DISSERTATIONES
MEDICINAE
UNIVERSITATIS
TARTUENSIS
263

SANDER PAJUSALU

Genome-wide diagnostics of Mendelian disorders: from chromosomal microarrays to next-generation sequencing

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
263

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
263

SANDER PAJUSALU

Genome-wide diagnostics of Mendelian disorders: from chromosomal microarrays to next-generation sequencing

UNIVERSITY OF TARTU
Press

Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in medicine) on 18th October 2017 by the Council of the Faculty of Medicine, University of Tartu, Estonia.

Supervisors: Professor Katrin Õunap, MD, PhD, Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Estonia

Associate Professor Tiia Reimand, MD, PhD, Department of Biomedicine, Institute of Biomedicine and Translational Medicine, and Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Estonia

Reviewers: Professor Pille Taba, MD, PhD, Department of Neurology and Neurosurgery, Institute of Clinical Medicine, University of Tartu, Estonia

Professor Pärt Peterson, PhD, Department of Biomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, Estonia

Opponent: Professor Helena Kääriäinen, MD, PhD, National Institute for Health and Welfare, Helsinki, Finland

Commencement: 13th December 2017

Publication of this dissertation is granted by the University of Tartu.

ISSN 1024-395X
ISBN 978-9949-77-625-2 (print)
ISBN 978-9949-77-626-9 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I Pajusalu S, Žilina O, Yakoreva M, Tammur P, Kuuse K, Mölter-Väär T, Nõukas M, Reimand T, Õunap K. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. Mol Syndromol. 2015 Sep;6(3):135–40.

II Pajusalu S, Kahre T, Roomere H, Murumets Ü, Roht L, Simenson K, Reimand T, Õunap K. Large Gene Panel Sequencing in Clinical Diagnostics – Results from 501 Consecutive Cases. Clin Genet. 2017 Apr 5. (Epub ahead of print)

III Pajusalu S, Reimand T, Õunap K. Novel homozygous mutation in KPTN gene causing a familial intellectual disability-macrocephaly syndrome. Am J Med Genet A. 2015 Aug;167A(8):1913–5.

IV Pajusalu S, Talvik I, Noormets K, Talvik T, Põder H, Joost K, Puusepp S, Piirsoo A, Stenzel W, Goebel HH, Nikopensius T, Annilo T, Nõukas M, Metspalu A, Õunap K, Reimand T. De novo exonic mutation in MYH7 gene leading to exon skipping in a patient with early onset muscular weakness and fiber-type disproportion. Neuromuscul Disord. 2016 Mar;26(3): 236–9.

Contribution of the author to the preparation of the original publications:
Paper I: Participation in the study design; collecting, analysing, and interpreting data; preparing figures and writing the manuscript.

Paper II: Participation in the study design; diagnostic evaluation of large gene panel sequencing of all cases; collecting, analysing, and interpreting summary data; statistical analysis; preparing figures and writing the manuscript.

Paper III: Collecting clinical data; reanalysis of chromosomal microarray analysis; bioinformatics and interpretation of whole-exome sequencing; preparing figures and writing the manuscript.

Paper IV: Collecting clinical data; participation in whole-exome sequencing interpretation and design of RNA studies; preparing some of the figures and writing the manuscript.

ABBREVIATIONS

aCGH	Array comparative genomic hybridization
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
AR	Autosomal recessive
ASD	Autism spectrum disorder
BAF	B-allele frequency
BAM	Binary version of a SAM file
bp	Base pair
BWA	Burrows-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
cDNA	complementary DNA
CFTD	Congenital fibre type disproportion
CMA	Chromosomal microarray analysis
CNV	Copy number variation
CT	Computed tomography
DD	Developmental delay
DGV	Database of Genomic Variants
DNA	Deoxyribonucleic acid
EEG	Electroencephalography
ExAC	Exome Aggregation Consortium
FISH	Fluorescence in situ hybridization
GATK	Genome Analysis Toolkit
gnomAD	The Genome Aggregation Database
HC	Haplotype Caller (a tool in GATK)
hg19	Human genome assembly version 19 (Feb 2009)
HGMD	Human Gene Mutation Database
HGVS	Human Genome Variation Society
HPO	Human Phenotype Ontology
IBD	Identity by descent
ID	Intellectual disability
IMPC	International Mouse Phenotyping Consortium
Indel	Insertion and/or deletion
IQ	Intelligence Quotient
kb	Kilobase (one thousand base pairs)
LCSH	Long contiguous stretch of homozygosity
LDM	Laing distal myopathy
LoF	Loss of function
Mb	Megabase (one million base pairs)
MCA	Multiple congenital anomalies
MLPA	Multiplex ligation-dependent probe amplification
MRI	Magnetic resonance imaging
mRNA	messenger RNA

MSM	Myosin storage myopathy
mTOR	Mechanistic target of rapamycin
NA	Not applicable
NGS	Next-generation sequencing
NIPT	Non-invasive prenatal testing
OFC	Occipitofrontal circumference
OMIM	Online Mendelian Inheritance in Man
PolyPhen	Polymorphism Phenotyping
PCR	Polymerase chain reaction
RPKM	Reads per thousand bases per million reads sequenced
SAM	Sequence alignment map
SD	Standard deviation
SIFT	Sorting Tolerant From Intolerant
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
UG	Unified Genotyper (a tool in GATK)
UPD	Uniparental disomy
VCF	Variant call format
VUS	Variant of unclear (clinical) significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
XL	X-linked

1. INTRODUCTION

Although the basic principles and contribution of inheritance in human diseases as well as other traits were previously elucidated by ancient physicians and scientists such as Hippocrates and Aristotle, the first modern principles of inheritance patterns were described by Gregor Mendel in the 19th century [Mendel 1866]. Sir Archibald Edward Garrod was the first to link a specific disorder to Mendelian inheritance pattern by correctly stating that alkaptonuria is a recessive disorder in 1902 [Garrod 1902]. It took another half a century before sufficient methods for analysing human chromosomes were developed and the correct number of chromosomes in human cells was identified as 46 [Tjio and Levan 1956]. The chromosomal aberration trisomy 21 (which includes an extra copy of chromosome 21) was discovered as the cause of Down syndrome in 1959 [Lejeune et al. 1959a; Lejeune et al. 1959b], making it the first genetic disease with known molecular aetiology. This led to the growth of clinical cytogenetics as a field in diagnostic medicine.

Methods used in molecular genetics began to evolve after the discovery of the structure of DNA in 1953 [Watson and Crick 1953]. Soon the central dogma of molecular biology was stated [Crick 1970; Crick 1958] and the genetic code deciphered [Nirenberg and Leder 1964]. The first gene associated with monogenic or Mendelian disorders was mapped to a specific locus in the human genome in 1983 when Huntington's disease was shown to be linked to a genetic marker on chromosome 4 [Gusella et al. 1983]. The exact mechanism of trinucleotide CAG-repeat expansion in the *HTT* gene took another ten years to be discovered [MacDonald et al. 1993].

Since the implementation of cytogenetic and molecular testing in clinical diagnostics, identification of genetic diagnosis in combination with proper patient counselling has been the main aim for clinical genetics services [Bowles Biesecker and Marteau 1999]. Identification of the specific aetiology of a patient's disorder allows appropriate genetic counselling for the family, and can be used for accurate risk predictions, prenatal diagnostics, estimating prognosis, and also for finding suitable treatment options [ACMG Board of Directors 2015]. Clinical genetics focuses on Mendelian disorders that represent a large group of diseases following either autosomal dominant (AD), autosomal recessive (AR) or X-linked (XL) inheritance patterns characterised by monogenic or monolocus causes and high penetrance [Antonarakis and Beckmann 2006]. Most Mendelian disorders can be classified as rare disorders, which are defined in the European Union as affecting less than one person per two thousand [Orphanet]. Most of the rare diseases listed in the Orphanet database are genetic in origin and according to the current knowledge and classification the total number of rare disorders is six to seven thousand, a large proportion of which the cause remains unknown.

In Estonia, chromosomal microarray analysis (CMA), genome-wide molecular assays that detect small DNA deletions and duplications in the submicros-

copic range (i.e., copy number variations [CNVs]), were first implemented in clinical practice in 2009. Since 2011, CMA has served as a first-tier diagnostic genetic testing method for patients with developmental delay (DD)/intellectual disability (ID), autism spectrum disorders (ASD) and/or multiple congenital anomalies (MCA). The use of CMA has been studied in Estonia by a few researchers. Dr Katrin Männik's PhD studies focused on the role of CNVs in patients with ID as well as in the general population [Männik 2012; Männik et al. 2011], whereas Dr Olga Žilina studied the diagnostic utility of CMA in the clinical setting in Estonia [Žilina 2014; Žilina et al. 2014a; Žilina et al. 2012; Žilina et al. 2014b]. Also, many case reports have been published based on the findings discovered by CMA [Leffler et al. 2016; Pajusalu et al. 2015a; Simenson et al. 2014; Vals et al. 2015; Õiglane-Šlik et al. 2014; Õunap et al. 2016]. The diagnostic role of long contiguous stretches of homozygosity (LCSHs), a frequent variant of unclear significance (VUS), has not been studied in Estonia before, and thus this study focuses partly on copy-number neutral LCSHs detected by CMA and aims to clarify their clinical utility in an outbred Estonian population.

Whole-exome sequencing (WES), which enables sequencing of all genes simultaneously, was first performed in a clinical setting for selected cases in 2013, but since 2014, the Estonian Health Insurance Fund began reimbursing WES for both proband-only and proband-parent trio approaches. Large gene panel or Mendeliome sequencing was introduced into clinical practice in 2015 in Estonia, and since then it has become one of the most commonly performed molecular genetic testing service at Tartu University Hospital. Several case reports, mainly resulting from Estonian research studies, of WES and next-generation sequencing (NGS) panel findings have been published [Maasalu et al. 2015; Reinson et al. 2016; Thompson et al. 2016; Vaher et al. 2014; Vals et al. 2014]. Due to reimbursement by the Estonian Health Insurance Fund that enables the use of NGS analyses in routine clinical diagnostics, it is important to evaluate the utility, diagnostic yield, and outcomes from the first years of practice. This study is the first to systematically evaluate the clinical utility of NGS investigations for diagnostics of Mendelian disorders in Estonia. In addition, reports on two interesting cases solved by WES are included in this study to illustrate crucial aspects of genome-wide diagnostics such as incorporating data from different genome-wide analyses and the necessity of functional experiments to achieve conclusive diagnosis.

2. LITERATURE REVIEW

2.1. Genetic analyses in diagnostics and discovery of Mendelian disorders

Since the early discoveries of gene-phenotype associations, novel genetic technologies and methods have facilitated and accelerated the discovery of new genetic disorders [Boycott et al. 2013]. The two early revolutionising methods enabling the discovery of DNA primary structure or the nucleotide sequences were Sanger sequencing [Sanger and Coulson 1975] and polymerase chain reaction (PCR) methods [Mullis et al. 1986], both still widely used in molecular diagnostic laboratories. The most transforming next step advancing the discovery of genes related to disease was the implementation of CMA and NGS into medical genetics research as well as diagnostics in the 2000s [Boycott et al. 2013]. Since the discovery of these methods, new gene-disease associations have been published at least weekly as of today and there are 3,733 genes with known phenotype-causing gene variants as well as 5,981 separate disease entities or phenotypes with known molecular basis in the OMIM database as of 22 April 2017.

Until recently, diagnostic efficiency in patients with suspected genetic disorders was very low, and mainly cases with clinically recognizable syndromes such as Down syndrome or Williams syndrome received confirmed diagnoses after genetic testing [Rauch et al. 2006]. Thus, after excluding Down syndrome cases, the diagnostic yield of conventional karyotyping for other subjects with DD/ID remained under 3% [Miller et al. 2010]. Fortunately, new genetic technologies like CMA that enable the detection of microdeletions and microduplications in the so-called submicroscopic range (i.e., smaller than 5 Mb in size) and NGS applications that are suitable for large scale mutation screening for both known and novel single nucleotide variants (SNVs) have been rapidly implemented in clinical diagnostics due to their vast potential in assisting diagnostics and patient care. For example, in the first large scale CMA study using first-tier CMA, potential pathogenic CNVs were detected in up to 25% of patients [Ahn et al. 2013]. Regarding NGS, the first proof-of-principle pilot study describing WES in 12 humans was published in 2009 [Ng et al. 2009], and then in 2011, Ambry Genetics, a commercial laboratory in the USA, launched clinical diagnostic WES services claiming to be the first in the world. Shortly the diagnostic yield of 25% was reported from routine clinical diagnostics [Yang et al. 2013]. In contrast to improved diagnostic yields, the rapid implementation of NGS testing has probably caused some problems [van El et al. 2013]. The lack of standardised algorithms for both laboratory experiments, bioinformatics, and interpretation have led to many different approaches on the reporting VUSs and the incidental findings between centres, as well as raised many other ethical and organizational concerns [van El et al. 2013].

Generally, genetic tests used in diagnostics of monogenic disorders can be divided into cytogenetic and molecular assays; the former detecting structural and copy-number variations in DNA molecules or chromosomes and the latter investigating aberrations within genes (Table 1). Both can be further divided by the resolution and scale of the tests [Katsanis and Katsanis 2013]. The scale ranges from targeted (only one or multiple loci being assessed) to genome-wide (whole genome scanned in one test). The choice of test for identifying the cause of the disease is dependent on the diagnostic hypothesis. For example, genotyping only one nucleotide is sufficient for diagnosing most cases of achondroplasia, the most common form of AD short-limb dwarfism [Rousseau et al. 1994; Shiang et al. 1994], whereas whole-exome studies are often needed for identification of disease causing variants in non-syndromic ID [Vissers et al. 2016].

Table 1. Cytogenetic and molecular DNA tests used in clinical diagnostics for Mendelian disorders.

Assay	Scale	Resolution	Primary mutation type targeted**
Cytogenetic tests			
Karyotyping	Genome-wide	>5 Mb	Aneuploidies, large structural variations
FISH	Targeted	>10 kb*	Microdeletions
CMA	Genome-wide	>100 kb (>1 kb)*	Microdeletions, microduplications
Molecular genetic tests			
PCR and restriction digest	Targeted	1 bp	SNVs, indels, CNVs*
Sanger sequencing	Targeted	1 bp	SNCs, indels
Repeat expansion assays	Targeted	1 repeat (3–4 bps)	Repeat expansions
Methylation-specific assays	Targeted	NA	Methylation profile
MLPA	Targeted	1 exon	CNVs
Mutation arrays	Wide-scale*	1 bp	Multiple SNVs, small indels
NGS panels	Wide-scale*	1 bp	SNVs, indels
WES	Genome-wide	1 bp	SNVs, indels

*Depends on a probe, microarray, etc., used.
**Other mutation types can be detected by some assays if a specific analysis is performed.

2.2. Chromosomal microarray analysis

2.2.1. Technology and types of chromosomal microarrays

Generally, there are two types of chromosomal microarrays widely used for detecting submicroscopic (i.e. smaller than 5Mb in size) chromosomal aberrations: a) array comparative genomic hybridisation (aCGH) and b) single nucleotide polymorphism (SNP)-arrays [Alkan et al. 2011; Emanuel and Saitta 2007]. aCGH makes use of two differently labelled genomic DNAs (a reference and a test sample) hybridized to a microarray [Pinkel and Albertson 2005]. Subsequently the signal ratios between reference and test samples are assessed for each measured probe and a copy number for each locus may be estimated (Figure 1a).

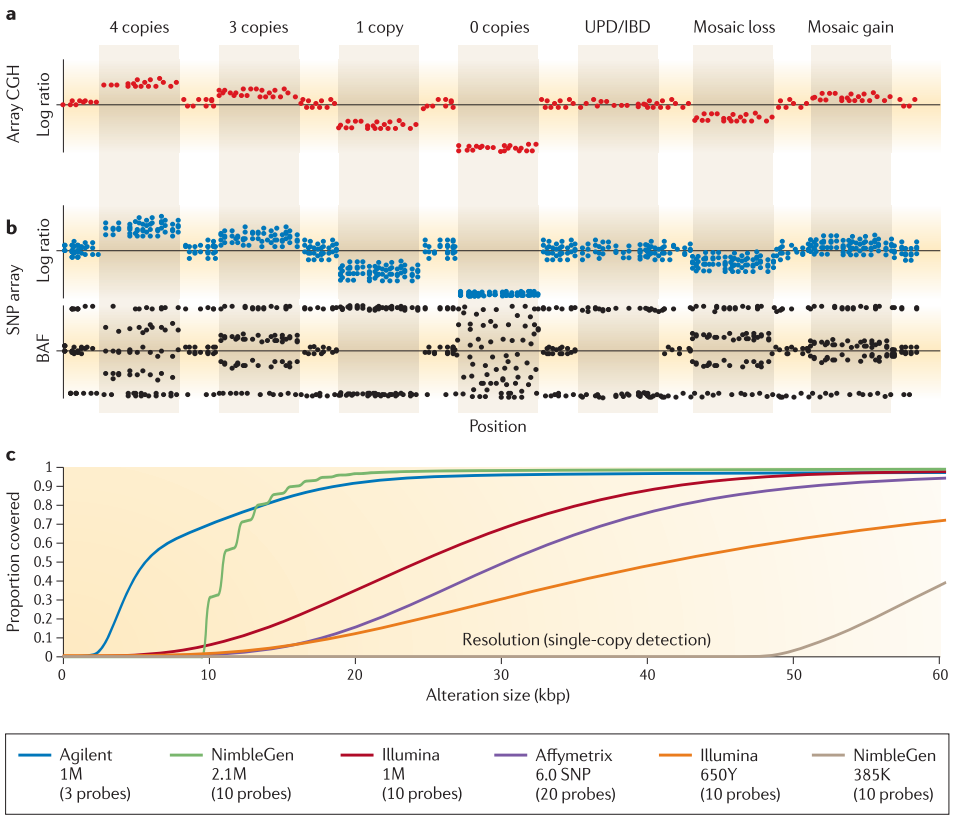


Figure 1. Array CGH versus SNP microarray detection. Parts A and B: visualization of aCGH and SNP-array outputs for different types of copy number alterations. Part C: comparison of different arrays based on their resolution and ability to detect CNVs of different sizes. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics 12(5), 363–376. Alkan, C., Coe, B. P., & Eichler, E. E. Genome structural variation discovery and genotyping., copyright 2011.

SNP-arrays enable simultaneous genotyping of hundreds of thousands to millions of SNPs, while two parameters are detected for each genotyped SNP: a signal intensity (Log ratio) and a B-allele frequency (BAF) (Figure 1b) [Alkan et al. 2011]. The signal intensity can be used to assess the copy number [Heinrichs and Look 2007]. BAF resembles a genotype and has a supporting role for copy number detection as deleted regions lose heterozygosity and duplicated regions have four possible genotypes instead of three as in disomic loci [Alkan et al. 2011]. The specific resolution and coverage is dependent on the probes used for aCGH design and the number as well as the content of SNPs selected for SNP-arrays (Figure 1c) [Alkan et al. 2011; Emanuel and Saitta 2007]. In Estonia, SNP-arrays have been used since the implementation of CMA into clinical diagnostics [Žilina et al. 2014b].

2.2.2. Clinical implementation and utility

The usage of CMA in clinical diagnostics is justified by the large proportion of disease causing CNVs that are smaller than 5–10 Mb, and thus undetectable by conventional karyotyping methods [Vissers et al. 2010]. These CNVs, 1 kb–5 Mb in size, are referred to as submicroscopic chromosomal rearrangements or as microdeletions and microduplications [Feuk et al. 2006; Rodriguez-Revenga et al. 2007]. Still, not all structural genomic variants can be detected by CMA. Thus, balanced, meaning copy-number neutral, chromosomal translocations and inversions as well as ring chromosomes and some other cytogenetic aberrations are still routinely detected by conventional karyotyping methods [South et al. 2013]. However, newer NGS-based methods can detect these types of variants [Redin et al. 2017; Vissers et al. 2010].

In many countries, including Estonia, CMA is the first-tier test used for clinical indications including but not strictly limited to DD/ID, ASD, and/or MCA [Ahn et al. 2013; Hochstenbach et al. 2009; Miller et al. 2010; Ozyilmaz et al. 2017; Žilina et al. 2014b; Vissers et al. 2010]. Besides postnatal cases, CMA has played an emerging role in prenatal diagnostics for high-risk pregnancies and foetal anomalies detected by ultrasound investigations [Oneda and Rauch 2017; Pons et al. 2017; Srebniak et al. 2017; Wapner et al. 2012]. Conventional karyotyping, however, has not been totally replaced in prenatal testing by CMA as of now [Oneda and Rauch 2017]. Non-invasive prenatal testing or NIPT, a technique which allows aneuploidy screening and in some cases also selected microdeletion testing non-invasively from maternal blood sample, has been widely implemented into prenatal genetic testing, but it has lower diagnostic yield than CMA due to missed microdeletions [Srebniak et al. 2017].

In most studies, the diagnostic yield of CMA in postnatal cases collected from routine clinical diagnostics ranges from 10–25%, depending on patient selection and classification criteria used for pathogenicity estimations of detected

CNVs [Ahn et al. 2013; Neill et al. 2010; Ozyilmaz et al. 2017; Žilina et al. 2014b]. To address the problem of nonuniformity in CNV interpretation criteria between diagnostic laboratories, the American College of Medical Genetics and Genomics (ACMG) has published guidelines for reporting CNVs in the clinical setting which are summarised in Table 2 [Kearney et al. 2011b]. The guidelines advocate for the classification of each CNV into one of the three main pathogenicity classes: 1) pathogenic, 2) uncertain clinical significance, and 3) benign. The class of uncertain clinical significance is further divided into three subclasses of 1) likely pathogenic, 2) likely benign, and 3) no subclassification [Kearney et al. 2011b].

Table 2. ACMG guidelines stating different aspects to be considered for clinical interpretation of CNVs. Adapted from [Kearney et al. 2011b].

Aspects of interpretation	Resources	Comments
Familiarity with well-known contiguous gene syndromes	OMIM, GeneReviews, DECIPHER	Well-known microdeletion/-duplication syndromes must be always mapped.
CNV size	Not applicable	Usually larger CNVs are more likely to be pathogenic, but very large benign CNVs exist.
Genomic content in CNV interval	OMIM	Check if CNVs encompass genes associated with known phenotypes that are caused by different types of mutations.
CNV frequency in databases	DGV, in-house databases	Important to assess clinical characterization of "normal" individuals.
Inheritance of the CNV	Not applicable	*De novo* state supports pathogenicity. If inherited, the carrier parent should be medically evaluated.

Based on their recurrence among both population-based cohorts of healthy individuals as well as affected patients, CNVs can be grossly divided into recurrent and non-recurrent categories [Lee et al. 2007]. Many recurrent CNVs emerge due to flanking segmental duplications that lead to increased mutation rates [Sharp et al. 2005]. A well-known example of recurrent CNV is microdeletion and microduplication of 16p11.2 region [Jacquemont et al. 2011; Zufferey et al. 2012; Walters et al. 2010]. Interestingly, microdeletion and microduplication of 16p11.2 show mirror phenotypes in body mass index [Jacquemont et al. 2011], but are also associated with many other clinical features [D'Angelo et al. 2016; Maillard et al. 2015; Shinawi et al. 2010]. The clinical interpretation of many recurrent CNVs is complicated by the variable expressivity and reduced penetrance for some phenotypes, which is illustrated by their presence in population-based cohorts [Männik et al. 2015]. In addition, recent studies have highlighted the role of maternally inherited CNVs as risk factors for neuro-

developmental disorders like autism in affected sons [Krumm et al. 2015]. In the prenatal setting, the lack of clarity for clinical significance of many CNVs detected by CMA brings even greater ethical and practical concerns, which makes proper genetic counselling essential [Oneda and Rauch 2017].

2.2.3. Long contiguous stretches of homozygosity

In addition to CNVs, copy-number neutral regions of homozygosity can be detected by CMA, but only if SNP-arrays are used [Alkan et al. 2011] (Figure 1b). These regions are often also described as areas, regions, or runs of homozygosity, but in this study, they are referred to as long contiguous stretches of homozygosity (LCSHs). Generally, LCSHs rarely cause disease on their own, with the exception of known uniparental disomy (UPD) syndromes, but they are still reported back to referring physicians as VUSs by some laboratories due to the possible role for recessive disorders [Žilina et al. 2014b; Wang et al. 2015]. The reporting policies regarding LCSHs vary between diagnostic centres and no internationally recognised guidelines exist. Still, 5 Mb is frequently used as a cut-off for LCSH reporting due to an observation that homozygous stretches shorter than 4 Mb are relatively common in outbred European populations [McQuillan et al. 2008]. The mechanism leading to the appearance of LCSHs can be either UPD, parental consanguinity or ancestral homozygosity [Kearney et al. 2011a]. If homozygosity is due to a chromosomal segment inherited from a common ancestor, it is referred to as identity by descent (IBD) [Wang et al. 2015].

In 1980, Professor Engel hypothesized that in rare instances both chromosomes could arise from only one parent, and proposed the term “uniparental disomy” or UPD to mark this cytogenetic abnormality [Engel 1980]. Later, UPD was first identified as the disease mechanism in a girl with cystic fibrosis by observing excessive homozygosity and lack of paternally inherited polymorphic markers on chromosome 7 [Spence et al. 1988]. In addition to loss of heterozygosity and thus increased risk for recessive disorders, another disease-causing mechanism of UPD is due to altered genomic imprinting, a phenomenon wherein gene expression is dependent on parental origin of a gene [Yamazawa et al. 2010]. It is important to note that UPD causes imprinting disorders only if it appears on certain chromosomes, namely 6, 7, 14, 15, 16, and 20 [Eggermann et al. 2015]. For example, maternal UPD of chromosome 15 causes Prader-Willi syndrome [Nicholls et al. 1989], whereas paternal UPD of chromosome 15 leads to Angelman syndrome [Nicholls et al. 1992]. Two types of UPD exist: in isodisomy, both chromosomes are identical to each other and represent copies of a single parental homologue, whereas in heterodisomy, both parental homologues are also present in an offspring [Yamazawa et al. 2010]. The mechanisms of UPD involve monosomy and trisomy rescues, gamete complementation, and post-fertilization mitotic errors [Engel 2006]. With respect to CMA, only isodisomic UPD regions can be detected from proband-

only analysis by SNP-arrays, whereas heterodisomy requires comparison of genotypes (BAFs) between a parent and an offspring [Conlin et al. 2010]. UPD should be suspected when there is one very long (>20 Mb) or multiple LCSHs restricted to one chromosome as isodisomic and heterodisomic segments can neighbour each other [Conlin et al. 2010; Wang et al. 2015]. If there is a suspicion of imprinting disorders but no isodisomic regions are detected, a specific test like methylation specific PCR or multiplex ligation-dependent probe amplification (MLPA) can be used in addition to parental SNP-array analysis followed by genotype comparison to exclude UPDs [Yamazawa et al. 2010].

Parental consanguinity is a well-known risk factor for recessive disorders [Bittles 2001]. SNP-arrays reveal parental consanguinity as multiple LCSHs located on multiple chromosomes with the total length exceeding 1.5% if the definition of parental consanguinity as "second cousin or closer" is used [Sund et al. 2013]. Moreover, the degree of consanguinity can be reliably assessed from SNP-array data by the proportion of the total length of LCSHs out of the total autosomal genome, i.e., 2800 Mb [Sund et al. 2013]. From a diagnostic perspective, identified or known parental consanguinity increases the chance of finding homozygous variants rather than compound heterozygous variants if sequencing investigations are carried out [Makrythanasis et al. 2014; Najmabadi et al. 2011]. Parental consanguinity is rare in Northern Europe including Estonia, which is also supported by the low number (4 out of 1191 patients) of multiple LCSHs detected among CMA samples in Estonia [Žilina et al. 2014b].

The clinical utility of LCSHs in cases where UPD and parental consanguinity is excluded, i.e., only one or two LCSHs detected with total length below 1% of the autosomal genome or 28 Mb, has been less studied. A study conducted in California (USA), found that 4% of patients with no reported CNVs had at least one LCSH over 5 Mb in size [Wang et al. 2015]. They also noted that five regions (two on the X-chromosome and another three on autosomes) were homozygous in multiple individuals and thus considered as polymorphic [Wang et al. 2015]. One can hypothesize that even a single LCSH can aid candidate gene identification, because recessive disease causing genes encompassed within LCSHs can harbour homozygous variants, just as in the case of multiple LCSHs in consanguineous families [Alkuraya 2010]. Wang et al. found nine individuals with a recessive candidate gene matching with the patient's phenotype, and in seven patients, pathogenic variants were detected by sequencing [Wang et al. 2015]. Homozygosity mapping, another method also derived from studies on consanguineous families, was used in a study in an outbred population of siblings with shared phenotypes and thus assumed recessive inheritance [Schuurs-Hoeijmakers et al. 2011]. They focused on recessive IDs, used a cut-off length of 1 Mb for shared homozygous regions, and concluded that the method could contribute to novel recessive ID gene discoveries [Schuurs-Hoeijmakers et al. 2011]. To aid in candidate gene identification, computational tools have been developed, e.g., Genomic Oligoarray and SNP array evaluation tool [Wierenga et al. 2013].

2.3. Next-generation sequencing

2.3.1. NGS technologies and applications

The term "next-generation sequencing" or NGS refers to many different high-throughput nucleotide sequencing technologies that differ from Sanger sequencing, which was virtually the only sequencing method for almost thirty years [Schuster 2008]. The first NGS methods, both published in 2005, described methods on how to massively parallelize sequencing reactions on either agarose thin layers or picotiter plates [Margulies et al. 2005; Shendure et al. 2005]. Since then, massively parallel sequencing has been served as a synonym for NGS more accurately describing the methodological difference from Sanger sequencing [Rogers and Venter 2005]. Currently used NGS technologies can be generally divided based on the length of single reads into short-read and long-read NGS, with the former being more widely applied, at least in human genetics [Goodwin et al. 2016]. One of the revolutionizing effects of NGS has been the tremendous drop in sequencing price over the last 15 years; the cost of sequencing one human genome was almost 100 million dollars in 2001, while one thousand dollars per genome was almost reached 15 years later by the end of 2015 [Wetterstrand 2016].

All sequencing experiments described in this study have been carried out using short-length sequencing technologies commercialised by Illumina (Illumina Inc., San Diego, CA, USA) and thus this technology is described in detail, although the methods are generalised when possible. Illumina technologies are currently the most widely used in sequencing studies including human DNA sequencing perhaps due to lowest per-base cost and highest throughput ability [Goodwin et al. 2016; Liu et al. 2012; van Dijk et al. 2014].

Generally, all DNA sequencing methods follow the following basic steps of library preparation that are required to be carried out before sequencing [van Dijk et al. 2014]:

1) Genomic DNA is extracted from the tissue of interest.
2) Long DNA molecules are fragmented into shorter molecules of desired length (50–500 nucleotides typically).
3) Adapters and indices are added to fragmented DNA molecules. For single-end sequencing, one sequencing primer is used, whereas for paired-end sequencing, two sequencing primers are used on both ends of the inserts. Oligonucleotide indices enable multiple samples to be sequenced during the same run, as demultiplexing can be easily performed afterwards using bio-informatic tools.
4) Some protocols need size-selection to eliminate inserts with too short or too long lengths as well as short free adapters.
5) Probes can be hybridized to enrich the library for desired targeted application if needed.
6) PCR is used for amplifying (enriching) the library.

In the case of Illumina sequencing, after the library has been prepared it can be transferred to a solid surface covered with adapter oligonucleotides binding to both ends of library inserts [van Dijk et al. 2014]. After another amplification step, the clusters containing approximately 1000 copies of single-stranded DNA are created to be sequenced [van Dijk et al. 2014].

The sequencing techniques used by Illumina platforms are based on basic principles similar to Sanger sequencing in which synthesis is terminated by blocking the ribose 3´-OH group and subsequent incorporation of the labelled deoxynucleotide into the synthesized polynucleotide is detected [Guo et al. 2008; Ju et al. 2006; Seo et al. 2005]. As termination of synthesis is reversible, there can be tens to hundreds of cycles involving termination of the synthesis followed by detection of the incorporated fluorescently labelled nucleotide, and thus single read-length is determined by the number of cycles performed [Goodwin et al. 2016]. There are both four-channel (e.g., HiSeq and MiSeq) and two-channel (e.g., NextSeq and MiniSeq) platforms produced by Illumina [Goodwin et al. 2016]. The possibility of using only two channels for detecting four possible combinations (red for C, green for T, red + green = yellow for A and no signal for G) has enabled desktop sequencers to efficiently shorten the time needed for sequencing runs [Goodwin et al. 2016; Neveling et al. 2016].

In human genomics, the most widely used NGS applications are for sequencing of targeted gene panels, whole exomes, or whole genomes [Sun et al. 2015]. Gene panel sequencing and WES rely on enrichment of targeted genomic regions [Hodges et al. 2007], and thus many different designs exist depending on the targets: from virtually few to all genes. Although whole genome sequencing (WGS) does not need an enrichment step during library preparation, the huge size of the genome makes WGS still many times more expensive than WES or gene panel sequencing [Sun et al. 2015]. Amplicon based methods, wherein targeted regions are PCR-amplified with specific primers rather than captured by hybridisation as in WES and large gene panel sequencing protocols, are another kind of very targeted NGS method used for applications where ultra-deep coverage is needed, e.g., somatic mutation testing in clinical oncology [Chang and Li 2013]. The choice of different application is guided by the diagnostic request and heterogeneity of tested disease; small targeted panels can be used when a reasonable number of genes are known to cause the phenotype (e.g., cardiomyopathy) [Akinrinade et al. 2015]. A genome-wide approach, however, is needed if the number of disorder-associated genes is quite large and novel disease gene discovery is likely, as is the case for IDs [Vissers et al. 2016].

2.3.2. Bioinformatics data processing

After raw sequencing reads are produced by the NGS platform, the data needs to be further processed before interpretation. Generally, for every NGS study the following bioinformatics processing steps must be carried out to generate adequate sets of detected variants ready for biological or clinical interpretation [Nielsen et al. 2011]:

1) Raw sequencing reads are mapped to the reference genome.
2) Mapped reads are sorted, duplicate reads removed, indels realigned, and base quality scores recalibrated.
3) Variants are called from pre-processed aligned reads.
4) Detected variants are filtered, genotypes refined, and genotype scores recalibrated.
5) The final set of detected variants are annotated with biological, and clinical information from population-based and clinical databases.

Currently, many different algorithms and bioinformatics tools exist for each data analysis step making the number of different possible pipelines (i.e., combinations of different data processing tools used subsequently usually in an automated fashion) almost infinite. Thus, multiple studies have highlighted the importance of evaluation and standardization of different pipelines [Brownstein et al. 2014; Hwang et al. 2015; Zook et al. 2014]. Moreover, the discordance between different pipelines is well known, and thus caution is needed when clinical NGS services are established [O'Rawe et al. 2013]. In different studies, the concordance of variant calling pipelines used on Illumina data have ranged from 57% to 99% [Cornish and Guda 2015; Laurie et al. 2016; O'Rawe et al. 2013]. However, as the newer versions of software have been developed, the concordance between different pipelines has increased [Hwang et al. 2015]. Generally, the concordance as well as sensitivity and specificity are higher for SNVs when compared to indels [Laurie et al. 2016].

Although *de novo* assembly is possible for human genomes and has the potential advantage in detecting genomic structural variants, the computational demand and non-suitability for targeted NGS applications like WES have reasoned for using a resequencing approach in clinical genomics [Li 2012]. Burrows-Wheeler Aligner (BWA) [Li and Durbin 2009], a widely used short-read aligner, was used for all experiments in the present study. BWA consists of three different algorithms (BWA-backtrack, BWA-SW, and BWA-MEM) that all make use of Burrows–Wheeler transform [Burrows and Wheeler 1994] and is meant for aligning short reads to large reference genomes like the human genome [Li and Durbin 2009]. Importantly, BWA is able to perform gapped alignment, supports paired-end sequencing, generates mapping quality metrics making, and outputs a file in SAM format, which makes it a powerful and convenient tool for the first step of bioinformatics processing of raw sequencing reads [Li and Durbin 2009].

After the reads have been aligned to a reference genome, a few steps need to be carried out to prepare the data file for the best quality variant identification. Best practice guidelines have been published for using the Genome Analysis Toolkit (GATK) [DePristo et al. 2011; Van der Auwera et al. 2013], which are widely accepted as the current gold standard in the field. First, the reads are sorted and SAM files are converted to binary BAM files to make the analysis faster [Van der Auwera et al. 2013]. Duplicate reads are marked to be ignored in subsequent steps, as they are likely to be produced from the same DNA molecule and thus do not add additional support for calling variants [Van der

Auwera et al. 2013]. This can be done by using Picard software package's tool MarkDuplicates, which compares 5' sequences of reads and marks lower quality reads with identical starting positions as duplicates. In addition, as variant calling depends on quality scores assigned to each base by sequencing platforms, these scores should be recalibrated (e.g., by using GATK base quality score recalibration pipeline) to address systematic technical errors, and thus achieve more accurate base quality scores [DePristo et al. 2011; Van der Auwera et al. 2013]. As indels are more difficult to detect and are prone to mapping discrepancies after alignment steps, the regions consisting probable indels can be realigned by tools like GATK IndelRealigner. However, as the current best practice tool for variant calling, GATK Haplotype Caller (HC), uses local haplotype reassembly for variant detection, indel realignment does not need to be performed beforehand.

Probably the most crucial step after read alignment is variant identification. Although there are specialised variant callers that only detect either SNVs or indels, most modern tools such as GATK HC and Unified Genotyper (UG) [McKenna et al. 2010], Platypus [Rimmer et al. 2014], and VarScan [Koboldt et al. 2009] can detect both simultaneously. While older tools such as GATK UG [McKenna et al. 2010] and samtools [Li et al. 2009] use simple read pileup for variant identification, the most current best practice tools like GATK HC and Platypus use local reassembly where regions of interest (i.e., regions containing a probable sequence alteration) are first identified and then local *de novo* assembly of the region identifies the most probable gene variant, thus improving variant calling accuracy especially for indels [Rimmer et al. 2014]. Typically, between 20,000–50,000 variants are identified using WES, depending on both laboratory protocols and bioinformatics pipelines [Gilissen et al. 2012]. The combination of BWA alignment and GATK HC variant calling results in a very high sensitivity and specificity for detecting SNVs (both >99.5%) for both WES and WGS experiments [Laurie et al. 2016]. Short indels were detected with a sensitivity and specificity of 98.5% using WGS; however for WES, the sensitivity was 96% for both deletions and insertions and the specificity was 72% and 87%, respectively, in the same study [Laurie et al. 2016]. Detection of other classes of variants such as CNVs and long indels is reviewed in Section 2.4.

After the variants are called, the genotypes can be filtered based on their quality estimates produced by the variant caller [Nielsen et al. 2011]. Also, adding pedigree information (familial prior probabilities) and known population genetic variance (population prior probabilities) can aid in genotype refining by calculating posterior genotype probabilities [Kojima et al. 2013; Van der Auwera et al. 2013]. Finally, the variant callset is annotated, which is another crucial step to be able to assess the molecular as well as clinical significance of variants [McCarthy et al. 2014; Salgado et al. 2016]. A few examples of widely used variant annotators are Annovar [Wang et al. 2010], snpEFF [Cingolani et al. 2012b], and Variant Effect Predictor [McLaren et al. 2016]. Importantly, discrepancies between variant annotators are well known, with the largest differences laying among splicing variants [McCarthy et al. 2014; Salgado et al.

2016]. Generally, annotations are either variant or gene level information associated with detected variant [Salgado et al. 2016]. Variant-level annotations may include:

- naming the variant according to HGVS nomenclature including a description of the change in nucleotide as well as amino acid;
- variant frequencies in population databases like ExAC [Lek et al. 2016] or 1000 genomes project [1000 Genomes Project Consortium et al. 2015], and in-house database;
- annotations from pathogenic variant databases like ClinVar or HGMD [Stenson et al. 2009];
- *in-silico* pathogenicity predictions like SIFT [Kumar et al. 2009], PolyPhen [Adzhubei et al. 2010] or CADD [Kircher et al. 2014];
- evolutionary conservation scores like PhyloP [Pollard et al. 2010].

In addition, gene-based annotations that link the gene to known disease (e.g., OMIM diseases), phenotypic features (e.g., Human Phenotype Ontology (HPO) terms [Kohler et al. 2017]) or biological information (e.g., gene ontology terms [Ashburner et al. 2000; The Gene Ontology Consortium 2015]) can be added to guide diagnostic interpretation of variants.

2.3.3. Clinical utility of NGS applications

The diagnostic yield of NGS has been of great interest since the introduction of NGS into clinical diagnostics, and many reports have been published on its clinical utility. The first report of diagnosing a genetic disease by WES was an article published in 2009 describing a patient with suspected Bartter syndrome caused by homozygous variant in *SLC26A3* detected by WES [Choi et al. 2009]. During the following years, most reports were on research studies focusing on particular syndromes or phenotypes and applying WES to associate new genes with disorders [Bilguvar et al. 2010; Gilissen et al. 2010; Ng et al. 2010a; Ng et al. 2010b].

In 2012, however, the first articles were published on groups of patients not selected for common phenotypes, but based on the clinical diagnostic setting. A group from Duke University School of Medicine reported six out of twelve patients receiving molecular confirmation to the diagnosis after genetic disorders were suspected, but with no shared phenotypes [Need et al. 2012]. Two large ID WES cohorts were also published in 2012. First, a study by Rauch et al. highlighted the role of *de novo* mutations as a cause of ID when they reported 16 cases out of 51 (31.4%) carried a *de novo* mutation in a known ID-gene and an additional 6 (11.8%) had *de novo* loss-of-function (LoF) variants in strong candidate genes [Rauch et al. 2012]. Second, a similar study by de Ligt et al. identified a molecular cause for 16 ID patients out of 100, and in addition, 22 patients were identified as carrying a strong candidate variants [de Ligt et al. 2012]. This study, similar to Rauch et al., highlighted the role of *de novo* mutations as a prominent cause for ID discovered in a clear majority of solved

cases after using parents-offspring trio approach in both cohorts [de Ligt et al. 2012; Rauch et al. 2012].

Since the first research projects, WES has been demonstrated as a valuable diagnostic tool as well [Ku et al. 2012]. The first study describing the clinical utility of WES as a routine test in non-selected patients was published only at the end of 2013 [Yang et al. 2013]. This study reported on 250 patients in whom WES was performed as a proband only approach and molecular diagnosis was made in 62 (25%) [Yang et al. 2013]. Like in the previous ID cohorts, they also noted a high percentage of *de novo* mutations, which were confirmed by conventional sequencing of parental samples after WES [Yang et al. 2013]. Since then, many diagnostic WES cohorts have been published (see Table 3 for summary). Notably, all comparable studies in large cohorts with unselected patients and not focusing on any disease group have reported very similar diagnostic yields ranging from 25% to 31%. Most of the diagnostic findings in patients without parental consanguinity are due to dominant heterozygous variants appearing *de novo* [Farwell et al. 2015; Lee et al. 2014; Yang et al. 2014]. In the consanguineous population, however, AR disorders are due to predominating homozygous variants [Trujillano et al. 2017]. Studies comparing the diagnostic yields between trio-sequencing and proband-only approaches generally show the advantage of trio sequencing due to discovered *de novo* mutations [Farwell et al. 2015; Lee et al. 2014; Retterer et al. 2016]. In these studies, trio exome sequencing had a diagnostic rate of 31–41% [Farwell et al. 2015; Lee et al. 2014; Retterer et al. 2016].

Table 3. Summary of large WES studies reporting on diagnostic yield in clinical setting in non-selected patients with variable indications for testing.

Study	Number of index patients	Diagnosed cases	Diagnostic yield	Comments
Yang et al. [2013]	250	62	25%	Proband-only approach
Farwell et al. [2015]	500	152	30%	Both trios and proband-only cases, trios showing higher yield.
Lee et al. [2014]	814	213	26%	Both trios and proband-only cases, trios showing higher yield.
Trujillano et al. [2017]	1000	307	31%	Mostly trios, 45.3% of the cases had parental consanguinity
Yang et al. [2014]	2000	504	25%	Proband-only approach
Retterer et al. [2016]	3040	876	29%	Proband-only, duo, trio and other designs, trios showing higher yield.

Many other diagnostic WES cohorts focusing on single disease groups have been published as well. For example, the diagnostic yield of WES was 51% in 266 Dutch patients with visual impairments [Haer-Wigman et al. 2017] and 33.5% in 200 patients with hearing impairments [Zazo Seco et al. 2017]. Other interesting comparisons have been made in the literature as well. For example, a study revealed a significantly higher diagnostic yield in patients having epilepsy than in those without epilepsy [Helbig et al. 2016].

Other NGS applications have not been studied as frequently in the clinical setting. For example, although used in many clinics, the diagnostic utility of large (i.e., covering thousands of genes) panels or Mendeliome sequencing has been less investigated. In a large study describing the use of custom large gene panels in patients from a highly consanguineous population, the diagnostic yield was 43% [Saudi Mendeliome Group 2015]. Also, different gene panels have been used efficiently in cohorts selected for different disease groups [Akinrinade et al. 2015; Ellingford et al. 2016; Poninska et al. 2016; Vega et al. 2016]. WGS has not been widely implemented into routine clinical practise yet because it is still more expensive than WES. Even more importantly, WGS is computationally very laborious and non-coding variants are mostly impossible to interpret in clinical settings due to their unpredictable effects [Sawyer et al. 2016]. A few studies, however, have been published demonstrating the benefit of WGS over WES. For example, 42% additional diagnostic yield was attributed to WGS in a cohort of severe ID due to the ability of WGS to detect the comprehensive spectrum of DNA variations including CNVs [Gilissen et al. 2014]. Another study estimated that 15% of variants discovered by WGS would have been missed by WES [Taylor et al. 2015]. There are no large diagnostic WGS cohorts reported in the literature as of the time of writing this thesis.

An additional aspect of diagnostic efficacy is the comparison of the clinical utility of NGS over traditional genetic tests. A pilot post hoc study demonstrated the significantly increased yield of WES compared to Sanger sequencing of single genes in patients with blindness, hearing impairments, mitochondrial disorders, and movement disorders [Neveling et al. 2013]. Another study showed a diagnostic yield of 29% for WES in children with rare disorders, whereas most remained unsolved after extensive screening by traditional methods involving single gene sequencing [Sawyer et al. 2016]. The authors concluded that the main reasons for patients being undiagnosed before WES were genetic heterogeneity and atypical presentations of underlying disorders [Sawyer et al. 2016]. Recently, a study was published assessing the clinical utility of WES versus conventional testing in paediatric patients with complex neurological disorders by performing both the standard diagnostic workup (including brain magnetic resonance imaging [MRI], muscle biopsies, and sequential single gene testing) and WES in parallel [Vissers et al. 2017]. They showed that a significantly increased proportion of conclusive diagnoses were obtained using WES compared to traditional methods (29.3% vs 7.3%) [Vissers et al. 2017]. Notably, the cost of the WES pipeline was not more expensive and may even reduce health-care costs at least for some cases [Vissers et al. 2017]. A different study

with a similar design (parallel use of standard and WES workup) carried out in 80 infants with suspected monogenic disorders revealed a diagnostic yield of 57.5% for singleton WES versus 13.75% for standard investigations [Stark et al. 2016]. They highlighted that in one third of the WES-diagnosed cases, clinical management was changed after the diagnosis.

No clear diagnostic guidelines have been published on using WES in clinical practice. One proposed position for NGS in the context of clinical genetics workup would be a second-tier test if more specific tests driven by clinical presentation remain negative or for nonspecific phenotypes, even a first-tier test along with CMA, depending on the phenotype (Figure 2) [Shashi et al. 2014]. The selection of whether to start testing by CMA or NGS should also be considered by taking the analysis of cost and technical details (ability to call CNVs from NGS data) into account [Shashi et al. 2014].

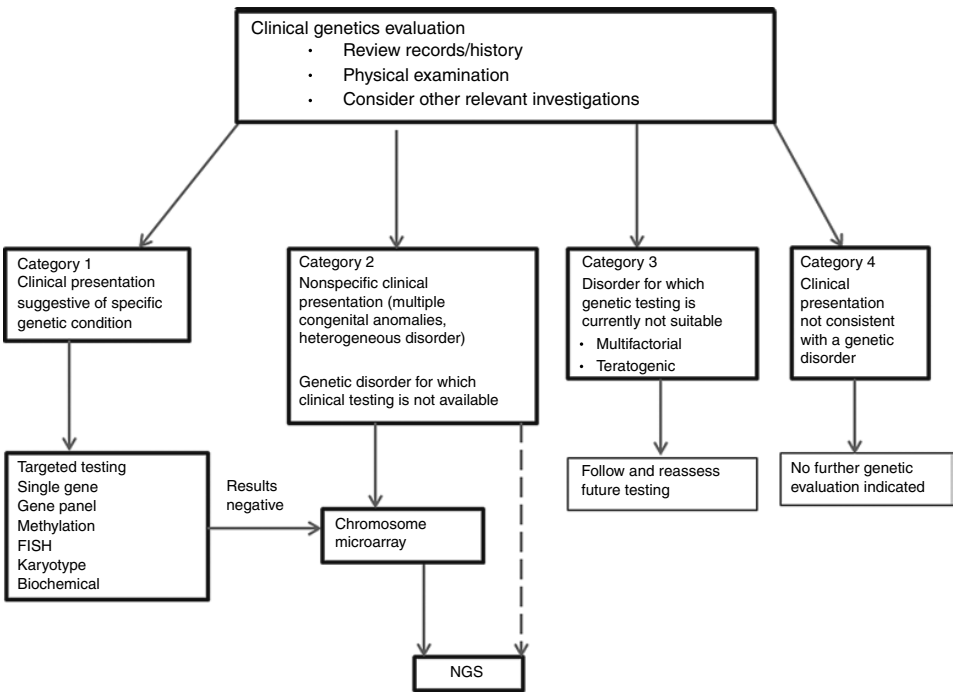


Figure 2. Algorithm to identify patients in a general genetics clinic most likely to benefit from NGS. Reprinted by permission from Macmillan Publishers Ltd: Genetics in Medicine 12(5), 363–376. Shashi V. et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders, copyright 2014.

The ACMG has also listed indications when to consider WES/WGS in the diagnostic setting [ACMG Board of Directors 2012]:

- The phenotype or family history data strongly implicate a genetic aetiology, but the phenotype does not correspond with a specific disorder for which a genetic test targeting a specific gene is available on a clinical basis.

- A patient presents with a defined genetic disorder that demonstrates a high degree of genetic heterogeneity, making WES or WGS analysis of multiple genes simultaneously a more practical approach.
- A patient presents with a likely genetic disorder, but specific genetic tests available for that phenotype have failed to arrive at a diagnosis.
- A fetus with a likely genetic disorder in which specific genetic tests, including targeted sequencing tests, available for that phenotype have failed to arrive at a diagnosis.

2.4. Mutation types: detection, classifications and relevance in clinical diagnostics

DNA variants can be divided according to their size into the following categories: SNVs, indels, CNVs, and large chromosomal aberrations (including aneuploidies) (Figure 3). Although different cut-off lengths have been proposed to differentiate between these classes, usually indels are considered to be 1 to 100 bp deletions and duplications, and CNVs represent the size range from 100 bp to 3 Mb [Zhang et al. 2009]. Also a cut-off from 50 bp [Sudmant et al. 2015] to 10 kb [Mills et al. 2006] has been suggested to separate indels from CNVs. Due to the wide spread use of targeted sequencing applications such as WES and NGS panels where exons are considered as targets for enrichment, an exon rather than a certain length in bps is considered as the measure for size for a CNV [Johansson et al. 2016; Krumm et al. 2012]. Thus, in the context of targeted resequencing studies, a simplification can be made as follows: SNVs represent substitutions of single nucleotides, indels are deletions and duplications smaller than one exon, and finally, CNVs can be defined as deletions and duplications from a single exon to multiple genes. Other forms of structural variations such as inversions, translocations, mobile element insertions, repeat expansions, and aneuploidies are all important components of genomic variations and are known to be associated with genetic disorders [Weischenfeldt et al. 2013], but are not the focus of this study.

SNVs as well as short indels representing the most prominent types of DNA variations causing Mendelian disorders and identified by NGS applications are detected by routine bioinformatics pipelines for NGS variant calling with high sensitivity and specificity as described in Section 2.3.2. Indels up to 30 bps in size can be detected by routine pipelines, but longer indels have to be called by specific variant callers [Marschall et al. 2013]. Tools developed for indel calling like Pindel, which uses a pattern growth algorithm [Ye et al. 2009], have superior sensitivity for indel detection compared to more widely used callers such as GATK HC and UG [Ghoneim et al. 2014; Marschall et al. 2013]. Currently, variant callers claim to be more universal and detect multiple mutation types (e.g., Platypus) [Rimmer et al. 2014]. Indel calling using WES and other enrichment based NGS applications is complicated by the fragmented nature of the data, and the non-uniform distribution of the reads [Karakoc et al. 2011].

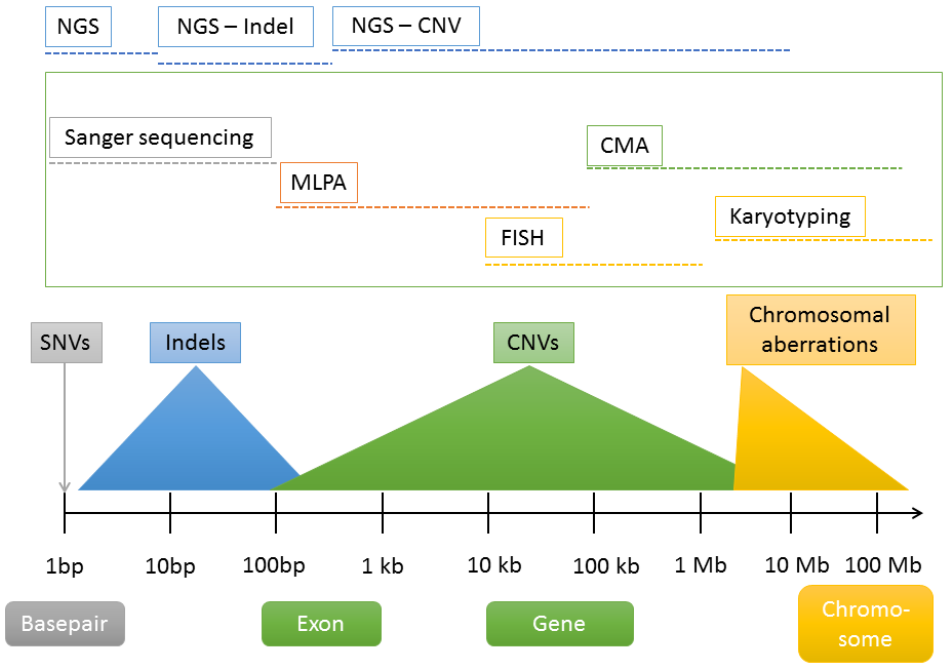


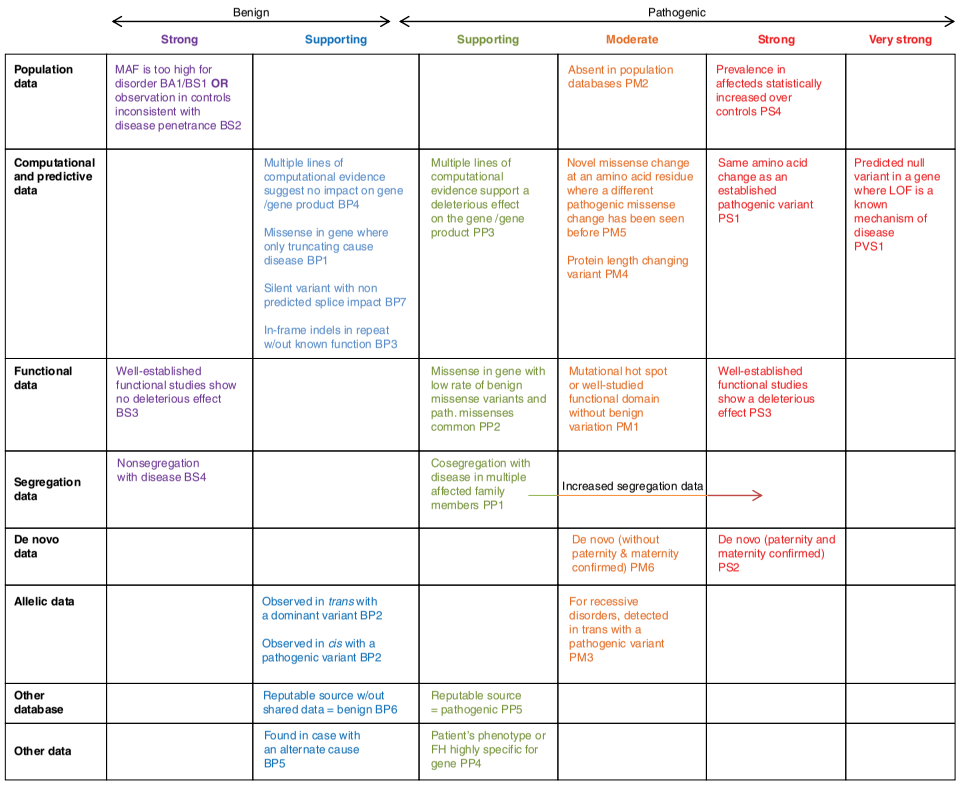
Figure 3. Classes of DNA variation based on size in basepairs and detection methods. Conventional diagnostic methods are grouped inside the green box whereas different NGS methods are presented at the top.

Studies assessing the clinical utility of using specialised indel callers for targeted NGS investigations in the diagnostic setting have not been published at the time of writing this thesis. Long indels are thought to be an under-represented group of genetic variations in many studies, and thus their contribution to human disease remains unknown [Hehir-Kwa et al. 2016]. We have demonstrated the ability to increase the diagnostic yield by approximately 1% in patients with ID by using specialized indel callers to detect variants in the size range of 20–200 bps [Pajusalu et al. submitted].

Regarding CNV calling, most widespread methods use read-depth analysis in which normalized per target (exon) read depths are compared to a panel of reference samples, and CNVs are detected by deviations from the average; higher read-depths indicate duplication and lower read-depths indicate deletions [Hehir-Kwa et al. 2015]. A tool called CoNIFER makes use of this approach and claims to detect CNVs consisting of at least three flanking exons [Krumm et al. 2012] whereas another tool called CoNVaDING [Johansson et al. 2016] enables even single exon deletion detection. In a recent large-scale (2,603 patients) diagnostic WES study, an added yield of 2% was reported for read-depth CNV screening [Pfundt et al. 2016]. Importantly, different disease groups show variance in detected number of causal CNVs; the CNV-associated diagnostic yield was highest for patients with hearing impairments (5.8%), complex

phenotypes (5.5%), and renal disorders (3.6%); however, no causal CNVs were detected in patients with sexual development disorders, craniofacial anomalies, metabolic disorders, or hereditary cancers [Pfundt et al. 2016]. In addition to read-depth based methods, other CNV calling algorithms have also been developed using one or combinations of the following principles: insert size abnormalities, split-read alignments, and *de novo* assembly [Marschall et al. 2013].

Another important classification of variants is the distinction between benign variations and pathogenic mutations. Although "mutation" should be used as a term for a permanent nucleotide change and a "polymorphism" is defined as a genetic variant appearing in at least 1% of the population, in medical literature, the terms are often used to describe pathogenic and benign variations, respectively [Richards et al. 2015]. To address this confusion, ACMG guidelines recommend using the term "variant" instead of both "mutation" and "polymorphism", with the following modifiers to be added for clarifying pathogenicity: (1) benign, (2) likely benign, (3) uncertain significance, (4) likely pathogenic, or (5) pathogenic [Richards et al. 2015]. The ACMG guidelines for variant interpretation consist of multiple criteria and rules for combining the criteria result in a variant classification scheme (Figure 4) [Richards et al. 2015]. The most important criteria that can be used for classifying most of each individual's genetic variation as benign is variant frequency in large population databases like ExAC [Lek et al. 2016]. Importantly, a newer database called gnomAD, which includes 123,136 exomes and 15,496 genomes from unrelated individuals, incorporates WGS data from more than two thousand Estonians that participated in a biobank at the Estonian Genome Centre at the University of Tartu, providing an important reference for variant interpretation. It must be highlighted that the pathogenicity or damaging effect of the variant does not always mean causality for the phenotype, which is especially important regarding the genes in which very little biological or clinical information is available [Richards et al. 2015]. Importantly, after adequate classification, only pathogenic and likely pathogenic variants should be used for clinical decision making [Richards et al. 2015].



	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Figure 4. ACMG guidelines for variant filtration: evidence framework and rules for combining criteria to classify sequence variants. Reprinted by permission from Macmillan Publishers Ltd: Genetics in Medicine 17(5), 405–423. Richards S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology, copyright 2015.

2.5. *KPTN* gene and its role in neurodevelopmental disorders

Kaptin, a protein encoded by the *KPTN* gene (OMIM 615620) located in chromosomal band 19q13.32, was first discovered as an actin-associated protein. It was identified by monoclonal antibody 2E4 originally raised against human platelet protein that eluted from F-actin with ATP [Bearer 1992]. The original study demonstrated that after stimulating undifferentiated cells with nerve growth factor, the 2E4 antibody stained mainly the neurite and growth cone, thus highlighting the role of *KPTN* in neurodevelopment [Bearer 1992]. In 1999, the *KPTN* gene was cloned [Bearer and Abraham 1999]. The discovery of kaptin being present at the tips of the elongating stereocilium, and thus having a probable role in stereocilia formation in addition to platelet activation [Bearer and Abraham 1999], led to the hypothesis that variants in *KPTN* could be associated with hearing loss [Bearer et al. 2000]. The *KPTN* gene was both functionally and positionally a great candidate gene for non-syndromic AD deafness (DFNA4 locus) [Bearer et al. 2000], supported by a report on a patient with a 0.8 Mb *de novo* deletion encompassing *KPTN* and having hearing impairment [Leal et al. 2009]. It was, however, only later found that the real causative genes for AD hearing loss at the 19q13 locus were *MYH14* [Donaudy et al. 2004] and *CEACAM16* [Zheng et al. 2011].

The first real patient-derived evidence of a disease-causing role for *KPTN* variants was published by Baple et al. only in 2014 when they described a large Amish pedigree where biallelic *KPTN* variants segregated with a syndrome of neurodevelopmental delay, macrocephaly, and seizures [Baple et al. 2014]. They identified altogether nine individuals from four related families carrying either homozygous nonsense variant c.776C>A p.Ser259* in four individuals or compound heterozygous variants with the same p.Ser259* variant and an additional 18-bp duplication (c.714_731dup p.Met241_Gln246dup) in five individuals [Baple et al. 2014]. The recurrent features in individuals with *KPTN*-related disorder (OMIM 615637, Mental retardation, AR 41) were mild to severe ID, increased occipitofrontal circumference (2.1–5.4 SDs), expressive and receptive language deficit, childhood hypotonia, seizures, and behavioural disturbances (stereotypies, repetitive speech, and anxiety) [Baple et al. 2014]. There were no characteristic dysmorphic features present in affected family members other than macrocephaly-associated frontal bossing and prominent chin [Baple et al. 2014]. Despite the small cohort, the authors suspected that nonsense-mutation homozygotes were more severely affected than those carrying in-frame duplication on one allele due to demonstrated retained in-vivo functionality of p.Met241_Gln246dup variant [Baple et al. 2014]. Besides the study by Baple et al. no more patients were reported in the medical literature with *KPTN*-related disorder prior to this study making it a very novel disease gene.

In addition to the clinical characterization of a novel AR ID syndrome, the authors performed functional studies to characterize both the physiological role of kaptin in neurons as well as the molecular and cellular effects of mutations

identified in affected individuals [Baple et al. 2014]. Kaptin was shown to be enriched in neuronal growth cones as well as discrete cortical sites during early development [Baple et al. 2014]. In mature neurons, kaptin was found at postsynapses as well [Baple et al. 2014]. The locations where kaptin is enriched were also high in F-actin content, which is expected due to the prior discovery of the association between kaptin and actin [Baple et al. 2014; Bearer 1992]. The p.Ser259* was demonstrated expectedly to result in degradation of the mutated transcript due to nonsense mediated decay, whereas p.Met241_Gln246dup was predicted to disrupt α-helix 4, but retained limited functionality *in vivo* [Baple et al. 2014].

As a part of the International Mouse Phenotyping Consortium (IMPC), a *Kptn* knock-out mouse model was generated and phenotypically characterized. *Kptn*-deficient mice showed increased body weight on a high-fat diet and normal hearing [White et al. 2013]. The IMPC web browser also lists other phenotypic alterations in addition to increased body weight and total body fat amount including: abnormal behaviour, increased circulating total protein and serum albumin levels, and increased circulating alkaline phosphatase and calcium levels. Furthermore, *Kptn*-knockout mice, similar to humans with *KPTN*-deficiency, show macrocephaly (personal communication, unpublished data).

Very recently, a study published in Nature demonstrated that kaptin is part of a larger protein complex that the authors called KICSTOR containing also ITFG2, C12orf66, and SZT2 [Wolfson et al. 2017]. The KICSTOR complex was shown to be a lysosome-associated negative regulator of mechanistic target of rapamycin complex 1 (mTORC1) signalling [Wolfson et al. 2017]. Thus, LoF variants in KICSTOR components may have similar consequences with mTOR activating mutations such as LoFs in other negative regulators (e.g., tuberous sclerosis complex genes *TSC1* and *TSC2*) [Laplante and Sabatini 2012]. Similar to *KPTN*-related syndrome, many mTOR activating disorders like tuberous sclerosis, neurofibromatosis, Noonan syndrome, and other rasopathies present clinical symptoms such as absolute or relative macrocephaly as well as increased risk for seizures and neurodevelopmental problems [Winden et al. 2015]. Importantly, the discovered link between *KPTN* and the mTOR pathway might facilitate the identification of novel treatment options for patients with *KPTN*-related disorders as drugs such as rapamycin are known to inhibit mTOR signalling [Wolfson et al. 2017].

2.6. *MYH7*-related disorders: genotype-phenotype associations

The *MYH7* gene (OMIM 160760), located in the chromosomal region 14q11.2, encodes a slow/beta heavy chain of myosin expressed in slow (type 1) muscle fibres and in the ventricles of the heart [Fiorillo et al. 2016]. In the OMIM database, *MYH7* pathogenic variants are associated with seven distinct disorders mostly following AD inheritance patterns (Table 4). In addition, some patients

carrying *MYH7* variants and suffering from muscle disorders reported in the literature do not fall into any OMIM-listed disorders, having instead muscle and/or clinical phenotypes characteristic of limb-girdle syndromes [Muelas et al. 2010] or congenital fibre type disproportion (CFTD) [Ortolano et al. 2011]. Still, Laing distal myopathy (LDM) and myosin storage myopathy (MSM) represent the main *MYH7*-associated skeletal muscle phenotypes, with the main difference being the pattern of affected muscles; LDM affects primarily distal muscles [Laing et al. 1995], whereas MSM as well as CFTD cause proximal muscle weakness [Ortolano et al. 2011; Tajsharghi and Oldfors 2013]. Regarding cardiomyopathies, heterozygous mostly missense *MYH7* mutations are the most prevalent cause of hypertrophic cardiomyopathy besides *MYBPC3* pathogenic variants [Maron et al. 2012]. In dilated cardiomyopathy patients, the genetic mutation profile is more heterogeneous, but *MYH7* variants are well documented, although not the most frequent cause of dilated cardiomyopathy, as well [McNally et al. 2013].

Table 4. *MYH7*-related disorders in the OMIM database.

Disorder	**Phenotype MIM number**	**Inheritance**
Cardiomyopathy, dilated, 1S	613426	AD
Cardiomyopathy, hypertrophic, 1	192600	AD
Laing distal myopathy	160500	AD
Left ventricular noncompaction 5	613426	AD
Myopathy, myosin storage, autosomal dominant	608358	AD
Myopathy, myosin storage, autosomal recessive	255160	AR
Scapuloperoneal syndrome, myopathic type	181430	AD

Importantly, a well described genotype-phenotype association exists between *MYH7* variants and associated phenotypes. In cardiomyopathy, significantly more variants are identified in the head and neck regions of the gene compared to the tail region, although generally cardiomyopathy causing variants are spread over the entire *MYH7* gene [Walsh et al. 2010]. In skeletal muscle disorders, however, disease causing variants almost exclusively reside in the tail region, with LDM-causing variants appearing in exons 32–36 and MSM-causing variants in very distal exons 37–40 [Udd 2009] (Figure 5). Although the first reports concluded that the phenotype of skeletal and cardiac muscle disorders do not overlap, there have been reports of patients having both skeletal and cardiac muscle involvement [Fiorillo et al. 2016].

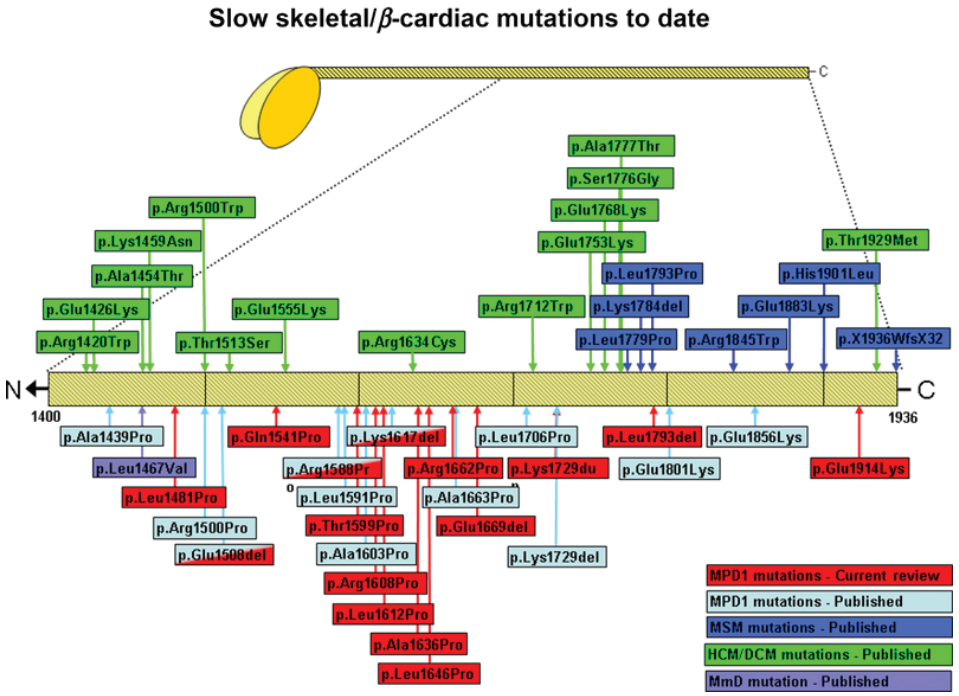


Figure 5. A schematic of the *MYH7* gene, indicating the position of gene mutations causing hypertrophic or dilated cardiomyopathy, myosin storage myopathy or Laing distal myopathy. HCM – hypertrophic cardiomyopathy, DCM – dilated cardiomyopathy, MSM – myosin storage myopathy, MPD1 – Laing early onset distal myopathy, MmD – multiminicore myopathy. Reprinted by permission from John Wiley and Sons: Human Mutation 35(7):868–879. Lamont PJ et al. Novel mutations widen the phenotypic spectrum of slow skeletal/beta-cardiac myosin (*MYH7*) distal myopathy. copyright 2014.

Most of the *MYH7* disease causing variants are missense and in-frame deletions, with both *de novo* and inherited mutations being described [Lamont et al. 2014]. LoF variants, however, appear to be non-pathogenic due to the relative high frequency in population databases. For example, combined allele frequency for LoF variants in the ExAC database is 1:3000. Moreover, based on the ExAC data, *MYH7* is under missense constraint (z = 6.54), but not under LoF constraint (probability of LoF intolerance or pLI = 0) [Lek et al. 2016]. Nevertheless, *MYH7* LoF variants have been reported to cause cardiomyopathies [Waldmuller et al. 2011]. Interestingly, multiple variants that change amino acid residues into prolines have been reported [Lamont et al. 2014]. Another recurrent missense variant is a charge changing substitution of glutamate to lysine, which is associated with mixed skeletal and cardiac muscle involvement [Udd 2009]. Additionally, two stop-codon altering mutations have been described [Fiorillo et al. 2016; Ortolano et al. 2011]. After we published a report on the first exon-skipping mutation [Pajusalu et al. 2016], other similar variants have been

detected [Fiorillo et al. 2016]. Remarkably, even in one family, the clinical phenotype caused by the *MYH7* variant may differ [Fiorillo et al. 2016]. Supporting this phenomenon, it has been also reported that muscle histology may differ at different ages, namely in a family with congenital fibre type disproportion, myosin storage was seen only in the oldest affected family member [Ortolano et al. 2011].

2.7. Summary of literature review

Genome-wide diagnostics of Mendelian disorders has been rapidly evolving since the implementation of CMAs and NGS-based analyses into routine clinical practice during the last decade. This has also accelerated the discovery of new genetic disorders and genotype-phenotype correlations. Nevertheless, many specific fields in modern genetic diagnostics remain understudied. For example, although very frequently reported after diagnostic CMA, the clinical value of detected LCSH has not been sufficiently investigated. Regarding NGS-based analyses, there have been many clinical WES cohorts described, but the utility of other NGS applications (e.g. Mendeliome sequencing) has been much less investigated. Although the clinical efficiency of genome-wide assays can be investigated only by studying large cohorts, some cases which may emerge from larger research studies or clinical diagnostics can provide specific novel information about rare disorders and thus remain of scientific interest.

3. AIMS OF THE PRESENT STUDY

1) To assess the diagnostic utility of single long contiguous stretches of homozygosity in patients without parental consanguinity in Estonia (Paper I).
2) To assess the diagnostic utility of large gene panel sequencing in the clinical diagnostic setting (Paper II).
3) To characterize and molecularly specify *KPTN*-related ID syndrome in two adult siblings (Paper III).
4) To investigate the molecular and clinical effect of a novel *de novo* mutation in the *MYH7* gene detected by trio whole exome sequencing (Paper IV).

4. MATERIALS AND METHODS

4.1. Study subjects

4.1.1. Cohort for CMA study and inclusion criteria (Paper I)

After excluding prenatal and non-affected parental samples, the primary study group for investigating the clinical role of single LCSHs consisted of 2110 consecutive patients analysed by CMA in our department during the years 2011 to 2014. The most patients received CMA as a first-tier genetic investigation due to MCA, ID, and/or ASD. Other indications and testing scenarios, however, were also included. Although CMA was first introduced to the cytogenetic diagnostics in 2009 in Estonia, the Estonian Health Insurance Fund started to reimburse the test in 2011. Thus, patients included in this study are all a part of routine diagnostics and not selected for severest phenotypes [Žilina et al. 2014b].

Patient inclusion criteria for the subsequent analysis was the presence of one or two LCSHs with a minimal length of 5 Mb per LCSH, but a total length not exceeding 28 Mb. Cases with the co-occurrence of pathogenic, likely pathogenic, or unclearly significant CNVs were excluded to focus solely on the clinical utility of LCSHs. The upper limits of LCSH length were used to eliminate parental consanguinity, as 28 Mb equals approximately 1% of the total autosomal genome, which is below the expected percentage for the subject's parents being second cousins [Sund et al. 2013]. Additionally, the maximum number of two LCSHs was used, as it is more likely that parental consanguinity results in multiple LCSHs on different chromosomes. We further excluded all patients carrying LCSHs with lengths greater than 25% of the whole chromosome on which they occurred to eliminate possible UPDs. Finally, to compile the main study cohort, we eliminated patients carrying only recurrent LCSHs, which we defined as being present in at least three unrelated patients. One patient from the final study group was investigated further and described in detail in Paper III and in the Sections 4.1.3 and 5.3 of this thesis.

4.1.2. Study group investigated by large NGS panel (Paper II)

The main characteristics of the NGS panel sequencing cohort are presented in Table 5. In summary, 501 consecutive unselected patients (children, adults, and 11 prenatal cases) referred for NGS gene panel sequencing due to various indications were included in this study. These tests were performed between April 2015 and August 2016. For many, but not all, patients, prior DNA-testing had been performed. Thus, screening for pathogenic large CNVs by CMA, for example, had been performed prior to NGS in most cases with ID/DD and/or MCA. The phenotypic details were collected from referral forms (open text with no predefined fields). Referring doctors were also asked to list candidate genes for the patient's disorder, or pick one or more predefined subpanels, e.g., ID,

epilepsy, metabolic disorders, etc. (11 to 600 genes), with no limit on the total number of genes/panels. In addition, referral forms describing only phenotypes were accepted – in these 14 cases, the candidate gene list was compiled by a laboratory molecular geneticist. In all cases, sequencing was performed on probands only. Parental samples, however, were collected and investigated only for single candidate variants for carrier status or familial segregation, if needed and available.

Table 5. The main characteristics of our study cohort, which included 501 first unselected cases for whom TruSight One panel (Illumina Inc.) sequencing was performed as a clinical diagnostic test. Adapted from [Pajusalu et al. 2017].

Total number of analysed cases	**501**
Children (0–18 years)	313
Adults (19 or more years)	177
Prenatal cases	11
Females/males (postnatal cases)	226/264
Median age in full years, range (postnatal cases)	11.5, 0–78
Referrals by department profiles	
Clinical genetics	374
Paediatrics (incl. child neurology, child haematology and others)	84
Paediatric (incl. neonatal) intensive care unit	3
Adult neurology	30
Others (ophthalmology, internal medicine, adult intensive care)	10

4.1.3. Clinical description of two siblings with *KPTN* mutations (Paper III)

The probands were a 32-year-old Estonain man and his 24-year-old sister with ID of unknown cause. They have two healthy older sisters. Although the parents did not report known consanguinity, they were born in the same rural Estonian parish. There are no known genetic disorders in the family. The family consulted a medical geneticist due to suspected genetic disorder in 1993 when the brother and sister were 10 and 2 years of age respectively, but at the time no cause for DD was found and conventional karyotyping revealed no abnormalities. Now, as adults, the affected sibs live in a special home for the intellectually disabled. They have basic self-care and communication skills. The level of ID in both sibs was classified as moderate, although formal IQ-assessments have not been performed.

The brother was born with normal growth parameters (occipitofrontal circumference [OFC] 34 cm) as the third child in the family. He had no significant early motor DD, and started regular school. During the first grades, DD was diagnosed and he was transferred to a special school. He had one episode of

generalized seizures at the age of 10 months during bronchitis, leading to a few months of anti-convulsive therapy. The seizures recurred only once afterwards at the age of 10 years when he had one episode of unprovoked generalized seizures during sleep. At 7 years of age, no abnormalities were found on computed tomography (CT) brain scan. Electroencephalography (EEG) was done at the age of 10 years, and generalized slowing of background activity (more pronounced over the posterior brain) without epileptiform discharges was noted. Since childhood, he has had anxiety and behavioural disorder with autistic features, stereotypic movements and utterances, and some self-aggression. Now, at the age of 32, he is of normal height and weight, but he has macrocephaly – his OFC is 63 cm (+4.5 SD). He has a prominent forehead, high palate, and slight microretrognathia. He still has behavioural disorder and his language abilities are poor with repetitive speech occurring.

The sister was born with normal growth parameters (OFC 37 cm) as the fourth child in the family. No DD was noted in the first years of life, but due to markedly delayed speech development, ID was diagnosed in preschool age, and therefore she went to a special school. She has never had seizures, although generalized slowing of background activity (more over the posterior brain) was described on EEG at the age of 3 years. Currently, at the age of 24 years, she has macrocephaly with OFC of 60 cm (+4 SD). Her speech abilities are low, but she can understand basic commands. She has also had high anxiety levels, but less than her brother and without remarkable behavioural disturbances. She has a prominent forehead, high palate and microretrognathia.

4.1.4. Clinical description of a patient with *MYH7*-related myopathy (Paper IV)

The patient is a boy of Estonian and Spanish origin with early-onset muscular hypotonia and delayed motor development. He was born after an uneventful pregnancy and delivery and had normal growth parameters. He was referred to a paediatric neurologist at 3.5 months of age. At the time, he had a complete lack of motor head control, and severe muscular hypotonia was observed in the neck and shoulder girdle but was less pronounced in the legs. ENMG at 3.5 months was interpreted as normal, but at 6 months the repeated test showed features of myopathy. At 6 months, he still could not control his head upon traction and had a positive dropped head sign, and generalized muscular hypotonia.

At 13 months of age, he had marked muscular weakness and hypotonia most prominently in the neck extensors, with a hypomimic face and mild bilateral ptosis. Despite the weakness of proximal lower limb and trunk muscles, he could put weight on his legs if supported, but his head control was remarkably poor and the dropped head sign was still present. Moderate calf hypertrophy was noticed. There were no weaknesses or functional deficits in gross-motor functions of the hands, but he lacked fine motor skills. Despite the hypotonia, his deep tendon reflexes were exaggerated but with no pathological reflexes. He

had a normal head circumference (47 cm). Serum creatine kinase and lactate levels were within normal range and spinal fluid analysis was also normal. Echocardiography as well as MRI scans of brain and spinal cord showed no pathologic findings. The muscle biopsy from the left anterior tibial muscle revealed fiber-type disproportion with no inclusions (Figure 6). Although he had marked motor developmental delay, his mental and social skills were assessed to be normal for his age. Clinically the phenotype could be classified as congenital myopathy.

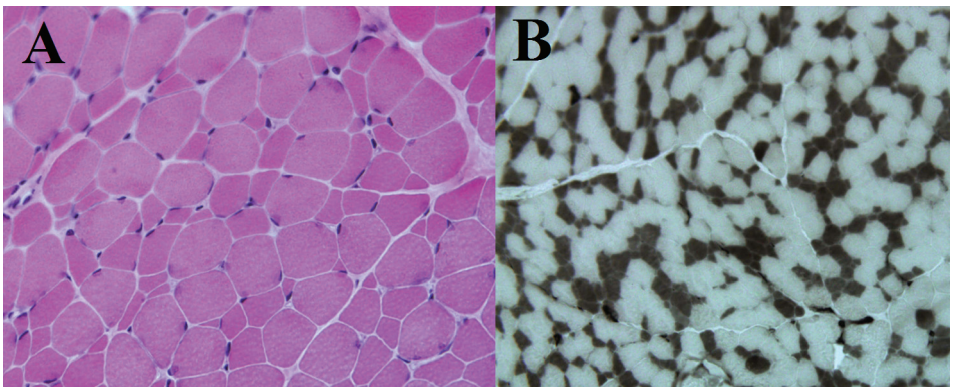


Figure 6. Cryosections of the patient's anterior tibial muscle. (A) Hematoxylin and eosin staining shows marked variation in fiber size, and (B) ATPase histochemical stain at pH 4.3 shows that type 1 muscle fibers (dark) are much smaller than type 2 muscle fibers (pale). Figure was prepared by Dr. Sanna Puusepp.

At the age of 21 months, the child is able to walk without support, but muscle hypotonia and weakness are still present, more pronounced in proximal muscles of lower limbs with Gower's sign and in the neck muscles. He uses orthosis to stabilize the trunk. He speaks two-word sentences and has no problems with breathing or swallowing. Repeated cardiac investigations showed no pathological findings.

4.2. Methods

Considering terminology, it must be noted that although according to ACMG guidelines, "variant" is preferred over "mutation" in many instances [Richards et al. 2015], in this thesis the term "mutation" is widely used to match the style with the four publications on which the dissertation is based.

4.2.1. The study of CMA-detected homozygous stretches (Paper I)

Genomic DNA was extracted from peripheral blood samples. CMA was performed on all samples using HumanCytoSNP-12 BeadChips (Illumina Inc.) at

the Estonian Genome Centre at the University of Tartu or Estonian Biocenter, both in Tartu, Estonia, and clinically interpreted at Tartu University Hospital. Genotypes were called by GenomeStudio software v2010.3 (Illumina Inc.). The GenomeStudio's cnvPartition plugin was used to detect LCSHs, with the minimum region size set to 5 Mb. All chromosomes of each sample were also visually inspected by a clinical cytogeneticist to eliminate false-positive and false-negative calls.

All LCSHs belonging to the subjects of the final study group were evaluated for clinical significance by searching for encompassed, AR disease-associated genes. This was done using the web-based Genomic Oligoarray and SNP array evaluation tool v.2.0 [Wierenga et al. 2013]. The search criterion used was "OMIM genes with recessive inheritance pattern". For every patient, the list of genes and associated phenotypes found by the software were compared with the patient's clinical information stated on the referral documents. If a candidate gene matching the patient's phenotype was found, WES was performed to find the causative mutation either from the identified candidate gene or any other genes not belonging to the LCSH. All probable pathogenic mutations identified by WES were confirmed by Sanger sequencing.

In the secondary analysis, we included all patients from the study group with LCSHs that did not encompass a good candidate gene and for whom WES was performed previously, although not as a part of this study. In cases of a homozygous pathogenic mutations, the CMA data were re-analysed to look for a <5 Mb LCSH in the region of the mutation.

4.2.2. Large gene panel sequencing and variant interpretation (Paper II)

After extraction, DNA was sent to either of two collaborating sequencing facilities (Asper Biotech or the Estonian Genome Centre at the University of Tartu, both located in Tartu, Estonia) in batches of 3, 9, 12, or 18 samples. Libraries were generated according to the manufacturer's protocols using TruSight One kits (Illumina Inc.). Sequencing was carried out on MiSeq or HiSeq platforms (Illumina Inc.) to mean sequencing depths of at least 70x and 150x, respectively. The bioinformatics analysis, including variant calling, was performed at sequencing facilities using their pipelines, which varied only in a few details between batches. In general, reads were aligned to reference genomes hg19 or b37 by Burrows-Wheeler Aligner [Li and Durbin 2009], and variants were called by GATK tools (UG or HC) [McKenna et al. 2010]. Variant call format (vcf) and bam files were transferred back to Tartu University Hospital for downstream analysis.

For the first few batches, variants from vcf files were annotated by VariantStudio (Illumina Inc.), but later this was shifted to an in-house variant annotation pipeline tailored to our needs. Annotations included, but were not limited to reference databases from ExAC [Lek et al. 2016] and 1000 Genomes

Project [1000 Genomes Project Consortium et al. 2015], and ClinVar pathogenicity annotations [Landrum et al. 2016], as well as HPO terms [Kohler et al. 2014] and OMIM disorders as gene-based annotations. Additionally, allele counts from our growing in-house database of variants detected among all NGS analyses (panels and WES) performed in our department (latest version used for this study consisted of 1011 samples) were annotated to every detected variant making it possible to exclude platform-specific false positive calls as well as to compare phenotypes of rare variant carriers.

CNVs were called using CoNIFER software [Krumm et al. 2012]. First, reads per thousand bases per million reads sequenced (RPKM) values were calculated for each sample separately. Second, all available samples from different batches were joined for CNV calling, depending on the sequencing platform (two different data sets of RPKM values for MiSeq and HiSeq samples). CNV detection and image generation for detected CNVs were carried out subsequently according to CoNIFER guidelines. In cases of high suspicion for single candidate genes (e.g., a single pathogenic mutation detected in a recessive gene), singular value decomposed Z-RPKM (SVD-ZRPKM) values were in addition evaluated manually, so as not to miss deletions smaller than three exons (i.e., below CNV detection size limit reported in original publication of CoNIFER).

In addition, coverage of requested genes was calculated using the GATK DepthOfCoverage tool. Sexes of the samples were estimated using the difference of expected versus observed heterozygosity on non-pseudoautosomal regions of chromosome X using VCFtools software [Danecek et al. 2011]. To detect discrepancies indicating suboptimal quality or sample swaps, sex estimations were subsequently compared to national identification codes (social security numbers), which state the sex as coded in the first digit.

As the first step of clinical interpretation of annotated variants, all non-requested genes were filtered out, thus eliminating the chance of detecting unsolicited findings. Variants were classified according to ACMG guidelines [Richards et al. 2015]. Alamut software (Interactive Biosoftware, Rouen, France) and Human Gene Mutation Database (HGMD® Professional) from BIOBASE Corporation [Stenson et al. 2009] were used to aid interpretation. Finally, class 5 (pathogenic) and class 4 (likely pathogenic) variants were reported back to the referring doctor. In addition, class 3 (VUS) variants were reported if variant classification was subject to change after additional studies (e.g., testing for segregation in family). For example, a novel heterozygous missense mutation in a dominant disease gene with multiple computational evidence of pathogenicity may have initially been reported as class 3, but after confirmation of *de novo* state, the variant could be reclassified as class 4. For compound heterozygous mutations, parental testing was performed by Sanger sequencing to confirm *trans*-position of the putative causal variants. Here, final variant classifications were used in assessing the clinical utility.

Most reported SNVs were confirmed by Sanger sequencing, including all low-quality variants (GATK quality score below 500). A clear majority of

higher quality variants were also confirmed among familial segregation analysis or parental carrier testing where index patients served as positive controls. All reported CNVs were confirmed by an alternative method, either MLPA or CMA, based on the size of the CNV and availability of the specific assay.

To assess whether diagnostic yield is dependent on the number of genes listed for analysis on the referral form, indicating the level of confidence for clinical hypothesis, two subgroups of the total study sample were created. The cut-off used was 10 genes or more for a large-panel group and less than 10 for a small-panel group. All samples for which no gene list was provided on the referral form were added to a large-panel subgroup. The statistical significance of the difference between diagnostic yields in two subgroups was tested using a two-sided Fisher's exact test. All statistical analyses were conducted in R version 3.3.1 [R Core Team 2016].

4.2.3. Genetic investigations performed in siblings with ID (Paper III)

To investigate probable genetic cause of the disorder, CMA on DNA samples from both sibs was performed using HumanCytoSNP-12 array (Illumina Inc.). To test the hypothesis of finding homozygous disease-causing mutation, WES was carried out on the brother's DNA sample. Library preparation and a sequencing run were performed by the company ServiceXS (Leiden, The Netherlands) using SureSelect XT Human All Exon v5 enrichment kit (Agilent Technologies, Santa Clara, CA) and HiSeq sequencer (Illumina Inc.). The fastq files were transferred to us. Subsequently, raw sequencing reads from fastq files were aligned to the hg19 reference genome using BWA [Li and Durbin 2009]. Bioinformatics processing, variant calling, and annotation were performed following GATK best practice guidelines [Van der Auwera et al. 2013] using Picard, GATK [DePristo et al. 2011; McKenna et al. 2010], Annovar [Wang et al. 2010] and SnpSift [Cingolani et al. 2012a] software. Variant interpretation was mostly focused on rare protein-damaging homozygous variants due to suspected shared ancestry in parents supported by LCSHs in the CMA results. The identified variant most likely to be causative for the phenotype for confirmed in the brother and tested for in the sister as well as in both parents by Sanger sequencing.

4.2.4. Molecular investigations in the patient with myopathy (Paper IV)

WES of the parents-offspring trio carried out in the Estonian Genome Centre at the University of Tartu. DNA libraries were performed using Nextera Rapid Capture Exome 37 Mb kit (Illumina Inc.) according to the manufacturer's protocols. The HiSeq 2500 (Illumina Inc.) platform was used for paired-end 2×100 bp sequencing. The bioinformatics data processing made use of BWA

[Li and Durbin 2009], which mapped the reads to the b37 reference genome; different Picard and GATK tools; and GenomeTrax from BIOBASE Corporation. The variant interpretation was focused on *de novo* and recessively inherited variants due to parents being healthy. Sanger sequencing was used for confirmation of the finding.

To investigate the detected mutation's effect on splicing, first, *in silico* analysis using both MutationTaster [Schwarz et al. 2014] and MutPred Splice [Mort et al. 2014] was performed. Subsequently, total RNA was extracted from the same muscle sample used for histologic analysis. After cDNA synthesis, PCR amplification was performed using three primers: a) MYH7_F1 (in exon 37): 5'-CATTAAGGACCTGCAGCACC-3', b) MYH7_R1 (in exon 39): 5'-AGCTTGTTGACCTGGGACTC-3', and c) MYH7_F2 (in exon 38): 5'-CGGAGGAGGACAGGAAAAAC-3'. Sanger sequencing was performed on PCR products extracted from gel fragments to check for skipping of exons.

4.3. Ethics

The study on the clinical utility of LCSHs (Paper I) was approved by the Research Ethics Committee of the University of Tartu (approval number 243/T-3, date 1/13/2015).

The part of the study which focused on NGS diagnostics (Paper II) was solely based on the results of clinical diagnostic investigations without any additional experiments or personalized data analysis performed for the study. Before the blood was collected, all patients, or as in the case of children, parents or legal guardians signed written informed consent forms approved by Tartu University Hospital or other local referring hospitals as subjects to routine clinical investigations. All samples were collected and analyses were performed according to national legislation and rules of Tartu University Hospital.

In both cases described in detail (Papers III and IV), the parents as legal guardians first signed written consents for performing the WES, and in addition, the consent for publishing photos in a scientific article.

5. RESULTS AND DISCUSSION

5.1. Single long contiguous stretches of homozygosity detected by diagnostic chromosomal microarray analysis and their clinical utility (Paper I)

During the study period, 172 (8.2%) patients out of 2,110 presented with an isolated finding of one or two LCSHs when considering the inclusion and exclusion criteria (other than recurrence). From a total of 172 patients, 161 patients carried a single LCSH and 11 patients carried two LCSHs, thus a total of 183 LCSHs were detected among the cohort.

As previously described, LCSHs represent the majority of VUSs reported after CMA in our diagnostic laboratory [Žilina et al. 2014b]. One of the most straightforward ways to decrease the number of VUSs is to identify recurrent LCSHs and to interpret them as benign polymorphisms, which can be left unreported depending on laboratory policies. In this study, we identified six different recurrent LCSHs totalling 52 events or 28.9% of all isolated LCSH findings (Table 6). Thus, we could clarify the clinical significance of nearly one-third of all LCSHs by classifying them as likely benign. Nevertheless, caution is still needed in classifying recurrent LCSHs as benign, because there is a possibility that shared haplotypes could be mutated in both parents. Moreover, shared recurrent LCSHs may indicate distant shared ancestry possibly increasing the risk for AR disorders, although this could not be investigated among this study due to lack of sufficient number of samples. Interestingly, out of six chromosomal regions where we identified a recurrence of a LCSH, only two matched with the five regions reported by Wang et al. [2014]. This could be attributed to different study populations, which clearly indicates the need for population-specific LCSH databases.

After excluding all recurrent LCSHs, 120 patients (5.7%) with 129 LCSHs remained in the final study sample (Table 7). Searching for genes associated with recessive disorders revealed a median of three genes per LCSH (range 0–15). In two cases, an appropriate candidate gene was discovered. First, in a 1-year-old girl with transfusion-dependent hemolytic anemia, an LCSH on chromosome 1 encompassed the *PKLR* gene known to cause pyruvate kinase deficiency leading to hemolysis and anemia. WES revealed a novel mutation in the *PKLR* gene that was predicted to cause an in-frame deletion of a single amino acid (NM_000298.5: c.1137_1139del p.(Lys380del)). Although enzyme analysis on the affected child was uninformative due to recurrent red blood cell transfusions, the pathogenicity of the mutation was still confirmed by detection of significantly decreased pyruvate kinase activity in both heterozygous parents.

Table 6. Recurrent, long contiguous stretches of homozygosity (LCSH) found in this study. Adapted from [Pajusalu et al. 2015c].

Cyto-bands	Maximal coordinates	Max size (Mb)	Mininmal overlapping coordinates	Minimal overlapping region size (Mb)	LSCH count (% of total 183)	First report
3p21.33–p21.1	chr3:43,805,245–54,093,664	10.3	chr3:48,787,219–50,483,437	1.7	17 (9.3%)	Wang et al. [2014]
6p22.3–p21.31	chr6:23,404,279–33,887,363	10.4	chr6:26,184,041– 30,071,279	3.9	9 (4.9%)	This report
7q11.21–q11.22	chr7:61,760,894–68,616,110	6.9	chr7:61,834,214–66,904,395	5.1	5 (2.7%)	This report
7q31.2–q31.33	chr7:116,266,792–124,008,463	7.7	chr7:117,783,250–122,990,979	5.2	3 (1.6%)	This report
11p11.2–q12.1	chr11:44,901,903–57,160,838	12.3	chr11:46,342,376–51,274,692	4.9	11 (6.0%)	Wang et al. [2014]
Xq13.1–q21.1	chrX:71,595,785–83,584,844	12.0	chrX:73,424,191–78,848,072	5.4	7 (3.8%)	This report
				TOTAL:	52 (28.4%)	

Second, in a 6-year-old boy with growth failure, DD, ataxia, and cerebellar atrophy, a LSCH on chromosome 5 was found to encompass the *SIL1* gene that is associated with Marinesco-Sjögren syndrome. WES revealed a homozygous one-base-pair duplication in the *SIL1* gene that was predicted to cause a translational frameshift, thereby resulting in a truncated protein; NM_022464.4: c.947dupT p.(Arg317Glufs*35). The mutation was first reported as disease-causing by Senderek et al. [2005]. The parents were confirmed to be heterozygous carriers by Sanger sequencing.

Table 7. Main characteristics of non-recurrent, long contiguous stretches of homozygosity (LCSH) detected in this study. Adapted from [Pajusalu et al. 2015c].

Total number of patients with non-recurrent LCSH (% of a total of 2,110 patients)	**120 (5.7%)**
Number of patients with one/two LCSHs	111/9
Non-recurrent LCSH count	129
Size range of single LCSHs	5–28 Mb
Size range of the sum of two LCSHs in one patient	11.1–25.1 Mb
Average size of a single LCSH	7.7 Mb
Number of genes causing a recessive disorder in a LCSH (median)	0–15 (3)
Number of cases with candidate recessive gene in a LCSH	2 (in both, the mutation was confirmed)
Total number of patients with non-recurrent LCSH (% of a total of 2,110 patients)	120 (5.7%)
Number of patients with one/two LCSHs	111/9
Non-recurrent LCSH count	129
Size range of single LCSHs	5–28 Mb

For patients for whom WES had been ordered outside this study, seven had been previously found to carry a LCSH by CMA that did not encompass a good candidate gene, and two of them had a recurrent LCSH. Of these seven patients, two received a definitive diagnosis by WES. First, WES in an adult man with ID and macrocephaly (reported in detail by our previous publication [Pajusalu et al. 2015b] and further discussed in Section 5.3) revealed a homozygous one-base-pair duplication in the *KPTN* gene (19q13.32) that is associated with a similar phenotype [Baple et al. 2014]. The patient also has a sister with a similar phenotype who carried the same homozygous mutation. A previous CMA found two LCSHs in the man (17.6 Mb in 1q25.3-q32.1 and 7.5 Mb in 14q13.3-q21.2), but only one of them (in 14q13.3-q21.2) was also present in his sister. However, the *KPTN* gene is not located in either of these LCSHs. Reanalysis of the CMA data revealed a 1.5-Mb LCSH on chromosome 19 that encompasses the *KPTN* gene and which is shared by both sibs (Figure 7). This LCSH was not reported after the initial CMA interpretation due to 5-Mb cut-off length

routinely used in our laboratory. The parents of the sibs did not report consanguinity, but were born in the same parish. Therefore, if there is a strong indication for recessive inheritance and distant consanguinity or common ancestry is likely, lowering the cut-off of the LCSH size can aid in identification of candidate genes. For example, a LCSH minimum cut-off length of 1 Mb was used in a study that focused on familial recessive IDs in outbred families [Schuurs-Hoeijmakers et al. 2011]. Second, in a girl carrying a LCSH on the X chromosome, compound heterozygous mutations were found in the *REN* gene (1p32.1), resulting in renal tubular dysgenesis. Thus, one should not presume that the finding of homozygosity is a general rule in patients with single LSCHs. Comprehensive analysis of WES data is needed to maximize the diagnostic yield.

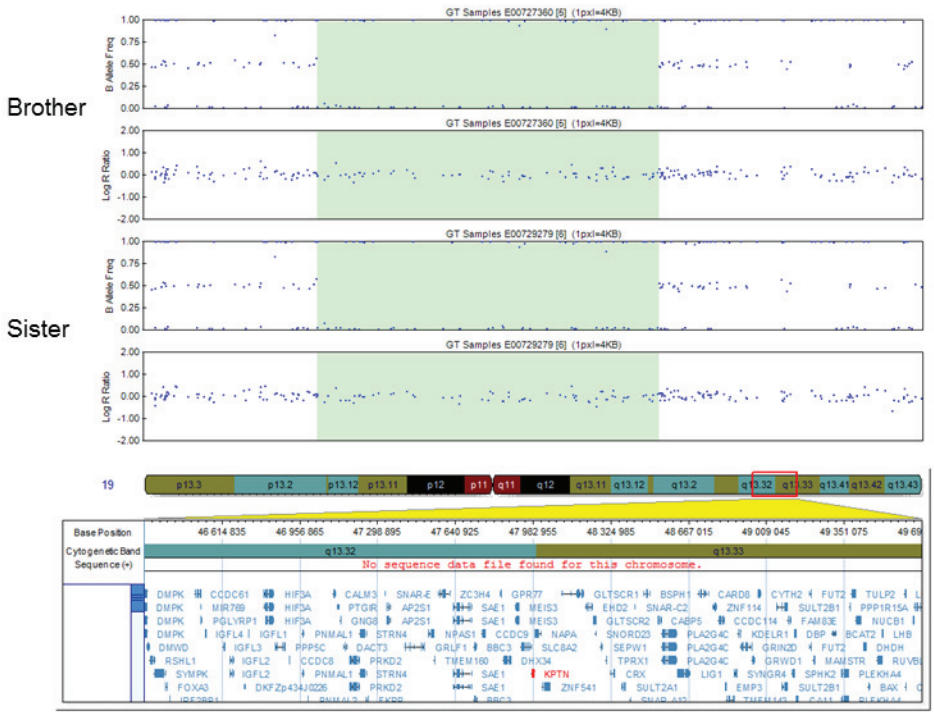


Figure 7. Graphical representation of the SNP-array data from both sibs in the chromosomal region encompassing the *KPTN* gene. 1.5 Mb shared region of homozygosity was detected after reanalyzing chromosomal microarray data. Figure was prepared using GenomeStudio v2011.1 software (Illumina Inc.).

During the last decade, studies have demonstrated the great utility of CMA in clinical diagnostics as a first-tier diagnostic cytogenetic test for patients with IDs, ASDs, and MCAs [Ahn et al. 2013; Hochstenbach et al. 2009; Miller et al. 2010; Vissers et al. 2010]. The primary goal of CMA is to find a causative microdeletion or microduplication. SNP-arrays have the advantage of also revealing genotype information, which allows detection of copy number neutral

chromosomal aberrations associated with an increased risk of AR disorders. Multiple LCSHs on different chromosomes that indicate close parental consanguinity, as well as LCSHs that cover the majority of single chromosome and are caused by UPD, have been studied during the last few years, and their clinical implications are thus better understood [Papenhausen et al. 2011; Wang et al. 2014]. Single LCSHs in patients without parental consanguinity are not infrequently found during routine diagnostics in centers where SNP array technology is in use, but their clinical significance remains unclear in most of the cases [Žilina et al. 2014b]. Reporting a VUS can cause anxiety in patients and frustration in referring doctors who are not specialists in the field of medical genetics [Coughlin et al. 2012; Žilina et al. 2014b]. Therefore, it is of great importance to work toward minimising the number of reported VUSs.

Although the homozygosity of a chromosomal region can be intuitively attributed to the increased risk of AR disorders, the clinical significance of single LCSHs remains unclear. In this study, only patients with one or two LCSHs, with total length not exceeding 28 Mb (i.e., 1% of autosomal genome), were analysed. As suspected, because only a very small proportion of the whole genome was covered, finding a candidate gene closely associated with a patient's phenotype was very rare. Nevertheless, if a well-matched candidate gene is found, the confirmation of pathogenic mutation is likely. This is illustrated by the described two patients with a plausible candidate gene inside the single LCSH and confirmed pathogenic variants in both. Wang et al. [2014] found nine patients with good candidate genes and a pathogenic variant was detected in seven of them. Although the study sample of Wang et al. was almost seven times larger than this study, and patients with parental consanguinity were also included, it can be concluded that the results are consistent with the previously published study [Wang et al. 2014] due to the very small number of cases with identified causal variants and the associated probability of a large statistical error. Despite the low success rate in this study, it can be can still recommended to look through detected, non-recurrent LCSHs for candidate genes, as this can lead to a molecular diagnosis. Additionally, if software tools such as the Genomic Oligoarray and SNP array evaluation tool [Wierenga et al. 2013], are used, the evaluation of LCSH regions is not very time-consuming, as the median number of genes causing a recessive disorder per LCSH was only three in our study group.

However, as genome-wide sequencing investigations are becoming more available and often already used as first-tier or the second test after inconclusive CMA results, the diagnostic value of LCSHs detected by CMA probably decreases, as rare homozygous variants can be easily prioritized without prior knowledge of LSCHs. Nevertheless, it could be hypothesised that if sequencing of coding regions does not reveal pathogenic variant in a good candidate gene located in LCSH, it may be worth doing additional studies on the gene to reveal non-coding variants. Thus, the information from detected LSCHs may remain important even in the NGS era.

5.2. Clinical utility of large gene panel sequencing in routine diagnostics of suspected Mendelian disorders (Paper II)

For 132 out of total 501 consecutive patients who received NGS gene panel testing during the study period, diagnosis for a specific genetic disorder was established (at least likely pathogenic variant consistent with patient phenotype). Summary characteristics of detected mutations are presented in Table 8.

Table 8. Summary of detected mutations. Adapted from [Pajusalu et al. 2017].

Total number of cases	**501**
Solved cases	132 (26.3%)
Partially solved cases	2 (0.4%)
Cases with reported VUS	43 (8.6%)
Patients with negative reports	323 (64.5%)
Total number of genetic disorders in solved cases	133
Types of causative mutations in solved cases:	
SNVs and indels	125 (94%)
Intragenic (exonic) deletions	3 (2.3%)
SNV + CNV compound	2 (1.5%)
Microdeletion	2 (1.5%)
X-chromosome monosomy	1 (0.75%)
Total number of reported causative variants	157
Number of recurrent mutations (total recurrences)	4 (9)
Number of non-recurrent genic mutations	146
Number of novel mutations	67 (46%)
Diagnosed disorders by inheritance	
Dominant	86 (65%)
Heterozygous mutation	84
Inherited	29
De novo	19
Unknown	36
Mosaic variant	2
Recessive	31 (23%)
Compound heterozygous mutations	23
Homozygous mutation	8
X-linked recessive	14 (10.5%)
Other	2 (1.5%)

VUS – variant of unknown significance. SNV – single nucleotide variant, indel – small insertion or deletion, CNV – copy number variant.

The overall diagnostic yield of NGS gene panel testing was revealed to be 26.3%. In addition, for two patients (0.4%), partial diagnosis was established, meaning that only some of the phenotype could be confidently explained by a detected class 4 or 5 variant. VUS or class 3 variants were reported for 43 patients (8.6%), and not clarified by the time of writing this manuscript. All reported mutations are listed in Table 9, case numbers in this text section follow the numeration in Table 9

In the 132 solved diagnostic cases, we established a diagnosis for 133 genetic disorders (Table 9), due to double diagnosis in one patient who had both hereditary ovarian cancer and congenital myotonia (Case #24). Out of these 133 disorders, 125 (94%) were caused by either SNVs or indels, three by intragenic CNVs, and two by a combination of CNV and SNV. In addition, two pathogenic microdeletions encompassing many genes and one monosomy of X chromosome were detected. X chromosome monosomy, or Turner syndrome (Case #126), in a girl with congenital liver disorder was incidentally discovered by mismatch in expected and observed sex estimation during quality control of the data, and later confirmed by conventional karyotyping.

After leaving out the Turner syndrome case, 86 of the 132 remaining disorders (65%) were dominant. Of all dominant cases, 84 were caused by germline heterozygous mutations, including four cases of dominant X-linked disorders in females (#51, #62, #120, #129), and two were caused by mosaic mutations. Of the 84 cases with dominant disorders (mosaics excluded), we possessed data from testing of both parents for 49 cases. Out of these 49 cases, 29 (59%) carried an inherited mutation segregating with the phenotype, 19 (39%) had a *de novo* mutation and one (Case #110) was identified as carrying a maternally inherited mutation in the *UBE3A* gene, consistent with the diagnosis of Angelman syndrome due to gene imprinting. In 31 cases (23%), an AR disorder was diagnosed and in most of such cases (23 patients), compound heterozygous mutations were detected, with the remaining eight being caused by homozygous mutations.

X-linked recessive disorders were diagnosed in 14 males (11%), 13 carrying hemizygous mutations, and one mosaic mutation causing Danon disease (Case #68). In the remaining case (#124), two rare, likely pathogenic, mutations on different alleles were found in the *MYH7* gene in a boy with early-onset cardiomyopathy. Both mutations were also apparent in his sister with a similar phenotype. Although mutations in *MYH7* are known to cause dominant cardiomyopathy, we cannot rule out the possibility of a modifying effect of the second mutation, thus causing a more severe phenotype.

Due to compound heterozygosity, 157 causative mutations were found in 132 solved cases (Table 9). Of these, 67 (43%) were not reported in the HGMD professional database, and thus were considered to be novel. We detected several recurring mutations. NM_000083.2(CLCN1):c.2680C>T p.(Arg894*) appeared in six patients, in either a homozygous or compound heterozygous state with another pathogenic mutation. Three other recurrent mutations appeared twice: NM_001171.5(ABCC6):c.3421C>T p.(Arg1141*), NM_000334.4(SCN4A):

c.4765G>A p.(Val1589Met) and NM_000401.3(EXT2):c.635G>C p.(Arg212Thr). Out of all recurrences, to our knowledge, only the SCN4A mutation causing congenital paramyotonia was due to close relatedness between patients. After removing recurrences and larger chromosomal events (microdeletions and monosomy of the X chromosome), 146 non-recurrent gene mutations remained, with the proportion of novel mutations increasing to 46%.

As there are no similar studies published on clinical utility of large gene panel sequencing in routine clinical setting, the comparison of the yield may be done only in the context of published WES cohorts. Despite many limitations, including the fixed and not entirely comprehensive content of the used panel, restriction of the analyses to requested genes only and sequencing only probands, the diagnostic yield in the presented cohort was comparable to previously published WES studies performed in unselected clinical cohorts [de Ligt et al. 2012; Farwell et al. 2015; Lee et al. 2014; Yang et al. 2013; Yang et al. 2014].

The diagnostic yield of our WES cohort of 68 families at the time of completing this study was 28%, with an additional 21% of families receiving a report containing VUS. Thus, the efficiency of WES is very similar at our centre if compared our results to other published clinical cohorts. It can be argued that our relatively high diagnostic yield is influenced by selection of cases – panel sequencing was generally favoured over WES in cases where a clearer clinical hypothesis was established prior to genetic testing. WES was available in parallel and was used in more complex cases.

The assumption of higher diagnostic yield in cases with a clear diagnostic hypothesis is supported by our observation that the subgroup of patients in whom less than ten genes were requested to be analysed showed a significantly higher yield than the larger panel subgroup. Namely, for 238 patients, one to nine genes were listed for analysis on referral forms, and a larger panel was requested for the remaining 263 samples. The diagnostic yield in the small-panel subgroup was 31.5%, compared to 21.7% in the large-panel subgroup. The difference was statistically significant ($p = 0.015$). The subgroups did not differ in age ($p = 0.7$) or sex ($p = 0.63$) distribution.

It should be noted that gene panel sequencing was also used for many cases with very clear indication for testing of one or two genes, such as in the cases of tuberous sclerosis, polycystic kidney disease, or Duchenne muscular dystrophy. This was mostly due to the absence of routine availability of separate single gene tests for these and many other disorders in Estonia, supported with the relatively low cost of gene panel sequencing.

Another difference from previously reported WES clinical cohorts, as well as our own experience with WES, was a significantly higher proportion of inherited autosomal dominant mutations, which outnumbered the *de novo* cases. In WES-based cohorts, inherited dominant mutations are rarely seen because of the selection bias towards severe paediatric cases with unaffected parents [Farwell et al. 2015; Lee et al. 2014]. The other obvious reason, however, could be the limitation of our single patient-based approach, where *de novo* mutations are confirmed only after separate testing of parents. Thus, some *de novo*

mutations could have been misinterpreted as VUSs in initial analysis due to lack of parental genetic information. Regarding AR disorders, causative compound heterozygous mutations appeared more often than did homozygous mutations, which is consistent with WES-based studies of outbred populations [Lee et al. 2014; Yang et al. 2014], as well as our previous study showing that homozygous stretches detected by CMA are rarely associated with genetic disorders in Estonia, discussed in Section 5.1 and in publication [Pajusalu et al. 2015c].

NGS-based methods allow screening for many types of mutations. In our cohort, we found, in addition to SNVs and indels, CNVs of different sizes, from single exon to over a Mb in length, and an incidental finding of Turner syndrome due to a discrepancy found during quality control when estimating sexes of samples. It is worth noting that in two cases, pathogenic CNVs and SNVs were found in a compound heterozygous state. The ability to simultaneously detect different types of mutations is a major advantage of NGS over Sanger sequencing-based assays. We cannot, however, recommend the routine usage of gene panel sequencing instead of CMA due to unknown sensitivity and dependence of gene panel targets for detecting microdeletions and microduplications. Nevertheless, for the disease groups where CNVs, especially non-coding CNVs, are an unlikely cause, the gene panel sequencing can be used as a preferred first-tier diagnostic test. In our cohort, CNVs contributed to genetic diagnosis in seven cases, increasing the diagnostic yield by 1.4%, which is comparable to a recent report on WES-based CNV detection in clinical diagnostic samples [Pfundt et al. 2016]. Despite being possible only in a few cases, the ability of NGS to obtain allelic information without additional testing of family members could be considered as another advantage of NGS, if compared to routine Sanger sequencing. For example, due to the proximity of causative mutations in patients #1 and #16 (Table 9), we were able to physically phase two heterozygous mutations into different alleles, allowing us to instantly clarify the trans-position of mutations (Figure 8 illustrates Case #1). Due to the observed high rate of previously unreported variants detected in our cohort (46%), the advantage over array-based mutation testing, which relies on known information about disease causing variations, is clearly recognizable.

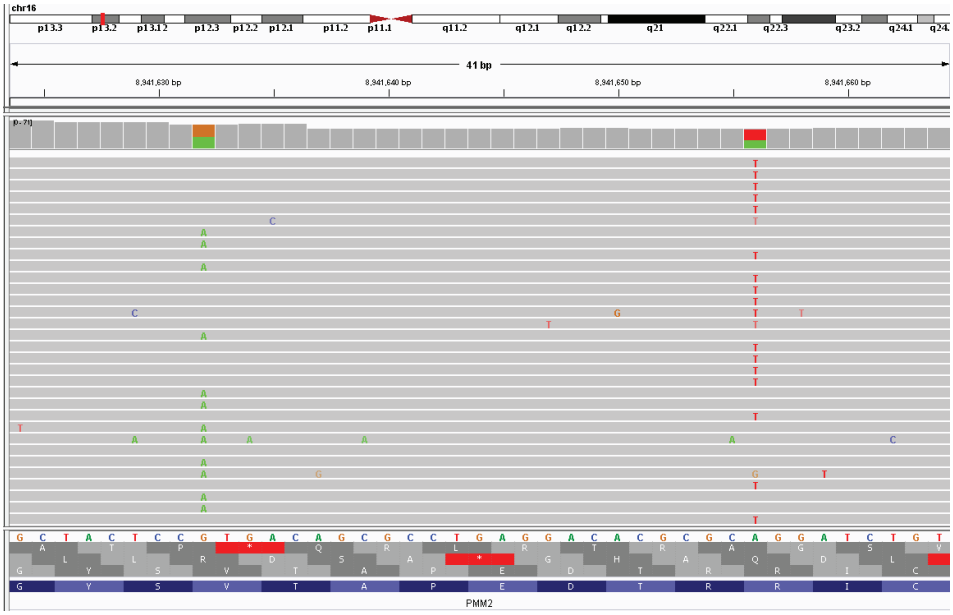


Figure 8. *PMM2* compound heterozygous mutations in Case #1 showing the ability to physically phase closely located gene variants. In this case, two heterozygous mutations appear on different chromosomes, as there are no reads encompassing both mutations. Figure was prepared using Integrative Genomics Viewer [Robinson et al. 2011].

Although having many pros, there are some disadvantages of large capture-based sequencing panels compared to either Sanger sequencing or smaller amplicon-based NGS panels. The major issue, which is no different from WES, is a decrease in sensitivity due to incomplete coverage of some genes or exons. Although large gene panels are generally able to obtain informative (i.e., over 20-fold) coverage of approximately 95% of targeted bases, the poor coverage of some genes remains a considerable drawback. To face this issue, we have made the information about average gene-based coverage (percentage of targeted bases with over 20-fold depth) available for referring doctors, guiding them in choosing the most appropriate test for their patients. In addition, we report back the coverage of requested genes calculated individually for every sample after the completion of test, thus providing an estimation of sensitivity. Due to its limitation regarding coverage, large gene panel sequencing, which can be referred to as a class C NGS test according to European guidelines [Matthijs et al. 2016], is not suitable for excluding pathogenic variation for many disorders. Nevertheless, the ability to efficiently screen for a large proportion of mutations justifies its use in clinical diagnostics.

Table 9. Mutations from 132 patients where large gene panel sequencing resulted in diagnostic findings. Variant classification follows the American College of Medical Genetics and Genomics (ACMG) guidelines [Richards et al. 2015]. Adapted from [Pajusalu et al. 2017] – supplementary information. **Phenotype /syndrome** – as listed on referral forms (ID – intellectual disability, DD – developmental delay, ASD – autism spectrum disorder, MCA – multiple congenital anomalies, IUGR – intra-uterine growth retardation, s. – syndrome). **Type** – type of mutation (MS – missense, SPL – splice site, NS – nonsense, FS – frameshift deletion or duplication, IF – in-frame deletion or duplication, CNV – copy number variation, Oth – other), **Zyg** – zygosity (HET – heterozygous, HOM – homozygous, MOS – mosaic, C-HET –compound heterozygous, HEM – hemizygous), **Inh** – observed inheritance (AD – autosomal dominant, AR – autosomal recessive, XL – X-linked, DN – *de novo*, IM – inherited mutation in imprinted gene, NA – no information about familial segregation), **Class** – ACMG class (5 – pathogenic, 4 – likely pathogenic, 3 – unclear significance), previously described pathogenic mutations listed in HGMD professional database, online version (as of Sep 28, 2016) are marked as KV – **K**nown disease causing **V**ariant. **Final diagnosis** – diagnosis established after gene panel sequencing (as in OMIM). CDG – congenital disorder of glycosylation, MODY – maturity-onset diabetes of the young.

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
1	Congenital disorder of glycosylation	*PMM2*	NM_000303.2:c.[691G>A];[715A>T] p.[(Val231Met)];[(Arg239Trp)]	MS; MS	C-HET	AR	KV; KV	PMM2-CDG
2	Marfanoid habitus, ID	*SRCAP*	NM_006662.2:c.9364del p.(Leu3122Cysfs*6)	FS	HET	NA	4	Floating-Harbor s.
3	CHARGE s.	*CHD7*	NM_017780.3:c.2699_2718del p.(Pro900Glnfs*8)	FS	HET	NA	4	CHARGE s.
4	Charcot-Marie-Tooth disease	*GJB1*	NM_001097642.2:c.319C>T p.(Arg107Trp)	MS	HEM	NA	KV	Charcot-Marie-Tooth neuropathy, XLD, 1
5	Marfan s.	*FBN1*	NM_000138.4:c.4delC p.(Arg2Valfs*16)	FS	HET	AD	5	Marfan s.
6	Early-onset obesity	*MC4R*	NM_005912.2:c.728G>A p.(Gly243Glu)	MS	HET	AD	4	Obesity, AD
7	Congenital myotonia	*CLCN1*	NM_000083.2:c.[1437_1450del];[2680C>T] p.[(Pro480Hisfs*24)];[(Arg894*)]	FS; NS	C-HET	AR	KV; KV	Myotonia congenita, AR
8	ASD, early puberty	*CHD8*	NM_001170629.1:c.2423_2424del p.(Arg808Lysfs*12)	FS	HET	AD	5	Autism, susceptibility to, 18
9	Porphyria	*HMBS*	NM_000190.3:c.445C>T p.(Arg149*)	NS	HET	AD	KV	Porphyria, acute intermittent
10	Muscular dystrophy	*DMD*	NM_004006.2:c.(1812+1_1813-1) _(6912+1_6913-1)del p.(Leu606_Val2305del) (ex16_ex47del)	CNV	HEM	DN	5	Duchenne muscular dystrophy

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
11	DD	*MVK*	NM_000431.3:c.[610G>C];[c.1000G>A] p.[(Asp204His)];[(Ala334Thr)]	MS; MS	C-HET	AR	4; KV	Mevalonic aciduria
12	Muscular dystrophy	*DMD*	NM_004006.2:c.2137C>T p.(Gln713*)	NS	HEM	NA	KV	Duchenne muscular dystrophy
13	Episodic ataxia	*CACNA1A*	NM_023035.2:c.757C>T p.(His253Tyr)	MS	HET	NA	KV	Episodic ataxia, type 2
14	Neonatal pulmonary hypertension	*FOXF1*	NM_001451.2:c.127_133del p.(Ile43Alafs*25)	FS	HET	DN	5	Alveolar capillary dysplasia with misalignment of pulmonary veins
15	Osteogenesis imperfecta	*COL1A2*	NM_000089.3:c.865G>A p.(Gly289Ser)	MS	HET	DN	4	Osteogenesis imperfecta, type IV
16	ID, speech impairment	*BCS1L*	NM_001257342.1:c.[232A>G];[245C>T] p.[(Ser78Gly)];[(Ser82Leu)]	MS; MS	C-HET	AR	KV; 4	Mitochondrial complex III deficiency, nuclear type 1
17	Usher s	*USH2A*	NM_206933.2:c.[(4627+1_4628-1)_(4987+1_4988-1)del];[11864G>A] p.[(Gly1543_Pro1662del)];[(Trp3955*)] (ex22_ex24del;Trp3955*)	CNV; NS	C-HET	AR	KV; KV	Usher syndrome, type 2A
18	Cerebral aneurysm	*COL3A1*	NM_000090.3:c.1691G>A p.(Gly564Asp)	MS	HET	NA	KV	Ehlers-Danlos s., type IV
19	Van der Woude s.	*IRF6*	NM_006147.3:c.120G>A p.(Trp40*)	NS	HET	AD	5	van der Woude s.
20	Cleft lip and palate, syndactyly	*TP63*	NM_003722.4:c.1028G>A p.(Arg343Gln)	MS	HET	DN	KV	Ectrodactyly, ecto-dermal dysplasia, and cleft lip/palate syndrome 3
21	Limited elbow extension	*LMX1B*	NM_002316.3:c.309C>G p.(Cys103Trp)	MS	HET	NA	KV	Nail-patella s.

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
22	Congenital myotonia	*SCN4A*	NM_000334.4:c.4765G>A p.(Val1589Met)	MS	HET	AD	KV	Paramyotonia congenita
23	Progressive muscular disorder	*DOK7*	NM_173660.4:c.1124_1127dup p.(Ala378Serfs*30)	FS	HOM	AR	KV	Myasthenic syndrome, congenital, 10
24	Ovarian cancer and congenital myotonia	*BRIP1*	NM_032043.2:c.806C>G p.(Ser269*)	NS	HET	AD	4	Breast cancer, early-onset
		SCN4A	NM_000334.4:c.4765G>A p.(Val1589Met)	MS	HET	AD	KV	Paramyotonia congenita
25	Polycystic liver disease	*PRKCSH*	NM_002743.3:c.1329del p.(Thr444Profs*20)	FS	HET	NA	4	Polycystic liver disease 1
26	Glanzmann thrombasthenia	*ITGA2B*	NM_000419.4:c.999-1G>A p.?	SPL	HET	NA	4	Glanzmann thrombasthenia
27	Craniosynostosis	*FGFR3*	NM_000142.4:c.749C>G p.(Pro250Arg)	MS	HET	NA	KV	Muenke s.
28	Defective function of polymorpho-nuclear neutrophils	*GUSB*	NM_000181.3:c.[511G>A];[1880G>A] p.[(Ala171Thr)];[(Trp627*)]	MS; NS	C-HET	AR	4; 5	Mucopolysaccha-ridosis VII
29	Ehlers-Danlos s. (type III)	*COL1A1*	NM_000088.3:c.4214G>A p.(Arg1405His)	MS	HET	AD	4	Ehlers-Danlos syndrome, classic
30	Ciliopathy	*RPGRIP1L*	NM_015272.4:c.[2407C>T];[2794_2795del] p.[(Gln803*)];[(Leu932Argfs*22)]	NS; FS	C-HET	AR	5; 5	Joubert syndrome 7
31	Torsion dystonia	*TOR1A*	NM_000113.2:c.907_909del p.(Glu303del)	IF	HET	NA	KV	Dystonia-1, torsion
32	Osler-Weber-Rendu disease	*ACVRL1*	NM_000020.2:c.1048del p.(Gly350Alafs*4)	FS	HET	NA	5	Osler-Weber-Rendu disease
33	Skeletal dysplasia, short stature	*FBN1*	NM_000138.4:c.5282C>T p.(Thr1761Ile)	MS	HET	DN	4	Geleophysic dysplasia 2
34	Congenital myotonia	*CLCN1*	NM_000083.2:c.[899G>A];[2680C>T] p.[(Arg300Gln)];[(Arg894*)]	MS; NS	C-HET	AR	KV; KV	Myotonia congenita, AR

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
35	Craniosynostosis	*TWIST1*	NM_000474.3:c.469G>T p.(Asp157Tyr)	MS	HET	NA	4	Saethre-Chotzen s.
36	Multiple midline defects	*MID1*	NM_000381.3:c.1508_1527dup p.(Glu510Thrfs*2)	FS	HEM	DN	5	Opitz GBBB s., type I
37	Dystonia	*ATP1A3*	NM_001256214.1:c.2878G>A p.Gly960Arg	MS	HET	DN	KV	Alternating hemiplegia of childhood 2
38	Allan-Herndon-Dudley s.	*SLC16A2*	NM_006517.4:c.1468G>A p.Gly490Arg	MS	HEM	XL	KV	Allan-Herndon-Dudley s.
39	Limb-girdle muscular dystrophy	*CAPN3*	NM_000070.2:c.[550del];[2092C>T] p.[(Thr184Argfs*36)];[(Arg698Cys)]	FS; MS	C-HET	AR	KV; KV	Muscular dystrophy, limb-girdle, type 2A
40	MODY	*GCK*	NM_033507.1:c.881T>C p.(Ile294Thr)	MS	HET	NA	KV	MODY, type II
41	Achondroplasia	*FGFR3*	NM_000142.4:c.1138G>A p.(Gly380Arg)	MS	HET	DN	KV	Achondroplasia
42	Congenital myotonia	*CLCN1*	NM_000083.2:c.950G>C p.(Arg317Pro)	MS	HET	AD	4	Myotonia congenita, AD
43	Microcephaly, DD	*EFTUD2*	NM_004247.3:c.2348-2A>G p.?	SPL	HET	NA	4	Mandibulofacial dysostosis, Guion-Almeida type
44	Strokes, hyper-cholesterolemia	*LDLR*	NM_000527.4:c.986G>A p.(Cys329Tyr)	MS	HET	NA	KV	Hypercholestero-lemia, familial
45	Craniosynostosis, DD, facial dysmorphism	*EHMT1*	NM_024757.4:c.3345C>A p.(Cys1115*)	NS	HET	DN	5	Kleefstra s.
46	Multiple epiphyseal dysplasia	*SLC26A2*	NM_000112.3:c.1957T>A p.(Cys653Ser)	MS	HOM	AR	KV	Epiphyseal dys-plasia, multiple, 4
47	Cerebellar atrophy	*SPTBN2*	NM_006946.2:c.1309C>T p.(Arg437Trp)	MS	HET	DN	4	Spinocerebellar ataxia 5
48	Neurofibromatosis	*NF1*	NM_001042492.2(NF1):c.5907_5908del p.(Arg1970Serfs*6)	FS	HET	AD	KV	Neurofibromatosis, type 1

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
49	Adreno-leukodystrophy	*ABCD1*	NM_000033.3:c.243dup p.(Leu82Alafs*113)	FS	HEM	XL	5	Adrenoleukody-strophy
50	MCA	*MID1*	NM_000381.3:c.1127T>C p.(Leu376Pro)	MS	HEM	XL	4	Opitz GBBB s., type I
51	Microcephaly, ID, short stature	*PDHA1*	NM_001173454.1:c.1014_1017dup p.(Arg340Leufs*13)	FS	HET	DN	5	Pyruvate dehydrogenase E1-alpha deficiency
52	MODY	*HNF1A*	NM_000545.5:c.1509C>G p.(Tyr503*)	NS	HET	AD	5	MODY, type III
53	Neuropathy	*MPZ*	NM_000530.7:c.341T>C p.(Ile114Thr)	MS	HET	AD	KV	Charcot-Marie-Tooth disease, type 1B
54	Immunodeficiency	*DCLRE1C*	NM_001033855.2:c.(?_-1)_(246+1_247_1)del p.(Met1_Ile82del) (ex1_ex3del)	CNV	HOM	AR	KV	Omenn s.
55	Polycystic kidney disease	*PKHD1*	NM_138694.3:c.[107C>T];[6992T>A] p.[(Thr36Met)];[(Ile2331Lys)]	MS; MS	C-HET	AR	KV; KV	Polycystic kidney and hepatic disease
56	Anophthalmia, Growth hormone deficiency	*OTX2*	NM_021728.3:c.133A>T p.(Lys45*)	NS	HET	NA	5	Microphthalmia, syndromic 5
57	Gait disturbance, hearing impairment	*MAN2B1*	NM_000528.3:c.[566C>A];[2248C>T] p.[(Pro189His)];[(Arg750Trp)]	MS; MS	C-HET	AR	4; KV	Mannosidosis, alpha-
58	Polycystic kidney disease	*PKD1*	NM_000296.3:c.12058_12059dup p.(Ala4021Glufs*18)	FS	HET	NA	5	Polycystic kidney disease, adult type I
59	Polycystic kidney and liver disease	*PKD1*	NM_000296.3:c.8311G>A p.(Glu2771Lys)	MS	HET	NA	KV	Polycystic kidney disease, adult type I
60	Coffin-Lowry s.	*RPS6KA3*	NM_004586.2:c.889_890del p.(Leu298Phefs*21)	FS	HEM	XL	KV	Coffin-Lowry s
61	Multiple exostoses	*EXT2*	NM_000401.3:c.635G>C p.(Arg212Thr)	MS	HET	NA	KV	Exostoses, multiple, type 2

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
62	Craniofacial asymmetry	*EFNB1*	NM_004429.4:c.2T>C p.(Met1?)	Oth	HET	XL	5	Craniofrontonasal dysplasia
63	Cleidocranial dysplasia	*RUNX2*	NM_004348.3:c.418G>T p.(Val140Phe)	MS	HET	AD	4	Cleidocranial dysplasia
64	Marfan s.	*FBN1*	NM_000138.4:c.2986T>C p.Cys996Arg	MS	HET	NA	KV	Marfan s.
65	46,XY female, gonadal dysgenesis	*WT1*	NM_024426.4:c.1432+5G>A p.?	SPL	HET	NA	KV	Frasier s.
66	DD, IUGR	*DYRK1A*	NM_001396.3:c.516+1G>A p.?	SPL	HET	DN	5	Mental retar-dation, AD 7
67	Lissencephaly	*VLDLR*	NM_003383.4:c.[1195del];[1666C>T] p.[(Glu399Asnfs*15)];[(Arg556*)]	FS; NS	C-HET	AR	5; 5	Cerebellar hypo-plasia and mental retardation with or without quadrupedal locomotion 1
68	Cardiomyopathy	*LAMP2*	NM_002294.2:c.467T>G p.(Leu156*)	NS	MOS	DN	KV	Danon disease
69	Cardiomyopathy	*MYH7*	NM_000257.3:c.1816G>A p.(Val606Met)	MS	HET	AD	KV	Cardiomyopathy, hypertrophic, 1
70	AD hearing impairment	*TECTA*	NM_005422.2:c.5597C>T p.(Thr1866Met)	MS	HET	AD	KV	Deafness, autosomal dominant 8/12
71	Paroxysmal exercise-induced dyskinesia	*SLC2A1*	NM_006516.2:c.971C>T p.(Ser324Leu)	MS	HET	AD	KV	GLUT1 deficiency s
72	Tuberous sclerosis	*TSC2*	NM_000548.3:c.2436C>G p.(Ser812Arg)	MS	MOS	DN	4	Tuberous sclerosis-2
73	Epilepsy, headaches	*TSC1*	NM_000368.4:c.1888_1891del p.(Lys630Glnfs*22)	FS	HET	DN	KV	Tuberous sclerosis-1
74	DD, progressive ataxia	*ATM*	NM_000051.3:c.[254C>A];[3802del] p.[(Ser85*)];[(Val1268*)]	NS; NS	C-HET	AR	5; KV	Ataxia-telangiectasia

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
75	Paraparesis	*DYNC1H1*	NM_001376.4:c.1869C>G p.(Phe623Leu)	MS	HET	AD	4	Spinal muscular atrophy, lower extremity-pre-dominant 1, AD
76	Aortic aneurysm	*ACTA2*	NM_001613.2:c.511T>A p.(Tyr171Asn)	MS	HET	AD	4	Aortic aneurysm, familial thoracic 6
77	Fetal hydrops	*PTPN11*	NM_002834.3:c.215C>A p.(Ala72Asp)	MS	HET	DN	4	Noonan syndrome 1
78	Tuberous sclerosis	*TSC2*	NM_000548.3:c.1831C>T p.(Arg611Trp)	MS	HET	DN	KV	Tuberous sclerosis-2
79	Myopathy	*RYR1*	NM_000540.2:c.[325C>T];[6721C>T] p.[(Arg109Trp)];[(Arg2241*)]	MS; NS	C-HET	AR	KV; KV	Central core disease
80	Hereditary multiple exostoses	*EXT2*	NM_000401.3:c.635G>C p.(Arg212Thr)	MS	HET	NA	KV	Exostoses, multiple, type 2
81	Pseudoxanthoma elasticum	*ABCC6*	NM_001171.5:c.[1171A>G];[3421C>T] p.[(Arg391Gly)];[(Arg1141*)]	MS; NS	C-HET	AR	KV; KV	Pseudoxanthoma elasticum
82	Pseudoxanthoma elasticum	*ABCC6*	NM_001171.5:c.3421C>T p.(Arg1141*)	NS	HOM	AR	KV	Pseudoxanthoma elasticum
83	CHARGE s.	*CHD7*	NM_017780.3:c.4393C>T p.(Arg1465*)	NS	HET	NA	KV	CHARGE s
84	Hemolytic anemia	*CDAN1*	NM_138477.2:c.[1200C>A];[2808C>A] p.[(Cys400*)];[(Phe936Leu)]	NS; MS	C-HET	AR	5; 4	Dyserythropoietic anemia, congenital, type Ia
85	Muscular dystrophy	*DMD*	NM_004006.2:c.1177C>T p.(Gln393*)	NS	HEM	NA	5	Duchenne muscular dystrophy
86	Tuberous sclerosis	*TSC2*	NM_000548.3:c.5227C>T p.(Arg1743Trp)	MS	HET	NA	KV	Tuberous sclerosis-2
87	Progressive muscular dystrophy	*COL6A1*	NM_001848.2:c.868G>A p.(Gly290Arg)	MS	MOS	DN	KV	Bethlem myopathy 1
88	Agamma-globulinemia	*BTK*	NM_000061.2:c.40T>C p.(Ser14Pro)	MS	HEM	XL	4	Agamma-globulinemia, X-linked 1

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
89	Macular (pattern) dystrophy	*PRPH2*	NM_000322.4:c.515G>A p.(Arg172Gln)	MS	HET	NA	KV	Macular dystrophy, patterned, 1
90	Cranioectodermal dysplasia	*SRCAP*	NM_006662.2:c.7330C>T p.(Arg2444*)	NS	HET	DN	KV	Floating-Harbor s
91	Neurofibromatosis	*NF1*	NM_001042492.2:c.1756_1759del p.(Thr586Valfs*18)	FS	HET	DN	KV	Neurofibromatosis, type 1
92	Haemolytic anaemia	*PKLR*	NM_000298.5:c.[761T>C];[1137_1139del] p.[(Leu254Ser)];[(Lys380del)]	MS; IF	C-HET	AR	4; 5	Pyruvate kinase deficiency
93	DD, microcephaly	Chr2	seq[GRCh37] del(2)(p24.1p23.3) chr2:g.(?_24,300,440)-(25,505,590_?)del	CNV	HET	NA	KV	2p24.1-2p23.3 microdeletion
94	Marfan s.	*FBN1*	NM_000138.4:c.6169C>T p.(Arg2057*)	NS	HET	AD	KV	Marfan s.
95	MODY	*GCK*	NM_000162.3:c.1315_1317del p.(Ile439del)	IF	HET	AD	4	MODY, type II
96	Tuberous sclerosis	*TSC1*	NM_000368.4:c.1525C>T p.(Arg509*)	NS	HET	NA	KV	Tuberous sclerosis-1
97	Lower limb muscular weakness	*LDB3*	NM_001080114.1:c.439G>A p.(Ala147Thr)	MS	HET	NA	KV	Myopathy, myofibrillar, 4
98	Optic nerve atrophy	*RPGRIP1*	NM_020366.3:c.[2554C>T];[(3238+1_3239-1)_(3339+1_3340-1)del] p.[(Arg852*)];[(Asp1080Glyfs*8)] (Arg852*;ex20del)	NS; CNV	C-HET	AR	KV; 5	Leber congenital amaurosis 6
99	Epileptic encephalopathy	*KCNQ2*	NM_172107.2:c.793G>A p.(Ala265Thr)	MS	HET	NA	KV	Epileptic encephalopathy, early infantile, 7
100	DD, ASD, epilepsy	*ANKRD11*	NM_001256182.1:c.2102_2104delinsCC p.(Ser701Thrfs*3)	FS	HET	AD	5	KBG s
101	Congenital myotonia	*CLCN1*	NM_000083.2:c.2680C>T p.(Arg894*)	NS	HOM	AR	KV	Myotonia congenita, AR

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
102	Muscular dystrophy	*DMD*	NM_004006.2:c.10171C>T p.(Arg3391*)	NS	HEM	NA	KV	Duchenne muscular dystrophy
103	Myotonic dystrophy	*ATP2A1*	NM_004320.4:c.[208T>C];[1966C>T] p.[(Cys70Arg)];[(Arg656*)]	MS; NS	C-HET	AR	4; 5	Brody myopathy
104	Leukodystrophy	*DARS2*	NM_018122.4:c.[228-21_228-20delinsC];[492+2T>C] p.[R76SfsX5];[?]	Oth; SPL	C-HET	AR	KV; KV	Leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation
105	DD, facial dysmorphism	*ARID1B*	NM_020732.3:c.5901_5904delinsTCACA p.(Lys1967Asnfs*30)	FS	HET	NA	5	Coffin-Siris syndrome 1
106	Congenital microcephaly, DD	*FOXG1*	NM_005249.4:c.958dup p.(Arg320Profs*135)	FS	HET	NA	5	Rett syndrome, congenital variant
107	Polycystic kidney disease	*PKD1*	NM_000296.3:c.6994_7000dup p.(Val2334Glyfs*88)	FS	HET	AD	KV	Polycystic kidney disease, adult type I
108	Congenital myotonia	*CLCN1*	NM_000083.2:c.2680C>T p.(Arg894*)	NS	HOM	AR	5	Myotonia congenita, AR
109	Hearing impairment	*MYO15A*	NM_016239.3:c.[4597-1G>A];[10181C>T] p.[?];[(Ala3394Val)]	SPL; MS	C-HET	AR	5; KV	Deafness, AR 3
110	Angelman s.	*UBE3A*	NM_000462.3:c.281C>G p.(Ser94*)	NS	HET	IM	5	Angelman s.
111	DD	*DEAF1*	NM_021008.3:c.658G>A p.(Gly220Ser)	MS	HET	DN	4	Mental retardation, AD 24
112	Charcot-Marie-Tooth disease	*MPZ*	NM_000530.7:c.233C>T p.(Ser78Leu)	MS	HET	AD	KV	Charcot-Marie-Tooth disease, type 1B
113	Proximal myopathy, myotonia	*CLCN1*	NM_000083.2:c.2680C>T p.(Arg894*)	NS	HOM	AR	KV	Myotonia congenita, AR
114	EEG abnormality	*NOTCH2*	NM_024408.2:c.5104C>T p.(Arg1702*)	NS	HET	AD	4	Alagille syndrome 2

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
115	Inclusion body myositis	*VCP*	NM_007126.3:c.464G>A p.(Arg155His)	MS	HET	NA	KV	Inclusion body myopathy with early-onset Paget disease and frontotemporal dementia 1
116	Sotos s.	*NSD1*	NM_022455.4:c.4928G>A p.(Cys1643Tyr)	MS	HET	DN	4	Sotos s. 1
117	Arthrogryposis, facial dysmorphism	*TNNI2*	NM_001145829.1:c.486G>C p.(Arg162Ser)	MS	HET	DN	4	Arthrogryposis multiplex congenita, distal, type 2B
118	Charcot-Marie-Tooth disease	*MPZ*	NM_000530.7:c.101C>T p.(Thr34Ile)	MS	HET	AD	KV	Charcot-Marie-Tooth disease, type 1B
119	Hailey-Hailey disease	*ATP2C1*	NM_014382.3:c.2082del p.(Ile695Phefs*13)	FS	HET	NA	5	Hailey-Hailey disease
120	ID, obesity	*KIAA2022*	NM_001008537.2:c.3709dup p.(Met1237Asnfs*24)	FS	HET	DN	5	Mental retardation, XL 98
121	Progressive muscle disorder	*COL6A1*	NM_001848.2:c.877G>A p.(Gly293Arg)	MS	HET	NA	KV	Collagen VI myopathy
122	Muscular disorder	*EMD*	NM_000117.2:c.650_654del p.(Leu217Profs*31)	FS	HEM	NA	5	Emery-Dreifuss muscular dystrophy 1, XL
123	Congenital myotonia	*CLCN1*	NM_000083.2:c.2680C>T p.(Arg894*)	NS	HOM	AR	KV	Myotonia congenita, AR
124	Cardiomyopathy	*MYH7*	NM_000257.3:c.[1447G>A];[5786C>T] p.[(Glu483Lys)];[(Thr1929Met)]	MS; MS	HET/ C-HET	AD/ AR	KV; KV	Cardio-myopathy, hypertrophic, 1
125	Tuberous sclerosis	*TSC1*	NM_000368.4:c.2054C>A p.(Ser685*)	NS	HET	NA	5	Tuberous sclerosis-1
126	DD, congenital liver disease	chrX	45,X	Oth	–	–	5	Turner s.

#	Phenotype/ syndrome	Gene	Mutation	Type	Zyg	Inh	Class	Final diagnosis
127	Leukodystrophy	*ARSA*	NM_000487.5:c.[542T>G];[769G>C] p.[(Ile181Ser)];[(Asp257His)]	MS; MS	C-HET	AR	KV; KV	Metachromatic leukodystrophy
128	Bleeding disorder	*VWF*	NM_000552.4:c.[2561G>A];[2372C>T] p.[(Arg854Gln)];[(Thr791Met)]	MS; MS	C-HET	AR	KV; KV	von Willebrand disease, type 2N
129	Seizures, hypotonia	*SLC6A8*	NM_005629.3:c.321_323del p.(Phe107del)	IF	HET	XL	KV	Cerebral creatine deficiency syndrome 1
130	Heart defect, cachexia	Chr22	seq[GRCh37] del(22)(q11.2) chr22:g.(?_18,899,041)_(21,304,140_?)del	CNV	HET	AD	KV	DiGeorge syndrome
131	Neuropathy in all limbs	*PMP22*	seq[GRCh37] del(17)(p12) chr17:g.(?_14,095,294)_(15,164,054_?)del (PMP22del)	CNV	HET	NA	KV	Hereditary neuropathy with liability to pressure palsies
132	Mucopoly-saccharidosis II	*IDS*	NM_000202.5:c.22C>T p.(Arg8*)	NS	HEM	NA	KV	Mucopoly-saccharidosis II

5.3. *KPTN*-related intellectual disability-macrocephaly syndrome (Paper III)

In the brother and sister with ID and macrocephaly, no significant CNVs were detected by CMA. However, two LCSHs greater than 5 Mb were found in the brother: chr1:182,559,510-200,136,582 and chr14:37,290,124-44,790,527 (hg19). The sister had only one large LCSH that was identical to the brother's LCSH in chromosome 14. Although WES from the brother's DNA sample did not reveal any mutations associated with the patient's phenotype in the shared LCSH, further search for rare homozygous deleterious mutations led to the discovery of a one-nucleotide duplication in exon 7 of the *KPTN* gene predicted to result in a frameshift: c.665dupA p.(Ser223Glnfs*18) (RefSeq NM_007059.2). The variant was not present in ExAC or ClinVar databases. Less than a year before this analysis was performed, *KPTN* mutations were associated with Mendelian disorders for the first time in a large Anabaptist kindred with many affected by ID-macrocephaly syndrome [Baple et al. 2014]. Sanger sequencing confirmed the homozygous variant in both affected sibs and heterozygous variant in both parents. Reanalysis of the CMA data revealed a 1.5 Mb LCSH encompassing the *KPTN* gene (Figure 7). Based on the occurrence of multiple LCSHs, we suspect distant parental consanguinity.

Baple et al. conducted extensive functional studies showing that both founder mutations causing the genetic disorder in the Anabaptist population result in loss of function of kaptin protein [Baple et al. 2014]. Kaptin associates with dynamic actin cytoskeletal structures in neurons and the loss of the association caused by loss-of-function mutations leads to neurodevelopmental alterations [Baple et al. 2014]. The authors proposed that *KPTN* mutations result in a clinically distinctive syndrome with the core phenotype of macrocephaly, global DD, behavioural abnormalities, and seizures [Baple et al. 2014]. Including patients, we can conclude that the cardinal features of the syndrome are macrocephaly and ID. It is also likely that there is an increased risk for seizures and behavioural problems, but these are not fully penetrant features. Based on our patients, the intellectual and language abilities tend to be more severe than motor development. Our patients and the Anabaptist patients did not share any dysmorphic features besides a prominent forehead or frontal bossing, which could be associated with macrocephaly. Like previously reported patients [Baple et al. 2014], ours did not have any documented hearing problems, thus the hypothesis about *KPTN* being a candidate gene for hearing loss [Bearer et al. 2000] is not supported. We also support the hypothesis that the *KPTN*-related syndrome is widespread and not restricted to the Anabaptist population [Baple et al. 2014].

The recent discovery of the KICSTOR protein complex links *KPTN*-related disorders to other mTOR-pathway associated disorders like tuberous sclerosis [Wolfson et al. 2017]. There are some clinical similarities in the described siblings to other mTOR-related disorders, mainly macrocephaly and increased risk for seizures and neurodevelopmental problems. As the KICSTOR complex

serves as negative regulator for the mTOR signalling pathway, *KPTN* LoF mutations may result in hyperactivation of the mTOR pathway. It can thus be hypothesized that rapamycin and similar known mTOR signalling inhibitors could have a positive effect on patients with *KPTN*-related disorders similarly to those affected by tuberous sclerosis. However, more studies investigating the pathogenetic mechanisms of *KPTN*-related disorders are needed before considering drug trials. Also, by the time of writing this thesis, the reported *KPTN*-related syndrome cases were limited to eleven patients from two kindreds and three pathogenic mutations, which is undoubtedly too few to make definitive conclusions on either the clinical spectrum of the disorder or the molecular effects of the mutations. Animal models would serve as excellent research subjects, supported by the macrocephaly present in *Kptn* knock-out mice (IMPC, personal communication, unpublished data) thus resembling the human phenotype. Further clinical reports, as well as studies on kaptin function and the mutation effects are needed to further delineate the *KPTN*-related syndrome.

Identification of the disease-causing mutation in the *KPTN* gene was straightforward in this case due to suspicion towards homozygous mutations based on LCSHs detected by CMA, and the ability of WES to detect virtually all rare coding homozygous mutations. Nevertheless, it must be noted that if the analysis would have been performed before the first report of *KPTN*-related syndrome [Baple et al. 2014], the conclusive interpretation would have been much more difficult to make. Also, if TruSight One panel (Illumina Inc.), which is widely used in our centre as a diagnostic test, would have been used in this patient, the case would have not been solved since the *KPTN* gene is not covered in the panel. This justifies the use of WES as the preferred assay in research settings.

5.4. Molecular and clinical phenotype of a novel *MYH7* mutation detected in a boy with congenital myopathy (Paper IV)

To investigate the molecular cause of a muscle disorder in a boy with congenital myopathy and fiber type disproportion, WES of the parents-offspring trio was carried out on the clinical diagnostic indications. *De novo* synonymous c.5655G>A, p.(Ala1885=) transition of the last nucleotide in exon 38 of *MYH7* gene (RefSeq NM_000257.3) was identified. The mutation was confirmed by Sanger sequencing. *In silico* analysis using both MutationTaster [Schwarz et al. 2014] and MutPred Splice [Mort et al. 2014] predicted that the variant disrupts splicing.

To clarify the mutation's effect on splicing, RNA studies were performed. After cDNA synthesis, PCR amplification using aforementioned primers detected three fragments with different sizes by agarose gel electrophoresis as expected for disrupted splicing: fragment #1 (393 bp) and #2 (297 bp) with

primers MYH7_F1 and MYH7_R1, and fragment #3 (199 bp) with primers MYH7_F2 and MYH7_R1. Sanger sequencing of extracted gel fragments #2 and #3 confirmed the skipping of exon 38 in fragment #2 (Figure 9). The change at the mRNA level can be described as NM_000257.3:r.5560_5655del and predicted on the protein level as an in-frame deletion NP_000248:p.(1854_1885del).

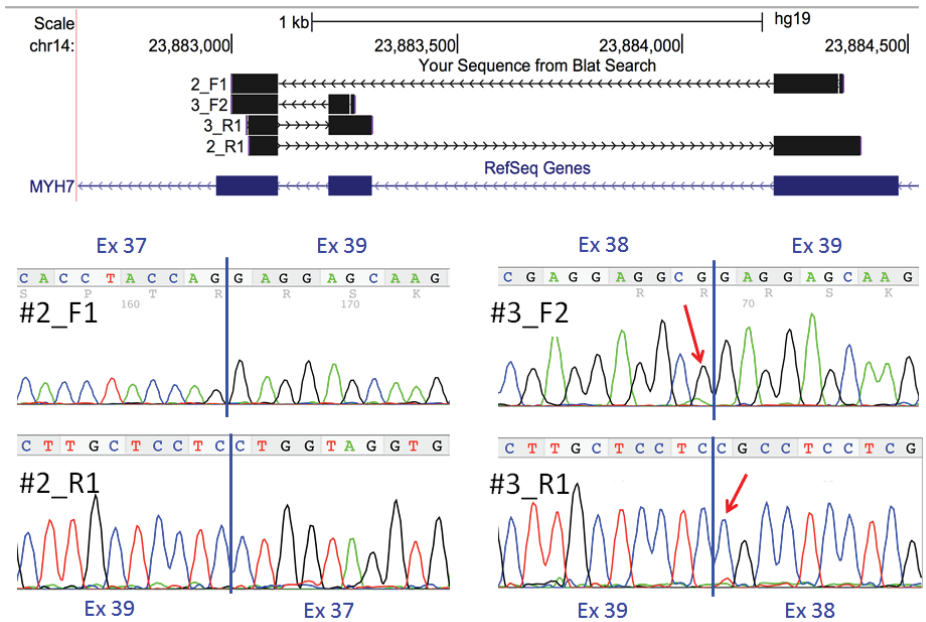


Figure 9. Results of RNA studies confirming exon-skipping. The upper part of the figure shows the alignment of muscle tissue cDNA PCR products (fragments #2 and #3) with the *MYH7* gene. The sequences of the exon-exon boundaries from the same fragments (#2 and #3) are presented in the lower part. It can be clearly seen that in fragment #2, exon 38 has been skipped, and the mutation detected at the DNA level is not present in fragment #3 where the c.5655G position is indicated with an arrow. Figure was prepared by Dr. Tarmo Annilo.

The *MYH7*-releated myopathies represent a wide spectrum of clinical phenotypes caused by mutations in the myosin-tail domain of the gene. In the era of genome-wide sequencing, the range of gene-associated phenotypes is predicted to widen even more because genetic testing does not focus only on known phenotype-genotype associations [Komlosi et al. 2014]. Thus, one can suppose that phenotypes once reported as distinct will show more overlap as new patients are reported [Ortolano et al. 2011].

The patient reported here presented early in infancy with muscular hypotonia and weakness in proximal muscles of the lower limbs and trunk (positive Gower's sign) and extremely prominent weakness in the neck extensor muscles (positive dropped head sign). It is important to emphasise that up to almost

2 years of age the disease has not progressed; on the contrary, his motor abilities are improving along with only lagging motor head control. His phenotype correlates more with MSM than LDM. Interestingly, exaggerated deep tendon reflexes were present in our patient, although the other patients with *MYH7*-related CFTD/MSM show absent or hypoactive tendon reflexes [Sobrido et al. 2005]. At 1 year and 9 months he had no cardiac abnormalities detected by echocardiography, consistent with a low prevalence of cardiomyopathy in *MYH7*-related skeletal muscle disorders [Tajsharghi and Oldfors 2013]. Because no myosin storage was found on muscle biopsy, we cannot diagnose the patient as having MSM. CFTD is not suggested as suitable for formal diagnosis because it represents a syndrome [Clarke 2011], but our case does not fit the diagnosis of either MSM or LDM. We therefore propose the general term *MYH7*-related myopathy to be the most accurate for cases like ours.

Although more severely affected, our patient shows clinical similarity with the patients from one family first described by Sobrido et al. [2005] and later with genetic investigations by Ortolano et al. [2011]. Among affected members of the kindred, myosin storage was seen only in the oldest patient while the younger patients had isolated CFTD, leading to the hypothesis that CFTD is an early histopathologic sign of MSM [Ortolano et al. 2011], which we cannot rule out in our patient either because the muscle biopsy was taken during infancy. We should note that our patient is one of the youngest ones reported and this could also explain some differences seen in the phenotype.

The described novel pathogenic mutation was the first single-nucleotide substitution in the *MYH7* gene demonstrated to cause in-frame skipping of exon 38, thus shortening the encoded protein by 32 amino acids. The unique type of mutation could be the cause for the clinical differences and relatively more severe phenotype compared to those of previously reported patients. After publishing our case, two next patients with mutations leading to skipping of exon 38 were described [Fiorillo et al. 2016]. Although the effect on mRNA was the same, the phenotypes differed between patients. Similarly to our patient, the dropped head sign due to axial muscles involvement was the leading feature in one case, but together with heart dysfunction; the other patient, however, presented with infantile onset of respiratory muscle impairment [Fiorillo et al. 2016]. Interestingly, neither of the two other patients had any findings of fibre type disproportion on muscle biopsy [Fiorillo et al. 2016]. Our finding that synonymous mutations disrupt splicing indicates that caution is needed when filtering out synonymous changes during WES data interpretation. Also, this report illustrates the importance of the RNA studies in the final interpretation of some cases of diagnostic WES where the clinical significance of detected mutations is difficult to predict. It is worth noting that although the locus of the reported transition is highly conserved in mammals, the dog has reference nucleotide A instead of the human reference G at this position (according to the Dog Sep. 2011 [Broad CanFam3.1/canFam3] assembly and transcript ENSCAFT00000027883). This appears not to cause altered splicing in dogs according to numerous mRNA and Expressed Sequence Tag (EST) sequences available in the GenBank.

The demonstrated skipping of exon 38 removes half of the assembly competence domain for the slow myosin protein [Sohn et al. 1997], indicating the phenotype seen in the reported patient could be caused by impaired myosin assembly into thick filaments. If the deletion completely prevented myosin dimerization, the pathogenetic mechanism could be described as haploinsufficiency. Loss-of-function mutations appear in the ExAC database with an allele frequency of approximately 1:3000, which might indicate that haploinsufficiency of the *MYH7* is not pathogenic. Nevertheless, both frameshift and nonsense mutations in the *MYH7* gene have been shown to cause cardiomyopathy [Waldmuller et al. 2011]. Also, as the described exon-skipping mutation is in-frame and does not remove the whole assembly competence domain, it probably does not result in complete loss of function, and rather is similar to in-frame deletions seen in Becker muscular dystrophy [Aartsma-Rus et al. 2006].

The clinical phenotype and the novel exon-skipping mutation of the patient expand the knowledge about *MYH7*-related disorders. Therefore, we advise suspecting mutations in the *MYH7* gene in patients with a dropped head sign and general hypotonia. More reports on patients with exonic deletions are needed to clarify the role of intragenic deletions in the muscular phenotype. Further functional studies on myosin can bring insight into pathogenetic mechanisms arising from proteins with in-frame deletions.

6. CONCLUSIONS

1. The results of a study concentrating on the diagnostic utility of single long contiguous stretches of homozygosity in patients without parental consanguinity based on a routine clinical diagnostic laboratory's CMA data (2,110 patients) from a four-year period were reported.
   1.1 In Estonia, 8.2% of patients carry one or two LCSH over 5 Mb in size, making LCSHs the most frequently reported class of VUSs from CMA.
   1.2 Six different recurrent LCSHs were identified resulting in 52 events or 28.9% of all isolated LCSH findings. Thus, we could clarify the clinical significance of nearly one-third of all LCSHs by classifying them as likely benign. Out of six recurrent LCSH-regions, only two were previously described in other populations, indicating the population specificity.
   1.3 Out of 129 non-recurrent LCSHs, two were identified to encompass a candidate gene for AR disorder matching with the clinical phenotype. In both cases, the homozygous disease-causing variant in the candidate genes were identified. Although finding a good candidate gene from single LCSHs is a rare event, systematic analysis of the genes inside LCSH can provide valuable clues for candidate gene prioritization.
   1.4 If no good candidate genes are identified in LCSH, it may still increase the probability of detecting homozygous mutations due to possible distant relatedness of parents, as illustrated by the case of siblings carrying a homozygous frameshift mutation in the *KPTN* gene.
2. Based on our centre's experience from the first 501 diagnostic cases in a whom large panel of 4,813 genes was sequenced and subsequently analysed based on virtual subpanels as requested by referring doctors, it can be concluded that large gene panel sequencing can make use of many capabilities of NGS for detection of a large variety of mutations.
   2.1 Out of 501 patients, 132 received a conclusive diagnosis after large gene panel sequencing. Thus, the diagnostic yield was 26.3%, making it comparable with large WES cohorts previously reported in the literature.
   2.2 From all reported disease-causing variants, 46% were novel (not reported before), advocating for sequencing whole coding regions of genes instead of testing for only known pathogenic variants in the diagnostics of rare disorders.
   2.3 The most common class of pathogenic variants were SNVs and small indels (94% of reported variants). However, as CNV calling was performed from NGS data, additional 7 cases were diagnosed either by detecting the single causative deletion or duplication or detecting CNV in compound heterozygous state with pathogenic SNV.
   2.4 The subgroup of patients in whom less than ten genes were requested to be analysed showed a significantly higher yield than the larger panel subgroup (31.5% vs 21.7%, $p = 0.015$). This indicates that good clinical

phenotyping can efficiently guide the selection of patients most likely to benefit from sequencing studies.

2.5 Our experience from a relatively small clinical laboratory demonstrates that large gene panels can be used for many different diagnostic scenarios in a cost- and time-efficient manner.

3. Clinical phenotype and molecular findings from two adult siblings with ID caused by a homozygous variant in the *KPTN* gene were described. This delineates the core phenotype of this novel AR genetic syndrome.

3.1 CMA revealed shared LCSH in both siblings. WES was performed in the brother to search for rare homozygous variants. A novel homozygous frameshift mutation c.665dupA p.(Ser223Glnfs*18) in the *KPTN* gene was detected, not encompassed in the shared LCSH. Sanger sequencing confirmed the sister as a homozygote for this variant as well.

3.2 This was the second case of *KPTN*-related ID-syndrome reported in the medical literature after the first Anabaptist kindred. Thus, we support the hypothesis of the first-describing authors that the *KPTN*-related syndrome is widespread and not restricted to Anabaptist population.

3.3 The cardinal features of the *KPTN*-related AR syndrome are macrocephaly and ID. Seizures and behavioural problems seem to be more frequent than in the general population, but these are not fully penetrant features. The intellectual and language disorders tend to be more severe than motor development.

4. Molecular investigations in a boy with congenital myopathy and fiber-type disproportion revealed a novel *MYH7* variant leading to the exon skipping in the mRNA, thus expanding the genotypic and phenotypic spectra of *MYH7*-related myopathies.

4.1 WES of the parents-offspring trio carried out on clinical diagnostic indications identified a novel *de novo* synonymous c.5655G>A, p.(Ala1885=) transition of the last nucleotide in exon 38. This mutation was confirmed by Sanger sequencing.

4.2 The studies performed on RNA extracted from the muscle tissue confirmed the molecular effect of the mutation as skipping of exon 38. This was the first report of a patient with an exon-skipping mutation in the *MYH7* gene, as other pathogenic mutations were missense or in-frame deletions of single amino acids.

4.3 The phenotypic spectrum of *MYH7*-related phenotypes is broad and was even further expanded by this report. Thus, the *MYH7* gene should be included in the candidate gene list in patients with a dropped head sign and general hypotonia.

4.4 From a diagnostics perspective, this case highlighted the needed caution in filtering out synonymous changes during WES data interpretation. Moreover, as illustrated here, the collaboration with research institutions with capabilities for RNA as well as other molecular investigations is crucial in solving some challenging diagnostic WES cases.

This study demonstrated the usefulness of genome-wide analyses for many diagnostic scenarios and revealed the diagnostic yields of different genetic tests in clinical settings. The implementation of modern NGS-techniques and CMA into clinical practice has greatly improved the diagnostics of Mendelian disorders in Estonia. The knowledge gained from this study can be further used to improve diagnostic algorithms for different diseases.

REFERENCES

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al*. 2015. A global reference for human genetic variation. Nature 526(7571):68–74.

Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJ, Den Dunnen JT. 2006. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve 34(2):135–144.

ACMG Board of Directors. 2012. Points to consider in the clinical application of genomic sequencing. Genet Med 14(8):759–761.

ACMG Board of Directors. 2015. Clinical utility of genetic and genomic services: a position statement of the American College of Medical Genetics and Genomics. Genet Med 17(6):505–507.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249.

Ahn JW, Bint S, Bergbaum A, Mann K, Hall RP, Ogilvie CM. 2013. Array CGH as a first line diagnostic test in place of karyotyping for postnatal referrals – results from four years' clinical application for over 8,700 patients. Mol Cytogenet 6(1):16.

Akinrinade O, Ollila L, Vattulainen S, Tallila J, Gentile M, Salmenpera P, Koillinen H, Kaartinen M, Nieminen MS, Myllykangas S *et al*. 2015. Genetics and genotype-phenotype correlations in Finnish patients with dilated cardiomyopathy. Eur Heart J 36(34):2327–2337.

Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. Nat Rev Genet 12(5):363–376.

Alkuraya FS. 2010. Autozygome decoded. Genet Med 12(12):765–771.

Antonarakis SE, Beckmann JS. 2006. Mendelian disorders deserve more attention. Nature Reviews Genetics 7(4):277–282.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29.

Baple EL, Maroofian R, Chioza BA, Izadi M, Cross HE, Al-Turki S, Barwick K, Skrzypiec A, Pawlak R, Wagner K *et al*. 2014. Mutations in *KPTN* cause macro-cephaly, neurodevelopmental delay, and seizures. Am J Hum Genet 94(1):87–94.

Bearer EL. 1992. An actin-associated protein present in the microtubule organizing center and the growth cones of PC-12 cells. J Neurosci 12(3):750–761.

Bearer EL, Abraham MT. 1999. 2E4 (kaptin): a novel actin-associated protein from human blood platelets found in lamellipodia and the tips of the stereocilia of the inner ear. Eur J Cell Biol 78(2):117–126.

Bearer EL, Chen AF, Chen AH, Li Z, Mark HF, Smith RJ, Jackson CL. 2000. 2E4/Kaptin (*KPTN*) – a candidate gene for the hearing loss locus, DFNA4. Ann Hum Genet 64(Pt 3):189–196.

Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tuysuz B, Caglayan AO, Gokben S *et al*. 2010. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature 467(7312):207–210.

Bittles A. 2001. Consanguinity and its relevance to clinical genetics. Clin Genet 60(2):89–98.

Bowles Biesecker B, Marteau TM. 1999. The future of genetic counselling: an international perspective. Nat Genet 22(2):133–137.

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 14(10):681–691.

Brownstein CA, Beggs AH, Homer N, Merriman B, Yu TW, Flannery KC, DeChene ET, Towne MC, Savage SK, Price EN *et al*. 2014. An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge. Genome Biol 15(3):R53.

Burrows M, Wheeler DJ. 1994. A block-sorting lossless data compression algorithm. Technical report 124 Palo Alto, CA: Digital Equipment Corporation.

Chang F, Li MM. 2013. Clinical application of amplicon-based next-generation sequencing in cancer. Cancer Genet 206(12):413–419.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S *et al*. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106(45):19096–19101.

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012a. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front Genet 3:35.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80–92.

Clarke NF. 2011. Congenital fiber-type disproportion. Semin Pediatr Neurol 18(4):264–271.

Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, Zackai EH, Deardorff MA, Krantz ID, Hakonarson H, Spinner NB. 2010. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum Mol Genet 19(7):1263–1275.

Cornish A, Guda C. 2015. A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference. Biomed Res Int 2015:456479.

Coughlin CR, 2nd, Scharer GH, Shaikh TH. 2012. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. Genome Med 4(10):80.

Crick F. 1970. Central dogma of molecular biology. Nature 227(5258):561–563.

Crick FH. 1958. On protein synthesis. Symp Soc Exp Biol. p 8.

D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, Hanson E, Maillard AM, Faucett WA, Mace A *et al*. 2016. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA Psychiatry 73(1):20–30.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C *et al*. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 367(20):1921–1929.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491–498.

Donaudy F, Snoeckx R, Pfister M, Zenner HP, Blin N, Di Stazio M, Ferrara A, Lanzara C, Ficarella R, Declau F *et al*. 2004. Nonmuscle myosin heavy-chain gene MYH14 is expressed in cochlea and mutated in patients affected by autosomal dominant hearing impairment (DFNA4). Am J Hum Genet 74(4):770–776.

Eggermann T, Perez de Nanclares G, Maher ER, Temple IK, Tumer Z, Monk D, Mackay DJ, Gronskov K, Riccio A, Linglart A *et al*. 2015. Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. Clin Epigenetics 7:123.

Ellingford JM, Barton S, Bhaskar S, O'Sullivan J, Williams SG, Lamb JA, Panda B, Sergouniotis PI, Gillespie RL, Daiger SP *et al*. 2016. Molecular findings from 537 individuals with inherited retinal disease. J Med Genet.

Emanuel BS, Saitta SC. 2007. From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. Nat Rev Genet 8(11):869–883.

Engel E. 1980. A new genetic concept: uniparental disomy and its potential effect, isodisomy. Am J Med Genet 6(2):137–143.

Engel E. 2006. A fascination with chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements. Eur J Hum Genet 14(11):1158–1169.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, Baxter RM, Zeng W, Mroske C, Parra MC *et al*. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. Genet Med 17(7):578–586.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet 7(2):85–97.

Fiorillo C, Astrea G, Savarese M, Cassandrini D, Brisca G, Trucco F, Pedemonte M, Trovato R, Ruggiero L, Vercelli L *et al*. 2016. MYH7-related myopathies: clinical, histopathological and imaging findings in a cohort of Italian patients. Orphanet J Rare Dis 11(1):91.

Garrod A. 1902. The incidence of alkaptonuria: a study in chemical individuality. The Lancet 160(4137):1616–1620.

Ghoneim DH, Myers JR, Tuttle E, Paciorkowski AR. 2014. Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. BMC Res Notes 7:864.

Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG *et al*. 2010. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. Am J Hum Genet 87(3):418–423.

Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A *et al*. 2014. Genome sequencing identifies major causes of severe intellectual disability. Nature 511(7509):344–347.

Gilissen C, Hoischen A, Brunner HG, Veltman JA. 2012. Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20(5):490–497.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333–351.

Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, Sano Marma M, Meng Q, Cao H, Li X *et al*. 2008. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proc Natl Acad Sci U S A 105(27):9145–9150.

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY *et al*. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306(5940):234–238.

Haer-Wigman L, van Zelst-Stams WA, Pfundt R, van den Born LI, Klaver CC, Verheij JB, Hoyng CB, Breuning MH, Boon CJ, Kievit AJ *et al*. 2017. Diagnostic exome sequencing in 266 Dutch patients with visual impairment. Eur J Hum Genet 25(5):591–599.

Hehir-Kwa JY, Marschall T, Kloosterman WP, Francioli LC, Baaijens JA, Dijkstra LJ, Abdellaoui A, Koval V, Thung DT, Wardenaar R *et al*. 2016. A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. Nat Commun 7:12989.

Hehir-Kwa JY, Pfundt R, Veltman JA. 2015. Exome sequencing and whole genome sequencing for the detection of copy number variation. Expert Rev Mol Diagn 15(8):1023–1032.

Heinrichs S, Look AT. 2007. Identification of structural aberrations in cancer by SNP array analysis. Genome Biol 8(7):219.

Helbig KL, Farwell Hagman KD, Shinde DN, Mroske C, Powis Z, Li S, Tang S, Helbig I. 2016. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet Med 18(9):898–905.

Hochstenbach R, van Binsbergen E, Engelen J, Nieuwint A, Polstra A, Poddighe P, Ruivenkamp C, Sikkema-Raddatz B, Smeets D, Poot M. 2009. Array analysis and karyotyping: workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. Eur J Med Genet 52(4):161–169.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ *et al*. 2007. Genome-wide in situ exon capture for selective resequencing. Nat Genet 39(12):1522–1527.

Hwang S, Kim E, Lee I, Marcotte EM. 2015. Systematic comparison of variant calling pipelines using gold standard personal exome variants. Sci Rep 5:17875.

Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, Martinet D, Shen Y, Valsesia A, Beckmann ND *et al*. 2011. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 478(7367):97–102.

Johansson LF, van Dijk F, de Boer EN, van Dijk-Bos KK, Jongbloed JD, van der Hout AH, Westers H, Sinke RJ, Swertz MA, Sijmons RH *et al*. 2016. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. Hum Mutat 37(5):457–464.

Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, Li X, Marma MS, Shi S, Wu J *et al*. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci U S A 103(52):19635–19640.

Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE. 2011. Detection of structural variants and indels within exome data. Nat Methods 9(2):176–178.

Katsanis SH, Katsanis N. 2013. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 14(6):415–426.

Kearney HM, Kearney JB, Conlin LK. 2011a. Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. Clin Lab Med 31(4):595–613, ix.

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance C. 2011b. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 13(7):680–685.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46(3):310–315.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25(17):2283–2285.

Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GC, Brown DL, Brudno M, Campbell J *et al*. 2014. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 42(Database issue):D966–974.

Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM *et al*. 2017. The Human Phenotype Ontology in 2017. Nucleic Acids Res 45(D1):D865–D876.

Kojima K, Nariai N, Mimori T, Takahashi M, Yamaguchi-Kabata Y, Sato Y, Nagasaki M. 2013. A statistical variant calling approach from pedigree information and local haplotyping with phase informative reads. Bioinformatics 29(22):2835–2843.

Komlosi K, Hadzsiev K, Garbes L, Martinez Carrera LA, Pal E, Sigurethsson JH, Magnusson O, Melegh B, Wirth B. 2014. Exome sequencing identifies Laing distal myopathy MYH7 mutation in a Roma family previously diagnosed with distal neuronopathy. Neuromuscul Disord 24(2):156–161.

Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Project NES, Quinlan AR, Nickerson DA, Eichler EE. 2012. Copy number variation detection and genotyping from exome sequence data. Genome Res 22(8):1525–1532.

Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He ZX *et al*. 2015. Excess of rare, inherited truncating mutations in autism. Nat Genet 47(6):582–588.

Ku CS, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R. 2012. Exome sequencing: dual role as a discovery and diagnostic tool. Ann Neurol 71(1):5–14.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4(7):1073–1081.

Laing NG, Laing BA, Meredith C, Wilton SD, Robbins P, Honeyman K, Dorosz S, Kozman H, Mastaglia FL, Kakulas BA. 1995. Autosomal dominant distal myopathy: linkage to chromosome 14. Am J Hum Genet 56(2):422–427.

Lamont PJ, Wallefeld W, Hilton-Jones D, Udd B, Argov Z, Barboi AC, Bonneman C, Boycott KM, Bushby K, Connolly AM *et al*. 2014. Novel mutations widen the phenotypic spectrum of slow skeletal/beta-cardiac myosin (MYH7) distal myopathy. Hum Mutat 35(7):868–879.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J *et al*. 2016. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44(D1):D862–868.

Laplante M, Sabatini DM. 2012. mTOR signaling in growth control and disease. Cell 149(2):274–293.
Laurie S, Fernandez-Callejo M, Marco-Sola S, Trotta JR, Camps J, Chacon A, Espinosa A, Gut M, Gut I, Heath S *et al*. 2016. From Wet-Lab to Variations: Concordance and Speed of Bioinformatics Pipelines for Whole Genome and Whole Exome Sequencing. Hum Mutat 37(12):1263–1271.
Leal T, Andrieux J, Duban-Bedu B, Bouquillon S, Breviere GM, Delobel B. 2009. Array-CGH detection of a de novo 0.8Mb deletion in 19q13.32 associated with mental retardation, cardiac malformation, cleft lip and palate, hearing loss and multiple dysmorphic features. Eur J Med Genet 52(1):62–66.
Lee C, Iafrate AJ, Brothman AR. 2007. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet 39(7 Suppl):S48–54.
Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M *et al*. 2014. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 312(18):1880–1887.
Leffler M, Puusepp S, Žilina O, Zhu Y, Kuuse K, Bain N, Burgess T, Õunap K, Field M. 2016. Two familial microduplications of 15q26.3 causing overgrowth and variable intellectual disability with normal copy number of IGF1R. Eur J Med Genet 59(4):257–262.
Lejeune J, Gauthier M, Turpin R. 1959a. Human chromosomes in tissue cultures. Comptes rendus hebdomadaires des séances de l'Académie des sciences 248(4):602.
Lejeune J, Turpin R, Gautier M. 1959b. Le mongolisme, premier exemple d'aberration autosomique humaine. Ann Genet 1(4):1–49.
Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB *et al*. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536(7616):285–291.
Li H. 2012. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. Bioinformatics 28(14):1838–1844.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.
Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364.
Maasalu K, Nikopensius T, Koks S, Noukas M, Kals M, Prans E, Zhytnik L, Metspalu A, Martson A. 2015. Whole-exome sequencing identifies de novo mutation in the COL1A1 gene to underlie the severe osteogenesis imperfecta. Hum Genomics 9:6.
MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N *et al*. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72(6):971–983.
Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S, Dukart J, Ferrari C, Conus P, Mannik K *et al*. 2015. The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. Mol Psychiatry 20(1):140–147.
Makrythanasis P, Nelis M, Santoni FA, Guipponi M, Vannier A, Bena F, Gimelli S, Stathaki E, Temtamy S, Megarbane A *et al*. 2014. Diagnostic exome sequencing to

elucidate the genetic basis of likely recessive disorders in consanguineous families. Hum Mutat 35(10):1203–1210.
Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al*. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380.
Maron BJ, Maron MS, Semsarian C. 2012. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. J Am Coll Cardiol 60(8):705–715.
Marschall T, Hajirasouliha I, Schonhuth A. 2013. MATE-CLEVER: Mendelian-inheritance-aware discovery and genotyping of midsize and long indels. Bioinformatics 29(24):3143–3150.
Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, Race V, Sistermans E, Sturm M, Weiss M *et al*. 2016. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet 24(1):2–5.
McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, Donnelly P. 2014. Choice of transcripts and software has a large effect on variant annotation. Genome Med 6(3):26.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303.
McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. Genome Biol 17(1):122.
McNally EM, Golbus JR, Puckelwartz MJ. 2013. Genetic mutations and mechanisms in dilated cardiomyopathy. J Clin Invest 123(1):19–26.
McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A *et al*. 2008. Runs of homozygosity in European populations. Am J Hum Genet 83(3):359–372.
Mendel J. 1866. Versuche über Plflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865. Abhandlungen: 3–47. Translated in: Druery, CT, and Bateson, W. 1901. Experiments in plant hybridization:1–32.
Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ *et al*. 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 86(5):749–764.
Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16(9):1182–1190.
Mort M, Sterne-Weiler T, Li B, Ball EV, Cooper DN, Radivojac P, Sanford JR, Mooney SD. 2014. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol 15(1):R19.
Muelas N, Hackman P, Luque H, Garces-Sanchez M, Azorin I, Suominen T, Sevilla T, Mayordomo F, Gomez L, Marti P *et al*. 2010. MYH7 gene tail mutation causing myopathic profiles beyond Laing distal myopathy. Neurology 75(8):732–741.
Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology: Cold Spring Harbor Laboratory Press. p 263–273.

Männik K. 2012. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. *Dissertationes Biologicae Universitatis Tartuensis 221*.Tartu: Tartu University Press.

Männik K, Mägi R, Mace A, Cole B, Guyatt AL, Shihab HA, Maillard AM, Alavere H, Kolk A, Reigo A *et al*. 2015. Copy number variations and cognitive phenotypes in unselected populations. JAMA 313(20):2044–2054.

Männik K, Parkel S, Palta P, Žilina O, Puusepp H, Esko T, Mägi R, Nõukas M, Veidenberg A, Nelis M *et al*. 2011. A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia. Eur J Med Genet 54(2):136–143.

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P *et al*. 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478(7367):57–63.

Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, McDonald MT, Meisler MH, Goldstein DB. 2012. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet 49(6):353–361.

Neill NJ, Torchia BS, Bejjani BA, Shaffer LG, Ballif BC. 2010. Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. Mol Cytogenet 3:11.

Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, Rodenburg RJ, Yntema HG, Spruijt L, Vermeer S *et al*. 2013. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat 34(12):1721–1726.

Neveling K, Tjwan Thung D, Beulen L, van Rens-Buijsman W, Gomes I, van den Heuvel S, Mieloo H, Derks-Prinsen I, Kater-Baats E, Faas BH. 2016. Validation of two-channel sequencing-by-synthesis for noninvasive prenatal testing of fetal whole and partial chromosome aberrations. Prenat Diagn 36(3):216–223.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC *et al*. 2010a. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42(9):790–793.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA *et al*. 2010b. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42(1):30–35.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461(7261):272–276.

Nicholls RD, Knoll JH, Butler MG, Karam S, Lalande M. 1989. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. Nature 342(6247):281–285.

Nicholls RD, Pai GS, Gottlieb W, Cantu ES. 1992. Paternal uniparental disomy of chromosome 15 in a child with Angelman syndrome. Ann Neurol 32(4):512–518.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12(6):443–451.

Nirenberg M, Leder P. 1964. RNA Codewords and Protein Synthesis. The Effect of Trinucleotides upon the Binding of sRNA to Ribosomes 145(3639):1399–1407.

O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE *et al*. 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med 5(3):28.

Oneda B, Rauch A. 2017. Microarrays in prenatal diagnosis. Best Pract Res Clin Obstet Gynaecol.

Ortolano S, Tarrio R, Blanco-Arias P, Teijeira S, Rodriguez-Trelles F, Garcia-Murias M, Delague V, Levy N, Fernandez JM, Quintans B *et al*. 2011. A novel MYH7 mutation links congenital fiber type disproportion and myosin storage myopathy. Neuromuscul Disord 21(4):254–262.

Ozyilmaz B, Kirbiyik O, Koc A, Ozdemir TR, Kaya OO, Guvenc MS, Erdogan KM, Kutbay YB. 2017. Experiences in microarray-based evaluation of developmental disabilities and congenital anomalies. Clin Genet.

Pajusalu S, Kahre T, Roomere H, Murumets Ü, Roht L, Simenson K, Reimand T, Õunap K. 2017. Large Gene Panel Sequencing in Clinical Diagnostics – Results from 501 Consecutive Cases. Clin Genet.

Pajusalu S, Pfundt R, Vissers LE, Kwint MP, Reimand T, Õunap K, Veltman JA, Hehir-Kwa JY. submitted. Identifying long indels in exome sequencing data of patients with intellectual disability.

Pajusalu S, Reimand T, Uibo O, Vasar M, Talvik I, Žilina O, Tammur P, Õunap K. 2015a. De novo deletion of HOXB gene cluster in a patient with failure to thrive, developmental delay, gastroesophageal reflux and bronchiectasis. Eur J Med Genet.

Pajusalu S, Reimand T, Õunap K. 2015b. Novel homozygous mutation in KPTN gene causing a familial intellectual disability-macrocephaly syndrome. Am J Med Genet A.

Pajusalu S, Žilina O, Yakoreva M, Tammur P, Kuuse K, Mölter-Väär T, Nõukas M, Reimand T, Õunap K. 2015c. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. Mol Syndromol 6(3):135–140.

Pajusalu S, Talvik I, Noormets K, Talvik T, Põder H, Joost K, Puusepp S, Piirsoo A, Stenzel W, Goebel HH *et al*. 2016. De novo exonic mutation in MYH7 gene leading to exon skipping in a patient with early onset muscular weakness and fiber-type disproportion. Neuromuscul Disord 26(3):236–239.

Papenhausen P, Schwartz S, Risheg H, Keitges E, Gadi I, Burnside RD, Jaswaney V, Pappas J, Pasion R, Friedman K *et al*. 2011. UPD detection using homozygosity profiling with a SNP genotyping microarray. Am J Med Genet A 155A(4):757–768.

Pfundt R, Del Rosario M, Vissers LE, Kwint MP, Janssen IM, de Leeuw N, Yntema HG, Nelen MR, Lugtenberg D, Kamsteeg EJ *et al*. 2016. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. Genet Med.

Pinkel D, Albertson DG. 2005. Array comparative genomic hybridization and its applications in cancer. Nat Genet 37 Suppl:S11–17.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20(1):110–121.

Poninska JK, Bilinska ZT, Franaszczyk M, Michalak E, Rydzanicz M, Szpakowski E, Pollak A, Milanowska B, Truszkowska G, Chmielewski P *et al*. 2016. Next-generation sequencing for diagnosis of thoracic aortic aneurysms and dissections: diagnostic yield, novel mutations and genotype phenotype correlations. J Transl Med 14(1):115.

Pons L, Till M, Alix E, Abel C, Boggio D, Bordes A, Caloone J, Raskin FC, Chatron N, Cordier MP *et al*. 2017. Prenatal microarray comparative genomic hybridization: Experience from the two first years of activity at the Lyon university-hospital. J Gynecol Obstet Hum Reprod 46(3):275–283.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Huffmeier U, Thiel C, Ruschendorf F *et al*. 2006. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am J Med Genet A 140(19):2063–2074.

Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N *et al*. 2012. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380(9854):1674–1682.

Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra CM, Abbott MA *et al*. 2017. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 49(1):36–45.

Reinson K, Õiglane-Šlik E, Talvik I, Vaher U, Õunapuu A, Ennok M, Teek R, Pajusalu S, Murumets Ü, Tomberg T *et al*. 2016. Biallelic CACNA1A mutations cause early onset epileptic encephalopathy with progressive cerebral, cerebellar, and optic nerve atrophy. Am J Med Genet A 170(8):2173–2176.

Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG *et al*. 2016. Clinical application of whole-exome sequencing across clinical indications. Genet Med 18(7):696–704.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E *et al*. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5):405–424.

Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Consortium WGS, Wilkie AO, McVean G, Lunter G. 2014. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 46(8):912–918.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29(1):24–26.

Rodriguez-Revenga L, Mila M, Rosenberg C, Lamb A, Lee C. 2007. Structural variation in the human genome: the impact of copy number variants on clinical diagnosis. Genet Med 9(9):600–606.

Rogers YH, Venter JC. 2005. Genomics: massively parallel sequencing. Nature 437(7057):326–327.

Rousseau F, Bonaventure J, Legeai-Mallet L, Pelet A, Rozet JM, Maroteaux P, Le Merrer M, Munnich A. 1994. Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. Nature 371(6494):252–254.

Salgado D, Bellgard MI, Desvignes JP, Beroud C. 2016. How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era. Hum Mutat 37(12):1272–1282.

Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94(3):441–448.

Saudi Mendeliome Group. 2015. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. Genome Biol 16:134.

Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, Smith A, Bedford HM, Bernard G, Bernier FP, Brais B *et al*. 2016. Utility of whole-exome

sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clin Genet 89(3):275–284.
Schuster SC. 2008. Next-generation sequencing transforms today's biology. Nat Methods 5(1):16–18.
Schuurs-Hoeijmakers JH, Hehir-Kwa JY, Pfundt R, van Bon BW, de Leeuw N, Kleefstra T, Willemsen MA, van Kessel AG, Brunner HG, Veltman JA *et al*. 2011. Homozygosity mapping in outbred families with mental retardation. Eur J Hum Genet 19(5):597–601.
Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11(4):361–362.
Senderek J, Krieger M, Stendel C, Bergmann C, Moser M, Breitbach-Faller N, Rudnik-Schoneborn S, Blaschek A, Wolf NI, Harting I *et al*. 2005. Mutations in SIL1 cause Marinesco-Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. Nat Genet 37(12):1312–1314.
Seo TS, Bai X, Kim DH, Meng Q, Shi S, Ruparel H, Li Z, Turro NJ, Ju J. 2005. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. Proc Natl Acad Sci U S A 102(17):5926–5931.
Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R *et al*. 2005. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77(1):78–88.
Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, Jiang YH, Xie P, Need A, Goldstein DB. 2014. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med 16(2):176–182.
Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309(5741):1728–1732.
Shiang R, Thompson LM, Zhu YZ, Church DM, Fielder TJ, Bocian M, Winokur ST, Wasmuth JJ. 1994. Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. Cell 78(2):335–342.
Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A *et al*. 2010. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet 47(5):332–341.
Simenson K, Õiglane-Šlik E, Teek R, Kuuse K, Õunap K. 2014. A patient with the classic features of Phelan-McDermid syndrome and a high immunoglobulin E level caused by a cryptic interstitial 0.72-Mb deletion in the 22q13.2 region. Am J Med Genet A 164A(3):806–809.
Sobrido MJ, Fernandez JM, Fontoira E, Perez-Sousa C, Cabello A, Castro M, Teijeira S, Alvarez S, Mederer S, Rivas E *et al*. 2005. Autosomal dominant congenital fibre type disproportion: a clinicopathological and imaging study of a large family. Brain 128(Pt 7):1716–1727.
Sohn RL, Vikstrom KL, Strauss M, Cohen C, Szent-Gyorgyi AG, Leinwand LA. 1997. A 29 residue region of the sarcomeric myosin rod is necessary for filament formation. J Mol Biol 266(2):317–330.
South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical G, Genomics Laboratory Quality Assurance C. 2013. ACMG Standards and Guidelines for constitutional cytogenomic microarray

analysis, including postnatal and prenatal applications: revision 2013. Genet Med 15(11):901–909.

Spence JE, Perciaccante RG, Greig GM, Willard HF, Ledbetter DH, Hejtmancik JF, Pollack MS, O'Brien WE, Beaudet AL. 1988. Uniparental disomy as a mechanism for human genetic disease. Am J Hum Genet 42(2):217–226.

Srebniak MI, Knapen M, Polak M, Joosten M, Diderich KEM, Govaerts LCP, Boter M, Kromosoeto JNR, van Hassel D, Huijbregts G *et al*. 2017. The influence of SNP-based chromosomal microarray and NIPT on the diagnostic yield in 10,000 fetuses with and without fetal ultrasound anomalies. Hum Mutat.

Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, Yeung A, Peters H, Mordaunt D, Cowie S *et al*. 2016. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genet Med 18(11):1090–1096.

Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. 2009. The Human Gene Mutation Database: 2008 update. Genome Med 1(1):13.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Hsi-Yang Fritz M *et al*. 2015. An integrated map of structural variation in 2,504 human genomes. Nature 526(7571):75–81.

Sun Y, Ruivenkamp CA, Hoffer MJ, Vrijenhoek T, Kriek M, van Asperen CJ, den Dunnen JT, Santen GW. 2015. Next-Generation Diagnostics: Gene Panel, Exome, or Whole Genome? Hum Mutat 36(6):648–655.

Sund KL, Zimmerman SL, Thomas C, Mitchell AL, Prada CE, Grote L, Bao L, Martin LJ, Smolarek TA. 2013. Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. Genet Med 15(1):70–78.

Zazo Seco C, Wesdorp M, Feenstra I, Pfundt R, Hehir-Kwa JY, Lelieveld SH, Castelein S, Gilissen C, de Wijs IJ, Admiraal RJ *et al*. 2017. The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. Eur J Hum Genet 25(3):308–314.

Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10:451–481.

Zheng J, Miller KK, Yang T, Hildebrand MS, Shearer AE, DeLuca AP, Scheetz TE, Drummond J, Scherer SE, Legan PK *et al*. 2011. Carcinoembryonic antigen-related cell adhesion molecule 16 interacts with alpha-tectorin and is mutated in autosomal dominant hearing loss (DFNA4). Proc Natl Acad Sci U S A 108(10):4218–4223.

Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M. 2014. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol 32(3):246–251.

Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, Mace A, Ferrari C, Kutalik Z, Andrieux J *et al*. 2012. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet 49(10):660–668.

Žilina O. 2014. Chromosomal microarray analysis as diagnostic tool: Estonian experience. *Dissertationes Biologicae Universitatis Tartuensis 264*.Tartu: Tartu University Press.

Žilina O, Kahre T, Talvik I, Õiglane-Šlik E, Tillmann V, Õunap K. 2014a. Mosaicism for maternal uniparental disomy 15 in a boy with some clinical features of Prader-Willi syndrome. Eur J Med Genet 57(6):279–283.

Žilina O, Reimand T, Zjablovskaja P, Männik K, Mannamaa M, Traat A, Puusepp-Benazzouz H, Kurg A, Õunap K. 2012. Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families. Am J Med Genet A 158A(1):254–256.

Žilina O, Teek R, Tammur P, Kuuse K, Yakoreva M, Vaidla E, Molter-Vaar T, Reimand T, Kurg A, Õunap K. 2014b. Chromosomal microarray analysis as a first-tier clinical diagnostic test: Estonian experience. Mol Genet Genomic Med 2(2):166–175.

Tajsharghi H, Oldfors A. 2013. Myosinopathies: pathology and mechanisms. Acta Neuropathol 125(1):3–18.

Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C *et al*. 2015. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet 47(7):717–726.

The Gene Ontology Consortium. 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res 43(Database issue):D1049–1056.

Thompson K, Majd H, Dallabona C, Reinson K, King MS, Alston CL, He L, Lodi T, Jones SA, Fattal-Valevski A *et al*. 2016. Recurrent De Novo Dominant Mutations in SLC25A4 Cause Severe Early-Onset Mitochondrial Disease and Loss of Mitochondrial DNA Copy Number. Am J Hum Genet 99(4):860–876.

Tjio JH, Levan A. 1956. The Chromosome Number of Man. Hereditas 42(1–2):1–6.

Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Koster J, Marais A, Paknia O, Schroder R, Garcia-Aznar JM, Werber M *et al*. 2017. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet 25(2):176–182.

Udd B. 2009. 165th ENMC International Workshop: distal myopathies 6–8th February 2009 Naarden, The Netherlands. Neuromuscul Disord 19(6):429–438.

Vaher U, Nõukas M, Nikopensius T, Kals M, Annilo T, Nelis M, Õunap K, Reimand T, Talvik I, Ilves P *et al*. 2014. De novo SCN8A mutation identified by whole-exome sequencing in a boy with neonatal epileptic encephalopathy, multiple congenital anomalies, and movement disorders. J Child Neurol 29(12):NP202–206.

Waldmuller S, Erdmann J, Binner P, Gelbrich G, Pankuweit S, Geier C, Timmermann B, Haremza J, Perrot A, Scheer S *et al*. 2011. Novel correlations between the genotype and the phenotype of hypertrophic and dilated cardiomyopathy: results from the German Competence Network Heart Failure. Eur J Heart Fail 13(11):1185–1192.

Vals MA, Kahre T, Mee P, Muru K, Kallas E, Žilina O, Tillmann V, Õunap K. 2015. Familial 1.3-Mb 11p15.5p15.4 Duplication in Three Generations Causing Silver-Russell and Beckwith-Wiedemann Syndromes. Mol Syndromol 6(3):147–151.

Vals MA, Õiglane-Šlik E, Noukas M, Shor R, Peet A, Kals M, Kivistik PA, Metspalu A, Õunap K. 2014. Coffin-Siris Syndrome with obesity, macrocephaly, hepatomegaly and hyperinsulinism caused by a mutation in the ARID1B gene. Eur J Hum Genet 22(11):1327–1329.

Walsh R, Rutland C, Thomas R, Loughna S. 2010. Cardiomyopathy: a systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic manifestations. Cardiology 115(1):49–60.

Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S *et al*. 2010. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature 463(7281):671–675.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J *et al*. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 11(1110):11 10 11–11 10 33.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends Genet 30(9):418–426.

van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, Howard HC, Cambon-Thomsen A, Knoppers BM, Meijers-Heijboer H *et al*. 2013. Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. Eur J Hum Genet 21(6):580–584.

Wang JC, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, El Naggar M, Kopita KA, Randolph LM, Chase JM *et al*. 2014. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. Eur J Hum Genet.

Wang JC, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, El Naggar M, Kopita KA, Randolph LM, Chase JM *et al*. 2015. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. Eur J Hum Genet 23(5):663–671.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16):e164.

Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM, Savage M, Platt LD, Saltzman D, Grobman WA *et al*. 2012. Chromosomal microarray versus karyotyping for prenatal diagnosis. N Engl J Med 367(23):2175–2184.

Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171(4356):737–738.

Vega AI, Medrano C, Navarrete R, Desviat LR, Merinero B, Rodriguez-Pombo P, Vitoria I, Ugarte M, Perez-Cerda C, Perez B. 2016. Molecular diagnosis of glycogen storage disease and disorders with overlapping clinical symptoms by massive parallel sequencing. Genet Med.

Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet 14(2):125–138.

Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcostsdata [Accessed: 02-05-2017]

White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, Bussell JN, Salisbury J, Clare S, Ingham NJ, Podrini C *et al*. 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell 154(2):452–464.

Wierenga KJ, Jiang Z, Yang AC, Mulvihill JJ, Tsinoremas NF. 2013. A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. Genet Med 15(5):354–360.

Winden KD, Yuskaitis CJ, Poduri A. 2015. Megalencephaly and Macrocephaly. Semin Neurol 35(3):277–287.

Vissers LE, de Vries BB, Veltman JA. 2010. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. J Med Genet 47(5):289–297.

Vissers LE, Gilissen C, Veltman JA. 2016. Genetic studies in intellectual disability and related disorders. Nat Rev Genet 17(1):9–18.

Vissers LE, van Nimwegen KJ, Schieving JH, Kamsteeg EJ, Kleefstra T, Yntema HG, Pfundt R, van der Wilt GJ, Krabbenborg L, Brunner HG *et al*. 2017. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. Genet Med.

Wolfson RL, Chantranupong L, Wyant GA, Gu X, Orozco JM, Shen K, Condon KJ, Petri S, Kedir J, Scaria SM *et al*. 2017. KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. Nature 543(7645):438–442.

Õiglane-Šlik E, Puusepp S, Talvik I, Vaher U, Rein R, Tammur P, Reimand T, Teek R, Zilina O, Tomberg T *et al*. 2014. Monosomy 1p36 – a multifaceted and still enigmatic syndrome: four clinically diverse cases with shared white matter abnormalities. Eur J Paediatr Neurol 18(3):338–346.

Õunap K, Pajusalu S, Žilina O, Reimand T, Žordania R. 2016. An 8.4-Mb 3q26.33-3q28 microdeletion in a patient with blepharophimosis-intellectual disability syndrome and a review of the literature. Clin Case Rep 4(8):824–830.

Yamazawa K, Ogata T, Ferguson-Smith AC. 2010. Uniparental disomy and human disease: an overview. Am J Med Genet C Semin Med Genet 154C(3):329–334.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z *et al*. 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 369(16):1502–1511.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C *et al*. 2014. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 312(18):1870–1879.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25(21):2865–2871.

WEB RESOURCES

Ambry Genetics. A privately-held healthcare company with the suite of genetic testing solutions for inherited and non-inherited diseases: http://www.ambrygen.com

ClinVar. Database of genomic variation and its relationship to human health: https://www.ncbi.nlm.nih.gov/clinvar/

DECIPHER. A DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources: https://decipher.sanger.ac.uk

DGV. A Database of Genomic Variants. http://dgv.tcag.ca

ExAC. The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community: http://exac.broadinstitute.org/

GATK. Variant discovery toolkit for high-throughput sequencing data: https://software.broadinstitute.org/gatk/

GenBank. National Institute of Health genetic sequence database, an annotated collection of all publicly available DNA sequences. https://www.ncbi.nlm.nih.gov/genbank/

GeneReviews. An international point-of-care resource for clinicians, provides clinically relevant and medically actionable information for inherited condition: https://www.ncbi.nlm.nih.gov/books/NBK1116

gnomAD. The Genome Aggregation Database, is a coalition of investigators seeking to aggregate and harmonize exome and genome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community: http://gnomad.broadinstitute.org/ .

IMPC. The International Mouse Phenotyping Consortium: http://www.mousephenotype.org

OMIM. Online Mendelian Inheritance in Man. An Online Catalog of Human Genes and Genetic Disorders: http://omim.org

Orphanet. The European Union portal for rare diseases and orphan drugs: http://www.orpha.net

Picard. A set of command line tools for manipulating high-throughput sequencing data: http://broadinstitute.github.io/picard/

SUMMARY IN ESTONIAN

Mendeliaarsete haiguste ülegenoomne diagnostika: kromosomaalsest mikrokiibi analüüsist järgmise põlvkonna sekveneerimiseni

20. sajandi teises pooles muutus tänu teaduse ja kliinilise meditsiini arengule võimalikuks geneetiliste haiguste tsüto- või molekulaargeneetiline diagnostika. Täpne geneetiline diagnoos on aga kliinilise geneetika kui arstliku eriala keskne eesmärk, mis omakorda võimaldab pakkuda patsiendile ja tema perekonnale adekvaatset nõustamist, hinnata kordusriske perele, teha sünnieelset diagnostikat ning üha sagedamini leida ka paremat ja personaliseeritumat ravi [ACMG Board of Directors 2015]. Kliinilises geneetikas keskendutakse peamiselt mendeliaarsete haiguste diagnostikale ja ravile. Mendeliaarseteks haigusteks nimetatakse arvukat gruppi kõrge penetrantsusega geneetilisi haiguseid, mis järgivad autosoom-retsessiivset, autosoom-dominantset või X-liitelist pärandumisviisi ja on põhjustatud ühe geeni või lookuse muteerumisest [Antonarakis and Beckmann 2006]. Enamik mendeliaarsetest haigustest on klassifitseeritavad kui harvikhaigused ehk nende esinemissagedus on alla ühe isiku 2000-st [Orphanet]. Enamik tänaseks kirjeldatud u 7000 harvikhaigusest on geneetilise etioloogiaga, kuid samas on paljude haiguste täpne molekulaarne tekkepõhjus siiani avastamata.

Kui klassikaliselt rajanes geneetiliste haiguste diagnostika võimalikult täpsel kandidaatgeenide valikul vastavalt patsiendi fenotüübile ja anamneesile, siis on uuemate ülegenoomsete analüüside juurutamine kliinilisse praktikasse võimaldanud uurida korraga ka suuri genoomipiirkondi ehk paljusid või ka kõiki geene korraga. Esimeseks kaasaegseks ülegenoomseks uuringuks kliinilises geneetikas oli kromosomaalne mikrokiibi analüüs (KMA), mis võimaldab uurida DNAd koopia-arvu muutuste (mikrodeletsioonide ja -duplikatsioonide) suhtes enam kui 100 korda täpsema resolutsiooniga kui klassikalisel mikroskopeerimisel põhineva kromosoomianalüüsiga. Eestis võeti KMA kliinilises diagnostikas kasutusele 2009. aastal, kuid alates 2011. aastast kuulub see Eesti Haigekassa tasustatavate teenuste nimekirja. KMA tulemuslikkust on uurinud Eestis dr Katrin Männik ja dr Olga Žilina oma doktoriuuringute raames [Männik 2012; Žilina 2014]. Dr Žilina uuringutest selgus, et KMA alusel leitakse täpne diagnoos 11%-l uuritutest. Samas esineb ka hulgaliselt ebaselge kliinilise tähendusega leide, millest kõige sagedasemad on pikad homosügootsed alad. Kromosoomiregiooni homosügootsus ei põhjusta haigust iseenesest, kuid võib suurendada autosoom-retsessiivsete haiguste riski. Pikkade homosügootsete alade kliinilist tähendust on siiani vähe uuritud, eriti riikides, kus vanemate lähisugulus on harv.

Veelgi uuemateks ülegenoomseteks uuringuteks on nn järgmise põlvkonna sekveneerimisel (ingl k *next-generation sequencing* ehk NGS) põhinevad geenianalüüsid. NGS tehnoloogia võimaldab korraga analüüsida suurt hulka geene

(eksoneid) sõltuvalt kasutatavast metoodikast. Enamlevinud on kogu eksoomi ehk kõigi valku kodeerivate geenipiirkondade (eksonite) sekveneerimine (ingl k *whole exome sequencing* ehk WES) ning erinevate geenipaneelide analüüsid. Geenipaneelide disaine on äärmiselt erinevaid, olles suunatud üksikutele kuni tuhandetele geenidele. Üheks levinumaks geenipaneeliks on u 5000 mendeliaarse haigusega seostatud geeni (mendelioomi) uuriv analüüs. Eestis on WES kliinilises kasutuses alates 2013. aastast (Haigekassa hinnakirjas alates 2014. aastast), mendelioomi sekveneerimisega alustasime Tartu Ülikooli Kliinikumis (TÜK) 2015. aasta aprillis. Kuigi teaduskirjanduses on avaldatud üksikuid juhukirjeldusi, ei ole süstemaatilist analüüsi NGS analüüside kliinilise tulemuslikkuse kohta Eestis seni tehtud. NGS metoodikate üks oluline aspekt on võimekus tuvastada nii teadaoleva patogeensusega kui ka varem kirjeldamata geenivariante ning avastada uusi genotüüp-fenotüüp seoseid. Keerukamad juhtumid uute mutatsioonide kliinilise tähenduse selgitamiseks vajavad aga tihti lahendamiseks rutiinsest diagnostikast välja jäävaid teaduslikke eksperimente.

Käesoleva uuringu eesmärgid

1. Hinnata pikkade homosügootsete alade kliinilist tähendust Eesti patsientidel, kellel ei esine vanemate lähisugulust (artikkel I).
2. Hinnata suure geenipaneeli ehk mendelioomi sekveneerimise tulemuslikkust kliinilises diagnostikas (artikkel II).
3. Kirjeldada ja molekulaarselt täpsustada *KPTN*-seoselist intellektipuude sündroomi täiskasvanud vennal ja õel (artikkel III).
4. Uurida trio-WES-il tuvastatud uudse *MYH7* geeni *de novo* mutatsiooni molekulaarset ja kliinilist tagajärge (artikkel IV).

Patsientide ja meetodite lühikirjeldus

Pikkade homosügootsete alade kliinilise tähenduse uurimiseks analüüsiti uuesti KMA andmeid 2110 patsiendil, kellele oli analüüs tehtud TÜK ühendlabori kliinilise geneetika keskuses aastatel 2011–2014. Lõplikusse uuringuvalimisse kaasati patsiendid, kellel esines üks või kaks pikka (üle 5 miljoni aluspaari ehk megabaasi) homosügootset ala kogupikkusega alla 28 megabaasi (st 1% kogu autosoomsest genoomist). Neil patsientidel uuriti homosügootsetes alades esinevate autosoom-retsessiivseid haigusi põhjustavate geenide sobivust patsiendi kliiniliste sümptomitega. Sobiva kandidaatgeeni leidmisel tehti WES, et uurida, kas patogeenne mutatsioon lokaliseerub homosügootse ala kandidaatgeeni või mõnesse teise geeni.

Geenipaneeli sekveneerimise tulemuslikkuse hindamiseks analüüsiti 501 järjestikkust patsienti, kellele oli aastatel 2015–2016 tehtud TÜK molekulaardiagnostika laboris sekveneerimisanalüüs, kasutades TruSight One paneeli (Illumina Inc.). Kõik analüüsid tehti diagnostiliste uuringutena ning kliinilised andmed koguti saatekirjadelt. Diagnostiline interpretatsioon lähtus arstide tellitud geenidest/paneelidest, mis moodustasid igale patsiendile personaalse virtuaalse alapaneeli ehk valiku kõigist paneeliga kaetud 4813 geenist. Nii välditi juhuleidude avastamist ja lähtuti konkreetsele diagnostilisele küsimusele vasta-

misest. Peale punktmutatsioonide ja lühikeste insertsioonide/deletsioonide ehk indelite kasutati ka NGS andmetest koopiaarvumuutuste tuvastamise algoritmi.

Lisaks on käesoleva uurimustöö osa kahe geenidiagnostika erinevaid aspekte illustreeriva haigusjuhu kirjeldus. Seni teadmata põhjusega intellektipuude ja makrotsefaaliaga õele-vennale tehti KMA ning pärast jagatud homosügootse ala tuvastamist vennale WES. Perekondlik segregatsioonianalüüs tehti Sanger sekveneerimisega. Teiseks haigusjuhuks oli kaasasündinud müopaatiaga 2-aastane poiss lihaskiutüüpide disproportsiooniga, kellel tehti geneetilise etioloogia selgitamiseks trio-WES (laps ja mõlemad vanemad). Tuvastatud sünonüümse *de novo* mutatsiooni molekulaarse efekti hindamiseks tehti lihaskoest eraldatud RNA PCR-amplifikatsioon ja Sanger sekveneerimine.

Peamised tulemused ja järeldused

1. KMA tehti rutiinse kliinilise uuringuna TÜK tsütogeneetika laboris neljaaastase uuringuperioodi vältel 2110 korral. Selles valimis uuriti üksikute homosügootsete alade tähendust patsientidel, kelle vanemad ei olnud sugulased.
   1.1 8,2% patsientidest kannavad ühte või kahte homosügootset ala pikkusega üle 5 megabaasi. See teeb homosügootsetest aladest kõige sagedasema KMA vastuses kajastatud ebaselge tähendusega leidude klassi.
   1.2 Leiti kuus patsientidel korduvat homosügootset ala (kokku 52 ehk 28,9% kõigist üksikutest homosügootsetest aladest). Seega õnnestus selgitada peaaegu kolmandiku homosügootsete alade tähendust, klassifitseerides need tõenäoliselt healoomulisteks variantideks. Kuuest korduvast homosügootsest regioonist ainult kaks olid varem kirjeldatud USAs läbi viidud uuringus, mis näitab populatsioonist tulenevat erisust.
   1.3 129 mittekorduvast homosügootsest alast kaks sisaldasid autosoomretsessiivset haigust põhjustavat kandidaatgeeni, mis sobis patsiendi kliinilise fenotüübiga. Mõlemal juhul tuvastati neis kandidaatgeenides haigust põhjustav homosügootne variant. Kuigi kliiniliselt sobiva kandidaatgeeni avastamine üksikust homosügootsest alast on harv juhus, võib süstemaatiline homosügootsetes alades esinevate geenide analüüs aidata kaasa sobivate kandidaatgeenide leidmisele.
   1.4 Kui kliiniliselt sobivat kandidaatgeeni üksikus pikas homosügootses alas ei tuvastata, võib see siiski tõsta homosügootse haigusseoselise mutatsiooni esinemise tõenäosust, arvestades vanemate kauge suguluse võimalikkust. Seda illustreeris juhtum, kus õel ja vennal tuvastati homosügootne raaminihke mutatsioon *KPTN*-geenis, mis ei asunud nendel mõlemal esinenud pikas homosügootses alas.
2. Tuginedes TÜK molekulaardiagnostika labori kogemusele esimese 501 diagnostilise juhtumi kohta, kellele tehti 4813 geeni paneelsekveneerimine ja sellele järgnenud analüüs vastavalt arstide tellitud alapaneelidele, saab järeldada, et laia paneeli kasutamine võimaldab ära kasutada mitmeid NGSi eeliseid, et tuvastada erinevaid mutatsioonitüüpe.

2.1 501 patsiendist 132-le õnnestus leida geenipaneeli sekveneerimisel põhjuslik diagnoos. Seega oli diagnostiline saagis 26,3%, mis on võrreldav varasemalt kirjanduses avaldatud suurte kliiniliste WES kohortide tulemuslikkusega.

2.2 Kõigist tuvastatud haigusseoselistest mutatsioonidest 46% olid uudsed (varem haigusseoseliste mutatsioonide andmebaasides kajastamata). See toetab haruldaste haiguste diagnostikas sekveneerimisanalüüside eelistamist võrrelduna vaid varasemalt teadaolevate mutatsioonide testimisele.

2.3 Kõige sagedasemad patogeensete geenivariantide tüübid olid punktmutatsioonid ja lühikesed indelid. Kuna kasutati ka NGS andmetest koopiaarvumuutuste tuvastamise algoritme, tuvastati lisaks punktmutatsioonidele seitsmel juhul haigusseoselise muutusena ka koopiaarvu muutus, kas üksikuna või punktmutatsiooniga liitheterosügootsena.

2.4 Diagnostiline saagis oli tunduvalt kõrgem nende patsientide hulgas, kellele oli tellitud 10 või vähema geeni analüüs võrrelduna suurema alapaneeliga tellimustega (vastavalt 31,5% ja 21,7%, $p = 0,015$). Sellest järeldub, et hea kliiniline fenotüpiseerimine võib aidata valida patsiente, kellel on kõige suurem tõenäosus tuvastada sekveneerimisanalüüsi abil diagnostiline leid.

2.5 TÜK molekulaardiagnostika labori kui maailma mastaabis suhteliselt väikese kliinilise labori kogemus näitab, et suure geenipaneeli sekveneerimist on võimalik kasutada paljude erinevate diagnostiliste juhtumite korral nii kulu- kui ka ajatõhusalt, laiendades laboris uuritavate geenide hulka märkimisväärselt.

3. Kahel intellektipuudega õel-vennal tuvastati *KPTN* geenis homosügootne geenivariant, mis võimaldas kirjeldada uue autosoom-retsessiivse geneetilise sündroomi tuumikfenotüüpi.

3.1 KMA tuvastas õel-vennal jagatud pika homosügootse ala. Haruldaste homosügootsete mutatsioonide leidmiseks tehti vennale WES, mis tuvastas *KPTN* geenis varem kirjeldamata homosügootse raaminihkemutatsiooni c.665dupA p.(Ser223Glnfs*18). *KPTN* geen ei asunud eelnevalt teada olevas pikas homosügootses alas. Sangeri järgi sekveneerimine kinnitas homosügootse leiu ka õel.

3.2 Antud juhtum oli maailmas teine *KPTN*-seoselise intellektipuudega patsientide kirjeldus pärast esmakirjeldust USAs Ohios amišite kogukonnas. *KPTN*-seoselise intellektipuude kirjeldamine Eestis toetab esmakirjeldajate hüpoteesi, et sündroomi esineb ka mujal maailmas väljaspool amišite isolaati.

3.3 *KPTN*-seoselise autosoom-retsessiivse sündroomi peamised tunnused on makrotsefaalia ja vaimse arengu mahajäämus. Epileptilised hood ja käitumishäired on arvatavasti sagedasemad kui üldrahvastikus, kuid on mittetäieliku penetrantsusega tunnused. Vaimse ja kõne arengu mahajäämus esineb neil patsientidel raskemas astmes kui motoorse arengu mahajäämus.

4. Kaasasündinud müopaatia ja kiutüüpide disproportsiooniga poisile tehtud molekulaargeneetilised analüüsid tuvastasid uudse mutatsiooni *MYH7* geenis, mille tulemusel jäetakse mRNA-s üks ekson vahele. See kirjeldus laiendas *MYH7*-seoseliste müopaatiate geneetilist ja kliinilist spektrit.
    4.1 WES, mis tehti lapsele ja tema vanematele diagnostilise trio-analüüsina, tuvastas patsiendil varem kirjeldamata uustekkese (*de novo*) sünonüümse mutatsiooni c.5655G>A, p.(Ala1885=). Muteerunud nukleotiid oli 38. eksoni viimane (RefSeq NM_000257.3). Mutatsiooni olemasolu kinnitus Sangeri järgi sekveneerimisega.
    4.2 Lihaskoest eraldatud RNA uuringud tõestasid mutatsiooni molekulaarse efektina 38. eksoni vahele jätmise. See oli esimene kirjeldus eksoni vahele jätmist põhjustavast mutatsioonist *MYH7* geenis, kuivõrd teised kirjeldatud haigusseoselised mutatsioonid on olnud *missense* muutused või ühe aminohappe kadu põhjustavad raaminihketa mutatsioonid.
    4.3 *MYH7* mutatsioonide põhjustatud fenotüüpide spekter on lai ning antud uuring võimaldas seda veelgi laiendada. *MYH7* geen tuleks lisada kandidaatgeenide nimekirja patsientidel, kellel esineb generaliseerunud hüpotoonia ja kaelalihaste nõrkusest tulenev raskus pea hoidmisel.
    4.4 Diagnostika vaatepunktist rõhutab antud juhtum, et sünonüümsete (neutraalsete) mutatsioonide väljafiltreerimisel WES interpretatsioonil peab suhtuma ettevaatusega. Samuti illustreerib juhtum RNA ja teiste bioloogiliste uuringute võimekusega teaduslaboritega koostöö vajalikkust keerukamate diagnostiliste WES analüüside lahendamisel.

Käesolev uuring näitas ülegenoomsete analüüside kasu mitmete erinevate diagnostiliste stsenaariumite korral ja võimaldas hinnata erinevate geneetiliste uuringute diagnostilist efektiivsust kliinilises praktikas. Uute NGS ja KMA analüüside juurutamine rutiinsete meditsiiniuuringutena on võimaldanud oluliselt parandada mendeliaarsete haiguste diagnostikat Eestis. Antud uuringu tulemusi saab tulevikus kasutada erinevate haiguste kliiniliste käsitlusjuhiste arendamiseks.

ACKNOWLEDGMENTS

My deepest gratitude goes to my two supervisors Professor Katrin Õunap and Associate Professor Tiia Reimand. I am especially thankful to Prof Katrin Õunap for inviting me to her department to work as a doctor's assistant during my fifth year of medical studies, and keeping me on track to become a specialized medical geneticist as well as a researcher since then. In addition, I cannot underestimate the importance of Dr Tiia Reimand in finalizing my studies, thanks to her great support and guidance.

I would like to thank the reviewers of this thesis Prof Pärt Peterson and Prof Pille Taba for taking the time to make many great comments and suggestions on how to improve this thesis.

The work presented in this thesis has been an effort from many co-authors who I would like to hereby thank: Tiia Reimand, Katrin Õunap, Olga Žilina, Maria Yakoreva, Pille Tammur, Kati Kuuse, Triin Mölter-Väär, Tiina Kahre, Hanno Roomere, Ülle Murumets, Laura Roht, Kristi Simenson, Inga Talvik, Klari Noormets, Tiina Talvik, Haide Põder, Kairit Joost, Sanna Puusepp, Andres Piirsoo, Werner Stenzel, Hans H. Goebel, Margit Nõukas, Tiit Nikopensius, Tarmo Annilo, and Andres Metspalu. I am truly grateful to all the patients and their parents who have participated in the studies and thus made this work possible.

Completing my clinical specialisation and PhD studies simultaneously would not have been possible without the constant support of my colleagues from the Department of Clinical Genetics at Tartu University Hospital. I received a lot of help regarding practicalities from the staff at the Dean's Office of the Faculty of Medicine, University of Tartu.

During my PhD studies, I had an invaluable opportunity to visit three world-class institutions and work there as a visiting PhD student under the supervision of many great scientists. First, I worked for five months at Radboud University Medical Center in Nijmegen, The Netherlands. For this wonderful experience, I am thankful to my supervisor Dr Jayne Hehir-Kwa and the group leader Prof Joris Veltmann. In addition, I am thankful to all the bioinformaticians in Radboudumc who introduced me to their exciting field. Second, I spent a month on the shores of beautiful Lake Geneva while working with Dr Katrin Männik and Prof Alexandre Reymond at the Center for Integrative Genomics at the University of Lausanne, Switzerland. My special thanks go to Dr Katrin Männik who taught me a lot about both written and unwritten laws and rules of science. And finally, I had the greatest two weeks experiencing the scientific excellence at the Broad Institute of MIT and Harvard in Cambridge, MA, USA working closely with Dr Monica Wojcik, Dr Monkol Lek, Prof Daniel MacArthur, as well as many other inspiring people.

I would also like to thank Dr Rael Laugesaar, Prof Tiina Talvik and Dr Anneli Kolk who supervised my very first clinical research project, where we investigated different aspects of paediatric sinovenous thrombosis. My sincere

thanks go also to Dr Ene Reimann and Prof Sulev Kõks who first introduced me to laboratory science as well as to next-generation sequencing technologies. For their support, I would like to thank Prof Aavo-Valdur Mikelsaar and Prof Maris Laan. In addition, I must thank Prof Eero Vasar for setting my goal on completing both residency and PhD studies before turning 30 and Prof Andres Metspalu who taught me, besides many other things, that it is always better to say "Thank you!" one time more than one time less.

I must express my deepest gratitude to my parents Renate and Karl for raising me in such a caring, supportive, and academic environment, and showing me that borders between different fields of science are arbitrary. Also, a big thanks to my older brother Mihkel to whom I have always looked up to and my little sister Katri for reminding me the good things about being young. Also, I am truly grateful to my grandparents, especially my maternal grandmother Ariadna who was the first to tell me about the laws of nature, and my paternal grandmother Linda who has taught me all about the true values in life. Last but not least, my sincerest thanks go to Kadi, who has stood by me during the entire journey of my PhD studies and residency and did her best at distracting me, which paradoxically gave me the true strength for completing everything on time.

Parts of this study were supported by Estonian Research Council grant PUT355, EU Social Fund through the Estonian Research Council (Grant IUT20-60), the Development Fund of the University of Tartu (grant SP1GVARENG), and the European Regional Development Fund to the Centre of Excellence in Genomics (EXCEGEN; grant 3.2.0304.11-0312). During my PhD studies, I received support from the University of Tartu ASTRA Project PER ASPERA (European Union, European Regional Development Fund), ePerMed project, (European Union H2020 grant 692145), national scholarship programs Kristjan Jaak and DoRa (Archimedes Foundation & Ministry of Education and Research of Estonia), Liisa Kolumbus scholarship and Andreas and Elmerice Traks scholarship (The University of Tartu Foundation), the Estonian Medical Association, and the European Society of Human Genetics.

PUBLICATIONS

CURRICULUM VITAE

Name:	Sander Pajusalu
Date of birth:	07.04.1988, Tartu, Estonia
Citizenship:	Estonian
Address:	2 L. Puusepa Street, 51014, Tartu, Estonia
Phone:	7 319 537
E-mail:	sander.pajusalu@kliinikum.ee

Education and employment

2004–2007	Hugo Treffner Gymnasium (Tartu, Estonia)
2007–2013	University of Tartu, Faculty of Medicine, Degree in Medicine (Diploma *cum laude*)
2013–2017	University of Tartu, Faculty of Medicine, Residency training in medical genetics
2014–	University of Tartu, Faculty of Medicine, PhD studies in medical genetics
2017–	Tartu University Hospital, United Laboratories, Department of Clinical genetics, Medical Geneticist
2017–	University of Tartu, Institute of Clinical Medicine, Junior Research Fellow

Visits to research/clinical institutions during PhD studies

Oct 2015–Jan 2016	Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. Supervisor: Jayne Hehir-Kwa, PhD. Group Leader: Prof. Joris Veltman.
Sept 2016	Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland Supervisor: Katrin Männik, PhD. Group Leader: Prof. Alexandre Reymond.
Oct 2016	Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. Supervisor: Jayne Hehir-Kwa, PhD. Group Leader: Prof. Joris Veltman
Feb 2017 (2 weeks)	Broad Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge (Boston), MA, United States. Group Leader: Prof. Daniel MacArthur.

Scientific work and professional organizations

Research fields:	Clinical genetics, bioinformatics, neurogenetics.
Publications:	Ten in international journals (five as the first author), two in Eesti Arst (Estonian Medical Journal), 20 published conference abstracts

Membership: Member of the European Society of Human Genetics
Member of management board of the Estonian Society of Medical Genetics
Member of the Estonian Junior Doctors Association
Member of the Estonian Society of Human Genetics.

List of publications

1. Vals, M.-A.; **Pajusalu, S**.; Kals, M.; Mägi, R.; Õunap, K. (2017). *The Prevalence of PMM2-CDG in Estonia Based on Population Carrier Frequencies and Diagnosed Patients.* JIMD Rep. Jul 7. doi: 10.1007/8904_2017_41. [Epub ahead of print]
2. **Pajusalu, S**.; Kahre, T.; Roomere, H.; Murumets, Ü.; Roht, L.; Simenson, K.; Reimand, T.; Õunap, K. (2017). *Large Gene Panel Sequencing in Clinical Diagnostics – Results from 501 Consecutive Cases.* Clinical Genetics, doi: 10.1111/cge.13031. [Epub ahead of print]
3. Vals, M.-A.; Morava, E.; Teeäär, K.; Žordania, R.; **Pajusalu, S.**; Lefeber, D.; Õunap, K. (2017). *Three families with mild PMM2-CDG and normal cognitive development.* American Journal of Medical Genetics Part A, 173 (6), 1620–1624.
4. Reinson, K.; Õiglane-Shlik, E.; Talvik, I.; Vaher, U.; Õunapuu, A.; Ennok, M.; Teek, R.; **Pajusalu, S**.; Murumets, Ü.; Tomberg, T.; Puusepp, S.; Piirsoo, A.; Reimand, T.; Õunap, K. (2016). *Biallelic CACNA1A mutations cause early onset epileptic encephalopathy with progressive cerebral, cerebellar, and optic nerve atrophy.* American Journal of Medical Genetics Part A, 170 (8), 2173–2176.
5. Õunap, K.; **Pajusalu, S**.; Zilina, O.; Reimand, T.; Žordania, R. (2016). *An 8.4-Mb 3q26.33-3q28 microdeletion in a patient with blepharophimosis-intellectual disability syndrome and a review of the literature.* Clinical Case Reports, 4 (8), 824−830.
6. La Piana, R; Cayami, FK; Tran, LT; Guerrero, K; van Spaendonk, R; Õunap, K; **Pajusalu, S**; Haack, T; Wassmer, E; Timmann, D; Mierzewska, H; Poll-Thé, BT; Patel, C; Cox, H; Atik, T; Onay, H; Ozkınay, F; Vanderver, A; van der Knaap, MS; Wolf, NI; Bernard, G. (2016). *Diffuse hypomyelination is not obligate for POLR3-related disorders.* Neurology, 86 (17), 1622−1626.
7. **Pajusalu, S**.; Talvik, I.; Noormets, K.; Talvik, T.; Põder, H.; Joost, K.; Puusepp, S.; Piirsoo, A.; Stenzel, W.; Goebel, HH.; Nikopensius, T.; Annilo, T.; Nõukas, M.; Metspalu, A.; Õunap, K.; Reimand, T. (2016). *De novo exonic mutation in MYH7 gene leading to exon skipping in a patient with early onset muscular weakness and fiber-type disproportion.* Neuromuscular Disorders, 26 (3), 236−239.
8. **Pajusalu, S**; Žilina, O; Yakoreva, M; Tammur, P; Kuuse, K; Mölter-Väär, T; Nõukas, M; Reimand, T; Õunap, K. (2015). *The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity.* Molecular Syndromology, 6, 134−139.

9. **Pajusalu, S**; Reimand, T; Õunap, K. (2015). *Novel homozygous mutation in KPTN gene causing a familial intellectual disability-macrocephaly syndrome.* American Journal of Medical Genetics Part A, 167 (8), 1913−1915.
10. **Pajusalu, S**; Reimand, T; Uibo, O; Vasar, M; Talvik, I; Žilina, O; Tammur, P; Õunap, K. (2015). *De novo deletion of HOXB gene cluster in a patient with failure to thrive, developmental delay, gastroesophageal reflux and bronchiectasis.* European Journal of Medical Genetics, 58, 336−340.
11. **Pajusalu, S**; Žordania, R; Jalas, T; Paju, R; Õunap, K; Reimand, T. (2013). *Congenital long QT syndrome.* [In Estonian] Eesti Arst (Estonian Medical Journal) 92(1):28–33.
12. Laugesaar, R; **Pajusalu, S**; Toome, L; Kolk, A; Tomberg, T; Ilves, P; Eelmäe, I; Talvik, I; Talvik, T. (2012). *Paediatric cerebral sinovenous thrombosis in Estonia.* [In Estonian] Eesti Arst (Estonian Medical Journal), 91(11), 604–609.

ELULOOKIRJELDUS

Nimi: Sander Pajusalu
Sünniaeg: 07.04.1988, Tartu
Kodakondsus: Eesti
Aadress: L. Puusepa 2, 51014, Tartu
Telefon: 7 319 537
E-post: sander.pajusalu@kliinikum.ee

Haridus- ja ametikäik

2004–2007	Hugo Treffneri Gümnaasium, kuldmedal
2007–2013	Tartu Ülikool, arstiteaduskond, arstiõpe (*cum laude*)
2013–2017	Tartu Ülikool, arstiteaduskond, meditsiinigeneetika residentuur
2014–	Tartu Ülikool, arstiteaduskond, doktoriõpe arstiteaduse (meditsiinigeneetika) erialal
2017–	Tartu Ülikooli Kliinikum, ühendlabor, kliinilise geneetika keskus, meditsiinigeneetik
2017–	Tartu Ülikool, kliinilise meditsiini instituut, nooremteadur

Külastused teadus- või kliinilistesse asutustesse doktoriõppe jooksul

Okt 2015–Jaan 2016	Inimesegeneetika osakond, Radboudi ülikooli meditsiinikeskus, Nijmegen, Holland. Juhendaja: Jayne Hehir-Kwa, PhD. Töögrupi juht: Prof Joris Veltman.
Sept 2016	Integreeritud genoomika keskus, Lausanne'i ülikool, Lausanne, Šveits Juhendaja: Katrin Männik, PhD. Töögrupi juht: Prof Alexandre Reymond.
Okt 2016	Inimesegeneetika osakond, Radboudi ülikooli meditsiinikeskus, Nijmegen, Holland. Juhendaja: Jayne Hehir-Kwa, PhD. Töögrupi juht: Prof Joris Veltman.
Veeb 2017 (2 näd.)	Mendeliaarse genoomika keskus, MITi ja Harvardi ülikooli Broadi instituut, Cambridge (Boston), MA, USA. Töögrupi juht: Prof Daniel MacArthur.

Teadus- ja erialane tegevus

Uurimisvaldkonnad:	Kliiniline geneetika, bioinformaatika, neurogeneetika.
Publikatsioonid:	10 rahvusvahelistes ajakirjades (5 esma-autorlust), 2 Eesti Arstis, 20 publitseeritud konverentsiettekande teesi.

Liikmelisus: Euroopa Inimesegeneetika Ühingu liige
Eesti Meditsiinigeneetika Selts, juhatuse liige
Eesti Nooremarstide Ühenduse liige
Eesti Inimesegeneetika Ühingu liige

Publikatsioonide nimekiri

1. Vals, M.-A.; **Pajusalu, S**.; Kals, M.; Mägi, R.; Õunap, K. (2017). *The Prevalence of PMM2-CDG in Estonia Based on Population Carrier Frequencies and Diagnosed Patients.* JIMD Rep. Jul 7. doi: 10.1007/8904_2017_41. [Avaldatud veebiväljandes enne trükki]
2. **Pajusalu, S**.; Kahre, T.; Roomere, H.; Murumets, Ü.; Roht, L.; Simenson, K.; Reimand, T.; Õunap, K (2017). *Large Gene Panel Sequencing in Clinical Diagnostics – Results from 501 Consecutive Cases.* Clinical Genetics, doi: 10.1111/cge.13031. [Avaldatud veebiväljandes enne trükki]
3. Vals, M.-A.; Morava, E.; Teeäär, K.; Žordania, R.; **Pajusalu, S.**; Lefeber, D.; Õunap, K. (2017). *Three families with mild PMM2-CDG and normal cognitive development.* American Journal of Medical Genetics Part A, 173 (6), 1620–1624.
4. Reinson, K.; Õiglane-Shlik, E.; Talvik, I.; Vaher, U.; Õunapuu, A.; Ennok, M.; Teek, R.; **Pajusalu, S**.; Murumets, Ü.; Tomberg, T.; Puusepp, S.; Piirsoo, A.; Reimand, T.; Õunap, K. (2016). *Biallelic CACNA1A mutations cause early onset epileptic encephalopathy with progressive cerebral, cerebellar, and optic nerve atrophy.* American Journal of Medical Genetics Part A, 170 (8), 2173–2176.
5. Õunap, K.; **Pajusalu, S**.; Zilina, O.; Reimand, T.; Žordania, R. (2016). *An 8.4-Mb 3q26.33-3q28 microdeletion in a patient with blepharophimosis-intellectual disability syndrome and a review of the literature.* Clinical Case Reports, 4 (8), 824−830.
6. La Piana, R; Cayami, FK; Tran, LT; Guerrero, K; van Spaendonk, R; Õunap, K; **Pajusalu, S**; Haack, T; Wassmer, E; Timmann, D; Mierzewska, H; Poll-Thé, BT; Patel, C; Cox, H; Atik, T; Onay, H; Ozkınay, F; Vanderver, A; van der Knaap, MS; Wolf, NI; Bernard, G. (2016). *Diffuse hypomyelination is not obligate for POLR3-related disorders.* Neurology, 86 (17), 1622−1626.
7. **Pajusalu, S**.; Talvik, I.; Noormets, K.; Talvik, T.; Põder, H.; Joost, K.; Puusepp, S.; Piirsoo, A.; Stenzel, W.; Goebel, HH.; Nikopensius, T.; Annilo, T.; Nõukas, M.; Metspalu, A.; Õunap, K.; Reimand, T. (2016). *De novo exonic mutation in MYH7 gene leading to exon skipping in a patient with early onset muscular weakness and fiber-type disproportion.* Neuromuscular Disorders, 26 (3), 236−239.
8. **Pajusalu, S**; Žilina, O; Yakoreva, M; Tammur, P; Kuuse, K; Mölter-Väär, T; Nõukas, M; Reimand, T; Õunap, K. (2015). *The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity.* Molecular Syndromology, 6, 134−139.

9. **Pajusalu, S**; Reimand, T; Õunap, K (2015). *Novel homozygous mutation in KPTN gene causing a familial intellectual disability-macrocephaly syndrome.* American Journal of Medical Genetics Part A, 167 (8), 1913−1915.
10. **Pajusalu, S**; Reimand, T; Uibo, O; Vasar, M; Talvik, I; Žilina, O; Tammur, P; Õunap, K. (2015). *De novo deletion of HOXB gene cluster in a patient with failure to thrive, developmental delay, gastroesophageal reflux and bronchiectasis.* European Journal of Medical Genetics, 58, 336−340.
11. **Pajusalu, S**; Žordania, R; Jalas, T; Paju, R; Õunap, K; Reimand, T. (2013). *Pärilik pika QT sündroom.* Eesti Arst 92(1):28–33.
12. Laugesaar, R; **Pajusalu, S**; Toome, L; Kolk, A; Tomberg, T; Ilves, P; Eelmäe, I; Talvik, I; Talvik, T. (2012). *Sinovenoosne tromboos lastel: kirjanduse ülevaade ja kogemus Eestis.* Eesti Arst 91(11), 604–609.

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS

1. **Heidi-Ingrid Maaroos**. The natural course of gastric ulcer in connection with chronic gastritis and *Helicobacter pylori*. Tartu, 1991.
2. **Mihkel Zilmer**. Na-pump in normal and tumorous brain tissues: Structural, functional and tumorigenesis aspects. Tartu, 1991.
3. **Eero Vasar**. Role of cholecystokinin receptors in the regulation of behaviour and in the action of haloperidol and diazepam. Tartu, 1992.
4. **Tiina Talvik**. Hypoxic-ischaemic brain damage in neonates (clinical, biochemical and brain computed tomographical investigation). Tartu, 1992.
5. **Ants Peetsalu**. Vagotomy in duodenal ulcer disease: A study of gastric acidity, serum pepsinogen I, gastric mucosal histology and *Helicobacter pylori*. Tartu, 1992.
6. **Marika Mikelsaar**. Evaluation of the gastrointestinal microbial ecosystem in health and disease. Tartu, 1992.
7. **Hele Everaus**. Immuno-hormonal interactions in chronic lymphocytic leukaemia and multiple myeloma. Tartu, 1993.
8. **Ruth Mikelsaar**. Etiological factors of diseases in genetically consulted children and newborn screening: dissertation for the commencement of the degree of doctor of medical sciences. Tartu, 1993.
9. **Agu Tamm**. On metabolic action of intestinal microflora: clinical aspects. Tartu, 1993.
10. **Katrin Gross**. Multiple sclerosis in South-Estonia (epidemiological and computed tomographical investigations). Tartu, 1993.
11. **Oivi Uibo**. Childhood coeliac disease in Estonia: occurrence, screening, diagnosis and clinical characterization. Tartu, 1994.
12. **Viiu Tuulik**. The functional disorders of central nervous system of chemistry workers. Tartu, 1994.
13. **Margus Viigimaa**. Primary haemostasis, antiaggregative and anticoagulant treatment of acute myocardial infarction. Tartu, 1994.
14. **Rein Kolk**. Atrial versus ventricular pacing in patients with sick sinus syndrome. Tartu, 1994.
15. **Toomas Podar**. Incidence of childhood onset type 1 diabetes mellitus in Estonia. Tartu, 1994.
16. **Kiira Subi**. The laboratory surveillance of the acute respiratory viral infections in Estonia. Tartu, 1995.
17. **Irja Lutsar**. Infections of the central nervous system in children (epidemiologic, diagnostic and therapeutic aspects, long term outcome). Tartu, 1995.
18. **Aavo Lang**. The role of dopamine, 5-hydroxytryptamine, sigma and NMDA receptors in the action of antipsychotic drugs. Tartu, 1995.
19. **Andrus Arak**. Factors influencing the survival of patients after radical surgery for gastric cancer. Tartu, 1996.

20. **Tõnis Karki**. Quantitative composition of the human lactoflora and method for its examination. Tartu, 1996.
21. **Reet Mändar**. Vaginal microflora during pregnancy and its transmission to newborn. Tartu, 1996.
22. **Triin Remmel**. Primary biliary cirrhosis in Estonia: epidemiology, clinical characterization and prognostication of the course of the disease. Tartu, 1996.
23. **Toomas Kivastik**. Mechanisms of drug addiction: focus on positive reinforcing properties of morphine. Tartu, 1996.
24. **Paavo Pokk**. Stress due to sleep deprivation: focus on $GABA_A$ receptor-chloride ionophore complex. Tartu, 1996.
25. **Kristina Allikmets**. Renin system activity in essential hypertension. Associations with atherothrombogenic cardiovascular risk factors and with the efficacy of calcium antagonist treatment. Tartu, 1996.
26. **Triin Parik**. Oxidative stress in essential hypertension: Associations with metabolic disturbances and the effects of calcium antagonist treatment. Tartu, 1996.
27. **Svetlana Päi**. Factors promoting heterogeneity of the course of rheumatoid arthritis. Tartu, 1997.
28. **Maarike Sallo**. Studies on habitual physical activity and aerobic fitness in 4 to 10 years old children. Tartu, 1997.
29. **Paul Naaber**. *Clostridium difficile* infection and intestinal microbial ecology. Tartu, 1997.
30. **Rein Pähkla**. Studies in pinoline pharmacology. Tartu, 1997.
31. **Andrus Juhan Voitk**. Outpatient laparoscopic cholecystectomy. Tartu, 1997.
32. **Joel Starkopf**. Oxidative stress and ischaemia-reperfusion of the heart. Tartu, 1997.
33. **Janika Kõrv**. Incidence, case-fatality and outcome of stroke. Tartu, 1998.
34. **Ülla Linnamägi**. Changes in local cerebral blood flow and lipid peroxidation following lead exposure in experiment. Tartu, 1998.
35. **Ave Minajeva**. Sarcoplasmic reticulum function: comparison of atrial and ventricular myocardium. Tartu, 1998.
36. **Oleg Milenin**. Reconstruction of cervical part of esophagus by revascularised ileal autografts in dogs. A new complex multistage method. Tartu, 1998.
37. **Sergei Pakriev**. Prevalence of depression, harmful use of alcohol and alcohol dependence among rural population in Udmurtia. Tartu, 1998.
38. **Allen Kaasik**. Thyroid hormone control over β-adrenergic signalling system in rat atria. Tartu, 1998.
39. **Vallo Matto**. Pharmacological studies on anxiogenic and antiaggressive properties of antidepressants. Tartu, 1998.
40. **Maire Vasar**. Allergic diseases and bronchial hyperreactivity in Estonian children in relation to environmental influences. Tartu, 1998.
41. **Kaja Julge**. Humoral immune responses to allergens in early childhood. Tartu, 1998.

42. **Heli Grünberg**. The cardiovascular risk of Estonian schoolchildren. A cross-sectional study of 9-, 12- and 15-year-old children. Tartu, 1998.
43. **Epp Sepp**. Formation of intestinal microbial ecosystem in children. Tartu, 1998.
44. **Mai Ots**. Characteristics of the progression of human and experimental glomerulopathies. Tartu, 1998.
45. **Tiina Ristimäe**. Heart rate variability in patients with coronary artery disease. Tartu, 1998.
46. **Leho Kõiv**. Reaction of the sympatho-adrenal and hypothalamo-pituitary-adrenocortical system in the acute stage of head injury. Tartu, 1998.
47. **Bela Adojaan**. Immune and genetic factors of childhood onset IDDM in Estonia. An epidemiological study. Tartu, 1999.
48. **Jakov Shlik**. Psychophysiological effects of cholecystokinin in humans. Tartu, 1999.
49. **Kai Kisand**. Autoantibodies against dehydrogenases of α-ketoacids. Tartu, 1999.
50. **Toomas Marandi**. Drug treatment of depression in Estonia. Tartu, 1999.
51. **Ants Kask**. Behavioural studies on neuropeptide Y. Tartu, 1999.
52. **Ello-Rahel Karelson**. Modulation of adenylate cyclase activity in the rat hippocampus by neuropeptide galanin and its chimeric analogs. Tartu, 1999.
53. **Tanel Laisaar**. Treatment of pleural empyema — special reference to intrapleural therapy with streptokinase and surgical treatment modalities. Tartu, 1999.
54. **Eve Pihl**. Cardiovascular risk factors in middle-aged former athletes. Tartu, 1999.
55. **Katrin Õunap**. Phenylketonuria in Estonia: incidence, newborn screening, diagnosis, clinical characterization and genotype/phenotype correlation. Tartu, 1999.
56. **Siiri Kõljalg**. *Acinetobacter* – an important nosocomial pathogen. Tartu, 1999.
57. **Helle Karro**. Reproductive health and pregnancy outcome in Estonia: association with different factors. Tartu, 1999.
58. **Heili Varendi**. Behavioral effects observed in human newborns during exposure to naturally occurring odors. Tartu, 1999.
59. **Anneli Beilmann**. Epidemiology of epilepsy in children and adolescents in Estonia. Prevalence, incidence, and clinical characteristics. Tartu, 1999.
60. **Vallo Volke**. Pharmacological and biochemical studies on nitric oxide in the regulation of behaviour. Tartu, 1999.
61. **Pilvi Ilves**. Hypoxic-ischaemic encephalopathy in asphyxiated term infants. A prospective clinical, biochemical, ultrasonographical study. Tartu, 1999.
62. **Anti Kalda**. Oxygen-glucose deprivation-induced neuronal death and its pharmacological prevention in cerebellar granule cells. Tartu, 1999.
63. **Eve-Irene Lepist**. Oral peptide prodrugs – studies on stability and absorption. Tartu, 2000.

64. **Jana Kivastik**. Lung function in Estonian schoolchildren: relationship with anthropometric indices and respiratory symptomas, reference values for dynamic spirometry. Tartu, 2000.
65. **Karin Kull**. Inflammatory bowel disease: an immunogenetic study. Tartu, 2000.
66. **Kaire Innos**. Epidemiological resources in Estonia: data sources, their quality and feasibility of cohort studies. Tartu, 2000.
67. **Tamara Vorobjova**. Immune response to *Helicobacter pylori* and its association with dynamics of chronic gastritis and epithelial cell turnover in antrum and corpus. Tartu, 2001.
68. **Ruth Kalda**. Structure and outcome of family practice quality in the changing health care system of Estonia. Tartu, 2001.
69. **Annika Krüüner**. *Mycobacterium tuberculosis* – spread and drug resistance in Estonia. Tartu, 2001.
70. **Marlit Veldi**. Obstructive Sleep Apnoea: Computerized Endopharyngeal Myotonometry of the Soft Palate and Lingual Musculature. Tartu, 2001.
71. **Anneli Uusküla**. Epidemiology of sexually transmitted diseases in Estonia in 1990–2000. Tartu, 2001.
72. **Ade Kallas**. Characterization of antibodies to coagulation factor VIII. Tartu, 2002.
73. **Heidi Annuk**. Selection of medicinal plants and intestinal lactobacilli as antimicrobil components for functional foods. Tartu, 2002.
74. **Aet Lukmann**. Early rehabilitation of patients with ischaemic heart disease after surgical revascularization of the myocardium: assessment of health-related quality of life, cardiopulmonary reserve and oxidative stress. A clinical study. Tartu, 2002.
75. **Maigi Eisen**. Pathogenesis of Contact Dermatitis: participation of Oxidative Stress. A clinical – biochemical study. Tartu, 2002.
76. **Piret Hussar**. Histology of the post-traumatic bone repair in rats. Elaboration and use of a new standardized experimental model – bicortical perforation of tibia compared to internal fracture and resection osteotomy. Tartu, 2002.
77. **Tõnu Rätsep**. Aneurysmal subarachnoid haemorrhage: Noninvasive monitoring of cerebral haemodynamics. Tartu, 2002.
78. **Marju Herodes**. Quality of life of people with epilepsy in Estonia. Tartu, 2003.
79. **Katre Maasalu**. Changes in bone quality due to age and genetic disorders and their clinical expressions in Estonia. Tartu, 2003.
80. **Toomas Sillakivi**. Perforated peptic ulcer in Estonia: epidemiology, risk factors and relations with *Helicobacter pylori*. Tartu, 2003.
81. **Leena Puksa**. Late responses in motor nerve conduction studies. F and A waves in normal subjects and patients with neuropathies. Tartu, 2003.
82. **Krista Lõivukene**. *Helicobacter pylori* in gastric microbial ecology and its antimicrobial susceptibility pattern. Tartu, 2003.

83. **Helgi Kolk**. Dyspepsia and *Helicobacter pylori* infection: the diagnostic value of symptoms, treatment and follow-up of patients referred for upper gastrointestinal endoscopy by family physicians. Tartu, 2003.
84. **Helena Soomer**. Validation of identification and age estimation methods in forensic odontology. Tartu, 2003.
85. **Kersti Oselin**. Studies on the human MDR1, MRP1, and MRP2 ABC transporters: functional relevance of the genetic polymorphisms in the *MDR1* and *MRP1* gene. Tartu, 2003.
86. **Jaan Soplepmann**. Peptic ulcer haemorrhage in Estonia: epidemiology, prognostic factors, treatment and outcome. Tartu, 2003.
87. **Margot Peetsalu**. Long-term follow-up after vagotomy in duodenal ulcer disease: recurrent ulcer, changes in the function, morphology and *Helicobacter pylori* colonisation of the gastric mucosa. Tartu, 2003.
88. **Kersti Klaamas**. Humoral immune response to *Helicobacter pylori* a study of host-dependent and microbial factors. Tartu, 2003.
89. **Pille Taba**. Epidemiology of Parkinson's disease in Tartu, Estonia. Prevalence, incidence, clinical characteristics, and pharmacoepidemiology. Tartu, 2003.
90. **Alar Veraksitš**. Characterization of behavioural and biochemical phenotype of cholecystokinin-2 receptor deficient mice: changes in the function of the dopamine and endopioidergic system. Tartu, 2003.
91. **Ingrid Kalev**. CC-chemokine receptor 5 (CCR5) gene polymorphism in Estonians and in patients with Type I and Type II diabetes mellitus. Tartu, 2003.
92. **Lumme Kadaja**. Molecular approach to the regulation of mitochondrial function in oxidative muscle cells. Tartu, 2003.
93. **Aive Liigant**. Epidemiology of primary central nervous system tumours in Estonia from 1986 to 1996. Clinical characteristics, incidence, survival and prognostic factors. Tartu, 2004.
94. **Andres, Kulla**. Molecular characteristics of mesenchymal stroma in human astrocytic gliomas. Tartu, 2004.
95. **Mari Järvelaid**. Health damaging risk behaviours in adolescence. Tartu, 2004.
96. **Ülle Pechter**. Progression prevention strategies in chronic renal failure and hypertension. An experimental and clinical study. Tartu, 2004.
97. **Gunnar Tasa**. Polymorphic glutathione S-transferases – biology and role in modifying genetic susceptibility to senile cataract and primary open angle glaucoma. Tartu, 2004.
98. **Tuuli Käämbre**. Intracellular energetic unit: structural and functional aspects. Tartu, 2004.
99. **Vitali Vassiljev**. Influence of nitric oxide syntase inhibitors on the effects of ethanol after acute and chronic ethanol administration and withdrawal. Tartu, 2004.

100. **Aune Rehema**. Assessment of nonhaem ferrous iron and glutathione redox ratio as markers of pathogeneticity of oxidative stress in different clinical groups. Tartu, 2004.
101. **Evelin Seppet**. Interaction of mitochondria and ATPases in oxidative muscle cells in normal and pathological conditions. Tartu, 2004.
102. **Eduard Maron**. Serotonin function in panic disorder: from clinical experiments to brain imaging and genetics. Tartu, 2004.
103. **Marje Oona**. *Helicobacter pylori* infection in children: epidemiological and therapeutic aspects. Tartu, 2004.
104. **Kersti Kokk**. Regulation of active and passive molecular transport in the testis. Tartu, 2005.
105. **Vladimir Järv**. Cross-sectional imaging for pretreatment evaluation and follow-up of pelvic malignant tumours. Tartu, 2005.
106. **Andre Õun**. Epidemiology of adult epilepsy in Tartu, Estonia. Incidence, prevalence and medical treatment. Tartu, 2005.
107. **Piibe Muda**. Homocysteine and hypertension: associations between homocysteine and essential hypertension in treated and untreated hypertensive patients with and without coronary artery disease. Tartu, 2005.
108. **Külli Kingo**. The interleukin-10 family cytokines gene polymorphisms in plaque psoriasis. Tartu, 2005.
109. **Mati Merila**. Anatomy and clinical relevance of the glenohumeral joint capsule and ligaments. Tartu, 2005.
110. **Epp Songisepp**. Evaluation of technological and functional properties of the new probiotic *Lactobacillus fermentum* ME-3. Tartu, 2005.
111. **Tiia Ainla**. Acute myocardial infarction in Estonia: clinical characteristics, management and outcome. Tartu, 2005.
112. **Andres Sell**. Determining the minimum local anaesthetic requirements for hip replacement surgery under spinal anaesthesia – a study employing a spinal catheter. Tartu, 2005.
113. **Tiia Tamme**. Epidemiology of odontogenic tumours in Estonia. Pathogenesis and clinical behaviour of ameloblastoma. Tartu, 2005.
114. **Triine Annus**. Allergy in Estonian schoolchildren: time trends and characteristics. Tartu, 2005.
115. **Tiia Voor**. Microorganisms in infancy and development of allergy: comparison of Estonian and Swedish children. Tartu, 2005.
116. **Priit Kasenõmm**. Indicators for tonsillectomy in adults with recurrent tonsillitis – clinical, microbiological and pathomorphological investigations. Tartu, 2005.
117. **Eva Zusinaite**. Hepatitis C virus: genotype identification and interactions between viral proteases. Tartu, 2005.
118. **Piret Kõll**. Oral lactoflora in chronic periodontitis and periodontal health. Tartu, 2006.
119. **Tiina Stelmach**. Epidemiology of cerebral palsy and unfavourable neurodevelopmental outcome in child population of Tartu city and county, Estonia Prevalence, clinical features and risk factors. Tartu, 2006.

120. **Katrin Pudersell**. Tropane alkaloid production and riboflavine excretion in the field and tissue cultures of henbane (*Hyoscyamus niger* L.). Tartu, 2006.
121. **Külli Jaako**. Studies on the role of neurogenesis in brain plasticity. Tartu, 2006.
122. **Aare Märtson**. Lower limb lengthening: experimental studies of bone regeneration and long-term clinical results. Tartu, 2006.
123. **Heli Tähepõld**. Patient consultation in family medicine. Tartu, 2006.
124. **Stanislav Liskmann**. Peri-implant disease: pathogenesis, diagnosis and treatment in view of both inflammation and oxidative stress profiling. Tartu, 2006.
125. **Ruth Rudissaar**. Neuropharmacology of atypical antipsychotics and an animal model of psychosis. Tartu, 2006.
126. **Helena Andreson**. Diversity of *Helicobacter pylori* genotypes in Estonian patients with chronic inflammatory gastric diseases. Tartu, 2006.
127. **Katrin Pruus**. Mechanism of action of antidepressants: aspects of serotoninergic system and its interaction with glutamate. Tartu, 2006.
128. **Priit Põder**. Clinical and experimental investigation: relationship of ischaemia/reperfusion injury with oxidative stress in abdominal aortic aneurysm repair and in extracranial brain artery endarterectomy and possibilities of protection against ischaemia using a glutathione analogue in a rat model of global brain ischaemia. Tartu, 2006.
129. **Marika Tammaru**. Patient-reported outcome measurement in rheumatoid arthritis. Tartu, 2006.
130. **Tiia Reimand**. Down syndrome in Estonia. Tartu, 2006.
131. **Diva Eensoo**. Risk-taking in traffic and Markers of Risk-Taking Behaviour in Schoolchildren and Car Drivers. Tartu, 2007.
132. **Riina Vibo**. The third stroke registry in Tartu, Estonia from 2001 to 2003: incidence, case-fatality, risk factors and long-term outcome. Tartu, 2007.
133. **Chris Pruunsild**. Juvenile idiopathic arthritis in children in Estonia. Tartu, 2007.
134. **Eve Õiglane-Šlik**. Angelman and Prader-Willi syndromes in Estonia. Tartu, 2007.
135. **Kadri Haller**. Antibodies to follicle stimulating hormone. Significance in female infertility. Tartu, 2007.
136. **Pille Ööpik**. Management of depression in family medicine. Tartu, 2007.
137. **Jaak Kals**. Endothelial function and arterial stiffness in patients with atherosclerosis and in healthy subjects. Tartu, 2007.
138. **Priit Kampus**. Impact of inflammation, oxidative stress and age on arterial stiffness and carotid artery intima-media thickness. Tartu, 2007.
139. **Margus Punab**. Male fertility and its risk factors in Estonia. Tartu, 2007.
140. **Alar Toom**. Heterotopic ossification after total hip arthroplasty: clinical and pathogenetic investigation. Tartu, 2007.

141. **Lea Pehme**. Epidemiology of tuberculosis in Estonia 1991–2003 with special regard to extrapulmonary tuberculosis and delay in diagnosis of pulmonary tuberculosis. Tartu, 2007.
142. **Juri Karjagin**. The pharmacokinetics of metronidazole and meropenem in septic shock. Tartu, 2007.
143. **Inga Talvik**. Inflicted traumatic brain injury shaken baby syndrome in Estonia – epidemiology and outcome. Tartu, 2007.
144. **Tarvo Rajasalu**. Autoimmune diabetes: an immunological study of type 1 diabetes in humans and in a model of experimental diabetes (in RIP-B7.1 mice). Tartu, 2007.
145. **Inga Karu**. Ischaemia-reperfusion injury of the heart during coronary surgery: a clinical study investigating the effect of hyperoxia. Tartu, 2007.
146. **Peeter Padrik**. Renal cell carcinoma: Changes in natural history and treatment of metastatic disease. Tartu, 2007.
147. **Neve Vendt**. Iron deficiency and iron deficiency anaemia in infants aged 9 to 12 months in Estonia. Tartu, 2008.
148. **Lenne-Triin Heidmets**. The effects of neurotoxins on brain plasticity: focus on neural Cell Adhesion Molecule. Tartu, 2008.
149. **Paul Korrovits**. Asymptomatic inflammatory prostatitis: prevalence, etiological factors, diagnostic tools. Tartu, 2008.
150. **Annika Reintam**. Gastrointestinal failure in intensive care patients. Tartu, 2008.
151. **Kristiina Roots**. Cationic regulation of Na-pump in the normal, Alzheimer's and CCK_2 receptor-deficient brain. Tartu, 2008.
152. **Helen Puusepp**. The genetic causes of mental retardation in Estonia: fragile X syndrome and creatine transporter defect. Tartu, 2009.
153. **Kristiina Rull**. Human chorionic gonadotropin beta genes and recurrent miscarriage: expression and variation study. Tartu, 2009.
154. **Margus Eimre**. Organization of energy transfer and feedback regulation in oxidative muscle cells. Tartu, 2009.
155. **Maire Link**. Transcription factors FoxP3 and AIRE: autoantibody associations. Tartu, 2009.
156. **Kai Haldre**. Sexual health and behaviour of young women in Estonia. Tartu, 2009.
157. **Kaur Liivak**. Classical form of congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Estonia: incidence, genotype and phenotype with special attention to short-term growth and 24-hour blood pressure. Tartu, 2009.
158. **Kersti Ehrlich**. Antioxidative glutathione analogues (UPF peptides) – molecular design, structure-activity relationships and testing the protective properties. Tartu, 2009.
159. **Anneli Rätsep**. Type 2 diabetes care in family medicine. Tartu, 2009.
160. **Silver Türk**. Etiopathogenetic aspects of chronic prostatitis: role of mycoplasmas, coryneform bacteria and oxidative stress. Tartu, 2009.

161. **Kaire Heilman**. Risk markers for cardiovascular disease and low bone mineral density in children with type 1 diabetes. Tartu, 2009.
162. **Kristi Rüütel**. HIV-epidemic in Estonia: injecting drug use and quality of life of people living with HIV. Tartu, 2009.
163. **Triin Eller**. Immune markers in major depression and in antidepressive treatment. Tartu, 2009.
164. **Siim Suutre**. The role of TGF-β isoforms and osteoprogenitor cells in the pathogenesis of heterotopic ossification. An experimental and clinical study of hip arthroplasty. Tartu, 2010.
165. **Kai Kliiman**. Highly drug-resistant tuberculosis in Estonia: Risk factors and predictors of poor treatment outcome. Tartu, 2010.
166. **Inga Villa**. Cardiovascular health-related nutrition, physical activity and fitness in Estonia. Tartu, 2010.
167. **Tõnis Org**. Molecular function of the first PHD finger domain of Autoimmune Regulator protein. Tartu, 2010.
168. **Tuuli Metsvaht**. Optimal antibacterial therapy of neonates at risk of early onset sepsis. Tartu, 2010.
169. **Jaanus Kahu**. Kidney transplantation: Studies on donor risk factors and mycophenolate mofetil. Tartu, 2010.
170. **Koit Reimand**. Autoimmunity in reproductive failure: A study on associated autoantibodies and autoantigens. Tartu, 2010.
171. **Mart Kull**. Impact of vitamin D and hypolactasia on bone mineral density: a population based study in Estonia. Tartu, 2010.
172. **Rael Laugesaar**. Stroke in children – epidemiology and risk factors. Tartu, 2010.
173. **Mark Braschinsky**. Epidemiology and quality of life issues of hereditary spastic paraplegia in Estonia and implemention of genetic analysis in everyday neurologic practice. Tartu, 2010.
174. **Kadri Suija**. Major depression in family medicine: associated factors, recurrence and possible intervention. Tartu, 2010.
175. **Jarno Habicht**. Health care utilisation in Estonia: socioeconomic determinants and financial burden of out-of-pocket payments. Tartu, 2010.
176. **Kristi Abram**. The prevalence and risk factors of rosacea. Subjective disease perception of rosacea patients. Tartu, 2010.
177. **Malle Kuum**. Mitochondrial and endoplasmic reticulum cation fluxes: Novel roles in cellular physiology. Tartu, 2010.
178. **Rita Teek**. The genetic causes of early onset hearing loss in Estonian children. Tartu, 2010.
179. **Daisy Volmer**. The development of community pharmacy services in Estonia – public and professional perceptions 1993–2006. Tartu, 2010.
180. **Jelena Lissitsina**. Cytogenetic causes in male infertility. Tartu, 2011.
181. **Delia Lepik**. Comparison of gunshot injuries caused from Tokarev, Makarov and Glock 19 pistols at different firing distances. Tartu, 2011.
182. **Ene-Renate Pähkla**. Factors related to the efficiency of treatment of advanced periodontitis. Tartu, 2011.

183. **Maarja Krass**. L-Arginine pathways and antidepressant action. Tartu, 2011.
184. **Taavi Lai**. Population health measures to support evidence-based health policy in Estonia. Tartu, 2011.
185. **Tiit Salum**. Similarity and difference of temperature-dependence of the brain sodium pump in normal, different neuropathological, and aberrant conditions and its possible reasons. Tartu, 2011.
186. **Tõnu Vooder**. Molecular differences and similarities between histological subtypes of non-small cell lung cancer. Tartu, 2011.
187. **Jelena Štšepetova**. The characterisation of intestinal lactic acid bacteria using bacteriological, biochemical and molecular approaches. Tartu, 2011.
188. **Radko Avi**. Natural polymorphisms and transmitted drug resistance in Estonian HIV-1 CRF06_cpx and its recombinant viruses. Tartu, 2011, 116 p.
189. **Edward Laane**. Multiparameter flow cytometry in haematological malignancies. Tartu, 2011, 152 p.
190. **Triin Jagomägi**. A study of the genetic etiology of nonsyndromic cleft lip and palate. Tartu, 2011, 158 p.
191. **Ivo Laidmäe**. Fibrin glue of fish (*Salmo salar*) origin: immunological study and development of new pharmaceutical preparation. Tartu, 2012, 150 p.
192. **Ülle Parm**. Early mucosal colonisation and its role in prediction of invasive infection in neonates at risk of early onset sepsis. Tartu, 2012, 168 p.
193. **Kaupo Teesalu**. Autoantibodies against desmin and transglutaminase 2 in celiac disease: diagnostic and functional significance. Tartu, 2012, 142 p.
194. **Maksim Zagura**. Biochemical, functional and structural profiling of arterial damage in atherosclerosis. Tartu, 2012, 162 p.
195. **Vivian Kont**. Autoimmune regulator: characterization of thymic gene regulation and promoter methylation. Tartu, 2012, 134 p.
196. **Pirje Hütt**. Functional properties, persistence, safety and efficacy of potential probiotic lactobacilli. Tartu, 2012, 246 p.
197. **Innar Tõru**. Serotonergic modulation of CCK-4- induced panic. Tartu, 2012, 132 p.
198. **Sigrid Vorobjov**. Drug use, related risk behaviour and harm reduction interventions utilization among injecting drug users in Estonia: implications for drug policy. Tartu, 2012, 120 p.
199. **Martin Serg**. Therapeutic aspects of central haemodynamics, arterial stiffness and oxidative stress in hypertension. Tartu, 2012, 156 p.
200. **Jaanika Kumm**. Molecular markers of articular tissues in early knee osteoarthritis: a population-based longitudinal study in middle-aged subjects. Tartu, 2012, 159 p.
201. **Kertu Rünkorg**. Functional changes of dopamine, endopioid and endocannabinoid systems in CCK2 receptor deficient mice. Tartu, 2012, 125 p.
202. **Mai Blöndal**. Changes in the baseline characteristics, management and outcomes of acute myocardial infarction in Estonia. Tartu, 2012, 127 p.

203. **Jana Lass**. Epidemiological and clinical aspects of medicines use in children in Estonia. Tartu, 2012, 170 p.
204. **Kai Truusalu**. Probiotic lactobacilli in experimental persistent *Salmonella* infection. Tartu, 2013, 139 p.
205. **Oksana Jagur**. Temporomandibular joint diagnostic imaging in relation to pain and bone characteristics. Long-term results of arthroscopic treatment. Tartu, 2013, 126 p.
206. **Katrin Sikk**. Manganese-ephedrone intoxication – pathogenesis of neurological damage and clinical symptomatology. Tartu, 2013, 125 p.
207. **Kai Blöndal**. Tuberculosis in Estonia with special emphasis on drug-resistant tuberculosis: Notification rate, disease recurrence and mortality. Tartu, 2013, 151 p.
208. **Marju Puurand**. Oxidative phosphorylation in different diseases of gastric mucosa. Tartu, 2013, 123 p.
209. **Aili Tagoma**. Immune activation in female infertility: Significance of autoantibodies and inflammatory mediators. Tartu, 2013, 135 p.
210. **Liis Sabre**. Epidemiology of traumatic spinal cord injury in Estonia. Brain activation in the acute phase of traumatic spinal cord injury. Tartu, 2013, 135 p.
211. **Merit Lamp**. Genetic susceptibility factors in endometriosis. Tartu, 2013, 125 p.
212. **Erik Salum**. Beneficial effects of vitamin D and angiotensin II receptor blocker on arterial damage. Tartu, 2013, 167 p.
213. **Maire Karelson**. Vitiligo: clinical aspects, quality of life and the role of melanocortin system in pathogenesis. Tartu, 2013, 153 p.
214. **Kuldar Kaljurand**. Prevalence of exfoliation syndrome in Estonia and its clinical significance. Tartu, 2013, 113 p.
215. **Raido Paasma**. Clinical study of methanol poisoning: handling large outbreaks, treatment with antidotes, and long-term outcomes. Tartu, 2013, 96 p.
216. **Anne Kleinberg**. Major depression in Estonia: prevalence, associated factors, and use of health services. Tartu, 2013, 129 p.
217. **Triin Eglit**. Obesity, impaired glucose regulation, metabolic syndrome and their associations with high-molecular-weight adiponectin levels. Tartu, 2014, 115 p.
218. **Kristo Ausmees**. Reproductive function in middle-aged males: Associations with prostate, lifestyle and couple infertility status. Tartu, 2014, 125 p.
219. **Kristi Huik**. The influence of host genetic factors on the susceptibility to HIV and HCV infections among intravenous drug users. Tartu, 2014, 144 p.
220. **Liina Tserel**. Epigenetic profiles of monocytes, monocyte-derived macrophages and dendritic cells. Tartu, 2014, 143 p.
221. **Irina Kerna**. The contribution of *ADAM12* and *CILP* genes to the development of knee osteoarthritis. Tartu, 2014, 152 p.

222. **Ingrid Liiv**. Autoimmune regulator protein interaction with DNA-dependent protein kinase and its role in apoptosis. Tartu, 2014, 143 p.
223. **Liivi Maddison**. Tissue perfusion and metabolism during intra-abdominal hypertension. Tartu, 2014, 103 p.
224. **Krista Ress**. Childhood coeliac disease in Estonia, prevalence in atopic dermatitis and immunological characterisation of coexistence. Tartu, 2014, 124 p.
225. **Kai Muru**. Prenatal screening strategies, long-term outcome of children with marked changes in maternal screening tests and the most common syndromic heart anomalies in Estonia. Tartu, 2014, 189 p.
226. **Kaja Rahu**. Morbidity and mortality among Baltic Chernobyl cleanup workers: a register-based cohort study. Tartu, 2014, 155 p.
227. **Klari Noormets**. The development of diabetes mellitus, fertility and energy metabolism disturbances in a Wfs1-deficient mouse model of Wolfram syndrome. Tartu, 2014, 132 p.
228. **Liis Toome**. Very low gestational age infants in Estonia. Tartu, 2014, 183 p.
229. **Ceith Nikkolo**. Impact of different mesh parameters on chronic pain and foreign body feeling after open inguinal hernia repair. Tartu, 2014, 132 p.
230. **Vadim Brjalin**. Chronic hepatitis C: predictors of treatment response in Estonian patients. Tartu, 2014, 122 p.
231. **Vahur Metsna**. Anterior knee pain in patients following total knee arthroplasty: the prevalence, correlation with patellar cartilage impairment and aspects of patellofemoral congruence. Tartu, 2014, 130 p.
232. **Marju Kase**. Glioblastoma multiforme: possibilities to improve treatment efficacy. Tartu, 2015, 137 p.
233. **Riina Runnel**. Oral health among elementary school children and the effects of polyol candies on the prevention of dental caries. Tartu, 2015, 112 p.
234. **Made Laanpere**. Factors influencing women's sexual health and reproductive choices in Estonia. Tartu, 2015, 176 p.
235. **Andres Lust**. Water mediated solid state transformations of a polymorphic drug – effect on pharmaceutical product performance. Tartu, 2015, 134 p.
236. **Anna Klugman**. Functionality related characterization of pretreated wood lignin, cellulose and polyvinylpyrrolidone for pharmaceutical applications. Tartu, 2015, 156 p.
237. **Triin Laisk-Podar**. Genetic variation as a modulator of susceptibility to female infertility and a source for potential biomarkers. Tartu, 2015, 155 p.
238. **Mailis Tõnisson**. Clinical picture and biochemical changes in blood in children with acute alcohol intoxication. Tartu, 2015, 100 p.
239. **Kadri Tamme**. High volume haemodiafiltration in treatment of severe sepsis – impact on pharmacokinetics of antibiotics and inflammatory response. Tartu, 2015, 133 p.

240. **Kai Part**. Sexual health of young people in Estonia in a social context: the role of school-based sexuality education and youth-friendly counseling services. Tartu, 2015, 203 p.
241. **Urve Paaver**. New perspectives for the amorphization and physical stabilization of poorly water-soluble drugs and understanding their dissolution behavior. Tartu, 2015, 139 p.
242. **Aleksandr Peet**. Intrauterine and postnatal growth in children with HLA-conferred susceptibility to type 1 diabetes. Tartu. 2015, 146 p.
243. **Piret Mitt**. Healthcare-associated infections in Estonia – epidemiology and surveillance of bloodstream and surgical site infections. Tartu, 2015, 145 p.
244. **Merli Saare**. Molecular Profiling of Endometriotic Lesions and Endometria of Endometriosis Patients. Tartu, 2016, 129 p.
245. **Kaja-Triin Laisaar**. People living with HIV in Estonia: Engagement in medical care and methods of increasing adherence to antiretroviral therapy and safe sexual behavior. Tartu, 2016, 132 p.
246. **Eero Merilind**. Primary health care performance: impact of payment and practice-based characteristics. Tartu, 2016, 120 p.
247. **Jaanika Kärner**. Cytokine-specific autoantibodies in AIRE deficiency. Tartu, 2016, 182 p.
248. **Kaido Paapstel**. Metabolomic profile of arterial stiffness and early biomarkers of renal damage in atherosclerosis. Tartu, 2016, 173 p.
249. **Liidia Kiisk**. Long-term nutritional study: anthropometrical and clinico-laboratory assessments in renal replacement therapy patients after intensive nutritional counselling. Tartu, 2016, 207 p.
250. **Georgi Nellis**. The use of excipients in medicines administered to neonates in Europe. Tartu, 2017, 159 p.
251. **Aleksei Rakitin**. Metabolic effects of acute and chronic treatment with valproic acid in people with epilepsy. Tartu, 2017, 125 p.
252. **Eveli Kallas**. The influence of immunological markers to susceptibility to HIV, HBV, and HCV infections among persons who inject drugs. Tartu, 2017, 138 p.
253. **Tiina Freimann**. Musculoskeletal pain among nurses: prevalence, risk factors, and intervention. Tartu, 2017, 125 p.
254. **Evelyn Aaviksoo**. Sickness absence in Estonia: determinants and influence of the sick-pay cut reform. Tartu, 2017, 121 p.
255. **Kalev Nõupuu**. Autosomal-recessive Stargardt disease: phenotypic heterogeneity and genotype-phenotype associations. Tartu, 2017, 131 p.
256. **Ho Duy Binh**. Osteogenesis imperfecta in Vietnam. Tartu, 2017, 125 p.
257. **Uku Haljasorg**. Transcriptional mechanisms in thymic central tolerance. Tartu, 2017, 147 p.
258. **Živile Riispere**. IgA Nephropathy study according to the Oxford Classification: IgA Nephropathy clinical-morphological correlations, disease progression and the effect of renoprotective therapy. Tartu, 2017, 129 p.

259. **Hiie Soeorg**. Coagulase-negative staphylococci in gut of preterm neonates and in breast milk of their mothers. Tartu, 2017, 216 p.
260. **Anne-Mari Anton Willmore**. Silver nanoparticles for cancer research. Tartu, 2017, 132 p.
261. **Ott Laius**. Utilization of osteoporosis medicines, medication adherence and the trend in osteoporosis related hip fractures in Estonia. Tartu, 2017, 134 p.
262. **Alar Aab**. Insights into molecular mechanisms of asthma and atopic dermatitis. Tartu, 2017, 164 p.