CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Helsingin yliopiston digitaalinen arkisto

MOLECULAR GENETICS OF THE GRACILE SYNDROME
(Growth Retardation, Aminoaciduria, Cholestasis, Iron overload, Lactacidosis and Early death)

Ilona Visapää

Department of Molecular Medicine, National Public Health Institute and
Department of Medical Genetics, University of Helsinki,
Helsinki, Finland,
and
Department of Human Genetics, David Geffen School of Medicine at UCLA,
Los Angeles, California, USA,
and
Hospital for Children and Adolescents, Helsinki University Central Hospital,
Helsinki, Finland.

Academic Dissertation

Helsinki University Biomedical Dissertations No. 19
Helsinki Biomedical Graduate School

To be publicly discussed with permission of the Medical Faculty of the University of Helsinki, in auditorium 2 of the Biomedicum Helsinki, Haartmaninkatu 8, on December 5th, 2002, at 12 noon.

Helsinki 2002

Supervised by

Professor Leena Peltonen-Palotie
National Public Health Institute
and University of Helsinki,
Helsinki, Finland
and
David Geffen School of Medicine at UCLA,
Los Angeles, California, USA

Reviewed by

Professor Pertti Aula
University of Helsinki
Helsinki, Finland

Docent Kirsi Huoponen
University of Turku
Turku, Finland

Copyright National Public Health Institute

Julkaisija – Utgivare – Publisher

Kansanterveyslaitos
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 47441, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 47441, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 47441, telefax +358 9 4744 8408

Publications of the National Public Health Institute, KTL A28/2002
ISBN 951-740-323-2 (Paperback)
ISBN 951-740-324-0 (PDF)
ISSN 0359-3584 (printed versions)
ISSN 1458-6290 (electronic versions)

Helsinki University Biomedical Dissertations No. 19
ISSN 1457-8433

Cosmoprint Oy
Helsinki 2002

CONTENTS

LIST OF ORIGINAL PUBLICATIONS **7**

ABBREVIATIONS **8**

SUMMARY **10**

REVIEW OF THE LITERATURE **12**

The GRACILE Syndrome **12**
Clinical Features 12
Histopathological findings 13
Iron metabolism 13
Mitochondrial investigations 14
Non-Finnish patients 14

The Human Genome Project **15**

Identification of human disease genes **18**

The Finnish disease heritage **21**
Founder effect 21
Genealogy of the GRACILE syndrome 23
Ancestral haplotypes and linkage disequilibrium 23
Molecular genetics of the Finnish disease heritage 25

Nuclear gene defects resulting in mitochondrial disease **29**

Mitochondrial functions **32**
Oxidative phosphorylation 32
Complex III of the respiratory chain and BCS1L 34
Formation of FeS clusters 35
ABCB7 and ABCB6, the human orthologs of yeast Atm1p 36
Other diseases with abnormalities of mitochondrial iron metabolism 37

AIMS OF THE STUDY **38**

SUBJECTS AND METHODS **39**

GRACILE families and patient samples **39**

Control samples **39**

DNA and RNA extraction **39**

Genotyping **39**

Linkage and LD analyses **40**

Radiation hybrid mapping **41**

Bioinformatics **41**

Sequencing and mutation screening **42**
Northern blot analysis **43**
Expression plasmid construction, cell culture, and transfections **43**
Immunofluorescence microscopy **43**
Western blot analysis **43**
Pulse chase analysis **44**
Yeast complementation studies **44**
Activity measurements of mitochondrial enzyme complexes **44**
RESULTS AND DISCUSSION **46**
Assignment of the GRACILE locus to chromosome 2 (I) **46**
Genome screen and linkage analysis 46
LD analyses and RH mapping of the critical region 47
Haplotype analysis 49
The mapping strategy and the importance of LD 50
Analysis of the positional candidate genes *ABCB6* and *BCS1L* (III, IV) **51**
Genes in the critical GRACILE region 51
Sequence analysis of ABCB6 and BCS1L 52
Northern blot analysis of ABCB6 and BCS1L 54
Positional exclusion of ABCB6 54
BCS1L sequence analysis in the British patients 54
Genomic structure of *BCS1L* (IV) **55**
Functional consequences of the S78G mutation on BCS1L (IV) **56**
Expression, targeting, and stability of the mutant BCS1L 56
Yeast complementation studies 58
Activity measurements of complex III in the Finnish GRACILE patients 58
Pathogenicity of the identified *BCS1L* mutations (IV) **61**
Phenotype heterogeneity of the *BCS1L* mutations (IV) **63**
DNA diagnostics of the GRACILE syndrome (I, II, IV) **67**
CONCLUDING REMARKS **68**
ACKNOWLEDGEMENTS **70**
ELECTRONIC DATABASE INFORMATION **73**
REFERENCES **74**

LIST OF ORIGINAL PUBLICATIONS

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

I Visapää I, Fellman V, Varilo T, Palotie A, Raivio KO, Peltonen L (1998) Assignment of the locus for a new lethal neonatal metabolic syndrome to 2q33-37. *American Journal of Human Genetics* 63:1396-1403

II Fellman V, Visapää I, Vujic M, Wennerholm UB, Peltonen L (2002) Antenatal diagnosis of hereditary fetal growth retardation with aminoaciduria, cholestasis, iron overload, and lactic acidosis in the newborn infant. *Acta Obstetricia et Gynecologica Scandinavica* 81:398-402

III Visapää I, Fellman V, Lanyi L, Peltonen L (2002) ABCB6 (MTABC3) excluded as the causative gene for the growth retardation syndrome with aminoaciduria, cholestasis, iron overload, and lactacidosis. *American Journal of Medical Genetics* 109:202-205

IV Visapää I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, Turnbull DM, Suomalainen A, Peltonen L (2002) GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in *BCS1L*. *American Journal of Human Genetics* 71:863-876

ABBREVIATIONS

AAA	ATPases associated with various cellular activities
ABCB6	human *ABCB6* gene
ABCB6	human ABCB6 protein
ABCB7	human *ABCB7* gene
ABCB7	human ABCB7 protein
ATM1	yeast *ATM1* gene
Atm1p	yeast Atm1p protein
ATP	adenosine triphosphate
bp	base pair
BAC	bacterial artificial chromosome
BCS1	yeast *BCS1* gene
BCS1L	human *BCS1L* (BCS1-like) gene
BCS1L	human BCS1L (BCS1-like) protein
BCS1p	yeast BCS1p protein
BLAST	basic local alignment search tool
cDNA	complementary DNA
cM	centiMorgan
COS-1 cells	African green monkey kidney cells
cR	centiRay
DNA	deoxyribonucleic acid
EST	expressed sequence tag
FeS	iron-sulfur
GRACILE	growth retardation, aminoaciduria, cholestasis, iron overload, lactacidosis and early death
HGP	Human Genome Project
kb	kilobase
IRE	iron response element
IRE-BP	IRE-binding protein
kD	kiloDalton
LD	linkage disequilibrium
LOD	logarithm of odds
Mb	megabase
mRNA	messenger RNA
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism

RH	radiation hybrid
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT	reverse transcriptase
SNP	single nucleotide polymorphism
tRNA	transfer RNA
θ	recombination fraction
UTR	untranslated region

In addition, the standard one-letter abbreviations of nucleotides and amino acids are used.

SUMMARY

The GRACILE syndrome is a metabolic disorder characterized by fetal growth retardation, lactic acidosis, iron overload, tubulopathy, cholestasis and death in early infancy. It has an autosomal recessive mode of inheritance and is part of the Finnish disease heritage. In Finland, the incidence is at least 1:47 000. In this study we localized and identified the gene underlying the GRACILE syndrome, which serves as a starting point for revealing the pathogenesis of this metabolic disease. This study has also produced DNA diagnostics for this fatal disease, enabling prenatal diagnosis and carrier testing for the Finnish GRACILE families.

The GRACILE gene was initially localized by a genome scan, using DNA samples from only nine patients, followed by conventional linkage analysis in eight nuclear families. The locus assignment to chromosome 2q33-37 was confirmed by LD analyses, and the critical markers were organized by RH mapping. Ancestral haplotype analysis revealed one founder haplotype shared by all GRACILE chromosomes and restricted the critical region between markers D2S2179 and D2S2244.

The physical length of the critical chromosomal region was 1.3 Mb, and it contained two interesting candidate genes, *ABCB6* and *BCS1L.* Lactic acidosis of the patients would suggest mitochondrial dysfunction, and both of these genes encoded proteins involved in mitochondrial respiratory function. In addition, *ABCB6* was known to be linked to mitochondrial iron metabolism and to maintenance of the stability of mitochondrial DNA. *ABCB6* was excluded as the causative gene for the GRACILE syndrome on the basis of the sequence, Northern blot and haplotype analyses. In *BCS1L*, however, all the Finnish GRACILE patients had a homozygous 232A>G mutation resulting in an S78G amino acid change. Sequencing of *BCS1L* in three British patients, whose symptoms closely resembled those of the Finnish patients, revealed five different mutations, including the Finnish S78G mutation. Pulse-chase experiments performed in COS-1 cells indicated that the S78G amino acid change results in instability of the polypeptide, and yeast complementation studies revealed a functional defect in the mutated BCS1L.

The function of Bcs1p, the yeast homolog of BCS1L, has been studied in detail. Bcs1p is an inner mitochondrial membrane protein and functions as a chaperone in the assembly of mitochondrial respiratory chain complex III. In humans, BCS1L has a related function, since both our British patients and previously described Turkish patients with *BCS1L* mutations unequivocally had a defect in complex III. In the Finnish GRACILE patients, however, the complex III

activities were normal or only slightly reduced. Presumably BCS1L has another, unknown but essential, cellular function, putatively involving iron metabolism.

REVIEW OF THE LITERATURE

The GRACILE Syndrome

GRACILE syndrome (OMIM 603358) is a lethal metabolic disorder with autosomal recessive inheritance. The acronym is composed of the first letters of the characteristic symptoms and findings: Growth Retardation, Aminoaciduria, Cholestasis, Iron overload, Lactic acidosis and Early death. The syndrome is the newest member of the Finnish disease heritage, a group of 36 monogenic diseases that are enriched in Finland (Norio et al. 1973; Norio 2000). The article by Norio et al. in 1973 mentioned five infants with fulminant lactic acidosis, the disease in these patients probably being the GRACILE syndrome. A comprehensive description of the disease entity, based on a series of 17 infants, was published by Fellman et al. in 1998. Since 1985, altogether 22 affected infants have been reliably diagnosed in 16 Finnish or Finnish-born families, and the male/female ratio of the patients has been 7/15. According to the 13 affected infants diagnosed in 1991-2000, the incidence of the GRACILE syndrome in Finland is at least 1/47,000.

Clinical Features

The growth of all the affected infants was severely retarded, the average birth weight being only 1700 g, corresponding to an SD score of -3.8 for gestational age. The average birth length was 43 cm, and head circumference 31 cm. Despite the extreme growth restriction, fetal distress was uncommon, and the patients were born near to or at term. At birth, the infants had a normal umbilical pH, and the first minute Apgar score was normal, but during the first day of life they developed severe lactic acidosis. On admission to intensive care, the mean arterial blood pH was 7.0 (normal range 7.35-7.43), the mean lactate 12.2 mmol/l (normal range 0.7-1.8), the mean pyruvate 120 μmol/l (normal range 40-70) and the mean lactate/pyruvate ratio 103 (normal <25). Despite intensive care and alkali therapy, half of the GRACILE patients died of fulminant acidosis between the ages of 1 and 12 days. The other half of the infants survived for 1-4 months. Their acidosis fluctuated, and their weight gain was poor (median 29 g/week). The patients did not have any dysmorphic features, but they had a similar worried-looking facial expression and wrinkled skin due to the absence of subcutaneous fat. (Fellman et al. 1998).

All the GRACILE patients had nonspecific Fanconi-type aminoaciduria. Low thrombo-test values and increased conjugated bilirubin and aminotransferase concentrations indicated mild to moderate hepatic dysfunction. The muscle tone was within the normal range, and no neurological symptoms or neuropathological abnormalities were found. Hemoglobin concentrations and reticulocyte counts

were normal, and no major cardiovascular, pulmonary or gastrointestinal problems were observed. (Fellman et al. 1998).

Histopathological findings

The pathological findings in 17 Finnish GRACILE patients have recently been described. The most prominent histopathological changes were present in the liver. In all except one case, the liver showed microscopic cholestasis. The extrahepatic bile ducts were open. The livers of the neonates were macroscopically unremarkable, but in most of them there was a paucity of interlobular bile ducts, which may be the reason for the cholestasis. In the patients who died at over 1 month of age, the livers showed a macroscopically green color and increased firmness, and the microscopic findings were steatosis and fibrosis. In most cases, the pancreas showed mild to moderate interstitial fibrosis and exocrine atrophy. Most of the patients had nephrocalcinosis, which is a relatively common finding in many disorders of infancy (Karlowicz and Adelman 1998). Tubular dysgenesis was observed in the kidneys of four patients, the amount of the proximal tubules being reduced to one-tenth of the controls. (Rapola et al. 2002).

Iron metabolism

One of the essential findings in the GRACILE syndrome is iron overload, including liver hemosiderosis, free plasma iron, and abnormal levels of proteins involved in iron transfer and storage (Fellman et al. 1998; Fellman et al. 2000). During the neonatal period, the livers of the affected infants contain extensive amounts of stainable iron both in the hepatocytes and in the Kupffer cells. After the age of one month, the stainable iron disappears or clearly decreases from the hepatocytes, but in the Kupffer cells moderate amounts of iron are still present. Electron microscopy of two liver biopsy samples revealed typical hemosiderin granules surrounded by numerous ferritin particles in the hepatocytes. Iron accumulation has not been detected in the parenchymal cells of any organ except the liver in the GRACILE patients, but iron-containing macrophages are typically frequent in the spleen, and mild or moderate amounts of siderotic macrophages have also been detected in the lymph nodes, thymus, lungs and pancreas. (Rapola et al. 2002).

GRACILE patients have free plasma iron and significantly decreased transferrin concentration in the serum, transferrin saturation being increased (Fellman et al. 1998; Fellman et al. 2000). The serum ferritin concentration is increased tenfold as compared to controls. It is not yet known whether the iron accumulation in this syndrome is the primary metabolic defect, or is secondary to the liver disease. The reduction in plasma transferrin concentration is unlikely to be the fundamental defect, since atransferrinemia and hypotransferrinemia typically

do not cause problems in newborns, the main finding later in life being anemia (Hayashi et al. 1993; Beutler et al. 2000).

GRACILE syndrome differs clearly from neonatal hemochromatosis, another lethal disorder characterized by liver hemosiderosis in the newborn period (Fellman et al. 1998; Rapola et al. 2002). In neonatal hemochromatosis, the hemosiderosis is abundant in several parenchymal organs in addition to the liver, but not in the reticuloendothelial cells (Knisely et al. 1987; Murray and Kowdley 2001). The liver failure is more fulminant than in the GRACILE patients, causing hypoalbuminemia, hypoglycemia, coagulopathy, ascites, hyperbilirubinemia and low transaminase levels in the newborn period. Instead, profound acidosis and severe fetal growth retardation are not typical of neonatal hemochromatosis. The inheritance pattern of neonatal hemochromatosis is not known. This rare disease has been reported in siblings with healthy parents, suggesting autosomal recessive inheritance, but also in half-siblings with the same mother, suggesting maternal inheritance (Verloes et al. 1996).

Two GRACILE patients have been treated by administrations of apotransferrin followed by exchange transfusions. The aim of the treatment was to reduce the toxic effects of free iron. Transferrin saturation and free plasma iron were decreased in both the treated infants, and no side effects were observed. However, the infants died later at the ages of ten and eight weeks. (Fellman et al. 2000).

Mitochondrial investigations

Lactic acidosis suggests mitochondrial dysfunction. The lactate/pyruvate ratio in the GRACILE patients is increased, and they do not have ketoacidosis or organic aciduria (Fellman et al. 1998). Thus the lactacidosis would be likely to result from a defect in the mitochondrial respiratory chain (Scriver et al. 2001a). However, no distinct abnormalities have been found in mitochondrial investigations of the GRACILE patients (Fellman et al. 1998). The polarographic oxygen consumption studies implied normal respiratory chain function, and measurements of the activity of the respiratory chain enzyme complexes revealed only a slight, nonspecific decrease in complex I. No deletion or depletion of mitochondrial DNA was found in the muscle biopsy samples analyzed by Southern blots. The mitochondria also appeared normal in both the liver and muscle specimens analyzed by electron microscopy (Rapola et al. 2002).

Non-Finnish patients

GRACILE patients with precisely the same symptoms and findings as described in the Finnish patients have not been reported elsewhere so far. However, three British patients have been described with a disease course closely resembling

that of the Finnish infants (Birch-Machin et al. 1989; Morris et al. 1995). The infants had moderate growth retardation, lactic acidosis, aminoaciduria and cholestasis, and died at the ages of 2, 42 and 105 days. However, there were a few differences from the Finnish disease entity. The British patients had muscle hypotonia, and one of them had seizures. In addition, a clear complex III defect was demonstrated in the enzyme activity measurements of the skeletal muscle mitochondria of the patients. Iron metabolism was not studied in the British cases.

The Human Genome Project

The Human Genome Project (HGP) is an international collaborative research program whose goal is the complete mapping and understanding of the human genes and genome. It is led by the National Human Genome Research Institute (NHGRI), which was established in 1989 at the National Institutes of Health (NIH) in USA. Other contributors are the United States Department of Energy (DOE) and numerous universities and other research facilities in the USA, the United Kingdom, France, Germany, Japan and China. The Human Genome Organization (HUGO), established in 1988, is an international organization coordinating the different national projects in order that the analysis of the human genome may be achieved as rapidly and effectively as possible. HUGO has also published several statements concerning the ethics of genetic research, for example concerning the collection of DNA samples, patenting of DNA sequences, and cloning. The HUGO Gene Nomenclature Committee approves and designates the symbols and names for novel genes. The goals for the HGP, initially defined in 1990, were: 1) mapping and sequencing the human genome, 2) mapping and sequencing the genomes of model organisms, 3) data collection and distribution, 4) ethical and legal considerations, 5) research training, 6) technology development, and 7) technology transfer. (NHGRI and HUGO web pages; Strachan and Read 1999a). (Internet addresses of the organizations and databases mentioned in this chapter are provided in the electronic database information section.)

The important achievements of HGP in the early 90's included production of dense genetic maps based on microsatellite markers, which were considerably less laborious to genotype than the RFLP markers used earlier (Weissenbach et al. 1992; Dib et al. 1996), and construction of the first physical maps covering the human genome (Cohen et al. 1993; Hudson et al. 1995). The main emphasis of the sequencing projects of the HGP was at first on the coding sequences, rather than on the genomic DNA. The cDNA libraries of different human tissues were randomly sequenced to produce expressed sequence tag (EST) sequences, which were subsequently mapped to the physical maps. By 1998, 42 000 ESTs representing 30 000 human transcripts were mapped to the human

genome by the international RH mapping consortium (Deloukas et al. 1998), and the goals for years 1998-2003 were to produce full-length cDNA sequences of the human genes, as well as the complete human genomic sequence (NHGRI web pages).

February 15th, 2001 was an important day in the history of the HGP. On that day the article "Initial sequencing and analysis of the human genome" was published in Nature by the International Human Genome Sequencing Consortium. The sequence was not in its final form: the coverage of the draft sequence was approximately 94% of the 3200 Mb human genome, and approximately 26% of the genomic sequence was in its finished form, including two completely sequenced chromosomes, numbers 21 and 22. The sequencing of the human genome had progressed faster than anticipated: over roughly fifteen months in 1999-2000 the coverage of the sequence had increased from 10% to more than 90%. The sequencing strategy used by HGP was "hierarchical shotgun sequencing". The genomic DNA of anonymous individuals was cloned as 100-200 kb fragments, mostly using bacterial artificial chromosome (BAC) vectors. BAC clones were mapped and organized to large clone contigs, and individual clones were selected for shotgun sequencing. Sequence fragments were then assembled to reconstruct the sequence of the whole genome. Since the BAC clone contigs were positioned along the chromosomes by anchoring them with markers from existing genetic and physical maps, gross errors in the sequence assembly were not likely to occur despite the many repetitive sequences (at least 50%) in the human genome. (International Human Genome Sequencing Consortium 2001).

At the same time as the initial analysis of the human genome was published by HGP in Nature, Science published the article "The Sequence of the Human Genome" by Venter et al. This sequence was produced by Celera, a commercial company, and the sequence has not been freely available even after the publication of the article, but the company sells licences to use the sequence database. Celera was able to produce the raw sequence data even faster than the HGP laboratories by the "whole genome shotgun sequencing" approach. The whole genome was sequenced randomly and assembled automatically, without laborious mapping of the genomic clones (Venter et al. 2001). However, Celera was able to utilize the sequencing and mapping results produced by HGP, since academic, publicly funded institutions of HGP daily released all the sequences produced to freely accessible internet databases. The competition about the initial sequencing of the human genome between HGP and Celera finally led to a joint press conference on June 26th 2000, when they both claimed to have sequenced 90% of the human genome, and to an agreement to publish the first analysis of the sequence simultaneously (NHGRI web page).

The biggest surprise in the analysis of the human genome was that it was estimated to contain only 30 000 – 40 000 protein-coding genes (previous estimates being near 100 000), which is only about twice as many as in the nematode worm *Caenorhabditis elegans* or the fruitfly *Drosophila melanogaster.* However, the full set of proteins encoded by the human genome is more complex than those of invertebrates. First, the genes are more complex, with more alternative splicing generating a larger number of protein products. Secondly, the protein domains and motifs are more innovatively combined in vertebrates, as compared with invertebrates. (International Human Genome Sequencing Consortium 2001).

Besides the sequencing of the human genome, much has been accomplished in the era of the sequencing of other organisms The sequencing of the first eucaryotic genome, the budding yeast *Saccharomyces cerevisiae* (genome size 12 Mb), was completed in 1996, and the genomes of *Caenorhabditis elegans* (97 Mb) and *Drosophila melanogaster* (120 Mb) were sequenced in 1998 and 2000, respectively (Goffeau et al. 1996; The C. elegans Sequencing Consortium 1998; Adams et al. 2000). Mapping of the mouse genome was performed with very much the same strategy as mapping of the human genome, but with every step following a couple of years later. The sequencing of the mouse genome is in progress (NCBI and Celera web pages). The genomes of the model organisms serve as an outstanding tool for the recognition and comparison of the genes and their regulatory elements, as well as the protein domains. They also provide tools for functional analyses of the genes and proteins. (Strachan and Read 1999a; International Human Genome Sequencing Consortium 2001).

The development of the laboratory methods used in molecular biology has also been very rapid during the last ten years. Although the basic method used in DNA sequencing is still the dideoxy sequencing approach, invented 25 years ago (Sanger et al. 1977), the development of the automated fluorescence-based systems and capillary electrophoresis has facilitated substantially both sequencing and genotyping. Besides stimulating the rapid development of laboratory techniques, the HGP has produced well organized databases for gene, genome and protein sequences (e.g. at NCBI, UCSC, and EBI) and numerous bioinformatics tools for processing of the vast amounts of sequence data.

Publication of the draft human genome sequence was only the starting point for understanding the function of the human genome, although the HGP has already produced several applications in medicine. The next goals are to close the gaps and produce a finished sequence covering all the chromosomes, to obtain a comprehensive collection of full-length human cDNA:s, and to identify the regulatory elements in the genome. More than two million single nucleotide polymorphisms (SNP), which play a crucial role in the identification of the genes responsible for complex diseases, have already been identified in the human

genome (dbSNP at NCBI web pages). The production of a more comprehensive SNP catalogue of the human genome, and further development of the array-based technologies for automated SNP genotyping are also challenges for the next few years. The ultimate goal is to move on to functional analyses of the proteins encoded by the genes (functional genomics and proteomics). (International Human Genome Sequencing Consortium 2001).

Identification of human disease genes

There are several ways to identify genes underlying diseases with mendelian inheritance. Usually the disease gene identification process is a mixture of several strategies, but the different approaches can be classified to four main categories: functional cloning, candidate gene approach, positional cloning and positional candidate approach (figure 1).

Functional cloning is the oldest method for disease gene identification. It was the only way to clone disease genes twenty years ago, when information about the human genome mapping was not available. Functional cloning is based on information about the basic biochemical defect causing the disease. The defective protein is purified, its amino acid sequence determined, and the cDNA of the disease gene is identified using oligonucleotides produced as a result of the amino acid sequence information. For example, identification of the hemophilia A gene followed this approach (Gitschier et al. 1984).

The candidate gene approach also requires some knowledge of the pathogenesis of the disease or, for example, an animal model for the disease. This strategy means hypothesizing the right gene based on its known function and the symptoms of the disease. The gene can be tested directly for mutations in the patients, or for linkage in families. For example a few rare mutations resulting in defects in mitochondrial enzymes have been identified using this strategy (Bourgeron et al. 1994; Bourgeron et al. 1995; de Lonlay et al. 2001). So far, however, successes based solely on the candidate gene approach, without any clue to the location of the disease gene, have been rare. Maybe in the future, when all the human genes and their functions are known, this strategy will be more generally used.

Development of the genetic and physical maps of the human genome enabled positional cloning. In this strategy the disease genes are identified through knowledge of their chromosomal location. If sufficient family material is available, the disease loci of Mendelian disorders can be identified with genetic mapping, and restricted with linkage disequilibrium or shared haplotype analyses in suitable populations. In some diseases, in which a clear functional defect is present in the patients' cell lines, it has been possible to localize the disease gene by functional

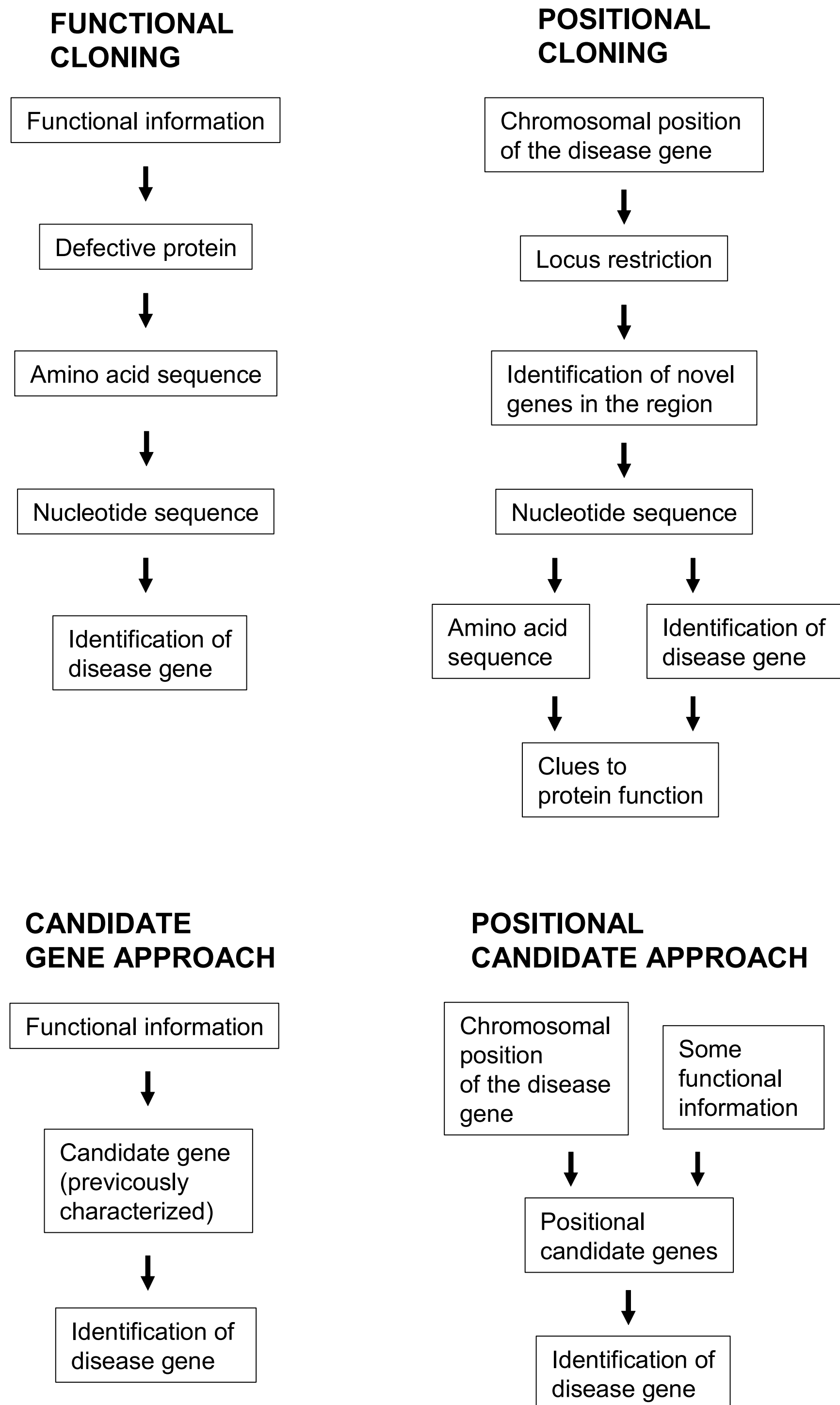


Figure 1
The principles of the four main strategies for identification of disease genes.

complementation, using microcell-mediated chromosome transfer, even though sufficient family material for linkage-based mapping has not been available (Cuthbert et al. 1995; Zhu et al. 1998; Seyda et al. 2001). Sometimes patients' chromosomal rearrangements provide useful short-cuts for locating a disease gene. Positional cloning of a disease gene, followed by characterization of the defective protein, may be the first step in the process to reveal the molecular pathogenesis of a genetic disorder. "Pure" positional cloning has traditionally meant identification of novel genes by different laboratory methods from genomic clone contigs (physical mapping). At the end of the 80s and in the beginning of the 90s, this laborious approach was used to identify dozens of disease genes, for example the causative genes for Duchenne muscular dystrophy and cystic fibrosis (Worton and Thompson 1988; Rommens et al. 1989). During the 90s, the positional cloning methods developed rapidly, and the number of disease genes identified on the basis of their position in the genome grew exponentially (Collins 1995).

The gene maps and sequence information produced by the HGP during recent years have gradually made the traditional physical mapping almost unnecessary. Nowadays, the predominant method of identifying disease genes is the positional candidate approach, a combination of the candidate gene approach and positional cloning. After assignment of the disease locus, it is usually unnecessary to identify novel genes in the region anymore. The sequences of positional candidate genes are available in the databases, and promising genes can be selected on the basis of their known function. If the exact function of a positional candidate gene or a predicted transcript is not known, information about the expression pattern of the gene, or about interesting homologies with other genes can be used as a guide to choosing suitable candidates. Especially amino acid sequence homologies to relevant proteins in model organisms can provide crucial clues to the identification of disease genes. Today also, positional cloning of disease genes can lead to the characterization of novel genes, but the cloning process, if needed, is nowadays carried out mostly "in silico", being enormously much faster than the laboratory methods used five or fifteen years ago. (Strachan and Read 1999b).

The number of identified disease genes is growing exponentially. As of October 1, 2001, the OMIM database contained 1,631 disorders in which the molecular basis had been identified at the DNA level. The total number of entries (genes, gene loci, allelic variants and phenotype descriptions of mendelian disorders) was 13,005. (Hamosh et al. 2002).

The Finnish disease heritage

The concept of the Finnish disease heritage was introduced some 30 years ago in an article entitled "Hereditary diseases in Finland; rare flora in rare soil", which described some twenty inherited disorders occurring in Finland much more frequently than elsewhere (Norio et al. 1973). Nowadays, the number of diseases belonging to this group has doubled (table 1). Most of the diseases also occur elsewhere in the world, but their incidences are especially high in Finland due to founder mutations enriched in the population. Unravelling of the molecular defects underlying the Finnish disease heritage has proceeded very rapidly during the last decade. Circumstances in Finland are very advantageous for genetic research: The quality of the health care system is high, which has made exact clinical diagnoses possible even for very rare conditions. The patients and their families usually have a positive attitude towards research. Furthermore, comprehensive church registers provide excellent possibilities for genealogical studies. However, probably the most important factor which has led to the rapid identification of the disease genes is the genetic structure of the Finnish population. (de la Chapelle and Wright 1998; Peltonen et al. 1999; Norio 2000).

Founder effect

The population of Finland is relatively young, and, for geographic and linguistic reasons, it has long remained rather isolated. There are several theories about the origin of the Finnish population. The recent analyses of Y chromosome haplotypes (Kittles et al. 1998), support a dual theory: The first settlers of Finland were Uralic speakers and arrived from the east some 4000 years ago. However, the majority of the genes in today's Finnish population originate from small founder populations of Indo-European speakers who arrived from the south in the beginning of the first millennium. These founder populations expanded and spread along the coastlines of southern and western Finland. In the beginning of the 16th century, the Finnish population comprised approximately 250 000 inhabitants, who were concentrated in the coastal regions. Inhabitation of the inland area of eastern, central and northern Finland began in the 16th century, and for that reason this area is called by geneticists the late-settlement region (figure 2A). A substantial part of this migration began from the small south-eastern area of South Savo. Inhabitation of the late-settlement region formed small, rural communities, which remained relatively stable and isolated because of the long distances and low population density. After the great famine and epidemics, which killed approximately one third of the population between the years 1690-1730, the Finnish population has expanded rapidly, growing from approximately 250 000 inhabitants to today's number of 5.2 million during three centuries. Since World War II, industrialization

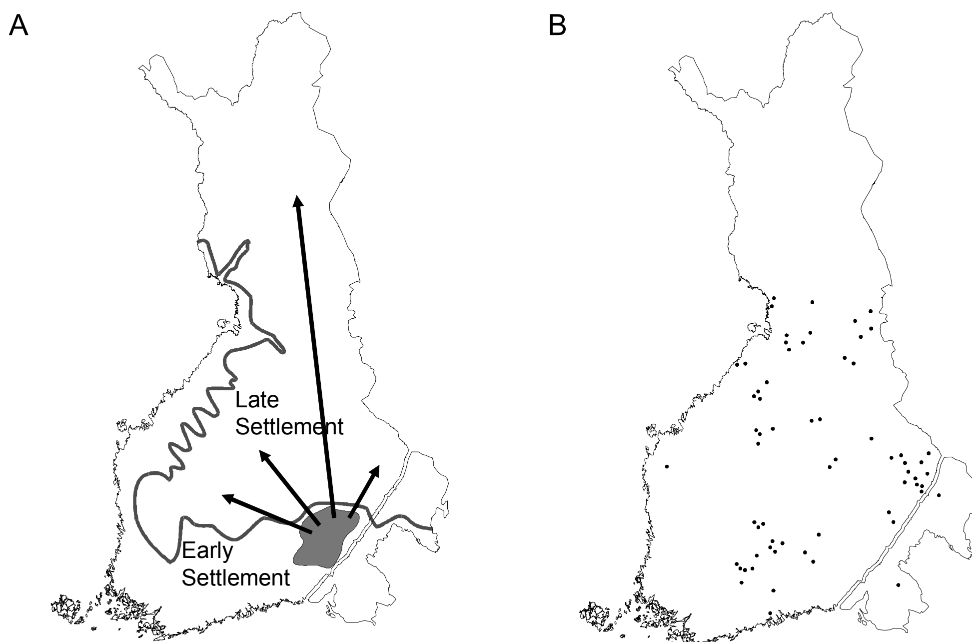


Figure 2
A) The internal migration movement of the 16th century from South Savo to the late settlement region of Finland, resulting in regional subisolates. B) Birthplaces of the great-grandparents of the GRACILE patients on the map of Finland. Modified from Varilo 1999.

has led to migration of the population from rural areas to urban communities, but there is still hardly any mixture between the subpopulations. (Norio 1981; de la Chapelle 1993; Peltonen et al. 1999; Varilo 1999; Norio 2000).

The population history of Finland explains why some diseases are enriched in Finland, whereas others, for example cystic fibrosis and phenylketonuria, are very rare as compared with other Caucasian populations. Expansion of an isolated population after a population bottleneck can cause remarkable changes in allele frequencies (compared with the situation in larger, constantly mixing populations), because any mutation present in the founder population can either be enriched or practically disappear. If the mutation does not affect reproduction – and the carrier state of a mutation causing an autosomal recessive trait usually does not – the allele frequencies depend on chance. These phenomena are called the founder effect and genetic drift.

In Finland there have been population bottlenecks both at the level of the population as a whole and on a smaller scale, especially in rural communities in the

sparsely populated areas of central, eastern and northern Finland. In many different Finnish diseases the distributions of the ancestors' birthplaces reflect the population movements in Finland (Varilo 1999; Norio 2000). Approximately half of the diseases belonging to the Finnish disease heritage have a typical pattern, the ancestors being distributed in the late-settlement region, but hardly at all in the coastal regions (Norio 2000). The mutations in this group were most probably spread by the internal migration in the 16th century, and it has been estimated that most of these mutations were introduced into the Finnish population some 30-50 generations ago (Varilo 1999). The second largest group includes six diseases: AGU, INCL, progressive myoclonus epilepsy, cartilage-hair hypoplasia, Batten disease and congenital nephrosis (Norio 2000). These are the most common disorders of the Finnish disease heritage, their incidences varying from 1:8000 to 1:19 000. In these six diseases, the ancestors' birthplaces are distributed fairly evenly across the whole country, but the late-settlement region is somewhat overrepresented as compared to the population density. The mutations causing these diseases had probably started to spread before the inhabitation of the inland regions, and it has been estimated that these mutations were introduced into the Finnish population some 70-120 generations ago (Peltonen et al. 1995; Varilo 1999). Only on the maps of two diseases, diastrophic dysplasia and the Meckel syndrome, does the distribution of the ancestors correspond to the population density of Finland, covering the whole country, and being concentrated in southern and western Finland (Norio 2000). These are relatively common disorders, and the major mutations are most likely very old. Variant late infantile ceroid-lipofuscinosis and progressive epilepsy with mental retardation represent the newest mutations, the ancestors' birthplace pattern showing tight clustering in small areas (Norio 2000).

Genealogy of the GRACILE syndrome

The GRACILE syndrome is a typical example of a disease spread by the migration of the 16th century to the late-settlement region of Finland (figure 2B). The pedigrees of the affected families have been traced back to the mid-19th century, and the oldest to the late 17th century. No tight clustering of the ancestors' birthplaces appears on the map of Finland, but almost all the ancestors were born in eastern or central Finland. In one family the parents were second cousins, and two earlier links between different nuclear families were established. (Fellman et al. 1998).

Ancestral haplotypes and linkage disequilibrium

In all the cloned Finnish disease genes, one major mutation is present in at least 70% of the disease chromosomes, and in most diseases one founder

mutation accounts for over 90% of the disease alleles (Peltonen et al. 1999; Varilo 1999). Thus, identical DNA fragments, inherited from a single common ancestor, are to be found in the disease chromosomes of most patients (figure 3). Each fragment contains not only the disease mutation, but also a unique allele combination of the genetic markers in the immediate vicinity of the disease gene. This allele combination is called the ancestral haplotype. Ancestral haplotype analysis means restriction of the critical chromosomal region for the gene hunt by determining the chromosomal interval, which is completely conserved in the disease chromosomes. The deviation of the marker allele frequencies in the disease chromosomes, compared to the general population, is called linkage disequilibrium (LD), or allelic association. Mathematical models have been developed to test the significance of LD in disease alleles (Terwilliger and Ott 1994; Terwilliger 1995).

Figure 3
The basis for ancestral haplotype and LD analyses. A founder chromosome (black) with a disease mutation alters with time because of recombinations in successive meioses. After many generations, only a short chromosomal segment in the immediate vicinity of the disease gene represents the original founder chromosome in a descendant. Identification of the conserved region in the disease chromosomes of the patients through use of ancestral haplotype and LD analyses indicates the chromosomal location of the disease gene.

It is also possible to estimate the age of the mutation and the distance between the disease gene and a certain marker, using LD-based statistical tools (Hästbacka et al. 1992; Lehesjoki et al. 1993).

Utilization of LD and haplotype analyses has greatly enhanced the positional cloning projects of the Finnish disease genes, since these methods enable restriction of the critical chromosomal region to a very small interval. The age of the Finnish population is ideal for these analyses: young enough for the LD to be detected with a reasonable marker density, but old enough for the ancestral recombinations to restrict the critical region. Usually the conserved haplotype present in all the disease alleles spans some hundreds of kilobases, and the chromosomal interval, where a significant LD ($p<0.0001$) can be demonstrated in two-point analyses, is roughly comparable to this region. However, a suggestive LD ($p<0.05$) can be detected over long chromosomal intervals, up to 13 cM. The distances demonstrating LD in the different Finnish diseases have been observed to be in good accord with the results of genealogical studies: the younger mutations show longer chromosomal intervals with LD than the older ones. (Peltonen et al. 1995; de la Chapelle and Wright 1998; Peltonen et al. 1999; Varilo 1999).

Molecular genetics of the Finnish disease heritage

The majority of the disease genes causing the Finnish diseases have been identified during the last decade, and at least the gene locus has been assigned for almost every disorder. Table 1 lists the diseases belonging to the Finnish disease heritage, the identified gene defects and gene loci, and the references. The methods used for identification of the disease genes have changed over the decade. In the beginning of the 90s, the predominating method was functional cloning and, for example, disease mutations resulting in gyrate atrophy and AGU were identified through knowledge of the defective protein. During the 90s, "pure" positional cloning, meaning identification of totally novel genes, was performed for example in the cloning of the genes responsible for diastrophic dysplasia and APECED. The strategy used in most projects has been a mixture of positional cloning and positional candidate approaches. Nowadays the positional candidate approach is practically the only method used.

Identification of the disease loci and mutations has made DNA diagnostics possible in most of the diseases belonging to the Finnish disease heritage. Rapid, reliable and specific diagnoses aid the handling of the patients, and the possibility of prenatal diagnostics has been a relief for many families suffering from the most devastating disorders. However, identification of the causative mutation is only the starting point of a process which aims at resolving the pathogenesis of a disease, and finally developing treatment. Functional studies of the defective proteins in the Finnish diseases have markedly contributed to cell

Table 1
Molecular genetics of the Finnish disease heritage. Included are those 36 diseases which Professor Reijo Norio listed as belonging to this entity in 2000 (Norio 2000). Almost all of these disorders are inherited by an autosomal recessive inheritance, with four exceptions: amyloidosis V and tibial muscle dystrophy are autosomal dominant traits, and choroideremia and retinoschisis are X-linked recessive traits.

IDENTIFIED GENE DEFECTS

Disease [OMIM number]	Defective protein	Reference (Finnish mutations)
Gyrate atrophy (GA) [258870]	Ornithine aminotransferase (OAT)	Mitchell at al. 1989
Amyloidosis V [105120]	Gelsolin (GSN)	Levy et al. 1990, Maury et al. 1990
Aspartylglucosaminuria (AGU) [208400]	Aspartylglucosaminidase (AGA)	Ikonen et al. 1991
Choroideremia [303100]	Rab escort protein 1 (REP1)	Sankila et al. 1992
Nonketotic hyperglycinemia (NKH) [605899]	Glycine decarboxylase (GLDC)	Kure et al. 1992
Diastrophic dysplasia [222600]	Solute carrier family 26, member 2 (SLC26A2)	Hästbacka et al. 1994
Infantile neuronal ceroid-lipofuscinosis (INCL) [256730]	Palmitoyl protein thioesterase 1	Vesa et al.1995
Batten disease [204200]	CLN3 protein	The international Batten disease consortium 1995
Hypergonadotrophic ovarian dysgenesis (ODG1) [233300]	FSH receptor (FSHR)	Aittomäki et al. 1995
Congenital chloride diarrhea [214700]	Solute carrier family 26, member 3 (SLC26A3)	Höglund et al. 1996
Progressive myoclonus epilepsy (EPM1) [254800]	Cystatin B (CSTB)	Pennacchio et al. 1996, Virtaneva et al. 1997
APECED (autoimmune polyendocrinopathy – candidiasis – ectodermal dystrophy) [240300]	Autoimmune regulator (AIRE)	Nagamine et al. 1997, The Finnish German APECED Consortium 1997
Congenital nephrosis [256300]	Nephrin	Kestilä et al. 1998
Variant late infantile neuronal ceroid-lipofuscinosis (vLINCL) [256731]	CLN5 protein	Savukoski et al. 1998
Retinoschisis [312700]	Retinoschisin	The retinoschisis consortium 1998

Lysinuric protein intolerance (LPI) [222700]	Solute carrier family 7, member 7 (SLC7A7)	Torrents et al. 1999, Borsani et al. 1999
Megaloblastic anemia 1 [261100]	Cubilin (CUBN)	Aminoff et al.1999
Progressive epilepsy with mental retardation (EPMR) [600143]	CLN8 protein	Ranta et al. 1999
Free sialic acid storage disease (Salla disease) [604369]	Solute carrier family 17, member 5 (SLC17A5)	Verheijen et al. 1999
Cornea plana congenita (CNA2) [217300]	Keratocan (KERA)	Pellegata et al. 2000
Mulibrey nanism [253250]	TRIM37	Avela et al. 2000
Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (PLOSL) [221770]	TYRO protein tyrosine kinase-binding protein (TYROBP) and TREM2	Paloneva et al. 2000, Paloneva et al. 2002
Cartilage-hair hypoplasia (CHH) [250250]	RNA component of RNase MRP (RMRP)	Ridanpää et al. 2001
Usher syndrome, type III [276902]	USH3A	Joensuu et al. 2001
Tibial muscle dystrophy (TMD) [600334]	Titin (TTN)	Hackman et al. 2002
GRACILE syndrome [603358]	BCS1-like (BCS1L)	This study

MAPPED LOCI

Disease [OMIM number]	Locus	Reference
Cohen syndrome [216550]	8q	Tahvanainen et al. 1994
Infantile onset spinocerebellar ataxia (IOSCA) [271245]	10q	Nikali et al. 1995
Meckel syndrome [249000]	17q	Paavola et al. 1995
Lethal congenital contracture syndrome (LCCS) [253310]	9q	Mäkelä-Bengs at al. 1998
Congenital lactase deficiency [223000]	2q	Järvelä et al. 1998
Muscle-eye-brain disease (MEB) [253280]	1p	Cormand et al. 1999
Hydrolethalus syndrome [236680]	11q	Visapää et al. 1999

DISEASE LOCUS NOT YET IDENTIFIED

Disease [OMIM number]	Reference (clinical characterization of the disease)
RAPADILINO syndrome [266280]	Kääriäinen et al. 1989
PEHO syndrome [260565]	Salonen et al. 1991
Lethal arthrogyposis with anterior horn cell disease	Vuopala et al. 1995

biology. Unravelling of the molecular pathology in rare diseases can also lead to a better understanding of the pathogenetic mechanisms in the common diseases. For example, research on AIRE, the protein underlying APECED, can lead to discoveries in the field of autoimmune disorders in general, and functional studies of the *CLN8* gene, mutated in progressive epilepsy with mental retardation, will most likely provide some clues to the molecular pathology of epilepsy. Two recent findings also provide interesting examples of the genetic mechanisms of Mendelian disorders: the gene mutated in cartilage-hair hypoplasia was discovered to encode a structural RNA molecule, instead of a protein, and the two genes underlying PLOSL, located in different chromosomes but resulting in identical phenotypes, have been found to encode proteins involved in the same cellular signaling complex.

Nowadays, prenatal diagnosis is possible in most of the severe Finnish diseases. Families with index cases are offered genetic counselling and the possibility of antenatal diagnostics. However, most of the children with autosomal recessive diseases are born to parents who are unaware of their carrier state. Since the majority of the disease mutations causing the Finnish diseases have now been identified, and the array-based technologies have made identification of point mutations relatively cheap when performed in large numbers, screening of these mutations on the population level will soon be technically possible. One DNA array has already been developed and successfully used to determine the carrier frequencies of several diseases in different Finnish subpopulations (Pastinen et al. 2001). In a questionnaire study evaluating the attitudes toward genetic testing among the general population in Finland, as well as among the relatives of the AGU patients, the attitudes were found to be favorable in general, but especially among the family members of the AGU patients (Hietala et al. 1995). In a pilot study, where carrier screening for the AGU mutations was offered to expectant couples in the framework of routine maternity health care in Helsinki, the participation rate was 95%, and the attitudes toward testing were mainly positive both before and after the gene test (Hietala 1998). Carrier screenings of fragile-X syndrome, AGU, INCL, and congenital nephrosis were offered to expectant couples in the pilot studies in Kuopio, where the participation rate was approximately 90% in all studies (Kallinen 2002). Thus, it seems that people in Finland mainly have a positive attitude toward the carrier screening of severe diseases.

However, all medical screening programs may cause anxiety in some participants, even in those who are found to be negative, and thus it is not self-evident that carrier screenings for very rare genetic diseases are justified. Combined screening for several mutations, using a DNA array, would especially be ethically problematic. On family-based screening, the genetic tests are always

offered to patients or their relatives with thorough genetic counselling. It would be difficult to offer individual counselling covering every disease included in a screening "package" to everyone participating in a population-based screening. However, everyone participating should be aware of what kind of decision-making might lie ahead during pregnancies, if both parents proved to be carriers of the same mutation. Ideally, the screenings should be arranged before pregnancy, so that couples who do not want an abortion would have the possibility to consider other options, for example in vitro fertilization with pre-implantation diagnosis, or adoption. Furthermore, numerous people who would receive carrier diagnoses but would not be at risk of having affected children, would need genetic counselling to avoid unnecessary anxiety and worries. (Hietala 1998; Hiilesmaa and Salonen 2000; Kere 2000; Kääriäinen 2000).

Nuclear gene defects resulting in mitochondrial disease

Mitochondrial diseases are a heterogeneous group of disorders, often characterized by morphological changes in the mitochondria, a defective respiratory chain and variable symptoms, ranging from severe metabolic disorders with onset in early infancy or childhood to late-onset adult myopathies. Mutations in mitochondrial DNA (mtDNA) are the most frequent cause of mitochondrial disease in adults, and during the last decade more than 100 pathogenic mutations have been identified in mtDNA. However, the mtDNA encodes only 13 of the 82 subunits of the different complexes of the respiratory chain, as well as two rRNA molecules, and 22 tRNA molecules. All the other mitochondrial proteins are encoded by nuclear genes, and, in pediatric patients, most mitochondrial disorders are caused by mutations in nuclear genes. During recent years, an increasing number of nuclear mutations causing mitochondrial diseases have been identified. The symptoms of these diseases vary, depending on which mitochondrial functions are disturbed, and not all of them cause morphological changes in the mitochondria. Most of the nuclear gene defects causing mitochondrial diseases are inherited as autosomal recessive traits, but some are autosomal dominant traits, and some are X-linked. (Shoubridge 2001; Zeviani 2001).

Table 2 lists the identified nuclear gene defects affecting the mitochondrial respiratory chain. These are mutations in genes encoding either the subunits of the respiratory chain enzyme complexes, the chaperones needed for assembly of the complexes, or the proteins involved in maintaining the stability of the mtDNA. The symptoms caused by these mutations are predominantly neurological. A typical example is Leigh disease, which is a fatal early-onset neurodegenerative disorder, characterized clinically by psychomotor retardation

Table 2
Diseases caused by nuclear gene defects affecting the mitochondrial respiratory chain, according to the reviews of Shoubridge 2001, Zeviani 2001, and Suomalainen and Kaukonen 2001.

DISEASE	DEFECTIVE PROTEIN
Defects in genes encoding structural components of the respiratory chain complexes	
Leigh syndrome or leukodystrophy and myoclonic epilepsy	NDUFV1 (Complex I)
Leigh syndrome	NDUFS1 (Complex I)
Cardiomyopathy and encephalomyopathy	NDUFS2 (Complex I)
Leigh syndrome	NDUFS4 (Complex I)
Leigh syndrome	NDUFS7 (Complex I)
Leigh syndrome	NDUFS8 (Complex I)
Leigh syndrome or late-onset optic atrophy, ataxia and myopathy	SDHA (Complex II)
Pheochromocytoma	SDHB (Complex II)
Hereditary paraganglioma	SDHC (Complex II)
Hereditary paraganglioma	SDHD (Complex II)
Defects in genes encoding assembly factors of the respiratory chain complexes	
Tubulopathy, hepatopathy and encephalopathy	BCS1L (Complex III)
Leigh syndrome	SURF1 (Complex IV)
Hepatic encephalopathy	SCO1 (Complex IV)
Early-onset hypertrophic cardiomyopathy with encephalopathy	SCO2 (Complex IV)
Leukodystrophy and proximal tubulopathy	COX10 (Complex IV)
Defects in genes altering the stability of mtDNA	
Autosomal dominant progressive external ophthalmoplegia (adPEO)	ANT1 (adenine nucleotide translocase1)
Autosomal dominant progressive external ophthalmoplegia (adPEO)	Twinkle
Progressive external ophthalmoplegia (both autosomal dominant and recessive forms, adPEO and arPEO)	DNA polymerase γ
Myoneurogastrointestinal encephalopathy (MNGIE)	Thymidine phosphorylase

Table 3 →
Examples of diseases caused by mutations in nuclear genes encoding mitochondrial proteins with various functions.

DISEASE [OMIM number]	DEFECTIVE PROTEIN	REFERENCE
X-linked sideroblastic anemia with ataxia [301310]	ABCB7 (involved in the maturation of the FeS cluster proteins)	Allikmets et al. 1999
X-linked sideroblastic anemia [301300]	ALAS2 (involved in heme biosynthesis)	Cox at al. 1994
Friedreich ataxia [229300]	Frataxin (involved in iron metabolism)	Patel and Isaya 2001
Mohr-Traneberg syndrome (deafness-dystonia syndrome), X-linked [304700]	DDP1 (involved in the protein import to mitochondria)	Koehler at al. 1999
Hereditary spastic paraplegia (SPG7) [602783]	Paraplegin (metalloprotease, involved in protein turnover)	Casari et al. 1998
Hereditary spastic paraplegia (SPG13) [605280]	HSPD1 (heat-shock 60-kd protein 1), mitochondrial chaperonin	Hansen et al. 2002
Autosomal dominant optic atrophy [165500]	OPA1 (may be involved in the control of mitochondrial morphology)	Delettre et al. 2000
Pyruvate dehydrogenase deficiency (symptoms vary from fatal lactic acidosis to chronic neurological dysfunction) [312170, 245348, 246900, 245349]	Different components of the pyruvate dehydrogenase complex (E_1, E_2, E_3, X-lipoate or pyruvate dehydrogenase phosphatase)	Scriver et al. 2001a
Ornithine transcarbamylase deficiency (hyperammonia with various neurological symptoms) [311250]	Ornithine transcarbamylase (enzyme of the urea cycle)	Tuchman et al. 1995
Methylmalonic aciduria [251000]	Methylmalonyl CoA mutase (involved in the metabolism of vitamin B12)	Ledley and Rosenblatt 1997
Fumarate hydratase deficiency (encephalopathy, dystonia, leucopenia and neutropenia), autosomal recessive inheritance [606812]	Fumarate hydratase (enzyme of the citric acid cycle)	Bourgeron et al. 1994
Dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer [150800, 605839]	Fumarate hydratase (enzyme of the citric acid cycle)	The multiple leiomyoma consortium 2002

and brainstem or basal ganglia dysfunction, and pathologically by bilateral lesions in the brainstem, basal ganglia, thalamus, and spinal cord. Interestingly, mutations in the subunits of complex II, as well as in fumarase, an enzyme of the citric acid cycle (The Multiple Leiomyoma Consortium 2002), have also been discovered to cause rare cancers. (Shoubridge 2001; Suomalainen and Kaukonen 2001; Zeviani 2001).

Table 3 lists some examples of nuclear mutations affecting miscellaneous mitochondrial proteins. The first three, involved in mitochondrial iron metabolism, are discussed in greater detail in the next paragraph.

Mitochondrial functions

The principal task of the mitochondria is to generate ATP by oxidative phosphorylation. These organelles also perform numerous metabolic reactions: the citric acid cycle, the urea cycle, the degradation of fatty acids, and the synthesis and catabolism of several amino acids. In addition, heme, FeS clusters and various other enzyme cofactors and hormones are synthesized in mitochondria. Mitochondria also actively take part in programmed cell death machinery and act as intracellular calcium stores.

Oxidative phosphorylation

An overview of human catabolic metabolism is presented in figure 4. The citric acid cycle, taking place in the mitochondrial matrix, includes a series of oxidation-reduction reactions that result in the oxidation of carbon fuels, usually in the form of acetyl CoA, to two molecules of carbon dioxide. These reactions also produce NADH and $FADH_2$, which are energy-rich molecules containing a pair of electrons with high transfer and reducing potential. NADH and $FADH_2$ are also formed in glycolysis and fatty acid oxidation. When the high-energy electrons of NADH and $FADH_2$ are used to reduce molecular oxygen to water, energy is liberated and used to generate ATP. This is called oxidative phosphorylation, and is the major source of ATP in aerobic organisms. Oxidative phosphorylation is performed by the respiratory chain, which is located in the inner mitochondrial membrane and consists of four enzyme complexes participating in electron transfer, and a fifth enzyme complex, ATP-synthase. The flow of electrons from NADH and $FADH_2$ through the respiratory chain to O_2 leads to the pumping of protons out of the mitochondrial matrix into the mitochondrial intermembrane space. The resulting uneven distribution of protons generates a pH gradient and a transmembrane electrical potential that creates a proton-motive force. ATP is synthesized when protons flow back to the mitochondrial matrix through ATP-

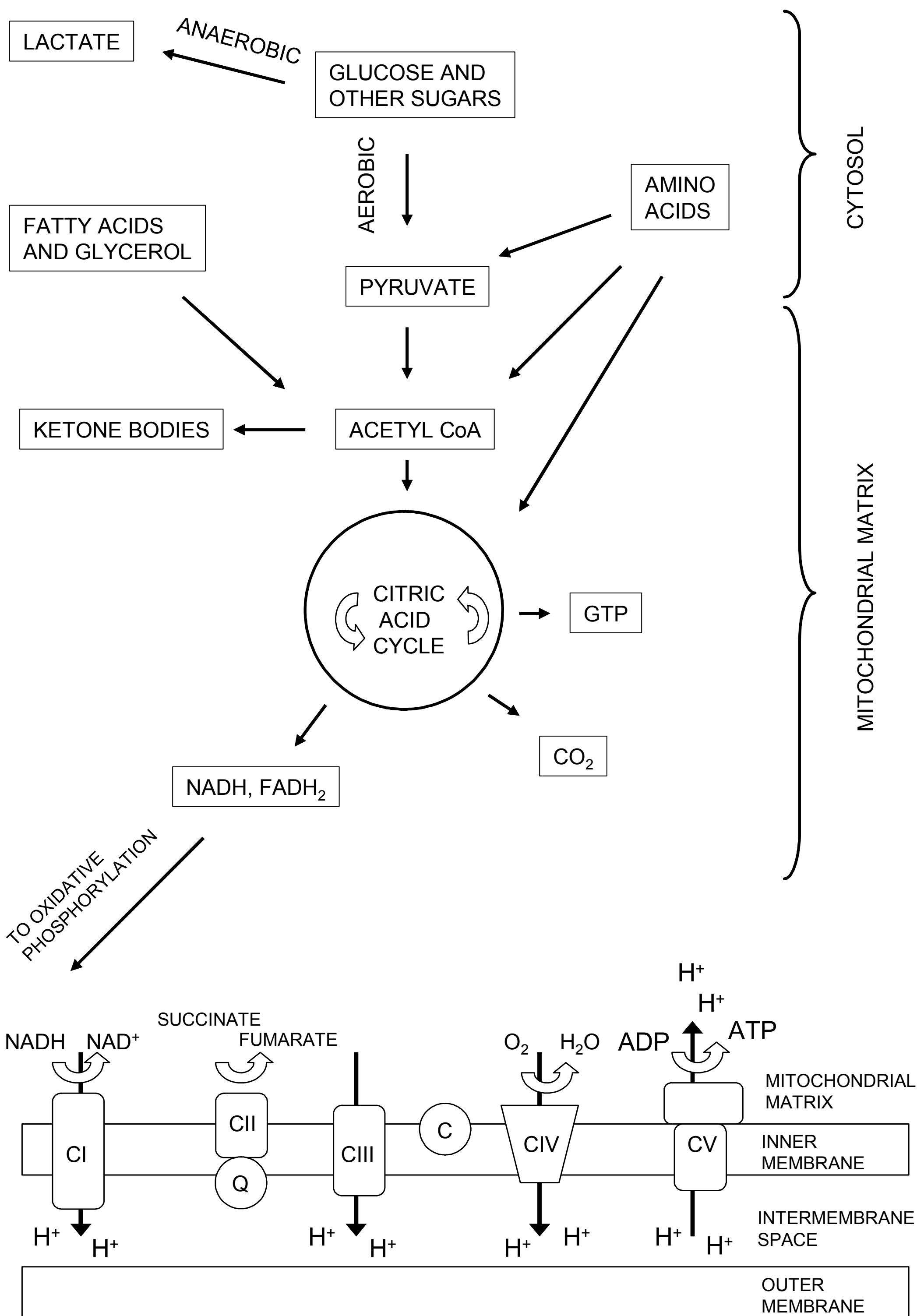


Figure 4
Simplified schematic presentation of the human catabolic metabolism. (CI-CIV = the mitochondrial respiratory chain complexes I-IV, CV = ATP synthase, Q = ubiquinone, C = cytochrome *c*)

synthase. ATP then serves as an energy source in various cellular reactions. (Berg et al. 2001b).

Complex III of the respiratory chain and BCS1L

Complex III of the respiratory chain is presented here in greater detail, since it is the crucial complex in the GRACILE syndrome. Complex III (cytochrome bc_1 complex or Q-cytochrome *c* oxidoreductase) of the respiratory chain is a dimer, with each monomer containing 11 subunits: core proteins I and II, six small subunits and three polypeptides involved in electron transfer: heme-containing cytochrome *b* and cytochrome c_1, and Rieske protein, which contains an FeS cluster. Cytochrome *b* is encoded by mtDNA, all other subunits being encoded by nuclear genes. Complex III catalyzes electron transfer between two mobile electron carriers, from ubiquinol to cytochrome *c*, and translocates protons from the mitochondrial matrix to the mitochondrial intermembrane space. (Berg et al. 2001b).

Isolated complex III defects are very rare, and are usually caused by point mutations in the cytochrome *b* gene, *MTCYB*. Mitochondrial point mutations are generally inherited maternally, whereas large rearrangements of mtDNA tend to be sporadic. *MTCYB* point mutations seem to be an exception to this rule, being mostly sporadic. They are usually heteroplasmic and restricted to the affected tissue. Typical manifestations of *MTCYB* mutations are exercise intolerance, proximal limb weakness and elevated lactate levels, sometimes associated with myoglobinuria (Andreu et al. 1999; Legros et al. 2001). More severe multisystem manifestations of *MTCYB* mutations, including symptoms of the central nervous system or cardiomyopathy, have also been reported (Valnot et al. 1999; Keightley et al. 2000; Wibrand et al. 2001).

No mutations in the other structural subunits of complex III have been described, but mutations in BCS1L, a chaperone needed for the assembly of complex III, have been reported (de Lonlay et al. 2001). Bcs1p protein, the yeast homolog of the human BCS1L (BCS1-like) has been studied in detail. It is an inner mitochondrial membrane protein with a single transmembrane domain (Fölsch et al. 1996). A short N-terminus lies in the intermembrane space, and the bulk of the protein is located in the matrix. Unlike most mitochondrial proteins, the Bcs1p does not include an N-terminal targeting signal, but a positively charged segment of amino acids, located immediately C-terminally to the transmembrane domain, acts as an internal targeting signal (Fölsch et al. 1996). Bcs1p belongs to the conserved AAA protein superfamily (AAA - ATPases Associated with a variety of cellular Activities), and acts as an ATP-dependent chaperone maintaining complex III in a competent state for the assembly of Rieske FeS and Qcr10p subunits (Nobrega et al. 1992; Cruciat et al. 1999).

The human *BCS1L* gene has been cloned by searching the EST database with the yeast Bcs1p protein sequence, and it has been shown that the human BCS1L polypeptide is imported unmodified to the inner mitochondrial membrane (Petruzzella et al. 1998). The function of the human protein is related to that of yeast Bcs1p, since the five patients described by de Lonlay et al. had four different missense mutations in the *BCS1L*, and decreased complex III activity in the tissues. The symptoms of the patients varied, the essential findings being metabolic acidosis, tubulopathy, hepatic dysfunction and encephalopathy (de Lonlay et al. 2001).

Formation of FeS clusters

FeS clusters are important cofactors of many of the proteins in mitochondria, cytosol and nucleus. Synthesis of FeS clusters is an important house-keeping function of mitochondria. The best-known FeS proteins are the FeS cluster -containing subunits of the respiratory chain complexes I, II and III, which serve as electron carriers. Recently, other FeS cluster-containing proteins, involved for example in the sensing of oxygen and iron, have also been studied intensively, as well as the mitochondrial proteins participating in the FeS cluster assembly. (Beinert and Kiley 1999; Lill et al. 1999; Beinert 2000; Lill and Kispal 2000; Muhlenhoff and Lill 2000).

Cytosolic aconitase is an interesting example of the FeS cluster proteins having sensing and regulatory functions. Mitochondrial aconitase, containing a FeS cluster, is an enzyme of the citric acid cycle, catalyzing the isomerization of citrate to isocitrate through steps of dehydration and hydration. Cytosolic aconitase, which shares a 30% amino acid homology with its mitochondrial counterpart, also contains an FeS cluster, but the cluster is relatively unstable, dissociating from the apoprotein in low-iron conditions. This change in conformation makes cytosolic aconitase an iron-response element -binding protein (IRE-BP).The IRE-BPs bind to the iron-responsive elements (IRE) of the 5'UTR of the ferritin mRNA, blocking the translation of this iron-storage protein. They also bind to the IREs in the 3'UTR of the transferrin receptor mRNA. This stabilizes the mRNA, leading to increased translation of the transferrin receptor, the protein responsible for iron uptake by the cells. The conformation change is reversible, and, in high-iron conditions, the IRE-BP again binds an FeS cluster, being thereafter unable to bind IREs. (Berg et al. 2001a; Walker et al. 2001).

Some ten mitochondrial proteins involved in the FeS cluster assembly machinery have been identified in yeast so far. These proteins share homologies with bacterial proteins performing the same task. Presumably, FeS proteins occurred early in evolution, and are present in the cells of virtually all species. Interestingly, synthesis of FeS proteins seems to be a more vital function of the

mitochondria than oxidative phosphorylation. The yeast strains with defects in the proteins involving the function of the respiratory chain usually have a petit phenotype – they grow on glucose using anaerobic metabolism and forming small colonies, but cannot grow on glycerol, since on non-fermentable carbon sources the yeast cells depend on aerobic metabolism. Yeast strains deleted for *NFS1, YAH1, JAC1* or *ARH1*, four essential genes involved in FeS cluster formation, cannot grow at all. (Lill et al. 1999; Lill and Kispal 2000; Muhlenhoff and Lill 2000).

ABCB7 and ABCB6, the human orthologs of yeast Atm1p

Mitochondria are needed not only for synthesis of the mitochondrial FeS proteins, but also for maturation of the extramitochondrial ones. The details of the formation of the cytosolic FeS proteins are not known yet, but the FeS clusters are probably formed in the mitochondria and transported to the cytosol, where they are incorporated to the apoproteins (Muhlenhoff and Lill 2000). Yeast protein Atm1p is a mitochondrial yeast protein presumably exporting FeS clusters from the mitochondria. Defects in the Atm1p result in defective assembly of cytosolic FeS proteins, deficiency of the holoforms but not of the apoforms of heme-bearing proteins, accumulation of iron in mitochondria, the instability of the mitochondrial DNA and the loss of oxidative respiration function (Leighton and Schatz 1995; Kispal et al. 1997; Kispal et al. 1999; Senbongi et al. 1999). Atm1p belongs to the protein superfamily of the ATP binding cassette (ABC) transporters, and has two human orthologs, ABCB7 and ABCB6 (Lill and Kispal 2001). (ABCB7 and ABCB6 are the systematic names of these proteins and the corresponding genes. Previously, ABCB7 was called ABC7 or hABC7, and the previous name of ABCB6 was MTABC3.) Both human genes, *ABCB7* and *ABCB6,* are able to partially reverse the phenotype of the yeast strains with Atm1p defects (Allikmets et al. 1999; Mitsuhashi et al. 2000).

ABCB7 has been mapped to the X-chromosome (Savary et al. 1997; Shimada et al. 1998). Missense mutation I400M in *ABCB7* results in X-linked sideroblastic anemia with ataxia (OMIM 301310), a recessive disorder characterized by in early childhood onset of non-progressive cerebellar ataxia and mild anemia with hypochromia and microcytosis (Allikmets et al. 1999). The disease causes iron accumulation in the mitochondria, but its pathogenesis is not yet well understood.

ABCB6 has been mapped to chromosome 2, and so far no human diseases associated with mutations in it have been reported. *ABCB6* has, however, been advocated in the literature as the gene most probably affected in the GRACILE syndrome (Mitsuhashi et al. 2000; Lill and Kispal 2001).

Other diseases with abnormalities of mitochondrial iron metabolism

Mitochondria play a central role in cellular iron metabolism. They are the site of several steps of heme biosynthesis, and contain large amounts of non-heme iron in the FeS clusters of the electron transport chain. However, in addition to X-linked sideroblastic anemia with ataxia, only two Mendelian disorders are known to be associated with impaired iron homestasis in mitochondria: X-linked sideroblastic anemia (OMIM 301300) and Friedreich ataxia (OMIM 229300).

X-linked sideroblastic anemia is caused by mutations in the gene encoding delta-aminolevulinate synthase 2, *ALAS2* (previous symbols *ALAS*, *ALASH*, *e-ALAS*) (Cox et al. 1994). ALAS2 carries out the first step of heme biosynthesis, catalyzing the condensation of glycine and succinyl coenzyme A to produce aminolevulinic acid (ALA). When heme biosynthesis is stalled by deficiency of ALAS2, iron accumulates in the mitochondria, resulting in a characteristic picture of ringed sideroblasts in the erythroid marrow. The anemia is usually normocytic and normochromic, but it may be microcytic and hypochromic, i.e. similar to other anemias that result from insufficient hemoglobin production. Many patients respond to pyridoxine, a cofactor of ALAS2. (Andrews 2000).

Friedreich ataxia is an autosomal recessive disorder characterized by spinocerebellar degeneration, cardiomyopathy and diabetes mellitus. The symptoms begin in childhood, and the disease leads to death in the fourth or fifth decade of life. Friedreich ataxia is caused by mutations in the frataxin gene, *FRDA* (other symbol *X25*) (Campuzano et al. 1996), the majority of the mutations being GAA-triplet repeat expansions in the first intron of the gene. Frataxin is a mitochondrial protein, the gene encoding it being located on chromosome 9q13. Decreased frataxin leads to accumulation of mitochondrial iron with failure of mitochondrial respiratory function. The yeast cells deleted for *YFH1* gene, the yeast homolog of frataxin, also accumulate iron in their mitochondria. Drastic increases in the concentrations of mitochondrial free iron have been reported upon inactivation of several yeast genes involved in the FeS cluster assembly. *YFH1* and frataxin might also play a role in FeS cluster formation, but their precise function is so far unknown. (Muhlenhoff and Lill 2000; Becker and Richardson 2001; Patel and Isaya 2001).

AIMS OF THE STUDY

This work was undertaken when the GRACILE syndrome had been newly discovered. The clinical characteristics of this fatal metabolic disorder were defined and the autosomal recessive inheritance was established, but the pathogenesis of the disease remained unknown. The aim of this study was to elucidate the molecular genetic background of the GRACILE syndrome, with the following specific aims:

1) To assign the GRACILE locus,
2) To restrict the critical chromosomal region,
3) To identify the causative gene either by positional cloning or by the positional candidate approach,
4) To provide functional evidence for the causative role of the identified mutation, and
5) From the clinical point of view, to produce DNA diagnostics for the GRACILE syndrome.

SUBJECTS AND METHODS

GRACILE families and patient samples

DNA samples from 19 Finnish GRACILE patients, 22 parents, 13 siblings and three half-siblings from 13 families were used for linkage and sequence analyses. Autopsy samples, muscle biopsy specimens or fibroblast cell lines of nine patients were used for RNA extractions and activity measurements of the mitochondrial enzyme complexes. In addition, DNA samples from three British infants (Morris et al. 1995) with a disease picture closely resembling that of the Finnish GRACILE patients were used for sequencing studies. This study was approved by the Ethical Committee of the Hospital for Children and Adolescents, Helsinki University Central Hospital, Finland, and parental informed consent was obtained.

Control samples

As controls for the linkage disequilibrium analyses, DNA samples of 46 individuals originating from eastern and central Finland were used. As controls in mutation screening we used 494 Finnish individuals, 70% of them originating from eastern and central Finland, and a control panel of 50 Caucasian individuals (HD50CAU, Coriell Cell Repositories). This control panel consisted of 50 Americans originating from different parts of Europe.

Control liver tissue samples used for RNA extraction were derived from two adults who had died in accidents, and the control fibroblast cell line used as an RNA control was also adult-derived.

DNA and RNA extraction

Genomic DNA of the deceased patients was extracted either from cultured fibroblasts using standard procedures, or from paraffin-embedded tissue samples by the rapid lysis method (Isola et al. 1994). DNA of the living patients and healthy family members was extracted from 2-20 ml peripheral blood samples according to standard procedures.

For RT-PCR sequencing, total RNA was extracted from deep frozen liver and kidney tissue samples, using the RNeasy Mini Kit (QIAGEN). Messenger RNAs for Northern blot analysis were extracted from fibroblast cell lines and liver tissue samples, using the Oligotex mRNA Direct Mini kit (QIAGEN).

Genotyping

A slightly modified Weber screening set (Sheffield et al. 1995) of 380 polymorphic microsatellite markers was used in the genome screen. The markers were amplified by PCR, and the PCR products were separated on an automated

laser fluorescence DNA sequencer (ALF Express, Pharmacia, Sweden). The microsatellite markers used in the dense mapping of the GRACILE locus originated from several internet databases. The primers were labeled with ^{32}P, and the PCR reactions and the size analysis of the PCR products were performed as described earlier (Aaltonen et al. 1993). Later, the exact allele sizes of the critical markers were determined using fluorescently labeled PCR primers, ABI377 sequencer and GENOTYPER program, version 2.0 (Applied Biosystems). Microsatellite genotyping for diagnostic purposes was performed with this system. The SNP markers flanking *ABCB6* were genotyped by direct sequencing.

Linkage and LD analyses

The principle of linkage analysis is to assign a disease locus to a specific chromosomal region by comparing the segregation of the alleles of genetic markers in families with affected individuals. If two loci (the marker and the disease locus) co-segregate in the pedigrees more often than they would by chance, they are said to be linked, meaning that they lie very close to each other in a chromosome. In practice, if the disease is inherited by autosomal recessive inheritance, a genetic marker is linked to the disease if affected siblings always inherit the same marker alleles from both parents, but the healthy siblings carry only one, or none, of these alleles. The unit for genetic distances is a centimorgan (cM), which corresponds to a recombination fraction of 1%, and implies that two loci 1 cM apart have approximately a 1% chance of recombination during meiosis. If the study sample is small, a marker can also seem linked by chance. The statistical significance of linkage is expressed by LOD (logarithm of odds) score. In linkage analysis two assumptions are tested: 1) the two loci are linked with a given recombination fraction (θ), or 2) they are not linked. The ratio of these likelihoods is called a likelihood ratio, and LOD score is the $\log_{10}$ of this ratio (Ott 1974; Terwilliger and Ott 1994). Thus a LOD score of 3.0 means roughly that it is 1000 times more likely that two loci are linked than that the observed allele segregation in the families would have arisen by chance. A LOD score around zero means that the marker is non-informative, or that the study sample is too small, and LOD scores under -2 indicate exclusion of the locus as the disease locus.

Two-point linkage analyses in this study were carried out with the MLINK option of the LINKAGE package computer programs, version FASTLINK 2.2 (Lathorp and Lalouel 1984), assuming autosomal recessive inheritance, complete penetrance and no phenocopies. The consanguinity of the parents of family 1 was taken into account in the linkage analyses.

The principle of LD is discussed in the review of the literature. Two-point linkage disequilibrium analyses of this study were carried out with the

DISLAMB computer program, and multipoint linkage disequilibrium analysis was performed with the DISMULT computer program (Terwilliger 1995).

Radiation hybrid mapping

The order and approximate distances of the genetic markers in the critical GRACILE region were determined by radiation hybrid (RH) mapping. This is a method based on cell hybrid panels created by fusing rodent and X-ray-treated human cells. A high dose of X-rays is used to fragment the human genomic DNA, and the broken chromosomal fragments are rescued by cell fusion with a recipient rodent cell line. A panel of individual fusion cell lines can then be analyzed for the presence or absence of different markers by PCR. Markers located very close to each other will be positive for the same, or almost the same, cell lines, since within a short distance the X-rays do not break the DNA many times. The markers can be ordered, and the physical distance separating them estimated, by statistical analyses. (Goss and Harris 1975; Cox et al. 1990; Boehnke et al. 1991).

In this study, we used Stanford medium(G3) and high(TNG) resolution RH panels. The markers were amplified by PCR and the products were separated by electrophoresis on agarose gels.

The RH mapping data were analyzed using the FORTRAN program RHMAXLIK of the RHMAP package version 3.0 (Lange et al. 1995; Lunetta et al. 1996). This program orders the markers and calculates marker distances in centirays (cR). It allows the use of data from several RH panels simultaneously, which was important in this study, because the ordered markers were too close to each other to be ordered with the Stanford G3 RH panel only, and too distant from each other to be ordered with the Stanford TNG RH panel alone. We used the data of both these panels and the branch and bound ordering option to determine the order and distances of the critical markers. The marker distances were converted from cR to cM by dividing the known total length (in cM) of the linked region proportional to the marker distances in cR.

Bioinformatics

The sequence databases and various programs for data processing provided on the web pages of different universities and organizations are essential tools in molecular genetics. The detailed web addresses of the internet databases mentioned here are provided in the electronic database information section.

For the fine mapping of the GRACILE region (carried out in 1997-1998), we utilized the marker maps of Genethon, the University of Southampton (Genetic location database, LDB), Whitehead Institute and Genome Database (GDB). The Stanford RH-panel raw data, available for some of the critical markers

on the Stanford Human Genome Center web pages, were utilized when the markers were arranged by RH-mapping.

"A new gene map of the human genome" at NCBI's web page, containing EST:s, mRNA:s and genes organized by RH mapping (Deloukas et al. 1998), was our first tool in searching for good positional candidate genes within the critical GRACILE region. The amount of the draft genomic sequence in the critical region, available in GenBank, increased exponentially in 1999-2000. While analyzing the positional candidate genes, we followed the development of the NCBI's "Map Viewer" and the "Human genome project working draft" of UCSC, updating new possible candidate genes and useful genetic markers. These are human genome assemblies providing locations of the markers and genes based on the actual sequence, but also providing links with the earlier physical and genetic maps.

In order to define the exact position of the *ABCB6* gene, we built physical maps *in silico*. The sequences of the genomic clones were searched at nucleotide and htgs (high throughput genomic sequence) databases with critical markers using BLAST programs (Altschul et al. 1990) at NCBI. New markers in the genomic clones were found by running electronic PCR (Schuler 1997) at NCBI. The sequence utilities of the Search Launcher at the web pages of Baylor College of Medicine, as well as the commercial SEQUENCHER computer program provided practical tools for the handling of the DNA sequences.

The genomic structure of the *BCS1L* gene was also determined using biocomputing tools. The genomic clone containing *BCS1L* was identified by a BLAST search with the *BCS1L* mRNA sequence. In order to find all the different splice variants of the gene we searched the human EST and nucleotide databases with the coding part of the *BCS1L* mRNA using BLAST programs. The different EST and mRNA sequences were then aligned with the genomic clone and divided into exons so that all the exon-intron boundaries followed the GT-AG rule (Mount 1982).

BLAST programs and protein sequence databanks at NCBI, as well as the MultAlin alignment program (Corpet 1988), were used to define the functional domains and conserved regions of the human BCS1L, by comparing the protein sequence with those of other organisms.

Sequencing and mutation screening

RT-PCR and genomic sequencing of the candidate genes were performed using BigDye terminator kit and ABI 377 and ABI 3700 automated sequencers (Applied Biosystems). The sequences were analyzed by the SEQUENCHER program.

Screening of the 232A>G (S78G) point mutation in the *BCS1L* gene in the Finnish GRACILE patients and their parents was carried out by direct sequencing. Mutation screening in siblings and control samples was performed using solid-phase minisequencing (Syvänen et al. 1993). For other mutations, the controls were screened by direct sequencing.

Northern blot analysis

Northern blotting was performed using standard procedures. Probes for *ABCB6* and *BCS1L* were produced by RT-PCR and labeled with ^{32}P, using a random primer DNA labeling system (Life Technologies). Hybridizations were carried out in Expresshyb hybridization solution (Clontech) as suggested by the manufacturer.

Expression plasmid construction, cell culture, and transfections

BCS1L cDNA with a c-terminal FLAG sequence was PCR-cloned to a pGEM®-T Easy vector (Promega). 232A>G (S78G) mutagenesis was carried out in pGEM-T Easy vector, using the QuickChange site-directed mutagenesis kit (Stratagene). Wild-type and mutant *BCS1L* cDNAs were then subcloned to the pCMV5 expression vector (Andersson et al. 1989). COS-1 cells obtained from the American Type Culture Collection were transiently transfected with the cDNA constructs, using Lipofectamine and PLUS reagent (Life Technologies), following the manufacturer's guidelines.

Immunofluorescence microscopy

For immunonofluorescence microscopy, COS-1 cells were plated on cover slips and transfected as described above. 48 h after transfection, the cells were incubated in DMEM without FBS, in the presence of 50 μg/ml of cycloheximide, for one hour to halt protein synthesis, and for an additional 30 minutes with cycloheximide and 200 nM MitoTracker Red CMXRos (Molecular Probes) for mitochondrial staining. Thereafter, the cells were fixed with methanol, blocked with 0.5% BSA/0.2% saponin, and incubated with anti-FLAG M2 antibody (Stratagene). The cells were washed with 0.5% BSA/0.2% saponin and stained with FITC-conjugated anti-mouse secondary antibody (Sigma). After washing with PBS, the cells were mounted in glycerol and viewed with an immunofluorescence microscope.

Western blot analysis

For Western blot analysis, COS-1 cells were lysed 48 h after transfection with lysis buffer supplemented with protease inhibitors. Samples were denatured and electrophoresed on a 10% SDS-PAGE gel. Western blotting was performed

according to standard procedures. The membrane was stained with anti-FLAG M2 antibody (Stratagene), the secondary antibody being peroxidase-conjugated anti-mouse IgG (Sigma). The signals were detected using ECL™ Western blotting detection reagents (Amersham Pharmacia Biotech).

Pulse chase analysis

The half lives of the wild-type and S78G mutant BCS1L polypeptides were determined by pulse chase experiments. Cells were labeled metabolically 36 h after transfection by starving them in methionine- and cysteine-free medium for one hour and thereafter labeling with both [^{35}S]methionine and [^{35}S]cysteine for one hour. Then the cells were incubated for 0, 12, 24, 36 or 48 h in chase media and lysed with lysis buffer supplemented with protease inhibitors. The lysed cells were immunoprecipitated with anti-FLAG M2 antibody (Stratagene) and Protein A/G Sepharose (Santa Cruz Biotechnology). Immunocomplexes were separated on 10% SDS-PAGE and visualized by fluorography. The results were analyzed by densitometry.

Yeast complementation studies

Human *BCS1L* cDNA, 480bp of yeast *BCS1* upstream sequence, and 210bp of yeast *BCS1* downstream sequence were PCR-cloned to the pBluescript® (Stratagene) vector. 232A>G (S78G) mutagenesis in *BCS1L* cDNA was carried out, using the QuickChange site-directed mutagenesis kit (Stratagene). The whole cassette containing yeast *BCS1* promoter, human *BCS1L* cDNA and yeast *BCS1* terminator was then subcloned to the high-copy yeast expression vector pRS425 (Sikorski and Hieter 1989). Yeast strain Y14211 (EUROSCARF, European *Saccharomyces cerevisiae* archives for functional analysis), deleted for *BCS1,* was transformed with the constructs. Transformations were performed by the lithium acetate procedure (Ito et al. 1983). Individual transformants were purified and checked for growth on YPEG plates (3% glycerol, 2% ethanol, 2% peptone, 1% yeast extract and 2% agar). To control the growth on glucose, the same transformants were also streaked on YPD plates (2% glucose, 2% peptone, 1% yeast extract and 2% agar), and on SD plates without leucine (2% glucose, 0.67% yeast nitrogen base, 2% agar, 20µg/ml histidine, uracil and tryptophan, and 30µg/ml adenine and lysine).

Activity measurements of mitochondrial enzyme complexes

Mitochondria were isolated from muscle biopsy specimens of five Finnish GRACILE patients, and from several autopsy-derived tissues obtained from four patients 2-7 h post mortem. The activities of respiratory chain enzymes (NADH:cytochrome *c* oxidoreductase (complex I + III), succinate:cytochrome *c*

oxidoreductase (complex II + III), succinate:ubiquinone oxidoreductase (complex II) and cytochrome *c* oxidase (complex IV)) were measured spectrophotometrically (Majander et al. 1995). The measurements were performed at the Department of Medical Chemistry of the University of Helsinki in Helsinki, Finland. A modified spectrophotometric assay (Rustin et al. 1994) including separate assessment of NADH:ubiquinone oxidoreductase (complex I) and ubiquinol:ferricytochrome-*c* oxidoreductase (complex III) was used to measure the enzyme activities from three deep-frozen liver samples and one muscle sample of the patients. These measurements were carried out at the Department of Pediatrics (Division of Pediatric Neurology and Metabolism) of the Ghent University Hospital in Ghent, Belgium.

RESULTS AND DISCUSSION

Assignment of the GRACILE locus to chromosome 2 (I)

The GRACILE locus was initially mapped by genome scan and linkage analysis, and the locus was confirmed and restricted by LD and haplotype analyses, using 11 markers ordered by RH mapping. We had DNA samples available from eight nuclear families, only four of which had two affected children (figure 5). In addition, DNA samples from two affected individuals from different families were used. These two samples were included in the LD analyses, but provided no information on linkage analysis. The parents of family 1 were second cousins, and based on genealogical studies the mother of family 4 and the father of family 5 were probably distant relatives. This number of families is not sufficient for disease locus assignment by conventional linkage analysis. However, since the GRACILE syndrome shows an autosomal recessive mode of inheritance, and the birthplaces of the families' ancestors were distributed in the late-settlement region of Finland, we assumed that all the families, or most of them, probably had a common ancestor and carried the same founder mutation. Therefore, we expected to be able to utilize LD and ancestral haplotype analyses for disease gene localization.

Genome screen and linkage analysis

The initial genome screen was carried out using DNA samples from only nine individuals: four affected sib pairs and the affected infant of family 5 (indicated by asterisks in figure 5). The affected infant of family 5 was included because the mother of family 4 and the father of family 5 were believed to be related, and consequently they may have shared a long chromosomal segment around the disease locus. The parents of family 1 were second cousins, and we assumed that the affected sibs in this family would be homozygous for the markers flanking the GRACILE locus. The analysis of the genotypes in the initial genome scan was carried out in a very simple and rapid way, by comparing the genotypes of the sibs within each family and looking for markers showing identical genotypes in sib pairs. After analyzing only 25 markers of the genome scan panel, we found two adjacent markers on chromosome 2q, D1S1649 (which, despite its name, maps to chr. 2) and D2S434, 5 cM apart, revealing identical genotypes in each of four sib pairs. In addition, marker D2S1363, 14 cM apart from marker D2S434, revealed identical genotypes in three sib pairs. In family 1, whose parents were second cousins, the sibs were homozygous for all these three markers. These markers were then genotyped in the complete family material. No obligatory recombinations could be identified between marker D2S434 and the GRACILE phenotype, the maximun LOD score being 3.0 at recombination fraction (θ) 0.

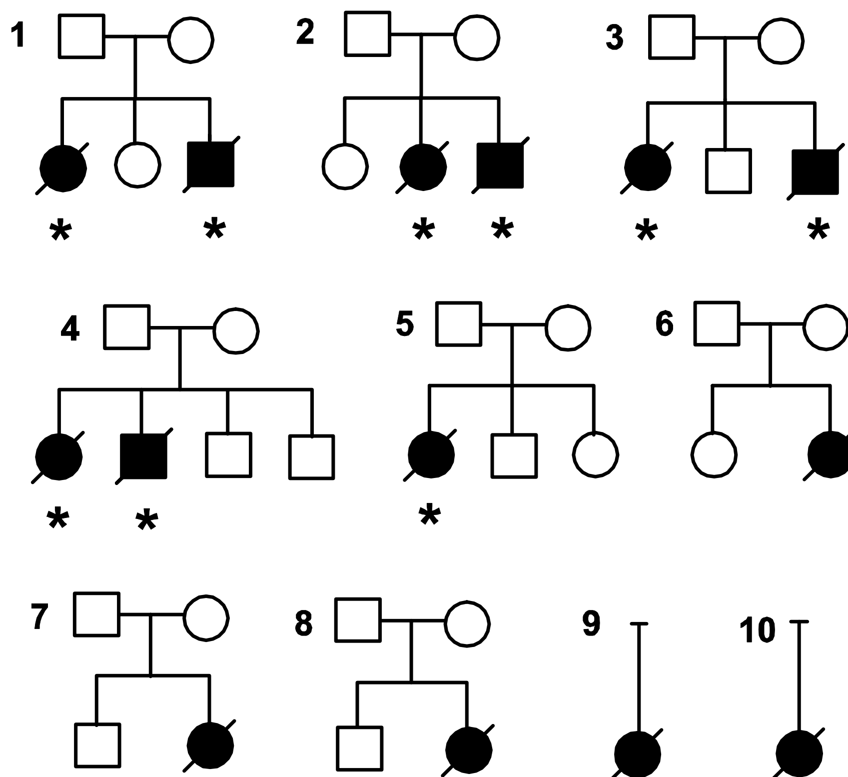


Figure 5
GRACILE pedigrees used in the linkage and LD studies (I). Asterisks indicate individuals whose DNA samples were analyzed in the initial genome screen.

The next step was to analyze several markers around marker D2S434 on chromosome 2q33-37 in the complete family material. Multiple markers between markers D2S164 and D2S2359 provided positive LOD scores (max. lod scores 0.91-3.88 at θ=0), suggesting a linkage on a 3 cM region (table 4). In the constructed haplotypes of the families, no obligatory recombination events could be detected between the GRACILE phenotype and a total of 11 markers on this 3 cM region. The affected infants of family 1 were homozygotes for all the markers over this region.

LD analyses and RH mapping of the critical region

To confirm the linkage and to restrict the critical chromosomal region, we performed LD analyses with all the markers that provided maximum two-point lod scores at recombination fraction 0. Since we had only 20 affected and 16 nonaffected chromosomes in our family material, we obtained additional non-affected chromosomes from regionally selected control samples. In total, we

Table 4
Pairwise linkage and LD analyses between the GRACILE syndrome and 15 marker loci in 2q33-37. Linkage analysis was carried out in eight families, and LD calculations were performed using allele data of 20 GRACILE and 108 control chromosomes.

Marker	Two-point linkage analysis, LOD Score at θ=0	Two-point LD analysis, p value
D2S301	$-\infty$	-
D2S164	$-\infty$	-
D2S2210	2.09	0.4
D2S434	3.00	0.05
D2S2249	3.57	0.002
D2S173	2.33	0.06
D2S104	2.63	0.002
D2S2179	2.16	0.002
D2S2250	3.88	1×10^{-12}
D2S433	2.29	2×10^{-9}
D2S1354	0.91	0.0003
D2S2244	2.54	5×10^{-7}
D2S163	3.30	0.005
D2S2359	$-\infty$	-
D2S120	$-\infty$	-

utilized 20 GRACILE and 108 control chromosomes in the linkage disequilibrium studies.

The p values of the LD analyses for each of the 11 markers on the linkage region on 2q33-37 are presented in table 4 (detailed results in I). Marker D2S2250 showed the most significant LD; all the GRACILE chromosomes carried the same allele 4, the proportion of which in the nondisease chromosomes was 21% (p value 10^{-12}). This confirmed the assignment of the GRACILE locus to the vicinity of this marker. Markers D2S433 and D2S2244 also provided significant p values in the LD analysis, and several flanking markers showed suggestive, although not significant, evidence of LD.

To be able to carry out multipoint LD analysis and to build the ancestral haplotypes correctly, we performed radiation hybrid mapping of the critical markers to confirm their order. The 11 linked markers were too close to each other to be ordered with the medium-resolution RH panel, but the distance between markers D2S2179 and D2S2250 turned out to be so long that not all the markers were linked in the high-resolution RH panel. We typed the markers in both panels and performed the RH data analysis with a computer program allowing the use of data from the two panels simultaneously. The most probable marker order and the approximate marker distances on the basis of this analysis are shown in figure 6A.

The total map length from D2S2210 to D2S163 was 60 $cR_{10\,000}$. Since 1 $cR_{10\,000}$ corresponds on average to 25 kb (Lunetta et al. 1996), the physical length of the 3 cM linkage region was estimated to be approximately 1.5 Mb, and that of the most critical 1 cM to be approximately 500 kb.

The multipoint LD analysis carried out with these markers resulted in a maximum lod score of 30.9 in the immediate vicinity of marker D2S2250 (figure 6A).

Haplotype analysis

The haplotypes of the GRACILE chromosomes are shown in table 5. The haplotypes of families 1-8 are presented as paternal and maternal disease chromosomes. In families 9 and 10 the parental data were missing, and consequently only the genotypes of the affected infants are shown. For these cases, we constructed the haplotypes on the basis of the expected common ancestors with the other affected individuals.

The 1 cM core haplotype cen-1-2-4-2-3-tel (D2S104-D2S2179-D2S2250-D2S433-D2S1354) was observed in 18/20 of the disease chromosomes. One of the haplotypes differing from the core haplotype had different alleles with markers D2S104 and D2S2179 and the other had a different allele with marker D2S433. The causative gene is probably located very close to the marker D2S2250, since all the GRACILE chromosomes carried the same allele 4, which was present in 21% of the control chromosomes, with this marker. In addition, with marker D2S1354, all the GRACILE chromosomes carried the same allele 3, but this allele was also very common (71%) in the control chromosomes. On the centromeric side of marker D2S2250, the critical DNA region was distinctly restricted by marker D2S2179, since the affected chromosome of the mother of family 4 was entirely different from the conserved GRACILE haplotype with marker D2S2179 and the markers centromeric from it. On the telomeric side of marker D2S2250, the border of the critical chromosomal region remained less definitive. With marker D2S433, the infant of family 10 was heterozygous for the core haplotype alleles 2 and 3. However, allele 3 differed from allele 2 by one tetranucleotide repeat only, and thus a marker mutation could not be excluded, especially since, with marker D2S2244, this infant was homozygous for allele 1, which showed strong linkage disequilibrium with the disease. Therefore the critical GRACILE region was not definitely restricted by marker D2S433, but lay between markers D2S2179 and D2S2244. Most probably, the causative gene was located in the immediate vicinity of marker D2S2250.

The length of the shared haplotype was comparable to that observed in infantile-onset spinocerebellar ataxia (IOSCA), another Finnish disease occurring mostly in the late-settlement region of eastern and central Finland (Varilo et al.

1996). This was in accordance with the results of the genealogical studies, suggesting that the GRACILE mutation was probably introduced into the Finnish population some 30-40 generations ago and spread to the late-settlement region with the population migration of the 16th century.

The mapping strategy and the importance of LD

In the initial genome screening, we used DNA samples from only nine individuals. This efficient screening method has previously been used successfully in the identification of other monogenic disease loci of the Finnish disease heritage (Mäkelä-Bengs et al. 1998; Pekkarinen et al. 1998). As in these earlier studies, we did not find a conserved haplotype shared by all affected individuals with the marker density utilized in the first stage of the genome scan, although a common ancestor in the distant past was suspected. However, it proved successful to monitor for shared genotypes between siblings within each family. If the families have common ancestors only a few generations back, a genome scan can be carried out very efficiently by monitoring for chromosomal regions, in which the affected individuals show homozygosity for genetic markers (homozygosity mapping) (Lander and Botstein 1987; Houwen et al. 1994; Nikali et al. 1995). This principle was also demonstrated in this study by family 1, in which the parents were second cousins, and the affected infants showed homozygosity of the genetic markers over a long chromosomal region. However, usually the chromosomal segments that are shared by most of the individuals carrying one founder mutation, are very short, and "pure" homozygosity mapping cannot be applied at the first stage of a genome scan. This is true among the diseases belonging to the Finnish disease heritage also. The conserved haplotypes of the disease alleles in patients from different nuclear families are very short (usually 0.5-2cM), although many of the founder mutations are relatively young (de la Chapelle and Wright 1998; Peltonen et al. 1999).

LD and ancestral haplotype analyses have facilitated all positional cloning projects of the Finnish disease heritage by allowing further restriction of the critical chromosomal region after conventional linkage analysis. In this study also, the linkage disequilibrium and ancestral haplotype analyses restricted the critical chromosomal region from the initial 3 cM to 1 cM. Even more importantly, these methods made locus assignment possible despite such a small number of families. The two-point lod scores 0.91-3.88 obtained with conventional linkage analysis were suggestive, but the p value of 10^{-12} with marker D2S2250 in the LD analysis, the lod score 30.9 in the multipoint LD analysis and the finding of the conserved haplotype in the GRACILE chromosomes unequivocally confirmed the locus assignment.

Table 5
Haplotypes of the GRACILE chromosomes in the 3 cM linkage region. Shading indicates the shared core haplotype.

Family	D2S2210	D2S434	D2S2249	D2S173	D2S104	D2S2179	D2S2250	D2S433	D2S1354	D2S2244	D2S163
1 P	2	4	4	2	1	2	4	2	3	1	6
1 M	2	4	4	2	1	2	4	2	3	1	6
2 P	3	2	2	3	1	2	4	2	3	2	6
2 M	1	3	2	5	1	2	4	2	3	1	4
3 P	3	2	2	3	1	2	4	2	3	1	6
3 M	3	4	2	3	1	2	4	2	3	2	6
4 P	3	2	2	3	1	2	4	2	3	1	6
4 M	1	4	5	1	2	3	4	2	3	2	5
5 P	3	2	2	3	1	2	4	2	3	1	1
5 M	3	4	4	3	1	2	4	2	3	1	4
6 P	2	4	4	2	1	2	4	2	3	1	3
6 M	2	2	2	3	1	2	4	2	3	2	6
7 P	2	4	4	2	1	2	4	2	3	1	5
7 M	3	2	2	3	1	2	4	2	3	1	6
8 P	3	2	2	3	1	2	4	2	3	2	6
8 M	3	3	2	3	1	2	4	2	3	2	6
9 O	1	5	2	3	1	2	4	2	3	2	2
9 O	2	4	4	2	1	2	4	2	3	1	3
10 O	3	2	5	3	1	2	4	2	3	1	6
10 O	2	4	4	2	1	2	4	3	3	1	4

Analysis of the positional candidate genes *ABCB6* and *BCS1L* (III, IV)

Genes in the critical GRACILE region

The critical GRACILE region, restricted by haplotype analysis, reached from D2S2179 to D2S2244. When the linkage was found, the length of this chromosomal region was estimated to be 1-1.5 cM and about 500-700 kb, based on rough genetic maps and RH mapping. At that time the Human Gene Map at NCBI's web page provided four cloned genes and 20 mapped EST:s on the 4 cM interval (D2S164-D2S163) including the critical GRACILE region. Now the genomic sequence of the critical region is almost complete, and in the NCBI's latest human genome assembly (MapViewer in August 2002), the physical interval is 1.3 Mb. It contains almost 30 known genes and several predicted, hypothetical

genes, representing a gene-rich region. Figure 6A shows the RH mapping -based marker map of the critical region four years ago, and figure 6B shows the present situation. The critical chromosomal interval contains two interesting candidate genes, *ABCB6* and *BCS1L*, both encoding mitochondrial proteins.

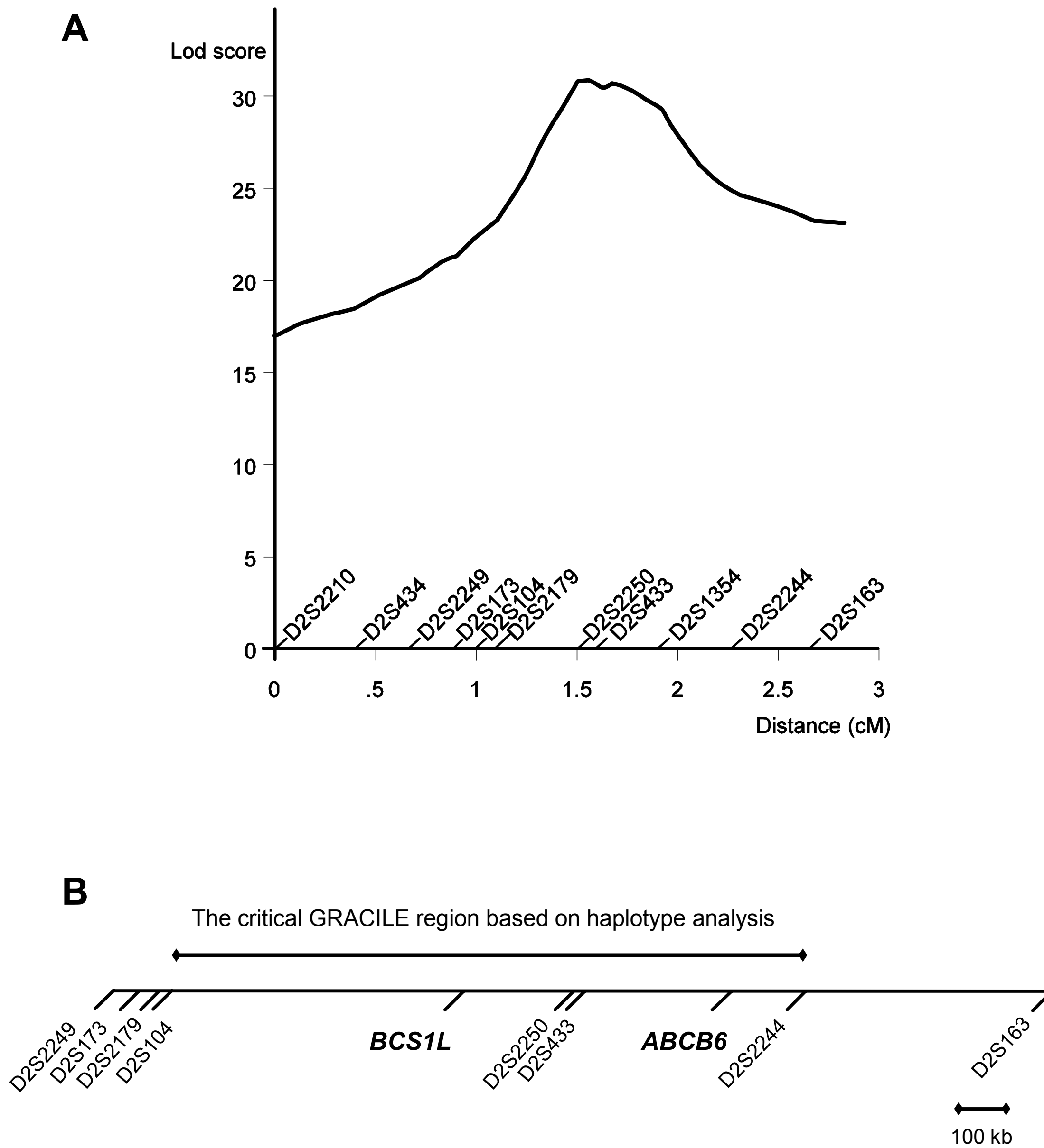
RH mapping with the Stanford medium resolution RH panel G3 localized the *BCS1L* to the immediate vicinity of the markers D2S2250 and D2S433, and ABCB6 between markers D2S433 and D2S2244. In addition, ABCB6 hit the same BAC clone, RP11-747C8, as D2S2244, implying that the gene was located closer to D2S2244 than to D2S433. During the sequencing of chromosome 2 (by HGP), these conclusions about the locations of *ABCB6* and *BCS1L* turned out to be correct. Thus *BCS1L* was located in the middle of the critical GRACILE region, whereas *ABCB6* lay on the very edge of it, being positionally not the most likely candidate, though theoretically possible. However, functionally *ABCB6* was exceptionally interesting. It was known to be involved in iron homeostasis, mitochondrial respiratory function and maintenance of the stability of mitochondrial DNA, and *ABCB6* had even been advocated in the literature as the most probable candidate gene for the GRACILE syndrome (Mitsuhashi et al. 2000; Lill and Kispal 2001). In contrast, *BCS1L* was functionally only a moderate candidate. It encoded a mitochondrial protein, whose yeast homolog Bsc1p was known to be essential for the function of the mitochondrial respiratory chain complex III. The symptoms of the GRACILE patients might be caused by a mitochondrial defect, but the activities of the respiratory chain complexes in the GRACILE patients had been shown to be within the normal range. We therefore analyzed both of these interesting candidates.

Sequence analysis of ABCB6 and BCS1L

We searched for *ABCB6* mutations in the patients' DNA samples by sequencing the coding region of the gene. In addition, 1300 bp of the genomic sequence upstream from the start codon and 500 bp of the genomic sequence downstream from the stop codon (including 5' and 3' UTRs) were sequenced. There were no possible mutations in these regions of the *ABCB6* in the patients, as compared with the controls and the reference sequences, but an SNP was identified in one patient. He had a heterozygous A>G nucleotide change, 218bp downstream from the stop codon, and 52bp downstream from the last base of exon 19.

The sequencing of the *BCS1L* covered the whole genomic sequence, including 2020 bp of the upstream, and 1040 bp of the downstream sequence. All the Finnish GRACILE patients had a homozygous 232A>G nucleotide change, causing an S78G amino acid change, in the second exon of *BCS1L*. All the parents were heterozygous for this variation, and the siblings and half-siblings of the patients were either homozygous for the normal allele, or heterozygous. There

Figure 6
A) Multipoint LD analysis of 11 markers in the 3 cM linkage region with 20 disease and 108 nondisease chromosomes. The marker order and distances are based on the RH mapping performed in 1998 (I). B) The order and distances between the critical markers and the analyzed positional candidate genes, *BCS1L* and *ABCB6,* as they were in the NCBI's Map viewer in August 2002.



were no other nucleotide changes in the *BCS1L* sequence in the patients. Of the 494 control Finns and 50 other Caucasians screened for the 232A>G nucleotide change, one Finnish control was a heterozygote, and all the other controls were homozygous for the normal allele.

Northern blot analysis of ABCB6 and BCS1L

Northern blot analyses of *ABCB6* and *BCS1L* were carried out in Finnish GRACILE patients to monitor the steady-state expression level of the genes and any possible differences in the transcript sizes. Hybridizations of both the *ABCB6* and *BCS1L* probes were carried out using fibroblast mRNA filters, and the *BCS1L* probe was also hybridized with a liver filter. No differences in the steady-state expression levels or transcript sizes were detected between the GRACILE patients and the controls with either of these genes (figures in III and IV).

Positional exclusion of ABCB6

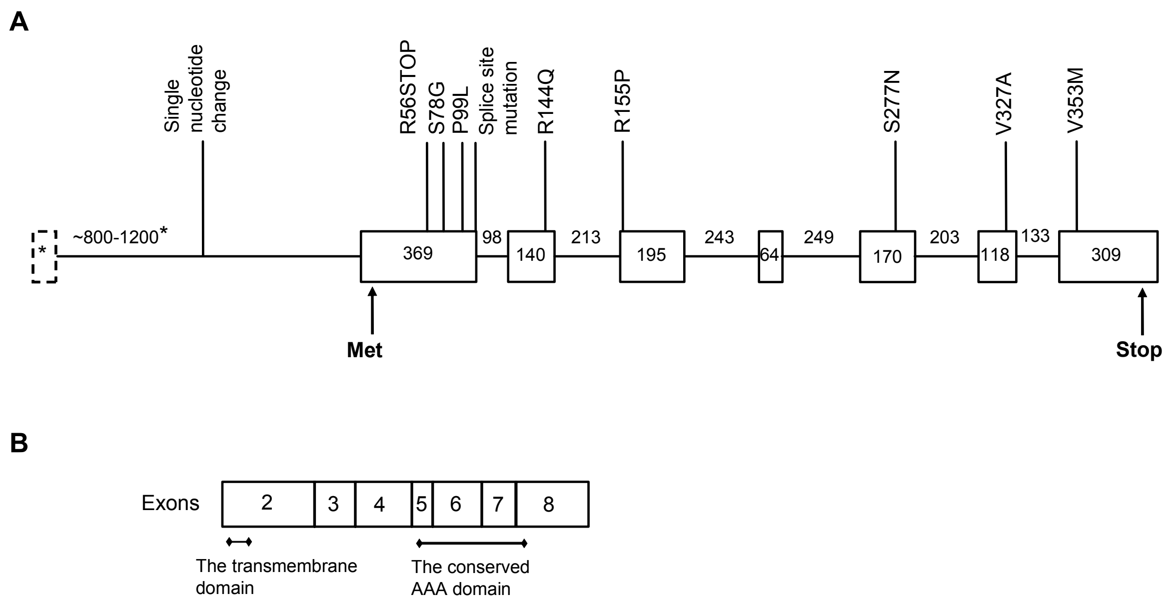
Since all the Finnish GRACILE families carried the same ancestral haplotype, the patients should be homozygous for the disease mutation and the genetic markers in its immediate vicinity. Thus, the heterozygous nucleotide variant found 218 bp downstream from the stop codon of *ABCB6* in one patient suggested that *ABCB6* could also be positionally excluded. To confirm the exclusion of *ABCB6*, we monitored for this SNP in the whole study sample of ten families. We also screened for other SNPs and found another informative SNP (NCBI ss# 12967), located 1543 bp upstream from the start codon of *ABCB6*. This SNP was also genotyped in all the families, and haplotypes were constructed using the genotype data of both sequenced SNPs and flanking microsatellite markers. Comparison of the haplotypes of the disease alleles clearly indicated that *ABCB6* lay telomeric to the critical GRACILE region (presented in table 1 in III).

BCS1L sequence analysis in the British patients

We also sequenced the *BCS1L* using the DNA samples of the three British infants (Morris et al. 1995), whose disease course closely resembled that of the Finnish patients with GRACILE syndrome. The different BCS1L mutations identified in this study and previously (de Lonlay et al. 2001) are summarized in figure 7A and table 6.

British patient 1 was a compound heterozygote for a premature stop codon at amino acid position 56 in the second exon, and for a V327A missense mutation in the seventh exon. Patient 2 had a heterozygous splice donor mutation changing the first G of the second intron to a T, and a heterozygous T>A single

Figure 7
A) Genomic structure of the *BCS1L* gene, and all identified mutations (IV and Lonlay et al. 2001). The sizes of the exons and introns are indicated as base pairs. The sizes of the first exon and the first intron (marked with asterisks) vary due to alternative splicing. B) The BCS1L polypeptide of 419 amino acid residues. The numbers indicate the polypeptide regions encoded by distinct exons of the *BCS1L* gene. Residues 9-32 correspond to the transmembrane region of the yeast BCS1p protein (Fölsch et al. 1996). Residues 224-344 represent the conserved AAA domain, according to the CD-Search (Marchler-Bauer et al. 2002) at NCBI's BLAST service (AAA-ATPases Associated with a variety of cellular Activities).



nucleotide change in the middle of the first intron, 588 bp upstream from the start codon of the BCS1L. Patient 3 was a compound heterozygote for two missense mutations: the Finnish S78G mutation in exon 2, and an R144Q mutation in exon 3. None of these mutations, including the intronic T>A change in patient 2, was found in 140 controls, 90 of whom were Finnish and 50 were other Caucasians.

Genomic structure of *BCS1L* (IV)

Both the mRNA sequence and the genomic structure of the *BCS1L* gene have been published earlier (Petruzzella et al. 1998; de Lonlay et al. 2001). However, the genomic sequence published by de Lonlay et al. was partially incomplete for the non-coding part of the first exons, which we determined using biocomputing tools. Comparison of the different EST and mRNA sequences

suggested that the 5'end of the *BCS1L* transcript can be alternatively spliced. The 15 EST or mRNA sequences found in public databases and Celera represented six different splice variants. Two of the variants were represented several times and had been found in clones originating from normal tissues, whereas the other four were rare and reported only in EST libraries of cancer tissues, or had a non-specified tissue origin in the databases. In most transcripts, the first exon was small (20-120 bp), and the size of the first intron varied from 800 to 1200 bp. The second exon was 369 bp, started 49 bp before the start codon, and contained an in-frame stop codon 42 bp upstream from the start codon. This exon and all six exons thereafter were identical in all the splice variants (figure 7A). A few clones contained an additional short exon in the middle of the first intron. The exons spanned 3.4 to 3.8 kb of the genomic sequence, and the size of the spliced transcript varied from 1.4 kb to 1.5 kb. The coding region was 1260 bp, encoding a polypeptide of 419 amino acids (Petruzzella et al. 1998).

The splice variants found by comparing the EST and mRNA clone sequences are not likely to affect the structure of the translated BCS1L polypeptide, since the second exon containing the start codon and all six exons thereafter are identical in all splice variants. It is also unlikely that alternative 5'UTRs would affect the translation initiation, because an in-frame stop codon lies 42 bp before the start codon, making earlier translation initiation impossible. The use of a methionine further downstream as an alternative start codon is not likely either, since the next methionine is residue 48, which is located downstream from the only transmembrane domain of the protein. Alternative splicing of the 5'UTR could, however, enable tissue-specific up- or downregulation of *BCS1L* expression. The occurrence of the splice variants in different tissue and cell types and their importance for the regulation of gene expression remains to be verified by experimental tools.

Functional consequences of the S78G mutation on BCS1L (IV)

Expression, targeting, and stability of the mutant BCS1L

It has previously been shown that the human BCS1L protein is targeted unmodified to the inner mitochondrial membrane in vitro (Petruzzella et al. 1998). To study the effect of the S78G mutation on the synthesis, targeting and stability of the BCS1L protein, we subcloned the coding region of both wild-type and mutant *BCS1L* cDNA:s into the pCMV5 expression vector with a C-terminal FLAG tag. Intracellular targeting of the wild-type and mutant BCS1L polypeptides was studied in transiently transfected COS-1 cells by immunofluorescence microscopy. Staining with the anti-FLAG antibody completely overlapped with the mitochondrial staining both in the cells transfected with the wild-type construct and

in those transfected with the mutant construct (figure in IV), indicating that the S78G mutation does not influence the mitochondrial targeting of the BCS1L protein in vitro.

Western blot analysis of the transiently transfected cell lysates revealed a single ~46 kDa polypeptide. There was no obvious difference in size or intensity between the signal from the wild-type and mutant polypeptides (figure in IV), suggesting that the S78G mutation does not affect the synthesis or posttranslational modifications of the BCS1L polypeptide. Pulse chase experiments were carried out to monitor the stability of the mutant BCS1L protein. The half-life of the FLAG-tagged wild-type BCS1L polypeptide was 41 h, whereas that of the S78G mutant was only 14 h, suggesting that the S78G mutation decreases the stability of the BCS1L polypeptide in vitro (figure 8).

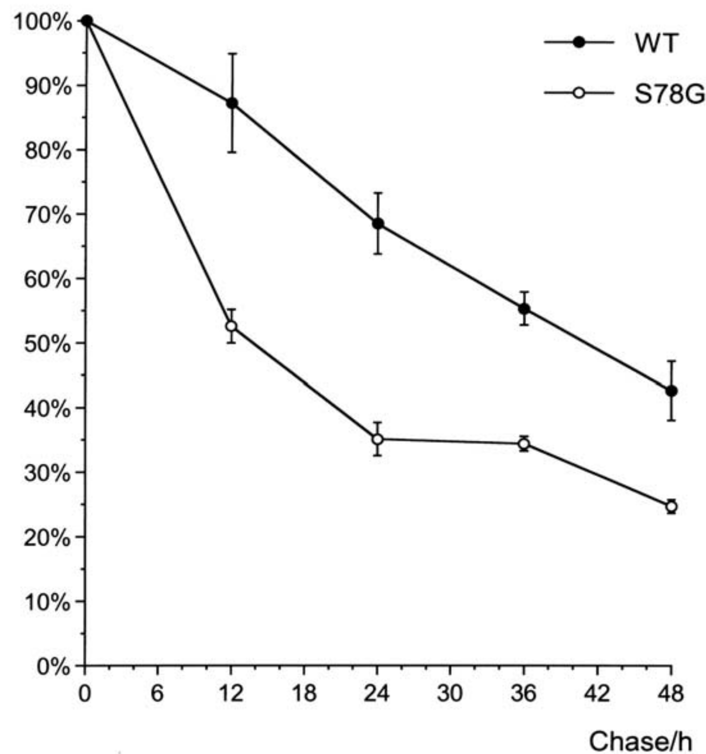


Figure 8
Stability of the wild-type and S78G mutant BCS1L polypeptides (IV). Transiently transfected COS-1 cells were pulse-labeled for one hour and chased for 0, 12, 24, 36 and 48 hours. The BCS1L polypeptides were immunoprecipitated, separated on 10% SDS-PAGE and visualized by fluorography. The radioactive bands were quantified by densitometry. The results are presented as percentages of the values at 0 h chase. The values on the curve are average values of duplicate experiments, with error bars showing the standard deviation.

Yeast complementation studies

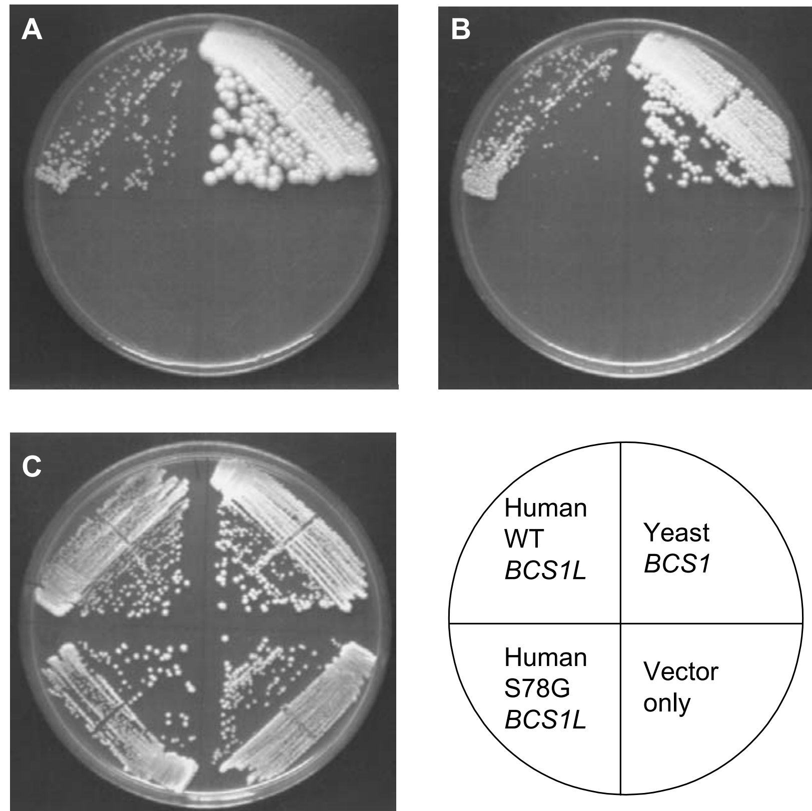
Yeast BCS1p protein has been shown to be essential for the assembly of the mitochondrial respiratory chain complex III (Nobrega et al. 1992; Cruciat et al. 1999). Yeast strains deleted for the *BCS1* gene cannot grow on glycerol, since on non-fermentable carbon sources, yeast cells are dependent on oxidative respiration. On glucose, however, *BCS1* deletion strains are able to grow utilizing anaerobic metabolism. Human BCS1L protein is highly homologous with yeast BCS1p protein (50% identity), and human *BCS1L* can partially rescue a yeast strain deleted for *BCS1* (de Lonlay et al. 2001). This facilitated the use of yeast complementation studies in monitoring the functional effects of the S78G mutation.

We tested the effect of the S78G mutation on the ability of human BCS1L to restore oxidative respiration in a *BCS1* deletion strain. The *BCS1* deletion yeast strain Y14211 was transformed with the wild type and 232A>G (S78G) mutant *BCS1L* cDNAs in a high-copy yeast expression plasmid pRS425. As a positive control, we also transformed the same strain with the wild-type yeast *BCS1* cDNA in the same vector, and as a negative control we used the plain vector. The purified transformants were streaked on YPEG (ethanol-glycerol), YPD and SD plates without leucine (vector selection), and incubated at 30ºC and 37ºC. On YPD and SD-leu plates all the transformants grew equally well at both temperatures, indicating that they were all viable on glucose. On YPEG plates, the yeast *BCS1* transformants were clearly growing after 48 h incubation. The wild-type human *BCS1L* transformants grew, but considerably more slowly, showing the first signs of growth after four days at 37ºC, and after six days at 30ºC. The S78G mutant and the vector-only transformants did not show any growth, even after incubation for two weeks (figure 9). This suggested that the S78G mutation disturbs the function of BCS1L.

Activity measurements of complex III in the Finnish GRACILE patients

One would expect that patients with *BCS1L* mutations would have decreased complex III activities in mitochondria. In the tissue samples of the three British patients (mutations identified in this study) and the five Turkish patients (mutations published earlier), the complex III activities have been shown to be decreased (Birch-Machin et al. 1989; Morris et al. 1995; de Lonlay et al. 2001). However, the Finnish GRACILE patients have been shown to have normal complex III activities in isolated mitochondria of all the tissues studied, and studies of oxygen consumption have also implied normal respiratory chain function (Fellman et al. 1998). Since it was surprising that the complex III activity seemed normal in the Finnish patients, we used all the available material and methods to

Figure 9
Yeast complementation experiment with *BCS1L* (IV). The *BCS1* deletion yeast strain was transformed with the wild-type and mutant *BCS1L* constructs in a high-copy yeast expression plasmid, pRS425. The wild type yeast *BCS1* cDNA in the same vector was used as a positive control, and the plain vector as a negative control. A) YPEG (ethanol-glycerol) plate incubated at 30ºC for 11 days. B) YPEG plate incubated at 37ºC for eight days. C) YPD (glucose) plate incubated at 30ºC for three days.



confirm it. The detailed results of the biochemical analyses are presented in the tables 3-5 in IV.

For respiratory chain enzyme activity measurements, fresh biopsy samples generally give the most reliable results, and we had appropriate age-matched controls available for muscle biopsy specimens. The respiratory chain enzyme activities measured in isolated mitochondria from skeletal muscle biopsies of five Finnish patients were within the normal range. Since these activities represented the values of indirect measurements of complex III, deduced from the combined activities with complex I and complex II, we also assessed complex III separately, using a deep frozen skeletal muscle sample from one patient. This sample showed moderately decreased complex III activity, being reduced by 1.5 SD from the controls' mean, in proportion to the citrate synthase activity. If this

reduction is not a post-mortem artefact due to a tissue preservation problem and reduced enzyme activities, then this mild decrease might have remained undetected in the indirect complex III analyses of the biopsy specimen.

Nuclear gene defects causing mitochondrial disorders often have a strictly tissue-specific expression (Papadopoulou et al. 1999). In the GRACILE syndrome the most relevant tissue for expressing complex III deficiency would be liver, since it is the major organ of dysfunction. The respiratory chain enzyme activities, including the separate assessment of complex III, were measured in tissue homogenates of deep frozen liver samples from three patients. In two of these samples all the mitochondrial enzyme activities were low, indicating a tissue preservation problem. However, in one liver sample the activities were in general within the normal range, and complex III also showed normal activity. Enzyme activity measurements carried out with isolated mitochondria from fresh liver specimens of four patients, obtained at autopsy shortly after death, also supported this finding by showing the presence of complex III activity. Unfortunately, appropriate controls for the fresh liver samples were not available, and thus fully reliable conclusions regarding these samples are not possible.

Respiratory chain enzyme activities were also measured in mitochondria isolated from myocardial, brain and kidney specimens of four patients. The tissue samples were obtained shortly after death, and the results did not indicate a complex III defect. Unfortunately, appropriate controls were available only for the myocardial samples and, regarding the other tissues, these results should be considered with caution.

Measurements of respiratory chain enzyme activities are highly sensitive to variations in experimental conditions (Thorburn 2000). They require fresh tissue specimens, or at least very well preserved deep frozen samples. The measurements should preferably be carried out using the relevant tissue, since respiratory chain enzyme defects are highly tissue-specific. The enzyme activities can show variation even within one tissue. Because healthy individuals show wide variation in the activities, the laboratories performing the measurements have to accumulate several matching controls. Especially regarding the complex III activity, measurements are notoriously difficult, and therefore some diagnostic laboratories routinely use combined (I+III, II+III, II and IV) measurements, whereas others consider separate measurement of complex III more accurate. For these reasons, it is not possible to draw definite conclusions concerning the complex III activity measurements of the Finnish GRACILE patients. If the patients were observed individually, and it were asked whether the results of respiratory chain enzyme activity measurements could explain the symptoms of the patients, the answer would definitively be that all the enzyme activities are within the normal reference range. However, when the Finnish GRACILE patients are viewed as a group, the

complex III activities have a tendency to be more often in the lower than in the higher half of the control range. Thus, we can quite certainly exclude a severe complex III defect in the Finnish patients, but not a mild reduction in activity. Even a mild defect could contribute to the lactic acidosis and other symptoms, but is very unlikely to explain them all. This suggests that BCS1L has some other, at present unknown, vital function in addition to complex III assembly.

Pathogenicity of the identified *BCS1L* mutations (IV)

Our conclusion that the S78G amino acid change in *BCS1L* really is a causative mutation, and not merely a harmless polymorphism, is based on sequence analyses of the Finnish and British patients and controls, as well as on the results of the pulse chase experiments and the yeast complementation studies.

The S78G mutation identified in the Finnish patients with GRACILE syndrome segregated without exception with the disease phenotype in all the families. Only one carrier of the mutation was found in the 544 control individuals screened, supporting the causative role of the mutation. The predicted effect of the serine-to-glycine change on the polypeptide is also considerable. Glycine, the smallest amino acid, enables the polypeptide backbone to make turns that are impossible with any other residue, and it lacks the hydroxyl group and the polarity of serine. S78 is also highly conserved across the species, implying that the amino acid residue is functionally important. (Comparisons between the amino acid sequences of different species were performed with the BLAST programs at NCBI and with the MultAlin alignment program (Corpet 1988).) On the other hand, both serine and glycine are small, uncharged amino acids, and a single amino acid change of this kind could also be a rare SNP occurring in the founder chromosome, located in the immediate vicinity of the real mutation and thus segregating with the disease phenotype. Therefore, additional evidence of the causative character of the S78G change was needed.

In a patient group with an identical or similar disease course, identification of different mutations in one gene can confirm the pathogenicity of the mutations. It was not possible to obtain this kind of proof by studying the Finnish GRACILE families, since they all carried the same founder mutation. However, sequence analysis of the three British patients (Morris et al. 1995), whose disease course closely resembled that of the Finnish GRACILE patients, revealed five different *BCS1L* mutations, including the Finnish S78G mutation.

The S78G mutation had functional consequences also. In the pulse chase experiments performed in COS-1 cells, the half-life of the mutant BCS1L was only one third of that of the wild-type polypeptide. This indicated that the amino acid change resulted in decreased stability of the protein. The effect could be

cumulative in vivo, leading to drastically decreased BCS1L levels in the patient tissues. On the other hand, a cellular feedback mechanism could increase the expression of the mutated BCS1L, and thus compensate for the accelerated degradation of the protein. Therefore the impact of the shortened half-life of the mutant BCS1L in the disease pathogenesis remains an open question.

The mutant human *BCS1L* was not able to rescue the yeast strain deleted for *BCS1*, whereas the wild-type human *BCS1L* was. This indicated a clear difference in function between the wild-type and the S78G mutant proteins. This result was interesting, since the growth defect in the *BCS1* deletion strain is probably caused by impaired complex III function, but the S78G mutation does not seem to affect complex III in human tissues. This suggests that BCS1L involves the mitochondrial respiratory function through some other mechanism also, in addition to complex III assembly. On the other hand, the wild-type human *BCS1L* was only partially able to complement the *BCS1* deletion yeast, and therefore the consequences of the S78G mutation on the respiratory chain may have appeared more serious in yeast complementation experiments than they are in vivo.

The pulse chase and yeast complementation experiments further supported the causative role of the Finnish S78G mutation. However, the identification of different mutations in the British patients, and especially the fact that one of the British infants was a compound heterozygote for the Finnish mutation and another mutation, provided the most convincing data confirming the pathogenic nature of the S78G mutation.

The pathogenicity of the individual British mutations has not yet been formally proved, but very likely the early stop mutation, the splice site mutation, and the R144Q and V327A missense mutations are pathogenic. These mutations were not present in the 140 controls screened. Unfortunately, no British controls were available, but screening of a control panel of 50 Caucasian individuals with ancestry in different parts of Europe, and additionally of 90 Finns, quite well excluded the possibility that the observed nucleotide variants were just common polymorphisms. The premature stop codon is a truncating mutation and can be considered pathogenic even without functional evidence. The functional significance of most of the other mutations is also obvious: The splice site mutation abolishes a classical splice donor consensus sequence, and is thus likely to result in intron retention and truncation of the protein. The R144Q mutation changes a basic and charged arginine into an acidic and uncharged glutamine. R144 is conserved in the mouse, and in both *Drosophila melanogaster* and *Saccharomyces cerevisiae* it is relatively well conserved, being replaced by lysine, another basic and charged amino acid. The V327A change appears the least dramatic of the current mutations, since both valine and alanine are small, nonpolar amino acids without functional groups. V327 is not strictly conserved either, since in *Saccharomyces cerevisiae* it

is replaced by threonine, a polar amino acid. However, since the early stop mutation found in the other allele of the same patient is an obligatory mutation, we consider this V327A substitution a likely mutation as well.

The importance of the heterozygous single nucleotide change in the first intron of *BCS1L* in British patient 2 is questionable. It was not present in 140 controls, implying that it does not represent a common SNP. However, the functional importance of the region is not supported by the phylogenetic conservation, since the 20 base pairs surrounding the nucleotide variant are not conserved in the mouse. In her other allele, the patient had the splice site mutation, which was heterozygous as well, but we did not identify any other possible mutations in the DNA of this infant. However, we cannot exclude the possibility of a regulatory mutation located far upstream.

Phenotype heterogeneity of the *BCS1L* mutations (IV)

All the Finnish GRACILE patients had the same homozygous *BCS1L* mutation, whereas the three British patients were compound heterozygotes for different mutations. The sequence analysis of the five patients with Turkish origin (de Lonlay et al. 2001) had previously revealed four other *BCS1L* mutations, all being missense mutations. The symptoms of the patients with the different *BCS1L* mutations are summarized in table 6. The common clinical characteristics of the patients included lactacidosis, tubulopathy, liver dysfunction and cholestasis. Fetal growth retardation was severe in the Finnish patients and moderate in the British infants, but only slight or non-existent in the Turkish patients. The Finnish patients had normal muscle tonus and no neurological or neuropathological abnormalities, whereas all the Turkish patients except one had severe neurological problems: One baby with the S277N mutation had encephalopathy with deafness and blindness, and her sister, who was alive at the age of nine years, had cerebral atrophy and severe psychomotor retardation. The Turkish infants with the P99L mutation had progressive neurological problems as well, including ventilation abnormalities of neurological origin and brainstem lesions compatible with Leigh syndrome. The British patients had muscle hypotonia and one of them developed seizures and coma at the age of one day. All the Finnish and British infants died before the age of five months, half of them already during the first days of life. The Turkish patients survived, on average, longer. One essential difference in the clinical data of the patient groups was that the respiratory chain complex III activities of the Finnish patients were within the normal range, whereas the other patients showed clearly decreased complex III activities in enzyme activity measurements. Unfortunately, iron metabolism was not studied in either the British or the Turkish patients. Thus, we do not yet know whether the abnormalities in iron metabolism

Table 6
Comparison of the symptoms caused by different *BCS1L* mutations.

BCS1L mutation(s) [a]	S78G	R56STOP and V327A	Splice site mutation and an intronic nucleotide change [b]
Number of patients	17 (11 families)	1	1
Origin of the patient(s)	Finnish	British	British
Male / female	5/12	m	f
Birth weight, g	median 1560	2020	1830
Birth length, cm	median 43	47	45
Head circum-ference, cm	median 31	31	31
Gestational age, weeks	median 38	38	39
Age at death	1 day – 4 months	2 days	1 month
Lactacidosis	yes	yes	yes
Tubulopathy	yes	yes	yes
Iron overload	yes	?	?
Cholestasis	yes	no	yes
Liver dysfunction	yes	?	yes
Muscle hypotonia	no	yes	yes
Neurological symptoms	no	seizures and coma	no
Complex III activity	normal [c]	decreased	decreased
References	Fellman et al. 1998, Rapola et al. 2002, IV	Birch-Machin et al. 1989, Morris et al. 1995, IV	Morris et al. 1995, IV

[a] If one mutation mentioned, the patients are homozygous for that mutation. If two different mutations, the patients are compound heterozygotes.
[b] Pathogenicity of these intronic nucleotide changes has not been confirmed.
[c] The enzyme activities are within the normal range, but a mild defect cannot be excluded.

observed in the Finnish patients were characteristic of the S78G mutation only, or whether the other mutations in *BCS1L* would also cause similar findings.

It is not uncommon for different mutations in one gene to cause different disease phenotypes. For example, different mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene can result in typical cystic fibrosis, characterized by severe lung disease, pancreatic exocrine insufficiency, an

S78G and R144Q	S277N	P99L	R155P and V353M
1	2 (siblings)	2 (not related)	1
British	Turkish	Turkish	Turkish
f	f	m / f	m
1840	2490 / 2940	2340 / 2620	3330
48	46 / 48	46 / 49	53
31	31	?	?
38	38 / 39	40 / 41	39
3 months	3 months / alive when 9 years old	6 months / 2 years	alive when 5 months old
yes	yes	yes	yes
yes	yes	yes	yes
?	?	?	?
yes	yes	?	?
yes	yes	yes	yes
yes	?	yes / ?	?
no	severe encephalopathy	Leigh Syndrome	no
decreased	decreased	decreased	decreased
Morris et al. 1995, IV	de Lonlay et al. 2001	de Lonlay et al. 2001	de Lonlay et al. 2001

increase in the concentration of sweat electrolytes, and male infertility, or solely male infertility with no other symptoms (Stuhrmann and Dork 2000). However, the BCS1L mutations cannot (at all) be grouped as “severe” and “mild” mutations. If the degree of fetal growth restriction or the survival time of the patients were the criteria, the Finnish and British mutations would be the most severe. However, considering the neurological problems, one British and four Turkish patients were seriously ill, whereas two British babies only had muscle hypotonia, and the Finnish infants were asymptomatic. Interestingly, muscle hypotonia and

neurological problems, for example encephalopathy, seizures and Leigh syndrome, are typical of mitochondrial disorders, and the occurrence of these symptoms correlated with decreased complex III activity in measurements of respiratory chain enzyme activities. One tempting explanation for this kind of variation in the symptoms and findings caused by *BCS1L* mutations would be that the human BCS1L protein has a dual function. Some BCS1L mutations essentially disturb the complex III assembly, whereas other mutations, for example the Finnish S78G mutation, severely damage the other, at present unknown, function of the BCS1L. Probably most of the mutations affect both functions to some extent. So far, *BCS1L* mutations are the only nuclear gene defects described that involve the function of complex III, and thus we cannot compare the consequences of the *BCS1L* mutations with isolated complex III defects. The published complex III defects (Andreu et al. 1999; Valnot et al. 1999; Keightley et al. 2000; Legros et al. 2001; Wibrand et al. 2001) have been caused by sporadic, heteroplasmic mutations in the mitochondrial cytochrome *b* gene, the situation being different compared to autosomal recessive inheritance.

What other type of function could BCS1L perform? The best known function of mitochondria is to supply cellular energy, but mitochondria perform numerous other tasks as well. One of the vital mitochondrial functions is the assembly of FeS clusters (Lill and Kispal 2000; Muhlenhoff and Lill 2000), which serve as electron carriers in the mitochondrial respiratory chain, and as important cofactors in several other proteins located in the mitochondria, cytosol and nucleus. Some ten mitochondrial proteins involved in the synthesis and transfer of the FeS clusters have been identified in *Saccharomyces cerevisiae* so far, but the details of this machinery are still largely unknown. It is an interesting question, whether BCS1L has a role to play in the synthesis or transport of FeS proteins in general, in addition to the fact that it is known to be necessary for correct assembly of the Rieske protein subunit, which also contains an FeS cluster, to the complex III.

Presumably, the function of BCS1L is linked with cellular iron metabolism, since at least the Finnish patients with the homozygous S78G mutation have a marked iron overload. Iron metabolism is tightly regulated, since iron is an essential cofactor in various cellular processes, but free iron is toxic because of its propensity to induce the formation of free radicals (Andrews 2000; Walker et al. 2001). A few mitochondrial diseases associated with disturbances in cellular iron metabolism have been described to date: X-linked sideroblastic anemia, X-linked sideroblastic anemia with ataxia and Friedreich ataxia (Andrews 2000). Like the GRACILE syndrome, they are caused by mutations in nuclear genes, but both their clinical and their histological picture are different. They are characterized by iron accumulation inside mitochondria, whereas, in electron microscopy of liver and skeletal muscle specimens, the mitochondria of GRACILE

patients appear normal (Rapola et al. 2002). The excess iron in the hepatocytes is, instead, stored in cytosolic hemosiderin granules and ferritin particles. BCS1L could possibly be involved in iron transport from the cytosol to the mitochondria. On the other hand, it is possible that the abnormalities of iron metabolism in GRACILE patients are secondary to the liver disease, and that there is no direct link between BCS1L and cellular iron traffic.

DNA diagnostics of the GRACILE syndrome (I, II, IV)

The assignment of the GRACILE locus already enabled linkage-based prenatal diagnostics for families with at least one previously affected child. Now the Finnish S78G mutation can be directly identified, and DNA diagnosis is possible also in families without a previously affected baby. A specific diagnosis can help to avoid unnecessary diagnostic tests and is psychologically important for the parents. In addition, the mutation-based DNA diagnostics enables reliable and easy carrier diagnosis for healthy siblings and other relatives, as well as for their spouses, making genetic counselling easier.

Prenatal diagnosis early in pregnancy is possible only if the carrier state of the parents is known in advance. Otherwise, the disease cannot be suspected until the end of the second trimester, when fetal growth retardation can be detected by ultrasound. However, even later in pregnancy a reliable DNA diagnosis of the GRACILE syndrome can be helpful, because it can prevent an unnecessary caesarean section by explaining the reason for the fetal growth retardation.

For non-Finnish patients, the sequence analysis of the *BCS1L* gene is a diagnostic option. So far, all Finnish GRACILE patients have had the same S78G founder mutation, but it is likely that other, rare mutations exist also in Finland. Thus, if BCS1L mutations are suspected on the basis of the patient’s symptoms, but the patient does not have the S78G mutation, or is heterozygous for it, sequencing of the *BCS1L* gene should be considered for Finnish patients also.

CONCLUDING REMARKS

In this study, we localized the causative gene for the GRACILE syndrome using linkage, linkage disequilibrium and ancestral haplotype analyses, and then identified the causative gene by analyzing the positional candidates in the critical chromosomal region. This project was started in 1996, when the use of microsatellite markers was routine and dense genetic maps were available in databases. Also, many genes and thousands of ESTs had been sequenced and localized to human chromosomes. Nevertheless, positional cloning projects of monogenic diseases were laborious. After initial localization of the GRACILE locus in 1997, we performed RH mapping of the critical markers to be able to determine exactly the critical chromosomal interval, and to create the basis for a physical map of the region. However, when the RH mapping was completed, the HGP was speeding up both the large-scale sequencing of the human genome and the identification of the genes, and traditional physical mapping in this project turned out to be unnecessary.

Nowadays, the positional candidate approach has practically replaced "pure" positional cloning. When a disease locus has been identified, the sequences of positional candidate genes are available in internet databases. If the sequence of a critical region is not complete, the physical map can be built by biocomputing methods, and new genes are cloned mostly "in silico". A huge amount of laboratory work can be avoided, as compared with the situation five years ago. Mapping and identification of a disease gene for a monogenic disorder can be perfectly straightforward and take only a few months. In many cases, however, it is still not easy. First, the sequencing of the human genome is not yet totally completed, for there are still gaps in the sequence. Secondly, although genes, especially their coding regions, can be identified fairly easily with gene prediction programs and by comparing genomic and EST sequences, the exact structure of many genes is not yet known. The regulatory regions of the genes, and their function, are for the most part still unidentified. In this study we observed several different splice variants of the 5'UTR of the *BCS1L*, but we do not know anything yet about the regulation or significance of the variable splicing. The same is true of many genes, concerning also the coding regions: the surprising result of the analysis of the initially sequenced human genome was the small total number of human genes, while the commonness of variable splicing contributes to the diversity of the human proteins.

Identification of the disease gene provides the starting point for revealing the pathogenesis of the GRACILE syndrome. Surprisingly, the respiratory chain complex III activities were normal or only slightly decreased in the Finnish GRACILE patients, despite the *BCS1L* mutation. This suggested that BCS1L has

another, as yet unknown, function, in addition to complex III assembly. Further studies of the BCS1L should address the interesting question of how the S78G mutation in BCS1L can cause a lethal disease without considerably affecting complex III function. So far, only two functional domains of BCS1L have been recognized (figure 7B). Residues 224-344 represent the conserved AAA domain, and residues 9-32 are hydrophobic and correspond to the transmembrane region of the yeast Bcs1p. Two of the missense mutations so far identified in BCS1L are located in the AAA domain, whereas four mutations, including the Finnish S78G mutation, lie between the transmembrane and AAA domains. The functional domains of this part of the protein remain to be characterized and could potentially be associated with the as yet uncharacterized function of BCS1L. Our other hypothesis is based on the pulse chase experiments, which revealed that the S78G mutation accelerates the degradation of BCS1L *in vitro*. If this were also to result in decreased protein levels in tissues *in vivo,* without the mutation interfering in protein conformation, the reduced amount of protein in the mitochondria might be sufficient to perform complex III assembly, whereas the other function could be compromised.

Identification of the gene defect underlying the GRACILE syndrome provides a starting point for more detailed studies of the function of BCS1L and proteins interacting with it in cellular energy and iron metabolism. From the clinical point of view, this study has produced DNA diagnostics for this lethal disease, enabling prenatal diagnosis and carrier testing in Finnish GRACILE families.

ACKNOWLEDGEMENTS

The work for this thesis was carried out between 1996-2002 in the Department of Molecular Medicine at the National Public Health Institute in Helsinki. I wish to thank the head of the Institute, Professor Jussi Huttunen, for the excellent research facilities. I also want to thank the David Geffen School of Medicine at UCLA, Los Angeles, for the excellent opportunity to carry out part of this work at the Department of Human Genetics in 2000-2002.

I wish to express my gratitude to all the people who have helped and supported me and made this study possible with their efforts, and especially I wish to thank:

My supervisor, Professor Leena Peltonen-Palotie, for personal guidance and for creating an "everything is possible" spirit in the lab. Despite her busy schedule, she somehow always managed to be available when I most needed her advice or support – by phone, email or fax, if we were on different continents. I have learned a lot from her ability to perceive the essential scientific findings among the details, and to separate important matters from secondary issues. I also greatly appreciate that she gave me the opportunity to work abroad and attend scientific conferences.

Docent Vineta Fellman, whose thorough work in characterizing the GRACILE syndrome and collecting the clinical data and patient samples made this study possible. She has continuously shown an interest in my work, encouraged me, and taken care of all the clinical issues in this project.

Docent Anu Wartiovaara (Suomalainen) for answering my numerous questions concerning cell biology and mitochondria. Being far away in Montreal – or on maternity leave – did not stop her from teaching me the principles of mitochondrial research and cheering me up with her long, thorough and funny emails.

Dr Jouni Vesa for teaching me the basic molecular biology methods. He explained all the procedures in detail to me, instructed me about everything on the bench step by step, and always took time to check how my cell cultures were doing, if I wasn't sure. And he made our lunch and coffee breaks at UCLA hilarious.

Dr Teppo Varilo for his efforts in the genealogical studies, and for advice whenever I was puzzled about issues concerning genealogy, carrier frequencies, etc.

Professor Greg Payne at UCLA for providing me with the possibility to carry out the yeast experiments in his lab, and all the nice people in his laboratory for help and company, especially Jenna, Giancarlo and Todd. Professor Marja Makarow and Dr Vijay Kumar for help with the planning of the yeast experiments.

Professor Douglass Turnbull, Dr Robert Taylor and Ayan Dasvarma for flexible and pleasant collaboration.

Professor Pertti Aula and Docent Kirsi Huoponen for reviewing this thesis in a tight schedule, and for their constructive suggestions. Dr Johanna Hästbacka for her valuable comments and encouragement as a member of my thesis committee.

Rector Kari Raivio for advice and for his interest in my work. Docent Anu Jalanko and docent Ismo Ulmanen for help and advice in numerous matters, both scientific and practical, and for creating a pleasant atmosphere in the lab. Professor Aarno Palotie and docent Ann-Christine Syvänen for helpful discussions along the way.

Dr Paulina Paavola, Docent Riitta Salonen, Anne Jokiaho and Lisa Lanyi for pleasant collaboration in my other project, the mapping of the hydrolethalus syndrome. Paulina took me under her wings when I started in the lab, making the beginning easier. Riitta has been a great support to me, as well as enjoyable company. She has also taught me a lot about clinical genetics. Anne's efficiency is incredible. She always works in the lab really fast, but without any mistakes and without seeming at all stressed. Lisa's capability to spread good humour in the lab is admirable, and I wish her good luck with the project in the future.

Sari Kivikko, Sari Mustala, Minna Partanen, Diane Ohkawahira and Anne Carson for their help with all the paperwork, and for making sure that everything works when Leena and her students travel between continents. Tero Hiekkalinna, Teemu Perheentupa, Joyce Tang, Aki Suomalainen, Jari Raikko and all the other computer support persons in Helsinki and at UCLA. Their help and patience have been irreplaceable.

All the people in our labs in Helsinki and at UCLA for their cheerful company, help and advice; Henna and Mira for sharing the ups and downs of the gene hunt with me; Naula for her constant readiness to help with everyday lab problems, and for reminding me every now and then that life is not that serious after all; Päivi P. for advice on several issues, as well as for many funny moments in and outside the lab; Arja for her flexible help with everything regarding the DNA extractions; Soili for her ability to always say something comforting to a stressed student; Tarja for being always good-humoured and for patiently answering my numerous questions concerning the PhD thesis bureaucracy; Pekka for making my life easier regarding the use of the pipetting robot and ABIs, and for always managing to rescue my lost ABI files from the network; Tuula for running perfect RNA gels; Tintti and Maikki for their great help with sequencing; Kathrin, Janna, Tutsa, and Lennu for helping me to get started in the lab at UCLA; Juha P. for getting us a nice apartment in Westwood; Jenni, Emma, Jesper, Ville, and Niklas for the memorable Sestri-parties; Joni and Elina for joyful moments in and

outside the lab in Los Angeles; Jyrki, Lasse, and Tomi for their jokes; Chris and Jenny for their parties; Johanna and Petra P. for explaining the basics of genetic mapping to me; Maija, Hannele, Anna-Liisa, and Marjo for company and support; Mervi E., Elli, Eija, Päivi T., Ritva, Anne V., Jaana, Anne N., Sisko and Paula for expert advice and assistance in different lab procedures.

All my friends, especially Päivi, Hanna, Raila, Piia, Aino, Heini, Heidi, Kati and Lissu, for encouragement, for making me forget my work and studies sometimes, and for writing me long emails while I was in USA.

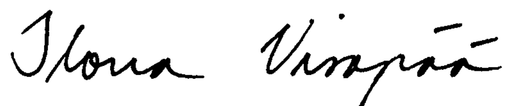
My parents for their love and support. My mother for long phone calls whenever I needed her to listen, and my father for his jokes, and for showing me that math and science are fun when I was a kid.

My husband Harri for his love, understanding and unfailing support. If I was tired or disappointed, he comforted me, and when things were going well in the lab, he was always ready to celebrate. And if I was busy and stressed, he offered practical help, covering everything from proof-reading my articles to bringing sandwiches to me in the lab.

All the families who participated in this study by giving blood samples and providing genealogical data.

This study has been supported financially by the Helsinki Biomedical Graduate School, the Ulla Hjelt Fund, the Academy of Finland, the Maud Kuistila Foundation, the Finnish Medical Society Duodecim and the Paulo Foundation.

Helsinki, October 2002



ELECTRONIC DATABASE INFORMATION

Internet databases are essential tools of molecular genetics, and an increasing number of references in the literature are nowadays provided as electronic references, instead of traditional printed ones. Internet addresses for the databases, programs and organizations mentioned in the text are provided below.

BLAST service at NCBI, http://www.ncbi.nlm.nih.gov/BLAST/
Celera, http://www.celera.com
Electronic PCR at NCBI, http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi
European Bioinformatics Institute (EBI), http://www.ebi.ac.uk/
GenBank at NCBI, http://www.ncbi.nlm.nih.gov/Genbank/
Genethon, http://www.genethon.fr/php/index.php
The Genetic Location Database (LDB) of University of Southampton, http://cedar.genetics.soton.ac.uk/public_html/ldb.html
Genome Data Base (GDB), http://gdbwww.gdb.org/
Human Genome Organization (HUGO), http://www.gene.ucl.ac.uk/hugo/
The Map Viewer at NCBI, http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch
Multalin alignment program, http://prodes.toulouse.inra.fr/multalin/multalin.html
National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/
National Human Genome Research Institute (NHGRI), http://www.genome.gov/
National Institutes of Health (NIH), http://www.nih.gov/
A new gene map of the human genome at NCBI (GeneMap'98), http://www.ncbi.nlm.nih.gov/genemap99/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
Sequence utilities of the Search Launcher at Baylor College of Medicine (BCM), http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html
The SNP database (dbSNP) at NCBI, http://www.ncbi.nlm.nih.gov/SNP/
Stanford Human Genome Center (SHGC), http://shgc-www.Stanford.edu/
University of California Santa Cruz (UCSC) Genome bioinformatics, http://genome.ucsc.edu/
Whitehead Institute, http://www-genome.wi.mit.edu/

REFERENCES

Aaltonen J, Komulainen J, Vikman A, Palotie A, Wadelius C, Perheentupa J, and Peltonen L. Autoimmune polyglandular disease type I: Exclusion map using amplifiable multiallelic markers in a microtiter well format. *Eur J Hum Genet* 1: 164-171, 1993

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, and Pfeiffer BD. The genome sequence of Drosophila melanogaster. *Science* 287: 2185-2195, 2000

Aittomäki K, Lucena JLD, Pakarinen P, Sistonen P, Tapanainen J, Gromoll J, Kaskikari R, Sankila EM, Lehväslaiho H, Engel AR, Nieschlag E, Huhtaniemi I, and de la Chapelle A. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotrophic ovarian failure. *Cell* 82: 959-968, 1995

Allikmets R, Raskind WH, Hutchinson A, Schueck ND, Dean M, and Koeller DM. Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A). *Hum Mol Genet* 8: 743-749, 1999

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. Basic local alignment search tool. *J Mol Biol* 215: 403-410, 1990

Aminoff M, Carter JE, Chadwick RB, Johnson C, Gräsbeck R, Abdelaal MA, Broch H, Jenner LB, Verroust PJ, Moestrup SK, de la Chapelle A, and Krahe R. Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. *Nat Genet* 21: 309-313, 1999

Andersson S, Davis DL, Dahlback H, Jornvall H, and Russell DW. Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J Biol Chem* 264: 8222-8229, 1989

Andreu AL, Hanna MG, Reichmann H, Bruno C, Penn AS, Tanji K, Pallotti F, Iwata S, Bonilla E, Lach B, Morgan-Hughes J, and DiMauro S. Exercise intolerance due to mutations in the cytochrome b gene of mitochondrial DNA. *N Eng J Med* 341: 1037-1044, 1999

Andrews N. Iron metabolism: iron deficiency and iron overload. *Annu Rev Genomics Hum Genet* 01: 78-98, 2000

Avela K, Lipsanen-Nyman M, Idänheimo N, Seemanova E, Rosengren S, Mäkelä TP, Perheentupa J, de la Chapelle A, and Lehesjoki AE. Gene encoding a new RING-B-box-Coiled-coil protein is mutated in mulibrey nanism. *Nat Genet* 25: 298-301, 2000

Becker E, and Richardson DR. Frataxin: its role in iron metabolism and the pathogenesis of Friedreich's ataxia. *Int J Biochem Cell Biol* 33: 1-10, 2001

Beinert H. Iron-sulfur proteins: ancient structures, still full of surprises. *J Biol Inorg Chem* 5: 2-15, 2000

Beinert H, and Kiley PJ. Fe-S proteins in sensing and regulatory functions. *Curr Opin Chem Biol* 3: 152-157, 1999

Berg JM, Tymoczko JL, and Stryer L. Biochemistry, Chapter 31, The control of gene expression. *W. H. Freeman and Company, USA*: 867-894, 2001a

Berg JM, Tymoczko JL, and Stryer L. Biochemistry, Part II, Transducing and storing energy. *W. H. Freeman and Company, USA*: 371-662, 2001b

Beutler E, Gelbart T, Lee P, Trevino R, Fernandez MA, and Fairbanks VF. Molecular characterization of a case of atransferrinemia. *Blood* 96: 4071-4074, 2000

Birch-Machin MA, Shepherd IM, Watmough NJ, Sherrat HS, Bartlett K, Darley-Usmar VM, Milligan DW, Welch RJ, Aynsley-Green A, and Turnbull DM. Fatal lactic acidosis in infancy with a defect of complex III of the respiratory chain. *Pediatr Res* 25: 553-559, 1989

Boehnke M, Lange K, and Cox DR. Statistical methods for multipoint radiation hybrid mapping. *Am J Hum Genet* 49: 1174-1188, 1991

Borsani G, Bassi MT, Sperandeo MP, De Grandi A, Buoninconti A, Riboni M, Manzoni M, Incerti B, Pepe A, Andria G, Ballabio A, and Sebastio G. SLC7A7, encoding a putative permease-related protein, is mutated in patients with lysinuric protein intolerance. *Nat Genet* 21: 297-301, 1999

Bourgeron T, Chretien D, Poggi-Bach J, Doonan S, Rabier D, Letouze P, Munnich A, Rotig A, Landrieu P, and Rustin P. Mutation of the fumarase gene in two siblings with progressive encephalopathy and fumarase deficiency. *J Clin Invest* 93: 2514-2518, 1994

Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgerois M, Viegas-Pequignot E, Munnich A, and Rotig A. Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nat Genet* 11: 144-149, 1995

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, and Cocozza S. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271: 1423-27, 1996

Casari G, De Fusco M, Ciarmatori S, Zeviani M, Mora M, Fernandez P, De Michele G, Filla A, Cocozza S, Marconi R, Durr A, Fontaine B, and Ballabio A. Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 93: 973-983, 1998

The C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* 282: 2012-2018, 1998

Cohen D, Chumakov I, and Weissenbach J. A first-generation physical map of the human genome. *Nature* 366: 698-701, 1993

Collins F. Positional cloning moves from perditional to traditional. *Nat Genet* 9: 347-350, 1995

Cormand B, Avela K, Pihko H, Santavuori P, Talim B, Topaloglu H, de la Chapelle A, and Lehesjoki AE. Assignment of the muscle-eye-brain disease gene to 1p32-p34 by linkage analysis and homozygosity mapping. *Am J Hum Genet* 64: 126-135, 1999

Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucl Acids Res* 16: 10881-10890, 1988

Cox DR, Burmeister M, Price ER, Kim S, and Myers RM. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250: 245-250, 1990

Cox TC, Bottomley SS, Wiley JS, Bawden MJ, Matthews CS, and May BK. X-linked pyridoxine-responsive sideroblastic anemia due to a Thr388-to-Ser substitution in erythroid 5-aminolevulinate synthase. *N Engl J Med* 330: 675-679, 1994

Cruciat CM, Hell K, Fölsch H, Neupert W, and Stuart RA. Bcs1p, an AAA-family member, is a chaperone for the assembly of the cytochrome bc_1 complex. *EMBO J* 18: 5226-5233, 1999

Cuthbert AP, Trott DA, Ekong RM, Jezzard S, England NL, Themis M, Todd CM, and Newbold RF. Construction and characterization of a highly stable human:rodent monochromosomal hybrid panel for genetic complementation and genome mapping studies. *Cytogenet Cell Genet* 71: 68-76, 1995

de la Chapelle A. Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet* 30: 857-865, 1993

de la Chapelle A, and Wright FA. Linkage disequilibrium mapping in isolated populations: The example of Finland revisited. *Proc Natl Acad Sci* 95: 12416-12423, 1998

de Lonlay P, Valnot I, Barrientos A, Gorbatyuk M, Tzagoloff A, Taanman JW, Benayoun E, Chretien D, Kadhom N, Lombes A, de Baulny HO, Niaudet P, Munnich A, Rustin P, and Rötig A. A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. *Nat Genet* 29: 57-60, 2001

Delettre C, Lenaers G, Griffoin JM, Gigarel N, Lorenzo C, Belenguer P, Pelloquin L, Grosgeorge J, Turc-Carel C, Perret E, Astarie-Dequeker C, Lasquellec L, Arnaud B, Ducommun B, Kaplan J, and Hamel CP. Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat Genet* 26: 207-10, 2000

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriquez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, Bentolila S, Bihoreau MT, Birren BB, Browne J, Butler A, Castle AB, Chiannilkulchai N, Clee C, Day PJR, Dehejia A,

Dibling N, Drouot N, Duprat S, Fizames C, Fox S, et al. A physical map of 30,000 human genes. *Science* 282: 744-746, 1998

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathorp M, Gyapay G, Morissette J, and Weissenbach J. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152-154, 1996

Fellman V, Bondsdorff L, and Parkkinen J. Exogenous apotransferrin and exchange transfusions in hereditary iron overload disease. *Pediatrics* 105: 398-401, 2000

Fellman V, Rapola J, Pihko H, Varilo T, and Raivio KO. Iron-overload disease in infants involving fetal growth retardation, lactic acidosis, liver haemosiderosis, and aminoaciduria. *Lancet* 351: 490-493, 1998

The Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet* 17: 399-403, 1997

Fölsch H, Guiard B, Neupert W, and Stuart RA. Internal targeting signal of the BCS1 protein: a novel mechanism of import into mitochondria. *EMBO J* 15: 479-487, 1996

Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, Eaton DH, Vehar GA, Capon DJ, and Lawn RM. Characterization of the human factor VIII gene. *Nature* 312: 326-330, 1984

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, and Oliver SG. Life with 6000 genes. *Science* 274: 563-567, 1996

Goss SJ, and Harris H. New method for mapping genes in human chromosomes. *Nature* 255: 680-684, 1975

Hackman P, Vihola A, Haravuori H, Marchand S, Sarparanta J, De Seze J, Labeit S, Witt C, Peltonen L, Richard I, and Udd B. Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *Am J Hum Genet* 71: 492-500, 2002

Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D, and McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 30: 52-55, 2002

Hansen JJ, Durr A, Cournu-Rebeix I, Georgopoulos C, Ang D, Nyholm Nielsen M, Davoine CS, Brice A, Fontaine B, Gregersen N, and Bross P. Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin hsp60. *Am J Hum Genet* 70: 1328-1332, 2002

Hayashi A, Wada Y, Suzuki T, and Shimizu A. Studies on familial hypotransferrinemia: unique clinical course and molecular pathology. *Am J Hum Genet* 53: 201-213, 1993

Hietala M. Prospects for genetic screening in Finland; Evaluation of the feasibility of carrier screening in primary health care using aspartylglucosaminuria carrier screening as the model. *PhD thesis, University of Turku, Turku, Finland*, 1998

Hietala M, Hakonen A, Aro AR, Niemelä P, Peltonen L, and Aula P. Attitudes toward genetic testing among the general population and relatives of patients with a severe genetic disease: a survey from Finland. *Am J Hum Genet* 56: 1493-1500, 1995

Hiilesmaa V, and Salonen R. Äitiyshuollon seulonnan mahdollisuudet ja ongelmat (in Finnish). *Duodecim* 116: 880-886, 2000

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, and Freimer NB. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat Genet* 8: 380-386, 1994

Hudson T, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Babtista R, Kruglyak L, Xu SH, Hu X, Colbert AME, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAngelis MM, and Ingalls KA. An STS-based map of the human genome. *Science* 270: 1919-1920, 1995

Hästbacka J, Chapelle Adl, Kaitila I, Sistonen P, Weaver A, and Lander E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2: 204-211, 1992

Hästbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, Coloma A, Lovett M, Buckler A, Kaitila I, and Lander ES. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 78: 1073-1087, 1994

Höglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, Karjalainen-Lindsberg ML, Airola K, Holmberg C, de la Chapelle A, and Kere J. Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. *Nat Genet* 14: 316-319, 1996

Ikonen E, Baumann M, Gron K, Syvänen A, Enomaa N, Halila R, Aula P, and Peltonen L. Aspartylglucosaminuria: cDNA encoding human aspartylglucosaminidase and the missense mutation causing the disease. *EMBO J* 10: 51-58, 1991

The International Batten Disease Consortium. Isolation of a novel gene underlying Batten disease, CLN3. *Cell* 82: 949-957, 1995

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921, 2001

Isola J, DeVries S, Chu L, Ghazvini S, and Waldman F. Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples. *Am J Path* 145: 1301-1308, 1994

Ito H, Fukuda Y, Murata K, and Kimura A. Transformation of intact yeast cells treated with alkali cations. *J Bacteriol* 153: 163-168, 1983

Joensuu T, Hämäläinen R, Yuan B, Johnson C, Tegelberg S, Gasparini P, Zelante L, Pirvola U, Pakarinen LM, Lehesjoki AE, de la Chapelle A, and Sankila EM. Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3. *Am J Hum Genet* 69: 673-684, 2001

Järvelä I, Sabri Enattah N, Kokkonen J, Varilo T, Savilahti E, and Peltonen L. Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene. *Am J Hum Genet* 63: 1078-1085, 1998

Kallinen J. Towards genetic testing in antenatal care. *PhD Thesis, University of Kuopio, Kuopio, Finland*, 2002

Karlowicz MG, and Adelman RD. What are the possible causes of neonatal nephrocalcinosis? *Semin Nephrol* 18: 364-367, 1998

Keightley JA, Anitori R, Burton MD, Quan F, Buist NRM, and Kennaway NG. Mitochondrial encephalomyopathy and complex III deficiency associated with a stop-codon mutation in the cytochrome b gene. *Am J Hum Genet* 67: 1400-1410, 2000

Kere J. Geenitesteissäkö kohtalomme avaimet? (in Finnish). *Duodecim* 116: 874-879, 2000

Kestilä M, Lenkkeri U, Männikkö M, Lamerdin J, McCready P, Putaala H, Ruotsalainen V, Morita T, Nissinen M, Herva R, Kashtan CE, Peltonen L, Holmberg C, Olsen A, and Tryggvason K. Positionally cloned gene for a novel glomerular protein - nephrin - is mutated in congenital nephrotic syndrome. *Mol Cell* 1: 575-582, 1998

Kispal G, Csere P, Guiard B, and Lill R. The ABC transporter Atm1p is required for mitochondrial iron homeostasis. *FEBS Lett* 418: 346-350, 1997

Kispal G, Csere P, Prohl C, and Lill R. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J* 18: 3981-3989, 1999

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, and Long JC. Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 62: 1171-1179, 1998

Knisely AS, Magid MS, Dische R, and Cutz. Neonatal hemochromatosis. *Birth Defects Orig Artic Ser* 23: 75-102, 1987

Koehler CM, Lauenberger D, Merchant S, Renold A, Junne T, and Schatz G. Human deafness dystonia syndrome is a mitochondrial disease. *Proc Natl Acad Sci* 96: 1817-1819, 1999

Kure S, Takayanagi M, Narisawa K, Tada K, and Leisti J. Identification of a common mutation in Finnish patients with nonketotic hyperglycinemia. *J Clin Invest* 90: 160-164, 1992

Kääriäinen H. Geeniseulontojen riskit ja haitat (in Finnish). *Duodecim* 116: 909-913, 2000

Kääriäinen H, Ryöppy S, and Norio R. RAPADILINO syndrome with radial and patellar aplasia/hypoplasia as main manifestations. *Am J Med Genet* 33: 346-351, 1989

Lander ES, and Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1987

Lange K, Boehnke M, Cox DR, and Lunetta KL. Statistical methods for polypoid radiation hybrid mapping. *Genome Res* 5: 136-150, 1995

Lathorp GM, and Lalouel JM. Easy calculations of Lod scores and genetic risks on small computers. *Am J Hum Genet* 36: 460-465, 1984

Ledley FD, and Rosenblatt DS. Mutations in mut methylmalonic acidemia: clinical and enzymatic correlations. *Hum Mutat* 9: 1-6, 1997

Legros F, Chatzoglou E, Frachon P, de Baulny HO, Laforet P, Jardel C, Godinot C, and Lombes A. Functional characterization of novel mutations in the human cytochrome b gene. *Eur J Hum Genet* 9: 510-518, 2001

Lehesjoki AE, Koskiniemi M, Norio R, Salvatore T, Sistonen P, Lander E, and de la Chapelle A. Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum Mol Genet* 2: 1229-1234, 1993

Leighton J, and Schatz G. An ABC transporter in the mitochondrial inner membrane is required for normal growth of yeast. *EMBO J* 14: 188-195, 1995

Levy E, Haltia M, Fernandez-Madrid I, Koivunen O, Ghiso J, Prelli F, and Frangione B. Mutation in gelsolin gene in Finnish hereditary amyloidosis. *J Exp Med* 172: 1865-1867, 1990

Lill R, Diekert K, Kaut A, Lange H, Pelzer W, Prohl C, and Kispal G. The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. *Biol Chem* 380: 1157-66, 1999

Lill R, and Kispal G. Maturation of cellular Fe-S proteins: an essential function of mitochondria. *Trends Biochem Sci* 25: 352-356, 2000

Lill R, and Kispal G. Mitochondrial ABC transporters. *Res Microbiol* 152: 331-340, 2001

Lunetta K, Boehnke M, Lange K, and Cox DR. Selected locus and multiple panel models for radiation hybrid mapping. *Am J Hum Genet* 59: 717-725, 1996

Majander A, Rapola J, Sariola H, Suomalainen A, Pohjavuori M, and Pihko H. Diagnosis of fatal infantile defects of the mitochondrial respiratory chain: age dependence and postmortem analysis of enzyme activities. *J Neurol Sci* 134: 95-102, 1995

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, and Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* 30: 281-283, 2002

Maury CP, Kere J, Tolvanen R, and de la Chapelle A. Finnish hereditary amyloidosis is caused by a single nucletide substitution in the gelsolin gene. *FEBS Lett* 276: 75-77, 1990

Mitchell GA, Brody LC, Sipilä I, Looney JE, Wong C, Engelhardt JF, Patel AS, Steel G, Obie C, Kaiser-Kupfer M, et al. At least two mutant alleles of ornithine delta-aminotransferase cause gyrate atrophy of the choroid and retina in Finns. *Proc Natl Acad Sci* 86: 197-201, 1989

Mitsuhashi N, Miki T, Senbongi H, Yokoi N, Yano H, Miyazaki M, Nakajima N, Iwanaga T, Yokoyama Y, Shibata T, and Seino S. MTABC3, a novel mitochondrial ATP-binding cassette protein involved in iron homeostasis. *J Biol Chem* 275: 17536-17540, 2000

Morris AAM, Taylor RW, Birch-Machin MA, Jackson MJ, Coulthard MG, Bindoff LA, Welch RJ, Howell N, and Turnbull DM. Neonatal Fanconi syndrome due to deficiency of complex III of the respiratory chain. *Pediatr Nephrol* 9: 407-411, 1995

Mount SM. A catalogue of splice junction sequences. *Nucleic Acids Res* 10: 459-472, 1982

Muhlenhoff U, and Lill R. Biogenesis of iron-sulfur proteins in eukaryotes: a novel task of mitochondria that is inherited from bacteria. *Biochim Biophys Acta* 1459: 370-382, 2000

The Multiple Leiomyoma Consortium. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat Genet* 30: 406-410, 2002

Murray KF, and Kowdley KV. Neonatal hemochromatosis. *Pediatrics* 108: 960-964, 2001

Mäkelä-Bengs P, Järvinen N, Vuopala K, Suomalainen A, Ignatius J, Sipilä M, Herva R, Palotie A, and Peltonen L. Assignment of the disease locus for the lethal congenital contracture syndrome to a restricted region of chromosome 9q34 by genome scan using five affected individuals. *Am J Hum Genet* 63: 506-516, 1998

Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, Krohn KJ, Lalioti MD, Mullis PE, Antonarakis SE, Kawasaki K, Asakawa S, Ito F, and Shimizu N. Positional cloning of the APECED gene. *Nat Genet* 17: 393-398, 1997

Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, and Peltonen L. Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. *Am J Hum Genet* 56: 1088-1095, 1995

Nobrega FG, Nobrega MP, and Tzagoloff A. BCS1, a novel gene required for the expression of functional Rieske iron-sulfur protein in Saccharomyces cerevisiae. *EMBO J* 11: 3821-3829, 1992

Norio R. Biocultural aspects of disease (ed. H R Rothschild), Chapter 12, Diseases of Finland and Scandinavia. *Academic press, New York*: 359-414, 1981

Norio R. Suomi-neidon geenit (in Finnish). *Otava, Keuruu*, 2000

Norio R, Nevanlinna HR, and Perheentupa J. Hereditary diseases in Finland: rare flora in rare soil. *Ann Clin Res* 5: 109-141, 1973

Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26: 588-97, 1974

Paavola P, Salonen R, Weissenbach J, and Peltonen L. The locus for Meckel syndrome with multiple congenital anomalies maps to chromosome 17q21-q24. *Nat Genet* 11: 213-215, 1995

Paloneva J, Kestilä M, Wu J, Salminen A, Bohling T, Ruotsalainen V, Hakola P, Bakker AB, Phillips JH, Pekkarinen P, Lanier LL, Timonen T, and Peltonen L. Loss-of-function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. *Nat Genet* 25: 357-361, 2000

Paloneva J, Manninen T, Christman G, Hovanes K, Mandelin J, Adolfsson R, Bianchin M, Bird T, Miranda R, Salmaggi A, Tranebjaerg L, Konttinen Y, and Peltonen L. Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *Am J Hum Genet* 71: 656-662, 2002

Papadopoulou LC, Sue CM, Davidson MM, Tanji K, Nishino I, Sadlock JE, Krishna S, Walker W, Selby J, Glerum DM, Coster RV, Lyon G, Scalais E, Lebel R, Kaplan P, Shanske S, De Vivo DC, Bonilla E, Hirano M, DiMauro S, and Schon E. Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene. *Nat Genet* 23: 333-337, 1999

Pastinen T, Perola M, Ignatius J, Sabatti C, Tainola P, Levander M, Syvänen AC, and Peltonen L. Dissecting a population genome for targeted screening of disease mutations. *Hum Mol Genet* 10: 2961-2972, 2001

Patel P, and Isaya G. Friedreich ataxia: from GAA triplet-repeat expansion to frataxin deficiency. *Am J Hum Genet* 69: 15-24, 2001

Pekkarinen P, Hovatta I, Hakola P, Järvi O, Kestilä M, Lenkkeri U, Adolfsson R, Holmgren G, Nylander P, Tranebjaerg L, Terwilliger J, Lönnqvist J, and Peltonen L. Assignment of the locus for PLO-SL, a frontal-lobe dementia with bone cysts, to 19q13. *Am J Hum Genet* 62: 362-372, 1998

Pellegata NS, Dieguez-Lucena JL, Joensuu T, Lau S, Montgomery KT, Krahe R, Kivelä T, Kucherlapati R, Forsius H, and de la Chapelle A. Mutations in KERA, encoding keratocan, cause cornea plana. *Nat Genet* 25: 91-95, 2000

Peltonen L, Jalanko A, and Varilo T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8: 1913-1923, 1999

Peltonen L, Pekkarinen P, and Aaltonen J. Messages from an isolate: lessons from the Finnish gene pool. *Biol Chem Hoppe-Seyler* 376: 697-704, 1995

Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, D'Amato E, Ramirez L, Faham M, Koskiniemi M, Warrington JA, Norio R, de la Chapelle A, Cox DR, and Myers RM. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). *Science* 271: 1731-1734, 1996

Petruzzella V, Tiranti V, Fernandez P, Ianna P, Carrozzo R, and Zeviani M. Identification and characterization of Human cDNAs specific to BCS1, PET112, SCO1, COX15, and COX11, five genes involved in the formation and function of the mitochondrial respiratory chain. *Genomics* 54: 494-504, 1998

Ranta S, Zhang Y, Ross B, Lonka L, Takkunen E, Messer A, Sharp J, Wheeler R, Kusumi K, Mole S, Liu W, Soares MB, Bonaldo MF, Hirvasniemi A, de la Chapelle A, Gilliam TC, and Lehesjoki AE. The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nat Genet* 23: 233-236, 1999

Rapola J, Heikkilä P, and Fellman V. Pathology of lethal fetal growth retardation syndrome with aminoaciduria, iron overload, and lactic acidosis (GRACILE). *Pediatr Pathol Mol Med* 21: 183-193, 2002

The Retinoschisis Consortium. Functional implications of the spectrum of mutations found in 234 cases with X-linked juvenile retinoschisis (XLRS). *Hum Mol Genet* 7: 1185-1192, 1998

Ridanpää M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, Mäkitie O, Kaitila I, and de la Chapelle A. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell* 104: 195-203, 2001

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui LC, and Collins FS. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 245: 1059-1065, 1989

Rustin P, Chretien D, Bourgeron T, Gerard B, Rötig A, Saudubray JM, and Munnich A. Biochemical and molecular investigations in respiratory chain deficiencies. *Clin Chim Acta* 228: 35-51, 1994

Salonen R, Somer M, Haltia M, Lorentz M, and Norio R. Progressive encephalopathy with edema, hypsarrhythmia, and optic atrophy (PEHO syndrome). *Clin Genet* 39: 287-293, 1991

Sanger F, Nicklen S, and Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Nat Acad Sci* 74: 5463-5467, 1977

Sankila EM, Tolvanen R, van den Hurk JA, Cremers FP, and de la Chapelle A. Aberrant splicing of the CHM gene is a significant cause of choroideremia. *Nat Genet* 1: 109-113, 1992

Savary S, Allikmets R, Denizot F, Luciani MF, Mattei MG, Dean M, and Chimini G. Isolation and chromosomal mapping of a novel ATP-binding cassette transporter conserved in mouse and human. *Genomics* 41: 275-278, 1997

Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, and Peltonen L. CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. *Nat Genet* 19: 286-288, 1998

Schuler GD. Sequence mapping by electronic PCR. *Genome Res* 7: 541-550, 1997

Scriver CR, Beaudet AL, Sly WS, and Valle D. The metabolic & molecular bases of inherited disease, Volume II, Chapter 100, Lactic acidemia: disorders of pyruvate carboxylase and pyruvate dehydrogenase. *The McGraw-Hill Companies, USA*: 2275-2291, 2001a

Scriver CR, Beaudet AL, Sly WS, and Valle D. The metabolic & molecular bases of inherited diseases, Volume II, Chapter 104, Oxidative phosphorylation diseases. *The McGraw-Hill Companies, USA*: 2367-2392, 2001b

Senbongi H, Ling F, and Shibata T. A mutation in a mitochondrial ABC transporter results in mitochondrial dysfunction through oxidative damage of mitochondrial DNA. *Mol Gen Genet* 262: 426-436, 1999

Seyda A, Newbold RF, Hudson TJ, Verner A, MacKay N, Winter S, Feigenbaum A, Malaney S, Gonzales-Halphen D, Cuthbert AP, and Robinson BH. A novel syndrome affecting multiple mitochondrial functions, located by microcell-mediated transfer to chromosome 2p14-2p13. *Am J Hum Genet* 68: 386-396, 2001

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, and Sunden SL. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet* 4: 1837-1844, 1995

Shimada Y, Okuno S, Kawai A, Shinomiya H, Saito A, Suzuki M, Omori Y, Nishino N, Kanemoto N, Fujiwara T, Horie M, and Takahashi E. Cloning and chromosomal mapping of a novel ABC transporter gene (hABC7), a candidate for X-linked sideroblastic anemia with spinocerebellar ataxia. *J Hum Genet* 43: 115-122, 1998

Shoubridge E. Nuclear genetic defects of oxidative phosphorylation. *Hum Mol Genet* 10: 2277-2284, 2001

Sikorski RS, and Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* 122: 19-27, 1989

Strachan T, and Read AP. Human Molecular Genetics, Chapter 13, Genome projects. *John Wiley & Sons, Inc., USA*: 295-314, 1999a

Strachan T, and Read AP. Human Molecular Genetics, Chapter 15, Identifying human disease genes. *John Wiley & Sons, Inc., USA*: 351-375, 1999b

Stuhrmann M, and Dork T. CFTR mutations and male infertility. *Andrologia* 32: 71-83, 2000

Suomalainen A, and Kaukonen J. Diseases caused by nuclear genes affecting mtDNA stability. *Am J Med Gen* 106: 53-61, 2001

Syvänen AC, Sajantila A, and Lukka M. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am J Hum Genet* 52: 46-59, 1993

Tahvanainen E, Norio R, Karila E, Ranta S, Weissenbach J, Sistonen P, and de la Chapelle A. Cohen syndrome gene assigned to the long arm of chromosome 8 by linkage analysis. *Nat Genet* 7: 201-204, 1994

Terwilliger J. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 56: 777-787, 1995

Terwilliger JD, and Ott J. Handbook of human genetic linkage. *John Hopkins University Press, Baltimore*, 1994

Thorburn DR. Practical problems in detecting abnormal mitochondrial function and genomes. *Hum Reprod* 15 (Suppl.2): 57-67, 2000

Torrents D, Mykkänen J, Pineda M, Feliubadalo L, Estevez R, de Cid R, Sanjurjo P, Zorzano A, Nunes V, Huoponen K, Reinikainen A, Simell O, Savontaus ML, Aula P, and Palacin M. Identification of SLC7A7, y+LAT-1, as the lysinuric protein intolerance gene. *Nat Genet* 21: 293-296, 1999

Tuchman M, Morizono H, Reish O, Yuan X, and Allewell NM. The molecular basis of ornithine transcarbamylase deficiency: modelling the human enzyme and the effects of mutations. *J Med Genet* 32: 680-688, 1995

Walker BL, Tiong JWC, and Jefferies WA. Iron metabolism in mammalian cells. *Int Rev Cytol* 211: 241-278, 2001

Valnot I, Kassis J, Chretien D, de Lonlay P, Parfait B, Munnich A, Kachaner J, Rustin P, and Rötig A. A mitochondrial cytochrome b mutation but no mutations of nuclearly encoded subunits in ubiquinol cytochrome c reductase (complex III) deficiency. *Hum Genet* 104:460-466, 1999

Varilo T. The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. *PhD thesis, National Public Health Institute, Helsinki, Finland*, 1999

Varilo T, Nikali K, Suomalainen A, Lönnqvist T, and Peltonen L. Tracing an ancestral mutation: genealogical and haplotype analysis of the infantile onset spinocerebellar ataxia locus. *Genome Res* 6: 870-875, 1996

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, and Lathorp M. A second-generation linkage map of the human genome. *Nature* 359: 777-778, 1992

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GLG, Nelson C, et al. The sequence of the human genome. *Science* 291: 1304-1350, 2001

Verheijen FW, Verbeek E, Aula N, Beerens CE, Havelaar AC, Joosse M, Peltonen L, Aula P, Galjaard H, van der Spek PJ, and Mancini GM. A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nat Genet* 23: 462-465, 1999

Verloes A, Temple IK, Hubert AF, Hope P, Gould S, Debauche C, Verellen G, Deville JL, Koulischer L, and Sokal EM. Recurrence of neonatal haemochromatosis in half sibs born of unaffected mothers. *J Med Genet* 33: 444-449, 1996

Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, Hofmann SL, and Peltonen L. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376: 584-587, 1995

Wibrand F, Ravn K, Schwartz M, Rosenberg T, Horn N, and Vissing J. Multisystem disorder associated with a missense mutation in the mitochondrial cytochrome b gene. *Ann Neurol* 50: 540-543, 2001

Virtaneva K, D'Amato E, Miao J, Koskiniemi M, Norio R, Avanzini G, Franceschetti S, Michelucci R, Tassinari CA, Omer S, Pennacchio LA, Myers RM, Dieguez-Lucena JL, Krahe R, de la Chapelle A, and Lehesjoki AE. Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nat Genet* 15: 393-396, 1997

Visapää I, Salonen R, Varilo T, Paavola P, and Peltonen L. Assignment of the locus for hydrolethalus syndrome to a highly restricted region on 11q23-25. *Am J Hum Genet* 65: 1086-1095, 1999

Worton RG, and Thompson MW. Genetics of Duchenne muscular dystrophy. *Annu Rev Genet* 22: 601-629, 1988

Vuopala K, Ignatius J, and Herva R. Lethal arthrogryposis with anterior horn cell disease. *Hum Pathol* 26: 12-19, 1995

Zeviani M. The expanding spectrum of nuclear gene mutations in mitochondrial disorders. *Semin Cell Dev Biol* 12: 407-416, 2001

Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, Cuthbert AP, Newbold RF, Wang J, Chevrette M, Brown GK, Brown RM, and Shoubridge EA. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. *Nat Genet* 20: 316-317, 1998