POSITIONAL CLONING OF

THE MULIBREY NANISM GENE (*MUL*)

Kristiina Avela

Department of Medical Genetics
Haartman Institute
University of Helsinki
and
Department of Molecular Genetics
Folkhälsan Institute of Genetics

Academic Dissertation

To be publicly discussed with the permission of the Faculty of Medicine, University of Helsinki, in the large lecture hall of the Haartman Institute, Haartmaninkatu 3, Helsinki, on September 15th, 2000, at 12 o'clock

Helsinki 2000

Supervised by

Anna-Elina Lehesjoki, M.D., Ph.D.
Docent
Department of Medical Genetics
Haartman Institute, University of Helsinki
and
Department of Molecular Genetics
Folkhälsan Institute of Genetics
Helsinki, Finland

Albert de la Chapelle, M.D., Ph.D.
Professor
Comprehensive Cancer Center, Division of Human Cancer Genetics
The Ohio State University, Columbus, Ohio, USA
and
Department of Molecular Genetics
Folkhälsan Institute of Genetics
Helsinki, Finland

Reviewed by

Anu Wartiovaara, M.D., Ph.D.
Docent
Montreal Neurological Institute, McGill University
Montreal, Quebec, Canada

Raimo Voutilainen, M.D., Ph.D.
Professor
Department of Pediatrics
University of Kuopio
Finland

Official opponent

Andrea Ballabio, M.D., Ph.D.
Professor, Naples Second University
Director, TIGEM Telethon Institute of Genetics and Medicine
Naples
Italy

ISBN 952-91-1902-X
ISBN 952-91-2438-4 (pdf version, http://ethesis.helsinki.fi)

To Mika

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I Avela K, Lipsanen-Nyman M, Perheentupa J, Wallgren-Pettersson C, Marchand S, Fauré S, Sistonen P, de la Chapelle A, Lehesjoki A-E.
Assignment of the Mulibrey Nanism Gene to 17q by Linkage and Linkage-Disequilibrium Analysis.
Am J Hum Genet 60: 896-902, 1997.

II Paavola P*, Avela K*, Horelli-Kuitunen N, Bärlund M, Kallioniemi A, de la Chapelle A, Palotie A, Lehesjoki A-E, Peltonen L.
High-Resolution Physical and Genetic Mapping of the Critical Region for Meckel Syndrome and Mulibrey Nanism on Chromosome 17q22-q23.
Genome Res 9: 267-276, 1999.

III Avela K, Lipsanen-Nyman M, Idänheimo N, Seemanová E, Rosengren S, Mäkelä TP, Perheentupa J, de la Chapelle A, Lehesjoki A-E.
Gene encoding a new RING-B-box-Coiled-coil protein is mutated in Mulibrey Nanism.
Nature Genet 25: 298-301, 2000.

* These authors contributed equally to this work.
Publication II also appears in the thesis of Paavola P (1998).
In addition unpublished work is presented.

ABBREVIATIONS

aa	amino acid
BAC	bacterial artificial chromosome
bp	base pair
CCM	chemical cleavage of mismatch
cDNA	complementary DNA
CEPH	Centre d'Etude du Polymorphisme Humain
CGH	comparative genomic hybridization
CHLC	the Cooperative Human Linkage Center
cM	centiMorgan
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
EST	expressed sequence tag
FISH	fluoresence in situ hybridization
HGP	human genome project
kb	kilobase
LD	linkage disequilibrium
LOD	logarithm of odds
LOH	loss of heterozygosity
Mb	megabase pairs
MKS	Meckel syndrome
mRNA	messenger RNA
MUL	Mulibrey Nanism
ORF	open reading frame
PAC	P1-derived artificial chromosome
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
PIC	polymorphism information content
RBCC	RING-finger-B-box-coiled coil
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
SNP	single nucleotide polymorphism
SSCP	single strand conformation polymorphism
STR	short tandem repeat
STS	sequence tagged site
TS	tumour suppressor
VNTR	variable number of tandem repeats
WICGR	Whitehead Institute Center for Genome Research
WT	Wilms´ tumour
YAC	yeast artificial chromosome

1. SUMMARY

Mulibrey Nanism (MUL) is an autosomal recessively inherited condition with growth retardation affecting several tissues of mesodermal origin.

MUL belongs to the diseases of the Finnish disease heritage, that is a number of disorders that occur more frequently in Finland than elsewhere. The molecular pathogenesis of MUL is not known.

In this study a positional cloning strategy was used to identify the causative gene for MUL. Initially the *MUL* locus was assigned to chromosome 17q by linkage analysis. The assignment of *MUL* was further refined to an 800 kb region by linkage disequilibrium and haplotype analysis.

A positional candidate gene strategy was employed since several genes, cDNA clones, and ESTs were localised to the *MUL* region.

The open reading frame of a cDNA clone, KIAA0898, was analysed by sequencing overlapping RT-PCR amplified fragments in four MUL patients representing five different haplotypes. Four independent MUL associated mutations were identified. All the mutations disrupt the reading frame predicting a truncated protein, thus providing strong evidence for KIAA0898 being the *MUL* gene.

MUL was identified to be a novel member of the RING-B-Box-Coiled-Coil (RBCC) family as the amino acid sequence has a RING finger, one B-box motive, and a coiled-coil domain. Furthermore, *MUL* was shown to be ubiquitously expressed corresponding well with the pleiotropic and developmental nature of the disorder.

The exact function of the MUL protein is presently unknown, but existing members of the RBCC family play a role in regulation of development and oncogenesis, both of which are disrupted in the MUL phenotype. The identification of the *MUL* gene is the first step towards the resolution of the molecular pathogenesis of Mulibrey Nanism.

2. INTRODUCTION

The Human Genome Project

The Human Genome Project (HGP), initiated in the mid 1980´s, is an international undertaking to co-ordinate the efforts needed for the sequencing of the human nuclear genome.

To achieve the whole sequence of the genome one needs high resolution genetic and physical maps. Much of this work has already been accomplished. The integrated genetic maps (e.g. Weissenbach et al. 1992, Gyapay et al. 1994, Murray et al. 1994, Dib et al. 1996) have on the average a marker density of one per 0.7 cM (Murray et al. 1994). Furthermore, several chromosomes and chromosomal regions are covered by physical maps (Chumakov et al. 1992, Foote et al. 1992, Cohen et al. 1993, Chumakov et al. 1995).

The major contributors in the ultimate effort of sequencing are the Department of Energy (DOE), Joint Genome Institute, USA; Baylor College of Medicine, Human Genome Sequencing Center, USA; The Sanger Centre, UK; Washington University, USA; and the Whitehead Institute for Biomedical Research/MIT Center for Genome Research, USA.

By February 2000 already 64% of the human genomic sequence was available via internet. About 17% of the genome is available as finished, high quality sequence, whereas the remaining 47% is available as “draft sequence” meaning that the sequence is of lower accuracy than the finished sequence and includes some gaps.

A draft sequence of the entire genomic sequence is estimated to be available by summer 2000, and a complete, high quality human genomic sequence is estimated to be available by 2003 (http://www.ornl.gov/hgmis/project/progress.html).

A major milestone within the HGP project was recently achieved as the first sequence of an entire human chromosome was completed decoding the sequence of chromosome 22 (Dunham et al. 1999).

Expressed sequence tags

During the last few years the HGP has put emphasis on the identification of expressed sequence tags (ESTs). ESTs represent coding regions of the genome and they are generated by sequencing short stretches of cDNA clones.

The major contributors in large-scale sequencing of cDNA clones are the Institute of Genomic Research (TIGR) and Merck Pharmaceuticals.

It has been estimated that only about 20% of human genes are expressed in such a restricted pattern or level that they are not likely to be found by analysing cDNA (Strachan and Read 1996).

By the end of 1999 altogether 1.1 million EST sequences have been produced from approximately 53,000 human genes. ESTs have provided a powerful tool for the identification of the disease genes, as 91% (December, 1997; http://www.ncbi.nlm.nih.gov/) of positionally cloned genes mutated in human diseases are represented by their exact matches with one or more ESTs in the EST division of GenBank (dbEST).

Consequences of HGP

Both great expectations and criticism have been raised due to the actions within the HGP. As the structure of every human gene will be revealed, individuals at risk of a hereditary disorder can be tested prenatally and

presymptomatically. The function and regulation of individual genes can be determined improving our knowledge on biological processes. Inevitably, gene therapy approaches will be increasingly applied. On the other hand, commercial exploitation of sequence data, e.g. patenting human genes has been objected by the public at large. The detection of disease associated mutations raises concern among citizens; will individuals be discriminated based on their genes by insurance companies or employers? These concerns have to be addressed e.g. by public debate and proper legislation.

Availability of data

The HGP has developed capabilities for collecting, storing, distributing, and analysing all the data produced. A free access to the data and facilities developed within the HGP is provided via internet for the whole scientific community. This has greatly promoted the work of academic research groups in the field of genetics throughout the world. Vice versa, academic research groups have also contributed by submitting their own data to public databases. Both have also happened in the positional cloning project of Mulibrey Nanism gene (*MUL*) described in this thesis. Free availability of genetic and sequence data should be acknowledged by everyone working in the field of genetics.

3. REVIEW OF THE LITERATURE

3.1. Finnish disease heritage

Basic concept

The concept of Finnish disease heritage refers to rare, monogenic disorders occurring more frequently in Finland than elsewhere (Norio et al. 1973).

Presently, the Finnish disease heritage includes some 35 disorders (reviewed by Peltonen et al. 1999). Most of these disorders are autosomal recessive (Norio 1981). However, the Finnish amyloidosis type V (Meretoja 1973) and tibial muscular dystrophy (Udd et al. 1992) represent an autosomal dominant pattern of inheritance, and two eye disorders, choroideremia (Shapira & Sitney 1943) and retinoschisis (Gieser & Falls 1961), are X-linked.

As a consequence of the same phenomena that has given raise to the Finnish disease heritage many hereditary disorders prevalent in other European populations, e.g. cystic fibrosis (Lowe et al. 1949) and Tay-Sachs disease (Petersen et al. 1983), are exceptionally rare or absent in the Finnish population (Norio et al. 1973, Norio 1981, Kere et al. 1994).

Finnish population history

The unique genome of the Finns has evolved as a consequence of the Finnish population history (Norio et al. 1973, de la Chapelle 1993, Sajantila et al. 1996, de la Chapelle & Wright 1998, Peltonen 1999, Peltonen et al. 1999, Varilo 1999), which is characterised by isolation, several population bottlenecks, and a restricted number of founders.

Finland was populated soon after the glacial period, some 10,000 years ago by settlers of unknown origin. For about 8,000 years Finland remained sparsely populated. During the first centuries A.D. a new wave of settlers originating from western European populations (Sajantila et al. 1995) immigrated to the southern and western coastal parts of Finland (early-settlement areas; Fig. 1).

The Finnish population remained isolated, partly due to geographical reasons. Finland is surrounded in the south and west by the Baltic Sea. In addition, Finland is located between two countries, Sweden and Russia, which have had a distinct culture and religion. Particularly, the Finns were isolated due to their language, as Finnish is not related to Swedish or Russian.

During the following centuries, wars, diseases and periods of famine inhibited the growth of the population producing population bottlenecks. The backlands of Finland (late-settlement areas; Fig. 1) were inhabited by people originating from South Savo during the 16th century as more land was needed to feed the increasing population. Moreover, the Swedish Crown favoured internal migration by lightening taxation to occupy the empty backlands close to Russia.

The small number of settlers introduced an additional internal population bottleneck. The new settlements were distantly located from each other and separated by forests and lakes forming sub-isolates within Finland. The random set of alleles introduced by the founders established the gene pool of the sub-isolate; a phenomenon called the founder effect. The random alleles introduced by the founders were enriched due to consanguineous marriages.

Meaning of birthplaces

Before the World War II, the majority of Finns got their living from agriculture and inhabited the same geographical region as their ancestors. Consequently, the birthplaces of the grandparents of presently living patients represent the founder region. The Finnish settlement history and the birthplaces of the grandparents of presently living patients can be utilised for a rough estimation of the age of the mutation (de la Chapelle & Wright 1998). If the birthplaces are exclusively scattered in the late-settlement region it can be hypothesised that the expansion of the mutation started after the 16th century, whereas if the birthplaces are detected also in the early-settlement region, the expansion of the mutation most likely has initiated before the 16th century (de la Chapelle & Wright 1998; Fig. 1).

An excellent possibility to study genetics

The Finnish disease heritage provides an excellent possibility to study human genetics. The high quality of the Finnish health care system enables the clinical investigation and diagnosis of rare hereditary disorders. Population records established in the 17th century provide reliable information on births, deaths and marriages for genealogical studies. Furthermore, the Finns have a positive attitude towards medical research. Not surprisingly, the molecular genetics of the diseases of the Finnish disease heritage have been studied extensively. Until now the gene locus is assigned for 33 disorders, and the underlying gene has been identified in 22 disorders (Table 1).

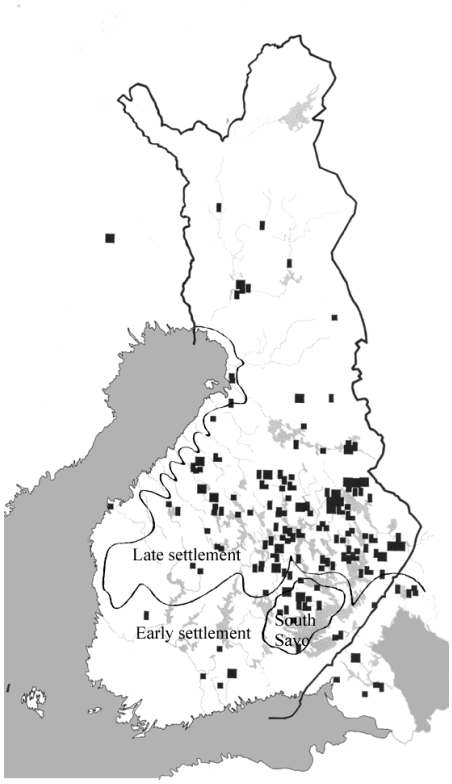

Fig. 1

Map of Finland presenting the early-settlement and late-settlement regions as well as the region of South Savo. The black squares represent the birthplaces of the grandparents of the parents of the MUL patients. The size of the square refers to the number of the grandparents. MUL is clustered in the Savo and North Carelia regions. (Adapted from Lipsanen-Nyman 1986 and de la Chapelle 1993.)

Table 1. The diseases of the Finnish disease heritage.

DISEASE, ABBREVIATION	LOCUS	DEFECTIVE GENE/PROTEIN	REFERENCE
The identified genes			
Aspartylglucosaminuria, AGU	4q	*AGA,* aspartylglucosaminidase	Ikonen et al. 1991
Autoimmune Polyendocrinopathy-	21q	*AIRE,* a novel gene featuring two zinc-finger domains	The Finnish-German APECED Consortium
Candidiasis-Ectodermal Dystrophy, APECED			1997, Nagamine et al. 1997
Choroideremia, CHM	Xq	RAB geranyl-geranyl transferase	Cremers et al. 1990
Congenital Chloride Diarrhoea, CLD	7q	*DRA,* down-regulated in adenoma	Höglund et al. 1996
Congenital Nephrosis, Finnish type; CNF	19q	*NPHS1,* a novel gene encoding a glomerular protein	Kestilä et al. 1998
Cornea plana congenita, CNA2	12q	*KTN,* keratocan, a leucine-rich proteoglycan	Pellegata et al. 2000
Diastrophic Dysplasia, DTD	5q	*DTDST,* diastrophic dysplasia sulphate transporter	Hästbacka et al. 1994
Familial Amyloidosis, Finnish type; FAF	9q	*GSN,* gelsolin	Levy et al. 1990, Maury et al. 1990
Gonadal dysgenesis, FSH-RO	2p	*FSHR,* follicle stimulating hormone receptor	Aittomäki et al. 1995
Gyrate Atrophy of Choroid and Retina, HOGA	10q	*OAT,* ornithine delta-aminotransferase	Mitchell et al. 1988
Hereditary Fructose Intolerance, HFI	9q	*ALDOB,* aldolase B	Cross et al. 1988
Infantile Neuronal Ceroid Lipofuscinosis, INC L	1p	*PPT,* palmitoyl protein transferase	Vesa et al. 1995
Lysinuric Protein Intolerance, LPI	14q	*SLC7A7,* a novel gene encoding a putative permease-related protein	Borsani et al. 1999, Torrents et al. 1999
Mulibrey Nanism, MUL	17q	*MUL,* a novel gene encoding a RBCC protein	III
Nonketotic hyperglycinemia, NKH	9p	P protein of glycine cleavage system	Kure et al. 1992
Polycystic lipomembranous osteodysplasia	19q	*TYROBP* , TYRO protein tyrosine kinase binding protein	Paloneva et al. 2000
with sclerosing leukoencephalopathy, PLOSL			
Progressive epilepsy with mental retardation, EPMR	8p	*CLN8,* a novel transmembrane protein	Ranta et al. 1999
Progressive myoclonus epilepsy, PME	21q	*CSTB,* cystatin B, a cysteine protease inhibitor	Pennacchio et al. 1996
Retinoschisis, RS	Xp	*XLRS1,* a novel gene encoding a putative cell adhesion protein	Sauer et al. 1997
Salla disease, SD	6q	*SLC17A5,* also known as *AST,* encoding a sialin	Verheijen et al. 1999
Selective intestinal malabsorption of vitamin B12, MGA1	10p	*CUBN,* cubilin, a gene encoding intrinsic factor-vitamin B12 receptor	Aminoff et al. 1999
Variant late-infantile neuronal ceroid lipofuscinosis, vINCL	13q	*CLN5,* a novel gene encoding a putative transmembrane protein	Savukoski et al. 1998
The mapped genes			
Cartilage-hair hypoplasia, CHH	9p		Sulisalo et al. 1993
Cohen syndrome, COH1	8q		Tahvanainen et al. 1994
Congenital lactase deficiency, CLD	2q		Järvelä et al. 1998
Growth retardation with acidosis	2q		Visapää et al. 1998
Hydrolethalus syndrome	11q		Visapää et al. 1999
Infantile-onset spinocerebellar ataxia, IOSCA	10q		Nikali et al. 1995
Lethal congenital contracture syndrome, LCCS	9q		Mäkelä-Bengs et al. 1998
Meckel syndrome, MKS	17q		Paavola et al. 1995
Muscle-eye-brain disese, MEB	1p		Cormand et al. 1999
Tibial muscular dystrophy, TMD	2q		Haravuori et al. 1998
Usher syndrome, type III; USH3	3q		Sankila et al. 1995
The unmapped loci			
Progressive encephalopathy with edema,			Salonen et al. 1991
Progressive encephalopathy with edema,			
RAPADILINO			Kääriäinen et al. 1989

3.2. Mulibrey Nanism

Mulibrey Nanism (MUL; MUscle-LIver-BRain-EYe Nanism; MIM No. 253250) was first described in Finland in the early 1970´s (Perheentupa et al. 1970 and 1973). It is a typical disorder of the Finnish disease heritage; the majority of the reported MUL patients are Finnish, and the pattern of inheritance is autosomal recessive.

MUL affects several tissues of mesodermal origin (Perheentupa et al. 1973) and is evidently caused by a pleiotropic gene (Perheentupa et al. 1973, Lipsanen-Nyman 1986, Lapunzina et al. 1995). The molecular pathogenesis of MUL is not known.

World-wide, some 75 MUL patients have been described, the majority from Finland, where the incidence of MUL is estimated to be 1:40,000 (Lipsanen-Nyman 1986). In Finland the disease is clustered in the Savo and North Carelia regions (Fig. 1) representing early and late settled areas (de la Chapelle 1993). Sporadic patients have been reported e.g. from Egypt (Thoren 1973), Canada (Cumming et al. 1976), USA (Voorhess et al. 1976), Spain, and Argentina (Lapunzina et al. 1995).

Clinical characteristics of MUL

MUL is characterised by growth failure of prenatal onset. The final height of MUL patients is on the average 33 cm below population means. MUL patients have typical dysmorphic features including triangular face, broad forehead, low nasal ridge, and telecanthus (Perheentupa et al. 1973, Myllärniemi et al. 1978; Fig. 2).

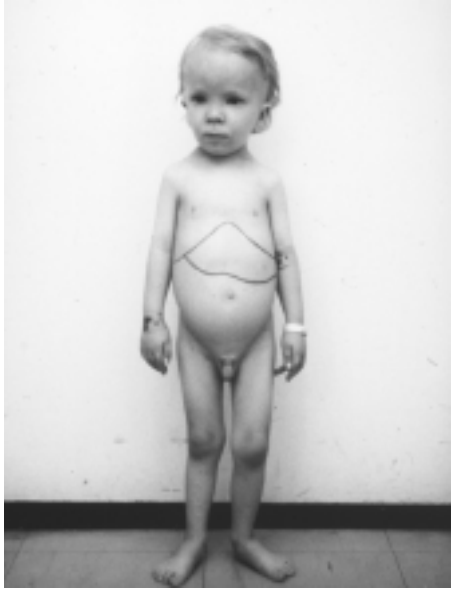

Fig. 2

A 4 year old Mulibrey Nanism patient. Notice the overall slenderness of the body. The borders of the enlarged liver are indicated with a marker. Also, the characteristic dysmorphic features, triangular face, prominent forehead, and low nasal bridge are detected.

MUL is also characterised by yellowish dots in the ocular fundi, J-shaped sella turcica, enlarged brain ventricles, prominent jugular veins, small voice, and muscle hypotonia (Perheentupa et al. 1973, Lipsanen-Nyman 1986, Lapunzina et al. 1995). Hypoplasia of various endocrine glands causing hormonal deficiencies is common e.g. growth hormone deficiency, hypocortisolism, hypothyroidism, and hypogonadism with infertility have been reported (Lenko et al. 1982, Lipsanen-Nyman 1986, Haraldsson et al. 1993). No signs of intellectual incapacity have been observed in MUL (Lipsanen-Nyman 1986). The clinical characteristics of MUL are summarised in Table 2.

Table 2. Summary of the findings in Mulibrey Nanism patients (adapted from Lapunzina et al. 1995).

In more than 66% of the patients	In at least 25% of the patients	In less than 25% of the patients
short stature	cutaneous naevi flammei	cardiomegaly
growth failure	dental malocclusion	cortical thickening of long bones
triangular face	missing/small sinuses	hypodontia of 2nd premolar
J-shaped sella turcica	prenatal growth failure	strabismus
depressed nasal ridge	pulmonary infections	large cerebral ventricles
broad nose	choroidal hypoplasia	ovarian tumour
yellowish dots in ocular fundi	ascites/edema	corneal dystrophia
prominent jugular veins	fibrous dysplasia of tibia	high set hyoid bone
hepatomegaly		simian crease
small voice		eosinophilia
frontal bossing		edema of vocal cords
pericardial constriction		coloboma of iris
muscle hypotonia		hypoplasia of dental enamel
		hydrops fetalis
		Wilms´ tumour
		hypoglycemia
		hyperammonemia

MUL and tumours

Interestingly, the risk for developing unclassified ovarian tumours and Wilms´ tumours (WT) is increased in MUL (Lipsanen-Nyman 1986). Two MUL patients with WT (Similä et al. 1980, Seemanová & Bartch 1999) and two siblings of MUL patients presenting some signs of MUL and suffering from WT (Lipsanen-Nyman 1986) have been reported. WT is one of the most common solid tumours of childhood. It occurs usually sporadically in 1 in 10,000 children and accounts for 8% of childhood cancers (Breslow & Beckwith 1982).

WT is associated with several genetic disorders e.g. the Beckwith-Wiedeman syndrome (Koufos et al. 1989), neurofibromatosis type-1 (Stay & Vawter 1977), the hereditary hyperparathyroid-jaw tumour syndrome (Szabo et al. 1995), the Simpson-Golabi-Behmel syndrome (Pilia et al. 1996), Li-Fraumeni syndrome (Hartley et al. 1993), the breast-ovarian cancer syndrome (Narod 1994), Bloom syndrome (Cairney et al. 1987), and the Perlman syndrome (Neri et al. 1984).

MUL and Silver-Russel syndrome

The phenotype of MUL overlaps to some extent with the phenotype of Silver-Russell syndrome (SRS; MIM No. 180860) often creating difficulties in the differential diagnosis of these disorders. The main characteristics of SRS are intrauterine growth retardation, asymmetry of the limbs, craniofacial disproportion, and triangular faces.

SRS is a genetically heterogeneous disorder. Both sporadic (Tanner et al. 1975) and familial (Fuleihan et al. 1971, Duncan et al. 1990) SRS cases have been reported. Approximately 10% of the SRS cases are associated with maternal uniparental disomy of chromosome 7 (Joyce et al. 1999). Recently, a maternally derived *de novo* duplication of 7p11.2-13 was identified in a SRS patient suggesting that overexpression of an imprinted gene in this region is responsible for SRS (Monk et al. 2000).

Additionally, two SRS patients with an identical translocation at 17q25 have been reported (Ramirez-Duenas et al. 1992, Midro et al. 1993) suggesting that a mutation in a gene located in 17q25 is also responsible for SRS. Association of SRS with abnormalities of chromosomes 8, 15 and 18 have also been reported (Chauvel et al. 1975, Scinzel et al. 1994, Rogan et al. 1996).

Furthermore, X-linked dominant inheritance has been suggested as no male-to-male transmission has been documented in SRS patients (Duncan et al. 1990).

3.3. Strategies for identifying disease genes

3.3.1. Functional cloning

In a functional cloning approach one utilises knowledge about the biochemical basis of a hereditary disorder in the identification of a disease gene. Either a purified protein or a functional gene assay must be available. If the protein can be isolated and the amino acid sequence determined, one can design *partially degenerate oligonucleotides* for cDNA library screening. Factor VIII gene, in which mutations cause hemophilia A, was identified using this method (Gitschier et al. 1984).

Alternatively, the amino acid sequence can be used to raise *specific*

antibodies against a protein or a fragment of it. The antibodies can be used to immunoprecipitate polyribosomes containing mRNA of an unidentified gene. The mRNA is isolated and converted to cDNA, and a specific cDNA clone is isolated from a cDNA library. This approach was applied in the identification of the phenylketonuria gene (Robson et al. 1982).

If the phenotype of a disease can be evaluated or measured from a cell culture and the defective gene product is unknown one can use *functional complementation* assays.

In the process of microcell mediated chromosome transfer (Cuthbert et al. 1995) all 22 autosomes and the X chromosome, one at a time, are incorporated into the genome of patient fibroblasts. The rescued phenotype of the hybrid fibroblast line determines the chromosomal localisation of the defective gene, which can be further defined by introducing deleted versions of the chromosome.

This method was applied in the cloning of the Leigh syndrome (LS) gene (Zhu et al. 1998). LS is a severe neurodegenerative disorder associated with systemic cytochrome c oxidase (COX) deficiency, which can be measured from cell culture extracts (Zhu et al. 1998).

3.3.2. Positional cloning

Positional cloning implies identifying a gene on the basis of its chromosomal location (Collins 1992, 1995). The number of inherited disease genes identified by positional cloning has been rapidly increasing during the last few years. By the year 1998 altogether 107 inherited disease genes were identified by *pure positional cloning* (http://genome.nhgri.nih.gov/clone/). As the Human Genome Project (HGP) proceeds, the pure positional cloning approach will be less frequently needed and the *positional candidate gene* approach will be applied (Collins 1995).

In the initial assignment of a disease locus comparative genomic hybridisation (CGH), loss of heterozygosity (LOH) analysis, chromosomal aberrations, and linkage analysis can be utilised. CGH and LOH analysis can only be applied in the identification of cancer genes, whereas chromosomal aberrations and linkage analysis can also be applied in the identification of a disease gene underlying any hereditary disorder.

3.3.2.1. Comparative genomic hybridisation

Comparative genomic hybridisation (CGH) allows the mapping of novel oncogenes and tumour suppressor (TS) genes (e.g. Hemminki et al. 1997; reviewed by Knuutila et al. 1998, 1999). CGH is a molecular cytogenetic method first reported by Kallioniemi et al. (1992) permitting screening of tumour genome for genetic alterations, gains or losses of DNA sequences.

Labelled tumour DNA and normal reference DNA are hybridised *in a competing manner* onto normal metaphase preparations. DNA copy number changes must be present in 50% of tumour sample cells in order for CGH to detect them (Kallioniemi et al. 1994). The sensitivity of the method is dependent on the size and magnitude of the copy number change. On the average, CGH detects DNA copy number changes of 5-10 Mb (Kallioniemi et al. 1994). From there on conventional tools of physical mapping

and mutation analysis are needed for the final identification of a cancer gene.

3.3.2.2. Loss of heterozygosity analysis

TS genes inhibit cell proliferation and their reduced transcription or deletion is known to be involved in malignant transformation (Knudson 1971, Friend et al. 1988). Locations for TS genes can be identified by loss of heterozygosity (LOH) analysis. Commonly the first mutation in a TS gene is a point mutation, whereas the second mutation involves a loss of a part of a chromosome. As a result of the second mutation one allele is lost at any DNA marker analysed close to a TS gene. Thus, an individual heterozygous at a locus is mutated into a hemizygote in tumour tissue. The screening of paired blood and tumour DNA samples with markers covering the whole genome provides a method for identifying TS gene loci (Vogelstein et al. 1989, Smith et al. 1992)

3.3.2.3. Chromosomal aberrations

In a process of identifying a gene for a hereditary disorder one can take advantage of chromosomal aberrations. If e.g. a balanced translocation is detected in a karyotype of a patient, it can be hypothesised that the gene disrupted during the translocation is the causative gene for the disorder. Thus, the characterisation of the chromosomal breakpoint regions pinpoints the critical gene region. Subsequently, conventional tools of physical mapping, gene identification, and mutation analysis are used.

This approach has been successfully exercised in identifying several X-linked genes such as the Duchenne muscular dystrophy gene (Monaco & Kunkel 1988) and the anhidrotic ectodermal dysplasia gene (Srivastava et al. 1996, Kere et al. 1996). This approach depends on rare events, which are not found for every disease gene identification project.

3.3.2.4. Linkage analysis

3.3.2.4.1. Basic concept

Linkage analysis offers a general approach for gene mapping. Any monogenic disorder locus can be assigned providing a sufficiently large panel of families and markers is available.

Linkage analysis locates a gene by its proximity to another locus on the same chromosome. Linkage analysis is based on recombinations that arise as a result of crossing-over events taking place during pachytene stage of meiosis I of germline cells. The arms of two homologous chromosomes tend to cross over allowing homologous chromosomes to exchange chromosomal material. Recombinations increase enormously the potential genetic diversity of gametes. Two loci far apart on a chromosome will most likely be separated in crossing-over, whereas two loci located close together on a chromosome will most likely stay together after a crossing-over event.

The frequency of recombination between two loci can be used as a rough measure of their distance. The distance between two loci is defined in units of centiMorgans (cM). If two loci are 1 cM apart, there is a 1% chance of recombination between these loci as the chromosome is passed from one generation to another.

The relation of genetic distance to physical chromosomal distance is not

linear. Firstly, recombination in autosomes is more frequent in females than males. Secondly, recombination frequency varies in different parts of the chromosome and seems to be greater at telomeres than near the centromere. On the average 1 cM corresponds to 1 Mb. (Morton 1955, Ott 1991).

3.3.2.4.2. Lod score

Basic concept
The proportion of recombination, out of all opportunities for recombination, is called the recombination fraction, denoted by the Greek letter theta (θ). For unlinked gene loci, the recombination fraction is equal to 0.5, while for linked loci it is less than 0.5.

The lod score method is a likelihood-based analysis testing whether an observed recombination fraction between two loci is significantly smaller than 0.5. The lod score (Z) is calculated as a $\log_{10}$ of the ratio: likelihood of linkage at θ/likelihood of no linkage. In practice, this ratio is computed for several values of θ. The most likely distance between the loci is the θ value, at which the highest positive lod score is obtained.

Positive lod scores suggest the presence of linkage, whereas negative lod scores do the opposite. A positive lod score of 3 or greater is considered significant evidence for linkage, since it indicates 10^{-3} or 1:1000 odds that the linkage observed did not happen by chance. A lod score less than –2 is commonly considered as a significant evidence against linkage, since it indicates 1:100 odds against linkage. (Morton 1955, Gelehrter and Collins 1990, Ott 1991).

Parametric and non-parametric method
The standard lod score analysis is called *parametric linkage analysis* since it requires a precise mode of inheritance, gene frequencies, and penetrance for each genotype. Parametric linkage analysis is suitable for localising disease genes for Mendelian disorders. In *two-point linkage analysis* one uses phenotype data and allele data at one marker locus for calculating lod scores, whereas in *multipoint linkage analysis* allele data from several loci are simultaneously considered improving the efficiency of analysis.

Non-parametric linkage analysis is a model-free method, which does not require knowledge about the inheritance pattern of the disease. Non-parametric methods ignore unaffected people, and look for alleles that are shared by those affected. *Sib pair analysis* uses the non-parametric method within nuclear families, whereas in *association studies* the non-parametric method is used at the population level. Non-parametric methods are used for searching genes predisposing to multifactorial diseases like diabetes (Mein et al. 1998).

3.3.2.4.3. Computer programmes

Presently, there are several computer programmes available for calculating the likelihood of linkage. LINKAGE (Lathrop et al. 1984, Ott 1995) is one of the most widely used programme packages consisting of a series of programmes for maximum likelihood estimation of recombination rates and analysis of genetic risks.

GENEHUNTER (Kruglyak et al. 1996) can be used for e.g. parametric linkage analysis, non-parametric multipoint linkage analysis, haplotype reconstruction, and sib pair analysis.

A multilocus linkage analysis programme VITESSE (O'Connell & Weeks 1995) uses a novel set-recoding scheme to recode each person's genotype and “fuzzy inheritance” to infer inheritance relations accelerating the calculation.

MAPMAKER/HOMOZ (Kruglyak et al. 1995) also allows rapid multipoint analysis, including homozygosity mapping. MAPMAKER/HOMOZ can be used in small consanguineous pedigrees with missing genotype information. The limiting factor in the analysis is the relatively small number of meiosis (about 16) that can be analysed at a time.

Simulation programmes e.g. SLINK (Weeks et al. 1980, Ott 1989), which is based on the LINKAGE programme, and SIMLINK (Boehnke 1986) can be used to estimate whether it is possible to localise a disease gene based on linkage analysis in a given set of families.

3.3.2.4.4. Polymorphic markers

The first polymorphic markers used for linkage analysis were *restriction fragment length polymorphisms (RFLPs)* (Botstein et al. 1980). RFLPs are often based on a single nucleotide variations at restriction sites. Initially, restriction digestion and Southern blotting were used in the analysis of RFLPs making the method laborious. Nowadays RFLPs can be typed by polymerase chain reaction (PCR).

Polymorphism information content (PIC) is a measure of the informativeness of a polymorphic marker. The number of alleles and the allele frequencies determine the PIC value. A marker with a PIC value of 0 is never informative, whereas a marker with a PIC value of 1 is always informative and most likely reveals whether a gamete is a recombinant or not.

The PIC value of the RFLPs is low, as the markers usually have only two alleles. The RFLPs can also be caused by variations in the number of short repeated sequences. They are referred to as *variable number of tandem repeats, VNTRs* (Nakamura et al. 1987). The analysis of VNTRs was initially based on restriction digestion and Southern blotting. The VNTRs are less frequently detected in the human genome than single nucleotide variations at restriction sites, but their PIC value is usually high.

The PCR method was introduced in the mid 1980´s (Mullis et al. 1986) providing an efficient method for genotyping *short tandem repeats (STRs)*, also called microsatellite markers (Litt & Luty 1989, Weber & May 1989). The microsatellite markers are prevalent and the PIC values are frequently high. They account for about 0.5% of the human genome and occur approximately once in every 30 kb (Strachan and Read 1996). Several CA-repeat microsatellite maps have been published (Weissenbach et al. 1992, Gyapay et al. 1994, Murray et al. 1994, Dib et al. 1996).

Single nucleotide polymorphism markers (SNPs) represent the latest generation of markers used for genotyping. An SNP is defined as a position at which two alternative bases occur at appreciable frequencies (Collins et al. 1997). SNPs are not prone to spontaneous mutations. The PIC values of SNPs are low compared with microsatellites.

Consequently, at least 2 to 5 times more SNPs will be required to achieve the same power as with microsatellite markers in pedigree based mapping studies (Kruglyak 1997). Furthermore,

the higher density of markers will also require a very high resolution physical map to assure the proper order of markers. This probably will not be accomplished until the full sequence of the human genome is known.

3.3.2.4.5. Family material

Families with two or more affected individuals are informative in a conventional linkage study, as they provide the possibility to know which alleles are inherited from which parent, and each meiosis can be scored as recombinant or nonrecombinant.

The enormous number of polymorphic markers and the automated genotyping systems (Ansorge et al. 1987, Karger 1996, Perlin et al. 1995) currently available allow effective genotyping. Actually, the main limitation in a high-throughput microsatellite genotyping is the required manual editing of allele calls (Pálsson et al. 1999).

This development has reduced significantly the number of families needed for a linkage study. Diastrophic dysplasia (DTD) was one of the first mapped disease loci of the diseases of the Finnish disease heritage. The DTD locus was mapped by linkage analysis with 13 multiplex families (Hästbacka et al. 1990), whereas more recently Cohen syndrome (Tahvanainen et al. 1994), Meckel syndrome (Paavola et al. 1995), and Muscle-eye-brain disease (Cormand et al. 1999) loci were assigned by linkage analysis with only five, five, and four multiplex families, respectively. Furthermore, the infantile-onset spinocerebellar ataxia locus was mapped with samples from only four patients but a different statistical approach, homozygosity mapping, was utilised (Nikali et al. 1995).

3.3.2.5. Linkage disequilibrium and haplotype analysis

Concept of linkage disequilibrium

Linkage disequilibrium and haplotype analysis can be used to refine the initial disease locus assignment in isolated populations.

The majority of patients suffering from a hereditary disease in an isolated population share the same mutation inherited from a common ancestor. Usually, the affecteds also share alleles at marker locus close to the disease gene. This non-random association of alleles at a linked locus is called linkage disequilibrium (LD). LD is seen when apparently independent chromosomes originate partly from the same ancestral chromosome. The extent of LD decreases exponentially with time, at a rate proportional to the recombination rate (Rannala & Slatkin 1998).

Families with only one affected individual, and data about the historical recombinations can be used in LD analysis increasing markedly the number of exploitable meiosis compared to linkage analysis.

Haplotype analysis

When a disease mutation and marker loci lie close to each other in the genome, a certain set of alleles of the markers will co-segregate with the disease mutation, and the disease phenotype, constituting a haplotype. The “ancestral” haplotype associated with the original founder mutation will be broken by historical recombinations that occurred during the centuries following the introduction of the mutation into the population.

If a long period of time has elapsed since the origin of the mutation, the haplotype will be broken by several, random historical recombinations, and subsequently, the length of the shared haplotype decreases (Fig. 3).

When analysing haplotype data, it must be remembered that markers, in particular microsatellites, do undergo spontaneous mutations leading to false interpretations. The high heterozygosity of a marker locus is associated with an increased risk of spontaneous mutation at the locus (Jeffreys et al. 1988, Bowcock et al. 1993).

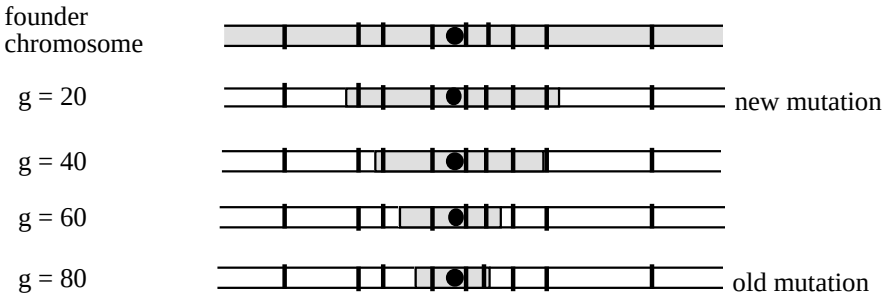

Fig. 3 The extent of linkage disequilibrium.

The mutation is marked with a black point, polymorphic markers are indicated with bold vertical lines, and the shaded area represents the shared haplotype in the disease chromosomes. The founder chromosome is shown above, and below that the disease chromosomes after generations (g) indicating the shortening of the shared haplotype. If the mutation is new the shared haplotype is detected over a longer region than in the case of an older mutation.
(Adapted from de la Chapelle et al. 1994)

LD and statistical formulas

Several statistical formulas for estimating the distance between a mutation and a marker locus on the basis of LD have been reported (Hästbacka et al. 1992, Lehesjoki et al. 1993, Terwilliger 1995, Kaplan et al. 1995, Xiong & Guo 1997, Rannala & Slatkin 1998).

Hästbacka et al. (1992) and Lehesjoki et al. (1993) applied the Luria-Delbrück formula originally designed for the analysis of bacterial cultures (Luria & Delbrück 1943) to study LD in the Finnish population. This analysis allows estimation of the recombination fraction between a disease locus and a marker, the degree of allelic homogeneity among disease causing alleles, and the mutation rate for nearby genetic markers.

Since the formula expects the population to grow exponentially with random mating, and some of the parameters have to be estimated according to the population history, the resulted estimates must be evaluated cautiously. The Luria-Delbrück analysis works well for recessive disorders that are primarily due to a single disease-causing mutation (Jorde 1995).

Terwilliger presented in 1995 a new likelihood-based approach to test LD. This method allows for multiple polymorphic marker loci to be studied simultaneously without reducing the analysis to a diallelic situation. Poorly matched case and control subjects may present a problem in association studies, as is reported in other approaches as well. While studying Mendelian disorders this problem is accounted for by treating non-transmitted alleles as the control subjects and transmitted alleles as the case subjects. Also this method has been applied in several disease gene identification projects (Nikali et al. 1995, Virtaneva et al. 1996, Aaltonen et al. 1997).

3.3.2.6.Physical mapping

After a disease gene locus has been refined by genetic mapping one usually initiates physical mapping of the region. There are several methods for physical mapping, which are usually applied simultaneously. The most commonly used methods are described below.

Radiation hybrids

Radiation hybrids (RH; Cox et al. 1990, Gyapay et al. 1996) provide a possibility for high-resolution physical mapping without a contig of clones. After fusing irradiated donor cells with recipient rodent cells (e.g. hamster), a panel of hybrid cell lines is produced. Typically, PCR reactions are carried out on each cell line of the panel using primers for an EST, a sequence tagged site (STS), or a polymorphic marker. Markers which are close to each other will be retained in the same cell line more often than markers which are further apart. The resolution of the radiation hybrid mapping is a function of the donor fragment size, which can be varied by altering the radiation dose.

A contig of overlapping clones

A contig of overlapping yeast and bacterial clones is an important tool in the process of disease gene identification. It is frequently constructed using yeast artificial chromosomes (YACs) (Schlessinger 1990, Burke et al. 1992), bacterial artificial chromosomes (BACs) (Shizuya et al. 1992), bacteriophage P1-based cloning systems (P1s) (Sternberg et al. 1990, Pierce et al. 1992), P1-derived artificial chromosomes (PACs) (Ioannou

et al. 1994), and cosmids (Collins & Hohn 1978).

The insert size of YACs ranges from 300 to 2000 kb. The YACs are frequently chimeric and rearranged making their usage laborious. A more stable and non-chimeric cloning system is provided by the BACs, P1s, and PACs, which have an insert size range of 70-300 kb. The cosmid vectors are again more unstable and have a 30% rate of chimerism. Nevertheless, cosmid vectors are suitable for applications requiring a small fragment of human DNA (30-44 kb) at a time.

Due to the efforts of several genome centers, e.g. Centre d'Etude du Polymorphisme Humain (CEPH) and Whitehead Institute Center for Genome Research (WICGR), a contig of YACs (Cohen et al. 1993, Chumakov et al. 1995) is usually available on the region of interest. If not available, YACs, as well as BACs, PACs, and cosmids can be conveniently screened by hybridisation or PCR from chromosome specific or whole genomic libraries (Shizuya et al. 1992, Ioannou et al. 1994).

STS content mapping

As a positive clone has been identified, the end sequences of the cloned DNA will be determined. A novel STS primer pair is designed from the unique end sequence. The presence of the STS in other clones of the contig is tested by PCR assay. Moreover, the STS can be utilised to screen a clone library for novel clones. This method is called STS content mapping.

Pulsed-field gel electrophoresis

The sizes of YACs, BACs, and PACs can be determined by pulsed-field gel electrophoresis (PFGE) (Schwartz & Cantor 1984). Based on the PFGE analysis it is possible to estimate the size of the entire contig of clones. PFGE is a modification of agarose gel electrophoresis. During a PFGE run the relative orientation of the gel and the electric field is periodically altered, so that the DNA molecules are intermittently forced to change their conformation and direction of migration. The time taken for a DNA molecule to re-orient itself is size-dependent. PFGE is able to size-fractionate DNA fragments up to several megabases.

Fiber-FISH

The *fiber-FISH method* has been shown to be a powerful method for contig construction (e.g. Klockars et al. 1996, Nikali et al. 1997). The fiber-FISH method (Heiskanen et al. 1995) is a modification of fluoresence *in situ* hybridisation (FISH; Pinkel et al. 1986, Lichter et al. 1988). Fiber-FISH enables the ordering of overlapping and non-overlapping clones and the estimation of sizes of overlaps and gaps between clones using clones of known localisation and size as reference probes in hybridisation. This way the size of the entire contig of clones can also be estimated.

The DNA fibers for fiber-FISH analysis are derived from agarose embedded proteinase K treated cells. A piece of an agarose block is melted on a microscopic slide, and the DNA fibers are spread over the slide manually by another slide. This easy and fast preparation of the target for hybridisation is one of the main advantages of fiber-FISH. On the other hand, some variability in the stretching occurs between slides due to the manual performance.

The probe labelling, hybridisation, and fluorescence microscopy are performed as in a conventional FISH method (Pinkel et al. 1986, Lichter et al. 1988).

Contig can be used for several purposes
The contig of yeast and bacterial clones can be utilised for *mapping* genes, ESTs, STSs, and polymorphic markers by PCR assay.

Furthermore, the contig of clones enables the *isolation of novel polymorphisms* that are useful in the refinement of the disease locus.

When required the genomic clones can also be exploited for the *identification of expressed sequences* and genes (see 3.3.2.7.).

If the critical gene region is small enough (<300 kb), it is feasible to sequence the whole region. *Large-scale sequencing* projects usually apply the shotgun sequencing method (Venter et al. 1998). Initially, the starting material, a BAC or a PAC clone, is partially digested with a rare-cutter enzyme. The partially overlapping fragments are shotgun-cloned into a suitable vector and the overlapping clones are randomly sequenced. The sequences are aligned by computer programmes.

3.3.2.7. Identification of expressed sequences

Novel expressed sequences can be identified from genomic clones by several methods of which the most frequently employed are CpG island identification, direct cDNA selection, and exon amplification.

If the genomic sequence of the region of interest is available, novel coding sequences can be identified by homology searches and exon prediction programmes.

None of the methods is able to detect all of the existing genes. To ensure as many as possible of the genes to be identified, usually several gene identification methods are applied simultaneously. At the end conventional cDNA library screening is often needed to construct the full length coding sequence of the gene.

CpG island identification
CpG islands (Bird 1986) are short regions (<1 kb) of genome with high GC content (>60%). In humans approximately 56% of the genes are associated with CpG islands, including housekeeping and widely expressed genes, but leaving a substantial number of genes out of reach of the method. The CpG islands often have recognition sites for rare-cutter restriction endonucleases, and thus they can be identified by restriction mapping (Larsen et al. 1992).

Direct cDNA selection
In direct cDNA selection a genomic clone or a part of it is PCR amplified, bound to a solid support, and hybridised to a mixture of PCR amplified cDNAs, e.g. a cDNA library (Lovett 1994). Several consequent hybridisations are carried out to increase the yield and specificity of the target cDNA. In the hybridisations repetitive DNA sequences in the genomic clone have to be carefully blocked. The expression level of a gene, the length of the exons, and the selection of cDNA library influence the feasibility of isolating it by direct cDNA selection.

Exon amplification
The majority of mammalian genes have more than one exon. This feature of a gene is exploited in the exon amplification method (Buckler et al.

1991, Church et al. 1994, Church & Buckler 1999).

A genomic fragment of interest (e.g. a cosmid) is cloned into a specific exon-amplification vector including a SV40 promoter and two vector exons between which the target genomic fragment is cloned. The construct is transfected to e.g. COS cells.

The process of transcription initiates from the SV40 promoter of the vector, and the produced pre-RNA molecule undergoes splicing by the RNA splicing machinery of the host cell.

The mature RNA molecule consists of the vector exons in both ends and of the captured exon(s) in between. This "fusion" RNA is isolated, and subsequently cDNA is synthesised. Vector-exon-specific PCR primers are utilised for the amplification of the trapped exon/s. As the method is based on splicing it is able to identify genes with at least two exons.

Homology searches

Similarity search programmes like BLAST (Basic Local Alignment Search Tool), Gapped BLAST and PSI-BLAST can be used to explore all the available sequence databases (Altschul et al. 1990 and 1997) for homology of the sequence of interest with known nucleotide sequences or proteins.

The programmes compare the input sequence in six reading frames against the databases. The matches to previously known genes, cDNAs, and ESTs are presented in the output according to their probability. Any significant gene, cDNA, or EST match, whether human or non-human in origin, indicates a gene associated sequence: a functional gene, a pseudogene or a gene fragment.

The coding DNA sequences are often well conserved between species. Many human genes show significant matches to the genes of evolutionarily distant species such as *Saccharomyces cerevisae*, *Drosophila melanogaster* and *Caenorhabditis elegans*. As most of the gene and protein families have already been identified, it is very rare to find a novel gene without any homology to previously identified gene(s).

Exon prediction

There are several exon prediction programmes available, e.g. GeneMark (Isono et al.1994), Genie (Henderson et al. 1997), GRAIL 1a & 2 (*G*ene *R*ecognition and *A*ssembly *I*nternet *L*ink; Milanesi et al. 1993), MZEF (Michael Zhang's Exon Finder; Zhang 1997), and Xpound (Thomas & Skolnick 1994).

Some of the programmes are based on identifying internal exons by looking for conservative sequences at the splice acceptor, splice donor, and branch sites as well as looking for long open reading frames (ORFs). Some prediction programmes look for additional structural gene elements like promoter sites, translation sites, polyadenylation signals, and stop codons.

The programmes currently available locate at best 80% of the internal coding exons, and 5% of the predictions do not overlap a real exon (Claverie 1997).

Genotator (Harris 1997) presents a new "workbench" for sequence analysis. It runs five different gene-finding programmes (GRAIL, GeneFinder, Xpound, GeneMark, and Genie), three homology searches (repetitive elements, ESTs, and translated coding regions), and searches for promoters, splice sites, and ORFs.

3.3.2.8.Mutation screening

Mutation screening is the final step in a positional cloning project. A candidate gene or a part of it is tested for sequence variations in a set of patients and control samples. The most commonly used mutation screening methods are briefly described.

Methods for mutation screening
Using *Southern blot analysis* (Southern 1975) size-fractionated DNA fragments of patients and controls are compared after hybridisation with a gene-specific probe. This method is able to detect only large genomic rearrangements: deletions, insertions, and duplications.

Northern blot analysis is a variant of Southern blotting, in which the target is RNA instead of DNA (see also 3.4.). Northern blot analysis is able to detect mutations that alter the level of expression or the size of transcript.

Single-strand conformational polymorphism (SSCP) analysis (Orita et al. 1989) is based on a tendency of single-stranded DNA to fold up according to the DNA sequence in a non-denaturing gel. Any variation in DNA sequence is seen in the gel as a mobility shift. SSCP analysis is suitable for relatively short PCR fragments (<250 bp). To maximize differential migration of DNA fragments in SSCP it is crucial to optimise the experimental conditions. Thus, not surprisingly, the sensitivity of SSCP for mutation detection has been reported to range from 30% to 100% (Michaud et al. 1992, Sarkar et al. 1992, Claustres et al. 1993, Sheffield et al. 1993).

Chemical cleavage of mismatches (CCM) (Moser & Dervan 1987, Grompe et al. 1989) detects mismatches, variations in DNA sequence, by a series of chemical reactions. This highly sensitive method can be applied to sequences of up to 1 kb. Since the necessary reagents are toxic the usage of CCM is inconvenient.

Denaturing gradient gel electrophoresis (DGGE) (Cariello & Skopek 1993) is based on migration of DNA duplexes on a chemical or a temperature gradient gel. Migration continues until the DNA duplexes reach a position in the gel where the strands melt and separate, after which the denatured DNA does not migrate much further. The melting temperature of the fragment is determined by the DNA sequence. Primers with long stretches Gs and Cs, that is GC-clamps, are used to maximize the sensitivity of the method for mutation detection. The need for GC-clamps in the primers makes DGGE an expensive procedure.

If a difference in Southern blotting, Northern blotting, SSCP, CCM, or DGGE is detected between the patients and the control individuals, one has to ultimately characterise the variation at nucleotide level by sequencing. As the sequencing facilities have improved during the last years, frequently one uses *direct sequencing* for primary mutation screening.

Mutation or polymorphism?
A sequence variation may be a polymorphism or a disease causing mutation.

A variation can be considered to be a disease causing mutation if one finds in the same gene at least two separate variations in different families affected by the same disease. A disease causing mutation also changes the amino acid code. If only one amino acid is changed its significance can be evaluated by

studying whether the amino acid is highly conserved between species. Highly conserved amino acids are functionally important and sequence variations abolishing them are likely to be mutations.

Additionally, in recessive disorders the variation should not be identified in a homozygous form in unaffected control individuals, and the heterozygous carrier frequency should not exceed the frequency expected by the disease incidence in the population.

The localisation and type of a sequence variation affects the probability of it being a disease causing mutation.

Deletions, insertions, and duplications in an exon are likely to be deleterious especially when causing a frameshift of the open reading frame (ORF). On the other hand, such variations are commonly detected in intronic sequences as polymorphisms.

A *single nucleotide variation* that causes a premature stop codon in an exon is likely to be deleterious, whereas if the consequence is merely a change of an amino acid, it may be a polymorphism as well as a disease causing mutation (see above). A single nucleotide variation in an intron is most likely a polymorphism, unless it produces a cryptic splice site or is localised in the consensus sequence of the splice acceptor or splice donor site.

On one hand *unstable trinucleotide repeats* located in 5´ and 3´ untranslated regions and coding regions are known to be associated with at least ten inherited neurological disorders (Ashley & Warren 1995, Warren 1996). On the other hand intronic microsatellite repeats (1-4 bps) are frequent and polymorphic. Interestingly, an intronic GAA triplet repeat expansion causes Freidreich´s ataxia (Campuzano et al. 1996).

A *minisatellite expansion* (12 bps) in the promoter region of the cystatin B gene, was reported to underlie EPM1, a recessively inherited neurodegenerative disorder (Lalioti et al. 1997, Virtaneva et al. 1997), introducing a novel type of human mutation.

3.4. Analysing gene expression

Gene expression studies provide valuable information about the function of a gene, as they reveal tissues, cell types, and developmental stages at which a gene is functioning.

Gene expression analysis is also useful in evaluating candidate genes. In a multi-organ phenotype the causative gene should be expressed ubiquitously (e.g. The Finnish-German APECED Consortium 1997). However, in a phenotype restricted to a particular tissue the underlying gene may be expressed only in the affected tissue (e.g. Höglund et al. 1996) or in most or all of the tissue types (e.g. Pennacchio et al. 1997).

Northern blot analysis is the basic method for studying gene expression. Either total RNA or mRNA is fractionated according to size using denaturing gel electrophoresis. RNA is subsequently transferred to a membrane which is hybridised with a labelled probe. If the studied gene is a single copy gene, a cDNA clone can be used as a probe. If the gene is a member of a gene family, the probe should be a PCR or a RT-PCR fragment from a gene sequence that is not conserved between family members. A Northern blot membrane consisting of different tissue types reveals the gross expression pattern, but may also detect different isoforms of RNA produced by alternative splicing.

Reverse-transcriptase PCR (RT-PCR) can also be used for studying different isoforms of RNA as well as for a rough quantitation of gene expression (Pykett et al. 1994). RT-PCR is a version of the standard genomic PCR reaction. First one isolates total or mRNA, and then converts it to cDNA using reverse transcriptase enzyme. The cDNA is used as a template for a PCR reaction with exon-specific primers. Extra amplification products in addition to the expected products may represent different isoforms of the RNA.

3.5. RING finger-B-box-coiled-coil proteins

In the positional cloning project described in this thesis mutations in a novel gene encoding a RING finger-B-box-coiled-coil (RBCC) protein were identified in MUL patients. In this chapter the current knowledge about RBCC proteins is reviewed.

The RBCC family consists of proteins that have a RING finger, and one or two B-box motifs followed by a coiled-coil domain (Freemont 1993, Saurin et al. 1996; Fig. 4a).

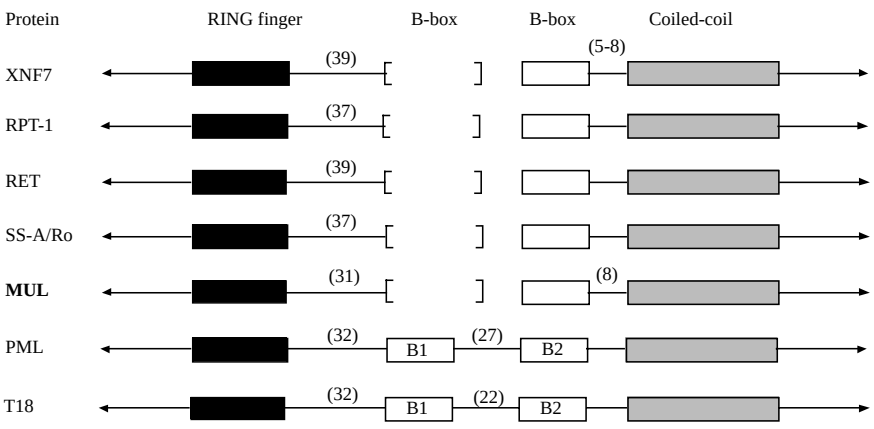

Fig. 4a Schematic representation of RING finger, B-box, and coiled coil domains in RBCC proteins.
The sequence spacing between domains is indicated in parentheses, and the symbol [ ] represents a gap.

RING finger proteins

The RBCC proteins constitute a subgroup in the family of RING finger proteins. The RING finger motif was found in the sequence of the human *RING1* (Really Interesting New Gene 1) gene proximal to the MHC region on chromosome six (Hanson et al. 1991). The RING finger is a cysteine rich motif, defined by the following sequence: Cys1-X_2-Cys2-$X_{(9\text{-}27)}$-Cys3-$X_{(1\text{-}3)}$-His1-X_2-Cys4-X_2-Cys5-$X_{(4\text{-}48)}$-Cys6-X_2-Cys7, where X can be any amino acid.

The RING finger motif uses eight ligands, Cys and His residues, to bind two zinc atoms in a unique cross-brace system (Borden et al. 1995a). In this system the first and third pairs of metal-binding ligands share one zinc atom and the second and fourth pairs share the second zinc.

RING finger proteins have been identified from human to plants to yeast to viruses but interestingly not in prokaryotes (Saurin et al. 1996). Presently, there are several hundred cDNAs encoding RING finger proteins in the GenBank database. *BRCA1*, the breast cancer predisposing gene, is one of the well-known RING finger genes (Wu et al. 1996).

Recent studies have shown that the RING finger domain plays an important role in mediating specific ubiquitylation events (Waterman et al. 1999, Yokouchi et al. 1999, Joazeiro et al. 1999, Lorick et al. 1999). It has been suggested that all RING proteins act as E3 ubiquitin protein-ligases involved in a variety of biological processes (Freemont 2000). This is in keeping with the functional diversity of the RING finger family proteins.

Nevertheless, most RING finger proteins have either unknown functions or functions not obviously related to ubiquitination. It has been proposed that RING-mediated ubiquitination targets the RING containing protein or associated protein for degradation in a regulated manner (Lorick et al. 1999).

B-box motif

A RBCC protein contains, C-terminal to the RING finger domain, one or two B-box motifs (Freemont 1993; Fig. 4a), which represents another cysteine-rich motif. The consensus motif of a B-box can be defined as Cys1-X_2-His1-X_7-Cys2-X_7-Cys3-X_2-Cys4-X_5-His2-X_2-His3 (Fig. 4b).

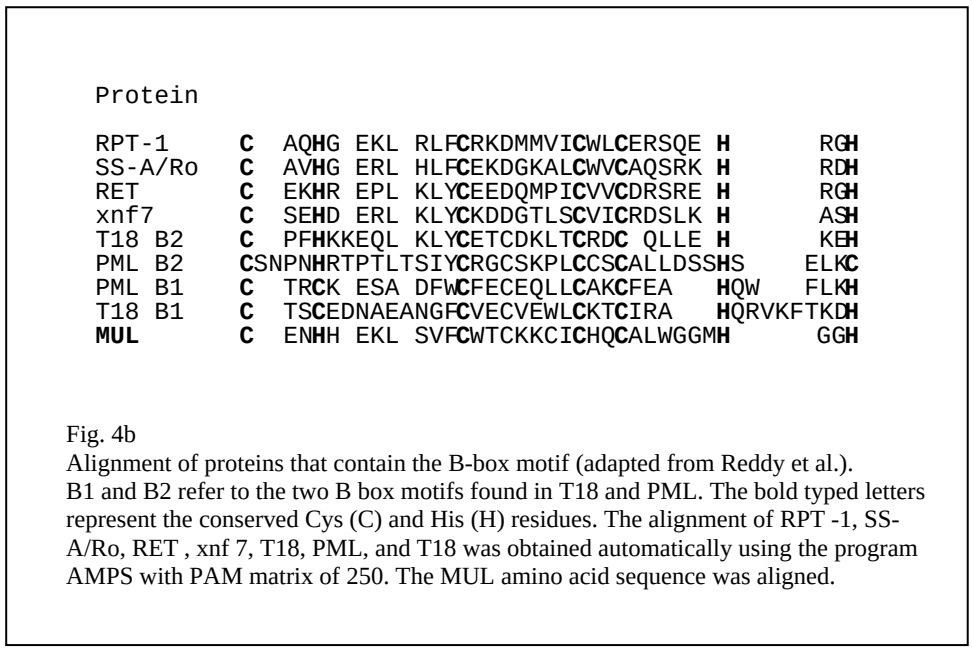

Fig. 4b
Alignment of proteins that contain the B-box motif (adapted from Reddy et al.). B1 and B2 refer to the two B box motifs found in T18 and PML. The bold typed letters represent the conserved Cys (C) and His (H) residues. The alignment of RPT -1, SS-A/Ro, RET , xnf 7, T18, PML, and T18 was obtained automatically using the program AMPS with PAM matrix of 250. The MUL amino acid sequence was aligned.

Although there are seven conserved ligands, Cys and His residues, it has been shown that only one zinc atom is bound per B-box (Borden et al. 1993, Bellini et al. 1995). According to the existing opinion the zinc-binding atoms of the B-box are: Cys1, His1, Cys4, and His2 (Borden et al. 1995b). The rest of Cys and His residues have been suggested to be conserved for functional reasons; they seem to be important in protein-protein interactions (Cao et al. 1997).

Unlike the RING finger, the B-box motif is highly conserved between family members (Fig. 4b). Furthermore, the sequence spacing between the RING finger and the first B-box is highly conserved (31-39 amino acids; Fig. 4a). This conservation suggests that RING and B-box domains may interact, forming an integrated fold. The second B-box is followed closely (5-8 amino acids) by a coiled-coil domain.

Coiled-coil domain

The coiled-coil domain is a predicted secondary structure - a bundle of α-helices that are wound into a superhelix (Lupas 1996a). The coiled-coils have a distinctive, repetitive pattern of hydrophobic and hydrophilic residues called the heptad repeat.

The long heptad repeats can be easily picked up by eye. For the detection of short heptad repeats at least two

computer programmes are available via the world wide web which take into account the residue distribution of known coiled-coils (*Coils*, Lupas 1996b; *Paircoil*, Berger et al. 1995).

Originally, the coiled-coil was described (Pauling & Corey 1953, Crick 1953) as the main structural element of fibrous proteins as keratin, myosin, and fibrinogen. It was only until the late 1980´s when coiled-coil proteins were identified as the dimerization element in the leucine zipper proteins (Landschulz et al. 1988).

In the 1990´s several transcriptional activators with short coiled-coils of two to five heptads have been shown to mediate dimerization, heterodimerization, and even to bind to the DNA (reviewed by Lupas 1996a).

Functions of RBCC proteins

The current members of the RBCC family have variable subcellular locations and functions. The RBCC proteins are thought to be involved in protein-protein and/or protein-nucleic acid interactions and in the formation of multi-protein complexes (Borden & Freemont 1996).

The RBCC family includes putative transcription factors, like the *Xenopus* nuclear factor (XNF7) (Borden et al. 1993), and the TIF proteins (transcriptional intermediary factors) TIF1α (Le Douarin et al. 1998) and TIF1β (Chang et al. 1998).

Interestingly, three RBCC genes, *ret* (Takahashi et al. 1988), *pml* as *pml-rara* (Borden et al. 1996), and *t18* (Le Douarin et al. 1995) have transformation potential when found as translocations in human and mice. In all of these fusions the zinc-finger and coiled-coil domains are retained.

Another RBCC member, SS-A/Ro, is a ribonucleoprotein and an autoantigen in patients with systemic lupus erythematosus and Sjögren syndrome (Chan et al. 1991).

RBCC proteins and human diseases

Familial Mediterranean fever (FMF) and Opitz syndrome (OS) are the only previously reported disorders in which mutations in genes encoding RBCC proteins are responsible for the primary defect.

The recessively inherited FMF characterised by dramatic episodes of fever and serosal inflammation, is caused by mutations in a novel RBCC gene, called *MEFV* (The French FMF Consortium 1997).

OS is a multi-organ disorder primarily affecting midline structures, and it was shown to be due to mutations in X-chromosomal *MID1* encoding the RBCC protein midin (Quaderi et al. 1997).

Midin is associated with microtubules throughout the cell cycle (Cainarca et al. 1999, Schweiger et al. 1999). Furthermore, midin is shown to mediate homodimerization and to exist in the form of large protein complexes (Cainarca et al. 1999).

Mutations found in OS patients are reported to reduce the ability of midin to bind microtubules. Instead, cytoplasmic clumps are formed (Cainarca et al. 1999, Schweiger et al. 1999). However, some of the mutants still possessing RBCC domain were shown to retain their ability to homodimerize and form multiprotein complexes (Cainarca et al. 1999).

4. AIMS OF THE PRESENT STUDY

• To determine the chromosomal localisation of the Mulibrey Nanism (*MUL*) locus by linkage analysis.

• To refine the *MUL* locus by linkage disequilibrium analysis.

• To cover the *MUL* region with a contig of yeast and bacterial clones.

• To localise candidate genes to the *MUL* region.

• To identify and characterise the *MUL* gene.

5. MATERIALS AND METHODS

5.1. Subjects

A total of 50 Finnish and 2 non-Finnish MUL patients were studied. The pedigrees of the Finnish multiplex families are shown in Fig. 5. The diagnosis of the MUL patients was based on the diagnostic criteria summarised in Table 2. The Finnish control individual DNA samples were provided by the Finnish Red Cross Blood Transfusion Service. The 100 non-Finnish control individual samples originated from CEPH. All the samples were obtained after informed consent.

5.2. DNA and RNA extraction

DNA was extracted from fresh or frozen venous blood using the non-enzymatic DNA extraction method (Sambrook et al. 1989). For RNA extraction EBV transformed lymphoblast cell cultures were established. Total RNA was extracted using the RNeasy RNA extraction kit (Qiagen) according to the manufacturer´s instructions. QuickPrep Micro mRNA Purification Kit (Amersham Pharmacia Biotech) was used for the extraction of polyA RNA according to the manufacturer´s instructions.

5.3. Microsatellite marker analysis

In the genome-wide search for linkage, 240 CA-repeat markers originating from Généthon were used covering the genome in approximately 20 cM intervals (Weissenbach et al. 1992, Gyapay et al. 1994, Dib et al. 1996). The PCR reactions were performed as described previously (Weber & May 1989). The initial systematic screening for linkage was carried out at Généthon, Paris, using a non-radioactive genotyping method (Vignal et al. 1993). Several amplification products from the same DNA sample, generated with different primer sets, were pooled and co-precipitated and co-electrophoresed in a single lane of a 6% polyacrylamide, 50% urea sequencing gel. After transfer to nylon membranes, hybridisation was performed with peroxidase labelled PCR primers. The alleles were detected using a chemiluminescence based detection system (ECL, Amersham Pharmacia).

In refining linkage, the following Généthon markers were used: D17S1799, D17S1607, D17S1606, D17S1853, D17S1604, D17S1811, D17S1855, and D17S948. Marker D17S1290 originated from the Cooperative Human Linkage Center (CHLC). Markers 272a-CA, 272b-CA, 132-CA, 58-CA, and 52-CA were identified by a CA-repeat isolation protocol (see below), and marker 95-CA was found in the end sequence of a PAC clone (95i19). All these markers were analysed by radioactive genotyping method described previously (e.g. Sulisalo et al. 1993).

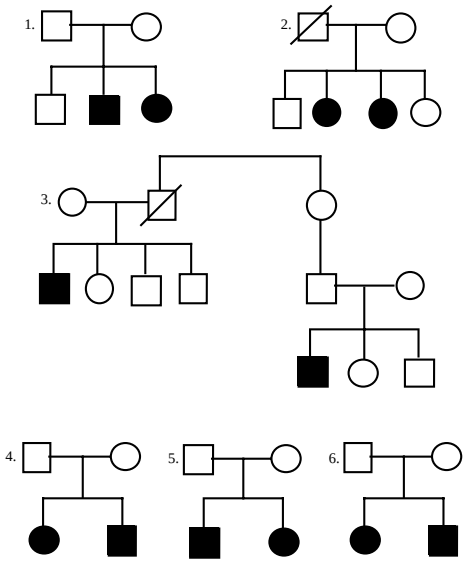

Fig. 5
The pedigrees of the six MUL multiplex families that were used in the initial linkage analysis.

5.4. Linkage analysis

By the simulation programme SLINK (Weeks et al. 1990, Ott 1989) the panel of six multiplex families (Fig. 5) was estimated to give an average lod score of 2.7 (θ=0.0) assuming a heterozygosity value of 0.70 for the markers.

In the initial screening for linkage, the films were scored for the presence or absence of recombinations by comparing familywise the alleles of affected individuals in multiplex families.

Two-point linkage analyses were carried out using the MLINK programme of the LINKAGE package OS/2-version 5.2 (Lathrop et al. 1984).

Multipoint linkage analysis was performed with VITESSE (O´Connell & Weeks 1995). Multipoint analysis was carried out under the assumption of a fixed order and fixed distances of the eight marker loci included in the analysis (Dib et al. 1996). Allele data for loci D17S1811 and D17S1855 were combined for the multipoint computation analysis with VITESSE while the other loci were considered separately. All calculations were done assuming complete penetrance and equal recombination frequency in males and females.

5.5. Linkage disequilibrium and haplotype analysis

To test whether there is a single allele with a significantly higher frequency on

disease bearing chromosomes than on control chromosomes, we used Fisher´s exact test with a Bonferroni´s correction for multiple testing (e.g. Hästbacka et al. 1992). The non-MUL bearing chromosomes of the parents were used as control chromosomes.

The Luria-Delbrück based calculation (Hästbacka et al. 1992, Lehesjoki et al. 1993) was used to estimate the genetic distance between the polymorphic markers and *MUL* under the following assumptions: the number of generations (g) since "founding", i.e. the beginning of spreading of the *MUL* mutation in the population, was 10-100; the mutation rate (μ) of the *MUL* locus was 5×10^{-6}, as the disease is rare world-wide; the overall *MUL* gene frequency (q) was 0.005 based on the Hardy-Weinberg equilibrium and the actual observed incidence of 1:40 000 of the disease phenotype in Finland (Lipsanen-Nyman 1986). Allelic excess was calculated by the formula: $p_{excess}=(p_{affected}-p_{normal})/(1-p_{normal})$, where p denotes allele frequency. The theta value (θ) or recombination between *MUL* and a marker locus was calculated using the formula $p_{excess}=(1-\mu g q^{-1})(1-\theta)^g$, where $(1-\mu g q^{-1})$ denotes the proportion of MUL chromosomes carrying the same ancestral mutation.

The significance of linkage disequilibrium between the disease and marker loci was also analysed using the DISLAMB programme (Terwilliger 1995). This programme applies a likelihood-based test for linkage disequilibrium and has only one degree of freedom, irrespective of the number of alleles at any given marker. The calculations are based on the parameter λ, which expresses the proportion of the increase of a certain allele in the disease chromosomes, relative to its population frequency. Confidence intervals for λ were calculated for 1-lod units.

The putative haplotypes were constructed manually, assuming a minimum number of recombinations in each family.

5.6. Construction of a contig of yeast and bacterial clones

ESTs, STSs, and YAC clones that were mapped between markers D17S957 and D17S1604 by WICGR and CEPH were utilised in the initial construction of the YAC contig (Hudson et al. 1995, Dib et al. 1996, Schuler et al. 1996; http://www-genome.wi.mit.edu).

Individual YAC clones were cultured, and total yeast DNA was extracted and stored in agarose beads. The chimerism of YAC clones was analysed by the metaphase-FISH technique (Bray-Ward et al. 1996), and clones mapping to only 17q22-q23 were used for further studies. The presence of STSs and ESTs in the YACs was confirmed by PCR amplification.

The STSs and ESTs present in the YACs were used for screening of PAC and BAC libraries by PCR assay. PCR-positive clones were picked and cultured in 100 ml Terrific Broth media or 400 ml Luria Bertani media, supplemented with 25 mg/ml of kanamycin (PACs) (Ioannou et al. 1994) or 12.5 mg/ml of chloramphenicol (BACs) (Shizuya et al. 1992). DNA was extracted using a Plasmid Maxiprep (tip 500) kit® (Qiagen) according to manufacturer's instructions.

The end fragments of the PAC and BAC clones were sequenced directly from the clone DNA with SP6 and T7 vector primers. New STSs were designed from the non-repetitive end sequences using the Primer3 software

(WICGR) and further used for the screening of genomic libraries. The localisation of STSs and ESTs in the PAC/BAC contig was determined by PCR amplification.

5.7. Isolation of novel microsatellite markers

Novel CA-repeats were isolated from the PAC and BAC clones using the following procedures. The PAC DNA was digested with Sau3A1 and ligated to pGEM7, transformed into DH5α cells and plated onto selective agar plates. BAC DNA was digested with RsaI and ligated to SmaI cut pUC18 vector (Pharmacia, Biotech), transformed into Epicurian Coli SURE® competent cells (Stratagene) and plated onto selective agar plates.

The colonies were transferred onto nylon filters (Hybond-N®, Amersham or Magna Chart®, MSI) and hybridised with a radioactively labelled $(GT)_{16}$ oligonucleotide at +65 °C overnight. After autoradiography, positive colonies were selected, and DNA isolated using a Plasmid Miniprep (tip 20) kit® (Qiagen), or Wizard Plus Miniprep kit® (Promega).

The inserts were sequenced with SP6 and T7 vector primers. Oligonucleotide primers flanking the CA-repeats were selected with Primer3 software (WICGR).

5.8. Fluorescence *in situ* hybridisation

The metaphase chromosome targets were achieved by culturing human peripheral blood lymphocytes according to standard protocols (Lemieux et al. 1992, Tenhunen et al. 1995). Agarose embedded lymphocytes were used as a source of free DNA fibers according to the method described previously (Heiskanen et al. 1994, 1996). All PAC and BAC clones used for the FISH experiments were labelled with biotin 11-dUTP (Sigma) and digoxigenin 11-dUTP (Boehringen Mannheim) by nick translation according to standard protocols.

The FISH procedure was carried out as described previously (Pinkel et al. 1986, Lichter et al. 1988, Aaltonen et al. 1997). Briefly, the hybridisation mixture contained 50% formamide and 10% dextran sulfate in 2xSSC. Repetitive sequences were suppressed with 10-fold excess of Cot-1-DNA (BRL). After overnight incubation at +37 °C the slides were washed at +44 °C three times in 50% formamide and 2xSSC, twice in 2xSSC, and once in 0.5xSSC. The slides were counterstained with DAPI (Sigma).

A multicolor image analysis was used for acquisition, display and quantitative analysis of fiber-FISH as described elsewhere (Heiskanen et al. 1996, Aaltonen et al. 1997).

5.9. Estimation of clone insert sizes by pulsed-field gel electrophoresis

NotI-digested BAC or PAC DNA was separated in 1% agarose using a work-shop made pulsed-field gel electrophoresis apparatus designed according to the principles described by Chu et al. (1986). The electrophoretic conditions were 120 V, 60 s pulse time for 48 h at 16 °C. The fragments were visualised by staining the gel with ethidium bromide and their sizes were estimated by comparing them with the size marker.

5.10. Mutation detection and screening

5.10.1. Sequencing

The PCR primers for the mutation analysis of *PNUTL2* were designed from the genomic sequence of a BAC clone, hRPK.112_H_10 that revealed the genomic organisation of *PNUTL2*. DNA was used as template for PCR amplification. The PCR primers for the mutation analysis of KIAA0898 were designed from the cDNA sequence, and subsequently the cDNA was used as template for PCR.

The purified PCR products (PCR Purification Kit, Qiagen) were sequenced using the PCR primers. Either the ABI PRISM Dye Terminator or the ABI PRISM Rhodamine cycle sequencing kit (Perkin-Elmer) was used according to the manufacturer's instructions. The sequencing reactions were run on an ABI 373 or 377 sequencer (Perkin Elmer).

5.10.2. SSCP analysis and non-denaturing-PAGE analysis

In SSCP analysis the denatured PCR amplified fragments were separated in 0.8xMDE™ gel (FMC) for 5-18 hours at 10 W and the alleles were visualised by silver staining. In the non-denaturing-PAGE analysis the PCR amplified fragments were run on a 12% polyacrylamide gel for 2-3 hours at 100 V, and the alleles were visualised with ethidium bromide.

5.11. Gene expression studies

5.11.1. Northern blot analysis

Two µg of polyA RNA or 10 µg of total RNA extracted from the EBV transformed lymphoblasts of MUL patients and non-MUL controls was run on a 1.5% agarose gel electrophoresis for 4-6 hours at 100 W. The RNA fragments were transferred to nylon membrane over night in 2xSSC. Commercially available multiple tissue northern blots were also analyzed (Human and Human II, Clontech). RT-PCR fragments and a β-actin control probe (Clontech) were used as probes. The probes were labelled with [α^{32}P]dCTP by random-priming using the RediPrimeII Kit (Amersham Pharmacia). After labelling the probe was purified with a Nick column (Amersham Pharmacia), and blocked with both human placental DNA (0.25 µg/µl) and salmon sperm DNA (0.5 µg/µl). The filters were hybridised for 2 hours at 65 °C in ExpressHyb solution (Clontech) and washed at 65 °C according to the manufacturer's instructions. After autoradiography for 2-6 hours at -80 °C X-ray films were developed.

5.11.2. RT-PCR analysis

cDNA was synthesised from lymphoblastoid RNA using M-MLV Reverse Transcriptase (Promega) and random hexamers (Promega). RT-PCR reactions were performed under standard PCR conditions and RT-PCR products were run on 1% agarose gel and visualised with ethidium bromide staining.

5.12. Computational sequence analyses

Sequence editing and aligning was performed using Sequencher 3.1 (Gene Codes Corporation). The nucleotide sequence databases and protein predictions programmes used in the analysis of KIAA0898 are shown in Table 3.

Table 3. The nucleotide sequence databases and protein prediction programmes used in the analysis of KIAA0898.

Nucleotide sequence databases	www-address
BlastN and BlastP comparisons	http://www.ncbi.nlm.nih.gov/BLAST/
The Unigene Resources	http://www.ncbi.nlm.nih.gov/UniGene/index.html
Protein prediction programmes	**www-address**
Pfam	http://pfam.wustl.edu/hmmsearch.shtml
ProfileScan	http://www.isrec.isb-sib.ch/software/PFSCAN_form.html
PSORT II	http://psort.nibb.ac.jp:8800/
Tmpred	http://www.isrec.isb-sib.ch/software/ TMPRED_form.html
SignalP V1.1	http://www.cbs.dtu.dk/services/SignalP/
Scanprosite	http://expasy.nhri.org.tw/sprot/scnpsit1.html
Kyte/Doolittle hydrophobicity plot	http://bioinformatics.weizmann.ac.il/hydroph/plot_hydroph.html

6. RESULTS AND DISCUSSION

6.1. Initial assignment of the *MUL* locus by linkage analysis (I)

A genome-wide search for linkage with 240 polymorphic markers was carried out in six multiplex families. After comparing familywise the alleles of the affecteds, the marker D17S787 at 17q was found to give a result suggestive of linkage: in five out of six families the alleles of the affecteds were identical, and in one family a recombination was detected. Four markers from four separate chromosomal regions showed identical alleles in four families. At these four loci, a recombination or an uninformative meiosis was detected in two families. More markers were genotyped around all the five regions and a linkage analysis was performed. The four regions with identical alleles were clearly excluded, whereas the 17q locus was shown to be linked to MUL (Fig. 6).

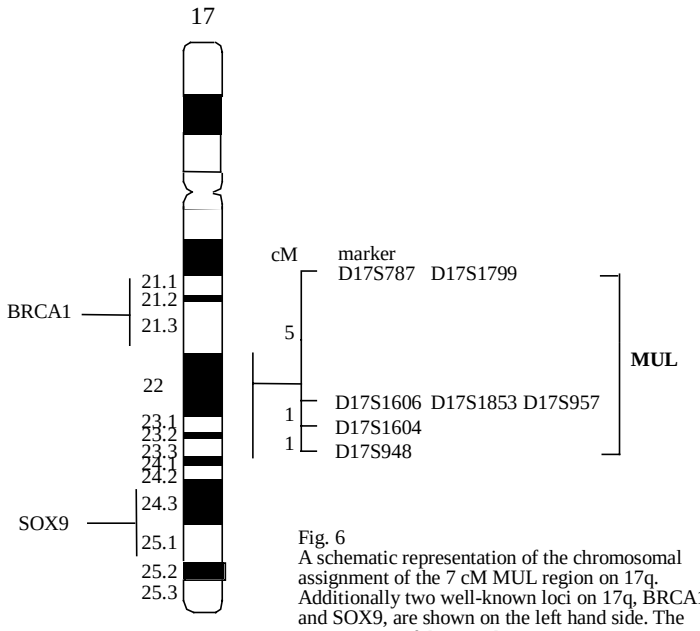

Fig. 6
A schematic representation of the chromosomal assignment of the 7 cM MUL region on 17q. Additionally two well-known loci on 17q, BRCA1 and SOX9, are shown on the left hand side. The genetic map of the initial 7 cM MUL region is shown on the right hand side.

Seven families with only one affected were included in the study. The critical *MUL* region was defined by recombinations to a 7 cM region between markers D17S1799 and D17S948 (Fig. 6). Within this region the highest two-point lod score of 4.46 (at θ=0.0) was obtained at locus D17S1604. In the multipoint linkage analysis with VITESSE a maximum LOD score of 5.01 was reached at loci D17S1606-D17S1853 and at locus D17S1604. The results strongly support the existence of *MUL* locus on chromosome 17q and provide strong evidence for genetic homogeneity in MUL in Finland.

6.2. Refinement of the *MUL* locus by linkage disequilibrium analysis (I and II)

The alleles at marker loci on the MUL chromosomes of 13 patients were compared with alleles on the normal chromosomes of the parents. A highly significant ($p<0.001$) allelic association was observed for the markers D17S1606, D17S1853 and D17S1604, the most common allele occurring in 74%, 85%, and 44% of the MUL chromosomes, but in 15%, 11%, and 4% of the normal chromosomes, respectively.

Luria-Delbrück analysis was used to estimate the genetic distances between the *MUL* gene and the marker loci demonstrating significant allelic association. The p_{excess} values for markers D17S1606, D17S1853 and D17S1604 were 0.70, 0.83, and 0.42 respectively. The genetic distance between *MUL* and three marker loci was calculated by assuming a single founding mutation, a *MUL* gene frequency (q) of 0.005, and a mutation rate (μ) of 5×10^{-6}. By these criteria the *MUL* gene resides closest to D17S1853. Assuming 50 generations since founding, the distance of *MUL* from D17S1853 was estimated to be 0.25 cM and assuming 25 generations, 0.65 cM. These estimates correspond to a *MUL* region of 500 kb and of 1300 kb, respectively, assuming that 1 cM corresponds to 1 Mb in the region.

The Luria-Delbrück analysis is considered to work well for recessive disorders that are primarily due to a single disease-causing mutation (Jorde 1995), as in the case of diastrophic dysplasia (Hästbacka et al. 1994). On the other hand, in progressive myoclonus epilepsy of Unverricht-Lundborg type that is caused by a enriched mutation in Finland, the analysis was slightly misleading (Virtaneva et al. 1996, Pennacchio et al. 1996). The Luria-Delbrück estimates are influenced by the relation of genetic and physical distances, that are known to vary substantially at different genomic regions.

The critical *MUL* region (defined below) has later been shown not to follow the average rule of 1 cM corresponding to 1 Mb. Instead, the 1 cM interval between markers D17S1606/D17S1853 (at the same genetic locus) and D17S1604 on the genetic map corresponds most likely to at least 2 Mb (see 6.5.). Furthermore, the Luria-Delbrück estimates are known to be subject to variation due to e.g. sampling error (Hästbacka et al. 1992).

The analysis of the novel markers isolated by us (see below) from the region between D17S1853 and D17S1604 revealed significant allelic association. Using the DISLAMB programme (Terwilliger 1995) the

strongest associations between *MUL* and an allele were detected at 95-CA, 272a-CA, 272b-CA, 132-CA, and 58-CA (λ-value > 0.95 with p-value < 10^{-5}; see also Figure 7).

6.3. Refinement of the *MUL* locus by haplotype analysis (unpublished, II, III)

The initial haplotype analysis in 26 MUL chromosomes assigned the *MUL* region between markers D17S1606/D17S1853 and D17S1604 (unpublished data).

Analysis of historical recombinations in haplotypes constructed from the novel markers (see below) in 42 MUL chromosomes refined the critical *MUL* region to a 1400 kb region between markers D17S1290 and 52-CA (Table 4; Fig. 7). Altogether five centromeric and four telomeric historical recombinations were detected in MUL chromosomes (Table 4).

After analysing a total of 100 chromosomes, the *MUL* locus was further refined by haplotype analysis to an approximately 800 kb between marker locus 95-CA and 58-CA (Table 4). Two patients were compound heterozygotes for the ancestral haplotype and a different haplotype which most likely represents a "minor" *MUL* mutation in the Finnish population (Table 4).

The 800 kb critical *MUL* region is relatively wide. In numerous other positional cloning projects the haplotype analysis has enabled to refine the critical gene region to less than 200 kb (Virtaneva et al. 1996, The Finnish-German APECED Consortium 1997, Kestilä et al. 1998). The relatively wide shared haplotype in MUL can be caused by a recent expansion of the ancestral mutation or a small number of chromosomes studied. Additionally, gene-rich regions, such as the *MUL* region, are sometimes less prone to crossing-over events.

number of disease chromosomes	cen	D17S1290	95-CA	272a-CA	272b-CA	132-CA	58-CA	52-CA	D17S1604	tel
n		7	4	4	7	3	1	6	4	
4		7	4	4	7	3	1	2	2	
2		5	4	4	7	3	1	6	4	
3		5	4	4	7	3	1	6	6	
1		5	6	4	7	3	1	6	n	
1		7	4	4	7	3	4	3	n	
1		7	4	4	7	3	5	3	n	
2		6	3	5	4	4	5	6	15	

Table 4. The haplotype analysis of the Finnish MUL patients. In the first row the major Finnish haplotype is defined. The haplotypes restricting the MUL gene region are shown below. The last row indicates the Finnish minor haplotype found in two patients in a compound heterozygote form. n= not determined.

6.4. Meckel syndrome was assigned to the same chromosomal region as *MUL* (II)

Interestingly, the locus for Meckel syndrome (MKS, MIM No. 249000) was assigned to chromosome 17q21-24 (Paavola et al. 1995) in the same genetic region as *MUL*. The phenotype of MKS is different from that of MUL as the disorder leads to death soon after birth. MKS is characterised by central nervous system malformations, polycystic kidneys, cystic and fibrotic changes of the liver, and polydactyly (Salonen 1984), none of which are seen in MUL. MKS also belongs to the disorders of the Finnish disease heritage (Salonen 1984). Globally MKS seems to be heterogeneous with a wide spectrum of phenotypes and locus heterogeneity (Roume et al. 1998).

The critical *MKS* region defined by haplotype analysis overlaps for 600 kb with the 800 kb *MUL* region (data not shown). However, the conserved haplotypes in the region were different in MUL and MKS patients (data not shown) excluding allelic homogeneity. The haplotype analysis leaves the possibility of different mutations in the same gene in MKS and MUL still open. There are several examples of genes responsible for more than one disease (reviewed by Romeo and McKusick 1994).

6.5. Construction of the physical map on 17q22-23 (II and III)

Six novel dinucleotide repeat markers were isolated from PAC (95i19, 52i20, and 58p18) and BAC (272g3 and 132i10) clones of the region. A CHLC marker, D17S1290, was assigned into the genomic clones within this region.

In addition, the detailed physical map enabled us to precisely localise four genes *SFRS1*, *MPO*, *BCL5*, and *RAD51c*; a total of 20 ESTs of both known and unknown homology; and two cDNA clones, KIAA0647 and KIAA0898.

The position of the PAC and BAC clones in relation to each other was also confirmed by the fiber-FISH method. Pulsed-field gel electrophoresis (PFGE) was used to determine the insert sizes of the PAC and BAC clones of the minimum tiling pathway. According to both the PFGE data and fiber-FISH, the size of the contig was estimated to be approximately 1400 kb (see Figure 7).

The *MKS-MUL* region was covered with bacterial clones at least two times and the overlaps were confirmed by both STS-content mapping and fiber-FISH analysis. Recently, a demand for at least five-times coverage has emerged in order to increase the reliability of the contigs. Even though the *MKS-MUL* contig does not meet this criteria, the double checking of overlaps by PCR and fiber-FISH and the high density of STS markers should account for the reliability of the contig.

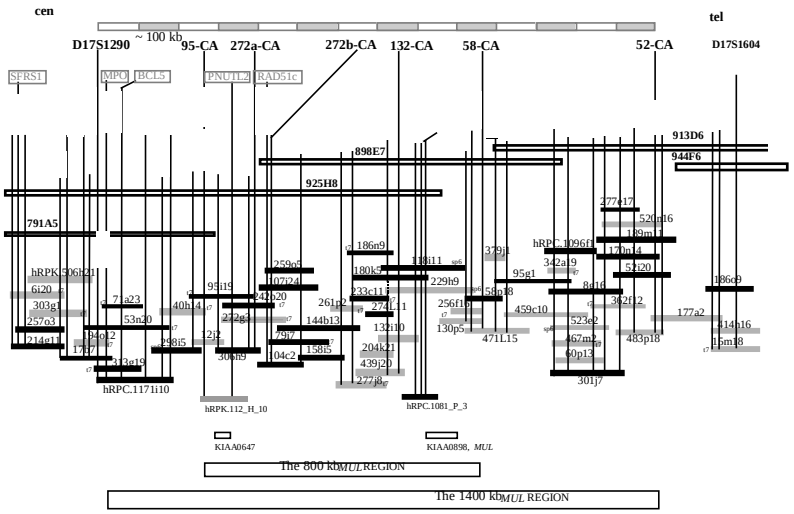

Fig. 7

The physical map on 17q22-23.
Non-chimeric YAC clones mapping to the region are shown as open bars. The PACs are indicated by black bars (214g11 is a P1 clone), and the BACs by gray bars. For the orientation of the clones either sp6 or t7 end of the clones are marked. The polymorphic markers used for the genetic mapping are shown in bold letters on the top, while the ESTs and genes are marked with gray text. Five sequenced bacterial clones are also shown in the map; hRPK.506h21, hRPC.1171i10, and hRPC.1096f1, hRPK.112_H_11, and hRPC.1081_P_3. The recently identified cDNA clones, KIAA0898 and KIAA0647, are shown below as open bars and are at least partly mapping to sequenced clones, hRPC.1081_P_3 and hRPK.112_H_11, respectively. The critical MUL region refined by haplotype analysis is shown below.

6.6. Exclusion and evaluation of candidate genes at 17q (I, II, III, and unpublished)

During the positional cloning effort of the *MUL* gene several positional candidate genes emerged. In some of these the putative role in MUL was not evident, whereas some were considered as good functional candidates. The former group includes the serine-arginine-rich splicing factor gene (*SFRS1*), the myeloperoxidase gene (*MPO*), and the B-cell leukemia/ lymphoma-5 gene (*BCL5*), and the latter the growth hormone (*GH*) gene complex, the *HOXB* gene cluster, the T-box transcription factor gene (*TBX2*), and the peanutlike-2 gene (*PNUTL2*). *SFRS1*, *MPO*, and *BCL5* were localised to the MUL clone contig but resided centromeric to 95-CA that defined the centromeric border of the critical region for *MUL*. Thus they were positionally excluded. The functionally relevant positional candidate genes were evaluated and their exclusion is described here in more detail.

Mutations involving different genes of *GH* complex (George et al. 1981)

have been observed in growth hormone deficiency (Phillips & Cogan 1994). Since one of the key findings in MUL is short stature, and since some of the MUL patients are growth hormone deficient (Lipsanen-Nyman 1986, Lapunzina et al. 1995, Lenko et al. 1982, Haraldsson et al. 1993) the *GH* locus assigned to 17q22-q24 (George et al. 1981) was considered a putative candidate for MUL. A microsatellite repeat polymorphism at the *GH* locus (Polymeropoulos et al. 1991) showed a recombination in family 13 excluding the *GH* locus as the *MUL* gene.

Mammalian homeobox genes (*HOX*) regulate the vertebrate embryogenesis (McGinnis & Krumlauf 1992). As MUL is a developmental disorder the *HOXB* cluster located in 17q21-22 (Solomon & Baker 1989) was considered a putative candidate. A microsatellite repeat marker (Deinard et al. 1992) at the *HOXB* locus (Acampora et al. 1989) did not reveal any recombinations in MUL families, but *HOXB* was excluded as a likely MUL candidate based on the absence of linkage disequilibrium at the marker.

The *TBX2* gene is expressed in mouse in many stages and tissues during the embryonic development (Chapman et al. 1996). *TBX2* was cloned and assigned to 17q23 (Campbell at al. 1995). The fiber-FISH hybridisation of a clone positive for *TBX2* on free DNA fibers showed that *TBX2* is located at least 2 Mb telomeric to the marker D17S1604. Thus *TBX2* could be positionally excluded as the causative gene for MUL.

The *PNUTL2* gene related to proteins involved in cytokinesis and development, was shown to map to the *MUL* region (Paavola et al. 1999). Therefore it was considered a good positional-functional candidate. The genomic sequencing in patients and controls did not reveal mutations in the exons or in the exon-intron boundaries of *PNUTL2* (unpublished data). Furthermore, the Northern blot analysis performed on lymphoblast mRNA from patients did not reveal size differences of the transcript or altered level of expression when compared to that of controls (unpublished data). According to this data *PNUTL2* was excluded as the *MUL* gene.

Eventually, one gene, *RAD51c* (Dosanjh et al. 1998), as well as two cDNA clones KIAA0647 (Ishikawa et al. 1998) and KIAA0898 (Nagase et al. 1998) resided in the *MUL* region. The possible role of *RAD51c* in MUL is difficult to evaluate as little is known of the function of the gene. *RAD51c* is possibly involved in recombinational repair of DNA damage and meiotic recombination.

KIAA0647 was identified from human brain cDNA libraries and is thought to be involved in signal transduction although the wide expression pattern suggests a more general function. The closest protein neighbours of KIAA0647 are myotubularins, suggesting a likely association of KIAA0647 with a muscular disease.

KIAA0898, also identified from human brain cDNA libraries, was categorised as “functionally unclassified” since only weak homologies to known gene products were identified. An analysis of the amino acid sequence of KIAA0898 revealed that the protein belongs to a group of zinc finger proteins, which are known to often act as transcription factors, rendering it to a putative candidate gene for MUL. Later

KIAA0898 was shown to be the *MUL* gene using mutation analysis.

6.7. Analysis of the *MUL* gene (III)

6.7.1. Mutation analysis of *MUL*

KIAA0898 is a 4,111 bp cDNA that includes an open reading frame (ORF) of 2892 bp thereby predicting a polypeptide of 964 amino acids (aa) starting from the first ATG codon with a Kozak consensus sequence (at position 46 in GenBank accession). The ORF of KIAA0898 was analysed in four MUL patients representing five different haplotypes by sequencing overlapping RT-PCR amplified fragments. Four frameshift mutations were identified in the patients indicating that KIAA0898 is the *MUL* gene.

The major Finnish mutation (Fin_{major}) is an A-to-G transition altering the consensus dinuclotide sequence of the 3´ splice site (AG) at position c.493-2 (Table 5). The next AG dinucleotide occurs after three base pairs and acts as the splice acceptor site producing a 5-bp deletion at c.493-497. The deletion leads to a premature STOP at codon 175 and thereby results in a predicted protein of 174 aa. The Fin_{major} cosegregates with the Finnish ancestral MUL haplotype, and is found in 98/100 of the Finnish MUL chromosomes. The carrier frequency of Fin_{major} varies regionally in Finland. In the high-risk region (Savo and North Carelia; Fig. 1), a relatively high carrier frequency of 1:24 was detected (Table 5). This is more than expected by the disease incidence. Possibly, the actual incidence of MUL in the high-risk region is higher than reported. The phenotype of MUL is variable and complex creating difficulties in the diagnosis. Especially, the "milder" MUL patients without cardiac involvement might be undiagnosed. In the rest of the country no carriers of the mutation were found in 99 individuals.

The second *MUL* mutation is a 1-bp deletion of a G at nucleotide c.2212 leading to a premature STOP at codon 768 (Table 5). We identified this mutation (Fin_{minor}) in one chromosome in each of two Finnish patients that are compound heterozygotes for the ancestral MUL haplotype and a different haplotype predicted to represent the "minor" *MUL* mutation in the Finnish population. No carriers of the Finnish minor mutation were detected in 316 Finnish control individuals, which is in agreement with the rare occurrence of the mutation in MUL patients (Table 5).

In a Czech patient homozygous for a "private" haplotype across the entire *MUL* region (data not shown) we identified on both chromosomes a 5-bp deletion of ACTTT at nucleotides c.838-842 (Table 5). No carriers of the Czech mutation were found in 190 CEPH control individuals. The fourth mutation is a 1-bp insertion of an A after nucleotide c.1346 (Table 5). The mutation disrupts the reading frame causing a premature STOP at codon 458. This mutation occurs in a homozygous form in an American patient suggesting that the parents are distantly related even though the disease chromosome haplotypes were not identical (data not shown). No carriers of the American mutation were detected in 200 CEPH control individuals.

Table 5. The MUL mutations.

Mutation	Predicted consequence	Carrier frequency
c.493-2A>G, Fin_{major}	frameshift, truncated 174 aa protein	1:24 (9:217) in the high risk MUL regions in Finland no carriers in 99 controls elsewhere in Finland
c.2212delG, Fin_{minor}	frameshift, truncated 767 aa protein	no carriers in 316 Finnish controls
c.838delACTTT, Czech	frameshift, truncated 334 aa protein	no carriers in 190 CEPH controls
c.1346insA, American	frameshift, truncated 457 aa protein	no carriers in 200 CEPH controls

6.7.2. Nucleic acid sequence analysis of *MUL*

BLAST comparisons revealed no significant homologies between *MUL* and previously known genes at the nucleotide level, whereas at the amino acid level several significant homologies appeared at approximately aa 1-230 including zinc-finger proteins and RING-finger proteins.

Unigene Homo Sapiens database included altogether 43 EST sequences for *MUL* mostly from the 3´end of the cDNA sequence. Two 5´ end ESTs were utilised to identify a STOP codon at position –157 before the first ATG. The EST sequences do not cover the full length cDNA of *MUL*.

Approximately one third of *MUL* cDNA, a 1.1 kb stretch from the 3´ end of the cDNA, maps to the clone hRPC.1081_P_3 that has been sequenced at the Whitehead Institute/MIT Center for Genome Research (Fig. 7). This 3´ end of *MUL* consists of seven exons ranging from 83 to 309 bps on a genomic region of approximately 36 kb.

We identified 15 homologous mouse ESTs and seven rat ESTs that were 84-93% and 84-90%, respectively, identical to the human gene, suggesting the existence of orthologous genes in mouse and rat.

6.7.3. Amino acid sequence analysis of MUL

A detailed analysis of MUL revealed that the protein belongs to a group of RBCC proteins that include a RING finger, one to two B-boxes, and a coiled-coil domain. In MUL the RING finger is located at aa 16-57, the B-box domain at aa 88-134, and the coiled coil region at aa 142-235 (Fig. 8).

In addition, MUL contains a NS domain at aa 232-421 (Fig. 8). This domain, which is also found in *C. elegans*, was identified originally in the N-terminus of the speckle-type POZ protein (SPOP) and appears to be required for localisation of the SPOP protein to nuclear speckles.

Two aspartate-glutamate-serine-rich sequences (DES) were identified between aa 416-550 and 895-919 with weak homologies to several transcriptional regulators (Fig. 8).

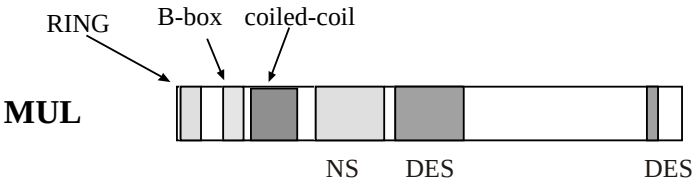

Fig. 8

The MUL protein belongs to the RBCC protein family.
The RING finger is located at aa 16-57, the B-box domain at aa 88-134, and the coiled-coil region at aa 142-235. Additionally a NS domain at 232-421, and two aspartate-glutamate-serine rich sequences (DES) were identified at aa 416-550 and 895-919.

6.7.4. Protein predictions of MUL

The putative structure of the protein encoded by *MUL* was analysed with several computer programmes. The molecular weight predicted from the 964 amino acids is 108 kD. The PSORT II programme predicted the localisation of the MUL protein to the nucleus (65.2%), cytoplasm (26.1%), and cytoskeleton (8.7%). Many of the RBCC proteins have been localised to the nucleus or the cytoplasm (Meroni et al. 1999).

Several potential modification sites, such as four N-glycosylation sites, 15 protein kinase C phosphorylation sites, 23 casein kinase II phosphorylation sites, two tyrosine kinase phosphorylation sites, 13 N-myristolation sites, a leucine-zipper pattern, and a myc-type helix-loop-helix (HLH) dimerization domain signature were identified using the Scanprosite programme.

Hydrophobicity predictions did not reveal strong hydrophobic regions in the MUL polypeptide suggesting that MUL is not likely a transmembrane protein.

6.7.5. Expression analysis of *MUL*

The tissue expression pattern of *MUL* was studied using a 507 bp RT-PCR fragment from the coding region (nucleotides 947-1452) to probe human multiple tissue Northern blots (Fig. 9).

In all tissues studied an RNA transcript of approximately 4.4 kb that corresponds to the full length mRNA was detected. Strong signals of approximately 3.9 kb from testis and 2.6 kb from heart were also detected. Furthermore, most tissues studied showed two signals around 2 kb.

On the other hand, RT-PCR reactions from human brain, heart, lymphoblastoid, placental, and testis RNA did not produce clear evidence of multiple RNA fragments (data not shown). These data suggest the presence of highly homologous genes rather than alternative splicing of *MUL*.

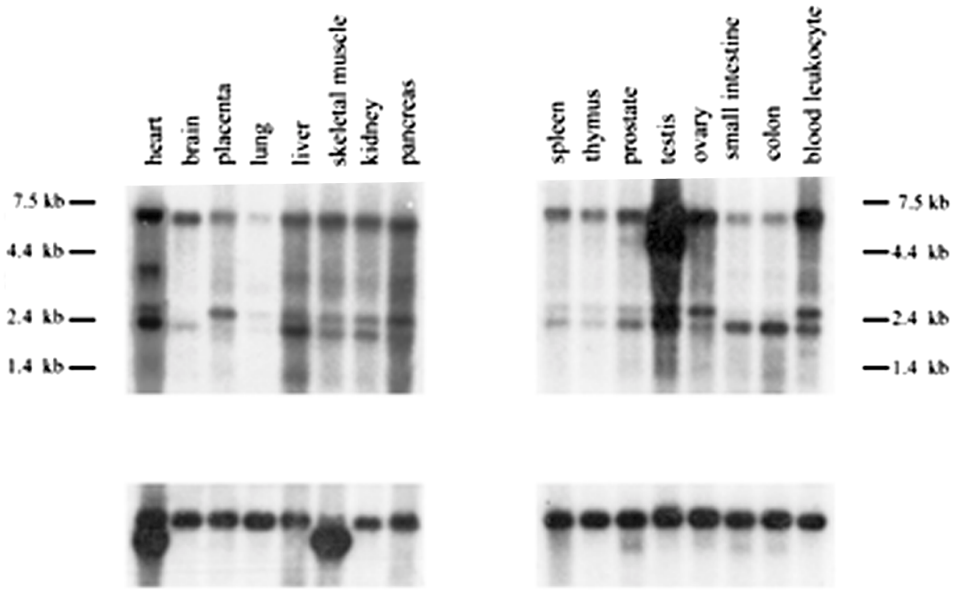

Fig. 9

Human multiple tissue northern blot analysis of *MUL* using a probe from the coding region. In all tissues tested a signal of approximately 4.4 kb is detected. In addition, a strong hybridization signal of 3.9 kb in testis and 2.6 kb in heart is seen. Most tissues also show two signals less than 2 kb in length. The lower panel provides loading controls following hybridization with a β-actin probe.

6.7.6. Evaluation of the consequences of the *MUL* mutations

Truncating mutations, like the ones detected in *MUL*, are described to have two kinds of consequences. On one hand, the mRNAs carrying premature chain termination codons are frequently degraded *in vivo* (Culbertson 1999). Thus, a complete loss of function of *MUL* is possible. On the other hand, a truncated protein lacking some of the normal C-terminal sequence is produced at some occasions (Lehrman et al. 1987). The effect will then depend on the stability of the polypeptide product, the extent of the truncation, and the functional importance of the lacking amino acids.

So far, functional data is available for one mutant RBCC protein, midin. The *MID1* gene encodes midin and is mutated in Opitz syndrome (OS) (Quaderi et al 1997). Midin co-localizes with microtubules, is able to form homodimers mediated by the RBCC motif, and forms large protein complexes within the cell (Cainarca et al. 1999). Both missense and truncating mutations which, except for one, retain the RBCC motif intact have been described in the *MID1* gene (Quaderi et al. 1997, Cainarca et al. 1999). Functional analysis of those midin forms resulting from truncating mutations in which the RBCC domain is intact, suggests only partial loss of *MID1* function as lowered levels of expression and retained ability to form multiprotein complexes and homodimers, but impaired microtubule binding have been identified (Cainarca et al. 1999). Thus, in OS the disease phenotype does not appear to arise directly from loss of function related to the tripartite RBCC motif. Interestingly, three out of four mutations in *MUL* also leave the RBCC motif intact. Whether a similar, partly functional truncated protein is formed in MUL remains to be studied.

7. CONCLUSIONS AND FUTURE PROSPECTS

In this study the gene for Mulibrey Nanism (MUL) was identified by positional cloning. Initially, the *MUL* locus was assigned to chromosome 17q by linkage analysis. Linkage disequilibrium and haplotype analysis were utilised in the subsequent refinement of the *MUL* locus to a 800 kb region. A cDNA clone, KIAA0898, was found out to be a positional and a putative functional candidate for MUL. Mutation analysis of KIAA0898 revealed four independent MUL associated mutations, thus providing strong evidence for KIAA0898 being the *MUL* gene.

In this positional cloning project the power of "cloning in silico" is well demonstrated. As the refined MUL region remained relatively wide, it would have been a major effort to identify the existing genes of the region. This was not necessary since several expressed sequences were already available for mutation analysis in the databases as ESTs, cDNA clones and genes.

Before this study the diagnosis of MUL was established exclusively on the basis of the complex and variable phenotype. The identification of the *MUL* gene facilitates the diagnostics of Mulibrey Nanism as DNA-based testing is now available. Furthermore, families with MUL patients have the access to prenatal diagnosis if they so wish. Genetic counselling of MUL is also improved.

The identification of the *MUL* gene is the first step in unravelling the molecular pathogenesis of MUL. The identification of the *MUL* gene also provides a possibility to develop animal models for MUL, which may eventually lead to more effective treatment including gene therapy. Additionally, it is now possible to study the function of mutated and wild-type MUL proteins in cell. As MUL is a developmental disorder the study of the normal function of the MUL protein will inevitably reveal important facts about the development of human beings.

8. ACKNOWLEDGEMENTS

This thesis work was carried out at the Department of Medical Genetics, Haartman Institute and at the Department of Molecular Genetics, Folkhälsan Institute of Genetics during 1994-2000. Part of this work was carried out in the Centre of Excellence in Disease Genetics of the Academy of Finland (project 44870, Finnish Centre of Excellence Programme (2000-2005)). I wish to express my sincere gratitude to all those who made the work possible, especially to:

The former and present heads of the institutes, Professors Juha Kere, Leena Palotie, Pertti Aula, and Albert de la Chapelle, for providing excellent working facilities.

Anna-Elina Lehesjoki, my supervisor, for never-failing interest in my work, encouragement and kindness. It has truly been a pleasure working under her professional and wise guidance.

Albert de la Chapelle, my other supervisor, for initially recruiting me and introducing the subject of positional cloning. The MUL meetings with professor de la Chapelle have given me inspiration and drive to complete this work.

Jaakko Perheentupa and Marita Lipsanen-Nyman, the clinical experts of the MUL team, for accurately characterising Mulibrey Nanism and thus setting solid grounds for this work, for pleasant collaboration, and for taking care of the contacts with the MUL families. Sinikka Lindh, for all the help in collecting blood samples. The MUL families, for participating in this study.

Anu Wartiovaara and Raimo Voutilainen, for thorough and beneficial pre-examination of this thesis.

My co-authors Carina Wallgren-Pettersson, Sylvie Marchand, Sabine Fauré, for splendid company and valuable help in France, Pertti Sistonen, for competent advice during the genetic mapping of MUL locus.

Paulina Paavola, for enjoyable and effective collaboration while constructing the huge physical map. Co-authors Leena Palotie, Mira Kyttälä, Aarno Palotie, Nina Horelli-Kuitunen, Anne Kallioniemi, and Maarit Bärlund, for their important contributions during the construction of the contig.

Niina Idänheimo, for excellent technical help and positive attitude in the lab. Eva Seemanová and Sally Rosengren, for providing the important non-Finnish MUL samples. Tomi Mäkelä, for his valuable contribution in the protein prediction analysis of the *MUL* gene.

Susanna Ranta, for friendship, and also for useful advice in the lab. Maria Aminoff-Backlund and Kimmo Virtaneva, for sharing their interest in genetics and “a room with a computer” at T2. Laura Huopaniemi, Tarja Joensuu, Bru Cormand, and Riikka Hämäläinen, for scientific and non-scientific discussions as well as for sharing the few O_2 molecules in a cheerful atmosphere in the offices at M97.

Also the rest of my colleagues with whom I've worked several years in the lab, Anne Rantala, Kirsi Alakurtti, Jose Dieguez, Jinmin Miao, Maaret Ridanpää, Katarina Pelin, Juha Kolehmainen, Ann-Liz Träskelin, Tanja Vakkilainen, as well as the new generation, Charlotta Diesen, Kati Donner, Liina Lonka, Riika Salmela, Johanna Tommiska, and Henna Tyynismaa, for pieces of advice; for recipes on molecular genetics as well as on cooking and baking; for amusing

company during lunch hours and coffee breaks; and for off-hour activities.

Aila Riikonen, for help in practical matters.

Monica Virtaneva, for the language revision of this thesis.

My parents, Irja and Arvo Patronen, for love, support and encouraging attitude towards my studies.

My dear husband, Mika Avela, most importantly for love and support, and for bringing huge amounts of joy and happiness into my life, but also for the sincere interest towards my work.

This study was financially supported by the Ulla Hjelt Fund of Finnish Foundation for Pediatric Research, Helsinki, Finland; Rinnekoti Research Foundation, Espoo, Finland; the Maud Kuistila's Foundation, Helsinki, Finland; the Finnish Medical Society Duodecim, Helsinki, Finland; and Helsinki University Central Hospital Research Fund, Helsinki, Finland.

Helsinki, March 2000

Kristiina Avela

9. REFERENCES

Aaltonen J, Horelli-Kuitunen N, Fan J-B, Björses P, Perheentupa J, Mayers R, Palotie A, Peltonen L. High-resolution physical and transcriptional mapping of the Autoimmune polyendocrinopathy–candidiasis–ectodermal-dystrophy locus on chromosome 21q22.3 by FISH. Genome Res 7: 820-829 (1997).

Acampora D, D´Esposito M, Faiella A, Pannese M, Migliaccio E, Morelli F, Stornaiuolo A, Nigro V, Simeone A, Boncinelli E. The human HOX gene family. Nucl Acids Res 17: 10385-10402 (1989).

Aittomäki K, Dieguez-Lucena JL, Pakarinen P, Sistonen P, Tapanainen J, Gromoll J, Kaskikari R, Sankila EM, Lehväslaiho H, Engel AR et al. Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. Cell 22: 959-968 (1995).

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

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402 (1997).

Aminoff M, Carter JE, Chadwick RB, Johnson C, Gräsbeck R, Abdelaal MA, Broch H, Jenner LB, Verroust PJ, Moestrup SK et al. Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. Nature Genet 21: 309-313 (1999).

Ansorge W, Sproat B, Stegemann J, Schwager C, Zenke M. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. Nucleic Acids Res 15: 4593-4602 (1987).

Ashley CT, Warren ST. Trinucleotide repeat expansion and human disease. Annu Rev Genet 29: 703-728 (1995).

Bellini M, Lacroix JC, Gall JG. A zinc-binding domain is required for targeting the maternal nuclear protein PwA33 to lampbrush chromosome loops. J Cell Biol 131: 563-570 (1995).

Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. Predicting coiled coils by use of pairwise residue correlations. Proc Natl Acad Sci USA 92: 8259-8263 (1995).

Bird AP. CpG-rich islands and the function of DNA methylation. Nature 321: 209-213 (1986).

Boehnke M. Estimating the power of a proposed linkage study: a practical computer simulation approach. Am J Hum Genet 39: 513-527 (1986).

Borden KL, Martin SR, O'Reilly NJ, Lally JM, Reddy BA, Etkin LD, Freemont PS. Characterisation of a novel cysteine/histidine-rich metal binding domain from Xenopus nuclear factor XNF7. FEBS Lett 335: 255-260 (1993).

Borden KL, Boddy MN, Lally J, O'Reilly NJ, Martin S, Howe K, Solomon E, Freemont PS. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. EMBO J 14:1532-1541 (1995a).

Borden KL, Lally JM, Martin SR, O'Reilly NJ, Etkin LD, Freemont PS. Novel topology of a zinc-binding domain from a protein involved in regulating early Xenopus development. EMBO J 14: 5947-5956 (1995b).

Borden KL, Freemont PS. The RING finger domain: a recent example of a sequence-structure family. Curr Opin Struct Biol 6: 395-401 (1996).

Borden KL, Lally JM, Martin SR, O'Reilly NJ, Solomon E, Freemont PS. In vivo and in vitro characterization of the B1 and B2 zinc-binding domains from the acute promyelocytic leukemia protooncoprotein PML. Proc Natl Acad Sci USA 93: 1601-1606 (1996).

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

Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331 (1980).

Bowcock A, Osborne-Lawrence S, Barnes R, Chakravarti A, Washington S, Dunn C. Microsatellite polymorphism linkage map of human chromosome 13q. Genomics 15: 376-386 (1993).

Bray-Ward P, Menninger J, Lieman J, Desai N, Mokady N, Banks A, Ward DC. Integration of the cytogenetic, genetic, and physical maps of the human genome by FISH mapping of CEPH YAC clones. Genomics 32: 1-14 (1996).

Breslow NE, Beckwith JR. Epidemiological features of Wilms´ tumor: results of the National Wilms´ tumor study. J Natl Cancer Inst 68: 429-436 (1982).

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88: 4005-4009 (1991).

Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. 1987. Biotechnology 24: 172-178 (1992).

Cainarca S, Messali S, Ballabio A, Meroni G. Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. Hum Mol Genet 8: 1387-1396 (1999).

Cairney AE, Andrews M, Greenberg M, Smith D, Weksberg R. Wilms tumor in three patients with Bloom syndrome. J Pediatr 111: 414-416 (1987).

Campbell C, Goodrich K, Casey G, Beatty B. Cloning and mapping of a human gene (TBX2) sharing a highly conserved protein motif with the Drosophila omb gene. Genomics 28: 255-260 (1995).

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271: 1423-1427 (1996).

Cao T, Borden KL, Freemont PS, Etkin LD. Involvement of the rfp tripartite motif in protein-protein interactions and subcellular distribution. J Cell Sci 110: 1563-1571 (1997).

Cariello NF, Skopek TR. Mutational analysis using denaturing gradient gel electrophoresis and PCR. Mutat Res 288: 103-112 (1993).

Chan EK, Hamel JC, Buyon JP, Tan EM. Molecular definition and sequence motifs of the 52-kD component of human SS-A/Ro autoantigen. J Clin Invest 87: 68-76 (1991).

Chang C-J, Chen Y-L, Lee S-C. Coactivator TIF1• interacts with transcription factor C/EBP• and glucocorticoid receptor to induce •1-acid glycoprotein gene expression. Mol Cell Biol 18: 5880-5887 (1998).

Chapman DL, Garvey N, Hancock S, Alexiou M, Agulnik S, Gibson Brown JJ, Cebra-Thomas J, Bollag RJ, Silver LM, Papaioannou VE. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Dev Dyn 206: 379-390 (1996).

Chauvel PJ, Moore CM, Haslam RHA. Trisomy-18 mosaicism with features of Russel-Silver syndrome. Dev Med Child Neurol 17: 220-224 (1975).

Chu G, Vollrath D, Davis RW. Separation of large DNA molecules by contour-clamped homogenous electric fields. Science 234: 1582-1585 (1986).

Chumakov I, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, Gervy P, LeGall I, Soularue P, Grinas L et al. Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359: 380-387 (1992).

Chumakov IM, Rigault P, Le Gall I, Belanné-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I et al. A YAC contig map of the human genome. Nature 377: 175-297 (1995).

Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. Nature Genet 6: 98-105 (1994).

Church DM, Buckler AJ. Gene identification by exon amplification. Methods Enzymol 303: 83-99 (1999).

Claustres M, Laussel M, Desgeorges M, Giansily M, Culard JF, Razakatsara G, Demaille J. Analysis of the 27 exons and flanking regions of the cystic fibrosis gene: 40 different mutations account for 91.2% of the mutant alleles in southern France. Hum Mol Genet 2: 1209-1213 (1993).

Claverie JM. Computational methods for the identification of genes in vertebrate genomic sequences. Hum Mol Genet 6: 1735-1744 (1997).

Cohen D, Chumakov I, Weissenbach J. A first generation physical map of the human genome. Nature 336: 698-701 (1993).

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

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

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 278: 1580-1581 (1997).

Collins J, Hohn B. Cosmids: a type of plasmid gene cloning vector that is packageable in vitro in bacteriophage heads. Proc Natl Acad Sci USA 75: 4242-4246 (1978).

Cormand B, Avela K, Pihko H, Santavuori P, Talim B, Topaloglu H, de la Chapelle A, 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).

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

Cremers FP, Brunsmann F, Berger W, van Kerkhoff EP, van de Pol TJ, Wieringa B, Pawlowitzki IH, Ropers HH. Cloning of the breakpoints of a deletion associated with choroideremia. Hum Genet 86: 61-64 (1990).

Crick FHC. Acta Crystallog 6: 685-697 (1953).

Cross NC, Tolan DR, Cox TM. Catalytic deficiency of human aldolase B in hereditary fructose intolerance caused by a common missense mutation. Cell 53: 881-885 (1988).

Culbertson MR. RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. Trends Genet 15: 74-80 (1999).

Cumming GR, Kerr D, Ferguson CC. Constrictive pericarditis with dwarfism in two siblings (Mulibrey nanism). J Pediatr 88: 569-572 (1976).

Cuthbert AP, Trott DA, Ekong RM, Jezzard S, England NL, Themis M, Todd CM, 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).

Deinard AS, Ruano G, Kidd KK. A dinucleotide repeat polymorphism at the HOX2B locus. Nucleic Acids Res 20: 1171 (1992).

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, Hästbacka J, Lehesjoki A-E, Sulisalo T, Kere J, Tahvanainen E, Sistonen P. Kytkentä ja kytkentäepätasapaino suomalaisessa tautiperinnössä (in Finnish). Duodecim 110: 654-664 (1994).

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

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380: 152-154 (1996).

Dosanjh MK, Collins DW, Fan W, Lennon GG. Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. Nucleic Acids Res 26: 1179-1184 (1998).

Duncan PA, Hall JG, Shapiro LR, Vibert BK. Three-generation dominant transmission of the Silver-Russell syndrome. Am J Med Genet 35: 245-250 (1990).

Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ et al. The DNA sequence of human chromosome 22. Nature 402: 489-495 (1999).

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

Foote S, Vollrath D, Hilton A, Page DC. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science 258: 60-66 (1992).

Freemont PS. The RING finger. A novel protein sequence motif related to the zinc finger. Ann New York Acad Sci 684: 174-192 (1993).

Freemont PS. RING for destruction? Curr Biol 10: 84-87 (2000).

The French FMF Consortium. A candidate gene for familial Mediterranean fever. Nature Genet 17: 25-31 (1997).

Friend SH, Dryja TP, Weinberg RA. Oncogenes and tumor-suppressing genes. N Engl J Med 318: 618-622 (1988).

Fuleihan DS, Der Kaloustian VM, Najjar SS. The Russell-Silver syndrome: report of three siblings. J Pediatr 78: 654-657 (1971).

Gelehrter TD, Collins FS. Principles of Medical Genetics, 1st publication 1990, Williams & Wilkins, Baltimore, Maryland 21202, USA, ISBN 0-683-03447-2, pp. 202-207.

George DL, Phillips JA, Francke U, Seeburg PH. The genes for growth hormone and chorionic somatomammotropin are on the long arm of human chromosome 17 in region q21-qter. Hum Genet 57: 138-141 (1981).

Gieser EP, Falls HF. Hereditary retinoschisis. Am J Ophthal 51: 1193-1200 (1961).

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

Grompe M, Muzny DM, Caskey CT. Scanning detection of mutations in human ornithine transcarbamoylase by chemical mismatch cleavage. Proc Natl Acad Sci USA 86: 5888-5892 (1989).

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J. The 1993-1994 Généthon human genetic linkage map. Nature Genet 7: 246-339 (1994).

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C et al. A radiation hybrid map of the human genome. Hum Mol Genet 5: 339-346 (1996).

Hanson IM, Trowsdale J. Colinearity of novel genes in the class II regions of the MHC in mouse and human. Immunogenet 34: 5-11 (1991).

Haraldsson A, van der Burgt CJAM, Weemaes CMR, Otten B, Bakkeren JAJM, Stoelinga GBA. Antibody deficiency and isolated growth hormone deficiency in a girl with Mulibrey nanism. Eur J Pediatr 152: 509-512 (1993).

Haravuori H, Mäkelä-Bengs P, Udd B, Partanen J, Pulkkinen L, Somer H, Peltonen L. Assignment of the tibial muscular dystrophy locus to chromosome 2q31. Am J Hum Genet 62: 620-626 (1998).

Harris NL. Genotator: a workbench for sequence annotation. Genome Res 7: 754-762 (1997).

Hartley AL, Birch JM, Tricker K, Wallace SA, Kelsey AM, Harris M, Jones PH. Wilms' tumor in the Li-Fraumeni cancer family syndrome. Cancer Genet Cytogenet 67: 133-135 (1993).

Heiskanen M, Karhu R, Hellsten E, Peltonen L, Kallioniemi O, Palotie A. High resolution mapping using fluorescence in situ hybridization (FISH) to extended DNA fibers prepared from agarose-embedded cells. Biotechniques 17: 928-934 (1994).

Heiskanen M, Hellsten E, Kallioniemi OP, Mäkelä TP, Alitalo K, Peltonen L, Palotie A. Visual mapping by fiber-FISH. Genomics 30: 31-36 (1995).

Heiskanen M, Kallioniemi O-P, Palotie A. Fiber-FISH: experiences and a refined protocol. Genet Anal Biomol Eng 12: 179-184 (1996).

Hemminki A, Tomlinson I, Markie D, Järvinen H, Sistonen P, Björkqvist A-M, Knuutila S, Salovaara R, Bodmer W, Shibata D et al. Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. Nature Genet 15: 87-90 (1997).

Henderson J, Salzberg S, Fasman KH. Finding genes in DNA with a hidden Markov model. J Comput Biol 4: 127-141 (1997).

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH et al. An STS-based map of the human genome. Science 270: 1945-1954 (1995).

Hästbacka J, Kaitila I, Sistonen P, de la Chapelle A. Diastrophic dysplasia gene maps to the distal long arm of chromosome 5. Proc Natl Acad USA 87: 8056-8059 (1990).

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genet 2: 204-211 (1992).

Hästbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Triverdi B, Weaver A et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positonal 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, Kere J. Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. Nature Genet 14: 316-319 (1996).

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

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong P. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet 4: 84-89 (1994).

Ishikawa K, Nagase T, Suyama M, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res 5: 169-176 (1998).

Isono K, McIninch JD, Borodovsky M. Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. DNA Res 1: 263-269 (1994).

Jeffreys AJ, Royle NJ, Wilson V, Wong Z. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332: 278-280 (1988).

Joazeiro CA, Wing SS, Huang J, Leverson JD, Hunter T, Liu YC. The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. Science 286: 309-312 (1999).

Jorde LB. Linkage disequilibrium as a gene mapping tool. Am J Hum Genet 56: 11-14 (1995).

Joyce CA, Sharp A, Walker JM, Bullman H, Temple IK. Duplication of 7p12.1-p13, including GRB10 and IGFBP1, in a mother and daughter with features of Silver-Russell syndrome. Hum Genet 105: 273-280 (1999).

Järvelä I, Sabri Enattah N, Kokkonen J, Varilo T, Savilahti E, 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).

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258: 818-821 (1992).

Kallioniemi OP, Kallioniemi A, Piper J, Isola J, Waldman FM, Gray JW, Pinkel D. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. Genes Chrom Cancer 10: 231-243 (1994).

Kaplan NL, Hill WG, Weir BS. Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56: 18-32 (1995).

Karger AE. Separation of DNA sequencing fragments using an automated capillary electrophoresis instrument. Electrophoresis 17: 144-151 (1996).

Kere J, Estivill X, Chillon M, Morral N, Nunes V, Norio R, Savilahti E, de la Chapelle A. Cystic fibrosis in a low-incidence population: two major mutations in Finland. Hum Genet 93: 162-166 (1994)

Kere J, Srivastava AK, Montonen O, Zonana J, Thomas N, Ferguson B, Munoz F, Morgan D, Clarke A, Baybayan P et al. X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. Nature Genet 13: 409-416 (1996).

Kestilä M, Lenkkeri U, Männikkö M, Lamerdin J, McCready P, Putaala H, Ruotsalainen V, Morita T, Nissinen M, Herva R et al. Positionally cloned gene for a novel glomerular protein-nephrin-is mutated in congenital nephrotic syndrome. Mol Cell 1: 575-582 (1998).

Klockars T, Savukoski M, Isosomppi J, Laan M, Järvelä I, Petrukhin K, Palotie A, Peltonen L. Efficient construction of a physical map by fiber-FISH of the CLN5 region: refined assignment and long-range contig covering the critical region on 13q22. Genomics 35: 71-78 (1996).

Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68: 820-823 (1971).

Knuutila S, Björkqvist AM, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123 (1998).

Knuutila S, Aalto Y, Autio K, Björkqvist AM, El-Rifai W, Hemmer S, Huhta T, Kettunen E, Kiuru-Kuhlefelt S, Larramendy ML et al. DNA copy number losses in human neoplasms. Am J Pathol 155: 683-694 (1999).

Koufos A, Grundy P, Morgan K, Aleck KA, Hadro T, Lampkin BC, Kalbakji A, Cavenee WK. Familial Wiedemann-Beckwith syndrome and a second Wilms tumor locus both map to 11p15.5. Am J Hum Genet 44: 711-719 (1989).

Kruglyak L, Daly M, Lander E. Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am J Hum Genet 56: 519-527 (1995).

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58: 1347-1363 (1996).

Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. Nature Genet 17: 21-24 (1997).

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

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

Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature 386: 847-851 (1997).

Landschulz WH, Johnson PF, McKnight SL. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240: 1759-1764 (1988).

Lapunzina P, Rodriguez JI, de Matteo E, Gracia R, Moreno F. Mulibrey Nanism: three additional patients and a review of 39 patients. Am J Med Genet 55: 349-355 (1995).

Larsen F, Gundersen G, Prydz H. Choice of enzymes for mapping based on CpG islands in the human genome. Genet Anal Tech Appl 9: 80-85 (1992).

Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81: 3443-3446 (1984).

Le Douarin B, Zechel C, Garnier JM, Lutz Y, Tora L, Pierrat P, Heery D, Gronemeyer H, Chambon P, Losson R. The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18. EMBO J 14: 2020-2033 (1995).

Le Douarin B, You J, Nielsen AL, Chambon P, Losson R. TIF1α: a possible link between KRAB zinc finger proteins and nuclear receptors. J Steroid Biochem Molec Biol 65: 43-50 (1998).

Lehesjoki A-E, Koskiniemi M, Norio R, Tirrito S, Sistonen P, Lander E, 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).

Lehrman MA, Schneider WJ, Brown MS, Davis CG, Elhammer A, Russell DW, Goldstein JL. The Lebanese allele at the low density lipoprotein receptor locus. Nonsense mutation produces truncated receptor that is retained in endoplasmic reticulum. J Biol Chem 262: 401-410 (1987).

Lemieux NM, Dutrillaux B, Viegas-Pequiqnot E. A simple method for simultaneous R - or G-banding and fluorescence in situ hybridization of small single-copy genes. Cytogenet Cell Genet 59: 311-312 (1992).

Lenko HL, Leisti S, Perheentupa J. The efficacy of growth hormone in different types of growth failure. An analysis of 101 cases. Eur J Pediatr 138: 241-249 (1982).

Levy E, Haltia M, Fernandez-Madrid I, Koivunen O, Ghiso J, Prelli F, Frangione B. Mutation in gelsolin gene in Finnish hereditary amyloidosis. J Exp Med 1: 1865-1867 (1990).

Lichter P, Cremer T, Chang Tang CJ, Watkins P, Manuelis L, Ward DC. Rapid detection of human chromosome 21 aberration by in situ hybridization. Proc Natl Acad Sci USA 85: 9664-9668 (1988).

Lipsanen-Nyman M. Mulibrey-nanismi. Academic dissertation, the University of Helsinki, Helsinki, Finland, ISBN 951-99732-8-1 (1986).

Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44: 397-401 (1989).

Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S, Weissman AM. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc Natl Acad Sci USA 96: 11364-11369 (1999).

Lovett M. Fishing for complements: finding genes by direct selection. Trends Genet 10: 352-357 (1994).

Lowe CU, May CD, Reed SC. Fibrosis of the pancreas in infants and children: a statistical study of clinical and hereditary features. Am J Dis Child 78: 349-374 (1949).

Lupas A. Coiled coils: new structures and new functions. Trends Biochem Sci 21: 375-382 (1996a).

Lupas A. Prediction and analysis of coiled-coil structures. Methods Enzymol 266: 513-525 (1996b).

Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28: 491-511 (1943).

Maury CP, Alli K, Baumann M. Finnish hereditary amyloidosis. Amino acid sequence homology between the amyloid fibril protein and human plasma gelsoline. FEBS Lett 15: 85-87 (1990).

McGinnis W, Krumlauf R. Homeobox genes and axial patterning. Cell 68: 283-302 (1992).

Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ et al. A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nature Genet 19: 297-300 (1998).

Meretoja J. Genetic aspects of familial amyloidosis with corneal lattice dystrophy and cranial neuropathy. Clin Genet 4: 173-185 (1973).

Meroni G, Reymond A, Fantozzi A, Cairo S, Riganelli D, Merla G, Zanaria E, Messali S, Luzi L, Guffanti A et al. Functional genomics of the B-box gene family reveals a possible role in subcellular compartmentalization. Am J Hum Genet Suppl 65: 420 (1999).

Michaud J, Brody LC, Steel G, Fontaine G, Martin LS, Valle D, Mitchell G. Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta-aminotransferase gene. Genomics 13: 389-394 (1992).

Midro AT, Debek K, Sawicka A, Marcinkiewicz D, Rogowska M. Second observation of Silver-Russel syndrome in a carrier of a reciprocal translocation with one breakpoint at site 17q25. Clin Genet 44: 53-55 (1993).

Milanesi L, Kolchanov N, Rogozin I, Kel A, Titov I. Sequence functional inference. In Bishop, MJ (ed.) Guide to Human genome computing. Academic Press, Cambridge, UK, 1993. pp. 249-312

Mitchell GA, Brody LC, Looney J, Steel G, Suchanek M, Dowling C, Der Kaloustian V, Kaiser-Kupfer M, Valle D. An initiator codon mutation in ornithine-delta-aminotransferase causing gyrate atrophy of the choroid and retina. J Clin Invest 81: 630-633 (1988).

Mitchison HM, O'Rawe AM, Taschner PE, Sandkuijl LA, Santavuori P, de Vos N, Breuning MH, Mole SE, Gardiner RM, Järvelä IE. Batten disease gene, CLN3: linkage disequilibrium mapping in the Finnish population and analysis of European haplotypes. Am J Hum Genet 56: 654-662 (1995).

Monaco AP, Kunkel LM. Cloning of the Duchenne/Becker muscular dystrophy locus. Adv Hum Genet 17: 61-98 (1988).

Monk K, Wakeling EL, Proud V, Hitchins M, Abu-Amero SN, Stanier P, Preece MA, Moore GE. Duplication of 7p11.2-p13, including GRB10, in Silver-Russell syndrome. Am J Hum Genet 66: 36-46 (2000).

Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet 7: 277-318 (1955).

Moser HE, Dervan PB. Sequence-specific cleavage of double helical DNA by triple helix formation. Science 238: 645-650 (1987).

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51: 263-273 (1986).

Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, Sheffield VC, Sunden S, Duyk GM et al. A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). Science 265: 2049-2054 (1994).

Myllärniemi S, Koski K, Perheentupa J. Craniofacial and dental study of Mulibrey Nanism. Cleft Palate J 15: 369-377 (1978).

Mäkelä-Bengs P, Järvinen N, Vuopala K, Suomalainen A, Ignatius J, Sipilä M, Herva R, Palotie A, Peltonen L. Assignment of the disease locus for 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 et al. Positional cloning of the APECED gene. Nature Genet 17: 393-8 (1997).

Nagase T, Ishikawa K, Suyama M, Kikuno R, Hirosawa M, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 5: 355-364 (1998).

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 27: 1616-1622 (1987).

Narod SA. Genetics of breast and ovarian cancer. Br Med Bull 50: 656-676 (1994).

Neri G, Martini-Neri ME, Katz BE, Opitz JM. The Perlman syndrome: familial renal dysplasia with Wilms tumor, fetal gigantism and multiple congenital anomalies. Am J Med Genet 19: 195-207 (1984).

Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, 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).

Nikali K, Isosomppi J, Lönnqvist T, Mao JI, Suomalainen A, Peltonen L. Towards cloning of a novel ataxia gene: refined assignment and physical map of the IOSCA locus (SCA8) on 10q24. Genomics 39: 185-191 (1997).

Norio R, Nevanlinna HR, Perheentupa J. Hereditary disease in Finland; Rare flora in rare soil. Annals of Clinical Research 5: 109-141 (1973).

Norio R. Diseases of Finland and Scandinavia; Biocultural Aspects of Diseases (ed. Rothschild HD). Academic Press, New York, 1981. pp. 369-415.

O´Connell JR, Weeks DE. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. Nature Genet 11: 402-408 (1995).

Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879 (1989).

Ott J. Computer-simulation methods in human linkage ananlysis. Proc Natl Acad Sci USA 86: 4175-4178 (1989).

Ott J. Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore and London, 1991.

Ott J. Linkage analysis with biological markers. Hum Hered 45: 169-174 (1995).

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

Paavola P. Molecular Genetics of Meckel syndrome. Academic Dissertation, the University of Helsinki, ISBN 951-740-103-5 (1998).

Paavola P, Horelli-Kuitunen N, Palotie A, Peltonen L. Characterization of a novel gene, *PNUTL2*, on human chromosome 17q22-23, and its exclusion as the Meckel syndrome gene. Genomics 55: 122-125 (1999).

Paloneva J, Kestilä M, Wu J, Salminen A, Böhling T, Ruotsalainen V, Hakola P, Bakker ABH, Phillips JH, Pekkarinen P et al. Loss of function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. Nature Genet 25: 357-361 (2000).

Pálsson B, Pálsson F, Perlin M, Gudbjartsson H, Stefánsson K, Gulcher J. Using quality measures to facilitate allele calling in high-throughput genotyping. Genome Res 9: 1002-1012 (1999).

Pauling L, Corey RB. Nature 171: 59-61 (1953).

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

Peltonen L. Positional cloning of disease genes: advantages of genetic isolates. Hum Hered 50: 66-75 (1999).

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

Pennacchio LA, Lehesjoki A-E, Stone N, Willour VL, Virtaneva K, Miao J, D´Amato E, Ramirez L, Faham M, Koskiniemi M et al. Mutations in the Gene Encoding Cystatin B in Progressive Myoclonus Epilepsy (EPM1). Science 271: 1731-1734 (1996).

Perheentupa J, Autio S, Leisti S, Raitta C. Mulibrey Nanism: dwarfism with muscle, liver, brain and eye involvement. Acta Paediatr Scand 59 Suppl: 74-75 (1970).

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

Perlin MW, Lancia G, Ng SK. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. Am J Hum Genet 57: 1199-1210 (1995).

Petersen GM, Rotter JI, Cantor RM, Field LL, Greenwald S, Lim JST, Roy C, Schoenfeld V, Lowden JA, Kaback M. The Tay-Sachs disease gene in North American Jewish populations: geographic variations and origin. Am J Hum Genet 35: 1258-1269 (1983).

Phillips JA, Cogan JD. Genetic basis of endocrine disease 6: molecular basis of familial human growth hormone deficiency. J Clin Endocr Metab 78:11-16 (1994).

Pierce JC, Sauer B, Sternberg N. A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. Proc Natl Acad Sci USA 89: 2056-2060 (1992).

Pilia G, Hughes-Benzie RM, MacKenzie A, Baybayan P, Chen EY, Huber R, Neri G, Cao A, Forabosco A, Schlessinger D. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. Nature Genet 12: 241-247 (1996).

Pinkel D, Straume T, Gray J. Cytogenetic analysis using quantitative, high sensitivity, fluorescence hybridization. Proc Natl Acad Sci USA 83: 2934-2938 (1986).

Polymeropoulos MH, Rath DS, Xiao H, Merril CR. A simple sequence repeat polymorphism at the human growth hormone locus. Nucl Acids Res 19: 689 (1991).

Pykett MJ, Murphy M, Harnish PR, George DL. The neurofibromatosis 2 (NF2) tumor suppressor gene encodes multiple alternatively spliced transcripts. Hum Mol Genet 3: 559-564 (1994).

Quaderi NA, Schweiger S, Gaudenz K, Franco B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S et al. Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. Nature Genet 17: 285-291 (1997).

Ramirez-Duenas ML, Medina C, Ocampo-Campos R, Rivera H. Severe Silver-Russell syndrome and translocation (17;20). Clin Genet 41: 51-53 (1992).

Rannala B, Slatkin M. Likelihood analysis of disequilibrium mapping, and related problems. Am J Hum Genet 62: 459-473 (1998).

Ranta S, Zhang Y, Ross B, Lonka L, Takkunen E, Messer A, Sharp J, Wheeler R, Kusumi K, Mole S et al. The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. Nature Genet 23: 233-236 (1999).

Reddy BA, Etkin LD, Freemont PS. A novel zinc finger coiled-coil domain in a family of nuclear proteins. Trends Biochem Sci 17: 344-345 (1992).

Robson KJ, Chandra T, MacGillivray RT, Woo SL. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. Proc Natl Acad Sci USA 79: 4701-4705 (1982).

Rogan PK, Seip JR, Driscoll DJ, Papenhausen PR, Johnson VP, Raskin S, Woodward AL, Butler MG. Distinct 15q genotypes in Russell-Silver and ring 15 syndromes. Am J Med Genet 62: 10-15 (1996).

Romeo G, McKusick VA. Phenotypic diversity, allelic series and modifier genes. Nature Genet 7: 451-453 (1994).

Roume J, Genin E, Cormier-Daire V, Ma HW, Mehaye B, Attie T, Razavi-Encha F, Fallet-Bianco C, Buenerd A, Clerget-Darpoux F et al. A gene for Meckel syndrome maps to chromosome 11q13. Am J Hum Genet 63: 1095-1101 (1998).

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T et al. Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5: 42-52 (1995).

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93: 12035-12039 (1996).

Salonen R. The Meckel syndrome: clinicopathological findings in 67 patients. Am J Med Genet 18: 671-689 (1984).

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

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning. A Laboratory Manual. 2nd edition. Cold Spring Harbour, 1989.

Sankila EM, Pakarinen L, Kääriäinen H, Aittomäki K, Karjalainen S, Sistonen P, de la Chapelle A. Assignment of an Usher syndrome type III (USH3) gene to chromosome 3q. Hum Mol Genet 4: 93-98 (1995).

Sarkar G, Yoon HS, Sommer SS. Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. Nucleic Acids Res 20: 871-878 (1992).

Sauer CG, Gehrig A, Warneke-Wittstock R, Marquardt A, Ewing CC, Gibson A, Lorenz B, Jurklies B, Weber BH. Positional cloning of the gene associated with X-linked juvenile retinoschisis. Nature Genet 17: 164-170 (1997).

Saurin AJ, Borden KLB, Boddy MN, Freemont PS. Does this have a familiar RING? Trends Biochem Sci 21: 208-213 (1996).

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

Schinzel AA, Robinson WP, Binkert F, Fanconi A. An interstitial deletion of proximal 8q (q11-q13) in a girl with Silver-Russell syndrome-like features. Clin Dysmorphol 3: 63-69 (1994).

Schlessinger D. Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. Trends Genet 245: 248-258 (1990).

Schuler GD, Boguski MS, Stewart EA, Stein LD, Guapay G, Rice K, White RE, Rodriquez-Tome P, Aggarwal A, Bajorek E et al. A gene map of the human genome. Science 274: 540-546 (1996).

Schwartz DT, Cantor CR. Separation of yeast chromosome-sized DNA:s by pulsed field gradient gel electrophoresis. Cell 37: 67-75 (1984).

Schweiger S, Foerster J, Lehmann T, Suckow V, Muller YA, Walter G, Davies T, Porter H, van Bokhoven H, Lunt PW et al. The Opitz syndrome gene product, MID1, associates with microtubules. Proc Natl Acad Sci USA 96: 2794-2799 (1999).

Seemanová E, Bartch O. Mulibrey Nanism and Wilms´ tumor. Am J Med Genet 85: 76-78 (1999).

Shapira TM, Sitney JA. Choroideremia. Am J Ophthal 26: 182-183 (1943).

Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, Simon M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escheria coli using an F-factor-based vector. Proc Natl Acad Sci USA 89: 8794-8797 (1992).

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 16: 325-332 (1993).

Similä S, Timonen M, Heikkinen E. A case of Mulibrey Nanism with associated Wilms´ tumor. Clin Genetics 17: 29-30 (1980).

Smith SA, Easton DF, Evans DG, Ponder BA. Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome. Nature Genet 2: 128-131 (1992).

Solomon E, Baker D. Report of the committee on the genetic constitution of chromosome 17. Cytogenet Cell Genet 51: 319-337 (1989).

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503-517 (1975).

Srivastava AK, Montonen O, Saarialho-Kere U, Chen E, Baybayan P, Pispa J, Limon J, Schlessinger D, Kere J. Fine mapping of the EDA gene: a translocation breakpoint is associated with a CpG island that is transcribed. Am J Hum Genet 58:126-132 (1996).

Stay EJ, Vawter G. The relationship between nephroblastoma and neurofibromatosis (Von Recklinghausen's disease). Cancer 39: 2550-2555 (1977).

Sternberg N. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as100 kilobase pairs. Proc Natl Acad Sci USA 87:103-107 (1990).

Strachan T, Read AP. Human Molecular Genetics, 1st publication, 1996, A John Wiley & Sons, Inc., Publication, ISBN 0-471-13373-6, p. 197, p. 347

Strathdee CA, Gavish H, Shannon WR, Buchwald M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. Nature 356: 763-767 (1992).

Sulisalo T, Sistonen P, Hästbacka J, Wadelius C, Mäkitie O, de la Chapelle A, Kaitila I. Cartilage-hair hypoplasia gene assigned to chromosome 9 by linkage analysis. Nature Genet 3: 338-341 (1993).

Szabo J, Heath B, Hill VM, Jackson CE, Zarbo RJ, Mallette LE, Chew SL, Besser GM, Thakker RV, Huff V et al. Hereditary hyperparathyroidism-jaw tumor syndrome: the endocrine tumor gene HRPT2 maps to chromosome 1q21-q31. Am J Hum Genet 56: 944-950 (1995).

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

Takahashi M, Inaguma Y, Hiai H, Hirose F. Developmentally regulated expression of a human "finger"-containing gene encoded by the 5' half of the ret transforming gene. Mol Cell Biol 8: 1853-1856 (1988).

Tanner JM, Lejarraga H, Cameron N. The natural history of the Silver-Russell syndrome: a longitudinal study of thirty-nine cases. Pediatr Res 9: 611-623 (1975).

Tenhunen K, Laan M, Manninen T, Palotie A, Peltonen L, Jalanko A. Molecular cloning, chromosomal localization and expression of the mouse aspartylglucosaminidase gene. Genomics 30: 244-250 (1995).

Terwilliger JD. 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).

Thomas A, Skolnick MH. A probabilistic model for detecting coding regions in DNA sequences. IMA J Math Appl Med Biol 11: 149-160 (1994).

Thorén C. So-called Mulibrey Nanism with pericardial constriction. Lancet 2: 731 (1973).

Torrents D, Mykkänen J, Pineda M, Feliubadalo L, Estevez R, de Cid R, Sanjurjo P, Zorzano A, Nunes V, Huoponen K et al. Identification of SLC7A7, encoding y+LAT-1, as the lysinuric protein intolerance gene. Nature Genet 21: 293-296 (1999).

Tuuteri L, Perheentupa J, Rapola J. The cardiopathy of Mulibrey Nanism, a new inherited syndrome. Chest 65: 628-631 (1974).

Udd B, Rapola J, Nokelainen P, Arikawa E, Somer H. Nonvacuolar myopathy in a large family with both late adult onset distal myopathy and severe proximal muscular dystrophy. J Neurol Sci 113: 214-221 (1992).

Varilo T. The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. Academic Dissertation, the University of Helsinki, ISBN 951-740-151-5 (1999).

Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. Shotgun sequencing of the human genome. Science 280: 1540-1542 (1998).

Verheijen FW, Verbeek E, Aula N, Beerens CE, Havelaar AC, Joosse M, Peltonen L, Aula P, Galjaard H, van Der Spek PJ et al. A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. Nature Genet 23: 462-465 (1999).

Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, Hofmann SL, Peltonen L. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. Nature 376: 584-588 (1995).

Vignal A, Gyapay G, Hazan J, Nguyen S, Dupraz C, Cheron N, Becuwe N, Tranchant M, Weissenbach J. Nonradioactive Multiplex Procedure for Genotyping of Microsatellite Markers. In: Methods in Molecular Genetics, vol. 1, Academic Press Inc, 1993. pp 211-221.

Virtaneva K, Miao J, Traskelin AL, Stone N, Warrington JA, Weissenbach J, Myers RM, Cox DR, Sistonen P, de la Chapelle A et al. Progressive myoclonus epilepsy EPM1 locus maps to a 175-kb interval in distal 21q. Am J Hum Genet 58: 1247-1253 (1996).

Virtaneva K, D'Amato E, Miao J, Koskiniemi M, Norio R, Avanzini G, Franceschetti S, Michelucci R, Tassinari CA, Omer S et al. Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. Nature Genet 15: 393-396 (1997).

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

Visapää I, Salonen R, Varilo T, Paavola P, 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).

Vogelstein B, Fearon ER, Kern SE, Hamilton SR, Preisinger AC, Nakamura Y, White R. Allelotype of colorectal carcinomas. Science 244: 207-211 (1989).

Voorhess ML, Husson GS, Blackman MS. Growth failure with pericardial constriction: The syndrome of Mulibrey Nanism. Am J Dis Child 130: 1146-1148 (1976).

Warren ST. The expanding world of trinucleotide repeats. Science 271: 1374-1375 (1996).

Waterman H, Levkowitz G, Alroy I, Yarden Y. The RING finger of c-Cbl mediates desensitization of the epidermal growth factor receptor. J Biol Chem 274: 22151-22154 (1999).

Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using polymerase chain reaction. Am J Hum Genet 44: 388-396 (1989).

Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. Am J Genet Suppl 47: A204 (1980).

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millaseau P, Vaysseix G. Lathrop M. A second-generation linkage map of the human genome. Nature 359: 794-801 (1992).

Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, Yang MC, Hwang LY, Bowcock AM, Baer R. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nature Genet 14: 430-440 (1996).

Xiong M, Guo S-W. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet 60: 1513-1531 (1997).

Yokouchi M, Kondo T, Houghton A, Bartkiewicz M, Horne WC, Zhang H, Yoshimura A, Baron R. Ligand-induced ubiquitination of the epidermal growth factor receptor involves the interaction of the c-Cbl RING finger and UbcH7. J Biol Chem 274: 31707-31712 (1999).

Zhang MQ. Identification of protein coding regions in the human genome by quadratic discriminant analysis. Proc Natl Acad Sci USA 94: 565-568 (1997).

Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, Cuthbert AP, Newbold RF, Wang J, Chevrette M et al. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. Nature Genet 20: 337-343 (1998).