Discovering the genetic cause of Mendelian disorders in the age of genomics:

the evolving capability of next-generation DNA sequencing

Cover design by: Proefschriftmaken || www.proefschriftmaken.nl
Printed by: Proefschriftmaken || www.proefschriftmaken.nl
Layout by: Scoutcollective || https://www.facebook.com/scoutcollective/
Published by: Proefschriftmaken || www.proefschriftmaken.nl
ISBN: 978-90-393-6854-1

Discovering the genetic cause of Mendelian disorders in the age of genomics:

the evolving capability of next-generation DNA sequencing

De genetische oorzaak van Mendeliaanse aandoeningen ontdekken in het genomica tijdperk:

de toenemende mogelijkheden van next-generation DNA sequencing

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 24 november 2017 des middags te 4.15 uur

door

Glen Ryan Monroe

geboren op 5 mei 1979
te New Haven, Connecticut
Verenigde Staten van Amerika

Promotor: Prof.dr. E.P.J.G Cuppen

Co-promotor: Dr. G.W. van Haaften

To my beloved grandmother,
who is always interested in my work,
is curious about genome sequencing
and mails me articles about George Church.
I come from great genes. All of my love!

Contents

1

Introduction

Modern healthcare has always relied upon technological innovation to study, diagnose, and ultimately treat patients better. Our increasing ability to see inside the body has allowed clinicians and researchers to understand fundamentally the complex interplay between tissues, organs, cells and molecules – and to recognize indications of disease. From the discovery of X-rays (1895) and nuclear magnetic resonance (1938) to the application of these discoveries and others in the clinic as the commercial X-ray machine, MRI scanner, ultrasound, and CT scanner, the ability to see inside the body – without cutting it open – has provided information on its state and guided care (Rabi et al. 1938; Rontgen 1896; Spiegel 1995). In the last 40 years, the ability to *see* the sequence of our DNA, and interpret when it is different from a healthy reference DNA sequence, has resulted in an expanding ability to understand how genetics contributes to health and disease. The technological innovations in DNA sequencing have resulted in an increased ability to analyze variation in patients' DNA and ultimately to provide a diagnosis. The impact of this technology has been particularly powerful in providing a diagnosis for those patients with rare genetic diseases.

RARE DISEASES, MENDELIAN DISORDERS, AND DETERMINING THE DIAGNOSIS

Rare diseases

A rare disease in Europe is defined as affecting fewer than 1 in 2,000 people (European Parliment 1999; Eurordis 2017). In the United States, a rare disease is defined as affecting less than 200,000 people total in the USA, an approximation that roughly equals the EU definition (United States House of Representatives 2002). This doesn't seem like a lot of people. In a country such as the Netherlands with a population of ~17 million, if a disease only affects 1 in 2,000 people there are less than 8500 people in the whole country with that particular rare disease - and often the prevalence is much lower. However, there are over 7000 rare diseases, collectively affecting a large portion of the population (Chong et al. 2015). In the USA, 25-30 million people are estimated to have a rare disease; similarly, in Europe a rare disease afflicts 6-8% of the population and potentially affects 30 million individuals (Costa et al. 1985; Eurordis 2017). Patients with rare diseases account for a disproportionate amount of healthcare resource use, with longer stays, more frequent visits, and a higher cost than other patient groups (Walker et al. 2017).

Mendelian disorders

A large portion of rare diseases has a genetic component. Nearly 20% of pediatric hospital admissions are directly attributable to a genetic cause (Kumar et al. 2001). Of those rare diseases that are genetic in origin, many are monogenic disorders – that is, a disorder caused by a defect in the DNA of a single gene (See Box 1).

BOX 1 | DNA, Genes and Genetic Variation. Deoxyribonucleic acid (DNA) is the molecular unit of inheritance (Avery et al. 1944). It consists of four nucleotides composed of different nucleobases: Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). First isolated as "nuclein" in 1869 and the molecular structure determined in 1953, DNA is structured in two linear strands that run antiparallel to one another in a double helix, with complementary nucleobases pairing to each other through hydrogen bonds (A pairs with T; G with C) (Dahm 2005; Miescher-Rüsch 1871; Watson and Crick 1953; Wilkins et al. 1953). In humans, DNA is compacted into chromosomes: 22 autosomes and the sex chromosomes X and Y. When you are born you inherit one haploid set of chromosomes from your father and one haploid set from your mother, for a total diploid genome of 46 chromosomes (44 autosomes and 2 sex chromosomes, XY determining if you will become a boy or XX if you will become a girl) consisting of a total of ~6.4 billion nucleotides (GRCh38.p10) (Genome Reference Consortium 2017). The DNA of these chromosomes is organized into specific, ordered sequences called genes. There are ~20,500 protein coding genes in human cells (Clamp et al. 2007). Genes primarily encode for proteins that are the functional effectors of the cell, performing cellular tasks. At each of the ~3.2 billion positions in the genome you inherit one **allele** from your father and one from your mother. These two alleles can both be the same as the reference genome (homozygous reference), one allele can differ from the reference while the other is the same as the reference (heterozygous variant) or both can differ from the reference (homozygous

(BOX 1 | *Continued*)

variant). The DNA composition from person to person varies, either as a single nucleotide variant level (SNV), a small insertion or deletion of <50 nucleotides (indel), a copy number variant (CNV), a copy neutral variant (translocations, inversions) or even larger, more complex structural changes (i.e. chromothripsis; chromosomal aneuploidies) (de Pagter et al. 2015). At the organism level, these variants result in the **phenotype**, the observable characteristics, of each one of us. If the DNA variant is deleterious and impairs or affects gene function, the gene may not be able to code for the correct protein at the right time, in the right tissue, or in the correct structure and may have an impact upon the development of an embryo or how our body functions. For some, this can ultimately lead to disease.

Of course, the majority of DNA variation does not result in disease. The DNA of two unrelated individuals is 99.5% identical (Gonzaga-Jauregui et al. 2012; Levy et al. 2007). Large-scale DNA sequencing projects have established that each of us is different from the reference genome at ~4.1-5 million positions, and have ~40,000 – 200,000 rare variants at a population minor allele frequency < 0.5% (The Genomes Project Consortium 2015). It is the interplay between DNA variation and our lifestyle, environment, and diet that results in the unique phenotype of each one of us.

Monogenic disorders are also known as **Mendelian disorders**, after the 19th century monk Gregor Mendel who worked at the Augustian St Thomas's Abbey in Brno, then part of the Austrian-Hungary Empire. In 1866 he published his "Versuche über Pflanzen-hybriden", a study based on his meticulous pea plant breeding experiments (Mendel 1996). Therein lay the fundamentals of inheritance – that traits could be passed to offspring, and that these traits (or at least the traits that he luckily studied) did not blend and segregated independently. Mendelian disorders are disorders in which the manifestation of disease follows Mendelian inheritance patterns. **Autosomal dominant** diseases are those diseases in which only one allele at a particular position needs to be a variant to get the disease (**heterozygous variant** at that position). **Autosomal recessive** diseases are caused by two variant alleles at the same genomic position (**homozygous variant** at that position) or by two deleterious alleles (one from the father, one from the mother) inherited at different positions but still affecting the same gene (**compound heterozygous** variants). Additionally, in humans, diseases can by inherited in a X-linked, Y-linked, or mitochondrial manner. The ascertainment of the exact number of Mendelian disorders is difficult and continually evolving, but the most recent estimate is that there are 7315 Mendelian disorders (Amberger et al. 2015; Chong et al. 2015). The genetic cause is known for approximately two-thirds of these phenotypes only (4963) (Chong et al. 2015; Online Mendelian Inheritance in Man 2017).

DISCOVERING THE GENETIC ETIOLOGY OF MENDELIAN DISORDERS

A prime aim of modern clinical genetics has been the identification of genes in which deregulation results in disease. By determining the genetic cause of a patient's disease and providing a diagnosis, clinical geneticists can inform the patient on their prognosis, on treatment options available, the chance of passing this disease to the patient's offspring, as well as informing the parents of their reproductive risks, or even testing in further relatives for the disease. Once the genetic basis of a Mendelian disorder has been determined, a patient with that suspected disorder can be genetically tested and the patient and his or her parents can receive a diagnosis and be managed accordingly. One of the ongoing goals in clinical genetics is discovering the genetic cause of all Mendelian disorders.

Traditional methods of identifying disease genes

With the discovery that DNA encoded what Francis Crick called "the secret of life" came the realization that variation in that DNA could be responsible for disease. By the 1990s, several hundred genes known to cause disease had been identified, with most of these identified not by genetic means but with preexisting biochemical knowledge of what was causing the disease (Collins 1992). These **functional cloning** methods relied on existing knowledge of the disease mechanism or result, which in many cases was not known.

Beginning in 1980, the construction of a genome-wide linkage map and subsequent **linkage analysis** was proposed to link each disease to a genomic locus (Botstein et al. 1980). Specific genetic markers (i.e. short tandem repeats, microsatellites, or single nucleotide polymorphisms (SNPs)) are inherited together with the causal disease variant due to their proximity close to one another; closer markers are inherited together more often than farther away markers. A rough approximation of the location of the gene responsible for the disease could be determined by comparing the genetic markers present in affected individuals of large families with highly penetrant diseases but not present in healthy family members. In the mid-1980s this strategy of **positional cloning** became useful and *CYBB* was the first gene to be identified as causal for a Mendelian disease (chronic granulomatous) without existing knowledge of the function or role of the gene or molecular pathology (Royer-Pokora et al. 1986). In a few years *CFTR* was found to be the causal gene for cystic fibrosis by strictly linkage mapping (Kerem et al. 1989; Riordan et al. 1989). A specific deletion of three nucleotides was established to be causal for disruption of CFTR in approximately 70% of cystic fibrosis patients.

The positional cloning approach was quite time-consuming and very rarely successful, particularly if a large region still existed following linkage analysis. This approach was

aided by the accumulation of knowledge of the transcripts that were present at particular locations in the genome, thereby giving clues on the functional unit (gene) involved so that a smaller, more specific genetic area could be examined for variation in patients. This would result in linkage analysis gradually progressing to the "**positional candidate**" approach (Collins 1995). In this approach, once the disease locus had been mapped, genes within that locus were prioritized based upon annotated function by previously discovered information. For instance, the discovery of the gene responsible for Marfan syndrome relied upon linkage mapping to a specific region, followed by prioritization of a gene (*FBN1*: fibrillin 1) whose known function in the body as a component of elastin could be explanatory for the disease pathology of Marfan syndrome, a genetic disorder affecting connective tissue (Collins 1995; Kainulainen et al. 1990; Magenis et al. 1991; Sakai et al. 1986). This approach sped up as the locations and function of more and more protein coding genes were determined and amassed into a well-annotated high-density transcript map. Then only those genes whose annotated function could be explanatory for the disease were prioritized, in contrast to the whole disease locus. The DNA sequences of these candidate genes in patients could then be determined to identify DNA variants, and a genotype-phenotype correlation made between the DNA variant and the disease.

The development of DNA sequencing greatly accelerated Mendelian disease gene discovery. **DNA sequencing** is the determination of the order of nucleotides of the DNA in a subject. The **Sanger sequencing** method of dideoxynucleotide chain termination, initially described by Sanger et al. in 1977 and subsequently automated, has been the predominant method of DNA sequencing over the course of the last 40 years (Botstein and Risch 2003; Hunkapiller et al. 1991; Maxam and Gilbert 1977; Sanger et al. 1977; Smith et al. 1986; Swerdlow et al. 1990). In this process, an amplified DNA template is incubated with a specific reaction mix of DNA polymerase and reaction buffer, DNA primers flanking the DNA region of interest, deoxynucleotides (dATP, dTTP, dCTP, dGTP), and, initially radioactively- and later fluorescently-tagged dideoxynucleotides (ddATP, ddTTP, ddCTP, ddGTP). This double stranded DNA template then undergoes a cyclic process of 1) denaturation into single stranded DNA, 2) annealing of the primers to the single stranded DNA, and 3) finally extension of the primers – incorporating the tagged dideoxynucleotides into the resulting new DNA molecules. Once a dideoxynucleotide DNA molecule is incorporated into the growing DNA strand the DNA strand can no longer be extended. This results in DNA molecules of varying sizes, all tagged at the end with a fluorescent dideoxynucleotide corresponding to the specific nucleotide at that position. As these fragments are run through a gel-filled capillary of a DNA sequencing instrument, the smaller molecules go through first followed by the larger molecules. A laser excites the fluorescent dideoxynucleotide as they progress through the gel, and the sequence is "read" into a Sanger sequencing chromatogram (Figure 1A) (Shendure and Ji 2008).

The ability to read the sequence of a patient's DNA, and the subsequent capacity to identify nucleotides in the DNA of patients that were different than healthy individuals, would prove to be essential in the discovery of specific genes whose dysfunction could be responsible for Mendelian disease. Over the next ten years after the *CYBB* discovery, 42 more genes would be identified in the same manner – dependent on highly penetrant diseases following a Mendelian inheritance pattern, large family pedigrees, and Sanger sequencing to identify DNA variants (Collins 1995).

The completion of the first draft of the human genome project (HGP) in 2000 and elucidation of the majority of the coding genes considerably aided disease gene identification efforts (Lander 2001; Venter et al. 2001). Sanger sequencing would prove to be very successful, and Craig Venter's team notably used automated Sanger sequencing in whole-genome shotgun sequencing assembly to produce their draft of the human genome in 2001 (Venter et al. 2001). Improvements in chemistry, automation and protocols resulted in read lengths of approximately 1000 bases and accuracy of up to 99.999% per base, making Sanger sequencing the "gold standard" of DNA sequencing (Shendure and Ji 2008). However, the success of the positional candidate strategy and Sanger sequencing in rare genetic diseases was hindered for those diseases that were sufficiently infrequent to only have one or a few families affected by the disease, or that the reduction in reproductive fitness limited the number of siblings that could be used for linkage analysis (Bamshad et al. 2011; Boycott et al. 2013). Small and limited number of families resulted either in the identification of a genomic region that was simply too large to investigate or only by large-scale Sanger sequencing, which became cost-prohibitive as the region to interrogate became larger (Botstein and Risch 2003). Francis Collins himself, director of the National institutes of Health (USA), would allude to the future of genetics by stating that "powerful methods to search for sequence variation over large numbers of candidate DNA segments…may still prevail" (Collins 1995).

The rise of next generation sequencing

With the completion of the human genome draft and its revision in 2000 and 2003, respectively, the sequence of the human genome was, for the most part, known (Human Genome Sequencing Consortium International 2004; Lander 2001; Venter et al. 2001). Newly developed "next-generation sequencing (NGS) technologies" could take advantage of the fact that there was a reference human genome and any DNA sequence generated could be compared back to this reference. In essence, as long as a piece of DNA could be isolated, sequenced, and be unique enough to be mapped correctly back to the human reference genome, its genomic location could be determined (**resequencing**) (Li and Durbin 2010). Mapped DNA sequences could then also be functionally annotated with what was known about that gene or region, and the sequence compared to the

reference genome to identify genetic variation. The positions present in the patient that were non-reference (**variant**) could be analyzed to find those specific variants that could be causal for disease; though identifying the exact variant causal for the disease was (and remains to be) a significant challenge.

The first NGS instrument developed, the 454 Life Sciences GS20 (2005), relied upon the pyrosequencing technique. In pyrosequencing, pooled DNA fragments are clonally amplified, extended, and pyrophosphate is released as the DNA strand extends. Detection of the amount of pyrophosphate after the incorporation of each specific dNTP reveals the DNA sequence (Margulies et al. 2005; Nyren et al. 1993; Ronaghi et al. 1996). This newfound ability to perform DNA sequencing on many different DNA fragments at the same time (**massively parallel sequencing**) would usher in the era of next generation sequencing, and the development of a diversity of NGS approaches. Currently, the predominant method of high-throughput NGS technology is based upon the cyclic enzymatic manipulation, extension, and high-resolution imaging of densely packed DNA to reveal the DNA sequence (Mitra and Church 1999; Mitra et al. 2003; Shendure and Ji 2008). In this process, 1) genomic DNA is sheared into smaller fragments, 2) a platform-specific DNA adaptor is ligated to the ends of the sheared genomic DNA fragments, 3) DNA fragments are clonally amplified and deposited into discrete wells or positions on the sequencing instrument, 4) the specific DNA sequences are extended as new nucleotides are added one at a time and 5) the DNA sequence is "read" at each position as a new nucleotide gets incorporated into the growing DNA strand (Figure 1B) (Shendure and Ji 2008). Whereas Sanger sequencing relied upon the presence of one specific amplified DNA template for DNA sequencing, NGS technologies offered massively parallel sequencing; that is, many different amplified DNA templates could all be DNA sequenced at the same time, on the same instrument, instead of only one DNA template (or 96 DNA templates for automated Sanger sequencing). This greatly increased the scalability of DNA sequencing, allowing larger regions of DNA to be sequenced than with Sanger sequencing – and at a resulting lower cost per nucleotide base (Shendure and Ji 2008).

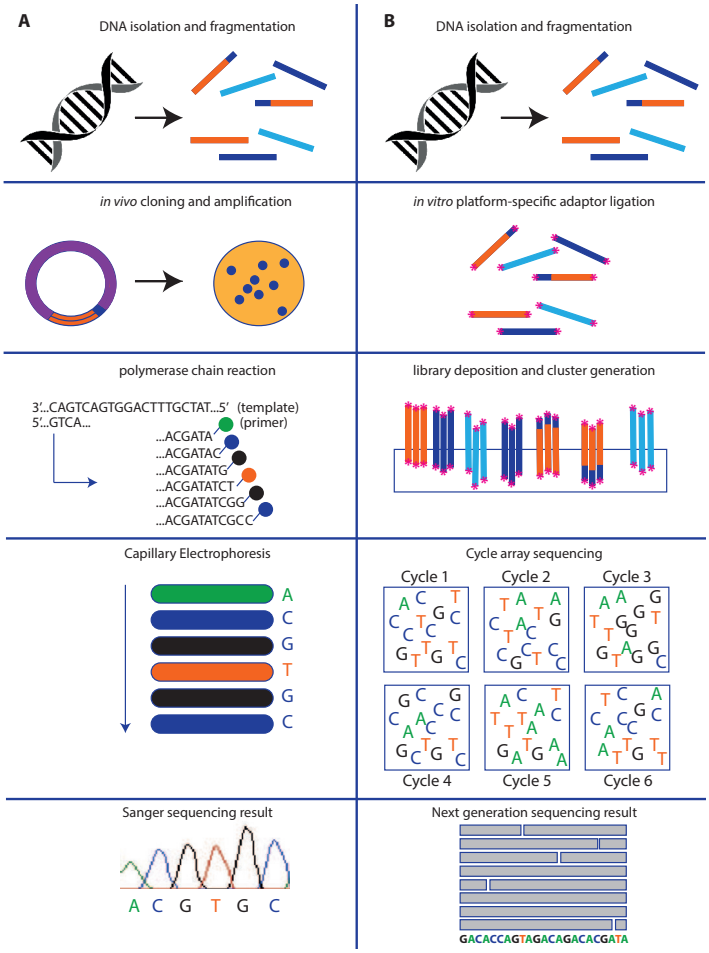


Figure 1 | Methods to sequence DNA. In high-throughput Sanger sequencing for genome assembly (A), DNA is isolated and fragmented, cloned into a vector and *E.coli* is transformed. A single bacterial colony is picked, allowed to replicate, and plasmid DNA isolated. This plasmid DNA is the template for the Sanger sequencing polymerase chain reaction. Incubating the template with dNTPs, ddNTPs, polymerase, reaction buffer and primers produces DNA template fragments of varying length, ending with a fluorescent ddNTP. These fragments are separated by size as they progress through the gel in capillary electrophoresis of a Sanger sequencing instrument. As they progress, a laser detects each base-specific fluorescent ddNTP. This results in a Sanger sequencing chromatogram, where the bases A, T, C, and G are labeled in a different color. In NGS (B), genomic DNA is isolated and fragmented. Next, sequencing platform-specific adaptors are ligated to the ends of the DNA. These adaptors allow the attachment to a surface or bead (dependent on NGS technology). On the surface (Illumina slide shown here), clusters are generated clonally from one DNA template. Cycle array sequencing occurs as the template is subject to rounds of primer hybridization, dNTP addition and imaging of the added nucleotide. Note that in Sanger sequencing only one template is DNA sequenced through a capillary. NGS, on the other hand, allows many different templates to be massively parallel sequenced at the same time. Figure adapted and expanded from (Shendure and Ji 2008).

Two publications in 2005 introduced the concept of NGS by different technical methods, and the potential of NGS was quickly realized to be transformative (Margulies et al. 2005; Shendure et al. 2005). This led to the development of various commercial instruments that have continually evolved over the following decade. These technologies would diversify and focus upon broadly-defined high-throughput applications (454 Life Sciences GS20; Illumina Genome Analyzer IIx; Applied Biosystems SOLiD; Illumina HiSeqX), long read-length generation (Pacific Biosciences Sequel; Nanopore MinION and PromethION) or "benchtop" use (Ion Torrent PGM and Proton; Illumina MiSeq) (Figure 2) (Bennett 2004; Bennett et al. 2005; Eid et al. 2009; Loman and Watson 2015; Margulies et al. 2005; Rothberg et al. 2011; Shendure and Aiden 2012; Shendure et al. 2005). The ideal instrument differs according to the application. High throughput machines are well-suited for the generation of large amounts of data, either from a patient standpoint (i.e. large cohorts) or for large sequence targets (i.e. the whole genome), though they are hindered by short read lengths, inability to resolve structural variants, and reduced accuracy compared to Sanger sequencing (Shendure and Ji 2008). Comparatively, long-read instruments have less total sequence output but have longer read length for each DNA sequence, allowing for better detection of structural variants, ability to sequence through complex or repetitive genomic regions, and resolution of haplotypes, though they currently have a relatively high error rate (Peters et al. 2012). Finally, "benchtop" instruments are well suited for small, but fast output production, at a more affordable initial investment and running cost than for the larger high-throughput or long-read instruments. In all applications, NGS offered scalability that was previously unavailable with Sanger sequencing, and two individual genomes, that of a Yoruba tribesman from Nigeria and James Watson, were sequenced in 2008 using the newly introduced NGS platforms of the Illumina GA1 and 454 Life Sciences FLX, respectively (Bentley et al. 2008; Shendure and Ji 2008; Wheeler et al. 2008).

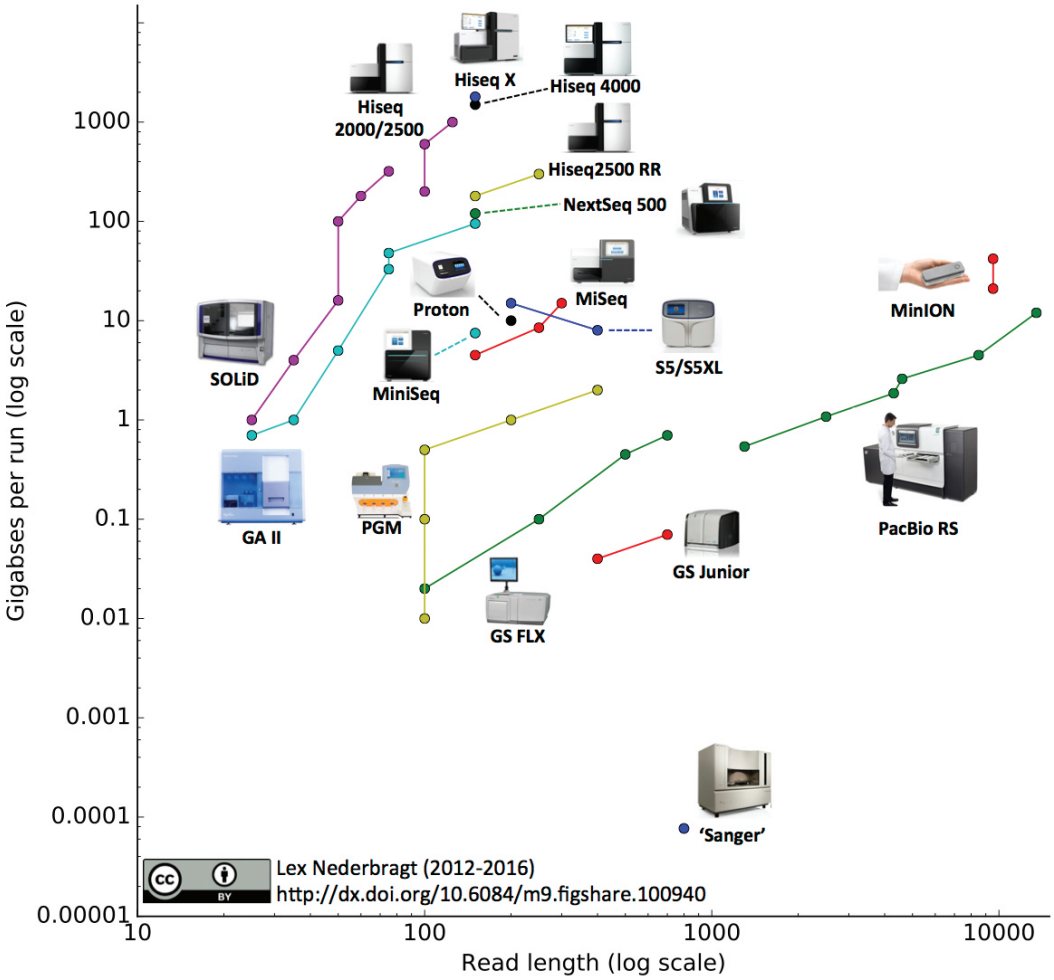


Figure 2 | NGS instruments, then and now. NGS instruments significantly improved on the output of the automated Sanger sequencing ABI 3730xl instruments used in the human genome project, and some instruments (i.e. the Pacific Biosciences RS, the Oxford Nanopore MinION) outperform Sanger sequencing on read length. Currently, each of the ten machines in a HiSeq X configuration can produce 15 genomes in 3 days. Figure by (Nederbragt 2017).

When first introduced, the initial NGS instruments offered great promise but also several distinct hurdles needed to be overcome. The first was read length – compared to Sanger sequencing read lengths of approximately 1000 bases, the length of NGS DNA sequences was extremely small; the initial machines had read lengths of only 50 bases or less (Bennett 2004; Margulies et al. 2005; Shendure et al. 2005). When mapping these sequences back to the reference genome, this posed a large disadvantage – short reads were less likely than longer reads to be assembled and aligned uniquely back to the genome, creating a bioinformatic challenge and need for better assembly and alignment algorithms (Gnerre et al. 2011; Shendure and Ji 2008). Repetitive regions posed a significant problem for mapping short-read NGS DNA sequences, whereas DNA sequences generated by traditional Sanger sequence could resolve some of these areas (Alkan et al. 2011; Shendure and Aiden 2012). Additionally, these newly generated NGS reads had a much higher error rate than Sanger sequencing due to errors introduced during sample preparation, PCR amplification and the sequencing cycling and imaging, creating a challenge in variant identification (Fox et al. 2014; Glenn 2011; Kanagawa 2003). This would have to be addressed by the development of stringent quality control measures of DNA reads and reliable variant algorithm creation (Shendure and Ji 2008). Finally, specific variant classes were technically harder to detect with this technology. Small indels, structural events, CNVs, tandem repeat expansions, transposition events, and copy-neutral events were all very difficult to resolve using short-read NGS technology, missing a large fraction of variants that could be causal for disease (Shendure and Aiden 2012).

First applications of next generation sequencing – targeted gene sequencing panels

The cost benefits and scalability of using NGS were rapidly apparent; in contrast to the sequencing of Craig Venter's entire genome by Sanger sequencing in 2001 at a cost of over $100 million dollars, James Watson's genome was successfully resequenced using NGS for less than $1 million (Venter et al. 2001; Wheeler et al. 2008). However, this high cost was prohibitive for individual patient genome sequencing. Instead, the first applications of NGS for Mendelian disorders focused upon targeted resequencing of specific genomic regions. As the human genome DNA sequence was known, specific genomic regions could be resequenced in patients with disease to identify variants that could be causal for the phenotype. These genomic areas of interest could be regions that had been identified by positional candidate linkage analysis but were too large for Sanger sequencing, or for disease genes that could be involved in the etiology of disease. **Homozygosity mapping**, identifying stretches of homozygous variants in related (consanguineous) families by SNP array, also sufficiently narrowed down the candidate gene region for NGS interrogation (Gilissen et al. 2012; Lander and Botstein 1987; Sundaramurthy et al. 2016). Methods to capture only these areas of a patient's DNA were developed and rapidly applied,

with most using some form of hybridization with a probe complementary to the DNA sequences of interest (Albert et al. 2007; Dahl et al. 2007; Harakalova et al. 2011; Porreca et al. 2007). Around this time, Sanger sequencing of one candidate gene in large patient groups had already proven to be a valid method to identify rare causal variants and thus confirm that dysfunction of that gene resulted in disease. Cohen et al. (2005) first Sanger sequenced a portion of *ANGPTL4*, a gene known in mice to be involved in lipid metabolism, in 3551 individuals and demonstrated that variants in this gene contribute to high plasma triglycerides (Cohen et al. 2005; Topol and Frazer 2007). However, only with the introduction of NGS was this approach economically feasible and possible for larger genomic regions and large patient groups.

Targeted gene sequencing panels were also ideal for diseases that were known to be **genetically heterogeneous** – that is, caused by multiple genetic loci. This approach is generally suited for diseases of which the genetic etiology is well established, or when an extensive "research" candidate gene list can be combined with other genes that may play a role in the same molecular pathway that is known or suspected to be perturbed in a disease. Targeted gene sequencing panels have been successful in sequencing many genes for ataxia, nuclear encoded mitochondrial genes, cardiomyopathies, congenital disorders of glycosylation, Joubert syndrome, and congenital anomalies of the kidney and urinary tract (CAKUT), amongst many others (Hoischen et al. 2010a; Jones et al. 2011; Kroes et al. 2016; Meder et al. 2011; Nicolaou et al. 2016; Shen et al. 2011). The development of NGS methods and the ability to interrogate larger total genomic regions at a cheaper cost than Sanger sequencing made candidate gene sequencing more appealing and cost-effective in larger patient cohorts. By 2006 variants in 1822 genes had been identified as causal for Mendelian disorders (Antonarakis and Beckmann 2006). However, there were still significant challenges to this approach.

The first challenge of this approach was of variant detection – the targeted gene sequencing panel had to include all genes that cause that disease; however, for the majority of diseases the genetic etiology is incompletely known. If a gene was not included in the targeted gene sequencing panel, variants in that gene would never be detected in the patient. For example, if a variant in gene C was causal for a patient's disease but was not included in the target gene sequencing panel of genes X, Y, Z in the patient, this variant in gene C would never be detected, and the patient diagnosis would not be known from this test. This approach thus could only provide a diagnosis for a subset of patients, as the other yet-unknown causal genes were not included in the targeted area.

The second challenge of this approach was in variant interpretation. The human genome has many variants between individuals (Box 1). The number of variants to interpret increased as the targeted gene sequencing panel became larger and included a greater

number of genes. Methods to filter through these variants and annotate certain variants as benign while assigning pathogenicity to others needed to be developed, as did the collection of large population databases of minor allele frequencies to identify rare or novel variants that may be causal. The problem with better and increased DNA sequencing capacity thus became one of not only detecting the causal variant in the sequencing data but also correctly interpreting that a specific variant may be causal, and of proving this with experiments showing that the variant impacted gene or protein function.

The challenge of variant interpretation would persist and only grow larger as the targeted genomic region increased in size – a massive challenge that we still encounter today (**Chapter 8**). However, to ensure that the causal variant was actually covered in the sequencing data there was a simple solution: sequence a larger portion of DNA rather than only a few tens or hundreds of genes. For this, a new method was developed to encompass all coding genes, interrogating the coding exonic region within the genome without *a priori* knowledge of what genes may cause the disease.

Whole-exome sequencing

In 2011 the genetic loci for ~3000 Mendelian disorders had been discovered, but a genetic etiology remained to be discovered for over half of the estimated known Mendelian disorders (Bamshad et al. 2011). The evolution of next generation sequencing progressed to not only capturing a portion of chosen coding genes, but to capture all protein-coding genes in the genome. This approach utilized high-density microarrays with tiling probes complementary to the more than 200,000 protein-coding exons (Hodges et al. 2007). Ng et al. (2009) refined and improved this technique by requiring less input DNA and using a capture method with a genomic footprint of 26.6 Mb to sequence the entire **exome** (all exons of all protein coding genes) of twelve individuals, including four patients with autosomal dominant Freeman-Sheldon syndrome (Figure 3) (Ng et al. 2009). This was the first demonstration of **whole-exome sequencing** (WES) in a group of patients with Mendelian disease. Deleterious variants in *MYH3*, a gene previously implicated in this disease, were detected and concluded to be causal in these patients for Freeman-Sheldon syndrome. In the same year, Choi et al. (2009) established that WES could be used to reach an accurate genetic diagnosis for a recessive disease using a different capture technique, the Roche/Nimblegen capture (Choi et al. 2009). A year later, Hoischen et al. (2010) and Ng et al. (2010) separately would demonstrate that WES could be used for variant detection and causal gene discovery in dominant Mendelian disorders; additionally, Bilguvar et al. (2010) would discover the genetic cause of a rare autosomal recessive syndrome (Bilguvar et al. 2010; Hoischen et al. 2010b; Ng et al. 2010a; Ng et al. 2010b). WES would also be used on affected and unaffected tissues of individuals with Proteus syndrome to identify a mosaic activating mutation (Lindhurst et al. 2011).

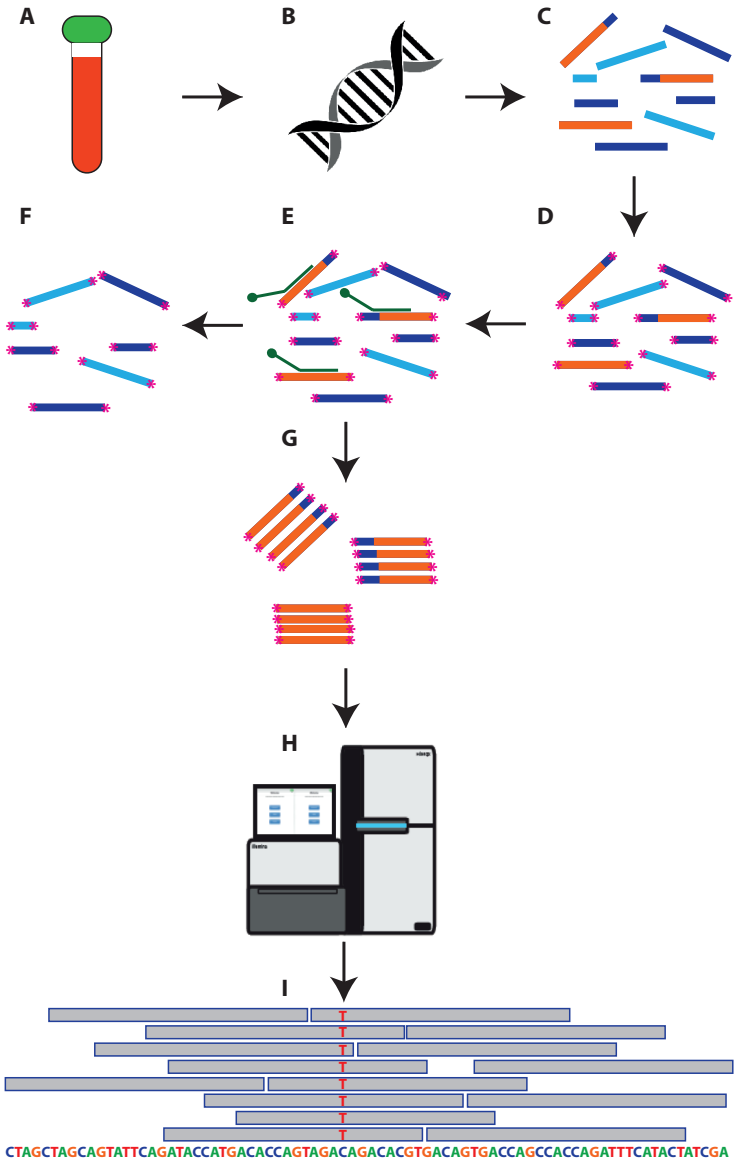


Figure 3 | Whole-exome sequencing. A) Blood is isolated from a patient and B) genomic DNA is extracted from peripheral lymphocytes. C) The genomic DNA is sheared, D) end-repaired, and platform specific sequences are ligated on the ends (pink astericks). E) Oligonucleotide probes (green) specific for the exonic regions (orange) capture only the exonic and intron/exon boundaries. F) Intergenic regions (light blue) and solely intronic regions (dark blue) are discarded. G) The exonic regions are amplified by PCR. H) The DNA library is then massively parallel sequenced on a NGS instrument and I) a variant is identified (T) that is in a homozygous state in the NGS data compared to the human reference genome (C). Figure adapted and expanded from (Seaby et al. 2016); Illumina HiSeq 2500 depiction courtesy of Myrthe Jager.

WES would prove to be highly effective in finding the genetic cause of many Mendelian diseases, particularly in cases where previous linkage analysis efforts had failed due to small family size, reduced penetrance, locus heterogeneity, and/or a reduction in fitness (Bamshad et al. 2011; Ng et al. 2010a). In particular, WES of patient DNA and DNA of the parents (**trio-WES**) is powerful for small families and in discovering all the DNA variants that exist within the patient and their parents – and in finding novel or rare homozygous variants or compound heterozygous variants within the patient that are causal for his or her disease (Dixon-Salazar et al. 2012; Singleton 2011). In a different variant filtering approach, trio-WES can also be used for detecting variants that are only present in the child and not present in either parent. These ***de novo* variants** can be causal for a range of diseases (See Box 2). Using trio-WES, highly penetrant *de novo* variants have been identified as the cause of a large proportion of intellectual disability patients (ID) (de Ligt et al. 2012; Vissers et al. 2010).

BOX 2 | *De novo* variants. The human genome is highly mutable, with a mutation rate per nucleotide per generation of approximately 0.97×10^{-8} - 1.20×10^{-8} (Conrad et al. 2011; Francioli et al. 2015; Gilissen et al. 2014; Goldmann et al. 2016; Kong et al. 2012; Rahbari et al. 2016; Roach et al. 2010). These mutations are caused by exposure to chemicals or the environment (i.e. mutagens or radiation) as well as internal processes such as reactive oxygen species and errors during DNA replication and persist due to limits on the diverse DNA repair pathways (Campbell and Eichler 2013; Crow 2000). Mutations arising in the sperm or the ova (germ-line mutations) of the parents that occur during gametogenesis can be passed on to offspring, even while those same variants are not present in the DNA of the rest of the cells in the parent. Additionally, post-zygotic DNA mutations can occur after fusion of the sperm and ova during the rapidly expanding and stressful time of clonal expansion of the zygote, resulting in mosaic variants in the offspring (Acuna-Hidalgo et al. 2016; Freed and Pevsner 2016). These new variants, called *de novo* variants, are a source of evolutionary change and have the potential to cause disease if they disrupt conserved and functional areas of the genome, particularly if they disrupt genes that are essential for development (Vissers et al. 2010).

The male germline has long been thought to be more mutagenic than the female germline due to the extensive cell divisions that occur during spermatogenesis, a proposal that has been confirmed in numerous studies (Figure 4) (Francioli et al. 2015; Goldmann et al. 2016; Haldane 1947; Kong et al. 2012; Wilkin et al. 1998). These sequential cell divisions mean that as the father ages more *de novo* variants occur, resulting in 2 additional *de novo* variants per year of father's age (Kong et al. 2012). Depending on the age of the father at conception this results in a range of 45 - 60 *de novo* SNVs in a newborn's genome that were not present in the genome of either the mother or the father (Kong et al. 2012). The *de novo* INDEL rate is approximately 1/6 of the SNV rate, determined from Sanger sequence studies due to the difficulty of detecting indels with NGS (Chen et al. 2009). The contribution of post-zygotic mutations to *de novo* variation is much smaller than male germline mutation, at ~5% in patients with autism spectrum disorder (Freed and Pevsner 2016). Each one of us has *de novo* variants when we are born; the genome is robust, and in many cases *de novo* variants do not have extreme impact upon development of the human embryo or function as an adult. However, if these variants disrupt essential amino acid residues or result in premature protein truncation or mRNA transcript degradation of a protein essential for development, the phenotypic effect may be profound. Additionally, as a group they are more deleterious than inherited variation because they have arisen in only one generation and have not been exposed to negative selection (Crow 2000).

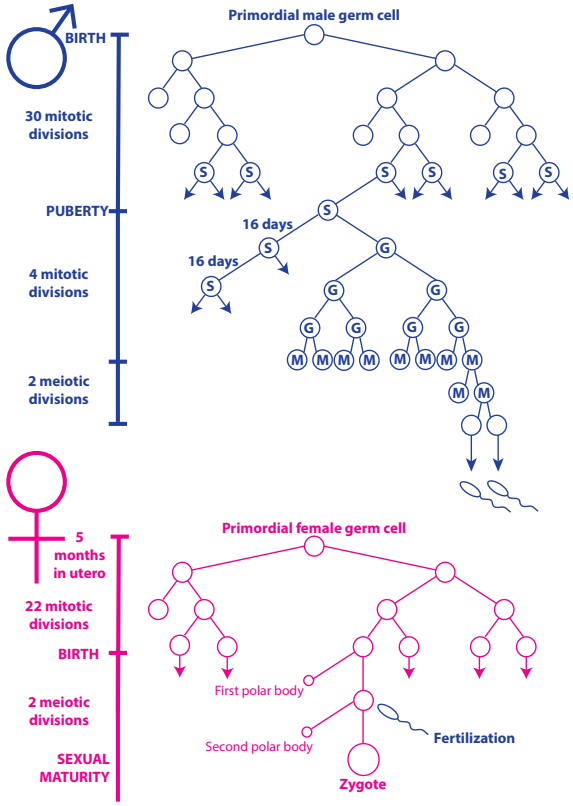


Figure 4 | Comparison of the number of cell divisions during oogenesis and spermatogenesis. Females are born with the full complement of ovum at birth, encompassing 23 chromosomal replications. Males continually produce sperm, with stem cells undergoing cell division every 16 days, so that there are new opportunities for mutation in each replicative round. S: stem cells. G: gonial cells. M: meiotic cells. Figure adapted from (Crow 2000; Vogel and Motulsky 1997).

Many *de novo* variants do not have an extreme impact upon gene function, but for those individuals with a developmental phenotype these variants are strong candidate variants to be explanatory for the disorder. *De novo* variants can be prioritized based upon novelty (if the variant has ever been reported before in healthy populations), gene function (if the gene's function is relevant for the disease etiology) and predicted consequence of the variant (does the variant affect protein function). This "short list' of candidate variants can be further followed up with laboratory experiments to evaluate the impact of the variant. *De novo* variants are thus strong candidates to be causal for a disorder, but need to be interpreted correctly in terms of variant recurrence, gene function, and predicted variant effect.

The contribution of *de novo* variation to severe intellectual disability is particularly striking, where the decrease in fitness usually prevents the variant to be passed to the next generation, while the prevalence of severe intellectual disability remains relatively constant (Willemsen and Kleefstra 2014). Trio-WES has enabled the detection of patient *de novo* variants in a rapidly growing list of genes that may be causal for intellectual disability and/or development disorders as well as for autism spectrum disorders, schizophrenia, and bipolar disorder amongst other diseases (de Ligt et al. 2012; Deciphering Developmental Disorders Study 2015, 2017; Fromer et al. 2014; Harakalova et al. 2012; Iossifov et al. 2014; Kataoka et al. 2016; O'Roak et al. 2011; Rauch et al. 2012; Singh et al. 2017; Vissers et al. 2010). This approach is mainly applicable to diseases with high locus heterogeneity, diseases where a dominant mode of inheritance is anticipated (as *de novo* variants affect only one allele), and sporadic cases where neither parent is affected by the disease (Bamshad et al. 2011). A recent large-scale analysis of the exomes of 4,293 developmental disorders patients and their parents estimated that 42% of these patients had a *de novo* variant present in the exome predicted to be damaging or result in loss-of-function, demonstrating the power of WES to detect *de novo* variants – and providing a strong candidate list of variants to follow up experimentally to establish causality (Deciphering Developmental Disorders Study 2017).

The biggest benefit of using WES is its size in comparison to sequencing the entire genome. By enriching only for the protein coding regions of the genome, the actual portion of the genome to be sequenced is limited to only ~1.5-2% (depending on the enrichment procedure probes used) reducing the number of total DNA sequencing reads required for adequate coverage of the whole-exome target. This significantly reduces sequencing costs while still targeting the most relevant areas, as approximately 85% of the currently known genetic causes for Mendelian disorders affect the coding regions of the genome, though this is undoubtedly inflated by ascertainment bias and will change as our ability to interpret variants in the noncoding region of the genome improves (Botstein and Risch 2003; Cooper et al. 2011; Kuhlenbäumer et al. 2011; Stenson et al. 2009).

However, WES, even with its iterative refinements since its introduction and its larger scope than the targeted gene sequencing panel approach, has technical limitations. The enrichment step to select for only the exome is more effective for some regions than others, resulting in 5-15% of the exome not covered adequately, though this has improved with enrichment kit refinements in the last few years (Biesecker 2010; Gilissen et al. 2012). Furthermore, WES enrichment only captures the coding regions and some regions that are known to be important; regions of the genome that are functionally relevant but are not included in the WES capture design simply cannot be sequenced and analyzed using WES. Additionally, WES is limited in its ability to detect structural variants, a group of variants that

due to the number of affected bases can have a bigger impact on the genome than SNVs, with the larger structural variants including CNVs, copy neutral events, mobile element insertions, large insertions and deletions, translocations, and complex events affecting as much as 13% of the genome (Conrad et al. 2010; English et al. 2015; Haraksingh and Snyder 2013). Finally, as more and more exomes and genomes are sequenced, the chance of detecting a *de novo* in the same gene in two patients with the same rare phenotype increases, bringing it with the danger of a false positive (See Box 3: Functional modeling: a necessity to establish causality) (MacArthur et al. 2014).

BOX 3 | Functional modeling: a necessity to establish causality. With the rise of NGS, detecting variants has become relatively easy; the challenge has transitioned to interpretation. In the case of *de novo* variants, each of us has 0-5 *de novo* variants in the coding region of our genome; in a patient, one of these *de novo* variants could be causal for their disease. But how to identify which variant is truly causal from those that do not cause a disease? Detected variants can be prioritized based upon statistical evidence of novelty or a low minor allele frequency in population variant databases; predicted deleterious effect at the protein level; annotated gene function related to disease etiology; or by predicted effect of the variant at the regulatory or splicing level. However, our knowledge of gene function is far from complete, and our ability to predict a variant's effect at the regulatory, transcription, or translation level is even more elementary. To actually assess the effect of a variant, we must look at it by functional modeling – experimentally showing that a variant has a different effect on the molecular, cellular or model organism level than the normal allele (MacArthur et al. 2014).

Experimental approaches to model variants are diverse and dependent on the type of variant and gene. Molecular experiments can be performed on patient RNA to establish that the variant results in aberrant splicing or inappropriate tissue expression. Experiments on patient cells or cells transfected with the patient variant can show the effects of a variant on a cellular level (i.e. do the cells correctly divide; do they have appropriate levels of cilia to receive and interpret developmental signals) that differ from normal. Finally, the ultimate test of a variant, particularly of a DNA variant that is expected to affect development, is to see if that variant results in a similar phenotype as the patient when the variant is expressed in a developing model organism. For instance, human patients with a defect in their leptin receptor (LEPR) get morbidly obese (Farooqi et al. 2007). Mice and rat models with this gene knocked out also have diabetes or are obese, respectively (Chua et al. 1996). The combination of these two facts links the genetic variant to the phenotype; without the functional knowledge of the variant, the genetic variant is only associated with the phenotype, and does not have to necessarily be causal. Ideally, the functional model should use the exact variant as seen in the patient, though approximations can also indicate the effect of the variant.

In our laboratory, we use the zebrafish as a model organism to model developmental and metabolic disorders. Zebrafish have the advantage over mice and rat models in that they are cheaper to maintain, have faster developmental times, need smaller space, and the embryo develops externally – allowing the effect of an expressed DNA variant on development to be immediately visible (Veldman and Lin 2008). The molecular toolkit for zebrafish experiments has rapidly expanded with CRISPR/Cas9 gene editing technology, and we are now relatively easily able to knock out specific genes, overexpress genes, as well as introduce patient variants and assess the phenotypic consequences in a developing embryo.

Functional tests at the molecular, cellular and model organism level are crucial to link the patient variant to the disease, and give us insight into the mechanistic dysfunction that is caused by the DNA variant present in the patient.

WES has been very successful as a technique when implemented in large patient group diagnostics, resulting in a diagnostic yield of ~25-30%; however, the remainder of these patients remain undiagnosed (Lee et al. 2014; Yang et al. 2013; Yang et al. 2014). The causal variant, if genetic, in these patients could lie beyond the coding exome, or not be able to be detected accurately with WES. For this a still larger investigative technique is needed, one encompassing all DNA: whole-genome sequencing.

Whole-genome sequencing

The ability to sequence the entire genome of a patient, at sufficient read depth, is the ultimate genetic test. James Lupski first demonstrated this on the genome scale by performing **whole-genome sequencing** (WGS) upon his own genome to discover the genetic cause of his Charcot-Marie-Tooth disease (Lupski et al. 2010). Unlike WES, WGS produces sequence information on the coding exome as well as the other ~98.5% of the genome that is noncoding but which many regions have functional relevance. Additionally, WGS results in more uniform coverage, better distributions of genotype quality and minor read ratio, and can cover exonic regions better than WES because there is no enrichment step (Belkadi et al. 2015; Turner et al. 2016). Gilissen et al. (2012) demonstrated that 42% of patients in an intellectual disability cohort with prior WES testing could now obtain a diagnosis by WGS, highlighting the technical limitations of WES as most causal variants were in the coding regions and originally missed, particularly for older exome enrichment methods or platforms (Gilissen et al. 2014). A year later, Taylor et al. (2015) estimated that 15% of their cases would have gone unresolved by using WES due to a variant a) lying outside the targeted exome, b) covered poorly by WES (<20X), or c) low alternate allele frequency (<3 reads), and they would not have been able to identify the 2 patients with causal noncoding variants (Taylor et al. 2015). Additionally, WGS can also be used for copy-neutral structural variant detection, a facet that is completely undetectable by WES (Korbel et al. 2007; Taylor et al. 2015). WGS thus provides a single genetic test that (if covered at sufficient read depth) is superior to WES, albeit at a higher cost. WGS use, interpretation and data storage are still being optimized, but it is also beginning to be introduced on a large-scale in the clinic (Bick et al. 2017; Stavropoulos et al. 2016).

Implementing WGS, similar to WES or targeted gene sequencing panels previously, and simply sequencing the whole genome does not solve all problems – in fact, it introduces new ones. Technically WGS is superior to all other NGS technologies, but the data generated is massive: each genome sequenced results in 80-90 GB, dependent on coverage. The ability to store this data in terms of hardware capacity is enormous and the cost also significant – recent estimates on storing one genome are approximately 30 euro per year (backed up in two separate locations), not including computing time for WGS processing or subsequent analysis files generation (Nijman 2017). Also from a computational standpoint, WGS data

processing, interpretation, and algorithm development for successful mapping, structural variant detection or data transition to a new build of the genome demands expertise and a bioinformatics core that is not available to smaller centers. Finally, sequencing the whole genome (as well as the whole exome) raises ethical and moral questions in the scope of secondary findings – those genetic findings that were not looked for, yet visible once an untargeted approach is undertaken (Kalia et al. 2017). Successfully dealing with patient wishes on this issue, while also reporting back variants that medical action could be taken, requires a careful, well-thought out variant interpretation pipeline and trained expertise.

WGS greatly enhances our ability to detect DNA variation. However, even more than smaller-scale NGS technologies its application shifts the focus from technical variant detection to one of variant interpretation and experimental functional validation. The latter is crucial; in a 2011 study of published severe "causal" variants, 27% of these variants did not have experimental evidence for pathogenicity or were common in the population upon sequencing more individuals (Bell et al. 2011).

Choosing the DNA variants that are candidates to be causal from this larger genetic data set becomes more challenging, as well as prioritizing those few variants that should be functionally followed up. Even with these challenges and wealth of data, as sequencing becomes more ubiquitous, the use and applications of WGS will only increase. Technical innovation has driven new Mendelian disease gene discoveries and the application in the clinic of new NGS instruments has empowered the diagnostics process (Figure 5). We have just entered the genuine genomic era.

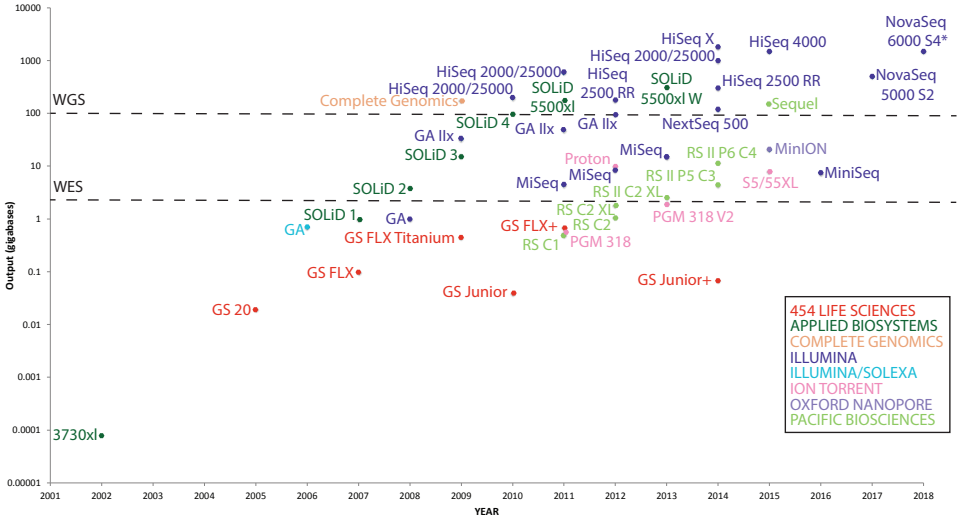


Figure 5 | Technical advances leading to the genomic era: new innovations drive research and diagnostic opportunities. The low-output of automated Sanger sequencing (2002) allowed only smaller genomic regions to be DNA sequenced. With the launch of the NGS instruments of 454 Life Sciences GS 20 (2005), the Solexa GA (2006), and the SOLiD 1 (2007), larger targeted sequencing panel projects could be undertaken. Improved chemistry and instrumentation increased base output of these machines and enabled routine WES with the Illumina GA IIx (2009) and the SOLiD 3 (2009). With the introduction of the Illumina HiSeq X platform (2014), large-scale human genome sequencing at low cost was possible and the "$1000 genome" had arrived. Key: 100 Gb = approximately 1 genome (3.2 Gb) at 30X. 3 Gb = 1 exome (.05 Gb) at 60X. NovaSeq 6000 S4: theoretical output indicated by *. Output values retrieved from product specification sheets compiled by (Nederbragt 2017) and adapted. The brand commercializing the platform is indicated by color.

THE GENOMIC ERA

The rapid introduction of NGS into research and clinical practice has far-reaching implications upon patient care. The opportunity to look into an individual's entire genome gives tremendous power to find diagnostic variants for patients and streamline healthcare tests. NGS has resulted in high diagnostic yields, diagnosing patients who have previously gone undiagnosed, even if the phenotype is observed in only one patient, over a broad spectrum of phenotypes (Choi et al. 2009; de Ligt et al. 2012; Need et al. 2012). Also, in some cases, NGS has resulted in the revision of the clinical diagnosis with the molecular diagnosis (Choi et al. 2009; Worthey et al. 2011). For the patient, a comprehensive all-in-one test has clear advantages – there is more chance of obtaining a diagnosis instead of with several sequential tests, and if WES or WGS is used a much larger genetic area is searched for causal variants. This significantly improves the number of diagnoses, thus providing access to care, management plans, and treatment options to more patients than before. This is particularly important given the fact that only one half of patients with neurological developmental disorders currently get a diagnosis with traditional diagnostic methods (Shashi et al. 2014). NGS, therefore, has the power to give patients better, faster care, and also eliminate redundant tests in the clinic by providing accurate diagnoses in a single test.

In the below chapters I illustrate the evolving capability of next generation sequencing in discovering the cause of Mendelian disorders, as we use more comprehensive and powerful methods to identify candidate variants. We then use molecular and cellular experiments, as well as the zebrafish as a model organism, to establish causality for the genetic variants that we identify. I then therefore discuss the next steps in broad scale NGS implementation and improvements needed in technology, the future challenges that we face in interpreting what lies inside of our genome and establishing the effect of DNA variation, and the changes that society will encounter as we move to a future where personal genomes are commonplace.

SCOPE: The research presented in this thesis identifies the genetic cause of a diverse range of Mendelian disorders using next generation sequencing. This work reflects the extremely rapid development of NGS, beginning with target gene panel sequencing in a research setting (**Chapters 2 and 3**) and only 4 years later progressing to our evaluation of diagnostic yield if whole-exome sequencing is implemented in the clinic and when using whole-genome sequencing for our research (**Chapters 7 and 8**). These discoveries show the evolving capability of NGS to uncover the genetic cause of Mendelian disorders and the revolution upon healthcare occurring as a result of its application.

In **Part I** we utilize NGS to discover genetic variants in patients with rare phenotypes for which no genetic cause is known. In **Chapter 2** and **Chapter 3** we use a targeted gene panel sequencing approach to identify causal variants by including known genes involved in disease etiology and candidate genes that could play a role in development of disease. In **Chapter 4** we move to a more comprehensive approach by utilizing WES to analyze the coding regions of all known protein coding genes and eventually establish a genetic diagnose for two brothers with Mohr syndrome.

In **Part II** we focus on rare metabolic diseases patients born of consanguineous parents and use regions of homozygosity to narrow our genetic search space for causal genes. We use targeted gene panel sequencing to evaluate all genes within the regions of homozygosity (**Chapter 5**) or identify candidate genes in those regions to Sanger sequence (**Chapter 6**).

Finally, in **Part III** we progress to the clinical utility of NGS by performing a study analyzing the diagnostic yield of trio-WES and concurrently evaluate cost savings if this technology was implemented early (**Chapter 7**). In **Chapter 8** we discuss the state-of-art of NGS and its implications for the patient, the clinic, and society. The expanding scope and use of NGS will enable more patients to obtain faster diagnoses and remodel the healthcare system, though there are still challenges to address. Finally, the ability to sequence our own genomes will impact individuals with the possibility of “personal genomes”, including my own, and what this may mean for society.

2

Familial Ehlers-Danlos syndrome with lethal arterial events caused by a mutation in *COL5A1*

American Journal of Medical Genetics Part A. 2015 Jun;167(6):1196-203. Epub 2015 Apr 2.

PMID: 25845371; DOI: 10.1002/ajmg.a.36997

Glen R. Monroe[1*], Magdalena Harakalova[1*], Saskia N. van der Crabben[1], Danielle Majoor-Krakauer[2], Aida M. Bertoli-Avella[2], Frans L. Moll[3], Björn I. Oranen[4], Dennis Dooijes[1], Aryan Vink[5], Nine V. Knoers[1], Alessandra Maugeri[6], Gerard Pals[6], Isaac J. Nijman[1], Gijs van Haaften[1], Annette F. Baas[1]

1 | Department of Medical Genetics, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands
2 | Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands
3 | Department of Vascular Surgery, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands
4 | Department of Vascular Surgery, Bethesda Hospital, Hoogeveen, The Netherlands
5 | Department of Pathology, University Medical Center Utrecht (UMCU), Utrecht, The Netherlands
6 | Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands
*Joint First Authors

ABSTRACT

Different forms of Ehlers-Danlos Syndrome (EDS) exist, with specific phenotypes and associated genes. Vascular EDS, caused by heterozygous mutations in the *COL3A1* gene, is characterized by fragile vasculature with a high risk of catastrophic vascular events at a young age. Classic EDS, caused by heterozygous mutations in the *COL5A1* or *COL5A2* genes, is characterized by fragile, hyperextensible skin and joint laxity. To date, vessel rupture in four unrelated classic EDS patients with a confirmed *COL5A1* mutation has been reported. We describe familial occurrence of a phenotype resembling vascular EDS in a mother and her two sons, who all died at an early age from arterial ruptures. Diagnostic Sanger sequencing in the proband failed to detect aberrations in *COL3A1*, *COL1A1*, *COL1A2*, *TGFBR1*, *TGFBR2*, *SMAD3* and *ACTA2*. Next, the proband's DNA was analyzed using a next-generation sequencing approach targeting 554 genes linked to vascular disease (VASCULOME project). A novel heterozygous mutation in *COL5A1* was detected, resulting in an essential glycine substitution at the C-terminal end of the triple helix domain (NM_000093.4:c.4610G>T; p.Gly1537Val). This mutation was also present in DNA isolated from autopsy material of the index's brother. No material was available from the mother, but the mutation was excluded in her parents, siblings and in the father of her sons, suggesting that the *COL5A1* mutation occurred in the mother's genome *de novo*. In conclusion, we report familial occurrence of lethal arterial events caused by a *COL5A1* mutation.

INTRODUCTION

Ehlers–Danlos Syndromes (EDSs) are a heterogeneous group of connective tissue diseases leading to abnormalities of the skin, joints, blood vessels and internal organs (de Paepe and Malfait 2012). Mutations have been identified in genes encoding fibrillar collagens or enzymes that modify these structural extracellular matrix proteins. Fibrillar collagens are trimeric molecules, consisting of three polypeptide chains (α-chains), which form a triple helix. Each α-chain consists of a repetition of (Gly-Xaa-Yaa) triplets, in which the presence of glycine is sterically essential for correct helix formation (Malfait et al. 2010). Mutations in type V and type III collagen cause classic or vascular EDS, respectively (de Paepe and Malfait 2012).

Patients with vascular EDS have a poor prognosis because of a predisposition to the rupture of arteries and hollow organs at a young age. In two separate large retrospective studies of 100 and 220 vascular EDS probands, two groups showed that 7-25% of the patients experienced a first major event by the age of 20 years, and up to 75-80% experienced a first major complication by the age of 40 years (Pepin et al. 2000). In the De Paepe and Malfait study, the majority (82%) of events were of arterial origin and involved aneurysms, dissections or ruptures of medium-sized abdominal vessels (mainly renal, iliac, femoral, mesenteric and hepatic). Of note, ruptures were not always preceded by aneurysmal dilatation. In contrast to classic EDS, the skin of vascular EDS patients is usually not hyperextensible, but thin and translucent, revealing a venous pattern over the chest and abdomen. Joint laxity is present, although usually not as extreme as in classic EDS, and mainly localized to the small joints of the hands. Other manifestations include easy bruising, congenital clubfoot or hip dislocation, pneumothorax and gingival recession (Pepin 1999). Patients with vascular EDS can have a characteristic facial appearance, with prominent eyes, a thin, pinched nose and small lips. However, the facial features may be very subtle or absent. To date more than 250 *COL3A1* mutations have been identified in vascular EDS patients, explaining over 90% of the patients (de Paepe and Malfait 2012).

The diagnostic triad of classic EDS consists of skin hyperextensibility, widened atrophic scars and joint hypermobility. In the majority of the patients a heterozygous mutation is identified in either the *COL5A1* or *COL5A2* gene (Symoens et al. 2012). Mutations resulting in haploinsufficiency (i.e. frameshift, splice-site or nonsense mutations) are expected to lead to a reduction of type V collagen. A minority of mutations represents missense mutations (mostly consisting of substitution of glycines in the triple helix domain) in either *COL5A1* or *COL5A2*, and those are expected to have a dominant-negative effect by producing an abnormal α-chain that is incorporated into the triple helix. Currently about 150 mutations in *COL5A1* and *COL5A2* have been identified in classic

EDS patients. There is no clear phenotype–genotype correlation. To date, major arterial complications (e.g. rupture and dissection) related to a *COL5A1* mutation was reported in four unrelated patients (Borck et al. 2010; de Leeuw et al. 2012; Karaa and Stoler 2013; Mehta et al. 2012). Similarly, an Ehlers-Danlos syndrome phenotype, with a propensity to arterial rupture in early adulthood has been described in a few patients with arginine to cysteine substitutions in the *COL1A1* gene (Malfait et al. 2007).

Here we present the first results of the VASCULOME project, in which we analyzed 554 genes known or suspected to be involved in vascular disease, in patients with an exceptional vascular phenotype (based on either early age of onset, extended vascular disease and/or a Mendelian inheritance pattern). We describe a family with a phenotype resembling vascular EDS, caused by a distinct molecular pathology: a novel *COL5A1* mutation.

MATERIALS AND METHODS

Family characterization and sample collection

The proband was referred to our Department of Medical Genetics at the University Medical Center Utrecht (UMCU), The Netherlands, by his vascular surgeon. He was examined for associated skeletal and skin features of connective tissue disorders. A pedigree was constructed. All relevant medical information from living and deceased family members was retrieved. After obtaining informed consent, blood and paraffin embedded tissue obtained from the autopsy of family members was collected and genomic DNA was isolated according to standard procedures.

Candidate gene analysis

Candidate genes for vascular disease were sequenced in the proband in the DNA diagnostic laboratory of the Free University Medical Center (VUMC), The Netherlands. The coding regions of *COL1A1*, *COL1A2*, *COL3A1*, *TGFBR1*, *TGFBR2*, *ACTA2* and *SMAD3* were analysed by Sanger sequencing using standard procedures. Primer information is available upon request. To investigate the presence of intragenic deletions or duplications in *COL3A1*, Multiplex Ligation-dependent Probe Amplification (MLPA) was performed (MRC-Holland, Amsterdam, Netherlands). Protein analysis on fibroblasts obtained from a skin biopsy of the proband was performed to detect collagen abnormalities.

VASCULOME gene selection

In order to detect the genetic cause in patients with an exceptional vascular phenotype,

based on either young age of onset, extent of the vascular disease and/or a Mendelian inheritance pattern, the VASCULOME was designed. This is a next generation sequencing panel, including genes that are either known or suspected to be involved in vascular disease and vessel wall integrity. From May until July 2011, a comprehensive literature search in the Online Mendelian Inheritance in Man (OMIM) database (http://omim.org/) and the PUBMED database (http://www.ncbi.nlm.nih.gov/pubmed) was performed to select known vascular disease genes. In addition, genes from relevant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (http://www.genome.jp/kegg/pathway.html) (i.e. vascular smooth muscle cell contraction, TFG-beta and NOTCH signaling pathway) and Gene Ontology (GO) terms (http://www.geneontology.org)(i.e. relaxation and contraction of vascular smooth muscle cell) were included. Specific genes expressed in the vascular tissue (TPM100) from the Interactome EST UniGene database (http://matrixdb.ibcp.fr/cgi-bin/interactome) were also annotated. The final VASCULOME panel consisted of 554 genes and is provided in Supplementary Table S1.

Targeted next-generation sequencing of the VASCULOME genes

The VASCULOME design consisted of 487,008 sixty-nucleotide-long probes, mapping to the coding sequences of all selected genes from the GRCh37/hg19 human genome build. The design had an average tiling density of 8 bp for both the positive and negative strand, and included 50 bp intronic flanks.

Sample fragment library preparation and genomic enrichment on a 1M custom microarray (Agilent Technologies, CA, USA) was performed as previously described (Harakalova et al. 2011). Briefly, 2 µg of purified gDNA (Qiagen, Hilden, Germany) was sheared into 150 bp fragments with a Covaris S2 sonicator (Covaris, Woburn, MA, USA), then blunt-ended and 5'phosphorylated, and short double-stranded adaptors complementary with the SOLiD next-generation sequencing platform were ligated to the ends. Next, the non-phosphorylated and non-ligated 3' ends were nick-translated, and 7 PCR cycles added unique bar code sequences to each sample. The proband sample (III:7, AN_004203) was then pooled into a sample cohort of 42 total patients, size-selected on a 2% agarose gel, and purified (Qiagen, Hilden, Germany). Samples were then enriched using a custom protocol (Harakalova et al. 2011) and amplified with 13 PCR cycles with primers complementary to the full SOLiD oligo sequences. The pooled samples were run as a full slide on the SOLiD 5500XL (Life Technologies, Carlsbad, CA, USA). Color space reads were mapped and aligned against GRCh37/hg19 reference genome using a custom pipeline based on the BWA software, and annotated as described previously (Harakalova et al. 2011). The criteria for variant detection coverage were set at 10 reads and a cut-off for heterozygote allele calls was set at 25-75%. Common alleles were filtered from rare or novel variants using the reference frequencies from the datasets of NCBI dbSNP Build 137 for Human (http://

www.ncbi.nlm.nih.gov/SNP), Exome Variant Server (EVS; http://evs.gs.washington.edu/EVS), 1000Genomes (http://www.1000genomes.org), Genome of the Netherlands (http://www.nlgenome.nl) and our in-house dataset. Variants were considered novel if not present in the above datasets and also compared to other patients in the cohort. For each patient, a final variant list was compiled including the genomic location of the variant, the amino acid change and the predicted effect on protein function for each separate transcript. The predicted effect on protein function and conservation score was calculated using prediction programs and algorithms (PolyPhen-2, SIFT, GERP) for variant prioritization. Confirmation of the selected SNVs and segregation analysis in the family were performed by Sanger sequencing. Primer information is available upon request.

RESULTS

Clinical Report

In 2009, at the age of 39 years, the index patient (III:7, AN_004203) visited our outpatient clinic. In the past, the proband had been diagnosed with vascular EDS on clinical grounds, similar to his only brother (III:5, AN_004202). Vascular EDS was also suspected in his mother (II:6, AN_004201; Figure 1 and Table 1). He had joint hyperlaxity in childhood, but no dislocations, abnormal wound healing, easy bruising or excessive scaring. At the age of 15 years he underwent emergency surgery for a ruptured left subclavian artery after he collapsed during a soccer game. Pathologic tissue examination revealed aneurysmal dilatation with loss of elastic fibers. At age 18 he underwent prophylactic resection of a coeliac artery aneurysm, which was complicated by severe bleeding. Since then, he had received regular computer tomography angiography (CTA) scans of the aorta and iliac arteries and no aneurysms had developed.

At physical examination, no typical facial features of vascular EDS were noticed (no prominent eyes, thin/pinched nose or small lips). Translucent skin was observed, which was not hyper-extensible. Scar tissue was not exceptionally broad and there were no skin changes on the knees or elbows. He had no varicose veins. He had hypermobility of his finger joints, but not of other joints. The Beighton score was four out of nine, due to hyper-extensibility of the thumbs and fifth fingers. He had no marfanoid habitus.

Four years after his visit his family informed us that the proband had suddenly died in the shower at age of 43. Prior to the event, he had complained of a severe headache. No trauma had taken place. The autopsy report stated that there was subarachnoidal bleeding which was likely the cause of death. In addition, a large retroperitoneal hemorrhage was detected, resulting from a rupture of the abdominal aorta, which may have contributed to his death.

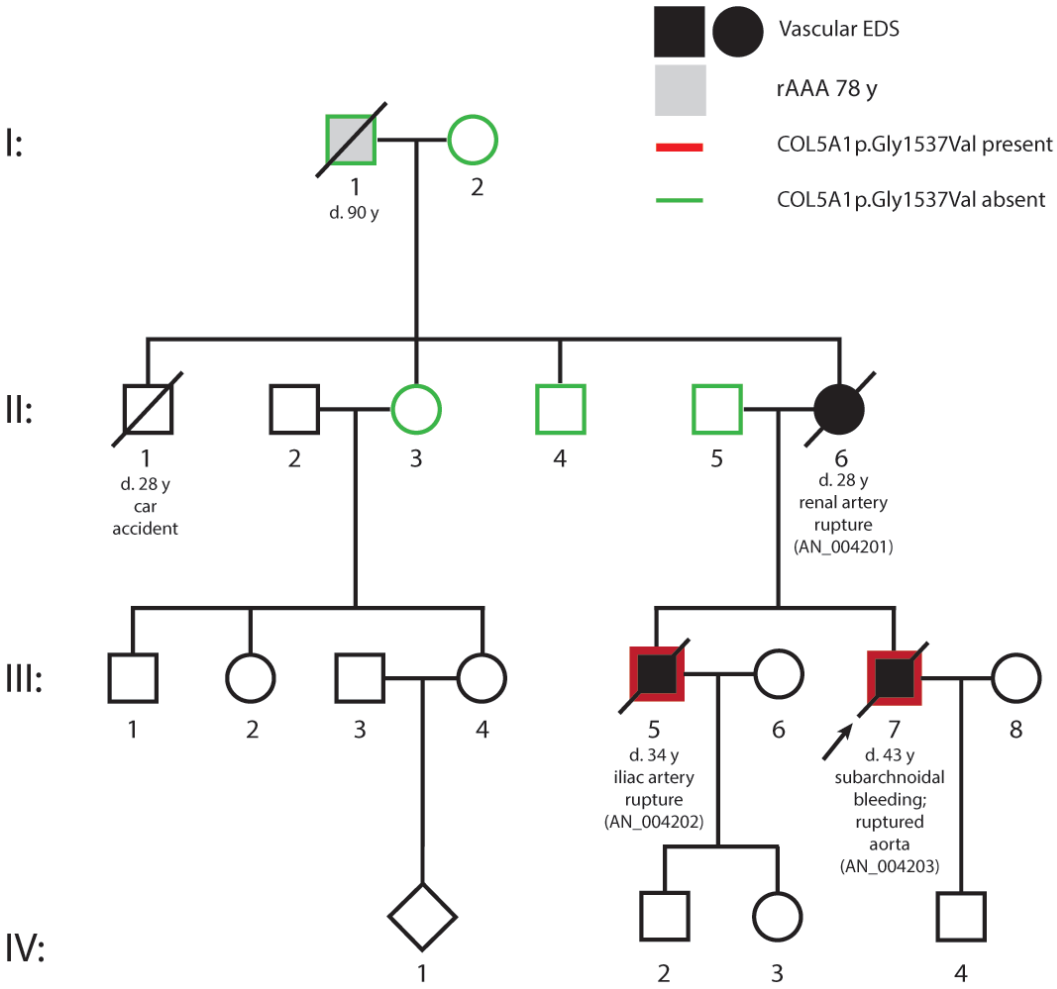


Figure 1 | Pedigree of the presented family.

The mother of our index patient (II:6) died suddenly in 1975 at age 28 while she underwent anesthesia for teeth extraction. Her widower reported that she had recurrent gingivitis. No pregnancy related complications were reported. An old autopsy report was retrieved, which revealed that the cause of death was a major hemorrhage from a tear in a renal artery. Furthermore, the report stated that all medium and large arteries showed massive degenerative changes with loss of elastic fibers.

The brother of our patient (III:5) had died in 2006 at the age of 34. Based on the medical history of his brother (III:7) and mother (II:6), and the fact that he had easy bruising and hypermobile joints, he had been clinically diagnosed with vascular EDS in childhood. He was diagnosed with osteoporosis in 2005 after he had had a vertebral fracture. He received annual magnetic resonance angiography (MRA) scans of the thorax and abdomen, and no aneurysms had been detected over the years. Prior to this death, he suddenly collapsed while sitting on the couch. Paramedics suspected that he had a large bleeding from a right iliac artery rupture, which was confirmed during emergency surgery. However, due to the fragile vasculature, it was impossible to stop the bleeding and he died during surgery. The autopsy report stated that the consistency of his vessels was comparable with that of rice paper, and the vessels showed massive degenerative changes with loss of elastic fibers (Figure 2).

The maternal grandfather of the index patient (I:1) had had emergency surgery for a ruptured abdominal aortic aneurysm (AAA) of 12 cm, at age 78. One year afterwards, he presented again in the emergency room with a ruptured false aneurysm that developed at the left side of the prosthesis. Seven years later, he was again admitted in the emergency room with a rupture of a false aneurysm on the right side of the prothesis. He died at the age of 90 years and no specific cause of death or event prior was reported (“death from old age”). The family reported no obvious skin or joint phenotype that could fit with EDS.

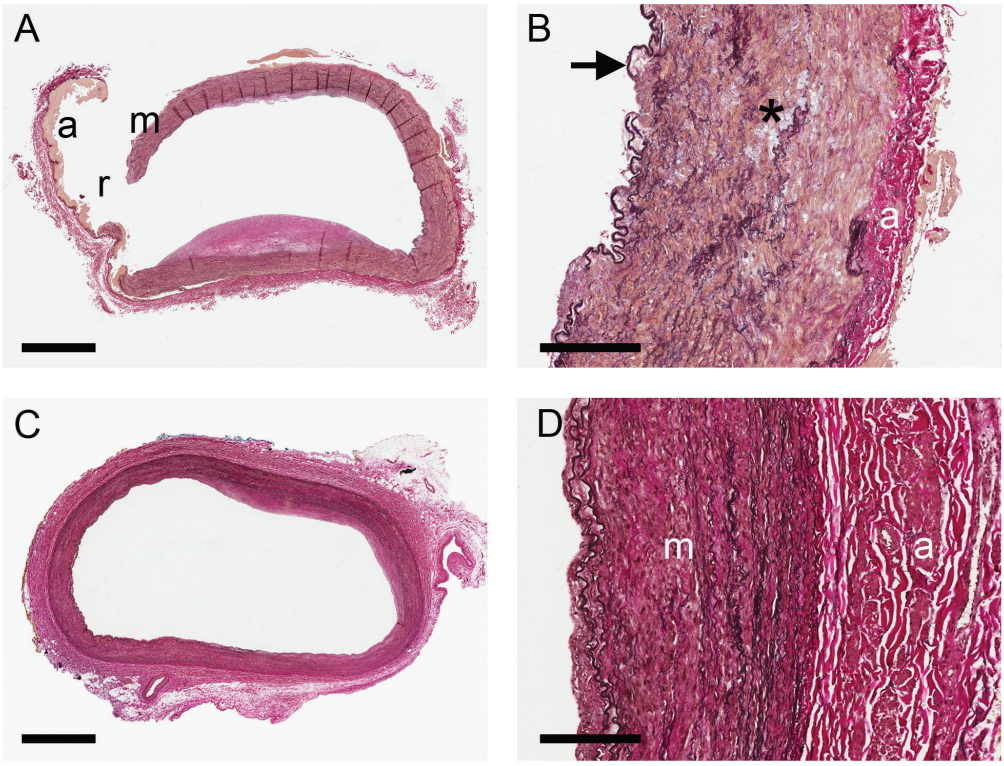


Figure 2 | Common iliac artery with rupture and control artery. A) Overview of the artery of the brother of the index patient (III:5) showing a rupture in the wall (r). Due to the rupture, the adventitia (a) is loose from the media (m). Elastin-van Gieson stain. Bar = 1 mm. B) Higher magnification of the arterial wall showing cystic degeneration of the media (indicated by *) with fragmentation of the elastic fibers (stained black). The arrow indicates the internal elastic lamina at the border of the intima and media. a, adventatia. Bar = 250 µm. C) Control artery of a 50-year-old individual. Bar = 1 mm. D) Higher magnification of the control artery without the cystic degeneration of the media (m). Bar = 250 µm.

Diagnostic analysis of *COL3A1* and other genes associated with vascular disease

In the proband's DNA, the coding region of *COL3A1* gene was analyzed first, but no mutation was detected. Subsequently, the coding regions of *COL1A1*, *COL1A2*, *TGFBR1*, *TGFBR2*, *ACTA2* and *SMAD3* genes were sequenced, but no abnormalities were found in these genes either. Furthermore, MLPA analysis of *COL3A1* detected no exonic deletions or duplications. Protein analysis by 14C-proline labeling in cultured fibroblasts showed no abnormalities of collagen type III; specifically there were no signs of over-modification and the ratio of collagen type I/ type III was normal (data not shown). Absence or presence of over-modification of collagen type V was inconclusive due to the low expression of *COL5A1*.

VASCULOME analysis

Average coverage of the VASCULOME target region was 137x, and 84% of the region was covered >=20x. After the filtering and prioritization process, the most obvious variant in the index patient and the only variant that segregated in the affected brother, was NM_000093.4:c.4610G>T resulting in a glycine substitution at the C-terminal side of the triple helix domain of *COL5A1*, p.(Gly1537Val) (Table 1) (Submitted to the Leiden Open Variant Database (LOVD) Ehlers-Danlos Syndrome Variant Database https://eds.gene.le.ac.uk/home.php?select_db=COL5A1). This variant was not described previously and was not present in available databases (dbSNP, Exome Variant Server, 1000Genomes, Genome of the Netherlands). Furthermore, the variant was predicted to be probably damaging by PolyPhen-2 and deleterious by SIFT software. Importantly, the affected amino acid belongs to a group of repetitive residue triplets consisting mainly of glycine-Xaa-Yaa, which is highly conserved among species. In general, glycine substitutions in the triple helix of collagen genes are known to lead to a disturbance of the triple helix formation, delayed processing and over-modification of the protein and are therefore considered to be pathogenic. Subsequently, segregation analysis of all novel variants found in the proband (III:7) was performed on DNA obtained from autopsy material of the index's brother (III:5), DNA isolated from blood of their father (II:5) and the parents and living siblings of their mother (I:1, I:2, II:3 and II:4) (Figure 1 and Supplementary Table S2).

The affected brothers (III:5 and III:7) only shared the *COL5A1* mutation. Neither one of the parents or tested siblings of the index's mother carried the *COL5A1* mutation, nor did the father of the brothers. Short Tandem Repeat genotyping confirmed parental relationships between the children and their father. This was all highly suggestive of a *de novo COL5A1* mutation in the mother (II:6), which was inherited by both her sons (III:5 and III:7), as the major cause of the severe familial vascular disease.

Table 1 | Clinical and molecular findings in patients with *COL5A1* mutations and a vascular phenotype, including the patients from previous reports and from the current study.

Article	Vascular Events	Age (yrs)	Skin phenotype	Joint phenotype	Other features	Mutation *COL5A1*
Borock et al., 2010	Rupture left common iliac artery	42	Easy bruising, pigmented scars, hyperextensibility		Inguinal hernias Hypertension	c.3184C>T p.(Arg1062*)
de Leeuw et al., 2012	Symptomatic superior mesenteric artery aneurysm	9*	Hyperextensibility Atrophic scars Easy bruising	Hypermobility		c.4691G>A p.(Gly1564Asp)
Mehta et al., 2012	Dissection iliac artery & iatrogenic dissection of contralateral artery	43	Hyperextensibiliy Atrophic scars	Hypermobility	Hypertension, tendon rupture	c.2185C>T p.(Gln729*)
Karaa and Stoler, 2013	Symptomatic aneurysms of superior-inferior mesenteric artery and celiac artery	11	Hyperextensibility, Atrophic scars	Hypermobility	Pectum excavatum, kyphoscoliosis, C1-C2 instability	c.2765G>A p.(Gly922Asp)
this report (mother)	Renal artery rupture	28*			Gingivitis	
this report (eldest son, proband)	Ruptured left artery subclavia A. coeliacus aneurysm Subarachnoidal bleeding & aortic rupture	15 18 43*	Tranlucent skin	Hypermobility		c.4610G>T p.(Gly1537Val)
this report (youngest son)	Right a. iliaca rupture	34*			Osteoporosis	c.4610G>T p.(Gly1537Val)

***Fatal vascular event**

DISCUSSION

In the current report, we describe a familial form of clinically diagnosed vascular EDS, leading to early death (28, 34 and 43 years) due to massive arterial bleeding. Also, the features observed at physical examination of the proband (translucent skin, hypermobility restricted to the finger joints and the absence of extreme skin fragility) resembled this type of EDS. However, while *COL3A1* mutations are usually found in vascular EDS patients, such a mutation could not be identified in the proband. Using a targeted next-generation sequencing approach, a pathogenic mutation was discovered in the *COL5A1* gene. *COL5A1* mutations are usually responsible for the classic form of EDS. In classic EDS patients, the connective tissue disease generally leads to an extreme joint and skin phenotype, but no major vascular events are commonly described. In contrast, since the mouse orthologue of *COL5A1* is expressed in the developing aorta (Roulet et al. 2007) and heterozygous knockout mice show decreased aortic stiffness (Wenstrup et al. 2006), murine studies suggest an important role for type V collagen in the vasculature. Importantly, four recent reports illustrate rupture, dissection and/or aneurysm in patients with a proven *COL5A1* mutation (Borck et al. 2010; de Leeuw et al. 2012; Karaa and Stoler 2013; Mehta et al. 2012). A summary of their clinical and molecular findings, as well as for the patients included in this present study, is provided in Table 1. The patient described by de Leeuw et al. (2012), who also died at young age due to complications of vascular disease, was shown to have a very similar mutation in COL5A1 as the patients described in this study: a substitution of a glycine at the C-terminal end of the triple helix domain (de Leeuw et al. 2012). Up until now, only a few glycine substitutions at the C-terminus end of the triple helix domain of COL5A1 have been described, with the mutation detected in the patient described by de Leeuw et al. (2012) and the one identified in this study as the closest to the C-terminus (Supplementary Figure S1), https://eds.gene.le.ac.uk/home.php?select_db=COL5A1) (Greenspan et al. 1991; Takahara et al. 1995). The number of glycine substitutions identified in this particular part of the triple helix is far too limited to establish a genotype-phenotype correlation. However, it is theoretically conceivable that this type of mutation in this part of the protein generates a more severe phenotype, because the assembly of the triple-helix starts at the C-terminus. Substitution of a glycine in this part of the protein may therefore strongly interfere with helix formation, and delay the triple helix assembly and lead to severe over-modification. Such phenotype-genotype correlation has been described for *COL1A1* in osteogenesis imperfect (Marini et al. 2007) where glycine substitutions nearby the C-terminus of the triple helix are associated with severe or lethal osteogenesis imperfecta, while glycine substitutions nearby the N-terminus of the triple helix as well as mutations leading to haploinsufficiency are associated with milder forms of the disease.

A second hypothesis to explain the severe vascular phenotype associated with the *COL5A1* mutation could be the presence of a genetic modifier in the described family. Remarkably, the grandfather of our proband suffered from a large, complicated AAA rupture, albeit at older age. An AAA is not uncommon in older males, with prevalence ranging from 5 to 8% (Pleumeekers et al. 1995; Sakalihasan et al. 2005). Pseudo-aneurysms occur in approximately 3% of treated patients (Hallett et al. 1997). Extremely large aneurysms, as detected in the grandfather of our proband (AAA of 12 cm), would be uncommon for vascular EDS, because the fragile vessel is expected to rupture before the artery reaches such an enlarged diameter. We therefore consider the clinical picture of the grandfather as a different phenotype, which is in keeping with the segregation analysis of the *COL5A1* mutation. However, it is conceivable that in this family a genetic modifier exists, which aggravated the vascular phenotype caused by the *COL5A1* mutation of the mother and her two sons. This modifier could also be responsible for the exceptionally large AAA in the grandfather and the complicated course following the repair, with the development and rupture of two false aneurysms indicative of some connective tissue weakness.

Our findings have important implications for the family. For the children of the proband and his brother who are in their early teens, it is now possible to perform DNA-diagnostics, in order to confirm or exclude the presence of the *COL5A*1 mutation. The impact of (presymptomatic) carriership will be extensively counseled. If the mutation is present, we will provide lifestyle advice (no contact sport, wear medic-alert) and consider treatment with celiprolol (Ong et al. 2010). If the mutation is absent, this will allow reassurance, although we may offer vascular imaging at low intervals in adulthood, because we cannot exclude the presence of genetic modifiers predisposing to vascular disease.

A general clinical consideration resulting from our study, and of the previous reports on major arterial complication in patients with *COL5A1* mutations, is that *COL5A1* should be analyzed in patients suspected of vascular EDS without a detected *COL3A1* mutation. Furthermore, classic EDS patients with *COL5A1* mutations may be at risk for vascular events. We do not recommend systematic counseling and screening of classic EDS patients for vascular disease at this stage, but clinicians should be aware of the possible complications. This is particularly true when glycine substitutions are identified nearby the C-terminal end of the triple helix, or when former medical charts or family history indeed reveal vascular events.

We conclude that the identified c.4610G>T p.(Gly1537Val) mutation in *COL5A1* can give a similar vascular EDS phenotype as a *COL3A1* mutation. This could be due to the particular position of this glycine substitution, and suggests a genotype/phenotype correlation. Alternatively, we hypothesize that in the presented family there may be a genetic modifier which, in synergy with the *COL5A1* mutation, induces a dramatic vascular phenotype.

ACKNOWLEDGEMENTS

We are grateful to the family that participated in our study. A.F. Baas was supported by a grant from the Dr. E. Dekker Program of the Netherlands Heart Foundation (2009T001). The authors declare no conflict of interest.

Supplementary Table S1 | List of genes included in the VASCULOME next-generation sequencing panel.

ABCA1	BMP2	COL6A3	GPI	LAMB1	MYH9	PLCB2	ROCK2
ABCC6	BMP4	COL6A6	GPR124	LAMB2	MYL10	PLCB3	RPS6KB1
ABCG5	BMP5	COL8A1	GPX1	LAMB3	MYL12A	PLCB4	RPS6KB2
ABCG8	BMP6	COL8A2	GRB2	LAMB4	MYL12B	PLCD1	RTN4
ABO	BMP7	CPA3	GREM1	LAMC1	MYL2	PLCD3	RXRA
AC005522.1	BMP8A	CR753892.3	GSK3B	LAMC2	MYL5	PLCE1	SAMHD1
AC233992.10	BMP8B	CR812478.5	HDAC1	LAMC3	MYL6	PLCL1	SCARB1
ACAT2	BMPR1A	CR933878.7	HDAC2	LDLR	MYL6B	PLK1	SH2B3
ACE	BMPR1B	CR942185.1	HDAC3	LEFTY1	MYL7	PLN	SKP1
ACE2	BMPR2	CREBBP	HDAC5	LEFTY2	MYL9	PLOD1	SLC27A4
ACTA2	BOLL	CTBP1	HDAC7	LFNG	MYLK	PLOD3	SLC2A10
ACTB	BRAF	CTBP2	HDAC9	LIPF	MYLK2	PLXND1	SMAD1
ACTG1	BX248519.2	CTGF	HES1	LMNA	MYLK3	PMF1	SMAD2
ACTG2	BX284686.4	CTNNA1	HHIPL1	LNPEP	MYO1E	PNLIP	SMAD3
ACTN1	BX284686.6	CTNNA2	HSPG2	LPA	MYOCD	PNLIPRP1	SMAD4
ACTN2	BX927239.2	CTNNA3	ID1	LPL	NCKAP1	PNLIPRP2	SMAD5
ACTN3	BX927320.4	CTNNB1	ID2	LRP1	NCL	PPAP2A	SMAD6
ACTN4	C17orf57	CTSA	ID3	LRP5	NCOR2	PPAP2B	SMAD7
ACVR1	CACNB4	CTSG	ID4	LRP6	NCSTN	PPAP2C	SMAD9
ACVR1C	CACNG1	CUL1	IFNG	LTBP1	NDP	PPARA	SMG6
ACVR2A	CACNG2	CXCL12	IGF1	LTBP2	NDUFV2	PPARG	SMTNL1
ACVR2B	CACNG3	CYP17A1	IGF1R	LTBP3	NF1	PPP1R12B	SMURF1
ACVRL1	CALD1	CYP46A1	IGF2	LTBP4	NLN	PPP1R14A	SMURF2
ADAM17	CALM1	CYP4A11	IGF2R	MAML1	NODAL	PPP2CA	SNF8
ADAMTS7	CALM2	CYP4A22	IGFBP1	MAML2	NOG	PPP2CB	SNW1

Supplementary Table S1 | *Continued.*

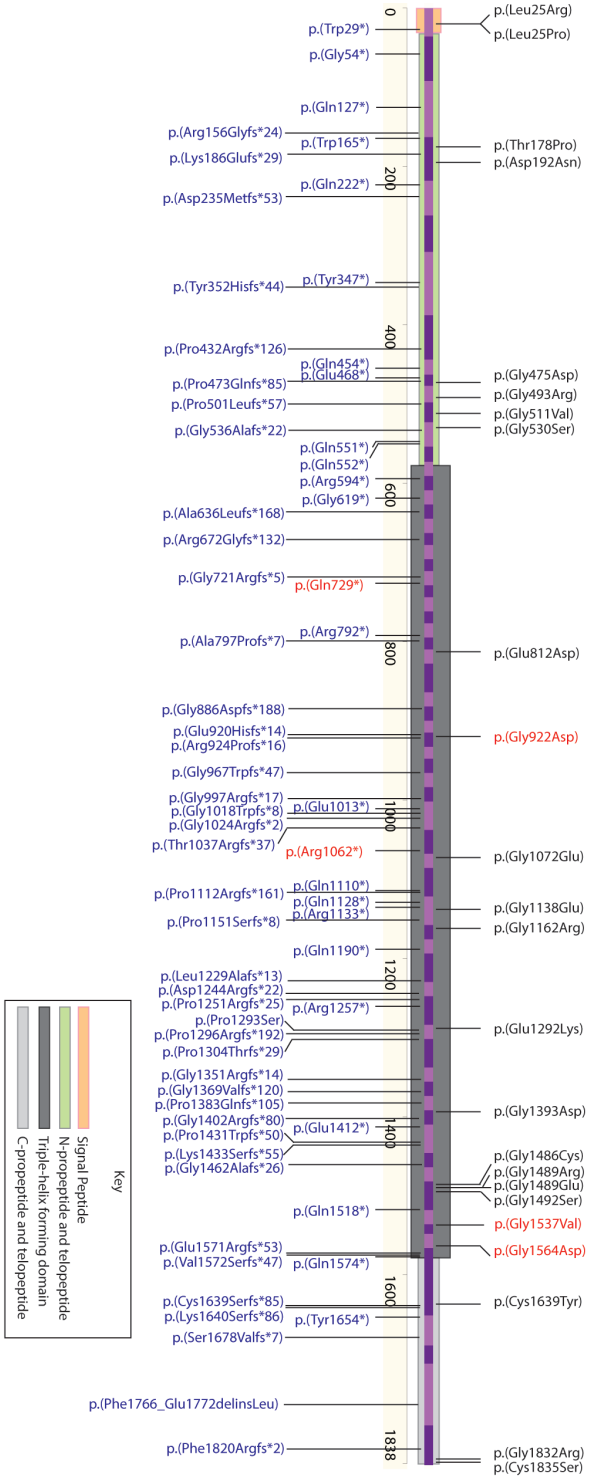
ADCY1	CALM3	DAB2IP	IGFBP2	MAML3	NOS3	PPP2R1A	SOX17
ADCY2	CALML3	DCN	IGFBP3	MAP2K1	NOTCH1	PPP2R1B	SP1
ADCY3	CALML5	DGAT1	IGFBP4	MAPK1	NOTCH2	PPP2R2A	SPECC1L
ADCY4	CALML6	DGAT2	IGFBP5	MAPK3	NOTCH3	PPP2R2C	SPP1
ADCY5	CASK	DHCR7	IGFBP6	MAPK8	NOTCH4	PPP2R2D	SRR
ADCY6	CASP3	DIAPH1	IGFBP7	MAPK9	NPC1L1	PPP3CB	SSR4
ADCY7	CASP7	DLL1	ITGA1	MARS2	NPPA	PPP3R1	STARD13
ADCY8	CASP8	DLL3	ITGA10	MAS1	NPPC	PRKAB1	TCF21
ADCY9	CASQ2	DLL4	ITGA11	MCM3	NPR1	PRKCH	TFDP1
ADM	CAV1	DTX1	ITGA2	MCM6	NPR2	PRKCI	TGFA
ADORA2A	CAV2	DTX3	ITGA2B	MED12	NR3C1	PRKCQ	TGFB1
ADORA2B	CAV3	DTX3L	ITGA3	MFNG	NRARP	PRKG2	TGFB2
ADRA1A	CCM2	DTX4	ITGA4	MGP	NRAS	PRKX	TGFB3
ADRA1B	CD36	DVL1	ITGA5	MIA3	NRG1	PRNP	TGFBR1
ADRA1D	CDC42	DVL2	ITGA6	MME	NRP1	PROK2	TGFBR2
AGPAT1	CDKN2A	DVL3	ITGA7	MMP1	NT5C2	PROX1	TGFBR3
AGPAT2	CDKN2B	E2F4	ITGA8	MMP10	NUMB	PRPF40A	THBS1
AGT	CDKN2B-AS1	E2F5	ITGA9	MMP12	NUMBL	PRRX2	THBS2
AGTR1	CEL	EFEMP2	ITGAV	MMP13	PARD3	PSEN1	THBS3
AGTR2	CHRD	EGFL7	ITGB1	MMP14	PCSK9	PSEN2	THBS4
AKT1	CIR1	EGFR	ITGB4	MMP19	PDCD10	PSENEN	THOP1
AL662828.2	CLPS	EGLN1	ITGB5	MMP2	PDGFRB	PSMA7	TIMP1
AL662847.2	CMA1	EIF1B	ITGB6	MMP3	PEMT	PTCRA	TIMP2
AL845464.5	CNNM2	ELN	ITGB7	MMP9	PHACTR1	PTEN	TIMP3
AL929587.5	COL11A1	ENG	ITGB8	MOGAT2	PITX2	PTGIR	TIMP4

Supplementary Table S1 | *Continued.*

AMH	COL11A2	EP300	JAG1	MOGAT3	PKD1	PTK2	TNC
AMHR2	COL14A1	FABP1	JAG2	MRAS	PKD2	PTPRJ	TNF
ANKS1A	COL15A1	FABP2	JM-JD7-PLA2G4B	MRPS6	PLA2G10	PTTG2	TNN
ANPEP	COL18A1	FBLN5	JUP	MRVI1	PLA2G12A	RAMP1	TNR
APH1A	COL1A1	FBN1	KARS	MTHFR	PLA2G12B	RAMP2	TNXB
APOA1	COL1A2	FBN2	KAT2A	MTTP	PLA2G1B	RAMP3	UBE2Z
APOA4	COL2A1	FBN3	KAT2B	MYC	PLA2G2A	RASD1	VCAN
APOA5	COL3A1	FHL1	KCNMA1	MYH1	PLA2G2C	RB1	VEGFA
APOB	COL4A1	FLNA	KCNMB1	MYH10	PLA2G2D	RB1CC1	VKORC1
APOC3	COL4A2	FN1	KCNMB2	MYH11	PLA2G2E	RBBP8	WDR12
APOE	COL4A3	FST	KCNMB3	MYH13	PLA2G2F	RBL1	ZC3HC1
ARAF	COL4A3BP	GDF5	KCNMB4	MYH14	PLA2G3	RBL2	ZFYVE16
ARHGEF1	COL4A4	GDF6	KCNU1	MYH15	PLA2G4A	RBPJ	ZFYVE9
ARHGEF11	COL4A5	GDF7	KDR	MYH2	PLA2G4B	RBPJL	ZNF259
ARHGEF12	COL4A6	GGCX	KRIT1	MYH3	PLA2G4E	RBX1	
ATP5G1	COL5A1	GNA12	LAMA1	MYH4	PLA2G5	REN	
ATP6V0A2	COL5A2	GOT2	LAMA2	MYH6	PLA2G6	RFNG	
AVP	COL5A3	GP5	LAMA3	MYH7	PLAT	RFTN2	
AVPR1A	COL6A1	GP6	LAMA4	MYH7B	PLAU	RHOA	
AVPR1B	COL6A2	GP9	LAMA5	MYH8	PLCB1	ROCK1	

Supplemental Table S2 | Novel variants found in the VASCULOME analysis of our index case, including segregation analysis with available DNA from family members.

Gene	cDNA Change	Amino Acid Change	I:1	I:2	II:3	II:4	II:5	III:5*	III:7*
NCL	NM_005381.2:c.777A>T	p.(Glu259Asp)	AA	AA	AA	AA	A/T	AA	A/T
LAMA4	NM_001105206.2:c.3572A>C	p.(His1191Pro)	AA	A/C	AA	A/C	AA	AA	A/C
RXRA	NM_002957.4:c.4314G>T	5' splice site region	GG	GG	GG	GG	G/T	GG	G/T
COL5A1	NM_000093.4:c.4610G>T	p.(Gly1537Val)	GG	GG	GG	GG	GG	G/T	G/T
PLCB3	NM_000932.2:c.3073G>A	p.(Glu1025Lys)	GG	G/A	GG	GG	GG	GG	G/A



Supplemental Figure S1 | *COL5A1* variants associated with EDS, as reported in the LOVD EDS variant database. Exon boundaries are designated with light/dark purple alternations. Amino acid numbers are given below the cDNA structure. Different domains of the protein are indicated with different colors. *COL5A1* variants associated with CLASSIC EDS are presented in black (missense variants) and blue (nonsense / frameshift / in-frame variants). *COL5A1* variants associated with vascular events are in red. Large duplications, exon deletions, exon skips, and splice site variants are not shown but can be viewed in the LOVD EDS database.

3

Heterozygous *KIDINS220/ARMS* nonsense variants cause spastic paraplegia, intellectual disability, nystagmus, and obesity

Human Molecular Genetics 2016 Jun 1;25(11):2158-2167.
Epub 2016 Mar 22.

PMID: 27005418; DOI:10.1093/hmg/ddw082

Dragana J. Josifova[1*], Glen R. Monroe[2,3*], Federico Tessadori[2,3,4*], Esther de Graaff[5], Bert van der Zwaag[2], Sarju G. Mehta[6], The DDD Study[7], Magdalena Harakalova[2], Karen J. Duran[2,3,] Sanne M.C. Savelberg[2,3,] Isaac J. Nijman[2,3], Heinz Jungbluth[8,9,10], Casper C. Hoogenraad[5], Jeroen Bakkers[4,11], Nine V. Knoers[2,3], Helen V. Firth[6,7], Philip L. Beales[12], Gijs van Haaften[2,3**], Mieke M. van Haelst[2**]

1 | Department of Clinical Genetics, Guys' & St. Thomas' Hospital, London SE1 7EH, United Kingdom
2 | Department of Genetics, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
3 | Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
4 | Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht 3584 CT, The Netherlands
5 | Division of Cell Biology, Faculty of Science, University of Utrecht, Utrecht 3584 CH, The Netherlands
6 | Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge CB2 0QQ, United Kingdom
7 | Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1RQ, United Kingdom
8 | Department of Paediatric Neurology, Evelina Children's Hospital, Guy's & St. Thomas' Hospital NHS Foundation Trust, London SE1 7EH, United Kingdom
9 | Randall Division of Cell and Molecular Biophysics, Muscle Signalling Section
10 | Department of Basic and Clinical Neuroscience, IoPPN, King's College, London WC2R 2LS, United Kingdom
11 | Department of Medical Physiology, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
12 | Genetics and Genomics Medicine Program, UCL Institute of Child Health, London WC1N 1EH, United Kingdom
*Joint First Authors**Joint Last Authors.

ABSTRACT

We identified *de novo* nonsense variants in *KIDINS220/ARMS* in three unrelated patients with spastic paraplegia, intellectual disability, nystagmus, and obesity (SINO). KIDINS220 is an essential scaffold protein coordinating neurotrophin signal pathways in neurites and is spatially and temporally regulated in the brain. Molecular analysis of patients' variants confirmed expression and translation of truncated transcripts similar to recently characterized alternative terminal exon splice isoforms of *KIDINS220*. *KIDINS220* undergoes extensive alternative splicing in specific neuronal populations and developmental time points, reflecting its complex role in neuronal maturation. In mice and humans, *KIDINS220* is alternatively spliced in the middle region as well as in the last exon. These full-length and *KIDINS220* splice variants occur at precise moments in cortical, hippocampal, and motor neuron development, with splice variants similar to the variants seen in our patients and lacking the last exon of *KIDINS220* occurring in adult rather than in embryonic brain. We conducted tissue-specific expression studies in zebrafish that resulted in spasms, confirming a functional link with disruption of the KIDINS220 levels in developing neurites. This work reveals a crucial physiological role of *KIDINS220* in development and provides insight into how perturbation of the complex interplay of *KIDINS220* isoforms and their relative expression can affect neuron control and human metabolism. Altogether, we here show that de novo protein-truncating *KIDINS220* variants cause a new syndrome, SINO. This is the first report of *KIDINS220* variants causing a human disease.

INTRODUCTION

Next Generation Sequencing (NGS) techniques provide a time and cost-efficient method to identify novel genes. We used this approach to investigate two patients presenting with an apparently unique combination of features, including obesity, spastic paraplegia, intellectual disability, nystagmus, and macrocephaly and identified different, *de novo*, heterozygous, nonsense variants in *KIDINS220* in each patient. A third patient with similar features was ascertained through DECIPHER (decipher.sanger.ac.uk) after the DDD study shared a *de novo* candidate variant identified by trio exome analysis (The Deciphering Developmental Disorders Study 2015; Wright et al. 2015).

KIDINS220 (Kinase D interacting substrate of 220 kDa), also known as ARMS (Ankyrin Repeat-rich Membrane Spanning), is a conserved scaffold protein that controls axonal and dendritic maturation (Iglesias et al. 2000; Kong et al. 2001). The spatiotemporal expression of different *KIDINS220* isoforms is finely tuned in the mammalian brain, and knockout *KIDINS220* animals show developmental central nervous system anomalies (Cesca et al. 2011; Cesca et al. 2012; Wu et al. 2009). Recently it was shown that distinctive neuronal populations express different KIDINS220 splice isoforms at specific times during development, including newly characterized alternative terminal exon splice isoforms (Schmieg et al. 2015). The expression of the complete range of these isoforms, with the corresponding domains for thus far poorly characterized interaction partners, may be crucial for proper neuronal development. We show here that our patients' variants result in truncated isoforms of KIDINS220 similar to an alternative terminal exon splice isoform expressed mainly during adulthood. Our patients' neurological phenotype may be caused by disruption of full-length KIDINS220 splice isoform repertoire expression and lack of essential interaction domains that are necessary during embryonic development of the neural network. We further show that perturbation of *KIDINS220* levels causes a spasticity phenotype in zebrafish. We therefore suggest that *KIDINS220* variants cause a new syndrome characterized by spastic paraplegia, intellectual disability, nystagmus, and obesity (SINO syndrome).

RESULTS

De novo KIDINS220 variants are implicated in patients with a unique phenotype of SINO

We performed NGS analysis in three unrelated patients with SINO (Figure 1 and Table 1, Supplement 1).

Figure 1 | Clinical features of patients. A) Patient 1 at one year of age. B) Patient 2 at one year of age. C) Patient 3 at 26 months of age. Note obesity, macrocephaly, brachycephaly and eye abnormalities.

Two patients were sequenced for a targeted gene panel of 582 genes to investigate the suspected genetic cause of syndromic and non-syndromic obesity (OBESITOME, UMC Utrecht, Netherlands) (Supplemental Table 1). We identified heterozygous *KIDINS220* variants, NM_020738.2:c.4050G>A, p.(Trp1350*), and NM_020738.2:c.4096C>T, p.(Gln1366*), in patient 1 and 2, respectively. Both variants result in a premature stop codon. Their *de novo* occurrence was confirmed by Sanger sequencing of both patients and their parents. There were no further *de novo* variants in other shared genes on the OBESITOME panel. Independently, whole-exome sequencing on patient 3 revealed a heterozygous *de novo* insertion of an A nucleotide, NM_020738.2:c.4520dup, resulting in a frameshift and predicted premature truncation of the protein p.(Leu1507Phefs*4) (Figures 2A and 2B) (The Deciphering Developmental Disorders Study 2015; Wright et al. 2015).

Table 1 | Clinical phenotype of the three patients with *KIDINS220 de novo* variants.

	Patient 1	Patient 2	Patient 3
Prenatal Scan	Dilatation of lateral ventricles	Dilatation of lateral ventricles	Dilatation of lateral ventricles
Polyhidramnious	-	+	?
Gestational age (weeks)	38	40	41
Weight (birth)	2.26kg (2nd centile)	3.65kg (75th centile)	3.28 (50th centile)
Weight (Infancy)	>99.6th centile	99.6th centile	98th centile
Weight (Puberty)	>99.6th centile	>99.6th centile	Overweight
Height (infancy)	91st – 98th centile	91st centile	75th centile
Height (Puberty)	91st centile	91st centile	?
OFC (birth)	25th -50th centile	?	?
OFC (infancy)	>>99.6th centile	99.6th centile	98th centile
OFC (puberty)	99.6th centile	>99.6th centile	25th – 50th centile
Feeding pattern	NI	NI	NI
Dysmorphic features	Brachyplagiocephaly Bossed forehead Deep set eyes	Brachyplagiocephaly Prominent forehead Deep-set eyes Crowded teeth	Brachyplagiocephaly Prominent forehead
Vision	Reduced acuity Hypermetropia Astigmatism	Squint Hypermetropia Astigmatism	Squint Hypermetropia Esotropia
Nystagmus	+	+	+
ERG	NI	?	?
VEP	Post retinal dysfunction	?	?
Development	Delayed	Delayed	Delayed
Neurology	Axial hypotonia Spastic paraplegia	Spastic paraplegia	Spastic paraplegia
Brain MRI	Dilated, lateral and third ventricles Normal fourth ventricle Reduced white matter bulk Mild delay in myelination Mild generalised atrophy	Dilated, third and lateral ventricles Normal fourth ventricle Reduced white matter bulk Mild generalised atrophy	Dilated lateral ventricles High riding 3rd ventricle Partial agenesis of corpus callosum
Bone age	?	Slightly delayed	?
Eating habits	NI	NI	NI
Development	Moderate global delay	Moderate global delay	Moderate global delay

NI: normal, +: present, -: absent,?: not reported, OFC: Occipito Frontal Circumference. ERG: Electroretinogram. VEP: Visual Evoked Potential.

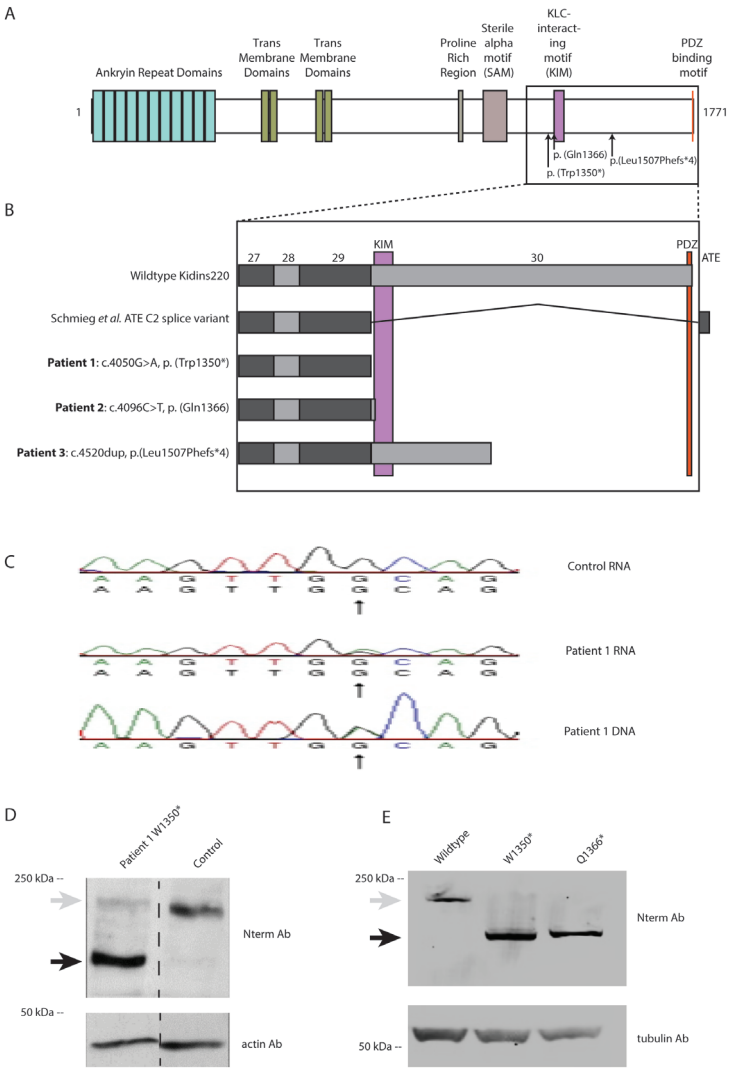


Figure 2 | *De novo* dominant nonsense variants in *KIDINS220*. A) Variants p.(Trp1350*) and p.(Gln1366*) result in truncated protein without the full KIM domain, and p.(Leu1507Phefs*4) results in a slightly longer protein containing the full KIM domain. B) Comparison of the wild-type *KIDINS220* exonic structure versus an alternate terminal ending splice variant described by Schmieg et al. (2015) (Schmieg et al. 2015) and the three patients described in this study. C) *KIDINS220* cDNA sequencing verifies that there is no mRNA nonsense mediated decay in patient 1. Western blot analysis of (D) total extracts from patient 1 versus a control fibroblast and (E) total HEK293T extracts expressing wild-type and variant (Trp1350* and Gln1366*) forms of KIDINS220. Expression of KIDINS220 was tested with antibodies against the N-terminal part of KIDINS220 (Nterm Ab). The truncated protein products caused by KIDINS220 variants are indicated with a black arrow and the expected full-length wild-type protein product by a grey arrow. Anti-actin was used as a loading control for the fibroblasts and anti-tubulin for the HEK293T cells.

Variant *KIDINS220* RNA expression results in truncated KIDINS220 in patient fibroblasts and HEK293T cells

All variants identified here fall within the last two exons of the gene (Figure 2A and 2B). Sanger sequencing on fibroblast cDNA from patient 1 showed that variant c.4050G>A resulted in proper mRNA expression (Figure 2C). This result was confirmed by Quantitative PCR, with the relative expression of *KIDINS220* before and after the variant c.4050G>A equal in patient fibroblast cDNA compared to control fibroblast cDNA, indicating that the shortened transcript escaped nonsense-mediated decay (NMD) (data not shown). The mammalian NMD pathway is unable to recognize premature termination codons in the last exon and the last ~55 bp of the penultimate exon (Miller and Pearce 2014). Shortened mRNA transcripts that escape NMD are translated into truncated protein products possibly lacking important C-terminal domains. Variants in the penultimate exon resulting in premature stop codons that escape NMD are reported in other syndromes, for example, autosomal dominant Robinow syndrome (White et al. 2015). Furthermore, we used cDNA sequencing to assess known alternative splicing of exon 25 and exon 26 in the middle region of *KIDINS220*, which has not yet been determined in fibroblast cells. Patient and control cDNA sequencing of the region flanking and containing exons 25 and 26 revealed that fibroblasts from both our patient 1 and control express *KIDINS220* with exon 26 removed by alternative splicing (Genbank Accession Number KJ812120), indicating that exon 26 is removed in all fibroblast cells (data not shown).

We analyzed total protein extract of fibroblasts of patient 1 by Western blot analysis and detected a 150 kDa band in addition to the 220 kDa wild-type KIDINS220 band, confirming expression of the truncated, Trp1350* protein (Figure 2D). The variants of patient 1 and 2 were also overexpressed separately in HEK293T cells to visualize the size of the truncated protein products, using human isoform containing both exon 25 and exon 26 (Genbank Accession Number KJ812119) (Schmieg et al. 2015). The presence of a 150 kDa protein coinciding with the p.(Trp1350*) variant of patient 1 and a 167 kDa protein coinciding with the p.(Gln1366*) variant of patient 2 were detected by Western blot. Only wild-type 220 kDa protein was identified in HEK293T extracts overexpressing the wild-type KIDINS220 (Figure 2E). The KIDINS220 protein product of these variants is strikingly similar to that produced by alternative terminal exon (ATE) splicing in mouse and humans (ATE C2) (Figure 2B) (Schmieg et al. 2015).

Motor neuron specific expression of *KIDINS220* variants induces spasms in zebrafish

Zebrafish is a powerful, versatile vertebrate model to study human motor neuron diseases (Babin et al. 2014; Patten et al. 2014). By using the Gal4/UAS approach (Halpern et al. 2008; Tessadori et al. 2012) we achieved motor neuron-specific expression of human *KIDINS220* variants in zebrafish embryos and analyzed their effect in vivo. Transgenic tg(Isl1BAC:GalFF) fish used here constitutively express Gal4 in motor neurons; when DNA constructs containing an upstream activation sequence (UAS) driving transcription are supplemented by microinjection, mosaic larvae are obtained in which Gal4 binds to the UAS sequence to express *KIDINS220* solely in motor neurons. Embryos tg(Isl1BAC:GalFF) microinjected with UAS:KIDINS220 wild-type or variant p.(Trp1350*) displayed spasms in the trunk at 5 days post fertilization (dpf), confirming a functional link between spasticity and disruption of KIDINS220 levels and/or function (Figure 3). This was slightly more evident in embryos expressing the mutant *KIDINS220*. Representative recordings of the spasm phenotypes can be found in the Supplemental material online.

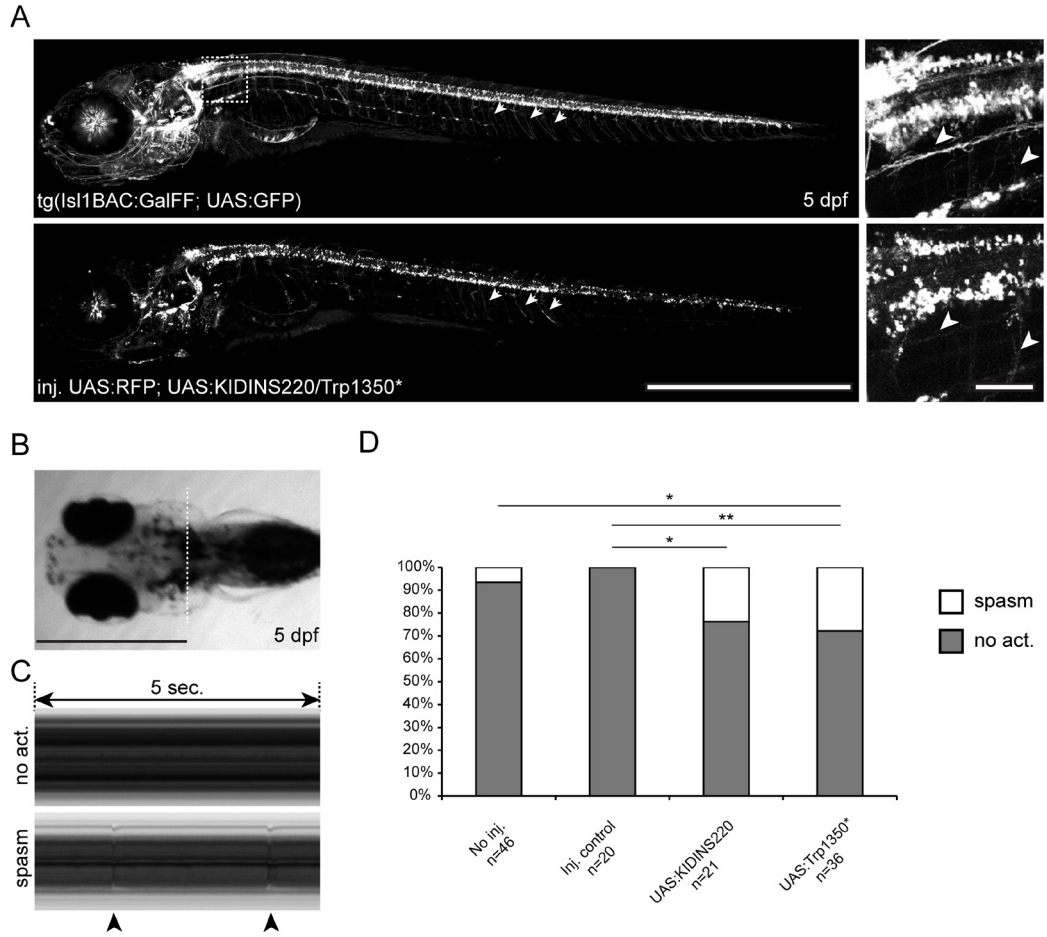


Figure 3 | Effect of *KIDINS220* and Trp1350* expression in zebrafish embryo motor neurons. A) Tissue-specific expression of *KIDINS220* variants in 5 dpf zebrafish embryos was achieved by microinjection of UAS:KIDINS220 and UAS:Trp1350* DNA in tg(Isl1BAC:GalFF; UAS:GFP) embryos (upper panel). A DNA UAS:RFP construct was co-injected (lower panel) to evaluate the efficiency of the microinjections. White arrowheads: axonic extensions of the motor neurons. B) 5 dpf zebrafish embryo (ventral view; anterior is left). White dotted line: position of kymographs presented in C. C) Visualization of spastic activity on thoracic kymographs. Representative kymographs of embryos showing no spastic activity (upper panel, no act.) and spastic activity (lower panel, spasm). Black arrowheads point at spastic events as visualized on the kymograph. Representative recordings are available as supplementary files. D) Quantification of spastic events in the different conditions of the study. Expression of *KIDINS220* wt and Trp1350* in motor neurons results in increased spastic activity. No act.: No spastic activity. *P<0.05; **P<0.01; Fisher's exact test. Scale bars: 1mm in A (left panels), B; 50 microns in A (right panels)

KIDINS220 variants do not result in different cellular localizations in Neuro-2A cells

During differentiation, KIDINS220 is transported to the neurite tips of PC12 neuronal cells, a process that is altered in KIDINS220 isoforms with an alternative terminal exon (Iglesias et al. 2000; Schmieg et al. 2015). We examined the cellular location of wild-type and truncated KIDINS220 in Neuro-2A cells, a cell line where KIDINS220 is endogenously expressed, albeit at lower levels than in PC12 cells (Iglesias et al. 2000). Following cotransfection of HA-tagged wild-type or p.(Gln1366*) *KIDINS220* as well as a mCherry reporter, Neuro-2A cells were differentiated with retinoic acid for a total of 5 days. Fixation was performed at day 5 and immunofluorescence performed with an anti-HA antibody. Differentiated cells expressing wild-type or p.(Gln1366*) exhibited localization of KIDINS220 throughout the entire cell body, as well as in the neurite tips. There was no difference in the cellular localization of the wild-type or variants (Figure 4).

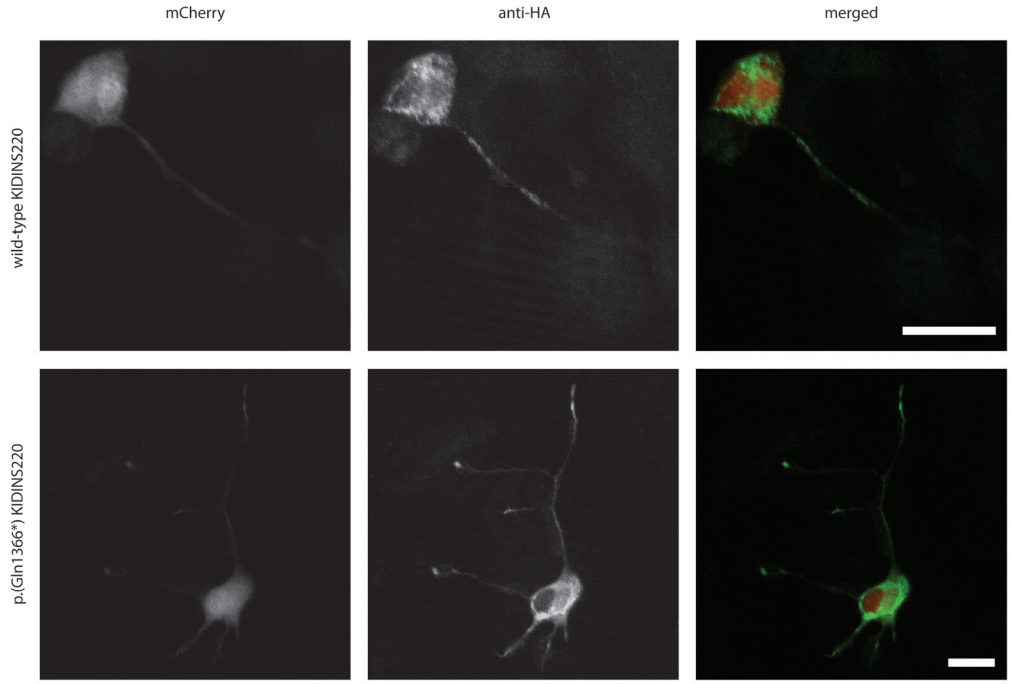


Figure 4 | Kidins220 truncated variant is able to localize to neurite tips in Neuro-2A cells. Representative image of transfected KIDINS220 wild-type and variant p.(Gln1366*) in differentiated Neuro-2A cells. Immunofluorescence of cotransfected mCherry (red), anti-HA-tagged wild-type or p.(Gln1366*) KIDINS220 (green) show that the truncated KIDINS220 variant is transported to developing neurite tips. Scale bar = 20 microns

DISCUSSION

KIDINS220 controls axonal and dendritic maturation in the developing mammalian brain, and is integral to the branching, stabilization, and complexity of cortical and hippocampal dendrites (Iglesias et al. 2000; Kong et al. 2001; Wu et al. 2009). The cytoplasmic N-terminus contains 11 ankyrin repeats. There are four transmembrane segments in the central part of the molecule (Arevalo et al. 2004). The C-terminal tail, also exposed to the cytoplasm, contains several protein-protein interacting domains: a proline-rich domain, Sterile Alpha Motif (SAM), Kinesin Light Chain (KLC) - Interactive Motif (KIM) and PDZ binding motif (Bracale et al. 2007; Kong et al. 2001). These domains enable the recruitment of cell-specific adaptors and effectors to activated Tropomyosin receptor Kinases (TrK) and are involved in functional interactions with other receptors such as the Glutamate, Ephrin and Vascular Endothelial Growth Factor receptors (Cesca et al. 2012; Lopez-Menendez et al. 2009). The KIM domain mediates intracellular trafficking of KIDINS220 through direct binding to KCL-1 and -2 and is important for a correct response to neurotrophic stimuli (Bracale et al. 2007; Neubrand et al. 2012). Expression of *KIDINS220* is developmentally regulated and differential splicing occurs in specific tissues such as the brain, heart, and skeletal muscle. The alternative splicing is complex and developmental occurrence of *KIDINS220* splice isoforms with the middle region and/or last exon removed varies in cortical, hippocampal and motor neurons (Schmieg et al. 2015). Schmieg et al. (2015) demonstrated in mice that *KIDINS220* with exon 30 included is expressed in embryonic tissue; in contrast, a similar *KIDINS220* splice variant lacking exon 30 is present solely in adult tissue. Motor neurons exhibited the highest variety of alternative splicing in the middle region, but did not exhibit any alternative terminal exon splicing at any time during development (Schmieg et al. 2015). Our patients' variants result in isoforms similar to *KIDINS220* splice variants with ATE splicing. The constitutive expression of truncated isoforms similar to ATE splice isoforms that are not usually expressed in the developing embryo or present in motor neurons may perturb the complex splice regulation necessary for proper neurite development and result in the phenotype seen in our patients.

In this manuscript, we show that *KIDINS220* nonsense variants are causal for a characteristic phenotype most likely due to disruption of the repertoire of splice isoforms available for developing neuronal populations. Interference in the delicate balance of *KIDINS220* isoform expression in the neurite tips will impede the various Neurotrophin, Glutamate, Ephrin and Vascular Endothelial Growth Factor receptor associations required for correct neuronal network development. The C-terminal region of KIDINS220 contains the KIM and PDZ domains that are required for protein interactions and transport. Lack of the KIM domain and its interaction with KCL-1 subunit KIF5A (Kinesin Family Member 5A) impairs protein transport to newly formed neurite tips (Bracale et al. 2007). Similarly, absence of

a PDZ domain will interfere with the interaction of KIDINS220 with SNX27 (Sorting Nexin Family Member 27), an important protein for transport to the plasma membrane (Steinberg et al. 2013). We here show in *KIDINS220* transfected Neuro-2a cells, both wild-type and KIDINS220 variant p.(Gln1366*) successfully localized to the differentiating neurite tips, indicating an alternate as-of-yet undetermined molecular pathology due to the variants seen in our patients.

Studies of Kidins220 during zebrafish and Xenopus Laevis embryogenesis showed dynamic expression in the nervous system including the hypothalamus, eye, branchial arches, heart and somites, partly correlating with the anomalies observed in our patients (Andreazzoli et al. 2012; Marracci et al. 2013). Mice completely lacking Kidins220 are not viable (Cesca et al. 2011; Cesca et al. 2012; Wu et al. 2009). The extensive neuronal cell death in the brain of mutant embryos results in abnormal ventricle enlargement. This observation appears concordant with the pre- and post-natal ventriculomegaly observed in our patients. Heterozygous mice display defects in dendritic growth and branching, indicating the role of Kidins220 in Brain-Derived Neurotrophic Factor (BDNF)-induced dendritic development (Cesca et al. 2012; Chen et al. 2012b; Wu et al. 2009). Although the heterozygous and wild-type mice have, at several age time points, comparable total Kidins220 protein levels in the brain, it seems likely that the temporal balance between Kidins220 isoforms in the brain is crucial for correct neural development (Cesca et al. 2011). Structural cardiac anomalies and defects of the vascular and immune system were also observed in *KIDINS220* -/- embryos (Cesca et al. 2011; Cesca et al. 2012), but are not seen in any of our patients, possibly due to different isoforms being required for normal heart and brain development (Schmieg et al. 2015).

Zebrafish is an excellent model for human motor neuron disorders (Babin et al. 2014). Motor neuron-specific expression in zebrafish using the wild-type or variant p.(Trp1350*) KIDINS220 resulted in spasms. The observed difference in the induction of spasms in zebrafish embryos expressing the wild-type or p.(Trp1350*) truncated KIDINS220 may reflect the complex regulation of *KIDINS220* tissue-specific splicing and expression. More conclusively, disruption of *KIDINS220* expression levels in motor neurons functionally links KIDINS220 deregulation to nervous system impairment. The specific features in our zebrafish model, in particular the presence of spasms that are not observed in affected humans, confirm the effects of KIDINS220 perturbation on neuronal function but are also likely to reflect anatomical differences in the central nervous system between zebrafish and humans. Specifically, the absence of corticospinal and rubrospinal tracts in zebrafish results in the replication of human lower motor neuron disorders more faithfully than human upper motor neuron disorders on the phenotypical level. KIDINS220 is expressed in cortical and hippocampal neurons, encompassing the upper motor neuron system, and

as such this dysfunction can be classified as an upper motor neuron disorder. Nonetheless, zebrafish have provided a suitable partial model for insights into the molecular and genetic mechanisms of other upper motor neuron disorders, such as hereditary spastic paraplegias (Allison et al. 2013; Fassier et al. 2010; Wood et al. 2006).

The pathophysiology of obesity is difficult to explain. KIDINS220 is known to act as a downstream substrate for protein kinase D and mediates multiple receptor signaling including BDNF and other neurotrophins. Since neurotrophins have been implicated in the molecular and cellular processes underlying body weight regulation, disruption of this finely tuned TrK pathway could possibly explain the obesity phenotype in these patients (Kernie et al. 2000; Liao et al. 2013). Increased expression of *KIDINS220* has been linked with perturbation of TrkA/p75 neurotrophin receptor complex association and overexpression of the differently spliced KIDINS220 ATE C2 leads to a substantial increase in TrkA expression (Chang et al. 2004; Schmieg et al. 2015). Further research is needed to provide insight into this interaction.

Not every loss-of-function variant in the last exon of *KIDINS220* causes the phenotype observed in our patients as illustrated by the presence of three *KIDINS220* variants that are present heterozygously in five healthy individuals in the Exome Aggregation Consortium database (Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org; accessed September 2015). These variants, resulting in p.(Glu1530Ter), p.(Arg1736Ter) and p.(Ser1740Ter), are shortly after the variant in our patient 3, pointing to a specific effect of the variants presented in this study.

Our patients manifest similarities with MOMO Syndrome (**M**acrosomia, **O**besity, **M**acrocephaly and **O**cular abnormalities; OMIM 157980), however prenatal ventricular dilatation and spastic paraplegia are not features of this condition. *KIDINS220* Sanger sequencing of ten MOMO patients did not yield any variants. We therefore suggest that our patients have a distinct syndrome characterized by **S**pastic paraplegia, **I**ntellectual disability, **N**ystagmus and **O**besity (SINO).

In conclusion, we show here that *de novo* heterozygous nonsense *KIDINS220* variants cause SINO syndrome. This is the first report of *KIDINS220* variants causing a human disease. Future research might provide more insights in the role of *KIDINS220* in the pathogenesis of spastic paraplegia and obesity.

MATERIALS AND METHODS

Clinical Features

The clinical features of the three patients are summarized in Figure 1 and Table 1 and case reports can be found in Supplement 1.

NGS and analytical pipeline

NGS on patients 1 and 2 was performed with fragment libraries prepared and enriched for the genomic regions of interest using a 1M custom OBESITOME microarray (Agilent Technologies, Santa Clara, CA, USA) as previously described (Harakalova et al. 2011; Monroe et al. 2015). The OBESITOME contained 582 obesity-related genes, with the genes selected on presumed obesity relevance from several databases, pathway sources, and research articles (Supplemental Table 1). The OBESITOME consisted of 55 samples that were pooled and run as a full slide on the SOLiD 5500XL (Life Technologies, Carlsbad, CA). Color space reads were mapped and aligned against GRCh37/hg19 reference genome using a custom pipeline based on the BWA software and annotated as described previously (Harakalova et al. 2011; Nijman et al. 2010). Mean sample coverage for the entire run of 55 samples on the OBESITOME was 100X, median coverage was 87X and 85% of genomic positions were covered by more than 20 reads. The mean coverage for patient 1 and 2 was 161X and 170X, respectively. The criteria for variant detection coverage were set at 10 unique reads, and a non-stringent cut-off for heterozygote allele calls was set at 25-75%. Variants were filtered against the reference frequencies from the datasets of NCBI dbSNP Build 137 for Human (https://www.ncbi.nlm.nih.gov/SNP), Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS), 1000Genomes (http://www.1000genomes.org), Genome of the Netherlands (http://www.nlgenome.nl) or our in-house dataset, with novel variants and rare variants less than a 0.5% minor allele frequency threshold retained for further analysis (databases accessed July 2014). For each patient, a final variant list was compiled including the location in the genomic sequence of the variant, the amino acid change and prioritized on the conservation score and predicted effect on protein function using prediction programs (GERP, PolyPhen-2, SIFT). Confirmation and segregation analysis in the family of the selected single nucleotide variants or small insertions and deletions was performed by Sanger sequencing.

Whole-exome sequencing and analysis for patient 3 was performed as described by the Deciphering Developmental Disorders (DDD) study (The Deciphering Developmental Disorders 2015; Wright et al. 2015). DNA of patient 3 was extracted from saliva samples and exome capture performed using Agilent Sureselect 55 MB Exome Plus (Agilent

Technologies, Santa Clara, CA, USA). Exomes were sequenced on the Illumina HiSeq (Illumina, San Diego, CA, USA) and *de novo* analysis performed using DeNovoGear (Ramu et al. 2013). *De novo* variants were validated using Sanger sequencing.

All variants have been submitted to the Leiden Open Variation Database (LOVD) located at http://databases.lovd.nl/shared/genes/KIDINS220 (April 2016).

KIDINS220 Sanger sequencing in a MOMO cohort and patient cDNA

To determine if variants within *KIDINS220* were causal for MOMO syndrome, the full coding regions and exon-intron boundaries of *KIDINS220* were analyzed in a cohort of 10 MOMO patients by Sanger sequencing.

To evaluate the impact of *KIDINS220* variants on RNA expression on our patients, total RNA was isolated from fibroblasts of patient 1 using TRizol Reagent (ThermofFisher Scientific, Waltham, MA, USA) and converted to cDNA with the High-Capacity cDNA Archive Kit (Life Technologies, Carlsbad, CA, USA). Material from patients 2 and 3 was not available. Sanger sequencing on the full cDNA of patient 1 was performed. Furthermore, Quantitative PCR was performed on cDNA of both patient and control fibroblasts using primers before and after the variant c.4050G>A to evaluate if expression of the mRNA was reduced compared to the control. Primer sequences:

5'-CCTGAAGACCCACGTTTCC-3'/5'-AAGCTGAAGTTGAGTGTGTAGG-3' and
5'-GCTCAGATGTCCCAGTTAGAAG-3'/5'-CTGATGAACTCTGACCCATGTAATA-3'.

Fibroblast and HEK293 protein preparation and immunoblotting

Fibroblasts of patient 1 and control samples were cultured, lysed, and Western blot analysis performed on the total protein using an antibody recognizing the N-terminus of KIDINS220 (Neubrand et al. 2010). Fibroblasts of patient 1 and control samples were cultured according to standard conditions. Frozen cell pellets were lysed at 4°C in 50mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton, 1% Na-deoxycholate, 0.1% SDS and 2mM EDTA pH8.0 in the presence of protease inhibitors (cOmplete, Roche, Basel, Switzerland) and phosphatase inhibitors (PhosSTOP, Roche, Basel, Switzerland) for 30 min, and sonicated for 10 seconds. Samples were spun for 5 minutes at 13.000rpm at 4°C.

Western blot analysis was then performed on the total protein extract of the patient fibroblasts using an antibody recognizing the N-terminus of KIDINS220 (Neubrand et al. 2010). The supernatant was diluted in 2x Sample buffer (8% SDS/25% Glycerol/0.05M Tris pH 6.8/200mM DTT/ Bromophenol Blue/H2O) and separated on 6% or 8% SDS-PAGE gels,

followed by blotting onto PVDF membranes (Biorad, Veenendaal, Netherlands). Blots were blocked with either 2% BSA/0.05% Tween/PBS and incubated overnight with the KIDINS220 antibody. Blots were washed with 0.05%Tween/PBS for 3 times 10 minutes each at room temperature and incubated with the appropriate secondary antibody. After washing three times in 0.05% Tween/ PBS and once in PBS, blots were developed with Pierce Enhanced Chemiluminescent Western Blotting Substrate (Life Technologies, Carlsbad, CA, USA).

To visualize the molecular weight of the truncated proteins of patients 1 and 2, we introduced the identified point variants in HEK293T cells and performed Western blot analysis on the total protein as described above. HEK293T cells were grown at 37°C and 5% CO2 in DMEM: HAMF10 (1:1) medium containing 10% FBS and 1% penicillin/streptomycin. The wild-type *KIDINS220* sequence corresponding to the long isoform of human *KIDINS220* (Genbank Accession Number KJ812119) was obtained from an expression construct by using full-length *KIDINS220* primers (primers available upon request) (Bracale et al. 2007). Wild-type *KIDINS220* was ligated into the TOPO entry plasmid pCR8/GW/TOP (Life Technologies BV, Bleiswijk, the Netherlands) and the mutations encoding for Trp1350* and Gln1366* were engineered into the *KIDINS220* expression construct using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA) and custom designed primers. The presence of the introduced mutations was confirmed by Sanger sequencing. HEK293T cells were transiently transfected with wild-type and variant DNA constructs using Polyethylenimin (Polyscience, Eppelheim, Germany), grown for 24 hours after transfection and prepared and analyzed as described earlier.

Zebrafish expression construct and microinjections in zebrafish embryos

The wild-type *KIDINS220* cDNA sequence or the patient variant Trp1350* were introduced into pCR8/GW/TOPO. The *KIDINS220* middle entry vectors were then used in a multi-site Gateway reaction (ThermoFisher Scientific, Waltham, MA, USA) together with p5E-UAS, p3E-IRES-EGFPpA and as destination vector pDestTol2pA3 (Kwan et al. 2007) to obtain UAS:KIDINS220 expression vectors. Circular plasmid DNA of UAS:KIDINS220 constructs was injected at 30 ng/µl in the presence of 25 ng/µl Tol2 mRNA into tg(Isl1BAC:GalFF; UAS:GFP) (Tessadori et al. 2012) embryos at the 1-cell stage together with 10 ng/µl pT2ZUASRFP as a fluorescent reporter (Asakawa et al. 2008). Healthy embryos displaying robust Isl1-specific RFP fluorescence at 5dpf were used for spastic activity recordings.

Zebrafish High Speed Imaging

Embryos at 5 dpf were mounted in 3% methylcellulose (Sigma-Aldrich, Zwijndrecht, the

Netherlands) prepared in E3 embryonic medium prior to imaging for 20-60 seconds with a Hamamatsu C9300-221 high speed CCD camera (Hamamatsu Photonics, Hamamatsu City, Japan) at 82-165 fps mounted on a Leica DM IRBE inverted microscope (Leica Microsystems GmbH, Wetzlar, Germany) at room temperature using Hokawo 2.1 imaging software (Hamamatsu Photonics GmbH, Herrsching am Ammersee, Germany). Image analysis was subsequently carried out with ImageJ (http://rsbweb.nih.gov/ij; accessed June 2015). Fischer's exact test was employed to determine significance between zebrafish groups with different construct injections.

Transfection and immunofluorescence of transfected Neuro-2A cells

GW1-MCS expression constructs with a N-terminal HA tag were created with wild-type or p.(Gln1366*) *KIDINS220* inserts. Neuro-2A cells were seeded at 10,000 cells/cm2 on poly-L-lysine coated 10 mm coverslips and grown for 24 hours in DMEM supplemented in 10% FCS and penicillin/streptomycin. The following day, cells were transfected in DMEM only using Lipofectamine2000 (Thermo Fischer Scientific, Waltham, MA, USA) with N-terminal HA-tagged wild-type or p.(Gln1366*) KIDINS220 in the GW1-MCS expression vector and cotransfected with a pcDNA/FRT/TO/mCherry expression construct. Cells were incubated for 4 hours at 37°C and 5% CO2 and medium removed, washed with PBS and replaced with differential medium (DMEM, 1% FCS, 5 uM retinoic acid). Cells were differentiated for 5 days, and differential medium was replaced after 2 and 4 days.

For fixation, medium was removed, the cells were washed once with PBS, and the coverslip treated with 4% PFA (paraformaldehyde) for 15 minutes at room temperature before being washed again with PBS. Cells were then permeabalized with 0.1% TritonX100 in PBS for 20 minutes, washed with PBS, blocked in 3% BSA/PBS (bovine serum albumin) for 20 minutes, and washed again. The coverslip was then incubated with 2 ug/ml rabbit Anti-HA-tag antibody Chip Grade in 3% BSA/PBS (ab9110; Abcam, Cambridge, UK) overnight at 4°C. Coverslips were washed three times in PBS and probed with 1:1000 polyclonal goat anti-rabbit CY5 (Jackson ImmunoResearch, West Grove, PA, USA) in 3% BSA/PBS for one hour at room temperature. Cells were washed 3 times in PBS and then stained in 1:5000 DAPI (Thermo Fisher Scientific, Waltham, MA, USA), in 3% BSA/PBS for 15 minutes, washed 3 times in PBS, and mounted using Vectashield Antifade Mounting Medium (Vector Laboratories, Burlingame, CA, USA). Coverslips were visualized using a TCS SP8 Confocal Laser Scanning Platform (Leica, Wetzlar, Germany).

Study Approval

Written informed consent from the parents of the all three patients was received prior to inclusion in this study.

Acknowledgements

The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. The research team acknowledges the support of the National Institute for Health Research (NIHR), through the Comprehensive Clinical Research Network. PLB is an NIHR Senior Investigator. The DECIPHER database: http://decipher.sanger.ac.uk. We are grateful to the patients and the parents of the patients who agreed to participate in this study.

Funding

The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [grant number WT098051]. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC).

Supplementary Videos are available at HMG online.

Supplement 1: Case reports of three SINO patients

Patient 1 is a 14-year-old boy, who was first seen by a clinical geneticist at the age of 12 months. He is the second of non-identical twins, product of the first pregnancy of a healthy, unrelated couple. His twin sister is in good health and of normal development. The 23 weeks' prenatal scan revealed dilated lateral ventricles, confirmed on foetal MRI scan at 32 weeks gestation. He was delivered at 38 weeks by elective caesarean section with a birth weight of 2.268kg (2nd centile) and head circumference 33.7cm (25th – 50th centile). He did not require resuscitation and spent the first 24 hours in Special Care Baby Unit for observation. In the first year of life, he presented with accelerated longitudinal growth,

weight gain, and rapid increase in head circumference. His twin sister maintained normal growth parameters on a very similar diet. He had nystagmus and poor visual behaviour. At the age of one year, his height was between 91st - 98th centile, weight above the 99.6th centile and head circumference, at 52.1cm, was significantly above the 99.6th centile. He had brachycephaly, bossed forehead, deep set eyes, full cheeks and small mouth with normal palate and teeth (Figure 1A). Hands and feet were in proportion, with tapering fingers. His motor development corresponded to a four-month-old infant, with some residual axial hypotonia. His visual behaviour remained poor. On ophthalmological assessment, he had a small astigmatic error, no squint, normal fundi, slightly larger optic discs with mild, bilateral temporal palor. Tone and reflexes in his lower limbs were increased. Brain MRI scan showed stable lateral and third ventricular enlargement when compared with the foetal MRI imaging, and a normal fourth ventricle. The white matter bulk was reduced, and there was mild delay in myelination on the background of a mild generalised brain atrophy. Spine MRI was unremarkable. Electroretinogram was normal; however, visually-evoked responses showed post retinal dysfunction. At the age of two years his growth parameters remained on the same trajectory, while his global developmental delay was becoming more apparent as he was still unable to sit unaided. His visual behaviour had improved. The patient continued to make slow, but steady progress. He was able to speak in sentences by four years of age. Neurologically, he had markedly increased tone in his lower limbs, brisk reflexes and equivocal plantars. Binocular vision was 6/48, equivalent to 0.90 Logmar crowded Kay pictures. Fundi were grossly normal. Nystagmus persisted and he was registered partially sighted. Despite a well-controlled diet and a good excercise regime the patient remained obese with rather generalised fat distribution. Interestingly, having entered puberty, he began to lose some of his excessive weight.

Patient 2 is a 15-year-old boy and is the only child of healthy, unrelated parents. Pregnancy was complicated by polyhydramnios. Prenatal scan showed dilated lateral ventricles at 20 weeks' gestation. He was born at term weighing 3.65kg (75th centile). Postnatal brain MRI confirmed dilatation of both lateral and the third ventricles. The fourth ventricle was normal. He was rapidly gaining weight and crossed the 99.6th centile by 3 months old, while his length was on the 91st centile and head circumference above the 99.6th centile. He was referred for a genetic opinion at one year of age. He had brachyplagiocephaly, a prominent and high forehead, deep-set eyes, crowded teeth and a normal palate. Ears were fleshy, but at normal position. He had a left convergent squint and nystagmus (Figure 1B). Ophthalmological assessment revealed binocular vision of 6/9.5, myopia, astigmatism, and a small alternating convergent squint. Follow-up brain MRI at the age of two years showed dilated, although stable, third and lateral ventricles (compared with the foetal MRI imaging). The fourth ventricle was normal. There was a lack of white matter bulk and mild cerebral atrophy. MRI of the spine was normal. Bone age was two years at

the chronological age of 2 years and 6 months. His motor skills were significantly more impaired due to his spastic paraplegia. His growth parameters remained on the same trajectory, with his weight and head circumference above the 99.6th centile and height on the 91st centile. He spoke in sentences by the age of four years. He was known to take normal amounts of food and to have a healthy diet.

Patient 3 is a 7-year-old boy and the second child of healthy unrelated parents. The genetics team first saw him at the age of 26 months. He was born at 41 weeks' gestation by spontaneous vaginal delivery weighing 3.28 kg (50th centile) with good Apgar scores of 9 and 10.at 1 and 5 minutes, respectively. There had been brief episodes of antepartum bleeding at 26 and 32 weeks' gestation. He was described as a "very hungry baby", feeding every 2 hours. His weight at 4 months was on the 25th centile but by 11 months it had increased to the 98th centile, with length on the 75th centile and head circumference on the 91st centile. He was referred to a dietician at the age of one year and subsequently his weight stabilised on the upper centiles of the normal range (Figure 1C). He smiled at 3-4 months, sat at one year and had a few single words by 2 years. He presented with global developmental delay and had markedly increased tone in his lower limbs with increased reflexes when examined at the age of 26 months. Ophthalmology review documented nystagmus and a squint with hypermetropia and right-sided esotropia. A cranial MRI scan at the age of 2.5 years revealed slight prominence of the lateral ventricles, a high riding 3rd ventricle and agenesis of the corpus callosum. Only the most anterior part of the genu of the corpus callosum had formed. He was diagnosed with spastic paraplegia. OFC measurement at age 6 years and 4 months was 53.0 cm (25th-50th centile). At the age of 7 years he attends a special needs school and is functioning at the level of a 4-year-old with regard to academic activities and has developed some challenging behaviour. He wears splints to improve his mobility and uses a rollator but is unable to walk independently. He has a prominent abdomen, with a truncal distribution of weight and poor core muscle strength.

Supplementary Table S1 | List of Ensembl gene IDs for all genes included on the OBESITOME. Genes from the following KEGG pathways were included: insulin signaling (map04910), adipocytokine signaling (map04920), type II diabetes mellitus (map04930), neurotrophin signaling (map04722) as well as the fat digestion and absorption pathway (map04975). Genes from the Online Mendelian Inheritance in Man (OMIM) with a related obesity phenotype were included. Finally, genes in research articles implicated in obesity through copy number variations (CNVs) or Genome Wide Association Studies (GWAS) were included. The OBESITOME design encompassed uniquely mapping coding regions of 582 genes from the genome build GRCh37/hg19 (http://www.ncbi.nlm.nih.gov/assembly/2758) as well as 50bp flanks into intronic regions, resulting in a total footprint size of 1.5 Mb.

ENSG00000116678	ENSG00000156510	ENSG00000097007	ENSG00000140992	ENSG00000174628	ENSG00000227642
ENSG00000174697	ENSG00000156515	ENSG00000099875	ENSG00000141200	ENSG00000174938	ENSG00000228333
ENSG00000050748	ENSG00000157017	ENSG00000100078	ENSG00000141510	ENSG00000174939	ENSG00000228892
ENSG00000102882	ENSG00000157764	ENSG00000100504	ENSG00000141522	ENSG00000174943	ENSG00000229593
ENSG00000104365	ENSG00000159346	ENSG00000100784	ENSG00000141564	ENSG00000174982	ENSG00000230143
ENSG00000107643	ENSG00000159399	ENSG00000100994	ENSG00000141837	ENSG00000174992	ENSG00000231321
ENSG00000109339	ENSG00000160014	ENSG00000101188	ENSG00000141934	ENSG00000175104	ENSG00000232157
ENSG00000115138	ENSG00000160691	ENSG00000102871	ENSG00000142273	ENSG00000175535	ENSG00000232280
ENSG00000133124	ENSG00000160883	ENSG00000102879	ENSG00000142875	ENSG00000175634	ENSG00000233301
ENSG00000169047	ENSG00000160999	ENSG00000102886	ENSG00000142949	ENSG00000175766	ENSG00000233715
ENSG00000176697	ENSG00000162409	ENSG00000102893	ENSG00000143171	ENSG00000175806	ENSG00000235712
ENSG00000181092	ENSG00000163281	ENSG00000103197	ENSG00000143801	ENSG00000177189	ENSG00000235758
ENSG00000184557	ENSG00000163932	ENSG00000103485	ENSG00000143921	ENSG00000177606	ENSG00000236271
ENSG00000185950	ENSG00000166441	ENSG00000103495	ENSG00000145349	ENSG00000179222	ENSG00000236493
ENSG00000051382	ENSG00000167191	ENSG00000103502	ENSG00000145384	ENSG00000180616	ENSG00000236873
ENSG00000100030	ENSG00000167193	ENSG00000103740	ENSG00000147955	ENSG00000180739	ENSG00000237953
ENSG00000105221	ENSG00000168702	ENSG00000104312	ENSG00000148384	ENSG00000181625	ENSG00000238045
ENSG00000105647	ENSG00000168916	ENSG00000104643	ENSG00000148408	ENSG00000181903	ENSG00000238496
ENSG00000105851	ENSG00000169032	ENSG00000104812	ENSG00000148660	ENSG00000181927	ENSG00000239065
ENSG00000109424	ENSG00000169252	ENSG00000105835	ENSG00000149922	ENSG00000181935	ENSG00000240001
ENSG00000109819	ENSG00000169885	ENSG00000106384	ENSG00000149923	ENSG00000181939	ENSG00000240913

Supplementary Table S1 | *Continued.*

ENSG00000117020	ENSG00000172260	ENSG00000106615	ENSG00000149925	ENSG00000182333	ENSG00000242193
ENSG00000117461	ENSG00000173039	ENSG00000108443	ENSG00000149926	ENSG00000182473	ENSG00000242198
ENSG00000121879	ENSG00000174775	ENSG00000108946	ENSG00000149927	ENSG00000182759	ENSG00000243708
ENSG00000132170	ENSG00000175161	ENSG00000108953	ENSG00000149929	ENSG00000183336	ENSG00000244405
ENSG00000137764	ENSG00000175426	ENSG00000109458	ENSG00000149930	ENSG00000183604	ENSG00000244409
ENSG00000140718	ENSG00000175564	ENSG00000109919	ENSG00000149932	ENSG00000183943	ENSG00000245719
ENSG00000141506	ENSG00000175567	ENSG00000110090	ENSG00000150594	ENSG00000184111	ENSG00000246710
ENSG00000142208	ENSG00000177885	ENSG00000110395	ENSG00000150907	ENSG00000184216	ENSG00000246820
ENSG00000145675	ENSG00000178188	ENSG00000110851	ENSG00000151067	ENSG00000184381	ENSG00000247372
ENSG00000148053	ENSG00000178363	ENSG00000110931	ENSG00000151090	ENSG00000185000	ENSG00000247735
ENSG00000159723	ENSG00000178372	ENSG00000111252	ENSG00000151247	ENSG00000185149	ENSG00000248332
ENSG00000166603	ENSG00000179295	ENSG00000111348	ENSG00000151726	ENSG00000185305	ENSG00000248881
ENSG00000171105	ENSG00000180008	ENSG00000111713	ENSG00000152256	ENSG00000185386	ENSG00000249848
ENSG00000171608	ENSG00000181929	ENSG00000112038	ENSG00000152270	ENSG00000185559	ENSG00000250616
ENSG00000181856	ENSG00000184304	ENSG00000112062	ENSG00000152495	ENSG00000185652	ENSG00000251724
ENSG00000185338	ENSG00000185634	ENSG00000112246	ENSG00000152804	ENSG00000185905	ENSG00000252778
ENSG00000198793	ENSG00000186951	ENSG00000113161	ENSG00000154415	ENSG00000185928	ENSG00000253230
ENSG00000204490	ENSG00000188778	ENSG00000113163	ENSG00000155846	ENSG00000186298	ENSG00000253697
ENSG00000206439	ENSG00000196396	ENSG00000114302	ENSG00000156711	ENSG00000186350	ENSG00000254206
ENSG00000223952	ENSG00000197594	ENSG00000114423	ENSG00000156873	ENSG00000186577	ENSG00000254457
ENSG00000228321	ENSG00000198668	ENSG00000116473	ENSG00000157005	ENSG00000186660	ENSG00000254634
ENSG00000228849	ENSG00000213281	ENSG00000116711	ENSG00000157388	ENSG00000186886	ENSG00000254665
ENSG00000228978	ENSG00000222040	ENSG00000117215	ENSG00000158786	ENSG00000187021	ENSG00000254804
ENSG00000230108	ENSG00000249325	ENSG00000117394	ENSG00000160789	ENSG00000187486	ENSG00000254841

Supplementary Table S1 | *Continued.*

ENSG00000232810	ENSG00000254154	ENSG00000117560	ENSG00000162302	ENSG00000187840	ENSG00000254900
ENSG00000254647	ENSG00000254685	ENSG00000117676	ENSG00000162407	ENSG00000187980	ENSG00000254921
ENSG00000002330	ENSG00000004660	ENSG00000118046	ENSG00000162512	ENSG00000188089	ENSG00000257506
ENSG00000006831	ENSG00000004846	ENSG00000118137	ENSG00000162889	ENSG00000188130	ENSG00000257691
ENSG00000008196	ENSG00000005249	ENSG00000118263	ENSG00000163558	ENSG00000188191	ENSG00000258130
ENSG00000011478	ENSG00000006071	ENSG00000118689	ENSG00000163581	ENSG00000188257	ENSG00000258150
ENSG00000043591	ENSG00000006283	ENSG00000119729	ENSG00000163586	ENSG00000188439	ENSG00000011143
ENSG00000067606	ENSG00000007264	ENSG00000119912	ENSG00000163794	ENSG00000188779	ENSG00000035862
ENSG00000076555	ENSG00000010310	ENSG00000119938	ENSG00000164176	ENSG00000188784	ENSG00000102239
ENSG00000082701	ENSG00000013364	ENSG00000119973	ENSG00000164398	ENSG00000189045	ENSG00000107317
ENSG00000099942	ENSG00000015520	ENSG00000120341	ENSG00000164776	ENSG00000196927	ENSG00000110148
ENSG00000100485	ENSG00000021645	ENSG00000120437	ENSG00000164924	ENSG00000197142	ENSG00000113966
ENSG00000100889	ENSG00000028137	ENSG00000122008	ENSG00000165029	ENSG00000197442	ENSG00000119401
ENSG00000100906	ENSG00000044446	ENSG00000122033	ENSG00000165059	ENSG00000197471	ENSG00000122507
ENSG00000104825	ENSG00000054282	ENSG00000122679	ENSG00000165140	ENSG00000197943	ENSG00000125124
ENSG00000106617	ENSG00000058404	ENSG00000123739	ENSG00000165699	ENSG00000198001	ENSG00000125863
ENSG00000106633	ENSG00000062282	ENSG00000123983	ENSG00000165862	ENSG00000198064	ENSG00000135312
ENSG00000107263	ENSG00000064300	ENSG00000124181	ENSG00000166225	ENSG00000198106	ENSG00000136931
ENSG00000109320	ENSG00000065675	ENSG00000125166	ENSG00000166391	ENSG00000198216	ENSG00000138686
ENSG00000110244	ENSG00000065882	ENSG00000125733	ENSG00000166402	ENSG00000198400	ENSG00000140287
ENSG00000111725	ENSG00000067113	ENSG00000125743	ENSG00000166405	ENSG00000198909	ENSG00000140463
ENSG00000112664	ENSG00000067177	ENSG00000126767	ENSG00000166407	ENSG00000200983	ENSG00000141027
ENSG00000113580	ENSG00000067182	ENSG00000127191	ENSG00000166436	ENSG00000202479	ENSG00000143799
ENSG00000115137	ENSG00000067225	ENSG00000127314	ENSG00000166444	ENSG00000204231	ENSG00000163093

Supplementary Table S1 | *Continued.*

ENSG00000115392	ENSG00000067560	ENSG00000127472	ENSG00000166484	ENSG00000204310	ENSG00000164326
ENSG00000115592	ENSG00000068366	ENSG00000128245	ENSG00000166681	ENSG00000205534	ENSG00000165269
ENSG00000115904	ENSG00000068976	ENSG00000128272	ENSG00000166913	ENSG00000205560	ENSG00000165533
ENSG00000117569	ENSG00000069764	ENSG00000128573	ENSG00000167114	ENSG00000206289	ENSG00000171497
ENSG00000118432	ENSG00000070808	ENSG00000130377	ENSG00000167194	ENSG00000206324	ENSG00000174483
ENSG00000120833	ENSG00000070831	ENSG00000130413	ENSG00000167371	ENSG00000206379	ENSG00000179941
ENSG00000121853	ENSG00000071242	ENSG00000130749	ENSG00000167744	ENSG00000206480	ENSG00000181004
ENSG00000122585	ENSG00000072062	ENSG00000130957	ENSG00000168610	ENSG00000206625	ENSG00000182636
ENSG00000124089	ENSG00000072133	ENSG00000131910	ENSG00000168918	ENSG00000207069	ENSG00000183395
ENSG00000124253	ENSG00000072310	ENSG00000132207	ENSG00000168970	ENSG00000207333	ENSG00000184502
ENSG00000126934	ENSG00000073009	ENSG00000132376	ENSG00000169169	ENSG00000207701	ENSG00000189058
ENSG00000129946	ENSG00000073060	ENSG00000132589	ENSG00000169203	ENSG00000207773	ENSG00000196498
ENSG00000130748	ENSG00000075391	ENSG00000132825	ENSG00000169592	ENSG00000208010	ENSG00000198707
ENSG00000131096	ENSG00000076984	ENSG00000133636	ENSG00000169627	ENSG00000212607	
ENSG00000131482	ENSG00000078061	ENSG00000134070	ENSG00000169692	ENSG00000213058	
ENSG00000131791	ENSG00000078900	ENSG00000134259	ENSG00000169710	ENSG00000213341	
ENSG00000132155	ENSG00000079277	ENSG00000134308	ENSG00000170027	ENSG00000213599	
ENSG00000133703	ENSG00000079435	ENSG00000134313	ENSG00000170044	ENSG00000213639	
ENSG00000134243	ENSG00000079616	ENSG00000135930	ENSG00000170835	ENSG00000213648	
ENSG00000135218	ENSG00000080224	ENSG00000136238	ENSG00000170890	ENSG00000214725	
ENSG00000135472	ENSG00000080815	ENSG00000137154	ENSG00000170893	ENSG00000220583	
ENSG00000137392	ENSG00000082556	ENSG00000137309	ENSG00000171132	ENSG00000221320	
ENSG00000138821	ENSG00000084674	ENSG00000137312	ENSG00000171530	ENSG00000222375	
ENSG00000143627	ENSG00000084710	ENSG00000138031	ENSG00000171606	ENSG00000223654	

Supplementary Table S1 | *Continued.*

ENSG00000143933	ENSG00000087088	ENSG00000138075	ENSG00000171791	ENSG00000224165
ENSG00000146232	ENSG00000087916	ENSG00000138092	ENSG00000172188	ENSG00000224687
ENSG00000148082	ENSG00000090238	ENSG00000138160	ENSG00000172531	ENSG00000224740
ENSG00000151353	ENSG00000090376	ENSG00000138308	ENSG00000172572	ENSG00000225339
ENSG00000152254	ENSG00000095015	ENSG00000138823	ENSG00000173273	ENSG00000226467
ENSG00000152359	ENSG00000095637	ENSG00000139515	ENSG00000173281	ENSG00000226585
ENSG00000153885	ENSG00000096968	ENSG00000140538	ENSG00000174482	ENSG00000227322

4

Compound heterozygous *NEK1* variants in two siblings with oral-facial-digital syndrome type II (Mohr syndrome)

European Journal of Human Genetics. 2016 Dec;24(12):1752-1760.
Epub 2016 Aug 17.

PMID: 27530628; DOI: 10.1038/ejhg.2016.103.

Glen R. Monroe[1,2*], Isabelle F.P.M. Kappen[1,3*], Marijn F. Stokman[1,2*], Paulien A. Terhal[1], Marie-José H. van den Boogaard[1], Sanne M.C. Savelberg[1,2], Lars T. van der Veken[1], Robert J.J. van Es[4], Susanne M. Lens[2,5], Rutger C. Hengeveld[2,5], Marijn A. Creton[4], Nard G. Janssen[4], Aebele B. Mink van der Molen[3], Michelle B. Ebbeling[6], Rachel H. Giles[7], Nine V. Knoers[1,2], Gijs van Haaften[1,2]

1 | Department of Genetics, University Medical Center Utrecht, Lundlaan 6, 3584 EA, Utrecht, The Netherlands
2 | Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
3 | Department of Plastic Surgery, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
4 | Department of Oral and Maxillofacial Surgery and Special Dental Care, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
5 | Department of Molecular Cancer Research, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
6 | Department of Ophthalmology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
7 | Department of Nephrology and Hypertension, University Medical Center Utrecht, Regenerative Medicine Center-Hubrecht Institute, University Medical Center Utrecht, Uppsalalaan 6, 3584CT Utrecht, The Netherlands
*Joint First authors

ABSTRACT

The oral-facial-digital (OFD) syndromes comprise a group of related disorders with a combination of oral, facial and digital anomalies. Variants in several ciliary genes have been associated with subtypes of OFD syndrome, yet in most OFD patients the underlying cause remains unknown. We investigated the molecular basis of disease in two brothers with OFD type II, Mohr syndrome, by performing single nucleotide polymorphism (SNP) array analysis on the brothers and their healthy parents to identify homozygous regions and candidate genes. Subsequently, we performed whole-exome sequencing (WES) on the family. Using WES we identified compound heterozygous variants c.[464G>C;1226G>A] in NIMA (Never in Mitosis Gene A) - Related Kinase 1 (*NEK1*). The novel variant c.464G>C disturbs normal splicing in an essential region of the kinase domain. The nonsense variant c.1226G>A, p.(Trp409*), results in nonsense-associated alternative splicing, removing the first coiled-coil domain of NEK1. Candidate variants were confirmed with Sanger sequencing and alternative splicing assessed with cDNA analysis. Immunocytochemistry was used to assess cilia number and length. Patient-derived fibroblasts showed severely reduced ciliation compared to control fibroblasts (18.0% vs. 48.9%, $p<0.0001$) but showed no significant difference in cilia length. In conclusion, we identified compound heterozygous deleterious variants in *NEK1* in two brothers with Mohr syndrome. Ciliation in patient fibroblasts is drastically reduced, consistent with a ciliary defect pathogenesis. Our results establish *NEK1* variants involved in the etiology of a subset of patients with OFD syndrome type II and support the consideration of including (routine) *NEK1* analysis in patients suspected of OFD.

INTRODUCTION

The oral-facial-digital (OFD) syndromes comprise a group of related disorders with a specific combination of oral, facial and digital anomalies. To date, fourteen phenotypically different subtypes have been reported, of which nine have been well characterized (Gurrieri et al. 2007; Toriello and Franco 1993). These subtypes show an autosomal recessive inheritance pattern, except for OFD type I, type VII and type VIII, which are X-linked forms of OFD (Gurrieri et al. 2007; Toriello and Franco 1993). Common characteristic findings associated with these syndromes include: hamartomas of the tongue and a lobulated tongue, midline cleft of the upper lip and/or palate, gingival frenulae, dental anomalies, brachydactyly, syndactyly, pre- or post-axial polydactyly and cystic renal disease (Gurrieri et al. 2007; Toriello and Franco 1993).

In 2001, the causative gene for OFD type 1 (OMIM 300170), *OFD1*, was identified (Ferrante et al. 2001). Functional studies have shown that the encoded protein has an important role in the formation of primary cilia (Macca and Franco 2009; Romio et al. 2004). Because cilia play an essential role in the morphogenesis and tissue homeostasis of multiple organ systems, gene variants interrupting cilia function manifest systemically and result in a diverse phenotypic picture (Fry et al. 2014; Powles-Glover 2014; Satir et al. 2010). In the following years, variants in several other genes involved in ciliogenesis or ciliary signaling have been implicated in specific subtypes of OFD, namely *TCTN3* for OFD type IV (OMIM 258860), *TMEM216* and *C5orf42* for OFD VI (OMIM 277170), *DDX59* for an unclassified autosomal recessive form of OFD with intellectual disability, *TBC1D32/C6ORF170* or *SCLT1* for OFD IX (OMIM 258865) and *C2CD3* for an atypical type of OFD with severe microcephaly and cerebral malformations (Adly et al. 2014; Lopez et al. 2014; Shamseldin et al. 2013; Thauvin-Robinet et al. 2014; Thomas et al. 2012). Variants in two of these genes, *TCTN3* and *TMEM216*, are also found in other ciliopathies such as Joubert syndrome and Meckel-Gruber syndrome (Thomas et al. 2012; Valente et al. 2010). These findings indicate that multiple OFD subtypes can be classified as ciliopathies. However, in the majority of patients with an autosomal recessive form of OFD the underlying genetic cause remains to be identified.

In the present study, we performed genetic analysis on two brothers from consanguineous parents who have features compatible with OFD type II (Mohr syndrome, OMIM 252100). As several other OFD subtypes have been shown to be ciliary diseases, we presumed that the brothers would also have a ciliopathy. The extreme genetic heterogeneity of ciliopathies and the large number of genes involved in ciliary pathways, as well as the large potential number of yet unknown ciliary proteins, prompted us to perform whole-exome sequencing (WES) to identify the disease related gene in the family (Huber and Cormier-

Daire 2012; Sun et al. 2015).

SUBJECTS AND METHODS

Patient material collection and consent

Blood was obtained from both brothers and their parents after informed consent. Skin fibroblasts of the eldest brother were obtained during a scheduled knee operation. Informed consent was obtained for whole-genome SNP array analysis and WES in the diagnostics laboratory. The medical ethical committee of the University Medical Center Utrecht (UMCU) approved the WES patient information and informed consent form.

SNP array analysis

Genomic DNA was extracted from peripheral leukocytes according to standard protocols. SNP array copy number profiling and analysis of regions of homozygosity were performed according to standard procedures using the Infinium Human Omni Express Exome BeadChip (Illumina, San Diego, CA, USA). Subsequently, visualizations of SNP array results and data analysis were performed using Nexus 7.5 software (BioDiscovery, Los Angeles, CA, USA). Human genome build Feb. 2009 GRCh37/hg19 was used. Results were classified with Cartagenia BENCH software (Cartagenia, Leuven, Belgium).

Whole-exome sequencing

Fragment libraries of the DNA samples of the two brothers and their parents were prepared (Kapa Biosystems, Wilmington, MA, USA), enriched using Agilent Sureselect V5 (Agilent, Santa Clara, CA, USA) and sequenced as a Rapid Run on the Illumina HiSeq 2500 Sequencer (Illumina, San Diego, CA, USA) at the Utrecht Sequencing Facility (Utrecht, Netherlands). A phased VCF was created and imported into Cartagenia BENCHlab NGS module (Cartagenia, Leuven, Belgium) for variant interpretation. Due to the consanguineous nature of the parents, homozygous variants were initially investigated, and subsequently compound heterozygote and X chromosome variants were also examined (Supplementary Methods S1). Candidate variants for both brothers were examined to detect the variant(s) that both brothers shared (Supplementary Table S2). Variants that cosegregated in both brothers were confirmed by Sanger sequencing (primer sequences available on request). Subsequently a literature search on the relevant variants was performed. Variants have been submitted to the Leiden Open Variation Database (LOVD) at http://databases.lovd.nl/shared/genes/NEK1 (LOVD DB-IDs NEK1_000002 and NEK1_000003).

Fibroblast RNA isolation and cDNA conversion

Patient and control fibroblasts were cultured with Dulbecco's Modified Eagle's Medium (DMEM) containing 10% Fetal Calf Serum (FCS) in standard cell culture conditions. Cells were washed with Phosphate Buffered Saline (PBS), harvested, and RNA extracted (RNeasy Mini, Qiagen, Germany). RNA was converted to cDNA using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA).

Variant screening of *NEK1* in an OFD type II cohort

To evaluate if *NEK1* variants were causal for OFD type II, *NEK1* was screened in a cohort of four patients suspected of having OFD and showing overlapping features with the presently described brothers (Supplementary Table S3). The full coding regions and exon-intron boundaries were analyzed by Sanger sequencing. Primer sequences are listed in Supplementary Methods S4.

Immunocytochemistry

To assess primary cilia length and morphology, control- and patient-derived fibroblasts were cultured on glass slides under standard cell culture conditions in DMEM containing 10% FCS. The formation of primary cilia was stimulated using serum-depleted medium (DMEM containing 0% FCS) for 24 hours prior to immunostaining. Cells were put on ice for 5 minutes before fixation with 4% paraformaldehyde (PFA) in PBS, and permeabilization with 0.1% Triton-X-100 in PBS. Cells were blocked with 3% bovine serum albumin (BSA) in PBS minimizing non-specific binding, followed by a 2.5-hour incubation with primary antibodies and 1-hour incubation with secondary antibodies at room temperature. For a full list of antibodies used, see Supplementary Methods S5. Coverslips were washed with PBS and mounted in Fluormount G (Cell Lab, Beckman Coulter, Brea, CA, US). Confocal imaging was performed using Zeiss Confocal laser microscope LSM700 (Carl Zeiss AG,

Oberkochen, Germany) and images were processed with ZEN 2011 software (Carl Zeiss AG, Oberkochen, Germany) and Volocity 6.3 software (PerkinElmer, Coventry, UK).

RESULTS

Patient description

The brothers showed similar features, summarized in Table 1 and Figure 1(A-H). They are the only children of healthy, consanguineous parents (fourth-degree relatives), as is depicted in Figure 1(I). The elder brother (LOVD individual id 00064295) was born after a gestation of 39 weeks with a birth weight of 3540 gram (0.03 SD), a birth length of 49 cm (-1.44 SD) and head circumference of 37.5 cm (+1.37 SD). The CT scan of the brain shortly after birth did not show any abnormalities. Histopathological examination of the tongue hamartomas showed polyp-like non-keratinizing stratified squamous epithelium. The underlying stroma contained mucoserous acini, fat and muscle cells and the tissue

Table 1 | Clinical phenotype of both brothers presenting with OFD type II, Mohr syndrome.

	Elder brother	Younger brother
Oral anomalies	Incomplete midline lip and alveolar cleft Submucous cleft palate Hyperplastic frenula Tongue hamartomas	Incomplete midline lip and alveolar cleft Submucous cleft palate Bifid tongue with hyperplastic frenula Tongue hamartomas
	General dental hypoplasia: Dental agenesis of the 3.1, 4.1	General dental hypoplasia: Dental agenesis of the 1.2, 2.2, and 2 lower incisors (3.1 and 4.1)
	Cupular shaped upper incisors	Central incisors and cuspids with talon cusps.
	Taurodontia with fused/conical short roots of the molars	Taurodontia with fused/conical short roots of molars
Facial anomalies	Maxillary hypoplasia	Maxillary hypoplasia
	Protruding ears, left more pronounced than right	
Limb anomalies	Brachydactyly, clinodactyly digiti V	Brachydactyly, syndactyly of fingers
	Bifid right hallux, broad left hallux	Bilateral broad hallux
	Mild mesomelic limb shortening in lower limbs	Mild mesomelic limb shortening in lower limbs
Other features	Progressive, mainly conductive hearing loss, right side more affected than left	Mild conductive hearing loss (15-20dB)
	Bilateral tortuosity of the retinal veins	Bilateral tortuosity of the retinal veins

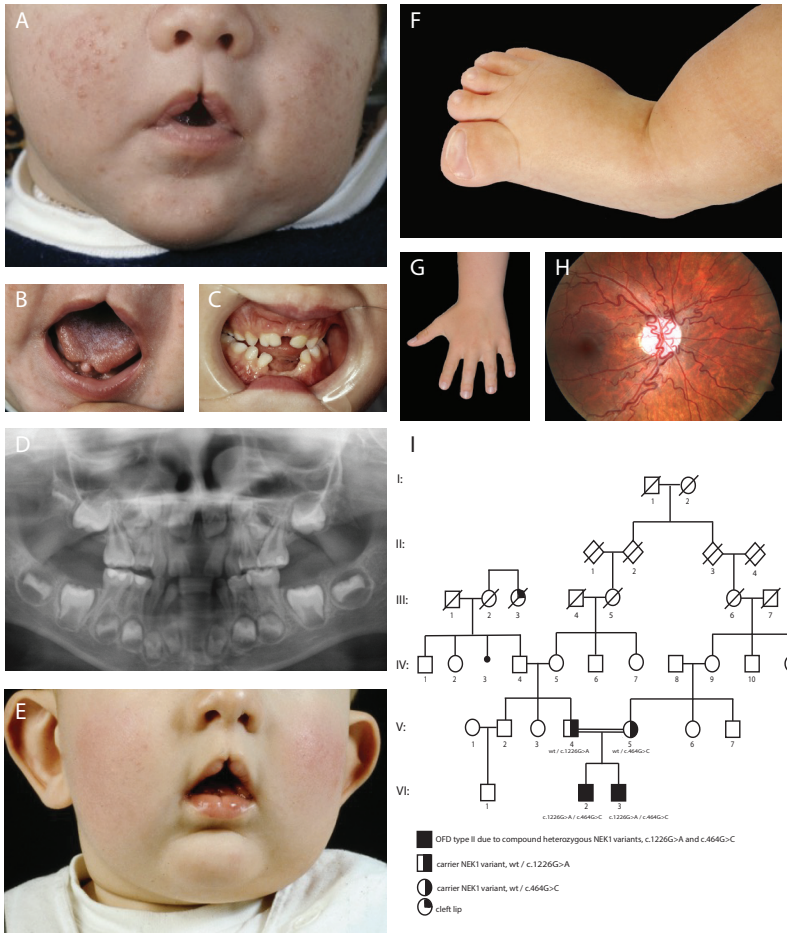


Figure 1 | Clinical features of both brothers* and pedigree of the family. A) Partial midline cheilognathoschisis present in the younger brother at the age of 1 month. B) Bilobulated tongue with tongue hamartoma and short frenulum of the younger brother at the age of 1 month. C) The intraoral situation of the younger brother (age 3.5 years old), shows a fusion of the two deciduous incisors 5.1 and 5.2. Also, the deciduous upper left lateral incisor (6.2) is congenitally missing. Furthermore, the typical hyperplastic frenula can be seen in the upper jaw. D) The orthopantomogram of the younger brother at the age of 5, shows that the permanent and deciduous central incisors of the lower jaw (3.1, 4.1, 7.1 and 8.1) are missing. The molar roots are conical and seem to be fused, while the pulp chambers are taurodontic. E) Incomplete midline cheilognathoschisis and mild protruding ears of the elder brother at the age of 8 months. F) Broad right hallux of the elder brother at the age of 8 months. G) Brachydactyly with partial cutaneous syndactyly present in the younger brother. Picture was taken at the age of 9.5 years. H) Tortuosity of the retinal veins of younger brother at the age of 16 years. I) Pedigree of the family. Circles represent females and squares represent males. Solid symbols indicate compound heterozygosity for the c.[464G>C;1226G>A] variants and half solid symbols indicate heterozygosity. Carriership was confirmed in both parents. * We obtained informed consent and approval of both parents and patients to include the present pictures.

The younger brother (LOVD individual id 00064296) was born after a gestation of 39 weeks with a birth weight of 3390 gram (-0.37 SD), a birth length of 44 cm (-3.59 SD) and a head circumference of 37.5 (+1.37 SD). A CT scan of the brain that was performed shortly after birth did not show abnormalities. The tongue hamartomas were resected shortly after birth and histopathological examination showed non-keratinizing stratified squamous epithelium, with the underlying stroma containing a mucoserous acinus and fat cells. At the age of 9.5 years a mild bilateral conductive hearing loss of 15-20 dB was detected. Both brothers had a normal psychomotor development and normal intelligence. An ultrasound of the kidneys was performed at 20 years in the elder brother and at 16 years in the younger brother; these did not show abnormalities. The elder brother had normal serum creatinine level when he was tested at age 11 years and both brothers do not have polyuria. Ophthalmological examination was also performed at the age of 20 in the eldest brother and 16 in the younger. Both brothers had bilateral tortuosity of the retinal veins. There were no signs of retinopathy. X-rays of the thorax, humerus, femur, ankle, and feet were made at different ages and showed mild shortening of the tibia and broad halluces in both brothers, with a bony duplication of the right hallux in the elder patient. Several dental panoramic x-rays were performed and showed the anomalies described in Table 1. Both brothers had orthodontic treatment before a surgically assisted prosthodontic rehabilitation. Both brothers developed velopharyngeal insufficiency for which they underwent a pharyngoplasty.

Based on the clinical features, specifically the congenital malformations in combination with the recessive inheritance pattern and the absence of cerebral malformations, intellectual disability, and severe tibia defects, OFD type II (Mohr syndrome) was diagnosed by two clinical geneticists (MJB, PAT).

SNP array analysis

SNP array analysis showed a normal male karyotype in both brothers and several homozygous regions indicative of distant consanguinity of the parents.

WES identifies biallelic variants in *NEK1*

Subsequently, the WES data of the parents and the two affected brothers were investigated. The genes implicated in the various OFD types were initially analyzed (*OFD1, TCTN3, TMEM216, C5orf42, DDX59, TBC1D32/C6orf170, SCLT1, C2CD3*) but no variants that could be considered causal were detected. The consanguineous relationship prompted us to investigate the shared homozygous regions in both boys for candidate homozygous variants but no plausible candidate genes were identified. After expanding our search in the whole exome and filtering for autosomal recessive and X-linked inheritance models,

several gene variants were identified in candidate genes (Supplementary Table S2). Based on population frequency, predicted effect of the variant on gene function and previous literature, variants in *NEK1* were the most likely candidates. Both brothers share the same novel compound heterozygous variants in *NEK1*, comprising an exonic transversion NM_012224.2:c.464G>C at the last base of exon 6 inherited from the mother and a transition variant NM_012224.2:c.1226G>A inherited from the father, c.[464G>C;1226G>A] (Figure 2; exons are numbered similarly as in Thiel et al. 2011) (Thiel et al. 2011).

Variants were validated and segregation was confirmed by Sanger sequencing. Both variants were novel and not present in the population frequency reference datasets of NCBI dbSNP Build 137 for Human, the Exome Variant Server (EVS), the 1000Genomes, the Genome of the Netherlands (GoNL) or our in-house dataset. The variant c.1226G>A results in a nonsense variant p.(Trp409*) in exon 15 and the resulting shortened mRNA transcript includes the full coding sequence for a functional kinase domain but lacks part of the coiled-coil domain. The effect of the exonic variant c.464G>C appeared to be a nonsynonymous coding change from serine to threonine, but we postulated that alternative splicing may also occur due to the position of this base so near to the canonical splice site. This position is conserved and classified as disease-causing by Alamut Visual (Alamut Interactive Biosoftware, Rouen, France) and predicted to result in abnormal splicing. Further Sanger sequencing of *NEK1* in four other patients with OFD type II did not reveal any additional variants in *NEK1* that could be considered causal.

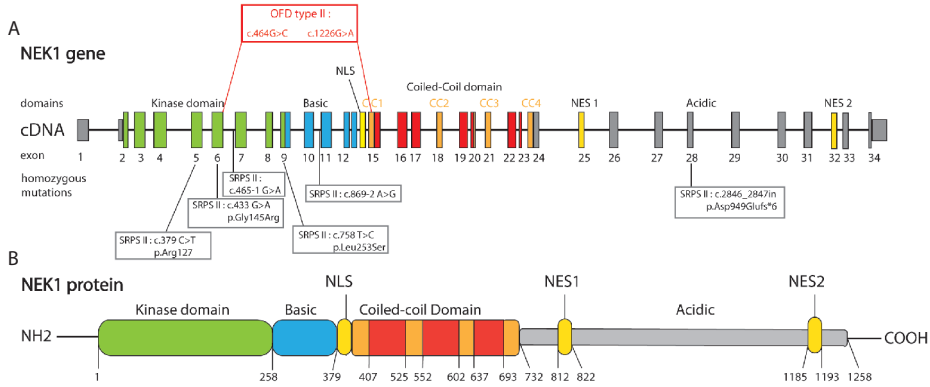
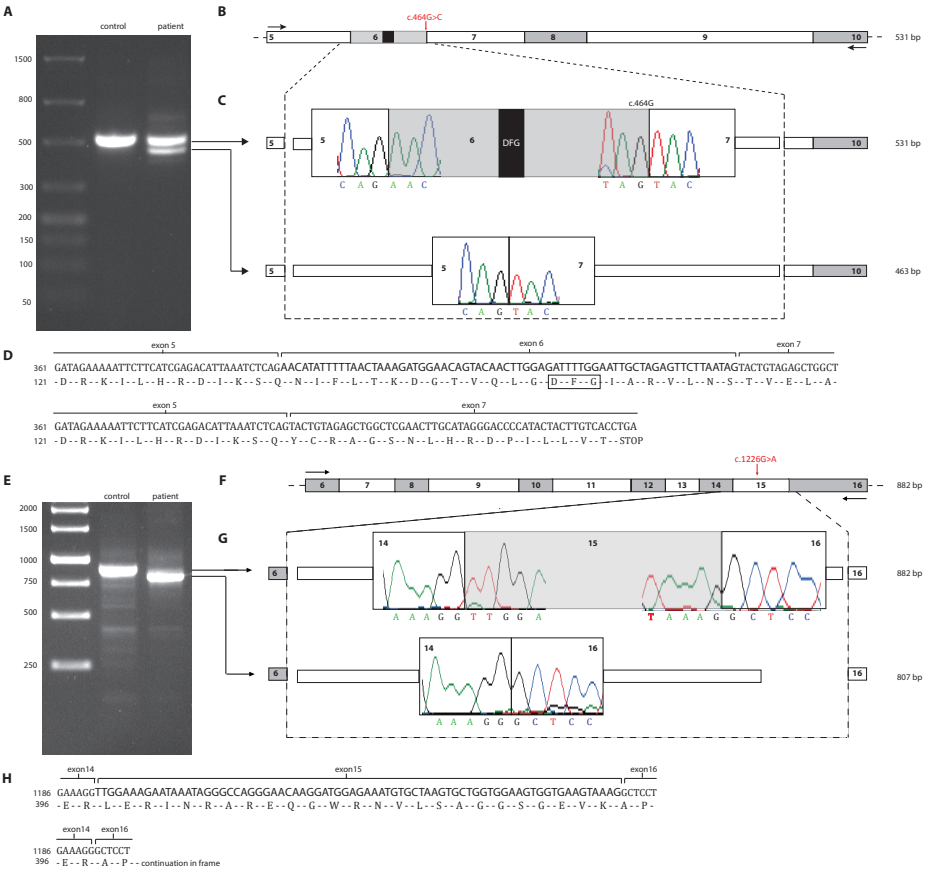


Figure 2 | *NEK1* genomic and protein structure. A) *NEK1* genomic structure including the kinase domain, basic domain, coiled-coil domains (cc1-cc4), nuclear localization signal (NLS) and nuclear exportation signal (NES1-2). Variants found in the patients presented in this study with OFD type II are indicated above, namely c.464G>C in the kinase domain (exon 6) and c.1226G>A in the coiled-coil domain 1 (CCD1, exon 15). NCBI reference transcript NM_012224.2 is used. B) NEK1 protein structure including the kinase, basic and coiled-coil domains, the NLS and NES.

cDNA sequencing confirms that both variants result in alternative splicing

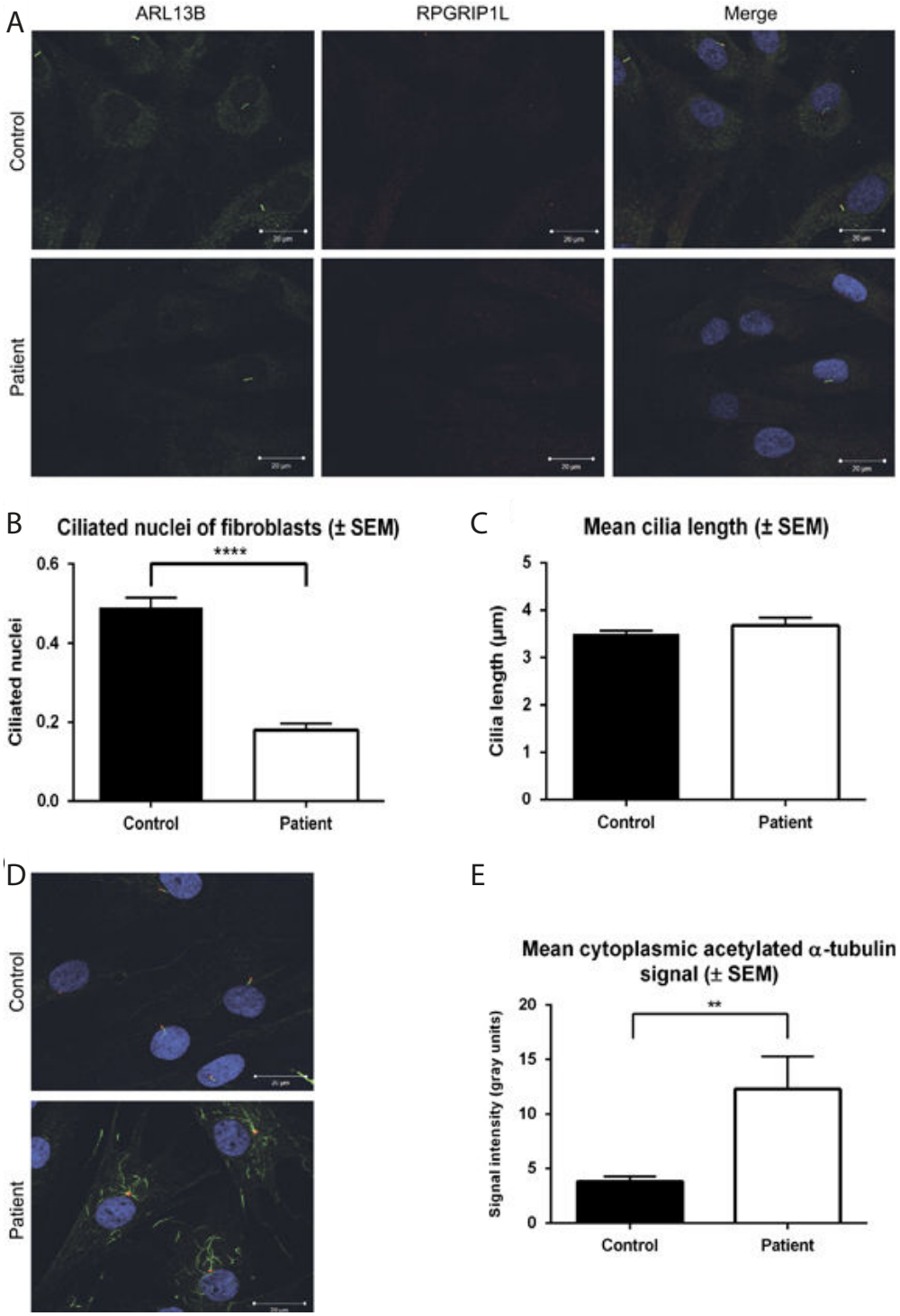
To assess the splicing impact of the variant c.464G>C, we designed primers covering several exons upstream and downstream of the variant in the cDNA of the elder brother (Supplementary Methods S6). Analysis of the resulting PCR products on 1% agarose gel identified a smaller product in addition to the expected product size, indicating an alternatively spliced product. Subsequent Sanger sequencing of the smaller product confirmed that the variant c.464G>C resulted in alternative splicing and exon skipping of exon 6 (Figure 3, A-D). The removal of exon 6 results in shift of the reading frame and introduces an early stop codon. The resulting transcript, r.397_464del, lacking exon 6, most likely undergoes nonsense-mediated decay due to the position of the premature stop codon. We observed that the correctly spliced product including exon 6 has no visible alternative allele C at position 464, demonstrating that this variant has no or minimal effect on coding changes from serine to threonine at this position.



Additionally, any transcripts escaping nonsense-mediated decay would lack exon 6 and the critical DFG motif (comprising the amino acids Asp-Phe-Gly) essential for the correct conformation of the ATP-binding site of the kinase domain NEK1 (Figure 3C) (Ung and Schlessinger 2015).

To assess the effect of nonsense variant c.1226G>A in the same cDNA, we designed primers covering several exons upstream and downstream of exon 15 (Supplementary Methods S7). We selected the allele containing the c.1226G>A variant using a primer in exon 6, which was confirmed by the absence of the c.464G>C variant. Sanger sequencing of this region showed that exon 15 is skipped in frame in nearly all transcripts (Figure 3, E-H). The resulting transcripts, r.1192_1266del, are near-full length. The loss of exon 15 removes the first coiled-coil domain of NEK1. The region containing coiled-coil domains 1 and 2 has been shown to interact strongly with ATRX and weakly with MRE11 and 53BP1 that are involved in dsDNA break repair and homologous recombination (Surpili et al. 2003). The coiled-coil domains 1 and 2 are also required for localization to the primary cilium, although the ciliary signal is strongest in transcripts containing coiled-coil domain 2 (White and Quarmby 2008).

Figure 3 | *NEK1* cDNA analysis of variant c.464G>C (A-D) and nonsense variant c.1226G>A, p.(Trp409*) (E-H). A) Primers flanking the patient variant c.464G>C were used for PCR on patient and control cDNA isolated from fibroblast mRNA. Following PCR amplification, the patient cDNA contained an extra product at 463 bp, indicating differential splicing. B) Visualization of the cDNA region amplified by PCR. Locations of the forward and reverse PCR primers are indicated with arrows. The position in exon 6 of the coding region for the DFG motif is indicated by a black box. The variant is located on the last base of exon 6. The expected full-length PCR product is 531 bp. C) Zoom in view of the splicing effect of c.464G>C. The wild-type allele G at position 464 of the cDNA (top) results in normal splicing, retaining exon 6, as seen in control cDNA and one allele of the patient cDNA. Sanger sequencing of splice junctions are shown. The product due to alternative splicing caused by variant c.464G>C (bottom) lacks exon 6, including the coding region for the DFG motif, and as a result is only 463 bp. D) Alternative splicing results in a frameshift and introduction of a premature stop codon. The wild-type allele G (top) results in correct splicing and translation. The splicing variant C (bottom) results in a frameshift due to the loss of exon 6 and introduces a premature stop codon 18 amino acids after the frameshift, presumably resulting in nonsense-mediated decay. E) Primers flanking the variant c.1226G>A were used for PCR on cDNA from mRNA isolated from patient and control-derived fibroblasts. Analysis of the PCR product on 1% Agarose gel showed that the patient cDNA transcript was 75 bp shorter than the control cDNA transcript (882 bp). F) Representation of the amplified cDNA region. Location of the forward and reverse PCR primers are indicated by arrows. The expected full-length PCR product is 882 bp. G) Detailed view of Sanger sequences surrounding the splice junctions caused by the nonsense variant in exon 15. Nonsense-associated alternative splicing causes skipping of exon 15 (75 bp) in the patient. H) Skipping of mutated exon 15 in the patient results in a continuation in the same reading frame. The nucleotide sequence containing the wild-type allele G results in correct splicing and translation (top). The nonsense variant c.1226G>A removes exon 15 (75 bp) from the transcript which continues in frame (bottom) resulting in a near-full length transcript.


A
ARL13B
RPGRIP1L
Merge
Control
Patient
B
Ciliated nuclei of fibroblasts (± SEM)
****
Ciliated nuclei
0.6
0.4
0.2
0.0
Control
Patient
C
Mean cilia length (± SEM)
Cilia length (µm)
5
4
3
2
1
0
Control
Patient
D
Control
Patient
E
Mean cytoplasmic acetylated α-tubulin signal (± SEM)
**
Signal intensity (gray units)
20
15
10
5
0
Control
Patient

Cilia number is reduced in patient-derived fibroblasts compared to controls

To determine the cellular effects of *NEK1* variants in the patients, we examined cilia number and length in primary skin fibroblasts from the elder brother and a healthy control using anti-acetylated α-tubulin and anti-ARL13B antibodies to mark the ciliary axoneme and anti-RPGRIP1L antibody to mark the basal body at the base of the cilium (Figure 4). Patient-derived fibroblasts showed a statistically significant reduction in cilia number compared to control fibroblasts (18.0% vs. 48.9%, $p<0.0001$). In addition, patient-derived fibroblasts showed a marked increase in cytoplasmic acetylated α-tubulin compared to control fibroblasts (12.3 vs. 3.8 gray units, $p=0.002$). There was no significant difference in mean cilia length (3.7 vs. 3.5 µm, $p=0.09$). Reconstitution of the cilia phenotype by transfecting fibroblasts with a full-length wild-type *NEK1* construct was attempted; however, robust results were not obtainable (data not shown).

◂ **Figure 4 | Patient-derived fibroblasts show a severe ciliation defect.** A) Immunofluorescence of ciliary axoneme (ARL13B, green), basal body (RPGRIP1L, red) and nucleus (Hoechst, blue) shows a severe reduction in cilia number but not in cilia length in patient- compared to control-derived fibroblasts. Scale bars represent 20 µm. B) An average of 48.9% of control-derived fibroblasts are ciliated compared to 18.0% of patient-derived fibroblasts (>800 cells analyzed). This difference is statistically significant (**** indicates two-tailed P value <0.0001, calculated using Mann-Whitney test). C) Mean cilia length in control-derived fibroblasts was 3.5 ± 0.1 µm and in patient-derived fibroblasts 3.7 ± 0.2 µm (± 200 and 50 cilia analyzed for control and patient, respectively). This difference is not statistically significant (two-tailed P value equals 0.09, calculated using Mann-Whitney test). D) Immunofluorescence of acetylated α-tubulin (green), pericentrin (red) and Hoechst (blue) shows a marked increase in cytoplasmic acetylated α-tubulin in patient- compared to control-derived fibroblasts. Scale bars represent 20 µm. E) Comparison of the intensity of cytoplasmic acetylated α-tubulin signal in control- and patient-derived fibroblasts (3.8 vs. 12.3 gray units, ** indicates P=0.002, calculated using Mann-Whitney test).

DISCUSSION

In this study, we identified compound heterozygous variants in *NEK1* in two brothers with OFD syndrome type II, Mohr syndrome. Homozygous variants in *NEK1* are associated with short-rib polydactyly syndrome (SRPS) Majewski type (OMIM 263520) (El Hokayem et al. 2012; Thiel et al. 2011), which in the past has been considered allelic to OFD (Mohr syndrome) (Franceschini et al. 1995; Silengo et al. 1987). Our results confirm that these syndromes may be part of the same phenotypic spectrum.

NEK1 is a known ciliary gene comprising a kinase domain in addition to a basic and coiled-coil domain for localization and protein interaction (Figure 2) (Meirelles et al. 2014; Shalom et al. 2008; White and Quarmby 2008). NEK1 kinase is a member of the NIMA-related kinase family and is localized at the cilium's basal body (Shalom et al. 2008). This kinase family is considered to play a role in cell cycle regulation and DNA damage repair, with some kinases having an additional role in the formation of the primary cilium (Fry et al. 2012; Shalom et al. 2008; White and Quarmby 2008).

Immunocytochemistry in patient-derived fibroblasts demonstrated severely reduced ciliation and an increase in cytoplasmic acetylated α-tubulin signal compared to control-derived fibroblasts. Our findings are consistent with the reduced number of normal cilia and the expression of long, thick microtubular extensions marked by acetylated α-tubulin in fibroblasts of a fetus with SRPS Majewski type resulting from *NEK1* variants (Thiel et al. 2011). This increase in cytoplasmic acetylated α-tubulin may be due to the inability of variant NEK1 to phosphorylate VHL, which stabilizes microtubules (Patil et al. 2013). A reduction in ciliated cells was also noted in embryonic fibroblasts of *kat2J* mice homozygous for a frameshift variant in the kinase domain (Shalom et al. 2008). Contrary to previous reports, we did not detect a difference in mean cilia length (Thiel et al. 2011). This could be related to using the anti-ARL13B antibody to visualize the ciliary axoneme instead of the anti-acetylated α-tubulin antibody, as increased microtubule staining prevented reliable measurements of cilia.

Decreased ciliation could negatively affect adequate signal transduction during morphogenesis and craniofacial development since cilia are involved in multiple signaling pathways such as Hedgehog and Wingless (Wnt) signaling (Satir et al. 2010; Veland et al. 2009; Zaghloul and Brugmann 2011). For example, variants in the NEK1 interaction partner KIF3A, a component of the intraflagellar transport system, lead to cilia defects resulting in an inappropriate increase in canonical Wnt signaling (Corbit et al. 2008). Although homozygous *Kif3a* null variants are lethal in mice (Marszalek et al. 1999), mice with conditional *Kif3a* deletions in cranial neural crest cells are characterized by impaired

skeletogenesis and proliferation defects in the facial midline, similar to the anomalies detected in our patients (Liu et al. 2014). Additionally, inactivation of *OFD1* is also associated with defective ciliation and subsequent disturbed Hedgehog and canonical Wingless signaling, showing overlap in the pathogenic disease mechanism (Bimonte et al. 2011; Ferrante et al. 2009; Ferrante et al. 2006; Romio et al. 2004). This might therefore explain the considerable overlap in phenotype between patients with OFD type I and our patients (Thauvin-Robinet et al. 2006).

In addition to localizing to the basal body of the cilium, NEK1 resides in the cilium, the cytoplasm and the nucleus (Hilton et al. 2009; Mahjoub et al. 2005; Shalom et al. 2008). There is a canonical nuclear localization signal in residues 355-378 within the coiled-coil domain of NEK1, and a cryptic nuclear localization signal within residues 258-321 of the basic domain (Figure 2) (Hilton et al. 2009). The coiled-coils 1 and 2 are essential for localization to the primary cilium (White and Quarmby 2008). However, in a study investigating the role of Nek1 in murine IMCD3 cells, a tagged-Nek1 construct was only retained in the ciliary compartment long enough to be visualized if the kinase domain (which normally targets the protein to the cytoplasm) was non-functional, and endogenous mNek1 could not be visualized in the primary cilium (White and Quarmby 2008). Cycling of NEK1 through the ciliary, cytoplasmic and nuclear compartments can partly explain why reconstitution assays yielded no robust results (White and Quarmby 2008). Transient expression of NEK1 at the basal body prevented assessment of transfected cells only, and therefore the rescue of the cilia phenotype could not be quantified. Moreover, rescue of the cilia phenotype was limited because NEK1 overexpression also inhibits ciliogenesis (Shalom et al. 2008; White and Quarmby 2008). Similar to other ciliary proteins with a dual role in the nucleus, NEK1 is involved in DNA damage response and cell-cycle control (Chaki et al. 2012; Chen et al. 2011). Expression and kinase activity of NEK1 are upregulated in irradiated cells and NEK1-null cells are hypersensitive to DNA damage by ionizing radiation (Chen et al. 2011; Polci et al. 2004). A portion of NEK1 can redistribute to form discrete nuclear foci at sites of DNA damage (Chen et al. 2011; Polci et al. 2004). The mechanism for this DNA damage response has not yet been fully established, but OFD1 has been linked to DNA repair through its interaction with the chromatin-remodeling complex TIP60 (Giorgio et al. 2007). NEK1-deficient primary renal tubular epithelial (RTE) cells of *kat2J* mice showed increased chromosome missegregation events and aneuploidy after mitosis (Chen et al. 2011). The DNA damage response of these ciliary proteins likely also has a role during morphogenesis when progenitor cells are under increased replication stress. Impairment of this response would therefore possibly result in tissue dysplasia (Chaki et al. 2012). Secondly, the impaired DNA damage response can underlie degenerative instead of developmental phenotypes in ciliopathies, e.g. nephronophthisis (Chaki et al. 2012; Slaats et al. 2015). Renal cysts were present in five out of eight reported lethal

NEK1-related SRPS type Majewski patients (El Hokayem et al. 2012; Thiel et al. 2011) and *NEK1* deficient *kat* and *kat2J* mice display slowly progressive polycystic kidney disease (Upadhya et al. 2000; Upadhya et al. 1999). Although our patients had a normal renal ultrasound at age 20 and 16 years respectively, we cannot exclude that a renal phenotype may develop at an older age.

The phenotype in our patients is mild compared to the previously described lethal cases with *NEK1* variants. This could reflect the extreme clinical heterogeneity associated with variants in ciliary genes; for example, in the ciliary gene *MKS1* biallelic null variants cause the embryonically lethal Meckel-Gruber syndrome while missense variants cause Joubert syndrome (Romani et al. 2014). In *NEK1*, missense, splice-site, nonsense and frameshift variants have been reported in the patients with SRPS (Supplementary Table S8) (Chen et al. 2012a; El Hokayem et al. 2012; Thiel et al. 2011).

Most variants were identified in the kinase domain of NEK1. In addition, one patient showed a homozygous frameshift variant in the C-terminal domain of *NEK*1 and in another patient digenic inheritance with *DYNC2H1* plays a role (Figure 2) (Chen et al. 2012a; El Hokayem et al. 2012; Thiel et al. 2011). The splice variant c.464G>C identified in the present study is predicted to result in nonsense mediated decay of the mRNA transcript due to the premature stop codon and thus no protein translation. The nonsense variant p.(Trp409*) removes the first coiled-coil domain. A possible explanation for the milder phenotype in our patients compared to previously published patients is basal exon skipping of exon 15 that contains the premature stop codon. Basal exon skipping is a mechanism whereby detrimentally mutated exons that begin and end in the same reading frame can be selectively spliced out of the final gene transcript (Drivas et al. 2015). This mechanism has been described in other ciliopathies (Drivas et al. 2015; Littink et al. 2010). Nonsense-associated alternative splicing results in a detectable level of near-full length *NEK1* transcript that contains the complete kinase domain and is expected to escape nonsense-mediated decay. The resulting protein likely has some residual kinase function. As the N-terminal kinase domain is required for cilia formation (Thiel et al. 2011; White and Quarmby 2008), the remaining 18% ciliation in patient-derived fibroblasts could suggest that some functioning kinase domain is (partly) preserved. Moreover, variants in modifier genes could play a role in the milder phenotype seen in our patients.

Similar to other ciliopathies, OFD type II is probably a genetically heterogeneous condition. Further studies in a cohort of patients with autosomal recessive types of OFD for variants in *NEK1* and genes coding for proteins interacting with NEK1, such as *KIF3A*, might reveal more insight into the underlying pathogenic mechanisms (NEK1 interacting proteins reviewed in (Surpili et al. 2003)). The fact that we found compound heterozygous variants

in a gene outside homozygous regions illustrates that consanguinity can be misleading and that WES or whole-genome sequencing can provide more information in cases in which homozygosity mapping does not lead to the causative variants.

In this study, we provide the first genetic and functional evidence that variants in *NEK1* can be involved in the etiology of OFD type II. Similar to the cellular phenotype seen in SRPS type Majewski, fibroblasts derived from our patient with OFD type II show a severe defect in cilia formation and an increase in cytoplasmic acetylated α-tubulin. Analysis of *NEK1* should be considered in patients with autosomal recessive OFD.

Acknowledgements

We would like to thank the patients and their parents for their participation in this study. We wish to thank the Utrecht Sequencing Facility, UMC Utrecht, for WES using the Illumina HiSeq 2500 and further data processing with the bioinformatics pipeline. We thank the Cell Microscopy Core at the Department of Cell Biology, Center for Molecular Medicine, UMC Utrecht for use of the Zeiss Confocal laser microscope LSM700 and Corlinda ten Brink for help with the Volocity 6.3 software. We thank Dr. R. Roepman and Dr. H. Arts for providing us with the anti-RPGRIP1L antibody and J. Hoogeboom (Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands) for providing us with clinical data on patients in the OFD type II cohort. This work was supported by the Dutch Kidney Foundation (KOUNCIL consortium: CP11.18 to NK and RG). The authors declare no conflict of interest.

Supplementary Methods S1: Whole-exome sequencing.

Parent-patients quatro WES was performed with an emphasis on autosomal recessive variant detection analysis for the causal variants. DNA libraries of the two patients and their parents were prepared using Kapa Biosystems reagents and enriched using Agilent Sureselect All exon V5 and a custom pooling protocol. Briefly, 1 ug of purified gDNA (Qiagen, Hilden, Germany) was sheared into 100-500 bp fragments with a Covaris S2 sonicator (Covaris, Woburn, MA, USA), blunt-ended, 5'phosphorylated, and A-tailed using Kapa Biosystems reagents (Kapa Biosystems, Wilmington, MA, USA). Adaptors containing the Illumina barcode sequences were ligated to each sample and amplified with 7 PCR cycles. Samples were finally quantified and were equimolarly combined into one pooled batch containing three samples. The fourth member of the quarto was enriched with two unrelated samples for a different project. Barcode blockers for the Illumina adaptor sequences and barcodes were added (sequences available upon request), and the pools were enriched according to the Agilent Sureselect V5 exome protocol (Agilent, Santa Clara, CA, USA), and finally PCR amplified 8 times. Two pools of three samples were then combined into a batch of six samples together and sequenced on a full flow cell as a Rapid Run on the Illumina HiSeq 2500 at the Utrecht Sequencing Facility (Utrecht, Netherlands). Sequencing reads were aligned using BWA and data was processed with GATK v3.1.1, according to best practice guidelines but skipping BQSR and applying the hard filters instead of VQSR (McKenna et al. 2010; Van der Auwera et al. 2013). Samples had mean bait coverage of 104X and 93% of the target bases were over 20X. A phased VCF was created and imported into Cartagenia BENCHlab NGS module (Cartagenia, Leuven, Belgium) for variant interpretation. Homozygous variants, compound heterozygous variants, and X chromosome variants were investigated. Variants were filtered out by presence above the minor allele frequency greater than 1.5% in the public population frequency databases of the NCBI dbSNP Build 137 for Human (https://www.ncbi.nlm.nih.gov/SNP), the Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS), the 1000 Genomes Project (http://www.1000genomes.org), the Genome of the Netherlands (GoNL) (http://www.nlgenome.nl), or our in-house dataset, as well as in the clinical variant database Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php). Variants were further prioritized by amino acid change or splicing effect, conservation score (GERP), and predicted effect on protein function using prediction algorithms (PolyPhen-2, SIFT) (Adzhubei et al. 2010; Cooper et al. 2005; Kumar et al. 2009).

Supplementary Table S2 | Homozygous and Compound Heterozygous Candidate Variants identified from Whole-exome sequencing.

Gene	*NPRL3*	*PRR35*	*PRR35*	*TPTE*	*TPTE*
Zygosity	Homozygous	Heterozygous	Heterozygous	Heterozygous	Heterozygous
Chr:Pos	16:138725	16:613344	16:615095	21:10941966	21:10944668
dbSNP	rs202015937	rs116903998	rs74459225	rs149843545	rs138088406
Ref	T	G	G	A	C
Alt	A	A	A	G	T
HGVS nucleotide	NM_001077350.2: c.1516A>T	NM_145270.2: c.50G>A	NM_145270.2: c.1504G>A	NM_199259.3: c.683T>C	NM_199259.3: c.512G>A
HGVS amino acid	p.(Asn506Tyr)	p.(Arg17Gln)	p.(Gly502Ser)	p.(Ile228Thr)	p.(Arg171Lys)
Coding effect	missense	missense	missense	missense	missense
MAF (ExAc)	0.0007	0.0038	0.0403	0.0050	0.0023
Sift	deleterious	deleterious	tolerated	deleterious	deleterious
PolyPhen	probably damaging	probably damaging	benign	probably damaging	probably damaging
Gene function	Nitrogen permease regulator-like 3. Component of the GATOR1 complex, an inhibitor of mTORC1 signaling. GTPase activator. Required for normal cardiovascular development. Highly conserved.	Proline Rich 35; very little known.	Proline Rich 35; very little known.	Transmembrane Phosphatase With Tensin Homology; signal transduction pathways of endocrine of spermatogenic function of testis.	Transmembrane Phosphatase With Tensin Homology; signal transduction pathways of endocrine of spermatogenic function of testis.
Interpretation	Known gene function cannot be linked to phenotype	rs74459225 not predicted to be deleterious; gene variants are probably not causal.	rs74459225 not predicted to be deleterious; gene variants are probably not causal.	Expression solely in the testes; not likely candidate	Expression solely in the testes; not likely candidate

Supplementary Table S3 | Clinical phenotype and genetic tests OFD type II cohort.

	Patient 1	Patient 2	Patient 3	Patient 4
Age	17	17	28	28
Oral anomalies	Frenulae from lip to gingiva Minimal median notch upper lip Notches gingiva Supernumerary incisor	Notches gingiva Lobulated tongue	Hyperplastic frenula Cleft tongue	Cleft lip, palate and uvula Palatal fibromas Ectopic tongue tissue Ankyloglossia
Facial anomalies		Broad nasal bridge Milae	Micrognathia	
Limb anomalies	Abnormal proximal phalanx and broad distal phalanx digitus 1 right hand Short middle phalanx digitus 5 right hand	Brachydactyly Syndactyly digiti 4 and 5 right hand Partial syndactyly digiti 2 and 3 left hand Broad digitus 1 left hand Bilateral syndactyly 2ND and 3RD toe	Unilateral preaxial polydactyly Bifid toes	Tarsal coalition left foot Pes planovalgus
Other features		Developmental delay (IQ 49) Left frontal interhemispheral cyst Corpus callosum agenesis	Developmental delay (IQ 74) Epilepsy Coarctatio aortae Bicuspid aortic valve Genital anomaly	Developmental delay Speech delay Hearing loss
Genetic findings	Normal female karyotype	Analysis of *OFD1*, *TCTN3* and other ciliopathy-associated genes[1] was normal	No information available	No information available

[1]Investigated ciliopathy-associated genes are: *AHI1, C5orf42, CC2D2A, CEP290, CEP41, INPP5E, KIF7, MKS1, NPHP1, RPGRIP1L, TCTN1, TMEM138, TMEM216, TMEM231, TMEM237, TMEM67, TTC21B* and *ZNF423.*

Supplementary Methods S4 | Primers used for NEK1 variant screening

	Forward	Reverse
Amplicon 1	5' GTTTCCTATGCTTTGGTTGG 3'	5' TTTACAGTGATCACCCATGC 3'
Amplicon 2	5' AAATTAAAAATATGGACTTAAAGGAC 3'	5' GTGAATGTGAATGCGATAGTG 3'
Amplicon 3	5' TCTTCATCTTCATGAATAGCC 3'	5' AGCAAAACATGGAGGTGTG 3'
Amplicon 4	5' AAAGGGGAGAGACCCAGAG 3'	5' TGTATTCATCCTCAATTTACCG 3'
Amplicon 5	5' TCCAGCTAGGTGACAGAGTG 3'	5' AGACTGAGCTGAAATTCAAGAG 3'
Amplicon 6	5' TTGTTGGATGTGTGTTTGTG 3'	5' TTATTCAGAAGGTAATATACTGACAC 3'
Amplicon 7	5' AAACTTTCACTGCCATGTTG 3'	5' ATTCACAGGTCAAGGGTGAG 3'
Amplicon 8	5' GGTGTGAGCCACTGTGC 3'	5' TTGAAGAGAAAATTTTACATAATGG 3'
Amplicon 9	5' TTTTGGGGTGACTTGTTAGG 3'	5' AAACCATTTTGACCACCATAG 3'
Amplicon 10	5' CCATCCAGGTCTATCAGGTC 3'	5' GCATTTCAGATAAGAGATACTCAGTC 3'
Amplicon 11	5' AGGGAGGCACTTCTATTCG 3'	5' TCTCTTCCTCTGTATCTATCTGC 3'
Amplicon 12	5' TTTCATGCTCTTTGGCATC 3'	5' GAGTATATGAATTTTCCTCTAATTTC 3'
Amplicon 13	5' CTTATAAACAATGAGGCTCCAG 3'	5' GATTAACAAAATCTTACACTTGAATC 3'
Amplicon 14	5' CCCAAATAACTTCCACATTG 3'	5' GAAGGCAAATGAAGTCTGTG 3'
Amplicon 15	5' TGCCATGGTTTTATCTATTTTC 3'	5' ACTGATGACGACAGTGGAAG 3'
Amplicon 16	5' GAATTTCTCTGGAACAAATGG 3'	5' GAAGATTTTGTAGGTCTGTAGGG 3'
Amplicon 17	5' TCAAAAGAAAAGAAACTGAAATC 3'	5' GGGTTTAACCTCCAAAGCAG 3'
Amplicon 18	5' TCCCAATTAACCAAAGCAAC 3'	5' CAGATTCTTCCGTAATTAAGTGTATC 3'
Amplicon 19	5' AATCAGGCTGAACTGGAAAG 3'	5' GTGAGAGCTCTTTGTTTAATCC 3'
Amplicon 20	5' CTTGTTTAGACACCGCCTTC 3'	5' TGTTTGTCACATTATGCTTTATTATC 3'
Amplicon 21	5' GAGTATAGTCACATAACAGCAACAC 3'	5' AAGCAGCAAGAAAGAGAAGG 3'
Amplicon 22	5' TGCCTACCTTTACTTCACCAC 3'	5' CATCTGTCTGGTTATTTTAAAGC 3'
Amplicon 23	5' GGGAAAAGAAATAAACAAAACAC 3'	5' TTGGGATTTTCAAATAAGTGC 3'
Amplicon 24	5' TTCTCTTTCTTGCTGCTTCC 3'	5' GACCCTAAATTTTTAATCTTGTTTG 3'
Amplicon 25	5' TGACTCCTGTAAGGCTTCTC 3'	5' GATACGTGCAACCATCACTC 3'
Amplicon 26	5' CAGCTAGGCATTTGATAAAGAC 3'	5' GAGGGAGTAAGCATGAGAGG 3'
Amplicon 27	5' AGACCATGGAAAGGAAAGTG 3'	5' CGTTTATTTAATACATCCCATC 3'
Amplicon 28	5' TCTGTCATTAGAAATAATTTTCTGG 3'	5' AGCTGTTGCAAATTTCTTAATAC 3'
Amplicon 29	5' TCTTTCTGCTGTTCTTCCAG 3'	5' TGGAAGTTGCTTTCTGTGAG 3'
Amplicon 30	5' ATACAGCCTGCCTGTGTAGC 3'	5' CTTGATGGTGTCCTTTGATG 3'
Amplicon 31	5' TCAGAGACTGGGAAGGAAAG 3'	5' CCCTATCCCTAAAGAAGTGG 3'
Amplicon 32	5' ATAAATGAGATGGGCAACTG 3'	5' TGAATGAAATCAAGTATGTCTGC 3'
Amplicon 33	5' GAGGCAGAGGTTGGAGTAAG 3'	5' CTCCTTTATTCCCGCTAATATC 3'
Amplicon 34	5' GATTAGGTTACAGAACAGCAAG 3'	5' TATGCAGTATTTGGCTTTCC 3'

Supplementary Methods S5: Immunocytochemistry and Statistical Analysis

Immunocytochemistry

The following primary antibodies were used to prepare coverslips for measurement of cilia number: anti-acetylated α-tubulin (mouse monoclonal, 1:10,000; Sigma T6793) and anti-ARL13B (rabbit polyclonal, 1:1,000; ProteinTech 17711-1-AP) (Figure S5.1). Anti-ARL13B (rabbit polyclonal, 1:1,000; ProteinTech 17711-1-AP) and anti-RPGRIP1L (guinea pig, 1:1,000; Eurogentec SNC039) were used to prepare coverslips to measure cilia length as hyper-stabilized acetylated α-tubulin prevented accurate measurement using anti-acetylated α-tubulin. Cells were washed with PBS prior to 1-hour incubation with the secondary antibody and Hoechst (1:5000; Invitrogen, H3570) at room temperature. The following secondary antibodies (Life Technologies, Forster City, CA, USA) were used: anti-mouse IgG Alexa Fluor 488 (1:500), anti-rabbit IgG Alexa Fluor 568 (1:500) and anti-guinea pig IgG Alexa Fluor 488 (1:500).

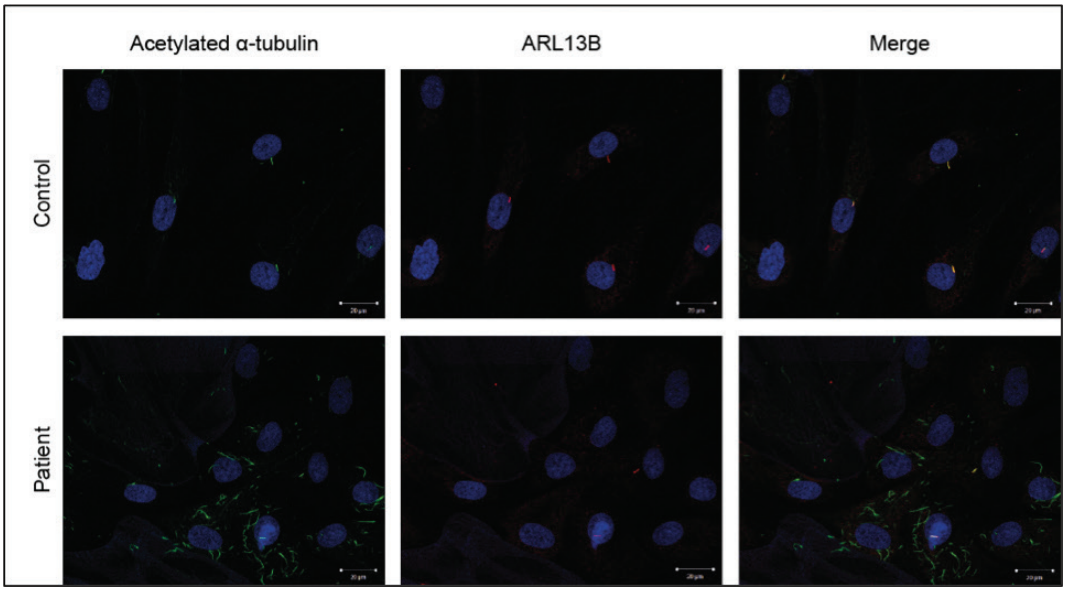


Figure S5.1 | Costaining of Acetylated α-tubulin and ARL13B. Immunofluorescence of acetylated α-tubulin (green), ARL13B (red) and Hoechst (blue). ARL13B signal corresponds to ciliary but not cytoskeletal acetylated α-tubulin. Scale bars represent 20 μm.

Statistical analysis

GraphPad prism 6.0 was used for cilia statistical analysis (©2015 GraphPad Software, La Jolla California USA). Equal nuclei density between patients and control cells was confirmed using the Mann-Whitney test. Outliers in nuclei density were removed using the modified Thompson tau test. To compare cilia number, approximately 800 nuclei were scored for patient- and control-derived fibroblasts in two technical replicates. GraphPad Prism 6.0 was used to perform Fisher's exact test and the Mann-Whitney test for non-parametric data.

Cilia length was measured using Image J. Approximately 50 cilia were measured in the patient-derived fibroblasts and 200 in the control. Outliers were removed using the ROUT method (Q=1%) in GraphPad Prism 6.0, and cilia length was analyzed using the Mann-Whitney test for non-parametric data.

Signal intensity cytoplasmic acetylated α-tubulin

To quantify the observed difference in cytoplasmic acetylated α-tubulin signal, ten 8-bit images of 10 x 10 µm (1024 x 1024 pixels) of the cytoplasm were taken using the same settings for control- and patient-derived fibroblasts. We measured signal intensity of acetylated α-tubulin for each image using Volocity 6.3 software. To control for acetylated α-tubulin background signal, we took ten 8-bit images of 10 x 10 µm of the intercellular space for each condition and subtracted the mean intensity of the background acetylated α-tubulin signal from the cytoplasmic acetylated α-tubulin signal. Relative intensities were analyzed in GraphPad prism 6.0. Outliers were removed using the ROUT method (Q=1%). Cytoplasmic acetylated α-tubulin signal intensity was analyzed using the Mann-Whitney test for non-parametric data.

Signal intensity ciliary ARL13B

We measured ARL13B signal intensity to exclude an effect of the *NEK1* mutation on ARL13B signal intensity inside the cilium, which could influence cilia count and measurements. We took 16-bit z-stacks of 21 x 21 µm (1024 x 1024 pixels) of nuclei and cilia, using the optimal interval and the same settings for control- and patient-derived fibroblasts (8 and 12 z-stacks respectively). We measured signal intensity of ARL13B inside the cilium for each composite image using Volocity 6.3 software. A region of interest (ROI) was defined around the cilium using the same intensity threshold for all images. Background signal was excluded. Intensity measurements were analyzed in GraphPad prism 6.0. Outliers were removed using the ROUT method (Q=1%).

There was no significant difference in ciliary ARL13B signal intensity between control- and patient-derived fibroblasts (15,456 vs. 14,371 gray units; two-tailed P-value =0.25, calculated using unpaired t-test) (Figure S5.2).

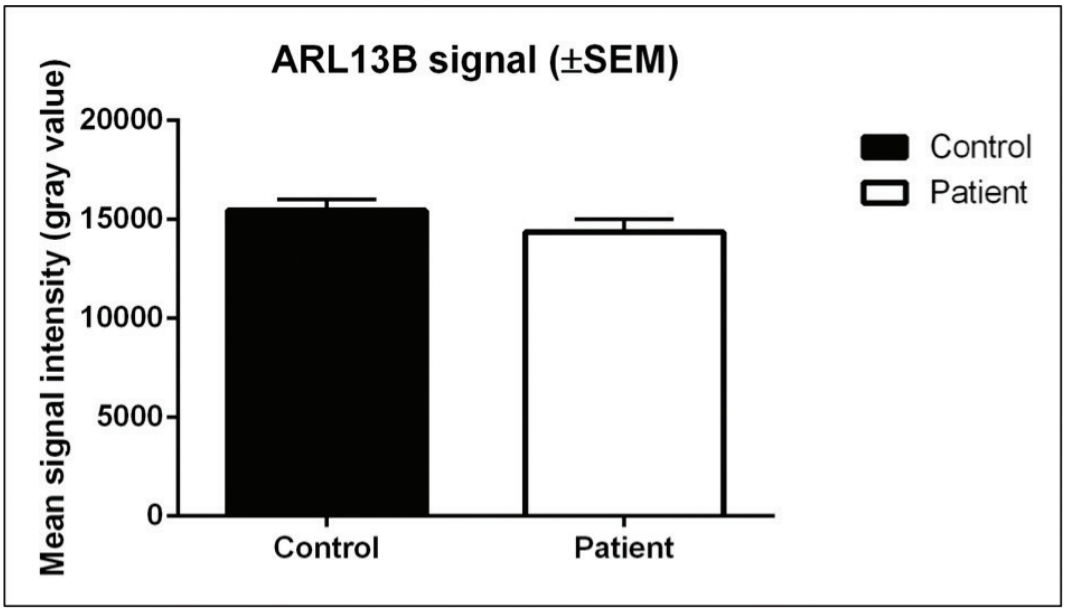


Figure S5.2 | Mean signal intensity ciliary ARL13B in control- and patient-derived fibroblasts.

Supplementary Methods S6: Splicing assessment of c.464G>C variant

To evaluate the splicing effect of the variant c.464G>C on the last base of exon 6, we designed primers upstream (exon 5) and downstream (exon 10) of the variant. Primers used were 5' CAGATATGTTTGGCCCTGAAA 3' and 5' AGGCTGTGATCCAAACTTCG 3'. RNA was isolated from patient and control fibroblasts and cDNA created as described. The following touchdown PCR program was used: 7' at 94°C, (30" at 94°C, 30" at 65-58°C in 0.5°C decreasing increments, 1' at 72°C) X 15, (30" at 94°C, 30" at 58°C, 1' at 72°C) X 27, 4' at 72°C, 8°C. Patient and control PCR products were run on a 1% agarose gel for 1 hour at 120 volts with a Lonza FlashGel DNA 50-1.5 kb marker. Sequencing of the PCR product was performed using standard Sanger sequencing.

Supplementary Methods S7: Splicing assessment of c.1226G>A variant

To evaluate the splicing effect of the variant c.1226G>A in exon 15, we designed primers upstream (exon 6) and downstream (exon 16) of the variant. Primers used were 5' TGGAACAGTACAACTTGGAGAT 3' and 5' GCTATAGTCCCTCCACTGCC 3'. RNA was isolated from patient and control fibroblasts and cDNA created as described. The following touchdown PCR program was used: 7' at 94°C, (30" at 94°C, 30" at 65-58°C in 0.5°C decreasing increments, 1' at 72°C) X 15, (30" at 94°C, 30" at 58°C, 1' at 72°C) X 27, 4' at 72°C, 8°C. Patient and control PCR products were run on a 1% agarose gel for 1 hour at 120 volts with a GeneRuler DNA 1 kb marker. Sequencing of the PCR product was performed using standard Sanger sequencing. The forward primer 5' TTCTCTCTCCTCAGC 3' was used to sequence the variant in exon 15 while the reverse primer 5' CATGCTTTTGAAGCTG 3' was used to confirm the absence of the c.464G>C mutation in exon 6.

Supplementary Table S8 | Previously reported variants in *NEK1*.

Author	Diagnosis	Zygosity	Nucleotide change	Exon/ intron	Type of variant	Amino Acid Change	Domain
El Hokayem et al. 2012, (case 1)	SRPS type II	ho	c.433G>A	Ex 6	missense	p.(Gly145Arg)	kinase
El Hokayem et al. 2012, (case 2)	SRPS type II	ho	c.758T>C	Ex 9	missense	p.(Leu253Ser)	kinase
El Hokayem et al. 2012, (case 3)	SRPS type II	ho	c.2847de-linsGG§	Ex 28	frameshift	p.(Asp949Glufs*11) §	C terminal
El Hokayem et al. 2012, (case 4)	SRPS type II	ho	c.379C>T	Ex 5	nonsense	p.(Arg127*)	kinase
El Hokayem et al. 2012, (case 9); Thiel et al. 2011	SRPS type II	ho	c.379C>T	Ex 5	nonsense	p.(Arg127*)	kinase
El Hokayem et al. 2012, (case 10); Thiel et al. 2011	SRPS type II	ho	c.869-2A>G	In 10	splice	-	basic
El Hokayem et al. 2012, (case 11); Thiel et al. 2011	SRPS type II	digenic	*NEK1*: c.1640dup *DYNC2H1*: c.11747G>A	Ex 18 Ex 82	frameshift missense	p.(Asn547Lysfs*2) p.(Gly3916Asp)	coiled- coil C terminal
Chen et al. 2012	SRPS type II	het	c.465-1G>A	In 7	splice	-	kinase
Present study	OFD type II		c.1226G>A c.464G>C	Ex 15 Ex 6	nonsense splice site	p.(Trp409*) -	coiled-coil kinase

Ho=homozygous, het=heterozygous, Ex = exon, In=Intron / *NEK1* NCBI reference sequence: NM_012224.2 / *DYNC2H1* NCBI reference sequence: NM_001080463.1

§ El Hokayem et al. 2012 original nomenclature is c.2846_2847insGG2847delT, p.(Asp949Glufs*6); notation in this manuscript corresponds to current HGVS standards

5

Monocarboxylate Transporter 1 Deficiency and Ketone Utilization

New England Journal of Medicine. 2014 Nov 13;371(20):1900-7.

PMID: 25390740 DOI: 10.1056/NEJMoa1407778.

Peter M. van Hasselt[1*], Sacha Ferdinandusse[4*], Glen R. Monroe[3], Jos P.N. Ruiter[4], Marjolein Turkenburg[4], Maartje J. Geerlings[3], Karen Duran[3], Magdalena Harakalova[3], Bert van der Zwaag[3], Ardeshir A. Monavari[7], Ilyas Okur[8], Mark J. Sharrard[9], Maureen Cleary[10], Nuala O'Connell[11], Valerie Walker[12], M. Estela Rubio-Gozalbo[5], Maaike C. de Vries[6], Gepke Visser[1], Roderick H.J. Houwen[2], Jasper J. van der Smagt[3], Nanda M. Verhoeven-Duif[3], Ronald J.A. Wanders[4], and Gijs van Haaften[3]

1 | the Division of Pediatrics, Department of Metabolic Diseases, and
2 | the Division of Pediatrics, Department of Pediatric Gastroenterology, Wilhelmina Children's Hospital, and
3 | the Center for Molecular Medicine, Department of Medical Genetics, University Medical Center Utrecht, Utrecht
4 | Laboratory Genetic Metabolic Diseases, Departments of Clinical Chemistry and Pediatrics, Academic Medical Center, Amsterdam
5 | the Division of Pediatrics, Department of Metabolic Diseases, and Laboratory Genetic Metabolic Diseases, Maastricht University Medical Center, Maastricht, and
6 | the Department of Pediatrics, Nijmegen Center for Mitochondrial Disorders, Radboud University Medical Center, Nijmegen — all in the Netherlands
7 | the National Centre for Inherited Metabolic Disorders, Children's University Hospital, Dublin, Ireland
8 | the Department of Pediatric Metabolism and Nutrition, Gazi University School of Medicine, Ankara, Turkey; and
9 | the Department of Paediatric Metabolic Medicine, Sheffield Children's Hospital, Sheffield
10 | the Department of Metabolic Medicine, Great Ormond Street Hospital NHS Foundation Trust, London
11 | Chemical Pathology, Department of Laboratory Medicine, Salisbury, and
12 | the Department of Clinical Biochemistry,Southampton General Hospital, Southampton — all in the United Kingdom.
*Joint First Authors

ABSTRACT

Ketoacidosis is a potentially lethal condition caused by the imbalance between hepatic production and extrahepatic utilization of ketone bodies. We performed exome sequencing in a patient with recurrent, severe ketoacidosis and identified a homozygous frameshift mutation in the gene encoding monocarboxylate transporter 1 (*SLC16A1*, also called *MCT1*). Genetic analysis in 96 patients suspected of having ketolytic defects yielded seven additional inactivating mutations in *MCT1*, both homozygous and heterozygous. Mutational status was found to be correlated with ketoacidosis severity, MCT1 protein levels, and transport capacity. Thus, MCT1 deficiency is a novel cause of profound ketoacidosis; the present work suggests that MCT1-mediated ketone-body transport is needed to maintain acid–base balance.

INTRODUCTION

Acetoacetate and 3-hydroxybutyrate are slightly acidic biomolecules that, together with acetone, are called ketone bodies and serve as the major circulating energy source during fasting. Ketone bodies are formed in the liver from the ultimate breakdown product of fatty acids —acetyl coenzyme A (CoA) — by coupling of two acetyl units in a three-step enzymatic process called ketogenesis. Ketone bodies are believed to undergo passive distribution to metabolically active tissues, where they are used as an energy source (Laffel 1999). Ketoacidosis, a pathologic state, occurs when ketone formation exceeds ketone utilization. The clinical consequences of ketoacidosis are exemplified by diabetic ketoacidosis, a condition that is marked by vomiting, osmotic diuresis, dehydration, and Kussmaul breathing and that may progress to decreased consciousness and, ultimately, death (Westerberg 2013).

Inborn errors of ketone utilization are manifested similarly; however, glucose levels in these types of ketoacidosis are normal or even low, in contrast to glucose levels in diabetic ketoacidosis (Sass 2012). Only two genetic causes of recurrent ketoacidosis are currently known: succinyl CoA oxoacid transferase (SCOT) deficiency (Online Mendelian Inheritance in Man [OMIM] database number, 245050) and mitochondrial acetoacetyl-CoA thiolase (ACAT1) deficiency (also called beta-ketothiolase deficiency; OMIM number, 203750). Both SCOT and ACAT1 are involved in ketolysis, the breakdown of ketone bodies into the key cellular energy source, acetyl CoA.

We performed targeted exome sequencing of homozygous genomic regions in a patient of consanguineous descent who had recurrent, severe ketoacidosis. A homozygous mutation was detected in the gene encoding monocarboxylate transporter 1 (*MCT1*). Subsequently, we evaluated a series of 96 patients with recurrent ketoacidosis, in whom known ketolytic defects had been ruled out enzymatically, to identify additional patients with mutations in *MCT1* or related genes.

METHODS

Study Participants and Genetic Studies

We performed targeted exome sequencing of homozygous genomic regions in the index patient and her family members. Regions of homozygosity in the index patient were determined with the use of a high-resolution single-nucleotide polymorphism (SNP) array. Coding parts of homozygous regions were then captured on a custom array and sequenced as described elsewhere (Harakalova et al. 2011; Nijman et al. 2010). To follow up on the findings from exome sequencing, we sequenced the entire coding region of *MCT1* in a series of 96 patients with ketoacidosis in whom known ketolytic defects had been ruled out because of the normal enzymatic activities of ACAT1 and SCOT. In addition, we performed Sanger sequencing of the related genes *MCT2 (SLC16A7)*, *MCT3 (SLC16A8)*, and *MCT4 (SLC16A3)*, plus the ancillary gene *BSG*, in these patients. Heterozygosity of mutations was confirmed by means of complementary DNA (cDNA) sequencing and genomic deletion analysis. Further details of sequencing and analysis are available in the Supplementary Appendix. Written informed consent for targeted whole-exome sequencing and Sanger sequencing was obtained from the parents of the index patient for themselves, the index patient, and her two siblings. Sanger sequencing of candidate genes in the cohort of patients with a suspected ketolytic defect and blood collection for functional studies were performed as a part of the diagnostic process (approved by the medical ethics committee of the University Medical Center Utrecht).

MCT1 Expression and Functional Studies

Erythrocyte lactate transport was measured essentially as described by Fishbein (1986) (Fishbein 1986). For the analysis of each blood sample from a patient, at least one control sample was included in the same experiment, with the same procedures used for blood collection, transport, and analysis. Control blood samples were taken at random from blood left over from diagnostic tests, which was used anonymously in this study. Immunoblotting was performed with fibroblast homogenates and an affinity-purified rabbit polyclonal antibody against MCT1 and MCT4. Fibroblast homogenates from 10 healthy volunteers were also used as controls. Further details on functional and expression studies are available in the Supplementary Appendix. Reference blood results from testing of children before and after 24 hours of fasting were obtained from published data (van Veen et al. 2011).

Statistical Analysis

All applied statistical tests were two-tailed unpaired t-tests. P values of less than 0.05 were considered to indicate statistical significance.

RESULTS

Case Report

The index patient had repetitive episodes of profound metabolic acidosis, with a blood pH below 7.00 on three occasions, all accompanied by massive urinary excretion of 3-hydroxybutyrate and acetoacetate, with plasma lactate and ammonia levels remaining in the normal range. Between episodes, results of blood gas analyses were normal. See the Supplementary Appendix for an extended case report.

Genetic Studies

SNP array analysis in the index patient revealed 36 homozygous genomic regions larger than 1 Mb, with 7 of the 36 larger than 10 Mb, which confirmed consanguinity. Targeted-exome sequencing of coding parts of these homozygous regions in the five family members yielded nine rare variants. Of these, a single-nucleotide insertion disrupting the reading frame of *MCT1* at asparagine 15 (c.41dupC; National Center for Biotechnology Information reference sequence number, NM_001166496.1) was the strongest candidate. The correct segregation of this mutation in the family was confirmed by means of Sanger sequencing. This variant was absent from multiple large variation databases (1000 Genomes, dbSNP, and the National Heart, Lung, and Blood Institute [NHLBI] GO Exome Sequencing Project Exome Variant Server [data release ESP6500]). The insertion of an extra nucleotide early in the gene sequence results in a frameshift and thus in a loss of MCT1 function.

MCT1 Mutations in Additional Patients with Ketoacidosis

MCT1 is one of the transmembrane transporters encoded by members of the SLC16 gene family; among these transmembrane transporters, MCT1, MCT2, MCT3, and MCT4 have been shown to transport monocarboxylates, including lactate and ketone bodies (Halestrap 2013). MCT1 requires the glycoprotein BSG (also called CD147) for proper subcellular expression (Kirk et al. 2000; Wilson et al. 2005). We therefore sequenced *MCT1* and related genes in a series of 96 patients with unexplained ketoacidosis. We identified seven additional suspected deleterious variants, all in *MCT1*, including six truncating mutations and one missense mutation (Figure 1A and Table 1). All truncating mutations

are expected to result in a loss of protein function. The missense mutation, p.Arg313Gln, targets a highly conserved residue that is essential for catalytic activity (Manoharan et al. 2006). In total, we identified 9 patients with mutations in *MCT1*, 3 homozygous and 6 heterozygous, including 2 siblings. None of these mutations were seen previously in population-based variation databases (1000 Genomes, dbSNP, NHLBI GO Genome Sequencing Project Exome Variant Server, Genome of the Netherlands) (Genome of the Netherlands Consortium 2014). Together, the genetic data firmly link defects in *MCT1* to ketoacidosis in our patients.

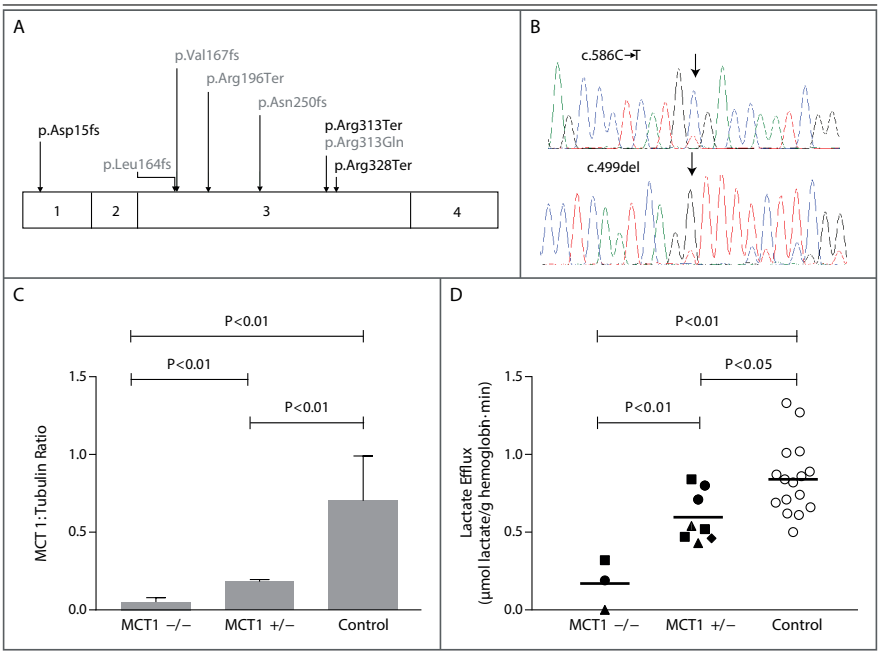


Figure 1 | MCT1 Deficiency in Patients with Episodes of Ketoacidosis. Panel A) shows the identified mutations in the monocarboxylate transporter 1 (MCT1) protein, with the four different exons indicated. The full-length MCT1 protein consists of 500 amino acids. Heterozygous mutations are indicated in gray, and homozygous mutations in black. Panel B) shows examples of the sequencing of complementary DNA in fibroblasts from patients with heterozygous mutations; the colors represent bases: green denotes A, red T, black G, and blue C. The arrows indicate the mutated residue. The majority of the transcript is unmutated *MCT1* sequence, and the minority is the mutated sequence as a result of nonsense-mediated decay. Panel C) shows the MCT1:tubulin ratio from a quantitative immunoblot with fibroblast homogenates from 4 patients with heterozygous truncating mutations as compared with 10 healthy controls. Panel D) shows lactate efflux from erythrocytes endogenously loaded with lactate. Efflux was measured for patients with a homozygous *MCT1* mutation and their heterozygous family members (solid circles denote p.Arg313Ter, squares p.Arg328Ter, and triangles p.Asp15fs), a heterozygous patient (p.Arg313Gln, denoted by the diamond), and controls (denoted by open circles). The lactate transport in the homozygous patients was significantly lower than in both controls and heterozygotes. The mean lactate transport activity in heterozygotes was also reduced. Horizontal bars indicate the mean transport activity.

MCT1 Expression

Next, we studied the effect of the mutations in *MCT1* at the protein level. Immunoblot analysis with fibroblast homogenates from patients with a homozygous *MCT1* mutation, performed with antibodies raised against a C-terminal peptide of MCT1, confirmed that predicted truncating mutations lead to an absence of the full-length protein (Figure S1A in the Supplementary Appendix). In five patients, we could detect only a single heterozygous truncating mutation in *MCT1* (Figure 1A), despite sequencing the complete coding region. Analysis of cDNA confirmed biallelic expression (Figure 1B). A custom multiplex amplification quantification analysis further ruled out the presence of exon-sized deletions in heterozygous patients. Immunoblot analysis showed significantly reduced levels of MCT1 protein in fibroblasts from patients with a heterozygous truncating mutation, relative to levels in controls (Figure 1C, and Figure S1B in the Supplementary Appendix). The immunoblot results confirmed that there was an MCT1 deficiency at the protein level. We tested fibroblasts to determine whether the MCT1 deficiency was compensated for by increased cellular expression of MCT4 but found no evidence for this (Figure S1C in the Supplementary Appendix).

Monocarboxylate Transport Assay

Next, we studied the effect of mutations in *MCT1* on monocarboxylate transport, using a previously described assay for MCT1 activity that measures lactate efflux from erythrocytes endogenously loaded with lactate (Fishbein 1986; Merezhinskaya et al. 2000). In the samples from patients with a homozygous mutation in *MCT1*, erythrocyte lactate transport was significantly reduced as compared with that in the control samples (Figure 1D). The mean lactate transport activity in erythrocytes from heterozygous carriers, both symptomatic and asymptomatic, was significantly reduced as compared with the transport activity in control samples but was significantly higher than that in homozygous patients (Figure 1D).

Table I | Spectrum of *MCT1* Mutations and Associated Phenotypes*

Variant	Effect	Inheritance	Phenotype	Patient Characteristic						
				Ancestry	Parental Relatedness	Sex	No. of Ketotic Events	Age at Diagnosis/ Last Event/ Last Follow-up	Intellectual Disability	Other Clinical Features
Homozygous										
c.41dupC	p.Asp15fs	Sporadic	Profound ketoacidosis	Syrian	Second-degree cousins	F	5	3 mo/ 7 yr 5 mo/ 8 yr	Moderate	Microcephaly, atrial septal defect, hypoplastic left pulmonary artery and main bronchus
c.937C->T	p.Arg313Ter	Sporadic	Profound ketoacidosis	Irish	None	F	4	1 yr 2 mo/ 7 yr 8 mo/ 21 yr	Moderate	Epilepsy
c.982C->T	p.Arg328Ter	Sporadic	Profound ketoacidosis	Turkish	Third-degree cousins	M	1	1 yr 11 mo/ 1 yr 11 mo/ 9 yr	Mild	Cleft palate, coughing attacks during exercise
Heterozygous										
c.586C->T	p.Arg196Ter	Sporadic	Cyclic vomiting, ketoacidosis with massive ketonuria	British	None	M	3	3 yr 11 mo/ 7 yr 6 mo/ 10 yr	None	Migraine
c.747_750del	p.Asn250fs	Familial	Ketoacidosis with massive ketonuria	British	None	M	10	3 yr 8 mo/ 14 yr/ 20 yr	None	Migraine
c.747_750del	p.Asn250fs	Familial	Ketoacidosis with massive ketonuria, pregnancy-related vomiting	British	None	F	5	6 yr 3 mo/ 17 yr/ 22 yr	None	-
c.499del	p.Val167fs	Sporadic	Exaggerated ketotic hypoglycemia, ketoacidosis with massive ketonuria	British	None	F	5	1 yr 6 mo/ 3 yr 9 mo/ 11 yr	None	-
c.490dupC	p.Leu164fs	Sporadic	Ketoacidosis with massive ketonuria	Dutch	None	M	4	3 yr 2 mo/ 7 yr 5 mo/ 10 yr	None	Short stature
c.938G->A	p.Arg313Gln#	Sporadic	Cyclic vomiting, ketoacidosis with massive ketonuria	Dutch	None	M	5	2 yr 6 mo/ 6 yr 5 mo/ 22 yr	None	Exercise intolerance

* The complementary DNA and protein annotations are based on National Center for Biotechnology Information reference sequence numbers NM_001166496.1 and NP_001159968.1, respectively.
# The variant was predicted by PolyPhen-2 to be “probably damaging” and by SIFT to be “deleterious.”

Clinical Symptoms

All patients presented with bouts of ketoacidosis provoked by fasting or infections in their first years of life (Table 1), as is illustrated for the index patient in Figure 2A, and Figure S2A in the Supplementary Appendix. The pH of the blood was normal between episodes. Excretion of urinary ketones under these circumstances ranged from normal to clearly elevated (Figure 2B). Ketoacidotic episodes were preceded by poor feeding and vomiting and were associated with dehydration, which was a consequence of osmotic diuresis and vomiting. Hypoglycemia was observed infrequently, and glucose levels usually remained in the normal range (Figure S2B and S2C in the Supplementary Appendix). In some patients, repetitive vomiting led to a diagnosis of ketotic vomiting or abdominal migraine. Profound metabolic acidosis was associated with decreased consciousness and insufficient respiratory drive (Figure 2C), which worsened the degree of acidosis. In all patients, treatment with intravenous glucose or dextrose (combined with bicarbonate) led to rapid clearance of metabolic acidosis. Early initiation of treatment appeared to prevent ketoacidosis. Similarly, ensuring an adequate caloric intake reduced the number of episodes. The frequency of ketoacidotic episodes appeared to decrease over time, and none of the patients had documented ketoacidosis after 7 years of age. Nevertheless, some patients continued to have marked ketonuria associated with (mild) infections. Patients with homozygous mutations in *MCT1* tended to present at a younger age ($P = 0.05$) and had more profound ketoacidosis (Figure 2D, and Figure S2D in the Supplementary Appendix). In addition, homozygous patients had mild-to-moderate developmental delay, whereas patients with milder deficiencies of *MCT1* had normal development.

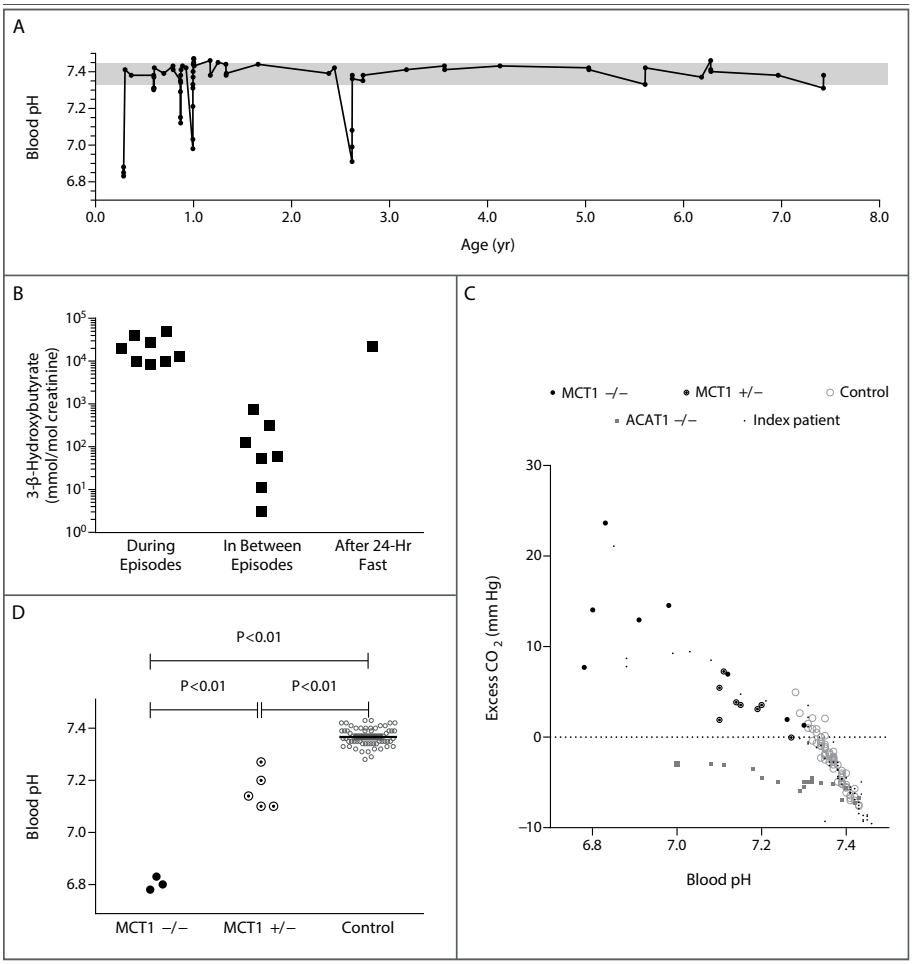


Figure 2 | Blood Test Results from Patients with MCT1 Deficiency. Panel A) shows the blood pH over time for the index patient (who had the p.Asp15fs mutation). The ketoacidotic events are clearly visible. An extended case report for this patient is available in the Supplementary Appendix. Panel B) shows the 3-β-hydroxybutyrate levels in urine, normalized to creatinine, in three MCT1-deficient patients. Control values are less than 10 mmol per mole of creatinine in between episodes and between 160 and 6400 mmol per mole of creatinine after 24 hours of fasting. Panel C) shows the carbon dioxide excess (actual carbon dioxide concentration minus expected carbon dioxide concentration) versus the pH in homozygous or heterozygous MCT1-deficient patients and controls. The expected carbon dioxide concentration was calculated according to Winter's formula; the horizontal dotted line indicates the absence of excess carbon dioxide. Data points for the index patient show clearly that blood values were within the normal range except during ketotic events. Values during a ketotic event in a patient with mitochondrial acetoacetyl-CoA thiolase (ACAT1) deficiency form a separate constellation from the values for MCT1-deficient patients. Panel D) shows the lowest recorded blood pH value during a ketotic event in homozygous and heterozygous MCT1-deficient patients and controls. The horizontal bar indicates the mean control value. The pH is lower in the homozygous and heterozygous patients than in controls and lower in homozygotes than in heterozygotes.

DISCUSSION

We identified MCT1 deficiency as a disorder of ketone utilization, expanding the spectrum of disorders leading to ketoacidosis from those of ketolysis to those of ketone delivery. This finding may aid in timely diagnosis of the disorder and allow for improved disease management. In addition, this observation implies that MCT1-mediated import of ketone bodies into extrahepatic tissues is essential during periods of catabolic stress in order to maintain acid–base balance.

MCT1 encodes a transmembrane protein that facilitates proton-linked transport of a range of monocarboxylate metabolites, such as lactate, pyruvate, and ketone bodies, across the cellular membrane (Halestrap 2013). Our observation of profound ketoacidosis in patients with MCT1 deficiency highlights the fact that facilitated transport of ketone bodies by MCT1 is essential during catabolic stress, when ketone turnover in the body is orders of magnitude higher than in the fed state. Heterozygous missense mutations in *MCT1* have previously been reported in association with suboptimal erythrocyte lactate transport resulting in muscle injury on exercise and heat exposure (Merezhinskaya et al. 2000). Our patient with a heterozygous missense mutation had exercise intolerance, which indicates a possible overlapping phenotype. Mutations leading to increased expression of MCT1 are associated with exercise-induced hyperinsulinism (Marquard et al. 2013; Otonkoski et al. 2007). Monocarboxylate transporters show tissue-specific variation in expression between species, which underlines the importance of insights gained in our study of *MCT*1 function in human disease (Merezhinskaya and Fishbein 2009). Homozygous *MCT1*-knockout mice die early during embryogenesis, in contrast to the patients with a complete loss of *MCT1* function in this study (Lee et al. 2012; Lengacher et al. 2013). Mouse placenta and human placenta express *MCT1* in opposite subcellular arrangements, which provides a possible explanation for the differences in embryonic lethality (Nagai et al. 2010). Disruption of *MCT1* in the central nervous system produces axon damage and neuronal loss in mice (Lee et al. 2012). The patients with homozygous loss-of-function mutations in *MCT1* described in our study have moderate intellectual disability; however, at this stage it remains unclear whether this is a direct effect of the absence of *MCT1* in the brain or caused by episodes of profound ketoacidosis.

The use of acidic biomolecules as an energy source presents a potential threat to acid–base homeostasis and therefore requires a careful balance between production and consumption. This aim is achieved indirectly, by linking the rate of ketogenesis with cellular energy status in extrahepatic tissues. Concerted action of glucagon and insulin released by the pancreas determines the rate of ketogenesis in the liver. The massive ketoacidosis in patients with MCT1 deficiency and other disorders of ketone utilization — SCOT deficiency

and ACAT1 deficiency — unveils the weak spot of this indirect feedback mechanism. In these disorders, the conversion of ketone bodies to acetyl CoA becomes rate limiting. As a consequence, ketone production is uncoupled from consumption, leading to profound metabolic acidosis. We found that a single heterozygous *MCT1* mutation can result in a deficiency of the transporter and in clinical symptoms. These findings suggest that *MCT1* deficiency is more prevalent than is apparent among previously described disorders of ketone utilization. The heterozygous family members of patients with homozygous MCT1 loss-of-function mutations had no history of ketoacidosis, which strongly suggests that heterozygous mutations — and possibly even homozygous mutations — in *MCT1* give rise to ketoacidosis only in conjunction with certain genetic and environmental factors. The varying number of episodes and long symptom-free intervals in both heterozygous and homozygous patients support this hypothesis. Future research will have to identify the factors involved in triggering the development of symptoms.

Supported by the Child Health program of the University Medical Center Utrecht.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org. We thank Aafke Terlingen for technical assistance.

SUPPLEMENTARY APPENDIX

Targeted whole-exome sequencing

Original genomic DNA was isolated from peripheral lymphocytes of the index patient and family members according to standard procedures. SNP array assays were performed on all five family members using a high-resolution SNP array HumanCytoSNP 12 BeadChip, (Illumina). Homozygous stretches larger than 1 MB were identified, using an in house developed algorithm (available upon request). Thirty-six homozygous stretches larger that 1 MB were identified. A custom capture array was designed targeting the exonic parts plus 50 bp flanks within the homozygous stretches (Nijman et al. 2010). The designed targeted area was 3.1 Mb, and probes could be designed for 92% of the requested genomic regions. Sequencing libraries were prepared as described previously and sequenced on the Solid 5500 platform with an average coverage ranging between 160 to 239 fold (Nijman et al. 2010). Mapping and variant analysis was performed as described previously, with the exception that a recessive inheritance model was applied (Harakalova et al. 2012). Initial filtering did not yield good candidate variants; therefore the filtering criteria were relaxed to allow for improved detection of small insertions and deletions. The used settings for filtering were: allele frequency <0.01, coverage >10 reads, percentage variant >50% for the patient, >10% for parents and <80% for unaffected siblings and exclusion of missense variants annotated as benign by PolyPhen-2.

Sanger sequencing, cDNA and deletion analysis

The complete coding regions of *MCT1/SLC16A1*, *MCT2/SLC16A7*, *MCT3/SLC16A8*, *MCT4 /SLC16A3* and *CD147/BSG* were sequenced in 96 patients suffering from episodes of severe ketoacidosis. The primer sequences are available upon request. Homozygous mutations were confirmed by verification of the heterozygous nature of the mutation in both parents. For the analysis of the mutation state in mRNA, cDNA was generated using High Capacity cDNA Reverse Transcription Kit (Life technologies) on RNA isolated from fibroblasts using Trizol reagent (Life technologies). Subsequently a PCR was performed using cDNA specific primers (sequence available upon request), followed by Sanger sequencing. To test for potential copy number variation in heterozygous patients, we performed an MLPA-‑assay in our ISO15189 accredited diagnostic lab according to the manufacturers protocols. Six multiplex amplification probes (Multiplicon) were designed covering the 5 exons of *SLC16A1*, with the largest exon 4 covered twice. Primer sequences are available upon request.

Expression and functional analysis

Erythrocyte lactate transport assay

Erythrocyte lactate transport was measured essentially as described by Fishbein (1986) (Fishbein 1986). Within 26 hours after drawing (EDTA) anticoagulated blood samples were incubated at 37°C for 2 h without stirring to load the erythrocytes with lactate. Erythrocytes were then packed, washed twice with buffer (134 mM NaCl, 4.8 mM KCl, 0.96 mM MgCl2, 10 mM sucrose, 10 mM MES pH 6) and subsequently diluted tenfold in incubation buffer (134 mM NaCl, 4.8 mM KCl, 0.96 mM MgCl2, 10 mM sucrose, 10 mM Tricine pH 7.5). At six 1-min intervals 0.5 ml aliquots were pipetted into iced buffer (134 mM NaCl, 4.8 mM KCl, 0.96 mM MgCl2, 10 mM sucrose, 10 mM MES pH 5) to stop the reaction. The samples were then spun (1 min, 20,000 x g at 4°C), washed with buffer (134 mM NaCl, 4.8 mM KCl, 0.96 mM MgCl2, 10 mM sucrose, 10 mM MES pH 6), spun again (1 min, 20,000 x g at 4°C) and lysed in 200 µl 0.59 M perchloric acid. Samples were thoroughly mixed, spun (5 min, 20,000 x g at 4°C) and the supernatant was used for measurement of lactic acid. Lactic acid concentration was determined by following the production of NADH spectrophotometrically at 340 nm for 30 min at 37°C in the presence of L-glutamate (50 mM in 400 mM NaCO3, pH10), NAD+ (2 mM), Glutamate-pyruvate transaminase (3.5 U/ml, Roche) and Lactate dehydrogenase (105 U/ml, Sigma). Specific activity was expressed as lactate efflux from the erythrocyte in time per g Hb. For each patient sample, at least one control sample was included within the same experiment, which was treated exactly the same from blood drawing until analysis. All analyses were also performed in the presence of 20 µM p-chloromercuribenzene suphonate (pCMBS), a potent inhibitor of MCT1, to verify that the measured lactate efflux was transporter dependent.

Immunoblot analysis

Cultured skin fibroblasts were homogenized in PBS containing protease inhibitors (Complete mini, Roche) and aliquots of 20 ug of total protein were separated a 4-12% gradient NuPAGE Bis-Tris gel (Invitrogen) in a MES buffer (pH 7.3) and transferred onto PVDF by semidry blotting. Membranes were blocked with normal goat serum (NGSe) for 1 h at room temperature and then incubated overnight at 4°C with affinity purified rabbit polyclonal antibody against MCT1 (Millipore, AB3538P) or MCT4 (Millipore AB3316P) diluted 1:100 and 1:500 respectively in 4% NGSe/PBS. As a loading control, the membranes were reprobed with a monoclonal antibody against α-tubulin (Molecular probes), using a 1:5000 dilution. Antigen-antibody complexes were visualized with IRDye 800CW goat anti-rabbit secondary antibody for MCT1 and IRDye 680RD donkey anti-mouse secondary antibody for tubulin using the Odyssey Infrared Imaging System (LI-COR Biosciences, Nebraska,

USA). Quantification of the signal intensities of the MCT1 and tubulin bands was done using AIDA Image analyzer software (Version 4.26, Raytest, Straubenhardt, Germany).

Calculations

The expected CO2 excess was calculated according to Winter's formula, where P(CO2) = (1.5 * HCO_3^-) + 8±2 where HCO_3^- is given in mEq / L and P(CO2) in mmHg.

Case Report

The proband, a girl, was the second child of consanguineous parents. At the age of 3.5 months she was admitted with respiratory distress due to a viral upper respiratory-tract infection and a wheeze. Unexpectedly, blood gas analysis revealed profound metabolic acidosis, pH 6.88, Base Excess (BE) - 28 mmol / L, and an elevated anion gap (32 mmol / L, normal < 15 mmol /L). Glucose was borderline normal at 3.4 mmol / L. Urinary ketostick showed strongly elevated ketone levels. She was transferred to an ICU and recovered rapidly after receiving sodium bicarbonate and IV glucose on this and subsequent occasions. Cardiac ultrasound showed a hemodynamically insignificant type II atrial septum defect, persistent ductus botalli and hypoplastic left pulmonary artery. In addition, subsequently, hypoplasia and malacia of the main left bronchus was observed on bronchoscopy. Abdominal ultrasound confirmed normal sized liver and spleen. On cerebral ultrasound increased echogenicity of caudate nucleus and lentiform nucleus was noted. Borderline microcephaly was observed on follow up (SD -2.7). MRI cerebrum was unremarkable, but subsequently a developmental delay became evident.

At an age of seven months she was admitted because of vomiting and refusal to eat. Metabolic acidosis (pH 7.30, BE - 10.6 mmol / L) cleared rapidly after a single correction with sodium bicarbonate. Nevertheless, strongly elevated urinary ketones were repeatedly observed, possibly due to insufficient breastfeeding. A similar episode at an age of nine months was triggered by a respiratory syncytial virus infection. At 12 months of age she was admitted for percutaneous closure of persistent ductus arteriosus. Despite regular fasting measures around this procedure, poor feeding resulted in vomiting, drowsiness and severe metabolic acidosis within a day, pH 6.98; BE - 23 mmol/L. The last significant episode occurred at 2.5 years and was triggered by poor feeding and otitis media acuta. Further episodes could be prevented with advice to avoid fasting and early admissions in case of poor feeding and vomiting. This included uneventful percutaneous closure of atrial septal defect, requiring general anesthesia at 7 years.

Metabolic investigations during bouts of ketoacidosis revealed massive urinary excretion of 3-hydroxybutyric acid (range 2100 - 12500 mmol /mol creatinine, normal < 8 mmol /mol creatinine) and 3-ketobutyric acid (range 3900 - 26500 mmol/mol creatinine, normal below 5 mmol /mol creatinine), without consistent other abnormalities. Plasma acylcarnitine profile occasionally showed elevated hydroxybutyrylcarnitine (range 0.05 - 1.51 µmol / L, normal <0.2 µmol / L).

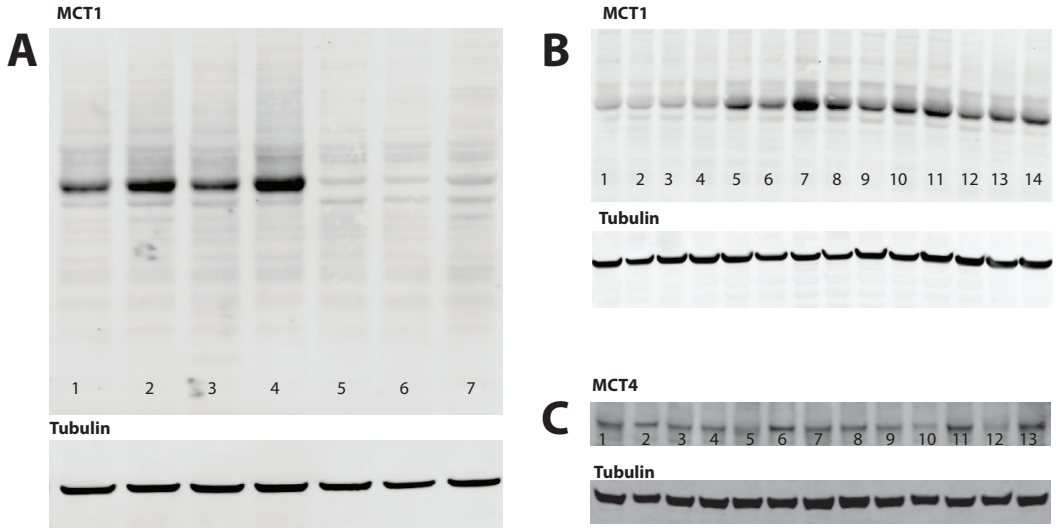


Figure S1 | Immunoblot results for patients with MCT1 deficiency and controls. Panel A and B show immunoblot results for fibroblast homogenates from MCT1 deficient patients, stained for MCT1 (45 kDa). Tubulin was used as a loading control. Panel A) shows results for patients with homozygous mutations and controls. Lanes were loaded as follows: 1 - 4 were controls, 5: p.Asp15fs, 6: p.Arg328Ter and 7: p.Arg313Ter. Panel B) shows results for patients with a heterozygous truncating mutation and controls, stained for MCT1. Lanes were loaded as follows: 1: p.Leu164ProfsTer46, 2: p.Asn250SerfsTer5, 3: p.Val167PhefsTer13, 4: p.Leu164ProfsTer46 and 5Noteworthy 14 were controls. Panel C) shows immunoblot results for fibroblast homogenates from MCT1 deficient patients, stained for MCT4 (50 kDa). Tubulin was used as a loading control. Lanes were loaded as follows: 1-3 homozygous mutations (1: p.Arg328Ter; 2: p.Arg313Ter; 3: p.Asp15fs); 4-7 heterozygous mutations (4: p.Arg196Ter; 5:p.Asn250SerfsTer5; 6: p.Val167PhefsTer13; 7. p.Leu-164ProfsTer46); 8-13: controls.

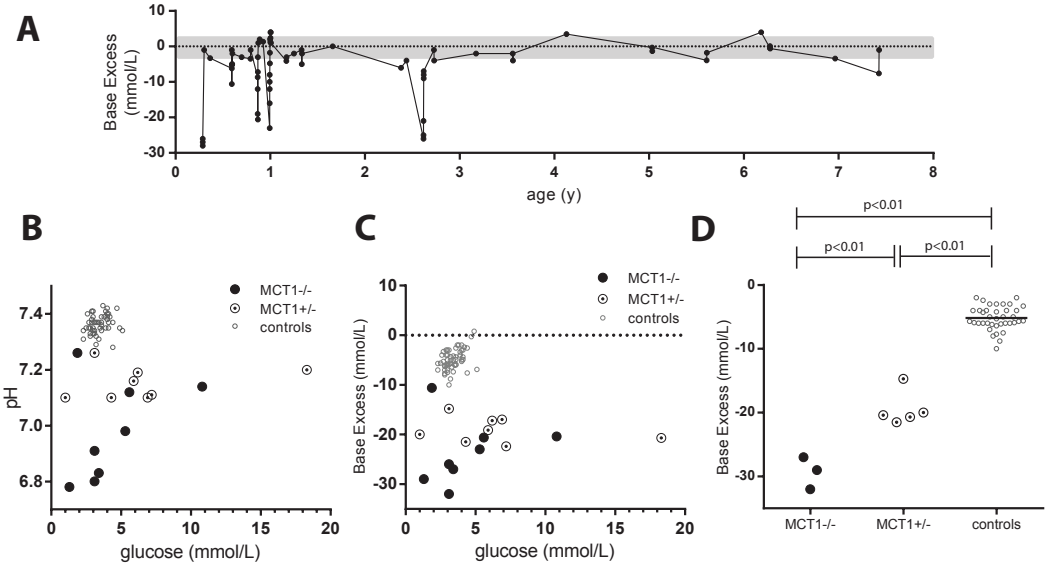


Figure S2 | Base excess in patients with MCT1 deficiency. Panel A) shows the base excess over time for the index patient. Panel B) and C) show the blood glucose versus pH (panel B) and blood glucose versus base excess (panel C) in homozygous (-/-) and heterozygous (+/-) *MCT1* deficient patients at presentation of a documented episode of catabolic stress. Panel D) shows the base excess for homozygous and heterozygous patients compared to controls. The origin of the control data is described in the legend of Fig 2.

6

Identification of a human D-lactate dehydrogenase deficiency

Manuscript in preparation

Glen R Monroe[1,2*], Albertien M. van Eerde[1,2*], Federico Tessadori[1,2,3], Karen J. Duran[1,2], Sanne M.C. Savelberg[1,2], Johanna C. van Alfen[4], Saskia N. van der Crabben[1], Paulien A. Terhal[1], Klaske D. Lichtenbelt[1], Johan Gerrits[1], Markus J. van Roosmalen[1,2], Mirjam van Aalderen[1], Bart G. Koot[5], Marlies Oostendorp[6], Marinus Duran[7], Gepke Visser[8], Tom de Koning[9], Francesco Calì[10], Paolo Bosco[10], Monique G.M. de Sain-van der Velden[1], Nine V. Knoers[1,2], Jeroen Bakkers[3,11], Nanda M. Duif-Verhoeven[1,2], Gijs van Haaften[1,2**], Judith J. Jans[1,2**]

1 | Department of Genetics, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
2 | Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
3 | Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht 3584 CT, The Netherlands
4 | Bartiméus, Institute for the Visually Impaired, Doorn, 3940 AB, The Netherlands
5 | Department of Pediatric Gastroenterology and Nutrition, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands
6 | Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
7 | Laboratory Genetic Metabolic Diseases, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands
8 | Department of Pediatric Gastroenterology and Metabolic Diseases, University Medical Center Utrecht, Utrecht 3584 CT, The Netherlands
9 | Section of Metabolic Diseases, Beatrix Children's Hospital, University Medical Center Groningen, Groningen 9713 GZ, The Netherlands
10 | Oasi Institute for Research on Mental Retardation and Brain Aging (IRCCS), Troina 94018, Italy
11 | Department of Medical Physiology, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands.
*Joint First authors, **Joint Last authors.

ABSTRACT

Human knockouts may provide direct insight into gene function. We identified a patient homozygous for a loss-of-function variant in *LDHD*, lactate dehydrogenase D. Lactate dehydrogenases catalyze the interconversion of pyruvate and lactate during anaerobic glycolysis, L-lactate being the form utilized in eukaryotic metabolism. D-lactate, the stereoisomer, is normally present physiologically at much lower levels with blood concentration approximately 100 times lower compared to L-lactate. In our patient, an abnormally high excretion of D-lactate and increased plasma concentration was detected as well as D-2-hydroxyisovaleric and D-2-hydroxyisocaproic acid. LDHD has been identified as a putative metabolizer of D-lactate but until now its function had not been shown *in vivo*. We confirmed the connection between LDHD and D-lactate metabolism by functional metabolic studies in zebrafish. We demonstrate that loss of function of LDHD in zebrafish results in increased concentration of D-lactate and both D-2-hydroxyacids with no other observable phenotype. Expressing human wild-type *LDHD* mRNA rescued the phenotype and restored D-lactate metabolism resulting in D-lactate reduction. In contrast, expression of patient variant *LDHD* mRNA was unable to restore LDHD function and resulted in high levels of D-lactate, thus confirming the role of LDHD in D-lactate metabolism and the loss-of-function effect of our patient's variant. Our work provides direct evidence that LDHD is essential for D-lactate metabolism in the human body.

INTRODUCTION

D-lactate is present at low concentration in human blood

In the human body lactate exists in two optical isomers: L-lactate and D-lactate (Figure 1). These optical isomers have identical physical and chemical properties but rotate polarized light in a different direction. L-lactate is present in blood at concentrations of 0.5-2.2 mmol/L and is produced from pyruvate during anaerobic glycolysis (Fabian et al. 2017; Kruse et al. 2011). D-lactate is present in blood at much lower levels than L-lactate (5-20 µmol/L in healthy adults) (Talasniemi et al. 2008). D-lactate is acquired exogenously by consumption of foods (particularly fermented products), by metabolism of chemical compounds such as propylene glycol, via production by intestinal bacteria, or endogenously as a result of methylglyoxal metabolism; a toxic product that is converted to D-lactate (Christopher et al. 1990; de Vrese et al. 1990; Jorens et al. 2004; Kalapos 1999; Motonaka et al. 1998).

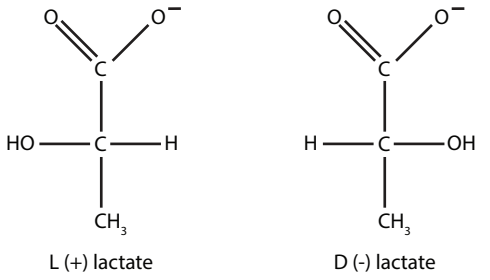


Figure 1 | Lactate optical isomers. D-lactate is the optical isomer of L-lactate. Note the relative position of the hydroxyl group and hydrogen atom.

D-lactic acidosis: causes and consequences

The human body has the ability to metabolize D-lactate (Cammack 1969; de Vrese et al. 1990; Oh et al. 1985; Tubbs 1965). In rare cases, D-lactate accumulates in the body and can result in metabolic acidosis if blood pH decreases below pH 7.35 (Oh et al. 1979). This condition is seen as a rare complication of short bowel syndrome, and can occur after removal of a section of the small intestine due to malignancy or disease (i.e. Crohn's disease), or following jejunoileal bypass surgery for treatment of obesity (Halperin and Kamel 1996; Halverson et al. 1984; Thompson et al. 2003; Uribarri et al. 1998). In this syndrome, the shortened small intestine impairs carbohydrate absorption and can lead to increased carbohydrate delivery to colonic bacteria. As a consequence, colonic bacteria can proliferate and eventually foster an acidic environment that may favor species that produce D-lactate. Development of D-lactate acidosis is dependent on the presence of these abnormal bacteria flora, and colonic pH below 6.5 stimulates D-lactate production (Caldarini et al. 1996; Petersen 2005; Uribarri et al. 1998). The accompanying clinical features include excretion of D-lactate in urine, and neurological symptoms such as altered mental status (from mild drowsiness to coma), slurred speech, ataxia and disturbance of gait (Uribarri et al. 1998). Management for D-lactate acidosis constitutes restoration of the acid-base balance by bicarbonate infusion, rehydration, antibiotic delivery to reduce the number and species of D-lactate producing bacteria, and subsequent carbohydrate reduction to prevent bacterial D-lactate overproduction (Fabian et al. 2017).

Identification of the human gene responsible for D-lactate metabolism

We report on a patient with increased excretion of D-lactate and elevated concentrations of D-lactate in plasma, but with no accompanying acidosis. The patient, originally described in Duran et al. (1977), was initially seen by a clinical geneticist at age 4 due to his delayed motor and mental development (Duran et al. 1977). Further biochemical assays to find the cause of his intellectual disability revealed the presence of large amounts of D-lactate in his urine. Upon genetic examination later in life by arrayCGH he was diagnosed with a *de novo* 11p13 deletion, known to cause intellectual disability (Riccardi et al. 1978). However, as increased D-lactate excretion is not a described feature of 11p deletion syndrome, we investigated if this perturbation of D-lactate metabolism could be due to a genetic cause. We identified a homozygous variant in a conserved region of *LDHD*, lactate dehydrogenase D, which has been identified as a putative lactate dehydrogenase in mammals but whose function in humans has not been elucidated (de Bari et al. 2002; Flick and Konieczny 2002).

The identification of the genetic defect in this patient allowed us to study the consequence of *LDHD* dysfunction and to demonstrate that this gene is responsible for D-lactate metabolism in the human body. By functional zebrafish metabolic studies, we could confirm that loss of function of LDHD leads to an inability to metabolize D-lactate sufficiently. In this report, we provide evidence for a role of LDHD in D-lactate metabolism in the human body.

RESULTS AND DISCUSSION

Clinical Synopsis

The patient, born of Sicilian parents of the same village, was born after an uneventful pregnancy. Postnatally he showed microcephaly, slanting of the eyelids, bilateral inguinal hernia, cryptorchidism and aniridia. At the age of one year his mental and motor development was evaluated as severely impaired. He had epilepsy until the age of 14. At the age of 40 he was re-evaluated. He had behavioral problems and blindness due to the aniridia, with later onset of cataract and glaucoma. He was no longer microcephalic but physical examination showed dymorphisms: slanting eyelids, a protruding lower lip, mildly dysplastic helices and patches of grey hair.

Metabolic analysis reveals abnormally high D-lactate excretion in urine

When initially investigated at the age of four, the urine of the patient contained a high concentration of lactate (7-29 mmol/24 hour), of which D-lactate was present almost exclusively instead of the common L-lactate (Duran et al. 1977). At the age of one month, urine had been collected and frozen. Retrospective analysis of this sample also revealed elevated D-lactate, indicating that he already had the condition at an early age. Both lactate isomers were present in blood (L-lactate: 1.7 mM; D-lactate: 0.6 mM) and in cerebrospinal fluid (L-lactate: 1.5 mM; D-lactate: 0.6 mM)(Duran et al. 1977). Several gut sterilizations to remove bacterial overgrowth that could be responsible for the increased D-lactate did not correct the D-lactate excretion, nor did a dietary modification of reduced carbohydrates.

Genetic diagnosis of 11p deletion syndrome

Routine array CGH was performed on DNA isolated from the patient on clinical suspicion of 11p deletion syndrome at the Genome Diagnostics Department of Genetics, University Medical Center Utrecht, the Netherlands. A paternal deletion 3q24 (260 kb) was detected in the patient, as well as a 11.13Mb *de novo* deletion spanning from 11p14.1 to 11p12, confirming the 11p deletion syndrome. Upon reexamination at the age of 40, a single-nucleotide polymorphism SNP array was performed to establish the minimal breakpoints

chr11:29651299 to chr11:40817882 of the *de novo* deletion (GRCh37/hg19 genome build). Moreover, several large regions of homozygosity were seen, indicating relatedness of the parents. The largest homozygous stretch contained an 11.1 MB region with 51 genes, encompassing the 11p13 region including PAX6 and WT1. The patient's intellectual disability, ophthalmologic features, cryptorchidism and seizures were consistent with 11p deletion syndrome - also known as WAGR syndrome (Wilms tumor; Aniridia; Genital anomalies; Retardation) - which was concluded to be likely causal for most of the patient's phenotype (Almind et al. 2009; Narahara et al. 1984; Riccardi et al. 1978).

Metabolic reanalysis confirms elevated D-lactate urine and plasma levels

Upon reexamination at the age of 40, the patient still had a dramatically increased excretion of D-lactate (mean: 1686 mmol/mol creatinine) compared to controls (Figure 2a). Plasma analysis revealed elevated D-lactate (0.7 mM) compared to controls (Figure 2b). Furthermore, elevated concentrations of D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid were observed in both urine and plasma (Figures 2c-f). The inability of repeated antibiotic treatment to resolve the elevated D-lactate levels made a bacterial origin unlikely. Additionally, the continued presence of increased D-lactate levels after 36 years of monitoring suggests that these observations were not due to a temporary metabolic disturbance. As elevated levels of D-lactate have not been reported in patients with 11p deletion syndrome, the dysfunction of D-lactate metabolism was likely due to another cause.

Genetic investigation identifies a homozygous loss-of-function *LDHD* variant

Diagnostic Sanger sequencing of *PDHX*, a gene in which homozygous or compound heterozygous variants are known to cause lactic acidemia (MIM #245349), detected no causal variants (Department of Genetics, Genome Diagnostics, University Medical Center Groningen, the Netherlands). Subsequently, we analysed the regions of homozygosity for possible candidate genes whose dysfunction could be causal for D-lactate excretion (Supplemental Table 1). The fifth largest stretch of homozygosity of 3.5 MB on chromosome 16 contained 28 protein coding genes including *LDHD* (Supplemental Table 2). *LDHD* has been previously identified as a putative D-lactate dehydrogenase that could be responsible for mammalian D-lactate metabolism but whose function in humans or other vertebrates had not been established *in vivo* (de Bari et al. 2002; Flick and Konieczny 2002). We performed Sanger sequencing of the exonic regions and intronic/exonic boundaries of the DNA isolated from blood of the patient and identified a LDHD homozygous nonsynonymous

variant NM_153486.2:c.1388C>T, p.(Thr463Met) (Figure 3a). Subsequent segregation analysis revealed that both parents were heterozygous carriers of the variant. The variant's predicted effect on protein function was classified as probably damaging (PolyPhen-2) and deleterious (Sorting Intolerant from tolerant; SIFT) (Adzhubei et al. 2010; Kumar et al. 2009). The variant is in a region that is highly conserved across multiple species (Figure 3b) and not present in large human population variant frequency databases such as the Genome of the Netherlands (http://www.genoomvannederland.nl), the Exome Variant Server (EVS; http://evs.gs.washington.edu/EVS), the 1000 Genomes (http://www.1000genomes.org), dbSNP Build 148 for Human (http://www.ncbi.nlm.nih.gov/SNP), or our in-house dataset (Accessed April 25, 2017).

Additionally, the variant is not present homozygously in any individual in the gnomAD database (Genome Aggregation Database, http://gnomad.broadinstitute.org/gene/ENSG00000166816) but it is present heterozygously in 27 of the 137,892 individuals who have coverage at this position in this database (Minor Allele Frequency: 0.0000979; Accessed April 25, 2017). Sanger sequencing of 200 additional individuals from the island of Sicily was performed to investigate the possibility that the NM_153486.2:c.1388C>T variant may be present at a higher frequency in the isolated Sicilian population but not represented in the above databases. However, no other individuals carrying this *LDHD* variant were detected.

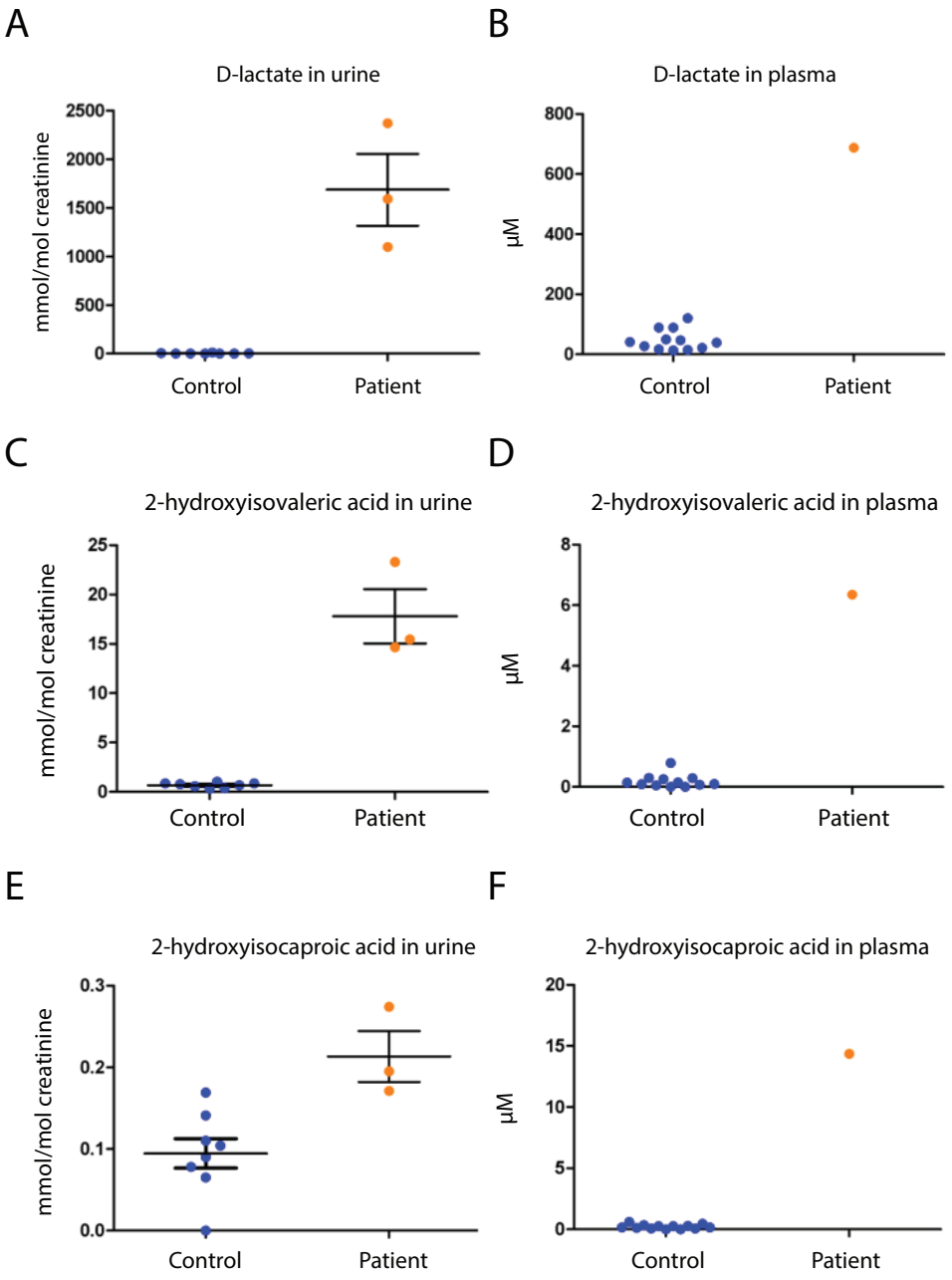


Figure 2 | Metabolic analysis reveals elevated D-lactate excretion in urine and increased D-lactate in plasma. The levels of D-lactate in A) urine and B) plasma are elevated in our patient compared to controls. Increased levels of D-2-hydroxyisovaleric acid (C-D) and D-2-hydroxyisocaproic acid (E-F) are also observed in urine and plasma, respectively.

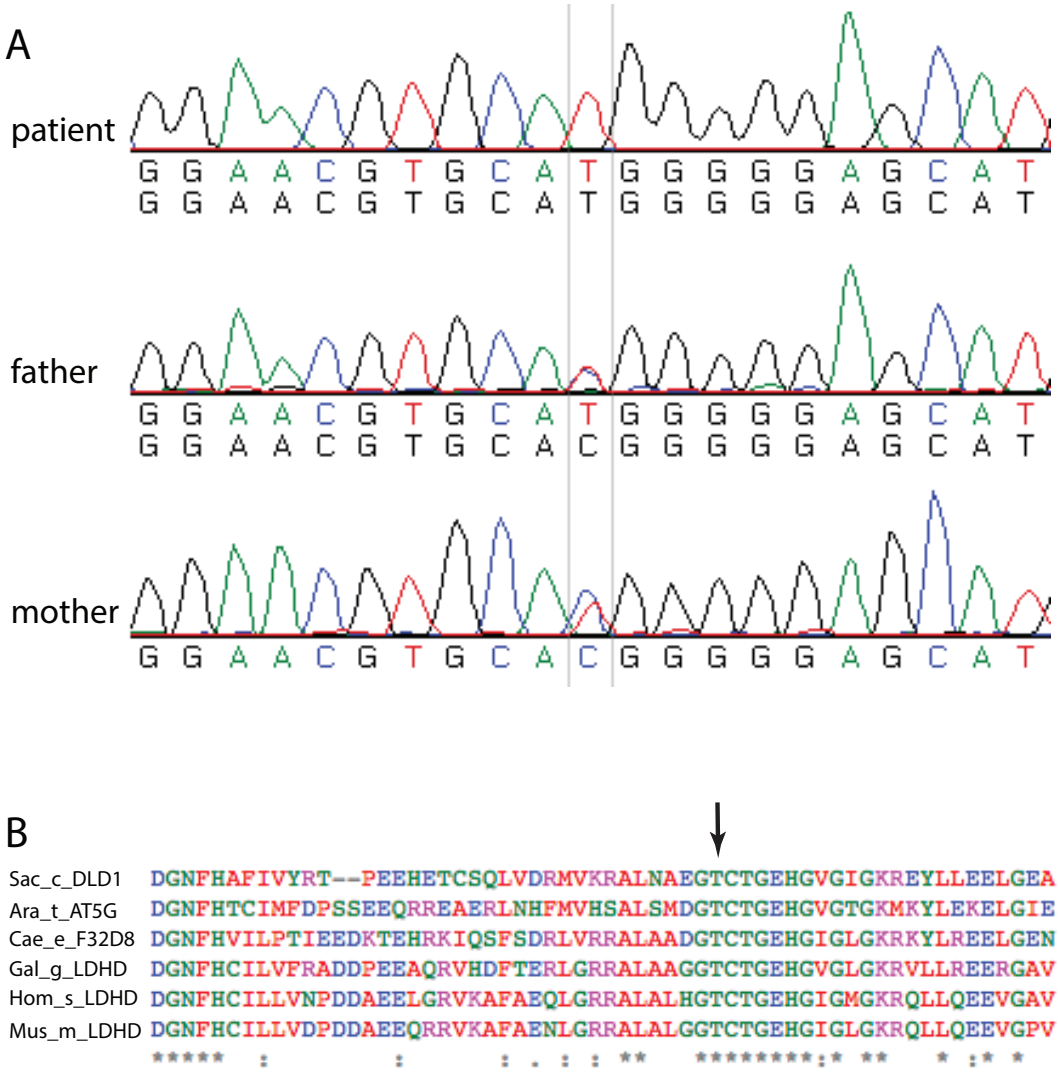


Figure 3 | Identification of a human knockout for *LDHD*. A) Sanger sequencing of the patient identifies a novel homozygous nonsynonymous variant NM_153486.2:c.1388C>T, p.(Thr463Met). Both the father and mother are heterozygous carriers of the variant. B) The variant encodes for the amino acid Methionine (M) instead of the normally present amino acid Threonine (T) in a region that is highly conserved across multiple species. (Abbreviations: Sac_c_DLD1: Saccharomyces_cerevisiae_DLD1; Ara_t_AT5G: Arabidopsis_thaliana_AT5G06580; Cae_e_F32D8: Caenorhabditis_elegans_F32D8.12; Gal_g_LDHD: Gallus_gallus_LDHD; Hom_s_LDHD: Homo_sapiens_LDHD; Mus_m_LDHD: Mus_musculus_LDHD)

The identification of a novel, homozygous *LDHD* variant in a highly conserved region prompted us to investigate LDHD function. Human and mouse *LDHD*/*Ldhd* have strong sequence similarity to yeast D-lactate dehydrogenases and LDHD is widely expressed in skeletal muscle, heart, liver and kidneys of humans (Flick and Konieczny 2002). The protein contains a N-terminal mitochondrial targeting sequence at amino acids 1-17, a predicted NAD+ binding site at amino acids 98-103 and a predicted FAD+ binding site at amino acids 171-181 (Flick and Konieczny 2002). Subcellular experiments have localized LDHD to the mitochondria in human embryonic kidney cells (HEK293); further experiments in rat livers have shown that it specifically localizes to the inner side of the inner membrane of mitochondria (de Bari et al. 2002; Flick and Konieczny 2002). In rat liver, three separate transporters stereospecifically move D-lactate from the cytosol into the mitochondria where it is oxidized to pyruvate (de Bari et al. 2002). In comparison to L-lactate dehydrogenase, the specificity of D-lactate dehydrogenase is evolutionary distinct and may have evolved specifically to remove the toxic aldehyde methylglyoxal (Cristescu et al. 2008; de Bari et al. 2002; Kochhar et al. 1992; Taguchi and Ohta 1991). However, the role of *LDHD* in human metabolism has not yet been identified *in vivo*.

Zebrafish ldhd^{sa15623} knockout line and rescue by human wild-type LDHD but not by variant LDHD

We utilized the zebrafish model organism to verify the effect of loss of function of *LDHD* and establish the role of LDHD in D-lactate metabolism. We acquired the *ldhdsa15623* line, through the European Zebrafish Resource Center (ZFIN ID: ZDB-GENE-030131-6140; EZRC at the Karlsruhe Institute of Technology), which carries a disrupted essential donor splice site following exon 3 (Allele *sa15623*; Zv9 chr25:g.13920446T>G) (Kettleborough et al. 2013). The modification at this splice site results in partial retention of the intron following exon 3 and the introduction of a premature stop codon. Since *ldhdsa15623* homozygous fish are adult viable and fertile, maternal and zygotic mutant embryos (henceforth referred to as ldhd-/-) could be obtained and were used throughout this study for phenotypical and metabolical evaluation for any ects due to ldhd deficiency. Phenotypically, 3dpf (days post fertilization) *ldhd*-/- embryos showed no abnormalities compared to wild-type (Figure 4a). However, metabolic analysis revealed elevated levels of D-lactate, D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid but not L-lactate in *ldhd*-/- larvae compared to wild-type. This confirms conservation of LDHD function in the zebrafish and demonstrates the biochemical phenotype of LDHD loss of function.

To study the effect of the missense variant present in our patient, we microinjected synthetic *LDHD* wild-type or *LDHD* Thr463Met variant mRNA into 1-cell embryos of either wild-type or *ldhd*-/- zebrafish. 3dpf embryos from the six different conditions were pooled

in separate groups of 15 embryos and metabolically investigated to evaluate D-lactate levels. Metabolic analysis of 3dpf wild-type zebrafish larvae showed physiological levels of D-lactate that were not altered by microinjections of the variant Thr463Met *LDHD* or wild-type *LDHD*, as the endogenous LDHD is sufficient to metabolize all of the D-lactate that is present in the developing embryo (Figure 4b). However, 3dpf *ldhd*-/- uninjected embryos showed a significant increase in D-lactate (mean: 27 μM). The increased D-lactate levels in *ldhd*-/- embryos were subsequently restored to baseline levels by mRNA microinjection of the wild-type *LDHD* sequence. In contrast, microinjection of mRNA of the variant Thr463Met *LDHD* did not rescue the metabolic phenotype, as the observed D-lactate levels remained similar to those of the uninjected *ldhd*-/- embryos (mean: 24 μM). This confirms that LDHD is responsible for D-lactate metabolism, and that the homozygous variant Thr463Met present in our patient results in loss of LDHD activity in zebrafish.

Additionally, uninjected *ldhd*-/- embryos and Thr463Met variant *LDHD*-injected *ldhd*-/- embryos showed increased levels of D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid, similar to our patient, suggesting an additional role for LDHD in the metabolism of these metabolites (Figure 4c and 4d). These metabolites were not observed at elevated levels in rescued *LDHD* wild-type-injected *ldhd*-/- embryos. Levels of L-lactate, L-2-hydroxyisovaleric acid and L-2-hydroxyisocaproic acid were not significantly altered among the different conditions, establishing that LDHD is a stereospecific enzyme for D-isomers and not solely D-lactate (Figure 4e, 4f, 4g).

L- and D-lactate production in human metabolism

L-Lactate is an essential metabolite in human metabolism, its formation being a means to generate NAD+ in times when oxygen is absent. In this reaction, pyruvate is reversibly reduced to L-lactate by the oxidation of NADH to NAD+ in a reaction performed by the ubiquitously expressed L-lactate dehydrogenases, including LDHA and LDHB (Gray et al. 2014; Sakai et al. 1987; Tsujibo et al. 1985). L-lactate thus is the main lactate isomer in human metabolism with D-lactate plasma physiological levels approximately 100 times lower (Hasegawa et al. 2003; Kruse et al. 2011; Talasniemi et al. 2008). Routine methods used to measure plasma lactate are based primarily on measuring the product of LDHL enzymatic action on L-lactate (Hollaar and Van der Laarse 1979; Seheult et al. 2016). Consequently, with most of these assays, even following a normal total lactate concentration result an increased anion gap may still be explained by the presence of elevated D-lactate. The detection of D-lactate requires specific assays based on bacterial LDHD or recently introduced mass spectrometry methods that can detect both isomers, both of which are not routinely applied (Henry et al. 2012; Marti et al. 1997; Scheijen et al. 2012).

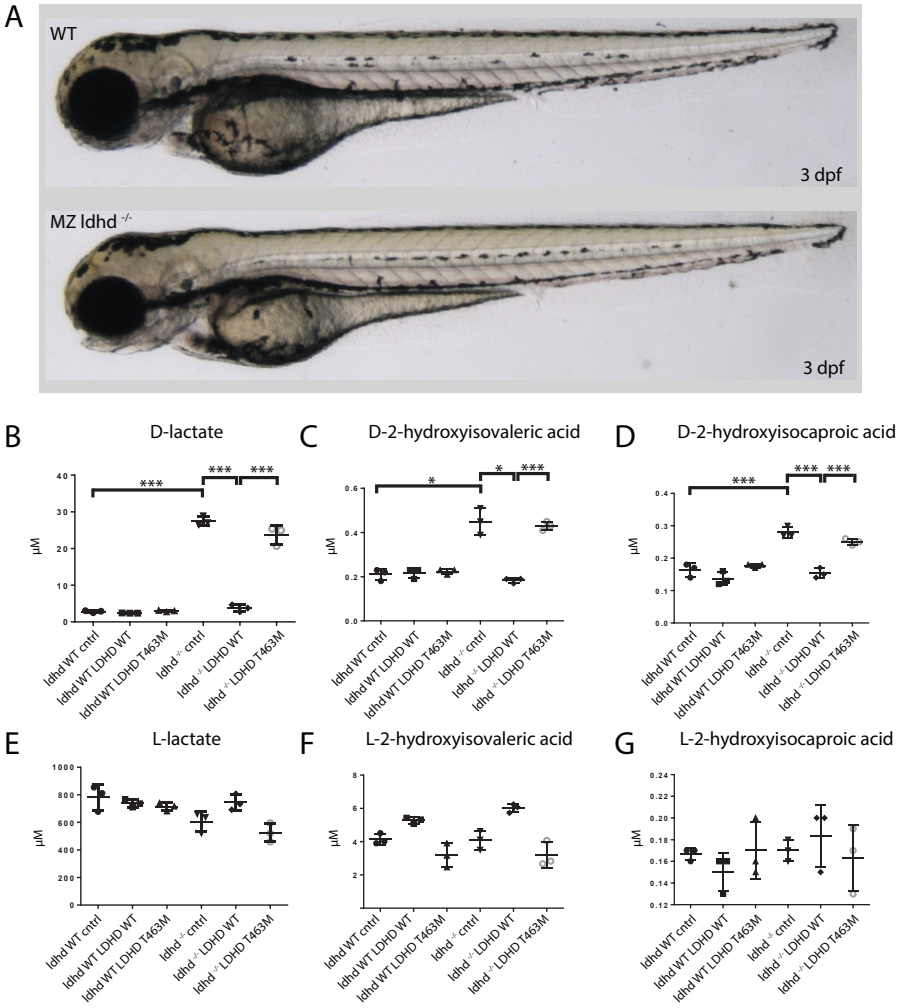


Figure 4 | Zebrafish metabolic studies. A) Wild-type zebrafish larvae (3dpf, upper panel) show no phenotype differences compared to maternal zygotic mutant *ldhd*-/- (lower panel). (B-G) Levels of D- and L- metabolites in response to LDHD activity. Wild-type zebrafish larvae *(ldhd* WT) and *ldhd*-/- zebrafish (*ldhd* -/-) at the 1 cell stage were subject to three conditions: uninjected (cntrl), microinjected with human wild-type *LDHD* RNA (*LDHD* WT), or microinjected with patient variant Thr463Met *LDHD* RNA (*LDHD* T463M). Metabolic assays were then performed to detect various metabolites present at 3dpf. Respectively, B) D- lactate, C) D-2-hydroxyisovaleric acid, D) D-2-hydroxyisocaproic acid, E) L-lactate, F) L-2-hydroxyisovaleric acid, and G) L-2-hydroxycaproic acid were assayed by mass spectrometry. The conditions where LDHD is nonfunctional results in statistically significant higher levels of the D-isomers of D-lactate, D-2-hydroxyisovaleric acid, and D-2-hydroxyisocaproic acid. No effect is seen in the L-isomer metabolites. Symbols: * is p-value < 0.05, ** is p-value < 0.01, *** is p-value < 0.005; all calculations by Students t-Test (two-sample unequal variance).

D-lactate metabolism and LDHD

The D-lactate ingested or produced in mammals is readily metabolized. Healthy individuals produce D-lactate as an end product of the methylglyoxal pathway. In this pathway, the highly reactive aldehyde methylglyoxal is detoxified to D-lactate in the cytosol of liver cells via the glyoxalase system (Kalapos 1999; Thornalley 1993). Humans and other mammals are efficient in clearing exogenously supplied D-lactate from the body (de Vrese et al. 1990; Giesecke et al. 1981; Oh et al. 1985). Humans metabolize around 90% of a D-lactate intravenous infusion of 1.0-1.3 meq/kg body weight/hr, eliminating only 10% of the D-lactate in their urine; this proportion decreases at higher infusion rates to 75% (Oh et al. 1985). The normal urine excretion is < 0.05 mmol/mol creatinine (Fabian et al. 2017).

In humans, this ability to metabolize D-lactate to pyruvate has previously been attributed to D-2-hydroxy acid dehydrogenase and mammals were thought to lack a functioning D-lactate dehydrogenase (Cammack 1969; Seheult et al. 2016; Tubbs 1965; Uribarri et al. 1998). The metabolism of D-lactate in lower order organisms is performed by isomer specific D-lactate dehydrogenases such as DLD1, AIP2, and YEL071W in Saccharomyces cerevisiae, and further D-lactate dehydrogenases have been identified in yeast and prokaryotes (Chelstowska et al. 1999; Flick and Konieczny 2002; Gleason and Nolan 1966; Olson and Massey 1979; Taguchi and Ohta 1991). Our identification of a human knockout of *LDHD* and resulting high levels of D-lactate in the urine and plasma indicates that LDHD has a central role in D-lactate metabolism and confirms *in vivo* that D-lactate is metabolized by LDHD, as previously proposed (de Bari et al. 2002; Flick and Konieczny 2002). This observation was confirmed in functional metabolic studies of zebrafish also demonstrating that loss of LDHD results in high physiological levels of D-lactate, an effect that could be rescued by wild-type *LDHD* but not by patient variant *LDH*D.

Neurological features of D-lactic acidosis patients in contrast to our patient

The first human cases of D-lactic acidosis were identified in the late 1970s, though D-lactic acidosis has long been known in ruminants in response to overfeeding (Dunlop and Hammond 1965; Duran et al. 1977; Oh et al. 1979). The current definition of D-lactic acidosis has been defined as D-lactate present at > 3mM in plasma. The majority of the cases identified thus far are secondary to short bowel syndrome (Uribarri et al. 1998). The number of these patients is estimated to be low due to underreporting, misdiagnosis, and attributing symptoms to other causes (Narula et al. 2000; Petersen 2005). In short bowel syndrome patients, repeated gut sterilizations often correct the D-lactic acidosis by reducing, or eliminating, the numbers of D-lactate producing bacteria. Additionally,

massive ingestion of 1,2-propylene glycol (a common food additive) has been linked to D-lactic acidosis (Christopher et al. 1990; Jorens et al. 2004).

Neurologic symptoms commonly identified in D-lactic acidosis patients include altered mental status, slurred speech, ataxia, inebriation appearance, abnormal gait and motor coordination, weakness, hostile or rude behavior, difficulty concentrating and memory deficits (Uribarri et al. 1998). It has been previously suggested that these neurologic symptoms may not be due to high concentrations of D-lactate, but instead may be caused by other specific metabolic products of the bacterial overgrowth. These products include neurotoxins such as mercaptans, aldehydes, alcohols, amines – several of which may function as false neurotransmitters (Halperin and Kamel 1996; Petersen 2005). Alternatively, it has been proposed that elevated D-lactate levels may alter brain cell metabolism of pyruvate and L-lactate and result in neurological symptoms (Ling et al. 2012; Petersen 2005). Monocarboxylate transporters, particularly MCT1, transport pyruvate and lactate into mitochondria; D-lactate, a less favorable substrate, can reduce L-lactate and pyruvate transport into the mitochondria by competitive inhibition (Brooks et al. 1999; Ling et al. 2012; McCullagh et al. 1996; Wang et al. 2006). Ling et al. (2012) demonstrated that this results in impaired transport of L-lactate and pyruvate into rat brain and heart cells, and postulated that this impaired transport of energy substrates can impair mitochondrial energy production and result in neurological effects (Ling et al. 2012).

Our patient presented with intellectual disability with behavioral problems as a feature of 11p deletion syndrome. In a clinical review of 54 WAGR cases, 72% of patients had intellectual deficit, with a smaller percentage of these patients presenting with diverse behavioral symptoms (Fischbach et al. 2005). One might speculate that the perturbed D-lactate metabolism in our patient has aggravated the severity of the intellectual deficit or behavioral problems caused by the 11p-deletion, but this can only be confirmed by finding other cases without alternative causes of mental disorder. Interestingly, in the human genetic variation database ExAC that includes 60,706 individual exomes, there is one other individual with a homozygous *LDHD* predicted loss of function variant: NM_153486.2:c.669G>A, p.(Trp223Ter). This database contains only individuals without severe pediatric disease; however, phenotypes are not publically available and we cannot rule out the possibility of this patient having neurological symptoms. If only a biochemical phenotype is associated with loss-of-function *LDHD*, it is possible that this metabolic phenotype is an under-recognized “condition”. Further research, or identification of a confirmed healthy individual with loss of function of *LDHD*, is needed to address this question.

Elevated levels of D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid

In addition to increased D-lactate levels, increased levels of D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid were observed in both our patient and ldhd-/- zebrafish larvae, suggesting a role for LDHD in the metabolism of these D-isomer metabolites as well. The origin of these metabolites in vertebrate metabolism in unknown. Increased levels of the L-forms of these metabolites have been reported in the urine of patients with Maple Syrup Urine Disease (MSUD, MIM #248600) (Jakobs et al. 1977; Tanaka et al. 1980), lactic acidosis, ketoacidosis, as well as diabetic ketoacidosis patients (Landaas and Jakobs 1977; Liebich and Forst 1984). The lactate dehydrogenase LDHA has experimentally been shown to be involved in the formation of the L-forms of these metabolites, particularly L-2-hydroxyisovaleric acid levels (Heemskerk et al. 2016). Heemskerk et al. (2016) show that LDHA can convert the ketoacid transamination products of valine, leucine, and isoleucine into L-2-hydroxyisovaleric acid, L-2-hydroxyisocaproic acid, and L-2-hydroxy-3-methylvaleric acid, respectively, and hypothesize that this function is necessary to prevent the accumulation of branched-chain ketoacids in hypoxic conditions. The increases of D-2-hydroxyisovaleric acid and D-2-hydroxyisocaproic acid that we have observed could suggest a role for LDHD in the metabolism of these D-isomer metabolites, but more research is needed to clarify the mechanism of this little known pathway. An intestinal bacterial origin of the D-2-branched-chain hydroxyacids cannot be ruled out completely. Spaapen et al. (1987) have demonstrated the presence of excessive amounts of D-2-hydroxyisocaproic acid in the urine of a patient with a short bowel syndrome, together with the typical D-lactate and, amongst others, the D-isomers of phenyllactic acid and 4-hydroxyphenyllactic acid which are metabolites of the amino acids phenylalanine and tyrosine, respectively (Spaapen et al. 1987). Their findings suggest that the formation of D-2-hydroxyacids from parent amino acids is a common intestinal bacterial phenomenon. All thus formed D-2-hydroxyacids appear to be metabolized by endogenous LDHD, which has a limited overall capacity. In short bowel patients this limited capacity may be overwhelmed by excessive bacterial metabolism and result in accumulation of these D-2-branched-chain hydroxyacids. Similarly, in our patients, the accumulation of these hydroxyacids may be a result of LDHD loss-of-function.

CONCLUSION

In this report, we show a patient with a homozygous *LDHD* missense variant resulting in loss of function that leads to a massive excretion of D-lactate in the urine as well as increased plasma concentration. Functional studies in *ldhd*-/- zebrafish establish that loss of LDHD activity results in increased D-lactate concentration that can be subsequently reduced with the introduction of wild-type *LDHD* but not by introduction of *LDHD* with the Thr463Met variant present in our patient. No other phenotype was observed in *ldhd*-/- zebrafish. We conclude that humans do have endogenous D-lactate activity other than D-2-hydroxy acid dehydrogenase, and that LDHD is essential for D-lactate metabolism in humans. Whether LDHD deficiency is a biochemical anomaly or whether a clinical phenotype is associated with it remains to be elucidated. This is the first *in vivo* identification of the role of LDHD in human D-lactate metabolism and demonstrates the value of human knockouts in revealing and characterizing gene function.

MATERIALS AND METHODS

Clinical case report

Our patient, originally described 40 years ago as a novel case of D-lactic aciduria was examined in the course of clinical reevaluation (Duran et al. 1977). The patient was born of Sicilian parents and presented to the clinic at the age of 4 years. Briefly, he had been born of an uneventful pregnancy at 40 weeks, with microcephaly (OFC 34.5 cm), slanting of the eyelids, bilateral inguinal hernia, and aniridia. Development quotient at the age of 1 year was 51 at Griffiths scale and motor development was also severely impaired.

At the age of 40, the family presented again for genetic counseling. By then he was known to carry a de novo deletion on chromosome 11p13. His medical history included a severe mental handicap with behavioral problems, cryptorchidism, blindness (aniridia, with later onset of cataract and glaucoma) and epilepsy until the age of 14. He had not developed a Wilms tumor.

Upon reexamination, he was normocephalic (OFC 57 cm), with down slanting eyelids, a protruding lower lip, mildly dysplastic helices and patches of greying hair. The second digits of his feet were longer than the halluces.

Follow-up studies were performed to investigate the breakpoints of the deletion.

Genetic Investigation

SNP array

A high-resolution SNP array was performed to establish the minimal regions of the 11p deletion and identify other regions of homozygosity (Supplemental Table 1). Genes present in the identified regions were annotated on function and literature consulted to establish if the gene could be a candidate for elevated D-lactate levels. Of the genes that were present in homozygous regions, *LDHD* was the most likely candidate as it had previously been identified as a putative D-lactate dehydrogenase (Supplemental Table 2) (de Bari et al. 2002; Flick and Konieczny 2002).

Sanger Sequencing of Candidate Gene in patient and Sicilian cohort

The exonic and exonic/intronic boundaries of *LDHD* were screened by Sanger sequencing (Primers available in Supplemental Table 3), identifying the variant NM_153486.2:c.1388C>T,

p.(Thr463Met). To determine if this variant was population specific in the isolated Sicilian community of the family, 200 additional individuals were screened for variants using Sanger sequencing. One hundred of these individuals originated from within a 100 km radius of the town of patient's birth in the province of Palermo; a further hundred individuals were from neighboring Sicilian areas outside of the 100 km radius.

Patient Metabolic Investigation

Patient metabolic measurements were initially measured as described by Duran et al. (1977) (Duran et al. 1977). Upon reexamination, patient metabolic measurements to detect D-lactate were performed as described below based upon Scheijen et al. (2012) with minor modifications and extensions of additional components (Scheijen et al. 2012).

Chemicals and Reagents

Chemicals and reagents were obtained from various suppliers. Sodium L (+) lactate, sodium D (-) lactate, D-2-hydroxyisovaleric acid, L-2-hydroxyisovaleric acid, L-2-hydroxyisocaproic acid, (+)-O,O -diacetyl-L-tartaric anhydride (DATAN) and ammonium formate were obtained from Sigma Aldrich (St. Louis, Missouri, USA). [13C3]-Sodium L (+) lactate was obtained from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Dichloromethane and acetic acid anhydrous were obtained from Merck (Kenilworth, NJ, USA). Acetonitril, methanol and formic acid (all at ULS/MS grade) were obtained from Biosolve (Dieuze, France). D-2-hydroxyisocaproic acid was obtained from Bachem AG (Bubendorf, Switzerland).

Plasma Sample Preparation

Twenty-five µL of plasma was added to 25 µL of internal standard solution (containing 434.75 µmol/L [13C3]-L-lactate). Samples were mixed thoroughly and subsequently deproteinized with a 600 µL mixture of methanol:acetonitrile (1:1, by volume) and centrifuged at 4°C for 5 minutes at 13000 rpm. The supernatant was pipetted into a reaction vial and evaporated completely under a gentle stream of nitrogen at a temperature of 50°C. Fifty µL of freshly made DATAN (50 mg/mL dichloromethane : acetic acid (4:1, by volume)) was added. The vial was capped, vortexed, and heated at 75°C for 30 minutes. After 30 minutes, the vial was allowed to cool down to room temperature, and the mixture was evaporated completely with a gentle stream of nitrogen. The derivatized residue was reconstituted with 150 µL Solvent A (1.5 mM ammonium formate (pH = 3.6)).

Urine Sample Preparation

Twenty-five µL of internal standard solution, 25 µL urine, and 300 µL of methanol were pipetted into a reaction vial. Samples were mixed thoroughly and evaporated completely under a gentle stream of nitrogen at a temperature of 50°C. Fifty µL of freshly made DATAN was added. The vial was capped, vortexed, and heated at 75°C for 30 minutes. After 30 minutes, the vial was allowed to cool down to room temperature, and the mixture was evaporated completely with a gentle stream of nitrogen. The derivatized residue was reconstituted with 300 µL Solvent A.

Chromatographic conditions

Samples were analysed by reversed phase LC-tandem MS using an Acquity UPLC BEH C18 analytical column (100 × 2.1 mm, 1.7 µm; Waters, Milford, MA, USA)(UPLC: ultra performance liquid chromatography). Detection was carried out using a Xevo TQ tandem mass spectrometer (Waters, Milford, MA, USA), which was operated in negative multiple-reaction-monitoring (MRM) mode. UPLC analysis was performed using a binary gradient at a flow of 0.5 mL/min using an Acquity UPLC (Waters, Milford, MA, USA). Solvent A was 1.5 mM ammonium formate (pH = 3.6), and solvent B was acetonitrile. A linear gradient was started at 0.5% B, and ramped to 3% B in 3 minutes, further ramped to 40 % in 2 minutes, held at 40 % B for 2 minutes and returned to 0.5 % B in 1 minute. The column was equilibrated for 1 minute at the initial composition. Injection volume was 10 µL, and column temperature was set at 40°C. Samples were kept at 6°C. Chromatograms were acquired and processed with Masslynx V4.1 SCN 843 (Waters, Milford, MA, USA).

Mass Spectrometry Conditions

Optimal conditions for all parents were found at a capillary voltage of 1.5 kV and a cone voltage of 10 V. The source and desolvation temperature were 150°C and 600°C, respectively. The cone gas flow and desolvation gas flow were 0 and 800 l/hour, respectively. To establish the most sensitive daughter ions, the collision energy was set at 8 eV with a collision gas flow of 0.15 mL/min (Table 1).

Component	Parent ion (m/z)	Daughter ion (m/z)	Collision energy (eV)	Dwell time (secs)
[13C3]-L-lactate	307.95	91.95	8	0.1
L/D-lactate	304.95	88.95	8	0.1
L/D-2-OH-isovaleric acid	333	117	8	0.1
L/D-2-OH-isocaproic acid	347	131	8	0.1

Calibration standard curves for all compounds were made in Milli-Q water. Calibration curves were obtained by linear regression of a plot of the analyte concentration versus the peak-area ratio of the analyte/internal standard area. For all the analytes, [13C3]-L-lactate was used as internal standard.

Zebrafish functional studies

Zebrafish ldhd$^{sa15623-/-}$ line

The ENU-mutagenized LDHD knockout zebrafish line *ldhd*$^{sa15623-/+}$ was obtained via EZRC (ZFIN ID: ZDB-GENE-030131-6140) to evaluate the D-lactate levels compared to wild-type zebrafish (ENU: 1-Ethyl-1-nitrosourea). The line contains the variant chr25:g.13920446T>G (Zv9 zebrafish genome build) that disrupts an essential splice site following exon 3, resulting in a knockout of LDHD (http://www.sanger.ac.uk/sanger/Zebrafish_Zmpgene/ENSDARG00000038845#sa15623). The heterozygous carriers were grown to adults and in-crossed to produce the *ldhd*-/- line, which is adult viable and fertile. In-crossing the *ldhd*-/- line provided the maternal and zygotic mutant embryos for this study.

Construct preparation and zebrafish microinjection

Expression constructs were created using the normal human sequence of *LDHD* wild-type and the *LDHD* Thr463Met variant to evaluate if the D-lactate phenotype could be introduced and/or subsequently rescued.

Human *LDHD* cDNA (accession number NM_194436.1) in the entry vector pCMV-ENTRY (Origene, Rockville, Maryland, USA) was inserted into the expression vector pCS2+/GW. Site-directed mutagenesis was performed to introduce the T463M variant by using primers 5-TCTCCACGGAACGTGCATGGGGGAGCA-3 and 5-TGCTCCCCCATGCACGTTCCGTGGAGA-3. These two plasmids were then linearized by NotI digestion and the reaction was cleaned up using the Qiaquick PCR purification kit (Qiagen, Hilden, Germany). RNA was synthesized from the linear DNA using the mMESSAGE mMACHINE SP6 transcription kit (Thermo Fisher Scientific, Waltham, MA, USA) and cleaned with the RNEasy mini kit (Qiagen, Hilden, Germany). 200 ng/ul mRNA of the *LDHD* wild-type or T463M *LDHD* variant was used for microinjections into 1-cell stage *ldhd*-/- or wild-type zebrafish embryos. A control group of uninjected wild-type zebrafish and *ldhd*-/- zebrafish was also included. At 3dpf, three batches of 15 embryos were collected for each of the six different conditions, pooled and stored at -80.

Zebrafish metabolic studies

Two-hundred µl methanol and 25 µL of internal standard solution (containing 434.75 µmol/L [13C3]-L-lactate) was added to the separate three batches for each of the six conditions. Next, sample extracts were homogenized with 0.5 mm zirconium oxide beads (product #ZrOB05, Next Advance, Inc., Averill Park, NY) using a bullet blender (model BBX24B-

CE, Next Advance, Inc., Averill Park, NY) for 2x5 min, speed 8 at 4°C. The extracts were centrifuged at 4 °C and 13,000 rpm for 5 min. The supernatant was pipetted into a reaction vial and evaporated completely under a gentle stream of nitrogen at a temperature of 50°C. Fifty µL of freshly made DATAN was added. The vial was capped, vortexed, and heated at 75°C for 30 minutes. After 30 minutes, the vial was allowed to cool down to room temperature, and the mixture was evaporated completely with a gentle stream of nitrogen. The derivatized residue was reconstituted with 150 µL Solvent A and chromatography and mass spectrometry were performed as described above.

Acknowledgements

We would like to thank the patient and his family for their participation in this study. We also wish to thank Dr. Mies M. van Genderen, the patient's ophthalmologist at Bartimeus, Institute for the Visually Impaired, Doorn, the Netherlands. AMvE is supported by The Dutch Kidney Foundation (KSTP12-010).

Supplemental Table S1 | Regions of homozygosity in the patient greater than 1 Mb. Genome build GRCh37/hg19.

Chromosome	Start Position	End Position	Size (base pair)
13	85802402	98498677	12696275
11	29651299	40817882	11166583
12	34512767	38651375	4138608
17	21770128	25568707	3798579
16	74873927	78363380	3489453
5	71489305	74242852	2753547
10	45556652	47593133	2036481
2	197129767	198962243	1832476
6	62415600	63936511	1520911
15	43344886	44847415	1502529
3	74479445	75974251	1494806
1	28211650	29665936	1454286
19	42001210	43372386	1371176
2	116014077	117364078	1350001
5	139188856	140519095	1330239
4	97935617	99214851	1279234
2	186218867	187481835	1262968
7	73909796	75166580	1256784
14	105904830	107146692	1241862
8	50265730	51469405	1203675
18	39796976	40996805	1199829
5	36716862	37892952	1176090
6	145800915	146972068	1171153
7	97136298	98233514	1097216
11	112040136	113133676	1093540
6	114860722	115944735	1084013
3	81988359	83039277	1050918
2	97595991	98629966	1033975
4	110000675	111030787	1030112
8	85474803	86493592	1018789

Supplemental Table S2 | Genes present in the regions of homozygosity of the patient that are greater than 1 MB. Genome build GRCh37/hg19.

Chromosome	Start position	Stop position	Gene
16	74907468	75034071	WDR59
16	75032928	75144892	ZNRF1
16	75145758	75150669	LDHD
16	75182390	75206134	ZFP1
16	75237994	75241083	CTRB2
16	75252898	75258822	CTRB1
16	75262928	75301951	BCAR1
16	75327596	75467383	CFDP1
16	75446582	75498604	RP11-77K12.1
16	75476952	75499395	TMEM170A
16	75510949	75529282	CHST6
16	75562430	75579326	RP11-77K12.7
16	75562433	75569145	CHST5
16	75572015	75590176	TMEM231
16	75600249	75611779	GABARAPL2
16	75630879	75657198	ADAT1
16	75661622	75682541	KARS
16	75681684	75795051	TERF2IP
16	75728367	75734089	AC025287.1
16	76311176	76593135	CNTNAP4
16	76587314	76669520	RP11-58C22.1
16	77224732	77236302	MON1B
16	77233294	77247112	SYCE1L
16	77281710	77469011	ADAMTS18
16	77756411	77776157	NUDT7
16	77822427	78014004	VAT1L
16	78056412	78100658	CLEC3A
16	78133310	79246564	WWOX

Supplemental Table S3 | Primer sequences for Sanger sequencing of *LDHD* exonic and intron/exon splice regions.

Primer name	Sequence
LDHD_1F	TGAACCTCGCCTCTCCTTTA
LDHD_2F	TGGAAAAGAGGCATTTCTGG
LDHD_3F	CTCGCTTGCTGTAGCCTCA
LDHD_4F	GATGGTCCCAGCACACCTT
LDHD_5F	GGTTCAGGGGACCTTCCTC
LDHD_6F	TTGTGGAGCAGAAAGCACTG
LDHD_7F	CATTGGTGAGCTGAGGATGG
LDHD_8F	AGGTATGCTGGGGTGGAGT
LDHD_9F	GACCCATCCTCAGCCTTGA
LDHD_10F	GAGAGGTGGGGAGGTGTCA
LDHD_11F	CTGCTGCAGGAGGAGGTG
LDHD_12F	CAGCGAGCCCACTGTATCTG
LDHD_1R	CCAAGGTCAGGTCACTTGGT
LDHD_2R	TGAGTGAGGAGGAAGGCAAC
LDHD_3R	ATTCGGTCCATATGCGTCA
LDHD_4R	CGTACAGAAGGCAGCCTCA
LDHD_5R	GAAGCCTAGCCTGCCAAAC
LDHD_6R	CAGGCATCCATCATGACTTC
LDHD_7R	GTGCCGTGCTGTCCAAAG
LDHD_8R	CTGCAGCCCTTGTTTCTAGC
LDHD_9R	GTTCTGCAAAAGCCTTGACC
LDHD_10R	GTCAGGGAACTTGTGGGCTA
LDHD_11R	ACCAGGTGAAGGGGGAAG
LDHD_12R	CCTCCTCCCCAGGCTATAAG

7

Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability

Genetics in Medicine. 2016 Sep;18(9):949-56. Epub 2016 Feb 4.

PMID: 26845106; DOI:10.1038/gim.2015.200

Glen R. Monroe[1,2*], Gerardus W. Frederix[3*], Sanne M.C. Savelberg[1,2], Tamar I. de Vries[1], Karen J. Duran[1,2], Jasper J. van der Smagt[1], Paulien A. Terhal[1], Peter M. van Hasselt[4], Hester Y. Kroes[1], Nanda M. Verhoeven-Duif[1,2], Isaäc J. Nijman[1,2], Ellen C. Carbo[1,2], Koen L. van Gassen[1], Nine V. Knoers[1,2], Anke M. Hövels[3], Mieke M. van Haelst[1], Gepke Visser[4**] and Gijs van Haaften[1,2**]

1 | Department of Genetics, Utrecht University Medical Center, Utrecht, The Netherlands
2 | Center for Molecular Medicine, Utrecht University Medical Center, Utrecht, The Netherlands
3 | Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
4 | Department of Metabolic Diseases, Wilhelmina Children's Hospital, Utrecht, The Netherlands
*Joint First authors; **Joint Last authors.

ABSTRACT

This study investigated whole-exome sequencing (WES) yield in a subset of intellectually disabled patients referred to our clinical diagnostic center and calculated the total costs of these patients' diagnostic trajectory in order to evaluate early WES implementation. We compared 17 patients' trio-WES yield with the retrospective costs of diagnostic procedures by comprehensively examining patient records and collecting resource use information for each patient, beginning with patient admittance and concluding with WES initiation. We calculated cost savings using scenario analyses to evaluate the costs replaced by WES when used as a first diagnostic tool. WES resulted in diagnostically useful outcomes in 29.4% of patients. The entire traditional diagnostic trajectory average cost was $16,409 per patient, substantially higher than the $3,972 trio-WES cost. WES resulted in average cost savings of $3,547 for genetic and metabolic investigations in diagnosed patients and $1,727 for genetic investigations in undiagnosed patients. The increased causal variant detection yield by WES and the relatively high costs of the entire traditional diagnostic trajectory suggest that early implementation of WES is a relevant and cost-efficient option in patient diagnostics. This information is crucial for centers considering implementation of WES and serves as input for future value-based research into diagnostics.

INTRODUCTION

The advent of next-generation sequencing—first in the research environment and recently in diagnostic laboratories—provides a powerful tool to interrogate the exome, or even the entire genome, without *a priori* knowledge in pediatric patients with a disorder that has a suspected underlying genetic cause (Choi et al. 2009; Ng et al. 2010b). The burden of genetic disease in child health is increasingly recognized, with an estimated 34% of inpatient children admissions having a clear genetic underlying cause, of which a significant and expanding proportion are identified as Mendelian disorders (Hall et al. 1978; McCandless et al. 2004). Whole-exome sequencing (WES) enables examination of the coding part of the genome for variants in genes linked to Mendelian disorders. Trio-based WES of patients with suspected rare, monogenic disorders and their healthy parents has previously been demonstrated as a particularly effective method to find the causal variant in cases suspected to be caused by a *de novo* variant or due to autosomal recessive inheritance. WES gives value to parents by allowing timely interventions and altered management and providing information necessary to make reproductive choices (Dixon-Salazar et al. 2012; Farwell et al. 2015; Iglesias et al. 2014; Lee et al. 2014; Srivastava et al. 2014). WES has a diagnostic advantage in situations where conventional single-gene or gene-panel tests may not be appropriate because a relevant genetic test has not yet been developed or because of genetic heterogeneity, incomplete or atypical clinical presentation, or lack of knowledge of the causal gene. In a research setting at the Sylvia Tóth Center (STC; Utrecht, The Netherlands), we performed trio-WES in a particularly difficult-to-diagnose patient population that could especially benefit from this emerging technology. The STC is a multidisciplinary center that specializes in diagnosing children with intellectual disability, and a large proportion of the diagnoses are attributed to genetic causes (Engbers et al. 2008).

Prior to admittance to the STC, patients are traditionally evaluated at peripheral hospitals in an iterative approach in which subsequent tests are introduced after initial diagnostic tests have negative results. This approach is effective in solving cases with an easily recognizable etiology, but other patients must undergo a “diagnostic odyssey,” subjected to numerous hospitalizations, diagnostic tests, and procedures over the course of many years— possibly with no diagnosis. In the Netherlands, these patients eventually end up in tertiary-level diagnostic facilities such as the STC.

The patients referred to the STC with intellectual disability seem highly suitable for a WES-based diagnostic approach because single-gene disorders account for at least one-quarter of intellectual-disability cases (Stromme 2000). The present study compared the effectiveness of WES with that of traditional diagnostic investigations. We provide

an extensive overview of the costs of the traditional diagnostic trajectory in this patient group. We also examine the cost savings that early implementation of WES would enable by rendering various genetic or metabolic tests unnecessary. The findings of this research can be used as input in the discussion regarding the societal, individual, and monetary value of next-generation sequencing (Phillips et al. 2014). A recent study assessed comprehensive costs of the traditional diagnostic pathway in a different group of patients, with complex pediatric neurological disorders (van Nimwegen et al. 2015). However, a clear determination of the diagnostic yield when actually performing WES compared with traditional diagnostic procedures and costs was not made. Reliable information on all resources used in the current clinical and diagnostic pathway is crucial for clinical diagnostic centers considering implementing WES as a diagnostic tool. In addition, this information could serve as input for future value-based research for the diagnostic trajectory.

In this study, we applied patient–parent trio-WES to 17 unsolved cases to identify the genetic cause of the patients' disorders and concurrently assessed the retrospective costs of diagnostic investigation for these patients using traditional methods. In 29.4% of the cases, WES analysis detected variants in genes recently discovered to cause Mendelian disorders or in genes mutated in patients with similar phenotypes. The estimated cost of WES is substantially lower than the mean cost ($16,409) of the traditional diagnostic trajectory for these patients. Early WES implementation would replace genetic and metabolic tests in patients who receive a diagnosis and replacement of genetic tests in patients who do not receive a diagnosis, resulting in cost savings of $3,547 and $1,727, respectively. This information suggests that WES should be considered a valid first tool in diagnostics in clinical diagnostic centers.

MATERIALS AND METHODS

Patient selection

Approximately 90 patients per year are seen at the STC in comprehensive one-day visits by a multidisciplinary team consisting of a clinical geneticist, a pediatric neurologist, a metabolic pediatrician, a physiotherapist, a psychologist, an ophthalmologist, a radiologist, and, on request, a psychiatrist and a dermatologist. From the 86 patients in the 2011 population, a subset of patients was selected that was suitable for WES based on the following criteria: the patients were born to healthy, unaffected parents; the parents were self-reported as nonconsanguineous; both parents could be contacted and were able to give consent; of genetic tests in patients who do not receive a diagnosis, and the patient was undiagnosed at the time of the study. Seventeen patients were then randomly

enrolled for inclusion in this study (Figure 1). All patients were intellectually disabled and as a cohort had wide and diverse phenotypes (Table 1). The parents of the patients signed informed consent for WES.

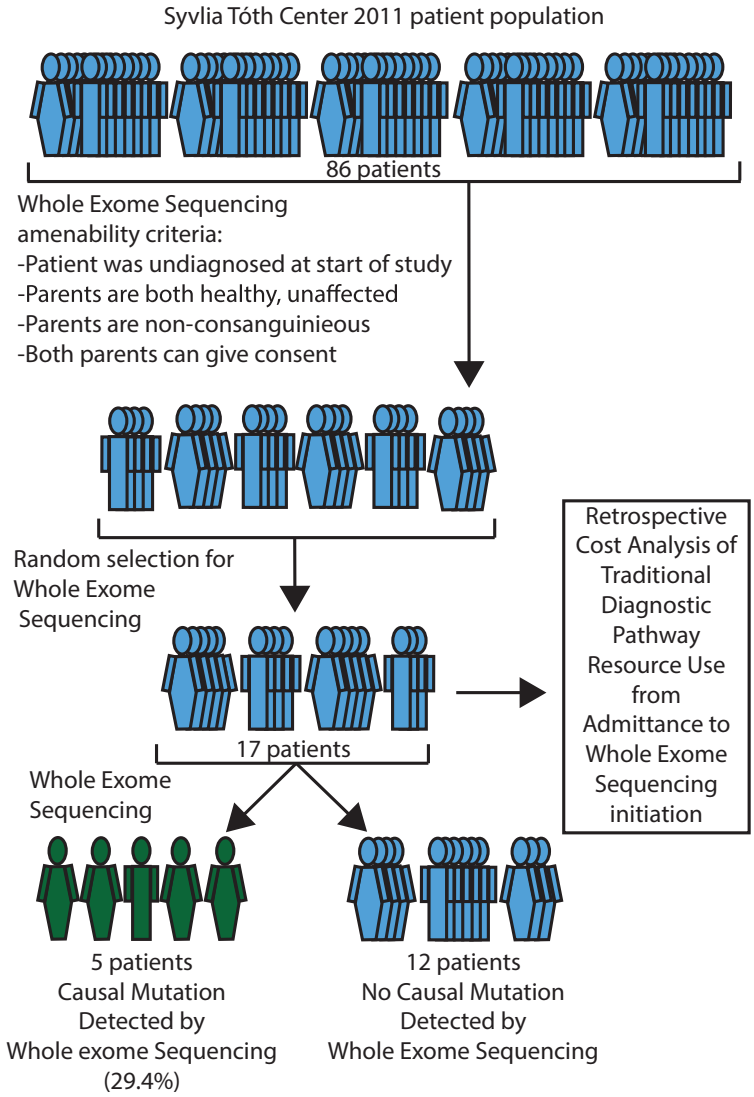


Figure 1 | Study outline. Of the 86 patients of the Sylvia Tóth Center 2011 patient population, WES was performed in a 17-patient subset. Of these 17 patients, the causal variant was detected in 5 (29.4%). Concurrently, a comprehensive retrospective cost analysis of the patients' individual traditional diagnostic pathway was performed.

Next-generation sequencing

Parent–patient trio-WES was performed with an emphasis on *de novo* variant detection analysis for the causal variants. DNA libraries for the 17 patients and their parents were prepared using Kapa Biosystems reagents (Kapa Biosystems, Wilmington, MA), enriched using Agilent Sureselect All Exon V5 (Agilent, Santa Clara, CA) and a custom pooling protocol (Supplementary Materials and Methods), and sequenced on a HiSeq 2500 (Illumina, San Diego, CA). All trios were sequenced to a desired minimal internal quality standard of 75× mean bait coverage for the patient, 65× mean bait coverage for the parent, and the percentage target bases over 20× >85%. Quality was often higher than this minimum requirement, with a mean coverage of 93× for the patients, 92× for the parents, and 90% for target bases over 20× (Supplementary Figure S1). Variant call files were imported into Cartagenia Bench Lab (https://cartagenia. com/) for additional variant interpretation, as well as into Combined Annotation Dependent Depletion (http:// cadd. gs.washington.edu/) for variant prioritization (Kircher et al. 2014). Emphasis in analysis using Cartagenia was placed initially on an anticipated de novo inheritance of a deleterious variant, but autosomal recessive, X-linked recessive, and compound heterozygote inheritance were subsequently investigated as well. Using the Cartagenia tool, de novo analysis was performed by filtering for Mendelian violations and prioritizing the variants using a classification tree by their presence in public population frequency databases, clinical variant databases, or our in-house data set (Supplementary Materials and Methods). Variants were further prioritized by amino acid change or splicing effect, evolutionary conservation score (genomic evolutionary rate profiling; GERP), and predicted effect on protein function using prediction algorithms (Polyphen-2, SIFT), and relevant literature (Adzhubei et al. 2010; Cooper et al. 2010; Kumar et al. 2009). Variants were initially validated and segregation analysis performed by Sanger sequencing in a research setting and subsequently validated, along with segregation analysis, in the Genome Diagnostics section at the University Medical Centre (UMC) Utrecht. Primer information is available upon request.

Table 1 | Patient information, length of diagnostic odyssey, and causal variant if identified.

Patient #	Phenotype	Sex	Age at admittance (years)	Causal Variant	Diagnostic Trajectory Duration (years)
1	Intellectual deficit, severe hypotonia, joint hypermobility, and obesity	Female	1.1	Not identified	6.5
2	Microcephaly, developmental delay, feeding problems and severe hypotonia	Female	0.8	*CTNNB1* – NM_001904.3:c.1925_1926delAG, p.(Glu642Valfs*5)	14.2
3	Microcephaly (-2.5 SD), an atrial septal defect, patent ductus arteriosus, submucous cleft palate with bifid uvula, dysmorphic features including epicanthic folds, and a small nose and mouth	Female	0.1	Not identified	5.7
4	Prenatal short stature (-3 SD), Intellectual deficit, dysmorphic facial features, obesity, radio ulnar synostosis, and hearing loss	Female	2.2	*ANKRD11* – NM_001256182.1: c.3382_3383delGA, p.(Asp1128Glnfs*41)	8.6
5	Intellectual deficit, microcephaly, short stature, epilepsy, brain anomalies, simian creases, and severe hypotonia	Male	0.0	Not identified	4.4
6	Apparently moderate nonsyndromic intellectual disability	Male	4.7	*ADNP* – NM_015339.2:c.2496_2499delTAAA, p.(Asn832Lysfs*81)	3.3
7	Macrocephaly (+2.5 SD), a prominent forehead with nevus flammeus, psychomotor retardation, and seizures. Asymmetry of the brain ventricles, asymmetric lesions in periventricular white matter and basal ganglia, polymicrogyria of perisylvian regions, cerebral atrophy and a relatively small vermis cerebelli	Female	0.8	Not identified	4.7
8	Severe syndromic intellectual disability	Female	0.6	*SMARCB1* NM_003073.3: c.1113C>G, p.(Asn371Lys)	5.8
9	Severe syndromic intellectual disability	Male	0.0	Not identified	4.4
10	Born preterm (GA 36) with an omphalocele. Apgar score of 2/7, and development delayed from birth. Cyclic neutropenia and trombopathy caused by a storage pool deficiency. Diminished fasting tolerance resulting in hypoglycaemia, disturbed gastric tract motility, nephropcalcinosis and delayed myelinisation. Muscle biopsy showed diminished ATP production and activity of multiple complexes but no mitochondrial DNA abnormalities	Male	1.0	Not identified	7.7
11	Mild nonsyndromic intellectual disability with unexplained cyanotic spells	Female	1.6	Not identified	12.0
12	Diminished length growth (-5.5 SD) from birth and stationary stiff joints. General severe delayed development. Facial dysmorphic features, blepharophimosis, small mouth, poor mimicry. Brachydactyly, generalized stiffness and contractures	Male	0.3	Not identified	16.2
13	Psychomotor delay, mainly as a result of spastic diplegia	Male	2.6	Not identified	3.4
14	Macrocephaly (+2.5 SD), hypertelorism, remarkable hyperlaxity of the finger joints, and relatively small ears. Fused cervical vertebrae C2-C6, bilateral total tarsal coalition, psychomotor retardation. MRI of the brain revealed a Chiari 1 malformation. Conductive hearing loss	Female	11.8	*CHD4* – NM_001273.2: c.3518G>T, p.(Arg1173Leu) (variant in candidate gene)	3.7
15	Intellectual deficit, short stature, macrocephaly, feeding problems, disturbed sleeping pattern, mobility problems requiring wheel chair use	Female	11.5	Not identified	3.9
16	Mild intellectual disability with a verbal IQ of 81, and a performance IQ of 56. Stocky build, brachydactyly (similar to his mother), a sandal gap, pes planus, and mild joint laxity	Male	7.6	Not identified	3.7
17	Intellectual deficit, dysmorphic facial features, short stature (<-2,5SD), and trichothillomania.	Female	4.7	Not identified	3.9

ATP, adenosine triphosphate; GA, gestational age; IQ, intelligence quotient; MRI, magnetic resonance imaging.

Retrospective cost analysis

For the cost analysis, we set up a retrospective bottom-up cost of illness study from a hospital perspective. Data collection began for each patient at the first visit to the UMC Utrecht until initiation of WES.

All available resource use data were retrospectively collected from the hospital information systems and patient records. The data consisted of all health-care-professional visits, hospitalizations, imaging, genetic tests, metabolic investigations, and biochemical investigations performed at the UMC Utrecht. These collective investigations constitute the traditional diagnostic pathway. Also, information about genetic tests performed at other genetics centers was collected through referral letters from general hospitals or other university medical centers. In addition, age at first visit and total length of diagnostic trajectory were recorded.

All individual units of care were then linked to their unit costs. Unit costs of resource use were derived from various sources. Reimbursement prices issued by the Dutch Healthcare Authority were used as individual unit costs for medical interventions, imaging and diagnostics, biochemical analysis, and surgeries. For inpatient days, health-care-professional visits, day admission, and blood products, national reference prices established by previous research were used (Tan et al. 2012). WES cost was estimated at $3,972 (3,600) per trio for ease of comparison with previous studies (van Nimwegen et al. 2015). This includes the costs of patient registration and blood draw, DNA isolation, sample preparation, exome enrichment, sequencing on an Illumina HiSeq 2500, interpretation, reporting of results, data storage, and infrastructure. All costs were indexed to 2014 levels using rates issued by the Dutch Healthcare Authority. Euros were converted to dollars using the 2 November 2015 official exchange rate of the European Central Bank (1 = $1.1032).

Cost scenario analyses

To appraise the impact of introducing WES early in the patient diagnostic trajectory, we created additional cost scenario analyses evaluating which procedures would be unnecessary if WES was performed initially. After WES was performed, we divided the patients into two groups—diagnosed and undiagnosed— and for each group we calculated the total amount of costs that could potentially be saved by a WES-first approach in diagnostics. For the diagnosed patients, WES would replace all genetic costs (except for array comparative genomic hybridization (aCGH) and single-nucleotide polymorphism (SNP) arrays) and all metabolic assessments. For undiagnosed patients, WES would

replace all genetic costs except aCGH/SNP array assessments. For both groups, aCGH/SNP arrays will continue to be performed to exclude copy-number variants. We created a third scenario for these two groups assuming that WES will result in a cost reduction of 50% for health-care visits, imaging, biochemical investigations, and patient day admission. Having WES results can reduce health-care visits by eliminating the need for additional visits to discuss negative genetic or metabolic results, to avoid redundant imaging or biochemical investigations, and to spend less time in patient day admission.

RESULTS

WES yield

We prioritized *de novo* variants, as the selected patients were born of healthy, nonconsanguineous parents. After filtering for only exonic and essential splice-site variants, a total of 38 *de novo* variants remained. We validated 32 of the variants by Sanger sequencing, resulting in a validation rate of 84%, with a mean value of 1.88 *de novo* variants for each patient and a range of 0 to 5 *de novo* variants per patient in the exome. A full list of *de novo* variants for each individual patient is provided in Supplementary Table S1. The value of 1.88 validated *de novo* variants per patient in our study is consistent with the 1.71–1.98 validated *de novo* variants per patient previously reported in larger intellectual-disability cohorts, thereby validating our sequencing and analysis pipeline (Hamdan et al. 2014; Rauch et al. 2012). Autosomal recessive, X-linked recessive, and compound heterozygous inheritance models were also investigated, but no variants that could explain the patients' phenotype were detected. The lack of homozygous autosomal recessive causal variants was in accordance with our initial selection of only patients whose parents were nonconsanguineous.

We considered a variant diagnostic only if other patients with a clear phenotypic overlap had been described and harbored variants with comparable predicted damaging effect in the same gene. Diagnostic variants were found in 5 of the 17 patients, resulting in a yield in known or candidate Mendelian disease genes in this cohort of 29.4%. Variants in *CTNNB1* (catenin (cadherin-associated protein), beta 1, 88 kDa), *ANKRD11* (ankyrin repeat domain 11), *ADNP* (activity-dependent neuroprotector homeobox), *SMARCB1* (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1), and *CHD4* (chromodomain helicase DNA binding protein 4) were causal for five patients' phenotype (Table 1). Patients 2 and 4 further delineated the phenotype of the previously identified *CTNNB1* haploinsufficiency syndrome and KBG syndrome, respectively (Kuechler et al. 2015; Ockeloen et al. 2014). Variants identified in *ADNP* (patient 6) and *SMARCB1* (patient 8) provided a diagnosis of ADNP syndrome and Coffin-

Siris syndrome. In addition, for patient 14 the *CHD4* variant has a strong likelihood of being causal given that similar missense variants within *CHD4* were previously reported in the DDDUK study (https://decipher. sanger.ac.uk), with a strong similarity between our patient's phenotype and the reported phenotype (Bragin et al. 2014; Firth et al. 2009; The Deciphering Developmental Disorders Study 2015). Further functional work and similar patients with variants in *CHD4* are required to establish causality. A clinical geneticist returned the WES results to the patients and parents.

Retrospective cost analysis

We performed a retrospective cost analysis for all 17 patients involved in this study. The complete data set is available on request. Patients first visited the UMC Utrecht at a median of 1.1 years of age, and an average of 3.0 years of age. The average duration of the diagnostic trajectory was 6.6 years (Table 1). For our cost analysis, we focused on the diagnostic trajectory costs: health-care visits, imaging, genetics, metabolic measurements, biochemical investigations, and patient day admission (Table 2 and Supplementary Table S2). Costs for WES were not included. On average, patients had 61 visits with health-care professionals, with the majority of these being visits to see a medical professional (for instance, a medical specialist, nurse, or physiotherapist), with a mean cost of $3,012 (median: $2,144; range: $435–$9,844). Patients underwent imaging, with the majority being X-ray and magnetic resonance imaging, an average of 16 times, with mean costs of $1,439 (median: $771; range: $0–$7,981). An average of seven genetic tests per patient were performed, with a mean cost of $6,588 (median: $5,745; range: $2,183–$20,476). The majority of these genetic tests comprise single-gene tests (65%). A mean of 1.5 aCGH/SNP array tests (as technology improved) were performed per patient (median: 1; range 1–3), at a mean cost of $1,361 (median: $890; range: $890–$2,670). An average of six metabolic tests were performed per patient, with a mean cost of $2,818

Table 2 | Overview of mean/median number and costs of traditional diagnostic trajectory.

	# (median)	# (mean)	Costs (median)	Costs (mean)
Health care visits	38	61	2 144	3 012
Imaging	8	16	771	1 439
Genetics	6	7	5 745	6 588
Metabolics	6	6	2 777	2 818
Biochemical investigations	5	28	355	2 034
Day admission	2	4	309	517
Total			14 153	16 409

Values are given in USD

(median: $2,777; range: $2,204–$4,343) and mean biochemical investigation costs of $2,034 (median: $355; range: $140–$15,457). On average, patients were admitted during the day four times, costing $517 (median: $309; range: $0–$3,243). The mean total costs per patient were $16,409, median costs were $14,153, ranging from $6,343 to $47,841 for the total diagnostic trajectory.

The costs per patient are shown in Figure 2. On average, the genetic costs were 42% of the mean total diagnostic trajectory costs, ranging from 11 to 76%. We were able to diagnose patients 2, 4, 6, 8, and 14 following WES. The total diagnostic trajectory until WES for diagnosed patients took, respectively, 14.2, 8.6, 3.3, 5.8, and 3.7 years. The total traditional diagnostic trajectory costs for these patients were, respectively, $23,173, $14,187, $6,343, $5,893, and $9,934.

Cost scenario analyses

In three hypothetical scenarios, we evaluated the impact of WES on cost savings with the assumption that WES would render certain diagnostic investigations unnecessary. If WES was performed as a first diagnostic approach for patients in our group who ultimately received a genetic diagnosis, cost savings of genetic testing and metabolic testing would average $4,986 (median: $5,342; range: $0–$10,684) and $2,533 (median: $2,446; range: $2,204–$2,866), respectively. For patients who did not receive a diagnosis following WES but for whom WES would replace genetic testing, savings would average $5,699 (median: $4,854; range: $890–$18,696). In the final scenario, we assumed that WES would result in a 50% reduction of number and cost in the categories of health-care visits, imaging, biochemical tests, and patient day admission. For the diagnosed patients, we calculated mean total savings of $1,660 and for the undiagnosed patients $4,269.

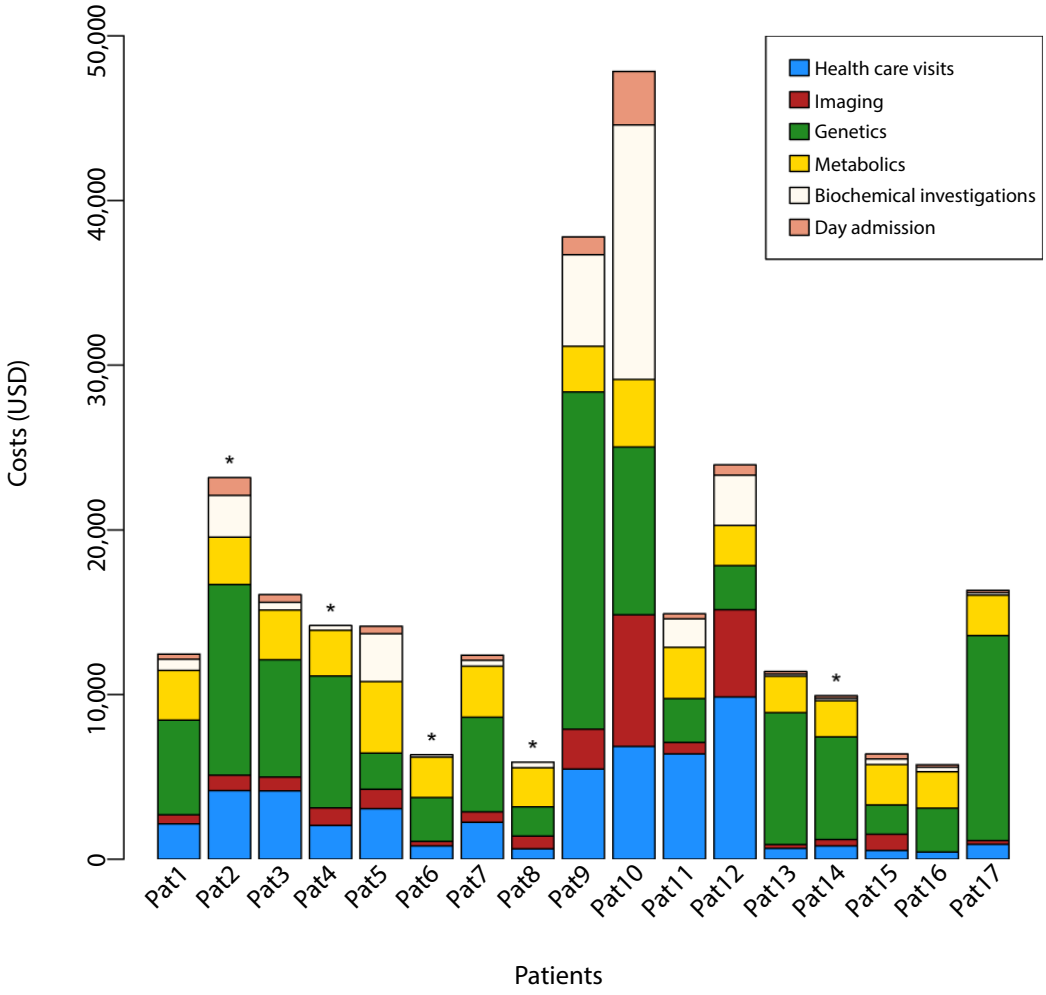


Figure 2 | Overview of the costs of the traditional diagnostic trajectory per patient. The cost of patient–parent trio-WES is not included. *Patients in whom the causal variant was identified.

DISCUSSION

Implementation of WES in clinical practice is at a turning point. The application of WES in diagnostics has transformed the clinic and allowed massive interrogation of the entire coding region of the genome. The diagnostic yield has been proven in numerous studies, and the cost continues to fall. However, information is lacking on the prospective yield using WES compared with the cost of patient diagnostics retrospectively using the traditional diagnostic pathway. An initial study focused on retrospective cost analysis of patient resource use in a cohort of neurological patients and the potential application of WES (van Nimwegen et al. 2015). The current study not only links the costs of patient diagnostics using traditional means but also reveals the higher yield and lower costs when actually implementing WES in a heterogeneous, difficult-to-diagnose patient group. This knowledge is essential for clinical diagnostic centers considering WES implementation so that the cost for increased diagnostic yield can be ascertained and considered if WES is advantageous.

The yield from this study further confirms that WES is useful in the clinic. Current single-gene disorder analyses or chromosomal microarray analyses have a diagnostic rate of ~13–14% in these patient populations, emphasizing the need for new molecular technologies (Henderson et al. 2014; Hochstenbach et al. 2009). Several previous reports using a trio-based WES strategy in large, specific patient populations reported diagnostic yields of ~25%, with some smaller studies reporting yields of up to 45–55% (de Ligt et al. 2012; Lee et al. 2014; Rauch et al. 2012; Wright et al. 2015; Yang et al. 2013). The yield in this study is comparable to the 25% from these large patient population yields, and it demonstrates the power of WES to identify the genetic cause of disease in cohorts of intellectually disabled patients. Recent studies have reported a diagnostic yield of 37% using trio analysis via the inclusion of novel genes for which a variant was observed in only one patient (Farwell et al. 2015). Our study's yield does not include variants in entirely novel genes without the support of additional patients with similar phenotypes or functional work to establish causality, as previous reports have demonstrated that "overinterpretation" of the exome can lead to false positives (Atwal et al. 2014). We expect that future essential patient data-sharing efforts and functional work will ultimately result in a higher yield for this patient cohort. Of note, the four previously known Mendelian genes that harbored diagnostic variants have only been known to be causal for these specific syndromes for less than 4 years, indicating the rapid progress in human genetics in the past few years. When this study was initiated, the underlying pathogenic mechanism for KBG (defects in *ANKRD11*) was still being elucidated and a diagnostic test had not yet been developed. Three of the genes—*ADNP*, *CTNNB1*, and *SMARCB1*—have only recently been associated with the respective syndromes. WES clearly has an advantage over traditional single-gene analysis

in situations when a specific molecular test does not exist or for syndromes for which the genetic cause has only recently been discovered. For the cases that are still unsolved, the WES results can be reanalyzed periodically to determine whether a detected *de novo* variant can be linked to new discoveries.

Our study confirms the long traditional diagnostic trajectory and the high costs for traditional diagnostic testing in this patient population. We found a mean cost of $16,409 per patient that is comparable to other total costs recently published, albeit for a different patient group (van Nimwegen et al. 2015). The largest proportion of costs (42%) was related to previously performed gene tests. A previous study examining WES yield in a cohort of 12 patients also estimated the high amount of resources used, with a single patient's laboratory investigations costing $22,000 (Need et al. 2012). High prices for additional diagnoses were also reported by Shashi et al. (2014) to be $25,000 per diagnosis if no diagnosis was obtained after a first visit (Shashi et al. 2014). Soden et al. (2104) recently estimated that negative diagnostic tests for a group of neurodevelopmental disorder patients cost $19,100 (Soden et al. 2014). In our study, patient 10 thus far has accrued a total of $47,841 in costs and still requires a diagnosis. For these patients, whole-genome sequencing is the next step in the diagnostic investigation to fully interrogate the genome.

Utilizing a WES-first approach would have immediate cost savings for diagnosed as well as undiagnosed patients. With a WES cost of $3,972 compared with a mean cost of $4,986 for genetic tests and $2,533 for metabolic tests, using WES would directly save, on average, $3,547 per patient who receives a diagnosis and a mean savings of $1,727 for patients who do not receive a diagnosis using WES.

Other savings may also be realized. Before WES was introduced, the differential diagnostic workup required additional investigations (e.g., cardiac or renal ultrasounds; skeletal X-rays). If no such anomalies are reported in a genetic condition that has been diagnosed by WES, then there is no need to perform these additional investigations. These additional savings demonstrate not only that obvious genetic or metabolic tests can become redundant if WES is the first diagnostic approach but also that a proportion of other patient procedures could be omitted.

This research demonstrates that implementing WES as a first diagnostic tool for patients with intellectual disability, even in a tertiary center population, could reduce health-care costs because WES could replace a large number of genetic and metabolic investigations. Costs of WES are markedly lower than the average total traditional diagnostic trajectory costs; indeed, the cost of trio-WES is already less than that of genetic investigations in all the patients in our study ($3,972 compared with $6,588). Minimally, WES would

avoid genetic and metabolic tests in patients who receive a diagnosis and genetic tests in patients who do not receive a diagnosis. An additional proportion of savings would be realized, at a proportion that will vary depending on which auxiliary procedures or costs the use of WES will partially replace. Importantly, these scenarios require an initial critical stratification by a physician specialized in diagnosing patients with intellectual disability. Such a physician is able to recognize patients who have a clear clinical presentation suggesting a known underlying genetic cause that is not detectable by WES (e.g., fragile X, trinucleotide repeat, or methylation disorders) and for whom WES would not be beneficial and a waste of resources. This indicates that WES should be considered first for the majority of cases in which a genetic condition is strongly suspected.

The current study has several limitations. Because we evaluated this procedure as a pilot project, the randomly selected sample size of 17 patients is limited. Definite conclusions regarding where WES should be implemented in the diagnostic pipeline cannot be drawn on the basis of such small numbers. However, the diagnostic yield was consistent with other studies, suggesting that this yield range is what can be expected if the study is expanded to include more patients. Moreover, the small number enabled us to disentangle individual diagnostic odysseys and their economic consequences.

The cost analysis study was as comprehensive as possible, but some additional costs may exist. Only direct medical costs were taken into account in this study, and resource use collection (except genetic tests) was restricted to the UMC Utrecht.

Also, if WES were applied earlier to our patient group, there would be a clear, quantifiable cost savings of genetic and metabolic tests in patients who receive a diagnosis and savings of genetic tests in patients who do not receive a diagnosis. The reduction of these costs represents the minimum savings that would be realized with earlier WES introduction. Additionally, the contribution of the auxiliary proportion of savings in the categories of health-care visits, imaging, biochemical investigations, and day admission is debatable. For instance, the 50% cost reduction that we assume may be lower for undiagnosed patients whom further diagnostic investigations are necessary. Indeed, investigations, metabolic or otherwise, can aid in the interpretation of WES results or elucidating the disease pathogenesis (Campeau et al. 2014; Patton et al. 1987). It is important for future analyses to consider these additional savings and correctly categorize diagnostic procedures that can be reduced by WES at the time of care to enable precise estimates of cost savings.

In conclusion, this study links the increase in diagnostically useful findings enabled by WES in intellectually disabled patients with the costs of traditional diagnostic investigations.

The increase in causal variant detection and speed of diagnosis and the lower cost of WES compared with traditional diagnostic investigations in this patient population suggest that WES should be implemented early in clinical diagnostic centers with similar patient populations.

ACKNOWLEDGMENTS

We are grateful to the patients and their parents for their participation in this study. This study makes use of data generated by the DECIPHER community. A full list of centers that contributed to the generation of the data is available at http://decipher.sanger.ac.uk and via e-mail from decipher@sanger.ac.uk. Funding for the DECIPHER project was provided by the Wellcome Trust.

DISCLOSURE

The authors declare no conflict of interest.

Supplementary Materials and Methods

Next generation sequencing

One ug of purified gDNA (Qiagen, Hilden, Germany) was sheared into 100-500 bp fragments with a Covaris S2 sonicator (Covaris, Woburn, MA, USA), blunt-ended, 5'phosphorylated, and A-tailed using Kapa Biosystems reagents (Kapa Biosystems, Wilmington, MA, USA). Adaptors containing the Illumina barcode sequences were ligated to each sample and amplified with 7 PCR cycles. Samples were finally quantified and each sample in the trio was equimolarly combined into one pooled batch. Barcode blockers for the Illumina adaptor sequences and barcodes were added (sequences available upon request), and the pool was enriched according to the Agilent Sureselect V5 exome protocol (Agilent, Santa Clara, CA, USA), and finally PCR amplified 8 times. Trio pools were then combined into batches of two trios together and sequenced on a full flow cell as a Rapid Run on the Illumina HiSeq 2500 at the Utrecht Sequencing Facility (Utrecht, Netherlands). Sequencing reads were aligned using BWA and data was processed with GATK v3.1.1, according to best practice guidelines but skipping BQSR and applying the hard filters instead of VQSR (McKenna et al. 2010; Van der Auwera et al. 2002). An in-house *de novo* script based on GATK PhaseByTransmission was used to provide a list of *de novo* variants. A nonphased (for *de novo* analysis) and phased VCF were then created for each trio.

Variant databases

NCBI dbSNP Build 137 for Human (https://www.ncbi.nlm.nih.gov/SNP)

The Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS)

The 1000 Genomes Project (http://www.1000genomes.org)

The Genome of the Netherlands (http://www.nlgenome.nl)

Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php)

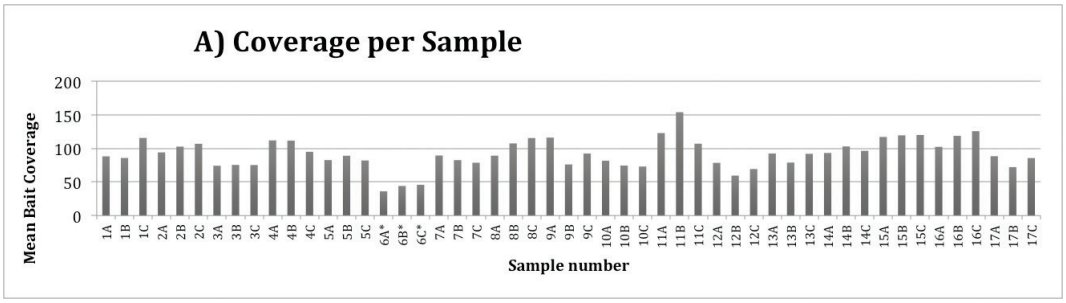


Figure S1 | NGS coverage of patients and parents of trio-WES. A) Sequencing coverage per sample. *The causal variant was found for patient 6 with sequencing coverage below desired minimal coverage requirement. No additional sequencing was necessary. B) Percentage bases covered to a depth of 20X per sample.

Table S1 | Sanger sequence validated *de novo* variants per patient.

Patient	Gene	HGVS nucleotide change	HGVS amino acid change	Coding effect
2	CTNNB1	NM_001904.3:c.1925_1926delAG	p.(Glu642Valfs*5)	frameshift
2	GPR158	NM_020752.2:c.2816_2817delAA	p.(Lys939Argfs*10)	frameshift
3	RP1L1	NM_178857.5:c.2365G>T	p.(Ala789Ser)	nonsynonymous
4	ZBTB9	NM_152735.3 :c.1065delA	p.(Gly357Valfs*15)	frameshift
4	JPH3	NM_020655.3 :c.1013A>T	p.(Tyr338Phe)	nonsynonymous
4	ANKRD11	NM_001256182.1:c.3382_3383delGA	p.(Asp1128Glnfs*41)	frameshift
5	REPIN1	NM_013400.3 :c.1177C>T	p.(His393Tyr)	nonsynonymous
5	FREM1	NM_144966.5:c.3899C>T	p.(Ser1300Leu)	nonsynonymous
5	DIAPH2	NM_006729.4:c.3226A>G	p.(Ile1076Val)	nonsynonymous
6	MATN2	NM_002380.3:c.2097T>C	n/a	synonymous
6	TEKT3	NM_031898.2:c.1046A>G	p.(Asn349Ser)	nonsynonymous
6	ADNP	NM_015339.2 :c.2496_2499delTAAA	p.(Asn832Lysfs*81)	frameshift
7	PATE4	NM_001144874.1:c.19C>T	n/a	synonymous
8	STAB1	NM_015136.2:c.955G>A	p.(Ala319Thr)	nonsynonymous
8	ZSCAN2	NM_181877.3:c.617T>C	p.(Val206Ala)	nonsynonymous
8	SLC35E4	NM_001001479.2:c.541G>A	p.(Glu181Lys)	nonsynonymous
8	SMARCB1	NM_003073.3 :c.1113C>G	p.(Asn371Lys)	nonsynonymous
10	TBL1XR1	NM_024665.4:c.1389C>G	p.(Asp463Glu)	nonsynonymous
10	FBF1	NM_001080542.1 :c.930C>T	n/a	synonymous
12	EVC	NM_153717.2:c.255G>A	n/a	synonymous
12	ELMSAN1	NM_194278.3 :c.1961A>C	p.(Tyr654Ser)	nonsynonymous
13	TMEM39A	NM_018266.2:c.258T>C	n/a	synonymous
13	APBA3	NM_004886.3 :c.1047C>A	p.(Phe349Leu)	nonsynonymous
14	FRAS1	NM_025074.6:c.5915C>T	p.(Ser1972Leu)	nonsynonymous
14	CHD4	NM_001273.2 :c.3518G>T	p.(Arg1173Leu)	nonsynonymous
15	KIFAP3	NM_001204517.1:c.2077G>A	p.(Ala693Thr)	nonsynonymous
15	HNRNPU	NM_031844.2:c.2299_2302delAACA	p.(Asn767Glufs*66)	frameshift
15	ZBTB49	NM_145291.3:c.1770C>T	n/a	synonymous
15	CDK3	NM_001258.2:c.735G>A	n/a	synonymous
15	HEXDC	NM_173620.2 :c.947C>T	p.(Pro316Leu)	nonsynonymous
17	MST1	NM_020998.3:c.1603C>G	p.(Arg535Gly)	nonsynonymous
17	OLFML2A	NM_182487.3:c.1666G>A	p.(Gly556Ser)	nonsynonymous

Table S2 | Unit costs of care procedures.

Procedure or visit performed	Cost (USD)	Supplemental information
Hospital visits		
Physician assistant	33	
Specialist	162	
Nurse	36	
Physiotherapist	43	
Orthopedist	43	
Audiologist	39	
Hospitalization	688	Per night stay
Radiology		
Ultrasound	94	
MRI	217	
X-ray	57	
ECG	41	
CT	183	
Genetics		
Mitochondrial DNA test	890	
Array CGH	890	
Single gene test	890	
Multiple gene test	2671	
MLPA	890	
Fibroblasts	202	
Metabolics		
Urine, metabolic	827	Initial test
	165	Additional tests
Plasma, metabolic	827	Initial test
	165	Additional tests
Cerebrospinal fluid, metabolic	827	Initial test
	165	Additional tests
Blood, enzyme	1211	Initial test
	243	Additional tests
Biochemical		
Blood basic	65	
Blood endocrine	71	
Blood virology	131	
Urine basic	75	
Cerebrospinal fluid basic	307	
Patient housing during the day	154	Per day

8

Discussion

NGS: the method for causal gene discovery in Mendelian disorders

The expanding breadth and decreasing cost of DNA sequencing techniques have allowed us to interrogate ever-larger portions of the genome in the search for a patient's diagnosis. The evolution from targeted gene panel sequencing to WES and WGS (**Chapters 2,3,5; Chapters 4,7; this chapter**) increases the chances of detecting the variant that may be causal, and NGS has rapidly been adopted to identify the genetic cause of rare diseases in pediatric and adult populations (Bick et al. 2017; Gilissen et al. 2014; Posey et al. 2016; Retterer et al. 2016; Stavropoulos et al. 2016; Wright et al. 2015; Yang et al. 2014). As an indication of its power, NGS is responsible for the overwhelming majority of new discoveries of the molecular cause of Mendelian disorders. Indeed, WES or WGS was used in 80% of such studies in the last 2.5 years and only 8% of studies solely utilized a traditional approach (Solomon et al. 2016).

With this wealth of data comes new challenges – the human genome is a massive place, and sifting through the 3.2 billion possible positions to identify the variant most likely to be causal is now the most difficult part that modern day genetic researchers face. Advancements in sequencing technology to decrease the error rate and provide longer reads coupled with the development of algorithms to detect complex events such as CNVs and structural variants is providing us with an even bigger but more complete set of variants. The fundamental challenge is no longer one of detection but of interpretation. However, the interpretation of genetic variants is hindered by lack of knowledge. We simply don't know precisely how the body works at the molecular and cellular level, nor the role of many genes and the effect of variation upon them – particularly in the unique genetic background of the individual. To compensate for this lack of knowledge, we use filtering methods to identify candidate variants that could be causal for the patient phenotype – though one must be aware of the limitations of each step. This significantly reduces the number of variants to consider. Then, widespread application and routine use of functional experiments and further 'omics' approaches such as RNA-sequencing (RNA-Seq) and metabolomics can be used to link DNA variation to biochemical processes.

The further incorporation of DNA sequencing in the clinic and the opportunities it entails will fundamentally change healthcare, both for patients to receive a diagnosis much faster than previously possible and for healthcare to provide better care at reduced cost. However, with this technology comes the significant challenges of ongoing technical limitations, correct variant interpretation, data storage costs, and secondary findings that remain to be addressed, as well as how healthcare and society cope with the knowledge that comes with personal genomes.

SUCCESS IS IN THE INTERPRETATION

We lack the time, resources, or ability to assess the effect of the 4.1 million to 5 million variant presents in the genome of a patient – particularly in the context of their specific genetic background (The Genomes Project Consortium 2015). To compensate for this lack of ability, we rely upon genetic filtering strategies to identify potentially causal variants in patients with a Mendelian disorder. We use the mode of inheritance, the novelty or rarity of the variant, and *in silico* variant effect predictions to reduce the variant candidate list. These strategies are frequently useful, though each step relies on assumptions that should always be questioned.

Mode of inheritance

A large portion of the success of WES and WGS sequencing in intellectual disability and Mendelian disorders is the fact that many of these disorders are due to *de novo* variants. Filtering for only *de novo* variants in the exome yields approximately 0 – 5 variants, a number that can be readily interpreted and experiments performed, particularly in light of the limited time and finances that can be dedicated. Similarly, in consanguineous families where an autosomal recessive inheritance is most likely, homozygosity mapping is beneficial in identifying regions that contain causal homozygous variants and has been used in combination with 40% of the WES or WGS studies to date (Najmabadi et al. 2011; Solomon et al. 2016; Walsh et al. 2010). Detection of *de novo* variants by trio-WES in **Chapter 7** allowed us to identify the causal variant in 5 patients with intellectual disability. Homozygosity mapping was helpful in identifying the genes to interrogate and the novel variants within those regions in **Chapters 5 and** 6, highlighting the role that traditional genetic mapping can still have in the era of NGS. However, the suspected mode of inheritance can be misleading. In **Chapter 4** we performed WES on two boys and their related parents, expecting a novel or rare homozygous variant to be causal for Mohr syndrome. Surprisingly, the actual genetic cause was due to compound heterozygous variants, a mode of inheritance only investigated after no suitable candidate homozygous variants were found. Inheritance modes that are less common but that can also be contemplated are maternal- or paternal imprinting, Y-linked inheritance, X-linked genes in the pseudo-autosomal regions, or genes that escape X-inactivation (Sobreira and Valle 2016). Thus the suspected mode of inheritance should be a guide, but not one to strictly abide by.

One diagnosis

Just as multiple modes of inheritance should be investigated in a patient, it should not be assumed that the patient phenotype is caused solely by one genetic cause. The contribution of two (or more) genetic diagnoses to a phenotype is an increasingly recognized interpretation complication. Two or more genetic causes could manifest in a blended phenotype when there is an overlapping phenotype; alternatively, if the features are non-overlapping a composite phenotype could result (Boycott and Innes 2017). The result in patients could be a more severe phenotype or phenotypic expansion of a known syndrome; when seen in the clinic, both should be viewed with caution and investigated more thoroughly because it could be the result of another molecular cause, perhaps still unknown (Posey et al. 2017). The contribution of multiple genetic mutations to the phenotype of the patient has thus far been not been extensively explored. Yang et al. (2013) first reported 4 of their 62 (6.5%) positive diagnoses had dual molecular diagnoses; subsequently larger studies have reported similar percentages but with a broad range of 1-7%, most likely dependent on the patient population (Eldomery et al. 2017; Farwell et al. 2015; Posey et al. 2017; Posey et al. 2016; Trujillano et al. 2017; Yang et al. 2013; Yang et al. 2014). For neurometabolic disorders, this proportion of a dual molecular diagnosis is even higher, at approximately 13-14% (Reid et al. 2016; Tarailo-Graovac et al. 2017). The prevalence of these dual diagnoses for diverse patient groups reveals that not only the suspected mode of inheritance but all modes of inheritance should be interrogated. The clinical question then moves from not only “what is the diagnosis of a patient?” to “is this the only molecular diagnosis for the patient?”

Population databases

The integration of large numbers of exomes and genomes into genetic population databases has enabled the identification of variants that are absent in a healthy population or that have a low minor allele frequency (MAF). Variant filtering strategies frequently compare a detected variant against these databases; those variants below a certain MAF threshold are retained for evaluation and may be causal for disease, as variants that are potentially pathogenic have been exposed to negative selection. In **Chapters 2 and 3**, variants were first prioritized upon novelty, and absent from population allele frequency variant databases such as GoNL, the 1000 Genomes, ES), or ExAC (Lek et al. 2016; NHLBI GO Exome Sequencing Project (ESP) 2011; The Genome of the Netherlands Consortium 2014; The Genomes Project Consortium 2015) This step removes the vast majority of detected variants, illustrating the value of such publically available population allele frequency variant databases. The collection of these datasets has traditionally been hindered by patient privacy concerns and researchers’ perception of ownership, in

addition to technical hurdles of standardizing NGS data processing and variant calling on data from many different centers, platforms and bioinformatic pipelines. The value of these databases cannot be overstated and have greatly aided variant interpretation effects.

Our interpretation ability for variants present in the noncoding genome is still hindered by the relatively small numbers of whole genomes that have been aggregated, though this is changing as more genomes are collected into databases such as gnomAD, Kaviar, and the Human Longevity genome database (Glusman et al. 2011; Lek et al. 2016; Telenti et al. 2016). Furthermore, sequencing efforts on a national stage will become integral in the interpretation of population-specific variants, particularly for ethnicities that are not well represented in the primarily Caucasian-dominated databases – such as the ongoing Genome Asia 100K project and the Korean Reference Genome Project (Fakhro et al. 2016; Genome Asia 100K 2017; Gudbjartsson et al. 2015; Seo et al. 2016; The UK10K Consortium 2015).

In silico tools

In silico variant effect prediction tools such as PolyPhen-2, SIFT, Condel, CADD, and Eigen have been useful in identifying which candidate variants to follow up experimentally (Adzhubei et al. 2010; Gonzalez-Perez and Lopez-Bigas 2011; Ionita-Laza et al. 2016; Kircher et al. 2014; Kumar et al. 2009). However, these tools can be individually highly discordant, biased upon their training variant set, and the single best strategy may be to compile a score based on multiple programs (Gonzalez-Perez and Lopez-Bigas 2011; Kircher et al. 2014; Miosge et al. 2015; Tennessen et al. 2012). Improvement on these tools is imperative. *In silico* tools at best lead in the direction of the causal variant; at worst, they either give a benign classification for a variant that is truly causal, or give a damaging prediction for a variant that is neutral (Manrai et al. 2016). Currently these *in silico* prediction programs may be helpful for variant prioritization but alone do not give sufficient evidence to assign definite pathogenicity, particularly for variants in newly identified candidate Mendelian disease genes (MacArthur et al. 2014). Thus, it is essential to evaluate the biological consequence of a genetic variant by demonstrating that it results in a different functional effect and could be explanatory for the disease. For this, we rely upon molecular or cellular experiments, and ultimately, for many developmental or metabolic phenotypes, on model organisms.

THE PROOF IS IN THE EXPERIMENTS

Functional tests and the zebrafish as a model organism

Experimental approaches on patient material, cell lines, or model organisms reveal abnormal effects of genetic variation and ultimately establish a phenotype-genotype correlation between the patient, their genetic variant, and their disorder. Molecular experiments were essential in demonstrating the aberrant splicing effects of the two variants present in our patients with Mohr syndrome, both of which were not expected from prediction programs; additionally, cellular experiments on patient fibroblasts established that these molecular defects had an impact at the cellular level by reducing the number of cilia (**Chapter 4**). To expand upon these molecular and cellular experiments and show abnormal function due to a variant at the organism level, we use the zebrafish to model developmental defects such as spasticity, a characteristic of the three children with SINO syndrome (**Chapter 3**) or for metabolic pathway disruption in a patient with LDHD deficiency and elevated D-lactate levels (**Chapter 6**).

To further study the effects of MCT1 deficiency upon metabolism (**Chapter 5**), we are utilizing CRISPR/Cas9 genome editing to introduce indels in the orthologous *MCT1* genes in zebrafish, *slc16a1b* and *slc16a1b*, effectively knocking out these two genes. This zebrafish line will be a functional model of *MCT1* loss-of-function, reflecting the physiological consequence of loss of this gene in humans. This line will allow us to investigate the metabolic consequences of *MCT1* loss-of-function at an organism level, and by experiments with different food intake (i.e. high fat) or experimental parameters (i.e. fasting), we hope to understand the diverse pathways and critical points that lead to ketoacidosis in our patients, to aid in future prevention of these episodes.

In research, the widespread possibility to utilize any molecular or cellular test at hand to establish causality for a patient variant is essential, and ultimately extensive functional tests, if chosen wisely, can prove pathogenicity. However, many of these experiments are patient-specific, costly, and take significant time; in the clinic, a more standardized functional test pipeline is needed to provide information necessary to interpret the large amounts of candidate variants from the routine use of WES or WGS. Establishing these standardized pipelines of cellular studies for a fast, reliable read-out of variant effect or the integration of a more integrated zebrafish facility will be crucial for interpretation patient variation on a large scale (Davis et al. 2014). For many developmental disorders, the zebrafish is ideally suited to reveal the consequences of variants in genes that are crucial for development or metabolism, and can be well integrated into diagnostic pipelines focusing on rare developmental disorders and metabolic diseases.

Complementary approaches to next generation DNA sequencing

Knowledge to interpret genetic variants does not only have to come from pedigrees, large population MAF databases, *in silico* prediction tools or model organism experiments. The integration of other sources of data to interpret DNA variants can be complementary and help to identify the causal variant – and be particularly useful when implemented in parallel with NGS.

RNA-Seq

If relevant patient material is available, RNA-Seq and the identification of altered transcripts due to candidate (or formerly unidentified) variants can detect specific genes that are alternatively spliced in different cell populations or tissues. By coupling this information with DNA variation this can increase diagnostic yield (Cummings et al. 2017; Gonorazky et al. 2016; Kremer et al. 2017). This approach is amenable to diseases that manifest themselves in easily biopsied tissue such as skin, or when necessary, muscle. However, for many Mendelian disorders the relevant tissue is not known or available. This approach is very applicable for patients with ketoacidosis but for whom no causative variant was found in *MCT1* (**Chapter 5**) or the two other known enzymes responsible for the inborn errors in ketolysis, SCOT or ACAT1 deficiencies. Identification of altered transcripts in MCT1 could reveal deep intronic or previously uncharacterized exonic splice variants; furthermore, the presence of altered transcripts in other genes could help lead to new genetic causes of this disease. Alternatively, if a specific candidate gene is suspected, cDNA sequencing of the transcripts present in a tissue could also indicate alternatively spliced products.

Metabolomics

Similar to RNA-Seq, metabolomics can give a functional readout of the consequences of genetic variation and can be used to give evidence for causality of a genetic variant. Metabolomics is the study of the metabolites (the substrates and products of enzymatic activity) that are present in a biological sample as a result of molecular and cellular processes. The metabolites reflect the actual molecular phenotype of a cell or tissue more accurately than RNA-Seq because it is a direct measurement of the consequence of biochemical activity, whereas the mRNA transcripts detected by RNA-Seq are subject to further post-transcriptional regulation. Metabolomics can constitute targeted assays for specific compounds with similar chemical properties (i.e. an organic acid screen or an acylcarnitine screen) or it can be untargeted, assaying all metabolites, regardless of their individual chemical properties. Measuring metabolite levels gives a readout of biochemical

processes in the sample and reflects the effect of genetic variation, and can be used in the identification of known diseases or even novel disease (Miller et al. 2015). By combining this knowledge with NGS, candidates in genes that could affect those metabolic pathways can be prioritized; or if genes identified by NGS are not known to cause an inborn error of metabolism (IEM), they can be prioritized and experimentally assessed to see if they result in the same metabolic phenotype (Guo et al. 2015).

Using NGS in combination with targeted metabolomics drastically improves diagnostic yield in patients with IEMs. Two recent studies using NGS of genes known to cause IEMs combined with diagnostic biochemical assays identifying specific biomarkers resulted in a diagnostic yield of 78% - 89% (Yubero et al. 2016) (Reid et al. 2016). This reflects the power of combining these two approaches to provide a diagnosis when both a genetic variant and perturbed metabolite is identified. Indeed, the diagnostic yield was much lower in patients for whom there was a clinical suspicion of a disorder but for whom no specific biomarker was identified (15.4% and 38%, respectively) (Reid et al. 2016; Yubero et al. 2016). This lower diagnostic yield highlights the limitations of each genetic and metabolic approach. By using only targeted metabolic assays, only those perturbed metabolites detectable in that specific assay are identified, missing all other metabolites that may be indicative of disease. Similarly, the above studies used targeted gene panel sequencing of only those genes already known to cause IEMs – potentially missing other genes that could be causal for the patient phenotype.

WES instead of a targeted gene panel sequencing approach results in a larger diagnostic yield. In a study combining WES and extensive biochemical testing, the total diagnostic yield of this group increased to 68% - showing the benefits of using a comprehensive genetic approach such as WES with targeted metabolomics (Tarailo-Graovac et al. 2016). Finally, the ultimate comprehensive approach would be an unbiased genetic approach such as WES or WGS in combination with an untargeted metabolomics approach. Yu et al. (2016) exemplified this approach by performing untargeted mass spectrometry and WES in a cohort of 1361 individuals to identify four genes with variants affecting serum metabolite levels (Yu et al. 2016). This approach, detecting all exonic variants and all metabolites, is the ideal to strive for - though combining these datasets, particularly in standard use, will be a challenge.

Regardless, the routine integration of NGS data and metabolomics data, targeted or untargeted, should be performed. Currently, centers assess these results independently, only combining the results in specific studies and not in the standard diagnosis pipeline. To maximize diagnostic yield for patients with IEM, these two technologies should be integrated into a merged diagnostic pipeline.

Conversely, metabolomics can also be used after a diagnostic genetic variant is identified to search for metabolite biomarkers that can be useful in future disorder recognition or confirmation of a diagnosis. Using an untargeted metabolomics approach and comparing metabolites in blood spots isolated from a patient with a homozygous *MCT1* loss-of-function variant resulting in an early frameshift p.(Asp15fs) (**Chapter 5**) versus control blood spots, we identified elevated 5-oxoproline (pyroglutamic acid) levels. MCT1 is a known transporter of 5-oxoproline, but until now a relationship between MCT1 dysfunction and elevated 5-oxoproline levels has not been demonstrated in humans (Sasaki et al. 2015). Metabolic evaluation of 5-oxoproline levels in body fluids of patients suspected of ketoacidosis due to MCT1 deficiency could quickly diagnose future patients. To further characterize this finding we performed targeted GC-MS (Gas Chromatography-Mass Spectrometry) analysis on urine of two patients with homozygous loss-of-function *MCT1* variants and one patient with a heterozygous loss-of-function variant. Both homozygote affected patients had on occasion elevated levels of 5-oxoproline in their urine. It was our hope that this biomarker could also be used to identify those patients with heterozygous *MCT1* variants and a risk of ketoacidosis from non-affected individuals carrying a heterozygous *MCT1* variant. Instead, the level of 5-oxoproline was not perturbed in urine of the heterozygous carrier and this particular biomarker would not be useful in heterozygous patient stratification. Nonetheless, this biomarker could be very useful in disease confirmation for patients with homozygous *MCT1* loss-of-function variants. Typical metabolic analysis of urine of these patients reveals ketonuria. However, the presence of ketones in urine is not specific for patients with homozygous *MCT1* variants. In fact, ketonuria is a physiological response upon fasting. The presence of ketonuria in combination of elevated 5-oxoproline could be used for faster recognition of MCT1 deficiency and indicate the necessity of genetic analysis of the gene. A specific assay is being developed to robustly detect this metabolite and quantify it in urine, including collecting values for ethnicity, sex and age-matched controls that are run on the same platform. Additionally, once this assay is operatory we can use it to further evaluate if 5-oxoproline levels are similarly elevated in the knock out zebrafish line as a consequence of loss-of-function of *MCT1*.

The variant is in the genome

DNA sequencing of a larger genomic area can be used when a targeted gene sequencing panel fails to detect a causal variant. In **Chapter 5** we identified a causal variant in 8 of 96 patients, all with a similar phenotype of recurrent ketoacidosis. To discover other variants or genes involved in this disease we performed WGS on a further 8 patients with ketoacidosis but with no causal *MCT1* variant. The interpretation of this dataset is challenging, as this disease can be caused by homozygous or heterozygous *MCT1* variants, with homozygous patients having a more severe phenotype. Notably, even though we performed WGS on

these patients our initial analyses focused upon the exome because variants within the coding region are expected to have the largest effect. Inborn errors of metabolism are primarily autosomal recessive diseases, requiring complete or near-complete absence of the protein, and we did not expect loss-of-function variants in the regulatory regions to be the most likely cause because the effect would most likely be more subtle. Thus far, no likely rare or novel homozygous exonic variant was identified in any single patient; additionally, no plausible deleterious variants in the same gene have been found in multiple patients. Single patient exonic analysis has revealed candidate variants in one patient that may ultimately implicate gain-of-function variants in the beta-oxidation pathway of fatty acid catabolism, leading to an increased production of ketones. In our patient, this could result in a physiological state that is more susceptible to ketoacidosis episodes in response to fasting. Currently we are in the process of quantifying gene expression within this pathway in patient fibroblasts to evaluate if this pathway is upregulated compared to controls; if so, it is possible that other genes within this pathway can be disrupted and could be causal for the remaining patient's phenotypes.

The fast-paced Mendelian disorder gene discovery and the importance of sharing

Identification of other phenotypically similar patients with potentially causal variants in the same gene is an important way to establish a genotype-phenotype connection. Collaborations with other research groups by sharing variants via databases (i.e. Genematcher) allow researchers to connect and empower their study with more patients, thus providing a more conclusive, comprehensive characterization of the disease (Sobreira and Valle 2016). However, historical competitiveness between research groups, researcher reluctance to share patient phenotype variant information, and the lack of widespread use of phenotype ontologies hinders the success of these sharing initiatives– hopefully something that researchers will become more comfortable with in the future as matchmaking efforts continue to be successful. Identifying other patients via the DECIPHER track of the DDD UK study (**Chapters 3**) or using Genematcher allowed us to include more patients and establish a clear genotype-phenotype relationship (Depienne et al. 2017; Firth et al. 2009; Sobreira and Valle 2016; Weiss et al. 2016). In the clinic, identifying other patients and combining this genotype-phenotype correlation with variant functional studies to establish causality enables a patient diagnosis. Valencia et al. (2105) reported that 33% of their patients with a diagnostic result were based on gene and phenotype correlations that were made only in the last 2 years (Valencia et al. 2015). Similarly, in **Chapter 7**, two of the patients with a candidate *de novo* variant that could not be considered causal at the time of the study were later included in a collaborative study and given a definitive diagnostic result after other patients with similar phenotypes and variants were identified

and functionally evaluated (Depienne et al. 2017; Weiss et al. 2016). Periodic reanalysis of patients' exomes with no previous diagnosis will additionally yield more diagnoses, as the genetic cause of more Mendelian disorders continues to be found (Wenger et al. 2017). While we are in this fast-paced time of discovery, it may be useful for healthcare providers to reanalyze exomes periodically to see if variants of unknown significance (VUSs) can be reclassified as now being causal for the patient disease (or, alternatively if new information reclassifies a previously causal variant to a VUS), though this is a significant burden on clinicians and should be undertaken only with care and deliberation (Valencia et al. 2015).

One solution could be to classify the reanalysis of a patient's WES or WGS data as an additional reimbursable clinical test. By categorizing the reanalysis as a reimbursable clinical product, initiated by a clinician when appropriate (i.e. at the patient's request; following a specified time period), the financial burden of providing the reanalysis would shift to insurance companies. This approach would require a critical health technology assessment including a cost analysis, as well as evaluation by the insurers, but it would make the decision to revisit NGS results for the patient much clearer and remove the hesitation by healthcare providers to reevaluate these results, particularly if there is clear clinical benefit. Alternatively, in light of the significant storage cost of genomic data and rapidly decreasing cost of DNA sequencing, it may be instead useful to simply perform DNA sequencing on the patient again. By making the patient eligible for WES or WGS and then reanalysis again within a specified time period, the consent, material, and perhaps even the reimbursement for this subsequent NGS test could be obtained upon the initial visit.

NGS AS A DISRUPTIVE TECHNOLOGY IN THE CLINIC

The application of NGS and the successful interpretation of variants by variant MAF databases, prediction tools, complementary experiments, and functional experiments can lead to causal variant identification in known or previously uncharacterized Mendelian disease genes. NGS is extremely powerful when used routinely for rare disease diagnosis in the clinic; however, the question remains of where in the diagnostic pathway NGS should be placed for optimal patient benefit.

Genetics first, ask questions later

Using NGS technologies, including WES or WGS, in a "genetics-first strategy" – before other diagnostic tests, but after careful clinical phenotyping and differential diagnosis – would have clear benefits for patients with rare genetic disease. Adopting a genetics-first approach is very useful for young patients, as these patients may not necessarily show

all the characteristic features of a syndrome until later in life – enabling a diagnosis years before it would have been possible with solely clinical methods (Iglesias et al. 2014). This approach is useful in rare diseases of genetic or clinical heterogeneity and those with phenotypic expansion, as these are the groups most likely to be impacted by sequential single gene or panel tests that do not identify causal genetic variants. Additionally, this strategy can provide a diagnosis in patients with an atypical clinical presentation, broadening the clinical spectrum, and arriving at a diagnosis faster (Neveling et al. 2013; Sawyer et al. 2016; Willemsen and Kleefstra 2014). A diagnosis for many patients results in a change of clinical management for the patient, family genetic testing, or recurrence risk evaluation (Stark et al. 2017). For those disorders that present these features later in life but that care can be given earlier if a diagnosis is made, making an earlier diagnosis can make a large difference. Most importantly, in patients whom have diseases that are treatable, having a diagnosis as soon as possible leads to immediate care options (Tarailo-Graovac et al. 2016). For many of these cases immediate effective treatment can only be possible by identifying the genetic variant as soon as possible in an unbiased DNA sequencing approach (Bainbridge et al. 2011; Tarailo-Graovac et al. 2017).

Such a route has benefits for diagnostic yield, though definitive conclusions on where the technology should be implemented are still lacking. In 2006, Rauch et al. first suggested WES should be used as a first-tier test for patients with intellectual disability; studies in larger patient groups also suggest that WES as a first-tier use would have saved significant time and healthcare cost in addition to improving diagnostic yield (Rauch et al. 2006; Retterer et al. 2016; Reuter et al. 2017; Trujillano et al. 2017). The current opinion is that WES in all of these scenarios could be useful and that most advantageous placement in the trajectory includes one where subsequent diagnostic test decisions are well thought out (Sawyer et al. 2016). This is particularly relevant considering the low diagnostic rate of traditional methods in confirming monogenic disorders. Parallel prospective studies contrasting traditional methods diagnostic yield versus WES diagnostic yield show a yield improvement from 7.3% - 13.75% using the traditional pathway compared to 29.7% - 57.5% using WES first (Stark et al. 2017; Vissers et al. 2017). With improving algorithms to also accurately detect copy number variants, WES can also replace SNP arrays to provide CNV, SNV, and indel detection in one test – this is very applicable in diseases where SNP arrays are not routinely performed, but that a CNV could be causal for the disorder (Pfundt et al. 2017). WES as a first-tier test thus has clear, demonstrable effects on improving diagnostic yield with comprehensive SNV and CNV detection. Though the CNV detection ability in WES is improving with better algorithms and should be implemented routinely in the clinic, WGS has a better sensitivity to detect smaller CNVs (Ellingford et al. 2017; Turner et al. 2016). WGS also detects more SNVs in coding regions than WES and detects structural variants, including copy-neutral events, though improvements in algorithms are

needed and there is inherent limitation in short-read NGS technology in detecting all kinds of structural variants due to its inability to span repetitive regions (Belkadi et al. 2015; Tattini et al. 2015). WGS will follow WES to become the test of choice in a genetics-first approach, as the interpretability of the noncoding genome increases and sequencing costs continue to decrease. Ultimately, implementing genetics as early in the process as possible yields clear benefits – not only for the increase in number of successful diagnoses, but also for the time saved for patients who have a clear diagnosis earlier.

A genetic first approach for cost savings

In the last 10 years, the cost of sequencing a genome have dropped 10,000-fold from $10 million in 2007 to approximately $1000 in 2017 (Hayden 2014; Phillips et al. 2015; van Nimwegen et al. 2016). As more data on cost accumulates, implementation of NGS will be facilitated by favorable cost/benefit ratio (Bowdin et al. 2016). Initial cost studies have focused upon patient populations that were highly selected, hard to solve, and enriched for patients at the end of the diagnostic odyssey – those without any more options and for whom WES or WGS was a last resort (**Chapter 7**) (Joshi et al. 2016). Due to previous extensive genetic testing in this group, it may not be surprising that application of a single test, trio-WES, resulted in cost savings. The widespread implementation of NGS as a first-tier test will result in WES or WGS applied to less highly selected populations, which may not have needed such extensive testing and in which the cost savings from NGS may be lower. Studies on early-WES implementation in these more general populations versus concurrent traditional care costs are scarce, though beginning to emerge. Scenario analysis of early WES implementation suggest that WES will result in cost savings compared to the traditional method and recent studies comparing prospective WES versus the traditional diagnostic pathway demonstrate that WES early in the diagnostic pathway reduces the cost per diagnosis (Sabatini et al. 2016; Schofield et al. 2017; Stark et al. 2017; Vissers et al. 2017). In the largest cost study to-date on patients undergoing WES, the implementation of WES as a genetic-first approach before other diagnostic tests would have resulted in a definitive cost reduction compared to the current diagnostic pathway – in all patients, even those for whom WES would not have resulted in a genetic diagnosis (Vrijenhoek et al. 2017). More of these studies are necessary, particularly for WGS implementation and cost/benefit analysis for different disease populations, but from initial assessments a genetics-first strategy, even on general intellectual disability populations, reduces costs while improving the diagnostic rate.

The realized cost-savings will be specific for each center and patient group application. For a thorough evaluation, costs that are often not well quantified within and beyond DNA sequencing must be taken into account as well as costs that are not encompassed in any

current study (Evans and Khoury 2013; Phillips et al. 2015; van Nimwegen et al. 2016). These include efficient usage of consumables, infrastructure costs, data storage costs, and interpretation time as well as incurred costs due to a secondary finding or therapy/treatment as a result of a WES genetic diagnosis. To allay concerns that the widespread availability of NGS will lead to over-diagnosis, over-treatment, and a resulting increase of demand on healthcare resources, prospective studies following up the cost of care after NGS-enabled use are needed (Diamandis and Li 2016; Solomon 2014). These studies are necessary to convince physicians, so that NGS can be more widely requested, as well as for insurers, so that NGS can be more routinely covered (Deverka and Dreyfus 2014). Ultimately NGS will become routinely adopted only if insurers are convinced of its value across different applications and will provide reimbursement the tests. Tellingly, in a survey of the five largest US insurance payers, covering 112 million customers, none of the policies covered WES or WGS (Phillips et al. 2017).

For patients, secondary findings will result in more monitoring, preventative treatment or even more genetic testing in relatives, though this has been shown to not result in dramatic cost increase in the only study yet performed on this effect (Vassy et al. 2017). Preliminary studies also indicate that the cost of care for large patient groups dramatically decreases following WES, indicating its perceived status as an “end-of-trajectory” test and that using WES does not result in more expensive subsequent care (Vrijenhoek et al. 2017). More prospective studies are needed to evaluate the long-term financial result of NGS, specifically for secondary findings, and alleviate fears that WES or WGS will result in drastically increased costs due to secondary finding testing in patients or relatives. Ultimately it will be the proven diagnostic advantage of NGS coupled with evidence-based publications on the cost-reduction of NGS, both within the diagnostic pathway and following NGS, that will result in the ubiquitous application, coverage, and payment of NGS.

ONGOING CHALLENGES

Though we currently are in the golden age of genetics for Mendelian disease, there remain significant gaps in our ability to detect all the genetic variation present in an individual. This, coupled with our incomplete knowledge of biology, results in decreased ability to interpret the variants present in a patient – and impairs our ability to provide a diagnosis. The result is that approximately half of all patients with rare disease do not receive a diagnosis, even by using WES or WGS and with extensive traditional diagnostic work up (Chong et al. 2015; Retterer et al. 2016; Stavropoulos et al. 2016; Yang et al. 2014). This failure to resolve all cases is due the technical limits of current widely used NGS short-read technology, incorrect assumptions or interpretation challenges, and simply unknown

biology. Whereas the past decade has seen an explosion in our ability to generate data, what is needed in the coming decade and future is ways to successfully interpret this variation.

Technical limitations of current NGS: short read length, structural variants detection, and the need for de novo assemblies

The domination by one company, Illumina, and its short-read technology in the NGS marketplace results in significant DNA sequencing limitations. GC-rich regions in WES remain hard to sequence as well as long homopolymer stretches in all NGS short-read technologies; additionally, short-reads have difficulty spanning large repeats, characterizing indels and do not resolve all structure variants (Ross et al. 2013; Snyder et al. 2010). These will need to be resolved either with longer read lengths, improved mapping algorithms, or a combination of the two. For WES and WGS, indel detection from short reads remains dramatically underassessed, and studies have estimated that 30-40% of indels may currently not be detected – primarily due to short tandem repeat regions that are problematic to map (Huddleston and Eichler 2016). The inability to accurately detect these variants decreases our chances of finding and interpreting the causal variant in a patient.

Similarly, structural variation remains an enormous challenge from NGS data, with up to 80% of longer structural variants (50 bases to 1 kb) missed by short read technology (Chaisson et al. 2015). The most straightforward to address these challenges is to incorporate long read sequencing technology, such as enabled by Oxford Nanopore Technology or Pacific Biosciences instrumentation, into sequencing pipelines. The ability to span complex regions in one single read immediately improves the resolution over these variants, and allows a much more detailed view of the variation present in a patient. A combination of sequencing approaches is the best solution to overcome the weaknesses of individual platforms, particularly true as long-read technology currently has a higher read error-rate (Koren et al. 2012; Ross et al. 2013). The implementation of this technology at low coverage on patients with an initial negative diagnosis from short-read NGS technology could improve diagnostic yield by detecting variants previously not accessible by short read NGS technology, as was recently reported for the first time in clinical practice by using low-coverage long-read Pacific Bioscience sequencing to detect a ~2 kb heterozygous deletion in a patient with a Mendelian disease (Merker et al. 2017). Additionally, the use of long reads has the advantage of characterizing haplotype information that is completely missed by short read sequencing. Haplotype information is important to identify variants that are present on the same DNA strand and can be combined in *cis* interactions, such as between an enhancer and a promoter (Kitzman et al. 2011; Peters et al. 2012).

Even when using long reads, the sequences are still mapped back to a reference genome. If the sequences are not present in the reference genome they are discarded as being unmappable. However, these reads may capture a structural variant unique to the individual that is very relevant but currently unmappable using a reference genome. For these variants, a *de novo* assembly of the patient genome is required. The best way to address this in the future may be a combination of long and short reads to build a de novo genome assembly (Chaisson et al. 2015; Erlich 2015). Zahir et al. (2017) recently used WGS and de novo assembly for 8 patients with intellectual disability, identifying a *de novo* 165 kb heterozygous deletion that was missed by clinical microarray (Zahir et al. 2017). However, the *de novo* assembly of genomes requires substantial computing power, particularly for routine use, and the computational, bioinformatic, and assembly challenges are currently beyond the ability of most of the largest computing centers. Most likely this will not be accomplished in the near future. In the meantime, the transition to the newest available genome build GRCh38 will aid in structural variant identification for those centers whose bioinformatic pipelines were based on the older GRCh37 genome build (Guo et al. 2017).

Finally, WES or WGS on any platform still has its shortcomings. The genome-benchmarking consortium Genome in a Bottle reported that only 75% of genes in ClinVar and OMIM were in high-confidence areas, suggesting that variant detection in the other 25% of the genome was inconsistent (Goldfeder et al. 2016). Also, there are some types of variants that NGS simply still cannot detect, highlighted in a recent study of WES in 150 trios, where the standard diagnostic pathway provided diagnoses for three patients that WES missed due to a duplication event, *FMR1* repeat expansion, or mosaic duplication (Vissers et al. 2017). Finally, NGS bioinformatic analysis has difficulty detecting genome mosaicism unless performed at a large read depth (Mirzaa et al. 2016; Sobreira and Valle 2016). This emphasizes the limitations of NGS as well as the necessity of a clear diagnostic pathway to clinically identify those patients for whom NGS would not be beneficial. Technical limitations of NGS remain, but if they are known and identified alternate diagnostic routes and technologies can partially compensate.

Variants affecting splicing

In addition to addressing ongoing NGS technical challenges, our successful interpretation of variants that affect splicing in the genome is only beginning. The contribution of splicing variants to human disease is estimated to be 1/3 of all variants; however, many of these variants are currently unidentified (Lim et al. 2011). Variants affecting the canonical splice sites of exon/intron junctions can be annotated as splicing variants with a high degree of confidence but variants in exonic regions that affect splicing are routinely missed or not

annotated as splicing variants. Furthermore, deep intronic variants may activate cryptic splice sites (Soemedi et al. 2017; van Haelst et al. 2015). For consanguineous families, homozygosity mapping data can reduce the genomic search areas in the coding and noncoding genome to interrogate for extremely rare or novel homozygous variants that could affect splicing (Khan et al. 2017). In non-consanguineous families, variants can be prioritized upon novelty or extreme rarity. Interpreting these rare variants with RNA-Seq can lead to the identification of causal splicing variants; without additional information, interpretation of these variants should begin with novelty and be supplemented with some level of support from in silico splicing prediction programs such as CryptSplice or Human Splicing Finder (Desmet et al. 2009; Khan et al. 2017; Lee et al. 2017). In **Chapter 4**, in our two individuals with Mohr syndrome, the heterozygous *NEK1* variant at position c. 464G >C was annotated as a nonsynonymous amino acid change; only upon cDNA sequencing were we able to establish that the genetic variant instead resulting in a splicing effect. A considerable proportion of variants annotated as disease-causing in fact cause disease not by coding level changes but by splicing changes – though this effect is not visible until functional experiments are performed (Soemedi et al. 2017). Considering the lack of predictability of variants that could affect splicing, when a very rare or novel variant is detected in the exonic region of a strong candidate gene its function should be assessed experimentally to determine if it has a splicing effect. Similarly, novel or very rare deep intronic variants should also be assessed in this manner, though identifying them remains harder due to the lack of evolutionary constraint in these regions as well as less robust population minor allele frequency information to filter against. Additionally, splicing is not affected solely at the SNV level – mechanisms such as nonsense-associated alternative splicing occur if a nonsense variant is present in the exon, removing an entire exon (**Chapter 4**). The occurrence of this mechanism is not well understood or predictable, and only can be assessed with functional tests.

Finally, synonymous variant evaluation and interpretation remains a largely ignored area of exploration. The majority of synonymous variants are not detrimental and exclusion can remove a large portion (~50%) of exonic variants) (Bamshad et al. 2011; Seaby et al. 2016). However, synonymous variants can affect splicing, transcription, and translation in many different ways and can have large effect (Manousaki et al. 2017; Sauna and Kimchi-Sarfaty 2011). In the context of rare disease, novel or rare synonymous variants should be evaluated critically if the known gene function could be involved in the disease etiology.

The noncoding genome

The biggest advantage of WGS over WES is its ability to cover the noncoding genome; however, our ability to interpret variants in the noncoding genome is still in its infancy. The

contribution of intronic sequence, intergenic regions, regulatory, promoter and enhancer elements, and microRNAs and lncRNAs to Mendelian disorders remain for the large part unassessed (Zhang and Lupski 2015). The developments of the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics databases have made functional elements of the noncoding genome identifiable (Bernstein et al. 2010; ENCODE Project Consortium 2012). Expanding on these regulatory datasets, 4C-seq, ChIA-PET, Capture-C, Hi-C, and ATAC-seq (and many other NGS-based techniques) have characterized the architecture of the genome, including chromatin loops to enable enhancer-promoter interactions and open chromatin (Belton et al. 2012; Buenrostro et al. 2015; Hughes et al. 2014; Li et al. 2010; van de Werken et al. 2012). A combinatorial approach to 1) identify rare variants in the noncoding genome, based on population databases of variant minor allele frequency, with 2) annotation of the roles and function of the noncoding genome at those variant positions is needed to prioritize candidate variants. This approach is suited for where variants are not identified in the exome, and will be applicable in the future to the *MCT1*-negative cohort and other patients for whom a diagnosis has not yet been possible.

The genetic singularity

The success and application of all sequencing technologies – from Sanger sequencing to NGS – has been the result of decreasing cost. The cost of Sanger sequencing decreased as Celera applied it on a massive scale to sequence the human genome; similarly, as NGS has been implemented on a massive scale, prices have plummeted. A genome in 2008 cost ~$1 million; now, with the introduction of the NovaSeq instruments by Illumina in 2017, the $100 genome in the next few years is within reach (Illumina Inc 2017; Wheeler et al. 2008). The question is no longer if this price can be obtained, it is simply when. This brings an interesting point in DNA sequencing that I will call the genetic singularity – when the costs of DNA sequencing and bioinformatic processing are equal to the costs of long term genome storage.

The price for bioinformatics processing on the genome varies considerable, and is an important financial consideration for centers without local computation infrastructure in place. In highly subsidized institutes with large processing infrastructure available that is "freely" accessible, current bioinformatic costs of processing the raw genome files are estimated at 25 euros (Nijman 2017). On the other hand, commercial companies have traditionally charged 10% more on top of the actual WGS price to perform bioinformatics analysis, leading to a rough approximation today of 150 euros for bioinformatics analysis per genome over the last few years – though this will be forced to change as DNA sequencing costs decrease.

However, one of the largest costs of WGS data generation is long-term storage. Diagnostic facilities store data for years to be able to return to the raw files that generated the diagnostic result. This cost also varies considerably depending on how secure and how the data is stored, but locally at the UMCU the current storage costs are ~30 euros for one genome per year (Nijman 2017). This includes storage of the Binary Alignment Map (BAM) files and smaller variant call files (VCF) or genome variant call files (g.VCF). As the cost of sequencing decreases, the price of bioinformatics analysis should in theory stay the same – the actual data storage size of the genome has remained relatively constant (3.2 billion bases x 30X: ~100 GB), the bioinformatic condensing of a genome has not been drastically optimized and implanted (many tools are not compatible with compressed CRAM format files), and the price of data storage is not currently decreasing radically (Nijman 2017). This will result in new data storage issues as more and more centers desire to store genomes to revisit diagnostic patient data or use it as a resource. This is not only in terms of physical data storage, but also in the cost; the cost of DNA storage at the UMCU for five years is 150 euros (assuming 30 euros/year per genome). Note that these values are only estimates and also highly-specific to the UMCU; a thorough microcosting analysis of all processing steps and infrastructure costs, both physical (i.e. electricity) and personnel-related (i.e. bioinformatic tool development) are needed for an accurate cost assessment. Regardless, at the UMCU there is a possibility that within the next decade the cost of DNA sequencing and bioinformatics processing will approach or be less than the cost of five-year genome data storage (Figure 1). This will change how we think about DNA sequencing – is it something that is performed once and result files are stored indefinitely? Or do we perform DNA sequencing on demand?

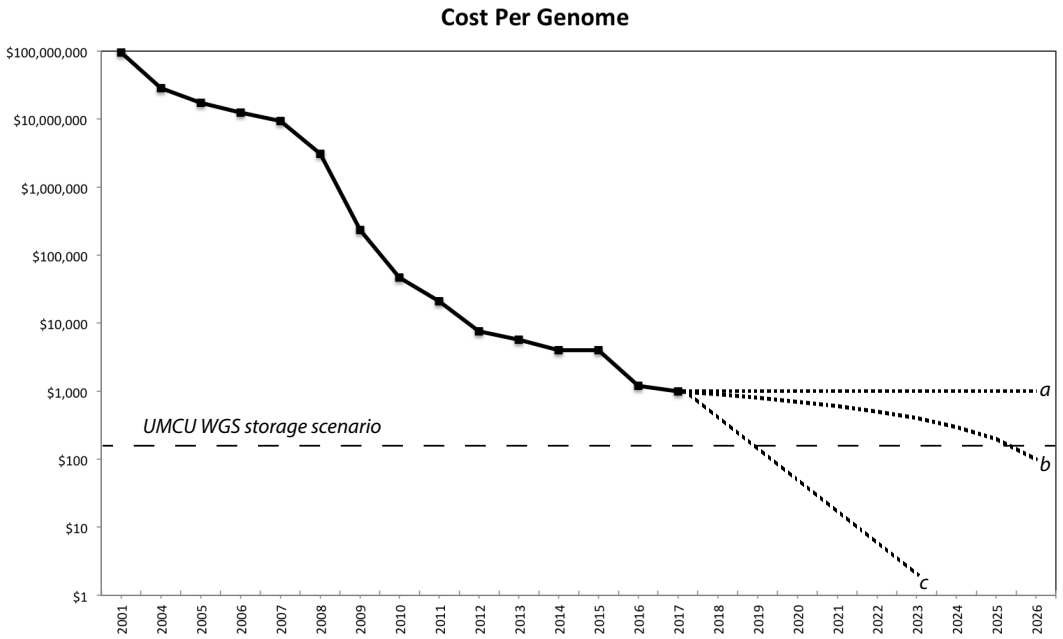


Figure 1 | The genetic singularity. Costs in genetic studies have historically been in data generation. Technological innovations such as the Illumina GA/GA IIx and SOLiD 2/3 (2008/2009), the Illumina HiSeq X (2014) and then only 3 years later with the NovaSeq (2017), have enabled dramatic decreases in sequencing costs. DNA sequencing costs approaching $100 in theory should be possible within the next few years. Here, future DNA sequencing costs are depicted in three scenarios: remaining constant at 2017 prices ("a"), decreasing at $100/year ("b"), or a technological innovation results in a drastic cost reduction ("c"). If DNA sequencing costs continue to decrease by $100/year, in roughly 8 years the cost of DNA sequencing and local processing will approximately equal five-year genome storage costs at the UMCU (150 euros: ~$175; UMCU WGS storage scenario). (1 euro = $1.17 9 Aug 2017)

If it costs more to store a genome for five years than to sequence and process the DNA again, a more viable solution will be to perform additional DNA sequencing as necessary, and not store the result files indefinitely. Solutions for this scenario will need to be addressed, and a critical evaluation of how much data needs to be stored will have to be performed. Perhaps the best solution will be to retain the VCF (or g.VCF) only and discard the larger BAM result files following patient analysis. As additional questions arise, it will be cheaper to re-perform the DNA sequencing as necessary, particularly because these raw files are rarely revisited. Additionally, improvements in sequencing technology and pipelines may improve variant interpretation upon subsequent patient DNA sequencing. Our perception of the necessity of data storage may change as we approach the limits of what is possible to economically retain.

PERSPECTIVES

As outlined above, NGS has great potential to be a disruptive technology in the clinic, and will improve patient care – particularly as we move to routine WGS. However, WGS by itself is not the solution for everything. Like every other technology, at the end of the day it is a tool – a tool to be used, but a tool that cannot do everything. WGS will enable more diagnoses but will also be hampered by its technical limitations and our interpretation due to lack of knowledge of basic molecular and cellular knowledge. The greater availability of this resource to the patient and general population, however, will challenge our concepts of clinical utility, change how we conceive of health and disease in an individual and demand new ways to think about healthcare.

Beyond clinical utility: personal utility

The decreasing cost of WGS will soon enable it to be broadly implemented, though caution and use of evidence-based cost studies should be performed to understand where in diagnostic pathways it should be placed for maximum patient and cost benefit. As WGS moves towards more general use, the notion of clinical utility must be revised to reflect the value of a diagnosis for a patient and their family. Clinical utility has traditionally been defined as the usefulness of genomic information to improve medical outcomes. Receiving a diagnosis for patients with rare disease often does not immediately translate to improving medical outcomes (Grosse and Khoury 2006). However, obtaining a diagnosis for many patients and parents does have clear benefits, including ending the diagnostic odyssey, long-term family planning, as well as knowledge of the disease prognosis and course (Botkin et al. 2010; Ravitsky and Wilfond 2006; Turner et al. 2008). A broader definition of utility must encompass what obtaining a diagnosis means for the patient and family beyond medical interventions, something referred to as “personal utility” (Berkenstadt et

al. 1999; Kohler et al. 2017). Studies have shown that patients, and individuals in direct to consumer testing, overwhelmingly desire to receive at least some genomic information – regardless of its implications (Bollinger et al. 2012). From a reimbursement standpoint, ending a diagnostic odyssey is as relevant as providing a diagnosis that changes treatment. This necessitates a change in mindset in requesting and delivering genetic results to patients, and justifying DNA sequencing for untreatable disorders – the value of a diagnosis is high. In the context of the clinic this means that there is a clear benefit to the individual to receive generated genetic information.

From bench to bedside to kitchen table: personal genomes

As DNA sequencing becomes available to everyone, there will be a clear need to interpret not only a patient's causal variant but also a healthy individual's genome. Only sixteen years ago, the first draft of the human reference genome was released. The first exome was sequenced only 8 years ago; as of today, the gnomAD database contains 123,136 exomes and 15,496 genomes, with thousands more sequenced in the clinic as routine diagnostic WES and WGS in laboratories around the world (Lek et al. 2016; Ng et al. 2009). Six years ago, in this laboratory, I was learning how to do targeted gene panel enrichments and optimizing exome sequencing; as of today, I, as well as other "healthy" individuals, have our own personal genomes, sequenced for less than a thousand dollars (~850 euros), stored on our personal computer. The progress from a human reference genome to personal genomes has occurred faster than anyone could have anticipated (Figure 2).

The value of a personal genome is hard to define; by definition, it is personal and the value itself depends on the individual. This requires a broad interpretation base. When DNA sequencing costs drop to $100, looking at one's own genome will no longer be a research hobby available to the privileged few, such as in my case; instead, it will be accessible to many, easily. Once a large proportion of the population has had their genome sequenced there will be a demand to interpret the results. For healthcare to stay relevant and be at the forefront of this disruptive technology it must not be reactive and wait for this situation to occur, but instead set up infrastructure (and a knowledge-base) to anticipate the need that will occur as genome sequencing becomes accessible and standard to healthy individuals. The alternative is that this niche will be filled with direct-to-consumer companies.

When established, such an infrastructure could deal with questions arising from genome sequencing. For instance, imagine a scenario with an individual sequencing their genome via a private company, and discovering a rare heterozygous frameshift variant in a muscle fiber gene. Upon further inspection, the individual discovers that variants within this gene

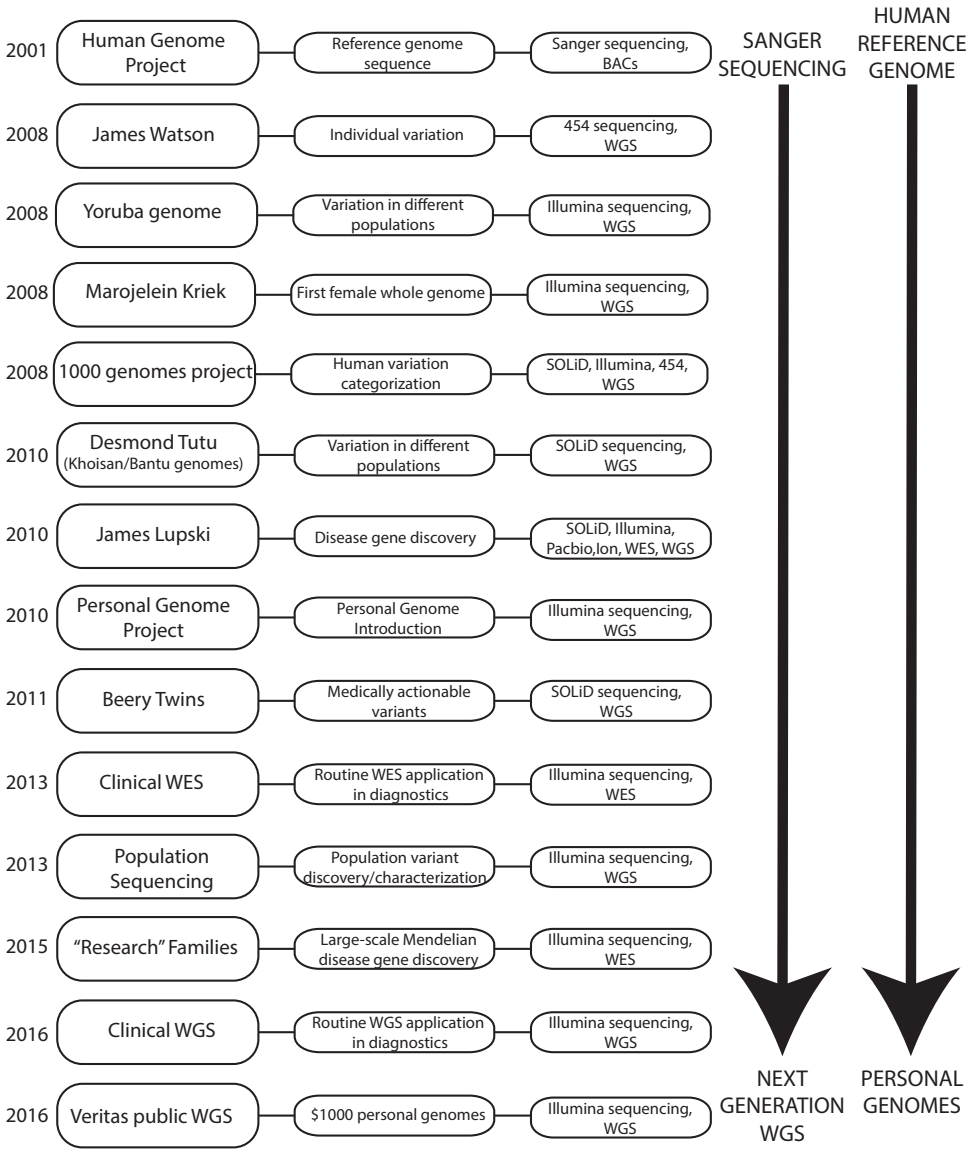


Figure 2 | From the human genome reference to personal genomes. Left column: exome or genome sequencing project, and year introduced. Middle column: significance for genetics, either from a clinical standpoint or a personal genome standpoint. Right column: technology utilized. Figure adapted and expanded from (Lupski 2016). (Bainbridge et al. 2011; Bentley et al. 2008; Chong et al. 2015; Church 2005; Consortium et al. 2015; English et al. 2015; Genome of the Netherlands Consortium 2014; Genomes Project Consortium et al. 2010; Gudbjartsson et al. 2015; Lander 2001; Leiden University Medical Center 2008; Lunshof et al. 2010; Lupski et al. 2013; Lupski et al. 2010; Schuster et al. 2010; Stavropoulos et al. 2016; Telenti et al. 2016; Venter et al. 2001; Wheeler et al. 2008; Yang et al. 2013; Yang et al. 2014)

have recently been linked to a cardiomyopathy. Additionally, this specific variant seen in the individual has been reported in one patient with a cardiomyopathy. Should this "healthy" individual be concerned? To evaluate the consequence of this variant, the individual would first go to their general practitioner for interpretation. The individual has no symptoms, nor does the family of the individual. However, due to the possible severity of this variant, the general practitioner may refer the individual to a clinical geneticist; also, the individual could be proactive and contact a clinical geneticist directly if the general practitioner did not think a referral was justified. It is also possible that the general practitioner would not think that this variant is a risk. In the current healthcare workflow, the result of this scenario is extremely unclear. This situation demonstrates the lack of a well-defined pathway to deal with genomic questions that will arise as more people have their genome sequenced. Even though questions such as this may be rare (or unheard of) at this point, they will become commonplace in the future. Rising genetic awareness will push a demand for a distinct response to genetic findings, resulting in more (preventative) care – or, depending on your viewpoint, more unnecessary care and overtreatment.

This scenario is not hypothetical; this variant (Figure 3) resides within my own genome. I have a heterozygous variant in the Obscurin gene (*OBSCN*; NM_001098623.2:c.20967del, p.(Ser6990Profs*82)). Heterozygous loss-of-function variants in this gene have been reported in patients with left ventricle non-compaction (LVNC), a rare heart cardiomyopathy, and this variant is seen in one patient reported (Marston 2017; Rowland et al. 2016).

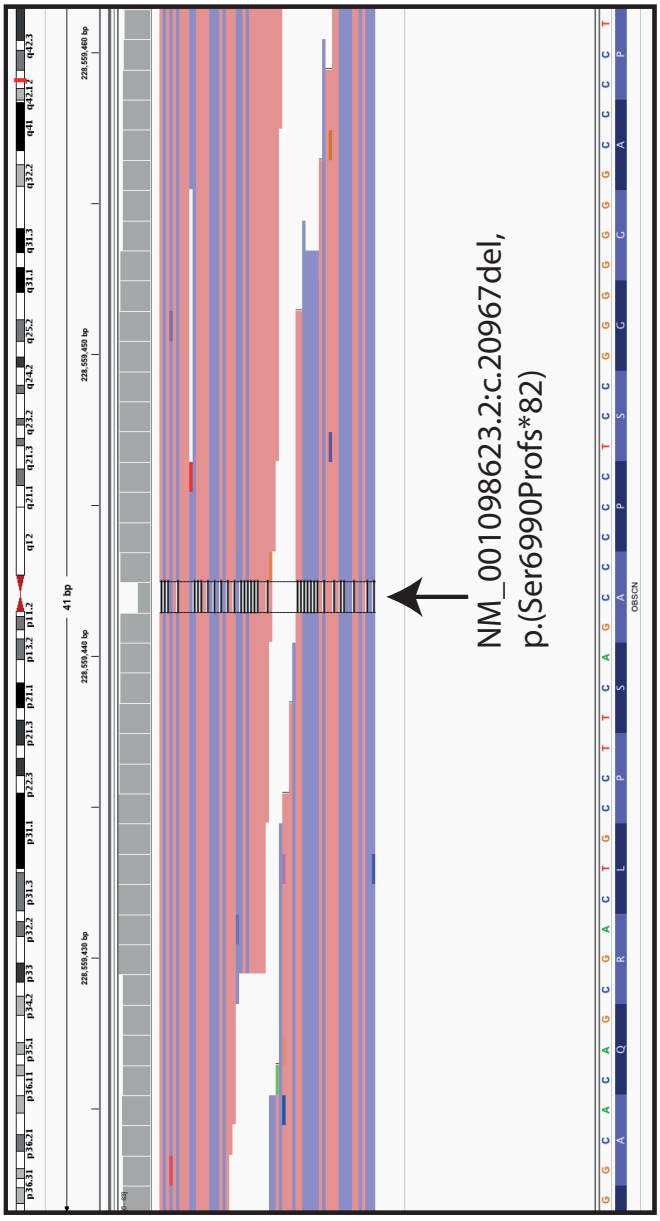


Figure 3 | *OBSCN* heterozygous frameshift variant within my genome. WGS data visualization of the variant NM_001098623.2:c.20967del, p.(Ser6990Profs*82) using the integrative genomics viewer (IGV) visualization tool. This variant results in a frameshift and predicted protein truncation of 900 amino acids (or, more likely, nonsense-mediated decay of the mRNA transcript). This exact variant has been reported in a patient with left ventricle non-compaction (Robinson et al. 2011; Rowland et al. 2016). Note that in Rowland et al. (2016) the variant was based upon NP_001258152, and annotated p.(Ser7947Profs*82) (Rowland et al. 2016)

The finding of variants such as this will only increase as more “healthy” people sequence their genomes. Solutions will need to be in place to cope successfully with variants that have enough evidence to follow up on. The interpretation of these deleterious variants in healthy individuals will be compounded by misannotation of benign variants as disease causing in databases such as HGMD or resilience to the disease in that individual due to protective variants – most of which are not known (HGMD: Human Gene Mutation Database) (Chen et al. 2016; Xue et al. 2012). In my case, I have the luxury of being able to consult with a clinical geneticist at the UMCU to interpret the impact of this variant and see if an echocardiogram or electrocardiogram to evaluate my heart’s structure and function is justified. However, I have no family history of heart problems and in ExAC there are 255 “healthy” people that are heterozygous for this exact variant (rs71180793; the overall MAF of this variant is 0.002, and 0.003 in the European population), though their phenotype is not known and the prevalence of LVNC in these individuals has not been assessed. Additionally, ExAC contains patient groups that are enriched for heart phenotypes.

This highlights the potential benefits and dangers of WGS in healthy individuals – for those that are prepared for the results, the findings represent an opportunity for preventative care. For those that are not, they may result in unnecessary anxiety. Regardless of the emotional consequences, the fact is that people will have their genomes sequenced as the price continues to decrease. The healthcare system must prepare for the inevitable increase in requests, justified or not, for increased questions, more follow-up, and perhaps further care. Secondary or unsolicited findings may result in a cost increase; if so, this must be anticipated and appropriate healthcare and insurance options developed and implemented.

CONCLUSION

Technology drives opportunity. NGS technologies have given us unprecedented ability to look inside our genes, and for patients with rare disease an unbiased, comprehensive test should be the first line of their care – for better, faster patient care at lower costs. NGS in research and the clinic has enabled the discovery of more Mendelian genes whose dysfunction results in disease. Importantly, this thesis is a compilation of success stories: for every patient that are we able to technically identify a variant, successfully interpret it as being a candidate for the disease and establish causality with functional studies, there are other patients whose diagnosis was not solved due to technical limitations, difficulty of interpretation, and the lack of time or resources for comprehensive functional studies. We have a long way to go. Even so, this technology has not yet achieved its pinnacle – just as I would have been amazed that WGS is commonplace 4 years ago, I am sure that in another 4 years there will be even bigger shifts in technology that will challenge our conception of health and disease. We are just at the beginning and those who do not imagine the ways in which DNA sequencing will change healthcare will be left behind when it does.

Bibliography

Acuna-Hidalgo R, Veltman JA, Hoischen A (2016) New insights into the generation and role of de novo mutations in health and disease. Genome biology 17: 241.

Adly N, Alhashem A, Ammari A, Alkuraya FS (2014) Ciliary genes TBC1D32/C6orf170 and SCLT1 are mutated in patients with OFD type IX. Human Mutation 35: 36-40.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nature Methods 7: 248-249.

Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, et al. (2007) Direct selection of human genomic loci by microarray hybridization. Nature Methods 4: 903-905.

Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nature Methods 8: 61-65.

Allison R, Lumb JH, Fassier C, Connell JW, Ten Martin D, et al. (2013) An ESCRT–spastin interaction promotes fission of recycling tubules from the endosome. The Journal of cell biology 202: 527-543.

Almind GJ, Brøndum-Nielsen K, Bangsgaard R, Baekgaard P, Grønskov K (2009) 11p Microdeletion including WT1 but not PAX6, presenting with cataract, mental retardation, genital abnormalities and seizures: a case report. Molecular Cytogenetics 2: 6.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic acids research 43: D789-D798.

Andreazzoli M, Gestri G, Landi E, D'Orsi B, Barilari M, et al. (2012) Kidins220/ARMS interacts with Pdzrn3, a protein containing multiple binding domains. Biochimie 94: 2054-2057.

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

Arevalo JC, Yano H, Teng KK, Chao MV (2004) A unique pathway for sustained neurotrophin signaling through an ankyrin-rich membrane-spanning protein. The EMBO journal 23: 2358-2368.

Asakawa K, Suster ML, Mizusawa K, Nagayoshi S, Kotani T, et al. (2008) Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. Proceedings of the National Academy of Sciences of the United States of America 105: 1255-1260.

Atwal PS, Brennan M-L, Cox R, Niaki M, Platt J, et al. (2014) Clinical whole-exome sequencing: are we there yet? Genetics in medicine 16: 717-719.

Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. The Journal of Experimental Medicine 79: 137-158.

Babin PJ, Goizet C, Raldua D (2014) Zebrafish models of human motor neuron diseases: advantages and limitations. Progress in Neurobiology 118: 36-58.

Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, et al. (2011) Whole-genome sequencing for optimized patient management. Science translational medicine 3: 87re83.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, et al. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nature reviews. Genetics 12: 745-755.

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, et al. (2015) Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proceedings of the National Academy of Sciences 112: 5473-5478.

Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, et al. (2011) Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. Science translational medicine 3: 65ra64-65ra64.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, et al. (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58: 268-276.

Bennett S (2004) Solexa Ltd. Pharmacogenomics 5: 433-438.

Bennett ST, Barnes C, Cox A, Davies L, Brown C (2005) Toward the 1,000 dollars human genome. Pharmacogenomics 6: 373-382.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53-59.

Berkenstadt M, Shiloh S, Barkai G, Katznelson MB, Goldman B (1999) Perceived personal control (PPC): a new concept in measuring outcome of genetic counseling. American journal of medical genetics 82: 53-59.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, et al. (2010) The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology 28: 1045-1048.

Bick D, Fraser PC, Gutzeit MF, Harris JM, Hambuch TM, et al. (2017) Successful Application of Whole Genome Sequencing in a Medical Genetics Clinic. Journal of pediatric genetics 6: 61-76.

Biesecker LG (2010) Exome sequencing makes medical genomics a reality. Nature Genetics 42: 13-14.

Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, et al. (2010) Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature 467: 207-210.

Bimonte S, De Angelis A, Quagliata L, Giusti F, Tammaro R, et al. (2011) Ofd1 is required in limb bud patterning and endochondral bone development. Developmental biology 349: 179-191.

Bollinger JM, Scott J, Dvoskin R, Kaufman D (2012) Public preferences regarding the return of individual genetic research results: findings from a qualitative focus group study. Genetics in medicine 14: 451-457.

Borck G, Beighton P, Wilhelm C, Kohlhase J, Kubisch C (2010) Arterial rupture in classic Ehlers-Danlos syndrome with COL5A1 mutation. American journal of medical genetics. Part A 152A: 2090-2093.

Botkin JR, Teutsch SM, Kaye CI, Hayes M, Haddow JE, et al. (2010) Outcomes of interest in evidence-based evaluations of genetic tests. Genetics in medicine 12: 228-235.

Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics 33 Suppl: 228-237.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics 32: 314-331.

Bowdin SC, Hayeems RZ, Monfared N, Cohn RD, Meyn MS (2016) The SickKids Genome Clinic: developing and evaluating a pediatric model for individualized genomic medicine. Clinical genetics 89: 10-19.

Boycott KM, Innes AM (2017) When One Diagnosis Is Not Enough. New England Journal of Medicine 376: 83-85.

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nature reviews. Genetics 14: 681-691.

Bracale A, Cesca F, Neubrand VE, Newsome TP, Way M, et al. (2007) Kidins220/ARMS is transported by a kinesin-1-based mechanism likely to be involved in neuronal differentiation. Molecular biology of the cell 18: 142-152.

Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth HV, et al. (2014) DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic acids research 42: D993-d1000.

Brooks GA, Brown MA, Butz CE, Sicurello JP, Dubouchaud H (1999) Cardiac and skeletal muscle mitochondria have a monocarboxylate transporter MCT1. Journal of applied physiology 87: 1713-1718.

Buenrostro J, Wu B, Chang H, Greenleaf W (2015) ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current protocols in molecular biology 109: 21.29.21-21.29.29.

Caldarini MI, Pons S, D'Agostino D, DePaula JA, Greco G, et al. (1996) Abnormal fecal flora in a patient with short bowel syndrome. An in vitro study on effect of pH on D-lactic acid production. Digestive diseases and sciences 41: 1649-1652.

Cammack R (1969) Assay, purification and properties of mammalian D-2-hydroxy acid dehydrogenase. The Biochemical journal 115: 55-64.

Campbell CD, Eichler EE (2013) Properties and rates of germline mutations in humans. Trends in genetics : TIG 29: 575-584.

Campeau PM, Kasperaviciute D, Lu JT, Burrage LC, Kim C, et al. (2014) The genetic basis of DOORS syndrome: an exome-sequencing study. The Lancet. Neurology 13: 44-58.

Cesca F, Yabe A, Spencer-Dene B, Arrigoni A, Al-Qatari M, et al. (2011) Kidins220/ARMS is an essential modulator of cardiovascular and nervous system development. Cell death & disease 2: e226.

Cesca F, Yabe A, Spencer-Dene B, Scholz-Starke J, Medrihan L, et al. (2012) Kidins220/ARMS mediates the integration of the neurotrophin and VEGF pathways in the vascular and nervous systems. Cell death and differentiation 19: 194-208.

Chaisson MJ, Wilson RK, Eichler EE (2015) Genetic variation and the de novo assembly of human genomes. Nature reviews. Genetics 16: 627-640.

Chaki M, Airik R, Ghosh AK, Giles RH, Chen R, et al. (2012) Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. Cell 150: 533-548.

Chang MS, Arevalo JC, Chao MV (2004) Ternary complex with Trk, p75, and an ankyrin-rich membrane spanning protein. Journal of neuroscience research 78: 186-192.

Chelstowska A, Liu Z, Jia Y, Amberg D, Butow RA (1999) Signalling between mitochondria and the nucleus regulates the expression of a new D-lactate dehydrogenase activity in yeast. Yeast 15: 1377-1391.

Chen CP, Chang TY, Chen CY, Wang TY, Tsai FJ, et al. (2012a) Short rib-polydactyly syndrome type II (Majewski): prenatal diagnosis, perinatal imaging findings and molecular analysis of the NEK1 gene. Taiwanese journal of obstetrics & gynecology 51: 100-105.

Chen JQ, Wu Y, Yang H, Bergelson J, Kreitman M, et al. (2009) Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Molecular biology and evolution 26: 1523-1531.

Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, et al. (2016) Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nature Biotechnology 34: 531-538.

Chen Y, Chen CF, Riley DJ, Chen PL (2011) Nek1 kinase functions in DNA damage response and checkpoint control through a pathway independent of ATM and ATR. Cell cycle 10: 655-663.

Chen Y, Fu WY, Ip JP, Ye T, Fu AK, et al. (2012b) Ankyrin repeat-rich membrane spanning protein (kidins220) is required for neurotrophin and ephrin receptor-dependent dendrite development. The Journal of neuroscience 32: 8263-8269.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 106: 19096-19101.

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, et al. (2015) The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. American journal of human genetics 97: 199-215.

Christopher MM, Eckfeldt JH, Eaton JW (1990) Propylene glycol ingestion causes D-lactic acidosis. Laboratory investigation; a journal of technical methods and pathology 62: 114-118.

Chua SC, Jr., Chung WK, Wu-Peng XS, Zhang Y, Liu SM, et al. (1996) Phenotypes of mouse diabetes and rat fatty due to mutations in the OB (leptin) receptor. Science 271: 994-996.

Church GM (2005) The Personal Genome Project. Molecular systems biology 1.

Clamp M, Fry B, Kamal M, Xie X, Cuff J, et al. (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proceedings of the National Academy of Sciences of the United States of America 104: 19428-19433.

Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, et al. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nature Genetics 37: 161-165.

Collins FS (1992) Positional cloning: Let's not call it reverse anymore. Nature Genetics 1: 3-6.

Collins FS (1995) Positional cloning moves from perditional to traditional. Nature Genetics 9: 347-350.

Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, et al. (2011) Variation in genome-wide mutation rates within and between human families. Nature Genetics 43: 712-714.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, et al. (2010) Origins and functional impact of copy number variation in the human genome. Nature 464: 704-712.

Consortium UK, Walter K, Min JL, Huang J, Crooks L, et al. (2015) The UK10K project identifies rare variants in health and disease. Nature 526.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, et al. (2011) A copy number variation morbidity map of developmental delay. Nature Genetics 43: 838-846.

Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, et al. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nature methods 7: 250-251.

Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, et al. (2005) Distribution and intensity of constraint in mammalian genomic sequence. Genome research 15: 901-913.

Corbit KC, Shyer AE, Dowdle WE, Gaulden J, Singla V, et al. (2008) Kif3a constrains beta-catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms. Nature Cell Biology 10: 70-76.

Costa T, Scriver CR, Childs B (1985) The effect of Mendelian disease on human health: a measurement. American journal of medical genetics 21: 231-242.

Cristescu ME, Innes DJ, Stillman JH, Crease TJ (2008) D- and L-lactate dehydrogenases during invertebrate evolution. BMC evolutionary biology 8: 268.

Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. Nature reviews. Genetics 1: 40-47.

Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, et al. (2017) Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Science translational medicine 9.

Dahl F, Stenberg J, Fredriksson S, Welch K, Zhang M, et al. (2007) Multigene amplification and massively parallel sequencing for cancer mutation discovery. Proceedings of the National Academy of Sciences 104: 9387-9392.

Dahm R (2005) Friedrich Miescher and the discovery of DNA. Developmental biology 278: 274-288.

Davis EE, Frangakis S, Katsanis N (2014) Interpreting human genetic variation with in vivo zebrafish assays. Biochimica et biophysica acta 1842: 1960-1970.

de Bari L, Atlante A, Guaragnella N, Principato G, Passarella S (2002) D-Lactate transport and metabolism in rat liver mitochondria. The Biochemical journal 365: 391-403.

de Leeuw K, Goorhuis JF, Tielliu IF, Symoens S, Malfait F, et al. (2012) Superior mesenteric artery aneurysm in a 9-year-old boy with classical Ehlers-Danlos syndrome. American journal of medical genetics. Part A 158A: 626-629.

de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, et al. (2012) Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. New England Journal of Medicine 367: 1921-1929.

de Paepe A, Malfait F (2012) The Ehlers-Danlos syndrome, a disorder with many faces. Clinical genetics 82: 1-11.

de Pagter MS, van Roosmalen MJ, Baas AF, Renkens I, Duran KJ, et al. (2015) Chromothripsis in healthy individuals affects multiple protein-coding genes and can result in severe congenital abnormalities in offspring. American journal of human genetics 96: 651-656.

de Vrese M, Koppenhoefer B, Barth CA (1990) D-lactic acid metabolism after an oral load of DL-lactate. Clinical Nutrition 9: 23-28.

Deciphering Developmental Disorders Study (2015) Large-scale discovery of novel genetic causes of developmental disorders. Nature 519: 223-228.

Deciphering Developmental Disorders Study (2017) Prevalence and architecture of de novo mutations in developmental disorders. Nature 542: 433-438.

Depienne C, Nava C, Keren B, Heide S, Rastetter A, et al. (2017) Genetic and phenotypic dissection of 1q43q44 microdeletion syndrome and neurodevelopmental phenotypes associated with mutations in ZBTB18 and HNRNPU. Human genetics 136: 463-479.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, et al. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research 37: e67.

Deverka PA, Dreyfus JC (2014) Clinical integration of next generation sequencing: coverage and reimbursement challenges. The Journal of law, medicine & ethics 42 Suppl 1: 22-41.

Diamandis EP, Li M (2016) The side effects of translational omics: overtesting, overdiagnosis, overtreatment. Clinical chemistry and laboratory medicine 54: 389-396.

Dixon-Salazar TJ, Silhavy JL, Udpa N, Schroth J, Bielas S, et al. (2012) Exome Sequencing Can Improve Diagnosis and Alter Patient Management. Science translational medicine 4: 138ra178.

Drivas TG, Wojno AP, Tucker BA, Stone EM, Bennett J (2015) Basal exon skipping and genetic pleiotropy: A predictive model of disease pathogenesis. Science Translational Medicine 7: 291ra297.

Dunlop RH, Hammond PB (1965) D-lactic acidosis of ruminants. Annals of the New York Academy of Sciences 119: 1109-1132.

Duran M, Van Biervliet JP, Kamerling JP, Wadman SK (1977) D-lactic aciduria, an inborn error of metabolism? Clinica chimica acta; international journal of clinical chemistry 74: 297-300.

Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. (2009) Real-Time DNA Sequencing from Single Polymerase Molecules. Science 323: 133-138.

El Hokayem J, Huber C, Couve A, Aziza J, Baujat G, et al. (2012) NEK1 and DYNC2H1 are both involved in short rib polydactyly Majewski type but not in Beemer Langer cases. Journal of medical genetics 49: 227-233.

Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, et al. (2017) Lessons learned from additional research analyses of unsolved clinical exome cases. Genome medicine 9: 26.

Ellingford JM, Campbell C, Barton S, Bhaskar S, Gupta S, et al. (2017) Validation of copy number variation analysis for next-generation sequencing diagnostics. European journal of human genetics 25: 719-724.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74.

Engbers HM, Berger R, van Hasselt P, de Koning T, de Sain-van der Velden MG, et al. (2008) Yield of additional metabolic studies in neurodevelopmental disorders. Annals of neurology 64: 212-217.

English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, et al. (2015) Assessing structural variation in a personal genome-towards a human reference diploid genome. BMC genomics 16: 286.

Erlich Y (2015) A vision for ubiquitous sequencing. Genome research 25: 1411-1416.

European Parliment (1999) Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. In: UNION TEPATCOTE (ed). Official Journal of the European Communities, Brussels, pp 0001-0005

Eurordis (2017) Rare Diseases Europe. http://www.eurordis.org. Accessed May 1 2017

Evans JP, Khoury MJ (2013) The arrival of genomic medicine to the clinic is only the beginning of the journey. Genetics in medicine 15: 268-269.

Fabian E, Kramer L, Siebert F, Hogenauer C, Raggam RB, et al. (2017) D-lactic acidosis - case report and review of the literature. Zeitschrift fur Gastroenterologie 55: 75-82.

Fakhro KA, Staudt MR, Ramstetter MD, Robay A, Malek JA, et al. (2016) The Qatar genome: a population-specific tool for precision medicine in the Middle East. Human Genome Variation 3: 16016.

Farooqi IS, Wangensteen T, Collins S, Kimber W, Matarese G, et al. (2007) Clinical and Molecular Genetic Spectrum of Congenital Deficiency of the Leptin Receptor. New England Journal of Medicine 356: 237-247.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, 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. Genetics in medicine 17: 578-586.

Fassier C, Hutt JA, Scholpp S, Lumsden A, Giros B, et al. (2010) Zebrafish atlastin controls motility and spinal motor axon architecture via inhibition of the BMP pathway. Nature neuroscience 13: 1380-1387.

Ferrante MI, Giorgio G, Feather SA, Bulfone A, Wright V, et al. (2001) Identification of the gene for oral-facial-digital type I syndrome. American Journal of Human Genetics 68: 569-576.

Ferrante MI, Romio L, Castro S, Collins JE, Goulding DA, et al. (2009) Convergent extension movements and ciliary function are mediated by ofd1, a zebrafish orthologue of the human oral-facial-digital type 1 syndrome gene. Human molecular genetics 18: 289-303.

Ferrante MI, Zullo A, Barra A, Bimonte S, Messaddeq N, et al. (2006) Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. Nature genetics 38: 112-117.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, et al. (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. The American Journal of Human Genetics 84: 524-533.

Fischbach BV, Trout KL, Lewis J, Luis CA, Sika M (2005) WAGR syndrome: a clinical review of 54 cases. Pediatrics 116: 984-988.

Fishbein WN (1986) Lactate transporter defect: a new disease of muscle. Science 234: 1254-1256.

Flick MJ, Konieczny SF (2002) Identification of putative mammalian d-lactate dehydrogenase enzymes. Biochemical and biophysical research communications 295: 910-916.

Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA (2014) Accuracy of Next Generation Sequencing Platforms. Next generation, sequencing & applications 1.

Franceschini P, Guala A, Vardeu MP, Signorile F, Franceschini D, et al. (1995) Short rib-dysplasia group (with/without polydactyly): report of a patient suggesting the existence of a continuous spectrum. American journal of medical genetics 59: 359-364.

Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, et al. (2015) Genome-wide patterns and properties of de novo mutations in humans. Nature Genetics 47: 822-826.

Freed D, Pevsner J (2016) The Contribution of Mosaic Variants to Autism Spectrum Disorder. PLoS genetics 12: e1006245.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, et al. (2014) De novo mutations in schizophrenia implicate synaptic networks. Nature 506: 179-184.

Fry AM, Leaper MJ, Bayliss R (2014) The primary cilium: guardian of organ development and homeostasis. Organogenesis 10: 62-68.

Fry AM, O'Regan L, Sabir SR, Bayliss R (2012) Cell cycle regulation by the NEK family of protein kinases. Journal of cell science 125: 4423-4433.

Genome Asia 100K (2017) Genome Asia 100K. http://www.genomeasia100k.com.

Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nature Genetics 46: 818-825.

Genome Reference Consortium (2017) Genome Reference Consortium. https://www.ncbi.nlm.nih.gov/grc/human. Accessed August 1 2017

Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.

Giesecke D, Fabritius A, Wallenberg Pv (1981) A quantitative study on the metabolism of d(–) lactic acid in the rat and the rabbit. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 69: 85-89.

Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, et al. (2014) Genome sequencing identifies major causes of severe intellectual disability. Nature 511: 344-347.

Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012) Disease gene identification strategies for exome sequencing. European journal of human genetics 20: 490-497.

Giorgio G, Alfieri M, Prattichizzo C, Zullo A, Cairo S, et al. (2007) Functional characterization of the OFD1 protein reveals a nuclear localization and physical interaction with subunits of a chromatin remodeling complex. Molecular biology of the cell 18: 4397-4404.

Gleason FH, Nolan RA (1966) D(-)-lactate dehydrogenase in lower fungi. Science 152: 1272-1273.

Glenn TC (2011) Field guide to next-generation DNA sequencers. Molecular ecology resources 11: 759-769.

Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC (2011) Kaviar: an accessible system for testing SNV novelty. Bioinformatics 27: 3216-3217.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America 108: 1513-1518.

Goldfeder RL, Priest JR, Zook JM, Grove ME, Waggott D, et al. (2016) Medical implications of technical accuracy in genome sequencing. Genome medicine 8: 24.

Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, et al. (2016) Parent-of-origin-specific signatures of de novo mutations. Nature Genetics 48: 935-939.

Gonorazky H, Liang M, Cummings B, Lek M, Micallef J, et al. (2016) RNAseq analysis for the diagnosis of muscular dystrophy. Annals of clinical and translational neurology 3: 55-60.

Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. Annual review of medicine 63: 35-61.

Gonzalez-Perez A, Lopez-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. American journal of human genetics 88: 440-449.

Gray LR, Tompkins SC, Taylor EB (2014) Regulation of pyruvate metabolism and human disease. Cellular and molecular life sciences 71: 2577-2604.

Greenspan DS, Cheng W, Hoffman GG (1991) The pro-alpha 1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-alpha 1(XI) collagen chain. The Journal of biological chemistry 266: 24727-24733.

Grosse SD, Khoury MJ (2006) What is the clinical utility of genetic testing? Genetics in medicine 8: 448-450.

Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, et al. (2015) Large-scale whole-genome sequencing of the Icelandic population. Nature Genetics 47: 435-444.

Guo L, Milburn MV, Ryals JA, Lonergan SC, Mitchell MW, et al. (2015) Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. Proceedings of the National Academy of Sciences of the United States of America 112: E4901-4910.

Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, et al. (2017) Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. Genomics 109: 83-90.

Gurrieri F, Franco B, Toriello H, Neri G (2007) Oral-facial-digital syndromes: review and diagnostic guidelines. American journal of medical genetics. Part A 143a: 3314-3323.

Haldane JB (1947) The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. Annals of eugenics 13: 262-271.

Halestrap AP (2013) The SLC16 gene family - structure, role and regulation in health and disease. Molecular aspects of medicine 34: 337-349.

Hall JG, Powers EK, McLlvaine RT, Ean VH (1978) The frequency and financial burden of genetic disease in a pediatric hospital. American journal of medical genetics 1: 417-436.

Hallett JW, Jr., Marshall DM, Petterson TM, Gray DT, Bower TC, et al. (1997) Graft-related complications after abdominal aortic aneurysm repair: reassurance from a 36-year population-based experience. Journal of vascular surgery 25: 277-284; discussion 285-276.

Halperin ML, Kamel KS (1996) D-lactic acidosis: turning sugar into acids in the gastrointestinal tract. Kidney international 49: 1-8.

Halpern ME, Rhee J, Goll MG, Akitake CM, Parsons M, et al. (2008) Gal4/UAS transgenic tools and their application to zebrafish. Zebrafish 5: 97-110.

Halverson J, Gale A, Lazarus C (1984) D-lactic acidosis and other complications of intestinal bypass surgery. Archives of internal medicine 144: 357-360.

Hamdan FF, Srour M, Capo-Chichi JM, Daoud H, Nassif C, et al. (2014) De novo mutations in moderate or severe intellectual disability. PLoS genetics 10: e1004772.

Harakalova M, Mokry M, Hrdlickova B, Renkens I, Duran K, et al. (2011) Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. Nature protocols 6: 1870-1886.

Harakalova M, van Harssel JJ, Terhal PA, van Lieshout S, Duran K, et al. (2012) Dominant missense mutations in ABCC9 cause Cantu syndrome. Nature Genetics 44: 793-796.

Haraksingh RR, Snyder MP (2013) Impacts of variation in the human genome on gene regulation. Journal of molecular biology 425: 3970-3977.

Hasegawa H, Fukushima T, Lee JA, Tsukamoto K, Moriya K, et al. (2003) Determination of serum D-lactic and L-lactic acids in normal subjects and diabetic patients by column-switching HPLC with pre-column fluorescence derivatization. Analytical and bioanalytical chemistry 377: 886-891.

Hayden EC (2014) Technology: The $1,000 genome. Nature 507: 294-295.

Heemskerk MM, van Harmelen VJA, van Dijk KW, van Klinken JB (2016) Reanalysis of mGWAS results and in vitro validation show that lactate dehydrogenase interacts with branched-chain amino acid metabolism. European journal of human genetics 24: 142-145.

Henderson LB, Applegate CD, Wohler E, Sheridan MB, Hoover-Fong J, et al. (2014) The impact of chromosomal microarray on clinical management: a retrospective analysis. Genetics in medicine 16: 657-664.

Henry H, Marmy Conus N, Steenhout P, Beguin A, Boulat O (2012) Sensitive determination of D-lactic acid and L-lactic acid in urine by high-performance liquid chromatography-tandem mass spectrometry. Biomedical chromatography : BMC 26: 425-428.

Hilton LK, White MC, Quarmby LM (2009) The NIMA-related kinase NEK1 cycles through the nucleus. Biochemical and biophysical research communications 389: 52-56.

Hochstenbach R, van Binsbergen E, Engelen J, Nieuwint A, Polstra A, et al. (2009) Array analysis and karyotyping: workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. European journal of medical genetics 52: 161-169.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, et al. (2007) Genome-wide in situ exon capture for selective resequencing. Nature Genetics 39: 1522-1527.

Hoischen A, Gilissen C, Arts P, Wieskamp N, van der Vliet W, et al. (2010a) Massively parallel sequencing of ataxia genes after array-based enrichment. Human Mutation 31: 494-499.

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, et al. (2010b) De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nature Genetics 42: 483-485.

Hollaar L, Van der Laarse A (1979) Interference of the measurement of lactate dehydrogenase (LDH) activity in human serum and plasma by LDH from blood cells. Clinica chimica acta; international journal of clinical chemistry 99: 135-142.

Huber C, Cormier-Daire V (2012) Ciliary disorder of the skeleton. American journal of medical genetics. Part C, Seminars in medical genetics 160c: 165-174.

Huddleston J, Eichler EE (2016) An Incomplete Understanding of Human Genetic Variation. Genetics 202: 1251-1254.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, et al. (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nature Genetics 46: 205-212.

Human Genome Sequencing Consortium International (2004) Finishing the euchromatic sequence of the human genome. Nature 431: 931-945.

Hunkapiller T, Kaiser RJ, Koop BF, Hood L (1991) Large-scale and automated DNA sequence determination. Science 254: 59-67.

Iglesias A, Anyane-Yeboa K, Wynn J, Wilson A, Truitt Cho M, et al. (2014) The usefulness of whole-exome sequencing in routine clinical practice. Genetics in medicine 16: 922-931.

Iglesias T, Cabrera-Poch N, Mitchell MP, Naven TJ, Rozengurt E, et al. (2000) Identification and cloning of Kidins220, a novel neuronal substrate of protein kinase D. The Journal of biological chemistry 275: 40048-40056.

Illumina Inc (2017) Illumina Introduces the NovaSeq Series - a New Architecture Designed to Usher in the $100 Genome. Business Wire, San Diego

Ionita-Laza I, McCallum K, Xu BIN, Buxbaum J (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nature Genetics 48: 214-220.

Iossifov I, O/'Roak BJ, Sanders SJ, Ronemus M, Krumm N, et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder. Nature 515: 216-221.

Jakobs C, Solem E, Ek J, Halvorsen K, Jellum E (1977) Investigation of the metabolic pattern in maple syrup urine disease by means of glass capillary gas chromatography and mass spectrometry. Journal of Chromatography B: Biomedical Sciences and Applications 143: 31-38.

Jones MA, Bhide S, Chin E, Ng BG, Rhodenizer D, et al. (2011) Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. Genetics in medicine 13: 921-932.

Jorens PG, Demey HE, Schepens PJ, Coucke V, Verpooten GA, et al. (2004) Unusual D-lactic acid acidosis from propylene glycol metabolism in overdose. Journal of toxicology. Clinical toxicology 42: 163-169.

Joshi C, Kolbe DL, Mansilla MA, Mason SO, Smith RJ, et al. (2016) Reducing the Cost of the Diagnostic Odyssey in Early Onset Epileptic Encephalopathies. BioMed research international 2016: 6421039.

Kainulainen K, Pulkkinen L, Savolainen A, Kaitila I, Peltonen L (1990) Location on Chromosome 15 of the Gene Defect Causing Marfan Syndrome. New England Journal of Medicine 323: 935-939.

Kalapos MP (1999) Methylglyoxal in living organisms: chemistry, biochemistry, toxicology and biological implications. Toxicology letters 110: 145-175.

Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, et al. (2017) Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genetics in medicine 19: 249-255.

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of bioscience and bioengineering 96: 317-323.

Karaa A, Stoler JM (2013) Ehlers Danlos Syndrome: An Unusual Presentation You Need to Know about. Case reports in pediatrics 2013: 764659.

Kataoka M, Matoba N, Sawada T, Kazuno AA, Ishiwata M, et al. (2016) Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. Molecular psychiatry 21: 885-893.

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, et al. (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245: 1073-1080.

Kernie SG, Liebl DJ, Parada LF (2000) BDNF regulates eating behavior and locomotor activity in mice. The EMBO journal 19: 1290-1300.

Kettleborough RNW, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, et al. (2013) A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 496: 494-497.

Khan AO, Becirovic E, Betz C, Neuhaus C, Altmuller J, et al. (2017) A deep intronic CLRN1 (USH3A) founder mutation generates an aberrant exon and underlies severe Usher syndrome on the Arabian Peninsula. Scientific reports 7: 1411.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics 46: 310-315.

Kirk P, Wilson MC, Heddle C, Brown MH, Barclay AN, et al. (2000) CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression. The EMBO journal 19: 3896-3904.

Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, et al. (2011) Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nature biotechnology 29: 59-63.

Kochhar S, Hunziker PE, Leong-Morgenthaler P, Hottinger H (1992) Evolutionary relationship of NAD(+)-dependent D-lactate dehydrogenase: comparison of primary structure of 2-hydroxy acid dehydrogenases. Biochemical and biophysical research communications 184: 60-66.

Kohler JN, Turbitt E, Biesecker BB (2017) Personal utility in genomic testing: a systematic literature review. European journal of human genetics 25: 662-668.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. (2012) Rate of de novo mutations and the importance of father/'s age to disease risk. Nature 488: 471-475.

Kong H, Boulter J, Weber JL, Lai C, Chao MV (2001) An evolutionarily conserved transmembrane protein that is a novel downstream target of neurotrophin and ephrin receptors. The Journal of neuroscience 21: 176-185.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318: 420-426.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, et al. (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nature Biotechnology 30: 693-700.

Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, et al. (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. Nature Communications 8: 15824.

Kroes HY, Monroe GR, van der Zwaag B, Duran KJ, de Kovel CG, et al. (2016) Joubert syndrome: genotyping a Northern European patient cohort. European journal of human genetics 24: 214-220.

Kruse O, Grunnet N, Barfod C (2011) Blood lactate as a predictor for in-hospital mortality in patients admitted acutely to hospital: a systematic review. Scandinavian journal of trauma, resuscitation and emergency medicine 19: 74.

Kuechler A, Willemsen MH, Albrecht B, Bacino CA, Bartholomew DW, et al. (2015) De novo mutations in beta-catenin (CTNNB1) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. Human genetics 134: 97-109.

Kuhlenbäumer G, Hullmann J, Appenzeller S (2011) Novel genomic techniques open new avenues in the analysis of monogenic disorders. Human Mutation 32.

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

Kumar P, Radhakrishnan J, Chowdhary MA, Giampietro PF (2001) Prevalence and patterns of presentation of genetic disorders in a pediatric emergency department. Mayo Clinic proceedings 76: 777-783.

Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, et al. (2007) The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Developmental Dynamics 236: 3088-3099.

Laffel L (1999) Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. Diabetes/metabolism research and reviews 15: 412-426.

Landaas S, Jakobs C (1977) The occurrence of 2-hydroxyisovaleric acid in patients with lactic acidosis and ketoacidosis. Clinica chimica acta; international journal of clinical chemistry 78: 489-493.

Lander ES (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236: 1567-1570.

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, et al. (2014) Clinical exome sequencing for genetic identification of rare Mendelian disorders. Jama 312: 1880-1887.

Lee M, Roos P, Sharma N, Atalar M, Evans TA, et al. (2017) Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. The American Journal of Human Genetics 100: 751-765.

Lee Y, Morrison BM, Li Y, Lengacher S, Farah MH, et al. (2012) Oligodendroglia metabolically support axons and contribute to neurodegeneration. Nature 487: 443-448.

Leiden University Medical Center (2008) First Female DNA Sequenced. ScienceDaily. https://www.sciencedaily.com/releases/2008/05/080526155300.htm. Accessed August 1 2017

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285-291.

Lengacher S, Nehiri-Sitayeb T, Steiner N, Carneiro L, Favrod C, et al. (2013) Resistance to Diet-Induced Obesity and Associated Metabolic Perturbations in Haploinsufficient Monocarboxylate Transporter 1 Mice. PloS one 8: e82505.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, et al. (2007) The diploid genome sequence of an individual human. PLoS Biol 5.

Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, et al. (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome biology 11: R22.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595.

Liao G-Y, Li Y, Xu B (2013) Ablation of TrkB expression in RGS9-2 cells leads to hyperphagic obesity. Molecular Metabolism 2: 491-497.

Liebich HM, Forst C (1984) Hydroxycarboxylic and oxocarboxylic acids in urine: products from branched-chain amino acid degradation and from ketogenesis. Journal of chromatography 309: 225-242.

Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proceedings of the National Academy of Sciences 108: 11093-11098.

Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, et al. (2011) A Mosaic Activating Mutation in AKT1 Associated with the Proteus Syndrome. New England Journal of Medicine 365: 611-619.

Ling B, Peng F, Alcorn J, Lohmann K, Bandy B, et al. (2012) D-Lactate altered mitochondrial energy production in rat brain and heart but not liver. Nutrition & Metabolism 9: 6-6.

Littink KW, Pott JW, Collin RW, Kroes HY, Verheij JB, et al. (2010) A novel nonsense mutation in CEP290 induces exon skipping and leads to a relatively mild retinal phenotype. Investigative ophthalmology & visual science 51: 3646-3652.

Liu B, Chen S, Johnson C, Helms JA (2014) A ciliopathy with hydrocephalus, isolated craniosynostosis, hypertelorism, and clefting caused by deletion of Kif3a. Reproductive toxicology 48: 88-97.

Loman NJ, Watson M (2015) Successful test launch for nanopore sequencing. Nature methods 12: 303-304.

Lopez E, Thauvin-Robinet C, Reversade B, Khartoufi NE, Devisme L, et al. (2014) C5orf42 is the major gene responsible for OFD syndrome type VI. Human genetics 133: 367-377.

Lopez-Menendez C, Gascon S, Sobrado M, Vidaurre OG, Higuero AM, et al. (2009) Kidins220/ARMS downregulation by excitotoxic activation of NMDARs reveals its involvement in neuronal survival and death pathways. Journal of cell science 122: 3554-3565.

Lunshof JE, Bobe J, Aach J, Angrist M, V. Thakuria J, et al. (2010) Personal genomes in progress: from the Human Genome Project to the Personal Genome Project. Dialogues in Clinical Neuroscience 12: 47-60.

Lupski JR (2016) Clinical genomics: from a truly personal genome viewpoint. Human genetics 135: 591-601.

Lupski JR, Gonzaga-Jauregui C, Yang Y, Bainbridge MN, Jhangiani S, et al. (2013) Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. Genome medicine 5: 57.

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, et al. (2010) Whole-Genome Sequencing in a Patient with Charcot–Marie–Tooth Neuropathy. New England Journal of Medicine 362: 1181-1191.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, et al. (2014) Guidelines for investigating causality of sequence variants in human disease. Nature 508: 469-476.

Macca M, Franco B (2009) The molecular basis of oral-facial-digital syndrome, type 1. American journal of medical genetics. Part C, Seminars in medical genetics 151c: 318-325.

Magenis RE, Maslen CL, Smith L, Allen L, Sakai LY (1991) Localization of the fibrillin (FBN) gene to chromosome 15, band q21.1. Genomics 11: 346-351.

Mahjoub MR, Trapp ML, Quarmby LM (2005) NIMA-related kinases defective in murine models of polycystic kidney diseases localize to primary cilia and centrosomes. Journal of the American Society of Nephrology 16: 3485-3489.

Malfait F, Symoens S, De Backer J, Hermanns-Le T, Sakalihasan N, et al. (2007) Three arginine to cysteine substitutions in the pro-alpha (I)-collagen chain cause Ehlers-Danlos syndrome with a propensity to arterial rupture in early adulthood. Human Mutation 28: 387-395.

Malfait F, Wenstrup RJ, De Paepe A (2010) Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. Genetics in medicine 12: 597-605.

Manoharan C, Wilson MC, Sessions RB, Halestrap AP (2006) The role of charged residues in the transmembrane helices of monocarboxylate transporter 1 and its ancillary protein basigin in determining plasma membrane

expression and catalytic activity. Molecular Membrane Biology 23: 486-498.

Manousaki D, Dudding T, Haworth S, Hsu YH, Liu CT, et al. (2017) Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. American journal of human genetics 101: 227-238.

Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, et al. (2016) Genetic Misdiagnoses and the Potential for Health Disparities. New England Journal of Medicine 375: 655-665.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380.

Marini JC, Forlino A, Cabral WA, Barnes AM, San Antonio JD, et al. (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Human Mutation 28: 209-221.

Marquard J, Welters A, Buschmann T, Barthlen W, Vogelgesang S, et al. (2013) Association of exercise-induced hyperinsulinaemic hypoglycaemia with MCT1-expressing insulinoma. Diabetologia 56: 31-35.

Marracci S, Giannini M, Vitiello M, Andreazzoli M, Dente L (2013) Kidins220/ARMS is dynamically expressed during Xenopus laevis development. The International journal of developmental biology 57: 787-792.

Marston S (2017) Obscurin variants and inherited cardiomyopathies. Biophysical reviews.

Marszalek JR, Ruiz-Lozano P, Roberts E, Chien KR, Goldstein LS (1999) Situs inversus and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II. Proceedings of the National Academy of Sciences of the United States of America 96: 5043-5048.

Marti R, Varela E, Segura RM, Alegre J, Surinach JM, et al. (1997) Determination of D-lactate by enzymatic methods in biological fluids: study of interferences. Clinical chemistry 43: 1010-1015.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America 74: 560-564.

McCandless SE, Brunger JW, Cassidy SB (2004) The burden of genetic disease on inpatient care in a children's hospital. American journal of human genetics 74: 121-127.

McCullagh KJ, Poole RC, Halestrap AP, O'Brien M, Bonen A (1996) Role of the lactate transporter (MCT1) in skeletal muscles. The American journal of physiology 271: E143-150.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20.

Meder B, Haas J, Keller A, Heid C, Just S, et al. (2011) Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. Circulation. Cardiovascular genetics 4: 110-122.

Mehta S, Dhar SU, Birnbaum Y (2012) Common iliac artery aneurysm and spontaneous dissection with contralateral iatrogenic common iliac artery dissection in classic ehlers-danlos syndrome. The International journal of angiology 21: 167-170.

Meirelles GV, Perez AM, de Souza EE, Basei FL, Papa PF, et al. (2014) "Stop Ne(c)king around": How interactomics contributes to functionally characterize Nek family kinases. World journal of biological chemistry 5: 141-160.

Mendel G (1996) Experiments in plant hybridization (1865).

Merezhinskaya N, Fishbein WN (2009) Monocarboxylate transporters: past, present, and future. Histology and histopathology 24: 243-264.

Merezhinskaya N, Fishbein WN, Davis JI, Foellmer JW (2000) Mutations in MCT1 cDNA in patients with symptomatic deficiency in lactate transport. Muscle & nerve 23: 90-97.

Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, et al. (2017) Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genetics in medicine.

Miescher-Rüsch F (1871) Ueber die chemische Zusammensetzung der Eiterzellen

Miller JN, Pearce DA (2014) Nonsense-mediated decay in genetic disease: friend or foe? Mutation Research/ Reviews in Mutation Research 762: 52-64.

Miller MJ, Kennedy AD, Eckhart AD, Burrage LC, Wulff JE, et al. (2015) Untargeted metabolomic analysis for the clinical screening of inborn errors of metabolism. Journal of inherited metabolic disease 38: 1029-1039.

Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, et al. (2015) Comparison of predicted and actual consequences of missense mutations. Proceedings of the National Academy of Sciences 112: E5189-E5198.

Mirzaa GM, Campbell CD, Solovieff N, Goold C, Jansen LA, et al. (2016) Wide spectrum of developmental brain disorders from megalencephaly to focal cortical dysplasia and pigmentary mosaicism caused by mutations of MTOR. JAMA neurology 73: 836-845.

Mitra RD, Church GM (1999) In situ localized amplification and contact replication of many individual DNA molecules. Nucleic acids research 27: e34.

Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM (2003) Fluorescent in situ sequencing on polymerase colonies. Analytical biochemistry 320: 55-65.

Monroe GR, Harakalova M, van der Crabben SN, Majoor-Krakauer D, Bertoli-Avella AM, et al. (2015) Familial Ehlers-Danlos syndrome with lethal arterial events caused by a mutation in COL5A1. American journal of medical genetics. Part A 167: 1196-1203.

Motonaka J, Katumoto Y, Ikeda S (1998) Preparation and properties of enzyme sensors for l-lactic and d-lactic acids in optical isomers. Analytica Chimica Acta 368: 91-95.

Nagai A, Takebe K, Nio-Kobayashi J, Takahashi-Iwanaga H, Iwanaga T (2010) Cellular expression of the monocarboxylate transporter (MCT) family in the placenta of mice. Placenta 31: 126-133.

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, et al. (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478: 57-63.

Narahara K, Kikkawa K, Kimira S, Kimoto H, Ogata M, et al. (1984) Regional mapping of catalase and Wilms tumor--aniridia, genitourinary abnormalities, and mental retardation triad loci to the chromosome segment 11p1305----p1306. Human genetics 66: 181-185.

Narula RK, El Shafei A, Ramaiah D, Schmitz PG (2000) D-lactic acidosis 23 years after jejuno-ileal bypass. American journal of kidney diseases 36: E9.

Nederbragt L (2017) Developments in high throughput sequencing - July 2016 edition. https://flxlexblog.wordpress.com/2016/07/08/developments-in-high-throughput-sequencing-july-2016-edition/. Accessed June 25 2017

Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, et al. (2012) Clinical application of exome sequencing in undiagnosed genetic conditions. Journal of Medical Genetics.

Neubrand VE, Cesca F, Benfenati F, Schiavo G (2012) Kidins220/ARMS as a functional mediator of multiple receptor signalling pathways. Journal of cell science 125: 1845-1854.

Neubrand VE, Thomas C, Schmidt S, Debant A, Schiavo G (2010) Kidins220/ARMS regulates Rac1-dependent neurite outgrowth by direct interaction with the RhoGEF Trio. Journal of cell science 123: 2111-2123.

Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, et al. (2013) A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Human Mutation 34: 1721-1726.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, et al. (2010a) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics 42: 790-793.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al. (2010b) Exome sequencing identifies the cause of a mendelian disorder. Nature Genetics 42: 30-35.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272-276.

NHLBI GO Exome Sequencing Project (ESP) (2011) Exome Variant Server. http://evs.gs.washington.edu/EVS.

Nicolaou N, Pulit SL, Nijman IJ, Monroe GR, Feitz WF, et al. (2016) Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. Kidney international 89: 476-486.

Nijman IJ (2017) personal communication.

Nijman IJ, Mokry M, van Boxtel R, Toonen P, de Bruijn E, et al. (2010) Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. Nature methods 7: 913-915.

Nyren P, Pettersson B, Uhlen M (1993) Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Analytical biochemistry 208: 171-175.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature Genetics 43: 585-589.

Ockeloen CW, Willemsen MH, de Munnik S, van Bon BW, de Leeuw N, et al. (2014) Further delineation of the KBG syndrome phenotype caused by ANKRD11 aberrations. European journal of human genetics.

Oh MS, Phelps KR, Traube M, Barbosa-Saldivar JL, Boxhill C, et al. (1979) D-Lactic Acidosis in a Man with the Short-Bowel Syndrome. New England Journal of Medicine 301: 249-252.

Oh MS, Uribarri J, Alveranga D, Lazar I, Bazilinski N, et al. (1985) Metabolic utilization and renal handling of D-lactate in men. Metabolism: clinical and experimental 34: 621-625.

Olson ST, Massey V (1979) Purification and properties of the flavoenzyme D-lactate dehydrogenase from Megasphaera elsdenii. Biochemistry 18: 4714-4724.

Ong KT, Perdu J, De Backer J, Bozec E, Collignon P, et al. (2010) Effect of celiprolol on prevention of cardiovascular events in vascular Ehlers-Danlos syndrome: a prospective randomised, open, blinded-endpoints trial. Lancet 376: 1476-1484.

Online Mendelian Inheritance in Man O (2017). https://omim.org/. Accessed June 26, 2017 2017

Otonkoski T, Jiao H, Kaminen-Ahola N, Tapia-Paez I, Ullah MS, et al. (2007) Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells. American journal of human genetics 81: 467-474.

Patil M, Pabla N, Huang S, Dong Z (2013) Nek1 phosphorylates Von Hippel-Lindau tumor suppressor to promote its proteasomal degradation and ciliary destabilization. Cell cycle 12: 166-171.

Patten SA, Armstrong GA, Lissouba A, Kabashi E, Parker JA, et al. (2014) Fishing for causes and cures of motor neuron disorders. Disease model & mechanisms 7: 799-809.

Patton MA, Krywawych S, Winter RM, Brenton DP, Baraitser M (1987) DOOR syndrome (deafness, onycho-osteodystrophy, and mental retardation): elevated plasma and urinary 2-oxoglutarate in three unrelated patients. American journal of medical genetics 26: 207-215.

Pepin M, Schwarze U, Superti-Furga A, Byers PH (2000) Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type. The New England journal of medicine 342: 673-680.

Pepin MB, PH. (1999) GeneReviews®. http://www.ncbi.nlm.nih.gov/books/NBK1494. 1993–2014

Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, et al. (2012) Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 487: 190-195.

Petersen C (2005) D-lactic acidosis. Nutrition in clinical practice : official publication of the American Society for Parenteral and Enteral Nutrition 20: 634-645.

Pfundt R, del Rosario M, Vissers LELM, Kwint MP, Janssen IM, et al. (2017) Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. Genetics in medicine 19: 667-675.

Phillips KA, Ann Sakowski J, Trosman J, Douglas MP, Liang SY, et al. (2014) The economic value of personalized medicine tests: what we know and what we need to know. Genetics in medicine 16: 251-257.

Phillips KA, Deverka PA, Trosman JR, Douglas MP, Chambers JD, et al. (2017) Payer coverage policies for multigene tests. Nature Biotechnology 35: 614-617.

Phillips KA, Pletcher MJ, Ladabaum U (2015) Is the "$1000 Genome" Really $1000? Understanding the Full Benefits and Costs of Genomic Sequencing. Technology and health care 23: 373-379.

Pleumeekers HJ, Hoes AW, van der Does E, van Urk H, Hofman A, et al. (1995) Aneurysms of the abdominal aorta in older adults. The Rotterdam Study. American journal of epidemiology 142: 1291-1299.

Polci R, Peng A, Chen PL, Riley DJ, Chen Y (2004) NIMA-related protein kinase 1 is involved early in the ionizing radiation-induced DNA damage response. Cancer research 64: 8800-8803.

Porreca GJ, Zhang K, Li JB, Xie B, Austin D, et al. (2007) Multiplex amplification of large sets of human exons. Nature methods 4: 931-936.

Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, et al. (2017) Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. The New England journal of medicine 376: 21-31.

Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, et al. (2016) Molecular diagnostic experience of whole-exome sequencing in adult patients. Genetics in medicine 18: 678-685.

Powles-Glover N (2014) Cilia and ciliopathies: classic examples linking phenotype and genotype-an overview. Reproductive toxicology 48: 98-105.

Rabi I, Zacharias J, Millman S, Kusch P (1938) A New Method of Measuring Nuclear Magnetic Moment. Physical Review 53: 318.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, et al. (2016) Timing, rates and spectra of human germline mutation. Nature Genetics 48: 126-133.

Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, et al. (2013) DeNovoGear: de novo indel and point mutation discovery and phasing. Nature Methods 10: 985-987.

Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, et al. (2006) Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. American journal of medical genetics. Part A 140: 2063-2074.

Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, et al. (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380: 1674-1682.

Ravitsky V, Wilfond BS (2006) Disclosing individual genetic results to research participants. The American journal of bioethics 6: 8-17.

Reid ES, Papandreou A, Drury S, Boustred C, Yue WW, et al. (2016) Advantages and pitfalls of an extended gene panel for investigating complex neurometabolic phenotypes. Brain.

Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, et al. (2016) Clinical application of whole-exome sequencing across clinical indications. Genetics in medicine 18: 696-704.

Reuter MS, Tawamie H, Buchert R, Hosny Gebril O, Froukh T, et al. (2017) Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders. JAMA psychiatry 74: 293-299.

Riccardi VM, Sujansky E, Smith AC, Francke U (1978) Chromosomal imbalance in the Aniridia-Wilms' tumor association: 11p interstitial deletion. Pediatrics 61: 604-610.

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066-1073.

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328: 636-639.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nature Biotechnology 29: 24-26.

Romani M, Micalizzi A, Kraoua I, Dotti MT, Cavallin M, et al. (2014) Mutations in B9D1 and MKS1 cause mild Joubert syndrome: expanding the genetic overlap with the lethal ciliopathy Meckel syndrome. Orphanet journal of rare diseases 9: 72.

Romio L, Fry AM, Winyard PJ, Malcolm S, Woolf AS, et al. (2004) OFD1 is a centrosomal/basal body protein expressed during mesenchymal-epithelial transition in human nephrogenesis. Journal of the American Society of Nephrology 15: 2556-2568.

Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P (1996) Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry 242: 84-89.

Rontgen WC (1896) On a new kind of rays. Science 3: 227-231.

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. (2013) Characterizing and measuring bias in sequence data. Genome biology 14: R51.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475: 348-352.

Roulet M, Ruggiero F, Karsenty G, LeGuellec D (2007) A comprehensive study of the spatial and temporal expression of the col5a1 gene in mouse embryos: a clue for understanding collagen V function in developing connective tissues. Cell and tissue research 327: 323-332.

Rowland TJ, Graw SL, Sweet ME, Gigli M, Taylor MR, et al. (2016) Obscurin Variants in Patients With Left Ventricular Noncompaction. Journal of the American College of Cardiology 68: 2237-2238.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, et al. (1986) Cloning the gene for an inherited human disorder-chronic granulomatous disease-on the basis of its chromosomal location. Nature 322: 32-38.

Sabatini LM, Mathews C, Ptak D, Doshi S, Tynan K, et al. (2016) Genomic Sequencing Procedure Microcosting Analysis and Health Economic Cost-Impact Analysis: A Report of the Association for Molecular Pathology. The Journal of molecular diagnostics 18: 319-328.

Sakai I, Sharief FS, Pan YC, Li SS (1987) The cDNA and protein sequences of human lactate dehydrogenase B. The Biochemical journal 248: 933-936.

Sakai LY, Keene DR, Engvall E (1986) Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. The Journal of cell biology 103: 2499-2509.

Sakalihasan N, Limet R, Defawe OD (2005) Abdominal aortic aneurysm. Lancet 365: 1577-1589.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74: 5463-5467.

Sasaki S, Futagi Y, Kobayashi M, Ogura J, Iseki K (2015) Functional characterization of 5-oxoproline transport via SLC16A1/MCT1. The Journal of biological chemistry 290: 2303-2311.

Sass JO (2012) Inborn errors of ketogenesis and ketone body utilization. Journal of inherited metabolic disease 35: 23-28.

Satir P, Pedersen LB, Christensen ST (2010) The primary cilium at a glance. Journal of cell science 123: 499-503.

Sauna ZE, Kimchi-Sarfaty C (2011) Understanding the contribution of synonymous mutations to human disease. Nature reviews. Genetics 12: 683-691.

Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, et al. (2016) Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clinical genetics 89: 275-284.

Scheijen JL, Hanssen NM, van de Waarenburg MP, Jonkers DM, Stehouwer CD, et al. (2012) L(+) and D(-) lactate are increased in plasma and urine samples of type 2 diabetes as measured by a simultaneous quantification of L(+) and D(-) lactate by reversed-phase liquid chromatography tandem mass spectrometry. Experimental diabetes research 2012: 234812.

Schmieg N, Thomas C, Yabe A, Lynch DS, Iglesias T, et al. (2015) Novel Kidins220/ARMS Splice Isoforms: Potential Specific Regulators of Neuronal and Cardiovascular Development. PloS one 10: e0129944.

Schofield D, Alam K, Douglas L, Shrestha R, MacArthur DG, et al. (2017) Cost-effectiveness of massively parallel sequencing for diagnosis of paediatric muscle diseases. Genomic Medicine 2: 4.

Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, et al. (2010) Complete Khoisan and Bantu genomes from southern Africa. Nature 463: 943-947.

Seaby EG, Pengelly RJ, Ennis S (2016) Exome sequencing explained: a practical guide to its clinical application. Briefings in functional genomics 15: 374-384.

Seheult J, Fitzpatrick G, Boran G (2016) Lactic acidosis: an update. Clinical chemistry and laboratory medicine.

Seo J-S, Rhie A, Kim J, Lee S, Sohn M-H, et al. (2016) De novo assembly and phasing of a Korean human genome. Nature 538: 243-247.

Shalom O, Shalva N, Altschuler Y, Motro B (2008) The mammalian Nek1 kinase is involved in primary cilium formation. FEBS letters 582: 1465-1470.

Shamseldin HE, Rajab A, Alhashem A, Shaheen R, Al-Shidi T, et al. (2013) Mutations in DDX59 implicate RNA helicase in the pathogenesis of orofaciodigital syndrome. American Journal of Human Genetics 93: 555-560.

Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, et al. (2014) The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genetics in medicine 16: 176-182.

Shen P, Wang W, Krishnakumar S, Palm C, Chi A-K, et al. (2011) High-quality DNA sequence capture of 524 disease candidate genes. Proceedings of the National Academy of Sciences 108: 6549-6554.

Shendure J, Aiden EL (2012) The expanding scope of DNA sequencing. Nature Biotechnology 30: 1084-1094.

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology 26: 1135-1145.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728-1732.

Silengo MC, Bell GL, Biagioli M, Franceschini P (1987) Oro-facial-digital syndrome II. Transitional type between the Mohr and the Majewski syndromes: report of two new cases. Clinical genetics 31: 331-336.

Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, et al. (2017) The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nature Genetics 49: 1167-1173.

Singleton AB (2011) Exome sequencing: a transformative technology. The Lancet. Neurology 10: 942-946.

Slaats GG, Saldivar JC, Bacal J, Zeman MK, Kile AC, et al. (2015) DNA replication stress underlies renal phenotypes in CEP290-associated Joubert syndrome. The Journal of clinical investigation 125: 3657-3666.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, et al. (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321: 674-679.

Snyder M, Du J, Gerstein M (2010) Personal genome sequencing: current approaches and challenges. Genes & development 24: 423-431.

Sobreira NL, Valle D (2016) Lessons learned from the search for genes responsible for rare Mendelian disorders. Molecular genetics & genomic medicine 4: 371-375.

Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, et al. (2014) Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Science translational medicine 6: 265ra168.

Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, et al. (2017) Pathogenic variants that alter protein code often disrupt splicing. Nature Genetics 49: 848-855.

Solomon BD (2014) Incidentalomas in genomics and radiology. The New England journal of medicine 370: 988-990.

Solomon BD, Lee T, Nguyen AD, Wolfsberg TG (2016) A 2.5 year snapshot of Mendelian discovery. Molecular genetics & genomic medicine 4: 392-394.

Spaapen LJ, Ketting D, Wadman SK, Bruinvis L, Duran M (1987) Urinary D-4-hydroxyphenyllactate, D-phenyllactate and D-2-hydroxyisocaproate, abnormalities of bacterial origin. Journal of inherited metabolic disease 10: 383-390.

Spiegel PK (1995) The first clinical X-ray made in America--100 years. American journal of roentgenology 164: 241-243.

Srivastava S, Cohen JS, Vernon H, Baranano K, McClellan R, et al. (2014) Clinical whole exome sequencing in child neurology practice. Annals of neurology 76: 473-483.

Stark Z, Schofield D, Alam K, Wilson W, Mupfeki N, et al. (2017) Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. Genetics in medicine.

Stavropoulos DJ, Merico D, Jobling R, Bowdin S, Monfared N, et al. (2016) Whole Genome Sequencing Expands Diagnostic Utility and Improves Clinical Management in Pediatric Medicine. NPJ Genom Med 1.

Steinberg F, Gallon M, Winfield M, Thomas EC, Bell AJ, et al. (2013) A global analysis of SNX27-retromer assembly and cargo specificity reveals a function in glucose and metal ion transport. Nature cell biology 15: 461-471.

Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, et al. (2009) The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. Human genomics 4: 69-72.

Stromme P (2000) Aetiology in severe and mild mental retardation: a population-based study of Norwegian children. Developmental medicine and child neurology 42: 76-86.

Sun Y, Ruivenkamp CA, Hoffer MJ, Vrijenhoek T, Kriek M, et al. (2015) Next-generation diagnostics: gene panel, exome, or whole genome? Human Mutation 36: 648-655.

Sundaramurthy S, Swaminathan M, Sen P, Arokiasamy T, Deshpande S, et al. (2016) Homozygosity mapping guided next generation sequencing to identify the causative genetic variation in inherited retinal degenerative diseases. Journal of human genetics 61: 951-958.

Surpili MJ, Delben TM, Kobarg J (2003) Identification of proteins that interact with the central coiled-coil region of the human protein kinase NEK1. Biochemistry 42: 15369-15376.

Swerdlow H, Wu S, Harke H, Dovichi NJ (1990) Capillary gel electrophoresis for DNA sequencing. Journal of Chromatography A 516: 61-67.

Symoens S, Syx D, Malfait F, Callewaert B, De Backer J, et al. (2012) Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. Human Mutation 33: 1485-1493.

Taguchi H, Ohta T (1991) D-lactate dehydrogenase is a member of the D-isomer-specific 2-hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in Escherichia coli of the D-lactate dehydrogenase gene of Lactobacillus plantarum. The Journal of biological chemistry 266: 12588-12594.

Takahara K, Hoffman GG, Greenspan DS (1995) Complete structural organization of the human alpha 1 (V) collagen gene (COL5A1): divergence from the conserved organization of other characterized fibrillar collagen genes. Genomics 29: 588-597.

Talasniemi JP, Pennanen S, Savolainen H, Niskanen L, Liesivuori J (2008) Analytical investigation: assay of D-lactate in diabetic plasma and urine. Clinical biochemistry 41: 1099-1103.

Tan SS, Bouwmans CA, Rutten FF, Hakkaart-van Roijen L (2012) Update of the Dutch Manual for Costing in Economic Evaluations. International journal of technology assessment in health care 28: 152-158.

Tanaka K, West-Dull A, Hine DG, Lynn TB, Lowe T (1980) Gas-chromatographic method of analysis for urinary organic acids. II. Description of the procedure, and its application to diagnosis of patients with organic acidurias. Clinical chemistry 26: 1847-1853.

Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, et al. (2016) Exome Sequencing and the Management of Neurometabolic Disorders. The New England journal of medicine 374: 2246-2255.

Tarailo-Graovac M, Wasserman WW, Van Karnebeek CD (2017) Impact of next-generation sequencing on diagnosis and management of neurometabolic disorders: current advances and future perspectives. Expert review of molecular diagnostics 17: 307-309.

Tattini L, D'Aurizio R, Magi A (2015) Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Frontiers in Bioengineering and Biotechnology 3: 92.

Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, et al. (2015) Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nature Genetics 47: 717-726.

Telenti A, Pierce LCT, Biggs WH, di Iulio J, Wong EHM, et al. (2016) Deep sequencing of 10,000 human genomes. Proceedings of the National Academy of Sciences 113: 11901-11906.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64-69.

Tessadori F, van Weerd JH, Burkhard SB, Verkerk AO, de Pater E, et al. (2012) Identification and functional characterization of cardiac pacemaker cells in zebrafish. PloS one 7: e47644.

Thauvin-Robinet C, Cossee M, Cormier-Daire V, Van Maldergem L, Toutain A, et al. (2006) Clinical, molecular, and genotype-phenotype correlation studies from 25 cases of oral-facial-digital syndrome type 1: a French and Belgian collaborative study. Journal of Medical Genetics 43: 54-61.

Thauvin-Robinet C, Lee JS, Lopez E, Herranz-Perez V, Shida T, et al. (2014) The oral-facial-digital syndrome gene C2CD3 encodes a positive regulator of centriole elongation. Nature genetics 46: 905-911.

The Deciphering Developmental Disorders S (2015) Large-scale discovery of novel genetic causes of developmental disorders. Nature 519: 223-228.

The Deciphering Developmental Disorders Study (2015) Large-scale discovery of novel genetic causes of developmental disorders. Nature 519: 223-228.

The Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nature Genetics 46: 818-825.

The Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526: 68-74.

The UK10K Consortium (2015) Genomics: The United Kingdom's 10K genome project. Nature methods 12: 1010-1010.

Thiel C, Kessler K, Giessl A, Dimmler A, Shalev SA, et al. (2011) NEK1 mutations cause short-rib polydactyly syndrome type majewski. American journal of human genetics 88: 106-114.

Thomas S, Legendre M, Saunier S, Bessieres B, Alby C, et al. (2012) TCTN3 mutations cause Mohr-Majewski syndrome. American Journal of Human Genetics 91: 372-378.

Thompson JS, Iyer KR, DiBaise JK, Young RL, Brown CR, et al. (2003) Short bowel syndrome and Crohn's disease. Journal of gastrointestinal surgery 7: 1069-1072.

Thornalley PJ (1993) The glyoxalase system in health and disease. Molecular aspects of medicine 14: 287-371.

Topol EJ, Frazer KA (2007) The resequencing imperative. Nature Genetics 39: 439-440.

Toriello HV, Franco B (1993) Oral-Facial-Digital Syndrome Type I. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, et al. (eds) GeneReviews(R). University of Washington, Seattle, Seattle (WA)

Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Koster J, et al. (2017) Clinical exome sequencing: results from 2819 samples reflecting 1000 families. European journal of human genetics 25: 176-182.

Tsujibo H, Tiano HF, Li SS (1985) Nucleotide sequences of the cDNA and an intronless pseudogene for human lactate dehydrogenase-A isozyme. European journal of biochemistry 147: 9-15.

Tubbs PK (1965) The metabolism of D-alpha-hydroxy acids in animal tissues. Annals of the New York Academy of Sciences 119: 920-926.

Turner G, Boyle J, Partington MW, Kerr B, Raymond FL, et al. (2008) Restoring reproductive confidence in families with X-linked mental retardation by finding the causal mutation. Clinical genetics 73: 188-190.

Turner Tychele N, Hormozdiari F, Duyzend Michael H, McClymont Sarah A, Hook Paul W, et al. (2016) Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. The American Journal of Human Genetics 98: 58-74.

Ung PM, Schlessinger A (2015) DFGmodel: predicting protein kinase structures in inactive states for structure-based discovery of type-II inhibitors. ACS chemical biology 10: 269-278.

United States House of Representatives (2002) Rare diseases act of 2002. In: Representatives Ho (ed). U.S. Government Publishing Office, Washington USA

Upadhya P, Birkenmeier EH, Birkenmeier CS, Barker JE (2000) Mutations in a NIMA-related kinase gene, Nek1, cause pleiotropic effects including a progressive polycystic kidney disease in mice. Proceedings of the National Academy of Sciences of the United States of America 97: 217-221.

Upadhya P, Churchill G, Birkenmeier EH, Barker JE, Frankel WN (1999) Genetic modifiers of polycystic kidney disease in intersubspecific KAT2J mutants. Genomics 58: 129-137.

Uribarri J, Oh MS, Carroll HJ (1998) D-lactic acidosis. A review of clinical presentation, biochemical features, and pathophysiologic mechanisms. Medicine 77: 73-82.

Valencia CA, Husami A, Holle J, Johnson JA, Qian Y, et al. (2015) Clinical Impact and Cost-Effectiveness of Whole Exome Sequencing as a Diagnostic Tool: A Pediatric Center's Experience. Frontiers in Pediatrics 3: 67.

Valente EM, Logan CV, Mougou-Zerelli S, Lee JH, Silhavy JL, et al. (2010) Mutations in TMEM216 perturb ciliogenesis and cause Joubert, Meckel and related syndromes. Nature genetics 42: 619-625.

van de Werken HJG, Landan G, Holwerda SJB, Hoichman M, Klous P, et al. (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nature methods 9: 969-972.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, et al. (2002) From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current protocols in bioinformatics. John Wiley & Sons, Inc.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, et al. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics 43: 11.10.11-33.

van Haelst MM, Monroe GR, Duran K, van Binsbergen E, Breur JM, et al. (2015) Further confirmation of the MED13L haploinsufficiency syndrome. European journal of human genetics 23: 135-138.

van Nimwegen KJ, Schieving JH, Willemsen MA, Veltman JA, van der Burg S, et al. (2015) The diagnostic pathway in complex paediatric neurology: a cost analysis. European journal of paediatric neurology 19: 233-239.

van Nimwegen KJ, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, et al. (2016) Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. Clinical chemistry 62: 1458-1464.

van Veen MR, van Hasselt PM, de Sain-van der Velden MGM, Verhoeven N, Hofstede FC, et al. (2011) Metabolic Profiles in Children During Fasting. Pediatrics 127: e1021-e1027.

Vassy JL, Christensen KD, Schonman EF, et al. (2017) The impact of whole-genome sequencing on the primary care and outcomes of healthy adult patients: A pilot randomized trial. Annals of Internal Medicine 167: 159-169.

Veland IR, Awan A, Pedersen LB, Yoder BK, Christensen ST (2009) Primary cilia and signaling pathways in mammalian development, health and disease. Nephron physiology 111: p39-53.

Veldman MB, Lin S (2008) Zebrafish as a developmental model organism for pediatric research. Pediatric research 64: 470-476.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304-1351.

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, et al. (2010) A de novo paradigm for mental retardation. Nature Genetics 42.

Vissers LE, van Nimwegen KJ, Schieving JH, Kamsteeg EJ, Kleefstra T, et al. (2017) A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. Genetics in medicine.

Vogel F, Motulsky AG (1997) Human Genetics; Problems and Approaches. Springer, Berlin

Vrijenhoek T, Middelburg EM, Monroe GR, van Gassen KLI, Geenen J, et al. (2017) The impact of whole-exome sequencing on medical decision-making - a costing study. . Submitted.

Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, et al. (2017) The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. Genetics in medicine 19: 546-552.

Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, et al. (2010) Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. American journal of human genetics 87: 90-94.

Wang Q, Lu Y, Yuan M, Darling IM, Repasky EA, et al. (2006) Characterization of monocarboxylate transport in human kidney HK-2 cells. Molecular pharmaceutics 3: 675-685.

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

Weiss K, Terhal PA, Cohen L, Bruccoleri M, Irving M, et al. (2016) De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. American journal of human genetics 99: 934-941.

Wenger AM, Guturu H, Bernstein JA, Bejerano G (2017) Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. Genetics in medicine 19: 209-214.

Wenstrup RJ, Florer JB, Davidson JM, Phillips CL, Pfeiffer BJ, et al. (2006) Murine model of the Ehlers-Danlos syndrome. col5a1 haploinsufficiency disrupts collagen fibril assembly at multiple stages. The Journal of biological chemistry 281: 12888-12895.

Westerberg DP (2013) Diabetic ketoacidosis: evaluation and treatment. American family physician 87: 337-346.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452: 872-876.

White J, Mazzeu JF, Hoischen A, Jhangiani SN, Gambin T, et al. (2015) DVL1 frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. American journal of human genetics 96: 612-622.

White MC, Quarmby LM (2008) The NIMA-family kinase, Nek1 affects the stability of centrosomes and ciliogenesis. BMC cell biology 9: 29.

Wilkin DJ, Szabo JK, Cameron R, Henderson S, Bellus GA, et al. (1998) Mutations in fibroblast growth-factor receptor 3 in sporadic cases of achondroplasia occur exclusively on the paternally derived chromosome. American journal of human genetics 63: 711-716.

Wilkins MH, Stokes AR, Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. Nature 171: 738-740.

Willemsen MH, Kleefstra T (2014) Making headway with genetic diagnostics of intellectual disabilities. Clinical genetics 85: 101-110.

Wilson MC, Meredith D, Fox JE, Manoharan C, Davies AJ, et al. (2005) Basigin (CD147) is the target for organomercurial inhibition of monocarboxylate transporter isoforms 1 and 4: the ancillary protein for the insensitive MCT2 is EMBIGIN (gp70). The Journal of biological chemistry 280: 27213-27221.

Wood JD, Landers JA, Bingley M, McDermott CJ, Thomas-McArthur V, et al. (2006) The microtubule-severing protein Spastin is essential for axon outgrowth in the zebrafish embryo. Human molecular genetics 15: 2763-2771.

Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, et al. (2011) Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genetics in medicine 13: 255-262.

Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, et al. (2015) Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. The Lancet 385: 1305-1314.

Wu SH, Arevalo JC, Sarti F, Tessarollo L, Gan WB, et al. (2009) Ankyrin Repeat-rich Membrane Spanning/Kidins220 protein regulates dendritic branching and spine stability in vivo. Developmental neurobiology 69: 547-557.

Xue Y, Chen Y, Ayub Q, Huang N, Ball Edward V, et al. (2012) Deleterious- and Disease-Allele Prevalence in Healthy Individuals: Insights from Current Predictions, Mutation Databases, and Population-Scale Resequencing. American journal of human genetics 91: 1022-1032.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, et al. (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. The New England journal of medicine 369: 1502-1511.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, et al. (2014) Molecular findings among patients referred for clinical whole-exome sequencing. Jama 312: 1870-1879.

Yu B, Li AH, Metcalf GA, Muzny DM, Morrison AC, et al. (2016) Loss-of-function variants influence the human serum metabolome. Science Advances 2: e1600800.

Yubero D, Brandi N, Ormazabal A, Garcia-Cazorla A, Perez-Duenas B, et al. (2016) Targeted Next Generation Sequencing in Patients with Inborn Errors of Metabolism. PloS one 11: e0156359.

Zaghloul NA, Brugmann SA (2011) The emerging face of primary cilia. Genesis 49: 231-246.

Zahir FR, Mwenifumbo JC, Chun H-JE, Lim EL, Van Karnebeek CDM, et al. (2017) Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. BMC genomics 18: 403.

Zhang F, Lupski JR (2015) Non-coding genetic variants in human disease. Human molecular genetics 24: R102-110.

SAMENVATTING

Technologische innovatie binnen de gezondheidszorg zorgt voor een continue verbetering van diagnostiek en patiëntenzorg. In het vakgebied van de klinische genetica is ons vermogen om de genetische oorzaken van ziekten te ontdekken drastisch veranderd door de snelle ontwikkeling en toepassing van technieken om varianten in het DNA te detecteren. We voorzien hiermee niet alleen patiënten van een diagnose, maar we vergroten ook onze inzichten in de moleculaire en cellulaire rol van deze genen door het karakteriseren van deze varianten. Daarmee vergroten we onze kennis van de humane fysiologie en de impact van genetische dysfunctie.

Het werk dat gepresenteerd wordt in dit proefschrift omvat de introductie, de snelle ontwikkeling en het gebruik van next generation sequencing (NGS) – een krachtige methode om de sequentie van DNA te bepalen. We gebruiken NGS om genetische variatie te ontdekken in patiënten met een vermoedelijke Mendeliaanse aandoening, beginnend met het sequencen van gerichte genpanels, wat uit te breiden is naar grotere genomische regio's met whole-exome sequencing (WES) en whole-genome sequencing (WGS) – allemaal binnen slechts enkele jaren na de ontwikkeling en introductie van NGS. Vervolgens maken we gebruik van moleculaire-, cellulaire- en diermodel-experimenten om te bevestigen dat de genetische varianten daadwerkelijk verantwoordelijk zijn voor de aandoening van de patiënt.

In Hoofdstuk 1 introduceer ik zeldzame aandoeningen en bespreek ik hun belasting op de menselijke gezondheid. Veel van deze zeldzame aandoeningen hebben een genetische oorzaak, en we richten ons op Mendeliaanse aandoeningen – aandoeningen waarbij een genetische variant in één gen de ziekte veroorzaakt. De geschiedenis van NGS wordt besproken en ik laat zien hoe de ontwikkeling van NGS ons vermogen om nieuwe genetische oorzaken van Mendeliaanse aandoeningen te ontdekken drastisch heeft verbeterd. Er moeten echter significante uitdagingen worden overwonnen in technische aspecten, ethiek en interpretatie voordat het volledige potentieel van deze technieken gerealiseerd kan worden. Het tijdperk van het genoom is aangebroken.

In Hoofdstuk 2 sequencen we een gericht genpanel met alle bekende en kandidaat genen voor vasculaire (dis)functie in een patiënt met klassiek Ehlers-Danlos syndroom en een slagaderlijke scheuring. We ontdekken een nieuwe heterozygote variant in COL5A1, waarvan voorspeld wordt dat deze de collageen structuur verstoort. Deze bevinding geeft inzicht in het risico voor de familie en kinderen van de nu overleden patiënt, evenals advies voor het monitoren van patiënten met vergelijkbare genetische varianten.

In Hoofdstuk 3 maken we ook gebruik van een gerichte benadering om alle bekende genen betrokken bij obesitas te sequencen in een kind met morbide obesitas. We ontdekken een de novo variant in KIDINS220, een neurologisch koppel-eiwit, en we identificeren twee andere patiënten met vergelijkbare fenotypes en de novo getrunceerde varianten in KIDINS220. Door deze variant tot overexpressie te brengen in zebravissen en het observeren van spasmes, bevestigen we een functionele link tussen deze variant en de spasticiteit in onze patiënten. We identificeren hiermee een nieuw syndroom, dat gekarakteriseerd wordt door spastische paraplegie, verstandelijke beperking, nystagmus en obesitas (SINO).

In Hoofdstuk 4 gaan we door naar WES bij twee patiënten met Mohr syndroom, en identificeren we compound heterozygote varianten in NEK1. Moleculaire experimenten laten zien dat deze NEK1 varianten resulteren in unieke splicing effecten: één variant introduceert een vroegtijdig stopcodon in het mRNA transcript, terwijl de andere verrassend leidt tot het mechanisme van nonsense-associated alternative splicing. Daarnaast lieten fibroblasten van de patiënt een verminderd aantal cellen met cilia zien vergeleken met controles. Daarmee wordt vastgesteld dat Mohr syndroom een ciliopathie is veroorzaakt door NEK1 varianten.

Hierna focussen we in Hoofdstuk 5 op aangeboren afwijkingen in het metabolisme. In patiënten met verwante ouders gebruiken we homozygote genomische regio's om de te onderzoeken gebieden met mogelijk de causale variant te specificeren. We gebruiken een gericht genpanel om alle genen binnen die homozygote regio's te sequencen en identificeren hiermee een nieuwe homozygote getrunceerde variant in MCT1. Door het screenen van andere patiënten met een vergelijkbaar fenotype stellen we vast dat MCT1 eiwit-truncerende varianten de oorzaak zijn van ketoacidose. Verder bepalen we het effect van deze varianten op het niveau van RNA, eiwit, en substraat-transport, waarmee we een nieuwe aangeboren afwijking in het metabolisme, ten gevolge van verminderde keton opname, in kaart brengen.

In Hoofdstuk 6 lossen we een oude onderzoeksvraag op door het Sanger-sequencen van LDHD, dat zich in een homozygote regio bevindt van een patiënt met verhoogd D-lactaat. We vinden een nieuwe, homozygote variant in een geconserveerde regio waarvan voorspeld wordt dat deze effect heeft op de gen-functie. Hierna gebruiken we de zebravis als metabool model om te laten zien dat een verlies van eiwit-functie resulteert in een biochemisch fenotype van verhoogd D-lactaat, vergelijkbaar met het fenotype in onze patiënt. We concluderen dat LDHD verantwoordelijk is voor het D-lactaat metabolisme in het menselijk lichaam.

In Hoofdstuk 7 gaan we door naar de klinische waarde van NGS door het uitvoeren van een onderzoek naar de diagnostische opbrengst van WES, in een cohort met patiënten met een verstandelijke beperking en een langdurige en dure diagnostische zoektocht. Tegelijkertijd evalueren we de kostenefficiëntie van deze technologie door middel van een uitgebreide kostenberekening van de traditionele diagnostiek voor deze patiënten. Uiteindelijk stellen we vast dat WES resulteert in kostenbesparing als het als eerste genetische test wordt uitgevoerd.

Tenslotte bespreek ik in Hoofdstuk 8 zowel het succes van NGS als de aanhoudende uitdagingen van de techniek en resultaat-interpretatie. Ook komt de noodzaak van uitgebreide functionele testen aan bod, die nodig zijn om de pathogeniciteit van de variant te evalueren en het oorzakelijk verband vast te stellen. De implementatie van NGS in de kliniek zal zorgen voor een hogere diagnostische opbrengst voor patiënten met een genetische aandoening, tegen een aanzienlijke kostenbesparing – mits het verstandig wordt gebruikt. Vervolgens bespreek ik hoe NGS, met de mogelijkheid tot het beschikken over je persoonlijke genoom, de samenleving zal transformeren, en hoe de gezondheidszorg gedwongen wordt zich hierop aan te passen om voorop te blijven in deze snel veranderende tijd. Het tijdperk van het genoom is hier, maar het is pas net begonnen.

SUMMARY

Technological innovation in healthcare has continually led to better patient diagnosis and care. In the field of clinical genetics, the development and rapid application of techniques to discover DNA variants in patients with a rare disease has revolutionized our ability to identify the genetic cause of disease. In addition to providing a diagnosis to the patient, the characterization of these variants has led to new insights into the molecular and cellular role of these genes, expanding our knowledge of human physiology and the impact of genetic dysfunction.

The work presented in this thesis encompasses the introduction, rapid improvement and use of next generation sequencing (NGS) – a powerful method to determine the sequence of DNA. We utilize NGS to discover genetic variation in patients with a suspected Mendelian disease, beginning with targeted gene sequencing panels and progressing to larger genomic regions with whole-exome sequencing (WES) and whole-genome sequencing (WGS) – all in only a few short years after NGS development. We then use molecular, cellular and model organism experiments to establish that genetic variants are responsible for the patient's disorder.

In Chapter 1, I introduce rare diseases and discuss their burden on human health. Many of these rare diseases have a genetic cause, and we focus upon Mendelian diseases – disorders for which a genetic variant in one gene causes the disease. I review the history of NGS and show how NGS development dramatically improved our ability to discover the genetic causes of Mendelian disorders. However, significant technical, ethical, and interpretation challenges remain for these technologies full potential to be realized. The era of the genome has arrived.

In Chapter 2 we use a targeted gene sequencing panel containing all known or candidate genes for vascular (dys)function in a classic Ehlers-Danlos syndrome patient with an arterial rupture. We detect a *COL5A1* novel heterozygous variant that is predicted to disrupt the collagen triple helix. This information provides risk guidance for the relatives and offspring of the now deceased patient, as well as monitoring advice in patients with similar genetic variants.

In Chapter 3 we also utilize a targeted approach to sequence all the known genes related to obesity in a child with morbid obesity. We detect a *de novo* variant in *KIDINS220*, a neuronal scaffold protein, and identify two other patients with similar phenotypes and *de novo* truncating *KIDINS220* variants. By overexpressing patient variant *KIDINS220* in zebrafish and observing spasms, we confirm a functional link of this variant to our

patients' spasticity and identify a new syndrome defined by spastic paraplegia, intellectual disability, nystagmus and obesity (SINO).

In Chapter 4 we progress to WES of two individuals with Mohr syndrome, identifying *NEK1* compound heterozygous variants. Molecular experiments establish that these *NEK1* variants result in unique splicing effects: one variant introduces an early stop codon in the mRNA transcript while the other variant results in the surprising mechanism of nonsense-associated alternative splicing. Additionally, patient-derived fibroblasts show a reduction of ciliated cells compared to controls, establishing that Mohr syndrome in these patients is a ciliopathy due to *NEK1* variants.

We then focus on inborn errors of metabolism and use regions of homozygosity in patients with related parents to narrow the genetic areas to investigate. **In Chapter 5** we use a targeted gene sequencing panel to sequence all the genes within the homozygous regions, identifying a novel homozygous *MCT1* truncating variant. By screening other patients with a similar phenotype, we establish that *MCT1* protein-truncating variants are causal for ketoacidosis. We then evaluate the effect of these variants on the RNA, protein, and substrate transport level to establish a new inborn error of metabolism due to decreased ketone uptake.

In Chapter 6 we solve a longstanding research question in a patient with elevated D-lactate levels by Sanger sequencing *LDHD*, a gene within this patient's region of homozygosity. We detect a novel, homozygous variant in a conserved genetic region that is predicted to affect gene function. We then use the zebrafish as a metabolic model to show that loss-of-function of this gene results in a biochemical phenotype of elevated D-lactate that is similar to our patient's phenotype. We conclude that LDHD is responsible for D-lactate metabolism in the human body.

In Chapter 7 we progress to the clinical utility of NGS by performing a study analyzing the WES diagnostic yield in an intellectual disability patient cohort of lengthy and expensive diagnostic odysseys. Concurrently, we evaluate the cost-effectiveness of this technology by performing an in-depth accounting of all costs in the traditional diagnostic pathway for these patients, and ultimately estimate that WES would result in cost savings if performed as a first-tier genetic test.

Finally, **in Chapter 8** I discuss the success of NGS as well as the ongoing technical and interpretation challenges, and of the necessity of widespread functional tests to evaluate variant pathogenicity to establish causality. The implementation of NGS in the clinic will result in a higher diagnostic yield for patients with a genetic disorder, at significant

cost savings – if utilized wisely. I then move to how NGS will transform society with the possibility of personal genomes, and how the healthcare system will be forced to adapt to stay ahead in this fast-moving time.

The era of the genome is here, but it is only just beginning.

LIST OF PUBLICATIONS

Lipstein N, Verhoeven-Duif NM, Michelassi FE, Calloway N, van Hasselt PM, Pienkowska K, van Haaften G, van Haelst MM, van Empelen R, Cuppen I, van Teeseling HC, Evelein AM, Vorstman JA, Thoms S, Jahn O, Duran KJ, **Monroe GR**, Ryan TA, Taschenberger H, Dittman JS, Rhee JS, Visser G, Jans JJ, Brose N. (2017) Synaptic UNC13A protein variant causes increased neurotransmission and dyskinetic movement disorder. Journal of Clinical Investigation 127: 1005-1018. doi: 10.1172/jci90259

Monroe GR*, Kappen IF*, Stokman MF*, Terhal PA, van den Boogaard MH, Savelberg SM, van der Veken LT, van Es RJ, Lens SM, Hengeveld RC, Creton MA, Janssen NG, Mink van der Molen AB, Ebbeling MB, Giles RH, Knoers NV, van Haaften G. (2016) Compound heterozygous NEK1 variants in two siblings with oral-facial-digital syndrome type II (Mohr syndrome). European Journal of Human Genetics 24: 1752-1760. doi: 10.1038/ejhg.2016.103

Leegwater PA, Vos-Loohuis M, Ducro BJ, Boegheim IJ, van Steenbeek FG, Nijman IJ, **Monroe GR**, Bastiaansen JW, Dibbits BW, van de Goor LH, Hellinga I, Back W, Schurink A. (2016) Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. BMC Genomics 17: 839. doi: 10.1186/s12864-016-3186-0

Weiss K, Terhal PA, Cohen L, Bruccoleri M, Irving M, Martinez AF, Rosenfeld JA, Machol K, Yang Y, Liu P, Walkiewicz M, Beuten J, Gomez-Ospina N, Haude K, Fong CT, Enns GM, Bernstein JA, Fan J, Gotway G, Ghorbani M, van Gassen K, **Monroe GR**, van Haaften G, Basel-Vanagaite L, Yang XJ, Campeau PM, Muenke M. (2016) De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. American Journal of Human Genetics 99: 934-941. doi: 10.1016/j.ajhg.2016.08.001

Monroe GR*, Frederix GW*, Savelberg SM, de Vries TI, Duran KJ, van der Smagt JJ, Terhal PA, van Hasselt PM, Kroes HY, Verhoeven-Duif NM, Nijman IJ, Carbo EC, van Gassen KL, Knoers NV, Hovels AM, van Haelst MM, Visser G, van Haaften G. (2016) Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. Genetics In Medicine 18: 949-56. doi: 10.1038/gim.2015.200

de Vries TI, **Monroe GR**, van Belzen MJ, van der Lans CA, Savelberg SM, Newman WG, van Haaften G, Nievelstein RA, van Haelst MM. (2016) Mosaic CREBBP mutation causes overlapping clinical features of Rubinstein-Taybi and Filippi syndromes. European Journal of Human Genetics 24: 1363-6. doi: 10.1038/ejhg.2016.14

Josifova DJ*, **Monroe GR***, Tessadori F*, de Graaff E, van der Zwaag B, Mehta SG, Harakalova M, Duran KJ, Savelberg SM, Nijman IJ, Jungbluth H, Hoogenraad CC, Bakkers J, Knoers NV, Firth HV, Beales PL, van Haaften G, van Haelst MM. (2016) Heterozygous KIDINS220/ARMS nonsense variants cause spastic paraplegia, intellectual disability, nystagmus, and obesity. Human Molecular Genetics 25: 2158-2167. doi: 10.1093/hmg/ddw082

Kroes HY, **Monroe GR**, van der Zwaag B, Duran KJ, de Kovel CG, van Roosmalen MJ, Harakalova M, Nijman IJ, Kloosterman WP, Giles RH, Knoers NV, van Haaften G. (2016) Joubert syndrome: genotyping a Northern European patient cohort. European Journal of Human Genetics 24: 214-20. doi: 10.1038/ejhg.2015.84

Nicolaou N, Pulit SL, Nijman IJ, **Monroe GR**, Feitz WF, Schreuder MF, van Eerde AM, de Jong TP, Giltay JC, van der Zwaag B, Havenith MR, Zwakenberg S, van der Zanden LF, Poelmans G, Cornelissen EA, Lilien MR, Franke B, Roeleveld N, van Rooij IA, Cuppen E, Bongers EM, Giles RH, Knoers NV, Renkema KY. (2016) Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. Kidney International 89: 476-86. doi: 10.1038/ki.2015.319

Slaats GG, Isabella CR, Kroes HY, Dempsey JC, Gremmels H, **Monroe GR**, Phelps IG, Duran KJ, Adkins J, Kumar SA, Knutzen DM, Knoers NV, Mendelsohn NJ, Neubauer D, Mastroyianni SD, Vogt J, Worgan L, Karp N, Bowdin S, Glass IA, Parisi MA, Otto EA, Johnson CA, Hildebrandt F, van Haaften G, Giles RH, Doherty D. (2016) MKS1 regulates ciliary INPP5E levels in Joubert syndrome. Journal of Medical Genetics 53: 62-72. doi: 10.1136/jmedgenet-2015-103250

Ducro BJ, Schurink A, Bastiaansen JW, Boegheim IJ, van Steenbeek FG, Vos-Loohuis M, Nijman IJ, **Monroe GR**, Hellinga I, Dibbits BW, Back W, Leegwater PA. (2015) A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses. BMC Genomics 16: 761. doi: 10.1186/s12864-015-1936-z

Monroe GR*, Harakalova M*, van der Crabben SN, Majoor-Krakauer D, Bertoli-Avella AM, Moll FL, Oranen BI, Dooijes D, Vink A, Knoers NV, Maugeri A, Pals G, Nijman IJ, van Haaften G, Baas AF. (2015) Familial Ehlers-Danlos syndrome with lethal arterial events caused by a mutation in COL5A1. American Journal of Medical Genetics Part A 167: 1196-203. doi: 10.1002/ajmg.a.36997

van Haelst MM, **Monroe GR**, Duran K, van Binsbergen E, Breur JM, Giltay JC, van Haaften G. (2015) Further confirmation of the MED13L haploinsufficiency syndrome. European

Journal of Human Genetics 23: 135-8. doi: 10.1038/ejhg.2014.69

van Hasselt PM*, Ferdinandusse S*, **Monroe GR**, Ruiter JP, Turkenburg M, Geerlings MJ, Duran K, Harakalova M, van der Zwaag B, Monavari AA, Okur I, Sharrard MJ, Cleary M, O'Connell N, Walker V, Rubio-Gozalbo ME, de Vries MC, Visser G, Houwen RH, van der Smagt JJ, Verhoeven-Duif NM, Wanders RJ, van Haaften G. (2014) Monocarboxylate transporter 1 deficiency and ketone utilization. New England Journal of Medicine 371: 1900-7. doi: 10.1056/NEJMoa1407778

Wiegerinck CL, Janecke AR, Schneeberger K, Vogel GF, van Haaften-Visser DY, Escher JC, Adam R, Thoni CE, Pfaller K, Jordan AJ, Weis CA, Nijman IJ, **Monroe GR**, van Hasselt PM, Cutz E, Klumperman J, Clevers H, Nieuwenhuis EE, Houwen RH, van Haaften G, Hess MW, Huber LA, Stapelbroek JM, Muller T, Middendorp S. (2014) Loss of syntaxin 3 causes variant microvillus inclusion disease. Gastroenterology 147: 65-68.e10. doi: 10.1053/j.gastro.2014.04.002

MANUSCRIPTS IN PREPARATION

Gueneau L*, Fish R*, Shamseddin HE*, Voisin N*, Them FTM*, Preiksaitiene E*, **Monroe GR***, Allias F, Ambosaidi Q, Ambrozaityte L, Cimbalistienė L, Delafontaine J, Guex N, Hashem MO, Kurdi W, Pippucci T, Pradervand S, Roechert B, van Hasselt PM, Wiederkehr M, Wright CF, DDD Study, Xenarios I, van Haaften GW, Shaw-Smith C, Schindewolf EM, Neerman-Arbez M, Chelly J, KucinskasV, Alkuraya FS, Reymond A. KIAA1109 variants are associated with a severe disorder of brain development and arthrogryposis. (in preparation)

ACKNOWLEDGEMENTS

And…a HUGE thank-you to everyone!

At the end of these four years I've found out more than ever that science is not about what you do on your own, it's about what you do together. If you are reading this, chances are very high that you have directly contributed to something within this work, or indirectly through work discussions, coffee discussions, or (very possibly) during a beer-or-two discussion. This thesis is a celebration of what we have done together. Thank you!

In the wise words of **Monique**, I'm "like furniture around here", so please be patient with my acknowledgements. There are many of you who have helped me, and even more good memories to mention.

First, thank you to the members of my reading committee, Prof. **Kors van der Ent**, Prof. **Irene van Langen**, Prof. **Hans Kristian Ploos van Amstel**, Prof. **Wouter de Laat** and Dr. **Eric Kalkhoven** for taking the time to critically read my work. Along the same lines, thank you to the members of my supervisory committee, **Jeroen Bakkers** and **Nine Knoers**, for your guidance these four years. **Nine**, thank you for always fostering such a good work environment; your extensive scientific input, personal guidance and leadership contributed in a large way to this thesis.

A big thanks to all those who read my thesis and took time to contribute helpful suggestions, critical comments, or to simply rip it apart so that I could build it up again. **Nayia**, **Kirsten**, **Joep**, **Ellen**, **Ies**, **Geert**, **Edwin** – thank you for all of your comments and suggestions, from the beginning phase to the very last exciting 24 hours. I value all of your opinions and knowledge highly, and your thoughts definitely made this a better work. The last minute critique by **Joep** also kept me on my toes and gave me a healthy dose of end-of-thesis thrill. I am also greatly indebted to my cover design team of **Iris**, **Rozemarijn**, **Sanne**, and **Monique** for their extensive brainstorming and for listening to my often-vague creative ideas, and for the 3-heads-are-better-than-one translation for my samenvatting by **Sanne**, **Iris** and **Rozemarijn**. Thank you!

For my time at the UMCU, I have to begin with thanking **Annette Baas** – thank you for first bringing me here, and leaving me in the capable hands of the van Haaften group. From welcoming me into the group at work as well as your home for dinner with Magdalena, you ushered in a new time in my life. Thank you for the opportunity, and it's been a pleasure working with you! Thank you to **Magdalena** for the first years and showing me everything there was to know about NGS and the fun life of array enrichments – and for pushing me

to challenge myself and consider doing a PhD. Your excitement and energy propelled me during those first few years. It was an honor to be one of your paranymphs, and it's been fantastic to see you develop from a PhD student into such a competent, independent researcher. Including me as a joint first author on the COL5A1 publication really motivated me to see what I was capable of.

Thank you to the members of the **Cuppen** group, both past and present, for taking me in like an adopted foreigner and welcoming me into the Hubrecht world at that time, which always seemed like a crazy contrast to the UMCU. Thank you, **Edwin**, **Ewart** and **Ewart**, **Nico**, **Pim**, **Marieke**, **Wensi**, **Lisanne**, **Mark**, **Robin**, **Ruben**, **Nicolle**, **Joep**, **Sander**, **Myrthe**, **Sebas**, **Francis**, **Sjors**, **Anna**, **Simone**, **Roel**, **Kim** and many others. I've enjoyed learning from all of you, listening to **Ewart de Bruin** sing karaoke (badly), hiking with **Nico**, concerts with **Ewart Kuijk**, **Pim**'s barrel-laugh, and borrels celebrating every publication that could be celebrated. Science should be fun, and there has been a lot to celebrate! **Ewart**: thanks for helping me through struggles with the robot, with exome enrichments, with how to open a beer on a crate – much appreciated! **Nicolle**, I've always enjoyed your energy, laugh and enthusiasm, and being in the lab with you early before anyone else yet arrived was always a good way to happily start the day. **Joep** – you are always up for anything, from dancing in a cupcake suit to brainstorming on how genetic knowledge will transform society – thanks for being one of the most sincere believers in the use and applications of genetic knowledge that I know. **Sander**: At this point we've hiked together in 8 countries thus far (!) – with many more to go, and had many dinners, festivities and innumerable longer-than-expected evenings. You know me quite well and surprisingly still continue to enjoy my company. Thank you for your friendship, thank you for your ridiculousness, and I'm looking forward to the return of the mustaches. Maybe no one else is, though.

Edwin – thank you for bringing me to the group so long ago, and for not taking me too seriously when I interviewed in a suit. I appreciate your trust in people and your critical eyes; more than once I've been amazed at your ability to ask a question I couldn't think of when you were busy on your tablet. You are an inspiration with your vision, and I really appreciate your guidance in the last few months of my PhD.

Monique, thank you for being the rock of the department in these years. I've been accustomed at this point to just point in your direction when anyone needs to know anything around the office, and you are always my first go-to for questions on anything (work, vegetarianism, life). You are also like furniture here and this place would fall apart without you. Thank you.

Thank you to all the many, many collaborators and coauthors on all of our joint publications. This thesis is as much a result of your work as of mine. **Mieke** and **Federico** – thank you for always pushing the KIDINS220 manuscript forward. **Mieke** you've helped me in so many ways, through guidance on so many of our copublications (CREBBP/MED13L/KIDINS220/STC) and I've always been caught up (and motivated) by your whirlwind energy. **Paulien** and **Marie-José**, thank you for your knowledge on the NEK1 manuscript and to **Isabelle** for all the writing contributions. **Marijn** and **Rachel** – so many cilia experiments, so little time. Marijn, thank you for the extensive experiments you performed for this publication, and the basal exon skipping is still one of the most surprising things I've heard of. You made that publication with your hard work and experiments! **Peter van Hasselt** and **Sacha Ferdinandusse**, it was great working with you on the MCT1 manuscript and our ongoing collaborative efforts on ketoacidosis research, in addition to **Ronald Wanders** and his departments' helpful insight. Thanks to **Hester Kroes** for introducing me to the genetics behind Joubert syndrome and her encouragement on finishing my thesis. From the veterinary department, thanks to **Frank "the tank" van Steenbeek** and **Hille Fieten** for convincing me that the dog is a useful model for human diseases – and should get more attention! Thank you to **Albertien**, **Federico**, **Sanne**, **Nanda**, **Judith**, **Paulien** and **Saskia** on the LDHD publication. For the STC manuscript – it was truly a joint collaborative effort if there ever was one. Thank you to **Mieke** and **Gepke** for their oversight, input and criticism, **Koen** for the always helpful variant filtering help, all the involved clinicians (**Jasper**, **Paulien**, **Peter**, **Hester**) for their attention to this work, and **Anke** and **Geert** for the cost analysis. **Geert** – this was my first exposure to the wide world of cost analyses, and I have to say that I really liked it. Thank you for your input, thank you for the coffee chats, and thank you for your friendship after this publication. I have to say that I never would have thought that talking about money could have been so interesting, but I'm thoroughly convinced now. It's been great seeing you grow into a leadership role, both at work and at home as a new father. You're great at both! To those outside of the research section, in particular **Patrick**, **Marc van Tuil**, **Esther**, **Jelena**, **Koen**, and **Marco** – I've appreciated working with you and learning from you.

I would like to give a big thank you to the metabolics group, who has been very patient with me in my beginning steps of getting into the complex and unfamiliar world of metabolics. Thank you to **Judith** and **Nanda** for explaining so much to me and correcting me when my logic went off in a different direction, and to **Lynne**, **Monique**, **Ruben**, and **Hanneke** and the rest of the group for your help and insight. Lynne, Ruben and Hanneke, I enjoyed talking science and life with you at the Masters Class, and from you I have gained new appreciation of gain-of-function variants, flux analysis, and untargeted metabolomics. Thank you also to **Johan** for the ever-continuing lactate measurements.

To the other groups in the research section of Genetics:

To the **Koeleman** group: **Bobby**, thank you for always being critical of my work and for showing met that it isn't always absolutely necessary to grow up; to **Ruben** and **Eric** for showing me the ropes around the lab and solving any problem that I may come up with. **Sakshi** your energy has always been welcome around the office, and **Wout** thank you for some work discussions but more than anything for introducing me to climbing. **Remy** and **Annabel**, I'm happy to know you both better after the Veluwe retreat and talking with **Remy** over the regular morning coffee. **Iris** – it was fantastic this year to get to know you better at the ESHG 2017 with Federico, getting our mic-pcr propaganda, learning from Fede's exercise routines, and exploring Copenhagen. You are always smiling, once-in-a-while sniffling, but regardless consistently cranking out a new paper every 3 weeks or so. Your motivation and focus on what you want was inspiring to me in the last few months as I considered my next career step. Thank you for your advice, your fun attitude, and the positive energy that you give to our office.

To the **Renkema** group: **Kirsten**, thanks for your leadership along the way, your input on functional networks and for looking my chapters over with a critical eye; I know with you that your door is always open to talk to, and that you bring the entire group closer by your attention to people. And I have to say that even though I am a kidney-outsider, my ears always now prick up whenever I hear kidney genetics come. **Edith**: it has been great to see you develop into a professional at NGS and cell culture, and I'm continually inspired by your trips and activity outside of work. **Albertien**: grant winner extraordinaire – thanks for the clinical input on our Monday meetings, as well as your help on the LDHD manuscript. **Rozemarijn**, – baking is just one of the many things that you excel at. From being our sailing captain to leading us around a forest or an escape room, you always have a feeling for the best way forward – including your advice to me to follow what I want to do in the future. Thank you for your always enthusiastic approach to the day, and for being one of the people whom obstacles simply cannot block the way for. Thanks for your encouragement on writing or pushing forward when writers block hit mid-thesis introduction. Thank you also to **Anukrati**, the recent addition: thanks for sneaking up behind me with your inquisitive questions on variant filtering and why Dutch or American people do the wacky things that they do. Finally, **Marijn**, thank you for so much – for your knowledge and love of cilia, for keeping me motivated to try and beat you to the office in the morning (though you always left after me), for consoling me after a bad presentation in Cardiff, for just being great. You are an incredible person and have shown that you are good at everything you do – research, clinical genetics, and motherhood. I'm looking forward to your defense!!!

A thank you to the rapidly expanding **de Ridder** group, quickly taking over the Stratenum: **Jeroen**, **Joep**, **Joske**, **Amin**, **Carl** and the rest of the group. Thank you for giving me a different viewpoint on my work, and giving new direction to this department. **Amin** it was great dominating Wipeout with you during the retreat this year (or at least falling on our faces together), and **Joske** I've told you already, but I would definitely vote for you for Prime Minister. Time to start running after you finish this PhD!

To the former **Paul de Bakker** and current **Folker Asselbergs** groups – thank you for opening my eyes to the wide world of complex genetics and cardiology. The super long lunches (I'm looking in your direction, **Vinicius**) with **Daniel** and **Laurent** always had some eye-opening discussion, even if it wasn't always about science. **Jessica** and **Sander**, it was great to spend time with you both on the conference to beat all others at the ESHG 2016! **Sara**, you set the bar high for other Americans to compare to, and I have yet to see a genetic problem that your enthusiasm and knowledge cannot overcome. To my favorite Brazilians: **Daiane** - where to begin?! From many concerts together to our epic trip with Sander and Sanne to hike in Norway, your desire to do new things and enjoy your time over here has certainly taken all of us to see and experience new things. From brigadeiros to listening to jazz on Kings Day to closing down Café Belgie, I have really enjoyed all of our fun times together. Thanks! **Vinicius** – from Brazil to the Netherlands to Iceland, married and now with a little girl. Life changes, but you seem to adapting quite well, and I want to thank you for your friendship, being one of your paranymphs in the most unstressed thesis defense I have seen, and most of all for your ability to put things into context: work hard, but appreciate your life as well. Selfishly I hope you return to the Netherlands soon, but I will also be ok if I have to go to Iceland to visit you, Gudrun and Saga.

A big thank you to **Ellen Carbo**, even though you have not contributed in any way to this thesis – WHAT?! Just kidding. Thank you for taking so much time to look at my thesis, for providing me with coffee and ice cream, the many concerts we've went to together (and sorry about the bad ones), and the fun (and noise) that you brought around the office. You survived the Tour du Mont Blanc with **Sander**, **Stef** and I, as well as your Scandinavian survival challenges; maybe 2018 is the year to see if the you can survive the Stratenum bridge? I look forward to many new hiking adventures and concerts with you, and thank you for helping me through any research or job dilemmas.

Terry, not surprisingly you get your own section – because, well, you are Terry! You are my favorite outside-the-box thinker, continually pushing what we should do, and how we should consider DNA sequencing outside of the context of only disease. You are inspiring, and I thank you for all the conversations, the input into my own work, the laughs, for including me in the Genetics Clinic of the Future. Keep pushing the envelope, and no

doubt we'll both look back a few years from now and all our far-flung ideas will be reality. I truly believe this.

Thank you to Ies, **Robert**, **Stef**, **Martin** and **Mark** for holding my hand and teaching me patiently about everything I could ever want to know about programming and bioinformatics. As far as I'm concerned you are all magicians, and thank you for solving any problem I have been working on for half a day in a few clicks of the mouse. You guys are great, and just like bioinformatics you all should get more credit for being behind the scenes awesome. I've also learned all I've ever wanted to you from our lunches about happy hardcore music, cars, or Vitesse. Thank you. Along the same line, thank you **Flip** for programming and computer help – at times, it is surprising that I can turn on my own computer without you. Flip, thanks for also inspiring me on the running, and even for getting **Mircea** out for a run on the beach with **Alessio** at the retreat in Bergen!

Rachel – thank you for your down to earth wisdom and your inspiration. As far as I'm concerned, you are the real Wonder Woman, and I always picture you behind a microscope, while writing a grant, at 1 am. **Glenn**, **Tim**, **Gisela**: my zebrafish and cilia friends. **Tim** – you and **Marijn**, thanks for introducing me to bouldering, and for being a fellow expat that we could look at Dutch culture together with the same eyes and complain about the music. **Glenn**, aka **G**: it's been a great time with you, and the Stratenum misses you around as well as your bad Christmas sweaters and Disco Lab Fridays. You were one of the first to make me appreciate the possibilities of the zebrafish, and it's nice to see you settled where you should be.

A big thanks to **Wigard's Wizards**. **Wigard**, thank you for first interviewing me back in the day, and who would have thought I'd ultimately end up in your group as a post-doc?! Thank for also ignoring any statements that may have been made in the past such as "I'm not interested in doing a PhD" or "I'm not interested in cancer research". Obviously, things change, for the better, and I'm looking forward to a post-doc in your group. Thank you for input and your creativity, your enthusiasm on the possibilities of technology, and for suggesting late night swims in Terschelling. To the rest of the Kloosterman group – you're stuck with me now, watch out! Thank you to the past members **Mirjam**, **Marlous**, and **Mark Pieterse**; the veterans **Mark van Roosmalen**, **Ivo**; the adoptee **Mircea**, and the (relatively) newbies **Alessio**, **Christina**, **Chris**, **Ellen**, **Jolena**, and **Jose**. I really appreciate that you've already taken me in as an honorary wizard on our trip to Regensburg, and looking forward to fun times and productive science as a real wizard. This would not have happened without **Ronald Zweemer**, whom I thank for the opportunity to discover new possibilities for cancer treatment.

Finally to all those who are currently in the room: sometimes super quiet and productive, sometimes loud and productive, sometimes just loud and fun. **Rozemarijn**, **Edith**, **Iris**, **Sanne**, **Karen**, **Helen** – I can never leave this room! Even though sometimes it gets uncomfortably hot from the heating (**Rozemarijn**). Thank you for making it such a pleasure to come to work everyday. **Marijn**, **Mirjam**, **Nayia**, **Chris** don't worry, your memories live on over here in the cumulated paraphernalia on the door, and **Christina** it was always fun watching you be scared by someone at least 5X a day. **Mirjam** thank you for your illustrator skills and for taking photos during my defense, and for your conversations and long dry cappuccino breaks at MiCafe.

A very big thanks to my supervisor **Gijs van Haaften** for your guidance all along the way. Your enthusiasm for science of any kind is inspiring, and thank you for teaching me so much about NGS, about writing, about functional experiments, about how to collaborate and how to be a leader. While I was here you rapidly showed to me how science should be done – meticulously, with many people, and with pleasure as we uncovered new candidate variants or potential functional proof. I am happy to have had you for a supervisor; thank you for guiding me these years.

To the rest of the **van Haaften** group: **Karen**, thank you for knowing everything that is to be known in the lab; as far as I'm concerned, if you can't do it, then it can't be done. I've enjoyed getting to know you, and appreciate your advice – you may not say it out loud all the time, but I get the impression that you see all of us a bit clearer than we see ourselves. I value your advice to myself on others on work and on life. Thank you for everything that you have contributed to in this work, both technically and with your solid knowledge or viewpoints. **Maarten** – the man with the cohort analysis plan. It was great having you in our group, seeing how quickly you adapted to the Mendelian disease field, and how you bring such a different dynamic to the group. Thanks for our brainstorms on variant filtering. We miss you over here, and anytime you want to speak in your American accent or talk about Italian sausages you are welcome to come on over again. **Federico**, the zebrafish guru: thanks for keeping Maarten in-line, and for knowing everything and more about the functional experiments; also for keeping up the good humor as we dealt with experimental, submission or revision stress (Kidins…). **Helen**, in the short time you've been here you've already submitted a paper and made a Cantu model zebrafish line; I think you're already more productive than me and its been nice to get to know you. I want to also thank all the students over the years – **Maartje**, **Annique**, **Ajmal**, **Olga**, **Moska**, **Marlies**, **Camilla**, **Laura**, **Anne** and **Maarten** – for doing such hard work and also challenging me to understand everything completely to satisfy their questions. Also I would like to thank our two famous Slovakian guests **Irena** and **David** for their fun time here and for teaching me about NIPD as it was being developed.

To my two paranymphs: Sanne and Nayia. Wow, its really here! With you women at my side during the defense I don't think that I even have to answer any questions – you know it all.

Sanne – when you first started, you took over my role as technician in the van Haaften group. Pretty much within a month you were an NGS master and I didn't have to do anything anymore. You are extraordinary in the laboratory, not only in performing experiments but planning them, researching them and thinking of the big picture. Nearly every chapter in my thesis is stamped with your hard work. Thank you for everything you've done, not only professionally but also in taking me to Oeteldonk carnival, to your crazy birthday parties, for dressing up for Kromkommer, for challenging yourself on crazy hikes in Norwegian mountains. We've laughed together, we've cried together (a bit, on occasion), and it has been a pleasure sharing this experience with you. Thank you, you bring sunshine to the group!

Nayia – where to begin? We both started at the same time here and shared the early days joys and frustrations of NGS; then, 4 years later, you were off to Cyprus as a doctor to return to your life there. So much of my time here was shared with you, from the Hubrecht borrels to concerts in town, to barbecues with your crazy Cypriote friends, to the honor of being your paranymph, to showing me around Cyprus – not once, but twice (so far!), to spending a week hiking together in the deep midgey Scottish highlands. Ironically you also entered into the Mendelian disease field; I think you should just leave Cyprus and come up back to the Netherlands again – actually, being a paranymph was just an excuse to get you here to stay! I miss not having you up here and thank you for knowing and caring so much about my work, about my life, and know that I value our friendship very highly. Thank you for everything. Now move back up here!

Thank you to both the **Monroe** and **Engelen** families. It means a lot to me that both families are here and that you both know each other so well. Thank you to the **Engelens** for always being interested in what I do and for being such an incredible adopted family over here. A thank you to my **Mom** and **Dad** for always being so supportive of whatever I do and always wanting what is best for me – regardless of where it is, as long as it makes me happy. You both are the rocks in my life and thank you for always being there behind me. To my sister **Becky** – you came over!!! This means so much to me to show you my life over here. Thank you even more that you are so good with keeping **Jamie**, **Van** and **Logan** updated with our life over here. Thanks to my cousin **Troy** and **Nate** for doing the interior design of the thesis; even though **Troy** is the busiest man alive but wouldn't have it any other way, and it was no small task. Thank you, T! And thank you **Nate** for your attention to detail and italicizing every small gene name. Thank you to my grandmother for her

enduring support and love and for not too often demanding that I return back to the USA.

And finally, an enormous thanks to **Chantal**. You've always supported me in everything I've done, always encouraged me to do just what I enjoy and not overthink it too much, and comforted me when things got stressful. Most of the time you know me better than myself, which I reluctantly admit after a few days. You've accepted long ago that you're married to a science nerd, and for this I am extremely grateful. You inspire me in your quiet self-confidence, your drive, and how you seem to effortlessly juggle a million things at once. Thank you for always being there, and also for letting me know that if things don't work out in my academic career there is always a carrot suit waiting with my name on it.

CURRICULUM VITAE

Glen Ryan Monroe was born on the 5th of May, 1979, in New Haven, CT, USA. In 2001 he completed his undergraduate studies in Molecular and Cellular Biology at the University of Connecticut in the USA. He then entered the commercial workforce at Genaissance Pharmaceuticals, New Haven, CT, a company performing large-scale population Sanger sequencing to characterize genetic diversity for pharmaceutical response studies. After several years working there, he left to explore other work and travel opportunities and ultimately ended at the South Pole to perform construction on a new telescope.

He returned to the USA and the genetics field in 2007 to work in the laboratory at Cogenics, New Haven, as a DNA sample preparation analyst. He would remain at this position until he transitioned into a supervising role for quality control of production and research Sanger sequencing. Working remotely, he was able to move to Europe and pursue a master's degree in Biotechnology at Wageningen University, where he would complete his master's degree in 2011. While there he performed an internship in the laboratory of Edwin Tijhaar, working on the expression and purification of antibodies to recombinant cytokines in the cell biology and immunology group of the animal sciences facility.

Upon completion of his master thesis in 2011, he began work in the laboratory of Gijs van Haaften at the University Medical Center Utrecht, NL, as a technician to implement and utilize next generation sequencing. There, in close collaboration with Edwin Cuppen and the Hubrecht Institute, he would specialize in targeted gene panel sequencing and optimize whole-exome sequencing strategies. Interest in the power of next-generation sequencing for use in research and diagnostics prompted him to begin his thesis studies in the group of Gijs van Haaften in 2013, focusing upon variant detection and interpretation in patients with rare Mendelian disorders and ultimately in patients with inborn errors of metabolism. He is intrigued by how changes in our DNA affect our health and is interested in using the genome to not only diagnose disease but to empower us as individuals with genetic knowledge to live a healthy life. Completing his doctoral studies in 2017, he currently resides in Utrecht and hopes in the next step of his career to continue to use the information within the genome to advance human health.

MY GENOME

If you are interested in exploring inside of my genome, I invite you to have a look at my downloadable BAM files and VCF on the Personal Genome Project website:

https://my.pgp-hms.org/profile/hu4FE12A.

ABBREVIATIONS

AAA: Abdominal Aortic Aneurysm
aCGH: array Comparative Genomic Hybridization
ARMS: Ankyrin Repeat-rich Membrane Spanning
ATE: Alternative Terminal Exon
BAM: Binary Alignment Map
BDNF: Brain-Derived Neurotrophic Factor
BE: Base Excess
bp: base pair
BSA: Bovine Serum Albumin
CADD: Combined Annotation Dependent Depletion
CAKUT: Congenital Anomalies of the Kidney and Urinary Tract
CCD1: Coiled Coil Domain 1
cDNA: complementary DNA
CNV: Copy Number Variant
CoA: Coenzyme A
CTA: Computer Tomography Angiography
DATAN: Diacetyl-L-Tartaric Anhydride
DDD: Deciphering Developmental Disorders
DMEM: Dulbecco's Modified Eagle's Medium
DNA: Deoxyribonucleic Acid
dpf: days post fertilization
EDS: Ehlers-Danlos Syndrome
EDTA: Ethylenediaminetetraacetic Acid
ENU: 1-Ethyl-1-Nitrosourea
ERG: Electroretinogram
ESP: Exome Sequencing Project
EVS: Exome Variant Server
ExAC: Exome Aggregation Consortium database
FBS: Fetal Bovine Serum
FCS: Fetal Calf Serum
GC-MS: Gas Chromatography-Mass Spectrometry
GERP: Genomic Evolutionary Rate Profiling
gmomAD: Genome Aggregation Database
GO: Gene Ontology
GoNL: the Genome of the Netherlands
GWAS: Genome Wide Association Study
g.VCF: genomic Variant Call File
HAMF10: Ham's F10 nutrient medium
HEK293: Human Embryonic Kidney 293 cell line
HGMD: Human Gene Mutation Database
HGP: Human Genome Project
ID: Intellectual Disability
IEM: Inborn Error of Metabolism
IGV: Integrative Genomics Viewer
INDEL: Insertion/Deletion
KEGG: Kyoto Encyclopedia of Genes and Genomes
KIDINS220: Kinase D Interacting Substrate of 220 kDa
KIF5A: Kinesin Family Member 5A
KIM: Kinesin Light Chain Interactive Motif

KLC: Kinesin Light Chain
LOVD: Leiden Open Variation Database
LVNC: Left Ventricular Non-compaction
MAF: Minor Allele Frequency
MLPA: Multiplex Ligation-dependent Probe Amplification
MOMO: Macrosomia, Obesity, Macrocephaly, and Ocular abnormalities
MRA: Magnetic Resonance Angiography
MRM: Multiple-Reaction-Monitoring
NES: Nuclear Exportation Signal
NGS: Next Generation Sequencing
NGSe: Normal Goat Serum
NIHR: National Institute for Health Research
NLS: Nuclear Localization Signal
NMD: Nonsense-Mediated Decay
OFC: Occipito Frontal Circumference
OFD: Oral-Facial-Digital
OMIM: Online Mendelian Inheritance in Man
PAGE: Polyacrylamide Gel Electrophoresis
PBS: Phosphate Buffered Saline
PC12: Rat Pheochromocytoma cell line
PCR: Polymerase Chain Reaction
PFA: Paraformaldehyde
PVDF: Polyvinylidene Difluoride
RFP: Red Fluorescent Protein
RNA: Ribonucleic Acid
RNA-seq: RNA-sequencing
ROI: Region Of Interest
RTE: Renal Tubular Epithelial
SAM: Sterile Alpha Motif
SDS: Sodium Dodecyl Sulfate
SIFT: Sorting Tolerant From Intolerant
SINO: Spastic paraplegia, Intellectual disability, Nystagmus and Obesity syndrome
SNP: Single Nucleotide Polymorphism
SNV: Single Nucleotide Variant
SNX27: Sorting Nexin Family Member 27
SRPS: Short-rib Polydactyly Syndrome
STC: Sylvia Tóth Center
Trk: Tropomyosin Receptor Kinase
UAS: Upstream Activation Sequence
UMCU: University Medical Center Utrecht
UPLC: Ultra Performance Liquid Chromatography
VCF: Variant Call File
VEP: Visual Evoked Potential
VUMC: Free University Medical Center
VUS: Variant of Unknown Significance
WAGR: Wilms tumor, Aniridia, Genital anomalies, Retardation
WES: Whole-exome Sequencing
WGS: Whole-genome Sequencing
Wnt: Wingless