General discussion
Heritable cardiovascular disease (CVD) often has a large impact on patients' lives, given the relatively young age at presentation, the potentially lifelong consequences, the occurrence of disease among close relatives, and the recurrence risk in (future) offspring. The aim of this thesis was to expand our knowledge regarding the pathogenesis of heritable CVD. This knowledge is essential for adequate genetic counseling of the patients and families involved, implementation of gene-tailored management strategies, and the development of future therapies.

Our studies focused on three main disease categories: thoracic aortic aneurysms and dissections, left-sided congenital heart disease, and pediatric cardiomyopathy. In this chapter, we highlight our main findings, and discuss the implications and suggestions for future research to further improve our understanding of the molecular mechanisms underlying these disorders.

Identification of new genes and pathways

The advent of next-generation sequencing (NGS) technologies has revolutionized the field of medical genetics, providing fast and cost-effective generation of genome-wide sequencing data. Whole exome sequencing (WES) and whole genome sequencing (WGS) have largely replaced traditional gene discovery approaches, such as positional cloning and Sanger sequencing of candidate genes. In recent years, NGS has also been widely adopted in the diagnostic setting. Patients with heritable CVD particularly benefit from these advancements, because these disorders display marked genetic heterogeneity and overlapping phenotypes [1]. Indeed, the yield of genetic testing in heritable CVD has significantly improved. For example, comprehensive genetic testing allows the identification of putative disease-causing variants in up to 60% of patients with hypertrophic cardiomyopathy [2, 3]. However, in certain types of CVD, such as thoracic aortic aneurysms (TAA) and dissections, the sensitivity of genetic testing remains low [4]. In this thesis, we exploited the power of NGS to identify new genes and pathways involved in CVD. The majority of CVD are thought to result from multifactorial inheritance, involving a complex interplay between multiple susceptibility genes and environmental factors. In our studies, we focused on monogenic forms of CVD. Insights obtained from the study of monogenic forms of CVD may ultimately improve our understanding of the molecular processes influencing more common and complex forms of the disease.

In the studies described in this thesis, different strategies were employed for finding new genes involved in both syndromic and non-syndromic CVD. The gene for Loeys-Dietz syndrome type 5, TGFB3, was identified using traditional linkage analysis and Sanger sequencing of candidate genes (Chapter 2.3). This strategy depends on the availability of large families with multiple affected family members and prior knowledge of the disease mechanism. Nowadays, Sanger
sequencing is often replaced by WES or WGS. The latter two can also be used for subsequent identification of additional patients, as genome-wide sequence data is readily accessible for retrospective analysis. However, linkage analysis remains a powerful tool to increase diagnostic power [5]. The alpha-kinase 3 (ALPK3) gene was identified using a combination of homozygosity mapping and exome sequencing in several consanguineous families with severe early-onset cardiomyopathy (Chapter 4.2). The homozygous regions detected by genome-wide genotyping were used to reduce the number of candidate variants, thereby greatly facilitating the identification of the causative gene. ALPK3 encodes a nuclear kinase that is essential for early differentiation of cardiomyocytes. Though the exact cellular and molecular mechanisms resulting in cardiomyopathy are yet unclear, the acquired knowledge offers a unique opportunity to explore new treatment strategies that aim to restore protein kinase activity and modification of target proteins. Whole exome sequencing was also applied in another family with early-onset cardiomyopathy with presumed autosomal recessive inheritance but no indication for parental consanguinity (Chapter 4.4). Here, filtering for rare homozygous or compound heterozygous variants resulted in the identification of a single candidate gene, ASNA1. Knockout of the corresponding gene in zebrafish resulted in a similar phenotype. ASNA1 codes for an ATPase required for post-translational insertion of tail-anchored proteins into the endoplasmic reticulum [6]. Our study provides the first evidence that this pathway is involved in human disease, and offers new insight into the molecular mechanisms underlying cardiomyopathy.

**Unsolved cases**

In a substantial proportion of patients and families under study, we were unable to detect the causal gene. The incomplete and age-dependent penetrance of CVD may hamper genetic studies by reducing the number of affected individuals, making large families (such as in Chapter 2.3) difficult to ascertain. Disease penetrance likely depends on a combination of genetic, epigenetic and environmental factors, blurring the distinction between monogenic and complex disorders [7]. Along the same line, hormonal factors or sex chromosomes may contribute to the marked gender differences (i.e. prevalence, presentation and progression) observed in CVD phenotypes [8]. Gene-based burden testing is an alternative approach for gene discovery in monogenic disorders with incomplete penetrance. However, the extreme genetic heterogeneity in CVD will require large sample sizes [9], with several hundreds or thousands of cases, that may exceed practical limits. In fact, the majority of CVD probably represent complex traits and therefore require different approaches to develop polygenic risk scores [10]. Genome-wide association studies (GWAS) are most effective in detecting common variants with small effect size. However, the genetic heterogeneity in CVD will limit the power to recover meaningful associations. Moreover, it has become clear that common variants, individually or in combination, explain only a small proportion of the heritability.
of human diseases [11]. Instead, to detect rare variants with large effect size, an analytical framework combining the aggregate burden of individual variants into a subset of genes and their corresponding pathways is probably more successful [12]. Subsequent integration of personal genomes and health records may help to predict clinical outcomes and assess the effectiveness of lifestyle modifications to reduce disease risk.

Other reasons that might explain the failure to identify a compelling candidate gene include: (i) incomplete capture of protein-coding regions (especially those with high GC content and large repeats) or low sequencing depths; (ii) inability to detect certain classes of variants, e.g. copy number variants (in particular single-exon deletions or duplications) and trinucleotide repeats; (iii) disease-causing variants residing outside the targeted exome sequence, e.g. deep intronic variants, variants in regulatory elements, and unannotated genes; (iv) non-Mendelian inheritance, e.g. somatic mosaicism, genomic imprinting, and complex inheritance; (v) the presence of phenocopies, e.g. hypertension-induced left ventricular hypertrophy and aortic dilatation; (vi) failure to correctly differentiate the disease-causing variant from the many rare non-pathogenic variants; (vii) limited availability of DNA samples due to CVD-associated premature mortality.

Considering the extreme genetic heterogeneity of CVD, we believe that studying individual families with CVD will remain the most powerful tool for identification of new disease genes for the foreseeable future. WES is now routinely performed in the diagnostic setting and has come to serve a dual role in genetic diagnosis (known genes) and gene discovery (new disease genes). WGS, on the other hand, offers more uniform coverage of the coding region of the genome, allowing more reliable detecting of single nucleotide and copy number variants. Thus, it is to be expected that WGS will further increase the diagnostic yield in patients with suspected heritable CVD. WGS also permits the identification of variants in non-coding regions, such as enhancers, that make up approximately 98% of the human genome. We do, however, still have very limited understanding of the function of these non-coding regions and their potential role in the pathogenesis of CVD. Experimental approaches to identify and functionally validate potential enhancers, such as chromatin immunoprecipitation combined with massively parallel reporter assays [13], can help to localize active enhancers involved in cardiac cells. WGS can then be used to look for disease-causing variants in these enhancers in patients with unexplained CVD.

### Challenges of variant interpretation

While the application of NGS has increased the diagnostic power, it also massively increased the number of inconclusive test results due to variants of unknown significance (VUS). In
In general, the number of VUS discovered even exceeds the increased diagnostic yield from expanded genetic testing. These problems particularly arise in genes that have not been extensively studied in the past (e.g. truncating variants in \textit{TTN}) \cite{14}, and genes with a controversial role in the disease (for example, see \textit{Chapter 4.1}). However, difficulties in variant interpretation also remain in well-studied genes, as illustrated by the large number of unclassified missense \textit{NOTCH1} variants in our cohort of patients with left-sided congenital heart defects (\textit{Supplemental Table 1} in \textit{Chapter 3.1}).

To overcome these challenges, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMG) published a classification framework for sequence variant interpretation \cite{15}. These guidelines have been widely adopted by the clinical genetics community. Disease- and gene-specific adaptations to the ACMG/AMP classification framework may further improve and standardize variant interpretation \cite{16, 17}. When assessing variant pathogenicity using these guidelines, it is important to keep several things in mind. First, variant interpretation requires specific knowledge of the gene involved (e.g. mutation hotspots, functional domains or alternative transcripts). With the application of genome-wide sequencing, molecular or genetic specialists may be confronted with variants in genes outside their field of expertise, which may complicate variant interpretation. Second, caution is needed in drawing conclusions from publicly available population databases, such as the Genome Aggregation Database (gnomAD). These cohorts may not contain suitable ethnically-matched controls in the context of a particular patient under study. For example, individuals of Turkish and Moroccan origin, which form an important fraction of the multi-ethnic patient population served by our hospital, are underrepresented in many large-scale control databases. Determining that a variant is absent or present in extremely low frequency in controls (weighted as moderate criterion in the ACMG/AMP guidelines) is therefore less informative in these cases. In addition, population databases may not only encompass healthy individuals. The NHLBI GO Exome Sequencing Project (ESP), for example, includes individuals with heart, lung and blood disorders, and is therefore less useful in studies focusing on CVD. Furthermore, individuals with asymptomatic CVD, e.g. normally functioning bicuspid aortic valve, or healthy individuals carrying CVD-causing variants with incomplete penetrance may also be included in population databases. Ideally, controls should be selected from the same source population as the case(s), and should undergo extensive cardiovascular screening to exclude potential subclinical phenotypes.

Sharing the available clinical and genetic data through international collaborations, such as the European Network on Rare Multisystemic Vascular Diseases (VASCERN), will greatly facilitate the interpretation of variants and genes of unknown significance. Data sharing should therefore receive high priority, while ensuring informed consent and protecting the privacy
of the participants involved. Only through these efforts, we will be able to exploit genetic data to its full potential.

**Expanding phenotypes**

CVD is characterized by extreme genetic heterogeneity, i.e. different genes and variants can cause identical phenotypes. Conversely, different phenotypes may result from the same genetic defect (pleiotropy). In this thesis, we further delineate the phenotypes associated with variants in *NOTCH1*, *PKP2* and *ALPK3*. Variants in *NOTCH1*, encoding the transmembrane receptor protein NOTCH1, have previously been associated with left-sided congenital heart disease (CHD), such as bicuspid aortic valve, aortic valve stenosis, coarctation of the aorta, and hypoplastic left heart syndrome (HLHS) [18]. We report on 14 families with a protein-truncating variant in *NOTCH1* and expand the phenotypic spectrum to include right-sided CHD, conotruncal heart defects, and TAA (Chapter 3.1). Penetrance was high: 75% of all carriers (probands excluded) had a cardiovascular malformation. Our findings indicate that genetic testing of *NOTCH1* is warranted in a broad range of CHD. In addition, in carriers of a pathogenic variant in *NOTCH1*, special attention should be paid to ascending aortic diameters during follow-up. Heterozygous variants in *PKP2*, encoding the desmosomal protein plakophilin-2, have previously been associated with arrhythmogenic right ventricular cardiomyopathy [19]. We report the first family with HLHS caused by a homozygous truncating variant in *PKP2* (Chapter 3.2). Our findings support that plakophilin-2 is essential for cardiac morphogenesis [20, 21], and that – at least in some cases – left ventricular hypoplasia may result from a primary defect in myocardial development [22]. We suggest *PKP2* for incorporation in diagnostic gene panels for HLHS. In Chapter 4.3, we expand the phenotypic spectrum of *ALPK3*-related cardiomyopathy. As confirmed in subsequent studies [23-25], a peculiar pattern of progression is observed: from dilated or mixed cardiomyopathy at birth to hypertrophic cardiomyopathy in childhood. Besides early-onset cardiomyopathy, patients often display dysmorphic facial features, short stature, and musculoskeletal abnormalities. Our findings provide important new details on the clinical course of *ALPK3*-related cardiomyopathy, that is essential for genetic counseling of the families involved.

We recommend adhering to a two-step approach for genetic testing in CVD, starting with targeted genetic analysis of known CVD genes, optionally followed by genome-wide analysis. Targeted analysis is still preferred whenever possible to prevent unexpected diagnosis, i.e. pathogenic findings not explaining the phenotype, and minimize the number of VUS. Extended analysis, though, may identify variants in genes in which the particular phenotype is rarely observed or genes that have not been associated with the particular phenotype in the past. These genes might nonetheless be good candidates, based on prior knowledge of biologic pathways and processes or results from animal studies, as illustrated by the aforementioned cases.
Towards effective management strategies

To ensure adequate referral and increase cost-effectiveness, we determined which patients would benefit most from genetic testing of known TAA genes (Chapter 2.1). We found that young age of onset (<50 years), positive family history and presence of syndromic features increase the likelihood of detecting a disease-causing variant, and therefore justify genetic counseling and testing. Without these inclusion criteria, the diagnostic yield was less than 10% [4]. Relatives of TAA patients without a disease-causing variant are, however, still at increased risk of developing TAA. This may be explained by shared genetic factors that are not detected by current diagnostic genetic tests (e.g. common variants of small effect size). As long as we are not able to accurately predict an individual’s risk for TAA in mutation-negative families, cardiovascular screening in first-degree relatives of TAA patients should be encouraged, based on specific characteristics on the index patient (Chapter 2.1). Large prospective cohort studies are essential to identify etiological factors that discriminate families with high and low risk for TAA, and subsequent development of evidence-based guidelines on family screening.

Though the yield is generally low, genetic testing for TAA is of particular importance as it helps guide medical and surgical management [26]. For example, in patients with Loeys-Dietz syndrome, vascular disease is not limited to the aortic root, requiring imaging of the complete arterial tree [27]. In addition, given the aggressive nature of vascular disease in Loeys-Dietz syndrome, prophylactic aortic root surgery is recommend at a threshold of 4.0 to 4.5 cm, depending on the specific subtype [28, 29]. In vascular Ehlers-Danlos syndrome, on the contrary, surgical intervention is often reserved for life-threatening situations only, given the high complication rates [30]. In other forms of CVD, such as cardiomyopathies, the effect of genetic testing on disease management is less evident. To date, only few correlations between specific genotypes and cardiac phenotypes have been established. As a result, genetic testing provides little prognostic information for the families involved. However, we believe that precise phenotyping in large patient cohorts could lead to refinements in risk stratification and eventually more individualized care.

The identification of new genes and pathways helps to unravel the complex molecular mechanisms underlying CVD. Finding effective treatment options will rely heavily on this expanding knowledge. Because most CVD are multifactorial, therapeutic interventions will need to target multiple genetic and non-genetic risk factors, e.g. combination drug therapy and lifestyle interventions. The generation of patient-derived induced pluripotent stem cells, which can be differentiated into cardiomyocytes, vascular smooth muscle cells and endothelial cells [31-33], provides researchers with opportunities to study the safety and efficacy of new drugs and treatment protocols at the variant level. The possibility to correct the disease-causing variant...
in patient-specific cells, and subsequent transfer of healthy cells back into the patient, holds great promise for the treatment of monogenic forms of CVD.

**Future directions**

Next-generation sequencing has greatly advanced the ability to identify new disease genes. Still, we are only just beginning to understand the molecular mechanisms underlying CVD. Further research should focus on resolving the missing heritability of complex forms of CVD, which include the majority of CVD patients. So far, however, most of our understanding of the pathogenesis of multifactorial CVD comes from studies on rare monogenic forms of CVD. Elucidating the genetic factors in monogenic forms of CVD should therefore receive our continued attention.

WES is now routinely performed in the diagnostic setting and has dramatically increased the diagnostic yield in families with CVD. With upgraded bioinformatics pipelines as well as rapid identification and publication of new disease genes and variants, reanalysis of WES data, though time-consuming, has shown to be very effective in unsolved cases [34, 35]. However, with steadily declining costs, WGS will soon replace most other genetic testing strategies as first-tier diagnostic test. Large-scale application of WGS will also enable to identify genetic variants that contribute to multifactorial forms of CVD. However, assessing their exact contribution and translating these findings into clinical practice will pose significant challenges for many years to come.

Next-generation sequencing often results in the identification of genes and variants of unknown significance. Careful evaluation and classification of these variants is needed, considering the potential impact on clinical decision making. International data sharing initiatives will greatly increase the likelihood of finding specific genes and variants that are relevant in CVD. In many cases, however, functional validation is needed to obtain definite evidence of pathogenicity. Future research should focus on the development and validation of sensitive, high-throughput assays to assess the functional impact of VUS in CVD genes. These assays may, for example, include *in vitro* enzymatic activity assays [36], contractility assays in smooth muscle(-like) cells [37], and CRISPR/Cas9-mediated genome editing in zebrafish [38]. Functional genetics will soon become part of routine diagnostics. The involvement of epigenetic mechanisms, including histone modification, DNA methylation and non-coding RNA, in the pathogenesis of CVD also warrants further investigation. Epigenetic changes offer a possible explanation for incomplete penetrance and variable expressivity, and provide interesting opportunities to develop personalized therapeutic interventions.
References


