



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1547*

Translating Cardiac and Cardiometabolic GWAS Using Zebrafish

BENEDIKT VON DER HEYDE



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019

ISSN 1651-6206
ISBN 978-91-513-0590-5
urn:nbn:se:uu:diva-378941

Dissertation presented at Uppsala University to be publicly examined in Humanistiska Teatern, Engelska Parken, Thunbergsv. 3H, Uppsala, Friday, 26 April 2019 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: PhD Shawn Burgess (National Human Genome Research Institute, National Institute of Health, Bethesda, USA).

Abstract

von der Heyde, B. 2019. Translating Cardiac and Cardiometabolic GWAS Using Zebrafish. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1547. 56 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0590-5.

Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiac and cardiometabolic traits. However, the trait-associated variants usually do not clearly point to causal gene(s), mechanism(s) or tissue(s). Model systems that allow for a comprehensive and quick candidate gene screening are necessary, ideally *in vivo*. The overall objective of my thesis is to establish large-scale, imaged-based screens in zebrafish embryos and larvae to examine candidate genes for their effects on heart rate and rhythm, as well as on early-onset atherosclerosis and dyslipidemia.

In Study 1, I prioritized 18 candidate genes in eight loci identified in a meta-analysis of GWAS for heart rate variability. Some of these genes were already known to be involved in cardiac pacemaking, whereas others require functional characterization.

In Study 2, I established an experimental pipeline to examine genetic effects on cardiac rate and rhythm and used it to characterize orthologues of six human candidate genes for heart rate and rhythm. I confirmed known effects of *rgs6* and *hcn4*, and established a role for *KIAA1755* in HRV.

In Study 3, I contributed to large-scale experiments to establish the zebrafish as a model system for early-onset atherosclerosis and dyslipidemia. Overfeeding and cholesterol-supplementation of the diet were shown to propel independent pro-atherogenic effects. Atherosclerotic burden was alleviated using commonly prescribed drugs in humans. Lastly, the effects of proof-of-concept genes known to be involved in lipid metabolism were examined and showed higher LDLc (*apoea*) and early-onset atherosclerosis (*apobb1*).

In Study 4, I characterized genes in GWAS-identified loci for triglyceride levels for a role in lipid metabolism and early-stage atherosclerosis. I identified three previously unanticipated genes that influence triglyceride levels in zebrafish larvae. Several additional genes influence other cardiometabolic risk factors. Interestingly, two genes showed trends towards lower triglycerides levels (*dock7* and *lpar2a*), with directionally opposite effects on vascular inflammation. This emphasizes that candidate genes need to be examined comprehensively to guide further mechanistic studies.

Keywords: GWAS, Zebrafish

Benedikt von der Heyde, Department of Immunology, Genetics and Pathology, Medicinsk genetik och genomik, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Benedikt von der Heyde 2019

ISSN 1651-6206

ISBN 978-91-513-0590-5

urn:nbn:se:uu:diva-378941 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-378941>)

How much is the fish?
- Hans Peter Geerdes

To my family

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I Nolte, IM., Munoz, ML., Tragante, V., Amare, AT., Jansen, R., Vaez, A., **von der Heyde, B.**, [159 co-authors], den Hoed, M., Snieder, H., de Geus, E. (2017) Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nature Communications*, 8:15805
- II **Von der Heyde, B.**, Emmanouilidou, A., Klingström, T., Mazzaferro, E., Vicenzi, S., Jumaa, S., Dethlefsen, O., Snieder, H., de Geus, E., Ingelsson, E., Allalou, A., Brooke, HL., den Hoed, M. (2019) Translating GWAS-identified loci for cardiac rhythm and rate using an *in vivo*, image-based, large-scale genetic screen in zebrafish. *Submitted manuscript*
- III Bandaru, MK., Emmanouilidou, A., Ranefall, P., **von der Heyde, B.**, Mazzaferro, E., Klingström, T., Masiero, M., Dethlefsen, O., Ledin, J., Larsson, A., Brooke, HL., Wählby, C., Ingelsson, E., den Hoed, M. (2019) Zebrafish larvae as a model system for systematic characterization of drugs and genes in dyslipidemia and atherosclerosis. *Submitted manuscript*
- IV **Von der Heyde, B.**, Masiero, M., Mazzaferro, E., Ranefall, P., Emmanouilidou, A., Klingström, T., Larsson, A., Wählby, C., Ingelsson, E., den Hoed, M. (2019) Characterising candidate genes for cardiometabolic risk factors in GWAS-identified loci for triglyceride levels using a high-throughput zebrafish screen. *Manuscript*

Reprints were made with permission from the respective publishers.

Contents

Introduction	11
The human genome	11
Genetic variation.....	12
Identifying candidate genes for simple and complex traits	13
Linkage studies	13
Genome-wide association studies	14
Translating GWAS-identified variants	15
The zebrafish genome.....	16
CRISPR-Cas9	17
A brief history	17
Cas9 as a tool for genome-editing	18
The human heart	20
The cardiac conduction system	21
Modulation of heart rate and rhythm	23
Heart rate variability	23
GWAS of cardiac rate and rhythm.....	24
The zebrafish heart.....	25
Zebrafish cardiac electrophysiology	26
Zebrafish as a model for system for cardiac traits	27
The vasculature.....	28
Lipid metabolism	28
Etiology of coronary artery disease	29
The genetics of CAD	30
Pharmacological intervention of CAD.....	31
The zebrafish as a model system for CAD	33
Present investigations	35
Aims.....	35
Summary.....	36
Study 1	36
Study 2	36
Study 3	38
Study 4	39
Concluding remarks and future perspective.....	40

Acknowledgements42

References44

Abbreviations

ANGPTL3	Angiopoietin-like 3
ANS	Autonomous nervous system
APO	Apolipoprotein
AV	Atrioventricular
BP	Base pair
BPM	Beats per minute
CAD	Coronary artery disease
CAS	CRISPR-associated sequence
CETP	Cholesteryl ester transfer protein
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
DPF	Days post fertilization
ECG	Electrocardiogram
EQTL	Expression quantitative trait locus
GTEX	Genotype tissue expression
GWAS	Genome-wide association study
HAPMAP	Haplotype map
HCN	Hyperpolarization-activated cyclic nucleotide-gated channel
HDLc	High density lipoprotein cholesterol
HMGR	3-hydroxy-3-methyl-glutaryl-coenzyme A reductase
HPF	Hours post fertilization
HRV	Heart rate variability
IDL	Intermediate density lipoprotein
INDEL	Insertion/deletion
LCAT	Lecithin-cholesterol acyltransferase
LD	Linkage disequilibrium
LDLc	Low density lipoprotein cholesterol
LPL	Lipoprotein lipase
MRNA	Messenger RNA
MV	Millivolt
PAM	Protospacer adjacent motive
PCR	Polymerase chain reaction
PCSK9	Proprotein convertase subtilisin/kexin type 9
RMSSD	Root mean square of successive differences

RNA	Ribonucleic acid
SA	Sinoatrial
SDNN	Standard deviation of NN intervals
SNP	Single nucleotide polymorphism
VAST	Vertebrate automated screening technology
VEP	Variant effect predictor
VLDL	Very low density lipoprotein

Introduction

The human genome

Every living organism's instruction manual for their building blocks is deoxyribonucleic acid (DNA). DNA consists of four bases: adenine, cytosine, guanine and thymine that are lined up by a sugar-phosphate backbone. The molecule is usually double-stranded, adenine base-pairing with thymine by forming two hydrogen bonds between each other, and guanine with cytosine, forming three hydrogen bonds. The biophysical properties of normal DNA give rise to the classic right-handed double helix. The information needed to orchestrate the successful assembly and maintenance of an organism is encoded in the sequence of the four bases. About 1.2% of the human genome is referred to as "coding", meaning that the genetic information is transcribed into messenger RNA (mRNA), which ultimately will be translated into proteins. A mRNA molecule can give rise to different versions of the protein, so-called isoforms, by a process called alternative splicing, thus serving a convenient way to produce temporally and spatially specific isoforms that may be needed. The estimate of coding genes in the most recent human genome build (GRCh38) is at 20,418 (Ensembl, accessed February, 2019). The remaining 98.8% of the genome are referred to as "non-coding". Non-coding sequences may be transcribed into non-coding RNAs, but are not translated into proteins. The current estimate of the number of non-coding genes is about 22,000 (Ensembl, accessed February, 2019). However, non-coding RNAs such as transfer RNAs, ribosomal RNAs or microRNAs may still be functional. Many are involved in development and homeostasis. Subsequently, many have been implicated in disease. Additionally, regulatory elements, repeat sequences, transposons, introns, telomers and pseudogenes belong to the non-coding category. Most researchers argue that a large proportion of the remaining 98.8% of the genome serve regulatory function (The ENCODE project consortium, 2007), while others consider this as overestimations (Palazzo & Gregory, 2014).

The first draft sequence of the human genome was published in 2001 (Lander et al., 2001; Venter et al., 2001), and the complete sequence was revealed shortly after (International Human Genome Sequencing Consortium, 2004). It was a milestone, enabling phylogenetic comparisons, to understand what makes us human and it fueled the hope of a better comprehension of the healthy and diseased body.

Genetic variation

The sequence of the human DNA bears a vast amount of complexity and variation. Genetic variation is vital for evolution, adaptation and diversity, and plays a significant role in traits and diseases. Variation occurs at an individual and at a population level. They can arise in every cell of the body, but only variants present in the germ cells will be inherited to the next generation.

The main sources of genetic variation are mutations and recombination. Mutations are changes in the DNA that are inherited from the parental generation, but also arise *de novo* in somatic cells, either due to errors in DNA replication or UV radiation, or insertion of transposable elements. The current estimate of the *de novo* mutation rate in humans is 10^{-8} per base pair (bp) per generation (1000 Genomes Project Consortium, 2010). Another source for genetic variation is recombination. During meiosis, fragments of the maternal and paternal homologous chromosomes may be exchanged, a so-called cross-over, thereby giving rise to novel chromosomal arrangements. Not every mutation has a functional consequence – it can be detrimental, neutral or provide an advantage to the organism.

At the sequence level, individuals vary in the form of single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and larger structural variations such as copy number variants, inversions and duplications.

By far, SNPs (the alteration of a single base pair) are the most common type of variant. Indels are the second most frequent variant and are defined to be insertions or deletions in the range from 2 bps up to several hundred bps. An indel can also describe an event where an insertion after a deletion occurred. Structural variants typically describe events that are in the range of kilobases. It has been estimated that an individual genome of European descent differs to the reference genome in around four million sites (1000 Genomes Project Consortium, 2015), largely constituted by SNPs and short indels.

There have been huge efforts to map and catalogue human variation from different populations. The haplotype map (HapMap) project was initiated shortly after sequencing of the first genome (International HapMap Consortium, 2003). Its aim was to map common genetic variants (allele frequency >5%), their allele frequencies and correlations to each other, in order to enable genetic association studies. To achieve this, 269 individuals from four populations were genotyped using different genotyping technologies, resulting in a vast increase of reported SNPs (International HapMap Consortium, 2005, 2007). It also fostered the development of high-throughput SNP genotyping technologies that are based on “tag-SNPs”. Some alleles on the same chromosome tend to be inherited together more often than by chance, which is designated as a haplotype block. Linkage disequilibrium (LD) describes the extent by which two alleles/markers are

inherited together. Thus, a tag-SNP can be used to mark a genomic region based on LD structure because the haplotype blocks are defined for a given population. This allows the capturing of neighboring variants that are not present on the genotyping chip, in order to minimize genotyping efforts. In parallel, this fostered the development of imputation; a method of statistically inferring genotypes of unobserved SNPs by comparison to a larger reference panel. Some of these advances ultimately enabled the first generation of genome-wide association studies (GWAS) that will be discussed later.

With the costs of sequencing decreasing over time and advances in sequencing technology, the 1000 genomes project was launched (1000 Genomes Project Consortium, 2010). The aim was to catalogue not only SNPs but also indels and structural variants, and deeper sequencing to detect low (allele frequency between 5 and 1%) and rare (allele frequency <1%) frequency variants. Genotypes were determined by a combined strategy of deep SNP genotyping, low coverage whole genome sequencing and deep exome sequencing. They reported in their final report the assembly of 2504 genomes from 26 populations, providing detailed insight into human genetic variation (1000 Genomes Project Consortium, 2015).

Identifying candidate genes for simple and complex traits

Linkage studies

For a long time, the most successful method to identify disease genes was linkage analysis. Mendelian disorders, also known as monogenic disorders, show familial aggregation. Linkage analysis is based on the fact that a genetic marker will co-segregate with the disease and follow a mendelian pattern of inheritance; hence the locus of the gene causing the disease can be identified. Additionally, the locus bearing the pathogenic gene has a high penetrance. Therefore, linkage analysis was extremely successful in identifying genes that underlie single-gene disorders, such as sickle-cell anemia (*HBB* gene), cystic fibrosis (*CFTR* gene) or huntington disease (*HTT* gene).

However, linkage analysis turned out to not be suitable for complex traits and diseases (Altmuller, Palmer, Fischer, Scherb, & Wjst, 2001). Complex traits arise from the interplay of genetics and environment, where many variants contribute with a small to modest effect, with low penetrance. Thus, the classic mendelian inheritance patterns do not apply.

GWAS were launched with the promise to detect association of variants with complex traits, with the hope to improve disease knowledge and prediction.

Genome-wide association studies

Following the huge efforts initiated by the HapMap project to map common genetic variation, the first GWAS were conducted. A GWAS is an observational study that tests the association of a genome-wide set of SNPs with a trait or disease, most commonly in humans, yet it can also be performed in any other species. Both dichotomous and quantitative traits can be examined, which informs the statistical method that needs to be used for analysis. For example, variants that manifest in a certain phenotype are more commonly found in cases rather than controls in a dichotomous outcome. Usually, the statistical power is higher in studies with continuous outcomes, due to their continuous nature. Over the years, GWAS has witnessed a drastic increase in sample size. Whereas in the beginning around a hundred individuals were analyzed (Klein et al., 2005), the sample size can now reach 1 – 1.5 million individuals (Nielsen et al., 2018). This increases the statistical power to detect variants with lower minor allele frequencies and smaller effects, and hence also improves understanding of the genetic architecture of a trait.

The initiative of international consortia fostered the data acquisition and generation of large cohorts. This bears the advantage that typically many parameters per individual are available that thus can be adjusted for, to reduce spurious associations. Subsequently, with increasing sample sizes, GWAS estimates of the SNP-based heritability improved, estimating how much phenotypic variance is explained by all SNPs.

It has also been demonstrated that the effect size of a variant on a trait in a population is not necessarily indicative of its molecular effect on the protein, illustrated by the identification of variants tagging proteins that already serve as drug targets in humans, for example 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (*HMGCR*), the molecular target of statins (Willer et al., 2013).

In the early days, many discovered associations suffered from “winners curse”, indicating that the effect size of the variant was overestimated, or in worst case a false positive association in the discovery analysis. Subsequently, a “multi-stage” analysis setup was introduced. This infers that first a discovery analysis is run (stage 1), followed by replication in an independent cohort (stage 2) to see if the association is true. Most commonly nowadays is the meta-analysis of GWAS. This requires that parameters between studies have been carefully harmonized, such as phenotype definition for example. Meta-analyses have been facilitated by imputation; hence regardless of which array is used in the respective cohort, data is obtained on the same variants across all studies using the same reference panel.

To date, GWAS have identified >100.000 unique SNP-trait associations (GWAS catalogue(Buniello et al., 2019), accessed February 2019), and have substantially aided the understanding of the genetic architecture of complex traits.

As the field continues to evolve, recent years saw developments such as GWAS of copy number variants for example (The Wellcome Trust Case Control Consortium, 2010). Additionally, methods that employ GWAS data, such as polygenic risk scores (calculations to predict an individual's disease risk/trait), pleiotropy (identification of variants that contribute to multiple traits) and mendelian randomization (investigation of the causal relationship between two traits using SNPs as instrumental variables) have been developed. Importantly, trait-associated variants are almost never indicative of the causal gene, mechanism and/or tissue, and thus warrant further studies.

Translating GWAS-identified variants

Once an association with a phenotype is detected, the association needs to be examined further to infer function. Since the consequences of variants can at most be predicted using bioinformatics approaches, integration of many layers of evidence is desirable to make the most educated guesses possible.

The genotype tissue expression (GTEx) database allows for cross-referencing if the SNP is as an expression quantitative trait locus (eQTL, a SNP that is associated with the expression of a gene), to see the variants' association with gene expression in a given tissue (GTEx Consortium, 2015). This may provide a first step in functional interpretation. However, eQTL SNPs have been shown to be promiscuous, being able to flag multiple genes in the same tissue (Battle, Brown, Engelhardt, & Montgomery, 2017). Similarly, eQTLs can act in *cis*, affecting the regulation of a gene nearby, or in *trans*, modifying expression of a distant gene. Using the Encyclopedia of DNA elements database (The ENCODE project consortium, 2007), it can be assessed whether a SNP is predicted to change a transcription-factor binding motif for example. Data from the Roadmap Epigenomics project can be utilized to investigate if a SNP coincides with an epigenetic mark (Kundaje et al., 2015). Hi-C information can be utilized to investigate if the SNP is in a distant motif that might bind to transcription elements via looping (Schmitt et al., 2016).

Functional interpretation of exonic variants is aided by tools such as PolyPhen2 (Shihab et al., 2013) and SIFT (Sim et al., 2012) that report the predicted effect of the variant on protein function by integrating sequence comparison and physical considerations in one algorithm. Variant annotation tools such as the variant effect predictor (VEP) (McLaren et al., 2016) and Annovar (H. Yang & Wang, 2015) have been developed that facilitate annotation of coding and non-coding variants incorporating information from PolyPhen2/SIFT, as well as overlap with transcription-factor binding site and regulatory features, amongst others.

The large number of disease-associated loci and the complications imposed by non-trivial selection of candidate genes for experimental follow-up have driven the development of candidate gene prioritization tools. Bioin-

formatic tools like DEPICT (Pers et al., 2015) and Fuma (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017) integrate results from the above mentioned databases, making the selection of candidate genes for experimental follow-up more straightforward and additionally provide the user with automated pathway and tissue-enrichment analysis. Ultimately, all bioinformatic tools provide predictions that can be used for candidate gene prioritization. This implies that the candidate genes need to be probed in a suitable (animal) model to confirm or refute the association. Ideally, the large number of GWAS prioritized candidate genes would be interrogated using a model system that allows for quick and comprehensive screening of these genes. The zebrafish possesses unique qualities to take up this challenging task.

The zebrafish genome

The zebrafish (*Danio rerio*), a tropical freshwater fish belonging to the class of teleost, has become a valuable tool in candidate disease studies and drug research (Lieschke & Currie, 2007; MacRae & Peterson, 2015). Zebrafish have a short generation time, larval transparency, easy genetic accessibility thanks to external fertilization and development and produce large numbers of offspring in short time. Commercially available high-throughput systems, such as the vertebrate automated screening technology (VAST) BioImager (Pardo-Martin et al., 2013), allow for image-based screening of up to 150 larvae per day. These factors combined make the zebrafish an attractive model to conduct high-throughput, image-based genetic screens to help translate GWAS-prioritized candidate genes.

A comprehensive analysis untangling the relationship between the human and the zebrafish genome showed that 71.4% of human genes have at least one zebrafish orthologue. Most of the orthologues have a one-to-one relationship, which is followed by one-to-many, indicating that one human gene can have multiple orthologues in zebrafish (Howe et al., 2013). This is likely due to the teleost-specific genome duplication that the common ancestor of the zebrafish has undergone about 340 million years ago (Meyer & Schartl, 1999). This is also reflected in the higher number of currently annotated protein coding genes in zebrafish (n=25,592, genome build GRCz11, February 2019). This implicates that some orthologues may have undergone neo- (the acquisition of a novel function) or subfunctionalization (division of functions to complement each other). In a study by LaFave and colleagues, high-throughput sequencing revealed about 17 million variants in a defined reference zebrafish strain (LaFave, Varshney, Vemulapalli, Mullikin, & Burgess, 2014). Genetic variation is very common and has been reported at a rate of 1 single nucleotide variant per 100 bases (Guryev et al., 2006). As opposed to mice, zebrafish are vulnerable to inbreeding which results in a decrease in

fitness. Hence, the immense genetic variation present in the zebrafish genome has to be considered when designing genetic experiments. In order to make use of the zebrafish as a model system, the ability to modify the genome fast and efficient would be beneficial. Initially, zebrafish mutants were generated by random chemical mutagenesis, followed by phenotyping and positional cloning to identify the mutated gene (Varshney & Burgess, 2014). Also, many strategies for insertional mutagenesis were developed (Varshney & Burgess, 2014). The zebrafish mutation project aimed to systematically mutate every protein-coding gene in the zebrafish genome, employing a chemical mutagenesis and large-scale sequencing strategy (Kettleborough et al., 2013). Targeted and precise genome editing became available through zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Varshney & Burgess, 2014). However, design of ZFNs and TALENs are laborious. The zebrafish is amenable to fast and efficient genome modification with the aid of clustered regularly interspaced short palindromic repeats (CRISPR) and Cas associated protein 9 (Cas9).

CRISPR-Cas9

A brief history

When Ishino and colleagues observed short repeat sequences in bacterial DNA for the first time in the late 1980s (Ishino, Shinagawa, Makino, Amemura, & Nakata, 1987), it remained elusive which impact this and subsequent discoveries would have. Due to increasing sequencing efforts in the beginning of the 2000s, it became apparent that short tandem repeats are widespread in bacteria and archaea (Mojica, Diez-Villasenor, Soria, & Juez, 2000). Shortly after, the clustered regularly interspaced short palindromic repeats were coined “CRISPR” due to their particular characteristics (Jansen, Embden, Gaastra, & Schouls, 2002). Simultaneously, a co-expressed cluster of genes directly adjacent to the CRISPR was studied, known as the CRISPR associated sequences (Cas) (Makarova, Aravind, Grishin, Rogozin, & Koonin, 2002). The genes were predicted to be involved in DNA repair. Around 2005, evidence was presented that the short spacer sequences are of foreign nature and integrated into the bacterial DNA (Mojica, Diez-Villasenor, Garcia-Martinez, & Soria, 2005). The parts of the puzzle were slowly pieced together with an article describing similarities of the CRISPR-Cas system to the eukaryotic RNA interference system that is used for regulation of gene expression and immunity by modulating RNA activity (Makarova, Grishin, Shabalina, Wolf, & Koonin, 2006). Two decades after the first discovery of CRISPR sequences, experimental evidence showed that the CRISPR-Cas system is part of an adaptive bacterial immune system to fight viruses and plasmids (Barrangou et al., 2007).

Bacteria can incorporate short sequences of foreign DNA into their own genome. These so called spacers are arranged sequentially and separated by common repeat sequences. Upon infection, the CRISPR cluster is transcribed, resulting in the pre-CRISPR RNA. Simultaneously, the Cas gene cluster is transcribed. These proteins (nucleases) are necessary for processing of the pre-CRISPR RNA into single, mature CRISPR RNAs. The CRISPR RNAs then guide the Cas-protein complex to the complementary sequence to initiate target cleavage, hence degrading the foreign DNA. Shortly thereafter, the protospacer adjacent motif (PAM) was described that is required for Cas-cleavage (Deveau et al., 2008). Furthermore, it was shown that Cas enzymes mediate the integration of novel foreign DNA into the CRISPR cluster (Barrangou et al., 2007). Since then, a variety of different Cas-systems have been described that rely on different PAM sequences and have different modes of target cleavage (Adli, 2018; Shmakov et al., 2015).

Cas9 as a tool for genome-editing

The possibility of adapting the CRISPR-Cas system for genome editing became available in 2012. Jinek and colleagues demonstrated that the endonuclease Cas9 isolated from *Streptococcus pyogenes* can be directed to any DNA sequence using a single guide RNA (gRNA) (Jinek et al., 2012). In order to achieve this, a target-specific CRISPR RNA was fused to the trans-encoded CRISPR RNA, which is essential for CRISPR RNA maturation and target cleavage (Deltcheva et al., 2011). Cas9 induces in complex with the gRNA double-strand breaks in the targeted sequence three bp upstream of the PAM, resulting in blunt end DNA. The host cell then attempts to repair the double-strand break via non-homologous end joining or homology directed repair. The repair often results in errors, thus altering the DNA sequence at the cleavage site. Thereby, the possibility of fast and efficient genome editing became available, as opposed to the laborious design of ZFNs or TALENs.

Having shown that Cas9 could be used as a fast and efficient enzyme for genome editing *in vitro*, significant efforts focused on unraveling the sensitivity and specificity of the approach. The first evidence that CRISPR-Cas9 works *in vivo* came from zebrafish experiments, in which ten different genes were targeted with a variable mutation frequency (Hwang et al., 2013). After the first *in vivo* demonstration in zebrafish, other organisms were also successfully genetically modified using CRISPR-Cas9 (Friedland et al., 2013; Wang et al., 2013). This sparked two pivotal questions. First, what determines gRNA efficiency? And secondly, how often will off-target cleavage occur?

Since then, numerous studies tried to disentangle the recipe for designing efficient gRNAs. In one of the first reports, evidence was provided that

gRNAs with >50% GC-content, a guanine directly adjacent to the PAM and a 5' GG displayed higher efficiency than randomly chosen sequences in zebrafish (Gagnon et al., 2014). Next, about 1920 gRNAs were studied in zebrafish larvae and it was observed that truncated gRNAs (1-2 nt shorter) were more effective than longer (20-22nt) gRNAs, nucleotide composition affects gRNA activity and guanine-rich gRNAs promote stability and thus activity (Moreno-Mateos et al., 2015).

Using a zebrafish codon-optimized version of Cas9, Jao and colleagues reached higher mutagenesis frequencies, showed that mutations can be inherited, and successfully conducted biallelic and multiplexed (several genes simultaneously) targeting (Jao, Wente, & Chen, 2013). This is especially desirable, since many genes have two orthologues and can thus be targeted in the same larva. Additionally, it reduces the number of animals that need to be injected and screened, and thereby cost and time. The approach was extended in 2015, when ten targets were co-injected with Cas9 using a cloning-free approach to generate gRNAs to enable high-throughput gene targeting (Varshney et al., 2015). It also paved the way for mutant screening in the F1 generation, as opposed to the F2 generation, as well as the attractive possibility to potentially follow-up on large numbers of GWAS-prioritized candidate genes.

On the quest to elucidate why some gRNAs *in vitro* predicted activity is not reflected *in vivo*, Thyme and colleagues conducted follow-up experiments (Thyme, Akhmetova, Montague, Valen, & Schier, 2016). They concluded that a portion of gRNAs fail due to refractory sequences, such as the binding motif of transcriptional repressor CTCF. Furthermore, gRNAs can fold into unfavorable secondary structures and hence affect Cas9 activity, which can largely be prevented by refolding the gRNAs. Also, there is considerable competition of the gRNAs for Cas9. A swift and convenient method to assess gRNA activity *in vivo* is the CRISPR somatic tissue activity test, which relies on quantifying the fragment length after fluorescent polymerase chain reaction (PCR) subsequent to an injection of the gRNAs that are to be tested (Carrington, Varshney, Burgess, & Sood, 2015).

To extend the CRISPR-Cas9 toolkit in zebrafish, constructs for spatial and temporal control of Cas9 have been designed (Ablain, Durand, Yang, Zhou, & Zon, 2015; Yin et al., 2015).

Similarly, many efforts have been carried out in cell culture studies. Mismatches between the gRNA and target sequence evoked different consequences dependent on the position of the mismatch in the sequence (Hsu et al., 2013; Pattanayak et al., 2013). Subsequent higher throughput studies discovered additional design rules (Doench et al., 2016; Doench et al., 2014).

Concentration and efficiency of gRNAs were shown to influence off-target effects *in vitro*. Off-target issues have been addressed by modulating Cas9 delivery, re-engineering Cas9 (nCas9 and dCas9), temporal and spatial

control of expression of Cas9 and modulating gRNA length, as discussed by Adli (Adli, 2018). The described findings prompted the development of gRNA design tools, such as ChopChop (Labun, Montague, Gagnon, Thyme, & Valen, 2016) and CRISPRscan (Moreno-Mateos et al., 2015).

As previously mentioned, the cell can repair the double strand break via homology directed repair or non-homologous end joining. Studies have shown that the choice of repair mechanism is dependent on availability of repair enzymes and cell cycle stage (Shibata, 2017; Shibata et al., 2011). It has also been demonstrated that the induced mutations are non-random and depend on target sequence composition (van Overbeek et al., 2016). A study based on these findings has generated a catalogue of gRNAs and their induced mutations in order to facilitate prediction of the Cas9 induced mutation pattern (Allen et al., 2018).

The CRISPR toolkit has been extended beyond genome editing (as reviewed by Adli (Adli, 2018)). The *Streptococcus pyogenes* Cas9 has been re-engineered for precise base pair editing (Gaudelli et al., 2017), and a catalytically ‘dead’ Cas9 (dCas9) has been employed for modulating gene expression (Qi et al., 2013). Various dCas9-fusion enzymes have been used for epigenome editing and live-chromatin imaging (Chen et al., 2013). In terms of using Cas9 as a therapeutic in humans, several points have been raised. First, the most commonly *Streptococcus pyogenes* Cas9 is relatively large. Therefore, smaller Cas9 versions have been identified in other bacterial species, with the trade off of requiring more complex PAM sequences (Kim et al., 2017). Second, it has been shown that a large proportion of humans might carry antibodies against Cas9 derived from *Streptococcus pyogenes* and *Staphylococcus aureus*, since they are common species to infect humans (Charlesworth et al., 2019). Identifying orthologous enzymes from other species might circumvent this problem.

In conclusion, whereas CRISPR-Cas9 cannot yet be used safely in the clinic, the system has emerged rapidly into a versatile genome-editing tool in model organisms, especially to study disease candidate genes.

The human heart

The steady beating of the heart facilitates the continuous blood flow to organs and tissues in order to ensure oxygen and nutrient supply. The human heart consists of four chambers: the left and right atrium as well as the left and right ventricle. Both atria and ventricles are separated by septa, the interatrial and interventricular septum, respectively. Similarly, the atrioventricular septum separates the atria from the ventricles. To ensure unidirectional flow of blood, two valves are located at the interface of the atria and ventricles (atrioventricular valves). Two semilunar valves are present at the exit of the heart to prevent backflow into the ventricles after contraction.

The heart is the first functioning organ in vertebrates. Initially, cells originating from the mesoderm form so-called cardiogenic cords. The cords further differentiate into two luminal endocardial tubes that show electrical activity. Merging of the two endocardial tubes gives rise to a beating tubular heart at 21 days post fertilization (dpf). The tubular heart differentiates rapidly into five distinct regions: the truncus arteriosus, bulbus cordis, primitive ventricle and atrium, and the sinus venosus. These regions will give rise to distinct features of the mature heart. The four-chambered partitioning of the heart is first visible at 28 dpf. The following weeks are characterized by generation and maturation of the septa and valves.

The heart can be affected by many pathological conditions, which are broadly categorized into congenital heart disease (structural cardiac abnormalities present at birth), cardiomyopathies (impaired cardiac muscle function), valvular heart disease (impaired valve function), and arrhythmogenic disorders (conditions that affect heart rate/rhythm). Diseases affecting the vasculature (such as coronary artery disease) that reinforce certain pathological cardiac conditions will be discussed later. Some of the above mentioned groups of morbidities might ultimately lead to heart failure.

Contraction of the atrial and ventricular cardiomyocytes needs to be initiated and carefully orchestrated to ensure efficient pumping of the blood.

The cardiac conduction system

Heart rhythm is generated in the sinoatrial (SA) node, a group of specialized cardiomyocytes located in the myocardium of the right atrium that generate the electrical impulse. The SA node is also commonly referred to as the pacemaker, or the pacemaker cells. To begin with, the SA node generates a pulse that facilitates simultaneous contraction of both atria. The electrical impulse travels via gap junctions that connect two adjacent cells. The signal travels to the atrio-ventricular (AV) node, which constitutes the electrical interface between atrium and ventricle. The AV node mediates a short delay to ensure complete clearance of blood from the atria to the ventricles. The signal is then propagated via the His-bundle to the Purkinje fibers, which innervate the ventricles. Ultimately, the ventricles contract and pump the blood into the pulmonary and systemic circulation. Both the AV-node and the His-bundle/Purkinje fibers have pacemaking capabilities, although at a much lower rate than the SA-node. This constitutes an emergency mechanism, in case the SA node ceases to function.

Without any external influence, the sinus rhythm is established at between 60-100 beats per minute (bpm). Abnormally high heart rate is referred to as tachycardia, whereas bradycardia denotes atypically low heart rate.

Myocyte contractions occur as a result of action potentials. An unequal distribution of intra- and extracellular ions is required to generate an action potential. Cardiac myocytes are normally negatively charged, while the ex-

tracellular environment is positively charged. Myocytes have a high intracellular concentration of K^+ , whereas Na^+ , Ca^{2+} and Cl^- are predominantly distributed in the extracellular space. The resulting potential energy is known as the membrane potential. Ion channels and pumps that are specific for certain ions maintain homeostasis.

It is important to distinguish between the specialized pacemaking cardiomyocytes and the “regular” contractile cardiomyocytes. Regular cardiomyocytes maintain a resting potential at about -90 millivolt (mV) by K^+ selective permeable ion channels and Na^+/K^+ exchange pumps. Upon a stimulus that is strong enough to reach the threshold, an action potential is generated, which results in the opening of Na^+ voltage gated channels and thus Na^+ influx. The voltage-gated Na^+ channels close upon reaching a membrane potential of about +30 mV. The depolarization is then followed by a plateau phase that is mainly mediated by the opening of slow Ca^{2+} channels, which makes the membrane potential slowly decrease from +30 mV to ~0 mV. The calcium influx mediates the release of Ca^{2+} that is stored in the sarcoplasmic reticulum, facilitating contraction of the cells. The long plateau phase also prevents excitation by premature subsequent action potentials. Lastly, the slow Ca^{2+} channels close, and repolarization is mediated by K^+ efflux.

The pacemaker cells are characterized by spontaneous depolarization. The membrane potential gradually rises from -60 mV to about -40 mV by steady Na^+ influx through voltage-gated channels. This culminates in the generation of an action potential, in which fast voltage gated Ca^{2+} channels open. At a membrane potential of +5 mV, repolarization is initiated by opening of K^+ channels. Hence, there is no plateau phase in the pacemaker cells. The spontaneous depolarization is attributable to voltage-gated channels belonging to the hyperpolarization-activated cyclic nucleotide-gated (HCN) channel family. These so-called “funny” channels are characterized by being nonselective, being permeable for both K^+ and Na^+ , and by hyperpolarization. Hence, the HCN channels will open upon hyperpolarization at around -60 mV, thereby facilitating influx of positively charged ions and thus generating a new action potential. Hence, the HCN channels are mediating the heart’s autorhythmicity (Ludwig et al., 1999; Stieber et al., 2003).

The electrical activity of the heart is most commonly measured using an electrocardiogram (ECG). The ECG is divided into three main waves, the P-wave, the QRS-wave, and the T-wave. The P-wave reflects the atrial depolarization. Subsequently, the length of the PR-segment indicates atrial and AV-node depolarization. The QRS-wave reflects ventricular depolarization, whereas the ST-segment serves as a readout for the depolarization length. Finally, the T-wave corresponds to repolarization of the ventricles. Many arrhythmogenic, structural and ischemic disorders can be diagnosed from the ECG. For example, sinoatrial arrests are characterized by a >2 second delay of the next action potential generated by the SA node. Atrial fibrillation, the rapid and irregular contractions of the atria that are associated with higher

risk of stroke and heart failure, is represented by irregularities and an absent P-wave in the ECG. Short and long QT-syndrome, as the name implies, are disorders characterized by atypically short or long QT-intervals, that significantly increase the risk of life-threatening arrhythmias and sudden cardiac death. Myocardial infarction, an ischemic condition in which blood flow to the heart is blocked, is reflected by alterations in the T-wave and ST-segment.

Modulation of heart rate and rhythm

Heart rate and rhythm are generated by the pacemaker cells, but may be modulated by the autonomous nervous system (ANS). The ANS is divided into the sympathetic and the parasympathetic nervous system. Both the sympathetic and parasympathetic system constantly modulate vital functions, and thus instigate alteration of heart rate. The function of the sympathetic system is commonly ascribed to energy mobilization, whereas the parasympathetic system serves as its counterweight, being responsible for restorative functions.

The medulla oblongata constitutes the “cardiovascular control center” in the brain. It receives and processes input from various visceral receptors, such as baroreceptors, chemoreceptors and proprioceptors to monitor the state of the body to maintain homeostasis. Baroreceptors are mechanoreceptors in the blood vessels responsible for monitoring blood pressure. Chemoreceptors respond to changes in metabolism, for example transmitting changes of metabolic byproducts as a consequence of physical activity. Proprioceptors transmit information related to position and movement. All signals travel via afferent nerves to the medulla oblongata that can initiate an appropriate alteration of heart rate via efferent nerves of the sympathetic and parasympathetic nervous system innervating the heart. An increase in heart rate is mediated through release of norepinephrine by the accelerans nerve. Norepinephrine binds to beta-1 adrenergic receptors in the cardiac tissue. Conversely, acetylcholine can be released from the vagus nerve to lower heart rate by binding to M₂ muscarinic receptors. Heart rate variability serves as a readout for the vagal tone of the ANS.

Heart rate variability

The inter-beat variation of consecutive heartbeats in time is referred to as heart rate variability (HRV). There is a strong inverse relationship between heart rate and HRV. HRV can be quantified using different methods. Most commonly, time-domain methods are used, e.g. the standard deviation of the normal-to-normal RR interval (SDNN) and the root mean square of successive heart beat interval differences (RMSSD). Frequency domain methods are used as well, like high and low frequency (Task Force of The European

Society of Cardiology and The North American Society of Pacing and Electrophysiology, 1996).

HRV can be used as a marker to assess autonomic imbalance, especially vagal tone that reflects the activity of the parasympathetic branch. Lower vagal tone, as reflected by lower HRV, has been associated with numerous morbidities such as diabetes (Schroeder et al., 2005), renal disease (Brotman et al., 2010) and hypertension (Goit & Ansari, 2016). Furthermore, it predicts mortality in patients having suffered from a myocardial infarction (Buccelletti et al., 2009). Higher risk was also conferred for incident coronary artery disease (CAD) in a cohort study from 1997 (Liao et al., 1997). Finally, lower HRV is associated with higher all-cause mortality in middle-aged (hazard ratio of 2.0) and elderly men (hazard ratio of 1.4) (Dekker et al., 1997). In epidemiological studies, a high heart rate has been associated with a higher prevalence of cardiovascular disease as well as with all-cause mortality (Dyer et al., 1980; Gillum, Makuc, & Feldman, 1991). Whether low HRV is causal for CAD or is merely a risk correlate remains unresolved.

GWAS of cardiac rate and rhythm

Before the advent of GWAS, mutations in several genes (mostly ion channels) that show familial aggregation were known to cause different forms of long QT-syndrome (*CACNA1c*, *KCNJ2*, amongst others), sick sinus syndroms (*SCN5A*, *HCN4*), or familial atrial fibrillation (*KCNQ1*, *KCNE2*) for example, as listed by Priori and colleagues in 2006 (Priori & Napolitano, 2006). To date, several GWAS have examined genetic associations with resting heart rate. The first GWAS examining heart rate was published in 2009 and reported the association of two loci (Cho et al., 2009). Following, in a 2010 GWAS one novel locus was reported to influence heart, as well as four affecting PR-interval and QRS-duration, respectively (Holm et al., 2010). A subsequent meta-analysis of GWAS identified six loci with resting heart rate in about 39.000 people (Eijgelsheim et al., 2010). With a huge increase in sample size (about 181.000 individuals), a meta-analysis by den Hoed and colleagues identified 14 novel loci influencing resting heart rate (den Hoed et al., 2013). In a recent GWAS, 46 additional novel loci were reported (Eppinga et al., 2016). Furthermore, exome studies contributed additional novel loci (van den Berg et al., 2017). GWAS have also been conducted examining electrophysiological traits such as PR-interval (van Setten et al., 2018), QT-interval (Arking et al., 2014), and morbidities arising from impaired cardiac electrophysiology, such as atrial fibrillation (Ellinor et al., 2012; Nielsen et al., 2018; Roselli et al., 2018).

Functional annotation of loci that were associated with electrophysiological traits were enriched for calcium signaling (Arking et al., 2014), cardiac signal transduction, development and ion channels (Ellinor et al., 2012; Nielsen et al., 2018; Roselli et al., 2018; van Setten et al., 2018). This is like-

ly reflecting a (partially) shared etiology, since many loci identified in a GWAS for PR-interval show substantial pleiotropy with other electrophysiological traits (van Setten et al., 2018). Studies investigating the heart rate response to exercise commonly flag loci harboring genes involved in neuronal development and adrenergic signaling (Ramirez et al., 2018; Verweij, van de Vegte, & van der Harst, 2018).

The genetic basis underlying heart rate variability remains largely elusive and has so far not been examined by GWAS. Heritability of HRV has been demonstrated to be 25%-71% in twin and family studies (Snieder, Boomsma, Van Doornen, & De Geus, 1997). Examining the genetic basis of HRV would provide meaningful information about vagal regulation of heart rate and rhythm.

The den Hoed heart rate GWAS meta-analysis not only conducted an extensive *in silico* candidate gene prioritisation, but also performed experimental follow-up in the fruitfly and the zebrafish (den Hoed et al., 2013). Their experimental results suggested a role for 20 out of 31 tested candidate genes in regulation of heart rate, by influencing processes such as cardiac development, signal transduction and pathological cardiac conditions (den Hoed et al., 2013). This highlights the strength of a combined GWAS/model system approach to identify novel genes and mechanisms regulating rate and rhythm.

Since murine models show differences in cardiac electrophysiology and rate (Poon & Brand, 2013), and are not amenable for high-throughput screening, the zebrafish provides unique qualities to rapidly screen a large number of candidate genes that have been prioritized for cardiac rate, rhythm as well as conduction.

The zebrafish heart

The zebrafish is a prominent model organism to study cardiac development and disease (Bakkers, 2011). The zebrafish heart consists of two chambers, a single atrium and ventricle, which are not separated. The sinus venosus leads into the atrium, whereas the ventricle is connected to the bulbus arteriosus. Comparable to humans, the zebrafish possess an AV-valve, and a bulboventricular valve. Although there are no “mature” His-bundles and Purkinje fibers described in the zebrafish, the evolutionary preceding ventricular trabeculae are thought to serve analogous functions (J. Liu et al., 2010).

Despite differences in morphology, cardiac development and the genes that play a role in this process are well conserved between the two species (Glickman & Yelon, 2002; Stainier, 2001). Myocardial precursor cells of a mesodermal origin migrate and fuse to produce a cardiac cone, which differentiates into a primitive cardiac tube at about 24 hours post fertilization (hpf). The cardiac tube shows peristaltic, rhythmic contractions. At around 30 hpf, the atrium and ventricle are distinguishable. Looping of the heart is

initiated at 33 hpf, while 37 hpf marks the time when cardiac valve formation is initiated (Poon & Brand, 2013). Hence, as described above, the sequence of the initial developmental stages of the zebrafish heart are comparable with the development of the mammalian heart. Zebrafish embryos (0-3dpf) and larvae (3-30dpf) possess another unique feature. Due to their small size, oxygen can diffuse passively into the tissue, and a functional cardiovascular system is thus not required for survival in early development (Pelster & Burggren, 1996). This allows survival in presence of severe cardiac/cardiovascular defects that would be fatal in other organisms.

Zebrafish cardiac electrophysiology

The electrophysiological properties of the cardiac conduction system in zebrafish are very similar to humans, as exemplified by the distinct P-, QRS-complex and T-wave of the zebrafish ECG (C. C. Liu, Li, Lam, Siu, & Cheng, 2016). Four distinct stages have been identified that describe the development of the cardiac conduction system (Chi et al., 2008). A linear electrical wave that travels along the cardiac tube is the hallmark of stage 1 (20-24 hpf). Stage 2 (36-48 hpf) marks the development of the delay in AV conduction, representing the onset of the cardiac conduction system. The development of the conduction network in the ventricle marks stage 3 (72-96hpf). Finally, stage 4 (until 21 dpf) displays the maturation of the conduction system into a fully developed apex-to-base excitation. These stages suggest the presence of a pacemaker in zebrafish. Indeed, pacemaker cells are present and have been pinpointed to the sinoatrial ring, the equivalent of the human SA node (Arrenberg, Stainier, Baier, & Huiskens, 2010; Tessadori et al., 2012).

The action potential and underlying driving ion currents show many similarities between humans and zebrafish. The major currents driving depolarization and repolarization are analogous in zebrafish and humans (Leong, Skinner, Shelling, & Love, 2010). The zebrafish action potential however lacks the fast phase-1 repolarization phase that is observed in human cardiomyocyte action potentials. In addition, there are some differences in ion channels and their respective sub-unit make-up (Vornanen & Hassinen, 2016). Furthermore, it has been demonstrated that until 3dpf, the action potential of contractile cardiomyocytes is driven by influx of Ca^{2+} ions, and subsequently transitions into an intermediate phenotype, the atrial action potential being driven by Ca^{2+} , and Na^{+} driving the ventricular action potential. In the mature zebrafish heart, action potentials in both chambers are completely driven by Na^{+} (Hou, Kralj, Douglass, Engert, & Cohen, 2014).

Zebrafish as a model for system for cardiac traits

Due to the ample analogies, many cardiac phenotypes have been examined using the zebrafish. Heart rate and heart rate variability have been quantified by recording videos using either bright-field or fluorescent imaging (Tg(*cmcl2:gfp*)) (Burns et al., 2005; Martin et al., 2019). *NKX2.5*, a gene that has been prioritized as a candidate in GWAS for heart rate (den Hoed et al., 2013) and atrial fibrillation (Nielsen et al., 2018) has been demonstrated to be necessary for maintaining cardiac chamber identity in early development in zebrafish (Harrington, Sorabella, Tercek, Isler, & Targoff, 2017). Further, *nkx2.5* loss-of-function mutants were characterized by higher heart rate and lower HRV as compared to wildtype controls. This is likely the result of impaired cardiac conduction, as a consequence of the cardiac chamber identity loss.

Variants near *CACNA1c* have been identified in heart rate GWAS (Eppinga et al., 2016) and mutations in the gene were shown to be involved in long QT syndrome (Priori & Napolitano, 2006). In the zebrafish, mutations in *cacna1c* influenced ventricular growth and thickening (Rottbauer et al., 2001). *KCNH2*, a gene known to be involved in human QT-syndrome (Priori & Napolitano, 2006), has been prioritized as a candidate gene in GWAS for QT interval and atrial fibrillation. Loss-of-function of the zebrafish orthologue (*kcnh2*) was shown to successfully model long-QT syndrome (Arnaout et al., 2007). In a screen conducted by Milan and colleagues, 15 genes were identified to be influencing cardiac repolarization (Milan et al., 2009). Furthermore, one of the identified genes coincides with a signal from a GWAS locus for QT-interval (*gins3*). Subsequent demonstration that knockdown of *gins3* resulted in shorter ventricular action potentials identifies *gins3* as a likely causal gene for this locus. Loss-of-function mutants of *trpm7* were displaying slower heart rate and sinoatrial pauses, as a consequence of *trpm7*s function to modulate *hcn4* expression (Sah et al., 2013). Variants close to *TRPM7* have been identified in GWAS for QT-interval (Arking et al., 2014). Last but not least, morpholino-mediated knockdown of *hcn4* resulted in bradycardia, sinoatrial pauses and has been demonstrated to serve as a zebrafish model for human sick sinus syndrome (Jou et al., 2017).

These examples illustrate the suitability of the zebrafish to examine GWAS-prioritized candidate genes for cardiac rate, rhythm and morbidities. High-throughput imaging systems (Pardo-Martin et al., 2013) combined with a multiplexed CRISPR-Cas9 set-up (Varshney et al., 2015) enable simultaneous examination of several candidate genes, which is essential to increase our understanding of disease etiology and potentially identify novel drug targets.

The vasculature

The cardiovascular system transports blood through vessels to ensure oxygen and nutrient supply, metabolic (waste) product and hormone transport; homeostasis can be mediated; and an immune response can be facilitated. In mammals, the circulation is divided into a systemic and a pulmonary circuit. Oxygenated blood travels through the body using the arterial system, deoxygenated blood flows back to the heart via veins. The heart then pumps the deoxygenated blood into the pulmonary circulation, which returns oxygenated blood to the heart, ready to be distributed into the systemic circuit again. The blood vessels are made up of three specific tissue layers: the tunica intima, media and externa. The cardiovascular system is subject to many diseases, coronary artery disease being the most common one. Cardiovascular diseases are the leading cause of death worldwide.

Lipid metabolism

Lipids serve vital functions in the body, for example as signaling molecules, energy storage and as a cell membrane component. Lipids can be ingested (exogenous path) or synthesized by the liver (endogenous path). In the following, I will explain the differences between the exogenous and endogenous path, as well as the specific lipoprotein particles that are needed for transport of the hydrophobic lipids in the circulation. The exogenous path covers the ingestion of dietary lipids, which are mainly constituted by triglycerides, cholesterol and phospholipids. Dietary triglycerides are first hydrolysed, and then re-synthesised in the enterocytes of the small intestine. The enterocytes form so-called nascent chylomicrons, which consist mainly of triglycerides, apolipoprotein (APO)B48 and cholesterol. Nascent chylomicrons are transported to the blood circulation via lymphatic vessels. As soon as they enter the circulation, chylomicrons are matured by uptake of further apolipoproteins, such as APOC2 and APOE from high-density lipoprotein (HDL). APOC2 is crucial for lipoprotein lipase (LPL) activation, which hydrolyses the triglycerides into free fatty acids, which are taken up by different tissues for energy usage and storage. The resulting chylomicron remnants are removed from the circulation by uptake from hepatocytes. This constitutes the exogenous path. In the endogenous path, triglycerides are synthesized in hepatocytes, assembled with ApoB100, and released into the blood stream as very low-density lipoprotein (VLDL) particles. In the bloodstream, VLDL particles acquire APOC2 and APOE from HDL particles. Similar to chylomicrons, LPL releases free fatty acids from VLDL, continuously shrinking particle size. Eventually, VLDL particles become intermediate-density lipoprotein (IDL). IDL particles have a roughly equal load of cholesterol and triglycerides. LPL then further shrinks IDL particles to low-density lipoprotein (LDL) particles that mainly carry cholesterol. LDL parti-

cles are cleared from the circulation by binding to low-density lipoprotein receptor (LDLR) on hepatocytes (parts of this section have been adapted from (von der Heyde, 2017).

In a process called reverse-cholesterol transport, cholesterol from peripheral tissues is transported back to the liver. This mainly occurs via HDL. APOA1, the main apolipoprotein of HDL, takes up cholesterol by interaction with ATP-binding cassette transporter 1. Subsequently, the cholesterol gets esterified by lecithin-cholesterol acyltransferase (LCAT). HDL can be then taken up directly by hepatocytes via binding to scavenger-receptor B1. Alternatively, cholesterol esters can be exchanged with LDL or VLDL for triglycerides, mediated by cholesteryl ester transfer protein (CETP).

Dyslipidemia, i.e. deviations of normal lipid levels in the blood, are a major driving factor of atherosclerosis, which ultimately leads to coronary artery disease.

Etiology of coronary artery disease

Coronary artery disease is the consequence of the building up of atherosclerotic plaques in the coronary vessels during the course of a lifetime (Lusis, 2000), which can have fatal consequences such as myocardial infarction. Atherosclerosis is characterized by chronic inflammation of the blood vessels, which is promoted by elevated lipoprotein levels. Due to their smaller size relative to other lipoprotein-particles, LDLc and chylomicron remnants can diffuse into the intima of the blood vessel and undergo oxidative modification. This initiates an inflammatory response of endothelial cells, which secrete pro-inflammatory and adhesion molecules. Attracted monocytes will adhere to the surface and differentiate into macrophages, which will scavenge the oxidized lipoproteins. Eventually, the macrophages will turn into lipid-loaded foam cells. Next, platelets accumulate at the site of inflammation. By secreting platelet-derived growth factor, platelets encourage smooth muscle cells to migrate into the intima. Smooth muscle cells secrete collagen and other extracellular matrix molecules, forming a fibrous cap. Rupture of the atherosclerotic plaque will trigger thrombosis, potentially resulting in a myocardial infarction, stroke or peripheral ischemia.

Epidemiological studies for CAD have identified male sex, age, hypertension, diabetes, obesity, smoking, a sedentary lifestyle, and an unhealthy diet as risk factors (Mack & Gopal, 2014). High concentrations of low-density lipoprotein cholesterol (LDLc) and triglyceride-rich lipoproteins in blood, as well as low concentrations of high-density lipoprotein cholesterol (HDLc) have also been associated with an adverse risk profile (Castelli, Anderson, Wilson, & Levy, 1992; Kannel, 1976).

The genetics of CAD

Researchers have tried to understand the molecular basis of CAD for decades. In the late 1980s, variants in *APOE*, *CETP* and *LPL* have been associated with atherosclerosis, while mutations identified in *LCAT* resulted in HDL deficiency (listed in (Scheuner, 2003)). Linkage studies identified further mutations in genes involved in lipid metabolism, such as *LDLR* (Lehrman et al., 1985), and *APOB* (Soria et al., 1989) in families affected by familial hypercholesteremia (i.e. pathologically elevated levels of LDL). Mutations in genes such as LDLR adaptor protein 1 (Garcia et al., 2001), as well as ATP-binding cassette sub-family G member 5 and 8 (Berge et al., 2000) were shown to cause autosomal recessive hypercholesterolaemia. Also, genes involved in processes other than lipid metabolism, such as matrix metalloproteinase 9 (extracellular matrix remodelling) or interleukin 6 (immune system) had been associated with CAD by the early 2000s (Scheuner, 2003). In 2003, proprotein convertase subtilisin/kexin type 9 (PCSK9) was linked to hypercholesterolaemia (Abifadel et al., 2003). In conclusion, monogenic mutations in these genes have large effects, but are not common in the population. In contrast, polygenic drivers of CAD are common but are characterized by smaller effects.

The first GWAS for CAD was published in 2007 and associated a locus on chromosome 9 (9p21), which resulted in a 30% increased risk of coronary heart disease in homozygous individuals carrying the risk allele (McPherson et al., 2007). Following, several GWAS were published reporting mostly single novel loci. Large international consortia were instrumental in detecting many novel CAD associations, such as the Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) consortium (Schunkert et al., 2011) (13 novel loci) and the Coronary Artery Disease (C4D) Genetics Consortium (2011) (5 novel loci). Lastly, the combined effort of both (CARDIoGRAMplusC4D) identified 15 novel loci in 2013 (C.ARDIoGRAMplusC4D Consortium et al., 2013) and additional 10 novel loci in 2015 (Nikpay et al., 2015). Since then, a recent GWAS using UK biobank data identified 64 novel loci (van der Harst & Verweij, 2018).

Thus, to date more than 160 genetic loci have now been robustly associated with CAD risk in adults of European descent. In addition, exome chip and whole-exome sequencing studies have recently provided information on rare variants that are not included on genotyping chips used by GWAS, due to their low minor allele frequency. This led to the identification of rare variants in *LDLR* and *APOA5* affecting CAD risk (Do et al., 2015). Pathway analysis of the CAD-associated loci have highlighted that a range of biological mechanisms are involved in CAD pathology, such as disturbances in lipid biology, vascular remodeling and vascular tone (Nikpay et al., 2015). However, translation of results from GWAS to function is complicated, since most variants are intergenic and pinpointing causal genes is not

straightforward. While the discovery of potential new pathways involved in CAD might open opportunities for novel therapeutic intervention, most loci remain uncharacterized until now.

As for CAD, GWAS and exome analysis have also examined genetic associations with blood lipids levels. In 2007, one locus was reported to be associated with triglyceride levels (Saxena et al., 2007). Subsequently, several novel loci were identified (Willer et al., 2008). Teslovich and colleagues reported 95 novel associations in 2010 (Teslovich et al., 2010). Similar to CAD, large consortia were initiated and reported another 62 novel loci (Willer et al., 2013). In a recent GWAS comprising over 300,000 individuals, 118 novel associations were revealed (Klarin et al., 2018). These studies highlight loci containing known genes contributing to dyslipidemias and their corresponding drug targets, as well as novel candidate genes for lipid biology. Thus, more than 350 loci have been identified as being associated with circulating lipids, harboring another huge reservoir of potential targets for pharmacological intervention.

A successful example of translating results from GWAS for CAD into causal transcript and function comes from sortilin 1 (*SORT1*) in locus 1p13. Results from experiments in a mouse model showed that Sort1 is involved in hepatic lipoprotein production. Sort1 deficiency results in impaired VLDL export from hepatocytes, thereby affecting CAD risk (Musunuru et al., 2010). Another successful example is A Disintegrin And Metalloproteinase With Thrombospondin Motifs 7 (*ADAMTS7*), which is involved in vessel wall remodelling. In vitro experiments demonstrated that upregulation of *ADAMTS7* increases its proteolytic activity and promotes migration of smooth muscle cells, thereby contributing to atherogenesis (Pu et al., 2013). Additionally, *Adamts7*^{-/-} mice showed less atherosclerotic burden than their littermate controls (Bauer et al., 2015). This opens the possibility of pharmaceutically targeting pathways other than lipid metabolism for prevention and treatment of CAD. In order to identify additional causal genes, systematic follow-up studies in a suitable *in vivo* model system are desirable.

Pharmacological intervention of CAD

As a result of CAD's devastating effects on society, the pharmaceutical industry has attempted to develop new therapies for many decades, but has failed to develop intrinsically novel treatments in the past three decades. Most commonly, cholesterol-lowering drugs like statins are still prescribed. Statins inhibit 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR), which catalyses the rate-limiting step in cholesterol-biosynthesis by the liver. Statins are often prescribed in combination with ezetimibe, which inhibits cholesterol-absorption in the intestine by targeting Niemann-Pick C1-like 1 (NPC1L1), thereby reducing hepatic delivery of cholesterol. Other treatment options for CAD include administering compounds that

interfere with platelet function (anti-platelet therapy (Meadows & Bhatt, 2007)), and blood pressure lowering medication (Olafiranye et al., 2011). However, currently available drugs have undesired side effects, such as a higher risk of diabetes (statins, ezetimibe) (Lotta et al., 2016) or bleeding (Bhatt, 2007) (anti-platelet compounds). In addition, they show large inter-individual differences in efficiency (blood pressure lowering medication). Recently, efforts have centered around PCSK9 (Dadu & Ballantyne, 2014). Blockage of PCSK9 results in increased recycling of low-density lipoprotein receptor (LDLR), thereby increasing internalization of LDL from the circulation. Antibodies against PCSK9 were successfully validated in clinical trials and reduced LDLc and incident CAD (Sabatine et al., 2017), yet also at the expense of a higher risk of diabetes (Schmidt et al., 2017). Therefore, finding novel targets that do not rely on LDL lowering is desirable.

Lower HDLc levels have been associated with adverse CAD progression (Castelli et al., 1992; Kannel, 1976). However, clinical trials aiming to increase HDLc and thereby lowering CAD were unsuccessful (Kingwell, Chapman, Kontush, & Miller, 2014).

Novel potential targets such as asialo-glycoprotein receptor 1, a receptor that is involved in the homeostasis of glycoproteins (Nioi et al., 2016), and cluster of differentiation 47, involved in the inflammatory system whose inhibition has been demonstrated to promote clearance of diseased cells in atherosclerotic lesions (Kojima et al., 2016), have shown promising results in early studies. However, general progress has been slow in identifying new drug targets and developing new therapeutics. A better understanding of the genetic basic and risk factors of CAD are anticipated to speed up this process, as illustrated by the PCSK9 example, which was first discovered in a sequencing effort as recently as 2003 (Abifadel et al., 2003).

Primary pharmacological treatment of high triglyceride levels is administration of fibrates, whose molecular target is peroxisome-proliferator activated receptor alpha. However, fibrates can lead to undesirable side-effects such as a higher risk for myopathy (Jacobson, 2009). Recent efforts to lower triglyceride and LDLc levels have focused on Angiopoietin-like protein 3 (ANGPTL3). ANGPTL3 was first described in a mouse model (Koishi et al., 2002), and has been associated with triglyceride, LDLc and total cholesterol levels by GWAS (Willer et al., 2013). ANGPTL3 is an inhibitor of LPL, and hence interest was sparked as a potential drug target. Indeed, targeting ANGPTL3 with antibodies (Dewey et al., 2017) or antisense-oligonucleotides (Graham et al., 2017) results in a successful reduction of triglyceride levels, LDLc and HDLc in humans, and decelerates atherosclerotic progression in mice. Similarly, targeting APOC3, a protein present on triglyceride-rich lipoproteins that inhibits action of LPL and hepatic uptake, has been shown promising results in clinical trials (Gaudet et al., 2015; Khetarpal et al., 2017). Triglyceride levels have been demonstrated to be causal for CAD risk in a mendelian randomization study (Do et al., 2013),

and targeting candidate genes prioritized for triglyceride levels might be another potential avenue to treat CAD without undesired risk on type-2-diabetes.

The zebrafish as a model system for CAD

As opposed to mammals, zebrafish have one cardiovascular circuit. The heart receives and pumps deoxygenated blood that becomes oxygenated in the gills.

The development of the circulatory system, as well as vessel assembly and formation are well conserved (Gore, Monzo, Cha, Pan, & Weinstein, 2012). Due to its excellent (live) *in vivo* imaging properties and availability of transgenic strains with fluorescently labelled cell types of interest, many vascular discoveries relevant to higher vertebrates were contrived in zebrafish, as reviewed by Hogan & Schulte-Merker (Hogan & Schulte-Merker, 2017).

Zebrafish lipoproteins and their metabolism show a remarkable similarity to their human counterparts (Fang, Liu, & Miller, 2014). All major apolipoprotein classes and lipoproteins are present with a high degree of homology (Babin et al., 1997; Stoletov et al., 2009). Similarly, enzymes essential for lipoprotein metabolism are present, including *CETP*, which is absent in mice. Additionally, intestinal absorption of lipids was shown to function similarly in zebrafish and humans as well (Fang et al., 2014). A noteworthy difference is that fish prefer lipids as an energy source, and hence apolipoproteins account for a higher proportion of total protein as compared to human plasma (Babin & Vernier, 1989). Key work by Yury Miller at UCSD showed that feeding zebrafish larvae on a cholesterol-enriched diet induced early-onset atherosclerosis in 15 dpf old larvae, and resulted in atherosclerotic lesions in adult zebrafish (Stoletov et al., 2009). Vascular lipid deposits can be stained with dyes such as monodansylpentane for example (H. J. Yang, Hsu, Yang, & Yang, 2012). Alternatively, specifically oxidized lipids can be visualized using a fluorescently labelled antibody (IK17) that is expressed under control of a heat-shock promoter (Fang et al., 2011). Small scale studies suggested that drugs commonly prescribed in humans such as ezetimibe (Baek, Fang, Li, & Miller, 2012) and atorvastatin (O'Hare et al., 2014) were effective in reducing cholesterol in zebrafish. In 2014, O'Hare and colleagues showed that LDLc levels were higher in embryos injected with morpholino oligonucleotides that suppress *ldlr* expression as compared with controls injected with a nonspecific morpholino (O'Hare et al., 2014). Furthermore, knockout of *apoc2* was shown to result in severe hypertriglyceridemia, analogous to human hypertriglyceridemia (C. Liu et al., 2015). Morpholino-knockdown of *angptl3* has been shown to result in decrease of hepatic cell proliferation, and lower vascular and hepatic lipid accumulation (Lee et al., 2014). In combination with transgenically-expressed fluorescent

markers on macrophages (Ellett, Pase, Hayman, Andrianopoulos, & Lieschke, 2011) and neutrophils (Renshaw et al., 2006), the initial immunological response can be visualized, suitable for examining early-onset atherosclerosis.

Identifying causal genes for CAD and its risk factors that promote atherosclerosis is anticipated to increase our understanding of disease etiology. It will likely also increase our understanding of the interplay between risk factors, and help identify novel drug targets. While proof-of-principle studies have been conducted in zebrafish, they are characterized by small sample sizes, and lipid measurements are typically based on pools of 25-100 larvae. Similarly, conclusions drawn from knockout studies are based on the comparison of few knockouts versus wildtype larvae. Thus, the characterization of zebrafish larvae as a model system for early-onset atherosclerosis should ideally be performed in large-scale.

Characterizing candidate genes prioritized for triglyceride levels is promising due to a) the causal relationship between triglyceride-rich lipoproteins and CAD risk that has been established in a mendelian randomization study, b) the positive results of ANGPTL3 inhibition on triglyceride levels in humans, and c) the increased risk of diabetes that has been associated with intake of LDL-lowering drugs. Therefore, identifying and characterizing further causal genes for triglyceride levels may further help in CAD treatment and prevention.

The zebrafish provides many advantages in the endeavor of large-scale systematic follow-up of candidate genes from GWAS-identified loci for CAD and dyslipidemia.

Present investigations

Aims

The main objective of this thesis was to prioritize candidate genes from GWAS for cardiac and cardiometabolic traits, and characterize their role in large-scale, CRISPR-Cas9 and image-based screens in zebrafish. This is essential, because the trait-associated variants only seldom point unequivocally to causal gene(s), mechanism(s) or tissue(s). With ever increasing numbers of trait-associated loci, model systems that allow for a comprehensive and quick characterization of candidate genes are necessary.

Specifically, the aims of the studies were to:

1. Identify variants associated with measures of heart rate variability and prioritize candidate genes for functional follow-up studies (Paper 1)
2. Characterize candidate genes prioritized in study 1 for their effect on heart rate and rhythm in a large-scale screen using zebrafish embryos and larvae (Manuscript 2)
3. Establish zebrafish larvae as a model system for early-onset atherosclerosis and dyslipidemia in a large-scale, CRISPR-Cas9 and imaged-based screen (Manuscript 3)
4. Examine the effect of candidate genes prioritized from GWAS for triglyceride levels, using the pipeline established in study 3 (Manuscript 4)

Summary

Study 1

Heart rate variability (HRV) is the inter-beat variation of consecutive cardiac contractions. Heart rate variability is modulated by the autonomous nervous system. Perturbations of the autonomic system are manifested in lower HRV. In turn, lower HRV has been associated with a higher risk of all-cause mortality. The genetic basis of HRV is largely unknown, but twin and family studies have confirmed it to be a heritable trait.

To elucidate the genetic basis of HRV, a meta-analysis of GWAS was performed in data from 53,174 individuals of European ancestry. The associations of the variants were analyzed with three different HRV measures. Associations in eight loci were detected, harboring a total of 17 SNPs. Genetic risk scores were computed and accounted for 0.9 – 2.6% of the variance in HRV. Estimations of the SNP-based heritability ranged from 10.8 and 13.2%.

Since heart rate and HRV are inversely associated, it was examined if the effect of the associations on HRV were driven by their effect on heart rate. Most of the associations showed some attenuation after adjusting for heart rate, yet the majority of associations with HRV remained unchanged. Next, the association of HRV-associated SNPs with resting heart rate was examined, resulting in the identification of eleven HRV lead SNPs that are also associated with heart rate.

Then, I annotated the GWAS-identified loci for HRV *in silico*, using different bioinformatics tools and databases. PolyPhen and SIFT were used to estimate the functional consequence of the variants on protein function; and I used RegulomeDB to distill information about transcription factor binding sites and chromatin state. To examine protein-protein interactions, GeneMANIA was used. Furthermore, I used gene prioritization tools such as MetaRanker, ToppGene, Endeavour and DEPICT. Synthesis of the functional annotation data resulted in the prioritization of 18 candidate genes.

In conclusion, genetic loci associated with HRV were identified. Some of these loci harbor genes that are already known to be involved in cardiac pacemaking, whereas some loci only contain genes with unknown relevance for HRV. To instigate functional follow-up studies, we prioritized 18 candidate genes.

Study 2

In study 2, we aimed to examine the effect of the candidate genes for heart rate and rhythm that were prioritized in the HRV GWAS. We selected six of the human candidate genes for follow-up. The six human genes have a total

of nine orthologues in zebrafish. Guide-RNAs against all targets were generated for CRISPR-Cas9-based genome editing and injected in multiplex in a transgenic background that allows visualization of the beating heart. The injected founders were raised to adulthood, and I used their offspring for phenotypic screening.

Larvae were anesthetized individually, and the beating atrium was recorded for 30s, to maximize the temporal resolution. Imaging was performed with the help of the Vertebrate Automated Screening Technology (VAST) BioImager that allowed for large-scale screening. The atria of the same embryos (2dpf) and larvae (5dpf) were recorded twice, since a shift in ionic currents that drives the myocardial action potential has been suggested to occur between 2 and 5 dpf. Hence, repeated measures enable us to capture effects of the mutations on different aspects of cardiac conduction. The acquired videos were analyzed automatically using a custom-written MatLab script. Larvae were then sequenced across all targeted sites, and transcript-specific dosage scores were calculated based on the predicted functional impact of the mutations on protein function. Finally, we conducted the association analysis.

We detected an altered heart rate in embryos and larvae with mutations in *rgs6* and *hcn4*; two genes that have already been implicated in cardiac rate and rhythm. Mice deficient in *Rgs6* have a higher risk of bradycardia and atrioventricular block, and a higher HRV. In addition, humans with loss-of-function variants in *RGS6* were characterized by higher HRV. Zebrafish embryos affected in *rgs6* showed a lower heart rate, independently of HRV.

HCN4 is well studied, and responsible for cardiac pacemaking. *Hcn4* deficient mice have a lower heart rate and die prenatally. Similarly, human individuals that harbor a heterozygous loss-of-function mutation in *HCN4* are characterized by bradycardia. In our study, *hcn4* affected larvae displayed a higher heart rate independently of HRV at 2dpf. At 5dpf, *hcn4* influenced heart rate and HRV, partly through its effect on the other trait. The direction of effect (higher) on heart rate is unexpected, and likely mediated by overcompensation by another *hcn4* orthologue present in the zebrafish genome, *hcn4l*. Since the gRNA directs Cas9 close to the sequence that encodes the ion-transporter domain in *hcn4*, we cannot rule out that a gain-of-function mutation was induced in a subset of larvae that instigate the increased heart rate.

We detected a higher HRV in embryos and larvae with mutations in *si:dkey-65j6.2*, one of the two orthologues for human *KIAA1755*. *KIAA1755* is a previously uncharacterized gene, and variants in *KIAA1755* have been identified in GWAS for heart rate and HRV. Lower expression of *KIAA1755* has been associated with higher HRV in humans, indicating that our results are directionally consistent with those obtained in humans.

In conclusion, we confirm known effects of *rgs6* and *hcn4*. Furthermore, we identify a novel likely causal gene that influences HRV, i.e. *KIAA1755*.

Study 3

In study 3, we aimed to establish zebrafish larvae as a model system for early-onset atherosclerosis and dyslipidemia. Atherosclerosis is the consequence of the building up of atherosclerotic plaques in the coronary vessels during the course of a lifetime. Atherosclerosis is characterized by chronic inflammation of the blood vessels, which is promoted by elevated lipoprotein levels. Rupture of the atherosclerotic plaque will trigger thrombosis, potentially resulting in a myocardial infarction, stroke or peripheral ischemia.

Proof-of-concept studies demonstrating that zebrafish larvae can be used to model early-onset atherosclerosis have been performed, but are typically characterized by small sample sizes, and pooled extracts of larvae to measure lipid levels. Therefore, we aimed to establish zebrafish larvae as a model system in a large-scale, enzymatic and image-based study using the VAST BioImager. Vascular lipid deposits were visualized using monodansylpentane, whereas the initial inflammatory response was visualized in a transgenic strain with fluorescent markers on macrophage and neutrophil-specific proteins. Lipid fractions were determined enzymatically using whole-body extracts of single larvae. Larvae were sequenced as described in Study 2.

First, in a dietary intervention, larvae were fed normal amounts and higher amounts of the same diet (overfeeding) with or without cholesterol supplementation. Data from more than 2000 larvae showed that overfeeding and cholesterol supplementation have independent pro-atherogenic effects.

Secondly, in a drug treatment intervention, >1000 larvae were overfed with a cholesterol-supplemented diet. In the treatment group, the diet was additionally supplemented with atorvastatin and ezetimibe. Larvae fed the drug-enriched diet had lower LDLc, total cholesterol and triglyceride levels, and less vascular inflammation and early-stage atherosclerosis, while glucose levels were higher. This resembles the observed effect of statin and ezetimibe treatment in humans.

Thirdly, we used the established imaging pipeline to examine the effects of CRISPR-Cas9 induced mutations in the zebrafish orthologues of human *LDLR*, *APOE* and *APOB*. This led to the observation that larvae affected in *apoea* showed higher LDLc levels, whereas larvae affected in *apobb.1* had more pronounced early-onset atherosclerosis based on the vascular-imaging traits.

In conclusion, study 3 facilitated the large-scale validation of zebrafish as a model system for early-onset atherosclerosis and dyslipidemia, which will likely aid in the characterization of candidate genes for coronary artery disease and dyslipidemia.

Study 4

In study 4, we followed up GWAS-identified loci for triglyceride levels. Elevated triglyceride levels have been causally linked to higher risk of cardiovascular disease. In recent years, drugs against ANGPTL3 and APOC3 have proven to be efficiently lowering triglyceride levels and risk of cardiovascular disease. Characterizing candidate genes from the 37 loci that have been robustly associated with triglyceride levels would not only elicit the biology of triglyceride metabolism, but also potentially identify novel drug targets.

We conducted candidate gene prioritization using DEPICT, which resulted in 37 candidates from 23 of the loci. Of these 37 genes, we identified 30 genes that have a total of 42 zebrafish orthologues. These 42 genes were targeted in five multiplexed lines using CRISPR-Cas9. Founders were raised and their 10 days old offspring were used for experiments as described in study 3.

We identified three genes that influence triglyceride levels. Mutations in *inhbc* and *znf335* resulted in higher triglyceride levels, whereas larvae affected in *arid1aa* had lower triglyceride levels. In addition, *arid1aa* affected larvae showed lower total cholesterol levels. Furthermore, we identified several genes that influence different cardiometabolic factors. Interestingly, two genes showed trends towards lower triglycerides levels (*dock7* and *lpar2a*), with directionally opposite effects on vascular inflammation. This suggests that the effect on vascular inflammation is independent of the gene's effect on triglyceride levels. Furthermore, *dock7* is in the same locus as *angptl3*. Although further studies are needed to confirm the effect of *dock7* on triglyceride levels, it might hint at the presence of another causal gene influencing triglycerides, in addition to *angptl3*.

In conclusion, I identified several genes that influence different cardiometabolic factors and highlight the importance of thoroughly dissecting genetic loci that have been associated with lipid levels in GWAS.

Concluding remarks and future perspective

GWAS have associated a plethora of loci with cardiac and cardiometabolic traits. While the number of associated loci are continuously increasing, the functional characterization of candidate genes has unfortunately not shown a similar trend. While GWAS significantly increased our understanding of pathways involved in complex traits, they probably raised more questions than they have answered so far. Recent reports examining the pleiotropy of traits (for example CAD) demonstrate that loci may influence several risk factors simultaneously (Webb et al., 2017). This informs trait and disease biology, but only rigorous examination of a given locus may provide clear answers as to how the genes influence the different traits. Another theory that has been proposed is the omnigenic view of complex traits (Boyle, Li, & Pritchard, 2017). The authors proposed the idea that a trait or disease is directly affected by a set of “core” genes, whereas signals from all over the genome outside the core genes indirectly modulate the core genes, as a consequence of the interconnectivity of all regulatory networks. Again, this emphasizes the need for high-throughput approaches, which enable rapid and comprehensive screening of candidate genes to promote knowledge of the genetic architecture of a trait or disease to refute or confirm their theory.

All candidate gene prioritization relies on the use of bioinformatic tools and huge databases. While these resources provide an immense aid in prioritizing candidate genes for functional characterization, their results remain prediction. The continuous effort of generating even larger databases in parallel with on-going development of bioinformatic tools will likely foster even more precise predictions. Similarly, exome sequencing efforts have the possibility to detect rare variants that exert a large effect on the phenotype of interest, with a – usually - more straightforward inference of causal genes.

Concerning the findings made in this PhD thesis, we detected interesting associations and trends with cardiac and cardiometabolic traits.

As for the cardiac traits, it would be very interesting to characterize tissue specific effects of *si:dkey-65j6.2* (i.e. *kiaa1755*) on heart rate variability. For example, *si:dkey-65j6.2* could be knocked out in specific cardiac and/or brain tissues using CRISPR-Cas9. Subsequently, transcriptomic studies could be conducted to identify deregulated genes upon *si:dkey-65j6.2* disruption, aiding mechanistic insights. Last but not least, it would be intriguing to acquire recordings of the whole beating heart, to examine if ventricular function or structure is affected.

As for the cardiometabolic traits, it would be interesting to generate tissue-specific knockouts of the genes that showed an effect. Furthermore, it would be interesting to phenotype genes that show an effect on lipid measurements and glucose in transgenic backgrounds that allow visualization of structures relevant to diabetes, such as fluorescently labeled beta-cells (Maddison & Chen, 2012). This combined phenotyping approach of screening larvae for early-onset atherosclerosis and diabetes may enable the identification of genes that lower LDLc levels, without increasing the risk for diabetes.

Acknowledgements

My first and foremost gratitude is directed to my main supervisor, Marcel. You accepted me in your lab as a PhD student after I quit my previous PhD position – I am deeply grateful for your belief in me and giving me that second chance. Thank you for the great continuous supervision and all the new skills I was able to learn from you, and making me a better scientist.

Many thanks to Anastasia, one of my two co-supervisors. You always had an open ear regardless if life was not going smooth inside or outside the lab. Thank you for being positive, supportive and all the skills you taught me in the lab.

Gratitude is also extended to my co-supervisor Erik. Your feedback was greatly appreciated. Thank you Tiffany, it was a blast working together with you. And of course, thank you for the great zebrafish, CRISPR and off-topic discussions. Manoj, we spent the last four years aside as PhD students in the lab, and it has been a lot of fun. Cheers! Of course, all current and previous members of the lab: Ci, Eugenia, Natalie, it has been a treat!

All fellow IGP-students and friends at BMC; Ida, Loora, Sanna, Marcus, Mitra and generally everybody at B11:4 – thanks to all of you. This of course extends to the whole IGP PhD council (fun times!) and all current and former people at Rudbeck: Diana, Marco, Svea, Vero, Yang, and all others. The former circle training group may not be forgotten under any circumstance: you know who you are!

To all my friends that I got to know while staying at the Max-Planck Institute previous to coming back to Uppsala: I miss you all!

Joao – my buddy and MacGyver. We need to drink beers soon and come up with business ideas; I will have a lot of free time soon. Also thank you Luciana, especially for always being positive.

Emma and Andreas, what would I have done without you two crazy people? From spontaneous midnight parties to philosophic discussions, we share so many epic fun times. I would have missed a great deal if I hadn't run into both of you.

Andrew, from being my supervisor in the lab during the master thesis to one of my closest friends, that is a progression that is not always given. Especially cherishing all the sauna and gaming nights. Thanks for always having an open ear and being a great buddy!

Julia und Philipp – ihr beiden Sahneschnitten! Ihr spielt beide einen wichtigen Teil in meinem Leben seit ein paar Jahren - mit euch beiden geht die Zeit immer viel zu schnell rum. Wir sehen uns ab jetzt öfter – pinky promise. Tina und Henning – meine Lieblingstottos! Wir kennen uns seit dem Master und seit dem seit ihr unaustauschbar! Mehr Besuche folgen ab bald.

Die Hannover Crew: Marek und Toby, wir kennen uns schon seit Urzeiten, danke fuer alles, es wuerde den Rahmen sprengen alles aufzuzählen. Bibi und Desi, wir haben unsere Freundschaft nach der Schule aufrecht erhalten und vertiefen können – danke euch! Lars mein Kuschelbär, muss ich mehr sagen? Anika und Finja, absolut grossartig das ihr mit im Boot seit. Michi, Chrissi, Bene, Henning, Vivi, Doro, Simone – wir kennen uns fast alle seit dem FSJ. Ihr rockt unfassbar, ich bin dankbar euch zu haben.

Jennifer, you filled the vacant position at my side about 2 ½ years ago. Thank you for enriching my life with so many great experiences, your endless support and perspective that I surely need by times. Till din familj – tack för att ni är alltid lättsam och varmhjärtat.

Last but not least, to my whole family. Ich bin euch arg dankbar, dass ihr mir alles in meinem Leben so einfach ermöglicht habt. Die ganzen Umzuege, mein spontanes Ende an meiner ersten Doktorandenstelle - ihr wart immer unterstuetzend da wenn ich irgendwas brauchte. Und auch ganz generell, ohne euch wäre alles um einiges komplizierter gewesen – vielen Dank!

References

- 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), 1061.
- 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68.
- Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M., . . . Boileau, C. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genetics*, 34(2), 154-156. doi:10.1038/ng1161
- Ablain, J., Durand, E. M., Yang, S., Zhou, Y., & Zon, L. I. (2015). A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Developmental Cell*, 32(6), 756-764. doi:10.1016/j.devcel.2015.01.032
- Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat Commun*, 9(1), 1911. doi:10.1038/s41467-018-04252-2
- Allen, F., Crepaldi, L., Alsinet, C., Strong, A. J., Kleshchevnikov, V., De Angeli, P., . . . Parts, L. (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nature Biotechnology*. doi:10.1038/nbt.4317
- Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. *American Journal of Human Genetics*, 69(5), 936-950. doi:10.1086/324069
- Arking, D. E., Pulit, S. L., Crotti, L., van der Harst, P., Munroe, P. B., Koopmann, T. T., . . . Newton-Cheh, C. (2014). Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nature Genetics*, 46(8), 826-836. doi:10.1038/ng.3014
- Arnaout, R., Ferrer, T., Huiskens, J., Spitzer, K., Stainier, D. Y., Tristani-Firouzi, M., & Chi, N. C. (2007). Zebrafish model for human long QT syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27), 11316-11321. doi:10.1073/pnas.0702724104
- Arrenberg, A. B., Stainier, D. Y., Baier, H., & Huiskens, J. (2010). Optogenetic control of cardiac function. *Science*, 330(6006), 971-974. doi:10.1126/science.1195929
- Babin, P. J., Thisse, C., Durliat, M., Andre, M., Akimenko, M. A., & Thisse, B. (1997). Both apolipoprotein E and A-I genes are present in a nonmammalian vertebrate and are highly expressed during embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, 94(16), 8622-8627.
- Babin, P. J., & Vernier, J. M. (1989). Plasma lipoproteins in fish. *Journal of Lipid Research*, 30(4), 467-489.
- Back, J. S., Fang, L., Li, A. C., & Miller, Y. I. (2012). Ezetimibe and simvastatin reduce cholesterol levels in zebrafish larvae fed a high-cholesterol diet. *Cholesterol*, 2012, 564705. doi:10.1155/2012/564705
- Bakkers, J. (2011). Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovascular Research*, 91(2), 279-288. doi:10.1093/cvr/cvr098

- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., . . . Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712. doi:10.1126/science.1138140
- Battle, A., Brown, C. D., Engelhardt, B. E., & Montgomery, S. B. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550*(7675), 204-213. doi:10.1038/nature24277
- Bauer, R. C., Tohyama, J., Cui, J., Cheng, L., Yang, J., Zhang, X., . . . Reilly, M. P. (2015). Knockout of Adamts7, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation*, *131*(13), 1202-1213. doi:10.1161/CIRCULATIONAHA.114.012669
- Berge, K. E., Tian, H., Graf, G. A., Yu, L., Grishin, N. V., Schultz, J., . . . Hobbs, H. H. (2000). Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science*, *290*(5497), 1771-1775.
- Bhatt, D. L. (2007). Intensifying platelet inhibition--navigating between Scylla and Charybdis *New England Journal of Medicine* (Vol. 357, pp. 2078-2081). United States.
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, *169*(7), 1177-1186. doi:10.1016/j.cell.2017.05.038
- Brotman, D. J., Bash, L. D., Qayyum, R., Crews, D., Whitsel, E. A., Astor, B. C., & Coresh, J. (2010). Heart rate variability predicts ESRD and CKD-related hospitalization. *Journal of the American Society of Nephrology*, *21*(9), 1560-1570. doi:10.1681/asn.2009111112
- Buccelletti, E., Gilardi, E., Scaini, E., Galiuto, L., Persiani, R., Biondi, A., . . . Silveri, N. G. (2009). Heart rate variability and myocardial infarction: systematic literature review and metanalysis. *European Review for Medical and Pharmacological Sciences*, *13*(4), 299-307.
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., . . . Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005-d1012. doi:10.1093/nar/gky1120
- Burns, C. G., Milan, D. J., Grande, E. J., Rottbauer, W., MacRae, C. A., & Fishman, M. C. (2005). High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nature Chemical Biology*, *1*(5), 263-264. doi:10.1038/nchembio732
- C.ARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., . . . Samani, N. J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics*, *45*(1), 25-33. doi:10.1038/ng.2480
- Carrington, B., Varshney, G. K., Burgess, S. M., & Sood, R. (2015). CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Research*, *43*(22), e157. doi:10.1093/nar/gkv802
- Castelli, W. P., Anderson, K., Wilson, P. W., & Levy, D. (1992). Lipids and risk of coronary heart disease. The Framingham Study. *Annals of Epidemiology*, *2*(1-2), 23-28.
- Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., . . . Porteus, M. H. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nature Medicine*, *25*(2), 249-254. doi:10.1038/s41591-018-0326-x

- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., . . . Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), 1479-1491. doi:10.1016/j.cell.2013.12.001
- Chi, N. C., Shaw, R. M., Jungblut, B., Huisken, J., Ferrer, T., Arnaout, R., . . . Stainier, D. Y. (2008). Genetic and physiologic dissection of the vertebrate cardiac conduction system. *PLoS Biology*, 6(5), e109. doi:10.1371/journal.pbio.0060109
- Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H., Ban, H.-J., . . . Park, M. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature Genetics*, 41(5), 527.
- Dadu, R. T., & Ballantyne, C. M. (2014). Lipid lowering with PCSK9 inhibitors. *Nature Reviews: Cardiology*, 11(10), 563-575. doi:10.1038/nrcardio.2014.84
- Dekker, J. M., Schouten, E. G., Klootwijk, P., Pool, J., Swenne, C. A., & Kromhout, D. (1997). Heart rate variability from short electrocardiographic recordings predicts mortality from all causes in middle-aged and elderly men - The Zutphen study. *American Journal of Epidemiology*, 145(10), 899-908.
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., . . . Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607. doi:10.1038/nature09886
- den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B. J., Peal, D. S., Evans, D. M., . . . Loos, R. J. (2013). Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nature Genetics*, 45(6), 621-631. doi:10.1038/ng.2610
- Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., . . . Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, 190(4), 1390-1400. doi:10.1128/jb.01412-07
- Dewey, F. E., Gusarova, V., Dunbar, R. L., O'Dushlaine, C., Schurmann, C., Gottesman, O., . . . Baras, A. (2017). Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. *New England Journal of Medicine*, 377(3), 211-221. doi:10.1056/NEJMoa1612790
- Do, R., Stitzel, N. O., Won, H. H., Jorgensen, A. B., Duga, S., Angelica Merlini, P., . . . Kathiresan, S. (2015). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*, 518(7537), 102-106. doi:10.1038/nature13917
- Do, R., Willer, C. J., Schmidt, E. M., Sengupta, S., Gao, C., Peloso, G. M., . . . Kathiresan, S. (2013). Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature Genetics*, 45(11), 1345-1352. doi:10.1038/ng.2795
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., . . . Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2), 184-191. doi:10.1038/nbt.3437
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., . . . Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology*, 32(12), 1262-1267. doi:10.1038/nbt.3026

- Dyer, A. R., Persky, V., Stamler, J., Paul, O., Shekelle, R. B., Berkson, D. M., . . . Lindberg, H. A. (1980). Heart rate as a prognostic factor for coronary heart disease and mortality: findings in three Chicago epidemiologic studies. *American Journal of Epidemiology*, 112(6), 736-749.
- Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P. I., Müller, M., Morrison, A. C., . . . Dörr, M. (2010). Genome-wide association analysis identifies multiple loci related to resting heart rate. *Human Molecular Genetics*, 19(19), 3885-3894.
- Ellett, F., Pase, L., Hayman, J. W., Andrianopoulos, A., & Lieschke, G. J. (2011). mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. *Blood*, 117(4), e49-56. doi:10.1182/blood-2010-10-314120
- Ellinor, P. T., Lunetta, K. L., Albert, C. M., Glazer, N. L., Ritchie, M. D., Smith, A. V., . . . Kaab, S. (2012). Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nature Genetics*, 44(6), 670-675. doi:10.1038/ng.2261
- Eppinga, R. N., Hagemeijer, Y., Burgess, S., Hinds, D. A., Stefansson, K., Gudbjartsson, D. F., . . . van der Harst, P. (2016). Identification of genomic loci associated with resting heart rate and shared genetic predictors with all-cause mortality. *Nature Genetics*, 48(12), 1557-1563. doi:10.1038/ng.3708
- Fang, L., Green, S. R., Baek, J. S., Lee, S. H., Ellett, F., Deer, E., . . . Miller, Y. I. (2011). In vivo visualization and attenuation of oxidized lipid accumulation in hypercholesterolemic zebrafish. *Journal of Clinical Investigation*, 121(12), 4861-4869. doi:10.1172/JCI57755
- Fang, L., Liu, C., & Miller, Y. I. (2014). Zebrafish models of dyslipidemia: relevance to atherosclerosis and angiogenesis. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 163(2), 99-108. doi:10.1016/j.trsl.2013.09.004
- Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiacovo, M. P., Church, G. M., & Calarco, J. A. (2013). Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods*, 10(8), 741-743. doi:10.1038/nmeth.2532
- Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Akhmetova, L., Pauli, A., . . . Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS One*, 9(5), e98186. doi:10.1371/journal.pone.0098186
- Garcia, C. K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., . . . Grishin, N. (2001). Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science*, 292(5520), 1394-1398.
- Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*, 551(7681), 464-471. doi:10.1038/nature24644
- Gaudet, D., Alexander, V. J., Baker, B. F., Brisson, D., Tremblay, K., Singleton, W., . . . Kastelein, J. J. (2015). Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. *New England Journal of Medicine*, 373(5), 438-447. doi:10.1056/NEJMoa1400283
- Gillum, R. F., Makuc, D. M., & Feldman, J. J. (1991). Pulse rate, coronary heart disease, and death: the NHANES I Epidemiologic Follow-up Study. *American Heart Journal*, 121(1 Pt 1), 172-177.
- Glickman, N. S., & Yelon, D. (2002). Cardiac development in zebrafish: coordination of form and function. *Seminars in Cell and Developmental Biology*, 13(6), 507-513.

- Goit, R. K., & Ansari, A. H. (2016). Reduced parasympathetic tone in newly diagnosed essential hypertension. *Indian Heart Journal*, 68(2), 153-157. doi:10.1016/j.ihj.2015.08.003
- Gore, A. V., Monzo, K., Cha, Y. R., Pan, W., & Weinstein, B. M. (2012). Vascular development in the zebrafish. *Cold Spring Harbor Perspectives in Medicine*, 2(5), a006684. doi:10.1101/cshperspect.a006684
- Graham, M. J., Lee, R. G., Brandt, T. A., Tai, L. J., Fu, W., Peralta, R., . . . Tsimikas, S. (2017). Cardiovascular and Metabolic Effects of ANGPTL3 Antisense Oligonucleotides. *New England Journal of Medicine*, 377(3), 222-232. doi:10.1056/NEJMoa1701329
- GTEX Consortium. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235), 648-660. doi:10.1126/science.1262110
- Guryev, V., Koudijs, M. J., Berezikov, E., Johnson, S. L., Plasterk, R. H., van Eeden, F. J., & Cuppen, E. (2006). Genetic variation in the zebrafish. *Genome Research*, 16(4), 491-497. doi:10.1101/gr.4791006
- Harrington, J. K., Sorabella, R., Tercek, A., Isler, J. R., & Targoff, K. L. (2017). Nkx2.5 is essential to establish normal heart rate variability in the zebrafish embryo. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 313(3), R265-r271. doi:10.1152/ajpregu.00223.2016
- Hogan, B. M., & Schulte-Merker, S. (2017). How to Plumb a Pisces: Understanding Vascular Development and Disease Using Zebrafish Embryos. *Developmental Cell*, 42(6), 567-583. doi:10.1016/j.devcel.2017.08.015
- Holm, H., Gudbjartsson, D. F., Arnar, D. O., Thorleifsson, G., Thorgeirsson, G., Stefansdottir, H., . . . Njølstad, I. (2010). Several common variants modulate heart rate, PR interval and QRS duration. *Nature Genetics*, 42(2), 117.
- Hou, J. H., Kralj, J. M., Douglass, A. D., Engert, F., & Cohen, A. E. (2014). Simultaneous mapping of membrane voltage and calcium in zebrafish heart in vivo reveals chamber-specific developmental transitions in ionic currents. *Frontiers in Physiology*, 5, 344. doi:10.3389/fphys.2014.00344
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., . . . Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498-503. doi:10.1038/nature12111
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., . . . Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), 827-832. doi:10.1038/nbt.2647
- Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., . . . Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, 31(3), 227-229. doi:10.1038/nbt.2501
- International HapMap Consortium. (2003). The international HapMap project. *Nature*, 426(6968), 789.
- International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299.
- International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931-945. doi:10.1038/nature03001

- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12), 5429-5433.
- Jacobson, T. A. (2009). Myopathy with statin-fibrate combination therapy: clinical considerations. *Nature Reviews: Endocrinology*, 5(9), 507-518. doi:10.1038/nrendo.2009.151
- Jansen, R., Embden, J. D., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565-1575.
- Jao, L. E., Wente, S. R., & Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), 13904-13909. doi:10.1073/pnas.1308335110
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), 816-821. doi:10.1126/science.1225829
- Jou, C. J., Arrington, C. B., Barnett, S., Shen, J., Cho, S., Sheng, X., . . . Tristani-Firouzi, M. (2017). A Functional Assay for Sick Sinus Syndrome Genetic Variants. *Cellular Physiology and Biochemistry*, 42(5), 2021-2029. doi:10.1159/000479897
- Kannel, W. B. (1976). Some lessons in cardiovascular epidemiology from Framingham. *American Journal of Cardiology*, 37(2), 269-282.
- Kettleborough, R. N., Busch-Nentwich, E. M., Harvey, S. A., Dooley, C. M., de Bruijn, E., van Eeden, F., . . . Stemple, D. L. (2013). A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature*, 496(7446), 494-497. doi:10.1038/nature11992
- Khetarpal, S. A., Zeng, X., Millar, J. S., Vitali, C., Somasundara, A. V. H., Zannoni, P., . . . Rader, D. J. (2017). A human APOC3 missense variant and monoclonal antibody accelerate apoC-III clearance and lower triglyceride-rich lipoprotein levels. *Nature Medicine*, 23(9), 1086-1094. doi:10.1038/nm.4390
- Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., . . . Kim, J. S. (2017). In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun*, 8, 14500. doi:10.1038/ncomms14500
- Kingwell, B. A., Chapman, M. J., Kontush, A., & Miller, N. E. (2014). HDL-targeted therapies: progress, failures and future. *Nat Rev Drug Discov*, 13(6), 445-464. doi:10.1038/nrd4279
- Klarin, D., Damrauer, S. M., Cho, K., Sun, Y. V., Teslovich, T. M., Honerlaw, J., . . . Assimes, T. L. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature Genetics*, 50(11), 1514-1523. doi:10.1038/s41588-018-0222-9
- Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., . . . Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, 308(5720), 385-389. doi:10.1126/science.1109557
- Koishi, R., Ando, Y., Ono, M., Shimamura, M., Yasuno, H., Fujiwara, T., . . . Furukawa, H. (2002). Angptl3 regulates lipid metabolism in mice. *Nature Genetics*, 30(2), 151-157. doi:10.1038/ng814
- Kojima, Y., Volkmer, J. P., McKenna, K., Civelek, M., Lusis, A. J., Miller, C. L., . . . Leeper, N. J. (2016). CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis. *Nature*, 536(7614), 86-90. doi:10.1038/nature18935

- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., . . . Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317-330. doi:10.1038/nature14248
- Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research*, *44*(W1), W272-276. doi:10.1093/nar/gkw398
- LaFave, M. C., Varshney, G. K., Vemulapalli, M., Mullikin, J. C., & Burgess, S. M. (2014). A defined zebrafish line for high-throughput genetics and genomics: NHGRI-1. *Genetics*, *198*(1), 167-170. doi:10.1534/genetics.114.166769
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921. doi:10.1038/35057062
- Lee, S. H., So, J. H., Kim, H. T., Choi, J. H., Lee, M. S., Choi, S. Y., . . . Kim, M. J. (2014). Angiopoietin-like 3 regulates hepatocyte proliferation and lipid metabolism in zebrafish. *Biochemical and Biophysical Research Communications*, *446*(4), 1237-1242. doi:10.1016/j.bbrc.2014.03.099
- Lehrman, M. A., Schneider, W. J., Sudhof, T. C., Brown, M. S., Goldstein, J. L., & Russell, D. W. (1985). Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science*, *227*(4683), 140-146.
- Leong, I. U., Skinner, J. R., Shelling, A. N., & Love, D. R. (2010). Zebrafish as a model for long QT syndrome: the evidence and the means of manipulating zebrafish gene expression. *Acta Physiologica (Oxford, England)*, *199*(3), 257-276. doi:10.1111/j.1748-1716.2010.02111.x
- Liao, D., Cai, J., Rosamond, W. D., Barnes, R. W., Hutchinson, R. G., Whitsel, E. A., . . . Heiss, G. (1997). Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study. The ARIC Study. Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology*, *145*(8), 696-706.
- Lieschke, G. J., & Currie, P. D. (2007). Animal models of human disease: zebrafish swim into view. *Nat Rev Genet*, *8*(5), 353-367. doi:10.1038/nrg2091
- Liu, C., Gates, K. P., Fang, L., Amar, M. J., Schneider, D. A., Geng, H., . . . Miller, Y. I. (2015). Apoc2 loss-of-function zebrafish mutant as a genetic model of hyperlipidemia. *Disease Models & Mechanisms*, *8*(8), 989-998. doi:10.1242/dmm.019836
- Liu, C. C., Li, L., Lam, Y. W., Siu, C. W., & Cheng, S. H. (2016). Improvement of surface ECG recording in adult zebrafish reveals that the value of this model exceeds our expectation. *Scientific Reports*, *6*, 25073. doi:10.1038/srep25073
- Liu, J., Bressan, M., Hassel, D., Huiskens, J., Staudt, D., Kikuchi, K., . . . Stainier, D. Y. (2010). A dual role for ErbB2 signaling in cardiac trabeculation. *Development*, *137*(22), 3867-3875. doi:10.1242/dev.053736
- Lotta, L. A., Sharp, S. J., Burgess, S., Perry, J. R., Stewart, I. D., Willems, S. M., . . . Wareham, N. J. (2016). Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. *JAMA*, *316*(13), 1383-1391. doi:10.1001/jama.2016.14568
- Ludwig, A., Zong, X., Stieber, J., Hullin, R., Hofmann, F., & Biel, M. (1999). Two pacemaker channels from human heart with profoundly different activation kinetics. *EMBO Journal*, *18*(9), 2323-2329. doi:10.1093/emboj/18.9.2323
- Lusis, A. J. (2000). Atherosclerosis. *Nature*, *407*(6801), 233-241. doi:10.1038/35025203

- Mack, M., & Gopal, A. (2014). Epidemiology, traditional and novel risk factors in coronary artery disease. *Cardiology Clinics*, 32(3), 323-332. doi:10.1016/j.ccl.2014.04.003
- MacRae, C. A., & Peterson, R. T. (2015). Zebrafish as tools for drug discovery. *Nat Rev Drug Discov*, 14(10), 721-731. doi:10.1038/nrd4627
- Maddison, L. A., & Chen, W. B. (2012). Nutrient Excess Stimulates beta-Cell Neogenesis in Zebrafish. *Diabetes*, 61(10), 2517-2524. doi:10.2337/Db11-1841
- Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B., & Koonin, E. V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Research*, 30(2), 482-496.
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, 1, 7. doi:10.1186/1745-6150-1-7
- Martin, W. K., Tennant, A. H., Conolly, R. B., Prince, K., Stevens, J. S., DeMarini, D. M., . . . Farraj, A. K. (2019). High-Throughput Video Processing of Heart Rate Responses in Multiple Wild-type Embryonic Zebrafish per Imaging Field. *Scientific Reports*, 9(1), 145. doi:10.1038/s41598-018-35949-5
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., . . . Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 122. doi:10.1186/s13059-016-0974-4
- McPherson, R., Pertsemliadis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., . . . Folsom, A. R. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science*, 316(5830), 1488-1491.
- Meadows, T. A., & Bhatt, D. L. (2007). Clinical aspects of platelet inhibitors and thrombus formation. *Circulation Research*, 100(9), 1261-1275. doi:10.1161/01.res.0000264509.36234.51
- Meyer, A., & Scharl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology*, 11(6), 699-704.
- Milan, D. J., Kim, A. M., Winterfield, J. R., Jones, I. L., Pfeufer, A., Sanna, S., . . . MacRae, C. A. (2009). Drug-sensitized zebrafish screen identifies multiple genes, including GINS3, as regulators of myocardial repolarization. *Circulation*, 120(7), 553-559. doi:10.1161/circulationaha.108.821082
- Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, 60(2), 174-182. doi:10.1007/s00239-004-0046-3
- Mojica, F. J., Diez-Villasenor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria *Molecular Microbiology* (Vol. 36, pp. 244-246). England.
- Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods*, 12(10), 982-988. doi:10.1038/nmeth.3543
- Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V., . . . Rader, D. J. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, 466(7307), 714-719. doi:10.1038/nature09266

- Nielsen, J. B., Thorolfsson, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., . . . Willer, C. J. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature Genetics*, 50(9), 1234-1239. doi:10.1038/s41588-018-0171-3
- Nikpay, M., Goel, A., Won, H. H., Hall, L. M., Willenborg, C., Kanoni, S., . . . Farrall, M. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, 47(10), 1121-1130. doi:10.1038/ng.3396
- Nioi, P., Sigurdsson, A., Thorleifsson, G., Helgason, H., Agustsdottir, A. B., Norddahl, G. L., . . . Stefansson, K. (2016). Variant ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. *New England Journal of Medicine*, 374(22), 2131-2141. doi:10.1056/NEJMoa1508419
- O'Hare, E. A., Wang, X., Montasser, M. E., Chang, Y. P., Mitchell, B. D., & Zaghoul, N. A. (2014). Disruption of *ldlr* causes increased LDL-c and vascular lipid accumulation in a zebrafish model of hypercholesterolemia. *Journal of Lipid Research*, 55(11), 2242-2253. doi:10.1194/jlr.M046540
- Olafiranye, O., Zizi, F., Brimah, P., Jean-louis, G., Makaryus, A. N., McFarlane, S., & Ogedegbe, G. (2011). Management of Hypertension among Patients with Coronary Heart Disease. *International Journal of Hypertension*, 2011. doi:10.4061/2011/653903
- Palazzo, A. F., & Gregory, T. R. (2014). The case for junk DNA. *PLoS Genet*, 10(5), e1004351. doi:10.1371/journal.pgen.1004351
- Pardo-Martin, C., Allalou, A., Medina, J., Eimon, P. M., Wahlby, C., & Fatih Yanik, M. (2013). High-throughput hyperdimensional vertebrate phenotyping. *Nat Commun*, 4, 1467. doi:10.1038/ncomms2475
- Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9), 839-843. doi:10.1038/nbt.2673
- Pelster, B., & Burggren, W. W. (1996). Disruption of hemoglobin oxygen transport does not impact oxygen-dependent physiological processes in developing embryos of zebra fish (*Danio rerio*). *Circulation Research*, 79(2), 358-362.
- Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H. J., Wood, A. R., Yang, J., . . . Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*, 6, 5890. doi:10.1038/ncomms6890
- Poon, K. L., & Brand, T. (2013). The zebrafish model system in cardiovascular research: A tiny fish with mighty prospects. *Glob Cardiol Sci Pract*, 2013(1), 9-28. doi:10.5339/gcsp.2013.4
- Priori, S. G., & Napolitano, C. (2006). Role of genetic analyses in cardiology: part I: mendelian diseases: cardiac channelopathies. *Circulation*, 113(8), 1130-1135. doi:10.1161/circulationaha.105.563205
- Pu, X., Xiao, Q., Kiechl, S., Chan, K., Ng, F. L., Gor, S., . . . Ye, S. (2013). ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *American Journal of Human Genetics*, 92(3), 366-374. doi:10.1016/j.ajhg.2013.01.012
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 1173-1183. doi:10.1016/j.cell.2013.02.022

- Ramirez, J., Duijvenboden, S. V., Ntalla, I., Mifsud, B., Warren, H. R., Tzanis, E., . . . Munroe, P. B. (2018). Thirty loci identified for heart rate response to exercise and recovery implicate autonomic nervous system. *Nat Commun*, 9(1), 1947. doi:10.1038/s41467-018-04148-1
- Renshaw, S. A., Loynes, C. A., Trushell, D. M., Elworthy, S., Ingham, P. W., & Whyte, M. K. (2006). A transgenic zebrafish model of neutrophilic inflammation. *Blood*, 108(13), 3976-3978. doi:10.1182/blood-2006-05-024075
- Roselli, C., Chaffin, M. D., Weng, L. C., Aeschbacher, S., Ahlberg, G., Albert, C. M., . . . Ellinor, P. T. (2018). Multi-ethnic genome-wide association study for atrial fibrillation. *Nature Genetics*, 50(9), 1225-1233. doi:10.1038/s41588-018-0133-9
- Rottbauer, W., Baker, K., Wo, Z. G., Mohideen, M. A., Cantiello, H. F., & Fishman, M. C. (2001). Growth and function of the embryonic heart depend upon the cardiac-specific L-type calcium channel alpha1 subunit. *Developmental Cell*, 1(2), 265-275.
- Sabatine, M. S., Giugliano, R. P., Keech, A. C., Honarpour, N., Wiviott, S. D., Murphy, S. A., . . . Investigators. (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *New England Journal of Medicine*, 376(18), 1713-1722. doi:10.1056/NEJMoa1615664
- Sah, R., Mesirca, P., Van den Boogert, M., Rosen, J., Mably, J., Mangoni, M. E., & Clapham, D. E. (2013). Ion channel-kinase TRPM7 is required for maintaining cardiac automaticity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(32), E3037-3046. doi:10.1073/pnas.1311865110
- Saxena, R., Voight, B. F., Lyssenko, V., Burt, N. P., de Bakker, P. I., Chen, H., . . . Daly, M. J. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829), 1331-1336.
- Scheuner, M. T. (2003). Genetic evaluation for coronary artery disease. *Genetics in Medicine*, 5(4), 269-285. doi:10.1097/01.gim.0000079364.98247.26
- Schmidt, A. F., Swerdlow, D. I., Holmes, M. V., Patel, R. S., Fairhurst-Hunter, Z., Lyall, D. M., . . . Sattar, N. (2017). PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol*, 5(2), 97-105. doi:10.1016/s2213-8587(16)30396-5
- Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., . . . Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep*, 17(8), 2042-2059. doi:10.1016/j.celrep.2016.10.061
- Schroeder, E. B., Chambless, L. E., Liao, D., Prineas, R. J., Evans, G. W., Rosamond, W. D., & Heiss, G. (2005). Diabetes, glucose, insulin, and heart rate variability: the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care*, 28(3), 668-674.
- Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., . . . Samani, N. J. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*, 43(4), 333-338. doi:10.1038/ng.784
- Shibata, A. (2017). Regulation of repair pathway choice at two-ended DNA double-strand breaks. *Mutation Research*, 803-805, 51-55. doi:10.1016/j.mrfmmm.2017.07.011
- Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., . . . Jeggo, P. A. (2011). Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO Journal*, 30(6), 1079-1092. doi:10.1038/emboj.2011.27

- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., . . . Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, 34(1), 57-65. doi:10.1002/humu.22225
- Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., . . . Koonin, E. V. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell*, 60(3), 385-397. doi:10.1016/j.molcel.2015.10.008
- Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(Web Server issue), W452-457. doi:10.1093/nar/gks539
- Snieder, H., Boomsma, D. I., Van Doornen, L. J., & De Geus, E. J. (1997). Heritability of respiratory sinus arrhythmia: dependency on task and respiration rate. *Psychophysiology*, 34(3), 317-328.
- Soria, L. F., Ludwig, E. H., Clarke, H. R., Vega, G. L., Grundy, S. M., & McCarthy, B. J. (1989). Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *Proceedings of the National Academy of Sciences of the United States of America*, 86(2), 587-591.
- Stainier, D. Y. (2001). Zebrafish genetics and vertebrate heart formation. *Nat Rev Genet*, 2(1), 39-48. doi:10.1038/35047564
- Stieber, J., Herrmann, S., Feil, S., Loster, J., Feil, R., Biel, M., . . . Ludwig, A. (2003). The hyperpolarization-activated channel HCN4 is required for the generation of pacemaker action potentials in the embryonic heart. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25), 15235-15240. doi:10.1073/pnas.2434235100
- Stoletov, K., Fang, L., Choi, S. H., Hartvigsen, K., Hansen, L. F., Hall, C., . . . Miller, Y. I. (2009). Vascular lipid accumulation, lipoprotein oxidation, and macrophage lipid uptake in hypercholesterolemic zebrafish. *Circulation Research*, 104(8), 952-960. doi:10.1161/circresaha.108.189803
- Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *European Heart Journal*, 17(3), 354-381.
- Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., . . . Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307), 707-713. doi:10.1038/nature09270
- Tessadori, F., van Weerd, J. H., Burkhard, S. B., Verkerk, A. O., de Pater, E., Boukens, B. J., . . . Bakkers, J. (2012). Identification and functional characterization of cardiac pacemaker cells in zebrafish. *PloS One*, 7(10), e47644. doi:10.1371/journal.pone.0047644
- The ENCODE project consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), 799-816.
- The Wellcome Trust Case Control Consortium. (2010). Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 464(7289), 713-720. doi:10.1038/nature08979
- Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., & Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun*, 7, 11750. doi:10.1038/ncomms11750

- van den Berg, M. E., Warren, H. R., Cabrera, C. P., Verweij, N., Mifsud, B., Haessler, J., . . . Munroe, P. B. (2017). Discovery of novel heart rate-associated loci using the Exome Chip. *Human Molecular Genetics*, 26(12), 2346-2363. doi:10.1093/hmg/ddx113
- van der Harst, P., & Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation Research*, 122(3), 433-443. doi:10.1161/circresaha.117.312086
- van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., . . . May, A. P. (2016). DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular Cell*, 63(4), 633-646. doi:10.1016/j.molcel.2016.06.037
- van Setten, J., Brody, J. A., Jamshidi, Y., Swenson, B. R., Butler, A. M., Campbell, H., . . . Sotoodehnia, N. (2018). PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun*, 9(1), 2904. doi:10.1038/s41467-018-04766-9
- Varshney, G. K., & Burgess, S. M. (2014). Mutagenesis and phenotyping resources in zebrafish for studying development and human disease. *Brief Funct Genomics*, 13(2), 82-94. doi:10.1093/bfpg/elt042
- Varshney, G. K., Pei, W., LaFave, M. C., Idol, J., Xu, L., Gallardo, V., . . . Burgess, S. M. (2015). High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Research*, 25(7), 1030-1042. doi:10.1101/gr.186379.114
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Zhu, X. (2001). The sequence of the human genome. *Science*, 291(5507), 1304-1351. doi:10.1126/science.1058040
- Verweij, N., van de Vegte, Y. J., & van der Harst, P. (2018). Genetic study links components of the autonomous nervous system to heart-rate profile during exercise. *Nat Commun*, 9(1), 898. doi:10.1038/s41467-018-03395-6
- von der Heyde, B. (2017). Zebrafish larvae as a model system for systematic image-based genetic screens of cardiac and cardiovascular risk factors and diseases. *Half time report*.
- Vornanen, M., & Hassinen, M. (2016). Zebrafish heart as a model for human cardiac electrophysiology. *Channels (Austin)*, 10(2), 101-110. doi:10.1080/19336950.2015.1121335
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153(4), 910-918. doi:10.1016/j.cell.2013.04.025
- Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA *Nat Commun* (Vol. 8).
- Webb, T. R., Erdmann, J., Stirrups, K. E., Stitzel, N. O., Masca, N. G., Jansen, H., . . . Kathiresan, S. (2017). Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. *Journal of the American College of Cardiology*, 69(7), 823-836. doi:10.1016/j.jacc.2016.11.056
- Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., . . . Abecasis, G. R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics*, 40(2), 161-169. doi:10.1038/ng.76

- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., . . . Abecasis, G. R. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, *45*(11), 1274-1283. doi:10.1038/ng.2797
- Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*, *10*(10), 1556-1566. doi:10.1038/nprot.2015.105
- Yang, H. J., Hsu, C. L., Yang, J. Y., & Yang, W. Y. (2012). Monodansylpentane as a blue-fluorescent lipid-droplet marker for multi-color live-cell imaging. *PloS One*, *7*(3), e32693. doi:10.1371/journal.pone.0032693
- Yin, L., Maddison, L. A., Li, M., Kara, N., LaFave, M. C., Varshney, G. K., . . . Chen, W. (2015). Multiplex Conditional Mutagenesis Using Transgenic Expression of Cas9 and sgRNAs. *Genetics*, *200*(2), 431-441. doi:10.1534/genetics.115.176917

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1547*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-378941



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019