TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. D OSA - TOM. 1024

MEDICA - ODONTOLOGICA

MOLECULAR DIAGNOSTICS OF RARE INHERITED SYNDROMES WITH A VIEW ON DIAGNOSTIC TEST DEVELOPMENT

by

Pia Pohjola

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2012

From: Department of Medical Biochemistry and Genetics
University of Turku, Finland

Supervised by: Professor Klaus Elenius, M.D
Department of Medical Biochemistry and Genetics
University of Turku, Finland

and

Research professor Helena Kääriäinen, M.D
National Institute for Health and Welfare,
Helsinki, Finland

Reviewed by: Adjunct professor Marketta Kähkönen, PhD
Fimlab laboratories
Laboratory of Genetics
Tampere, Finland

and

Jukka Moilanen, M.D
Department of Clinical Genetics,
University of Oulu,
Oulu, Finland

Dissertation opponent: Adjunct professor Riitta Salonen-Kajander, M.D, PhD
Department of Medical Genetics
Family Federation of Finland
Helsinki, Finland

ISBN 978-951-29-5071-3 (PRINT)
ISBN 978-951-29-5072-0 (PDF)
ISSN 0355-9483
Painosalama Oy, Turku, Finland 2012

Pertulle ja Laurille

ABSTRACT

Pia Pohjola

MOLECULAR DIAGNOSTICS OF RARE INHERITED SYNDROMES WITH A VIEW ON DIAGNOSTIC TEST DEVELOPMENT

Department of Medical Biochemistry and Genetics, University of Turku
Annales Universitatis Turkuensis
Turku 2012

CHARGE syndrome, Sotos syndrome and 3p deletion syndrome are examples of rare inherited syndromes that have been recognized for decades but for which the molecular diagnostics only have been made possible by recent advances in genomic research. Despite these advances, development of diagnostic tests for rare syndromes has been hindered by diagnostic laboratories having limited funds for test development, and their prioritization of tests for which a (relatively) high demand can be expected.

In this study, the molecular diagnostic tests for CHARGE syndrome and Sotos syndrome were developed, resulting in their successful translation into routine diagnostic testing in the laboratory of Medical Genetics (UTUlab). In the CHARGE syndrome group, mutation was identified in 40.5% of the patients and in the Sotos syndrome group, in 34%, reflecting the use of the tests in routine diagnostics in differential diagnostics. In CHARGE syndrome, the low prevalence of structural aberrations was also confirmed. In 3p deletion syndrome, it was shown that small terminal deletions are not causative for the syndrome, and that testing with array-based analysis provides a reliable estimate of the deletion size but benign copy number variants complicate result interpretation. During the development of the tests, it was discovered that finding an optimal molecular diagnostic strategy for a given syndrome is always a compromise between the sensitivity, specificity and feasibility of applying a new method. In addition, the clinical utility of the test should be considered prior to test development: sometimes a test performing well in a laboratory has limited utility for the patient, whereas a test performing poorly in the laboratory may have a great impact on the patient and their family. At present, the development of next generation sequencing methods is changing the concept of molecular diagnostics of rare diseases from single tests towards whole-genome analysis.

Key words: CHARGE syndrome, Sotos syndrome, 3p deletion syndrome, *CHD7*, *NSD1*, genetic testing, molecular diagnostics

TIIVISTELMÄ

Pia Pohjola

HARVINAISTEN PERIYTYVIEN OIREYHTYMIEN MOLEKYYLIDIAGNOSTIIKKA JA NÄKÖKULMA DIAGNOSTISTEN TESTIEN KEHITTÄMISEEN

Lääketieteellinen biokemia ja genetiikka, Turun yliopisto
Annales Universitatis Turkuensis
Turku 2012

CHARGE-oireyhtymä, Sotosin oireyhtymä ja 3p deleetio -oireyhtymä ovat esimerkkejä harvinaisista periytyvistä oireyhtymistä, jotka on tunnettu kliinisinä kokonaisuuksina vuosikymmeniä, mutta joiden molekyylidiagnostiikka on tullut mahdolliseksi vasta genomitutkimuksen edistymisen myötä viime vuosina. Harvinaisten oireyhtymien molekyylidiagnostiikan kehittymistä on kuitenkin hidastanut diagnostisten laboratorioiden rajallinen mahdollisuus kehittää uusia testejä ja priorisointi testeihin, joilla voidaan olettaa olevan suurehko kysyntä. Lisäksi tutkimuslaboratoriossa kehitetyt testit eivät välttämättä päädy diagnostisiksi testeiksi.

Tässä tutkimuksessa kehitettiin molekyyligeneettinen testi CHARGE-oireyhtymälle ja Sotosin oireyhtymälle, ja siirrettiin oireyhtymien testaus osaksi lääketieteellisen genetiikan palvelulaboratorion (UTULAB) valikoimaa. 3p deleetio -oireyhtymälle ei oireyhtymän harvinaisuuden vuoksi kannattanut kehittää spesifistä testiä, vaan tarvittaessa testauksessa tehdään yhteistyötä molekyylisytogeneettisten laboratorioiden kanssa. Sotosin oireyhtymän tutkituista näytteistä mutaatio oli todettavissa 34 %:lla ja CHARGE-näytteistä mutaatio todettiin 40.5 %:lla. Lisäksi CHARGE-oireyhtymän kohdalla todennettiin koko geenien kattavien ja geenin sisäisten rakenteellisten mutaatioiden vähäinen esiintyvyys. 3p-oireyhtymän kohdalla todettiin, että pienet, terminaaliset deleetiot eivät aiheuta oireyhtymälle tyypillistä kliinistä kuvaa. Lisäksi todettiin, että potilaiden deleetioiden koko pystyttiin analysoimaan tarkasti mikrosirupohjaisella menetelmällä, mutta tulosten tulkintaa vaikeuttavat harvinaiset, ei-patogeeniset kopiolukuvariantit. Molekyylidiagnostisten testien kehittämistyössä todettiin, että testausmenetelmän valinta on aina kompromissi testien sensitiivisyyden, spesifisyyden ja uusien menetelmien käyttöönottamisen välillä, ja sen lisäksi joudutaan pohtimaan testien kliinistä hyödyllisyyttä. Osa testeistä toimii hyvin laboratoriossa, mutta niiden hyöty potilaalle on pieni, kun taas toisten testien hyödyt ovat suuret, vaikka testien toiminta laboratoriossa on heikompaa. Tulevaisuudessa harvinaisten tautien molekyylidiagnostiikka muuttunee yksittäisten geenien testauksesta kohti koko genomin tutkimusta uuden sukupolven sekvensointimenetelmien yleistyessä.

Avainsanat: CHARGE-oireyhtymä, Sotosin oireyhtymä, 3p deleetio -oireyhtymä, *CHD7*, *NSD1*, geenitestaus, molekyylidiagnostiikka

TABLE OF CONTENTS

ABBREVIATIONS

aCGH	array comparative genomic hybridization
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BDGP	Berkeley Drosophila Genome Project
cDNA	complementary DNA
CHD	congenital heart defect
CHD7	chromodomain helicase binding protein 7
CNV	copy number variant
DHPLC	denaturing high-performance liquid chromatography
DNA	deoxyribonucleic acid
EBV	Epstein-Barr virus
EQA	External Quality Assessment
ELSI	Ethical, legal and social issues
ES	exome sequencing
FISH	fluorescence *in situ* hybridization
HGVS	human genome variation society
HGMD	Human Gene Mutation Database
HMM	Hidden Markov Model
HMTase	histone methyltransferase
HUCH	Helsinki University Central Hospital
LNA	Locked Nucleic Acid
LCA	Leber Congenital Amaurosis
LCR	low copy repeat
MLPA	multiplex ligation-dependent probe amplification
MR	mental retardation
mRNA	messenger RNA
NAHR	non-allelic homologous recombination
NHGRI	National Human Genome Research Institute
NGS	next generation sequencing
NMD	nonsense-mediated mRNA decay
NORD	National Organization for Rare Disorders
NSD1	nuclear receptor binding SET domain 1
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
PHA	phytohemagglutin
PHD	plant homeodomain
PWWP	proline-tryptophan-tryptophan-proline

RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
SNP	single nucleotide polymorphism
UBCA	unbalanced chromosome abnormality
UCSC	University of California, Santa Cruz
UV	unclassified variant
WBDD	Winter-Baraitser dysmorphology database

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals I-IV.

I. Vuorela P, Ala-Mello S, Saloranta C, Penttinen M, Pöyhönen M, Huoponen K, Borozdin W, Bausch B, Botzenhart EM, Wilhelm C, Kääriäinen H, Kohlhase J: Molecular analysis of the CHD7 gene in the CHARGE syndrome: Identification of 22 novel mutations and evidence for a low contribution of large CHD7 deletions. Genetics in Medicine 2007; 9(10):690-694.

II. Vuorela PE, Penttinen MT, Hietala MH, Laine JO, Huoponen KA, Kääriäinen HA: A familial CHARGE syndrome with a CHD7 nonsense mutation and new clinical features. Clinical Dysmorphology 2009 Oct; 17(4):249-53.

III. Pohjola* P, de Leeuw N, Penttinen M, Huoponen K, Kääriäinen H: Terminal 3p deletions in two families–correlation between molecular karyotype and phenotype. American Journal of Medical Genetics A. 2010 Feb;152A(2):441-6.

IV. Pohjola* P, Peippo M, Penttinen M, Huoponen K, Elenius K, Kääriäinen H: Translation of a research based genetic test of a rare syndrome into clinical service testing with Sotos syndrome as an example. Accepted for publication in Genetic Testing and Molecular Biomarkers.

The original publications have been reproduced with the permission of the copyright holders. *Pia Pohjola (née Vuorela)

1 INTRODUCTION

Rare inherited syndromes have been recognized as clinical entities for decades without the knowledge of their etiology. Although diagnoses today are still largely based on the clinical examination and family history, during the past decade, molecular diagnostics has become an important part of the diagnostics workflow and has developed into an independent field of medicine. In the rare syndromes with fairly severe symptoms, such as CHARGE syndrome, Sotos syndrome and 3p deletion syndrome, patients are usually the first in their family to suffer from the symptoms, naturally generating high anxiety in the other family members. Finding the correct diagnosis – either by clinical examination or by molecular diagnostics – has an undeniable benefit: confirmation of a diagnosis is extremely important for the psychosocial health of a family, and the stressful search for diagnosis with different, often invasive, methods ends. In addition, consistent follow-up for possible future complications can be effectively arranged and family planning offered.

Despite the advances in the development of molecular genetic analysis methods and the growing knowledge of causative mutations, there are still a number of patients without the correct diagnosis and a number of syndromes without clarification of the genetic background. A recent major breakthrough in methodology, massively parallel sequencing, holds great promise and may be the method by which the etiology of even the rarest of syndromes could be dissected. The future will show what impact these methods will have on the field of molecular diagnostics.

2 REVIEW OF THE LITERATURE

2.1 Overview of rare diseases and syndromes

According to the Orphanet database of rare diseases "There is no disease so rare that it does not deserve attention. Rare diseases are rare, but rare disease patients are numerous" (http://www.orpha.net). To date, 6,000-8,000 rare diseases are recognized, 75% of them affect children and 30% of patients die before the age of 5 (http://www.eurordis.org). Rare diseases are often life-threatening or chronically debilitating, and for most of the diseases there is no effective treatment. About 80% of the rare diseases have a genetic origin and the remaining 20% are the result of infection, allergies and environmental causes (http://www.orpha.net).

There is no commonly accepted definition for a rare disease. Because the true prevalence of many rare diseases remains unclear, the European Union defines rare diseases as life-threatening or chronically debilitating diseases which are of such low prevalence that special combined efforts are needed to address them. As a guide, prevalence is considered low if it is less than 5 per 10,000. This definition is also used in Finland (http://ec.europa.eu/health/rare_disease/policy/index_en.htm, www.harvinaiset.fi). Although only a handful of people suffer from a certain disease, it has been estimated that in Europe, altogether nearly 30 million people are affected by diseases fitting into this category.

'A disease' can be broadly defined as a pathological condition of a part, organ, or system of an organism resulting from various causes, such as infection, genetic defect, or environmental stress, and characterized by an identifiable group of signs or symptoms and whose etiology, pathology, and prognosis may be known or unknown (http://www.thefreedictionary.com). 'A **syndrome'** can be defined as a broad error of morphogenesis in which the simultaneous presence of more than one malformation or functional defect is known or assumed to be a result of a single etiology, and use of the term implies that the group of malformations have been repeatedly seen in a fairly consistent and unique pattern (Cassidy and Allanson 2010).

The term '**genetic test'** can be defined as an analysis of human genetic material which detects the presence of, or risk for disorders with a strong heritable component (Javaher *et al*. 2008) Another definition for 'a genetic test' is a test performed in genetic testing laboratories (cytogenetics, molecular genetics and biochemical genetics) as part of genetic services (Eurogentest Unit 3, recommendations for genetic counseling in genetic testing). '**Diagnostic testing'** refers to a situation where

a genetic test is triggered by a patient presenting clinical signs or symptoms indicating a genetic disorder. In this case, the test is performed to confirm, refine or exclude a clinical diagnosis (Javaher et al. 2008). This study follows the definitions of Javaher *et al.*

The term 'syndrome' comes from the Greek words *syn* (together) and *dromos* (run). Rare heritable syndromes have been recognized for centuries, with the first description of trisomy 13 dating back as far as the 17th century (Jones 2005). Until 1945, clinical descriptions of new syndromes were published rarely in the literature. During the 1960s and 70s, several new syndromes were reported annually, representing the "Golden Age" of describing new syndromes (Figure 1). Since then, the number of new syndrome reports has declined and the research focus has been shifted to dissecting the genetic cause of the syndromes. In the 80s and 90s, the genetic background of many syndromes was revealed, for example those of most of the syndromes belonging to the Finnish Disease Heritage. Recently, methodological developments, such as massively parallel sequencing technologies, have facilitated in-depth genomic analyses, and reports of the mutational background of many previously genetically uncharacterized syndromes have thereby started to emerge. Similarly, these findings have led to more precise classification of known syndromes and revelation of their subtypes.

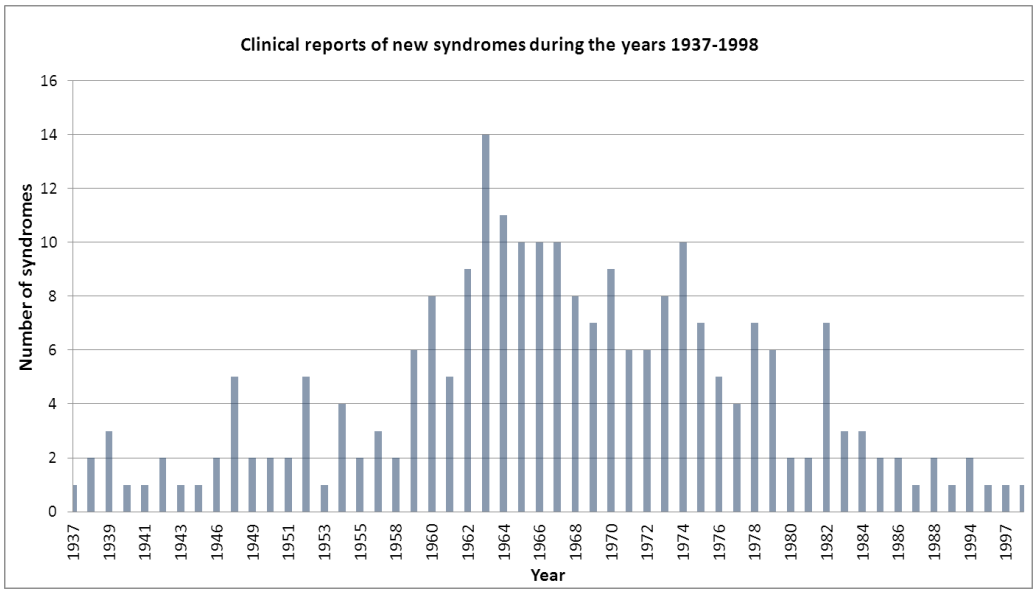

Figure 1: Number of reported syndromes annually during the years 1937-1998. Data collected from book "Smith's recognizable patterns of human malformation" (Jones 2005).

As the number of syndromes is in the thousands, different databases have been developed to aid in the syndrome diagnostics. For instance, the Winter-Baraitser Dysmorphology Database (WBDD) currently contains information on over 4,700 dysmorphic, multiple congenital anomaly and mental retardation syndromes (http://www.lmdatabases.com). Another dysmorphology syndrome database is Possum (http://www.possum.net.au/), in addition to which, for instance, OMIM (www.ncbi.nlm.nih.gov/omim), NORD (www.rarediseases.org), Orphanet (www.orpha.net), GeneTests (www.genetests.org) and Finnish Harvinaiset (www.harvinaiset.fi) collect and share information about rare diseases. GeneTests and Orphanet also provide information about laboratories offering genetic testing for rare syndromes.

Although individual syndromes can be extremely rare, due to their large number, they represent a significant part of medicine. It has been estimated that around 1% of the neonates have multiple anomalies, and in approximately half of these, a recognizable syndrome can be diagnosed. In many cases a syndrome is suspected when an infant has single or multiple severe anomalies, growth deviation from the normal range or a motor and/or mental developmental delay is noticed [Aula, Kääriäinen and Palotie (eds.) 2007].

Many syndromes have been named after the clinician who first described it, for example Weaver syndrome (Weaver et al. 1974), or the name is an acronym, such as Mulibrey nanism, which is coined from **mu**scle, **li**ver, **br**ain and **ey**es (Perheentupa *et al*. 1973).

2.2 A brief history of the molecular diagnostics of inherited diseases

Molecular diagnostics can be defined as the use of molecular biological techniques to expand scientific knowledge of the natural history of the diseases, identify people who are at risk of acquiring specific diseases and diagnose infectious and other human diseases at the nucleic acid level (Tsongalis and Coleman 2002). Although molecular diagnostics is also used to detect infectious diseases, in the following the emphasis is on the molecular diagnostics of inherited syndromes.

Molecular diagnostics is a relatively young field of medicine, and its development has closely followed the breakthrough inventions of its methodology (Demidov 2003). The basis for molecular genetics and diagnostics was already set in 1865 when the Austrian monk Gregor Mendel discovered the laws of inheritance. Until 1953, when the structure of DNA was revealed by Watson and Crick, human genetics developed as a basic science with clinical observation in families, biostatistics and population-based mathematical analyses. However, during this time the Men-

delian inheritance for several disorders, such as albinism and brachydactyly, was defined (Rimoin and Hirschhorn 2004).

The next half decade witnessed the rapid evolution of molecular genetics with its culmination in 2003, when the complete sequence of the human genome was published. During this era, the genetic code was described by Nirenberg in 1966, and technological inventions such as *in situ* hybridization in 1969, Southern blotting and sequencing techniques in 1975, development of the FISH method in 1980, PCR in 1983, and introduction of array technologies in the 1990s have had a tremendous impact on the development of molecular diagnostics in terms of resolution, turnaround time, reliability, number of tests offered, and value of the tests to the patients (Nirenberg *et al*. 1966, Gall and Pardue 1969, Sanger and Coulson 1975, Southern 1975, Mullis *et al*. 1986, Bauman *et al*. 1980).

The first reporting of the correct human chromosome number dates back to 1956, when Tjio and Levan reported that the correct human chromosome number is 46 instead of 48, as had been the assumption for many decades (Tjio and Levan 1956). Soon thereafter, reports describing chromosome abnormalities in Down, Turner and Klinefelter syndromes were published as well as reports of XXX females and XYY males, creating the field of medical cytogenetics (Ford *et al*. 1959, Jacobs *et al*. 1959, Jacobs and Strong 1959, Lejeune, Turpin and Gautier 1959, Sandberg *et al*. 1961). The first fetal karyotype analyses from the amniotic samples in 1966 initiated routine prenatal diagnostics (Steele and Breg 1966, Aula 1981). The invention of chromosome banding in 1970 allowed identification of individual chromosomes for the detection of translocations, deletions, duplications and inversions (Caspersson, Zech and Johansson 1970).

In 1967, chromosome loss in somatic cell hybrids between mouse and human cells was discovered, laying the foundations for chromosome mapping. The first locus mapped was the thymidine kinase locus to chromosome 17 (Weiss and Green 1967). Later, map-based gene discovery (positional cloning) and linkage analysis became the leading methods elucidating the molecular basis of genetic diseases. The first disease gene, the huntingtin, was mapped to chromosome 4 in 1983 using polymorphic markers (Gusella *et al*. 1983). The first human disease gene to be cloned was the gene for chronic granulomatous disease, followed soon by the Duchenne muscular dystrophy gene and the gene responsible for cystic fibrosis (Royer-Pokora *et al*. 1986, Koenig *et al*. 1987, Riordan *et al*. 1989). Diagnostic laboratories immediately adopted linkage analysis as a diagnostic tool, enabling the diagnosis of diseases segregating in families.

In 1966, the studies on congenital nephrosis of the Finnish type by Reijo Norio created the concept of the Finnish Disease Heritage (Norio 1966). Today, this refers to a

group of 40 rare single-gene diseases which are more common in Finland than other parts of the world (Norio 2003). The molecular genetic research of these diseases has been ongoing in the former National Public Health Institute, present National Institute for Health and Welfare, and other research laboratories since 1987, and to date, the genes and mutations behind all these diseases have been identified. In each disease, nearly all cases are caused by a founder mutation, which has facilitated the development of diagnostic tests for these diseases and diagnostic services.

Side by side with the advances in molecular diagnostics genetic counseling also developed. The term 'genetic counseling' was introduced in 1947, and three requirements for genetic counseling were delineated: knowledge of human genetics, respect for sensitivities, attitudes and reactions of clients, and teaching and providing genetic information to the fullest extent known (Reed 1955). In Finland, the Family Federation started genetic counseling already in 1951, and was followed in the 70s and 80s by the clinical genetic units in University hospitals. At that time, the diagnoses were based on the clinical examination and presence of typical features, and recurrence risks in the families were concluded from the family relationships and inheritance mode. Later, development of molecular diagnostics enabled the confirmation of clinical diagnoses and carrier status analyses of the symptomless relatives and prenatal diagnostics, and changed the view of genetic counseling. Today, genetic counseling comprehensively includes the correct indications for the correct test, pre-test counseling, consenting to the test, taking and sending the sample with adequate clinical information, performing high quality tests correctly, interpreting the result to the clinicians, post-test counseling, other post-test actions such as informing relatives and organizing possible follow-up and fair reimbursement policies (Eurogentest unit 3, http://www.eurogentest.org).

Molecular diagnostic laboratories originated from a need to serve patients with the findings generated in the research laboratories. In Finland, the first diagnostic prenatal amniotic fluid analyses were performed from 1975 on in the diagnostic unit operating in the pediatric, pathological and gynecological clinics of the former Helsinki University Central Hospital (HUCH; Aula 1978). In 1977, the first official prenatal diagnostic unit was established at the HUCH gynecology clinic (Aula 1981). As the need for tests was growing rapidly, another unit was established in the Oulu University Central Hospital, followed by units in the Kuopio, Tampere and Turku University Central Hospitals in the mid-1980s.

Development of molecular genetics in the late 1980s led to the founding of diagnostic genetic units in all university centers in Finland. In the universities of Helsinki, Oulu and Turku, departments of medical genetics and genetic technology had DNA diagnostic laboratories from the mid-1980s onwards (Aula 1991, Vesa

Juvonen, personal communication), and Tampere followed later in the mid-1990s (Kalle Simola, personal communication). Linkage analyses as diagnostic methods were rapidly substituted by direct mutation detection techniques, and today several hundred monogenic diseases can be detected with a specific gene test. In addition to university diagnostic laboratories, private companies and foundations, such as Yhtyneet Medix laboratories, the Minerva Foundation Institute for Medical Research and the Foundation for Pediatric Research, have also had a large role in the diagnostics of rare diseases.

With the development of new and improved methods, molecular diagnostics has taken an important role in disease diagnostics. The revelation of the first complete draft of the human genome in the early 2000 opened the gates of molecular diagnostics to medicine (Figure 2). The current number of diseases with a molecular genetic test available listed in the GeneTests database is 2,226, and a specific cytogenetic test is available for 133 diseases, with 392 clinical laboratories performing molecular genetic tests and 77 performing cytogenetic tests (www.genetests.org, accessed 5.3.2012).

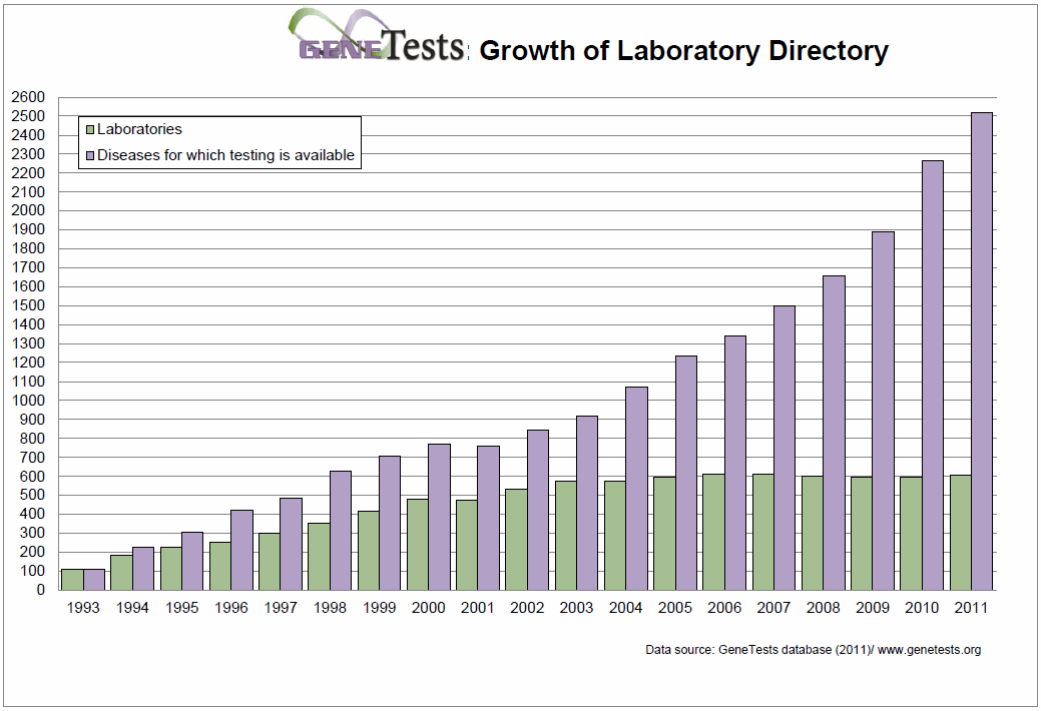

Figure 2: Number of laboratories performing gene tests and number of diseases with genetic tests available during the years 1993-2011. In: GeneTests: Medical Genetics Information Resource (database online). About GeneTests: Data available to the Public. Copyright, University of Washington, Seattle. 1993-2011. Available at http://www.genetests.org. Accessed 5.3.2011.

2.3 Diagnostic tests and test development

In the genomics era, molecular diagnostic tests for rare syndromes are developed continuously in research laboratories, and the growing knowledge of new causative mutations creates a need for developing new diagnostic tests for routine diagnostic laboratories. Every diagnostic laboratory, however, has to evaluate which tests can be adopted from the research setting for diagnostics and which are the factors that dictate the applicability of a given test for the purposes of a diagnostic laboratory. In most laboratories, instrumentation is not the limiting factor, but the challenge is to determine the most cost-effective test, taking into consideration the utility of the test for patients and their families. Different factors influence the translation of a research-based test into a diagnostic tool: diagnostic laboratories have limited funds and personnel for test development and they prioritize tests for which a (relatively) high demand can be expected. On the other hand, researchers may not have close contact with diagnostic laboratories willing to proceed with test translation (Das, Bale and Ledbetter 2008, Faucett *et al*. 2008, Grody and Richards 2008).

In the past, diagnostic tests have drifted from the research laboratories to routine diagnostics based on the demand from the clinicians. However, quality issues in diagnostic laboratories concerning the development of new tests have become more stringent, with requirements for laboratories to gain accreditation or participate in External Quality Assessment rounds (EQA), and to conduct the tests according to defined standards. Over recent years, guidelines and programs have been developed to aid the test implementation process (Das *et al*. 2008, Faucett *et al*. 2008, Jennings *et al*. 2009, Mattocks *et al*. 2010).

Previously, finding a laboratory performing a diagnostic test for a rare syndrome used to be difficult, but today, however, due to the combined efforts of the Orphanet and Eurogentest projects (www.orpha.net; www.eurogentest.org), clinicians can easily find the European laboratories performing a certain test in the order of their level of quality as shown by accreditation, and with notes on whether the laboratory in question participates in EQAs. The Genetest database (www.genetests.org) provides information about laboratories in the United States performing genetic tests and their level of quality. However, the dilemma of identifying the correct gene to be tested still remains: many conditions are genetically heterogeneous, and taking several gene tests one after the other is slow and expensive.

The components of a genetic test are analytical validity, clinical validity, clinical utility, and ethical, legal and social issues (ELSI) (Sanderson *et al*. 2005). 'Analytical validity' refers to the ability of the test to reliably measure the genotype of

interest, whereas 'clinical validity' is the ability of the test to detect or predict the phenotype of interest. 'Clinical utility' is the likelihood of the test leading to an improved outcome, assessment of risks and benefits of testing as well as economic evaluation. Ethical, legal and social issues depend on the type and context of the test used.

When a test appears to be ready for use in a clinical application, its analytical and clinical validity, and most importantly, clinical utility have to be demonstrated (Grody and Richards 2008). The analytical validity may be the most straightforward to determine, as the methods used in the research laboratory and the diagnostic laboratory may be much the same. However, the research laboratories have the flexibility to develop and test new methods, but translation of these methods into routine diagnostic tools requires a thorough validation process (Figure 3). When developing a test for a rare disease, a concern is to obtain a sufficient number of control samples for analytical validation. Determination of the clinical validity is usually more complicated, as the information about the correlation between the genotype and phenotype of a given syndrome is usually inadequate and only evolves with time. In particular, the mild end of the spectrum in many rare diseases is revealed only after the routine testing has begun.

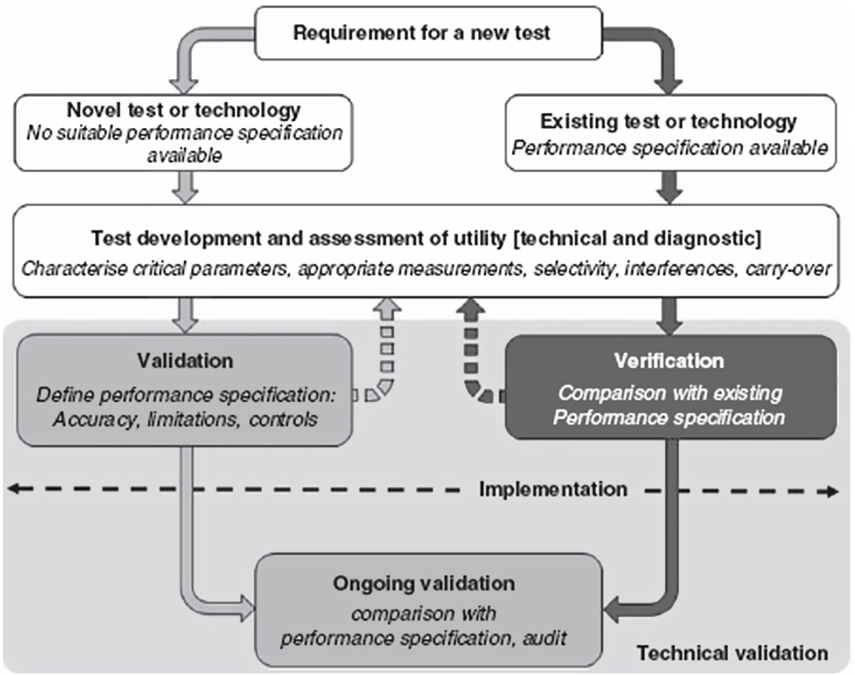

Figure 3. Implementation of a new molecular genetic test (from Mattocks *et al*. 2010). Reprinted with permission from copyright holder.

The major challenge is to balance clinical validity, clinical utility and cost-benefit issues: in some cases, a test performs superbly in the laboratory, but is not viable from the clinical or economic point of view. For instance, testing for non-syndromic hearing loss is an example of such testing, where many genes can have many different mutations and testing for only a subset of these has limited utility. On the other hand, some tests are limited in their validity, but nevertheless have great impact on patient and family management. An example of this kind of situation is hereditary hemochromatosis, where the disease genotype does not necessarily predict the presence of disease, but if the disease is diagnosed, effective treatment is readily available. Therefore, the requirements for a test should be defined in the context of its impact on the clinical setting while bearing in mind that the laboratory genetic test is only one of the many components of an overall evaluation (Eurogentest Unit 3, http://www. eurogentest.org).

2.4 Utility of molecular diagnostic testing

Due to the rarity of most syndromes, most clinicians meet patients with a certain syndrome only rarely, and identifying a given syndrome may therefore be difficult. In many syndromes, clinical phenotypes overlap, and experienced clinicians in different medical specialties are needed to establish the correct diagnosis. Several laboratory and diagnostic tests may also be required. The rarer the syndrome, the more difficult it is to find experts with experience of the treatment or families and patients with the same syndrome.

Upon asking the CHARGE families for the most stressful aspect of living with the syndrome, the delay in reaching the correct diagnosis was the sole answer (CHARGE syndrome support group day 2008, personal discussions with the families). For some of the diseases, testing is only available through research projects, and it may take years for the results to reach patients. For the rarest diseases, there may not be any kind of testing available. It has been estimated that between 30 to 40 percent of children with special needs remain without a correct diagnosis (National Human Genome Research Institute, http://www.genome.gov/17515951).

When molecular diagnostics becomes available for a given disease, there is an immediate benefit for the patients. A correct diagnosis based on a genetic test may render unnecessary other diagnostic procedures that may include risky and/or costly measures, guide clinical management including therapy, and facilitate specific genetic counseling. In cases of an affected child in the family, molecular diagnostics may enable prenatal diagnostics and provide the parents with the choice of not having a second affected child should they so decide. In addition, finding the

correct diagnosis can be of significant psychological impact. Molecular diagnostics also enables the finding of patients with mild or atypical symptoms, which have importance in assessing the recurrence risk of the diseases in their families (Javaher *et al*. 2008).

2.5 Changes in the human genome

2.5.1 Small scale mutations

The human genome is subject to variety of different types of heritable changes (mutations), and the study of mutations in human genes is of utmost importance for our understanding of the pathophysiology of inherited diseases and optimization of diagnostic testing. 'A mutation' can be defined as an alteration in DNA sequence, which occurs in the protein coding part of the gene or in the intronic and promoter regions, and is a change that is associated with disease (Maddalena *et al*. 2005). At the simplest level, changes can be divided into nucleotide level and chromosome level mutations, but the present advances in molecular cytogenetic methods have blurred this classification of changes as chromosomal abnormalities or molecular defects. In humans, the overall per-generation mutation rate is exceptionally high, but the mutation rate in the human germline is lower than for any other species (Lynch 2010).

Small scale mutations can be grouped into different mutation classes according to their effect on the DNA sequence: base substitutions, also known as point mutations, deletions and insertions. Base substitutions cause a single base in the DNA to be replaced by another, and the effect of this mutation depends on the location and the degree to which the mutation alters the function of the gene or its product. A *nonsense* mutation is a base substitution in which a codon specifying an amino acid is replaced by a translation stop codon, while in *missense* mutations the altered codon specifies another amino acid. *Splice site* mutations can alter the canonical splice site sequences or activate cryptic splice sites, resulting in mRNA altered in length, whereas in *frameshift* mutations, the reading frame is shifted due to an insertion, deletion, splice site alteration or whole-exon deletion or duplication. Small insertions add one or more nucleotides to the DNA sequence and may therefore alter splicing of the reading frame, and small deletions remove one or more nucleotides with consequences similar to the insertions. Mutations can arise spontaneously due to the errors in the replication, or be induced by radiation and chemicals (Strachan and Reed 2010).

In addition to the above-mentioned mutations, instability of trinucleotide repeats is another mechanism causing hereditary diseases. The trinucleotide repeats are

usually polymorphic in the population, but in some cases the repeats exist in a plastic premutation state which is prone to expansion even further, thereby generating a disease-causing mutation. An example of such a syndrome is fragile X syndrome, in which the number of repeats in a premutation is 50-200 and in a full mutation over 200 (Kremer *et al*. 1991, Devys *et al*. 1992).

Single nucleotide substitutions are the most frequent pathologic mutations in the human genome, followed by small deletions and insertions (less than 20 bp) and gross deletions, as recorded by the Human Gene Mutation Database (www. HGMD, Figure 4). Studies have shown that G and C nucleotides are approximately twice as mutable as A and T, and transitions (purine→purine, pyrimidine→pyrimidine) are twice as probable as transversions (purine→pyrimidine, pyrimidine→purine) (Gojobori, Li and Graur 1982, Krawczak, Ball and Cooper 1998). Within the disease-causing genes, pathogenic mutations are distributed unevenly, with an overrepresentation in conserved domains and underrepresentation in variable regions (Miller *et al*. 2003). Currently, HGMD lists 4,411 genes with over 100 000 pathogenic mutations, but these are likely to be only a small fraction of all disease-causing mutations (the "mutome") yet to be identified (Chen, Férec and Cooper 2010).

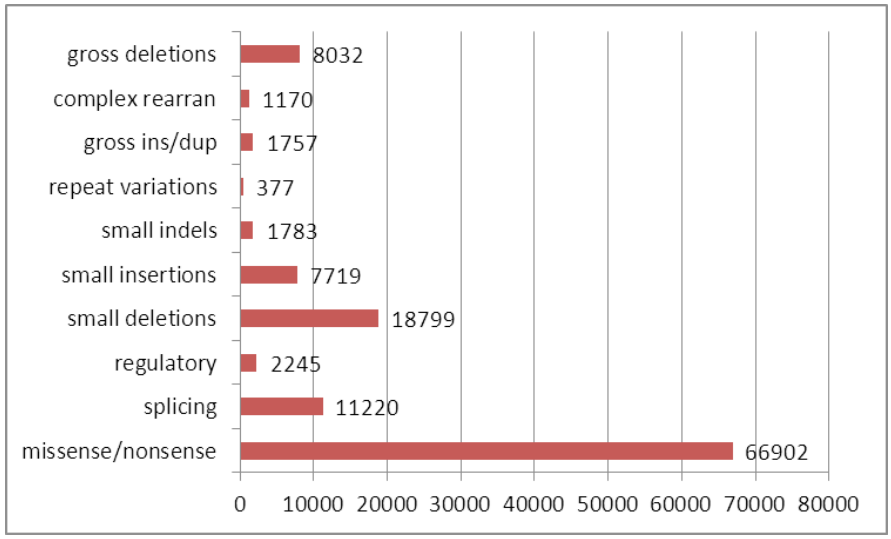

Figure 4. The spectrum of different mutations in the Human Gene Mutation Database (professional version). Accessed 30.3.2012.

2.5.2 Chromosome abnormalities

Recently, high-resolution genome-wide analyses of the human genome sequence have revealed extensive submicroscopic structural variation in addition to previ-

ously known single nucleotide variation (SNVs). This structural variation involves copy number variants (CNVs), such as deletions, duplications, triplications, insertions, and unbalanced cryptic translocations, as well as balanced genomic inversions. Recent studies have revealed human genomes differing more as a consequence of this structural variation than due to single-base-pair differences (Iafrate *et al*. 2004, Sebat *et al*. 2004, Redon *et al*. 2006, Conrad *et al*. 2010).

CNVs can lead to genomic disorders known as microdeletion and microduplication syndromes, which are common diseases affecting between 1 in a 100 to 1 in a 1,000 newborns, and that are often sporadic (Shaffer and Lupski 2000). Several mechanisms mediate the formation of structural rearrangements and of these, nonallelic homologous recombination (NAHR) has been extensively characterized (Ou *et al*. 2011). NAHR results in recurrent rearrangements which include the same genomic interval occurring in unrelated individuals. Rearrangements are a consequence of the underlying architecture of the genome, in which the rearranged interval is flanked by paralogous repeat sequences or low-copy repeats (LCRs, also known as segmental duplications) with high sequence identity. To date, 37 genomic regions have been reported to be associated with genomic disorders caused by NAHR (Liu *et al*. 2012). On the other hand, replication mechanisms are a major contributor to nonrecurrent genomic rearrangements wherein the rearrangement size, genomic extent, and breakpoint position at a genetic locus can differ amongst unrelated subjects.

The clinical consequences of CNVs are determined by the size of the genomic rearrangement, the total number and status of genes within the CNV, and the mode of inheritance. The majority of disease-causing CNVs contain one or more dosage-sensitive/haploinsufficient genes that produce an abnormal phenotype via a decreased or increased amount of encoded protein. Clinically relevant CNVs are collected in databases such as the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER; https://decipher.sanger.ac.uk/information) and in the European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA; http://www.ecaruca.net).

In addition to microdeletions and microduplications, intragenic deletions and duplications have been recognized in the pathogenesis of several syndromes. Intragenic aberrations can account for 3-10% of causative changes in different syndromes (Kirchhoff *et al*. 2007). These small scale rearrangements involve from one to a few exons, and special methods such as MLPA are needed in the detection.

2.6 Diagnostic methods

The essential part of molecular diagnostics of rare syndromes is formed by different molecular detection methods, as the detection of the variants solely relies on them. Currently, more than 50 different molecular detection methods are in use, and new methods are being constantly developed (Tsongalis and Silverman 2006). However, only a limited number of methods are suitable for diagnostic use in terms of reliability, repeatability, amount of sample required and cost-effectiveness. The selection of the method to be employed is dependent on the nature of variants to be detected, for example the change of a single nucleotide as opposed to thousands of nucleotides, the expected number of unique variants and the number of samples to be analyzed. In the following, some of the most common methods in molecular diagnostics are presented.

2.6.1 Molecular methods

After its invention in 1975 (Sanger and Coulson 1975), *Sanger sequencing,* also known as the chain termination method, has been one of the leading methods in molecular diagnostics and research. The order of nucleotides in DNA is solved by using modified dideoxynucleotides and capillary electrophoresis, enabling the detection of point mutations, small insertions and deletions, but large-scale mutations such as balanced rearrangements and microdeletions remain undetected.

Polymerase chain reaction (PCR) is a basic DNA amplification method widely used in molecular biology that is based on the ability of the thermostabile polymerase enzyme function in high temperatures (Mullis et al. 1986). Using sequence-specific primers, the DNA sequence of interest is amplified million-fold, and the amplification products are stained and visualized using agarose gel electrophoresis. Further development of the method is *real-time quantitative PCR*, in which the number of gene copies in a sample can be deduced by comparing its yield with that of control samples with a known number of copies. The amplification process is visualized using fluorescent dyes and results generated as the run proceeds without the need for post-run applications. Real-time PCR is widely used in gene expression analyses, analyzing single nucleotide polymorphisms, and detecting mutations affecting the copy numbers of the genes.

Restriction fragment length polymorphism (RFLP) is a simple point mutation detection method based on the endonuclease enzymes digesting DNA at the sites of their recognition sequences. Sequences of interest are amplified with PCR, digested with a suitable enzyme, and visualized on an agarose gel. Mutations creating or destroying recognition sequences affect the length of the digestion products, causing an

aberrant fragment pattern. RFLP has been commonly used in the mutation detection and SNP analyses, but has largely been replaced by sequencing.

Denaturing high-performance liquid chromatography (DHPLC) is a point mutation detection method based on the formation of homo- and heteroduplexes, and the different attachment properties of these duplexes to a solid surface (Oefner and Underhill, 1998). The melting curves generated in the chromatograms of normal and mutation samples have a different elution profile. As no labeling or purification of samples is needed, the method can be easily automated. DHPLC has been applied to SNP detection and mutation screening of large genes.

In 2002, *multiplex ligation-dependent probe amplification (MLPA)* was developed to detect larger rearrangements, i.e. deletions and duplications from single exons to regions of multiple genes (Schouten *et al*. 2002). Using a set of region-specific probes, the target DNA is amplified with PCR and the generated products are quantitated with capillary electrophoresis. MLPA can be used for diagnosing any disease with exonic or chromosomal rearrangements in a high-throughput fashion, and it is therefore widely used in diagnostic and research laboratories.

A recent breakthrough in methodology has been the *next generation sequencing (NGS)* techniques. Also known as massively parallel sequencing, these techniques allow the sequencing of millions of DNA molecules simultaneously, and sequencing of the entire human genome can now be performed in a short period of time with a reasonable cost. The application of NGS to genome research has led to the discovery of mutations behind known syndromes with unexplained etiology, such as Coffin-Siris syndrome (Tsurusaki *et al*. 2012), Kabuki syndrome (Ng *et al*. 2010), Weaver syndrome (Tatton-Brown *et al*. 2011, Gibson *et al*. 2012) and Floating Harbor syndrome (Hood *et al*. 2012). These techniques have opened new possibilities in the diagnostic field, but currently there are still challenges, such as the large amount of computational power and number of personnel required, the high number of detected variants and the economic issues that hinder the application of NGS to routine diagnostics.

2.6.2 Cytogenetic and molecular cytogenetic methods

The karyotype analysis is an informative and still commonly-used method based on the different staining properties of the chromosomal regions producing a distinct pattern of light and dark bands in each chromosome. The method allows the identification of each chromosome, rapid detection of numerical chromosomal aberrations, such as trisomy 21 (Down syndrome) and gross rearrangements that are larger than 3 Mb, such as deletions and translocations. *Fluorescence in situ hybridization (FISH)* is based on the hybridization of fluorescently labeled probes specific

to the chromosome region of interest, ranging from the smallest probes recognizing only one gene (locus-specific FISH) to the largest sets of probes hybridizing to all chromosomes (chromosome painting, multicolor FISH). Locus-specific FISH has been widely used in syndrome diagnostics, whereas whole-chromosome painting has been efficient in detecting complex rearrangements in, for example, cancer.

In recent years, the use of different types of *arrays* has become a common diagnostic method due to their high resolution and wide genomic coverage. In arrays, probes for the target of interest are spotted onto a glass slide and labeled samples are hybridized to the probes. Depending on the nature of the probes (BACs, oligonucleotides or SNPs), the smallest variants detected can be only a few kilobases in length, and also the breakpoints of rearrangements can be effectively defined. *Array comparative genomic hybridization (aCGH)* merges traditional chromosome analysis with molecular diagnostics. The detection of copy number variants is based on the competitive hybridization of samples onto the probes (Pinkel *et al*. 1998). *SNP arrays,* which were originally developed for high-throughput genotyping purposes, allow the quantitation of up to one million SNPs simultaneously, and in addition to detecting copy number variation also allow haplotype analyses. Arrays are used in diagnostics as a first-line method in situations where patients have nonspecific symptoms, for example an intellectual disability or developmental disorder (Forsström 2012), and their use has led to the identification of new microdeletion and microduplication syndromes, such as the 17q21.3 microdeletion syndrome (Koolen *et al*. 2006).

2.7 Clinical features and genetic background of the syndromes studied

2.7.1 The clinical presentation of CHARGE syndrome

CHARGE syndrome (OMIM #214800) is an autosomal dominant congenital anomaly syndrome with an incidence between 1:8,500 and 1:10,000 (Issekutz *et al*. 2005, Blake *et al*. 1998). In 1979, Hall and Hittner described patients presenting with coloboma, choanal atresia and congenital heart defects, and suggested an association between these features (Hall 1979, Hittner *et al*. 1979). In 1981, the acronym CHARGE was introduced by Pagon and colleagues, summarizing the main clinical features: **c**oloboma, congenital **h**eart defects, choanal **a**tresia, **r**etardation of growth and/or development, genital **a**nomalies and **e**ar anomalies (Pagon *et al*. 1981). It first became known as the CHARGE association, and after the molecular etiology was revealed, as CHARGE syndrome.

In addition, several other features with variable frequency have also been described in CHARGE syndrome patients. These include temporal bone anomalies, charac-

teristic facial features, facial nerve palsy, cleft lip/palate (sometimes replaced with choanal atresia), abnormal olfactory bulb development, hearing loss, behavioral problems and delayed puberty, immune deficiency and skeletal abnormalities (Graham *et al*. 2005, Pinto *et al*. 2005, Aramaki *et al*. 2006, Jongmans *et al*. 2006, Lalani *et al*. 2006, Delahaye *et al*. 2007, Sanlaville and Verloes 2007).

The current clinical criteria for the diagnosis of CHARGE syndrome were defined by Blake *et al* in 1998 with an update by Verloes in 2005 (Table 1, Blake *et al*. 1998, Verloes 2005). According to Blake, patients likely to have CHARGE syndrome harbor either four major symptoms (choanal atresia, coloboma, characteristic ears and cranial nerve anomalies) or three major and three of the minor symptoms including cardiovascular malformations, genital hypoplasia, cleft lip/palate, tracheoesophageal fistula, growth deficiency, developmental delay and distinctive facial features. In Verloes's criteria, three major symptoms are needed to diagnose CHARGE, and he divided the clinical presentation into typical, partial or atypical CHARGE based on the number of the patient's major and minor symptoms. The main differences between the two criteria are that Verloes included semicircular canal malformations as a major criterion and his criteria are less sex and age dependent. In both sets of criteria, coloboma and choanal atresia must be present for a patient to have CHARGE syndrome.

Table 1. Current diagnostic criteria for CHARGE syndrome according to Blake and Verloes.

	Major criteria	**Minor criteria**	**Inclusion rule**
Blake	1. Coloboma, microphthalmia 2. Choanal atresia or stenosis 3. Characteristic external ear, middle/inner ear malformations, mixed deafness 4. Cranial nerve dysfunction	1. Genital hypoplasia 2. Developmental delay 3. Cardiovascular malformations 4. Growth deficiencies 5. Orofacial cleft 6. Tracheoesophageal-fistula 7. Characteristic face	Four majors OR three majors +three minors
Verloes	1. Ocular coloboma 2. Choanal atresia 3. Hypoplasia of semicircular canals	1. Rhombencephalic dysfunction 2. Hypothalamo-hypophyseal dysfunction 3. Malformation of the ear 4. Malformation of mediastinal organs 5. Mental retardation	Typical CHARGE: three majors OR two majors + two minors Partial CHARGE: two majors +one minor Atypical CHARGE: two majors but no minors OR one major + two minors

Table modified from Sanlaville and Verloes (2007) and Bergman *et al*. (2011).

The CHARGE syndrome phenotype shows overlapping features with those described in the 22q11.2 deletion syndrome (congenital heart disease, cleft palate, hearing loss and learning difficulties) and Kallmann syndrome (hypogonadotropic hypogonadism and anosmia).

2.7.2 Mutations in the CHD7 gene

Some light was shed on the molecular etiology of CHARGE syndrome in 2004, when Vissers *et al*. reported patients with overlapping microdeletions in the 8q12 region spanning the large *CHD7* gene (chromodomain helicase DNA binding protein 7, Vissers *et al*. 2004)þ. Subsequent sequencing of the *CHD7* gene revealed heterozygous intragenic mutations in 10 patients, and further studies have confirmed mutations in the *CHD7* gene as the major cause of CHARGE syndrome.

To date, hundreds of different heterozygous point mutations have been described in the *CHD7* gene (Vissers *et al*. 2004, Aramaki *et al*. 2006, Jongmans *et al*. 2006, Lalani *et al*. 2006, Sanlaville *et al*. 2006, Delahaye *et al*. 2007, Wincent *et al*. 2008, Bartels *et al*. 2010, Bilan *et al*. 2012). The number of recurrent mutations is low, with nearly all patients carrying a unique mutation. The vast majority (~70%) of mutations are truncating mutations, clearly indicating haploinsufficiency as the pathogenic mechanism. A minority of the mutations are missense (10%) and splice site (13%) mutations (Zentner *et al*. 2010).

Mutations have been detected throughout the coding sequence without clear mutational hotspots, although exons 2, 8, 31 and 34 have been suggested to be more susceptible to mutations due to the fact that in several studies many mutations have accumulated in these exons (Bilan *et al*. 2012). Bergman *et al*. reported in their study in 2011 that exons 2, 3, 30, and 31 harbor 34% of all the detected mutations and exons 8, 12, 26, 30 and 36 show a remarkably high number of mutations related to their size (19% of mutations, 9% of size) (Bergman *et al*. 2011). Moreover, IVS25 splicing mutations and c.2504_2508delATCTT have been described in several publications; therefore, these locations could act as hotspots (Bilan *et al*. 2012).

Microdeletions or small intragenic rearrangements represent a minority (5%) of the disease-causing mutations in CHARGE syndrome (Bergman *et al*. 2008, Wincent *et al*. 2008, (Wincent, Schulze and Schoumans 2009). However, the number of these mutations may be underrepresented, since structural aberrations have not been analyzed in all of the reported series.

Mutations usually occur *de novo*, but rare occasions of familial parent-to-child transmission have been described (Lalani *et al*. 2006, Delahaye *et al*. 2007, Jongmans *et al*. 2008, Pauli *et al*. 2009). In the familial cases, the affected parent has presented

with very mild symptoms or has been asymptomatic, and some have been carriers of gonadal/somatic mosaicism (Jongmans *et al*. 2008, Pauli *et al*. 2009, Pauli *et al*. 2012). A recent study suggested that mutations are predominantly of paternal origin (Pauli *et al*. 2012). No evidence of genotype-phenotype correlation has been observed even in large groups of patients, and monozygotic twins with identical mutation have been reported to present with variable phenotypes, as well as unrelated patients with identical mutations (Jongmans *et al*. 2006, Lalani *et al*. 2006, Jongmans *et al*. 2008). The reasons for the variable phenotypes in families are currently unknown.

Rarely, *CHD7* mutations have been detected in patients with Kallman syndrome, DiGeorge syndrome and Omenn-like syndrome (Ogata *et al*. 2006, Sanka *et al*. 2007, Gennery *et al*. 2008). In addition, in a clinically diagnosed patient a *SEMA3E* mutation has been described (Lalani *et al*. 2004).

Of patients clinically diagnosed or suspected of having CHARGE syndrome, 40%-70% have an identifiable mutation in the *CHD7* gene, whereas in rest of the patients the diagnosis remains purely clinical (Aramaki *et al*. 2006, Jongmans *et al*. 2006, Lalani *et al*. 2006, Zentner *et al*. 2010, Bergman *et al*. 2011). The features which almost always are present with the *CHD7* mutation are ocular colobomas, external ear anomalies, semicircular canal hypoplasia, cranial nerve dysfunction and delayed attainment of motor milestones. As opposed to these, heart defects, choanal atresia/stenosis, genitourinary abnormalities, clefting and tracheoesophageal fistula are as common in patients with and without mutations (Zentner *et al*. 2010, Bergman *et al*. 2011).

After the molecular analysis of CHARGE syndrome became available, it has become evident that the mild end of the symptom spectrum is only now starting to be revealed, and these patients cannot be diagnosed using strict clinical criteria. In the study of a large group of patients done by Bergman *et al* (2011), even 17% of the *CHD7* mutation positive patients were only mildly affected and could not be diagnosed using strict clinical criteria.

2.7.3 The CHD7 protein and its function

The CHD7 protein consists of 2,997 amino acids and belongs to a CHD family of 9 ATP-dependent chromatin remodelers contributing to changes in chromatin structure during transcription, recombination, repair and replication (Woodage *et al*. 1997, Hall and Georgel 2007, Marfella and Imbalzano 2007). The CHD family is characterized by N-terminal tandem chromodomains and an SNF2-like domain in the central section of the protein, and can be further divided into three subgroups

based on additional functional domains and functional partners. Together with CHD6, CHD8 and CHD9, CHD7 belongs to group III, which is defined by additional paired BRK, SANT-like, CR and DNA-binding domains (Marfella and Imbalzano 2007).

The expression studies carried out with mouse embryos revealed variable relative expression levels in different tissues (Bosman *et al*. 2005, Lalani *et al*. 2006). The highest expression levels were in those tissues in which congenital abnormalities are frequently detected in CHARGE syndrome: the outflow tract of the heart, facioacoustic preganglion complex, optic vesicle, brain and olfactory pit. Sanlaville *et al*. (2006) studied the gene expression in 10 human fetuses with truncating *CHD7* mutations and noticed that the expression is ubiquitous until day 20, after which it is gradually restricted to the tissues affected in CHARGE syndrome. In the global gene expression microarray analysis, a distinct expression pattern was seen in individuals with *CHD7* mutation compared to those without (Lalani *et al*. 2006). Hurd *et al*. (2011) demonstrated that deletion of *CHD7* in the developing otocyst results in cochlear hypoplasia and complete absence of the semicircular canals and the christae.

The specific functions of the CHD7 protein are largely unknown, but, based on the domain composition, it has been thought to play an essential role in regulating chromatin structure and gene expression. Recent findings have revealed its involvement in the regulation of neural crest formation and neural stem cell development (Bajpai *et al*. 2010). In another study, CHD7 was shown to be a cofactor of the SOX2 protein, which is also an essential regulator of neural stem cells. The study revealed that SOX2 and CHD7 have physical interaction and overlapping genome-wide binding sites, and they regulate a set of common target disease-causing genes including *JAG1*, *GLI2*, *GLI3* and *MYCN*, creating a novel CHD7-SOX2 pathway (Engelen *et al*. 2011).

2.7.4 Sotos syndrome

In 1964, Sotos *et al*. described five patients with excessively rapid growth, acromegalic features and a non-progressive cerebral disorder with mental retardation, and defined the condition as a syndrome based on the similarity of the cases and the obvious absence of other recognized causes of overgrowth (Sotos *et al*. 1964). Cole and Hughes (1994) analyzed 79 patients with Sotos syndrome and concluded that a typical facial gestalt, pattern of overgrowth, bone age and developmental delay are the major features of Sotos syndrome (Cole and Hughes 1994).

The autosomal dominant Sotos syndrome belongs to a heterogeneous group of overgrowth syndromes characterized by either isolated or generalized overgrowth

(Neri and Moscarda 2009). There are tens of syndromes characterized by tall stature, many of them extremely rare, but those that a clinician usually has to consider as differential diagnostic possibilities include Bannayan-Riley-Ruvalcaba syndrome, Beckwith-Wiedemann syndrome, Weaver syndrome, Simpson-Golabi-Behmel syndrome and Perlman syndrome. In addition to overgrowth, these syndromes are associated with a risk for malignancy and cognitive impairment. Finding the correct diagnosis may be complicated with the clinical overlap in these syndromes, but the advances in dissecting the molecular basis of these syndromes have significantly aided the differentiation diagnostics.

The most prominent clinical features associated with Sotos syndrome are facial dysmorphism, cognitive impairment and childhood overgrowth, which are present in more than 90% of the patients with molecularly confirmed Sotos syndrome (Tatton-Brown *et al*. 2005). Facial characteristics include a high and broad forehead with high hairline, down-slanting palpebral fissures, and a pointed chin. The facial features are more evident between the ages of one and six, and become less prominent in adulthood (Fickie *et al*. 2011). The cognitive status varies from normal to severely impaired with most patients having mild to moderate difficulties. The overgrowth at birth may be evident, but is not proportionally increased, and the final height usually remains within the normal range.

Other features associated with the syndrome are advanced bone age, abnormalities in cranial imaging, jaundice and difficulty of feeding, neonatal hypotonia, cardiac and renal anomalies, seizures and scoliosis (Tatton-Brown *et al*. 2005). Previously, Sotos syndrome was also considered a cancer predisposition syndrome, but there is no strong evidence either way.

2.7.5 The NSD1 gene and mutations

A connection between Sotos syndrome and the *NSD1* gene was reported in 2002, when a Japanese group described a patient with translocation disrupting the *NSD1* gene (Kurotaki *et al*. 2002). *NSD1* is a large gene consisting of 23 exons located in chromosomal band 5q35.3. Mutations in the *NSD1* gene can be of all types, including heterozygous point mutations, partial gene deletions and hemizygous microdeletions of variable size containing the entire *NSD1* gene. The probable underlying pathogenic mechanism is haploinsufficiency, as most of the point mutations and deletions apparently cause a 50% reduction of the expression.

A difference in the mutation type frequency between the Japanese and non-Japanese patients has been observed: in 50% of the Japanese patients a common microdeletion of 1.9 Mb can be detected, whereas in non-Japanese patients microdele-

tions are more variable in size and account for only 10% of all mutations (Kurotaki *et al*. 2002, Douglas *et al*. 2003, Kamimura *et al*. 2003, Rio *et al*. 2003, Türkmen *et al*. 2003, Tatton-Brown *et al*. 2005). This difference is caused by a low copy repeat (LCR) inversion polymorphism surrounding the *NSD1* gene more frequently found in the Japanese population, and probably predisposing them to the common microdeletion (Kurotaki *et al*. 2003, Kurotaki *et al*. 2005, Visser *et al*. 2005a, Visser *et al*. 2005b). Microdeletions are predominantly of paternal origin (Miyake *et al*. 2003, Tatton-Brown *et al*. 2005)

Around 90% of the mutations in non-Japanese patients are intragenic missense, nonsense, frameshift and splice site mutations. Hundreds of different mutations have been described throughout the gene, and no hotspot for the mutations has been discovered, instead the mutations are scattered throughout the gene (Douglas *et al*. 2003, Kamimura *et al*. 2003, Nagai *et al*. 2003, Rio *et al*. 2003, Türkmen *et al*. 2003, Tatton-Brown *et al*. 2005, Saugier-Veber *et al*. 2007). Most of the mutations are nonsense or frameshift mutations causing a premature stop codon, and structural aberrations account for 5-15% of the mutations (Tatton-Brown *et al.* 2005, Fagali *et al*. 2009). Most patients carry a novel mutation, but rare cases of familial transmission have been reported. Studies of familial cases have suggested that they are more likely due to a missense mutation and these familial mutations tend to be located outside the functionally important SET domain (Douglas *et al*. 2003, Höglund *et al*. 2003, Kurotaki *et al.* 2003, Tatton-Brown *et al*. 2005, Saugier-Veber *et al*. 2007).

An *NSD1* mutation can be detected in 90% of the clinically diagnosed Sotos patients, with a small group of clinically diagnosed patients remaining without a molecularly confirmed diagnosis (Tatton-Brown *et al*. 2005, Saugier-Veber *et al*. 2007). Analysis of the promoter region of the *NSD1* gene and the *NSD2* and *NSD3* genes did not reveal any abnormalities to explain the syndrome in the latter group (Visser *et al*. 2006, Douglas *et al*. 2005). To date, mutations in other genes are not known to be associated with Sotos syndrome.

With their analysis of 116 classical Sotos syndrome patients, Saugier-Veber *et al*. (2007) suggested that patients with non-truncating mutations or in-frame deletions might have a less severe phenotype. When comparing patients with mutations and microdeletions, it seems that patients with deletions have a more severe learning disability, a coarser facial gestalt and less pronounced overgrowth (Douglas *et al*. 2003, Tatton-Brown *et al*. 2005). However, the learning disability and growth retardation are common features in microdeletions throughout the genome (Devriendt and Vermeesch 2004). All features present in patients with microdeletions were also present in patients with other mutations, and the size of the dele-

tion did not correlate with the clinical phenotype (Tatton-Brown *et al*. 2005). This suggests that genes other than *NSD1* have only a minor effect on the symptoms of the Sotos syndrome.

NSD1 mutations are not a frequent finding in other syndromes, suggesting that they are tightly Sotos syndrome related. In rare cases, however, *NSD1* mutations have been detected in cases of Beckwith-Wiedemann, Nevo and Weaver syndromes, familial overgrowth and certain malignancies (Douglas *et al*. 2003, Baujat *et al*. 2004, van Haelst *et al*. 2005, Kanemoto *et al*. 2006, Morishita and di Luccio 2011), but it is unknown whether these cases represent rare clinical features belonging to Sotos syndrome. Previously, Weaver syndrome, also belonging to overgrowth syndromes, was considered allelic to Sotos syndrome due to the highly overlapping clinical features and rare *NSD1* mutational findings (Douglas *et al*. 2003). Recent reports, however, revealed the *EZH2* gene (enhancer of zeste, drosophila homolog 2) mutations are causative of the Weaver syndrome (Tatton-Brown *et al*. 2011, Gibson *et al*. 2012). The clinical overlap between these syndromes is probably due to the fact that both *NSD1* and *EZH2* are histone methyltransferases playing a major role in embryonic development.

2.7.6 The NSD1 protein

NSD1 belongs to a family of three NSD HMTase proteins with NSD2 (MMSET/WHSC1) and NSD3 (WHSC1L1). The *NSD2* deletion is essential in the development of Wolf-Hirschorn syndrome, and *NSD3* is amplified in breast cancer cell lines (Stec *et al*. 1998, Angrand *et al*. 2001). The proteins encoded by these genes share a high degree of similarity despite NSD2 and NSD3 lacking the longest exon 5 in the 5' portion of the protein and nuclear receptor interaction domains found in NSD1. The protein product of the *NSD1* gene consists of 2696 amino acids and is characterized by several functional domains, including two PWWP (proline-tryptophan-tryptophan-proline), five PHD (plant homeodomain), C5HC3 and SET (suvar, enhancer of zeste, trithorax) domains (Huang *et al*. 1998).

Little is known about the specific functions of the NSD1 protein, but based on the functional domain composition it has been thought to play a role in post-translational chromatin regulation in a wide range of functions. The profound developmental defect seen in Sotos patients results in a 50% dosage of wild-type NSD1, indicating that the target genes are highly sensitive to the NSD1 dosage (Pasillas, Shah and Kamps 2011). Studies performed with Sotos syndrome mutant proteins have shown that missense mutations concentrating on the SET domain cause a great or nearly complete reduction of the histone lysine methyltransferase activity of the protein, with the loss of activity being the biochemical basis of Sotos syndrome

(Qiao *et al*. 2011). Normally, the SET domain catalyzes methylation of histone H3 Lys36 (H3K36) and histone H4 Lys20 (H4K20). H3K36 methylation is a mark of expressed genes and is associated with the suppression of intragenic transcription initiation in expressed genes, and H4K20 methylation is associated with gene repression and activation, mitosis and DNA damage checkpoint signaling (Rayasam *et al*. 2003, Lee and Shilatifard 2007).

A similar study performed with mutant proteins showed that mutations in the PHD domains of NSD1 disrupted the binding of the protein to the methylated H3K4 and H3K9, and, in addition, disrupted binding to the NZIP1, in which mutations have been shown to cause a Sotos-like phenotype (Nielsen *et al*. 2004, Pasillas *et al*. 2011).

2.7.7 3p deletion syndrome

The 3p deletion syndrome is a rare but fairly well recognized contiguous gene syndrome (Verjaal and De Nef 1978). The underlying cause of this syndrome is deletions in the distal region of the short arm of chromosome 3.

The syndrome is characterized by low birth weight, growth retardation, and microcephaly with typical facial features. Facial features include flat occiput, triangular face, bushy eyebrows with synophrys, narrow forehead, hypertelorism, ptosis, epicanthal folds, upslanting palpebral fissures, narrow nose, long philtrum and downturned mouth. The ears are small and malformed with occasional preauricular pits. Most patients have developmental delay varying from mild to severe. In addition, patients suffer from cardiovascular and urinary tract abnormalities, sacral dimple, abnormal dermatoglyphics and hearing loss. The severity of the symptoms varies tremendously (Verjaal and De Nef 1978, Merrild *et al*. 1981, Witt, Biedermann and Hall 1985, Benini *et al*. 1999, Malmgren *et al*. 2007, Fernandez *et al*. 2008).

Most cases occur sporadically, but in some families a parent with normal appearance or only with mild symptoms has transmitted the deletion to their offspring (Tazelaar *et al*. 1991, Knight *et al*. 1995, Shrimpton, Jensen and Hoo 2006, Takagishi *et al*. 2006, Cuoco *et al*. 2011). In these cases, the offspring has been more seriously affected, apart from the mother-daughter duos described by Knight *et al* (1995), where both were unaffected.

2.7.8 Deletions in the terminal 3p region

Causative deletions occur in the 3p26-p25 region. Most deletions are terminal and relatively large, 4.5-12 Mb in size, while a minority of the deletions are interstitial with a size ranging from 1.6 Mb to 10 Mb and rare cases of small, ~1Mb terminal

deletions have also been described (Cargile *et al*. 2002, Malmgren *et al*. 2007, Shuib *et al*. 2009, Gunnarsson and Foyn Bruun 2010, Cuoco *et al*. 2011). Proximal breakpoints are not clustered in any meaningful way, indicating the absence of a breakpoint hotspot. The largest deletions contain several dozens of genes, whereas the smallest interstitial deletion comprised of only 24 known genes. No clear maternal or paternal inheritance has been shown.

The variable size of causal deletions has prompted several attempts to define the causative genes and minimal critical region required for the symptoms, which are currently unknown. Based on their suggested functional roles, a number of genes residing in the distal 3p area have been suggested as contributing to the mental retardation (MR) in del3p patients: *CHL1* (also known as *CALL*), *CNTN4*, *CNTN6, CRBN* and *SRGAP3/MEGAP*. A number of the patients suffer from sensorineural hearing loss, and by fine mapping of the deletion breakpoints and comparing the data with auditory test results McCullough *et al*. (2007) suggested gene *ATB2B2* in 3p25.3 to be responsible for the hearing loss (McCullough *et al*. 2007). This was supported by Gunnarsson and Foyn Bruun (2011), whose patient had normal hearing and distal deletion compared to the patients of McCullough and their group. A congenital heart disease (CHD) is also a frequent finding in the 3p deletion patients, and the *CAV3*, *SLC6A1*, *HRH1* and *ATG7* genes have been suggested as candidates for CHD (Green *et al*. 2000, Malmgren *et al*. 2007, Gunnarsson and Foyn Bruun 2010).

3 THE AIMS OF THE STUDY

This study focuses on the development of molecular diagnostics tests for three rare inherited syndromes, the CHARGE, Sotos and 3p deletion syndromes. These three syndromes were chosen for the study for slightly different reasons: the resolution of the genetic background of CHARGE and Sotos syndromes had recently begun, which made them good, novel candidates for the test development, whereas for 3p deletion syndrome there was a longer-standing need to develop a test with good resolution. Prior to this study, molecular diagnostic tests for CHARGE and Sotos syndromes were not offered in Finland, and only non-specific chromosome analyses were available for 3p deletion. Of the aspects of the tests, analytical and clinical validity and clinical utility are discussed, whereas ELSI issues are beyond the scope of this study.

The specific aims of this study were:

I. to develop molecular diagnostic tests for the syndromes in this study

II. to clarify the mutational backgrounds of the groups of patients and the relevance of the detected variants to the phenotype

III. to suggest a molecular diagnostic protocol for the syndromes

4 SUBJECTS AND METHODS

4.1 Subjects and their clinical features

4.1.1 Ethical issues (I, II, III, IV)

The ethically difficult issue in this study was that most of the patients were minors and/or had developmental delay so that they could not consent to giving samples. However, the clinical examinations performed and the blood samples taken for the molecular or chromosomal analyses were performed for the clinical diagnostic purposes of those patients and their families. In each family, the patient, or usually the parents, had consented to these investigations. In addition, an informed consent to publish clinical details and photographs was obtained from the patients and/or parents in CHARGE and 3p deletion cases. In the case of Sotos syndrome, informed consents and clinical data were obtained from the patients in the first group; the second group consisted of clinical patients who had consented to be tested. The study for CHARGE and Sotos syndromes was approved by the ethics committee of the Hospital District of Southwest Finland, whereas the basis for the study in the case of 3p families was clarification of the diagnosis, and the approval of ethics committee was therefore not needed.

In all cases, the results of the tests were returned to the referring clinicians, who were responsible for organizing the genetic counseling for the patients.

4.1.2 CHARGE patients (I, II)

Samples from 18 patients were collected in Clinical Genetic Units of Turku University Central Hospital and Helsinki University Central Hospital. In addition, samples of 56 German patients were analyzed in the German collaborator laboratory. There were no pre-selection criteria other than the clinical evaluation according to the expertise of the referring clinicians.

4.1.3 A CHARGE family (II)

The CHARGE family consisted of an affected father and his son. In addition, there was a second pregnancy in the family which was terminated at the 23rd gestational week because of cardiac anomalies of the male fetus detected in the ultrasound scan. The father was subjectively healthy, and was not examined after the birth of his affected son. Although his findings did not fulfill the criteria of the clinical diagnosis of the CHARGE syndrome the recurrence of the disorder in this family made it possible that he was a carrier, as he had some features suggesting mild CHARGE-like

syndrome while the mother had no such features (Table 2). In addition, the father's brother also had very mild symptoms possibly suggesting CHARGE-syndrome. The parents of the father were not available for clinical investigation. The clinical findings in this family are described in detail in II.

Table 2. Clinical findings of the CHARGE family members.

	Patient			
Feature	**Proband**	**Fetus**	**Father**	**Father's brother**
Congenital heart disease				
Atricular/ventricular septal defect	+	+	-	-
Ductus arteriosus	+	+	-	-
Rightsided aorta	-	+	-	-
Pulmonary trunk stenosis	-	+	-	-
Pulmonary vascular atresia	-	+	-	-
Dysmorphic features				-
Abnormal skull shape	+	-	-	-
Square face	+	+	-	-
Hypertelorism	+	-	-	-
Broad nasal bridge	+	-	-	-
Small chin	+	-	-	-
Dysplastic ears	+	+	+	+
Upslanting palpebral fissures	-	-	+	-
Cleft lip/palate	+	+	-	-
Ocular findings				-
Bilateral retinal coloboma	+	-	-	-
Retinal lesion	-	-	+	-
Unilateral microphthalmia	-	+	-	-
Dysplastic papilla	+	-	-	-
Other features				-
Absense of olfactory bulbs	+	+	N/A	-
Hypoplastic optic nerves	N/A	+	-	-
Conductive hearing loss	N/A	-	+	+
Hypolobated lungs	-	+	-	-
Stenosis of right bronchus	-	+	-	-
Symmetric liver	-	+	-	-
Extrahepatic duct	+	-	-	-
Intestinal malrotation	-	+	-	-
Unilobular thymus	-	+	-	-
Hypoplastic scrotum	-	+	-	-
Micropenis	+	-	-	-
Feeding difficulties	+	N/A	-	-
Hypoplastic nails	+	N/A	-	-
Fetal asphyxia	+	N/A	-	-

N/A= not analysed

4.1.4 Sotos patients (IV)

The Sotos patient samples were analyzed in two groups. The first group consisted of 13 samples from patients clinically diagnosed with classical Sotos syndrome by experienced clinical geneticists. The second group consisted of 174 samples referred for diagnostic testing to our laboratory (DNA diagnostic laboratory of the University of Turku, Department of Medical Biochemistry and Genetics, UTUlab) as we had notified the genetics and pediatrics community of our intention to develop testing suitable for routine clinical testing of Sotos syndrome. This group of patients was more heterogeneous than the first group; the symptoms were more variable and were not fully consistent with the clinical picture of the Sotos syndrome.

4.1.5 3p deletion syndrome patients (III)

Two families were involved in this study. In family A, the mother had minor signs suggestive of the syndrome, but these were noticed only after the proband was diagnosed with the syndrome. The symptoms of the proband were consistent with 3p deletion syndrome. His sister had selective mutism and facial features resembling those of the brother, but otherwise she was normal (Table 3). In family B, the mother was considered healthy and normal, and the symptoms of the proband were atypical of the syndrome (Table 4). The maternal grandfather has symptoms similar to the proband. The clinical findings of these families are presented in detail in III.

Table 3. Clinical findings of 3p deletion family A.

	Family A		
	Proband	**Mother**	**Sister**
short upward turned nose	+	+	+
long philtrum	+	+	+
highly arched palate	+	-	-
mildly dysmorphic auricles	+	-	-
preauricular tags	-	+	-
hypotonia	+	-	-
laryngomalacia	+	-	-
speech delay	+	-	-
developmental delay	+	-	-
sandal caps	+	-	-
plantar furrows	+	-	-
clubbed fingers	+	-	-
selective mutism	-	-	+

Table 4. Clinical findings of 3p deletion family B.

	Family B		
	Proband	**Mother**	**Maternal grandfather**
microcephaly	+	-	+
hypotelorism	+	-	-
low forehead	+	+	-
long, thin and pointed nose	+	+	-
thin and fair hair	+	-	+
motor delay	+	-	-
learning disability	+	-	-
temper tantrums	+	-	-
hyperpigmentation	+	-	-
tapering fingers	+	-	-
nocturnal enuresis	+	-	-
hypothyroidism	+	-	-

4.2 Methods

4.2.1 DNA extraction (I, II, III, IV)

For the *CHD7* mutation analysis, DNA was extracted from peripheral blood lymphocytes, fibroblast or liver cells, and for 3p deletion and *NSD1* analyses DNA was extracted from peripheral blood lymphocytes. DNA was extracted from cell cultures with standard proteinase K/phenol-chloroform extraction and from blood cells with the Nucleon BACC3 Genomic DNA Extraction kit (GE Healthcare, Buckingham-shire, UK). In the CHARGE family, a paternal buccal cell sample was collected and DNA extracted with the BuccalAmp DNA Extraction Kit (Epicentre Biotechnologies, Madison, WI, United States).

4.2.2 PCR and sequencing (I, II, III, IV)

In the CHARGE syndrome analysis, primers for the *CHD7* gene were designed with the Primer3 software. For analyzing the *NSD1* gene in Sotos syndrome, primers according to Douglas *et al*. (2003) were used with minor modifications. In 3p deletion syndrome, for the *CNTN4* and *CRBN* genes, primers were designed with the Primer 3 software (Rozen and Skaletsky 2000). For the *CHL1* gene analysis, primers according to Frints *et al*. (2003) were used, except for exons 3, 7, 8, 13, 14, 17, 18 and 28, for which new primers were designed with the Primer3 program.

All PCR reactions were performed in a standard 25 µl reaction volume with 100 ng of template DNA, and the PCR products were purified with an ExoSAP enzyme

mix (USB Corporation, Cleveland, Ohio, USA). All exons were sequenced in both directions using the BigDye Terminator v3.1 cycle sequencing kit and ABI Prism 3100/3130 DNA sequencers (Applied Biosystems, Foster City, CA, USA) according to manufacturer's protocol.

4.2.3 Cell culturing and mRNA analysis (I, IV)

In CHARGE and Sotos syndromes, in order to study the pathogenic role of the splice site mutations at the RNA level, patient and control-derived PHA-stimulated and EBV-induced lymphocytes were cultured in RPMI medium, and the total RNA was extracted using an RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Primers for the cDNA amplification were designed with the Primer3 software. Reverse transcription and cDNA amplification were performed in a one-tube reaction with the Titan One Tube RT-PCR kit (Roche Diagnostics, Mannhein, Germany) using 100 ng of the total RNA. The amplified cDNA fragments were sequenced with the ABI 3100 capillary DNA sequencer.

4.2.4 Restriction fragment length analysis (RFLP, I, II, IV)

In the CHARGE family, the detected familial mutation was directly analyzed by means of the restriction enzyme analysis from the samples of the family members and the fetus. Exon 21 of the *CHD7* gene was amplified with PCR and digested using the Tsp45I restriction enzyme (New England Biolabs, MA, USA) in a 20 µl reaction volume with 1 µg of amplified DNA. Fragments were separated on a 2.5% agarose gel.

To study possible somatic mosaicism in the father of the CHARGE family, the mutation was analyzed from the buccal cell and lymphocyte DNA samples with the restriction enzyme analysis as described above. To estimate the relative amount of mutation in the lymphocyte and buccal cell samples, the signal intensities of the restriction fragments were evaluated visually from the agarose gel.

RFLP was also used for studying the *de novo* nature of the mutations. In the test set-up phase, novel mutations were screened in 200 ethnically-matched control chromosomes with RFLP, and parental samples were analyzed when available.

4.2.5 In silico analyses *(I, IV)*

The functional effects of the splice site mutations detected in the CHARGE and Sotos syndrome patients were analyzed with the Splice Site Predictor program by Berkeley Drosophila Genome Project (BDGP, http://www.fruitfly.org). Pathogenicity prediction of the detected missense mutations was performed using the

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2) and Pmut (http://mmb.pcb.ub.es/PMut/) software. Mutation nomenclature was checked with the Mutalyzer sequence variant nomenclature checker v.1.0.1 (https://mutalyzer.nl/) according to the Human Genome Variation Society (HGVS) guidelines.

4.2.6 Locked Nucleic Acid PCR (LNA, II)

In the CHARGE family, somatic mosaicism was analyzed from the lymphocyte samples of the grandparents and paternal uncle with allele-specific PCR. PCR was performed with Locked Nucleic Acid primers (Sigma Aldrich, MO, USA) specific to either the normal C or the mutated T allele in a 25 μl reaction volume with 100 ng of template DNA. Fragments were separated on a 1.5% agarose gel.

4.2.7 Denaturing High-Performance Liquid Chromatography (DHPLC, IV)

To pre-screen the mutations in the *NSD1* gene, DHPLC analysis was applied with the Transgenomics Wave® (Transgenomic, Omaha, NE, USA) machinery in the Centre for Biotechnology, Turku. After the amplification of the PCR fragments (see section 4.2.2), a test run was performed in 50°C with 10 μl of the samples to check the quality and amount of the amplified products. Rehybridization was performed by heating the samples to 95°C and decreasing the temperature by 0.5°C/cycle to a final temperature of 20°C. Analysis of PCR products was performed with temperatures specific to each fragment, and fragments displaying variable elution profiles were sequenced (see section 4.2.2) in order to clarify the nature of the variant.

4.2.8 Karyotype analysis and Fluorescence **In Situ** *Hybridization (FISH, II, III)*

In the CHARGE family, the proband's karyotype was analyzed with the Giemsa staining method (450-band resolution) and subtelomere FISH, using the ToTelVysion assay (Abbott Molecular, Inc., Downers Grove, IL) according to the manufacturer's instructions to exclude large chromosomal aberrations. In the 3p deletion families, the primary deletion analysis was performed on metaphase (400 band resolution) and prometaphase chromosomes (800 band resolution) using the Giemsa staining method on cultured phytohemagglutinin (PHA)-stimulated peripheral blood lymphocytes. A subsequent subtelomere FISH analysis was performed using the ToTelVysion assay according to the manufacturer's instructions.

4.2.9 Multiplex Ligation-dependent Probe Amplification (MLPA, I, III, IV)

MLPA was used in CHARGE and Sotos syndromes to detect new copy number variants, and in 3p deletion syndrome to confirm previously detected deletions. In CHARGE syndrome, the P201 SALSA kit was used, containing probes for the *CHD7* gene. In Sotos syndrome, the P026B SALSA kit was used, which has probes specific to *NSD1* (all kits MRC Holland, Amsterdam, The Netherlands). In the 3p deletion families, the P208 telomere-6 SALSA kit containing probes for a 4.4 Mb area in the 3p telomere was used. The reactions were performed in a 0.25-fold volume of the reagents and with a 3-hour hybridization modified from the manufacturer's protocol with 100 ng of genomic DNA. The MLPA PCR products were separated using an ABI 310 capillary DNA sequencer (Applied Biosystems, Foster City, CA, USA), and the chromatograms were analyzed with the GeneMarker software (SoftGenetics, State College, Pennsylvania, USA). For each sample, relative probe signals were defined by dividing each peak area measured by the area of the combined peak areas in that sample, and the relative values were compared to those of control samples.

4.2.10 SNP array (III)

In 3p deletion family A, the breakpoints and sizes of the deletions were further characterized using genome-wide Single Nucleotide Polymorphism (SNP) array analysis using the Affymetrix NspI 250k SNP array platform, which contains 25-mer oligonucleotides representing a total of 262,264 SNPs. The array experiments were performed according to protocols provided by the manufacturer (Affymetrix, Inc., Santa Clara, CA, USA). Copy-number estimates were determined using the Copy Number Analyzer for the Affymetrix GeneChip mapping (CNAG) software package (version 2.0). The normalized ratios were analyzed for loss and gain of regions by a standard Hidden Markov Model (HMM), which was optimized in order to maximize the detection of the known validated copy-number aberrations, while minimizing the false-positive rate, as described before (de Vries *et al*. 2005). The set HMM parameters were 0 for N=2, -0.38 for N=1, 0.3 for N=3, and 0.55 for N=4. An average of five or more consecutive SNPs showing a single copy-number loss (N=1) and an average of seven or more consecutive SNPs showing a single copy-number gain (N=3) provided 95% confidence of representing a true copy-number variation (Hehir-Kwa *et al*. 2007).

4.2.11 Real-time quantitative PCR (I)

In our laboratory, real-time quantitative PCR was used to confirm the results of the MLPA analysis in CHARGE syndrome. Primers were designed with the Beacon Designer software (PREMIER Biosoft International, Palo Alto, California, USA) for

exon 31 in gene *CHD7* and for exon 8 in the control gene, *GAPDH,* to amplify 159-bp and 162-bp fragments, respectively. Reactions were performed in triplicate using 12.5 ng of DNA in a 25 µl reaction with one-fold iQ SYBRGreen supermix and the iCycler thermal cycler (Bio-Rad, Hercules, California, USA). Four-fold dilutions ranging from 25 ng to 0.0976 ng were used to prepare the standard curve, and the relative difference in the copy number ratio of *CHD7* and *GAPDH* between the control and patient samples was analyzed with the Pfaffl method (Pfaffl 2001).

5 RESULTS AND DISCUSSION

5.1 Development of the molecular diagnostic tests (I, III, IV)

The initiative for the molecular diagnostic test development for these syndromes originally came from clinicians working with the patients. The development process was begun in the research laboratory, but the final aim of the test development was to translate the tests into diagnostic service testing. However, the approach of the research laboratory and diagnostic laboratory to mutation detection is fundamentally different. Research laboratories have the flexibility to test and use new methods without existing knowledge on sensitivity and specificity of the methods, whereas diagnostic laboratories need to evaluate their methods thoroughly prior to use within the frames of existing quality standards. This difference was to be considered during the test development.

From a molecular diagnostic point of view CHARGE syndrome and Sotos syndrome share similar features: both syndromes are caused by mutations in large gene, all types of mutations can be expected to be detected, and, in both, a negative result from a diagnostic test does not exclude the presence of the syndrome. Compared to the previous, 3p deletion syndrome is a different entity and its molecular diagnostics is more straightforward due to the fact that it is caused by only different-sized deletions which are expected to be found in the 3p area. Interpretation of the results, however, is complicated with the possibility of the presence of rare benign CNVs in the same chromosomal area.

Optimally, there should be a robust and relatively cheap method for mutation detection in all the diseases/situations where molecular diagnostic testing is considered clinically useful. In practice, in the diagnostics of syndromes with large genes and high number of unique/novel mutations, such as the CHARGE and Sotos syndromes, the choice is often whether to use direct sequencing (higher costs, higher sensitivity) or pre-screening methods (lower costs but reduced sensitivity). In addition, the presence of structural aberrations may require the use of additional methods. In syndromes caused solely by structural aberrations, the choice has usually been between microarrays and FISH-based tests. In recent years, MLPA has also become a common method in detecting structural aberrations.

In the pilot-testing phase of Sotos syndrome, to compare performance and costs, intragenic mutations were sought in parallel with direct sequencing and pre-screening method DHPLC (Figure 5). DHPLC has been effectively used in the screening of *NSD1* mutations and mutations in several other genes (Kosaki, Udaka and Okuyama

2005, Melchior, Schwartz and Duno 2005). Based on the results, direct sequencing was chosen as the method to be used in Sotos syndrome diagnostics. The major disadvantages of DHPLC were the requirement of a large amount of PCR product for the analysis and several melting temperatures for each fragment, and in consequence, prolongation of the analysis time. The melting temperature is an important determinant of the sensitivity of the DHPLC, because insufficient melting of the fragments will leave part of the variant undetected. Another disadvantage was the inherent problem of DHPLC, the requirement of subsequent sequencing of each fragment showing aberrant elution profiles to clarify the nature of the detected variant. *NSD1* has 23 polymorphisms scattered in 9 exons, and pathological mutations could have been masked by the elution profile differences caused by polymorphism in these exons. Consequently, these exons had to be sequenced in any case. Compared to sequencing, the amount of laboratory work per sample was excessive in DHPLC.

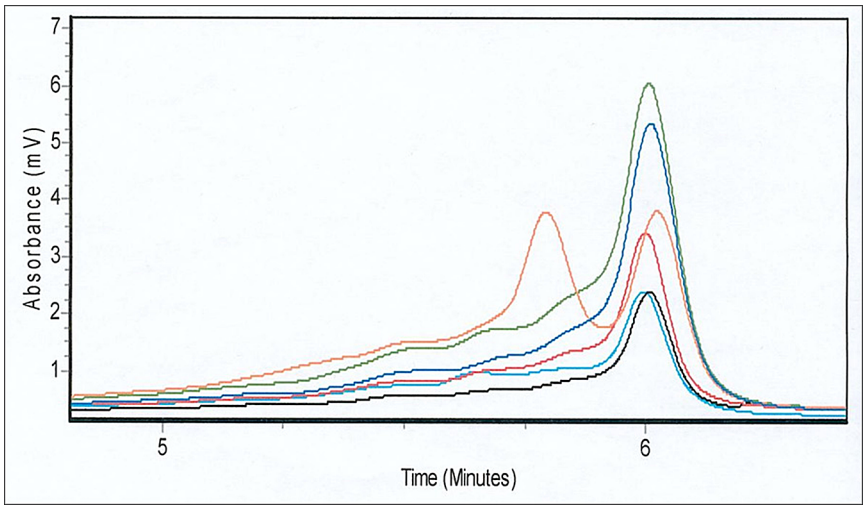

Figure 5. The elution profile of the DHPLC analysis of exon 18 of the *NSD1* gene. The sample with a mutation is marked with orange. The patient has a heterozygous C→T transition resulting in an amino acid change in the functionally important SET domain. In the figure, the sample with a mutation is illustrated by two peaks, whereas the samples without a mutation are displayed as single peaks.

In addition, DHPLC was not routinely used in the diagnostics laboratory, and there were no assessments of the specificity and sensitivity of the method in the diagnostic setting. As in rare diseases the number of positive samples is low in the pilot phase, determination of the analytical validity and sensitivity of the used method may be difficult, and the use of previously validated methods may be advisable (Mattocks *et al*. 2010). Similarly, the setting up and maintenance of the DHPLC ma-

chinery was not deemed expedient and suitable for the methodological strategy of the diagnostics laboratory. Even though costs of direct sequencing of the large *NSD1* gene exceed the costs of pre-screening, by choosing sequencing as the method to be used in diagnostics, a compromise was made between the costs, the work load and the ability of the laboratory to set up a new test.

Because a small percentage of the *NSD1* mutations are intragenic deletions or microdeletions, another method was needed to detect these variants, as sequencing is not capable of detecting large rearrangements. For this purpose, the alternative methods considered were MLPA and array CGH. A fundamental difference between these methods is that the MLPA test for Sotos covers only the *NSD1* gene, whereas array CGH covers the whole genome. Arrays are often used in the diagnostics of patients with developmental delay and malformations as a primary test, especially when no particular syndrome is suspected. Because the tests set up in this study were specifically aimed at Sotos syndrome diagnostics, MLPA was chosen. In addition, MLPA is capable of detecting exonic deletions and duplications. Other reasons were the simple set up and robustness of the method, allowing the adaptation of the method also for other diseases tested in the diagnostic laboratory. Arrays are not a part of the methods used in the diagnostic laboratory, which also supported the choice of MLPA.

The development of the CHARGE syndrome test followed that of Sotos syndrome testing, with the exception that, based on the experiences generated from the *NSD1* testing, the sequencing analysis was directly applied. The *CHD7* gene is even larger than the *NSD1* gene, and similar difficulties were expected to arise with the pre-screening methods. Although intragenic rearrangements and deletions covering the *CHD7* gene are rare (Bergman *et al*. 2008), the MLPA method is also applied to enhance the sensitivity of the testing.

After the development phase of the tests for CHARGE and Sotos syndromes, both of these tests were translated to diagnostic service testing for the Laboratory of Medical Genetics (UTUlab).

For 3p deletion syndrome, the initial aim was to develop a FISH-based test with a series of probes covering 15 Mb from the distal 3p arm to resolve the size of the deletions of the patients. However, at the time the development was about to begin, array CGH was becoming common in the microdeletion syndrome diagnostics, and it was therefore decided to use the array. The MLPA method was also considered for the molecular diagnostics of 3p deletion syndrome, but the probes in the available telomere kits did not cover the entire area in the 3p arm where deletions have been described. For this reason, the array-based analysis was chosen.

5.2 Molecular genetic findings

Table 5 summarizes the molecular genetic findings detected in the patients in this study. Details of the findings are discussed in the following sections.

Table 5. Characteristics of small scale mutations and structural aberrations detected in this study.

CHARGE syndrome [a] **30/74 (40.5%)**	Small scale mutations	Missense Nonsense Frameshift Splice site UV	4/30 (13%) 9/30 (30%) 11/30 (37%) 4/30 (13%) 2/30 (7%)
	Deletions	Whole gene Intragenic	No detected deletions
Sotos syndrome [a] **59/174 (34%)**	Small scale mutations	Missense Nonsense Frameshift Splice site UV	17/59 (29%) 15/59 (25%) 19/59 (32%) 2/59 (3%) 3/59 (5%)
	Deletions	Whole gene Intragenic	4/59 (7%) 2/59 (3%)
3p deletion syndrome	Family A deletion	8.99 Mb, 3p26.3-p25.3 [b]Proband, mother, sister	
	Family B deletion	1.1 Mb, 3p26.3 [b]Proband, mother	

[a] Number of detected mutations versus number of tested samples. [b] Affected family members. UV= unclassified variant

5.2.1 Mutations in the CHD7 gene (I)

In the study of CHARGE syndrome, samples of 74 patients were analyzed and a total of 30 (40.5%) intragenic mutations were identified (Table 5). 18 patients were analyzed in the University of Turku, at the former Department of Medical Genetics and 56 patients in Germany, Praxis für Humangenetic, Freiburg. Of the detected 30 mutations, 21 were novel and 9 have been published previously. The majority (20/30, 67%) were truncating mutations, either nonsense (9) or frameshift (11) mutations, causing a premature stop codon. In addition to these, six missense variants (6/30, 20%) and four splice site mutations (4/30, 13%) were detected, correlating well with the previously reported frequencies (Zentner *et al*. 2010).

The truncating mutations were scattered throughout the coding area. The recent suggestion of mutational hotspots in exons 2, 3, 8, 12, 26, 30, 31 and 36 was not supported (Bergman *et al*. 2011, Bilan *et al*. 2012), as 80% of the mutations in our series occurred outside these exons. One of the frameshift mutations, c.8962dupG (p.D2988fsX1), was of interested, as it was *de novo* and occurred only 33 bp upstream of the 3´ end of the coding sequence in a patient fulfilling the clinical criteria for CHARGE syndrome. As known, mutations introducing premature translation termination codons occurring >50 basepairs upstream of the 3' end are likely to subject the nascent mRNA to nonsense-mediated mRNA decay (NMD). In the case of the CHARGE patient, the minimally shortened mRNA possibly escapes the NMD and may lead to residual or altered function of the protein. Alternatively, because only one aberrant amino acid is introduced into the truncated CHD7 protein, this mutation may indicate an as yet unknown functional domain in the most 3' region of CHD7. The nonsense mutation p.R2284X, previously published by Felix *et al*. (2006) was identified in two unrelated patients in this study. Parental samples were available for analysis in 12 cases with truncating mutations, and in all of these the mutations were confirmed to have occurred *de novo*.

Four splice site mutations were detected, of which three were within the consensus splice sites and one 17 bp upstream of the 5´ end of exon 26 (c.5405-17G→A). An mRNA analysis of this mutation revealed the formation of a cryptic splice site and, as a result, an insertion of five codons into the cDNA in-frame. The likely effect of this mutation on the protein is the disruption of the three-dimensional structure and therefore a dysfunctional protein. For splice site mutations, the parental analysis confirmed the *de novo* occurrence in three cases out of the four; in one case (c.5895-2A→G), the parental samples were not available.

Six missense variants were found scattered throughout the entire length of the gene. Two variants in the functional domains, p.I1028V in the SNF2-related helicase/ATPase domain and p.Q1395H in the HELICc domain, are highly likely to be pathogenic mutations due to their destroying effect on the function on these domains. These mutations were also confirmed to have occurred *de novo*, and, in addition, the p.I1028V mutation had previously been published in two studies (Vissers *et al*. 2004, Jongmans *et al*. 2006). In the case of three other missense variants (p.W983G, p.D1596G and p.R2319C), parental samples were not available, but these three mutations were not detected in the 200 control chromosomes. In addition, p.D1596G and p.R2319C have been previously published (Felix *et al*. 2006, Jongmans *et al*. 2006, Bilan *et al*. 2012), supporting their pathogenic role. In the case of p.C1643W, the maternal sample was available but showed no mutation. Thus, in the case of p.W983G and p.C1643W pathogenicity remained elusive and they were classified as variants of unknown significance.

In addition to the putative pathogenic mutations, 10 new polymorphisms in the *CHD7* gene were also detected. Of these, five missense changes were found either in patients with a truncating mutation or in one unaffected parent.

Although *CHD7* was originally discovered in a CHARGE patient carrying a translocation disrupting the gene (Vissers *et al*. 2004), further studies have shown that structural aberrations are a minority of causative mutations (Bergman *et al*. 2008, Wincent and Schoumans 2009). In this study, deletions encompassing the *CHD7* gene were not detected, confirming the low presence of structural aberrations containing the *CHD7* gene.

5.2.1.1 Results of the CHARGE family (II)

Familial cases in the CHARGE syndrome are rare. In all cases with confirmed parent-to-child transmission of mutations, there has been a considerable intrafamilial heterogeneity with the affected parent presenting with mild features, resembling the situation in the family in this study (Lalani *et al*. 2006, Delahaye *et al*. 2007, Jongmans *et al*. 2008). In a few familial cases, the unaffected parents have been carriers of somatic or gonadal mosaicism (Jongmans *et al*. 2008, Pauli *et al*. 2009). Previously, mental retardation was considered explanatory for the rare parent-to-child transmission, but it is now known that intellectual skills range from normal to severe mental retardation, suggesting that other factors prevail. The more likely explanation is the reduced fertility due to the gonadotropin deficiencies (hypogonadotropic hypogonadism).

In the proband of the family in this study, the karyotype, subtelomere FISH and MLPA analyses were normal, but in the sequence analysis a nonsense mutation, p.Q1599X (c.4795C→T), was detected in exon 21 of the *CHD7* gene. In the RFLP analysis, the same mutation was detected in the samples of the fetus and the father. In addition to peripheral blood lymphocytes, the mutation was also detected in the buccal cell sample of the father; thus, somatic mosaicism could not be detected based on these two tissues. In both tissues, the restriction fragment signal intensities were equal, indicating the same relative amount of mutated DNA in both of the paternal samples.

Although the mutation was detected in both the buccal cells and peripheral blood cells of the father, the restriction fragment signal intensities between the samples of the father (peripheral blood) and the fetus and the proband were not equal. In recurrent analyses, the intensities of normal fragments in the samples of the fetus and proband were weaker than in the sample of the father (Figure 6). This may indicate that the father had a larger amount of normal allele than the fetus and the proband,

explaining his milder symptoms. Other explanations for this discrepancy could be that in the PCR, the alleles multiply unequally in different samples. The samples of the fetus and proband originate from the liver cells and fibroblasts, whereas the sample from the father is from blood lymphocytes. It is also possible that the digestion is interfered with by different types of proteins in different samples. However, the visual analysis of the RFLP fragments is only semiquantitative, and a qualitative analysis should be applied to rule out possible mosaicism in the father.

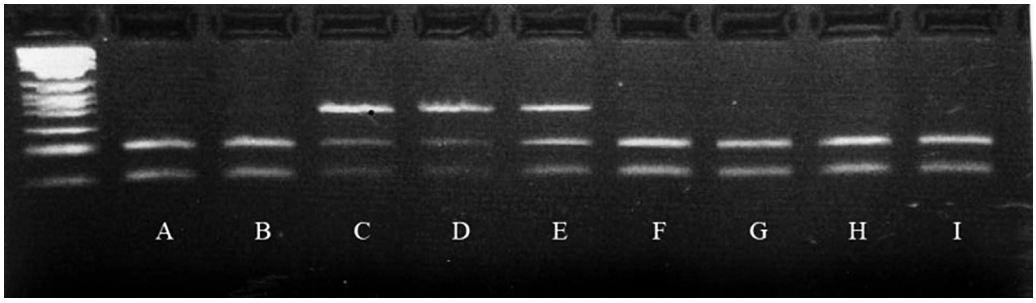

Figure 6. Result of a restriction enzyme analysis of the samples from the CHARGE family. A: paternal grandmother, B: paternal grandfather, C: fetus, D: proband, E: father, F: mother, G: father's brother, H and I: controls. Where there is no mutation, the normal alleles have been digested into two fragments. In the case of the father, the fetus and the proband, only the normal allele has been digested when the mutated allele has remained undigested.

The samples of the paternal grandparents were available for analysis, and the presence of the mutation in their samples was analyzed with RFLP and allele-specific PCR. The RFLP analysis indicated absence of the mutation, and the allele-specific PCR confirmed the result, when only the normal C allele but not the mutated T allele was amplified in their samples (data not shown). This indicated that they were not mutation carriers and that the mutation had occurred *de novo* in the father. The same mutation was neither detected in the restriction enzyme or allele-specific PCR analysis of the lymphocyte sample of the father's brother. The sequencing and MLPA analysis of the *CHD7* gene did not reveal other mutations in his *CHD7* gene.

The father had only mild symptoms (characteristic ears, hearing loss and facial asymmetry) fulfilling two major features according to Blake and also one minor according to Verloes. The proband fulfilled two of the major (coloboma and characteristic ears, cranial nerve anomalies were not studied) and four of the minor criteria (cleft palate, congenital heart disease, dysmorphic features and genital hypoplasia) according to Blake, and the fetus presented with similar features. According to Verloes's criteria, the proband had one major feature (coloboma) and two minor (abnormal external ear and congenital heart disease). However, the

cleft palate can sometimes be replaced by choanal atresia, and mental retardation could be replaced by CNS abnormalities in the fetus and the proband. Using strict clinical criteria, the father would have not been diagnosed with CHARGE syndrome.

There are two possible explanations for the clinical findings in this family. First, the father's brother had symptoms resembling those of the father, and because he did not have the *CHD7* mutation, it can be assumed that the hearing loss and external ear dysmorphism are not related to the *CHD7* mutation. In this case, the father could be interpreted as a symptomless CHARGE carrier. The other possibility is that the father and his brother have ear malformations and hearing loss because of two independent reasons: in the father's case because of the p.Q1599X mutation, whereas in the father's brother, the reason remains unexplained. Thus, the father would be an extremely mildly manifesting carrier. Supporting the latter explanation, Lalani *et al*. (2006) reported a family where a mother, proband and the proband's sister presented with clinical features of CHARGE syndrome but only the proband carried a mutation. Possible explanations for this kind of situation could be mutations in the regulatory sequences of the gene which are not detected by *CHD7* sequencing, mutations in the genes that regulate the expression of the *CHD7* gene or mutation in another gene causing hearing loss and external ear dysmorphism. Regardless, our results show that a healthy or nearly healthy parent of a CHARGE child may carry a *CHD7* mutation.

5.2.2 Sotos syndrome mutations (IV, unpublished)

The pilot testing included 13 patients diagnosed with classical Sotos syndrome, and 12 of these patients were discovered to have a pathogenic mutation in the *NSD1* gene (92%). In one patient, no mutation or deletion was detected (8%). Of the second group of 161 patients referred to routine diagnostics and suspected or diagnosed of having Sotos syndrome, pathogenic mutations were detected in 49 patients (30%). In all, 59 pathogenic mutations and two possibly pathogenic variants were detected in 174 patients with a mutation rate of 34% (Table 5). Mutation detection rates in both groups are consistent with the previously published rates from the groups of selected clinical Sotos patients (Tatton-Brown *et al*. 2005, Saugier-Veber *et al*. 2007) and unselected groups referred to diagnostic laboratories (Melchior *et al*. 2005).

Six deletions (~10%) were detected using MLPA. However, structural aberrations are likely to be underrepresented in the group referred to routine diagnostics, because not all samples were analyzed with MLPA (76/174). In the test set-up phase, all samples with negative sequencing results were analyzed with

MLPA, but in the routine laboratory MLPA was recommended after a negative result and performed if requested by the referring clinician. Four of the deletions discovered were microdeletions covering the entire gene and two were intragenic deletions of different sizes. One of the intragenic deletions consisted of only a single exon, 14. However, the single nucleotide variants in the probe binding sites can affect the hybridization of the probes, producing a result similar to an exonic deletion. In this case, absence of a variant in the probe binding site was confirmed with sequencing. The other intragenic deletion covered exons 3-23. Fifty-five of the mutations were intragenic mutations and of these, 34 (62%) were truncating mutations causing a premature stop codon. Two mutations affected splice sites: one of these occurred in the canonical splice site of exon 10, probably resulting in skipping of exon 10 from the mRNA. The other was a G→A transition 14 basepairs upstream of exon 21, and the mRNA analysis revealed the activation of a strong cryptic splice site and insertion of 13 basepairs into the mRNA. As a result, a premature stop codon was inserted 26 codons downstream of the mutation site.

Altogether, 19 missense changes were detected. They were considered pathogenic if they occurred in functional domains and changed a conserved amino acid. Seventeen changes conformed to these requirements and were putative pathogenic missense mutations. Two detected changes, p.Cys1661Arg and p.Tyr1870Cys, occurred outside the functional domains, and their pathogenicity was unclear. Computational pathogenicity prediction was used to evaluate their possible deleterious effects, and in this analysis the mutations were predicted to be pathogenic. The p.Tyr1870Cys change occurs between the SET and PWWP domains, and may cause structural changes in these two domains, and, in addition, the altered tyrosine residue is conserved between the orthologs. The patient's father was not carrying the same variant, but the mother's sample was not available for analysis, and thus the change remained unclassified. In p.Cys1661Arg, the cysteine residue is conserved between orthologs, and replacement of cysteines in an amino acid sequence usually causes disruption of disulfide bonds and structural changes. However, the parental samples were not available for studying the transmission of the mutation, and therefore the pathogenicity also remained elusive.

Thirty-one of the intragenic mutations were novel and 21 were recurrent mutations described in previous studies. Recurrent mutations were also identified in unrelated patients in this study: p.Arg604* was detected in three patients and p.Ser1128Phefs*2 in two patients. One familial mutation was detected, p.Ser299Tyrfs*21 (family previously described in Höglund *et al*. 2003).

5.2.3 Deletions in the 3p patients (III)

In the families participating in this study, the diagnosis of 3p deletion syndrome was originally set based on the findings detected in karyotype and FISH analyses. The clinical findings of one of the two families, however, were inconsistent with the 3p deletion syndrome and prompted further studies to explain both this inconsistency and the intrafamilial variability seen in the other family. Inheritance of the 3p deletions has been described in a few cases (Tazelaar *et al*. 1990, Knight *et al*. 1995, Sklower-Brooks *et al*. 2002, Shrimpton *et al*. 2006, Tagakishi *et al*. 2006, Cuoco *et al*. 2011).

In Family A, karyotype analysis revealed a hemizygous terminal deletion of the short arm of chromosome 3 in the proband, his mother and his sister (Table 5). The chromosomes of the father and maternal grandparents were available for analysis, and showed normal results. Based on a subsequent genome-wide SNP array analysis, the proximal breakpoint appeared to be in 3p25.3, located 2.5 kb upstream of the SRGAP3 gene. The final karyotype was 46,XY,del(3)(p25.3).arr snp 3p26.3p25.3 (SNP_A-1971271 -> SNP_A-1971700) x 1 mat. The deleted segment was 8.99 Mb in size containing 19 reference sequence genes (UCSC Feb 2009 Assembly), and the size of the deletion was the same in the mother, sister and the proband. In order to explain the more severe features of the proband, sequencing analysis of the *CRBN*, *CHL1*, and *CNTN4* genes was performed but did not reveal any pathogenic mutations. No other copy number variants (CNVs) with potential clinical relevance were detected in the genome-wide SNP array analysis.

In family A, despite the large, cytogenetically visible deletion of approximately 9 Mb, all the affected members had different and relatively mild symptoms. The proband had minor limb anomalies, dysmorphic features and a learning disability, but the mother had only mildly dysmorphic features. The proband's sister also had mildly dysmorphic features and she had selective mutism, which has not been described in association with deletions of 3p before. Her mutism could have been a psychological reaction to her brother's condition and therefore unrelated to 3p deletion syndrome, but this remains unresolved.

In 2005, Barber reviewed 130 families with directly transmitted unbalanced chromosome abnormalities (UBCAs). In 23% of the families, the affected proband and the phenotypically normal family member had the same UBCA, and 23% of these were deletions with the average size of 7.5 Mb. In 59% of the families, UBCAs consistently had mild consequences, and, of these, 49% were deletions with the average size of 10.9 Mb. As an explanation for such intrafamilial phenotypic variability, the role of modifier alleles, undetected differences at the molecular level, parental mosaicism and unmasking of a recessive allele were suggested. The existence of masked alleles in Family A was excluded, but the possibility of somatic mosaicism

in the mother remained unsolved, because only a peripheral blood sample was analyzed. The other above-mentioned explanations also remained unproven.

In family B, a routine karyotype analysis did not reveal any abnormalities in the sample of the proband. As the suspicion of a chromosomal etiology was strong, a subtelomere FISH analysis was performed and it revealed a hemizygous terminal deletion with the proximal breakpoint in 3p26 in the samples of the proband and his mother. SNP array analysis defined the size of the deletion as 1.1 Mb with the breakpoint in 3p26.3, located 4.9 kb upstream of the *CNTN6* gene. The final karyotype was 46,XY,arr snp 3p26.3(SNP_A-1971271 ->SNP_A4238977) x 1 mat. The size of the deletion was the same in both the mother and the proband, containing only one reference sequence gene, *CHL1*. The subsequent sequencing analysis of *CHL1* showed no recessive mutations. Samples of the maternal grandparents, father and sister were not available for analysis, as these family members wished not to be tested. Other copy number variants (CNVs) with potential clinical relevance were not detected in the genome-wide SNP array analysis.

In the proband of family B, the clinical features were inconsistent with 3p deletion syndrome, since his facial features were atypical of 3p deletion and, in addition, he had skin pigmentation, a clinical feature not reported in connection with 3p deletion syndrome before. Only his unspecific features, mild learning difficulties, microcephaly and growth retardation were consistent with the syndrome phenotype. A possible explanation for the symptoms seen in family B is that the proband had two separate clinical conditions. The 3p deletion could be responsible for the mild mental retardation while the other symptoms, including the skin pigmentation, could have a different etiology. In fact, the possibility of an X-chromosomal condition is supported by the history of the maternal grandfather having similar symptoms to the proband, excluding the learning difficulties. However, the grandfather has not been investigated clinically, and there was no sample available to examine for his possible 3p deletion.

Cuoco *et al*. (2011) reported a family with an 895 kb terminal deletion in the father and two affected brothers presenting with cafe-au-lait spots and other features aspecific for 3p deletion syndrome. Based on the clinical features, they concluded that such small terminal deletions are not causative for 3p deletion syndrome, supporting the evidence from family B in our study. In fact, a search of the 1000 Genomes database (http://browser.1000genomes.org/index.html) retrieved an 865 kb copy number variant, further supporting the nonpathogenic role of small terminal deletions (Figure 7). The 1.1 Mb deletion seen in Family B has not been reported in CNV databases, and the reason may be the low incidence of the entire 1.1Mb deletion in the general population. However, in correlation with the results of the Cuoco *et al*. (2011), it is likely a rare familial variant. A further genome-wide sequencing analysis in this family could resolve the etiology of the symptoms.

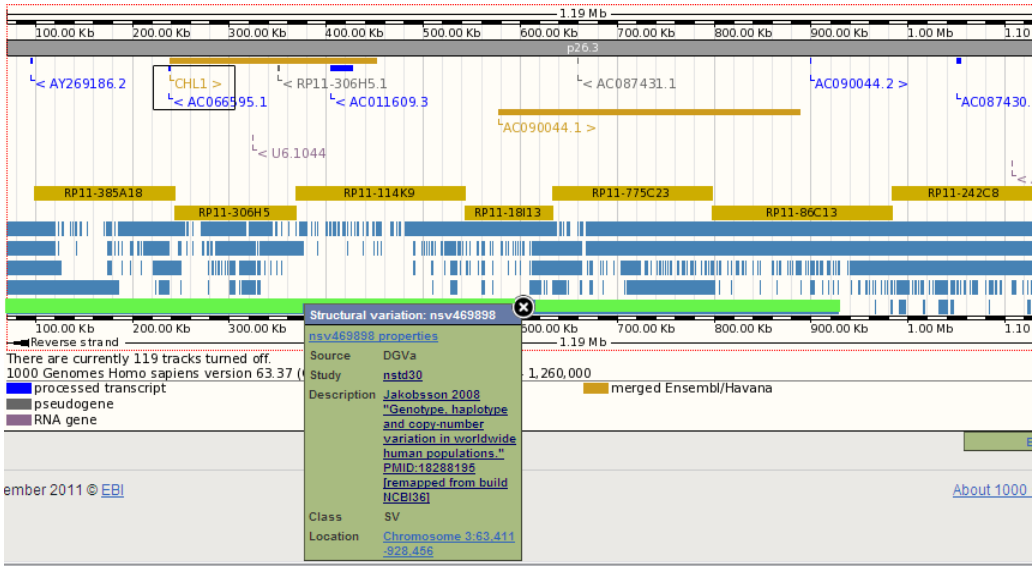

Figure 7. A figure showing the 865 kb copy number variant in the terminal 3p region (highlighted with green). Modified from the Ensembl/1000 Genomes browser (http://browser.1000genomes.org/Homo_sapiens/Location/Overview?db=core&r=3%3A70001-1260000, accessed 29.3.2012)

5.2.4 Genotype-phenotype correlation in 3p deletion patients

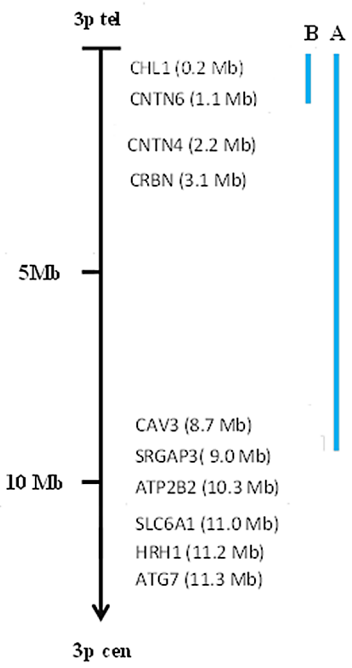

Figure 8. A schematic representation of the deletions found in families and relative locations of the genes suggested to contribute to the phenotype. A denotes Family A, and B denotes Family B.

Several genes have been suggested to contribute to the clinical picture of 3p deletion syndrome (Figure 8). For the mental retardation, the *CHL1*, *CNTN6*, *CRBN* and *SRGAP3* genes have been proposed as candidates. The protein products of the two most distal genes, *CHL1* and *CNTN6,* act as neural cell adhesion molecules, and *CHL1* has been connected to non-syndromic mental retardation (Frints *et al*. 2003). In family B, the deleted region contains only the *CHL1* gene, and the breakpoint is located 4.9 kb upstream of the *CNTN6,* possibly interfering with its regulatory region. Based on the evidence from this and previous studies, deletion of the *CHL1* and *CNTN6* genes cause only a mild mental deficit or no symptoms at all, and is not sufficient to cause the severe MR seen in patients (Dijkhuizen *et al*. 2006, Shrimpton *et al*. 2006, Cuoco *et al*. 2011). Thus, a loss of the more proximal genes in 3p seems to be required for the clinical features seen in 3p deletion syndrome.

Fernandez *et al*. (2004) described a patient with MR and autism spectrum disorder with a *de novo* balanced translocation disrupting the *CNTN4* gene located 3Mb from the telomere. The *CNTN4* gene belongs to a superfamily of immunoglobulins and function as a neural adhesion molecule guiding axon molecules in the developing neural system. The patient described by Dijkhuizen *et al*. (2006) had severe MR and complex rearrangement, the *CNTN4* gene was deleted in addition to the nearby *CRBN* gene, and they suggested that the features seen in 3p deletion syndrome are due to these two genes. *CRBN* is highly expressed in the brain, with a role in memory and learning. A family with nonsyndromic MR and with a *CRBN* nonsense mutation has been previously described (Higgins *et al*. 2004). Further supporting the role of *CNTN4* in the pathogenesis, it has been shown to connect with autism spectrum disorder (Roohi *et al*. 2009, Cottrell *et al*. 2011).

The findings of Endris *et al*. (2002) and Shuib *et al*. (2009) argue against the concept of a critical region containing *CHL1*, *CNTN4* and *CRBN* in the development of MR (Endris *et al*. 2002, Shuib *et al*. 2009). The patient described by Endris *et al*. had severe mental retardation and a translocation disrupting the *SRGAP3/MEGAP* gene located 9 Mb from the 3p telomere. By analyzing the deletions of 14 patients, Shuib *et al*. defined the candidate region for MR as containing only the same gene. *SRGAP3* guides neuronal and leukocyte migration through the Roundabout transmembrane receptors and is highly expressed in fetal and adult brain tissue (Endris *et al*. 2002). Gunnarsson and Foyn Bruun (2010) further supported these results, with their patient having profound mental retardation and small 1.6 Mb interstitial deletion retaining two copies of *CNTN4* but loss of the other copy of *SRGAP3*.

The results in this study are concordant with these observations. The proband in Family A had only mild learning difficulties, and the other clinical features observed were consistent with 3p deletion syndrome. The proximal breakpoint of the dele-

tion was located 2.5 kb upstream of *SRGAP3*, but the effect on the expression of the gene is difficult to assess. Presuming that *SRGAP3* is intact, the results from this family further confirm the role of *CHL1*, *CNTN4* and *CRBN* in causing mild symptoms and *SRGAP3* being responsible for more severe MR.

As an attempt to locate the candidate gene for congenital heart disease (CHD), Green *et al*. (2000) studied ten patients, of which five had CHD, and showed that these patients also had most proximal breakpoints. They defined the candidate critical region as being between 10.6 Mb and 11.5 Mb in p25.3, but were unable to show any candidate genes within this region. Malmgren *et al*. (2007) narrowed the critical region further down to 0.45 Mb, containing only the *SLC6A1*, *HRH1* and *ATG7* genes, of which none is an obvious candidate for CHD based on their functions. Gunnarsson and Foyn Bruun (2010) concluded that *CAV3* could be responsible for CHD, but it is located distal to the region defined by Green *et al*. and Malmgren *et al*. In this study, *CAV3* was also deleted in the proband of family A, but he had no CHD, thus arguing against *CAV3* being a candidate for CHD.

The *ATP2B2* gene located 10.3 Mb from the telomere has been suggested as a candidate for the hearing loss seen in some patients (McCullough *et al*. 2007, Gunnarsson and Foyn Bruun 2010). In concordance with this, the deletion in Family A was distal to the ATP2B2 gene, and the hearing of the proband was normal.

However, the role of other genes in the terminal 3p region has not been elucidated, and they may contribute to the clinical picture with a mechanism yet unknown.

5.3 Diagnostic considerations (I, II, III, IV)

5.3.1 CHARGE syndrome

In CHARGE syndrome, the diagnostics is most challenging. In previous reports, mutations have been detected in 58%-90% of the studied patients fulfilling the clinical criteria in preselected groups (Jongmans *et al*. 2006, Lalani *et al*. 2006). The group in our analysis was heterogeneous, consisting of patients with both clinically diagnosed and suspected CHARGE syndrome; in fact, in most cases there was insufficient clinical data available to evaluate whether patients fulfilled the current clinical criteria. The mutation detection percentage of 40.5% in our analysis reflects the situation in the diagnostic laboratory: when a clinician diagnoses the patient with or suspects them of having CHARGE syndrome, the mutation can be molecularly confirmed in less than 50% of the patients. Presumably, mutations are detected even less frequently if testing is used to exclude CHARGE syndrome in patients with only some features suggestive of CHARGE. A fairly similar mutation detection per-

centage (32%) in a diagnostic laboratory was recently reported by Bartels *et al*. (2010).

The strategy of molecular diagnostic testing of CHARGE syndrome depends on the clinical diagnosis: if the clinical diagnosis is strongly suggestive of CHARGE syndrome, the molecular analyses should begin with a sequencing analysis to detect point mutations in the *CHD7* gene because of the low prevalence of intragenic and microdeletions (Figure 9). If the result is negative, analyses should nevertheless be continued with MLPA to exclude structural aberrations. If the result is still negative, an array-based analysis can be applied to detect other chromosomal abnormalities, and differential diagnoses such as Kallmann syndrome and 22q11.2 deletion syndrome can be considered. If the patient has only minor or atypical symptoms, then the array analysis should be considered as a first line analysis and if returned negative, the mutation testing for *CHD7* should be applied. Recent evidence shows that patients with whole-gene deletions might have an atypical phenotype (Wincent and Schoumans 2009), but more studies are needed to confirm these observations.

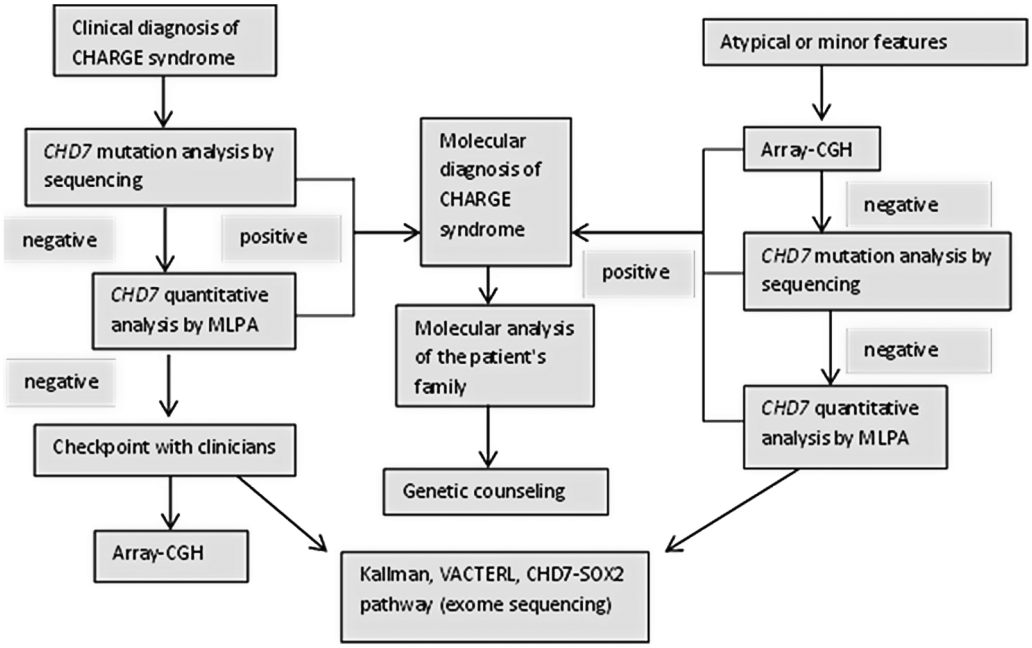

Figure 9. A proposal for the molecular diagnostic protocol for CHARGE syndrome. Modified from Bilan *et al*. (2012).

Inner ear malformations are a distinct feature of CHARGE syndrome and are present in 95% of the patients with a clinical diagnosis (Lalani *et al*. 2006, Zentner *et al*. 2010, Bergman *et al*. 2011). The prevalence of inner ear malformations in the group of this study is unknown, because clinical data were available for only a sub-

set of patients, and in those the inner ear malformation had not been specifically investigated.

A recent finding that *CHD7* acts as a cofactor for *SOX2* in a novel pathway groups together five syndromes with overlapping features (Figure 10). Mutations in the *SOX2*, *MYCN*, *JAG1* and *GLI3* genes cause anopthalmia, Feingold syndrome, Alagille syndrome and Pallister-Hall syndrome, respectively. In a mouse Chd7$^{+/-}$ model, *JAG1* has a dramatically reduced expression level, suggesting that low levels of *CHD7* down-regulate *JAG1* (Engelen *et al*. 2011). Heterozygosity of *JAG1* is known to cause semicircular canal malformations. In addition to these, it also causes cholestasis, cardiac defect, skeletal and ophthalmologic abnormalities and the characteristic facial features seen in Alagille syndrome (Li *et al*. 1997). Heterozygosity for *SOX2*, *CHD7* or *MYCN* are the only known causes of trachea-esophageal malformations, and SOX2, CHD7 and *GLI3* cause pituitary and genital abnormalities. Testing for mutations in the above-mentioned genes in the atypical cases of CHARGE syndrome may be considered.

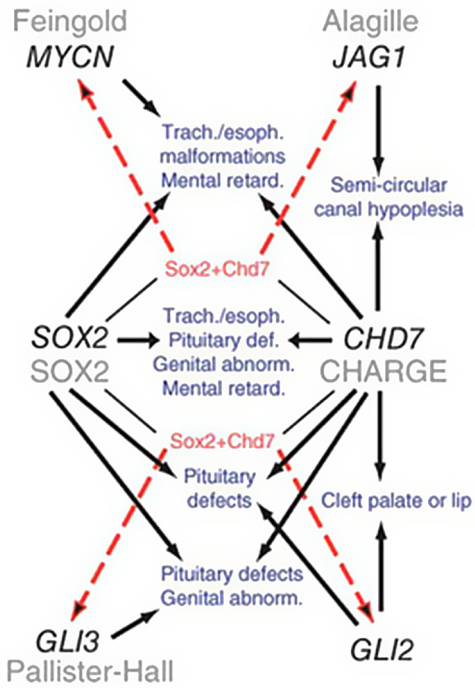

Figure 10. The *CHD7-SOX2* pathway and the relationship between the genes and clinical phenotypes. Figure from Engelen *et al*. (2011). Reprinted with permission from the copyright holder.

The findings of the family in our series underline the clinical utility of the molecular genetic testing of CHARGE syndrome, and the importance of testing the parents even in the case of a truncating mutation. To this family, effective family planning could be offered, and there was no need for the other diagnostic procedures. In addition, the father could be informed and followed up for possible future complications.

5.3.2 Sotos syndrome

Given that 85-90% of the mutations in non-Japanese patients are point mutations, Sotos syndrome testing should be begun with an *NSD1* sequencing analysis, and to improve the sensitivity of the test, followed directly by an MLPA analysis to detect structural aberrations in those patients who test negative for mutations. In fact, instead of offering the sequencing and MLPA analyses separately as before, the test package in our laboratory now includes both analyses automatically. When considering deletions, as the test developed in this study aims particularly at confirming or excluding Sotos syndrome, the detection of an *NSD1* deletion is sufficient, and from this point of view, the size of the deletion is irrelevant, because clinical features seen in patients are primarily caused by *NSD1* haploinsufficiency. However, an array-based analysis can be further used to define the size of the deletion more precisely if the symptoms of the patient indicate a large deletion, because large genomic rearrangement will alter the prognosis to the extent that definition of the individual genomic status is to be recommended. In addition, the test specifically for Sotos syndrome does not reveal other possible genomic imbalances leading to different diagnoses. If the clinical features are suggestive of Sotos, an array-based analysis can be considered as a first line analysis.

The clinical sensitivity of the test, according to the clinical criteria of Cole and Hughes (1994), is approximately 90%. In about 10% of the clinically diagnosed patients, the pathogenic mutation probably locates outside the *NSD1* coding area or is located in another currently unrecognized gene and therefore remains undetected. The clinical specificity of the test is approximately 95%, because a small number of patients with atypical features carry a mutation in the *NSD1* gene, and their mutation will not be sought and detected if strict clinical criteria are used (Tatton-Brown *et al*. 2005). The analytical validity of the Sotos syndrome testing is close to 100%, if both sequencing and MLPA are used.

Due to the overlapping features of overgrowth syndromes, the differential diagnostics can be challenging in some cases. Particularly, fragile X -patients with tall stature and Weaver syndrome patients can have overlapping features with those with Sotos syndrome. In these cases, the typical Sotos patients should be first tested for

NSD1 mutations and then for deletions with MLPA, which would exclude an atypical case of fragile X syndrome with a tall stature. Then the testing should proceed to *NSD1* sequencing and/or array analysis based on the probability of the syndrome. In addition, mutations of *EZH2* causing Weaver syndrome could be tested (if available) for differential diagnostics.

The criteria for clinical utility of Sotos syndrome testing are clearly fulfilled (Javaher *et al*. 2008). The *NSD1* defect is confirmatory, and its absence highly suggestive of another diagnosis. Confirmation of a diagnosis is extremely important for the psychosocial health of a family, and the stressful search for a diagnosis with different, often invasive, methods ends. The natural history of Sotos syndrome is fairly well known and recommendations for its follow-up exist, which are important issues in the counseling of the family (Cole, 2005).

5.3.3 3p deletion syndrome

The main question in 3p deletion syndrome concerns the size of the deletion and which genes contribute to the phenotype. To date, the smallest deletion associated with the typical 3p deletion syndrome was interstitial and only 1.6 Mb in size, and was located between 8.3 Mb and 9.8 Mb in the 3p arm (Gunnarsson and Foyn Bruun 2010). In this study, both families were originally diagnosed with 3p deletion syndrome based on FISH results. Our results confirmed, however, that small terminal deletions are likely to be rare copy number variants, and the deletion of *CHL1* is not sufficient to cause the clinical features of the syndrome.

When the causative interstitial deletions may be less than 2 Mb in size, the use of subtelomere FISH as a diagnostic method is not recommended as it leaves these deletions undetected. MLPA has a better resolution, but currently available kits do not cover a sufficient region of the distal 3p for the smallest deletions to be detected. Thus, an array-based analysis is the primary choice for the diagnostics of the 3p deletion. However, as seen also in this study, the presence of benign rare copy number variants is a general problem in the interpretation of array results, and special care should be taken when variants are identified as pathogenic. In general, benign copy number variants are usually smaller than pathogenic variants, with the smallest ones just larger than 50 bp (Alkan, Coe and Eichler 2011). Different databases collecting both benign and pathogenic variants can be used in the assessment of the pathogenicity of the variants in patients (see section 2.5.2)

When a sample of a patient with a nonspecific dysmorphism/malformation /developmental delay is analyzed with an array and the finding indicates a structural aberration, the reverse problem may appear. One has to decide which symptoms can

be explained with the arrangement, which are unrelated to the detected malformation, and what is to be expected in the future. In the future, when exome sequencing becomes a common routine diagnostic method, the focus of the clinical diagnostics may be shifted from pre-test assessment to post-test evaluation to decide the true pathogenic variant(s) behind the symptoms of the patients (Hennekam and Biesecker 2012). Recent analyses have revealed that unexpected findings may appear in the "wrong" genes. In a study by Majewski *et al*. (2011), a patient diagnosed with Leber's congenital amaurosis (LCA) was not carrying mutations in any of the large number of loci responsible for this genetically heterogeneous phenotype. Instead, she was homozygous for a well-known Zellweger Syndrome mutation at *PEX1* (Majewski *et al*. 2011). In another case, exome sequencing revealed congenital chloride diarrhea by detecting a mutation in *SLC26A3* in a case suspected of being Bartter's syndrome (Choi *et al*. 2009).

The clinical utility of the 3p deletion testing is fulfilled, as with the diagnosis rehabilitation can be offered to the patients and follow-up organized for the growth and cardiac problems. In addition, patients whose deletion extends to the *VHL1* gene can be followed up for the possible development of Von Hippel-Lindau disease. In genetic counseling, the transmission of deletions from phenotypically normal or mildly affected parents to their offspring creates challenges, especially if the abnormalities are sporadic and detected prenatally, as even large deletions can cause only minor deficits and the phenotype of the offspring cannot be predicted solely on the basis of chromosomal findings. The growing understanding of genotype–phenotype correlations may clarify some situations of this type in the future.

6 INSIGHTS INTO THE FUTURE

In little over 60 years, the field of molecular diagnostics has seen a huge development from the first reports of whole chromosome abnormalities to sequencing of the complete human genome and solving the genetic backgrounds of thousands of syndromes. Even though we now have robust methods for use in the laboratories in order to analyze our genomes and mutations, such as Sanger sequencing and different types of arrays, the power of these methods is nonetheless limited. At present, the field of molecular diagnostics is experiencing the next fundamental change in the form of Next Generation Sequencing techniques (NGS) or more precisely, exome sequencing (ES).

In molecular genetic research, NGS has already been shown to be a valuable tool for mining the human genome, and when considering ES in routine diagnostics, many laboratories have begun to implement it as a part of their routine, and are thereby pioneering this work. Consequently, the question in routine diagnostics has already shifted from the ”should we?” to ”how can we implement it in the best possible way?” ES promises to be the method in diagnostics by which the etiology of even the rarest of syndromes could be dissected, but the success of this technique raises questions: is there any longer a need for targeted diagnostic testing or will all testing be eventually translated into exome sequencing? The main challenge for ES in diagnostics is the essential problem of how to translate the massive amount of data produced into relevant information within the frames of routine diagnostics (Figure 11). However, recent examples of ES in the diagnostics of Charcot-Marie-Tooth disease and many other diseases imply that this challenge can be overcome (Montenegro *et al*. 2011), and that in the near future, the molecular diagnostics of rare syndromes will be fundamentally different.

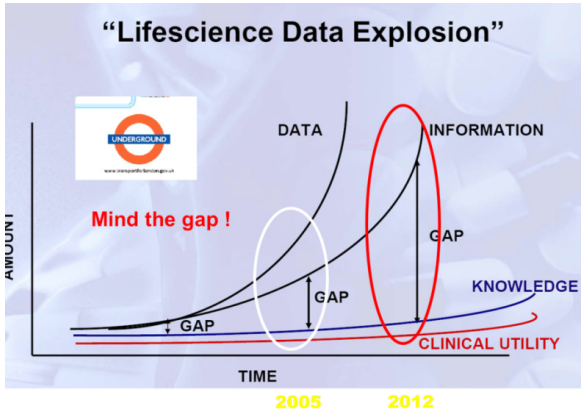

Figure 11. Mind the gap! Figure reprinted by courtesy of Prof. Milan Macek.

7 CONCLUDING REMARKS

I. Molecular diagnostic tests were developed for all the syndromes in the study and are currently used in the routine diagnostics service. When developing a test for a given syndrome, one has to first consider whether the test is worth developing, that is, is there a need for such testing? For validity and utility, methods covering all types of mutations known to cause the syndrome should be applied but with a reasonable cost. Both CHARGE syndrome and Sotos syndrome testing have been requested often, confirming the need for testing for these syndromes in a local setting.

II. The genetic background of the CHARGE and Sotos syndromes were studied for the first time in the Finnish population, and the results represent previously published characteristics in terms of mutation types, frequency of mutations and detection percent. In 3p deletion syndrome, the results show that small, terminal deletions are not sufficient to cause the syndrome, but are likely to be benign variants, and further molecular diagnostic analyses are recommended for patients with small terminal deletions.

III. As shown by the negative mutation test result of the father's brother in the CHARGE family, occasionally molecular genetic analyses reveal unexpected findings, underlining the significance of confirming a clinical diagnosis by a genetic test. Only in patients with the most typical syndromic features could the clinical diagnosis be considered "definite" without a diagnostic test. Most patients, however, show only some features suggesting a syndrome, and confirming or ruling out the etiologic diagnosis rests mainly on the molecular diagnostics.

IV. The molecular diagnostics of rare syndromes is in the process of change, as the next generation sequencing techniques are becoming common as methods. Whether the single gene testing as a primary test becomes unnecessary remains to be seen, but mutation testing of the patient's family members with conventional methods may still be needed. In addition, although diagnostic methods will be changed, the same challenge of deciding the true pathogenic variant amongst many possibilities and explaining the significance of unexpected findings of relatives still remains.

ACKNOWLEDGEMENTS

This study was carried out at the former Department of Medical Genetics and the present Department of Medical Biochemistry and Genetics, University of Turku during the years 2005-2011. I would like to thank the former and present heads of the department, Professor Emerita Marja-Liisa Savontaus, Research Professor Helena Kääriäinen, Professor Klaus Elenius and Professor Johanna Schleutker for providing excellent working facilities.

I wish to express my profound gratitude to my principal supervisor Helena Kääriäinen for taking this PhD road till the end with me. With your enthusiasm, expertise, effort and encouragement I was able to finish the thesis. I owe my gratitude to my supervisor Klaus Elenius for taking over for the last meters and for the valuable comments on the fourth article and the thesis. I warmly thank Adjunct Professor Kirsi Huoponen for her encouragement to start my PhD studies and sharing her valuable technical expertise. Marja-Liisa Savontaus has my deepest gratitude, your enthusiasm and inspiring lessons at the beginning of my biology studies convinced me to study medical genetics.

I acknowledge the reviewers Marketta Kähkönen and Jukka Moilanen for their critical and valuable comments for my thesis. Minna Pöyhönen and Tiina Alitalo are warmly thanked for participating in my supervisory committee, and Pertti Aula is warmly thanked for kindly helping me with the historical part of my thesis. Damon and Maaria Tringham are warmly appreciated for the language revision of the thesis.

I offer my sincere gratitude to the patients and families involved in this study for their invaluable participation and making this study possible.

I owe my deep gratitude to all clinicians working with families and patients for their expertise and support, especially to Marja Hietala and Maila Penttinen. I also would like to thank Marja for the hours spent in teaching and for making my hours of duty much easier and joyful. I am grateful to all collaborators and authors of the articles for their contribution. I owe my gratitude to the lab secretary Pia Tahvanainen for always giving me a helping hand and keeping things running. I would like to thank the present and former employees of the DNA and chromosome diagnostic laboratories for their help and expertise, as well as the personnel of the clinical genetics unit at the Turku University Central Hospital. Special thanks for all of the students who participated in the lab work of the study, especially Katrin Hahn.

The years spent in X-boksi were full of joy thanks to my colleague PhD students Maaria, Maija, Johanna, Petra, Kati, Marie and Minna and all the present and former members of ”Tutkijat”. I thank you for the ever-so-lively discussions on scientific and not-so-scientific topics, the fun life outside the lab and the support you have given me over the years. All my friends in the outside world are truly appreciated for being there.

My warmest gratitude goes to my grandmother, mum, dad, and brother for support and for always being there for me. My husband and best friend Perttu, thank you for giving me the strength when the rocks on the way seemed to fall on me, and always making me laugh. Lauri, my little ”Puuha-Pete”, with you all the problems are just a faint memory. My unborn ”Nappula”, I can’t wait to see you here. I love you all.

This study was financially supported by the EVO fundings of the Turku University Central Hospital, the Turku University Foundation, the Emil Aaltonen Foundation, the Finnish Concordia Fund, the Medical Faculty of the University of Turku and EuroGentest2, an EU-FP7 supported coordination action (contract number HEALTH-F4-2010-261469).

This work is dedicated to all patients who participated in the study.

Turku, June 2012

Pia Pohjola

REFERENCES

ELECTRICAL REFERENCES:

1000Genomes browser: http://browser.1000genomes.org

DEPICHER: http://decipher.sanger.ac.uk/

ECARUCA: http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp

European Commission on Rare Diseases:
http://ec.europa.eu/health/rare_diseases/policy/index_en.htm

Eurogentest: http://www.eurogentest.org

Eurogentest Unit 3: http://www.eurogentest.org/unit5/

European Organization of Rare Diseases (EURORDIS): http://www.eurordis.org/

GeneTests: http://www.genetests.org

Harvinaiset-verkosto: http://www.harvinaiset.fi

Human Gene Mutation Database (HGMD): http://www.hgmd.org/

London Medical Databases (Winter-Baraitser): http://www.lmdatabases.com/

Mutalyzer: https://mutalyzer.nl/

National Human Genome Research Institute (NHGRI): http://www.genome.gov/17515951

National Organization of Rare Diseases (NORD): http://www.rarediseases.org

Orphanet Database of Rare Diseases: http://www.orpha.net

Online Mendelian Inheritance in Man (OMIM):
http://www.ncbi.nlm.nih.gov/omimSamankaltaisia

Pmut: http://mmb.pcb.ub.es/PMut/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/

Possum Dysmorpohlogy Database: http://www.possum.net.au

Thefreedictionary: http://www.thefreedictionary.com/

PUBLICATIONS:

Alkan, C., B. P. Coe & E. E. Eichler (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet,* 12**,** 363-76.

Angrand, P. O., F. Apiou, A. F. Stewart, B. Dutrillaux, R. Losson & P. Chambon (2001) NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics,* 74**,** 79-88.

Aramaki, M., T. Udaka, R. Kosaki, Y. Makita, N. Okamoto, H. Yoshihashi, H. Oki, K. Nanao, N. Moriyama, S. Oku, T. Hasegawa, T. Takahashi, Y. Fukushima, H. Kawame & K. Kosaki (2006) Phenotypic spectrum of CHARGE syndrome with CHD7 mutations. *J Pediatr,* 148**,** 410-4.

Aula, P. (1978) Lapsivesitutkimukset Suomessa. *Duodecim***,** 9-11.

--- (1981) Lapsivesitutkimus sikiödiagnostiikassa. *Duodecim***,** 1339-52.

--- (1991) Periytyvien tautien DNA-diagnostiikka Suomessa. *Suomen Lääkärilehti,* 46**,** 2288-2295.

Aula, P., H. Kääriäinen & A. Palotie (eds) (2007) *Perinnöllisyyslääketiede*, 3. *painos.* Helsinki: Duodecim. 368 pp.

Bajpai, R., D. A. Chen, A. Rada-Iglesias, J. Zhang, Y. Xiong, J. Helms, C. P. Chang, Y. Zhao, T. Swigut & J. Wysocka (2010) CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature,* 463**,** 958-62.

Barber, C. (2005) Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet,* 42(8):609-29.

Bartels, C. F., C. Scacheri, L. White, P. C. Scacheri & S. Bale (2010) Mutations in the CHD7 gene: the experience of a commercial laboratory. *Genet Test Mol Biomarkers,* 14**,** 881-91.

Baujat, G., M. Rio, S. Rossignol, D. Sanlaville, S. Lyonnet, M. Le Merrer, A. Munnich, C. Gicquel, V. Cormier-Daire & L. Colleaux (2004) Paradoxical NSD1 mutations in Beckwith-Wiedemann syndrome and 11p15 anomalies in Sotos syndrome. *Am J Hum Genet,* 74**,** 715-20.

Bauman, J. G., J. Wiegant, P. Borst & P. van Duijn (1980) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp Cell Res,* 128**,** 485-90.

Benini, D., L. Vino, S. Vecchini & V. Fanos (1999) 46, XY, del (3) (pter-->p25) syndrome: further delineation of the clinical phenotype. *Eur J Pediatr,* 158**,** 955-7.

Bergman, J. E., N. Janssen, L. H. Hoefsloot, M. C. Jongmans, R. M. Hofstra & C. M. van Ravenswaaij-Arts (2011) CHD7 mutations and CHARGE syndrome: the clinical implications of an expanding phenotype. *J Med Genet,* 48**,** 334-42.

Bergman JE, I. de Wijs, M.C. Jongmans, R.J. Admiraal, L.H. Hoefsloot & C.M. van Ravenswaaij-Arts (2008) Exon copy number alterations of the CHD7 gene are not a major cause of CHARGE and CHARGE-like syndrome. *Eur J Med Genet.* 51(5):417-25.

Bilan, F., M. Legendre, V. Charraud, B. Manière, D. Couet, B. Gilbert-Dussardier & A. Kitzis (2012) Complete screening of 50 patients with CHARGE syndrome for anomalies in the CHD7 gene using a denaturing high-performance liquid chromatography-based protocol: new guidelines and a proposal for routine diagnosis. *J Mol Diagn,* 14**,** 46-55.

Blake, K., S. Davenport, B. Hall, M. Hefner, R. Pagon, M. Williams, A. Lin & J. J. Graham (1998) CHARGE association: an update and review for the primary pediatrician. *Clin Pediatr (Phila),* 37**,** 159-73.

Bosman, E. A., A. C. Penn, J. C. Ambrose, R. Kettleborough, D. L. Stemple & K. P. Steel (2005) Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet,* 14**,** 3463-76.

Cargile, C. B., D. L. Goh, B. K. Goodman, X. N. Chen, J. R. Korenberg, G. L. Semenza & G. H. Thomas (2002) Molecular cytogenetic characterization of a subtle interstitial del(3)(p25.3p26.2) in a patient with deletion 3p syndrome. *Am J Med Genet,* 109**,** 133-8.

Caspersson, T., L. Zech & C. Johansson (1970) Analysis of human metaphase chromosome set by aid of DNA-binding fluorescent agents. *Exp Cell Res,* 62**,** 490-2.

Cassidy, S. B. & J. E. Allanson. 2010. *Management of Genetic Syndromes, 3rd edition*. Wiley-Blackwell, Hoboken, New Jersey. 984 pp.

CHARGE syndrome support group: personal discussions with families. 14.6.2008 Jyväskylä

Chen, J. M., C. Férec & D. N. Cooper (2010) Revealing the human mutome. *Clin Genet,* 78**,** 310-20.

Choi, M., U. I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloğlu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane & R. P. Lifton (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A,* 106**,** 19096-101.

Cole, T.R. (2005) Sotos Syndrome. In: Cassidy SB and Allanson JE (eds). Management of Genetic Syndromes, 2nd ed. Hoboken, New Jersey, pp 527-538

Cole, T. R. & H. E. Hughes (1994) Sotos syndrome: a study of the diagnostic criteria and natural history. *J Med Genet,* 31**,** 20-32.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C. H. Ihm, K. Kristiansson, D. G. Macarthur, J. R. Macdonald, I. Onyiah, A. W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N. P. Carter, C. Lee, S. W. Scherer, M. E. Hurles & W. T. C. C. Consortium (2010) Origins and functional impact of copy number variation in the human genome. *Nature,* 464**,** 704-12.

Cottrell, C. E., N. Bir, E. Varga, C. E. Alvarez, S. Bouyain, R. Zernzach, D. L. Thrush, J. Evans, M. Trimarchi, E. M. Butter, D. Cunningham, J. M. Gastier-Foster, K. L. McBride & G. E. Herman (2011) Contactin 4 as an autism susceptibility locus. *Autism Res,* 4**,** 189-99.

Cuoco, C., P. Ronchetto, S. Gimelli, F. Béna, M. T. Divizia, M. Lerone, M. Mirabelli-Badenier, M. Mascaretti & G. Gimelli (2011) Microarray based analysis of an inherited terminal 3p26.3 deletion, containing only the CHL1 gene, from a normal father to his two affected children. *Orphanet J Rare Dis,* 6**,** 12.

Das, S., S. J. Bale & D. H. Ledbetter (2008) Molecular genetic testing for ultra rare diseases: models for translation from the research laboratory to the CLIA-certified diagnostic laboratory. *Genet Med,* 10**,** 332-6.

de Vries, B. B., R. Pfundt, M. Leisink, D. A. Koolen, L. E. Vissers, I. M. Janssen, S. Reijmersdal, W. M. Nillesen, E. H. Huys, N. Leeuw, D. Smeets, E. A. Sistermans, T. Feuth, C. M. van Ravenswaaij-Arts, A. G. van Kessel, E. F. Schoenmakers, H. G. Brunner & J. A. Veltman (2005) Diagnostic genome profiling in mental retardation. *Am J Hum Genet,* 77**,** 606-16.

Delahaye, A., Y. Sznajer, S. Lyonnet, M. Elmaleh-Bergès, I. Delpierre, S. Audollent, S. Wiener-Vacher, A. L. Mansbach, J. Amiel, C. Baumann, D. Bremond-Gignac, T. Attié-Bitach, A. Verloes & D. Sanlaville (2007) Familial CHARGE syndrome because of CHD7 mutation: clinical intra- and interfamilial variability. *Clin Genet,* 72**,** 112-21.

Demidov, V. V. (2003) DNA diagnostics in the fifty-year retrospect. *Expert Rev Mol Diagn,* 3**,** 121-4.

Devriendt, K. & J. R. Vermeesch (2004) Chromosomal phenotypes and submicroscopic abnormalities. *Hum Genomics,* 1**,** 126-33.

Devys, D., V. Biancalana, F. Rousseau, J. Boue, J. L. Mandel & I. Oberle (1992) Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. *Am. J. Med. Genet,* 43: 208-216

Dijkhuizen T, T. van Essen, P. van der Vlies P, J.B. Verheij, B. Sikkema-Raddatz, A.Y. van der Veen, K.B. Gerssen-Schoorl, C.H. Buys & Kok K (2006) FISH and array-CGH analysis of a complex chromosome 3 aberration suggests that loss of CNTN4 and CRBN contributes to mental retardation in 3pter deletions. *Am J Med Genet A.* 140(22):2482-7.

Douglas, J., K. Coleman, K. Tatton-Brown, H. E. Hughes, I. K. Temple, T. R. Cole, N. Rahman & C. O. Collaboration (2005) Evaluation of NSD2 and NSD3 in overgrowth syndromes. *Eur J Hum Genet,* 13**,** 150-3.

Douglas, J., S. Hanks, I. K. Temple, S. Davies, A. Murray, M. Upadhyaya, S. Tomkins, H. E. Hughes, T. R. Cole & N. Rahman (2003) NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. *Am J Hum Genet,* 72**,** 132-43.

Endris, V., B. Wogatzky, U. Leimer, D. Bartsch, M. Zatyka, F. Latif, E. R. Maher, G. Tariverdian, S. Kirsch, D. Karch & G. A. Rappold (2002) The novel Rho-GTPase activating gene MEGAP/ srGAP3 has a putative role in severe mental retardation. *Proc Natl Acad Sci U S A,* 99**,** 11754-9.

Engelen, E., U. Akinci, J. C. Bryne, J. Hou, C. Gontan, M. Moen, D. Szumska, C. Kockx, W. van Ijcken, D. H. Dekkers, J. Demmers, E. J. Rijkers, S. Bhattacharya, S. Philipsen, L. H. Pevny, F. G. Grosveld, R. J. Rottier, B. Lenhard & R. A. Poot (2011) Sox2 cooperates with Chd7 to regulate

genes that are mutated in human syndromes. *Nat Genet,* 43**,** 607-11.

Fagali, C., F. Kok, P. Nicola, C. Kim, D. Bertola, L. Albano & C. P. Koiffmann (2009) MLPA analysis in 30 Sotos syndrome patients revealed one total NSD1 deletion and two partial deletions not previously reported. *Eur J Med Genet,* 52**,** 333-6.

Faucett, W. A., S. Hart, R. A. Pagon, L. F. Neall & G. Spinella (2008) A model program to increase translation of rare disease genetic tests: collaboration, education, and test translation program. *Genet Med,* 10**,** 343-8.

Fernandez, T. V., I. J. García-González, C. E. Mason, G. Hernández-Zaragoza, V. C. Ledezma-Rodríguez, V. M. Anguiano-Alvarez, R. E'Vega, M. Gutiérrez-Angulo, M. L. Maya, H. E. García-Bejarano, M. González-Cruz, S. Barrios, R. Atorga, M. G. López-Cardona, J. Armendariz-Borunda, M. W. State & N. O. Dávalos (2008) Molecular characterization of a patient with 3p deletion syndrome and a review of the literature. *Am J Med Genet A,* 146A**,** 2746-52.

Fickie, M. R., P. Lapunzina, J. K. Gentile, N. Tolkoff-Rubin, D. Kroshinsky, E. Galan, E. Gean, L. Martorell, V. Romanelli, J. F. Toral & A. E. Lin (2011) Adults with Sotos syndrome: review of 21 adults with molecularly confirmed NSD1 alterations, including a detailed case report of the oldest person. *Am J Med Genet A,* 155A**,** 2105-11.

Ford, C. E., K. W. Jones, P. E. Polani, J. C. De Almeida & J. H. Briggs (1959) A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet,* 1**,** 711-3.

Forsström, L. 2012 Molecular diagnosis of developmental disorders. Thesis. In *Department of Medicine*. Helsinki: University of Helsinki.

Frints, S. G., P. Marynen, D. Hartmann, J. P. Fryns, J. Steyaert, M. Schachner, B. Rolf, K. Craessaerts, A. Snellinx, K. Hollanders, R. D'Hooge, P. P. De Deyn & G. Froyen (2003) CALL interrupted in a patient with non-specific mental retardation: gene dosage-dependent alteration of murine brain development and behavior. *Hum Mol Genet,* 12**,** 1463-74.

Félix, T. M., B. C. Hanshaw, R. Mueller, P. Bitoun & J. C. Murray (2006) CHD7 gene and non-syndromic cleft lip and palate. *Am J Med Genet A,* 140**,** 2110-4.

Gall, J. G. & M. L. Pardue (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci U S A,* 63**,** 378-83.

Gennery, A. R., M. A. Slatter, J. Rice, L. H. Hoefsloot, D. Barge, A. McLean-Tooke, T. Montgomery, J. A. Goodship, A. D. Burt, T. J. Flood, M. Abinun, A. J. Cant & D. Johnson (2008) Mutations in CHD7 in patients with CHARGE syndrome cause T-B + natural killer cell + severe combined immune deficiency and may cause Omenn-like syndrome. *Clin Exp Immunol,* 153**,** 75-80.

Gibson, W. T., R. L. Hood, S. H. Zhan, D. E. Bulman, A. P. Fejes, R. Moore, A. J. Mungall, P. Eydoux, R. Babul-Hirji, J. An, M. A. Marra, D. Chitayat, K. M. Boycott, D. D. Weaver, S. J. Jones & F. C. Consortium (2012) Mutations in EZH2 cause Weaver syndrome. *Am J Hum Genet,* 90**,** 110-8.

Gojobori, T., W. H. Li & D. Graur (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol,* 18**,** 360-9.

Graham, J. M., B. Rosner, E. Dykens & J. Visootsak (2005) Behavioral features of CHARGE syndrome (Hall-Hittner syndrome) comparison with Down syndrome, Prader-Willi syndrome, and Williams syndrome. *Am J Med Genet A,* 133A**,** 240-7.

Green, E. K., M. D. Priestley, J. Waters, C. Maliszewska, F. Latif & E. R. Maher (2000) Detailed mapping of a congenital heart disease gene in chromosome 3p25. *J Med Genet,* 37**,** 581-7.

Grody, W. W. & C. S. Richards (2008) New quality assurance standards for rare disease testing. *Genet Med,* 10**,** 320-4.

Gunnarsson, C. & C. Foyn Bruun (2010) Molecular characterization and clinical features of a patient with an interstitial deletion of 3p25.3-p26.1. *Am J Med Genet A,* 152A**,** 3110-4.

Gusella, J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace & A. Y. Sakaguchi (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature,* 306**,** 234-8.

Hall, B. (1979) Choanal atresia and associated multiple anomalies. *J Pediatr,* 95**,** 395-8.

Hall, J. A. & P. T. Georgel (2007) CHD proteins: a diverse family with strong ties. *Biochem Cell Biol,* 85**,** 463-76.

Hehir-Kwa, J. Y., M. Egmont-Petersen, I. M. Janssen, D. Smeets, A. G. van Kessel & J. A. Veltman (2007) Genome-wide copy number profiling

on high-density bacterial artificial chromosomes, single-nucleotide polymorphisms, and oligonucleotide microarrays: a platform comparison based on statistical power analysis. *DNA Res,* 14**,** 1-11.

Hennekam, R.C. & L.G. Biesecker (2012) *Hum Mut* 33:884-886

Higgins J.J., J. Pucilowska, R.Q. Lombardi & J.P. Rooney (2004) *Neurology*, 63(10):1927-31.

Hittner, H., N. Hirsch, G. Kreh & A. Rudolph (1979) Colobomatous microphthalmia, heart disease, hearing loss, and mental retardation -a syndrome. *J Pediatr Ophthalmol Strabismus,* 16**,** 122-8.

Hood R.L., M.A. Lines, S.M. Nikkel, J. Schwartzentruber, C. Beaulieu, M.J. Nowaczyk, J. Allanson, C.A. Kim, D. Wieczorek, J.S. Moilanen, D. Lacombe, G. Gillessen-Kaesbach, M.L. Whiteford, C.R. Quaio, I. Gomy, D.R. Bertola, B. Albrecht, K. Platzer, G. McGillivray, R. Zou, D.R. McLeod, A.E. Chudley, B.N. Chodirker, J. Marcadier; FORGE Canada Consortium, J. Majewski, D.E. Bulman, S.M. White & Boycott K.M. (2012) Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome. *Am J Hum Genet*, 90(2):308-13.

Huang, N., E. vom Baur, J. M. Garnier, T. Lerouge, J. L. Vonesch, Y. Lutz, P. Chambon & R. Losson (1998) Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J,* 17**,** 3398-412.

Hurd, E. A., M. E. Adams, W. S. Layman, D. L. Swiderski, L. A. Beyer, K. E. Halsey, J. M. Benson, T. W. Gong, D. F. Dolan, Y. Raphael & D. M. Martin (2011) Mature middle and inner ears express Chd7 and exhibit distinctive pathologies in a mouse model of CHARGE syndrome. *Hear Res,* 282**,** 184-95.

Höglund, P., N. Kurotaki, S. Kytölä, N. Miyake, M. Somer & N. Matsumoto (2003) Familial Sotos syndrome is caused by a novel 1 bp deletion of the NSD1 gene. *J Med Genet,* 40**,** 51-4.

Iafrate, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer & C. Lee (2004) Detection of large-scale variation in the human genome. *Nat Genet,* 36**,** 949-51.

Issekutz, K., J. J. Graham, C. Prasad, I. Smith & K. Blake (2005) An epidemiological analysis of CHARGE syndrome: preliminary results from a Canadian study. *Am J Med Genet A,* 133**,** 309-17.

Jacobs, P. A., A. G. Baikie, W. M. Brown, T. N. MacGregor, N. Maclean & D. G. Harnden (1959) Evidence for the existence of the human "super female". *Lancet,* 2**,** 423-5.

Jacobs, P. A. & J. A. Strong (1959) A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature,* 183**,** 302-3.

Javaher, P., H. Kaariainen, U. Kristoffersson, I. Nippert, J. Sequeiros, R. Zimmern & J. Schmidtke (2008) EuroGentest: DNA-based testing for heritable disorders in Europe. *Community Genet,* 11**,** 75-120.

Jennings, L., V. M. Van Deerlin, M. L. Gulley & C. o. A. P. M. P. R. Committee (2009) Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med,* 133**,** 743-55.

Jones, K. (2005) *Smith's Recognizable Patterns of Human Malformation, 6th edition*. Elsevier Health Sciences. 976 pp.

Jongmans, M., R. Admiraal, K. van der Donk, L. Vissers, A. Baas, L. Kapusta, J. van Hagen, D. Donnai, T. de Ravel, J. Veltman, A. Geurts van Kessel, B. De Vries, H. Brunner, L. Hoefsloot & C. van Ravenswaaij (2006) CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J Med Genet,* 43**,** 306-14.

Jongmans, M. C., L. H. Hoefsloot, K. P. van der Donk, R. J. Admiraal, A. Magee, I. van de Laar, Y. Hendriks, J. B. Verheij, I. Walpole, H. G. Brunner & C. M. van Ravenswaaij (2008) Familial CHARGE syndrome and the CHD7 gene: a recurrent missense mutation, intrafamilial recurrence and variability. *Am J Med Genet A,* 146A**,** 43-50.

Juvonen, V. Personal discussion. 17.2.2012.

Kamimura, J., Y. Endo, N. Kurotaki, A. Kinoshita, N. Miyake, O. Shimokawa, N. Harada, R. Visser, H. Ohashi, K. Miyakawa, J. Gerritsen, A. M. Innes, L. Lagace, M. Frydman, N. Okamoto, R. Puttinger, S. Raskin, B. Resic, V. Culic, K. Yoshiura, T. Ohta, T. Kishino, M. Ishikawa, N. Niikawa & N. Matsumoto (2003) Identification of eight novel NSD1 mutations in Sotos syndrome. *J Med Genet,* 40**,** e126.

Kanemoto, N., K. Kanemoto, G. Nishimura, T. Kamoda, R. Visser, O. Shimokawa & N. Matsumoto (2006) Nevo syndrome with an NSD1 deletion: a variant of Sotos syndrome? *Am J Med Genet A,* 140**,** 70-3.

Kirchhoff, M., A. M. Bisgaard, T. Bryndorf & T. Gerdes (2007) MLPA analysis for a panel of

syndromes with mental retardation reveals imbalances in 5.8% of patients with mental retardation and dysmorphic features, including duplications of the Sotos syndrome and Williams-Beuren syndrome regions. *Eur J Med Genet,* 50**,** 33-42.

Knight, L. A., M. H. Yong, M. Tan & I. S. Ng (1995) Del(3) (p25.3) without phenotypic effect. *J Med Genet,* 32**,** 994-5.

Koenig, M., E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener & L. M. Kunkel (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell,* 50**,** 509-17.

Koolen, D. A., L. E. Vissers, R. Pfundt, N. de Leeuw, S. J. Knight, R. Regan, R. F. Kooy, E. Reyniers, C. Romano, M. Fichera, A. Schinzel, A. Baumer, B. M. Anderlid, J. Schoumans, N. V. Knoers, A. G. van Kessel, E. A. Sistermans, J. A. Veltman, H. G. Brunner & B. B. de Vries (2006) A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet,* 38**,** 999-1001.

Kosaki, K., T. Udaka & T. Okuyama (2005) DHPLC in clinical molecular diagnostic services. *Mol Genet Metab,* 86**,** 117-23.

Krawczak, M., E. V. Ball & D. N. Cooper (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am J Hum Genet,* 63**,** 474-88.

Kremer, E. J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland & R. I. Richards. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 252: 1711-1714

Kurotaki, N., N. Harada, O. Shimokawa, N. Miyake, H. Kawame, K. Uetake, Y. Makita, T. Kondoh, T. Ogata, T. Hasegawa, T. Nagai, T. Ozaki, M. Touyama, R. Shenhav, H. Ohashi, L. Medne, T. Shiihara, S. Ohtsu, Z. Kato, N. Okamoto, J. Nishimoto, D. Lev, Y. Miyoshi, S. Ishikiriyama, T. Sonoda, S. Sakazume, Y. Fukushima, K. Kurosawa, J. F. Cheng, K. Yoshiura, T. Ohta, T. Kishino, N. Niikawa & N. Matsumoto (2003) Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion. *Hum Mutat,* 22**,** 378-87.

Kurotaki, N., K. Imaizumi, N. Harada, M. Masuno, T. Kondoh, T. Nagai, H. Ohashi, K. Naritomi, M. Tsukahara, Y. Makita, T. Sugimoto, T. Sonoda, T. Hasegawa, Y. Chinen, H. A. Tomita Ha, A. Kinoshita, T. Mizuguchi, K. Yoshiura Ki, T. Ohta, T. Kishino, Y. Fukushima, N. Niikawa & N. Matsumoto (2002) Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet,* 30**,** 365-6.

Kurotaki, N., P. Stankiewicz, K. Wakui, N. Niikawa & J. R. Lupski (2005) Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. *Hum Mol Genet,* 14**,** 535-42.

Lalani, S., A. Safiullah, S. Fernbach, K. Harutyunyan, C. Thaller, L. Peterson, J. McPherson, R. Gibbs, L. White, M. Hefner, S. Davenport, J. Graham, C. Bacino, N. Glass, J. Towbin, W. Craigen, S. Neish, A. Lin & J. Belmont (2006) Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am J Hum Genet,* 78**,** 303-14.

Lalani, S. R., A. M. Safiullah, L. M. Molinari, S. D. Fernbach, D. M. Martin & J. W. Belmont (2004) SEMA3E mutation in a patient with CHARGE syndrome. *J Med Genet,* 41**,** e94.

Lee, J. S. & A. Shilatifard (2007) A site to remember: H3K36 methylation a mark for histone deacetylation. *Mutat Res,* 618**,** 130-4.

Lejeune, J., R. Turpin & M. Gautier (1959) [Mongolism; a chromosomal disease (trisomy)]. *Bull Acad Natl Med,* 143**,** 256-65.

Li, L., I. D. Krantz, Y. Deng, A. Genin, A. B. Banta, C. C. Collins, M. Qi, B. J. Trask, W. L. Kuo, J. Cochran, T. Costa, M. E. M. Pierpont, E. B. Rand, D. A. Piccoli, L. Hood & N. B. Spinner (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genet,* 16: 243-251

Liu, P., C. M. Carvalho, P. Hastings & J. R. Lupski (2012) Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev*.

Lynch, M. (2010) Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A,* 107**,** 961-8.

Maddalena, A., S. Bale, S. Das, W. Grody, S. Richards & A. L. Q. A. Committee (2005) Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. *Genet Med,* 7**,** 571-83.

Majewski, J., Z. Wang, I. Lopez, S. Al Humaid, H. Ren, J. Racine, A. Bazinet, G. Mitchel, N. Braverman & R. K. Koenekoop (2011) A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: novel use of genomic diagnostics and exome sequencing. *J Med Genet,* 48**,** 593-6.

Malmgren, H., S. Sahlén, K. Wide, M. Lundvall & E. Blennow (2007) Distal 3p deletion syndrome: detailed molecular cytogenetic and clinical characterization of three small distal deletions and review. *Am J Med Genet A,* 143A**,** 2143-9.

Marfella, C. G. & A. N. Imbalzano (2007) The Chd family of chromatin remodelers. *Mutat Res,* 618**,** 30-40.

Mattocks, C. J., M. A. Morris, G. Matthijs, E. Swinnen, A. Corveleyn, E. Dequeker, C. R. Müller, V. Pratt, A. Wallace & E. V. Group (2010) A standardized framework for the validation and verification of clinical molecular genetic tests. *Eur J Hum Genet,* 18**,** 1276-88.

McCullough, B. J., J. C. Adams, D. J. Shilling, M. P. Feeney, K. C. Sie & B. L. Tempel (2007) 3p--syndrome defines a hearing loss locus in 3p25.3. *Hear Res,* 224**,** 51-60.

Melchior, L., M. Schwartz & M. Duno (2005) dHPLC screening of the NSD1 gene identifies nine novel mutations--summary of the first 100 Sotos syndrome mutations. *Ann Hum Genet,* 69**,** 222-6.

Merrild, U., S. Berggreen, L. Hansen, M. Mikkelsen & K. Henningsen (1981) Partial deletion of the short arm of chromosome 3. *Eur J Pediatr,* 136**,** 211-6.

Miller, M. P., J. D. Parker, S. W. Rissing & S. Kumar (2003) Quantifying the intragenic distribution of human disease mutations. *Ann Hum Genet,* 67**,** 567-79.

Miyake, N., N. Kurotaki, H. Sugawara, O. Shimokawa, N. Harada, T. Kondoh, M. Tsukahara, S. Ishikiriyama, T. Sonoda, Y. Miyoshi, S. Sakazume, Y. Fukushima, H. Ohashi, T. Nagai, H. Kawame, K. Kurosawa, M. Touyama, T. Shiihara, N. Okamoto, J. Nishimoto, K. Yoshiura, T. Ohta, T. Kishino, N. Niikawa & N. Matsumoto (2003) Preferential paternal origin of microdeletions caused by prezygotic chromosome or chromatid rearrangements in Sotos syndrome. *Am J Hum Genet,* 72**,** 1331-7.

Montenegro, G., E. Powell, J. Huang, F. Speziani, Y. J. Edwards, G. Beecham, W. Hulme, C. Siskind, J. Vance, M. Shy & S. Züchner (2011) Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family. *Ann Neurol,* 69**,** 464-70.

Morishita, M. & E. di Luccio (2011) Cancers and the NSD family of histone lysine methyltransferases. *Biochim Biophys Acta,* 1816**,** 158-63.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn & H. Erlich (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol,* 51 Pt 1**,** 263-73.

Nagai, T., N. Matsumoto, N. Kurotaki, N. Harada, N. Niikawa, T. Ogata, K. Imaizumi, K. Kurosawa, T. Kondoh, H. Ohashi, M. Tsukahara, Y. Makita, T. Sugimoto, T. Sonoda, T. Yokoyama, K. Uetake, S. Sakazume, Y. Fukushima & K. Naritomi (2003) Sotos syndrome and haploinsufficiency of NSD1: clinical features of intragenic mutations and submicroscopic deletions. *J Med Genet,* 40**,** 285-9.

Neri, G. & M. Moscarda (2009) Overgrowth syndromes: a classification. *Endocr Dev,* 14**,** 53-60.

Ng, S. B., A. W. Bigham, K. J. Buckingham, M. C. Hannibal, M. J. McMillin, H. I. Gildersleeve, A. E. Beck, H. K. Tabor, G. M. Cooper, H. C. Mefford, C. Lee, E. H. Turner, J. D. Smith, M. J. Rieder, K. Yoshiura, N. Matsumoto, T. Ohta, N. Niikawa, D. A. Nickerson, M. J. Bamshad & J. Shendure (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet,* 42**,** 790-3.

Nielsen, A. L., P. Jørgensen, T. Lerouge, M. Cerviño, P. Chambon & R. Losson (2004) Nizp1, a novel multitype zinc finger protein that interacts with the NSD1 histone lysine methyltransferase through a unique C2HR motif. *Mol Cell Biol,* 24**,** 5184-96.

Nirenberg, M., T. Caskey, R. Marshall, R. Brimacombe, D. Kellogg, B. Doctor, D. Hatfield, J. Levin, F. Rottman, S. Pestka, M. Wilcox & F. Anderson (1966) The RNA code and protein synthesis. *Cold Spring Harb Symp Quant Biol,* 31**,** 11-24.

Norio, R. (2003) Finnish Disease Heritage III. *Hum Genet,*112:470-526

Norio, R. (1966) Heredity in the congenital nephrotic syndrome. A genetic study of 57 finnish families with a review of reported cases. *Ann Paediatr Fenn,* 12**,** Suppl 27:1-94.

Oefner, P.J. & P.A. Underhill (1998) DNA mutation detection using denaturing high-performance liquid chromatocraphy (DHCPL).Current protocols in human genetics, suppl. 19 (7.10.1-7.10.12).

Ogata, T., I. Fujiwara, E. Ogawa, N. Sato, T. Udaka & K. Kosaki (2006) Kallmann syndrome phenotype in a female patient with CHARGE

syndrome and CHD7 mutation. *Endocr J,* 53**,** 741-3.

Ou, Z., P. Stankiewicz, Z. Xia, A. M. Breman, B. Dawson, J. Wiszniewska, P. Szafranski, M. L. Cooper, M. Rao, L. Shao, S. T. South, K. Coleman, P. M. Fernhoff, M. J. Deray, S. Rosengren, E. R. Roeder, V. B. Enciso, A. C. Chinault, A. Patel, S. H. Kang, C. A. Shaw, J. R. Lupski & S. W. Cheung (2011) Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. *Genome Res,* 21**,** 33-46.

Pagon, R., J. J. Graham, J. Zonana & S. Yong (1981) Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *J Pediatr,* 99**,** 223-7.

Pasillas, M. P., M. Shah & M. P. Kamps (2011) NSD1 PHD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human sotos syndrome. *Hum Mutat,* 32**,** 292-8.

Pauli, S., L. Pieper, J. Häberle, P. Grzmil, P. Burfeind, M. Steckel, U. Lenz & H. W. Michelmann (2009) Proven germline mosaicism in a father of two children with CHARGE syndrome. *Clin Genet,* 75**,** 473-9.

Pauli, S., N. von Velsen, P. Burfeind, M. Steckel, J. Mänz, A. Buchholz, W. Borozdin & J. Kohlhase (2012) CHD7 mutations causing CHARGE syndrome are predominantly of paternal origin. *Clin Genet,* 81**,** 234-9.

Perheentupa, J., S. Autio, S. Leisti, C.Raitta, L. Tuuteri (1973) Mulibrey nanism, an autosomal recessive syndrome with pericardial constriction. *Lancet,* 302:351-355 (orig. vol 2)

Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res,* 29**,** e45.

Pinkel, D., R. Segraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W. L. Kuo, C. Chen, Y. Zhai, S. H. Dairkee, B. M. Ljung, J. W. Gray & D. G. Albertson (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet,* 20**,** 207-11.

Pinto, G., V. Abadie, R. Mesnage, J. Blustajn, S. Cabrol, J. Amiel, L. Hertz-Pannier, A. M. Bertrand, S. Lyonnet, R. Rappaport & I. Netchine (2005) CHARGE syndrome includes hypogonadotropic hypogonadism and abnormal olfactory bulb development. *J Clin Endocrinol Metab,* 90**,** 5621-6.

Qiao, Q., Y. Li, Z. Chen, M. Wang, D. Reinberg & R. M. Xu (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem,* 286**,** 8361-8.

Rayasam, G. V., O. Wendling, P. O. Angrand, M. Mark, K. Niederreither, L. Song, T. Lerouge, G. L. Hager, P. Chambon & R. Losson (2003) NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J,* 22**,** 3153-63.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. González, M. Gratacòs, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer & M. E. Hurles (2006) Global variation in copy number in the human genome. *Nature,* 444**,** 444-54.

Reed, S. C. (1955) *Counseling in Medical Genetics*. Philadelphia and London: W.B. Saunders Co. 268 pp.

Rimoin, D. L. & K. Hirschhorn (2004) A history of medical genetics in pediatrics. *Pediatr Res,* 56**,** 150-9.

Rio, M., L. Clech, J. Amiel, L. Faivre, S. Lyonnet, M. Le Merrer, S. Odent, D. Lacombe, P. Edery, R. Brauner, O. Raoul, P. Gosset, M. Prieur, M. Vekemans, A. Munnich, L. Colleaux & V. Cormier-Daire (2003) Spectrum of NSD1 mutations in Sotos and Weaver syndromes. *J Med Genet,* 40**,** 436-40.

Riordan, J. R., J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic & J. L. Chou (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science,* 245**,** 1066-73.

Roohi, J., C. Montagna, D. H. Tegay, L. E. Palmer, C. DeVincent, J. C. Pomeroy, S. L. Christian, N. Nowak & E. Hatchwell (2009) Disruption of contactin 4 in three subjects with autism spectrum disorder. *J Med Genet,* 46**,** 176-82.

Royer-Pokora, B., L. M. Kunkel, A. P. Monaco, S. C. Goff, P. E. Newburger, R. L. Baehner, F. S. Cole, J. T. Curnutte & S. H. Orkin (1986) Cloning the gene for an inherited human disorder--chron-

ic granulomatous disease--on the basis of its chromosomal location. *Nature,* 322**,** 32-8.

Rozen, S. & H. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* 132**,** 365-86.

Sandberg, A. A., G. F. Koepf, T. Ishihara & T. S. Hauschka (1961) An XYY human male. *Lancet,* 2**,** 488-9.

Sanderson, S., R. Zimmern, M. Kroese, J. Higgins, C. Patch & J. Emery (2005) How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. *Genet Med,* 7**,** 495-500.

Sanger, F. & A. R. Coulson (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol,* 94**,** 441-8.

Sanka, M., N. Tangsinmankong, M. Loscalzo, J. W. Sleasman & M. J. Dorsey (2007) Complete DiGeorge syndrome associated with CHD7 mutation. *J Allergy Clin Immunol,* 120**,** 952-4.

Sanlaville, D., H. C. Etchevers, M. Gonzales, J. Martinovic, M. Clément-Ziza, A. L. Delezoide, M. C. Aubry, A. Pelet, S. Chemouny, C. Cruaud, S. Audollent, C. Esculpavit, G. Goudefroye, C. Ozilou, C. Fredouille, N. Joye, N. Morichon-Delvallez, Y. Dumez, J. Weissenbach, A. Munnich, J. Amiel, F. Encha-Razavi, S. Lyonnet, M. Vekemans & T. Attié-Bitach (2006) Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development. *J Med Genet,* 43**,** 211-217.

Sanlaville, D. & A. Verloes (2007) CHARGE syndrome: an update. *Eur J Hum Genet,* 15**,** 389-99.

Saugier-Veber, P., C. Bonnet, A. Afenjar, V. Drouin-Garraud, C. Coubes, S. Fehrenbach, M. Holder-Espinasse, J. Roume, V. Malan, M. F. Portnoi, N. Jeanne, C. Baumann, D. Héron, A. David, M. Gérard, D. Bonneau, D. Lacombe, V. Cormier-Daire, T. Billette de Villemeur, T. Frébourg & L. Bürglen (2007) Heterogeneity of NSD1 alterations in 116 patients with Sotos syndrome. *Hum Mutat,* 28**,** 1098-107.

Schouten, J. P., C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens & G. Pals (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res,* 30**,** e57.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Månér, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg & M. Wigler (2004) Large-scale copy number polymorphism in the human genome. *Science,* 305**,** 525-8.

Shaffer, L. G. & J. R. Lupski (2000) Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu Rev Genet,* 34**,** 297-329.

Shrimpton, A. E., K. A. Jensen & J. J. Hoo (2006) Karyotype-phenotype analysis and molecular delineation of a 3p26 deletion/8q24.3 duplication case with a virtually normal phenotype and mild cognitive deficit. *Am J Med Genet A,* 140**,** 388-91.

Shuib, S., D. McMullan, E. Rattenberry, R. M. Barber, F. Rahman, M. Zatyka, C. Chapman, F. Macdonald, F. Latif, V. Davison & E. R. Maher (2009) Microarray based analysis of 3p25-p26 deletions (3p- syndrome). *Am J Med Genet A,* 149A**,** 2099-105.

Simola, K. personal discussion. 20.2.2012

Sklower-Brooks S., J. Pappas, C. Duncan, M. Genovese, H. Gu & E.C. Jenkins (2002) Deletion 3p25 in a mother and child. *Am J Hum Genet* 71:301 (Abstract #751).

Sotos, J. F., P. R. Dodge, D. Muirhead, J. D. Crawford & N. B. Talbot (1964) Cerebral gigantism in childhood. A Syndrome of excessively rapid growth and acromegalic features and a nonprogressive neurologic disorder. *N Engl J Med,* 271**,** 109-16.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol,* 98**,** 503-17.

Stec, I., T. J. Wright, G. J. van Ommen, P. A. de Boer, A. van Haeringen, A. F. Moorman, M. R. Altherr & J. T. den Dunnen (1998) WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet,* 7**,** 1071-82.

Steele, M. W. & W. R. Breg (1966) Chromosome analysis of human amniotic-fluid cells. *Lancet,* 1**,** 383-5.

Strachan, T. & A. Reed (2010) *Human molecular genetics, 4th edition*. Garland Science, 781 pp.

Takagishi, J., K. A. Rauen, T. Drumheller, B. Kousseff & M. Sutcliffe (2006) Chromosome 3p25 deletion in mother and daughter with minimal phenotypic effect. *Am J Med Genet A,* 140**,** 1587-93.

Tatton-Brown, K., J. Douglas, K. Coleman, G. Baujat, T. R. Cole, S. Das, D. Horn, H. E. Hughes, I. K. Temple, F. Faravelli, D. Waggoner, S. Turkmen, V. Cormier-Daire, A. Irrthum, N. Rahman & C. O. Collaboration (2005) Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. *Am J Hum Genet,* 77**,** 193-204.

Tatton-Brown, K., S. Hanks, E. Ruark, A. Zachariou, S. e. V. Duarte, E. Ramsay, K. Snape, A. Murray, E. R. Perdeaux, S. Seal, C. Loveday, S. Banka, C. Clericuzio, F. Flinter, A. Magee, V. McConnell, M. Patton, W. Raith, J. Rankin, M. Splitt, V. Strenger, C. Taylor, P. Wheeler, K. I. Temple, T. Cole, J. Douglas, N. Rahman & C. O. Collaboration (2011) Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. *Oncotarget,* 2**,** 1127-33.

Tazelaar, J., J. Roberson, D. L. Van Dyke, V. R. Babu & L. Weiss (1991) Mother and son with deletion of 3p25-pter. *Am J Med Genet,* 39**,** 130-2.

Tjio, J. H. & A. Levan (1956) The chromosome number of man. *Hereditas***,** 1-6.

Tsongalis, G. J. & W. B. Coleman (2002) *Molecular diagnostics: a training and a study guide*. Washington D.C.: AACC Press. 374 pp.

Tsongalis, G. J. & L. M. Silverman (2006) Molecular diagnostics: a historical perspective. *Clin Chim Acta* 369(2):188-192

Tsurusaki, Y., N. Okamoto, H. Ohashi, T. Kosho, Y. Imai, Y. Hibi-Ko, T. Kaname, K. Naritomi, H. Kawame, K. Wakui, Y. Fukushima, T. Homma, M. Kato, Y. Hiraki, T. Yamagata, S. Yano, S. Mizuno, S. Sakazume, T. Ishii, T. Nagai, M. Shiina, K. Ogata, T. Ohta, N. Niikawa, S. Miyatake, I. Okada, T. Mizuguchi, H. Doi, H. Saitsu, N. Miyake & N. Matsumoto (2012) Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet,* 44**,** 376-8.

Türkmen, S., G. Gillessen-Kaesbach, P. Meinecke, B. Albrecht, L. M. Neumann, V. Hesse, S. Palanduz, S. Balg, F. Majewski, S. Fuchs, P. Zschieschang, M. Greiwe, K. Mennicke, F. R. Kreuz, H. J. Dehmel, B. Rodeck, J. Kunze, S. Tinschert, S. Mundlos & D. Horn (2003) Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. *Eur J Hum Genet,* 11**,** 858-65.

van Haelst, M. M., J. J. Hoogeboom, G. Baujat, H. T. Brüggenwirth, I. Van de Laar, K. Coleman, N. Rahman, M. F. Niermeijer, S. L. Drop & P. J. Scambler (2005) Familial gigantism caused by an NSD1 mutation. *Am J Med Genet A,* 139**,** 40-4.

Verjaal, M. & M. B. De Nef (1978) A patient with a partial deletion of the short arm of chromosome 3. *Am J Dis Child,* 132**,** 43-5.

Verloes, A. (2005) Updated diagnostic criteria for CHARGE syndrome: a proposal. *Am J Med Genet A,* 133**,** 306-8.

Visser, R., T. Hasegawa, N. Niikawa & N. Matsumoto (2006) Analysis of the NSD1 promoter region in patients with a Sotos syndrome phenotype. *J Hum Genet,* 51**,** 15-20.

Visser, R., O. Shimokawa, N. Harada, A. Kinoshita, T. Ohta, N. Niikawa & N. Matsumoto (2005a) Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. *Am J Hum Genet,* 76**,** 52-67.

Visser, R., O. Shimokawa, N. Harada, N. Niikawa & N. Matsumoto (2005b) Non-hotspot-related breakpoints of common deletions in Sotos syndrome are located within destabilised DNA regions. *J Med Genet,* 42**,** e66.

Vissers, L. E., C. M. van Ravenswaaij, R. Admiraal, J. A. Hurst, B. B. de Vries, I. M. Janssen, W. A. van der Vliet, E. H. Huys, P. J. de Jong, B. C. Hamel, E. F. Schoenmakers, H. G. Brunner, J. A. Veltman & A. G. van Kessel (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet,* 36**,** 955-7.

Weaver, D. D., C. B. Graham, I. T. Thomas & D. W. Smith (1974) A new overgrowth syndrome with accelerated skeletal maturation, unusual facies, and camptodactyly. *J Pediatr,* 84**,** 547-52.

Weiss, M. C. & H. Green (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc Natl Acad Sci U S A,* 58**,** 1104-11.

Wincent, J., E. Holmberg, K. Strömland, M. Soller, L. Mirzaei, T. Djureinovic, K. Robinson, B. Anderlid & J. Schoumans (2008) CHD7 mutation spectrum in 28 Swedish patients diagnosed with CHARGE syndrome. *Clin Genet,* 74**,** 31-8.

Wincent, J., A. Schulze & J. Schoumans (2009) Detection of CHD7 deletions by MLPA in

CHARGE syndrome patients with a less typical phenotype. *Eur J Med Genet,* 52**,** 271-2.

Witt, D. R., B. Biedermann & J. G. Hall (1985) Partial deletion of the short arm of chromosome 3 (3p25----3pter). Further delineation of the clinical phenotype. *Clin Genet,* 27**,** 402-7.

Woodage, T., M. Basrai, A. Baxevanis, P. Hieter & F. Collins (1997) Characterization of the CHD family of proteins. *Proc Natl Acad Sci U S A,* 94**,** 11472-7.

Zentner, G. E., W. S. Layman, D. M. Martin & P. C. Scacheri (2010) Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *Am J Med Genet A,* 152A**,** 674-86.