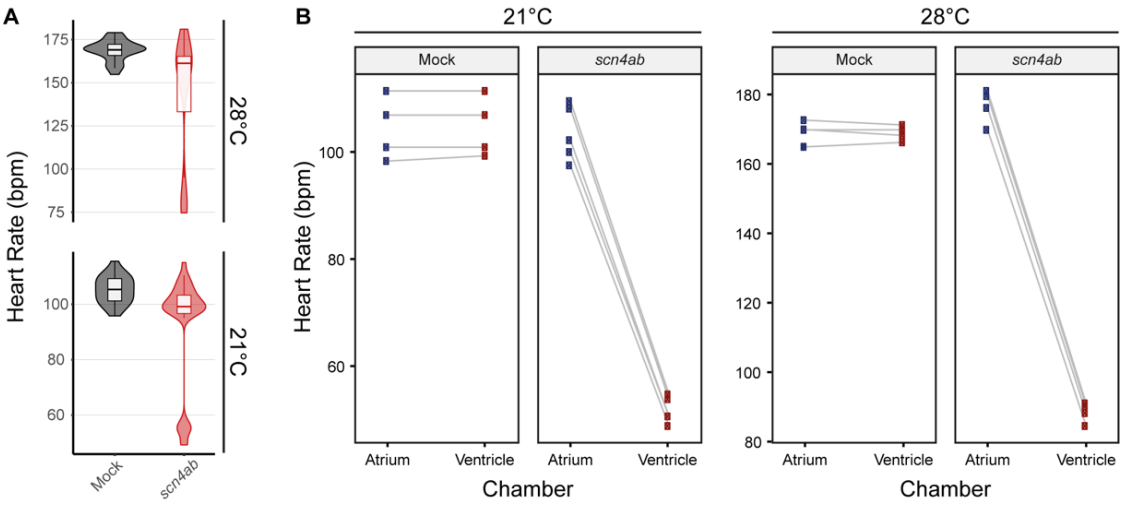


Figure 2.21 Medaka *scn4ab* crispants exhibit AV-block arrhythmia phenotype
A) Comparative heart rate analysis of GFP-injected (Mock; dark grey) and *scn4ab* crispant embryos from (Figure 2.16) reveals a population of *scn4ab* crispants displaying half the average heart rate. **B)** Paired plots reveal 2:1 AV-block arrhythmia phenotype via scoring of heart rates at both temperatures in individual chambers separately within the same embryos from **(A)** (atrium in blue, and ventricle in red). Figure adapted from [144].

crispants displayed AV-blocks of various severities; from mild (regular heart beats with occasional ventricle beat skipping), to moderate (consistent 2:1 atrial to ventricular contractions), to severe (3:1 or more atrial to ventricular contractions; up to 10:1 even 30:1 observed in some embryos).

2.2.6 Medaka *scn4ab* crispants phenocopy *scn4ab* mutant arrhythmia phenotypes

The specificity and relevance of phenotypes observed in F0 crispants in relation to a mutant line was then addressed by raising the *scn4ab* mutants to homozygosity. *scn4ab* was chosen due to its specific and easily scorable arrhythmia phenotype. The SLEDGE-hammer approach followed by standard mismatch cleavage assays such as T7EI was considered for cost-effective rapid genotyping, as to avoid the complete reliance on sequencing-based approaches. The T7EI assay, however, can only provide a qualitative assessment of presence of multiple alleles such as for crispants, heterozygous or homozygous mutants of different alleles (compound heterozygous mutants). Wild-type versus homozygous mutants, as well as heterozygous versus compound heterozygous mutants cannot be distinguished using the T7EI assay. To address this issue, I designed and implemented a Tri-primer PCR approach, using standard locus primers used for *scn4ab* and in addition a third primer (p3) binding at the targeted CRISPR/Cas9 cutting site (Figure 2.22A). Therefore, in addition to normal full locus amplification, p3-mediated amplification can only occur in wild-type alleles resulting in two observable amplified bands, while alleles with InDel mutations introduced at the cut site would prevent p3 binding and amplification, resulting in a single amplified band representing the full locus. Subjecting individual embryos to Tri-primer PCR as well as standard PCR followed by

T7EI approaches would result in specific combinations, thus the ability to distinguish all possible genotyping outcomes (Figure 2.22B). Interestingly, performing Tri-primer PCR on wild-type embryos resulted in an unexpected preference and amplification of only the shorter fragment (p3-locus) while the full locus fragment appears to be completely omitted, as for heterozygous and homozygous/compound heterozygous mutants, the expected amplified band distributions were observed (Figure 2.22C). This convenient finding enabled me to distinguish between wild-type, heterozygous and homozygous/compound heterozygous mutants using a single Tri-primer PCR reaction, thus further streamlining the genotyping procedure.

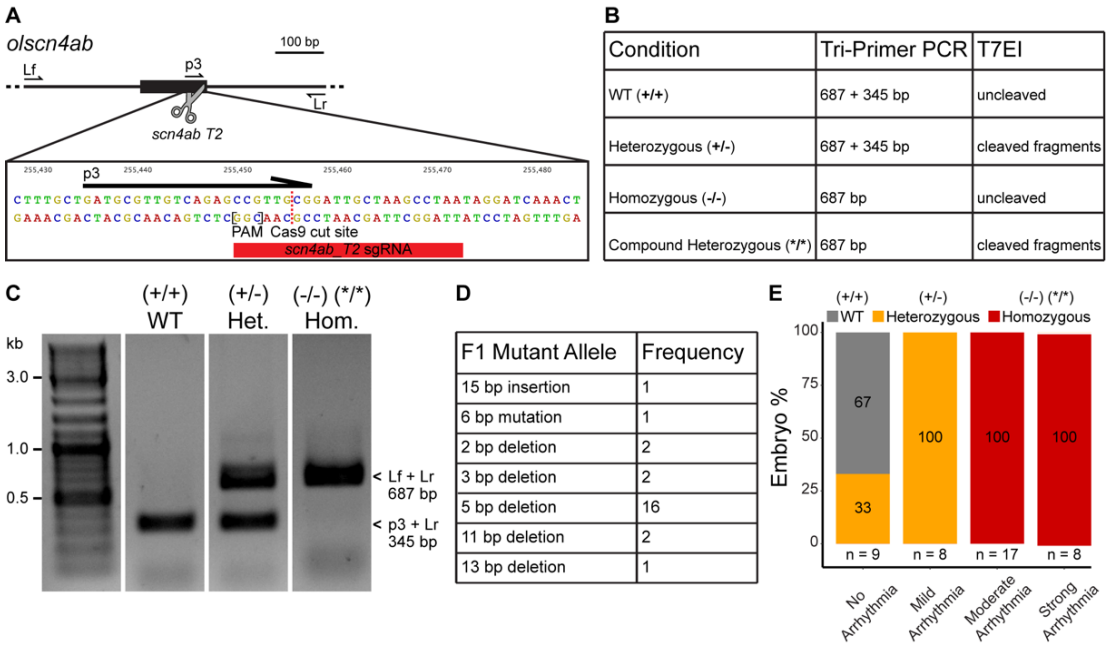


Figure 2.22 AV-block arrhythmia phenotype in medaka *scn4ab* crispants indistinguishable from homozygous *scn4ab* mutant

A) Schematic representation of the Tri-Primer PCR approach enabling the differentiation between the different genotypic statuses in an F1 and F2 mutant setting. The sgRNA *scn4ab_T2* (red rectangle) was designed to target the *scn4ab* locus in medaka (*olscn4ab*) via CRISPR/Cas9-mediated DSB (scissors) resulting in NHEJ-mediated InDel mutations at the Cas9 cut site (dotted red line). A third primer (p3) was designed as to bind around the Cas9 cut site only in a wild-type allele and not in a mutated allele. **B)** Theoretical scheme for precise genotype determination under various genetic conditions using both of Tri-Primer PCR and T7EI mismatch assays. Compound heterozygous corresponds to a homozygous mutant specimen carrying different mutations. **C)** Representative example of wild-type embryos (WT), heterozygous mutant (Het.) and homozygous/compound heterozygous mutant (Hom.) *scn4ab* embryos genotyped using the Tri-Primer PCR approach. Note: in wild-type embryos, strong preference for the amplification of the shorter fragment (primers p3/Lr) allows direct differentiation of wild-type and heterozygous mutants without reliance on T7EI mismatch assay. **D)** Mutant alleles of heterozygous *scn4ab* mutants (F1, n = 25) determined by sanger sequencing and TIDE analysis reveal a remarkably common mutation (5 bp deletion with 64% incidence rate). **E)** Proportion of embryo genotypes (wild-type, heterozygous and homozygous/compound heterozygous mutant; determined by Tri-Primer PCR) within phenotypically classified *scn4ab* F2 embryos (no, mild, moderate and strong arrhythmia) reflects the specificity of the arrhythmia phenotypes.

Having raised *scn4ab* crispants to adulthood, and determined founders carrying a mutation in the germline, crispants were outcrossed to wild-types of the Cab strain and the resulting F1 embryos were raised to adulthood and genotyped by sequencing to

identify the mutant allele for the F2 crossing strategy. Notably, 64% (16/25) of heterozygous mutant F1 fish from different F0 founders show a specific 5 base pair deletion despite the countless NHEJ-mediated InDel mutations possible (Figure 2.22D). Heterozygous F1 founder fish were then incrossed to generate F2 homozygous mutants, distinguishable from wild-type and heterozygous siblings by Tri-primer PCR (Figure 2.22C). The arrhythmia phenotype observed in homozygotes resembled that of F0 crispants, strengthening the power of F0 experiments as fast and efficient validation tool. Furthermore, *scn4ab* F2 embryos displayed varying strengths of the arrhythmia similar to the F0 population. Genotyping embryos from different phenotypic categories (no arrhythmia, mild, moderate and strong arrhythmia) by Tri-primer PCR confirmed the correlation between homozygous *scn4ab* mutants and the arrhythmia phenotype (Figure 2.22E). Only homozygous mutants exhibited moderate (consistent 2:1 AV-block; 17/17) or strong (≥3:1 AV-block; 8/8) arrhythmias, while mild arrhythmias (occasional 2:1 AV-block) were observed only in heterozygous mutants (8/8). Some heterozygous mutants and all wild-type siblings displayed no arrhythmia phenotypes. These results demonstrate the specificity of the observed F0 phenotype when comparing it to a stable mutant, and in turn further showcased the simplicity and relevance of performing rapid gene validations already in the F0 generation. Moreover, the high specificity of the F0 phenotypes demonstrated here paves the way towards performing drug screening using crispants, without the requirement for establishing and raising mutant lines.

3

Discussion

3.1 Fast and Efficient high-throughput genotyping using the SLEDGE-hammer approach

In the era of CRISPR gene editing, a continuous rise of high-throughput genetic screens can be observed [113–115]. A wide range of methods for subsequent genetic analysis are available, among commonly used approaches are mismatch cleavage assays (e.g. Surveyor [117–119] and T7EI assays [120]), HRMA [121], Sanger sequencing in combination with TIDE analysis [126] and next generation sequencing (NGS). However, gDNA extraction still represents a common caveat with respect to increased throughput of genotyping approaches. Advances in nucleic extraction methods have been made in different directions, particularly the binding of nucleic acids to cellulose-based surfaces (e.g. dipsticks) allowed the bypass of the tedious incubation steps by direct amplification off the membrane [137]. Although the dipsticks are incredibly useful for sample collections in remote areas or outside laboratory perimeters, the method is non-compatible with high-throughput approaches. In this study a workaround has been developed, initially exploited the nucleic acid binding properties of cellulose, to place cellulose filter discs inside standard pipette tips (Figure 2.1A), thus allowing parallel handling by the use of multichannel pipettes (Figure 2.6A).

Most genotyping methods involving fish embryos are invasive and require embryo sacrifice, by mechanical grinding using a pestle or mortar, due to the stiffness of the chorion (especially for medaka) which blocks access of the lysis buffer to the tissues for successful disintegration and release of nucleic acids. The mechanical grinding step so far was performed in individual samples. Therefore, with 96-well plates in perspective, the hammer was custom made to circumvent manual grinding of embryos (Figure 2.5), which would be extremely time consuming and resource intensive. Using the hammer, all embryos on a plate can be ground simultaneously with a single push. An alternative to the hammer would be to use magnetic beads for grinding [141], however this would require extra steps such as addition, incubation on shaking platform and removal of beads, a stock of beads and related instruments to process multiple plates. Continuous developments are

undergoing towards establishment of non-invasive and non-lethal genotyping protocols for fish embryos either using a rugged-surfaced chip [169] or direct incubation in microtiter plates containing DNA collection solution [170], however these methods require already hatched embryos (hatchlings) and are not compatible with unhatched embryos. An attempt using the filter-in-tips to non-invasively extract environmental DNA (eDNA) from hatched larvae was unsuccessful (data not shown). However, it was performed by just exposing the filter-in-tip to fish water with a hatchling swimming in it. It can be worth trying again eDNA extraction but using DNA collection solution as reported by Zhang et al. [170]. Nevertheless, establishing non-invasive genotyping methods is a necessity as it would largely facilitate the early selection of founders to generate stable lines.

The gDNA extraction procedure, similar to the cellulose dipsticks, takes about 30 seconds per sample (Figure 2.1B), thus already a substantial improvement (both in terms of time and manual handling) compared to our standard Blin-Stafford approach, which takes 2 hours. For the first step, tissue lysis, an SDS-based lysis buffer was employed. Other lysis solutions can be also used such as Alkaline lysis (NaOH-based), however they require later neutralization steps, and in some instances lacked consistency in the performance. For nucleic acid binding and elution, a 10-second binding contact of lysate with the tip resulted in the highest number of consecutive elutions possible with a single tip (Figure 2.2C), with six elutions possible compared to only two elutions using a 5-second binding period (data not shown). Longer binding contacts were not tested, as to avoid unnecessary prolongation of the protocol. Moreover, a 100 µl lysate is enough volume to use for at least 12 separate binding/elution attempts (Figure 2.2D), thus in a conservative estimate of 12 attempts with four successful elutions per attempt, a single medaka embryo lysate can be used for at least 48 reactions, enough for amplification of multiple loci for CRISPR-mediated multiplex targeting validations.

It was later surprisingly apparent that the retention of the lysate to the tips surface plays a major, but not necessarily definitive, role in the nucleic acid extraction process. The standard tips work as reliably as the filter-in-tips, significant differences between the two tips were not observed (Figure 2.4). In a few cases (e.g., low retention tips), the addition of the cellulose filter resulted in an improved yield (Figure 2.4C). Reliability of using the standard tips also extends to downstream Sanger sequencing and TIDE analysis (i.e., identification of *scn4ab* mutant alleles, Figure 2.22D). Thus, the pipette tip alone is sufficient enough for nucleic acid binding for purification and amplification purposes, without affecting the quality for downstream purposes. Although the retention capacity of standard, filter-less tips (i.e., number of consecutive elutions possible) was not assessed, it can be foreseen that the filter could act as a nucleic acid reservoir, allowing for multiple successive elutions with a single filter-in-tip.

The key advantage of the tip approach is the compatibility with multi-channel pipettes, which allows processing of multiple samples simultaneously, thus further cutting down processing and handling times. In a previous genotyping setting, unless a pipetting robot was available, embryos on a microtiter plate would be transferred from the plates into

separate tubes to be individually ground and lysed with a pestle. Each individual sample would then go through nucleic acid extraction via the Blin-Stafford approach, with various pipetting, incubating, precipitating and centrifuging steps. Although this approach does produce a very high yield of gDNA, it can take a whole day-long to process an entire plate, and for many applications where the sample is analysed only once or a few times (e.g., validation of CRISPR targeting), such high gDNA yield is not always necessary. Conversely, using tips mounted on a multi-channel pipette, processing times can be drastically sped up, depending on the capacity of the multi-channel pipette. From 6 minutes of processing time for a 96-well plate using an 8-channel pipette (Figure 2.6A), down to only 30 seconds using a 96-channel pipette. Moreover, due to the simplicity of our approach (only a standard tip is required), it would be compatible with automated pipetting robots already in existence, which go up to 1536 samples on a single plate, thus truly enabling high-throughput genotyping.

The SLEDGE-hammer approach was established and extensively used in this thesis combined with the use of T7EI mismatch cleavage assay [120] or Sanger sequencing followed by TIDE analysis [126] as downstream allele detection analysis. Both approaches, although widely used in laboratories, still have limitations regarding accuracy and sensitivity of mutant allele detection. T7EI assays are commonly used due to their low cost and simplicity to produce a quick qualitative assessment, or semi-quantitative when using a defined gDNA input, of whether an amplified locus has been largely modified (e.g., InDel mutations). However, they lack the sensitivity to detect single nucleotide changes and do not offer a more detailed insight into the nature of the InDel editing events. On the other hand, Sanger sequencing followed by TIDE analysis provide a good and accurate overview on the nature and frequency of InDel mutations in a sample [171], but they depend on the input amplicon's quality and therefore require more processing time post-amplification and have higher costs. Additionally, many other approaches can be used for multiplex genotyping such as real-time PCR (qPCR) using conditional fluorescent probes [172,173] and NGS. Thus, depending on the nature, aim and size of the gene editing application (e.g. InDel versus Base-Editing), a suitable genotyping strategy needs to be considered.

Although Zou et al. were able to extract RNA from plant tissues and amplify it via RT-PCR [137]. In the current study, an attempt at extracting RNA with the SLEDGE-hammer approach was made which turned unsuccessful (data not shown). Furthermore, the applicability of the SLEDGE-hammer protocol with other downstream allele detection analyses, importantly with qPCR and NGS applications, was not yet tested. Therefore, while the SLEDGE-hammer protocol is reliably applicable with mismatch cleavage assays (i.e., T7EI) as well as with Sanger sequencing and TIDE analysis, it would be worth investing time in optimizing the protocol to fully expand its range of applications.

Nevertheless, the SLEDGE-hammer approach is efficient and reproducible, its sensitivity matches the lab standard, high yield Blin-Stafford approach, demonstrated by the detection of the same allele distribution profiles when genotyping *oca2* crispants (Figure

2.2B). The broad applicability of the approach across kingdoms (Figure 2.3), in addition to the compatibility with upscaling (Figure 2.6 and Figure 2.7) and automated systems (e.g., pipetting robots), makes it extremely useful for high-throughput screens as well as for stock maintenance in laboratories or stock centres, where genotyping is a daily routine.

3.2 Highly efficient CRISPR/Cas9 mutagenesis boosts F0 screening applications

The emergence of CRISPR/Cas9 and its grand promise for targeted gene editing has revolutionised the biological and medical fields. It has gone through continuous advancements over the years to significantly boost its efficiency [100] and to increase its versatility with respect to PAM recognition [174]. The heiCas9 was developed by adding an optimized artificial nuclear localization signal to a standard Cas9, which facilitates the immediate localization of Cas9 to the nucleus and thus greatly increased chances of rapid gene targeting already at the early stages after fertilization [101]. To assess whether the improved targeting efficiency of the heiCas9 would permit relevant rapid gene knock-out in injected generation medaka, an artificially inserted *GFP* gene in a transgenic cardiac reporter line as well as the endogenous eye pigmentation gene *oca2* [145] were chosen as initial targets (Figure 2.8 and Figure 2.9). That is, due to their easily observable and scorable phenotypes (loss of green fluorescence and loss of eye pigment, respectively). The abolished GFP fluorescence when injecting in a 1-cell stage embryo, with the complete loss of the wild-type allele clearly demonstrated the rapid and high editing efficacy of the heiCas9, so efficient that only when injecting at the 4-cell stage, was a mosaic GFP expression observed (Figure 2.8). This was further substantiated by the prominent loss of eye and skin pigmentation in *oca2* crispant embryos observed in this study (Figure 2.9), which requires successful bi-allelic mutagenesis [107]. The impressive increased efficacy of the heiCas9 over other improved Cas9 variants (up to 27-fold) was further reported and showcased in [101]. Thus, the *hei*-tag has immensely improved *in vivo* targeting efficiency of the Cas9 and consequently provided the foundation for the relevant F0 mutagenesis approach developed and implemented in this study.

3.3 Different factors impact on Cas9-mediated efficacy

Many factors influence the successful and efficient Cas9-mediated mutagenesis, PAM recognition represents one of the most influential Cas9 properties. The reported ability of Cas9 to recognise NAG PAMs [98,99], even if to a much lower extent than NGG PAMs, increased flexibility with regarding to the choice of sgRNA design for targeting.

However, in my hands none of the NAG PAMed-sgRNAs yielded Cas9-mediated InDel mutations *in vivo* in medaka (data not shown). The sole NGG PAM recognition by the Cas9 still poses as a large limitation and strongly restricts the choice of the locus to be targeted. Recent developments towards relaxing Cas9 PAM requirements have led to structure-guided engineered NGN PAMed (SpG) followed by a near-PAMless variant (SpRY), enabling unconstrained targeting [174]. As much as abolishing PAM requirements increases the Cas9 targeting range, it similarly increases the chances of off-targeting. Thus, caution must be taken while designing sgRNAs for PAMless Cas9 variants to avoid misleading off-target effects.

Besides PAM specifications, other factors such as chromatin accessibility [146,147,149] and gRNA secondary structures [148,149] may influence Cas9-mediated genome editing efficacy. This was indeed observed throughout this study, where a number of sgRNAs did not lead to Cas9-mediated DSBs despite having an NGG PAM (data not shown). Moreover, significant variations in medaka *oca2* targeting efficiencies were revealed when using two different sgRNAs (*oca2_T1* and *oca2_T3*) (Figure 2.9). Even though both sgRNAs target the same exon at roughly 18 bases apart (Figure 2.9A), *oca2_T3* yielded a considerably larger fraction of near-complete pigment-less crispants compared to *oca2_T1* (Figure 2.9B). Peculiarly, predictions of both off-targets using CCTop [102] as well as sgRNA efficacy using CRISPRater [175] leaned more preferably towards the less effective *oca2_T1* (9 predicted off-targets; Medium CRISPRater score 0.71) rather than the more effective *oca2_T3* (17 predicted off-targets; Medium CRISPRater score 0.68). Notably, targeting *oca2* using both sgRNAs combined yielded pigment-less embryos at lower efficiencies compared to when using *oca2_T3* alone (data not shown). Although it is widely believed, that using multiple sgRNAs targeting a gene of interest would increase chances of a gene knock-out. The observations made in this study imply the possible occurrence of competitive binding resulting from introducing multiple sgRNAs, therefore leading to an overall decrease of effectiveness across all target sites. Despite that both guides have an NGG PAM, the only notable difference between them is the GGG PAM present the highly efficient *oca2_T3* compared to the AGG PAM in the less efficient *oca2_T1* (Figure 2.9A). If such a minor difference within the NGG PAM can impact so much on the activity remains questionable. Therefore, a systematic comparison between the two *oca2* sgRNAs when used alone or in combination could reveal interesting insights regarding the impact of the number of sgRNAs employed, as well as the sequence surrounding their respective PAMs on editing effectiveness in an *in vivo* situation.

3.4 In house sgRNAs are more effective than synthetic cr/tracr RNAs

Due to the oncoming high demands of guide RNAs for gene validations, the use of commercially available crRNAs and tracrRNAs seemed an attractive opportunity to circumvent the days-long process of synthesizing our own sgRNAs in house. However, the large difference in efficacies when targeting *oca2* in medaka evident by the compared

levels of pigmentation loss (Figure 2.10), clearly omitted such possibility, at least for now. According to the manufacturer's instructions, crRNA and tracrRNA counterparts are to be duplexed by incubation prior to ribonucleoprotein (RNP) complex formation with heiCas9. However, using crRNA and tracrRNA as singletons without prior incubation resulted in a much stronger targeting effect (Figure 2.10). High variations of phenotypes were observed, some embryos displayed extensive pigmentation loss comparable to our sgRNA employed but at much lower ratios. Additionally, singleton crRNA and tracrRNA resulted in an increased lethality rate compared to duplexed cr/tracrRNA or in house sgRNAs (data not shown). Thus, our in-house sgRNAs have proven themselves to be indispensably efficient at a much lower costs (~15 euros/sgRNA excluding work-hours) compared to the cr/tracr RNA counterparts (~90 euros/crRNA plus more than 100 euros for a tracrRNA stock). Synthetic sgRNAs are also available in the market, however, they were not considered owing to their very high costs (~200 euros/sgRNA).

In sum, within my initial trial, in house sgRNAs proved to be more efficient and with lower lethality rates than the synthetic cr/tracrRNAs (Figure 2.10). Still, careful considerations regarding a proper and full assessment of the synthetic gRNA alternatives need to be taken and when can they become beneficial despite their higher costs. In this study, the cr/tracrRNAs were tested with a modified Cas9 (heiCas9) in mRNA form and not a protein form. Although the cr/tracrRNAs are advertised as chemically modified to increase stability and protection against cellular nucleases, there is another (~150% more expensive) available "crRNA XT" variant with added chemical modifications to further increase its stability and protection, marketed specifically for *Cas9* mRNA applications. The very low efficiency of the duplexed or the high variability of the singleton cr/tracrRNAs employed may therefore arise from different sources. Compatibility issues may arise with one or multiple components of the *hei* tag. Moreover, the use of heiCas9 in mRNA and not in protein form may have led to delayed RNP complex formation, and therefore an extensive exposure of cr/tracrRNAs to nucleases resulting in their degradation. Considering a future direction with abundant gene and SNP validations, a continuous supply of sgRNAs will be required, which will solely take up extensive man-power. Therefore, a re-evaluation of cr/tracrRNAs would be valuable, considering the different suppliers as well as the various variants and their compatibility with our *heiCas9* mRNA.

3.5 F0 medaka crispants phenocopy relevant mutant phenotypes

F0 mutagenesis approaches are still viewed with scepticism due to the fear from misleading and unspecific phenotypic interpretations. Nevertheless, improvements in CRISPR/Cas9 on-targeting efficiency with negligible off-target effects [176] allowed F0 approaches to gain significant momentum with respect to large throughput functional validation of candidate genes [105,106,177]. Further exemplified by the validation of

trabeculation-associated genes in medaka [178]. In this study, medaka *nkx2-5* crispants effectively phenocopied zebrafish *nkx2-5* mutants with respects to cardiac morphological phenotypes (Figure 2.13). They exhibited similarly enlarged heart chambers [151], which also notably resembled heart phenotypes observed in heterozygous mutant *Nkx2-5* mice [179]. Moreover, other than the observable loss of GFP expression in the cardiac dual *cmlc2::eGFP cmlc2::H2A-mCherry* reporter line (Figure 2.8), and the prominent loss of eye and body pigments in *oca2* crispants (Figure 2.9), medaka *scn4ab* crispants exhibited highly penetrant AV-block arrhythmias exceeding 90% of injected embryos (Figure 2.21). The arrhythmia phenotypes observed in medaka *scn4ab* crispants were indistinguishable from the arrhythmia phenotypes of *scn4ab* homozygous mutant fish (Figure 2.22). Thus, the results reported in this study clearly demonstrate the reliability of the F0 targeting approach for large-scale screening purposes, and how the mosaicism in crispants allows investigation of essential genes that would be otherwise lethal in a homozygous state. Additionally, the results reveal the potential to perform discovery drug or compound screening directly on crispant embryos with a highly penetrant phenotype (i.e., arrhythmias).

3.6 F0 CRISPR screen is a reliable platform for rapid identification of clinically relevant genes

For the discovery of potential genetic determinants of CVDs, human genome wide associations studies have been performed extensively over the past two decades [12,32–40,180–182]. Although human GWAS typically yield tens of thousands of CVD-associated polymorphisms in coding and non-coding regions, these are all correlative bioinformatic associations in need of experimental validation. The richness of the GWAS output provides ample opportunities to identify novel players with cardiac relevance and validate potential genetic markers. Nevertheless, polymorphisms in genes with known function and links to cardiac development and function are usually addressed further. Polymorphisms in uncharacterized genes, however, are often untouched despite their potential clinical relevance. There is a lack of rapid and reliable experimental pipelines for validation of a large number of genes, and generating mutant lines to address each candidate gene is time-consuming and unfeasible. In this study, this was addressed by developing an F0 functional gene validation pipeline in medaka, combining highly efficient CRISPR/Cas9-mediated mutagenesis with the established high-throughput heart rate assay (Figure 2.11).

The F0 validation pipeline proved reliable in the medaka *nkx2-5* crispant embryos, which not only consistently phenocopied previously discussed morphological, but also functional phenotypes such as an elevated heart rate phenotype [183] (Figure 2.13D-F). The specificity of the *nkx2-5* phenotypes observed could be verified by lack of phenotypes in

mock-injected or in *oca2* crispants (Figure 2.12B and Figure 2.13), which not only proved the robustness of the medaka embryos to extensive manual handling, but also confirmed the *oca2* as a bona fide negative control gene. Remarkably, the validation of the randomly selected genes (Table 2.1) further demonstrated the accuracy of our approach by detecting heart rate phenotypes only in heart-relevant genes with either experimental (*cdc42*) or GWAS-related (*ogdh*) connection to heart phenotypes (Figure 2.14 and Figure 2.15). Additionally, by applying our assay in a more targeted manner across the hGWAS set of genes (Table 2.2), 16 genes were identified, of which 11 genes had little to no functional connection to the heart so far and were experimentally validated here for the first time (Figure 2.16-Figure 2.18). The new light that my findings shed onto these genes is substantiated by the emerging studies on two of the uncharacterized genes (*CNOT1* [164,165], and *CMYA5* [184]) and their roles in heart structure and function.

Interestingly, an evident overall inclination towards significantly elevated heart rate phenotypes is observable for 85% of the genes, while only 15% showed a significant depression of heart rates compared to control siblings (Figure 2.17). The increase in heart rate could result as a direct effect of cells trying to compensate for non/less-functional cardiac cells or for the loss of efficient contractility, in order to maintain embryo survival. Depending on the extent of gene editing variations in the degree of manifestation can arise. Nevertheless, the inherent low inter-individual variability in the highly isogenic medaka [61] represented a key property for the reliable detection of subtle but relevant heart rate phenotypes in a genetically mosaic setting.

The GRASP database [156] proved to be an extremely useful resource for the search and selection of associations to various phenotypes. Here in this study, human coding SNPs with heart-related phenotypes were selected based on the significance of the association and the associated gene, as to focus on novel uncharacterized genes. The significance of the selected associations ranged between $3.4x10^{-2}$ and $3.6x10^{-26}$. Priority was given to associations with highest significance. To determine the possible presence of significance bias, some genes were chosen even though having a low significance value. This was done as association strength can be misleading due to the inherently large genetic variance in humans and their uncontrollable environment. Notably, there was no bias in the selection of candidates according to their phenotypic associations, meaning that the subcategorization of selected associations into “non-heart rate” or “heart-rate” associated-phenotypes, was only done post validation and analysis. Thus, embryonic medaka phenotypes matched the class of adult human hGWAS effects despite the evolutionary distance from fish to humans, which ultimately attests the clinical relevance of our approach at filtering down GWAS-candidates, and consequently the identification of novel genetic markers with promising diagnostic and predictive potential for cardiovascular diseases in humans.

3.7 Redirecting the spotlight onto novel uncharacterized genes

The F0 CRISPR screen developed and employed in this study provides a starting point towards the discovery of novel genetic CVD markers. A total of 11 genes, aside from the positive control genes (*TTN, NACA, CASQ2, KCNH2* and *SCN5A*), which clearly responded in the assay, have little to no previous evidence to affect cardiac function and were identified in this study to cause significant heart rate phenotypes in medaka, which placed them under the spotlight for further deeper investigations (Figure 2.18A). Some have been previously implied in a broad range of biological functions such as cellular transport (*ABCB1* [185], and *TRAPPC12* [167]), anion transporter (*SLC17A3* [186]), angiogenesis (*MINAR1* [187], and *PADI2* [188]), non-sense mediated RNA decay (*SMG6* [189,190]) and modulation of signalling pathways (*VEPH1* [191], *TRAPPC12* [192], and *GIGYF1* [193]). Many of the genes (*UFSP1*, *CNOT1*, *SSPO*, *CCDC141*, *HOMEZ*, *CMYA5*, and *MYRF*), however, have not been characterized yet and their biological functions remain unknown. Thus, follow up investigations of the identified genes through generation of stable mutant lines are required for a deeper understanding of their biological functions. Novel insights can be gathered on the impact of these novel genes on cardiac gene regulatory networks, cardiac histology and development, as well as cardiac function under physiological state, environmental (i.e., external temperature) or physical stress (i.e., exercise).

Medaka *trappc12* crispants exhibited significant heart rate phenotypes, with a 3-5% increased mean heart rate compared to mock siblings at both measured temperatures (Figure 2.16 and Figure 2.17). Furthermore, 26% of *trappc12* crispants displayed morphological heart looping phenotypes (Figure 2.19). TRAPPC12 (Trafficking Protein Particle Complex Subunit 12) is a part of the multi-subunit complex TRAPPIII whose functions are not fully understood, it is proposed to be involved in endoplasmic reticulum/Golgi transport [194,195]. However, there is increasing evidence that each subunit of TRAPPIII might be involved in different non-overlapping function, shown by the connection of certain individual subunits to various diseases, collectively named as TRAPPopathies [196]. TRAPPC12 has been implied in previous reports to play a role in mitosis [197], and ciliogenesis [192], but no connection to cardiac development or function has been previously reported. It appears to be ubiquitously lowly expressed in medaka [168], so low that it was undetectable in an *in situ* hybridization trial attempted in this study (data not shown). Interestingly, primary cilia and ciliogenesis have been reported to play important roles in heart development as well as in the pathogenesis of congenital heart diseases [198–201]. Whether the observed impact of *trappc12* loss of function in heart structure and performance is related to these mechanisms or is yet another unidentified role of *trappc12* is still to be unravelled.

Medaka *myrf* crispants also displayed a significant elevation (~7%) of mean heart rates at both temperatures compared to mock siblings (Figure 2.16 and Figure 2.17), with

additional morphological defects such as heart looping defects and tubular chambers (data not shown). MYRF (myelin regulatory factor) is a membrane-bound transcription factor, evolutionarily conserved throughout the animal kingdom and essential for myelin development in vertebrates [202,203]. The very interesting aspect about *MYRF* is its GWAS association to diverse human disorders, including congenital heart defects [204,205], despite its unknown role outside of myelin formation. There is prior evidence linking the autonomous nervous system to cardiovascular function [206,207] and cardiovascular diseases [208–210]. Thus, the cardiac phenotypes observed in this study might reflect an aberrant cardiac innervation from faulty myelination due to loss of MYRF. Taken together, *myrf* and *trappc12* both yielded prominent heart rate as well as cardiac morphological phenotypes, and thus represent the most promising candidates identified in this screen. Although the characterizations could not be completed in the time frame of this study, further investigations are currently ongoing in homozygous settings.

3.8 The importance of other cardiac phenotypes beyond the heart rate

The heart rate is an important indicator for overall cardiac performance and a clinically relevant phenotype. It is widely recognised as a risk predictor for the onset of cardiovascular diseases [7,10,12,40], but also for their prevention through early modulation [23,211]. Consequently, in addition to its ease of quantification and interpretation in high-throughput, a major attention has been put on the heart rate as the phenotypic readout throughout this study. Nevertheless, phenotypes beyond the heart rate such as global developmental phenotypes in addition to CVD-relevant cardiac malformations and blood clots were observed. For some of the targeted genes this was assessed in a little more detail (Figure 2.20). A clear observation was the higher incidence of globally affected embryos when targeting genes with broad and important roles in embryonic development (e.g., *smg6* [189,190]*, abcb4* [212]*, nkx2-5* [213]). The presence of such developmentally delayed embryos posed a risk of skewing the heart rate comparisons, which may be potentially misleading. To circumvent this problem, a developmental focusing filter, excluding embryos looking younger than stage 28, is applied in the analysis phase. However, developmental focusing, although deemed important, in only few cases (e.g., *minar1* and *smg6*) significantly altered the outcome of the analysis (Figure 2.17), but this would largely depend on the percentage of randomly selected embryos exhibiting such strong delays.

In this study, the assessment of crispants exhibiting blood clots and morphological cardiac phenotypes, such as enlarged/diminished heart chambers and heart looping defects, was performed manually by counting the incidences. This was performed manually due to the lack of an available detection algorithms at the time, which on the one hand is time consuming, and on the other hand also can introduce a selection bias due to the personal assessment of the phenotype by the observer. However, it has revealed a lack of an

apparent connection between morphological and functional heart phenotypes. This was evident by a number of genes with a high incidence rate of blood clots (e.g. *rgs3a*, *scmh1*, *clcnk*) or cardiac morphological defects (e.g. *nubp2*, *rgs3a*) that did not exhibit significant heart rate phenotypes, and vice versa (Figure 2.20). These observations therefore emphasize the importance of scoring other CVD-relevant phenotypes beyond the heart rate.

Phenotypic scoring of various cardiac parameters such as chambers size, volume and shape during systole/diastole, stroke volume, ejection fraction and blood flow parameters (e.g., speed and viscosity) would be possible nowadays with the continuous development of automated detection algorithms [214]. Automated detection and quantification of cardiac parameters (including heart rate) in bright field illumination is still a significant challenge that can be substantially eased by the use of fluorescent cardiac reporter line, due to the diminished background noise influencing segmentation parameters. Performing the F0 CRISPR screen in a reliable transgenic cardiac reporter line (i.e., *cmlc2::eGFP*) would have been ideal, however, the line was not available in sufficient size to continuously deliver the large number of embryos required to perform such screens. With the ongoing expansion of this transgenic line in the lab, or alternatively the generation of additional reporter lines in sufficient numbers, via highly efficient CRISPR/Cas9-mediated endogenous protein tagging [109,142], high-throughput F0 CRISPR screens in reporter lines are within reach.

3.9 The heart rate assay: a highly versatile tool for quantitative phenotyping

The established heart rate assay was developed using medaka unhatched embryos, which are naturally restrained by their chorions (Figure 1.2). Thus, enabling heart rate phenotyping without the need for orientation or sedation [72]. Overall, the heart rate assay delivered great performance and quantitative extraction of clinically relevant heart rate phenotypes without the need of sophisticated setups for image acquisitions or specialized personnel to use the *HeartBeat* software.

A major improvement to look forward to is the full automation of heart detection and heart rate quantification. The current *HeartBeat* software reliably does so in a semi-automated manner, which still requires manual control checks by the user with occasional manual adjustment of the parameters in case of detection errors, which takes an experienced user roughly 40 minutes per full 96-well plate. Background fluctuations of movements such as embryo twitching, fin movements or blood flow may result in incorrect heart detection and segmentation, and therefore leading to potentially misleading results if uncorrected by the user. Current developments, employing artificial intelligence (AI) based machine learning algorithms, are ongoing to increase the accuracy

and detection of the software while transitioning to full automation thus significantly saving the user's screen time to be invested elsewhere.

The *HeartBeat* software could reliably detect the heart region and accurately quantify the heart rate, an additional unused feature of this version is its ability to calculate beat to beat variability, which is another important marker for cardiovascular diseases [215,216]. However, technical limitations with respect to the current imaging machine sampling rates of 13 fps in addition to imaging "hiccups" due to technical processing power-mediated delays do not provide sufficient resolution to reliably quantify beat-to-beat variability. Furthermore, the software was developed and optimized for heart rate extraction and quantification from image sequences taken under bight-field illumination. This eliminated the reliance on reporter lines, of which sufficient numbers may not always be available for high-throughput applications. Still, the use of transgenic reporter lines, *cmlc2::eGFP* for instance, would further improve and facilitate feature detection due to the almost complete removal of unspecific background signal.

In sum, the assay is highly versatile with endless potential. Higher sampling rates of at least 30 fps would enable acquisition of image sequences with high quality and temporal resolution [72,217], and the use of transgenic reporter lines with labelled blood or cardiac cells would minimize background noise. Such setup can then be combined with an AI-improved or "smarter" feature detection algorithm to enable automated detection and quantification of additional cardiac and blood flow parameters. The ability to detect and quantify not only heart rate but also heart rate variability, chamber size, systolic and diastolic volumes, stroke volume and ejection fraction will expand this assay to a broadly applicable, multifunctional high-throughput quantitative phenotyping assay.

3.10 Perspectives of F0 CRISPR screens for future characterization of candidate genes

Throughout this study, the F0 validation assay has undergone various iterations, which greatly improved many of its aspects. However, there are still improvements to be made and considerations to be taken. It became apparent that the injection volume into the cell and embryo handling can differ significantly between investigators, leading to variation in the results. Some investigators observed a heart rate phenotype in negative control *oca2* crispants, but also in mock-injected embryos when compared to un-injected siblings (data not shown). Thus, for this study it was fundamental that only one investigator (myself) performed the injections to ensure consistency of the results. This technical variation can easily be overcome in the future with injection practice and routine check-up of the controls (un-injected versus mock-injected embryos or mock-injected versus *oca2*

crispants or other negative control). Ideally, each gene should be targeted by two investigators in order to ensure the reproducibility of the observed phenotype.

Although medaka of the Cab strain are highly isogenic and showed low inter-individual variability, still some variability in the heart rate results was observed, sometimes even significant, between different experimental plates run on different days. While medaka fish are raised in a tightly controlled environment, still day-to-day differences influencing on the egg quality can be expected such as embryo handling, feeding, the cumulative aging of fish, or even the different generation of fish employed. Accordingly, it was determined that including mock-injected siblings in each experimental plate as controls was the best strategy to account for both embryo batch differences and the manual handling throughout the procedure. Moreover, to exclude any influence of the fish generation or the *heiCas9* mRNA on the heart rate, a control run comparing heart rates of *oca2* crispants to mock-injected siblings was performed with every change of fish generation or *heiCas9* mRNA batch.

Medaka heart rates have shown to be highly adaptable to environmental stimuli, such as temperature. Therefore, as an added dimension to the F0 validation assay in the form of gene and environment interaction, crispant heart rates were scored at two different temperatures (high 28°C and low 21°C). This was done to evaluate such adaptability, and potentially uncover phenotypes that would not be observed when imaging at a single set temperature. It is noteworthy to mention, that the phenotypes observed in crispants manifest in a genetically mosaic setting, and therefore, depending on the mutated gene and the nature of the NHEJ-mediated InDel mutations, the penetrance and strength of phenotypes may vary. However, as anticipated, the different temperatures increased the detection range of heart rate phenotypes in the assay. Some candidates yielded significant heart rate phenotypes at both measured temperatures, which could be considered as an indicator for relatively strong penetrance. However, the majority of identified genes exhibited significant heart rate phenotypes at only one of the temperatures, with no clear predisposition to one temperature over another (Figure 2.18). The specificity of the temperature causing heart rate phenotypes and the mechanism behind it remains eluded and could not be addressed within the timeframe of this study.

A second technical replicate was performed for most genes with a heart rate phenotype (data not shown). Overall, significant heart rate phenotypes were consistently detected in both replicates, except for some genes (*abcb4*, *veph1*, *ufsp1* and *minar1*) when targeted more than once resulted in significant heart rate phenotypes on one attempt but not in the other. This occurred to a lesser extent and these genes were then flagged as “border-line” genes. Interestingly, the significant heart rate phenotypes were detected in some cases under different temperature conditions in the two different replicates. For *casq2* and *padi2* on their first targeting attempt yielded significant heart rate phenotypes at both 21 and 28°C for *casq2* and only at 28°C for *padi2*, while on the second replicate attempt *casq2* yielded a phenotype only at 21 but not at 28°C, and *padi2* yielded a phenotype at both temperatures. *sspo* on the other hand, even though on both attempts yielded a significant

phenotype only at 28°C, *sspo* crispants interestingly exhibited once an overall heart rate elevation and another overall heart rate depression. Thus, the different temperatures applied to the assay should be, for now, considered a mean to increase the overall detection range of heart rate phenotypes. Nevertheless, a deeper investigation into the role of external temperature on the heart function of crispants is required to enrich our understanding of gene-environment interactions. Phenotyping crispants at additional temperatures could give us insights into the role of candidate genes in heart adaptability. Alternatively, raising the fish at different temperatures before phenotyping would provide new clues regarding the external environment in which phenotypes are more or less likely to occur.

Two or more sgRNAs targeting different regions of candidate genes are usually employed for CRISPR-mediated gene knock-out studies [106,107,218,219] aiming to achieve full gene knock outs with higher efficacy. Alternatively, a rather less aggressive approach was employed in this study to validate the GWAS candidate genes. Only one sgRNA was used to introduce random InDel mutations at a single exon in the medaka orthologous gene. The concept behind this approach was to minimize excessive global or lethal phenotypes that may result from the extensive disruption of important or broadly acting genes, which would have severely impaired the heart rate analysis. This approach turned out beneficial, as by minimally disrupting gene functions in mosaic settings severely dampened strong global phenotypes, which in turn aided in the assessment of cardiac-specific phenotypes in otherwise normally looking embryos.

Nevertheless, all genes yielding heart rate phenotypes in this screen represent an opportunity for the identification of novel genetic markers relevant for CVD diagnosis and prediction. For the border line genes, a repeated validation replicate would be required to include/exclude them as potential candidates. For the already known cardiac genes, the extensive background knowledge on their structure and function readies them for a deeper level of actual SNP validations using highly efficient base editors [220,221]. This can also be applied to validate CVD-associated SNPs in novel candidate genes, in which deleterious mutations such as protein truncation or missense variants are predicted. Indeed, this has been showcased by [222], where highly efficient base editors were used to mimic pathological SNPs in F0, in doing so they not only replicated clinically relevant phenotypes in well-studied genes, but also validated SNPs in a subset of novel candidate congenital heart disease-associated genes. However, owing to the substantial number of coding CVD-associated SNPs, precise SNP-validations without any prior knowledge of the relevance of the corresponding gene to the heart would be quite time and resource consuming. There is a large imbalance between the high output of predicted and associated SNPs/genes and the low level of understanding of the biological functions of the majority of identified genes. Mainly caused by the increasing affordability of whole genome sequencing techniques and the lack of subsequent rapid functional validation assays. Therefore, functional gene validations are crucial to filter down the excessive GWAS candidates, enabling to prioritize the most promising and relevant genes for follow up investigations. Such investigations of poorly or uncharacterized genes can be

performed through a gradual but brute disruption of individual genes by employing multiple sgRNAs targeting various exons, enhancing the chances of obtaining noticeable phenotypes. Afterwards, a comprehensive assessment of the cardiac phenotypes obtained can be performed in the injected generation (e.g., heart rate, as well as other cardiac parameters as previously discussed), with promising candidate genes knock outs raised to homozygosity for more precise characterization.

3.11 A new era: multiplex CRISPR targeting for a polygenic assessment of candidate genes

Although many heart-relevant genes have been previously identified to directly cause heart development or functional defects in random mutagenesis screens [30,31], the incidences of mendelian inheritance of single genes causing CVDs are largely uncommon [26,27]. Instead, increasing evidence, largely substantiated by human CVD-relevant GWAS, suggests a non-mendelian polygenic nature of CVDs. Complex CVDs could be caused by a large number of genes with small effects acting in a vast intricate network. Indeed, the number of genes might be too large and with too small effect sizes to be studied in an individual setting. Therefore, a rather more complicated approach, would be to use multiplex CRISPR-mediated targeting for investigation and validation of multiple associated genes (or SNPs) in a combinatorial fashion [105,177,223]. Raising multigenic crispants followed by incrossing yields embryos with unique combinatorial mutations at the desired loci. Phenotyping and consequent genotyping of such mutants would offer interesting and novel insights into the coherent influence of the different genes onto cardiac function. However, preliminary tests are required in order to assess the maximum possible number of genes co-targeted that would result in a substantial portion (ideally >50%) of viable embryos to be able to carry the incrosses, while also maintaining a high degree of germline-transmission of the mutations [223]. Another important aspect to consider when applying this multiplex approach is the genotyping strategy of these multigenic crispant (F0) or mutant (≥F1) embryos. Such multiplex approaches are mainly coupled with NGS sequencing of targeted loci, allowing high-throughput sample processing. However, they are time-consuming and not cost efficient for low number of targets (less than 20) [116]. Fluorescent qPCR can be employed as an alternative. Remarkably, a novel color mixing strategy using multiplexed probes modified with fluorophores and quenchers was developed, which exponentially increased the number of variants that can be detected using qPCR approaches [224].

Although substantial efforts are being invested in the validation of disease-associated coding SNPs, these coding SNPs represent only a minor fraction of the total output of human GWAS, while the majority of disease-associated SNPs (>90%) lie within non-coding regions of our genome [41]. These regions contain key gene regulatory elements

such as promoters, insulators and enhancers that are yet to be identified and their target genes assigned [42,43,225]. Thus, another exciting exploratory phase lies ahead, validating non-coding SNPs using highly efficient base editors with a primary focus on regions conserved between humans and medaka. In turn revealing novel regulatory elements involved in CVD development and pathogenesis.

3.12 F0 CRISPR assay as a platform for compound screening in crispants

High-throughput drug and small molecule screens have led to the discovery of many potential candidate drugs for various diseases [226]. Most drug screens are usually performed in an *in vitro* setting such as cell-based or biochemical assays, yielding thousands of potential candidates showing a phenotype at the cellular level. However, such knowledge can be irrelevant in an *in vivo* situation, as bioavailability and pharmacokinetic properties could restrict drug access to target site [227]. Using animal models for *in vivo* screens offer a substantial advantage of directly revealing candidates with actual physiological relevance. Teleost fish models boast numerous economical, logistical and physiological advantages over other models, which make them highly regarded for cardiovascular research, particularly in the context of high-throughput applications such as drug screens and toxicological studies [56,57,80–82]. Particularly, teleost fish represent a cost-efficient vertebrate model, able to bridge the gap between high-throughput *in vitro* applications lacking physiological relevance and lower throughput *in vivo* models (e.g., rodents) with high human relevance [58,59,228]. Although zebrafish still is more popular compared to other teleosts such as medaka, relatively longer medaka development within the chorion (6-8 days) with significantly less pronounced movements compared to zebrafish (2-3 days) presents a unique opportunity to investigate long term drug exposure effects on embryos without the need for sedation.

A proof of concept for the use of medaka embryos for drug screening purposes was previously demonstrated by measuring medaka heart rate response when exposed to cardioactive drugs [72]. Moreover, another pilot was previously performed, where embryos from three different strains of inbred medaka (HdrR-II, Cab and HO5) were subjected to the same known cardioactive drug, Verapamil, and exhibited significantly different strain-specific responses [217]. This pilot has only teased the potential of using highly isogenic medaka strains to perform pharmacogenomics studies. With further optimizations, a compound screen can be applied across the entire medaka inbred panel (of more than 100 strains), resulting in a plethora of clinically relevant insights into the role of genetics in drug response.

Another pilot has been attempted in this study using the heart rate assay to assess symptomatic treatment of arrhythmia phenotypes using anti-arrhythmic drugs in the background of *scn4ab* homozygous mutants (data not shown). Known anti-arrhythmic

drugs with various modes of action were tested: Quinidine (Class Ia, intermediate sodium and potassium channel blocker), Mexiletine (Class Ib, fast sodium channel blocker), Flecainide (Class Ic, slow sodium channel blocker), Propranolol (Class II, beta-blocker), Amiodarone and Vernakalant (Class III, potassium channel blockers), and Verapamil (Class IV, calcium channel blocker). Although the 2:1 AV-block arrhythmia phenotypes in *scn4ab* F0 crispants and homozygous F2 mutants were indistinguishable, *scn4ab* homozygous mutants were primarily chosen for the pilot due to the constant supply of arrhythmic embryos without the need of injections. Only embryos with moderate arrhythmia were considered for the compound assay, as genotyping randomly selected embryos consistently showed that only homozygous *scn4ab* mutant embryos exhibit moderate and strong arrhythmias. Unfortunately, due to various complications and limitations with respect to time and analysis, the pilot results were inconclusive within the timeframe of this study. The drugs appeared to have an overall depressive effect on the heart rate. Most drugs actually worsened the arrhythmias with prolonged exposure, but no specific-arrhythmia alleviation was so far observed.

The binary assessment of the arrhythmia phenotype (an embryo classified as arrhythmic or not arrhythmic) requires a direct view of the heart. This was a substantial challenge mainly due to the different embryo orientations, especially under bright field illumination, the presence of background signal obscured the view on the heart. Thus, there are various imaging and screening parameters that need further optimization. Using transgenic cardiac reporter lines, for instance, would immensely improve arrhythmia detection from unideal orientations. Due to the numerous applications and opportunities for novel compound discoveries, it would be worthwhile developing such a platform that combines genetic validation and drug discovery. This would ultimately enable the co-identification of novel genetic players and interacting compounds with rescuing power.

4

Conclusion

Cardiovascular diseases, the leading cause of death globally, can be potentially prevented through early diagnosis and intervention. Genetic factors are being widely recognised for their crucial contribution to the onset and development of cardiovascular diseases. Consequently, continuous expeditions are being undertaken through the human genome in the form of genome wide association studies (GWAS), constantly mining for novel genetic markers with diagnostic and predictive power for cardiovascular diseases. The extensive data produced by the human GWAS although provides a valuable resource for identification of such markers, their functional relevance of associations is concealed by the lack of mechanistic insights of the correlated genes. The pressure for mechanistic data and the absence of gene validation assays offering rapid functional insights has led to a negative loop of rediscovery, discriminating against novelty and reliance on known and established genes.

In this thesis, I have addressed this blind spot of discovery by establishing an experimental pipeline for the rapid functional validation of candidate genes associated to clinically-relevant heart phenotypes in human GWAS. By that, I could identify novel genes with diagnostic and predictive potential of human cardiovascular diseases. By combining the established high-throughput heart rate assay to a reverse genetic validation approach via CRISPR/Cas9-mediated mutagenesis of understudied and neglected human GWAS candidates *in vivo* in medaka F0, I identified 11 novel candidates with a newly discovered role in heart rate. Therefore, not only did I redirect the spotlight onto this understudied domain of candidates so far neglected, but I also provided first experimental evidence connecting them to cardiac development and function. Moreover, I have developed a rapid high-throughput genotyping assay perfectly complementing the functional gene validation assay to achieve individual embryo genotype-phenotype correlations.

Several aspects of cardiac function were revealed to be closely conserved despite the evolutionarily distance between humans and medaka, the average resting heart rate, and most importantly the high correlation between the respective phenotypes observed in medaka crispant embryos and the associated phenotype in adult human GWAS. This deep functional conservation emphasizes the high relevance of medaka embryos for the straight forward genetic validation and identification of novel predictive genetic markers for cardiovascular diseases in humans. The high-throughput gene validation pipeline

developed and the findings in this study therefore serve as the foundation for future investigations towards the deeper characterization of identified novel genes, the functional validation of additional disease-associated human GWAS candidate genes and SNPs in monogenetic but also in polygenic contexts. The assays developed and presented here are highly versatile with almost unlimited potential. With proper investments into software development using machine-learning, a multitude of cardiovascular-relevant phenotypes can be simultaneously extracted and quantified in a single experimental platform. Consequently, the effects of single and multiple gene mutations can be more intricately evaluated, potentially revealing novel insights into the functional gene networks controlling heart development, function and disease.

5

Materials & Methods

5.1 Materials

5.1.1 Fish lines

All lines were created by the CRISPR/Cas9 system (CR) by microinjection in one-cell stage medaka embryos, except for the transgenic HdrR-II dual *cmlc2* reporter line, which was generated by Tol2 injection [217].

Table 5.1: Stocks and transgenic lines used in this thesis.
All Cab stocks are derived from the same original Cab stock [48].

Name of fish line	Internal stock number	Source
Medaka Cab F67-F74	7524, 7796, 8072, 8368, 8617, 8813, 9170 and 9406	Laboratory stock
Zebrafish AB F3	7331	Laboratory stock
HdrR T2(*cmlc2::EGFP cmlc2::H2A-mCherry*) F1	9341	Laboratory stock
CR(SCN5A_T2)	8761	This thesis
CR(SCN5A_T2)_cF1	8934	This thesis
CR(SCN5A_T2)_F1c1	9526	This thesis

5.1.2 Plasmids

Table 5.2: Plasmids used in this thesis.
Plasmid stock numbers refer to the internal database of the Wittbrodt Laboratory at COS.

Plasmid name	Stock number	Source
DR274 sgRNA backbone (T7)	3632	Laboratory stock
pCS2+ Inv X_Cas9	5197	Laboratory stock
DR274(VEPH1_T1)	5246	This thesis
DR274(BAG3_T1)	5247	This thesis
DR274(CASQ2_T1)	5249	This thesis

Plasmid name	Stock number	Source
DR274(CASQ2_T4)	5250	This thesis
DR274(CCDC141_T1)	5251	This thesis
DR274(CLCNKA_T1)	5253	This thesis
DR274(CLCNKA_T3)	5254	This thesis
DR274(EDN1_T1)	5255	This thesis
DR274(EDN1_T4)	5256	This thesis
DR274(GIGYF1_T1)	5257	This thesis
DR274(GIGYF1_T3)	5258	This thesis
DR274(HOMEZ_T1)	5259	This thesis
DR274(HOMEZ_T2)	5260	This thesis
DR274(KCNH2_T1)	5261	This thesis
DR274(KCNH2_T2)	5262	This thesis
DR274(NKX2-5_T4)	5265	This thesis
DR274(NKX2-5_T5)	5266	This thesis
DR274(SCN5A_T1)	5267	This thesis
DR274(SCN5A_T2)	5268	This thesis
DR274(UFSP1_T1)	5269	This thesis
DR274(UFSP1_T2)	5270	This thesis
DR274(ABCB1_T7)	5475	This thesis
DR274(CEP85L_T5)	5477	This thesis
DR274(CLCNKA_T10)	5479	This thesis
DR274(CMYA5_T1)	5480	This thesis
DR274(CNOT1_T5)	5483	This thesis
DR274(HCN4_T5)	5484	This thesis
DR274(MAML3_T19)	5487	This thesis
DR274(MYRF_T1)	5488	This thesis
DR274(NUBP2_T4)	5491	This thesis
DR274(PIEZO1_T1)	5492	This thesis
DR274(PLG_T1)	5494	This thesis
DR274(RGS3_T6)	5499	This thesis
DR274(RNF207_T2)	5501	This thesis
DR274(SCMH1_T2)	5503	This thesis
DR274(SH2B3_T1)	5504	This thesis
DR274(SLC17A3_T5)	5507	This thesis

Plasmid name	Stock number	Source
DR274(SMG6_T2)	5509	This thesis
DR274(TTN_T1)	5510	This thesis
DR274(ZFHX3_T1)	5512	This thesis
DR274(cabp4_T9)	5515	This thesis
DR274(CDC42l_T2)	5516	This thesis
DR274(duox_T3)	5519	This thesis
DR274(EML6_T2)	5521	This thesis
DR274(GIT2_T3)	5523	This thesis
DR274(mus81_T7)	5525	This thesis
DR274(OGDH_T4)	5527	This thesis
DR274(or124-2_T1)	5528	This thesis
DR274(plekha8_T9)	5531	This thesis
DR274(ttl_T1)	5532	This thesis
DR274(ATP8B4_ROI_T2)	5703	This thesis
DR274(TRAPPC12_ROI_T2)	5706	This thesis
DR274(SSPO_ROI3_T4)	5711	This thesis
DR274(MINAR1_ROI3_T1)	5714	This thesis
DR274(PPP1R9A_ROI_T7)	5715	This thesis
DR274(XYLB_ROI2_T1)	5719	This thesis
DR274(COL9A1b_ROI2_T2)	5722	This thesis
DR274(PADI2_ROI2_T4)	5725	This thesis
DR274(KCNH2_ROI_T1)	5727	This thesis
DR274(OR5AU1_ROI2_T2)	5731	This thesis
DR274(GRID2_ROI2_T8)	5734	This thesis
DR274(NACA_ROI3_T1)	5738	This thesis

5.1.3 Oligonucleotides

All oligonucleotides were ordered from Eurofins Genomics (*previously available from a common laboratory oligonucleotide stock).

Table 5.3: Oligonucleotides for annealing.

Number	Name	Sequence (5'–3')
JW7001	BAG3_T1_F	TAGgGCAGGGCGCCGTTCACCC
JW7002	BAG3_T1_R	AAACGGGTGAACGGCGCCCTGC

Number	Name	Sequence (5'–3')
JW7005	CASQ2_T1_F	TAggTTGTCTGGGATGGGTTCG
JW7006	CASQ2_T1_R	AAACCGAACCCATCCCAGACAA
JW7007	CASQ2_T4_F	TAggTCTACCACGAACCCATCC
JW7008	CASQ2_T4_R	AAACGGATGGGTTCGTGGTAGA
JW7009	CCDC141_T1_F	TAggAGGAACCTGAAGGTTGTC
JW7010	CCDC141_T1_R	AAACGACAACCTTCAGGTTCCT
JW7013	CLCNKA_T1_F	TAggCCCCAAGCAGCCGACCAC
JW7014	CLCNKA_T1_R	AAACGTGGTCGGCTGCTTGGGG
JW7015	CLCNKA_T3_F	TAggATGAGCCACTCCTTCCCC
JW7016	CLCNKA_T3_R	AAACGGGGAAGGAGTGGCTCAT
JW7017	EDN1_T1_F	TAggTCCGTCGGCAGCCGGCAT
JW7018	EDN1_T1_R	AAACATGCCGGCTGCCGACGGA
JW7019	EDN1_T4_F	TAGgCGACGGAGTCTGCGCGGA
JW7020	EDN1_T4_R	AAACTCCGCGCAGACTCCGTCG
JW7021	GIGYF1_T1_F	TAggTGAACGCCGGAGTTAACG
JW7022	GIGYF1_T1_R	AAACCGTTAACTCCGGCGTTCA
JW7023	GIGYF1_T3_F	TAGgATGAACCGGCATGAACGC
JW7024	GIGYF1_T3_R	AAACGCGTTCATGCCGGTTCAT
JW7025	HOMEZ_T1_F	TAGgTACCAGCAGGTGCGAGAT
JW7026	HOMEZ_T1_R	AAACATCTCGCACCTGCTGGTA
JW7027	HOMEZ_T2_F	TAggAACCAATCTCGCACCTGC
JW7028	HOMEZ_T2_R	AAACGCAGGTGCGAGATTGGTT
JW7029	KCNH2_T1_F	TAggACGGAAGTTATCATTCCG
JW7030	KCNH2_T1_R	AAACCGGAATGATAACTTCCGT
JW7031	KCNH2_T2_F	TAggTGATAACTTCCGTCTTCG
JW7032	KCNH2_T2_R	AAACCGAAGACGGAAGTTATCA
JW7033	NKX2-5_T4_F	TAGgCGCGGGTCCTCTTCTCCC
JW7034	NKX2-5_T4_R	AAACGGGAGAAGAGGACCCGCG
JW7035	NKX2-5_T5_F	TAggGACAGACCCAAGCCCCGG
JW7036	NKX2-5_T5_R	AAACCCGGGGCTTGGGTCTGTC
JW7037	SCN5A_T1_F	TAggGAGCCGTTGCGGATTGCT
JW7038	SCN5A_T1_R	AAACAGCAATCCGCAACGGCTC
JW7039	SCN5A_T2_F	TAggAGGCTTAGCAATCCGCAA
JW7040	SCN5A_T2_R	AAACTTGCGGATTGCTAAGCCT

Number	Name	Sequence (5'–3')
JW7041	UFSP1_T1_F	TAGgTGGAGAGCATCGCAGTCC
JW7042	UFSP1_T1_R	AAACGGACTGCGATGCTCTCCA
JW7043	UFSP1_T2_F	TAggGAGCTCCATCGGCACTTC
JW7044	UFSP1_T2_R	AAACGAAGTGCCGATGGAGCTC
JW7045	VEPH1_T1_F	TAGGCTCTGGTGGAGGTGTCCC
JW7046	VEPH1_T1_R	AAACGGGACACCTCCACCAGAG
JW7865	CLCNKA_T10_F	TAggCTCCACTGGAGTAGTTTT
JW7866	CLCNKA_T10_R	AAACAAAACTACTCCAGTGGAG
JW7869	RNF207_T2_F	TAggGACTTGTTTGTACTGATC
JW7870	RNF207_T2_R	AAACGATCAGTACAAACAAGTC
JW7871	PLG_T1_F	TAGGCAACGGGGCCAATTATCG
JW7872	PLG_T1_R	AAACCGATAATTGGCCCCGTTG
JW7875	MYRF_T1_F	TAggTTATTGGAGTCCATATTG
JW7876	MYRF_T1_R	AAACCAATATGGACTCCAATAA
JW7881	RGS3_T6_F	TAGgTCAAGTCACAGTCCAAGA
JW7882	RGS3_T6_R	AAACTCTTGGACTGTGACTTGA
JW7885	MAML3_T19_F	TAggATGTAAGGTGTCATCATA
JW7886	MAML3_T19_R	AAACTATGATGACACCTTACAT
JW7889	SCMH1_T2_F	TAggTGACCTGCGCTGTACGAG
JW7890	SCMH1_T2_R	AAACCTCGTACAGCGCAGGTCA
JW7891	TTN_T1_F	TAGgGCTAGCTGTCAAAGCCAT
JW7892	TTN_T1_R	AAACATGGCTTTGACAGCTAGC
JW7895	SH2B3_T1_F	TAggGAAGTCCGTCGCTGCAAT
JW7896	SH2B3_T1_R	AAACATTGCAGCGACGGACTTC
JW7899	ZFHX3_T1_F	TAggCTGAGCGCACCCTGCCTG
JW7900	ZFHX3_T1_R	AAACCAGGCAGGGTGCGCTCAG
JW7905	SMG6_T2_F	TAggAGATAAAACCAAGGGCGT
JW7906	SMG6_T2_R	AAACACGCCCTTGGTTTTATCT
JW7907	CMYA5_T1_F	TAGgAGAACAGGTAATTCTCGT
JW7908	CMYA5_T1_R	AAACACGAGAATTACCTGTTCT
JW7917	CNOT1_T5_F	TAGgCAAGTGTTGACAATACTG
JW7918	CNOT1_T5_R	AAACCAGTATTGTCAACACTTG
JW7921	CEP85L_T5_F	TAggCCCAGGACTTCTCAGATA
JW7922	CEP85L_T5_R	AAACTATCTGAGAAGTCCTGGG

Number	Name	Sequence (5'–3')
JW7923	HCN4_T5_F	TAggACTCCCTTCGAACTTGTG
JW7924	HCN4_T5_R	AAACCACAAGTTCGAAGGGAGT
JW7929	ABCB1_T7_F	TAggGTCCCTGAGTGTGAAGAG
JW7930	ABCB1_T7_R	AAACCTCTTCACACTCAGGGAC
JW8267	NUBP2_T4_F	TAGgACTGAATGTGGCACTGTT
JW8268	NUBP2_T4_R	AAACAACAGTGCCACATTCAGT
JW8269	PIEZO1_T1_F	TAggGGGAGCCCGGATGTAGTT
JW8270	PIEZO1_T1_R	AAACAACTACATCCGGGCTCCC
JW8275	SLC17A3_T5_F	TAGgCTTGGCGCCGATTGTGAC
JW8276	SLC17A3_T5_R	AAACGTCACAATCGGCGCCAAG
JW8764	OGDH_T4_F	TAggAAGCTGGACTTGGCCTCA
JW8765	OGDH_T4_R	AAACTGAGGCCAAGTCCAGCTT
JW8768	cabp4_T9_F	TAGgATTTGACTACGATGCAGA
JW8769	cabp4_T9_R	AAACTCTGCATCGTAGTCAAAT
JW8770	CDC42l_T2_F	TAggAGCGGAGAGAAACTGGCT
JW8771	CDC42l_T2_R	AAACAGCCAGTTTCTCTCCGCT
JW8776	duox_T3_F	TAggTTCGGCACGCTTTCTCTA
JW8777	duox_T3_R	AAACTAGAGAAAGCGTGCCGAA
JW8780	EML6_T2_F	TAggGCGCTGTTCGCACGCTAA
JW8781	EML6_T2_R	AAACTTAGCGTGCGAACAGCGC
JW8784	GIT2_T3_F	TAggAACGCCTTCGAAACACGG
JW8785	GIT2_T3_R	AAACCCGTGTTTCGAAGGCGTT
JW8788	mus81_T7_F	TAggGAGGATGGACGACCTCTG
JW8789	mus81_T7_R	AAACCAGAGGTCGTCCATCCTC
JW8790	or124-2_T1_F	TAGgCCACGTGGTCTCGTACGC
JW8791	or124-2_T1_R	AAACGCGTACGAGACCACGTGG
JW8796	plekha8_T9_F	TAGgGGTGAAGCTTTCATAACA
JW8797	plekha8_T9_R	AAACTGTTATGAAAGCTTCACC
JW8798	ttl_T1_F	TAGgTGGTAAATTACTACAGAG
JW8799	ttl_T1_R	AAACCTCTGTAGTAATTTACCA
JW9322	ATP8B4_ROI_T2_F	TAggAGTTCTACGATAACACCC
JW9323	ATP8B4_ROI_T2_R	AAACGGGTGTTATCGTAGAACT
JW9328	TRAPPC12_ROI_T2_F	TAggACACGCAGTGCCTCAAGC
JW9329	TRAPPC12_ROI_T2_R	AAACGCTTGAGGCACTGCGTGT

Number	Name	Sequence (5'–3')
JW9338	SSPO_ROI3_T4_F	TAGgGCACCAAGTCATGTGGTT
JW9339	SSPO_ROI3_T4_R	AAACAACCACATGACTTGGTGC
JW9344	MINAR1_ROI3_T1_F	TAGgTGCCGTCGGCGACGCGTA
JW9345	MINAR1_ROI3_T1_R	AAACTACGCGTCGCCGACGGCA
JW9346	PPP1R9A_ROI_T7_F	TAggGTTCTTAGGAGATTCGGG
JW9347	PPP1R9A_ROI_T7_R	AAACCCCGAATCTCCTAAGAAC
JW9354	XYLB_ROI2_T1_F	TAggTCCAGTGTCAGTACCTAC
JW9355	XYLB_ROI2_T1_R	AAACGTAGGTACTGACACTGGA
JW9360	COL9A1b_ROI2_T2_F	TAGgCGGACGCGTAGGCATTCC
JW9361	COL9A1b_ROI2_T2_R	AAACGGAATGCCTACGCGTCCG
JW9366	PADI2_ROI2_T4_F	TAggACTGGTATTACGGTTAAG
JW9367	PADI2_ROI2_T4_R	AAACCTTAACCGTAATACCAGT
JW9370	KCNH2_ROI_T1_F	TAggTCACTGCTGGGAGAACCG
JW9371	KCNH2_ROI_T1_R	AAACCGGTTCTCCCAGCAGTGA
JW9428	OR5AU1_ROI2_T2_F	TAGgCATGGACTCTTGCCTGTA
JW9429	OR5AU1_ROI2_T2_R	AAACTACAGGCAAGAGTCCATG
JW9434	GRID2_ROI2_T8_F	TAggAGGCTACGGCTAAGGGTC
JW9435	GRID2_ROI2_T8_R	AAACGACCCTTAGCCGTAGCCT
JW9442	NACA_ROI3_T1_F	TAggGGTCTTAGGCAAGTAACG
JW9443	NACA_ROI3_T1_R	AAACCGTTACTTGCCTAAGACC

Table 5.4: Primers for sequencing in other non-medaka organisms.

Organism	Name	Sequence (5'–3')
**Arabidopsis thaliana*	tubulin_F	AGTAGTTTAAGGACCTACTTCG
	tubulin_R	GAGCCTTACAACGCTACTCTGTCTGTC
**Chironomous riparius*	germ cell-less_F	CTTTATTAGCGTTCGGTCGTG
	germ cell-less_R	CCAACGTAACATTATCCTTCGC
**Drosophila melanogaster*	twist_F	CAATTTGAGCAATGGCCGGAAGGA
	twist_R	ACTGCTGCTGCTGGTTGTTGTAGA
**Mus musculus*	trpm4 P1	GTTTGATGTCTCCTTCAGTCG
	trpm4 P2	GAGTTCCTGTCCTCCTAAAGG
	trpm4 P3	ACCTACAGGAAACCTCGGGG
**Danio rerio*	drOca2_F	ACAGGTGCTGTATAATTGGACCAT
	drOca2_R	AAAGAGTGGTCATAAACGGCTACT

Table 5.5: Primers for amplification and sequencing in medaka.

Number	Name	Sequence (5'–3')
JW7100	BAG3_F	CTTCGCTTAGGAGCCAGTCC
JW7101	BAG3_R	CAGTATGGTCTGAACCCGCC
JW7102	CASQ2_F	CGAACAGGTCAAACCGTGTG
JW7103	CASQ2_R	GGAACTTGCACAAACGGACC
JW7104	CCDC141_F	GGCTACGGATGAAGGAGCTC
JW7105	CCDC141_R	TCCTGGCTAAGTGAAGCAGC
JW7106	CLCNKA_F	GATTCATGCTGTGTTCCCGC
JW7107	CLCNKA_R	CCACTGAGTTGGTCGCTTTG
JW7108	EDN1_F	TCTCCGTGCTGTCAGTGTTC
JW7109	EDN1_R	ATGCTGGTTGCCATGGAGTC
JW7110	GIGYF1_F	ACAGGAGATAGCAGCAGCAG
JW7111	GIGYF1_R	CGTCCCTGATTGTAGATTTGCG
JW7112	HOMEZ_F	GCCAAGGAAGACCAAAGAGC
JW7113	HOMEZ_R	TCACAAACCTCGCTCTAGGC
JW7114	KCNH2_F	CAACCTCTACACTCGCCCAG
JW7115	KCNH2_R	ACCTCATCGCCACTTGAGTG
JW7118	NKX2.5_F	TAGCTTTGAGGCCTGCAGAC
JW7119	NKX2.5_R	CGCATAGTTCGTGTTGCAGG
JW7120	SCN5A_F	TCACCACTTCTGCAGGATGG
JW7121	SCN5A_R	GAGGCGTAGTGATCTGACGG
JW7122	UFSP1_F	TGGGAAAGGACTAGAGGAGGG
JW7123	UFSP1_R	GACACTCGCTTCCAGGACAC
JW7124	VEPH1_F	TCTGTCAGAGCGGATGCAAG
JW7125	VEPH1_R	AGTCTCTGATCGGGATGCAAC
JW7780	SCN5A_T2_Fwd2	CGTTGTCAGAGCCGTTGCG
JW7781	SCN5A_T2_Rev2	ATCCTATTAGGCTTAGCAATCCG
JW7931	ABCB1_F	CCTCTGCAGAAGCTGGATGG
JW7932	ABCB1_R	TTTGCGATGGCGTGTGAGCG
JW7933	CEP85L_F	AGGAGTGAGCTGAGGATGGG
JW7934	CEP85L_R	GCATCCTCACTCTAGCCGGG
JW7935	CMYA5_F	TTTGTGGGAGCTCGTCTGGG
JW7936	CMYA5_R	GCCACGCTGAGATGGAGCCC

Number	Name	Sequence (5'–3')
JW7937	CNOT1_F	CCGGTTGTGGCGCCTTGAGG
JW7938	CNOT1_R	GCTTTGGGTCTGGCACTGCG
JW7939	HCN4_F	GACCCGTCAAGTGCCACAGG
JW7940	HCN4_R	CACACCCTCCTTGTCCCTTG
JW7941	MAML3_F	CGGCAGCCATGTTGTCGTAC
JW7942	MAML3_R	AAGCGGGTTATGTGTGCTGC
JW7943	MYRF_F	GGTCCACCAACCGCCTGCC
JW7944	MYRF_R	AGCATCCTAGCATTGCAGCC
JW7945	PLG_F	CACAAACACAGCCGCACGCC
JW7946	PLG_R	GTGGGCGTGTCCGAGTCGAC
JW7949	RGS3_F	TGAGCTGTGTTCGCCCACTC
JW7950	RGS3_R	GGAGGGCTTCAACTGTGAGG
JW7951	RNF207_F	GCAGAAGCCTCCCATTGACG
JW7952	RNF207_R	CCTCTGGCTGACCGCTCCTG
JW7953	SCMH1_F	ACTCTGTTCGAGCGCCTCCC
JW7954	SCMH1_R	AGGCGCATCCTGCAGCTTCC
JW7955	SH2B3_F	TTTCTGCCGGCATCTGCTCC
JW7956	SH2B3_R	CCTGCTGGTCGTTGTCCGTC
JW7957	SMG6_F	CCCAGCATCAGTGGAGGTGC
JW7958	SMG6_R	CCCTGAGCCTCAGTCCCAGC
JW7959	TTN_F	CAGAGCTGGACGGACATCGG
JW7960	TTN_R	GGAGCGCTGACGTCTGCCTC
JW7961	ZFHX3_F	GGCTGGGTGGTGGGTGTCAG
JW7962	ZFHX3_R	TGTGCGGCTAGGTGGTGGAC
JW8277	NUBP2_F	ACTGATGAGCCAGCCAACCC
JW8278	NUBP2_R	TGGTCTGTGGCCACATGGTG
JW8279	PIEZO1_F	GCAGCAGCAGGAAGGTCAGG
JW8280	PIEZO1_R	CCGCTGCATTAGACCACTGC
JW8281	SLC17A3_F	CCGCTGTGCCCAGGAGCAAG
JW8282	SLC17A3_R	AGGAAACGACCCGTGACCAC
JW8802	OGDH_F	GAACACTCGTCTGGCCGGCC
JW8803	OGDH_R	TACAGCAAGTCCCGCCCACG
JW8804	cabp4_F	GTGCCATCTTGCCATGCTGC
JW8805	cabp4_R	ATCCTGCTGCTCCTGGTCCC

Number	Name	Sequence (5'–3')
JW8806	CDC42l_F	TCCAGAGCTGCGAGGATGGC
JW8807	CDC42l_R	CCAGCCGCTCTGTTCTGGGC
JW8808	duox_F	GACGCTGTCAGCTCGCACTG
JW8809	duox_R	GGTCCACTCATCGCTCCCTC
JW8810	EML6_F	ACGCCACGGACAGCATGGTC
JW8811	EML6_R	TCACTGACAGCTTTGAAGGTTGACG
JW8812	GIT2_F	TTGCTGTCGCCGCTGCTCTG
JW8813	GIT2_R	GGGCTCATGGGACCTTATGTTGG
JW8814	mus81_F	GAAGACGACGAGACGGCCGG
JW8815	mus81_R	TGACCCGAACAGTACCTTTGTG
JW8816	or124-2_F	TGAGGCTCAGACCCTCCTGG
JW8817	or124-2_R	ATGCGGTCCGGATGAGCTGG
JW8818	plekha8_F	CTGGCACCTTCCTGTCCAGC
JW8819	plekha8_R	ACGGAAGTCGCCCAAGTCCG
JW8820	ttl_F	CCGGGCGGGACCAACAACTG
JW8821	ttl_R	GGAGTCAGACAGCTCTGGGC
JW9376	ATP8B4_ROI_F	ACCTACAGTCAAAGGGAAATGAGC
JW9377	ATP8B4_ROI_R	ACGCTGCCTCGTCTGGTGAC
JW9382	TRAPPC12_ROI_F	GTGCCTGCTGTACCTGGGCC
JW9383	TRAPPC12_ROI_R	ACCACCACCCTCTCTCCCAGC
JW9392	SSPO_ROI3_F	TGCAGTCGAGGGTCAGTGGTCG
JW9393	SSPO_ROI3_R	CACAGGATCGCGGGCAAGAAGG
JW9396	MINAR1_ROI2&3_F	CGGTCAGCATGAACTAAGCTTCGG
JW9397	MINAR1_ROI2&3_R	TGGAGACGGACAGTGATTCCAGTGG
JW9398	PPP1R9A_ROI1&2_F	CGAGGTCCAACCGAGGCAGC
JW9399	PPP1R9A_ROI1&2_R	TCTCTGTCAGATCTGAGCCGGG
JW9404	XYLB_ROI2_F	CCTGCAGCAACATGGCAGTGTG
JW9405	XYLB_ROI2_R	ACCATCACTGTGGTCAATGGCTGC
JW9410	COL9A1b_ROI2_F	AGGGACTCACTTACCGGAAGCCC
JW9411	COL9A1b_ROI2_R	CTGGTCCTCAGGGAGTGGCAGG
JW9416	PADI2_ROI2_F	GCATGACACTACCTGAGAATGAAAC
JW9417	PADI2_ROI2_R	AGCCTCAGTCATCTCAGTAAAGC
JW9420	KCNH2_ROI_F	TCGCCCAGGCAAGTCCAACG
JW9421	KCNH2_ROI_R	ACCTCATCGCCACTTGAGTGTG

Number	Name	Sequence (5'–3')
JW9444	OR5AU1_ROI2_F	GCATTCAGCAGTTCTTCTGTCTTC
JW9445	OR5AU1_ROI2_R	GGCTTTGATAGATCTGTGTGGAAGC
JW9448	GRID2_ROI2_F	ACCTACAGACCTGTAACTCCTTCGC
JW9449	GRID2_ROI2_R	ACCTGGAGCACCTACACCACAGG
JW9456	NACA_ROI3_F	TCGCTCTCCTCCTGTACTGTAGGG
JW9457	NACA_ROI3_R	TCTGAGTCTACTGAAGCCAGCC
*JW3292	OCA2_F	GTTAAAACAGTTTCTTAAAAAGAACAGGA
*JW3293	OCA2_R	AGCAGAAGAAATGACTCAACATTTTG
*JW2051	rx2_F	TGCATGTTCTGGTTGCAACG
*JW2052	rx2_R	AGGGACCATACCTGACCCTC
*JW1408	Cr_gfp_T3_R	AAACGCTCGACCAGGATGGGCA
*JW1451	actb_F	CAGCAACGACTTCGCACAAA
*JW1452	actb_R	CAGGGGCAATTCTCAGCTCA
*ye	dnmtY_F	ACAGGTAAACCAGAAAAACTA
*ll	dnmtY_R	AACTAATTCATCCCCATTCC
*JW1458	gfp_noStart_F	GTGAGCAAGGGCGAGGAGCT
*JW1459	gfp_noStop_R	CTTGTACAGCTCGTCCATGC
*JW8829	DR274_sgRNAtemplate_F	AAAAGCACCGACTCGGTGCCACT
*JW8830	DR274_sgRNAtemplate_R	GAGGTCAGGTATGATTTAAATGGTCAGTATTGA

5.1.4 RNAs

Table 5.6: sgRNAs used in this thesis.
For respective plasmid stock numbers, refer to Table 5.2. All sgRNAs target medaka genes unless specified otherwise (*Dr = Danio rerio*). Protospacer adjacent motif (PAM) in target sequence enclosed in square brackets "[]". (*previously available from a common laboratory stock).

Name	Target	Sequence [PAM]
OH_sgRNA 3	BAG3_T1	GAGCAGGGCGCCGTTCACCC[GGG]
OH_sgRNA 5	CASQ2_T1	TGTTGTCTGGGATGGGTTCG[TGG]
OH_sgRNA 6	CASQ2_T4	TTTCTACCACGAACCCATCC[CAG]
OH_sgRNA 7	CCDC141_T1	CCAGGAACCTGAAGGTTGTC[AGG]
OH_sgRNA 9	CLCNKA_T1	TTCCCCAAGCAGCCGACCAC[CAG]
OH_sgRNA 10	CLCNKA_T3	ACATGAGCCACTCCTTCCCC[AAG]
OH_sgRNA 11	EDN1_T1	ACTCCGTCGGCAGCCGGCAT[CAG]
OH_sgRNA 12	EDN1_T4	GCCGACGGAGTCTGCGCGGA[GGG]
OH_sgRNA 13	GIGYF1_T1	CATGAACGCCGGAGTTAACG[CAG]

Name	Target	Sequence [PAM]
OH_sgRNA 14	GIGYF1_T3	GAATGAACCGGCATGAACGC[CGG]
OH_sgRNA 15	HOMEZ_T1	GCTACCAGCAGGTGCGAGAT[TGG]
OH_sgRNA 17	KCNH2_T1	AGACGGAAGTTATCATTCCG[CAG]
OH_sgRNA 18	KCNH2_T2	AATGATAACTTCCGTCTTCG[TAG]
OH_sgRNA 19	NKX2.5_T4	GCCGCGGGTCCTCTTCTCCC[AGG]
OH_sgRNA 20	NKX2.5_T5	CGGACAGACCCAAGCCCCGG[AGG]
OH_sgRNA 21	SCN5A_T1	CAGAGCCGTTGCGGATTGCT[AAG]
OH_sgRNA 22	SCN5A_T2	TTAGGCTTAGCAATCCGCAA[CGG]
OH_sgRNA 23	UFSP1_T1	GCTGGAGAGCATCGCAGTCC[AGG]
OH_sgRNA 24	UFSP1_T2	AGGAGCTCCATCGGCACTTC[CAG]
OH_sgRNA 25	VEPH1_T1	GGCTCTGGTGGAGGTGTCCC[AGG]
OH_sgRNA 27	CLCNKA_T10	TTCTCCACTGGAGTAGTTTT[TGG]
OH_sgRNA 28	RNF207_T2	TAGACTTGTTTGTACTGATC[TGG]
OH_sgRNA 29	PLG_T1	GGCAACGGGGCCAATTATCG[AGG]
OH_sgRNA 31	RGS3_T6	GATCAAGTCACAGTCCAAGA[TGG]
OH_sgRNA 32	MAML3_T19	TCATGTAAGGTGTCATCATA[GGG]
OH_sgRNA 33	SCMH1_T2	TCTGACCTGCGCTGTACGAG[TGG]
OH_sgRNA 34	TTN_T1	GAGCTAGCTGTCAAAGCCAT[GGG]
OH_sgRNA 35	SH2B3_T1	AAGAAGTCCGTCGCTGCAAT[CGG]
OH_sgRNA 36	ZFHX3_T1	AGCTGAGCGCACCCTGCCTG[AGG]
OH_sgRNA 37	SMG6_T2	AAAGATAAAACCAAGGGCGT[AGG]
OH_sgRNA 38	CMYA5_T1	GTAGAACAGGTAATTCTCGT[TGG]
OH_sgRNA 41	CEP85L_T5	AACCCAGGACTTCTCAGATA[GGG]
OH_sgRNA 42	HCN4_T5	AAACTCCCTTCGAACTTGTG[AGG]
OH_sgRNA 43	ABCB1_T7	ATGTCCCTGAGTGTGAAGAG[TGG]
OH_sgRNA 44	NUBP2_T4	GAACTGAATGTGGCACTGTT[AGG]
OH_sgRNA 45	PIEZO1_T1	TTGGGAGCCCGGATGTAGTT[AGG]
OH_sgRNA 49	SLC17A3_T5	GTCTTGGCGCCGATTGTGAC[AGG]
OH_sgRNA 50	MYRF_T1	CCTTATTGGAGTCCATATTG[TGG]
OH_sgRNA 52	CNOT1_T5	GCCAAGTGTTGACAATACTG[AGG]
OH_sgRNA 53	OGDH_T4	CAAAGCTGGACTTGGCCTCA[GGG]
OH_sgRNA 54	cabp4_T9	GAATTTGACTACGATGCAGA[TGG]
OH_sgRNA 55	CDC42_T2	AGAGCGGAGAGAAACTGGCT[CGG]

Name	Target	Sequence [PAM]
OH_sgRNA 56	duox_T3	CATTCGGCACGCTTTCTCTA[AGG]
OH_sgRNA 57	EML6_T2	CTGCGCTGTTCGCACGCTAA[AGG]
OH_sgRNA 58	GIT2_T3	TAAACGCCTTCGAAACACGG[AGG]
OH_sgRNA 59	mus81_T7	AAGAGGATGGACGACCTCTG[TGG]
OH_sgRNA 60	or124-2_T1	GCCCACGTGGTCTCGTACGC[CGG]
OH_sgRNA 61	plekha8_T9	GCGGTGAAGCTTTCATAACA[CGG]
OH_sgRNA 62	ttl_T1	GCTGGTAAATTACTACAGAG[GGG]
OH_sgRNA 77	ATP8B4_ROI_T2	TAAGTTCTACGATAACACCC[TGG]
OH_sgRNA 80	TRAPPC12_ROI_T2	CAACACGCAGTGCCTCAAGC[TGG]
OH_sgRNA 85	SSPO_ROI3_T4	GTGCACCAAGTCATGTGGTT[GGG]
OH_sgRNA 88	MINAR1_ROI3_T1	GTTGCCGTCGGCGACGCGTA[GGG]
OH_sgRNA 89	PPP1R9A_ROI_T7	TGGTTCTTAGGAGATTCGGG[TGG]
OH_sgRNA 93	XYLB_ROI2_T1	TTTCCAGTGTCAGTACCTAC[AGG]
OH_sgRNA 96	COL9A1b_ROI2_T2	GACGGACGCGTAGGCATTCC[AGG]
OH_sgRNA 99	PADI2_ROI2_T4	AGACTGGTATTACGGTTAAG[AGG]
OH_sgRNA 101	KCNH2_ROI_T1	CATCACTGCTGGGAGAACCG[GGG]
OH_sgRNA 105	OR5AU1_ROI2_T2	GTCATGGACTCTTGCCTGTA[TGG]
OH_sgRNA 108	GRID2_ROI2_T8	AAAGGCTACGGCTAAGGGTC[TGG]
OH_sgRNA 112	NACA_ROI3_T1	TTGGTCTTAGGCAAGTAACG[GGG]
*TT_sgRNA_57	Oca2_ex8_T1	GAAACCCAGGTGGCCATTGC[AGG]
*TT_sgRNA_58	Oca2_ex8_T3	TTGCAGGAATCATTCTGTGT[GGG]
*TT_sgRNA_sg40	rx2_T5	GCATTTGTCAATGGATACCC[TGG]
*TT_sgRNA_sg144	*Dr*_Oca2_T1	GTACAGCGACTGGTTAGTCA[TGG]
*AC_sgRNA_18	GFP_BE_C71_T1	AGCACTGCACGCCGTAGGTC[AGG]
crRNA_OCA2_ sg58_IDT	OCA2_ex8_T3_IDT_cr RNA	UUGCAGGAAUCAUUCUGUGUG[UUU]UAG AGCUAUGCU

Table 5.7: mRNAs.
Plasmid numbers are from the internal database of the Wittbrodt Laboratory at COS and indicate the plasmids used to generate the transcription template.

Name	Plasmid stock number	Source
pCS2+ Inv *X_Cas9*	5197	Laboratory Stock
pCS2+ *GFP*	883	Laboratory Stock

5.1.5 Chemicals

Table 5.8: Chemicals and reagents used in this thesis.

Chemical/Reagent	Supplier
2-Propanol	Sigma-Aldrich
2,3-Butanedione 2-monoxime (BDM)	Abcam
Adenosine triphosphate (ATP)	Thermo Fisher Scientific
Agar	Carl Roth
Agarose	Sigma-Aldrich
Agarose Low Melt	Carl Roth
Ampicillin	Carl Roth
Bacto Tryptone	Gibco
Bromphenol blue	Sigma-Aldrich
Calcium chloride dihydrate ($CaCl_2 \cdot 2\ H_2O$)	AppliChem
Deoxyadenosine triphosphate (dATP)	Thermo Fisher Scientific
Dimethyl sulfoxide (DMSO)	Carl Roth
Disodium hydrogen phosphate (Na_2HPO_4)	Sigma-Aldrich
Disodium hydrogen phosphate dihydrate ($Na_2HPO_4 \cdot 2\ H_2O$)	Sigma-Aldrich
Ethanol 70 % (denatured)	Carl Roth
Ethanol 96 % (denatured)	Carl Roth
Ethanol 99 %	Sigma-Aldrich
Ethidium Bromide solution (EtBr)	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth
Formamide	Sigma-Aldrich
Glacial acetic acid 99 %	Merck
Glucose	Sigma-Aldrich
Glycerol	Merck
Glycine	Sigma-Aldrich
Hydrogen Chloride (HCl)	Merck
Hydrogen peroxide (H_2O_2)	Sigma-Aldrich
Hypochlorite solution (bleach)	Danklorix
Kanamycin	Carl Roth
Methanol (MeOH)	Carl Roth
Methylcellulose	Sigma-Aldrich
Methylene blue trihydrate	Sigma-Aldrich
N-Phenylthiourea (PTU)	Sigma-Aldrich

Chemical/Reagent	Supplier
Nuclease-free water	Sigma-Aldrich
Orange G	Sigma-Aldrich
Paraformaldehyde (PFA)	Sigma-Aldrich
Polyethylene glycol 4000 (PEG-4000)	Thermo Fisher Scientific
Potassium acetate (KAc)	AppliChem
Potassium chloride (KCl)	AppliChem
Potassium dihydrogen phosphate (KH_2PO_4)	Merck
Potassium hydrogen phosphate (K_2HPO_4)	Merck
Potassium hydroxide (KOH)	Merck
Proteinase K	Roche
Red sea salt	Red Sea
rNTPs (ATP, CTP, GTP, UTP)	Roche
Sodium acetate (NaAc)	Grüssing
Sodium chloride (NaCl)	Sigma-Aldrich
Sodium citrate	Sigma-Aldrich
Sodium dodecyl sulphate sodium solution (SDS) 20%	Carl Roth
Sodium hydrogen carbonate ($NaHCO_3$)	Merck
Sodium hydroxide (NaOH)	Sigma-Aldrich
Sodium tetraborate ($Na_2B_4O_7$)	Sigma-Aldrich
Sucrose	Carl Roth
Tricaine (MS-222)	Sigma-Aldrich
Tris base	Carl Roth
Triton X-100	Sigma-Aldrich
Tween 20	Sigma-Aldrich
Xylene cyanol	Serva
Yeast Extract	Carl Roth

5.1.6 Molecular materials

Table 5.9: Molecular materials used in this thesis.

Material	Supplier
Deoxynucleotide Triphosphates Set (dNTPs)	Thermo Fisher Scientific
DNA Purple loading dye (6x)	NEB
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific

Material	Supplier
Mach1T1 chemically competent *E. coli*	Thermo Fisher Scientific
RNA Loading Dye 2x Rapid	Thermo Fisher Scientific

5.1.7 Enzymes

Table 5.10: Enzymes and corresponding buffers used in this thesis.

Enzyme/Buffer	Supplier
Alkaline phosphatase (1 U/µl)	Roche
BamHI-HF (20 U/µl)	NEB
BsaI-HF (20 U/µl)	NEB
CutSmart buffer (10x)	NEB
DraI-FD	Thermo Fisher Scientific
FastDigest buffer (10x)	Thermo Fisher Scientific
FastDigest green buffer (10x)	Thermo Fisher Scientific
Hatching enzyme	lab made
IDT Duplex Buffer	Integrated DNA Technologies
KpnI-HF (20 U/µl)	NEB
NEB Buffer 2 (10x)	NEB
NotI-HF (20 U/µl)	NEB
Q5 High-Fidelity DNA Polymerase (2 U/µl)	NEB
Q5 Reaction Buffer (5x)	NEB
T4 DNA Ligase (5 U/µl)	Thermo Fisher Scientific
T4 DNA Ligase Buffer (10x)	Thermo Fisher Scientific
T7 Endonuclease I (10 U/µl)	NEB
TURBO DNase I (2 U/µl)	Thermo Fisher Scientific
XhoI (20 U/µl)	NEB

5.1.8 Kits

Table 5.11: Kits used in this thesis.

Kit	Company
innuPREP PCRpure Kit	Analytik Jena
MEGAscript T7 Transcription Kit	Thermo Fisher Scientific
MinElute Gel Extraction Kit	QIAGEN
Monarch DNA Gel Extraction Kit	NEB
QIAPrep Spin Miniprep Kit	QIAGEN

Kit	Company
QIAquick Gel Extraction Kit	QIAGEN
QIAquick Nucleotide Removal Kit	QIAGEN
QIAquick PCR Purification Kit	QIAGEN
RNeasy Mini Kit	QIAGEN

5.1.9 Consumables

Table 5.12: Consumables used in this thesis.

Consumable	Company
96-well plate, U-bottom (#268152)	Thermo Fisher Scientific
Borosilicate capillary glass (GC100TF-15)	Harvard Apparatus
Borosilicate capillary glass GC100F-10	Harvard Apparatus
Cell saver tips 200 µl	Biozym
Filter paper (#3030917)	Whatman
Filter Tips 10 µl, 20 µl, 200 µl, 1.25 ml	Starlab
Filter tips TipOne RPT 10 µl, 20 µl, 200 µl	Starlab
Gas permeable adhesive foil (moisture barrier seal 96) (4ti-0516/96)	4titude
Glass beads	Carl Roth
Glass dishes for microscopy (P35G-1.5-10-C)	MatTek
Glass vials	Carl Roth
Injection moulds	homemade
Latex gloves	SemperGuard
LDPE Pasteur pipettes 3.3 ml fine tip (LW4060)	Carl Roth
Micro pestles 0.5/1.5 ml	Laborversand Hartenstein
Micro pestles 1.5/2.0 ml	Eppendorf
Microloader tips	Eppendorf
Nitrile gloves	Starlab
Parafilm	Bemis
PCR lids	Sarstedt
PCR stripes	Sarstedt
PCR tubes	Kisker-Biotech
PE Pasteur pipettes 3 ml (2030/08)	Carl Roth
Petri dishes 60 mm	Greiner Bio-One
Petri dishes 90 mm	Sarstedt

Consumable	Company
Pipette tips 10 µl, 20 µl, 200 µl, 1.25 ml	Kisker
Pipette tips 1250 µl for multichannel pipette	Starlab
PP Tubes 15 ml, 50 ml	Sarstedt
Reaction tubes 1.5 ml, 2 ml	Sarstedt
Safe-lock tubes 1.5 ml, 2 ml	Eppendorf
Sandpaper 1000 grit	Bauhaus
Scalpel blades	Carl Roth
Serological pipettes 2 ml, 5 ml, 10 ml, 25 ml	Sarstedt

5.1.10 Equipment

Table 5.13: Equipment used in this thesis.

Equipment	Company
Bacterial Shaker INNOVA 44	New Brunswick Scientific
Centrifuges 5417C, 5425, 5430R, 5810R	Eppendorf
Desktop computer for visualization (Windows 8.1, Intel Xeon E5-2620 v3 (2x), 256 GB DDR3, NVIDIA GeForce GTX Titan X, 22 TB storage)	Custom
Digital Sight DS-Ri1	Nikon
DS-Fi2 camera	Nikon
FemtoJet express	Eppendorf
Fish incubators	Heraeus and RuMed
Forceps 110 mm, straight	NeoLab
Forceps 5, 55 Inox stainless steel	Dumont
Freezer -20°C	Liebherr
Freezer -80°C	Thermo Fisher Scientific
Fridge 4°C	Liebherr
Gel electrophoresis chambers and combs	homemade and Peqlab
Gel iX20 Imager	INTAS
HIVE data module	ACQUIFER Imaging GmbH
Imaging Machines IM03 and IM04	ACQUIFER Imaging GmbH
Incubator 32°C, 37°C, 60°C	Binder
InjectMan NI2	Eppendorf
Leica TCS SP8	Leica
MacBook Pro (2.7 GHz Intel Core i7, 16 GB 2133 MHz LPDDR3/Radeon Pro 455 2048 MB/Intel HD Graphics 530 1536 MB)	Apple

Equipment	Company
Microinjector 5242	Eppendorf
Micropipette Puller P-30	Sutter Instrument
Micropipette Puller P-97	Sutter Instrument
Microwave R-939	Sharp
Milli-Q (Q-POD)	Merck Millipore
Mini-centrifuge	Sarstedt
Minishaker DRS-12	Neolab
Multichannel pipette	Eppendorf
Multipette plus	Eppendorf
PCR C100 Touch Thermal Cycler	Bio-Rad
pH-Meter	Sartorius
PipetBoy	Gilson
Pipette 2 µl, 10 µl	Starlab
Pipettes 20 µl, 200 µl, 1 ml	Gilson
Power supply Power-PAC Basic	Bio-Rad
Rotating arm	homemade
Scale	Sartorius
Scale Extend ED224S	Sartorius
Shakers CAT S 20, DRS-12	NeoLab
Spectrophotometer DS-11FX+	DeNovix
Stereo microscope SZX7	Olympus
Stereomicroscope Stemi SV11	Zeiss
Stereomicroscope Zeiss Stemi 2000	Zeiss
Synology RS4017xs+	Synology
Synology RX1217RP	Synology
Thermomixer Compact	Eppendorf
Thermomixer F1.5	Eppendorf
UV transilluminator	Vilber Lourmat
Vortex Genie 2	Scientific Industries
Zeiss Axio Imager M1	Zeiss

5.1.11 Solutions and Buffers

Table 5.14: Buffer solutions for fish husbandry, handling, and genotyping.

Solution	Ingredients	Final concentration
ERM (10x) (Embryo Rearing Medium)	NaCl	17 mM
	KCl	0.4 mM
	$CaCl_2 \cdot 2\ H_2O$	0.27 mM
	$MgSO_4 \cdot 7\ H_2O$	0.66 mM
	HEPES pH 7.3	17 mM
	pH = 7	
Medaka hatching medium	Methylene blue trihydrate in 1x ERM	2 mg/l
Tricaine (20x)	Tricaine	4 g/l
	$Na_2HPO_4 \cdot 2\ H_2O$	10 g/l
	pH 7-7.5	
PFA stock solution (32%)	PFA	500 g
	Millipore H_2O	ad 1562 ml
PFA stock solution (16%)	32 % PFA stock	200 ml
	Millipore H_2O	ad 360 ml
	solve at 60°C, adjust to pH 7	with 1 M NaOH
	Millipore H_2O	ad 400 ml, store at 4°C
1-phenyl-2-thiourea (PTU) stock solution (50x)	N-Phenylthiourea Grade-I 98%	0.33 g
	Millipore H_2O	ad 200 ml
Zebrafish medium (E3)	Red sea salt	300 mg/l
Zebrafish hatching medium	Methylene blue trihydrate in 1x E3	1 mg/l

Table 5.15: Buffer solutions for Bacterial work.

Solution	Ingredients	Final concentration
Ampicillin solution	Ampicillin	100 mg/ml
	50% EtOH	
Kanamycin solution	Kanamycin	50 mg/ml
	Water	
LB agar Plates	Bacto-Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	10 g/l
	Agar	15 g/l

Solution	Ingredients	Final concentration
	Ampicillin or Kanamycin	100 µg/ml 50 µg/ml
LB-Medium	Bacto-Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	10 g/l
	Ampicillin or Kanamycin (freshly added)	100 µg/ml 50 µg/ml
TB-medium	Bacto-Tryptone	12 g/l
	Yeast extract	24 g/l
	Glycerol	0.4 %
	KH_2PO_4	2.13 g/l
	K_2HPO_4	12.54 g/l

Table 5.16: Solutions and buffers for work with DNA and RNA.

Solution	Ingredients	Composition
EtBr bath	EtBr in 1x TAE	0.0002% (v/v)
Fin-Clip buffer	Tris-HCl pH 8.5	100 mM
	EDTA pH 8	10 mM
	NaCl	200 mM
	SDS	2 %
Fin-Clip buffer with Proteinase K	Fin-Clip buffer	95%
	Proteinase K (20 mg/ml)	5 %
Oligo annealing buffer	Tris pH 7.5-8	10 mM
	NaCl	30 mM
Orange G loading dye (10x)	Orange G	2 mg/ml
	Glycerol	33% (v/v)
P1 buffer (Stored at 4°C)	Glucose	50 mM
	Tris-HCl (pH 8.0)	25 mM
	EDTA (pH 8.0)	10 mM
	RNase A	100 µg/ml
P2 buffer	NaOH	0.2 M
	SDS	1 % (w/v)
P3 buffer	CH3COOK	3 M
	pH 5.5	
Tris-Acetate-EDTA	Tris base	242 g/l

Solution	Ingredients	Composition
buffer (TAE) (50x)	Glacial acetic acid 99%	5.71 % (v/v)
	EDTA	50 mM
	pH 8.5	
2x RNA loading dye	Bromphenol blue	0.25 %
	Xylene cyanol	0.25 %
	SDS	0.025 %
	EDTA pH 8.0	5 mM
	Formamide	95 %

5.1.12 Software

Table 5.17: Software used in this thesis.

Software	Source
Adobe Illustrator	Adobe
Anaconda Navigator	Anaconda
CCTop	[102]
dplyr	[229]
Fiji	[230]
FileMaker Pro	FileMaker Inc.
Geneious	Biomatters Ltd.
ggplot2	[231]
ggpubr	[232]
GraphPad Prism	GraphPad Software Inc.
LAS X	Leica, Inc.
Microsoft Office	Microsoft
Papers 4	Digital Science Research & Solutions Inc.
Python	Python Software Foundation
R	[233]
RStudio	[234]
Spyder	Spyder Project Contributors
tidyverse	[235]

5.2 Methods

5.2.1 Fish husbandry

Fish are maintained in closed stocks in constantly recirculating systems at 28°C on a 14 h light/10 h dark cycle at Heidelberg University. Medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) husbandry (permit number 35–9185.64/BH Wittbrodt, Regierungspräsidium Karlsruhe) is performed according to local animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr. 1) in accordance with European Union animal welfare guidelines [236]. Mutant fish lines were generated under the authorizing permit number 35-9185.81/G-145/15 Wittbrodt. The fish facility is under the supervision of the local representative of the animal welfare agency. Embryos of medaka and of zebrafish were used at stages prior to hatching. Medaka were raised and maintained as previously described [237]. Prior to fin-clipping, adult medaka of the Cab strain were anesthetised in 0.01% Tricaine solution (Sigma-Aldrich, A5040-25G), and all efforts were made to minimize suffering.

5.2.2 Candidate gene selection

For the randomised gene selection, ten numbers between 1 and 23622 were generated using an online number generator, corresponding to the Ensembl IDs of annotated medaka coding genes in Ensembl [152] (Table 2.1). A random exon was then chosen for CRISPR/Cas9 targeting.

The genome-wide repository of associations between SNPs and phenotypes (GRASP v2.0) [156] was used for the targeted selection of human heart-GWAS (hGWAS) candidate genes. “Heart” and “Heart rate” were chosen in the search field as the respective categories for all heart- and heart rate-related phenotypes associated in human GWAS. Associations were filtered to include only coding SNPs, located in exons (i.e. SNP functional class = exons). Consequently, a list of candidate genes for the functional validation assay was compiled (Table 2.2). Uncharacterized genes, or genes with no prior experimental link to heart function were prioritized, however, some cardiac-relevant genes were selected as positive controls. The medaka orthologous gene for each human hGWAS candidate gene was identified using Ensembl [152]. For human genes with no medaka orthologue annotation in Ensembl, the medaka orthologue was identified by BLASTing the human protein sequence using the “tblastn” function of the NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Ensemble (http://www.ensembl.org/Multi/Tools/Blast) online tools. The region of interest (ROI) for CRISPR targeting on medaka genes was determined by aligning the medaka and human protein sequences in Geneious 8.1.9 (https://www.geneious.com). Primarily, ROIs surrounding the locus with the most significant phenotype-associated SNP (highest *p*-value of phenotype association) were highlighted for single guide RNA (sgRNA) design. For some genes with no direct overlap between the human SNP region and medaka gene,

another highly accessible (i.e. not surrounded by repeat regions) or conserved ROI was then chosen for targeting.

5.2.3 sgRNA target site selection

sgRNAs used in this thesis are all listed in Table 5.6. CCTop was used to design sgRNAs as previously described in [102]. Briefly, gene ROIs were pasted into the CCTop query, and initially NRG (*Streptococcus pyogenes*) PAM type was searched for. However, after finding out that NAG PAMs did not work in medaka *in vivo* (i.e. did not lead to Cas9-mediated double strand breaks), NGG (*Streptococcus pyogenes*) PAM type was specifically selected for. sgRNA target sites were selected based on number of potential off-target sites and their corresponding mismatches. Priority was given to sgRNAs with no off-target sites, otherwise sgRNAs with at least 3 nucleotide mismatches were selected. sgRNAs with only/mostly non-exonic off-target sites (i.e. intronic or intergenic) were preferably selected over sgRNAs with exonic off-target sites. Subsequently, for each sgRNA its corresponding oligo pair with 5' substitution was ordered. Cloning of sgRNA templates and *in vitro* transcription was performed as described [102].

5.2.4 sgRNA backbone T7 vector preparation

Template DR274 sgRNA backbone T7 vector plasmid #3632 (addgene #42250) was linearized by digestion with BsaI-HF (NEB) overnight at 37°C. Digest was run on 1% agarose/TAE gel and the respective linearized plasmid band (2147 bp) was extracted using innuPREP PCRpure Kit (Analytik Jena). Purified linearized vector was diluted to 40 ng/µl.

5.2.5 Oligonucleotide annealing

For each sgRNA, its respective oligonucleotide-pair listed in Table 5.3 were mixed in PCR tubes. 18 µl nuclease-free H_2O were mixed with 20 µl annealing buffer (10mM Tris-HCl pH 7.5-8; 30mM NaCl) and 1 µl of each 100 µM oligonucleotide. Oligonucleotide annealing was done in a thermocycler with the following program:

Table 5.18: Oligo annealing thermocycler program.

Temp. (°C)	Time
95°C ramp down to 70°C	0.1°C/sec
70°C hold	10 min
70°C ramp down to 65°C	0.1°C/sec
65°C hold	10 min
65°C ramp down to 60°C	0.1°C/sec
60°C hold	10 min
60°C ramp down to 10°C	0.1°C/sec

Annealed oligonucleotides were then diluted to 0.075 pmol/µl with nuclease-free H_2O.

5.2.6 Ligation of oligos into DR274 vector

Ligation reaction was set up as described in Table 5.19 and incubated at room temperature for 10 min.

Table 5.19: Ligation after oligonucleotide annealing reaction.

Component	Volume per 10 µl ligation reaction
10x T4 DNA ligation buffer	1 µl
0.075 pM annealed oligo product	1 µl
0.025 pM linearized vector	0.5 µl
T4 DNA ligase (5 U/µl)	1 µl
H_2O	ad 10 µl

5.2.7 Transformation of chemically competent cells

MachT1 chemically competent cells (50 µl) were thawed on ice and 3.5 µl ligated DNA were then added and mixed by pipetting. The mix was incubated on ice for 10 min, followed by a 42°C heat shock for 30 s. The mix was let to cool on ice for 2 min, 400 µl TB medium were then added and the tube was incubated for 1 hour at 37°C and 200 rpm. After incubation, the mix was spread onto pre-warmed LB plates containing the respective antibiotic. Plates were incubated at 37°C overnight, and single colonies were used for plasmid preparation.

5.2.8 Plasmid preparation

Minipreps

For sgRNA synthesis, an initial small-scale plasmid preparation (Miniprep) was performed. Single bacterial clones were inoculated in 4 ml LB medium containing the respective antibiotic and incubated overnight at 37°C and 200 rpm. 2 ml of the resulting bacterial culture was centrifuged in 2 ml tubes for 2 min at 12000 rpm. The supernatant was discarded and the pellet was resuspended in 200 µl P1 buffer (4°C). 200 µl P2 buffer were then added, and the contents were mixed by inverting the tube. 200 µl P3 buffer were subsequently added and tubes mixed again by inversion. Tubes were centrifuged for 20 min at 14000 rpm at 4°C, the resulting supernatant was transferred to a new 1.5 ml tube. 500 µl 2-Propanol were added and the tubes were thoroughly mixed by shaking and then centrifuged for 20 min at 14000 rpm at 4°C. After discarding the supernatant, the pellet was washed with 500 µl EtOH (70 %) and centrifuged for 5 min at 14000 rpm at 4°C. The supernatant was discarded and the pellet was left to dry before dissolving it in 50 µl nuclease-free water. After the Miniprep, templates for sgRNA *in vitro* transcription (IVT) were either prepared by a larger scale plasmid preparation ("Big Mini") or by direct amplification of DNA plasmid via Q5 PCR.

Big Minis

For the larger scale plasmid preparation ("Big Mini"), the QIAprep Spin Miniprep Kit was used. 20 ml LB medium containing the respective antibiotic were inoculated with 100 µl of the respective bacterial suspension used for the Miniprep mentioned above, and incubated at 37°C and 200 rpm overnight. Purification of plasmid DNA was performed according to manufacturer's instructions with the following changes: two 2 ml tubes were used per plasmid and centrifuged 2x each with 2 ml bacterial culture (therefore a total of 8 ml bacterial culture was used per plasmid preparation). The supernatants of both tubes were later merged into the same spin column. 50 µl nuclease-free water were used to elute the DNA. Afterwards, 10 µg plasmid DNA were digested with DraI-FD overnight at 37°C.

Direct PCR amplification

In order to save time, the necessity for DNA amplification via Big Mini for sgRNA IVT purposes could be circumvented by direct amplification of DNA plasmid from Minipreps. Eluted Miniprep DNA was diluted 1:1000 and directly used as a template for the PCR amplification reaction as follows:

Table 5.20: sgRNA IVT template via Q5 PCR

Component	Volume per 50 µl reaction
5x Q5 reaction buffer	10 µl
10 mM dNTP mix	1 µl
10 µM JW8829 DR274_sgRNAtemplate_F	1 µl
10 µM JW8830 DR274_sgRNAtemplate_R	1 µl
Mini Prep DNA (1 ng)	x µl
Q5 High Fidelity DNA Polymerase (2 U/µl)	0.2 µl
H_2O	ad 50 µl

Digested/amplified plasmid was run on a fresh 1.5% Agarose/TAE gel and the 300 bp band was extracted and purified using the innuPREP PCRpure Kit (Analytik Jena).

5.2.9 sgRNA *in vitro* transcription and clean up

50 – 150 ng purified sgRNA template were transcribed *in vitro* using MEGAscript T7 Transcription Kit (Thermo Fisher Scientific) according to manufacturers' protocol. After transcription, RNA sample was subjected to 1 µl TurboDNase (2 U/µl) for 15 min at 37°C, and then purified using RNeasy Mini Kit (QIAGEN) according to manufacturers' protocol. RNA quality was finally checked with a test gel and the concentration adjusted to 150 ng/µl.

All sgRNAs were initially tested after synthesis for successful *in vivo* targeting via pooled injections into medaka embryos, followed by genotyping (Sequencing or T7 Endonuclease I assay).

For testing of synthetic crRNA/tracrRNAs, crRNA was designed as to target the exact same locus as our in-house *oca2_sg58_T3* sgRNA, and ordered in addition to the tracrRNA component from Integrated DNA Technologies (IDT).

5.2.10 Microinjection

Injections were performed in medaka and zebrafish embryos at the one-cell stage. Only for CRISPR/Cas9-mediated targeting of *gfp* in *cmlc2::eGFP cmlc2::H2A-mCherry* transgenic reporter line were medaka embryos injected both at one or four-cell stage. Males were separated from females the night before injections. The next morning, fish were put together for mating and embryos were collected after 10-15 min. In the meantime, injection plates consisting of 9 cm petri dish filled with 1.5 % Agarose/H_2O with grooves created using an injection mould for embryo mounting and positioning. The plate was then covered with ice-cold 1x ERM (medaka) or E3 (zebrafish). Embryos were detached with two 10µl pipette tips and aligned into the grooves of the injection plate. Injection needles were freshly pulled from borosilicate glass capillaries with a needle puller prior to injection. Needles were filled with 2-3 µl injection mix and opened by scratching over the chorion. A pressure injector was used and depending on each needle opening, a standby pressure between 80-300 hPa (= P3) and an injection pressure between 700-1100 hPa (= P2) was set. The needle was inserted through the chorion into the cytoplasm of one-cell stage embryos and a small amount of the corresponding injection mixture (~10 % of the cell volume by visual assessment) was injected. After injections, embryos were raised in 1x ERM/E3 at 28°C. The following day, successfully injected embryos were screened for using the GFP injection tracer, and then raised in 1x ERM/E3 or Hatching medium at 28°C to appropriate stages for further analysis. Injection mix components are listed in Table 5.21. Exceptionally, for *gfp-rx2* tagging, a biotinylated PCR donor fragment was used as previously described in [108].

For testing of synthetic crRNA/tracrRNA duplex from IDT, the ordered crRNA (2 nmol) and tracrRNA (5 nmol) were resuspended in 20 and 50 µl IDT duplex buffer, respectively, according manufacturer's suggestions to end up with 100 pmol/µl stock solution. From that stock, two different working solutions (A and B) were prepared to compare efficiencies between (A) direct addition of crRNA/tracrRNA components individually into the injection mix without prior duplex formation (Table 5.22); and (B) pre-incubation of crRNA/tracrRNA components for duplex formation which is then added to the injection mix (Table 5.23). For working solution (A): 0.92 µl from original stock solution (100 pmol/µl) were diluted in 19.08 µl nuclease-free water to obtain a 4.6 pmol/µl working solution for each of crRNA and tracrRNA components. For working solution (B): 0.92 µl from original stock solution (100 pmol/µl) were diluted in 19.08 µl IDT duplex buffer to obtain a 4.6 pmol/µl working solution for each of crRNA and tracrRNA components. 5µl each of crRNA (B) and tracrRNA (B) were mixed in a tube for

crRNA:tracrRNA duplex formation by incubating at 95°C for 5 mins and then allowing the mix to cool to room temperature.

Table 5.21: Injection mixes used for Cas9 injections using sgRNAs.

Component	Final concentration
hei*Cas9* mRNA	150 ng/µl
sgRNAs	15 ng/µl (0.46 pmol/µl)
gfp mRNA injection tracer	10 ng/µl
(optional) biotinylated PCR donor fragment	5 ng/µl
RNase-free H_2O	ad 10 µl

Table 5.22: Injection mix (A) used for Cas9 injections using IDT crRNA/tracrRNAs

Component	Final concentration
hei*Cas9* mRNA	150 ng/µl
gfp mRNA injection tracer	10 ng/µl
crRNA	0.46 or 1.4 pmol/µl
tracrRNA	0.46 or 1.4 pmol/µl
RNase-free H_2O	ad 10 µl

Table 5.23: Injection mix (B) used for Cas9 injections using IDT crRNA/tracrRNAs

Component	Final concentration
hei*Cas9* mRNA	150 ng/µl
gfp mRNA injection tracer	10 ng/µl
crRNA:tracrRNA duplex	0.46 or 1.4 pmol/µl
RNase-free H_2O	ad 10 µl

5.2.11 Sample preparation for imaging

Embryo preparation for in vivo heart rate assay

One day prior to imaging (3 dpf), medaka embryos were transferred from methylene blue-containing medium into plain ERM and incubated at 28°C. On day of imaging, individual medaka embryos (4 dpf) were loaded using Cell-Saver tips to a 96 curved (U) bottom well plate (Nunc, Thermofisher #268152) with 150-200 µl ERM per well and sealed using gas-permeable adhesive foil (4titude, Wotton, UK, 4ti-0516/96).

Embryo preparation for in vivo confocal imaging

For the live confocal microscopy of the reporter lines, injected embryos were treated with 5x phenylthiourea (PTU) in ERM solution from 4 dpf onwards to avoid pigmentation. On the day of imaging (7 dpf), PTU solution was washed away with ERM, embryos were

rolled on fine sand paper to increase permeability for the hatching enzyme for embryo de-chorionation. De-chorionated embryos were then treated with 50 mM 2,3-butanedione monoxime (BDM) in 1x Tricaine solution until the heart beat was de-coupled (~40 mins), yielding fully dilated heart chambers. Embryos were then mounted on Matek dishes using 1.5% low-melting agarose with 85 mM BDM in 3x Tricaine solution. To avoid dehydration between and during imaging sessions, mounted samples were covered a solution contanining 30 mM BDM in 1x Tricaine.

5.2.12 Imaging

Documentation of medaka embryos and hatchlings

Images and videos were acquired on an SMZ18 stereomicroscope, embryos and hatchlings were mounted on plastic dishes in 3% methylcellulose in ERM. Hatchlings were sedated in 0.01% Tricaine solution before mounting. The settings of the DS-Fi2 camera were as follows: Mode: “Normal”; Resolution: “Fast (Focus): 1280 x 960 Normal”, “Quality (Capture): 1280 x 960 Normal - 8-bit”, and “High Quality Capture option checked”; Exposure times: “Auto” or “40-80 ms”; Analog gain: “2.0x”. AVI acquisition settings were as follows: Live: “Fast”; Duration: “10 s”; Size: “Original”, with MJPEG compression. Magnifications: 3-5.

Automated microscopy of fish embryos

Plates were automatically imaged using an ACQUIFER Imaging Machine (ACQUIFER GmbH, Pforzheim, Germany) as described in [72]. Medaka embryos (4 dpf) were imaged at 21 and 28°C with a 30-minute equilibration period before each measurement.

For heart rate assay-related experiments, images were acquired in brightfield illumination using 130 z-slices (dz = 0 µm) and a 2x Plan UW N.A. 0.06 objective (Nikon, Düsseldorf, Germany) to capture the centered embryo. Integration times were fixed with 80 % relative white LED intensity and 10 ms exposure time. Therefore, the whole 96-well plate was captured, with image sequences (videos) of entire microwells of ~10 seconds with 13 frames per second (fps). More details can be found in [72].

For the SLEDGE-Hammer related experiments, images were acquired in brightfield illumination using 9 z-slices (dz = 100 µm) and a 4x Plan UW N.A. 0.06 (Nikon, Düsseldorf, Germany) to capture the centered embryo. Integration times were fixed with 100% relative white LED intensity and 50 ms exposure time. GFP channel was used on rx2 specimens at 30% relative LED intensity and 200 ms exposure time.

Confocal in vivo imaging of medaka embryos

All confocal *in vivo* microscopy images were acquired at a Leica TCS SP8 with 10x dry or 20× oil objective, z-stacks of 200-300 µm were acquired with a z-step of 5 µm or 1 µm for 10x and 20x acquired images, respectively.

5.2.13 Image processing, data and statistical analysis

Heart rate quantifications were obtained using the *HeartBeat* software, details surrounding the images pre-processing, and principles of the *HeartBeat* sofrware are described in [72]. Images were processed via Fiji image processing software. Statistical analyses and data graphical presentations were performed using R [233], RStudio [234] and respective R packages listed in Table 5.17. Statistical tests used and the corresponding parameters and significances are indicated in their respective sections of the results. Figures and illustrations were done using Adobe Illustrator CS6.

5.2.14 Nucleic acid extraction

Standard lab protocol: Blin-Stafford followed by DNA precipitation

Individual or pooled embryos (2-10) or a piece of an adult tail fin are loaded into a tube with 100 µl Fin-Clip lysis buffer and 5 µl Proteinase K (20 mg/ml) and lysed by grinding using pestles. The lysate tubes were then incubated either over night at 50°C or at 60°C for at least 1 hour. 200 µl nuclease-free water are then added and mixed by inverting the tube several times. Afterwards, Proteinase K was inactivated by incubating the lysate at 95°C for at least 20 minutes. 200 µl supernatant were transferred to a new tube, to which 20 µl sodium acetate (3 M) and 600 µl absolute ethanol (100%) were added, and the tubes were then vigorously shaken. The tubes were then centrifuged for 30 mins at 14000 rpm at 4°C. The supernatant was discarded, residual ethanol was removed using a pipette tip and the pellet was air dried for 10 minutes before dissolving it in 30 µl 1x TE buffer.

SLEDGE-Hammer protocol

Initially before discovering the nucleic acid binding properties of the standard pipette tips, filter-in-tips were prepared by punching 2 mm filter discs from Whatman Cellulose paper (3030–917 Grade 3MM) using a small paper puncher (Harris Uni-Core 2.0). The filter discs were transferred to standard yellow 200 µl pipette tips, yielding our filter-in-tips. Since that finding, standard pipette tips were directly employed for this approach. For individual processing, specimens were transferred to 1.5 or 2 ml tubes with 50-100 µl Fin-Clip lysis buffer. The ratio of the lysis buffer to amount of material should be carefully considered to avoid viscous lysates, which can easily contaminate the pipette itself. For single embryos (stages older that 32), no less than 100 µl lysis buffer was used, and viscous lysates can be thinned out by adding nuclease-free water with discretion.

For high-throughput processing, samples (medaka and zebrafish embryos) were individually transferred to 96 U-well plates with 50 µl Fin-Clip lysis buffer. Fin-Clip lysis buffer (0.4 M Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 0.15 M NaCl, 0.1% SDS in Milli-Q water) was used for all samples except for *M. musculus* ear punches and *M. musculus* embryonic stem cell samples, where 1% SDS lysis buffer (0.1 M Tris-HCl pH 8.0, 5 mM EDTA pH 8.0, 0.2 M NaCl, 1% SDS in Milli-Q water) was used instead.

Individual embryos in tubes or 96-well plates were lysed by mechanical grinding using a clean pestle or the ‘hammer’, respectively. Before use, the ‘hammer’ is always pre-

cleaned by incubation in hypochlorite solution (1:10 dilution in water) for at least 15 mins followed by 5 mins incubation in Milli-Q water. After lysis, direct transfer of nucleic acids from lysate to prepared PCR mix was performed using the filter-in-tips or standard tips (individual or using a multi-channel pipettor for upscaled processing) as follows: 1) Binding: pipette in 25 – 50 µl lysate, wait ~10 seconds before releasing lysate back into the tube for storage. 2) Washing: washing pipette tip to remove residual lysis buffer by pipetting Milli-Q water in and out. One wash step is essential and sufficient, however with higher SDS concentration, more washing steps are required (at least 4 washing steps for 1% SDS lysates). 3) Elution: pipette in the prepared PCR reaction mix, wait ~10 seconds before releasing the mix back for amplification.

5.2.15 Polymerase chain reaction (PCR)

PCRs were carried out as described in Table 5.24. For genotyping of *scn4ab* F2 mutants, a Tri-Primer PCR approach was employed to distinguish embryos with a wild-type, heterozygous and homozygous mutant genotypic background.

Table 5.24: PCR mix for a 50 µl reaction.

Component	Volume per 50 µl reaction
5x Q5 reaction buffer	10 µl
10 mM dNTPs	1 µl
10 µM forward primer	1 µl
10 µM reverse primer	1 µl
10 µM Primer3 (optional)	1 µl
template DNA (< 1 µg)	variable
Q5 High Fidelity DNA Polymerase (2 U/µl)	0.2 µl
nuclease-free H_2O	ad 50 µl

PCR program ("Q5 genotyping") was used as described in Table 5.25. Respective annealing temperatures were individually calculated using the NEB Tm calculator online tool (http://tmcalculator.neb.com). Extension times were individually calculated as to allow 30 seconds per kilo base (kb) template.

Table 5.25: PCR "Q5 genotyping" thermal cycling program.

Step	Temp. (°C)	Time
Initial Denaturation	98°C	1 min
30x cycles	98°C	10 sec
	*anneal temperature	30 sec
	72°C	*extension time
Final extension	72°C	2 min
Cooling	12°C	5 min

After amplification, 10 µl of the amplified PCR product was loaded with 2 µl 6x orange loading dye on a 1 - 1.5% Agarose/1xTAE gel for verification.

5.2.16 Genotyping

After successful PCR amplification of the desired locus, genotyping was performed either via T7EI mismatch assay, or via Sanger sequencing of the PCR amplicon followed by downstream TIDE analysis [126], depending on the desired detail of allele information.

T7 Endonuclease I assay (T7EI)

For a quick qualitative confirmation of CRISPR-mediated targeting, the T7EI mismatch assay was usually employed as follows: 10 µl of un-purified amplified PCR product were transferred to a fresh PCR tube with 2 µl NEB Buffer 2 and 7.5 µl Milli-Q water. The tube was then incubated in the thermocycler running the "T7_EndoNucI" program described in Table 5.26 for heteroduplex formation.

Table 5.26: Heteroduplex formation using "T7_EndoNucI" thermal cycling program.

Step	Temp. (°C)	Ramp Rate	Time
Initial Denaturation	95°C		5 min
Annealing	95 - 85°C	-2°C/sec	10 sec
	85 - 25°C	-0.1°C/sec	20 min
Cooling	4°C		15 min

Afterwards, 0.5 µl T7 Endonuclease I (NEB) were added and the tube incubated at 37°C for 30 min. Samples were then loaded on 1 - 2% Agarose/TAE gels depending on the expected band sizes.

Sanger Sequencing and TIDE analysis

For more detailed insights into the CRISPR-edited alleles, the PCR reaction was either directly purified using either of PCR purification kits listed in Table 5.11 following the respective kits protocol. Alternatively, the PCR products were first run on 1 – 1.5% Agarose/TAE gels and the desired bands subsequently excised and extracted using either of Gel extraction kits listed in Table 5.11 following the respective kits protocol. The purified/extracted amplicons were then sequenced at Eurofins Genomics, and the resulting sequence files (.ab1) were uploaded on the TIDE online platform (http://shinyapps.datacurators.nl/tide/) [126] for the examination of alleles distributions using the default pre-set parameters, with exception of the "Indel size range" which was increased in the presence of InDels larger than 10 bp.

5.2.17 96-well "Hammer"

The hammer entails five components. A 6 mm aluminium top plate (material 3.3211) milled to size, drilled and countersunk to fit ISO 10642 screws and acts as a backstop for the pins. A second, 6 mm perforated aluminium plate (material 3.3211) milled to size. The holes on the second plate are drilled, reamed and countersunk on a CNC milling

machine to achieve the demanded tolerances. An aluminium handle (material 3.3211) turned to size, drilled and threads are cut on a lathe, with a knurled OD for better grip. Finally, 96 Stainless-steel pins (material 1.4301) made out of ISO 2338, which received a radius to fit the well ground. For proper alignment during the assembly, the pins are placed in a 96 U-well plate and properly aligned in the perforated plate then glued in with a solvent resistant adhesive. The handle and the perforated plate are then screwed to the top plate.

5.2.18 Agarose gel electrophoresis

6x orange loading dye was added to DNA samples and mixed before loading into wells of 1 – 1.5 % agarose/1x TAE gels placed in 1x TAE-filled gel chambers. Electrophoresis was performed at 90 – 130 V for 20 – 40 min depending on the gel and DNA fragment size. The agarose gel was subsequently stained by incubation for at least 10 min in an ethidium bromide bath. The stained gels were examined and imaged using a UV light-based (λ = 254 nm) gel documentation system.

5.2.19 Gel Extraction

DNA bands of interest were cut out of the agarose gel with a clean scalpel. Subsequent DNA extraction was performed using either of gel extraction kits listed in Table 5.11 according to respective manufacturer's protocol. The DNA was finally eluted with 20 - 30 µl pre-warmed nuclease-free H_2O.

5.2.20 Sanger sequencing

Mini prep plasmid DNA and purified PCR amplicons were sent to Eurofins Genomics for sequencing following the respective instructions listed on their website.

5.2.21 DNA restriction

0.5 µl of the respective restriction enzymes, 1x buffer (dependent on the restriction enzymes) and 3–5 µg of DNA were mixed and incubated for at least 1 h at 37°C. For test restriction digests, 5 µl Miniprep were digested with 0.2 µl restriction enzyme. The products of restriction digests were subsequently analysed by gel electrophoresis.

References

1. Virani, S. S. *et al.* Heart Disease and Stroke Statistics—2020 Update. *Circulation* 141, e139–e596 (2020).

2. Ghebre, Y. T. *et al.* Vascular Aging: Implications for Cardiovascular Disease and Therapy. *Transl Medicine* 06, (2016).

3. Magyar, K. From hypertension to heart failure. *World J Hypertens* 5, 85 (2015).

4. Lawes, C. M., Hoorn, S. V., Rodgers, A. & Hypertension, for the I. S. of. Global burden of blood-pressure-related disease, 2001. *Lancet* 371, 1513–1518 (2008).

5. Podrez, E. A. *et al.* Platelet CD36 links hyperlipidemia, oxidant stress and a prothrombotic phenotype. *Nat Med* 13, 1086–1095 (2007).

6. Kolloch, R. *et al.* Impact of resting heart rate on outcomes in hypertensive patients with coronary artery disease: findings from the INternational VErapamil-SR/trandolapril STudy (INVEST). *Eur Heart J* 29, 1327–1334 (2008).

7. DYER, A. R. *et al.* HEART RATE AS A PROGNOSTIC FACTOR FOR CORONARY HEART DISEASE AND MORTALITY: FINDINGS IN THREE CHICAGO EPIDEMIOLOGIC STUDIES. *Am J Epidemiol* 112, 736–749 (1980).

8. Gillum, R. F., Makuc, D. M. & Feldman, J. J. Pulse rate, coronary heart disease, and death: The NHANES I Epidemiologic Follow-up Study. *American Heart Journal* 121, 172–177 (1991).

9. Diaz, A., Bourassa, M. G., Guertin, M.-C. & Tardif, J.-C. Long-term prognostic value of resting heart rate in patients with suspected or proven coronary artery disease. *Eur Heart J* 26, 967–974 (2005).

10. Böhm, M. *et al.* Heart rate as a risk factor in chronic heart failure (SHIFT): the association between heart rate and outcomes in a randomised placebo-controlled trial. *The Lancet* 376, 886–894 (2010).

11. Nauman, J., Janszky, I., Vatten, L. J. & Wisløff, U. Temporal Changes in Resting Heart Rate and Deaths From Ischemic Heart Disease. *Jama* 306, 2579–2587 (2011).

12. Eppinga, R. N. *et al.* Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nature Genetics* 48, 1557–1563 (2016).

13. Collins, R. *et al.* Blood pressure, stroke, and coronary heart disease Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 335, 827–838 (1990).

14. Wang, J.-G., Staessen, J. A., Gong, L. & Liu, L. Chinese Trial on Isolated Systolic Hypertension in the Elderly. *Arch Intern Med* 160, 211–220 (2000).

15. Verdecchia, P. *et al.* Angiotensin-Converting Enzyme Inhibitors and Calcium Channel Blockers for Coronary Heart Disease and Stroke Prevention. *Hypertension* 46, 386–392 (2005).

16. Castilla-Guerra, L. & Fernández-Moreno, M. del C. Update on the Management of Hypertension for Secondary Stroke Prevention. *Eur Neurol* 68, 1–7 (2012).

17. Wierzbicki, A. S., Poston, R. & Ferro, A. The lipid and non-lipid effects of statins. *Pharmacol Therapeut* 99, 95–112 (2003).

18. Collaboration, C. T. T. (CTT). Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. *Lancet* 376, 1670–1681 (2010).

19. Barter, P. J. & Rye, K.-A. New Era of Lipid-Lowering Drugs. *Pharmacol Rev* 68, 458–475 (2016).

20. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34, 154–156 (2003).

21. Koren, M. J. *et al.* Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 as monotherapy in patients with hypercholesterolaemia (MENDEL): a randomised, double-blind, placebo-controlled, phase 2 study. *Lancet* 380, 1995–2006 (2012).

22. Giugliano, R. P. *et al.* Efficacy, safety, and tolerability of a monoclonal antibody to proprotein convertase subtilisin/kexin type 9 in combination with a statin in patients with hypercholesterolaemia (LAPLACE-TIMI 57): a randomised, placebo-controlled, dose-ranging, phase 2 study. *Lancet* 380, 2007–2017 (2012).

23. Beere, P. A., Glagov, S. & Zarins, C. K. Retarding effect of lowered heart rate on coronary atherosclerosis. *Science* 226, 180 (1984).

24. Beere, P. A., Glagov, S. & Zarins, C. K. Experimental atherosclerosis at the carotid bifurcation of the cynomolgus monkey. Localization, compensatory enlargement, and the sparing effect of lowered heart rate. *Arteriosclerosis Thrombosis J Vasc Biology* 12, 1245–1253 (1992).

25. Hopkins, P. N. & Hunt, S. C. Genetics of hypertension. *Genet Med* 5, 413–429 (2003).

26. Cambien, F. & Tiret, L. Genetics of Cardiovascular Diseases. *Circulation* 116, 1714–1724 (2007).

27. Kathiresan, S. & Srivastava, D. Genetics of Human Cardiovascular Disease. *Cell* 148, 1242–1257 (2012).

28. Jarauta, E., Bea-Sanz, A. M., Marco-Benedi, V. & Lamiquiz-Moneo, I. Genetics of Hypercholesterolemia: Comparison Between Familial Hypercholesterolemia and Hypercholesterolemia Nonrelated to LDL Receptor. *Frontiers Genetics* 11, 554931 (2020).

29. A, L., M. *et al.* Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* 227, 140–146 (1985).

30. Stainier, D. Y. *et al.* Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* 123, 285–292 (1996).

31. Ding, Y. *et al.* Trapping Cardiac Recessive Mutants via Expression-Based Insertional Mutagenesis Screening. *Circ Res* 112, 606–617 (2013).

32. Cordell, H. J. *et al.* Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot. *Hum Mol Genet* 22, 1473–1481 (2013).

33. Villard, E. *et al.* A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J* 32, 1065–1076 (2011).

34. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* 41, 677–687 (2009).

35. Saxena, R. *et al.* Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science* 316, 1331–1336 (2007).

36. Sandhu, M. S. *et al.* LDL-cholesterol concentrations: a genome-wide association study. *Lancet* 371, 483–491 (2008).

37. Eijgelsheim, M. *et al.* Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet* 19, 3885–3894 (2010).

38. Newton-Cheh, C. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature Genetics* 41, 399–406 (2009).

39. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nature Genetics* 1–11 (2010).

40. Hoed, M. den *et al.* Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders. *Nature Publishing Group* 45, 621–631 (2013).

41. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 337, 1190–1195 (2012).

42. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet* 21, 3255–3263 (2012).

43. Gloss, B. S. & Dinger, M. E. Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Medicine* 50, 1–8 (2018).

44. Perenthaler, E., Yousefi, S., Niggl, E. & Barakat, T. S. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. *Front Cell Neurosci* 13, 352 (2019).

45. Duun, R., Palle *et al.* Functional Validation of Candidate Genes Detected by Genomic Feature Models. *G3 Genes Genomes Genetics* 8, 1659 (2018).

46. Baranski, T. J. *et al.* A high throughput, functional screen of human Body Mass Index GWAS loci using tissue-specific RNAi Drosophila melanogaster crosses. *PLOS Genetics* 14, e1007222-18 (2018).

47. Dooley, K. & Zon, L. I. Zebrafish: a model system for the study of human disease. *Curr Opin Genet Dev* 10, 252–256 (2000).

48. Wittbrodt, J., Shima, A. & Schartl, M. MEDAKA — A MODEL ORGANISM FROM THE FAR EAST. *Nature Reviews Genetics* 3, 53–64 (2002).

49. Schartl, M. Beyond the zebrafish: diverse fish species for modeling human disease. *Disease Models & Mechanisms* 7, 181–192 (2014).

50. Kirchmaier, S., Naruse, K., Wittbrodt, J. & Loosli, F. The Genomic and Genetic Toolbox of the Teleost Medaka (Oryzias latipes). *Genetics* 199, 905–918 (2015).

51. Gut, P., Reischauer, S., Stainier, D. Y. R. & Arnaout, R. Little Fish, Big Data: Zebrafish as a Model for Cardiovascular and Metabolic Disease. *Physiol Rev* 97, 889–938 (2017).

52. Loosli, F. *et al.* A genetic screen for mutations affecting embryonic development in medaka fish (Oryzias latipes). *Mech Develop* 97, 133–139 (2000).

53. Oxendine, S. L., Cowden, J., Hinton, D. E. & Padilla, S. Adapting the medaka embryo assay to a high-throughput approach for developmental toxicity testing. *Neurotoxicology* 27, 840–845 (2006).

54. Lessman, C. A. & Tuan, R. S. The developing zebrafish (Danio rerio): A vertebrate model for high-throughput screening of chemical libraries. *Birth Defects Research Part C: Embryo Today: Reviews* 93, 268–280 (2011).

55. Spomer, W., Pfriem, A., Alshut, R., Just, S. & Pylatiuk, C. High-Throughput Screening of Zebrafish Embryos Using Automated Heart Detection and Imaging. *J Laboratory Automation* 17, 435–442 (2012).

56. L., Y., Krystle, M., I., Gregory, D., R., Tara & C., V., David. High-Content Screening Assay for Identification of Chemicals Impacting Cardiovascular Function in Zebrafish Embryos. *Environ Sci Technol* 47, 11302–11310 (2013).

57. Kithcart, A. & MacRae, C. A. Using Zebrafish for High-Throughput Screening of Novel Cardiovascular Drugs. *Jacc Basic Transl Sci* 2, 1–12 (2017).

58. Gehrig, J., Pandey, G. & Westhoff, J. H. Zebrafish as a Model for Drug Screening in Genetic Kidney Diseases. *Frontiers Pediatrics* 6, 183 (2018).

59. Westhoff, J. H. *et al.* In vivo High-Content Screening in Zebrafish for Developmental Nephrotoxicity of Approved Drugs. *Frontiers Cell Dev Biology* 8, 583 (2020).

60. Hyodo-Taguchi, Y. & Egami, N. Establishment of inbred strains of the medaka Oryzias latipes and the usefulness of the strains for biomedical research. *Zoological Science* 2, (1985).

61. Spivakov, M. *et al.* Genomic and Phenotypic Characterization of a Wild Medaka Population: Towards the Establishment of an Isogenic Population Genetic Resource in Fish. *G3: Genes|Genomes|Genetics* 4, 433–445 (2014).

62. Leger, A. *et al.* Genomic variations and epigenomic landscape of the Medaka Inbred Kiyosu-Karlsruhe (MIKK) panel. *Biorxiv* 2021.05.17.444424 (2021) doi:10.1101/2021.05.17.444424.

63. Fitzgerald, T. *et al.* The Medaka Inbred Kiyosu-Karlsruhe (MIKK) Panel. *Biorxiv* 2021.05.17.444412 (2021) doi:10.1101/2021.05.17.444412.

64. Nemtsas, P., Wettwer, E., Christ, T., Weidinger, G. & Ravens, U. Adult zebrafish heart as a model for human heart? An electrophysiological study. *Journal of Molecular and Cellular Cardiology* 48, 161–171 (2010).

65. Staudt, D. & Stainier, D. Uncovering the Molecular and Cellular Mechanisms of Heart Development Using the Zebrafish. *Annu Rev Genet* 46, 397–418 (2012).

66. Liu, C. C., Li, L., Lam, Y. W., Siu, C. W. & Cheng, S. H. Improvement of surface ECG recording in adult zebrafish reveals that the value of this model exceeds our expectation. *Sci Rep-uk* 6, 25073 (2016).

67. Lin, M.-H. *et al.* Development of a rapid and economic in vivoelectrocardiogram platform for cardiovascular drug assay and electrophysiology research in adult zebrafish. *Sci Rep-uk* 8, 15986 (2018).

68. Yonekura, M. *et al.* Medaka as a model for ECG analysis and the effect of verapamil. *J Pharmacol Sci* 137, 55–60 (2018).

69. Bowley, G. *et al.* Zebrafish as a tractable model of human cardiovascular disease. *Brit J Pharmacol* (2021) doi:10.1111/bph.15473.

70. Cutie, S. & Huang, G. N. Vertebrate cardiac regeneration: evolutionary and developmental perspectives. *Cell Regen* 10, 6 (2021).

71. P, D., J., Borne,P, van de, Linkowski,P & Cauter,E, V. Quantitative analysis of the 24-hour blood pressure and heart rate patterns in young men. *Hypertension* 18, 199–210 (1991).

72. Gierten, J. *et al.* Automated high-throughput heartbeat quantification in medaka and zebrafish embryos under physiological conditions. *Scientific Reports* 1–12 (2020).

73. Stainier, D. Y., Lee, R. K. & Fishman, M. C. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Dev Camb Engl* 119, 31–40 (1993).

74. Y.R., S., Didier & C., F., Mark. The zebrafish as a model system to study cardiovascular development. *Trends Cardiovas Med* 4, 207–212 (1994).

75. B., K., Charles, W., B., William, R., K., Seth, Bonnie, U. & F., S., Thomas. Stages of embryonic development of the zebrafish. *Dev Dynam* 203, 253–310 (1995).

76. S, G., Nathalia & Deborah, Y. Cardiac development in zebrafish: coordination of form and function. *Semin Cell Dev Biol* 13, 507–513 (2002).

77. Miura, G. I. & Yelon, D. A Guide to Analysis of Cardiac Phenotypes in the Zebrafish Embryo. in vol. 101 161–180 (2011).

78. Collins, M. M. *et al.* Early sarcomere and metabolic defects in a zebrafish pitx2c cardiac arrhythmia model. *Proc National Acad Sci* 116, 24115–24121 (2019).

79. Iwamatsu, T. Stages of normal development in the medaka Oryzias latipes. *Mechanisms of Development* 121, 605–618 (2004).

80. Milan, D. J., Peterson, T. A., Ruskin, J. N., Peterson, R. T. & MacRae, C. A. Drugs That Induce Repolarization Abnormalities Cause Bradycardia in Zebrafish. *Circulation* 107, 1355–1358 (2003).

81. Finn, J. *et al.* Effects of propranolol on heart rate and development in Japanese medaka (Oryzias latipes) and zebrafish (Danio rerio). *Aquat Toxicol* 122, 214–221 (2012).

82. Puybareau, E., Genest, D., Barbeau, E., Léonard, M. & Talbot, H. An automated assay for the assessment of cardiac arrest in fish embryo. *Comput Biol Med* 81, 32–44 (2017).

83. Guerra, A. *et al.* Distinct myocardial lineages break atrial symmetry during cardiogenesis in zebrafish. *Elife* 7, e32833 (2018).

84. Chan, P. K., Lin, C. C. & Cheng, S. H. Noninvasive technique for measurement of heartbeat regularity in zebrafish (Danio rerio) embryos. *Bmc Biotechnol* 9, 11 (2009).

85. Pylatiuk, C. *et al.* Automatic Zebrafish Heartbeat Detection and Analysis for Zebrafish Embryos. *Zebrafish* 11, 379–383 (2014).

86. Luca, E. D. *et al.* ZebraBeat: a flexible platform for the analysis of the cardiac rate in zebrafish embryos. *Sci Rep-uk* 4, 4898 (2014).

87. Huang, W.-C. *et al.* Combined Use of MS-222 (Tricaine) and Isoflurane Extends Anesthesia Time and Minimizes Cardiac Rhythm Side Effects in Adult Zebrafish. *Zebrafish* 7, 297–304 (2010).

88. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636–646 (2010).

89. Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther* 5, 1 (2020).

90. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Bio* 14, 49–55 (2013).

91. O'Driscoll, M. & Jeggo, P. A. The role of double-strand break repair — insights from human genetics. *Nat Rev Genet* 7, 45–54 (2006).

92. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology+* 151, 2551–2561 (2005).

93. 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, 653–663 (2005).

94. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821 (2012).

95. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nat Rev Microbiol* 9, 467–477 (2011).

96. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* 13, 722–736 (2015).

97. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827–832 (2013).

98. Wenyan, J., David, B., David, C., Feng, Z. & A, M., Luciano. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31, 233 (2013).

99. Yilan, Z. *et al.* Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep-uk* 4, 5405 (2014).

100. Zhang, J.-H. *et al.* Improving the specificity and efficacy of CRISPR/CAS9 and gRNA through target specific DNA reporter. *J Biotechnol* 189, 1–8 (2014).

101. Thumberger, T. *et al.* hei-tag: a highly efficient tag to boost targeted genome editing. *Biorxiv* 2021.05.27.445956 (2021) doi:10.1101/2021.05.27.445956.

102. Stemmer, M. *et al.* CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS ONE* 10, e0124633-11 (2015).

103. Cui, Y., Xu, J., Cheng, M., Liao, X. & Peng, S. Review of CRISPR/Cas9 sgRNA Design Tools. *Interdiscip Sci Comput Life Sci* 10, 455–465 (2018).

104. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819–823 (2013).

105. Shah, A. N., Davey, C. F., Whitebirch, A. C., Miller, A. C. & Moens, C. B. Rapid reverse genetic screening using CRISPR in zebrafish. *Nature Methods* 12, 535–540 (2015).

106. Wu, R. S. *et al.* A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. *Dev Cell* 46, 112-125.e4 (2018).

107. Lischik, C. Q., Winkler, C., Adelmann, L. & Wittbrodt, J. Enhanced in vivo-imaging in medaka by optimized anaesthesia, fluorescent protein selection and removal of pigmentation. *PLoS ONE* 14, e0212956-19 (2019).

108. Gutierrez-Triana, J. A. *et al.* Efficient single-copy HDR by 5' modified long dsDNA donors. *eLife* 7, 1–15 (2018).

109. Seleit, A., Aulehla, A. & Paix, A. Endogenous protein tagging in medaka using a simplified CRISPR/Cas9 knock-in approach. *Biorxiv* 2021.07.29.454295 (2021) doi:10.1101/2021.07.29.454295.

110. Carroll, K. J. *et al.* A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proc National Acad Sci* 113, 338–343 (2016).

111. Sevinç, G. *et al.* A patient-based medaka alg2 mutant as a model for hypo-N-glycosylation. *Development* 148, (2021).

112. Ding, Q. *et al.* Permanent Alteration of PCSK9 With In Vivo CRISPR-Cas9 Genome Editing. *Circ Res* 115, 488–492 (2014).

113. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491 (2014).

114. Kweon, J. & Kim, Y. High-throughput genetic screens using CRISPR–Cas9 system. *Archives of Pharmacal Research* 1–10 (2018).

115. Covarrubias, S. *et al.* High-Throughput CRISPR Screening Identifies Genes Involved in Macrophage Viability and Inflammatory Pathways. *Cell Reports* 33, 108541 (2020).

116. Bennett, E. P. *et al.* INDEL detection, the 'Achilles heel' of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Res* 48, gkaa975- (2020).

117. A., O., Catherine, R., B. M., Colleen, K., G., Andrew & T., Y., Anthony. Mutation detection using a novel plant endonuclease. *Nucleic Acids Res* 26, 4597 (1998).

118. Peter, Q. *et al.* Mutation detection using Surveyor™ nuclease. *Biotechniques* 36, 702 (2004).

119. Maxim, P. *et al.* Recombinant nucleases CEL I from celery and SP I from spinach for mutation detection. *Bmc Biotechnol* 7, 29 (2007).

120. D., M., Robert, Jason, K. & Jeffrey, S. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat Genet* 9, 177 (1995).

121. Thomsen, N., Nichols, J., Ali, R. G., Ahmed, J. N. & Arkell, R. M. High Resolution Melt Analysis (HRMA); a Viable Alternative to Agarose Gel Electrophoresis for Mouse Genotyping. *PLoS ONE* 7, e45252-7 (2012).

122. 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).

123. Yang, Z. *et al.* Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res* 43, e59–e59 (2015).

124. Mock, U., Hauber, I. & Fehse, B. Digital PCR to assess gene-editing frequencies (GEF-dPCR) mediated by designer nucleases. *Nat Protoc* 11, 598–615 (2016).

125. Bell, C. C., Magor, G. W., Gillinder, K. R. & Perkins, A. C. A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. *Bmc Genomics* 15, 1002 (2014).

126. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42, e168–e168 (2014).

127. Hsiau, T. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *Biorxiv* 251082 (2019) doi:10.1101/251082.

128. Ali, N., Rampazzo, R. de C. P., Costa, A. D. T. & Krieger, M. A. Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics. *Biomed Res Int* 2017, 1–13 (2017).

129. Arthur, C. L. & Pawliszyn, J. Solid phase microextraction with thermal desorption using fused silica optical fibers. *Anal Chem* 62, 2145–2148 (1990).

130. Meng, L. & Feldman, L. A rapid TRIzol-based two-step method for DNA-free RNA extraction from Arabidopsis siliques and dry seeds. *Biotechnol J* 5, 183–186 (2010).

131. Chomczynski, P. & Rymaszewski, M. Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *Biotechniques* 40, 454–458 (2006).

132. N, B. & W, S., D. A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res* 3, 2303 (1976).

133. Hsiao, K. M. *et al.* Application of FTA® sample collection and DNA purification system on the determination of CTG trinucleotide repeat size by PCR-based southern blotting. *Journal of Clinical Laboratory Analysis* 13, 188–193 (1999).

134. Jangam, S. R., Yamada, D. H., McFall, S. M. & Kelso, D. M. Rapid, Point-of-Care Extraction of Human Immunodeficiency Virus Type 1 Proviral DNA from Whole Blood for Detection by Real-Time PCR ▿ . *J Clin Microbiol* 47, 2363–2368 (2009).

135. Kim, J. *et al.* A PCR reactor with an integrated alumina membrane for nucleic acid isolation. *Analyst* 135, 2408–2414 (2010).

136. McFall, S. M. *et al.* A simple and rapid DNA extraction method from whole blood for highly sensitive detection and quantitation of HIV-1 proviral DNA by real-time PCR. *Journal of Virological Methods* 214, 37–42 (2015).

137. Zou, Y. *et al.* Nucleic acid purification from plants, animals and microbes in under 30 seconds. *PLOS Biology* 15, e2003916-22 (2017).

138. Masaru, M. *et al.* DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417, 559 (2002).

139. Hammouda, O. T., Böttger, F., Wittbrodt, J. & Thumberger, T. Swift Large-scale Examination of Directed Genome Editing. *PLoS ONE* 14, e0213317-11 (2019).

140. Alexa, B. *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* 143, 2025 (2016).

141. Berensmeier, S. Magnetic particles for the separation and purification of nucleic acids. *Appl Microbiol Biot* 73, 495–504 (2006).

142. Arturo, G.-T., Jose *et al.* Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* 7, e39468 (2018).

143. DYER, A. R. *et al.* HEART RATE AS A PROGNOSTIC FACTOR FOR CORONARY HEART DISEASE AND MORTALITY: FINDINGS IN THREE CHICAGO EPIDEMIOLOGIC STUDIES. *American Journal of Epidemiology* 112, 736–749 (1980).

144. Hammouda, O. T., Wu, M. Y., Kaul, V., Thumberger, T. & Wittbrodt, J. In vivo identification and validation of novel potential predictors for human cardiovascular diseases. *bioRxiv* 2021.02.03.429563 (2021).

145. Fukamachi, S. *et al.* Conserved function of medaka pink-eyed dilution in melanin synthesis and its divergent transcriptional regulation in gonads among vertebrates. *Genetics* 168, 1519–1527 (2004).

146. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* 12, 823–826 (2015).

147. Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res* 44, 6482–6492 (2016).

148. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* 7, 11750 (2016).

149. Jensen, K. T. *et al.* Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *Febs Lett* 591, 1892–1901 (2017).

150. Zakariyah, A. F., Rajgara, R. F., Veinot, J. P., Skerjanc, I. S. & Burgon, P. G. Congenital heart defect causing mutation in Nkx2.5 displays in vivo functional deficit. *J Mol Cell Cardiol* 105, 89–98 (2017).

151. Targoff, K. L. *et al.* Nkx genes are essential for maintenance of ventricular identity. *Development* 140, 4203–4213 (2013).

152. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Research* 48, D682–D688 (2020).

153. Qian, L. *et al.* Tinman/Nkx2-5 acts via miR-1 and upstream of Cdc42 to regulate heart function across species. *J Cell Biology* 193, 1181–1196 (2011).

154. Li, J. *et al.* Essential role of Cdc42 in cardiomyocyte proliferation and cell-cell adhesion during heart development. *Dev Biol* 421, 271–283 (2017).

155. Thanassoulis, G. *et al.* Genetic Associations with Valvular Calcification and Aortic Stenosis. *New Engl J Medicine* 368, 503–512 (2013).

156. Eicher, J. D. *et al.* GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Research* 43, D799–D804 (2015).

157. Itoh-Satoh, M. *et al.* Titin Mutations as the Molecular Basis for Dilated Cardiomyopathy. *Biochemical and Biophysical Research Communications* 291, 385–393 (2002).

158. Park, C. Y. *et al.* skNAC, a Smyd1-interacting transcription factor, is involved in cardiac development and skeletal muscle growth and regeneration. *Proc National Acad Sci* 107, 20750–20755 (2010).

159. Faggioni, M. & Knollmann, B. C. Calsequestrin 2 and arrhythmias. *American Journal of Physiology-Heart and Circulatory Physiology* 302, H1250–H1260 (2012).

160. Gianulis, E. C. & Trudeau, M. C. Rescue of Aberrant Gating by a Genetically Encoded PAS (Per-Arnt-Sim) Domain in Several Long QT Syndrome Mutant Human Ether-á-go-go-related Gene Potassium Channels*. *J Biol Chem* 286, 22160–22169 (2011).

161. Huttner, I. G. *et al.* A transgenic zebrafish model of a human cardiac sodium channel mutation exhibits bradycardia, conduction-system abnormalities and early death. *J Mol Cell Cardiol* 61, 123–132 (2013).

162. Zaklyazminskaya, E. & Dzemeshkevich, S. The role of mutations in the SCN5A gene in cardiomyopathies. *Biochimica Et Biophysica Acta Bba - Mol Cell Res* 1863, 1799–1805 (2016).

163. Benson, M. A. *et al.* Ryanodine receptors are part of the myospryn complex in cardiac muscle. *Scientific Reports* 1–12 (2017).

164. Yamaguchi, T. *et al.* The CCR4-NOT deadenylase complex controls Atg7-dependent cell death and heart function. *Sci Signal* 11, eaan3638 (2018).

165. Lisa, E. *et al.* Silencing of CCR4-NOT complex subunits affects heart structure and function. *Dis Model Mech* 13, dmm044727 (2020).

166. Rathjens, F. S. *et al.* Preclinical evidence for the therapeutic value of TBX5 normalization in arrhythmia control. *Cardiovascular Research* (2020).

167. Jemima, B., Deepali, B., Karin, R. & Susan, F.-N. TRAPP complexes in membrane traffic: convergence through a common Rab. *Nat Rev Mol Cell Bio* 11, 759–763 (2010).

168. Li, Y. *et al.* Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. *Genome Res* 30, gr.258871.119 (2020).

169. Lambert, C. J. *et al.* An automated system for rapid cellular extraction from live zebrafish embryos and larvae: Development and application to genotyping. *Plos One* 13, e0193180 (2018).

170. Zhang, X., Zhang, Z., Zhao, Q. & Lou, X. Rapid and Efficient Live Zebrafish Embryo Genotyping. *Zebrafish* 17, 56–58 (2020).

171. F., S., Monica, T., P., Samuel, P., F., Colin, P., C., Jon & M., P.-M., Shondra. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep-uk* 8, 888 (2018).

172. Chen, S. X., Zhang, D. Y. & Seelig, G. Conditionally fluorescent molecular probes for detecting single base changes in double-stranded DNA. *Nat Chem* 5, 782–789 (2013).

173. Xie, N. G. *et al.* High-throughput variant detection using a color-mixing strategy in real-time PCR reactions. *Biorxiv* 2021.06.30.450651 (2021) doi:10.1101/2021.06.30.450651.

174. Saba, P. *et al.* MIC-Drop: A platform for large-scale in vivo CRISPR screens. *Science* 373, 1146–1151 (2021).

175. Maurice, L. *et al.* Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res* 46, 1375–1385 (2018).

176. Höijer, I. *et al.* Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. *Genome Biol* 21, 290 (2020).

177. Heyde, B. von der *et al.* Translating GWAS-identified loci for cardiac rhythm and rate using an in vivo image- and CRISPR/Cas9-based approach. *Sci Rep-uk* 10, 11831 (2020).

178. Meyer, H. V. *et al.* Genetic and functional insights into the fractal structure of the heart. *Nature* 1–25 (2020).

179. Furtado, M. B. *et al.* Point mutations in murine Nkx2-5 phenocopy human congenital heart disease and induce pathogenic Wnt signaling. *Jci Insight* 2, e88271 (2017).

180. Divers, J. *et al.* Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes. *Bmc Genet* 18, 105 (2017).

181. Feitosa, M. F. *et al.* Gene discovery for high-density lipoprotein cholesterol level change over time in prospective family studies. *Atherosclerosis* 297, 102–110 (2020).

182. Wang, Z. *et al.* The impact of growth differentiation factor 15 on the risk of cardiovascular diseases: two-sample Mendelian randomization study. *Bmc Cardiovasc Disor* 20, 462 (2020).

183. Harrington, J. K., Sorabella, R., Tercek, A., Isler, J. R. & Targoff, K. L. Nkx2.5 is essential to establish normal heart rate variability in the zebrafish embryo. *Am J Physiology-regulatory Integr Comp Physiology* 313, R265–R271 (2017).

184. Tsoupri, E. *et al.* Myospryn deficiency leads to impaired cardiac structure and function and schizophrenia-associated symptoms. *Cell Tissue Res* 1–22 (2021) doi:10.1007/s00441-021-03447-2.

185. Fung, K. L. & Gottesman, M. M. A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function. *Biochimica Et Biophysica Acta Bba - Proteins Proteom* 1794, 860–871 (2009).

186. Ishibashi, K., Matsuzaki, T., Takata, K. & Imai, M. Identification of a New Member of Type I Na/Phosphate Co-Transporter in the Rat Kidney. *Nephron Physiol* 94, p10–p18 (2003).

187. Ho, R. X.-Y. *et al.* MINAR1 is a Notch2-binding protein that inhibits angiogenesis and breast cancer growth. *J Mol Cell Biol* 10, 195–204 (2018).

188. Mehrdad, K. *et al.* Identification of Padi2 as a novel angiogenesis-regulating gene by genome association studies in mice. *PLOS Genetics* 13, e1006848 (2017).

189. Huntzinger, E., Kashima, I., Fauser, M., Saulière, J. & Izaurralde, E. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *Rna* 14, 2609–2617 (2008).

190. Eberle, A. B., Lykke-Andersen, S., Mühlemann, O. & Jensen, T. H. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nature Structural & Molecular Biology* 16, 49–55 (2008).

191. Brown, T. J., Kollara, A., Shathasivam, P. & Ringuette, M. J. Ventricular Zone Expressed PH Domain Containing 1 (VEPH1): an adaptor protein capable of modulating multiple signaling transduction pathways during normal and pathological development. *Cell Commun Signal* 17, 116 (2019).

192. Zhang, C., Li, C., Siu, G. K. Y., Luo, X. & Yu, S. Distinct Roles of TRAPPC8 and TRAPPC12 in Ciliogenesis via Their Interactions With OFD1. *Frontiers Cell Dev Biology* 8, 148 (2020).

193. Giovannone, B. *et al.* Two Novel Proteins That Are Linked to Insulin-like Growth Factor (IGF-I) Receptors by the Grb10 Adapter and Modulate IGF-I Signaling*. *J Biol Chem* 278, 31564–31573 (2003).

194. Bassik, M. C. *et al.* A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility. *Cell* 152, 909–922 (2013).

195. Kim, J. J., Lipatova, Z. & Segev, N. TRAPP Complexes in Secretion and Autophagy. *Frontiers Cell Dev Biology* 4, 20 (2016).

196. Sacher, M., Shahrzad, N., Kamel, H. & Milev, M. P. TRAPPopathies: An emerging set of disorders linked to variations in the genes encoding transport protein particle (TRAPP)-associated proteins. *Traffic* 20, 5–26 (2019).

197. Milev, M. P. *et al.* TRAMM/TrappC12 plays a role in chromosome congression, kinetochore stability, and CENP-E recruitment. *J Cell Biol* 209, 221–234 (2015).

198. Willaredt, M. A., Gorgas, K., Gardner, H. A. R. & Tucker, K. L. Multiple essential roles for primary cilia in heart development. *Cilia* 1, 23 (2012).

199. Koefoed, K., Veland, I. R., Pedersen, L. B., Larsen, L. A. & Christensen, S. T. Cilia and coordination of signaling networks during heart development. *Organogenesis* 10, 108–125 (2013).

200. Klena, N. T., Gibbs, B. C. & Lo, C. W. Cilia and Ciliopathies in Congenital Heart Disease. *Csh Perspect Biol* 9, a028266 (2017).

201. Gabriel, G. C., Young, C. B. & Lo, C. W. Role of cilia in the pathogenesis of congenital heart disease. *Semin Cell Dev Biol* 110, 2–10 (2020).

202. Emery, B. *et al.* Myelin Gene Regulatory Factor Is a Critical Transcriptional Regulator Required for CNS Myelination. *Cell* 138, 172–185 (2009).

203. Koenning, M. *et al.* Myelin Gene Regulatory Factor Is Required for Maintenance of Myelin and Mature Oligodendrocyte Identity in the Adult CNS. *J Neurosci* 32, 12528–12542 (2012).

204. Jin, S. C. *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet* 49, 1593–1601 (2017).

205. Huang, H., Zhou, F., Zhou, S. & Qiu, M. MYRF: A Mysterious Membrane-Bound Transcription Factor Involved in Myelin Development and Human Diseases. *Neurosci Bull* 37, 881–884 (2021).

206. Longhurst, J. C. Cardiac receptors: Their function in health and disease. *Prog Cardiovasc Dis* 27, 201–222 (1984).

207. Hanna, P. *et al.* Cardiac neuroanatomy - Imaging nerves to define functional control. *Autonomic Neurosci* 207, 48–58 (2017).

208. Floras, J. S. Sympathetic activation in human heart failure: diverse mechanisms, therapeutic opportunities. *Acta Physiol Scand* 177, 391–398 (2003).

209. Kumar, R. *et al.* Brain axonal and myelin evaluation in heart failure. *J Neurol Sci* 307, 106–113 (2011).

210. Shen, M. J. & Zipes, D. P. Role of the Autonomic Nervous System in Modulating Cardiac Arrhythmias. *Circ Res* 114, 1004–1021 (2014).

211. Jr, H. C. M., McMahan, C. A. & Gidding, S. S. Preventing Heart Disease in the 21st Century. *Circulation* 117, 1216–1227 (2008).

212. Liu, X. ABC Family Transporters. *Adv Exp Med Biol* 1141, 13–100 (2019).

213. Harvey, R. P. NK-2Homeobox Genes and Heart Development. *Dev Biol* 178, 203–216 (1996).

214. Schutera, M. *et al.* Machine Learning Methods for Automated Quantification of Ventricular Dimensions. *Zebrafish* 16, 542–545 (2019).

215. Sessa, F. *et al.* Heart rate variability as predictive factor for sudden cardiac death. *Aging Albany Ny* 10, 166–177 (2018).

216. Mulkey, S. B. *et al.* Heart rate variability is depressed in the early transitional period for newborns with complex congenital heart disease. *Clin Auton Res* 30, 165–172 (2020).

217. Gierten, J. Genome-wide analysis of quantitative cardiac traits in medaka inbred strains. (2019).

218. Zuo, E. *et al.* One-step generation of complete gene knockout mice and monkeys by CRISPR/Cas9-mediated gene editing with multiple sgRNAs. *Cell Res* 27, 933–945 (2017).

219. Kim, B. H. & Zhang, G. Generating Stable Knockout Zebrafish Lines by Deleting Large Chromosomal Fragments Using Multiple gRNAs. *G3 Genes Genomes Genetics* 10, g3.401035.2019 (2020).

220. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424 (2016).

221. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464–471 (2017).

222. Cornean, A. *et al.* Precise in vivo functional analysis of DNA variants with base editing using ACEofBASEs target prediction. *Biorxiv* 2021.07.26.453883 (2021) doi:10.1101/2021.07.26.453883.

223. Varshney, G. K. *et al.* A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish. *Nat Protoc* 11, 2357–2375 (2016).

224. Xie, N. G. *et al.* High-throughput variant detection using a color-mixing strategy. *Biorxiv* 2021.06.30.450651 (2021) doi:10.1101/2021.06.30.450651.

225. Feng, Z. & R., L., James. Non-coding genetic variants in human disease. *Hum Mol Genet* 24, R102–R110 (2015).

226. Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nature Publishing Group* 1–8 (2011).

227. Atanasov, A. G. *et al.* Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances* 33, 1582–1614 (2015).

228. A., M., Calum & T., P., Randall. Zebrafish as tools for drug discovery. *Nat Rev Drug Discov* 14, 721–731 (2015).

229. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A Grammar of Data Manipulation*. (2020).

230. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682 (2012).

231. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (2016).

232. Kassambara, A. *ggpubr: "ggplot2" Based Publication Ready Plots*. (2020).

233. Team, R. C. *R: A Language and Environment for Statistical Computing*. (2020).

234. Team, Rs. *RStudio: Integrated Development Environment for R*. (2019).

235. Wickham, H. *et al.* Welcome to the {tidyverse}. *Journal of Open Source Software* 4, 1686 (2019).

236. Bert, B. *et al.* Considerations for a European animal welfare standard to evaluate adverse phenotypes in teleost fish. *The EMBO Journal* 35, 1151–1154 (2016).

237. Koster, R., Stick, R., Loosli, F. & Wittbrodt, J. Medaka spalt acts as a target gene of hedgehog signaling. *Development* 124, 3147 (1997).

Publications

Publications

Hammouda, O. T., Böttger, F., Wittbrodt, J., & Thumberger, T. (2019). Swift Large-scale Examination of Directed Genome Editing. *PLoS ONE*, *14*(3), e0213317-11.

Gierten, J., Pylatiuk, C., **Hammouda, O. T.**, Schock, C., Stegmaier, J., Wittbrodt, J., Gehrig, J., & Loosli, F. (2020). Automated high-throughput heartbeat quantification in medaka and zebrafish embryos under physiological conditions. *Scientific Reports*, 1–12.

Submitted manuscripts

Hammouda, O. T., Wu, M. Y., Kaul, V., Thumberger, T., & Wittbrodt, J. (2021). In vivo identification and validation of novel potential predictors for human cardiovascular diseases. *BioRxiv*, 2021.02.03.429563.

Leger, A., Brettell, I., Monahan, J., Barton, C., Wolf, N., Kusminski, N., Herder, C., Aadepu, N., Becker, C., Gierten, J., **Hammouda, O. T.**, Hasel, E., Lischik, C., Lust, K., Suzuki, R., Tavhelidse, T., Thumberger, T., Tsingos, E., Watson, P., … Fitzgerald, T. (2021). Genomic variations and epigenomic landscape of the Medaka Inbred Kiyosu-Karlsruhe (MIKK) panel. *BioRxiv*, 2021.05.17.444424. https://doi.org/10.1101/2021.05.17.444424

Fitzgerald, T., Brettell, I., Leger, A., Wolf, N., Kusminski, N., Monahan, J., Barton, C., Herder, C., Aadepu, N., Gierten, J., Becker, C., **Hammouda, O. T.**, Hasel, E., Lischik, C., Lust, K., Suzuki, R., Tsingos, E., Tavhelidse, T., Thumberger, T., … Loosli, F. (2021). The Medaka Inbred Kiyosu-Karlsruhe (MIKK) Panel. *BioRxiv*, 2021.05.17.444412. https://doi.org/10.1101/2021.05.17.444412

Acknowledgments

First and foremost, I want to thank **Jochen** for everything. Thank you for giving me this great opportunity to join your lab and work on this exciting project. Thank you for providing an excellent infrastructure and a stimulating working environment. Thank you for your great supervision and support throughout my thesis. Thank you for giving me all the opportunities to learn and grow scientifically and personally, and for always challenging me to become the best version of myself. Thank you for your presence in the lab and for always being there for us with the needed dose of optimism to overcome the challenges we face. Thank you for your genuine interest in our personal life and health. Thank you for being such a great role model to look up to, both in terms of a terrific scientist as well as an amazing mentor. My respect for you as a person, scientist and mentor goes way beyond this paragraph. It has been an honor and pleasure to work in your lab. This PhD is one of the very few things in my life where I would blindly do it all over again.

I want to thank **Prof. Dr. med. Johannes Backs** for being my second supervisor, and for his great input, support and scientific advice throughout my PhD as a TAC member. I would also like to thank **Dr. Eileen Furlong**, **Dr. Felix Loosli** and **Jun.-Prof. Dr. Steffen Lemke** for their excellent and diverse input and for providing external perspectives of my project as valuable TAC members. I would also like to extend my gratitude to **Jun.-Prof. Dr. Steffen Lemke and Jun.-Prof. Dr. Lazaro Centanin** for agreeing to be my examiners.

I want to thank **Dr. Rolf Lutz**, **Martina Galvan** and **Sandra Martini** from HBIGS for always being happy to help, and for providing us graduate students with a great infrastructure and various interesting and very useful opportunities for self-development.

I want to thank **Prof. Dr. Ewan Birney** and all the **Birney Group** members for the great and stimulating collaboration. The IndiGene project for me is truly an eye opener, the ambition, the (very)long-term vision, the persistence, the optimism, the trust and the implementation. A text book example of how a great collaboration can have endless potential to grow.

I am deeply grateful for **Ute Volbehr, Frederike Seibold** and **Kevin Verweyen** and past members of the Sekretariat office for always seamlessly taking care of our paperwork and for always helping with all the organizational queries with great enthusiasm.

I want to thank **Tanja Kellner** for all the valuable advice she gave me throughout the years and for all the help she patiently provided. I also would like to thank **Beate Wittbrodt**, for all her help, advice as well as the tips and tricks for doing my *in situs*. I extend my gratitude towards the Fish team **Erik Leist, Antonino Saraceno Marzena**

Majewski and **Sven Erny** for the excellent fish husbandry and for always taking extra care of our precious fish.

I want to acknowledge the great students I had the chance to supervise, **Meng Yue Wu**, **Verena Kaul** and **Gasser Elmissiery**, thank you all very much for your hard work and for giving me the opportunity to transfer my knowledge and passion to you.

The 5th floor at COS, all past and present members, from all three groups has been like a family to me. So many great friends I have made over the years, so many great moments and happy memories from all the COS parties, defense and paper parties as well as game and movie nights. I especially thank **Natalia Sokolova, Bettina Welz, Julian Stolper, Ann-Kathrin Heilig, Naima Ruhland, Karen Groß, Leonie Adelmann, Sevinc Gücüm, Risa Suzuki, Philip Watson, Clara Becker, Arturo Gutierrez, Jørgen Benjaminsen** and **Lucie Zilova** as well as all members of the Wittbrodt lab for the awesome working environment and stimulating discussions.

I am deeply grateful to **Erika Tsingos** and **Venera Weinhardt** for supporting me with my project during the Covid lockdown and for the great discussions and advice that helped shape my story. I am also extremely grateful to **Alex Cornean, Natalia Sokolova, Bettina Welz, Philip Watson** and **Linda Manhart** for their great and extensive feedback on my thesis and for finding the time despite it happening around the summer holiday period.

I am immensely grateful to **Jakob Gierten** for his help and valuable teachings when I first joined the lab, for being a great collaborator and colleague, and for the many stimulating discussions we had on our projects. Your hard work ethics and organization are really admirable.

Thomas Thumberger thank you very much for your continuous help and support throughout my thesis. Thank you for your endless and contagious enthusiasm and the good vibes you bring to the lab. Thank you for your patience with me during the various geneious and troubleshooting sessions. I am also very thankful for our collaboration on the SLEDGE-hammer protocol.

Tinatini Tavhelidse thank you for being such a great listener, you were always there whenever I was down, disappointed, frustrated or just in a whiny mood due to PCR problems and always patient. But you were also always there whenever I needed advice or help with stuff in the lab or the various protocols, especially during the *in situ* periods. I have rarely laughed so much like that one time (you know what I'm talking about). Talking about *in situs*, I would like to thank **Natalia Sokolova** again for putting up with me and my many questions, also for the emotional support during the hard times.

Alex Cornean and **Colin Lischik** you get your own paragraph and special mention ;). Thank you very much for the great working environment, for the stimulating discussions,

both scientific and other random topics. Thank you for always being there whenever I need advice, help or just to vent off. Thank you for being amazing friends both in and outside of the lab. This extends also to **Lena Appel** and **Miriam Lischik** and little **Marlene-Sophia Lischik**. I really enjoyed all our jogging sessions, game and GoT nights and Joe Molese dinners. I really hope we stay in touch and perhaps one day do that trip to Egypt together.

I am eternally grateful to my Egyptian family here in Germany. It has all started with **Sherif Ismail**, **Ali Seleit** and **Mostafa ElMaghraby**, thank you very much for all the amazing times we spent together, all the laughs we had and great moments we shared, all the billiard, GoT, CoD Zombie, Tarneeb and board game sessions we had, and for always being there for me. The family has been slowly expanding since then, and now it has fully grown into the "فيسبادن الباسلة" family, with an added new generation. Thank you all very much for all the great memories and for all the great times we had and for being an amazing group even if we are spread across cities. I also want to thank **Sven Reislöhner**, **Federica Fiorentino, Alessandro Greco** and **Chiara Di Ponzio** for all the great times and all the fun we had.

I want to thank my amazing friends in Egypt, starting with my oldest friend **Tarek ElGhadban** and his little sister **Nour ElGhadban** (yes little :P)**,** I have known you both since you were born and consider you like a brother and sister more than friends. I also thank my school friends "نادي الرجال السري", whom I've known since Kinder Garten, my friends from Uni: "Bus 73", "El Shella", "La Maison Blue" and "Forbidden" groups, as well as my friends from Squash. You are all such great and loyal friends, thank you all for always being there for me, for always making my stay in Egypt fun and memorable and for always staying in touch. I also want to thank my Squash buddies here in Heidelberg "Montags-Squash" for all the great games we played over the years.

I am and will always be eternally grateful and thankful to my family, **Tarek Hammouda, Giuliana Cirillo** and my awesome brother **Youssef Hammouda**. I am **where** I am today because of you, and I am **who** I am today because of you, so thank you for your unconditional love and support, thank you for always supporting and believing in me and for providing me with all the opportunities and the best education I needed to succeed in life. Thank you for always being there for me at my best and worst times. I know I don't really show it or express it often but I do love you so much and I couldn't have asked for better parents and brother. My gratitude also extends to the rest of my family in Egypt as well as in Italy, with special mention of my cousins "Shabab el3ela".

I am also deeply grateful and thankful to my second family, thank you **Amgad Abdelhamid, Hanaa Khalil, Mahmoud Abdelhamid and Moustafa Abdelhamid** for all the love and support you gave me throughout the years. Thank you very much for including me into your family as one of your own, and for always being there for me. Thank you for all the great memories we had, and many more to come, I really couldn't have asked for a better family. My gratitude also extends to the rest of the family in Egypt

as well as in Sweden, with special mention of the “El haشaysha” group of an interesting mixture of close/distant cousins and friends.

For you, my dearest love, **Shaymaa Abdelhamid**, even a whole PhD thesis would not merely be enough to describe how of an awesome, caring, loving and supportive wife you are. From the very beginning, you have been nothing but supportive throughout all the decisions that I have made for my life and career. When I got the chance to study abroad, even if it would lead to a long-distance relationship, you encouraged me to pursue my dream. When I had the opportunity to switch to a better program, you knew it was better for me and even if it meant that I had to start over prolonging our long-distance relationship for an additional 2 years you still encouraged me to do it without hesitation. After we got married and I received acceptance for a PhD I wasn’t totally convinced of, even if at the time both of us would risk being unemployed, you still fully supported me and my decision to skip it to look for another project I would be really passionate about, and so I did and it landed me here where I am now and that is because of **YOU**. Thank you so much for always believing in me, for always supporting me, for always loving me at my best and worst. Thank you for always being there for me whenever I need someone to share my best and worst moments, and whenever I needed emotional and mental support. Thank you for always taking care of me, especially during the hard writing phase, even with our little baby, you still took excellent care of me. You have no idea how much your “30 activities before 30” meant to me, how much it helped me power through the dark depressing writing times and how much of an amazing gift this was. We have been together now for eight years, shifted from friends to parents with our little **Rayyan Hammouda** “**Ruru**”**.** I know I am not always the best as expressing myself, but I want you to know that I love you more than anything, and that you are the best thing that has happened in my life and I couldn’t have done all of this without you. Thank you **Shaymaa Abdelhamid** for everything.

Declaration

Herewith I declare that I prepared the PhD Thesis "Novel genetic predictors of human cardiovascular diseases validated by CRISPR in medaka" on my own and with no other sources and aids than quoted.

Heidelberg, 2021

List of Figures

List of Tables

Appendix

For Appendix Tables 1-3 below, Ns correspond to number of scored embryos per condition. Before DF = before developmental focusing (i.e. all scored embryos). After DF = after developmental focusing (i.e. only scored embryos looking older than stage 28).

Appendix Table 1. Number of successfully scored embryos in Figure 2.12 and Figure 2.13

Figure	Panel/Condition	Ns before DF		Ns after DF	
		21C	28C	21C	28C
Figure 2.12	B) WT	35	34	35	34
	B) Mock	23	20	23	20
Figure 2.13	D) Mock	22	23	22	23
	D) *nkx2-5_T4*	34	34	31	31
	D) *oca2_T3*	35	35	35	35
	E) Mock	19	23	19	23
	E) *nkx2-5_T4*	34	36	31	33
	F) Mock	22	21	22	21
	F) *nkx2-5_T5*	33	33	31	31

Appendix Table 2. Number of successfully scored embryos in Figure 2.14 and Figure 2.15

Gene	Ns before DF		Ns after DF		Ns mock before -> after DF	
	21C	28C	21C	28C	21C	28C
cdc42	32	32	29	30	22	20
ogdh	31	33	31	33		
duox	32	33	27	28	19	21
git2	35	33	33	31		
mus81	27	33	25	31	20 -> 18	21 -> 20
or124-2	34	34	30	30		
plekha8	35	31	29	26	21	23
ttl	33	30	28	24		
cabp4	34	34	34	34	21	21
eml6	36	34	34	32		

Appendix Table 3. Number of embryos successfully scored in Figure 2.16 and Figure 2.17

Gene	Ns before DF		Ns after DF		Ns mock before -> after DF	
	21C	28C	21C	28C	21C	28C
bag3	32	33	32	33	22	24
edn1	35	34	35	34		
zfhx3	33	35	33	35	23	22
nubp2	36	34	36	34		
minar1	33	33	30	30	24	22
ppp1r9a	32	34	32	34		
atp8b4	35	34	32	31	24	23
padi2	31	36	31	36		
sspo	33	32	33	32	22	22
scn4ab	35	32	35	32		
ufsp1	33	32	33	32	23	24
veph1	34	35	34	35		
clcnk	34	31	34	31	22	23
rgs3a	26	28	21	23		
myrf	34	34	33	33	24 -> 23	23 -> 22
cnot1	36	35	22	22		
trappc12	35	35	34	34	23	24
or5au1	34	33	34	33	22	22
naca	33	32	27	27		
abcb4	36	35	36	35	23	22
smg6	36	35	30	29		
maml3	35	35	34	34	23	24
plg	35	36	33	34		
sh2b3	35	36	35	36	24	24
rnf207b	35	36	33	34		
slc17a3	36	35	36	35	22	24
scmh1	33	33	33	33		
hcn4	35	36	33	34	23	24
grid2	30	34	30	34	23	20
casq2	59	56	53	50	25 -> 24	24 -> 23
ccdc141	36	34	36	34	22	23
gigyf1	34	35	34	35		
homeza	32	35	32	35	19	23
cep85l	35	36	35	36	22	24
piezo1	35	33	35	33		
ttn.2	34	35	34	35	23	24
cmya5	33	34	33	34		
xylb	35	33	35	33	23	22

col9a1b	35	35	34	34		
kcnh2	34	36	34	36	21 -> 20	21