Zebrafish models for large-scale genetic screens in dyslipidemia and atherosclerosis

Validation and application

MANOJ KUMAR BANDARU
Abstract

Hundreds of loci have been robustly associated with circulating lipids, atherosclerosis and coronary artery disease; but for most loci the causal genes and mechanisms remain uncharacterized. The overall aim of my thesis is to develop and validate novel in vivo model systems that are suitable for high-throughput, image-based genetic screens in coronary artery disease and related traits, and use these model systems to systematically characterize positional candidate genes.

In Study I, I developed an experimental pipeline to validate the suitability of zebrafish larvae as a model system for systematic, large-scale characterization of drugs and genes associated with dyslipidemia and atherosclerosis. Using this pipeline, I showed that five days of overfeeding and cholesterol supplementation have independent pro-atherogenic effects in zebrafish larvae, which could be diminished by concomitant treatment with atorvastatin and ezetimibe. CRISPR-Cas9-induced mutations in orthologues of proof-of-concept genes resulted in higher LDL cholesterol levels (apoE), and more early stage atherosclerosis (apoB.1). Finally, the pipeline helped me to identify putative causal genes for circulating lipids and early-stage atherosclerosis (LPAR2 and GATAD2A).

In Study II, I characterized cardiometabolic traits in apoc2 mutant zebrafish larvae and found that, similar to humans, larvae with two non-functional apoc2 alleles have higher whole-body levels of triglycerides and total cholesterol, and more vascular lipid deposition than larvae without mutations in apoc2. Interestingly, apoc2 mutant larvae also had lower glucose levels after adjusting for triglyceride levels, suggesting that therapeutic stimulation of apoc2 to prevent hypertriglyceridemia may result in hyperglycemia. Still, zebrafish larvae with mutations in apoc2 can be a useful model to identify and characterize additional causal genes for triglyceride metabolism.

In Study III, I examined the effects of mutations in pcsk9 on atherosclerosis and diabetes-related traits in nearly 5,000 zebrafish larvae. Similar to the loss-of-function mutations in PCSK9 in humans, larvae with mutations in pcsk9 had lower LDLc levels and were protected from early-stage atherosclerosis. Interestingly, mutations in pcsk9 also resulted in fewer pancreatic β-cells in 10 days old larvae, which suggests the higher risk of diabetes in humans with mutations in PCSK9 may result from a direct effect on the beta cell.

Based on these large-scale proof-of-concept studies, my thesis confirms that zebrafish larvae can be used for large-scale, systematic genetic screens in dyslipidemia and early-stage atherosclerosis.

Keywords: zebrafish, dyslipidemia, atherosclerosis, genetic screens, high-cholesterol diet, APOE, APOB, LDLR, APOC2, PCSK9, LPAR2, GATAD2A

Manoj Kumar Bandaru, Department of Immunology, Genetics and Pathology, Medicinsk genetik och genomik, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Manoj Kumar Bandaru 2019

ISSN 1651-6206
urn:nbn:se:uu:diva-397715 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-397715)
To Nannamma.
"A proper goodbye from your grandson".
List of Papers

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


1. Introduction .................................................................................................................. 11
  1.1. Coronary artery disease, survey of the field .......................................................... 11
    1.1.1. Global significance ......................................................................................... 11
    1.1.2. Pathophysiology ......................................................................................... 11
    1.1.3. Risk factors .................................................................................................. 13
    1.1.4. Genetics ........................................................................................................ 14
    1.1.5. Treatment ..................................................................................................... 17
    1.1.6. Mining GWAS loci for novel treatments ..................................................... 18
    1.1.7. Candidate gene prioritization ........................................................................ 18
  1.2. Zebrafish for translational research ........................................................................ 19
    1.2.1. Need for a new animal model ....................................................................... 19
    1.2.2. Scope for genetic screens using zebrafish model ......................................... 20
    1.2.3. Previous small-scale studies on using zebrafish to model dyslipidemia and atherosclerosis ................................................................. 20

2. Aims .............................................................................................................................. 25
  2.1. General aim ........................................................................................................... 25
  2.2. Specific aims ......................................................................................................... 25
    Study I ..................................................................................................................... 25
    Study II .................................................................................................................... 25
    Study III .................................................................................................................. 25

3. Study summaries .......................................................................................................... 26
    Study I ..................................................................................................................... 28
    Study II .................................................................................................................... 30
    Study III .................................................................................................................. 32

4. Concluding remarks .................................................................................................... 34

5. Future perspectives ..................................................................................................... 35

6. References .................................................................................................................. 39
Abbreviations

APOB  Apolipoprotein B
APOC-II Apolipoprotein C-II
APOE  Apolipoprotein E
CAD    Coronary artery disease
CETP   Cholesterol ester transfer protein
Chr    Chromosome
CI     Confidence interval
CRISPR Clustered regularly interspaced short palindromic repeats
DNA    Deoxyribonucleic acid
dpf    Days post-fertilization
EGFP   Enhanced green fluorescent protein
eQTL   Expression quantitative trait loci
GATAD2A GATA Zinc Finger Domain Containing 2A
GMIP   GEM Interacting Protein
gRNA   Guide ribonucleic acid
GWAS   Genome-wide association studies
HCD    High-cholesterol diet
HDLc   High-density lipoprotein cholesterol
HMGCR  3-hydroxy-3-methyl-glutaryl-coenzyme A reductase
LD     Linkage disequilibrium
LDLc   Low-density lipoprotein cholesterol
LDLR   Low-density lipoprotein receptor
LPAR2  Lysophosphatidic Acid Receptor 2
MAF    Minor allele frequency
LPL    Lipoprotein lipase
MPEG1  Macrophage Expressed Gene 1
MPO    Myeloperoxidase
NPC1L1 Niemann-Pick C1-Like Protein 1
oxLDL  Oxidized low-density lipoprotein
PCR    Polymerase chain reaction
PCSK9  Protein convertase subtilisin/kexin type 9
SD     Standard deviation
SNP    Single nucleotide polymorphism
TALEN Transcription activator-like effector nucleases
TM6F2  Transmembrane 6 Superfamily Member 2
VAST   Vertebrate automated screening technology
1. Introduction

1.1. Coronary artery disease, survey of the field

1.1.1. Global significance

Coronary artery disease (CAD) reached epidemic proportions during mid-20\textsuperscript{th} century\textsuperscript{1} and still continues to be the most common disease worldwide that causes disability and death. According to the World Health Organization statistics for 2015, an estimated 7.4 million people died from CAD, representing 13\% of all deaths globally\textsuperscript{2}. Although age-adjusted mortality due to CAD has gradually decreased in developed countries, it was an underlying cause of 1 of every 7 deaths during 2014\textsuperscript{3}. It is projected that by 2030, deaths due to CAD might increase by 14.9\% in men and 13.1\% in women\textsuperscript{4}. This illustrates the need for implementing more effective measures to identify and prevent the risk factors (primordial prevention) as well as to prevent clinical manifestations of CAD in the high risk group (primary prevention) and the affected group (secondary prevention). While the primordial, primary and secondary prevention of CAD have significantly improved, thanks to contributions from pathological, therapeutic, and epidemiological studies, the pharmaceutical industry has failed to develop conceptually new medication in the last three decades. Such developments are essential to reduce the projected consequences of CAD on morbidity and mortality in the decades to come.

1.1.2. Pathophysiology

In CAD, the coronary arteries supplying blood and nutrients to the heart become hard and narrow due to accumulation of lipids and immunogenic elements that form plaques in the artery wall (Figure 1). Functional studies in murine model systems and relevant cell types, as well as clinicopathological studies of autopsies\textsuperscript{5-7} have elucidated that atherosclerosis is the main underlying pathophysiological process of CAD. Early-stages of atherosclerosis manifests when cholesterol-rich apolipoprotein-B (apoB)-containing lipoproteins retain in the arterial wall\textsuperscript{8-10}. Normally, lipoproteins <70 nm in diameter flux into and out of the arterial wall\textsuperscript{11}. However, when their concentrations in plasma raise above the normal levels, a fraction of the lipoproteins that enter the arterial wall are retained\textsuperscript{12, 13}. The retained lipoproteins undergo modifica-
tions such as lipolysis, proteolysis and oxidization leading to endothelial activa-
tion and initiation of inflammatory processes such as recruitment of plate-
lets, neutrophils and monocytes to the scene\textsuperscript{14}, \textsuperscript{15}. Monocytes transmigrate across the arterial wall, where they proliferate and differentiate into macro-
phages, and take up the oxidized lipids, forming foam cells\textsuperscript{16}. In advance-stage atherosclerosis, apoptosis and necrosis of the foam cells, together with local accumulation of extracellular lipids, calcium and other debris can lead to fatty streak formation. This in turn triggers the recruitment of smooth muscle cells to form a fibrous cap over the necrotic core\textsuperscript{17}. Over time, intimal calcification, neo-vascularization of growing plaques, and degradation of the fibrous cap by proteases can increase the influx of pro-inflammatory T cells into atherosclerotic plaques, making them acute and unstable. Rupturing of such unstable plaques can ultimately lead to thrombosis and myocardial infarction\textsuperscript{18}. Despite some knowledge on how atherosclerosis develops, a precise understanding of its underlying molecular processes is still lacking. For example, we are far from understanding how the lipid loading into macrophages precisely affects macrophage biology and atherosclerotic processes\textsuperscript{19}. Understanding the events at each particular stage will likely shed light onto the process as a whole, and could potentially uncover new drug targets.

\textbf{Figure 1.} Etiology of coronary artery disease. The coronary arteries that normally carry blood and oxygen to the heart become hard and narrow in coronary artery disease, due to the formation of plaques consisting of lipids, macrophages and smooth muscles cells. Advanced plaques upon fissure can completely occlude the coronary arteries. Image was modified from Wikimedia Commons\textsuperscript{20}. 

![Coronary artery in normal condition](image1.png) ![Coronary artery with developing plaque](image2.png)

- Coronary arteries supplying blood and oxygen to heart
- Artery cross-section
- Normal blood flow
- Abnormal blood flow
- Lumen
- Narrowed lumen
- Plaque

\textcopyright 2023
1.1.3. Risk factors

Epidemiological studies have provided insights into the risk factors associated with the development of CAD. Multiple large-scale, longitudinal epidemiological studies have highlighted a range of risk factors, including behavioral habits (i.e. smoking, alcohol use, high-fat diet, and a sedentary lifestyle), medical conditions (i.e. diabetes, hypertension, dyslipidemia, and obesity), lipid components in blood (i.e. higher cholesterol, higher triglycerides, higher LDL cholesterol and lower high-density lipoprotein cholesterol levels in the circulation) and physical traits (i.e. higher age, male gender and short stature) (reviewed in ref. 21) (Figure 2). All the above risk factors are regarded as conventional risk factors of CAD, and all except age and sex are influenced by genetic factors. For example, over 250 genetic loci have now been identified as being robustly associated with diabetes, largely through insulin resistance and beta-cell dysfunction22-24. Our understanding of risk factors has helped develop risk prediction models that allow selection of individuals at risk for preventive strategies. However, up to 20% of CAD patients have no conventional risk factors and 40% have only one25. Therefore, to further improve the sensitivity of risk prediction models, contemporary studies are focusing on identifying molecular biomarkers that characterize pathological signs of the disease26,27.

Figure 2. Conventional risk factors associated with increased risk for CAD. All these risk factors are influenced by genetic factors and account for part of CAD heritability.
1.1.4. Genetics

Several epidemiological studies provided the clues for familial aggregation of CAD\textsuperscript{28, 29}. Subsequent studies in twins have shown that the heritability of CAD is estimated at 57% for men and 38% for women\textsuperscript{30, 31}. In 2004, the INTERHEART study confirmed that a positive family history of CAD is a risk factor, independently of age, sex, smoking and geographical region. After adjusting for genetically influenced conventional risk factors, the risk due to family history was only slightly attenuated, implying that additional genetic factors that directly influence disease risk play a role in disease susceptibility\textsuperscript{32}. These findings strengthened the arguments in favor of trying to improve our understanding of the genetic basis of CAD. Numerous genomic and genetic studies have since been employed to dissect the genetic factors underlying CAD.

Genome-wide linkage studies

Efforts to understand the genetic basis of CAD began in the 1980s, using a genome-wide linkage approach. Microsatellite markers (short tandem repeat sequences) were genotyped in related individuals, particularly in sibling pairs of CAD or myocardial infarction affected families, to locate the disease susceptibility loci flanked by the markers. There studies were not very successful\textsuperscript{33}, and the identified loci varied across families from different geographical regions, e.g. 13q12-13 in Iceland\textsuperscript{34}, 2q21.2-22 and Xq23-26 in Finland\textsuperscript{35}, 14q32.2 in Germany\textsuperscript{36}, 1p34-36, 3q13 and 5q31 in the US\textsuperscript{37}, and 2p11 and 17p11-17q21 in the UK\textsuperscript{38}. Fine mapping of locus 13q12-13 with additional markers led to the identification of arachidonate 5-lipoxygenase activating protein (ALOX5AP), which represents the only gene identified by an early-stage linkage study that we now know is indeed robustly linked to CAD risk\textsuperscript{34}. The lack of success in identifying susceptible loci or genes of statistically unequivocal linkage can largely be attributed to: 1) the extensive genetic heterogeneity of CAD; 2) the inability of the small set of microsatellite markers (~400 markers), which are spaced as far as ~10 centiMorgan apart; 3) inadequate statistical power due to small sample size; 4) absence of genotypic data for an adequate number of affected and unaffected family members\textsuperscript{39}; and 5) the low penetrance of effect for common variants.

Candidate-gene association studies

Most progress in understanding the genetic basis of CAD has come from association studies. Early association studies in the late 1980s attempted to identify genes associated with quantifiable risk factors of CAD (e.g. plasma lipoproteins, blood glucose) and rare single-gene (Mendelian) disorders that involve premature CAD (e.g. due to familial hypercholesterolemia and dyslipidemia). Such studies commenced by selecting putative candidate genes, either by their location in a region of linkage (for example, ALOX5AP at 13q12-13), or on the basis of other evidence that they might affect disease
risk— for example, genes involved in lipid metabolism were hypothesized to play a role in atherosclerosis. The selected candidate genes were compared between patients and controls for the allele frequencies of a set of variants within or near the genes. Such efforts demonstrated the importance of variants in genes encoding LDL receptor (LDLR)\textsuperscript{40}, cholesterol ester transfer protein (CETP)\textsuperscript{41}, apolipoprotein B (APOB)\textsuperscript{42} and apolipoprotein E (APOE)\textsuperscript{43} for lipid metabolism and CAD pathogenesis. However, few variants have been consistently replicated across studies. This is at least in part due to differences in study design and very limited statistical power (i.e. no more than a few hundred cases per study) in most studies\textsuperscript{44, 45}. Moreover, the biased approach of examining only a subset of SNPs within a subset of genes using a small sample size increases the likelihood of missing common and rare variants with low to modest effect sizes\textsuperscript{45}. This issue was circumvented to some extent by comprehensive sequencing of candidate genes in a large population (comprising patients and controls) to identify both common (minor allele frequency – MAF >5%, i.e. 1 in 20) and rare (MAF <1%) variants that have functional consequences. Deep sequencing of PCSK9 (encoding proprotein convertase subtilisin/kexin type 9 serine protease), for example, has led to identification of rare variants that reduce LDLc by 40% and confer protection from CAD\textsuperscript{46}. Another candidate-gene sequencing effort found that individuals with low levels of HDLc were significantly more likely to harbor rare nonsynonymous variants in ABCA1, APOA1, and LCAT, which had previously been implicated in dyslipidemia\textsuperscript{47}. However, hypothesis-driven candidate gene association studies mostly pointed towards obvious disease pathways, thereby limiting the scope for identifying new pathways and developing new therapies, with the exception of PCSK9.

Genome-wide association studies
In addition to candidate-gene association studies that screen for a set of genetic variants within pre-selected genes, genetic epidemiologists have also employed comprehensive, hypothesis-free screening for genetic variants across the genome that are associated with CAD or CAD-associated traits. This approach has become known as the genome-wide association (GWA) approach. In GWA studies (GWAS), genomes of tens of thousands of CAD cases and CAD-free controls can now be examined for a few million common SNPs (typically MAF >1%). GWAS are largely based on the accomplishments of the Human Genome Project\textsuperscript{48}, HapMap Project\textsuperscript{49}, and 1000 Genomes Project\textsuperscript{50}, which provided a reference of 3.2 billion nucleotide base pairs of the human genome and 11 million relatively common single nucleotide polymorphisms (SNPs) that serve as high density genetic markers for the entire genome\textsuperscript{51}. Also, the significant progress of genotyping arrays\textsuperscript{52} that simultaneously type a large number of common variants (now up to 5 million SNPs) in a quick and efficient process have made GWAS more feasible. In 2007, three GWAS in data from up to 23,000 individuals for ~1 million SNPs reported a
strong association of SNPs on chromosome 9p21.3 with CAD risk\textsuperscript{53-55}. Following these initial publications, numerous studies have confirmed the currently poorly understood role of the chromosome 9p21.3 region on CAD risk. The next round of GWAS for CAD identified 11 more SNPs, where each additional copy of the risk allele conferred a 30-40\% higher risk of CAD\textsuperscript{56}. At the same time, additional GWAS and follow-up studies reported 11 other loci that were robustly associated with risk of CAD\textsuperscript{57-60}. Findings in these initial GWAS studies demonstrated that common susceptibility variants for CAD carried minimal incremental risk, and additional loci would require very large sample sizes - in the order of tens of thousands to hundreds of thousands - to be uncovered\textsuperscript{61}. Recognizing the need for larger samples sizes, researchers that previously published successful results independently started collaborating, to form the CARDIoGRAM consortium (comprising >20,000 cases and >60,000 controls) and the C4D Genetics Consortium (>15,000 cases and >15,000 controls) in 2011. Subsequent efforts confirmed associations for nearly all previously reported loci and uncovered 17 new loci through meta-analysis of all available results across consortia\textsuperscript{62, 63}. Through further collaboration between these two consortia, and testing associations for up to 2.5 million SNPs (85\% with MAF >1\%), the CARDIoGRAMplusC4D consortium (>60,000 cases and >120,000 control) reported an additional 15 previously unanticipated loci in 2013\textsuperscript{64}, and eight more loci in 2015, by interrogating 6.7 million common (MAF >5\%) and 2.7 million low-frequency (0.5\% < MAF < 5\%) variants\textsuperscript{65}. More recently, two separate studies querying additional samples and SNPs have identified a further 20 previously unanticipated loci\textsuperscript{66, 67}. Altogether, these studies have resulted in the identification of 78 CAD-associated loci in largely European and, to a lesser extent, South Asian populations. Approximately 20\% of the loci are located near genes with known roles in metabolism of LDLc (e.g. \textit{LDLR}, \textit{APOB}, \textit{APOE}, \textit{PCSK9} and \textit{HMGCR}) or triglyceride-rich lipoproteins (e.g. \textit{APOA1}-\textit{APOC3}-\textit{APOA4}, \textit{APOA3}, and \textit{APOC5}), reinforcing the key roles of these pathways in the development of CAD, and providing internal validation of GWAS findings\textsuperscript{68}. An additional 10-20\% of loci are related to blood pressure, inflammation, cellular proliferation and/or vascular remodeling\textsuperscript{68}. For most other loci (~60\%), the causal genes and mechanisms in disease susceptibility have not yet been identified\textsuperscript{69}.

\textbf{Exome sequencing studies}

Although GWAS for CAD have identified many susceptibility loci harboring variants (MAF >1\%) with modest effect sizes, the majority of these variants are located outside the protein-coding regions (exons), making it difficult to pinpoint causal genes and affected pathways\textsuperscript{68, 70}. As exonic variants most often exert functional consequences, and the translation of their association to causal genes would be more straightforward, geneticists started to test associations of rare variants within protein-coding regions of the genome (exome). Such exome association studies typically involve sequencing of exomes from...
individuals experiencing clinical forms of CAD alongside controls, followed by mapping of protein-altering variants (e.g. nonsynonymous, splice-site, and nonsense single-nucleotide substitutions) and testing their association with CAD risk. In 2014, a collaborative endeavor of many exome sequencing studies as a part of the Exome Sequencing Project has contributed to the development of an exome genotyping chip that includes a subset of putative functional exonic variants (~240,000) with MAF >0.1%, selected from data in >12,000 individuals. Since then, many studies employed both exome sequencing - to find novel variants - and genotyping to validate findings. A study by the Exome Sequencing Project sequenced exomes of 18,666 genes in about 4000 individuals and identified three rare loss-of-function variants (0.7% MAF) in \textit{APOC3} (apolipoprotein C-III) that were associated with triglyceride levels. Carriers of these rare variants had a 40% lower risk of CAD. Another exome sequencing effort with data from ~5000 early-onset MI cases and MI-free controls identified rare coding variants in \textit{LDLR} (0.5% MAF) and \textit{APOA5} (apolipoprotein A-V) (0.5% MAF). \textit{LDLR} mutation carriers had higher plasma LDLc, whereas \textit{APOA5} mutation carriers had higher plasma triglyceride levels, increasing the risk of MI by 4.2- and 2.2-fold respectively. Another finding was related to 15 rare loss-of-function variants (0.2% MAF) in \textit{NPC1L1} (Niemann–Pick C1-like 1), which were associated with lower LDLc and a 53% relative reduction of CAD risk. Despite the success of these studies in identifying rare variants, many exclusively pointed to previously established genes for lipoprotein metabolism. However, additional discoveries from whole-exome sequencing are expected as larger sample sizes become available.

1.1.5. Treatment

Adherence to a healthy lifestyle is highly recommended to both prevent and manage CAD. As Elliot Joslin stated: “Genetics loads the gun, but the environment pulls the trigger”. CAD became more manageable after therapeutic studies had developed a range of medical and surgical procedures. Medical therapies to prevent disease progression and recurrent cardiac events primarily include medication to lower lipids (e.g. using statins and ezetimibe), lower blood pressure (e.g., atenolol, ramipril, valsartan) and prevent platelet coagulation (e.g. aspirin, clopidogrel, warfarin). Surgical revascularization procedures like stent implantation and bypass grafting are commonly employed during clinical manifestations of CAD (i.e. thrombosis, angina, myocardial infarction) to prevent sudden cardiac death. Because of a substantial risk for recurrence of cardiac events, medical therapy is intensified following a surgical procedure. However, it must be noted that some medical therapies can cause side effects. For example, lipid lowering drugs (statins) were shown to increase the risk of diabetes in recent clinical trials and genetic association studies, and currently used anticoagulants increase the risk of cerebral
bleeding\textsuperscript{82}. The combination of an increasing global incidence and undesirable side effects of available therapies presents the need for novel treatment strategies for CAD.

1.1.6. Mining GWAS loci for novel treatments

As mentioned earlier, the majority of GWAS-identified loci for CAD are not known to be related to conventional risk factors or pathways. For example, loci 9p21.3 and 6p24.1 have been strongly associated with CAD in several independent GWAS (reviewed in \textsuperscript{69}), but the causal genes and mechanisms that influence disease risk remain incompletely understood for 9p21.3\textsuperscript{83} and unknown for 6p21.4\textsuperscript{69}. Identifying and characterizing the causal genes in such loci will increase our understanding of CAD pathophysiology, and may lead to novel drug targets.

In theory, translating GWAS loci to genes and disease mechanisms seems straightforward: first, one identifies promising candidate genes within the locus; next, one characterizes the candidate genes to identify the putative causal genes; and finally, one examines pathways and mechanisms through which putative genes act. However, in practice there are challenges at every stage. Firstly, while candidate genes can be identified using available bioinformatics tools, each tool uses different databases (e.g. expression quantitative trait loci and epigenetic marks) and algorithms to pinpoint promising candidate genes, and hence, multiple genes may be predicted to be causal within each locus. To not throw away the child with the bath water, it is advisable to employ a range of bioinformatics tools and select all top candidate genes for experimental follow-up. Secondly, animal models are required to characterize promising candidate genes, but the traditionally used murine model systems cannot provide the throughput that is required for systematic characterization. Finally, putative causal genes identified using an experimental model system have to be further validated and characterized in higher order vertebrate models to confirm how perturbation of a gene affects processes at a molecular, cellular, physiological, and whole organismal level in the context of the development of CAD.

1.1.7. Candidate gene prioritization

A range of bioinformatics tools permit systematic analysis of GWAS results (i.e. using summary statistics) to identify positional candidate genes. These tools operate by: 1) identifying the putative causal variants; and/or 2) identifying the genes that are functionally related to genes in other identified loci. During the former step, both the significantly associated variants and the variants that are in high linkage disequilibrium (LD) with the significantly associated variants are individually evaluated for their functional implications,
such as amino acid changes, presence in known regulatory regions of the genome, and LD with expression quantitative trait loci using the publicly available genomic data in Encyclopedia of DNA elements (ENCODE), RoadMap Epigenomics and Genotype-Tissue Expression (GTEx) projects\textsuperscript{84}. Variants with possible functional consequences are considered as putative causal variants and are taken forward to next step. Tools like GenomeRunner\textsuperscript{85}, RiVIERAbeta\textsuperscript{86} and FineMap\textsuperscript{87} facilitate effective assessment of putative causal variants. In the next step, genes affected by the putative causal variants are determined with tools like Sherlock\textsuperscript{88}, RegulomeDB\textsuperscript{89}, and MetaXcan\textsuperscript{90}, which evaluate the possible associations of variants at a transcriptomic and translational level – such as altering protein function, transcript abundance, splice sites, and regulation of expression of genes located further away from lead SNPs\textsuperscript{91}. Some tools (e.g. MetaXcan\textsuperscript{90}, DEPICT\textsuperscript{92} and Pascal\textsuperscript{93}) additionally provide tissues, cell types and pathways by which putative causal variants and candidate genes are anticipated to influence the disease-associated traits. However, as mentioned earlier, it is recommended to employ a range of bioinformatics tools and select top candidate genes across approaches. Since the results from bioinformatic annotation remain prediction, these results should be used to inform further functional studies, e.g. in animal models; not as providing an absolute truth.

1.2. Zebrafish for translational research

1.2.1. Need for a new animal model

Traditionally, candidate genes within a disease-associated locus have been investigated by knock-out or over-expression of the corresponding orthologous genes in an appropriate animal model, and subsequent characterization in the context of disease-associated phenotypes. For many years, mouse models have served as the premier vertebrate model system to perform such genetic studies, including in cardiovascular diseases\textsuperscript{94}. The apolipoprotein E (\textit{ApoE}) and low-density lipoprotein (\textit{Ldlr}) knockout mice are well-established mouse models that are commonly used in cardiovascular disease research\textsuperscript{95,96}. However, mouse models do not facilitate the large-scale screening for candidate genes in GWAS-identified loci, for reasons like deficient library of knockout models, difficulties of embryonic manipulations, slow development and life cycles, slow development of atherosclerosis, and expensive maintenance costs.
1.2.2. Scope for genetic screens using zebrafish model

During the last decade, the zebrafish has become increasingly popular to study human diseases, since they resemble humans in their development and metabolism\(^97\), and have a well-annotated genome with orthologues for at least 70% of human genes\(^98\). Embryonic genetic manipulation has been challenging for decades, but recent developments in multiplexed mutagenesis using Clustered Regulatory Interspaced, Short Palindromic Repeats (CRISPR) and CRISPR associated systems (Cas9) can now be efficiently applied in zebrafish to introduce mutations in up to eight genes simultaneously\(^99\). In this context, it is helpful that zebrafish eggs are fertilized and develop externally, and re-implantation of embryos after mutagenesis is not required. The after-effects of the induced mutations and the disease progression at cellular, physiological and whole-organismal levels can be conveniently monitored in zebrafish larvae thanks to their innate optical transparency until ~30 days post-fertilization (dpf). The availability of transgenic lines with fluorescently labelled cells, and the possibility to administer desired reagents (e.g. fluorescent dyes) through bodily infusion are additional strengths of this vertebrate model system. Moreover, recent improvements in automated handling, positioning and orienting of zebrafish larvae imply that relevant cellular processes can now be automatically imaged and quantitatively analyzed in high-throughput\(^100\), making full use of the advantage of the zebrafish’s high fecundity (~300 offspring per week).

1.2.3. Previous small-scale studies on using zebrafish to model dyslipidemia and atherosclerosis

Small-scale studies suggested that zebrafish larvae may be used to model dyslipidemia and early-stage atherosclerosis. Firstly, larvae fed on a high-cholesterol diet for 10 days become hypercholesterolemic, and accumulate lipids and oxidation-specific epitopes in their vasculature, much like in early-stage atherosclerosis in humans\(^101-103\). Secondly, additional suppression of \(l dl_r\) expression during the first 8 days of larval development drastically increased \(L DL_c\) levels, indicating that the role of \(l dl_r\) in \(L DL_c\) metabolism is conserved in zebrafish\(^104\). Thirdly, zebrafish mutants with nonfunctional apolipoprotein CII (APOCII) fed on a normal diet were hypertriglyceridemic, and accumulated lipids and lipid-laden macrophages in their vasculature, much like patients with \(A P O C I I\) deficiency\(^105\). Fourthly, treating hypercholesterolemic and hypertriglyceridemic zebrafish with commonly prescribed lipid-lowering drugs, i.e. statins and/or ezetimibe, ameliorated cholesterol and triglyceride levels\(^104, 106, 107\), and high resolution imaging showed lower levels of vascular lipid deposition in zebrafish larvae\(^107\). An overview of the studies performed so far, including the sample size, is shown in Table 1.
<table>
<thead>
<tr>
<th>Findings</th>
<th>Interventions</th>
<th>Samples</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>High cholesterol diet (HCD)-induced dyslipidemia</td>
<td>HCD (3 M)</td>
<td>11 adults</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>HCD (14 D)</td>
<td>Homogenates of 20 larvae (n=3)</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>HCD (14 D)</td>
<td>Homogenates of 20 larvae (n=6)</td>
<td>106</td>
</tr>
<tr>
<td>Hypercholesterolemia-induced vascular lipid deposition</td>
<td>HCD+BODIPY-C11 (10 D)</td>
<td>4 larvae</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>HCD (3 M)</td>
<td>9 adults</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>HCD+BODIPY-C11 (14 D)</td>
<td>4 larvae</td>
<td>102</td>
</tr>
<tr>
<td>HCD-induced oxidation of lipoproteins</td>
<td>HCD (3 M)</td>
<td>8 adults</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>HCD (14 D)</td>
<td>Homogenates of 50 larvae (n=3)</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>HCD (14 D)</td>
<td>6 larvae</td>
<td>102</td>
</tr>
<tr>
<td>Recruitment of monocytes to the vascular lipid deposits</td>
<td>HCD (10 D)</td>
<td>12 larvae</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>HCD (14 D)</td>
<td>Homogenate of 50 larvae</td>
<td>108</td>
</tr>
<tr>
<td>Lipid-lowering drugs reduce HCD-induced dyslipidemia</td>
<td>HCD+Simvastatin (14 D)</td>
<td>Homogenates of 20 larvae (n=2)</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>HCD+Ezetimibe (14 D)</td>
<td>Homogenates of 20 larvae (n=3)</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>HCD (10 D), HCD+Atorvastatin (5 D)</td>
<td>Not available</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>HCD (10 D), HCD+Ezetimibe (5 D)</td>
<td>Not available</td>
<td>107</td>
</tr>
<tr>
<td><em>apoc2</em> mutant zebrafish as a model for hypertriglyceridemia</td>
<td><em>apoc2</em> knockout; Normal diet (10 D)</td>
<td>3 larvae</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td><em>apoc2</em> knockout; Normal diet (3 M)</td>
<td>6 adults</td>
<td></td>
</tr>
<tr>
<td>Hypertriglyceridemia-induced early-stage atherosclerosis</td>
<td><em>apoc2</em> knockout; HCD+BODIPY (10 D)</td>
<td>8 larvae</td>
<td>105</td>
</tr>
<tr>
<td>Role of disrupted <em>ldlr</em> in hyperlipidemia</td>
<td><em>ldlr</em> knockout; HCD (2 D)</td>
<td>Homogenates of 100 larvae (n=2)</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td><em>ldlr</em> knockout</td>
<td>4 larvae</td>
<td>109</td>
</tr>
</tbody>
</table>
### Table 1

<table>
<thead>
<tr>
<th>Findings</th>
<th>Interventions</th>
<th>Samples</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role of disrupted <em>ldlr</em> in vascular lipid deposition</td>
<td><em>ldlr</em> knockdown; HCD (2 D)</td>
<td>45 larvae</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td><em>ldlr</em> knockout; HCD+BODIPY (5 D)</td>
<td>15 larvae</td>
<td>109</td>
</tr>
<tr>
<td><em>Lxra</em> mutant zebrafish as model for hypercholesterolemia</td>
<td><em>nr1h3</em> knockout; HCD (21 D)</td>
<td>3 to 5 adults</td>
<td>110</td>
</tr>
<tr>
<td>Activity of <em>cetp</em> increases in hypercholesterolemic zebrafish</td>
<td>HCD+Turmeric extract (5 W)</td>
<td>80 adults</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>HCD+Grape extract (5 W)</td>
<td>70 adults</td>
<td>112</td>
</tr>
<tr>
<td>Expression analysis of apolipoproteins- A, B, &amp; E in zebrafish</td>
<td>In-situ hybridization from 1to 6 D</td>
<td>3 larvae</td>
<td>113</td>
</tr>
</tbody>
</table>

Abbreviations: HCD- high cholesterol diet; apocc2- apolipoprotein CII; ldlr- low-density lipoprotein receptor; Lxra- liver x receptor alpha; nr1h3- nuclear receptor subfamily 1, group H, member 3; cetp- cholesterol ester transfer protein
Most of the studies highlighted in Table 1 took advantage of available reporter lines (Table 2), in conjunction with using lipid labelling dyes to characterize early-stage atherogenic processes. Advancing further, Fang et al. engineered a transgenic zebrafish that conditionally expresses a fluorescently labelled human monoclonal antibody IK17, which specifically binds to and labels oxidized LDL. The studies listed in Table 1 also used various enzymatic assays, immunoassays, and chromatography techniques on homogenates of 25 to 100 larvae per sample to characterize dyslipidemia and atherosclerosis in zebrafish larvae. So far, gene expression or function have been altered using morpholino-based downregulation or TALEN-based mutagenesis techniques.

On the need for further validation and method development and refinement

Like humans, zebrafish are genetically heterogeneous. As a result, data from a large number of larvae is needed to draw meaningful conclusions. Although the aforementioned proof-of-principle studies have so far provided promising results, they were based on observations in fewer than 25 larvae per condition (Table 1). This was at least in part because mounting larvae in low melting agarose for imaging is time-consuming. In addition, whole-body cholesterol, triglyceride and protein levels were usually quantified in homogenates of 20-100 pooled larvae (Table 1). While suitable and efficient for dietary and drug treatment interventions, pooling larvae for phenotypic characterization is not optimal in genetic interventions, where – depending on the approach – sequencing of individual larvae may be required. Next, results from downregulation of gene expression using morpholinos have to be interpreted with caution as in many studies, morphant phenotypes could not be confirmed in mutants. While TALEN-based mutagenesis is a better alternative to morpholinos in this aspect, it lacks in the ability to edit multiple genes simultaneously. This might be necessary when there are multiple zebrafish orthologues for a human gene, and all of them have to be targeted together to ensure that no genetic compensation occurs for the loss of one or the other.

With the recent developments in imaging and image-quantification technologies, it is now possible to acquire and quantify image-based traits in a large number of larvae. For example, with a Vertebrate Automated Screening Technology (VAST) Bioimager coupled to a fluorescence microscope, larvae can automatically be positioned and oriented within the microscope’s field of view, and fluorescent images of larvae can be acquired in a high-throughput. The acquired images can subsequently be quantified in batch mode using image-analysis software like CellProfiler. Furthermore, lipid-analyzers such as Mindray BS-380 that are designed to analyze lipid fractions in human plasma samples can be used to quantify whole-body levels of LDLc, HDLc, triglycerides, total cholesterol and glucose at single larval level. Finally, using the state-of-the art multiplexed CRISPR-Cas9 mutagenesis, it is now possible
to target up to at least nine genes simultaneously\textsuperscript{99,118,119}. By combining these technical advances in imaging, image-analysis and mutagenesis, it is now possible to target multiple zebrafish orthologues together and examine their effects in a large number of larvae.

Hence, confirmation of initial findings, an improved resolution of quantitative readouts, and a higher throughput are desirable before we can conclude if zebrafish larvae can be used as a model system for large-scale characterization of candidate genes for dyslipidemia, atherosclerosis and CAD.

<table>
<thead>
<tr>
<th>Table 2. Transgenic zebrafish and fluorescent tracers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tools</strong></td>
</tr>
<tr>
<td><strong>Transgenic zebrafish</strong></td>
</tr>
<tr>
<td>fli1:EGFP*</td>
</tr>
<tr>
<td>lyz:EGFP, lyz:DsRed2</td>
</tr>
<tr>
<td>mpeg1:EGFP</td>
</tr>
<tr>
<td>mpo:EGFP*</td>
</tr>
<tr>
<td>hsp70-IK17:EGFP*</td>
</tr>
<tr>
<td><strong>Lipoprotein/lipid tracers and enzyme substrates</strong></td>
</tr>
<tr>
<td>Dil-LDL</td>
</tr>
<tr>
<td>cholesterol BODIPY-C11</td>
</tr>
<tr>
<td>NBD-cholesterol</td>
</tr>
<tr>
<td>BODIPY-C12</td>
</tr>
<tr>
<td>PED6</td>
</tr>
<tr>
<td>Monodansylpentane*</td>
</tr>
</tbody>
</table>

Table adapted from\textsuperscript{103}

Abbreviations: hsp- heat-shock protein; CE- cholesterol ester; PLA\textsubscript{2}- phospholipase A\textsubscript{2}

*Tools used in this thesis
2. Aims

2.1. General aim
My thesis is based at the intersection of the challenges for functional validation of candidate genes for coronary artery disease, that is, to develop and validate novel in vivo model systems that are suitable for high-throughput, image-based genetic screens in coronary artery disease and related traits, and use these model systems to systematically characterize positional candidate genes.

2.2. Specific aims
Study I
To validate zebrafish as a model system for dyslipidemia, atherosclerosis and coronary artery disease using a dietary intervention, a treatment intervention with lipid lowering drugs, and a multiplexed, CRISPR-Cas9-based genetic intervention of proof-of-concept genes. Further, to characterize four candidate genes in a GWAS-identified, pleiotropic locus on chr 19p13.11. (Manuscript in revision)

Study II
To confirm or refute the role of apolipoprotein CII in hypertriglyceridemia and early-stage atherosclerosis in zebrafish larvae; and to examine if apolipoprotein CII represents a promising target for therapeutic intervention. (Manuscript)

Study III
To examine the effects of mutations in proprotein convertase subtilisin/kexin type 9 (PCSK9) on dyslipidemia, atherosclerosis and diabetes-related traits in zebrafish larvae, and assess if these effects are consistent with loss-function mutations in PCSK9 in humans and with PCSK9 inhibitors in humans. (Manuscript)
3. Study summaries

Study design, methods and main results

An overview of each of the Studies, I-III, is presented in Table 3.
Table 3. Overview of Studies I, II and III

<table>
<thead>
<tr>
<th>Study</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
</table>
| **Interventions and sample sizes** | *Diet – overfeeding, cholesterol supplementation; n=2366*  
*Drugs – atorvastatin plus ezetimibe; n=1063*  
*Zebrafish orthologues of proof-of-concept genes – APOE, APOB, LDLR; n=383*  
*Zebrafish orthologues of candidate genes – LPAR2A, GATA2A, TM6SF2, GMIP; n=547* | Zebrafish orthologue of apolipoprotein CII (*apoc2*); n=384 | Zebrafish orthologue of proprotein convertase subtilisin/kinase type 9 (*pcsk9*); n=1121 |
| **Traits quantified** | *Whole-body lipid and glucose levels*  
*Vascular early-stage atherogenic traits (incl. accumulation of lipids, macrophages, neutrophils and oxidized LDL)*  
*Body size* | *Whole-body lipid and glucose levels*  
*Vascular early-stage atherogenic traits (incl. accumulation of lipids, macrophages and neutrophils)*  
*Body size*  
*Lipid deposition in liver* | *Whole-body lipid and glucose levels*  
*Vascular early-stage atherogenic traits (incl. accumulation of lipids, macrophages and neutrophils)*  
*Body size*  
*Liver size and lipid deposition in liver* |
| **Analysis** | *Overfeeding vs. Normal feeding*  
*Cholesterol supplementation vs. No cholesterol*  
*Drug treatment vs. No treatment*  
*For each targeted gene, 2 vs.0 mutated alleles*  
*For each targeted gene, effect of each additional mutated allele*  
*Effect of each additional mutated allele across all targeted genes* | *2 vs. 0 mutated apoc2 alleles (with and without including additional wildtype larvae from other screens)*  
*Effect of each additional mutated apoc2 allele (with and without including additional wildtype larvae from other screens)* | *Effect of each additional mutated pck9 allele by including additional wildtype larvae from other screens* |
Study I

Zebrafish larvae as a model system for systematic characterization of drugs and genes in dyslipidemia and atherosclerosis

In this study, I performed a large-scale validation of the zebrafish model system for dyslipidemia and atherosclerosis using 1) a dietary intervention; 2) a treatment intervention with lipid lowering drugs; and 3) genetic intervention of proof-of-concept genes. Finally, I used the model system to characterize four candidate genes in GWAS-identified pleotropic locus on chr 19p13.11.

In the dietary intervention, atherogenic potential of overfeeding and dietary cholesterol supplementation was examined in nearly 2000 transgenic larvae with fluorescently labelled macrophages, neutrophils, oxidized LDL and/or vascular endothelium. In the treatment intervention with the LDLc lowering drugs atorvastatin and ezetimibe, nearly 1000 transgenic larvae were overfed on a cholesterol supplemented diet with or without the drug treatment. In the genetic intervention of proof-concept genes - APOE, APOB and LDLR were targeted together in zebrafish with labelled macrophages and neutrophils using a multiplexed CRISPR-Cas9 approach. Similarly, the four candidate genes – LPAR2, GATAD2A, TM6SF2 and GMIP in the GWAS-identified locus were genetically intervened together. In both genetic interventions, offspring (F1 generation) from the CRISPR founders were examined. All larvae were first visualized and quantified for early-stage atherosclerotic traits at 10 days post-fertilization using a semi-automated imaging set up (i.e. A Vertebrate Automated Screening Technology [VAST] BioImager and fluorescent microscope) and objective image quantification pipelines. Subsequently, whole-body lipid and glucose levels were quantified at the single larval level using enzymatic assays, and CRISPR-Cas9 target sites in each larva were sequenced using paired-end sequencing. Finally, associations between the exposures (i.e. diet, drugs, or genes) and effects on the quantified traits were analyzed.

The results show that five days of overfeeding and cholesterol supplementation have independent pro-atherogenic effects, including elevated total cholesterol and triglyceride levels, more vascular deposition of lipids and oxidized LDL, and more co-localization of lipids with macrophages and neutrophils. Combined treatment with the lipid lowering drugs for five days protected larvae from the pro-atherogenic effects of overfeeding and cholesterol supplementation. CRISPR-Cas9-induced mutations in orthologues of proof-of-concept genes resulted in higher LDLc levels (apoea), and more early stage atherosclerosis (apobb.1). Finally, our pipeline helped to identify putative causal genes for circulating lipids and early-stage atherosclerosis (LPAR2 and GATAD2A) (Figure 3).
In conclusion, our pipeline facilitates systematic, *in vivo* characterization of drugs and candidate genes to increase our understanding of disease etiology, and can likely help identify novel targets for therapeutic intervention.

**Figure 3.** The effect of overfeeding and cholesterol supplementation (n>2000), treatment with atorvastatin and ezetimibe (n>1000), and mutations in *apoea, apobb.1, lpar2a* and *gatad2ab* on whole-body lipid and glucose levels, vascular atherogenic traits and body size. Across all forest plots, circles represent the effect sizes and whiskers represent 95% confidence intervals. In the forest plots for *apoea* and *apobb.1*, open circles and dotted lines represent the effect of two functionally knocked-out alleles vs. two unmodified alleles, and full circles and filled lines represent the additive per mutated allele effect.
Study II

Apoc2 mutant zebrafish: a model for hypertriglyceridemia and early-stage atherosclerosis

A zebrafish model system for hypertriglyceridemia can be useful to systematically characterize positional candidate genes for triglyceride metabolism. Recent studies performed elsewhere suggested that zebrafish with mutations in apolipoprotein CII (apoc2) have elevated triglyceride levels and more early-stage atherosclerosis. However, results were based on data from only three wildtype and three mutant larvae. Hence, in Study II, I generated apoc2 founders using CRISPR-Cas9, and phenotypically characterized their offspring to: 1) further validate our pipeline for systematic characterization of candidate genes for dyslipidemia; 2) confirm or refute a role for apoc2 in hypertriglyceridemia and early-stage atherosclerosis in zebrafish larvae; and 3) examine if apolipoprotein CII represents a promising target for therapeutic intervention.

I targeted apoc2 in transgenic zebrafish with fluorescently labelled macrophages and neutrophils and examined the larvae (F1 generation) from an incross of founders. Larvae were fed on regular dry food enriched with 4% extra cholesterol from the age of 5 days post-fertilization (dpf) until 9 dpf. At 10 dpf, 384 larvae were imaged to visualize body size, vascular atherogenic traits and lipid accumulation in the liver, followed by semi-automated quantification of whole-body lipoprotein and glucose levels and paired-end sequencing of the CRISPR-targeted site.

Of the 384 sequenced F1 larvae, 174, 119 and 61 larvae carried 0, 1 and 2 mutated apoc2 alleles. Compared with the 174 larvae free of mutations in apoc2, the 35 larvae with anticipated complete loss of apoc2 had higher whole-body levels of triglycerides, HDLc and total cholesterol, and a tended to have lower whole-body glucose levels. Mutant larvae also tended to have more vascular lipid deposition. Interestingly, the trends for lower whole-body glucose levels and more vascular lipid deposition in larvae with anticipated loss of functional apoc2 reached significance when larvae (n=3812) from other screens, in which apoc2 was not experimentally perturbed were included as additional wildtype controls (Figure 4).

Thus, our large-scale study confirms the role of apoc2 in hypertriglyceridemia and early-stage atherosclerosis. While apoc2 mutant zebrafish model can be used as a genetic background to identify and characterize causal genes for triglyceride metabolism, independent and opposite effects on triglycerides and glucose suggest that APOCII may not be suitable as a target for prevention and treatment of coronary artery disease.
Figure 4. The effect of mutations in *apoc2* (n=384) on whole-body lipid and glucose levels, vascular atherogenic traits, metabolic traits and body size. Forest plot on the left includes 174 wildtype larvae identified within the *apoc2* genetic screen. Forest plot on the right includes additional wildtype larvae from other screens in which *apoc2* was not experimentally perturbed. In both forest plots, open circles and the dotted lines represent the effect of two functionally knocked-out alleles vs. two unmodified alleles, and full circles and filled lines represent the additive per mutated allele effect. Circles represent the effect sizes and whiskers represent 95% confidence intervals.
Study III

*Image-based, in vivo characterization of cardiometabolic consequences of mutations in pcsk9*

Inhibitors of proprotein convertase subtilisin/kexin type 9 (PCSK9) have emerged as a new class of drugs, reducing LDLc by 50–60% in several clinical settings, i.e. more than the maximal dose of statins plus ezetimibe combined typically accomplish. However, Mendelian randomization studies and clinical studies suggest that the LDLc lowering effect of PCSK9 inhibition may be accompanied by an elevated risk of incident diabetes. In this study, we mimicked the mechanistic action of PCSK9 inhibitors in humans by inducing mutations in *pcsk9* in zebrafish and examined the effects of such mutations on dyslipidemia, early-stage atherosclerosis and diabetes-related traits in nearly 5000 zebrafish larvae.

We targeted *pcsk9* using CRISPR-Cas9 mutagenesis to generate founder fish (F₀ generation) in two reporter backgrounds: 1) with fluorescently labelled macrophages and neutrophils; and 2) with fluorescently labelled insulin expressing cells and hepatocytes. At 10 dpf, 2x384 F₁ larvae in the former background were imaged for vascular atherogenic traits, and another 353 F₁ larvae in the later background were imaged for diabetes-related traits. Furthermore, body size and whole-body lipid and glucose levels were quantified in all larvae. Larvae were finally sequenced at the CRISPR-targeted site in *pcsk9* using paired-end sequencing.

Of the total 1121 screened F₁ larvae, 15, 24 and 1019 larvae carried 0, 1 and 2 mutated *pcsk9* alleles respectively. Given the small number of larvae with 0 or 1 mutated alleles, we used comparable data from other screens in which *pcsk9* was not experimentally perturbed as wildtype controls, that is, larvae from a dietary intervention (n=2366), drug treatment intervention with atorvastatin and ezetimibe (n=1063), genetic screen for *APOE, APOB* and *LDLR* (n=383), and a genetic screen for *APOC2* (n=384), i.e., larvae characterized in Studies I and II. Compared with nearly 2500 larvae free from mutations in *pcsk9* that were included in Studies I and II, each additional mutated allele in *pcsk9* was associated with lower LDLc levels, on average. Each additional mutated allele in *pcsk9* was also associated with less vascular lipid deposition, less co-localization of macrophages with lipids, more co-localization of macrophages with neutrophils and fewer pancreatic β-cells (*Figure 5*).

In conclusion, our findings in *pcsk9* mutant larvae are in line with results from people carrying loss-of-function *PCSK9* mutations, and are also in line with the effects of PCSK9 inhibitors in humans. Furthermore, the direct effect of mutations in *pcsk9* on the number of pancreatic β-cells, suggests that the higher risk of diabetes in humans with mutations in *PCSK9* may result from a direct effect on the beta cell.
Figure 5. The effect of each additional mutated \textit{pcsk9} allele (n=1121) on whole-body lipid and glucose levels, vascular atherogenic traits, metabolic traits and body size. Given the small number of wildtype larvae (n=15) within the \textit{pcsk9} genetic screen, wildtype larvae from other screens in which \textit{pcsk9} was not experimentally perturbed were include in the association analysis. Full circles and filled lines represent the additive per mutated allele effect.
4. Concluding remarks

The three large-scale proof-of-concept studies described in this thesis confirm that zebrafish larvae can be used to model dyslipidemia and early-stage atherosclerosis. These studies also highlight the necessity to screen a large number of zebrafish larvae in order to draw firm conclusions on the role of genes in complex disease-related traits. Further, I demonstrated how the model system can be used to characterize candidate genes in GWAS-identified loci. This approach presents an opportunity to increase our understanding of the molecular causes of cardiometabolic diseases, and reduce the hundreds of candidate genes in GWAS-identified loci to a more feasible number for: 1) further in-depth characterization using higher order animal models; 2) targeted association studies and genotype-based recall efforts using exome-sequencing data in humans; and 3) adequately powered, target-based small molecule screens.
5. Future perspectives

**Further characterization of the putative causal genes—**LPAR2 and GATD2A. In Study I, the preliminary characterization of candidate genes in a pleiotropic locus on chr 19p13.11 highlighted a role for LPAR2 in cholesterol metabolism and early-stage atherosclerosis, and for GATD2A in cholesterol metabolism. However, these findings were based on a small number of mutant larvae (11 heterozygous mutants for lpar2a and 32 for gatad2ab). To draw firm conclusions on the role of these genes, founders for these genes were generated again, and the screening of their offspring is currently being finalized. Perhaps, as demonstrated in Study III, the combined phenotyping for early-stage atherosclerosis and diabetes related traits may also shed light on the potential of these genes for being useful therapeutic targets.

**Optimizing mutagenesis efficiency for an adequate distribution of genotypes.** In genetic screen for two of the genes in Study I (apoba and apobb.1), and for pesc9 in Study III, there were many larvae with two mutated alleles, and few wildtype controls, reflecting the high mutagenic activity of the CRISPR guide RNAs used for the mutagenesis, in spite of pre-testing and selecting the CRISPR guide RNAs that are more likely to induce mutations in one of the two alleles of a gene (heterozygous mutations) in the founder fish. In order to have an adequate number of wildtype controls in future genetic screens, alongside embryos injected with highly efficient guide RNAs, some non-injected siblings will be raised to adulthood in separate tanks. Then, phenotypic characterization on F1 larvae from in-crosses of injected and non-injected fish should result in a higher statistical power.

**Genetic screens in apoc2 or apobb.1 knockout background at 5 days post fertilization (dpf).** As larvae with mutations in apobb.1 (Study I) and apoc2 (Study II) developed hypertriglyceridemia and early-stage atherosclerosis after 4 days of feeding on high-cholesterol diet, it would be interesting to examine if these traits manifest already at 5 dpf, i.e. without a dietary intervention. If so, then performing genetic screens in one of these mutant backgrounds would allow phenotypic characterization of larvae at 5 dpf without having to feed them on high-cholesterol diet. This would eliminate any batch effects that may prevail during the feeding or maintenance of larval tanks, as well as speed up the screening process.
Acknowledgements

This work has been accomplished with the tireless efforts of so many. I would like to take this opportunity to express my gratitude to:

My Main Supervisor Marcel den Hoed, for the opportunities, endless support and help. Your enthusiastic guidance, persistent encouragement and amazing patience through my Master’s thesis to Ph.D. thesis helped me to get this far. Thanks for bringing out the best in me almost always.

My co-supervisor Anastasia Emmanouilidou, thanks for setting up the lab and making it a pleasant playground. I am truly grateful for your guidance throughout the experimental part of this work. Thanks for being supportive all these years and correcting me when necessary.

My co-supervisor Erik Ingelsson, I am sincerely thankful for your feedback on the work and availability despite your busy schedule.

Carolina Wählby and Petter Ranefall, thanks for your help in establishing the image analysis pipelines and constructive feedback on manuscripts. Anders Larsson, thanks for kindly analyzing all the samples in this work. Your efforts are highly appreciated. Collaborating with all of you was a pleasant experience. To Hannah Brooke, a special thanks for your input with the statistical analysis and data visualization.

Benedikt, thanks for programming the pipetting robot and making the sample preparation easy. Also, thanks for being that friend with whom I can talk beyond work. I thoroughly enjoyed your company in- and out of the office. Eugenia, you have been a very big helping hand in my projects. Many thanks for sharing the work load, and at times, for taking the entire load on you. Also, thanks for developing the variant-calling pipeline. I feel lucky to have such understanding fellow Ph.D. students.

Tiffany, the CRISPR-mutant lines would not come to life without you. Thanks for the help in generating them and also for teaching me to inject. Joao, thanks for your help in setting up our own fish room. Your DIY ideas are impressive! Ci and Natalie, ex-members in the group, thanks for your intellectual input during our weekly group meetings. Sharing office-space with all of you has been fun.
A big thank you goes to the whole Zebrafish Facility at EBC for their support during the beginning of this work: Johan, Katarina, Judith, Beata and everyone else I had the privilege to meet.

To the Epidemiology Group- Tove, Mwenya, Chris, Markus and Stefan; and to the Cell Group- Susanne, Casimiro and Naomi, I am grateful for the many hours of scientific and social engagement.

The many students that joined the group to conduct their thesis projects kindly helped in fish husbandry and other laboratory work. Thanks to each one of you- Lisa, Mauro, Sitaf, Silvia, Chrissy, Mehran, David, Adriana, Shreya, Aida, Taya, Anne, Marta, Endrina, Tiarne, Azizah, Bobby, Tessa and Amine. Also, thanks to Lynsey for taking care of the fish when there were only a few people in the group.

Departments –IMV and IGP have provided an excellent infrastructure for my research and professional development. I particularly thank Anna Foyer, Elin Ekberg, Ulrica Bergström and Helene Norlin for the administrative support.

My extended gratitude to those who balanced the work with the other aspects of life:

To friends in Uppsala who have made life funnier with gatherings💬💭, sports💪🏼, festivals🎆🎄 and parties🎉ワイン. Thanks to each one of you- Sunil, Sethu, Srinivas, Sudharsan, Sandeep, Hari, Javeed, Brajesh, Srinu T, Raghu, Mahesh, Varun, Gopi, Ram, Phani, Prakash anna, Chandu anna, Kiran anna, Ramesh anna, Srisailam anna, Soudarya, Anusha, Christine, Saritha, Ammulu, Prathusha, Indu and Sowmya. Huda, Sharanya, Sony, Sandhya and Geetha akka, special thanks for serving delicious food always😋.

To my buddies from Bachelors and School- Abhishek, Charan and Rupa thanks for being in touch despite your busy lives and for your concern on my wellbeing. Naresh, I will forever be grateful for your help in getting the admission in Uppsala University, thanks.

Finally, to my family. Dad and Mum, thanks for pushing your boundaries throughout your life to give me a comfortable life🏠. Souji, Pranay, Raja, Pinne and Babaye, heartfelt thanks for filling the son’s shoes and taking care of Dad and Mum in times when I am supposed to be there🏠. Kanaka, my better helf full, I can’t thank you enough for your love, support and sacrifice above and beyond the call of duty شكراً. Thanks for taking care and for always having my back👨‍❤️‍👨. And to our❤️, Vismai for being such an easy kid to handle.
6. References


57 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.


116 Malzahn A, Lowder L, Qi Y. Plant genome editing with TALEN and CRISPR. Cell Biosci. 2017;7:21-.


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”.)