Summary

The work presented in this thesis describes the development and application of new techniques for detecting small variations (mutations) in genomic DNA that underlie various disorders. These techniques include High Resolution Melting Curve Analysis (HR-MCA) followed by Sanger sequencing, targeted, X-exome and whole exome capture followed by Next Generation Sequencing (NGS).

Chapter 2 describes the use and implementation of HR-MCA followed by Sanger sequencing as a pre-sequencing routine diagnostic scanning method for all exons and exon-intron junctions of the DMD gene. After validating the technique, we screened a group of 22 unrelated DMD/BMD patients and 11 females in which deletions and duplications of the gene had been excluded. Seventeen different pathogenic mutations were found in the screened group, of which ten were novel. Our results show that HR-MCA is a powerful and inexpensive diagnostic pre-sequencing scanning method to detect small mutations in BMD/DMD patients and carriers.

In chapter 3 we describe the application of array-based sequence capture (385K NimbleGen arrays) to enrich the exons and immediate intron flanking sequences of 112 genes, which are potentially involved in mental retardation and congenital malformation. Captured material was sequenced using Illumina technology and a data analysis pipeline was built. Our data show that: 1) An array-based sequence capture followed by Illumina sequencing, offers a versatile tool for successfully selecting sequences of interest from a total human genome. 2) All known variants were reliably detected. 3) Although overall coverage varied considerably, it was reproducible per region and facilitated the detection of large deletions and duplications (CNVs), including a partial deletion in the B3GALTL gene from a patient sample. 4) There is room for improvement of the methodology for ultimate diagnostic application, in particular with respect to array design that can obtain a more even coverage of the targeted regions.

In chapter 4, we performed X-exome capture followed by Illumina (Genome Analyzer II) sequencing in two probands from Dutch and Italian families with Terminal Osseous Dysplasia (TOD). TOD is an X-linked dominant male-lethal disease, characterized by terminal skeletal dysplasia, pigmentary defects of the skin, and recurrent digital fibroma during infancy. Previous linkage studies have mapped the disease-causing gene to Xq27.3-q28. After analyzing the data,

we identified a silent variant at the last nucleotide of exon 31 of the FLNA gene in both patients. The same variant c.5217G>A was also found in another four unrelated cases but not in 400 control X chromosomes, the 1000 Genomes, or in the database for FLNA gene variants. In families, this variant co-segregated with the disease. Our data show that due to nonrandom X chromosome inactivation, the mutant allele was not expressed in patient fibroblasts. RNA expression of the mutant allele was detected only in cultured fibroma cells obtained from material that had been surgically removed 15 years ago. The variant activates a cryptic splice site, removing the last 48 nucleotides from exon 31. At the protein level, this results in a loss of 16 amino acids (p.Val1724_Thr1739del), predicted to remove a sequence at the surface of filamin repeat 15. Our data show that TOD is caused by single unique recurrent mutation in the FLNA gene.

In chapter 5, we have used whole exome sequencing to identify pathogenic mutations causing autosomal recessive Spinocerebellar ataxia type 7 (SCAR7). The locus of SCAR7 has been linked to chromosome band 11p15. We have now identified the causative gene for SCAR7 by exome sequencing in the index family. One missense and one splice site mutation were found in the *TPP1* gene which co-segregated with the disease. The same mutations were found in an unrelated patient with a similar phenotype. Affected individuals showed low activity of tripeptidyl peptidase1, the protein coded by *TPP1*, the gene known to cause the infantile form of Neuronal Ceroid Lipofuscinosis (CLN2). However, the patients that we studied had none of the findings that are characteristic for CLN2: epilepsy, ophthalmic abnormalities, curvilinear bodies in the skin biopsy tissue. Also, the slow progressive evolution of the disease until old age of the patients is clearly different from the relentless progression in infancy known for CLN2.

In chapter 6, we studied the genetic cause of Chudley-McCullough Syndrome (CMS). We sequenced the exomes of three patients with CMS from two unrelated Dutch families from the same village and identified the same homozygous frameshift *GPSM2* variant c.1473delG in all patients. This variant was confirmed by Sanger sequencing in all affected patients and in a heterozygous form in their parents. We have shown that this variant arises from a shared, rare haplotype. Our data confirm the recent finding of Doherty *et al*., who reported *GPSM2* variants as a cause of CMS. The c.1473delG mutation in *GPSM2* associated with CMS appears to be an ancient founder mutation brought to North America by early Mennonite settlers originating from

Western Europe. Furthermore, we have established an LOVD database for *GPSM2* containing all variants thus far described.

In chapter 7, we describe the successful application of whole exome sequencing for finding the genetic cause of Faciocapulohumeral dystrophy type 2 (FSHD2). FSHD is the third most common myopathy, which is characterized by progressive and irreversible weakness of the facial, shoulder and upper arm muscles. In the majority of cases, the FSHD type 1 (FSHD1) is caused by contraction of the D4Z4 repeat array on a specific permissive allele on chromosome 4. This leads to local chromatin relaxation and stable expression of the D4Z4-encoded *DUX4* retrogene in skeletal muscle. In FSHD2, the myopathy results from chromatin relaxation and stable *DUX4* expression but without D4Z4 array contraction. To determine the genetic cause of FSHD2 and to identify the locus controlling the D4Z4 hypomethylation, we performed whole exome sequencing of twelve individuals from seven unrelated FSHD2 families: Five with dominant segregation of the hypomethylation and two with sporadic hypomethylation. Different mutations in *SMCHD1* (Structural Maintenance of Chromosomes flexible Hinge domain containing ***1)*** were identified in all affected individuals except in one family. We used Sanger sequencing to confirm the presence of these mutations and included 12 additional unrelated families with FSHD2 from whom DNA or RNA was available. We identified heterozygous out-of-frame deletions, heterozygous missense, and splice-site mutations in *SMCHD1* in 15/19 (79%) families. Mutations in *SMCHD1* substantially reduce the *SMCHD1* protein levels in skeletal muscle, which leads to contraction-independent *DUX4* expression. Furthermore, we found that mutations in *SMCHD1*, which is on chromosome 18, segregates independently of the FSHD-permissive *DUX4* allele on chromosome 4. This results in a digenic inheritance pattern in affected individuals. FSHD2 occurs exclusively in individuals who inherit both the *SMCHD1* mutation and a normal-sized D4Z4 array on a chromosome 4 haplotype permissive for *DUX4* expression. This showed that *SMCHD1* is an epigenetic modifier of the D4Z4 metastable epiallele and is a key genetic determinant of FSHD2 disorder.

Finally in chapter 8 we have discussed the pros and cons of all the techniques that we have presented in this thesis. These methods have made a significant contribution to accurate molecular diagnosis and to quick identification of disease causing genes.