Helsinki University Biomedical Dissertations No. 102

The molecular basis of Marinesco-Sjögren syndrome

Anna-Kaisa Anttonen

Department of Medical Genetics
Folkhälsan Institute of Genetics
and
Neuroscience Center

University of Helsinki

ACADEMIC DISSERTATION

To be publicly discussed, with the permission of
the Medical Faculty of the University of Helsinki,
in the Seth Wichmann lecture hall of Women's Hospital,
Haartmaninkatu 2, Helsinki,
on February 8th 2008 at 12 noon.

Helsinki 2008

Supervised by

Professor Anna-Elina Lehesjoki, M.D., Ph.D.
Folkhälsan Institute of Genetics and
Neuroscience Center
University of Helsinki
Helsinki, Finland

Reviewed by

Docent Tiina Paunio, M.D., Ph.D.
Department of Molecular Medicine
National Public Health Institute and
Department of Psychiatry
Helsinki University Central Hospital
Helsinki, Finland

Docent Helena Pihko, M.D., Ph.D.
Department of Pediatric Neurology
Hospital for Children and Adolescents
Helsinki University Central Hospital
Helsinki, Finland

Official opponent

Professor Graeme Black, MB ChB DPhil
University Department of Medical Genetics
St Mary's Hospital and
Centre for Molecular Medicine
University of Manchester
Manchester, UK

ISSN 1457-8433
ISBN 978-952-10-4510-3 (paperback)
ISBN 978-952-10-4511-0 (PDF)
http://ethesis.helsinki.fi
Helsinki University Printing House
Helsinki 2008

To Mikko, Emil, and Eljas

TABLE OF CONTENTS

1 LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to by their Roman numerals in the text. In addition some unpublished data are presented.

I) Mahjneh I, Anttonen AK, Somer M, Paetau A, Lehesjoki AE, Somer H, Udd B. Myopathy is a prominent feature in Marinesco-Sjogren syndrome – A muscle computed tomography study. *J Neurol.* 2006 Mar;253(3):301-6. Epub 2005 Sep 15.

II) Anttonen AK, Mahjneh I, Hämäläinen RH, Lagier-Tourenne C, Kopra O, Waris L, Anttonen M, Joensuu T, Kalimo H, Paetau A, Tranebjaerg L, Chaigne D, Koenig M, Eeg-Olofsson O, Udd B, Somer M, Somer H, Lehesjoki AE. The gene disrupted in Marinesco-Sjögren syndrome encodes SIL1, an HSPA5 cochaperone. *Nat Genet.* 2005 Dec;37(12):1309-11. Epub 2005 Nov 13.

III) Anttonen AK, Siintola E, Tranebjaerg L, Iwata NK, Bijlsma EK, Meguro H, Ichikawa Y, Goto J, Kopra O, Lehesjoki AE. Novel *SIL1* mutations and exclusion of functional candidate genes in Marinesco-Sjögren syndrome. *Eur J Hum Genet.* 2008 *in press*

2 ABBREVIATIONS

ATP	adenosine triphosphate
bp	base pair
c.	coding DNA reference sequence position
CCFDN	congenital cataracts facial dysmorphism and neuropathy
cDNA	complementary DNA
CEPH	Centre d'Etude du Polymorphisme Humain
CK	creatine kinase
CMRD	chylomicron retention disease
CNS	central nervous system
COS-1	african green monkey kidney cells
CT	computed tomography
del	deletion
DNA	deoxyribonucleic acid
dup	duplication
E	embryonic day
ER	endoplasmic reticulum
ERAD	ER-associated degradation
EST	expressed sequence tag
fs	frameshift
GO	the Gene Ontology (database)
HA	hemagglutinin
HSP	heat shock protein
kb	kilobase
lod	logarithm of the odds
Mb	megabase
MRI	magnetic resonance imaging
mRNA	messenger RNA
MSS	Marinesco-Sjögren syndrome
NCBI	National Center for Biotechnology Information
nt	nucleotide
OMIM	online Mendelian inheritance in man
ORF	open reading frame
P	postnatal day
p.	protein reference sequence position
PCR	polymerase chain reaction
PTC	premature termination codon
RNA	ribonucleic acid
RT-PCR	reverse transcriptase PCR
SD	standard deviation
UCSC	University of California Santa Cruz
UPR	unfolded protein response
UTR	untranslated region
wz	woozy mouse mutant

3 ABSTRACT

Marinesco-Sjögren syndrome (MSS) is a rare autosomal recessive neurodegenerative disorder characterized by cerebellar ataxia due to cerebellar cortical atrophy, infantile- or childhood-onset bilateral cataracts, progressive myopathy, and mild to severe mental retardation. Additional features include hypergonadotropic hypogonadism, various skeletal abnormalities, short stature, and strabismus. The neuroradiologic hallmarks are hypoplasia of both the vermis and cerebellar hemispheres. The histopathologic findings include severe cerebellar atrophy and loss of Purkinje and granule cells. The common pathologic findings in muscle biopsy are variation in muscle fiber size, atrophic fibers, fatty replacement, and rimmed vacuole formation. The presence of marked cerebellar atrophy with myopathy distinguishes MSS from another rare syndrome, the congenital cataracts, facial dysmorphism, and neuropathy syndrome (CCFDN).

Previously, work by others had resulted in the identification of an *MSS* locus on chromosome 5q31. A subtype of MSS with myoglobinuria and neuropathy had been linked to the *CCFDN* locus on chromosome 18qter, at which mutations in the *CTDP1* gene had been identified. We confirmed linkage to the previously identified locus on chromosome 5q31 in two Finnish families with eight affected individuals, reduced the critical region by fine-mapping, and identified *SIL1* as a gene underlying MSS. We found a common homozygous founder mutation in all Finnish patients. The same mutation was also present in patient samples from Norway and Sweden. Altogether, we identified eight mutations in *SIL1*, including nonsense, frameshift, splice site alterations, and one missense mutation. *SIL1* encodes a nucleotide exchange factor for the endoplasmic reticulum (ER) resident heat-shock protein 70 chaperone GRP78. GRP78 functions in protein synthesis and quality control of the newly synthesized polypeptides. It senses and responds to stressful cellular conditions.

We showed that in mice, SIL1 and GRP78 show highly similar spatial and temporal tissue expression in developing and mature brain, eye, and muscle. Studying endogenous proteins in mouse primary hippocampal neurons, we found that SIL1 and GRP78 colocalize and that SIL1 localizes to the ER. We studied the subcellular localization of two mutant proteins, a missense mutant found in two patients and an artificial mutant lacking the ER retrieval signal, and found that both mutant proteins formed aggregates within the ER. Well in line with our findings and the clinical features of MSS, recent work by Zhao *et al.* showed that a truncation of SIL1 causes ataxia and cerebellar Purkinje cell loss in the naturally occurring woozy mutant mouse. Prior to Purkinje cell degeneration, the unfolded protein response is initiated and abnormal protein accumulations are present. MSS thus joins the group of protein misfolding and accumulation diseases. These findings highlight the importance of *SIL1* and the role of the ER in neuronal function and survival.

The results presented in this thesis provide tools for the molecular genetic diagnostics of MSS and give a basis for future studies on the molecular pathogenesis of MSS. Understanding the mechanisms behind this pleiotropic syndrome may provide insights into more common forms of ataxia, myopathy, and neurodegeneration.

4 INTRODUCTION

Cells are the building blocks of human life, and almost all human cells have a nucleus which contains the genetic material. The 46 human chromosomes and mitochondrial DNA comprise the genome, which combines the genetic information from previous generations and makes up the blueprint for a human being. Cells are complex factories of life, and all their processes have to meet certain criteria for the entity to function properly. A cell needs quality control mechanisms to identify and prevent mistakes during its essential activities: the chromosomal DNA, the transcribed mRNA as well as the newly synthesized proteins are subject to continuous proofreading.

Approximately 6330 Mendelian monogenic syndromes have been characterized in man (9th January, 2008; www.ncbi.nlm.nih.gov/Omim/mimstats.html). In a little over half of them the molecular basis, the defective gene, is known. Each syndrome has a unique combination of phenotypic features, and each syndrome differs from another by only one or few of those features. When comparing the number of human protein coding genes, currently estimated to be 25000, to the number of genes implicated in disease, it has been argued that the majority of single gene defects that give rise to disease have not been identified yet. Although in recent years there have been many advances in genetics finding the right diagnosis for a patient with a rare disorder is still demanding, as is finding a major genetic risk factor for a common disease.

In 1931 the Romanian neurologist Georges Marinesco published with his associates a novel syndrome in four individuals from a single family (Marinesco *et al.* 1931). Later, a Swedish professor Torsten Sjögren saw four similar cases and identified 10 additional patients belonging to three families, carried out thorough genetic statistical analysis and concluded that the disorder is hereditary and transmitted in an autosomal recessive manner (Sjögren 1950). The syndrome with cerebellar ataxia, congenital cataracts, delayed motor development and varying degrees of mental retardation is thus called Marinesco–Sjögren syndrome (MSS), even though it was probably first described in 1904 by Ernest Emil Moravscik (Alter *et al.* 1968; Superneau *et al.* 1985). Unfortunately, Moravscik chose to write to a Hungarian journal, *Orvosi Hetilap*, which may have restricted his audience (Brogdon *et al.* 1996). MSS is a rare disorder: although it was characterized over a century ago, there are only 200 published cases world wide, including the four patients from Finland that were published by Riitta Herva and her colleagues in 1987.

Marinesco-Sjögren syndrome has been one of the syndromes described in *Mendelian Inheritance in Man* since the book's first edition (McKusick 1966). Today, the online version of the book, OMIM (Online Mendelian Inheritance in Man; www.ncbi.nlm.nih.gov/Omim), also links the syndrome with its defective gene *SIL1*, encoding a nucleotide exchange factor.

5 REVIEW OF THE LITERATURE

5.1 Autosomal recessive cerebellar ataxias

Cerebellar ataxias are a heterogeneous group of neurodegenerative disorders. The hereditary ataxias are all characterized by motor incoordination due to cerebellar dysfunction and the group consists of a large number of autosomal dominant ataxias and ataxias with autosomal recessive, X-linked, or mitochondrial inheritance. Different types of hereditary ataxia have overlap in their clinical presentation and it may be difficult to distinguish between them even when the family history is known (Opa *et al.* 2002). The dominant forms share certain characteristics: the onset is in adulthood, anticipation (i.e. the tendency of a disease to become more severe or to start at earlier ages in successive generations) may be present, which is explained by the fact that many forms are due to repeat expansion mutations. Autosomal recessive ataxias are usually early onset presenting before the age of 20 years (Harding 1983). A comparison of selected autosomal recessive cerebellar ataxia types is presented in Table 1.

Friedreich's ataxia (FRDA, MIM 229300) is world wide the most common form of autosomal recessive ataxia, but in Finland its incidence is much lower (Juvonen *et al.* 2002). FRDA is clinically characterized by slowly progressive ataxia, dysarthria, absent reflexes in lower limbs, sensory loss, and pyramidal signs. Skeletal deformities, like scoliosis and high arch foot, and cardiomyopathy are found in the majority of the patients, and they have increased frequency of impaired glucose tolerance and diabetes. In a typical case, symptoms begin in late childhood or adolescence and the death typically occurs in the fourth decade. Some patients have atypical presentation with onset after 25 years, retained tendon reflexes, and absence of extensor plantar response (Durr *et al.* 1996). 98% of patient alleles have a GAA repeat expansion in the first intron of the frataxin (*FXN*) gene (Campuzano *et al.* 1996) encoding a mitochondrial anti-apoptotic protein, which prevents mitochondrial damage and reactive oxygen species production and is involved in iron homeostasis (Condo *et al.* 2006). Point mutations in *FXN* are rare, and depending on the mutation, they can cause indistinguishable or atypical phenotypes when combined with the repeat expansion (Delatycki *et al.* 2000).

Ataxia-telangiectasia (AT, MIM 208900) is the second most frequent autosomal recessive ataxia. The hallmark features of AT include progressive cerebellar ataxia with onset typically before three years of age, telangiectases in skin and conjunctiva of the eye, immune defects, and predisposition to malignancy (Perlman *et al.* 2003). Oculomotor apraxia is frequently present. Magnetic resonance imaging (MRI) shows cerebellar atrophy and serum α-fetoprotein levels are usually elevated. AT results from mutations in the ataxia telangiectasia mutated (*ATM*) gene (Savitsky *et al.* 1995) encoding a serine/threonine protein kinase (Canman *et al.* 1998). The large size of *ATM* with no mutation hotspots makes genetic diagnostics challenging.

Ataxia with isolated vitamin E deficiency (AVED; MIM 183090) is caused by mutations in the α-tocopherol transfer protein gene, *α-TTP*, leading to impaired incorporation of α-tocopherol into lipoproteins and inefficient recycling of plasma vitamin E (Ouahchi *et al.* 1995). The age of onset is commonly around

Table 1. Characteristics of selected autosomal recessive cerebellar ataxias. Modified from (Breedveld *et al.* 2004).

Disease	Locus	Gene *	Protein **	Age at onset (years)	Distinguishing features	OMIM number	References
Ataxia-telangiectasia (AT)	11q22.3	*ATM*	Serine-protein kinase ATM	0 - 20	Telangiectasia, immune deficiency, cancer, elevated AFP	208900	(Savitsky *et al.* 1995)
Ataxia with vitamin E deficiency (AVED)	8q13.1-q13.3	*TTPA*	Alpha-tocopherol transfer protein TTPA	2 - 52 (<20)	As FRDA, but cardiomyopathy and diabetes are rare; head titubation	183090	(Ouahchi *et al.* 1995)
Friedreich's ataxia (FRDA)	9q13	*FXN*	Frataxin, mitochondrial FRDA	4 - 40	Cardiomyopathy, diabetes	229300	(Campuzano *et al.* 1996)
Infantile onset spinocerebellar ataxia (IOSCA)	10q24	*C10orf2*	Twinkle protein, mitochondrial PEO1	~ 1	Oftalmoplegia, hypacusis, athetosis, peripheral neuropathy, epilepsy	271245	(Nikali *et al.* 2005)
Marinesco-Sjögren syndrome (MSS)	5q31	*SIL1*	Nucleotide exchange factor SIL1	Childhood	Cataract, myopathy, mental retardation, hypergonadotropic hypogonadism	248800	Study II
Mitochondrial recessive ataxia syndrome (MIRAS)	15q24	*POLG*	DNA polymerase subunit gamma-1 DPOG1	5 - 50	Myoclonus, epilepsy	-	(Hakonen *et al.* 2005)

AFP: α-fetoprotein
* Gene abbreviations are according to the HUGO Gene Nomenclature Committee (HGNC) recommendations (www.genenames.org).
** Protein names and abbreviations are according to the UniProt database (beta.uniprot.org).

10 years, and in its classical form AVED is very similar to FRDA. The absence of cardiomyopathy and diabetes in AVED distinguishes it from FRDA. Diagnosis is based on low levels of vitamin E in serum. The condition is treatable with vitamin E supplementation (Koenig 2003).

Some autosomal recessive ataxias have specific geographical distributions. In Finland, the most relevant ones are infantile-onset spinocerebellar ataxia (IOSCA, MIM 271245), mitochondrial recessive ataxia syndrome (MIRAS), and progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1, MIM 254800). The onset of IOSCA is between 9 and 18 months of age and the characteristic features include atrophy of the cerebellum, brain stem and spinal cord, ataxia, athetosis, loss of deep tendon reflexes, and sensory axonal neuropathy (Koskinen *et al.* 1994; Nikali *et al.* 2005). Other symptoms include ophthalmoplegia, hearing loss, optic atrophy, and female primary hypergonadotropic hypogonadism. Recessive mutations in the *C10orf2* gene (chromosome 10 open reading frame 2; also called *PEO1*) encoding a mitochondrial helicase Twinkle underlie IOSCA, whereas dominant mutations are associated with progressive external ophthalmoplegia (Spelbrink *et al.* 2001). MIRAS is a recently identified ataxia syndrome (Rantamäki *et al.* 2001) which has been associated with mutations in the polymerase γ (*POLG*) gene involved in the replication and repair of mitochondrial DNA (Van Goethem *et al.* 2004; Hakonen *et al.* 2005). MIRAS manifests at variable age of onset (5-41 years) with heterogenous symptoms: ataxia, peripheral neuropathy, dysarthria, mild cognitive impairment, involuntary movements, psychiatric symptoms, and epileptic seizures. In EPM1 ataxia is a cardinal feature, but it is usually present at later stages of the disease with the main clinical presentation of stimulus-sensitive myoclonus and generalized tonic-clonic seizures (Lehesjoki *et al.* 2004).

5.2 Marinesco-Sjögren syndrome

5.2.1 Clinical manifestations

Marinesco-Sjögren syndrome (MSS, MIM 248800) is an autosomal recessive disorder with marked phenotypic variability (Williams *et al.* 1996; Lagier-Tourenne *et al.* 2003; Slavotinek *et al.* 2005). The syndrome is a panethnic disease with 200 published cases (Orphanet Reports Series 2007), but exact prevalence figures are not available. Clinically, the patients described in the literature form a heterogeneous entity, and can be divided into classical MSS and MSS-like groups. Typically, classical MSS is characterized by cerebellar atrophy with ataxia, early-onset cataracts, mild to severe mental retardation, and myopathy. Additional features include hypergonadotropic hypogonadism, short stature, various skeletal abnormalities, and strabismus (Williams *et al.* 1996; Lagier-Tourenne *et al.* 2002; Slavotinek *et al.* 2005). The common clinical features of MSS are summarized in Table 2.

The cerebellar signs dominate the central nervous system (CNS) symptoms in MSS (Williams *et al.* 1996; Lagier-Tourenne *et al.* 2002; Slavotinek *et al.* 2005). Delayed motor milestones are among the first clinical signs. Gait or truncal ataxia is one of the most frequent symptoms. Dysarthria, nystagmus, and intention tremor have been noted in many cases. In classical MSS neuroimaging studies such as MRI and computed tomography (CT) show cerebellar atrophy, which is usually more pronounced in the vermis than in the hemispheres (Katafuchi *et al.* 1990; McLaughlin *et al.* 1996; Georgy *et al.* 1998; Harting *et al.* 2004). The cerebellar fissures are widened and the fourth ventricle

Table 2. Clinical features of MSS.

	(Slavotinek *et al.* 2005) 75 cases	(Williams *et al.* 1996) 125 cases
Cerebellar signs		
Truncal / limb ataxia	72%	>98% cerebellar dysfunction
Dysarthria	45%	
Nystagmus	36%	
Intention tremor	16%	
Ophthalmological features		
Cataracts	91%	98%
Strabismus	44%	
Other neurological signs		
Truncal / limb hypotonia	53%	89% hypotonia, muscle weakness or atrophy
Muscle atrophy	51%	
Muscle weakness	36%	
Ortopedic manifestations		
Pes planus / planovalgus	28%	46% skeletal anomalies
Genu valgum	20%	
Scoliosis	19%	
Kyphoscoliosis	16%	
Joint contractures	17%	
Pectus deformities	9%	

Adapted from a table by Slavotinek *et al* (Slavotinek *et al.* 2005), where following cases were included in the review: (Alter *et al.* 1962; Skre *et al.* 1977; Hakamada *et al.* 1981; Alexianu *et al.* 1983; Walker *et al.* 1985; Herva *et al.* 1987; Superneau *et al.* 1987; Sewry *et al.* 1988; Komiyama *et al.* 1989; Bromberg *et al.* 1990; Katafuchi *et al.* 1990; Tachi *et al.* 1991; Torbergsen *et al.* 1991a; Kodama *et al.* 1992; Zimmer *et al.* 1992; Ishikawa *et al.* 1993; Brogdon *et al.* 1996; McLaughlin *et al.* 1996; Williams *et al.* 1996; Farah *et al.* 1997; Muller-Felber *et al.* 1998; Aguglia *et al.* 2000; Reinker *et al.* 2002).

is enlarged (Katafuchi *et al.* 1990; Harting *et al.* 2004). Atrophy of the cerebral cortex and the brainstem have been reported (Bromberg *et al.* 1990; Katafuchi *et al.* 1990), as well as atrophy of the pons and medulla oblongata (Sasaki *et al.* 1996). A T2-hyperintense cerebellar cortex has recently been reported in three patients with MSS, who also presented reduced N-acetylaspartate and elevated myo-inositol metabolite ratios in proton magnetic spectroscopy of the cerebellum (Harting *et al.* 2004). One patient with MSS showed decreased glucose metabolic rate in the thalamus with positron emission tomography (Bromberg *et al.* 1990). Four patients had experienced epileptic seizures (Goto *et al.* 1990; McLaughlin *et al.* 1996). MSS patients have variable degrees of developmental delay, and the slow progression of neurological deterioration led to the search for a metabolic defect (McLaughlin *et al.* 1996). Lysosomal inclusions in cultured fibroblasts have been observed in one report (Walker *et al.* 1985), but the finding has not been confirmed (Herva *et al.* 1987).

The neuromuscular manifestations described so far show considerable variation. Hypotonia, slowly progressive muscle weakness, and atrophy suggesting chronic myopathy have been the most frequent findings (Andersen 1965; Alter *et al.* 1968; Herva *et al.* 1987; Superneau *et al.* 1987). Electromyography (EMG) typically shows only myopathic features (Chaco 1969; Komiyama

et al. 1989; Torbergsen *et al.* 1991b). Demyelinating polyneuropathy (Zimmer *et al.* 1992), axonal polyneuropathy (Skre *et al.* 1976) or both (Alexianu *et al.* 1983; Müller-Felber *et al.* 1998) have also been described. Serum creatine kinase (CK) concentrations are normal (Superneau *et al.* 1987; Torbergsen *et al.* 1991a) or moderately increased, usually 2–4 fold the upper normal limits.

The bilateral cataracts have been stated as congenital, but there are several reports on rapid postnatal development of lens opacities (Herva *et al.* 1987; Ishikawa *et al.* 1993; McLaughlin *et al.* 1996; Farah *et al.* 1997; Slavotinek *et al.* 2005). In only a few cases has the type of cataract been stated. In addition to variable single reports, posterior capsular opacities have been reported (Ishikawa *et al.* 1993; Farah *et al.* 1997). The eyes are typically operated on in the first decade of life. Strabismus is a frequent finding and optic atrophy was present in three cases (Dotti *et al.* 1993).

Radiography of the bones may help to determine the extent of skeletal involvement. In the literature, the most distinguishable skeletal features are shortening of metacarpals and metatarsals, shortening of phalanges, and deformation of the sternum. In addition, small posterior cranial fossa, scoliosis and kyphoscoliosis, scalloped vertebral bodies, gracile bones, pes planovalgus, and valgus deformities of elbows and hips are present (Brogdon *et al.* 1996; Reinker *et al.* 2002).

Although atypical findings like optic atrophy, seizures, hearing loss, and peripheral neuropathy have been reported, it is unknown if these are rare manifestations of MSS or features of a distinct disorder (Williams *et al.* 1996; Lagier-Tourenne *et al.* 2003; Slavotinek *et al.* 2005).

5.2.2 Natural history

MSS patients are usually born after uncomplicated pregnancies. Muscular hypotonia is typically noted in early infancy. The development of a child sometimes deteriorates after febrile illness (Slavotinek *et al.* 2005). Muscular weakness is already noticed during the first decade of life in the distal muscles of the extremities, and some patients are never able to walk without assistance (Alter *et al.* 1968; Herva *et al.* 1987; Superneau *et al.* 1987). Later, patients show truncal ataxia, dysdiadochokinesia, and dysarthria. The motor functions of the patients worsen progressively, followed by stabilization of the functional state at varying ages and at variable degrees of severity.

Patients show delayed developmental milestones and their mental performance varies usually from mild to moderate mental retardation (Lagier-Tourenne *et al.* 2003). Although many of the patients are severely handicapped in their adulthood, the life span of patients with MSS is near to normal.

5.2.3 Histopathologic findings

Under light microscopy variation in muscle fiber size, atrophic fibers, fat accumulation, and rimmed vacuole formation are observed (Herva *et al.* 1987; Sewry *et al.* 1988; Suzuki *et al.* 1997). Rimmed vacuoles were most evident in samples taken at older ages (Herva *et al.* 1987). In electron microscopy, autophagic vacuoles with myeloid bodies beneath the sarcolemma or near the nucleus, and a perinuclear dense double-membrane structure suggested to be specific to MSS, are seen (Herva *et al.* 1987; Sewry *et al.* 1988; Sasaki *et al.* 1996).

Sural nerve biopsy has been performed on four MSS patients. The findings have been

presence of segmental demyelination (Hakamada *et al.* 1981; Alexianu *et al.* 1983; Farah *et al.* 1997), and features of chronic axonal degeneration (Farah *et al.* 1997), which was proposed to be secondary to the segmental demyelination (Alexianu *et al.* 1983). One case was reported to have normal findings in sural nerve biopsy (Mahloudji *et al.* 1972).

Autopsy findings in three patients have shown severe cerebellar atrophy with loss of Purkinje and granule cells (Todorov 1965; Mahloudji *et al.* 1972; Skre *et al.* 1977). The remaining Purkinje cells were mostly vacuolated or binucleated. The brain lesions were almost exclusively limited to the cerebellum, although in one case cell loss and atrophy were present in pons and medulla (Mahloudji *et al.* 1972). The cerebellar Purkinje and granule cell loss, however, does not extend to the IX and X lobules (nodulus, flocculus, and paraflocculus ventralis) (Todorov 1965).

There is only one report of histological findings in ovary and testis (Bassoe 1956). The ovarian tissue showed fibrous changes and only a few scarred follicles were observed. Examination of the testis showed large aggregations of Leydig cells, scanty testicular canals with few Sertoli cells, and most of the tissue was replaced by a hyaline substance.

5.2.4 Differential diagnosis

Other syndromes presenting with cerebellar atrophy and ataxia (see above) should be considered as differential diagnoses. This is especially true in a case of an affected child without family history, when all MSS features, e.g. cataract and hypergonadotropic hypogonadism, may not be present (McLaughlin *et al.* 1996). In MSS, the degree of mental retardation may vary as well as the myopathic features.

The congenital cataracts, facial dysmorphism, and neuropathy syndrome (CCFDN, MIM 604168) shares with MSS the key features of ataxia, cataracts, psychomotor delay and hypogonadism (Tournev *et al.* 1999). The presence of more severe mental retardation, marked cerebellar atrophy leading to severe ataxia, myopathy with specific features on muscle biopsy, and absence of peripheral neuropathy and microcornea distinguishes MSS from CCFDN (Lagier-Tourenne *et al.* 2002).

Some recent reports have described disorders clinically similar to, but genetically different from, MSS. An X-linked syndrome, cataract, ataxia, short stature, and mental retardation (CASM, MIM 300619), affects males while female carriers present only cataracts (Guo *et al.* 2006). A Marinesco-Sjögren-like syndrome was reported in two siblings sharing most of the clinical features found in MSS but without marked cerebellar atrophy (Schulz *et al.* 2007). As additional symptoms the patients presented external ophthalmoplegia and dysphagia.

There are some other rare, mainly autosomal recessive disorders that resemble MSS. The clinical entities include the Cataract-ataxia-deafness-retardation syndrome (MIM 212710), which differs from MSS by the presence of sensorineural hearing loss and polyneuropathy (Begeer *et al.* 1991), and familial Danish dementia (FDD, MIM 117300), a dominant disorder where cataracts and ataxia occur later in life than in MSS, and where the patients have dementia or psychosis (Vidal *et al.* 2000). VLDLR-associated cerebellar hypoplasia (Dysequilibrium syndrome; MIM 224050) found in Hutterites shares many features with MSS including cerebellar atrophy, cataracts, psychomotor delay, short stature, strabismus,

and hypotonia, but lacks progressive myopathy (Boycott *et al.* 2005). In aniridia, cerebellar ataxia, and mental deficiency (Gillespie syndrome; MIM 206700) two main MSS features, cataracts and myopathy, are not present and the patients have aniridia (Nelson *et al.* 1997). Ataxia-microcephaly-cataract syndrome (MIM 208870) has been reported in only one family, in which ataxia-telangiectasia was also present (Ziv *et al.* 1992). In addition, several other syndromes share some of the main clinical features with MSS, but they are distinguishable due to additional features. For example, in pseudohypoparathyroidism (PHP, MIM 103580) mental retardation, cataracts, short stature, and short metacarpals are present, but the patients have neither ataxia nor myopathy (Brogdon *et al.* 1996).

5.2.5 Molecular genetics

Two loci have been linked to MSS. Lagier-Tourenne et al. (Lagier-Tourenne *et al.* 2003) used a homozygosity mapping strategy in two large consanguineous families of Turkish and Norwegian origin with the strict clinical features of MSS. They localized the *MSS* locus to a 9.3 cM interval on chromosome 5q31 between markers *D5S1995* and *D5S436*.

A second locus on chromosome 18qter was identified in three Roma families with the phenotype of MSS with myoglobinuria and demyelinating neuropathy (Merlini *et al.* 2002). This locus was first described in families with CCFDN syndrome (Angelicheva *et al.* 1999), which is one of the differential diagnoses for MSS. When a homozygous mutation leading to aberrant splicing of the *CTDP1* gene was shown to underlie CCFDN, a panel of 14 non-Roma patients with MSS was screened and no mutations were found (Varon *et al.* 2003). It is not clear whether the original reports of a subtype of MSS with myoglobinuria and neuropathy (Müller-Felber *et al.* 1998; Merlini *et al.* 2002) actually described patients with CCFDN rather than MSS.

5.3 The Human Genome Project

The Human Genome Project (HGP), which achieved its main goal of sequencing >98% of the gene-containing part of the human genome in April 2003, has changed the field of genetics completely. This historic project started in 1990 as an international effort, which aimed at identifying all human genes, determining the complete sequence of the human genome, and making the data accessible for further biological studies. One of the first milestones of the HGP was the construction of genetic and physical maps (Collins *et al.* 2003). The finished euchromatic sequence of the human genome is estimated to contain 20000-25000 genes, a surprisingly low number (The International Human Genome Sequencing Consortium 2004). A task still remains in sequencing the 20% of the human genome that lies within heterochromatin (Stein 2004).

Many have stated that the complete sequence of the 3 billion bases in the human genome is just the end of the beginning. As part of the HGP, parallel studies were carried out on selected model organisms, such as fruit fly (*Drosophila melanogaster*) and house mouse (*Mus musculus*), to help to develop the technology and interpret human gene function. Comparative genomics and defining the function of the known genes are major challenges in the near future. New large-scale projects that have been launched include the HapMap project, which has just completed the second phase haplotype map cataloguing common patterns of human DNA sequence variation (The International HapMap Consortium 2003; 2005; Frazer *et al.* 2007),

the ENCODE (Encyclopedia of DNA elements) project (The ENCODE Consortium 2004; Birney *et al.* 2007; Weinstock 2007), in which the goal is to identify the functional elements within the genome, and the Knockout Mouse Project aiming to generate a public resource of mouse embryonic stem cells containing a null mutation in every gene in the mouse genome (Austin *et al.* 2004).

5.4 Strategies in identifying human disease genes

During the past 25 years, disease gene identification has changed enormously with the help of new techniques. For each project, the "method of choice" is chosen based on such factors as the family material available, mode of inheritance, and the background information on the defective pathway. In the early years of disease gene identification, functional cloning, which is based on the knowledge of the basic biochemical defect in a particular disease, and the candidate gene approach utilizing partial functional information about the disease were the only methods available (Collins 1995). Later, pure positional cloning, i.e. cloning a gene identifying first its chromosomal location following the laborious identification of the genes in the region, was widely used. Currently, disease genes are mainly identified using the positional candidate gene approach relying on a combination of mapping the disease to the right chromosomal region and searching interesting candidates within this interval. Now that the human sequence is available from databases, the major constraint is to collect family material suitable for Mendelian disease gene identification. Identifying susceptibility genes for common complex diseases remains, however, much more difficult.

5.4.1 Functional cloning and other position-independent approaches

Prior to the availability of comprehensive genetic maps, disease genes were identified through knowledge of the defective protein product. This method, referred as functional cloning, required information about the basic molecular pathogenesis of the disease in the absence of information about the chromosomal position of the gene involved (Collins 1992). After determining the peptide sequence of the protein involved, the cDNA sequence was translated, oligonucleotide probes were synthesized and the corresponding gene isolated from a cDNA library. Functional cloning was successfully used in identifying genes such as the coagulation factor VIII (*F8*) gene in hemophilia A and the phenylalanine hydroxylase (*PAH*) gene in phenylketonuria (Gitschier *et al.* 1984; Kwok *et al.* 1985). Nevertheless, the method was difficult to extend to majority of Mendelian diseases because in most heritable traits there is no pre-existing information about the disease mechanism (Collins 1992).

Identifying disease genes through knowing the protein product is not the only position independent approach. Well-characterized animal models can be helpful. If a defective mouse gene is cloned, the human homolog becomes a natural candidate, as shown by the example of Waardenburg-Shah syndrome with pigmentary abnormalities, sensorineural deafness, and Hirschsprung disease. The defect in the *Sox10* (sex determining region Y-box 10) gene was first identified in mice with aganglionic megacolon, which is reminiscent of Hirschsprung disease (Herbarth *et al.* 1998). Later, human patients with Waardenburg-Shah syndrome were shown to have mutations in *SOX10* (Pingault *et al.* 1998). More frequently, the knowledge of a mouse mutant is combined with that of its human counterpart after mapping the disease

locus, which was the case in Waardenburg syndrome type I (Tassabehji *et al.* 1992). The human locus on chromosome 2q was compared to the corresponding region on mouse chromosome 1 and the disease-causing mutations were identified in the paired box 3 (*PAX3*) gene after showing that *PAX3* was deleted in mice with a similar phenotype. Comparing the mRNA expression pattern of patient and control samples in array experiments may also lead to disease gene identification if the defective gene has reduced expression. A recent example is Arts syndrome with mental retardation, early-onset hypotonia, ataxia, delayed motor development, hearing impairment, and optic atrophy. Arts syndrome is linked to Xq22.1-q24 and oligonucleotide microarray expression profiling revealed reduced expression levels of the phosphoribosyl pyrophosphate synthetase 1 (*PRPS1*) gene, in which patient mutations were shown (de Brouwer *et al.* 2007).

5.4.2 *Position-dependent gene identification*

Positional cloning, first introduced in 1986 (Royer-Pokora *et al.* 1986), allowed the isolation of disease genes without direct information about the basic functional defect. Instead, the identification of a defective gene is based on its chromosomal map position. In positional cloning the defective gene is first assigned to a chromosomal position using linkage analysis. The candidate interval is narrowed using additional markers and finally the disease gene is identified among the genes in the region (Collins 1992). Among the first examples are the identification of the cytochrome b beta subunit (*CYBB*) gene in chronic granulomatous disease (Royer-Pokora *et al.* 1986) and the retinoblastoma 1 (*RB1*) gene (Friend *et al.* 1986). Pure positional cloning is practically not used anymore and has been replaced by approaches utilizing *in silico* information (Collins 1995). As the human sequence and almost all of the transcripts within a region are now available, positional candidate gene analysis and positional functional candidate gene analysis combining the positional cloning, candidate gene, and functional candidate gene analysis methods are at present the most effective approaches (Figure 1).

5.4.3 *Linkage analysis – defining the gene location*

The success of a positional cloning project relies greatly on the size of the candidate region, which is usually identified with the help of linkage analysis. First, families where the disease of interest is segregating are collected and the phenotype is evaluated carefully. Families with two or more patients are suitable for linkage analysis, but the size of the families and number of families needed vary depending on the mode of inheritance and analysis method used. The likelihood of obtaining positive linkage can be predicted utilizing statistical analysis.

If there is no known locus for the disease, the affected and healthy family members are genotyped with multiple polymorphic markers covering the genome evenly. The goal is to identify one or more markers that segregate together with the disease within each family. Traditionally, microsatellite markers, short polymorphic tandem DNA repeats, are used in genome-wide approaches (Weissenbach *et al.* 1992; Ellegren 2004). If the family material is large enough, 350-400 informative microsatellites covering the genome evenly are enough for detecting linkage. Single nucleotide polymorphisms (SNPs) are nowadays used for genome-wide scans, too. Although individually SNPs are less informative than microsatellites, the higher

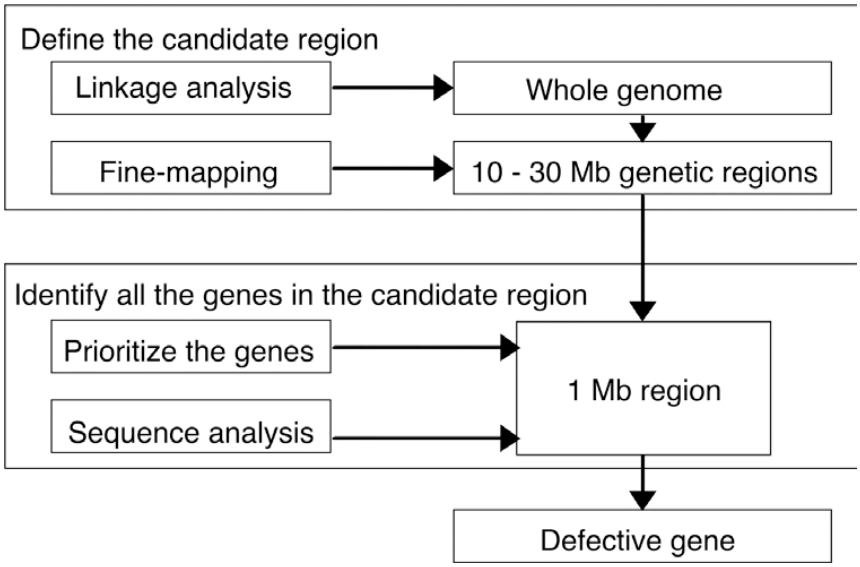

Figure 1. Disease gene identification through positional candidate gene analysis. The first step is to assign the disease locus to a chromosomal region using linkage analysis. Next, the candidate region is narrowed further by fine-mapping. Positional candidate genes are prioritized and evaluated for mutations cosegregating with the disease phenotype.

density of SNPs gives finer resolution and array-based SNP genotyping techniques are more high-throughput than genotyping microsatellite markers. A genome-wide scan using 6000-500000 SNPs may also require less fine-mapping.

Homozygosity mapping is a method utilizing a consanguineous family structure to map an autosomal recessive disease. The term autozygosity denotes homozygosity for markers identical by descent, which means inherited from a defined common ancestor. The application was first introduced by Lander and Botstein (Lander *et al.* 1987), and has been used in mapping the genes for diseases that are more common in population isolates (Peltonen *et al.* 1999). In Finland, the population structure due to national and regional isolation has lead to unique assortment of hereditary diseases, denoted as the Finnish Disease Heritage (FDH) (Norio 2003a). In the majority of FDH diseases one founder mutation accounts for almost all disease alleles thus allowing the exploitation of homozygosity mapping (Norio 2003b).

Linkage analysis is based on the fact that two closely related loci on the same chromosome are inherited together more often than would be expected by chance. The recombination fraction (θ) indicates the probability that a recombination, that is a crossing-over event during meiosis, is observed between the two loci. Two loci are genetically linked when θ is less than 0.5 (or 50%). The genetic distance between two loci is measured in centiMorgans (cM). If two loci are 1 cM apart from each other, there is a 1% chance of recombination between these loci as the chromosome is passed to the next generation. In other words, 100 meioses are required to observe one recombination. The genetic distance is not linear compared to physical distance, but in general 1 cM corresponds to 1 Mb. The recombination fractions differ between females and males, and in most mammals female genetic maps are longer than those for males (Kong *et al.* 2002; Coop *et al.* 2007). Certain areas in the genome are more prone to recombination than others and it has been estimated that 80% of crossing over events take place in 10-20% of genomic sequence, in so called

recombination hotspots (Coop *et al.* 2007). Despite the highly identical DNA sequence between human and chimpanzee, the observed locations of recombination hotspots do not overlap between these two species suggesting that recombination patterns must evolve very rapidly (Winckler *et al.* 2005).

A statistical method is used to evaluate the results obtained in a linkage study. Calculation of the lod score (Z; logarithm of odds) represents the most efficient statistical test for proving that the result is not simply due to chance (Morton 1955). The lod score gives a logarithm based likelihood ratio of a pedigree on two alternative hypotheses: is linkage with a recombination fraction of θ more likely or is it more likely that the two loci are not linked (θ = 0.5) (Terwilliger *et al.* 1994)? In a set of families the overall lod score is obtained by simply adding up individual lod scores for each family. The lod score calculation can be represented as a formula:

$$Z = \log 10 \frac{(1-\theta)^n \theta^r}{0.5^{n+r}}$$

n = number of non-recombinant offspring
r = number of recombinant offspring

For a disease with autosomal inheritance lod scores of 3.0 or higher are taken as evidence of linkage with a 5% chance of error (p = 0.05), and linkage can be rejected if lod score is < -2.0. Lod scores between -2 and +3 do not give conclusive results. In parametric, model based linkage analysis requiring detailed information on the mode of inheritance, gene frequencies, and penetrance, computer programs such as MLINK (Lathrop *et al.* 1984) and Genehunter (Kruglyak *et al.* 1996) are widely used. MLINK is part of the LINKAGE program package, which analyzes limited numbers of marker loci and is not suitable for multipoint analyses. Genehunter allows more effective multipoint analyses, but pedigrees need to be of moderate size (Nyholt 2002). However, model based analysis is usually not optimal for complex, nonmendelian diseases, in which nonparametric linkage analysis may be the method of choice.

Although straightforward, a positional cloning project may also have some pitfalls. It requires suitable family material, and is sensitive to misdiagnosis. Locus heterogeneity within individual families can be avoided using large families in autosomal dominant disorders and homozygosity mapping in autosomal recessive diseases, if possible. Microsatellite markers distributed across the genome at intervals of 10 cM may fail to identify a disease locus, but this problem could be overcome by using a denser set of SNP markers (Chiang *et al.* 2006).

5.4.4 Chromosomal abnormalities

In some cases it is possible to localize the disease gene with the help of associated cytogenetic abnormalities. For severe dominant diseases, in which the genetic cause is *de novo*, chromosome aberrations are especially helpful while linkage analysis is not applicable (Breuning *et al.* 1993). In balanced translocations or inversions, the breakpoints may directly pinpoint the defective gene. For larger, cytogenetically visible deletions and microdeletions detected with high-resolution comparative genomic hybridization analysis, the whole deleted area rather than a specific breakpoint becomes the candidate region and focus of the search. Both balanced translocations and visible deletions were utilized in identifying the *DMD* gene underlying Duchenne and Becker muscular dystrophies (Monaco *et al.* 1986; Bodrug *et al.* 1987). A more recent example is identification of the *DYX1C1* gene underlying

susceptibility to dyslexia (Taipale *et al.* 2003).

5.4.5 Ancestral haplotypes and linkage disequilibrium

After the initial assignment using a genome wide scan, the candidate locus is fine-mapped with as many informative polymorphic markers as possible. Microsatellite markers or SNPs that cover the region at high density are selected from the databases. The aim of this analysis is to identify the key recombinations and reduce the size of the candidate region.

After exploiting all recombinants, the candidate region may still be large. If a founder mutation is probable due to population history, the search for a shared ancestral haplotype in the families with no known relationship may be a reasonable approach. Linkage disequilibrium (LD) refers to co-occurrence of certain alleles, at loci that are close to each other, more often than it would be expected by chance. The use of is based on the hypothesis that the patients have inherited the mutation or susceptibility allele from a common ancestor. As a result, the alleles of the neighboring loci on the same chromosome are also identical and form a specific ancestral haplotype. LD mapping has been successfully used in Finland (Hästbacka *et al.* 1992; Lehesjoki *et al.* 1993) and is a valuable tool for fine-mapping (de la Chapelle *et al.* 1998).

5.4.6 Gene identification

Before the complete human genome sequence became available, the isolation and examination of all the genes contained in the candidate locus was a major effort. Now, the entire human sequence is available in electronic format except for some gaps in difficult areas (The International Human Genome Sequencing Consortium 2004) and the identification of genes in the region of interest is much easier. There are three widely used genome browsers available on-line: the University of California, Santa Cruz (UCSC) Human Genome Browser Gateway (genome.ucsc.edu/cgi-bin/hgGateway), the European Bioinformatics Institute's Ensembl (www.ensembl.org), and the National Center for Biotechnology Information (NCBI) MapViewer (www.ncbi.nlm.nih.gov/mapview). Genome browsers make the sequence data easy to access and for example a list of annotated genes in a given region can be obtained within seconds (Wolfsberg *et al.* 2002).

The genes in the candidate region are evaluated and prioritized with respect to their known or predicted function and tissue expression profiles before picking candidate genes for sequencing. Appropriate expression patterns include the tissues that are affected, but need not be restricted to those tissues. Databases contain several different types of information about gene expression including microarray expression data and *in situ* hybridization data for the mouse brain (Allen Brain Atlas; www.brain-map.org). The function of a gene may be uncovered by comparing Gene Ontology terms (GO; www.geneontology.org) such as biological process, molecular function and cellular component, as well as the transmembrane domains, and predicted functional relationships.

5.4.7 Mutation screening

The final step in identifying a disease gene is to detect disease-associated mutations in patient samples. There are several mutation detection methods with varying sensitivities and costs including direct sequencing, Southern blotting, as well as PCR-based

mutation scanning methods such as single-strand conformational polymorphism analysis. Sequencing is nowadays used as a primary method because of its decreased costs. However, sequencing of only the coding regions of a gene is not an optimal method if the mutation lies in an intronic region. Alternative methods, such as MLPA (multiplex ligation-dependent probe amplification), quantitative PCR, or Southern blotting, are also required to detect larger chromosomal aberration e.g. deletions, insertions or duplications.

Distinguishing between a mutation that changes the phenotype and a harmless polymorphism can be a difficult task. After detecting a genomic variant in a patient, the segregation of the change is studied. If the change is a mutation, it should co-segregate with the disease in the family. Mutations causing premature termination codons (PTC; e.g. nonsense mutations) (Maquat 2004), frameshift, or deletion of the whole gene almost certainly impact gene function. If a missense variation changes a conserved or functionally important amino acid, and the amino acid change is nonconservative, i.e. replacing a polar with a nonpolar amino acid (Grantham 1974), the variant is more likely to be pathogenic. Substitutions changing conserved splice site sequences lead to exon-skipping and altered splicing, which can have an impact on the expression or function of the gene. Silent variants affecting the splicing regulatory elements are more difficult to recognize, but may have deleterious effects (Cartegni *et al.* 2002; Pagani *et al.* 2004; Wang *et al.* 2007). In addition, a mutation should not be detected in a panel of healthy control individuals or if detected, it should not exceed the estimated carrier frequency in the population. In recessive diseases the variant should not be detected in homozygous form in control samples.

5.4.8 Functional studies

The ultimate proof showing that the detected variant is a pathogenic mutation comes from functional studies. The type of studies applicable is dependent on the known or predicted function of the gene. Sequence analysis of cDNA can be utilized to detect changes that have an impact on the splicing of the transcript (Varon *et al.* 2003). The mutation may alter the subcellular localization of the protein product or lead to aggregate formation (Huynh *et al.* 2003). If the mutated gene encodes for example an ion channel protein the effect of the variation may be evaluated by expressing the mutated protein in a cellular model and determining the channel properties, as has been performed for mutations in the voltage-gated sodium channel SCN1A (Spampanato *et al.* 2001). The defective gene may encode an enzyme whose activity can be measured in patient cells (Manya *et al.* 2007). If the mutation disrupts known or predicted functional domains it is more likely to be pathogenic (Escayg *et al.* 2000), and the mutation may impair protein's interactions with other proteins (Laezza *et al.* 2007). Lastly, the variant may be introduced into an animal model whose altered phenotype would be proof of pathogenicity (Davis *et al.* 2007).

5.5 Molecular chaperones

Molecular chaperones have an important task in assisting the proper folding and assembly of other proteins hence preventing inappropriate interactions between polypeptides and the accumulation of misfolded proteins. Chaperones also regulate several other cellular processes, including protein targeting, transport, recognition of misfolded proteins and targeting to degradation, and signal transduction (Ellgaard *et al.* 2003; Muchowski *et al.* 2005).

Conditions of cellular stress increase the synthesis of certain molecular chaperones, the heat shock proteins (HSPs), as a response to increases in the amount of misfolded proteins. HSPs are classified into six main families on the basis of their molecular mass: HSP100, HSP90, HSP70, HSP60, HSP40, and small heat-shock proteins (sHSPs) (Muchowski *et al.* 2005). Different family members function in different subcellular compartments, for example stress-70 protein (GRP75) in mitochondria, heat shock cognate 71 kDa protein (HSP7C) in cytosol, and 78 kDa glucose-regulated protein (GRP78) in the ER.

5.5.1 Diseases associated with defective chaperones

Some inherited diseases, many of which are described in more detail below, are caused by mutations in general chaperone proteins. In addition, mutations in a few substrate-specific chaperones have been identified (Barral *et al.* 2004; Muchowski *et al.* 2005).

Mutations in genes encoding sHSPs cause several diseases, some of which have clinical overlap with MSS. α-crystallin is composed of two subunits, αA-crystallin and αB-crystallin, and mutations in these two cause either dominantly or recessively inherited congenital cataracts (Litt *et al.* 1998; Pras *et al.* 2000; Berry *et al.* 2001; Mackay *et al.* 2003). In addition to lens, αB-crystallin is also expressed in muscle and other tissues, and mutations in this subunit cause myofibrillar myopathies as well (Vicart *et al.* 1998; Selcen *et al.* 2003). Mutations in two other sHSPs, HSP22 and HSP27, are associated with axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy (Evgrafov *et al.* 2004; Irobi *et al.* 2004).

Two mitochondrial chaperones have been associated with human disease. Mutations in heat shock 60kDa protein 1 (*HSPD1*), encoding mitochondrial chaperonin, cause a rare form of autosomal dominant pure hereditary spastic paraplegia, SPG13 (Hansen *et al.* 2002). An autosomal resessive syndrome, dilated cardiomyopathy with ataxia (DCMA), characterized by early onset dilated cardiomyopathy with conduction defects, non-progressive cerebellar ataxia, testicular dysgenesis, growth failure, and increase in urine organic acids is associated with mutation of *DNAJC19* gene encoding mitochondrial import inner membrane translocase subunit TIM14 (Davey *et al.* 2006).

One rare form of cerebellar ataxia with restricted geographic distribution, autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), is also most likely a chaperone-associated disease. The defective gene, the intronless 13-kb *SACS* gene on chromosome 13q11, encodes a novel protein, sacsin, with repeated heat-shock domains suggesting a chaperone function (Engert *et al.* 2000). The main clinical findings include slowly progressive ataxia first noted at gait initiation, spasticity of the limbs, slurred speech, and nystagmus (Bouchard *et al.* 1998). Interestingly, patients with ARSACS present atrophy of the superior cerebellar vermis and absence of Purkinje cells, features that resemble the cerebellar findings of MSS.

5.5.2 Chaperones in neurodegenerative diseases

Pathogenesis of many neurodegenerative disorders, like Alzheimer's disease, Parkinson's disease, Huntington's disease, and spinocerebellar ataxias, is characterized by the accumulation of intra- or extracellular protein aggregates leading to neurodegeneration (Muchowski 2002).

Molecular chaperones are often associated with the aggregates as they try to overcome the toxic effects of aberrantly folded proteins until the capacity of the chaperone machinery is overloaded (Barral *et al.* 2004). Thus, chaperones provide the first line of defense against misfolded aggregation-prone proteins and cellular dysfunction. Recent studies in animal models of human diseases have indicated that chaperones are among the most potent suppressors of neurodegeneration (Selkoe 2003; Muchowski *et al.* 2005). Molecular chaperones are therefore regarded as promising therapeutic targets for protein conformational disorders.

5.6 Molecular chaperones of the ER

5.6.1 ER function

The ER is a specialized subcellular compartment providing a major protein folding and maturation site for secreted, plasma membrane and secretory pathway organelle (ER, Golgi, lysosomes, and secretory vesicles) proteins (Ma *et al.* 2004; Ni *et al.* 2007). The ER differs from other folding compartments, e.g. the cytosol and mitochondria, by nature of the proteins that fold in the ER, by translocation machinery that segregates proteins across the membrane, by physiological conditions such as high oxidizing potential, occasionally high Ca2+ concentration, and/or by the presence of carbohydrates and glycosylation machinery (Brooks 1999; Stevens *et al.* 1999; Hendershot 2004). The ER possesses two major chaperone systems: the calnexin/calreticulin chaperone system assisting the folding of glycoproteins (Ellgaard *et al.* 2003) and the GRP78 chaperone system recognizing and binding to proteins containing hydrophobic residues (Ma *et al.* 2004; Ni *et al.* 2007). The ER contains at least 10 general chaperones (Table 3) and some protein-specific chaperones whose expression is tissue-specific (Stevens *et al.* 1999).

5.6.2 The GRP78 chaperone system

GRP78 (also called BiP for binding protein, HSPA5 for heat shock 70 kDa protein 5, and Kar2p in budding yeast) is the primary peptide-binding chaperone, which has many substrates and recognizes unfolded regions on proteins containing hydrophobic residues (Hendershot 2004). GRP78, which is encoded by the *HSPA5* gene, generally localizes to the ER but may have a different subcellular localization under stressful conditions (Sun *et al.* 2006). Its expression levels are tightly controlled and elevated in response to physiological or pathological stress, e.g. cell differentiation and accumulation of misfolded proteins (Hebert *et al.* 2007).

GRP78 has numerous functions in the ER (Figure 2). The first step in the biosynthesis of secretory pathway protein involves targeting it to the ER. The N-terminal ER-targeting sequence enters the ER lumen through a translocon co-translationally. GRP78 maintains the permeability barrier between cytosol and the ER by sealing the inactive translocon from the luminal side (Figure 2a). Once inside the ER the protein associates with GRP78 and starts folding (Figure 2b). Glycoproteins displaying N-linked glycans in their very N-terminal parts are, however, not targeted to GRP78 as they fold with the assistance of calnexin/calreticulin chaperone system. One of the major functions of GRP78 involves participating in identifying and degrading proteins that fail to mature properly (Figure 2c). This protein quality control system is called ER-associated degradation (ERAD). GRP78 binds calcium and contributes to calcium homeostasis in the ER (Figure 2d). Moreover, GRP78 is involved

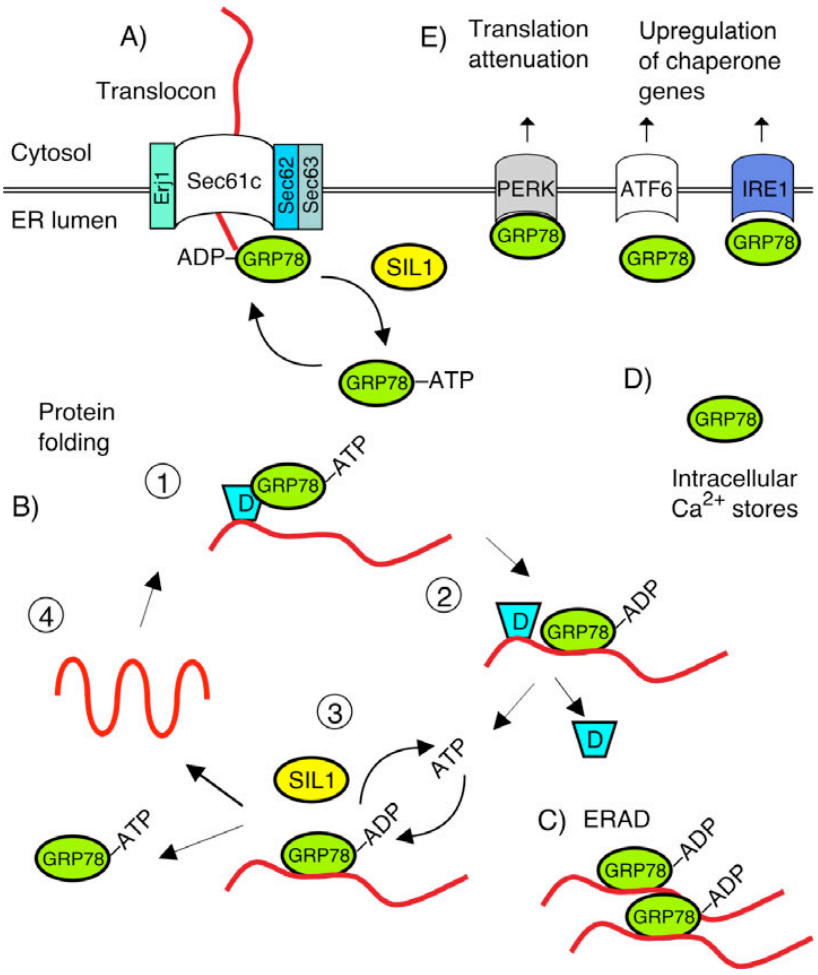

Figure 2. Functions of the ER localized chaperone GRP78. A) Nascent proteins are co-translationally translocated across the ER-membrane, where GRP78 has been shown to participate in sealing the translocon to maintain the permeability barrier between the ER and cytosol. GRP78 may require its ATPase activity during protein translocation (Zimmermann *et al.* 2006). B) The ATPase cycle of GRP78 regulates its binding and release from nascent protein chains: 1) Chaperone DJB11 binds to the unfolded region of a nascent protein and recruits ATP-bound GRP78 to the substrate. 2) DJB11 stimulates the ATPase activity of GRP78: ATP is hydrolyzed to ADP locking the ADP-bound GRP78 onto the unfolded substrate. 3) Nucleotide exchange factor, such as SIL1, catalyzes the release of ADP allowing the rebinding of ATP to GRP78. 4) The ATP-bound GRP78 is released from the substrate, which is ready for another round of assembly until the folding process is completed. C) GRP78 recognizes proteins that fail to fold properly and targets them to ER-associated degradation (ERAD). D) GRP78 has a role in maintaining ER calcium stores. E) GRP78 also participates in ER stress signaling, as in normal physiological conditions GRP78 is bound to ER transmembrane proteins IRE1, PERK, and ATF6 keeping them inactive. Modified from (Hendershot 2004).

in the regulation of signal transduction pathway called the unfolded protein response (UPR). In physiological conditions GRP78 binds to stress-sensing proteins located in the ER membrane and keeps them in an inactive state. If unfolded proteins accumulate in the ER, GRP78 is released from the stress-sensing proteins to interact with unfolded proteins (Figure 2e). Like all HSP70 family members, GRP78 binds both adenosine triphosphate (ATP) and adenosine diphosphate (ADP), which regulate its chaperoning functions with the exception of calcium storage (shown in more detail in Figure 2b). The GRP78 ATPase cycle is controlled by two other chaperone groups: HSP40 chaperones catalyze the

hydrolysis of ATP to ADP, which locks GRP78 to unfolded protein, and nucleotide exchange factors SIL1 and HYOU1 induce ADP release so that ATP can be substituted back (Tyson *et al.* 2000; Chung *et al.* 2002).

5.6.3 Cochaperone SIL1

Human SIL1 (also known as BAP for BiP-associated protein) is a ~54 kDa N-glycosylated protein with N-terminal ER targeting and C-terminal ER retrieval sequences (Chung *et al.* 2002). It acts as an adenine nucleotide exchange factor for GRP78 and promotes the exchange of bound nucleotides by stimulating the dissociation of ADP and the subsequent uptake of ATP. SIL1 was originally characterized in yeast (Boisrame *et al.* 1998; Tyson *et al.* 2000), and its nucleotide exchange factor activity can be replaced by another ER resident protein, HYOU1 (Steel *et al.* 2004; Weitzmann *et al.* 2006). SIL1 is evolutionarily conserved, with 85%, 50%, and 23% identical orthologs in mouse, zebrafish, and fruit fly, respectively (Figure 3). Interestingly, in SIL1 protein the ER retrieval motif has undergone several changes during evolution and the motif is not as highly conserved as it is in e.g. GRP78 or other ER resident chaperones (Study III Figure 2). Recently, a truncation of *Sil1* was shown to underlie the ataxia in woozy (*wz*) mutant mice (Zhao *et al.* 2005). Prior to cerebellar Purkinje cell degeneration, the *wz* mice develop ubiquitinated protein accumulations

Table 3. The main chaperones in the ER. Modified from (Stevens *et al.* 1999; Ellgaard *et al.* 2003; Hebert *et al.* 2007).

Chaperone family / Protein	Function
HSP70	
GRP78	Protein folding, ERAD, UPR
GrpE-like	
SIL1	Nucleotide exchange factor for GRP78
HYOU1	Possibly protein folding, nucleotide exchange factor for GRP78
HSP40	
DNJC1 (ERdj1)	Cofactor for GRP78
SEC63 (ERdj2)	Protein translocation across the ER membrane
DJB11 (ERdj3)	Cofactor for GRP78
DNJB9 (ERdj4)	Cofactor for GRP78
DJC10 (ERdj5)	Cofactor for GRP78
Hsp90	
GRP94	Folding - limited substrates
Lectins	Folding of N-linked monoglycosylated proteins
Calnexin	
Calreticulin	
EDEM1, EDEM2, EDEM3	
Peptidyl-prolyl isomerases	Possibly peptidyl-prolyl bond formation
Thiol-disulphide oxidoreductases	Oxidation, isomerization and reduction of disulphide bonds

in the ER and nucleus of these neurons. In order to restore ER homeostasis and degradation of the accumulations, unfolded protein response is induced in Purkinje cells, but the dysfunction of Sil1 leads to specific cell death with the exception of cells in cerebellar lobules IX and X.

```
Hs   MAPQSLPSSRMAPLGMLLGLLMAACFTFCLSHQNLKEFALTNPEKSSTKETERKETKAEE
Mm   MAPQHLPSTRMASPGMLLGLLLTSCLTLCLSCQNSNNFALTNPEKSIHQESDTKETREEE
     **** ***:***. *******:::*:*:*** ** ::**********  :*:: ***: **

Hs   ELDAEVLEVFHPTHEWQALQPGQAVPAGSHVRLNLQTGEREAKLQYEDKFRNNLK----G
Mm   ELDTEILEVFHPTQEWQTLQPGQAVPAGSHVRMNLQTGVNEVKLQQEDKFQNNLKGFKRG
     ***:*:*******:***:**************:***** .*.*** ****:****    *

Hs   KRLDINTNTYTSQDLKSALAKFKEGAEMESSKEDKARQAEVKRLFRPIEELKKDFDELNV
Mm   RRLDINANTYTSQDLKSALAKFKEGTEMENSKDELARQATVKQLFRPIEELKKEFDELNV
     :*****:******************:***.**:: **** **:**********:******

Hs   VIETDMQIMVRLINKFNSSSSSLEEKIAALFDLEYYVHQMDNAQDLLSFGGLQVVINGLN
Mm   VLETDMQIMVRLINKFNSSSSSLEEKVAALFDLEYYVHQMDNAQDLLSFGGLQVVINGLN
     *:************************:**********************************

Hs   STEPLVKEYAAFVLGAAFSSNPKVQVEAIEGGALQKLLVILATEQPLTAKKKVLFALCSL
Mm   STEPLVKEYAAFVLGAAFSSNPKVQVEAIEGGALQKLLVILATNQPLPAKKKVLFALCSL
     ******************************************:***.************

Hs   LRHFPYAQRQFLKLGGLQVLRTLVQEKGTEVLAVRVVTLLYDLVTEKMFAEEEAELTQEM
Mm   LRHFPYAQQQFLKLGGLQVLRSLVQEKSAKVLAVRVVTLLYDLVTEKMFAEEEAELTQDS
     ********:************:*****.::*****************************:

Hs   SPEKLQQYRQVHLLPGLWEQGWCEITAHLLALPEHDAREKVLQTLGVLLTTCRDRYRQDP
Mm   SPEKLQQYRQVQLLPGLQEQGWCEITAQLLALPEHDAREKVLQTLGALLTTCRDRYRQDL
     ***********:***** *********:******************.************

Hs   QLGRTLASLQAEYQVLASLELQDGEDEGYFQELLGSVNSLLKELR
Mm   QLSRTLGRLQAEYQALASLELQEGEDDGYFRELLASINSLMKELR
     **.***. ******.*******:***:***:***.*:***:****
```

Figure 3. Conservation of SIL1 amino acid sequence in human (Hs) and mouse (Mm). Human SIL1 and the mouse ortholog show 85% identical amino acid sequence. The putative ER retention tetrapeptide in the C-terminus is highlighted in gray.

6 AIMS OF THE STUDY

The aim of this study was to gain understanding into the molecular basis of Marinesco-Sjögren syndrome. The ultimate aim of the research is to understand the pathogenesis of Marinesco-Sjögren syndrome, which will be the basis for improved therapies.

The specific aims of this study were to:

1. Localize and refine the *MSS* locus in Finnish patients with MSS.

2. Identify and initially characterize the gene(s) and gene product(s) underlying MSS in Finnish as well as foreign patients.

3. Further describe the clinical features of the MSS patients to gain better understanding of the variable clinical characteristics of MSS.

7 MATERIALS AND METHODS

7.1 Ethical considerations

The initial research plan was approved by the Institutional Review Board of the Kainuu Central Hospital (Dnro 22/77/14.5.2003). The HUS Ethics Committee for Gynecology and Obstetrics, Otology, Ophthalmology, Neurology, and Neurosurgery approved an extension to the research plan (Dnro 350/E9/06).

7.2 MSS families and control individuals

In total, eight Finnish families participated in this study. DNA samples were available from 14 patients with MSS and 15 family members. Two of the families, M1 and M5, had been earlier described (Herva *et al.* 1987). Through collaboration we collected 23 families from different European countries, the USA, Australia, and Japan. For the foreign families DNA samples were available from 34 patients with MSS and 42 family members. RNA samples were obtained from six patients and four parents. The pedigrees of the MSS families with *SIL1* mutations are shown in Figure 4.

Fibroblast cell cultures were available from two Finnish and three Swedish and Norwegian patients and diagnostic muscle specimens from one Swedish and two Finnish patients.

The consecutive patient numbering used in study I corresponds to patient codes used in study II and this thesis as follows: 1: M603, 2: M116, 3: M115, 4: M110, 5: M503, 6: M113, 7: M114, 8: M403, and 9: M404.

The control samples used in this study were from the Centre d'Etude du Polymorphisme Humain (CEPH), unrelated Finnish blood donors, and unrelated healthy control subjects of Turkish and Japanese origin.

7.3 Methods

The methods used in this thesis are summarized in Table 4. Each method is described in more detail in the corresponding article.

Table 4. Methods used in this study.

Method	Study
DNA extraction	I, II, III
Polymerase chain reaction (PCR)	I, II, III
Microsatellite marker analysis	I, II
Linkage analysis	I, II
Haplotype analysis	I, II
Direct sequencing	II, III
Mutation analysis	II, III
In silico sequence analysis (BLAST, RepeatMasker, Primer 3, ClustalW, AceView)	II, III
RNA isolation	II, III
RT-PCR	II, III
Semiquantitative RT-PCR	II
Northern blot analysis	II
Western blot analysis	II
Immunohistochemical stainings	II
Immunofluorescence stainings	III
Fluorescence and bright field microscopy	II, III
Recombinant DNA techniques (cloning)	II, III
Site-directed mutagenesis	III
Transient DNA transfections	III

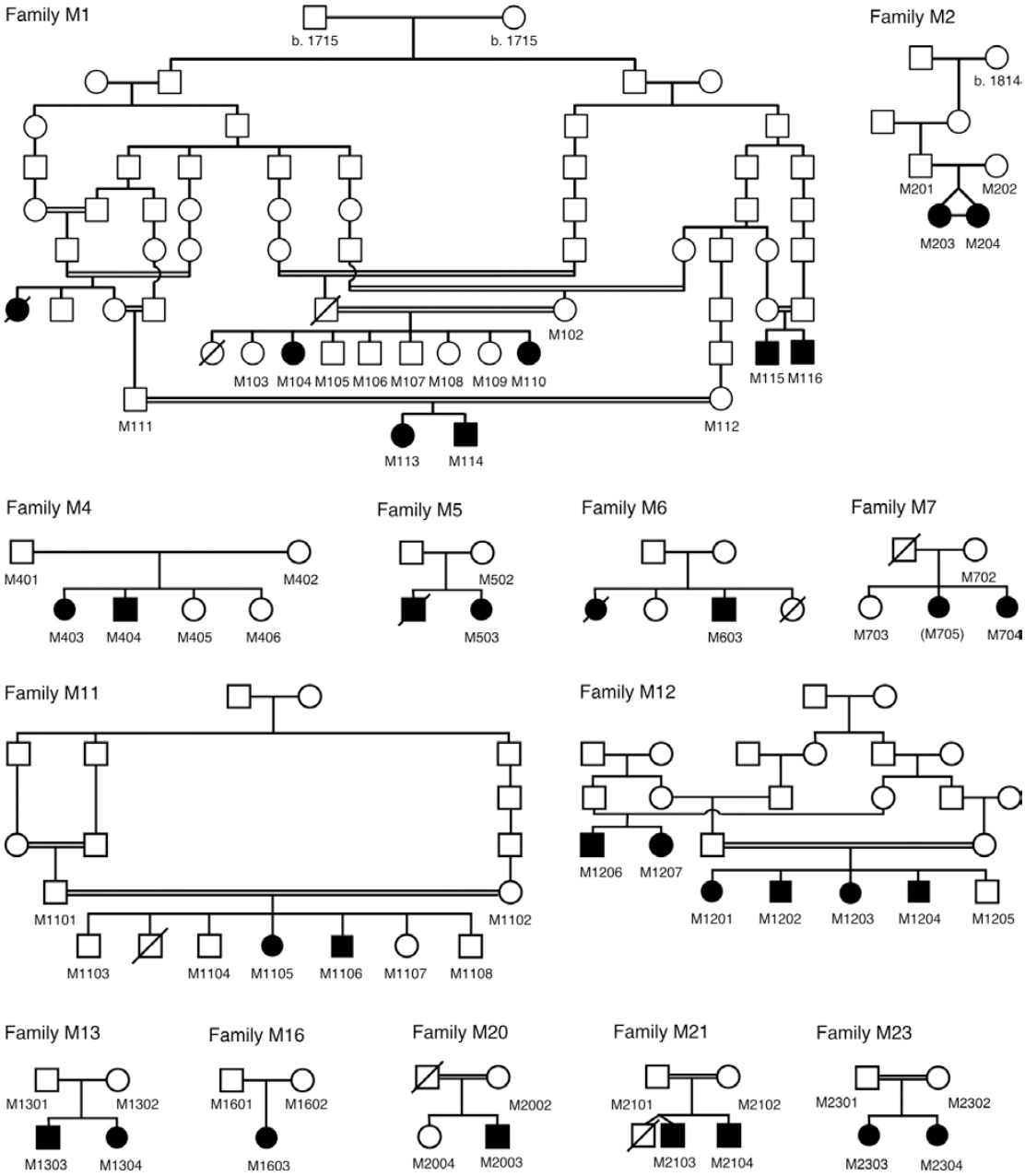

Figure 4. The MSS families with *SIL1* mutations. The affected individuals are indicated with filled symbols. For simplicity the individuals in the older generations in families M1, M2, M11, and M12 are not marked as deceased. Families M1, M4, M5, and M6 are from Finland. Family M2 lives in Sweden, but the paternal grandmother was born in Finland. The three Norwegian families are M7, M12, and M13. M11 and M20 are of Turkish origin. Family M16 is French and families M21 and M23 are Japanese. The DNA samples and patient records (affected persons only) were available from numbered individuals. Patient records were available for one additional patient shown in parenthesis.

7.3.1 Genotyping and linkage analysis (I, II, unpublished)

Genomic DNA was isolated using standard methods. The segregation of the MSS phenotype with alleles of fluorescently labeled microsatellite markers was evaluated for two chromosomal loci. Six polymorphic markers covering the chromosome 18qter locus and ten markers covering the locus on 5q31 were studied. The marker order, intermarker distances, and primer sequences were obtained from the Marshfield genetic map (Broman *et al.* 1998), the deCODE genetic map (Kong *et al.* 2002), and NCBI sequence based STS map (NCBI Build 35.1). The PCR-amplified products were separated with an ABI 3730 DNA Analyzer (Applied Biosystems) and determined allele sizes were determined using the GeneMapper v5.0 program (Applied Biosystems). The haplotypes were constructed manually. The pair-wise lod scores were calculated with the help of MLINK program from the LINKAGE package (Lathrop *et al.* 1984) assuming autosomal recessive inheritance, complete penetrance and a disease allele frequency of 0.001. For the 5q31 locus, the marker allele frequencies were determined from 42 Finnish control chromosomes and for the 18qter locus equal allele frequencies were assumed.

7.3.2 Mutation analysis (II, III)

The gene content of the interval spanning *D5S479* and *D5S2116* was examined with the NCBI Map Viewer, the Ensembl Human Genome Browser, and the UCSC Human Genome Browser. The exons and exon-intron boundaries of *SIL1*, the ten positional candidate genes on chromosome 5q31 (*CTNNA1*, *EGR1*, *GFRA3*, *HBEGF*, *HSPA9B*, *KLHL3*, *MYOT*, *NRG2*, *SPOCK1*, and *WNT8A*), and the three functional candidate genes (*AARS*, *HSPA5*, and *HYOU1*) were amplified from genomic DNA with intronic primers designed with Primer3 (Rozen *et al.* 2000) or ExonPrimer programs. The amplified products were purified using the PCR Product Pre-Sequencing Kit (USB Corporation) and both strands were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI 3730 DNA Analyzer. Sequences were analyzed using the Sequencher program (version 4.2 to 4.7; Gene Codes Corporation). Unrelated control individuals were screened for the identified mutations by sequencing.

The mutations implied to affect the splicing of *SIL1* were analyzed from cDNA. Patient and control RNA was extracted from freshly obtained peripheral blood using the PAXgene Blood RNA System (PreAnalytiX) or from fibroblast cultures with the RNeasy mini kit (Qiagen, Hilden, Germany). The reverse transcriptase (RT) PCR reaction was performed with the High-Capacity cDNA Archive Kit (Applied Biosystems). The cDNA was amplified with exonic primers and the products were sequenced either directly (see above), after purification from a gel (QIAquick Gel Extraction Kit; QIAGEN) or after TA-cloning into the pCRII vector using the TA Cloning Kit (Invitrogen).

7.3.3 Gene expression studies (II)

Commercial multiple tissue Northern blots (Human Fetal II, Human Brain II, and Human Brain V; BD Biosciences Clontech) were hybridized with a 1519 bp PCR-generated *SIL1* cDNA probe covering the whole coding region as well as with *GAPDH* (BD Biosciences Clontech). The probes were labeled with ^{32}P-dCTP using RediPrimeII kit (Amersham Pharmacia) and blocked with human placental (0.5 mg/ml) and salmon sperm (0.25 mg/ml) DNA before hybridization. Hybridizations were carried out for 2 h at 65°C in

ExpressHyb solution (BD Biosciences Clontech), and filters were washed at 60°C according to the manufacturer's instructions. Semiquantitative RT-PCR was performed from BD™ MTC Multiple Tissue cDNA Panels (Human I, Human II, and Human Fetal; BD Biosciences Clontech) according to the manufacturer's instructions. A 930 bp *SIL1* fragment (c.584_*+127) was PCR amplified and separated by gel electrophoresis.

7.3.4 Immunohistochemistry (II)

Frozen sections from muscle biopsies taken for diagnostic examinations from patients and controls were stained with a polyclonal goat anti-human SIL1 antibody (Abcam Ltd) detecting a C-terminal epitope. Sections were briefly fixed in –20°C acetone after which the protocol described below was followed with the exception that no antigen retrieval was used. Goat anti-human SIL1 antibody (Abcam Ltd) was used at a 1:200 dilution (stock 0.50 mg/ml).

Paraffin-embedded sections of C57BL mouse tissues of embryonic day 12.5, 15.5, and 18.5, and of postnatal day 5, 14, 60, and 180 were deparaffinized, hydrated, and treated with 10 mmol/l citric acid in microwave oven for 5 min for endogenous antigen retrieval. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide for 5 min, and nonspecific binding was blocked with 1.5% normal serum in 0.1% Tween-phosphate-buffered saline (PBS) for 30–60 min. The primary antibodies were incubated overnight at 4°C in 0.1% Tween-PBS. Goat anti-human SIL1 antibody (Abcam Ltd) was used at a 1:400 dilution (stock 0.50 mg/ml) and rabbit anti-human HSPA5 antibody (Sigma) at a 1:1500 dilution (stock 10.3 mg/ml). Western blot analyses were performed to verify specific cross-reactivity of both antibodies with mouse tissue (data not shown). An avidin-biotin immunoperoxidase system (Vectastain Elite ABC Kit, Vector Laboratories) was utilized to visualize the bound antibody with diaminobenzidine (Sigma) as chromogen. The sections were counterstained with hematoxylin. Non-immune serum or PBS was used instead of the primary antibody for negative controls.

7.3.5 The expression plasmid and site-directed mutagenesis (III)

The ORF of the human *SIL1* cDNA was cloned in frame into the aminoterminal hemagglutinin (HA) tag containing pAHC expression vector (kindly provided by T. Mäkelä, University of Helsinki, Finland). Two nucleotide changes, c.1370T>C (p.Leu457Pro) and c.1372A>T (p.Lys458X), were introduced into the pAHC-SIL1 construct by site-directed mutagenesis using the Quick Change Site-Directed Mutagenesis Kit (Stratagene). The primer sequences are available on request. All constructs were verified by sequencing the entire coding regions of the inserts.

7.3.6 Cell culture and transfections (III)

COS-1 cells (American Type Culture Collection) were cultivated in Dulbecco's modified Eagle's medium (Lonza Walkersville Inc.) supplemented with 10% fetal calf serum (PromoCell GmbH), penicillin and streptomycin, and 1x GlutaMAX (Gibco, Invitrogen). The cells (1.5x 105) were plated onto 6-well plates on coverslips one day prior to transfection, and were transfected with 2 μg of wild-type or mutated pAHC-SIL1 using FuGENE 6 Transfection Reagent (Roche) according to the manufacturer's instructions. For immunofluorescence analysis, protein production was inhibited by two hours incubation with cycloheximide 18 hours after

transfection and cells were fixed with 4% paraformaldehyde.

Hippocampi of embryonic day 16.5 C57BL mice were dissected out in ice-cold 100 mM PBS without Ca2+/Mg2+ (PBS, pH 7.4) supplemented with 20 mM glucose. Tissue was trypsinized by adding 20 ng/ml DNAseI and 0.1%/0.4 mM trypsin-EDTA and incubating for 5 minutes at 37°C. Trypsin was inhibited with fetal calf serum and the tissue was further mechanically dissociated. Four hippocampi were pooled and resuspended in a Neurobasal medium (Gibco, BRL) supplemented with 1x B27 (Gibco), glutamine and antibiotics. The cells were plated on poly-D-lysine (Sigma-Aldrich) coated 58 mm dishes with coverslips and cultured for 8-14 days. The cells were fixed in 4% paraformaldehyde in PBS for 20 min.

7.3.7 Immunofluorescence analysis (III)

Transfected COS-1 cells were permeabilized with 0.1% Triton X-100 in PBS for 15 min or with 0.2% saponin in PBS supplemented with 0.5% bovine serum albumin for 30 min. Mouse or rabbit anti-HA (Covance Research Products, clone 16B12, 1:16000; and Santa Cruz Biotechnology, Y-11, 1:500, respectively) antibodies were used to detect the N-terminus of the HA-tagged SIL1 protein. To differentiate subcellular compartments, rabbit anti-GRP78 (Sigma-Aldrich Co, 1:100) and mouse anti-PDI (Stressgen, Assay Designs Inc., 1:50) antibodies were used for ER, rabbit anti-p58 (kindly provided by R. Pettersson, Ludwig Institute for Cancer Research, Sweden) was used for ERGIC, sheep anti-TGN46 (Serotec, 1:200) and rabbit anti-Giantin (BioSite, 1:1000) were used for Golgi apparatus, and mouse anti-Lamp-1 (DSHB, H4A3; developed by August, J.T and Hildreth, J.E.K and obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biological Sciences, Iowa City, IA, USA; 1:100) was used for lysosomes. Secondary antibodies were Cy2- or Cy3-conjugated anti-rabbit IgG, Cy2-conjugated anti-goat IgG, and Cy2- or Cy3-conjugated anti-mouse IgG (Jackson ImmunoResearch, 1:200).

The hippocampal neurons were permeabilized with 0.1% Triton X-100 in PBS containing 0.5% BSA. The primary antibodies were goat anti-SIL1 (Abcam Ltd, 1:5), rabbit anti-GRP78 (Sigma-Aldrich, 1:300), and rabbit anti-PDI (Stressgen, 1:50). The secondary antibodies were Cy2-conjugated anti-goat-IgG and Cy3-conjugated anti-rabbit-IgG (Jackson ImmunoResearch, 1:250).

7.4 Web resources

Online Mendelian Inheritance in Man (OMIM) cataloging human genes and genetic disorders is available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.
Three genome browsers, the Ensembl Human Genome Browser (http://www.ensembl.org/Homo_sapiens/), NCBI MapViewer (http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606), and UCSC Human Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), were used in this study.

Microsatellite marker allele numbering was determined from the CEPH database (http://www.cephb.fr/cephdb/php/eng/). The primers used in sequencing were designed with either Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) or Exon Primer (http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html) programs.

Gene names are according to HUGO Gene Nomenclature Committee (http://www.genenames.org/). The recommendations for the

description of sequence variants published by the Human Genome Variation Society (HGVS) (http://www.hgvs.org/mutnomen/) were followed in describing the mutations.

The GO database (http://www.geneontology.org/) was utilized in prioritizing the positional candidate genes on chromosome 5q31. Alternatively spliced *SIL1* transcripts were identified using NCBI AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html). *In silico* analysis of the c.1030-9G>A mutation was performed with two exon prediction programs. GENSCAN is available at http://genes.mit.edu/GENSCAN.html and GrailEXP Exon Prediction Program (Perceval) at http://compbio.ornl.gov/grailexp/. *SIL1*, *HSPA5*, and *HYOU1* expression patterns based on the GNF (Genomics Institute of the Novartis Research Foundation) Human Gene Expression Atlas 2 database were analyzed with the UCSC gene sorter (http://genome.ucsc.edu/cgi-bin/hgNear). Expressed sequence tags (ESTs) for *SIL1* and *HSPA5* in different ocular regions were found in the NEIBank EST database (http://neibank.nei.nih.gov/index.shtml). Sequence alignments were made with ClustalW (http://www.ebi.ac.uk/clustalw/) and Multiple Alignment using Fast Fourier Transform (MAFFT) (http://www.ebi.ac.uk/mafft/).

8 RESULTS AND DISCUSSION

8.1 Refinement of the *MSS* locus (I, II, unpublished)

Before the initiation of this study families having a subtype of MSS with myoglobinuria and neuropathy had been linked to a locus on 18qter (Merlini *et al.* 2002), which was initially characterized as a locus for the congenital cataracts with facial dysmorphism and neuropathy (CCFDN) syndrome (Angelicheva *et al.* 1999). Another locus spanning a 9.3 cM interval on chromosome 5q31 had been reported for classical MSS (Lagier-Tourenne *et al.* 2003) shortly after we started our study. We pursued the candidate locus approach to evaluate whether one of the known loci was underlying MSS in Finland. For linkage analysis, DNA samples were available from 8 affected individuals, 5 parents, and 8 healthy siblings in two families, M1 and M4. In addition, two families with samples available from one affected individual were identified. Due to the size and multiple consanguinity loops in family M1, we chose the MLINK program for lod score calculations.

8.1.1 Exclusion of the candidate locus on chromosome 18qter

We genotyped six polymorphic microsatellite markers, *D18S844*, *D18S462*, *D18S461*, *D18S1122*, *D18S1095*, and *D18S70*, and constructed the haplotypes manually. Assuming that a common ancestor and one homozygous mutation underlies MSS in Finland, we expected to detect homozygosity across the linked region. Instead, we found that two pairs of affected siblings (M115 and M116; M403 and M404) did not share haplotypes on 18qter (Figure 5) and another pair of siblings, M104 and M110, shared haplotypes only partially. Moreover, the two-point lod scores were < -2 for each marker (Table 5) excluding linkage to this region.

8.1.2 Mapping MSS to 5q31 in the Finnish families

We selected 18 polymorphic microsatellite markers spanning the chromosome 5q31 locus (Table 6). We confirmed linkage to this locus in family M1, in which meiotic and historical recombinations defined a 3.52-Mb region between markers *D5S479* and

Table 5. Results of two-point linkage analysis on chromosome 18qter.

theta / marker	0.000	0.001	0.010	0.050	0.100	0.200	0.300	0.400	cM (Marshfield[1])
D18S844	-inf	-3.599	-1.440	0.088	0.517	0.526	0.295	0.100	116.44
D18S462	-inf	-3.883	-1.622	-0.058	0.362	0.401	0.229	0.082	120.05
D18S461	-inf	-0.595	1.306	2.252	2.272	1.704	0.978	0.364	120.05
D18S1122	-inf	-1.148	0.882	2.071	2.179	1.627	0.879	0.287	122.61
D18S1095	-inf	-5.232	-2.753	-0.741	-0.037	0.305	0.230	0.073	124.11
D18S70	-inf	-6.595	-3.578	-1.070	-0.173	0.290	0.240	0.081	126.00

[1] (Broman *et al.* 1998)

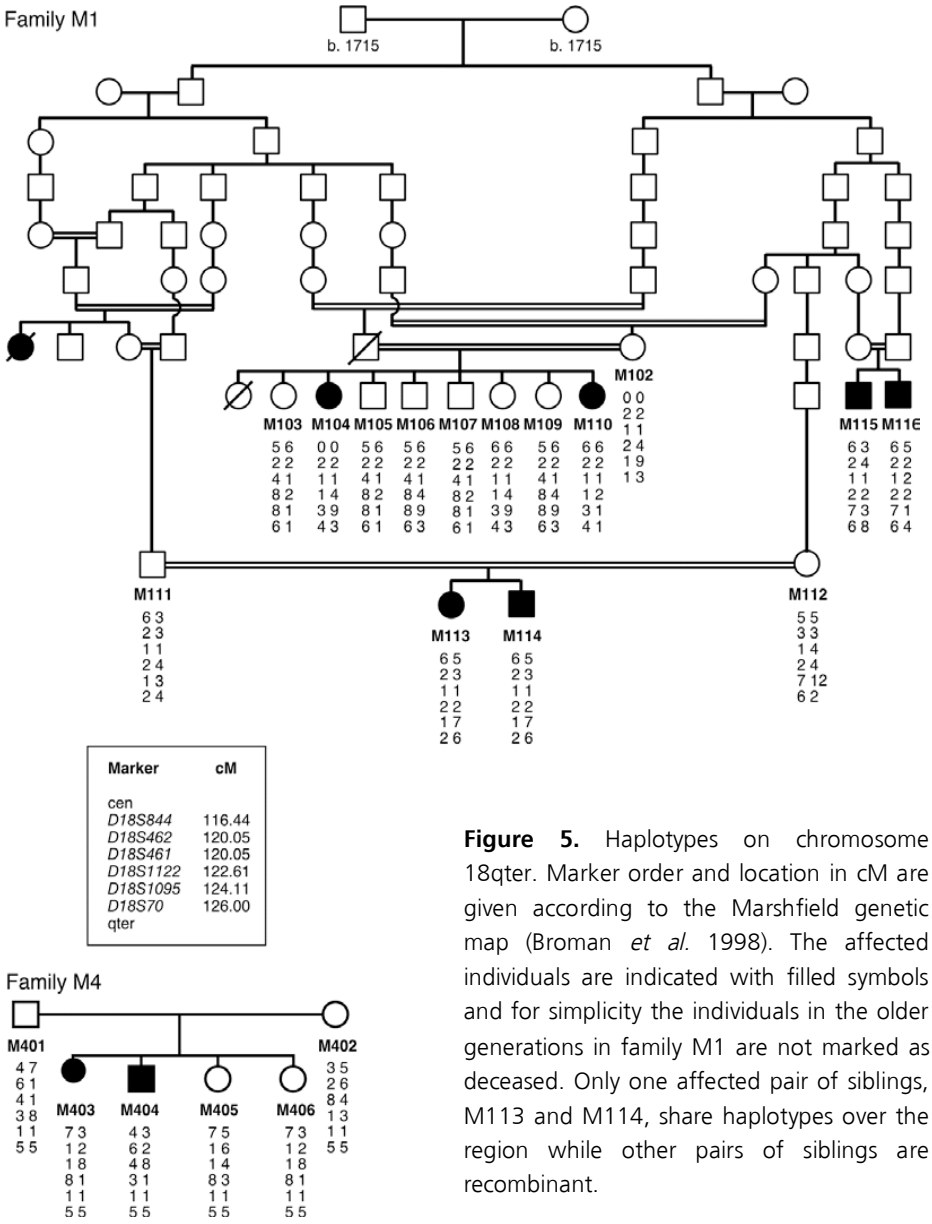

Figure 5. Haplotypes on chromosome 18qter. Marker order and location in cM are given according to the Marshfield genetic map (Broman *et al.* 1998). The affected individuals are indicated with filled symbols and for simplicity the individuals in the older generations in family M1 are not marked as deceased. Only one affected pair of siblings, M113 and M114, share haplotypes over the region while other pairs of siblings are recombinant.

D5S2116 (Figure 6; Study II Supplementary Figure 2). Four additional Finnish patients from three unrelated families M4, M5, and M6 shared the 3.52-Mb haplotype with patients from the M1 pedigree, suggesting a founder effect (Study II Supplementary Figure 2). The highest two-point lod score of 7.095 (at θ=0.000) was obtained with *D5S414* for families M1 and M4 (Table 6).

8.2 Analysis of the positional candidate genes (II, unpublished)

The haplotypes were constructed manually and this analysis excluded *SAR1B* (also called *SARA2*) (Jones *et al.* 2003) as a positional candidate gene. Mutations in *SAR1B* had been identified in patients with chylomicron retention disease (CMRD, MIM 246700), including one family with CMRD and MSS. However, no *SAR1B* mutations had been found in additional 5q31-linked MSS families (Lagier-Tourenne *et al.* 2003). The parents of the two patients with CMRD and MSS were from a small village in Southern Italy, but were not reported to be consanguineous (Aguglia *et al.* 2000). Nevertheless, the

Table 6. Results of two-point linkage analysis on the *MSS* locus 5q31.

theta / marker	0.000	0.001	0.010	0.050	0.100	0.200	0.300	0.400	cM (Marshfield[1])	(deCODE[2])
D5S1995	-inf	-1.956	-0.045	0.962	1.090	0.825	0.456	0.156	137.95	135.81
D5S479	-inf	2.371	3.229	3.297	2.822	1.699	0.773	0.205	141.27	138.82
D5S1983	1.625	1.619	1.559	1.306	1.019	0.561	0.255	0.078	141.82	138.96
D5S414	7.095	7.078	6.917	6.200	5.294	3.505	1.888	0.669	141.82	-
D5S500	4.352	4.339	4.218	3.686	3.039	1.850	0.888	0.266	140.72	138.96
D5S476	2.913	2.902	2.809	2.404	1.926	1.104	0.503	0.140	140.72	139.38
D5S2009	1.134	1.129	1.086	0.902	0.670	0.383	0.174	0.051	142.92	139.38
D5S2116	1.157	1.177	1.296	1.374	1.218	0.783	0.398	0.130	142.92	139.91
D5S658	-inf	1.257	2.320	2.838	2.646	1.796	0.896	0.259	142.92	-
D5S1979	-inf	1.010	2.069	2.575	2.374	1.552	0.751	0.216	144.06	-

[1] (Broman *et al.* 1998), [2] (Kong *et al.* 2002)

D5S642	*D5S1995*	*D5S816*	*D5S479*	*D5S1983*	*D5S414*	*D5S500*	*D5S476*	*D5S2009*	*rs13385*	*D5S2116*	*D5S658*	*D5S2010*	*D5S1979*	No. of disease chromosomes	Patient no.
5	5	2	3	1	8	5	2	1	1	1	3	9	3	5	
1	5	2	3	1	8	5	2	1	1	4	2	5	5	5	
2	2	6	2	1	8	5	2	1	1	4	2	5	5	1	M104
1	5	2	3	1	8	5	2	1	2	1	5	6	6	1	M113
1	2	1	2	2	7	9	2	1	0	8	2	1	8	1	M203

Figure 6. Haplotypes across the *MSS* locus in families M1 and M2. In the M1 family four different haplotypes were present. The majority of the patients shared the two haplotypes depicted on the first two rows. Historical recombinations restricted the candidate region between markers *D5S642* and *D5S2116*. Patient M104 had a meiotic recombination further narrowing the region to 3.52 Mb between markers *D5S479* and *D5S2116*. A meiotic recombination present in patient M113 did not initially narrow the region until a heterozygous SNP rs13385 was detected while sequencing the *HBEGF* gene. The *MSS* locus was refined to between markers *D5S500* and *rs13385* according to the "Finnish" disease chromosome present in patients M203 and M204.

patients were homozygous for the markers across the refined *MSS* locus (Jones *et al.* 2003) and it therefore remained possible that they had two separate recessive diseases.

8.2.1 Genes in the critical region

According to the sequence browsers, the 3.52-Mb region contained 45 genes or pseudogenes (Figure 7 and Table 7). We selected genes for sequencing based on their tissue expression, GO annotations, and known or predicted function. Ten genes, highlighted with gray in Table 7, were excluded by sequencing the exons and exon-intron boundaries from genomic DNA of patient M113, her parents, patient M203, and one Finnish control individual.

Table 7. Positional candidate genes in the 3.52-Mb *MSS* locus on chromosome 5q31.

Symbol	Description
SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1
KLHL3	kelch-like 3 (Drosophila)
HNRPA0	heterogeneous nuclear ribonucleoprotein A0
NPY6R	neuropeptide Y receptor Y6 (pseudogene)
MYOT	myotilin
PKD2L2	polycystic kidney disease 2-like 2
C5orf5	chromosome 5 open reading frame 5
WNT8A	wingless-type MMTV integration site family, member 8A
NME5	non-metastatic cells 5, protein expressed in (nucleoside-diphosphate kinase)
BRD8	bromodomain containing 8
KIF20A	kinesin family member 20A
CDC23	CDC23 (cell division cycle 23, yeast, homolog)
GFRA3	GDNF family receptor alpha 3
CDC25C	cell division cycle 25C
FAM53C	family with sequence similarity 53, member C - hypothetical protein *
JMJD1B	jumonji domain containing 1B - hypothetical protein *
REEP2	receptor accessory protein 2 - hypothetical protein *
EGR1	early growth response 1
ETF1	eukaryotic translation termination factor 1
HSPA9B	heat shock 70kDa protein 9B (mortalin-2)
LOC391836	similar to ribosomal protein L10a
CTNNA1	catenin (cadherin-associated protein), alpha 1, 102kDa
LOC401210	hypothetical gene supported by AK022326 - discontinued **
SIL1	**SIL1 homolog, endoplasmic reticulum chaperone (S. cerevisiae)**
MATR3	matrin 3
LOC441111	LOC441111 - discontinued **
PAIP2	poly(A) binding protein interacting protein 2
SLC23A1	solute carrier family 23 (nucleobase transporters), member 1
PACAP	proapoptotic caspase adaptor protein
LOC389333	LOC389333
LOC389334	LOC389334 - discontinued **
LOC202051	hypothetical protein LOC202051
DNAJC18	DnaJ (Hsp40) homolog, subfamily C, member 18 - hypothetical protein *
LOC340061	hypothetical protein LOC340061
LOC401211	similar to nuclear receptor coactivator 4
UBE2D2	ubiquitin-conjugating enzyme E2D 2 (UBC4/5 homolog, yeast)
CXXC5	CXXC finger 5
PSD2	PSD2 pleckstrin and Sec7 domain containing 2 - hypothetical protein *
NRG2	neuregulin 2
PURA	purine-rich element binding protein A
C5orf32	chromosome 5 open reading frame 32
PFDN1	prefoldin 1
HBEGF	heparin-binding EGF-like growth factor
SLC4A9	solute carrier family 4, sodium bicarbonate cotransporter, member 9
ANKHD1	ANKHD1 ankyrin repeat and KH domain containing 1

The double line on the left indicates the final *MSS* locus refined by the shared haplotype between the Swedish twins and Finnish patients as well as a meiotic recombination observed in patient M113.

* Current gene names are given to hypothetical genes that have been characterized since the release of NCBI Build 35.1 (August 2004).

** These genes have been discontinued since the release of NCBI Build 35.1.

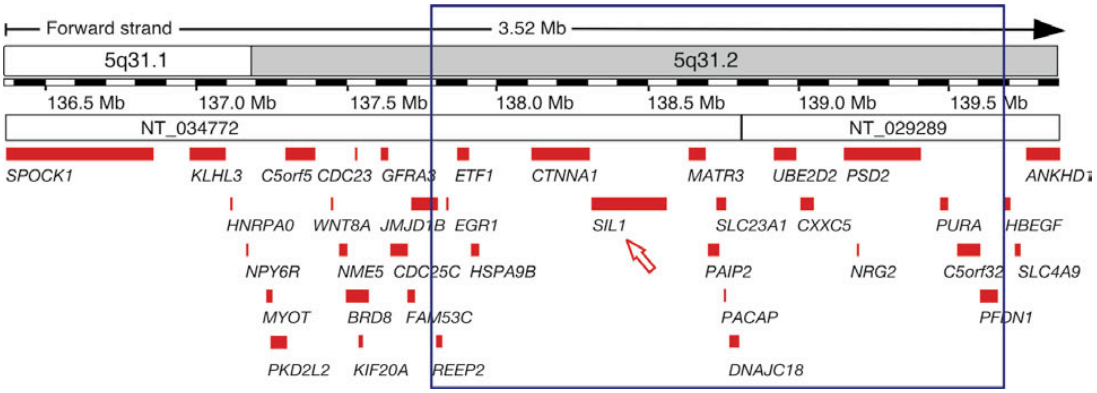

Figure 7. Genes in the 5q31 candidate region. The original 3.52-Mb candidate region is shown. The upper bar shows the chromosomal band and each gene is depicted as one solid bar. Blue box indicates the final candidate region between *D5S500* and rs13385. *SIL1* is marked with a red arrow. The picture was modified from the Ensembl genome browser view of the region.

At the same time as we performed the candidate gene analysis we found that a pair of Swedish monozygotic twins with MSS (M203 and M204), whose paternal great grandmother was born in Finland, shared the Finnish founder haplotype on the paternal chromosome between markers *D5S500* and *D5S2116*. This defined the Finnish haplotype to a region of 1.98 Mb and excluded 17 genes (Figures 6 and 7, Table 7). In addition, M113 had a recombinant SNP (rs13385) in the 3' UTR of the *HBEGF* gene, which further narrowed the critical region and excluded three more genes.

8.2.2 Mutations in SIL1 underlie MSS

The 11th gene selected for sequencing was *SIL1*, the ER-resident cochaperone of GRP78. Human *SIL1* is a 1923 bp transcript encoding a 461 amino acid protein. The main transcript has ten exons, of which nine are coding, and spans a 251.7-kb genomic region (Figure 8). One additional 5' noncoding exon (marked as exon 1a) is found at least in placental tissue. All Finnish patients were homozygous for a 4-nucleotide duplication in exon 6 compatible with the predicted founder effect. As expected, the Swedish patients M203 and M204 were compound heterozygotes for the Finnish founder mutation and another mutation changing the conserved donor splice site of intron 6. The analysis of further families with MSS from Norway, France, and Turkey revealed additional mutations, confirming that *SIL1* mutations underlie MSS. Refining the candidate region using the ancestral Finnish haplotype and identifying the founder mutation in *SIL1* is yet another example of exploiting the founder effect in disease gene identification (see Section 5.4.5).

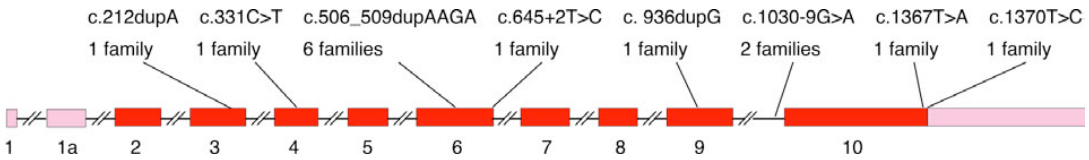

Figure 8. The structure of *SIL1* showing the MSS-associated mutations identified in this study. The main variant of *SIL1* is transcribed from ten exons. The boxes represent the exons. The coding region is in red while untranslated regions are in pink. The positions of the mutations are shown with lines.

8.3 *SIL1* mutations in patients with MSS (II, III, unpublished)

We identified a total of eight mutations in the *SIL1* gene in 29 patients of different ethnic origins (Figure 8 and Table 8). The mutations were distributed quite evenly throughout *SIL1*. Two mutations, c.331C>T and c.1367T>A, were nonsense changes creating PTCs: p.Arg111X and p.Leu456X, respectively. Three mutations were small duplications (c.212dupA, the Finnish founder mutation c.506_509dupAAGA, and c.936dupG) leading to frameshifts and PTCs (p.His71GlnfsX5, p.Asp170GlufsX4, and p.Leu313AlafsX39, respectively). Two mutations affect splice sites (c.645+2T>C and c.1030-9G>A) and their consequences are discussed in Section 8.3.2. Only one mutation, c.1370T>C, was a missense change (p.Leu457Pro). The Swedish patients were compound heterozygous for c.506_509dupAAGA and c.645+2T>C, while all the other patients carried homozygous mutations.

All the mutations segregated with the disease phenotype in the respective families. One carrier for the c.506_509dupAAGA mutation was identified among 96 Finnish controls. In 54 Japanese controls, we detected one carrier for the c.936dupG mutation. CEPH controls were negative for all the mutations. The detection of mutation carriers among healthy unrelated controls may imply that the carrier

Table 8. MSS-associated mutations in *SIL1*.

Location	Nucleotide change	Predicted consequence	Ethnic origin	Reference
Exon 3	c.212dupA	p.His71GlnfsX5	France	II
Exon 4	c.331C>T	p.Arg111X	Turkey Iran, Turkey Italy	II (Senderek *et al.* 2005) (Annesi *et al.* 2007)
Exon 6	c.506_509dupAAGA	p.Asp170GlufsX4	Finland, Sweden, Norway	II
Intron 6	c.645+2T>C	p.Val186_Gln215del p.Ala152_Gln215del	Sweden	II
Exon 9	c.936dupG	p.Leu313AlafsX39	Japan	III
Intron 9	c.1030-9G>A	p.Phe345AlafsX9	Norway	III
Exon 10	c.1367T>A	p.Leu456X	Turkey	III
Exon 10	c.1370T>C	p.Leu457Pro	Japan	III
Exon 3	c.178C>T	p.Glu60X	Vietnam	(Senderek *et al.* 2005)
Exon 4	c.346delG	p.Gly116fsX42	Vietnam	(Senderek *et al.* 2005)
Intron 6	c.645+1G>A	skipping of exon 6	Turkey	(Senderek *et al.* 2005)
Exon 9	c.947_948insT	p.Leu316fsX36	Germany, Russia	(Senderek *et al.* 2005)
Intron 9	c.1029+1G>A	skipping of exon 9	Bosnia	(Senderek *et al.* 2005)
Intron 9	c.1030-18G>A	p.Met344fsX13	Germany	(Senderek *et al.* 2005)
Exon 10	c.1249C>T	p.Gln417X	Mali	(Senderek *et al.* 2005)
Exon 10	c.1312C>T	p.Gln438X	Egypt	(Karim *et al.* 2006)
Exon 10	c.1366delT	p.Leu456fsX2	Russia	(Senderek *et al.* 2005)

frequency is relatively high at least in certain parts of Finland and Japan.

At the same time as study II was published, Senderek *et al.* independently identified nine *SIL1* mutations causing MSS (Senderek *et al.* 2005). In addition to these mutations, others have reported single MSS families with *SIL1* mutations (Karim *et al.* 2006; Annesi *et al.* 2007), summing up the total number of mutations to 17 (Table 8).

8.3.1 The geographical distribution of the Finnish founder mutation

As discussed above, all the Finnish patients were homozygous for c.506_509dupAAGA resulting in a frameshift at codon 170 and a stop codon after three novel amino acids. The Swedish patients had this mutation in their paternal allele. The largest Finnish family M1 originated from Kainuu and the three smaller families (M4, M5, and M6) were from Ostrobothnia (Figure 9). The paternal great grandmother of the Swedish family M2 was born in Lapland, Finland. Although it is likely that these families are distantly related to each other, we were unable to find a relationship between the families from parish records. In addition, patients in one Norwegian family, M12, were homozygous for the c.506_509dupAAGA mutation. Interestingly, although the Norwegian family had no knowledge of Finnish ancestors, it originated from the Finnmark county (Skre *et al.* 1976), which is known for receiving immigrants from northern Finland. Moreover, the patients in family M12 shared the 2–1 haplotype (markers *D5S479* and *D5S2009;* see Figure 6) with Finnish and Swedish patients as well as an allele A of a coding SNP in *SIL1* exon 3 (rs3088052). It thus remains possible that the c.506_509dupAAGA mutation arose only once and that these families share a common ancestor.

8.3.2 Consequences of the splice site mutations

The effect of the splice site mutation c.645+2T>C was analyzed by RT-PCR from the two compound heterozygote patients and

Figure 9. Origin of the Nordic MSS families. The ancestries of the Finnish families and the paternal great grandmother of the Swedish family were from Ostrobothnia, Kainuu, and Lapland (filled circles). The Norwegian family with the same homozygous mutation originated from Finnmark county (filled circle), but had no known ancestry in Finland. The other Norwegian mutation was detected in two families from northern Norway (Troms and Nordland counties) (gray circles). The origin of the Swedish family with the splice site mutation is indicated with a gray triangle. This figure is based on an original by Teppo Varilo and has been modified with his kind permission.

their mother who carried this mutation (Study II Figure 1). The c.506_509dupAAGA allele transcript showed the expected length product. Two abnormal sized variants were identified from the splice site mutant allele. In the shorter variant, expressed at higher levels, exon 6 is skipped, predicting an in-frame deletion of 64 amino acids (p.Ala152_Gln 215del). In the longer variant, a cryptic splice site in the middle of exon 6 is used, predicting an in-frame deletion of 30 amino acids (p.Val186_Gln215del). The transcript lacking the whole exon 6 seems to be normally expressed at low levels as it is seen in controls (Study II Figure 1). The longer variant, however, is not normally present, which implies the functional importance of the deleted 30 amino acids. Interestingly, this region of SIL1 is rich in serine residues. This could point to important phosphorylation sites necessary for normal functioning of SIL1. Furthermore, the putative model of SIL1–GRP78 interaction suggests that the major interaction site is composed of *SIL1* exons 6 and 9 and that exon 10 provides a minor interaction site (Senderek *et al.* 2005).

The c.1030-9G>A mutation was detected in two apparently unrelated families from northern Norway (Figure 9). We confirmed by RT-PCR that the mutation activates a cryptic splice site and leads to the insertion of 7 nucleotides into the mature mRNA (Study III Figure 1). The insertion results in a frameshift at codon 345 and truncation of the protein product after eight novel amino acids (p.Phe345AlafsX9).

8.3.3 Most SIL1 mutations predict premature termination codons

It is noteworthy that 13 of the 17 MSS-associated *SIL1* mutations described so far (Table 8) predict a truncated SIL1 protein likely to make it non-functional, or for the transcript or the protein to be degraded. Eight of these mutations introduce PTCs within the amino acids encoded by the last exon of *SIL1* either through a frameshift or a nonsense change. Since exon-exon junctions in the transcript downstream of the PTC are essential in initiating nonsense-mediated mRNA decay (Maquat 2004), these eight mutations are more likely to produce truncated polypeptides than PTC-causing mutations that are located more upstream.

The four remaining mutations predict either in-frame deleted SIL1 variants due to splice site mutations or a missense change. The pathogenicity of the in-frame deleted SIL1 variants could be explained by incorrect folding and degradation or absence of important functional domains. So far, only one missense mutation has been detected in *SIL1*, c.1370T>C (p.Leu457Pro) changing a leucine just before the ER retrieval signal at position 457 to a cyclic proline (Study III Figures 1 and 2). As proline has a built-in bend in the backbone, this substitution may create a turn in the polypeptide chain (see Section 8.5.2) hiding the ER retrieval signal located in the C-terminus of SIL1.

In addition to the Finnish founder mutation, two mutations, c.331C>T and c.947_948insT, have each been identified in more than one family (Table 8). Recently, the c.331C>T mutation was reported in patients with CMRD and MSS (see Section 8.2) proving that the patients have two recessive diseases due to mutations in two genes, *SAR1B* and *SIL1*, localized on chromosome 5q31.

8.3.4 Genetic heterogeneity in classical MSS

SIL1 mutations are not detected in all MSS patients presenting the main features of cerebellar atrophy and ataxia, early-onset cataracts, myopathy, and mental retardation

(Study III, Senderek *et al.* 2005). Genetic heterogeneity, e.g. disease associated mutations at more than one locus, is a rather common phenomenon in inherited disorders, and is most likely present in MSS too. As cDNA analysis has not been performed for all MSS patients without *SIL1* mutations, we can't rule out the possibility that some of the mutation negative patients may have deleterious changes affecting the intronic regions that would not be detected by analyzing the coding regions of *SIL1* from genomic DNA.

8.4 Expression analysis of SIL1 (II)

Previous Northern analysis of *SIL1* expression in adult human tissues had shown a 1.8-kb transcript with highest levels in secretory organs, a relatively high level in skeletal muscle, and low expression in the brain (Chung *et al.* 2002). Based on database ESTs and RT-PCR experiments (see below), alternative splicing is relatively frequent in *SIL1*. The 3' end of *SIL1* contains two polyadenylation sites. Both of these sites may be used, as supporting EST sequences are found in the AceView database (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html).

8.4.1 Expression of SIL1 in patient muscle tissue

We studied the expression of SIL1 in skeletal muscle biopsies of MSS patients and controls using an antibody detecting a C-terminal epitope "ELLGSVNSLLKELR". In control muscle cross sections, SIL1 immunoreactivity appeared as discrete subsarcolemmal crescents often in close vicinity to nuclei (Study II Figure 1c), which is the location of rough ER in myofibers. In Finnish patients, homozygous for p.Asp170GlufsX4, no definite subsarcolemmal immunostaining was observed, compatible with the predicted truncation of the polypeptide and loss of the epitope. Whether these patients express a truncated protein remains unknown. In the compound heterozygous Swedish patients (M203 and M204) with two abnormally sized in-frame deleted variants, p.Ala152_Gln 215del and p.Val186_Gln215del, the immunostaining was similar to the controls, indicating that the in-frame deleted variant(s) from the splice site mutation allele are translated to proteins. Compatible with the loss-of-function effect of this mutation, it is thus likely that the exon 6 encoded amino acids are critical for normal SIL1 function. Although the exon 6 in-frame deleted variant is present at low levels in control individuals, the total loss of normal variant seems to be pathogenic, which is also supported by the finding of another splice site mutation leading to skipping of exon 6 (Senderek *et al.* 2005). Interestingly, there was no evidence of SIL1 positive aggregate formation in the muscle expressing the exon 6 in-frame deleted variants, contrary to the findings in the *wz* mouse mutant and overexpressed mutant SIL1 proteins in COS-1 cells (see Section 8.5). It thus remains possible that the in-frame deleted variant is stable and nonaggregating, but nonfunctional due to conformational changes.

8.4.2 Expression of SIL1 in human tissues

We investigated *SIL1* expression in human fetal and adult tissues with emphasis on the brain. Northern analysis showed *SIL1* expression in all adult brain regions, but not in fetal brain (Study II Supplementary Figure 4a). RT-PCR analysis, however, revealed *SIL1* expression also in fetal brain indicating lower expression in developing brain (Study II Supplementary Figure 4b). Since SIL1 acts as an adenine nucleotide exchange factor for

GRP78 (see Section 5.6.3), we wanted to see whether the expression pattern of these two proteins and their respective genes, *SIL1* and *HSPA5*, overlap. According to the Human Gene Expression Atlas 2 database, *SIL1* and *HSPA5* to have highly similar tissue expression, with highest expression in liver, pancreatic islets, lung, prostate, thyroid, and placenta, and relatively low expression in neuronal tissues. Searching for the NEIBank Expressed Sequence Tags showed *SIL1* and *HSPA5* expression in fetal whole eye and in adult human lens, retinal pigment epithelium, and choroid.

8.4.3 Expression of SIL1 in mouse tissues

Given the cerebellar atrophy, congenital cataract, and progressive myopathy of MSS patients, we investigated the temporal and spatial expression of SIL1 and GRP78 (denoted as Hspa5 in Study II Supplementary Figure 5) in mouse tissues. In fetal tissues (embryonal day 12.5 to 18.5), SIL1 and GRP78 exhibit similar expression patterns in cells of developing retina, lens, and brain (Study II Supplementary Figure 5). From postnatal day 5 onwards both SIL1 and GRP78 are expressed in cortical neurons especially on layers I-IV throughout the brain, in cerebellar Purkinje and granule cells, and in hippocampal neurons (Study II Supplementary Figure 5). In hippocampus, SIL1 expression was somewhat stronger in the pyramidal neurons of the CA3 region, while GRP78 was more uniformly expressed in all regions. In addition to the CNS, similar expression of SIL1 and GRP78 was seen in several other tissues including muscle, developing bone, cartilage, and gonads (data not shown).

8.4.4 SIL1 tissue expression patterns and manifestations of MSS

The tissue and subcellular expression patterns of *SIL1*/SIL1 and *HSPA5*/GRP78 are highly similar, suggesting related functions *in vivo* for the two proteins, in agreement with their demonstrated interaction in cellular models (Boisrame *et al.* 1998; Tyson *et al.* 2000; Chung *et al.* 2002). Interestingly, no clinical symptoms have been reported from tissues showing the highest *SIL1* expression levels such as liver and pancreas, suggesting that certain cells and tissues, including cerebellar neurons, skeletal muscle and lens cells, lack compensatory mechanisms or are more sensitive to disturbed cellular homeostasis. It is possible that other nucleotide exchange factors can compensate for the loss of SIL1 in the non-vulnerable tissues. In line with this prediction, Sil1p is non-essential for viability in yeast (Tyson *et al.* 2000), where Lhs1p acts as a nucleotide exchange factor for Kar2p in a mutually exclusive manner to Sil1p (Steel *et al.* 2004), accounting for the lethality of Δlhs1Δsil1 double mutants (Tyson *et al.* 2000). The human *LHS1* homologue, *HYOU1* has a tissue expression pattern partially different from that of *SIL1*, e.g. *HYOU1* is almost absent from skeletal muscle, and the HYOU1 protein might thus compensate for SIL1 in some tissues.

It is noteworthy that the distribution of affected tissues in MSS shows similarities with the tissue distribution in some other disorders. The group of O-mannosylation disorders, Walker-Warburg syndrome (WWS), Fukuyama congenital muscular dystrophy (FCMD), and muscle-eye-brain disease (MEB), are similar to MSS with respect to the main affected tissues including brain, muscle, and eye (van Reeuwijk *et al.* 2005a). Interestingly, yeast Sil1 has been suggested to have a genetic interaction with Pmt2 (Schuldiner *et al.* 2005), which corresponds to the human

POMT2 gene. *POMT2* together with *POMT1* is responsible for the protein O-mannosyltransferase enzyme activity, which is defective in a subset of patient with WWS (van Reeuwijk *et al.* 2005b). It remains to be seen whether SIL1 has a more specific role related to processing of O-mannosylated proteins.

8.5 Subcellular localization of SIL1 and its mutants (III)

Previously, exogenously expressed SIL1 had been demonstrated to localize to the ER in transfected COS-1 cells (Chung *et al.* 2002). This finding is in accordance with the known SIL1 function as a nucleotide exchange factor for GRP78 and the presence of N-terminal ER targeting and C-terminal ER retrieval sequences in SIL1. Just prior to the publication of *SIL1* mutations underlying MSS, a truncation of *Sil1* was shown to cause ataxic phenotype in *wz* mutant mouse (Zhao *et al.* 2005) (see also Section 5.6.3). Interestingly, the affected *wz* mutant Purkinje cells showed ubiquitin-positive protein accumulations within the ER and the nucleus.

8.5.1 SIL1 localizes to the endoplasmic reticulum in cultured mouse neurons

We studied whether endogenous SIL1 localizes to the ER in cultured mouse primary hippocampal neurons. Using immunofluorescence analysis, we found SIL1 to predominantly localize to the ER, which was indicated by colocalization with the ER marker PDI (Study III Figure 4). SIL1 also colocalizes with GRP78 in neurons, however, no colocalization was detected with a Golgi marker (data not shown). This finding was well in line with previous localization studies in transiently transfected COS-cells (Chung *et al.* 2002; Zhao *et al.* 2005).

8.5.2 Subcellular localization of the wild-type and mutant SIL1 proteins

We chose two mutations for the subcellular localization studies. The missense substitution p.Leu457Pro detected in patients changes a leucine to a cyclic proline and may thus influence the tertiary structure of SIL1 by creating a turn in polypeptide and hiding the assumed ER retrieval tetrapeptide present at the very end of the protein (Figure 3). Since several mutations creating a premature stop codon in exon 10 have been described (Table 8), we decided to create an artificial mutation deleting only the predicted ER retrieval signal at the very end of the protein product (p.Lys458X).

We transiently transfected COS-1 cells with wild-type and mutant SIL1 with HA-tags in order to find out whether mutant SIL1 had normal subcellular localization. The exogenous wild-type SIL1 protein colocalized with both the ER and the Golgi apparatus markers (Study III Figure 3) in our experiments. Given the ER localization of the endogenously expressed SIL1, overproduction of the protein may lead to its escape from the ER into the Golgi apparatus, a phenomenon that has been seen with some other proteins possessing weaker ER retrieval signals (Lewis *et al.* 1992).

We hypothesized that the missense mutation (p.Leu457Pro) and the mutation deleting the ER retrieval motif (p.Lys458X) would lead to defective retrieval of the mutant protein from later stages of the secretory pathway. Both mutant proteins clearly showed altered localization compared to wild-type SIL1: in cells with higher expression levels, mutant proteins formed aggregates that showed ER localization while in cells with lower

expression levels mutant SIL1 was present in the ER without aggregate formation (Study III Figure 3). The aggregates showed strongest overlap with GRP78 and PDI markers and they did not overlap with any of the ER-Golgi intermediate compartment, Golgi apparatus, or lysosome markers (Study III Figure 3 and data not shown). In some cells, GRP78 staining was more pronounced in the aggregates, suggesting that it is present in these structures as would be expected knowing its function in targeting misfolded proteins to degradation (see Section 5.6.2).

To summarize, the pathogenicity of the Leu457Pro and Lys458X mutant proteins is explained both by abnormal subcellular localization and by structural changes making them prone to aggregate formation. The same mechanisms could occur with other mutations leading to truncation of the SIL1 protein within exon 10. It is interesting to note that the *wz* mutant mouse also shows protein accumulations prior to cerebellar Purkinje cell degeneration. The accumulations are ubiquitinated and show positive immunostaining for several ER chaperones including GRP78 and GRP94. However, the biochemical composition of the aggregates present in *wz* Purkinje cells remains unknown (Zhao *et al.* 2005).

8.6 Exclusion of functional candidate genes in MSS (III)

As previously discussed, MSS is genetically heterogeneous and *SIL1* mutations are not found in all patients with typical clinical findings. Given the functional interaction between SIL1 and GRP78 and the fact that HYOU1 can substitute for SIL1 function, we chose to study the respective genes, *HSPA5* and *HYOU1*, for mutations in patients without *SIL1* mutations. *Hspa5* and *Hyou1* mouse knockout models show embryonic lethality (Ni *et al.* 2007), but we reasoned that other than loss-of-function mutations could play a role in MSS. Subsequently, the *AARS* gene was also selected as a possible functional candidate because of the similarity between the cerebellar phenotypes of the *Sil1* and *Aars* mouse models (Zhao *et al.* 2005; Lee *et al.* 2006). Analyzing a panel of 18 patients with MSS, we found no disease-causing alterations within the coding regions of *HSPA5*, *HYOU1*, and *AARS*. A possibility remains that in a subset of patients a mutation is located in the intronic regions of these genes and was therefore not identified by sequencing genomic DNA. Nevertheless, according to our data, *HSPA5*, *HYOU1*, and *AARS* are unlikely to be major genes underlying MSS.

8.7 Clinical findings in patients with *SIL1* mutations (I – III, unpublished)

The approximately 200 patients with MSS reported so far present marked variability in their clinical features. Identification of *SIL1* as the gene underlying MSS is the first step towards the classification of MSS and related disorders likely to be clinically and genetically heterogeneous. Moreover, the molecular genetic characterization of *SIL1* mutations underlying classical MSS allows definition of more precise diagnostic criteria. Even with the current data, the patients may be classified into two groups. In classical MSS all the hallmark features are present as well as a variety of additional features. In the MSS-like group some, but not all, of the cardinal features are present together with several atypical symptoms. In some patients clear distinction cannot be made; one hallmark feature may not exist or some atypical symptoms may be present. The boundaries between the two groups are likely to change in the future as the number of *SIL1* mutation positive cases grows and additional genes

underlying MSS are described.

In the present study, detailed clinical information was available from 29 patients with MSS and proven *SIL1* mutations. In addition, clinical information was available on patient M705, the sister of M704, for whom no DNA analysis was performed (Table 9). All these 30 patients had classical MSS. Although more thorough clinical characterization of this group is needed, some notes about the classical MSS features can be made. Cerebellar atrophy was present in all the patients for whom brain imaging was performed. With

Table 9. Comparison between the clinical features of *SIL1* mutation positive patients with MSS and patients described earlier in the literature.

	Patients with *SIL1* mutations 30 cases	(Slavotinek *et al.* 2005) 75 cases
Cerebellar signs		
Truncal / limb ataxia	100% (29/29)	72%
Dysarthria	100% (6/6)	45%
Nystagmus	69% (20/29)	36%
Intention tremor		16%
Cerebellar atrophy	100% (26/26)	
Ophthalmological features		
Cataracts	100% (30/30)	91%
Strabismus	89% (24/27)	44%
Other neurological signs		
Truncal / limb hypotonia	100% (29/29)	53%
Muscle atrophy		51%
Muscle weakness	100% (9/9)	36%
Myopathic changes	92% (23/25)	
Mental retardation (mild to profound)	100% (30/30)	
Ortopedic manifestations	62% (16/26)	
Pes planus / planovalgus		28%
Genu valgum		20%
Scoliosis	69% (6/9)	19%
Kyphoscoliosis		16%
Joint contractures		17%
Pectus deformities		9%
Endocrinological signs		
Hypergonadotropic hypogonadism	90% (19/21)	
Growth		
Short stature	60% (18/30)	
Microcephaly	7% (2/28)	

brain MRI, cerebellar atrophy was more pronounced in the vermis than in the hemispheres (Study II Supplementary Figure 1). Patient M603 showed a T2-hyperintense cerebellar cortex (data not shown), which was recently reported as one neuroradiological feature of MSS (Harting *et al.* 2004). Ataxia and hypotonia were noted in all patients.

In study II, cataracts were diagnosed in all 21 patients after the age of 2.5 years and in some patients a rapid progression of cataracts was evident. Prior to development of bilateral cataracts, patients M203, M204 and M503 (families M2 and M5) underwent eye examination with normal results. Cataracts appeared at six years of age in M1603 (family M16) and at 4.5 and 6.5 years of age in patients M1105 and M1106 (family M11). Eye examination 3-6 months before the onset of cataracts revealed microcrystalline opacities in only one of the patients.

In study I, we suggested that the progressive myopathy characterized by marked muscle atrophy together with fat and connective tissue replacement is a hallmark feature of MSS. The duration of the disease does not alone explain the variation in the severity of the muscle involvement. Moreover, studying nine Finnish patients homozygous for the same *SIL1* mutation we were able to show that the severity of the muscle symptoms can vary from moderate to severe. Mental retardation is another symptom with remarkable variation in severity and most patients are mildly or moderately affected.

The additional features often present in classical MSS include hypergonadotropic hypogonadism, orthopedic manifestations, short stature, and strabismus. Many of the additional symptoms seemed to be more frequent in our patient series than in previous reports. For example, skeletal abnormalities were present in 62% of the patients with *SIL1* mutations. Although no definite analysis was possible due to the small number of cases in study I, a tendency of more severe scoliosis and longer disease duration together with more pronounced myopathy was noted.

According to the available data, no clear correlation exists between the site of the *SIL1* mutation and the clinical phenotype of patients with MSS. Only one feature shows putative correlation with the location of the mutation or mutation type. The two Japanese patients (M2103 and M2104) with the missense mutation p.Leu457Pro in exon 10 have normal levels of FSH, LH, and testosterone, and show no sign of hypogonadism. Interestingly, there is a third patient without hypergonadotropic hypogonadism and he is a compound heterozygote for a mutation in exon 10 (p.Leu456fsX2) and another frameshift (p.Leu316fsX36) (Senderek *et al.* 2005). On the contrary, in a family described by Karim *et al.* where the mutation is located more upstream (p.Gln438X), two of the cases show either primary amenorrhea or hypogonadism (Karim *et al.* 2006). In the case of the p.Leu457Pro mutation, the lack of gonad phenotype is not easily explained given the fact that the missense mutant protein seemed to be prone to aggregation (see Section 8.5.2). One possible explanation is that the terminal mutations prone to aggregation maintain some level of function, which would be enough to prevent gonadal dysfunction. Another possibility is that other genetic or environmental factors, like febrile illnesses, play a role in the pathogenesis. It remains to be seen whether the other patients with C-terminal mutations or in-frame deletions have hypergonadotropic hypogonadism.

In study III we did not identify *SIL1* mutations in all patients fulfilling the clinical diagnosis showing that classical MSS is heterogeneous. Moreover, no *SIL1* mutations were identified in patients belonging to the MSS-like group

and lacking more than one hallmark feature as well as having several atypical features. Our findings are well in line with previous reports noting heterogeneity in classical MSS (Senderek *et al.* 2005) and lack of *SIL1* mutations or linkage to 5q31 in the atypical, MSS-like group (Lagier-Tourenne *et al.* 2003; Schulz *et al.* 2007).

In conclusion, this study demonstrates that *SIL1* mutations are associated with the classical MSS phenotype including cerebellar atrophy and ataxia, cataracts, myopathy, and psychomotor delay. The clinical symptoms of patients with *SIL1* mutations are rather similar (Table 9) and they form a clinically definable subgroup among the heterogeneous group of MSS patients described earlier in the literature. The most evident differences are that hypotonia, skeletal abnormalities, and hypergonadotropic hypogonadism are more frequent features of MSS than previously reported. At odds with some previous reports, the cataracts may present in early childhood rather than being congenital. This is important to notice when considering the differential diagnosis for other disorders resembling MSS.

8.8 MSS – a protein misfolding and accumulation disease?

Several pieces of information suggest that MSS pathogenesis is linked to abnormal protein quality control in the ER. Chaperone GRP78 has multiple roles in processing proteins targeted to the secretory pathway (Hendershot 2004). As cochaperone, SIL1 regulates the ATPase cycle of GRP78. Most of the *SIL1* mutations described so far are predicted to lead to nonfunctional or absent SIL1 and are thus likely to result in decreased rates of GRP78 hydrolysis and abnormal protein translocation and folding (Figure 10). The accumulation of unfolded proteins in the ER causes ER stress, and if stressful conditions are not alleviated this may lead to cell death (Ma *et al.* 2004). In addition to the prior pathogenetic mechanisms, our results in the mutant SIL1 expression studies imply that missense changes and some truncations in the SIL1 polypeptide may lead to conformational alterations making the polypeptide prone to aggregation and thus nonfunctional.

The hypothesis of defective protein folding in MSS is supported by data from the *wz* mouse. In the *wz* Purkinje cells loss-of-function of SIL1 causes induction of the unfolded protein response and abnormal accumulation of proteins, which is followed by neurodegeneration (Zhao *et al.* 2005). The accumulations in *wz* Purkinje cells contain several ER chaperones, including GRP78 and GRP94, but it remains to be studied whether truncated SIL1 could also be present. Another interesting aspect is the survival of a subset of cerebellar Purkinje cells in MSS and in the *wz* mouse mutant. The vestibulocerebellum, in which the Purkinje cells do not degenerate, is the phylogenetically oldest part of the cerebellum that receives input from the vestibular nuclei in the brainstem. The Purkinje cells in the vestibulocerebellum must therefore have alternative chaperones to SIL1. Establishing the differences between the Purkinje cell populations may help to clarify the pathogenesis of MSS and give rise to future treatment options.

MSS may thus be added to the growing number of protein misfolding and accumulation disorders due to defective molecular chaperone function (Barral *et al.* 2004; Muchowski *et al.* 2005). Mutations in other chaperone encoding genes have been associated with hereditary cataracts (Litt *et al.* 1998) and neuromuscular disorders (Vicart *et al.* 1998; Hansen *et al.* 2002; Selcen *et al.* 2003; Evgrafov *et al.* 2004; Irobi *et al.* 2004). The autosomal recessively inherited MSS is the first human disorder linked to the most highly

conserved HSP70 chaperones and, unlike many chaperone-associated disorders, it displays a wide spectrum of symptoms including the involvement of the central nervous system, skeletal muscle, lens, bone, and the gonads.

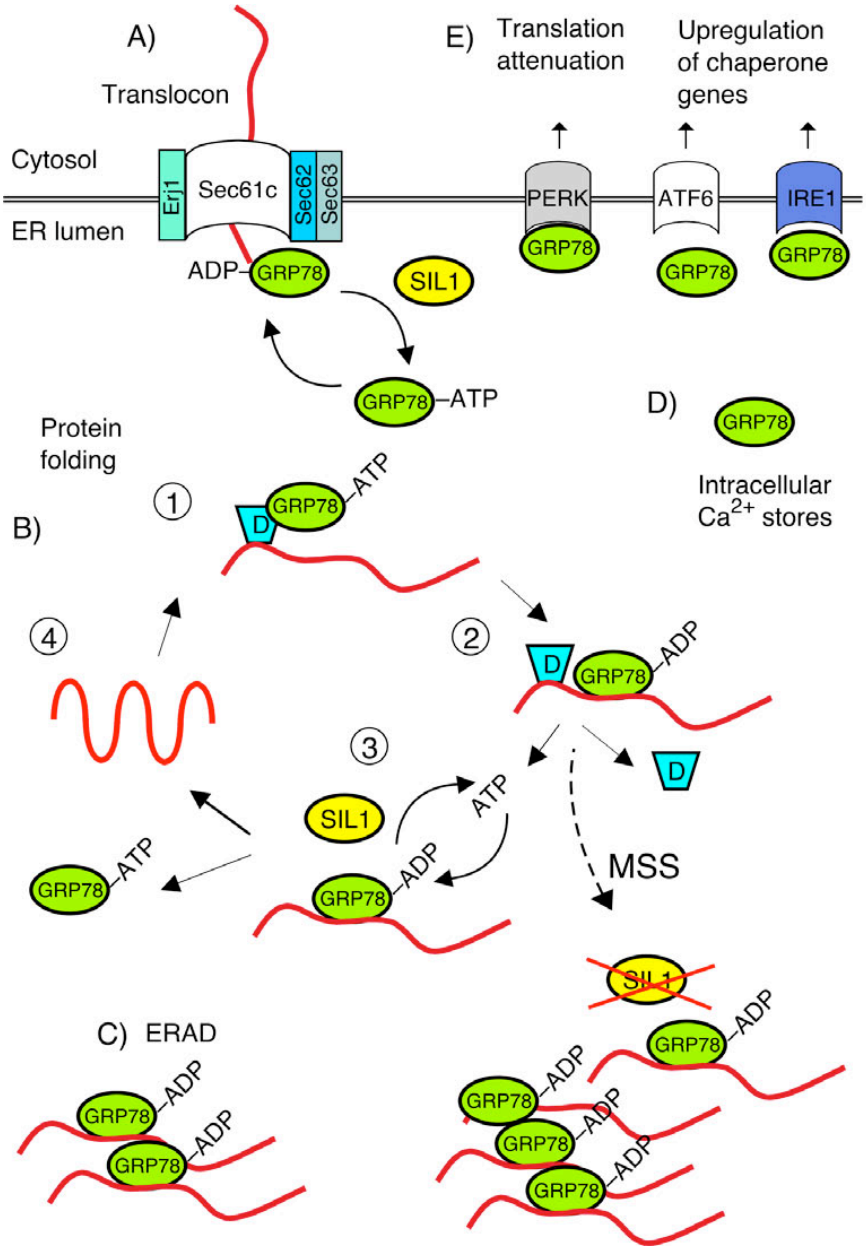

Figure 10. Role of SIL1 in the pathogenesis of MSS. In normal cells the ATPase cycle of GRP78 regulates its binding and release from nascent protein chains as is shown in Figure 2 (see Chapter 5.6.2). Nucleotide exchange factor SIL1 catalyzes the release of ADP allowing the rebinding of ATP on GRP78 (3). In cells lacking SIL1 the ATPase cycle of GRP78 is disrupted (MSS). GRP78 can still bind and lock onto the unfolded substrates, but without SIL1 GRP78 fails to be released and recycled, which may lead to accumulation of unfolded substrates.

9 CONCLUDING REMARKS AND FUTURE PROSPECTS

In this thesis, the identification and characterization of *SIL1*, the first and most common gene underlying MSS, is described. This gives a basis for the molecular genetic diagnostics of MSS. However, future work is needed to give a more comprehensive view of the clinical findings in patients with *SIL1* mutations and establishing diagnostic criteria for classical MSS and MSS-like syndromes.

Most of the *SIL1* mutations underlying MSS result in PTCs and are predicted to lead to absence of the SIL1 protein due to mRNA decay. A few mutations are assumed to lead to absence of functional SIL1 due to loss of important functional domains, e.g. the in-frame deletions of exons 6 and 9, or due to aggregate formation, which was seen with the p.Leu457Pro missense mutation. In addition, aggregate formation leading to nonfunctional SIL1 is predicted to be the pathogenetic mechanism of the mutations leading to PTC within the last exon of *SIL1*.

Not all patients with a clinical MSS diagnosis have detectable mutations in the *SIL1* gene implying genetic heterogeneity. Functional candidate gene analysis failed to identify additional genes underlying MSS when focusing on the close chaperone partners of SIL1, namely GRP78 and HYOU1, and the similarity of the cerebellar phenotype of the *Aars* mouse model. Further analysis of potential functional candidate genes together with identification and genome wide analysis of extended or consanguineous families without *SIL1* mutations is needed to characterize other genes underlying MSS. The characterization of other MSS genes may couple them with *SIL1* function and elucidate the defective pathway.

Although SIL1 is a protein with established nucleotide exchange factor function, little is known about its role in multiple tasks carried out by GRP78. As alternative nucleotide exchange factors exist in the ER, SIL1 could participate in a subset of GRP78 tasks or be substrate specific. The experimental work done on model organisms such as yeast may give some clues to the basic biochemical pathways that are defective in MSS and possibly link MSS to certain protein modification processes, such as O-mannosylation. More importantly, establishing the differences between the degenerating and surviving *wz* mouse Purkinje cell populations may help to clarify the pathogenesis of MSS. Although the *wz* mice do not show symptoms in other MSS-associated organs, the animal model is a good basis for studying future treatment options in more detail.

The relevance of accumulating proteins in the pathogenesis of MSS needs clarification. On one hand the mutant SIL1 proteins are prone to aggregation and on the other hand the absence of SIL1 leads to accumulation of as yet unknown proteins in the *wz* mouse. Furthermore, only some human cells and tissues seem to be vulnerable to the loss-of-function of SIL1. Unraveling the interaction partners and potential substrates of SIL1 as well as defective pathways in MSS may also give some basic information about neuronal function and survival, and aid in the understanding of other protein folding diseases.

10 ACKNOWLEDGEMENTS

This thesis work was carried out at the Department of Medical Genetics, Folkhälsan Institute of Genetics, and Neuroscience Center during 2003-2008. The former and current heads of the department, Professors Leena Palotie, Kristiina Aittomäki, Päivi Peltomäki, Albert de la Chapelle, Anna-Elina Lehesjoki, and Heikki Rauvala, are warmly thanked for providing excellent research facilities. The Helsinki Biomedical Graduate School is acknowledged for giving me the opportunity to start scientific work side by side with the Medical school and for financial support. This study was supported by Biomedicum Helsinki Foundation, Emil Aaltonen Foundation, Finnish Medical Foundation, and Neurologiasäätiö.

I am deeply grateful to my supervisor, Professor Anna-Elina Lehesjoki, for introducing me to the interesting field of neurogenetics. Her encouraging and kind guidance has meant tremendously much to me during these years.

Docent Tiina Paunio and Docent Helena Pihko, the official pre-examiners of my thesis, are warmly thanked for their valuable comments and kind advice. I would like to express my gratitude to my thesis committee members, Docents Tiina Paunio and Pentti Tienari, for having time to evaluate my work, and their positive and supportive attitude when the thesis project seemed like it wasn't proceeding.

I am grateful to Docent Ibrahim Mahjneh for being the clinical expert in this project. I especially admire his enthusiasm towards all aspects of MSS. Docent Mirja Somer is warmly thanked for her kind help in the clinical analyses as well as suggesting this interesting project to me. Docent Hannu Somer is thanked for his indispensable contribution in the beginning of this project. Professor Hannu Kalimo, Docent Anders Paetau, and Docent Bjarne Udd are thanked for always finding time for my questions in their tight schedules and sharing their vast knowledge in the field of muscle disorders. Professor Markku Heikinheimo is thanked for his valuable help and advice. Docent Riitta Herva and Dr. Aune Hirvasniemi are warmly acknowledged for setting the foundation for the present study.

I am most indebted to Riikka Hämäläinen for her friendship and advice during this project. I would like to extend my warmest thanks to Eija Siintola for her skillful help and inspiring discussions regarding this and other projects. Outi Kopra is especially thanked for her efficient contribution and admirable expertise. I wish to thank Tarja Joensuu for giving a helping hand in the busiest time of this project and for having solutions to almost every lab-related problem! I own my deepest thanks to Laura Waris for her help in the beginning of this project.

My foreign co-authors, Lisbeth Tranebjaerg, Clotilde Lagier-Tourenne, Orvar Eeg-Olofsson, Michel Koenig, Emilia Bijlsma, Denys Chaigne, Nobue Iwata, Hiroyuki Meguro, Yaeko Ichikawa, and Jun Goto are thanked for being extremely helpful and pleasant to collaborate with.

I am grateful to Hanna Olanne for excellent help in the lab and to Sinikka Lindh for tracing the ancestry of the patients and keeping track of all the samples. Ann-Liz Träskelin is thanked for her help as well as for being such a joyful and down to earth person keeping spirits up every day. Teija Toivonen, Taru Jokinen, Paula Hakala, and Mira Aronen

are warmly thanked for their expert technical assitance. Aila Riikonen, Minna Maunula, Maija Puhakka, and Jaana Welin-Haapamäki are warmly thanked for secretarial help. The expertise of Leena Toivanen saved me a lot of time while trying to locate older references.

Jodie Painter is thanked for carefully editing the language of this thesis as well as for enjoyable conversations and sharing the frustration caused by linkage programs.

Since I first started working at FIG in 1997, I have had many friends and colleagues to whom I own my deepest thanks. I especially wish to thank Kati Donner, Jukka Kallijärvi, Maria Lehtinen, and Kirsi Alakurtti for discussions and help concerning this project. My long-time roommates at FIG Juha Isosomppi, Anna Naukkarinen, Saara Tegelberg, Hanna Västinsalo, and Maria Kousi are thanked for unforgettable company: I'm proud to be the "mentally oldest" member of our late office team.

I warmly thank all my colleagues at the Department of Clinical Genetics and Department of Medical Genetics. I wish to express my gratitude to Docent Kristiina Aittomäki for sharing her expertise and giving me the opportunity to learn new things in her department. Docents Minna Pöyhönen and Hannele Laivuori are thanked for their sincere help and shared interest in teaching.

I wish to express my gratitude to persons involved in my past thesis projects, especially Docent Mirja Somer and Dr. Johanna Hästbacka in the PEHO project, as well as Docent Eija Gaily, Dr. Elina Liukkonen, and Dr. Auli Sirén in the epilepsy project.

My dear friends Laura Tahkokallio, Henriikka Utriainen, Maija Seppänen, and Riikka Mäkinen are thanked for their friendship and sharing similar views of life. Although we do not see each other that often, our long friendship means a lot to me.

I am grateful to my friends sharing the paranormal class in Medical school. Without you the studies and the life after would have been less fun! I own my deepest thanks to the "girls gang" for the interesting conversations during our restaurant excursions and to the "whole gang" for sharing various events together. Pamela Hellevuo, Jaana Karhunen, Katri Aro, Satu Heliövaara-Peippo, and Maria Halonen are thanked for their joyful company during our maternity leaves. Riina Oksjoki is warmly thanked for sharing her positive attitude and the ups and downs of a basic research PhD.

I own my deepest gratitude to my parents Riitta and Reino Pusa for their love, help, and never-ending faith in me. My sister Eevis is thanked for always supporting me and being an admirably kind and positive person. My parents in-law Kirsti and Esa Anttonen are thanked for enjoyable company and giving their help in baby-sitting.

And lastly, I am most grateful to Mikko, Emil, and Eljas for bringing so much love into my life. Without Mikko's help, faith and encouragement I could have never finished this work. Emil's amazing observations and Eljas' smile make every day worth living.

Espoo, January 2008

11 REFERENCES

Aguglia U, Annesi G, Pasquinelli G, Spadafora P, Gambardella A, Annesi F, Pasqua AA, Cavalcanti F, Crescibene L, Bagala A, Bono F, Oliveri RL, Valentino P, Zappia M and Quattrone A (2000). Vitamin E deficiency due to chylomicron retention disease in Marinesco-Sjögren syndrome. *Ann Neurol* 47(2): 260-4.

Alexianu M, Christodorescu D, Vasilescu C, Dan A, Petrovici A, Magureanu S and Savu C (1983). Sensorimotor neuropathy in a patient with Marinesco-Sjogren syndrome. *Eur Neurol* 22(3): 222-6.

Alter M and Kennedy W (1968). The Marinesco-Sjogren syndrome. Hereditary cerebello-lental degeneration with mental retardation. *Minn Med* 51(7): 901-6.

Alter M, Talbert OR and Croffead G (1962). Cerebellar ataxia, congenital cataracts, and retarded somatic and mental maturation. Report of cases of Marinesco-Sjogren syndrome. *Neurology* 12: 836-47.

Andersen B (1965). Marinesco-Sjoegren Syndrome: Spinocerebellar Ataxia, Congenital Cataract, Somatic and Mental Retardation. *Dev Med Child Neurol* 47: 249-57.

Angelicheva D, Turnev I, Dye D, Chandler D, Thomas PK and Kalaydjieva L (1999). Congenital cataracts facial dysmorphism neuropathy (CCFDN) syndrome: a novel developmental disorder in Gypsies maps to 18qter. *Eur J Hum Genet* 7(5): 560-6.

Annesi G, Aguglia U, Tarantino P, Annesi F, De Marco EV, Civitelli D, Torroni A and Quattrone A (2007). SIL1 and SARA2 mutations in Marinesco-Sjogren and chylomicron retention diseases. *Clin Genet* 71(3): 288-9.

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, Grieder FB, Heintz N, Hicks G, Insel TR, Joyner A, Koller BH, Lloyd KC, Magnuson T, Moore MW, Nagy A, Pollock JD, Roses AD, Sands AT, Seed B, Skarnes WC, Snoddy J, Soriano P, Stewart DJ, Stewart F, Stillman B, Varmus H, Varticovski L, Verma IM, Vogt TF, von Melchner H, Witkowski J, Woychik RP, Wurst W, Yancopoulos GD, Young SG and Zambrowicz B (2004). The knockout mouse project. *Nat Genet* 36(9): 921-4.

Barral JM, Broadley SA, Schaffar G and Hartl FU (2004). Roles of molecular chaperones in protein misfolding diseases. *Semin Cell Dev Biol* 15(1): 17-29.

Bassoe HH (1956). Familial congenital muscular dystrophy with gonadal dysgenesis. *J Clin Endocrinol Metab* 16(12): 1614-21.

Begeer JH, Scholte FA and van Essen AJ (1991). Two sisters with mental retardation, cataract, ataxia, progressive hearing loss, and polyneuropathy. *J Med Genet* 28(12): 884-5.

Berry V, Francis P, Reddy MA, Collyer D, Vithana E, MacKay I, Dawson G, Carey AH, Moore A, Bhattacharya SS and Quinlan RA (2001). Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. *Am J Hum Genet* 69(5): 1141-5.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermuller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB,

Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B and de Jong PJ (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146): 799-816.

Bodrug SE, Ray PN, Gonzalez IL, Schmickel RD, Sylvester JE and Worton RG (1987). Molecular analysis of a constitutional X-autosome translocation in a female with muscular dystrophy. *Science* 237(4822): 1620-4.

Boisrame A, Kabani M, Beckerich JM, Hartmann E and Gaillardin C (1998). Interaction of Kar2p and Sls1p is required for efficient co-translational translocation of secreted proteins in the yeast Yarrowia lipolytica. *J Biol Chem* 273(47): 30903-8.

Bouchard JP, Richter A, Mathieu J, Brunet D, Hudson TJ, Morgan K and Melancon SB (1998). Autosomal recessive spastic ataxia of Charlevoix-Saguenay. *Neuromuscul Disord* 8(7): 474-9.

Boycott KM, Flavelle S, Bureau A, Glass HC, Fujiwara TM, Wirrell E, Davey K, Chudley AE, Scott JN, McLeod DR and Parboosingh JS (2005). Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification. *Am J Hum Genet* 77(3): 477-83.

Breedveld GJ, van Wetten B, te Raa GD, Brusse E, van Swieten JC, Oostra BA and Maat-Kievit JA (2004). A new locus for a childhood onset, slowly progressive autosomal recessive spinocerebellar ataxia maps to chromosome 11p15. *J Med Genet* 41(11): 858-66.

Breuning MH, Dauwerse HG, Fugazza G, Saris JJ, Spruit L, Wijnen H, Tommerup N, van der Hagen CB, Imaizumi K, Kuroki Y, van den Boogaard MJ, de Pater JM, Mariman EC, Hamel BC, Himmelbauer H, Frischauf AM, Stallings R, Beverstock GC, van Ommen GJ and Hennekam RC (1993). Rubinstein-Taybi syndrome caused by submicroscopic deletions within 16p13.3. *Am J Hum Genet* 52(2): 249-54.

Brogdon BG, Snow RD and Williams JP (1996). Skeletal findings in Marinesco-Sjogren syndrome. *Skeletal Radiol* 25(5): 461-5.

Broman KW, Murray JC, Sheffield VC, White RL and Weber JL (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63(3): 861-9.

Bromberg MB, Junck L, Gebarski SS, McLean MJ and Gilman S (1990). The Marinesco-Sjogren syndrome examined by computed tomography, magnetic resonance, and 18F-2-fluoro-2-deoxy-D-glucose and positron emission tomography. *Arch Neurol* 47(11): 1239-42.

Brooks DA (1999). Introduction: molecular chaperones of the ER: their role in protein folding and genetic disease. *Semin Cell Dev Biol* 10(5): 441-2.

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M and Pandolfo M (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271(5254): 1423-7.

Canman CE, Lim DS, Cimprich KA, Taya Y, Tamai K, Sakaguchi K, Appella E, Kastan MB and Siliciano JD (1998). Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* 281(5383): 1677-9.

Cartegni L, Chew SL and Krainer AR (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3(4): 285-98.

Chaco J (1969). Marinesco-Sjogren syndrome with myopathy. *Confin Neurol* 31(6): 349-51.

Chiang AP, Beck JS, Yen HJ, Tayeh MK, Scheetz TE, Swiderski RE, Nishimura DY, Braun TA, Kim KY, Huang J, Elbedour K, Carmi R, Slusarski DC, Casavant TL, Stone EM and Sheffield VC (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc Natl Acad Sci U S A* 103(16): 6287-92.

Chung KT, Shen Y and Hendershot LM (2002). BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. *J Biol Chem* 277(49): 47557-63.

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

Collins FS (1995). Positional cloning moves from perditional to traditional. *Nat Genet* 9(4): 347-50.

Collins FS, Morgan M and Patrinos A (2003). The Human Genome Project: lessons from large-scale biology. *Science* 300(5617): 286-90.

Condo I, Ventura N, Malisan F, Tomassini B and Testi R (2006). A pool of extramitochondrial frataxin that promotes cell survival. *J Biol Chem* 281(24): 16750-6.

Coop G and Przeworski M (2007). An evolutionary view of human recombination. *Nat Rev Genet* 8(1): 23-34.

Davey KM, Parboosingh JS, McLeod DR, Chan A, Casey R, Ferreira P, Snyder FF, Bridge PJ and Bernier FP (2006). Mutation of DNAJC19, a human homologue of yeast inner mitochondrial membrane co-chaperones, causes DCMA syndrome, a novel autosomal recessive Barth syndrome-like condition. *J Med Genet* 43(5): 385-93.

Davis RE, Swiderski RE, Rahmouni K, Nishimura DY, Mullins RF, Agassandian K, Philp AR, Searby CC, Andrews MP, Thompson S, Berry CJ, Thedens DR, Yang B, Weiss RM, Cassell MD, Stone EM and Sheffield VC (2007). A knockin mouse model of the Bardet-Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. *Proc Natl Acad Sci U S A* 104(49): 19422-7.

de Brouwer AP, Williams KL, Duley JA, van Kuilenburg AB, Nabuurs SB, Egmont-Petersen M, Lugtenberg D, Zoetekouw L, Banning MJ, Roeffen M, Hamel BC, Weaving L, Ouvrier RA, Donald JA, Wevers RA, Christodoulou J and van Bokhoven H (2007). Arts syndrome is caused by loss-of-function mutations in PRPS1. *Am J Hum Genet* 81(3): 507-18.

de la Chapelle A and Wright FA (1998). Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc Natl Acad Sci U S A* 95(21): 12416-23.

Delatycki MB, Williamson R and Forrest SM (2000). Friedreich ataxia: an overview. *J Med Genet* 37(1): 1-8.

Dotti MT, Bardelli AM, De Stefano N, Federico A, Malandrini A, Vanni M and Guazzi GC (1993). Optic atrophy in Marinesco-Sjogren syndrome: an additional ocular feature. Report of three cases in two families. *Ophthalmic Paediatr Genet* 14(1): 5-7.

Durr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, Mandel JL, Brice A and Koenig M (1996). Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med* 335(16): 1169-75.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5(6): 435-45.

Ellgaard L and Helenius A (2003). Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* 4(3): 181-91.

Engert JC, Berube P, Mercier J, Dore C, Lepage P, Ge B, Bouchard JP, Mathieu J, Melancon SB, Schalling M, Lander ES, Morgan K, Hudson TJ and Richter A (2000). ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat Genet* 24(2): 120-5.

Escayg A, De Waard M, Lee DD, Bichet D, Wolf P, Mayer T, Johnston J, Baloh R, Sander T and Meisler MH (2000). Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. *Am J Hum Genet* 66(5): 1531-9.

Evgrafov OV, Mersiyanova I, Irobi J, Van Den Bosch L, Dierick I, Leung CL, Schagina O, Verpoorten N, Van Impe K, Fedotov V, Dadali E, Auer-Grumbach M, Windpassinger C, Wagner K, Mitrovic Z, Hilton-Jones D, Talbot K, Martin JJ, Vasserman N, Tverskaya S, Polyakov A, Liem RK, Gettemans J, Robberecht W, De Jonghe P and Timmerman V (2004). Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat Genet* 36(6): 602-6.

Farah S, Sabry MA, Khuraibet AJ, Anim JT, Quasrawi B, Al-Khatam S, Al-Busairi W, Hussein JM, Khan RA and Al-Awadi SA (1997). Marinesco-Sjogren syndrome in a Bedouin family. *Acta Neurol Scand* 96(6): 387-91.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R and Stewart J (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164): 851-61.

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM and Dryja TP (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323(6089): 643-6.

Georgy BA, Snow RD, Brogdon BG and Wertelecki W (1998). Neuroradiologic findings in Marinesco-Sjogren syndrome. *AJNR Am J Neuroradiol* 19(2): 281-3.

Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, Eaton DH, Vehar GA, Capon DJ and Lawn RM (1984). Characterization of the human factor VIII gene. *Nature* 312(5992): 326-30.

Goto Y, Komiyama A, Tanabe Y, Katafuchi Y, Ohtaki E and Nonaka I (1990). Myopathy in Marinesco-Sjogren syndrome: an ultrastructural study. *Acta Neuropathol (Berl)* 80(2): 123-8.

Grantham R (1974). Amino acid difference formula to help explain protein evolution. *Science* 185(4154): 862-4.

Guo X, Shen H, Xiao X, Dai Q, Li S, Jia X, Hejtmancik JF and Zhang Q (2006). Cataracts, ataxia, short stature, and mental retardation in a Chinese family mapped to Xpter-q13.1. *J Hum Genet* 51(8): 695-700.

Hakamada S, Sobue G, Watanabe K, Kumagai T, Hara K and Miyazaki S (1981). Peripheral neuropathy in Marinesco-Sjogren syndrome. *Brain Dev* 3(4): 403-6.

Hakonen AH, Heiskanen S, Juvonen V, Lappalainen I, Luoma PT, Rantamaki M, Goethem GV, Lofgren A, Hackman P, Paetau A, Kaakkola S, Majamaa K, Varilo T, Udd B, Kaariainen H, Bindoff LA and Suomalainen A (2005). Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *Am J Hum Genet* 77(3): 430-41.

Hansen JJ, Durr A, Cournu-Rebeix I, Georgopoulos C, Ang D, Nielsen MN, Davoine CS, Brice A, Fontaine B, Gregersen N and Bross P (2002). Hereditary spastic paraplegia SPG13 is associated with a mutation in the gene encoding the mitochondrial chaperonin Hsp60. *Am J Hum Genet* 70(5): 1328-32.

Harding AE (1983). Classification of the hereditary ataxias and paraplegias. *Lancet* 1(8334): 1151-5.

Harting I, Blaschek A, Wolf NI, Seitz A, Haupt M, Goebel HH, Rating D, Sartor K and Ebinger F (2004). T2-hyperintense cerebellar cortex in Marinesco-Sjogren syndrome. *Neurology* 63(12): 2448-9.

Hebert DN and Molinari M (2007). In and out of the ER: protein folding, quality control, degradation, and related human diseases. *Physiol Rev* 87(4): 1377-408.

Hendershot LM (2004). The ER function BiP is a master regulator of ER function. *Mt Sinai J Med* 71(5): 289-97.

Herbarth B, Pingault V, Bondurand N, Kuhlbrodt K, Hermans-Borgmeyer I, Puliti A, Lemort N, Goossens M and Wegner M (1998). Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci U S A* 95(9): 5161-5.

Herva R, von Wendt L, von Wendt G, Saukkonen AL, Leisti J and Dubowitz V (1987). A syndrome with juvenile cataract, cerebellar atrophy, mental retardation and myopathy. *Neuropediatrics* 18(3): 164-9.

Huynh DP, Yang HT, Vakharia H, Nguyen D and Pulst SM (2003). Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. *Hum Mol Genet* 12(13): 1485-96.

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A and Lander E (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2(3): 204-11.

Irobi J, Van Impe K, Seeman P, Jordanova A, Dierick I, Verpoorten N, Michalik A, De Vriendt E, Jacobs A, Van Gerwen V, Vennekens K, Mazanec R, Tournev I, Hilton-Jones D, Talbot K, Kremensky I, Van Den Bosch L, Robberecht W, Van Vandekerckhove J, Broeckhoven C, Gettemans J, De Jonghe P and Timmerman V (2004). Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nat Genet* 36(6): 597-601.

Ishikawa T, Kitoh H, Awaya A and Nonaka I (1993). Rapid cataract formation in Marinesco-Sjogren syndrome. *Pediatr Neurol* 9(5): 407-8.

Jones B, Jones EL, Bonney SA, Patel HN, Mensenkamp AR, Eichenbaum-Voline S, Rudling M, Myrdal U, Annesi G, Naik S, Meadows N, Quattrone A, Islam SA, Naoumova RP, Angelin B, Infante R, Levy E, Roy CC, Freemont PS, Scott J and Shoulders CC (2003). Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nat Genet* 34(1): 29-31.

Juvonen V, Kulmala SM, Ignatius J, Penttinen M and Savontaus ML (2002). Dissecting the epidemiology of a trinucleotide repeat disease - example of FRDA in Finland. *Hum Genet* 110(1): 36-40.

Karim MA, Parsian AJ, Cleves MA, Bracey J, Elsayed MS, Elsobky E and Parsian A (2006). A novel mutation in BAP/SIL1 gene causes Marinesco-Sjogren syndrome in an extended pedigree. *Clin Genet* 70(5): 420-3.

Katafuchi Y, Kosai K, Ohtaki E, Yamashita Y, Horikawa M, Terasawa K and Nonaka I (1990). Cerebral cortex and brainstem involvement in Marinesco-Sjogren syndrome. *Ann Neurol* 27(4): 448-9.

Kodama S, Komatsu M, Miyoshi M, Nakao H, Sakurai T and U M (1992). Marinesco-Sjogren syndrome with reduced cytochrome c oxidase in muscle. *Kobe J Med Sci* 38(4): 245-54.

Koenig M (2003). Rare forms of autosomal recessive neurodegenerative ataxia. *Semin Pediatr Neurol* 10(3): 183-92.

Komiyama A, Nonaka I and Hirayama K (1989). Muscle pathology in Marinesco-Sjogren syndrome. *J Neurol Sci* 89(1): 103-13.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR and Stefansson K (2002). A high-resolution recombination map of the human genome. *Nat Genet* 31(3): 241-7.

Koskinen T, Santavuori P, Sainio K, Lappi M, Kallio AK and Pihko H (1994). Infantile onset spinocerebellar ataxia with sensory neuropathy: a new inherited disease. *J Neurol Sci* 121(1): 50-6.

Kruglyak L, Daly MJ, Reeve-Daly MP and Lander ES (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58(6): 1347-63.

Kwok SC, Ledley FD, DiLella AG, Robson KJ and Woo SL (1985). Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochemistry* 24(3): 556-61.

Laezza F, Gerber BR, Lou JY, Kozel MA, Hartman H, Craig AM, Ornitz DM and Nerbonne JM (2007). The FGF14(F145S) mutation disrupts the interaction of FGF14 with voltage-gated Na+ channels and impairs neuronal excitability. *J Neurosci* 27(44): 12033-44.

Lagier-Tourenne C, Chaigne D, Gong J, Flori J, Mohr M, Ruh D, Christmann D, Flament J, Mandel JL, Koenig M and Dollfus H (2002). Linkage to 18qter differentiates two clinically overlapping syndromes: congenital cataracts-facial dysmorphism-neuropathy (CCFDN) syndrome and Marinesco-Sjogren syndrome. *J Med Genet* 39(11): 838-43.

Lagier-Tourenne C, Tranebaerg L, Chaigne D, Gribaa M, Dollfus H, Silvestri G, Betard C, Warter JM and Koenig M (2003). Homozygosity mapping of Marinesco-Sjogren syndrome to 5q31. *Eur J Hum Genet* 11(10): 770-8.

Lander ES and Botstein D (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236(4808): 1567-70.

Lathrop GM and Lalouel JM (1984). Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 36(2): 460-5.

Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P and Ackerman SL (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* 443(7107): 50-5.

Lehesjoki AE and Kälviäinen R (2004). Unverricht-Lundborg Disease. GeneReviews at GeneTests: Medical Genetics Information Resource [database online] 2007/09/18 [cited 2007/11/15]; Available from: http://www.genetests.org.

Lehesjoki AE, Koskiniemi M, Norio R, Tirrito S, Sistonen P, Lander E and de la Chapelle A (1993). Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum Mol Genet* 2(8): 1229-34.

Lewis MJ and Pelham HR (1992). Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* 68(2): 353-64.

Litt M, Kramer P, LaMorticella DM, Murphey W, Lovrien EW and Weleber RG (1998). Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Hum Mol Genet* 7(3): 471-4.

Ma Y and Hendershot LM (2004). ER chaperone functions during normal and stress conditions. *J Chem Neuroanat* 28(1-2): 51-65.

Mackay DS, Andley UP and Shiels A (2003). Cell death triggered by a novel mutation in the alphaA-crystallin gene underlies autosomal dominant cataract linked to chromosome 21q. *Eur J Hum Genet* 11(10): 784-93.

Mahloudji M, Amirhakimi GH, Haghighi P and Khodadoust AA (1972). Marinesco-Sjögren syndrome. Report of an autopsy. *Brain* 95(4): 675-80.

Manya H, Bouchet C, Yanagisawa A, Vuillaumier-Barrot S, Quijano-Roy S, Suzuki Y, Maugenre S, Richard P, Inazu T, Merlini L, Romero NB, Leturcq F, Bezier I, Topaloglu H, Estournet B, Seta N, Endo T and Guicheney P (2007). Protein O-mannosyltransferase activities in lymphoblasts from patients with alpha-dystroglycanopathies. *Neuromuscul Disord* doi:10.1016/j.nmd.2007.08.002.

Maquat LE (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* 5(2): 89-99.

Marinesco G, Draganesco S and Vasiliu D (1931). Nouvelle maladie familiale caractérisée par une cataracte congénitale et un arrêt du développement somato-neuro-psychique. *Encéphale* 26: 97-109.

McKusick VA (1966). 2330 Marinesco-Sjögren syndrome. *Mendelian inheritance in man; catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes*. Baltimore,, Johns Hopkins Press**:** 206.

McLaughlin JF, Pagon RA, Weinberger E and Haas JE (1996). Marinesco-Sjogren syndrome: clinical and magnetic resonance imaging features in three children. *Dev Med Child Neurol* 38(7): 636-44.

Merlini L, Gooding R, Lochmuller H, Muller-Felber W, Walter MC, Angelicheva D, Talim B, Hallmayer J and Kalaydjieva L (2002). Genetic identity of Marinesco-Sjogren/myoglobinuria and CCFDN syndromes. *Neurology* 58(2): 231-6.

Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM and Kunkel LM (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* 323(6089): 646-50.

Morton NE (1955). Sequential tests for the detection of linkage. *Am J Hum Genet* 7(3): 277-318.

Muchowski PJ (2002). Protein misfolding, amyloid formation, and neurodegeneration: a critical role for molecular chaperones? *Neuron* 35(1): 9-12.

Muchowski PJ and Wacker JL (2005). Modulation of neurodegeneration by molecular chaperones. *Nat Rev Neurosci* 6(1): 11-22.

Müller-Felber W, Zafiriou D, Scheck R, Patzke I, Toepfer M, Pongratz DE and Walther U (1998). Marinesco Sjögren syndrome with rhabdomyolysis. A new subtype of the disease. *Neuropediatrics* 29(2): 97-101.

Nelson J, Flaherty M and Grattan-Smith P (1997). Gillespie syndrome: a report of two further cases. *Am J Med Genet* 71(2): 134-8.

Ni M and Lee AS (2007). ER chaperones in mammalian development and human diseases. *FEBS Lett* 581(19): 3641-51.

Nikali K, Suomalainen A, Saharinen J, Kuokkanen M, Spelbrink JN, Lönnqvist T and Peltonen L (2005). Infantile onset spinocerebellar ataxia is caused by recessive mutations in mitochondrial proteins Twinkle and Twinky. *Hum Mol Genet* 14(20): 2981-90.

Norio R (2003a). Finnish Disease Heritage I: characteristics, causes, background. *Hum Genet* 112(5-6): 441-56.

Norio R (2003b). The Finnish Disease Heritage III: the individual diseases. *Hum Genet* 112(5-6): 470-526.

Nyholt DR (2002). GENEHUNTER: your 'one-stop shop' for statistical genetic analysis? *Hum Hered* 53(1): 2-7.

Opa P and Zoghbi HY (2002). The hereditary ataxias. *Emery and Rimoin's Principles and Practice of Medical Genetics*. Rimoin DL, J. M. Connor, R. E. Pyeritz, B. R. Korf, Churchill Livingstone**:** 3109-3123.

Orphanet Reports Series (2007). The prevalence of rare diseases: A bibliographic study October 2007 - Issue 6. [cited 2007/11/15]; Available from: http://www.orpha.net/orphacom/cahiers/reports-orphanet.htm.

Ouahchi K, Arita M, Kayden H, Hentati F, Ben Hamida M, Sokol R, Arai H, Inoue K, Mandel JL and Koenig M (1995). Ataxia with isolated vitamin E deficiency is caused by mutations in the alpha-tocopherol transfer protein. *Nat Genet* 9(2): 141-5.

Pagani F and Baralle FE (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 5(5): 389-96.

Peltonen L, Jalanko A and Varilo T (1999). Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8(10): 1913-23.

Perlman S, Becker-Catania S and Gatti RA (2003). Ataxia-telangiectasia: diagnosis and treatment. *Semin Pediatr Neurol* 10(3): 173-82.

Pingault V, Bondurand N, Kuhlbrodt K, Goerich DE, Prehu MO, Puliti A, Herbarth B, Hermans-Borgmeyer I, Legius E, Matthijs G, Amiel J, Lyonnet S, Ceccherini I, Romeo G, Smith JC, Read AP, Wegner M and Goossens M (1998). SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* 18(2): 171-3.

Pras E, Frydman M, Levy-Nissenbaum E, Bakhan T, Raz J, Assia EI, Goldman B and Pras E (2000). A nonsense mutation (W9X) in CRYAA causes autosomal recessive cataract in an inbred Jewish Persian family. *Invest Ophthalmol Vis Sci* 41(11): 3511-5.

Rantamäki M, Krahe R, Paetau A, Cormand B, Mononen I and Udd B (2001). Adult-onset autosomal recessive ataxia with thalamic lesions in a Finnish family. *Neurology* 57(6): 1043-9.

Reinker K, Hsia YE, Rimoin DL, Henry G, Yuen J, Powell B and Wilcox WR (2002). Orthopaedic manifestations of Marinesco-Sjogren syndrome. *J Pediatr Orthop* 22(3): 399-403.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole FS, Curnutte JT and Orkin SH (1986). Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature* 322(6074): 32-8.

Rozen S and Skaletsky HJ (2000). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Krawetz S, S. Misener, Humana Press**:** 365-386.

Sasaki K, Suga K, Tsugawa S, Sakuma K, Tachi N, Chiba S and Imamura S (1996). Muscle pathology in Marinesco-Sjogren syndrome: a unique ultrastructural feature. *Brain Dev* 18(1): 64-7.

Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle DA, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali SR, Simmons A, Clines GA, Sartiel A, Gatti RA, Chessa L, Sanal O, Lavin MF, Jaspers NG, Taylor AM, Arlett CF, Miki T, Weissman SM, Lovett M, Collins FS and Shiloh Y (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 268(5218): 1749-53.

Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS and Krogan NJ (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123(3): 507-19.

Schulz S, Vielhaber S, Muschke P, Mohnike K, Gooding R and Wieacker P (2007). Congenital cataract, ataxia, external ophthalmoplegia and dysphagia in two siblings. A Marinesco-Sjogren-like syndrome. *Neuropediatrics* 38(2): 88-90.

Selcen D and Engel AG (2003). Myofibrillar myopathy caused by novel dominant negative alpha B-crystallin mutations. *Ann Neurol* 54(6): 804-10.

Selkoe DJ (2003). Folding proteins in fatal ways. *Nature* 426(6968): 900-4.

Senderek J, Krieger M, Stendel C, Bergmann C, Moser M, Breitbach-Faller N, Rudnik-Schoneborn S, Blaschek A, Wolf NI, Harting I, North K, Smith J, Muntoni F, Brockington M, Quijano-Roy S, Renault F, Herrmann R, Hendershot LM, Schroder JM, Lochmuller H, Topaloglu H, Voit T, Weis J, Ebinger F and Zerres K (2005). Mutations in SIL1 cause Marinesco-Sjogren syndrome, a cerebellar ataxia with cataract and myopathy. *Nat Genet* 37(12): 1312-4.

Sewry CA, Voit T and Dubowitz V (1988). Myopathy with unique ultrastructural feature in Marinesco-Sjogren syndrome. *Ann Neurol* 24(4): 576-80.

Sjögren T (1950). Hereditary congenital spinocerebellar ataxia accompanied by congenital cataract and oligophrenia. *Confin Neurol* 10: 293-308.

Skre H, Bassoe HH, Berg K and Frovig AG (1976). Cerebellar ataxia and hypergonadotropic hypogonadism in two kindreds. Chance concurrence, pleiotropism or linkage? *Clin Genet* 9(2): 234-44.

Skre H and Berg K (1977). Linkage studies on Marinesco-Sjogren syndrome and hypergonadotropic hypogonadism. *Clin Genet* 11(1): 57-66.

Slavotinek A, Goldman J, Weisiger K, Kostiner D, Golabi M, Packman S, Wilcox W, Hoyme HE and Sherr E (2005). Marinesco-Sjogren syndrome in a male with mild dysmorphism. *Am J Med Genet A* 133(2): 197-201.

Spampanato J, Escayg A, Meisler MH and Goldin AL (2001). Functional effects of two voltage-gated sodium channel mutations that cause generalized epilepsy with febrile seizures plus type 2. *J Neurosci* 21(19): 7481-90.

Spelbrink JN, Li FY, Tiranti V, Nikali K, Yuan QP, Tariq M, Wanrooij S, Garrido N, Comi G, Morandi L, Santoro L, Toscano A, Fabrizi GM, Somer H, Croxen R, Beeson D, Poulton J, Suomalainen A, Jacobs HT, Zeviani M and Larsson C (2001). Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat Genet* 28(3): 223-31.

Steel GJ, Fullerton DM, Tyson JR and Stirling CJ (2004). Coordinated activation of Hsp70 chaperones. *Science* 303(5654): 98-101.

Stein LD (2004). Human genome: end of the beginning. *Nature* 431(7011): 915-6.

Stevens FJ and Argon Y (1999). Protein folding in the ER. *Semin Cell Dev Biol* 10(5): 443-54.

Sun FC, Wei S, Li CW, Chang YS, Chao CC and Lai YK (2006). Localization of GRP78 to mitochondria under the unfolded protein response. *Biochem J* 396(1): 31-9.

Superneau D, Wertelecki W and Zellweger H (1985). The Marinesco-Sjogren syndrome described a quarter of a century before Marinesco. *Am J Med Genet* 22(3): 647-8.

Superneau DW, Wertelecki W, Zellweger H and Bastian F (1987). Myopathy in Marinesco-Sjogren syndrome. *Eur Neurol* 26(1): 8-16.

Suzuki Y, Murakami N, Goto Y, Orimo S, Komiyama A, Kuroiwa Y and Nonaka I (1997). Apoptotic nuclear degeneration in Marinesco-Sjogren syndrome. *Acta Neuropathol (Berl)* 94(5): 410-5.

Tachi N, Nagata N, Wakai S and Chiba S (1991). Congenital muscular dystrophy in Marinesco-Sjogren syndrome. *Pediatr Neurol* 7(4): 296-8.

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K and Kere J (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 100(20): 11553-8.

Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P and Strachan T (1992). Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* 355(6361): 635-6.

Terwilliger JD and Ott J (1994). Handbook of Human Genetic Linkage, The John Hopkins University Press.

The ENCODE Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306(5696): 636-40.

The International HapMap Consortium (2003). The International HapMap Project. *Nature* 426(6968): 789-96.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437(7063): 1299-320.

The International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431(7011): 931-45.

Todorov A (1965). [Marinesco-Sjogren syndrome. 1st anatomo-clinical study]. *J Genet Hum* 14(2): 197-233.

Torbergsen T, Aasly J, Borud O, Lindal S and Mellgren SI (1991a). Mitochondrial myopathy in Marinesco-Sjogren syndrome. *J Ment Defic Res* 35 (Pt 2): 154-9.

Torbergsen T, Stalberg E, Aasly J and Lindal S (1991b). Myopathy in Marinesco-Sjogren syndrome: an electrophysiological study. *Acta Neurol Scand* 84(2): 132-8.

Tournev I, Kalaydjieva L, Youl B, Ishpekova B, Guergueltcheva V, Kamenov O, Katzarova M, Kamenov Z, Raicheva-Terzieva M, King RH, Romanski K, Petkov R, Schmarov A, Dimitrova G, Popova N, Uzunova M, Milanov S, Petrova J, Petkov Y, Kolarov G, Aneva L, Radeva O and Thomas PK (1999). Congenital cataracts facial dysmorphism neuropathy syndrome, a novel complex genetic disease in Balkan Gypsies: clinical and electrophysiological observations. *Ann Neurol* 45(6): 742-50.

Tyson JR and Stirling CJ (2000). LHS1 and SIL1 provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. *Embo J* 19(23): 6440-52.

Van Goethem G, Luoma P, Rantamaki M, Al Memar A, Kaakkola S, Hackman P, Krahe R, Lofgren A, Martin JJ, De Jonghe P, Suomalainen A, Udd B and Van Broeckhoven C (2004). POLG mutations in neurodegenerative disorders with ataxia but no muscle involvement. *Neurology* 63(7): 1251-7.

van Reeuwijk J, Brunner HG and van Bokhoven H (2005a). Glyc-O-genetics of Walker-Warburg syndrome. *Clin Genet* 67(4): 281-9.

van Reeuwijk J, Janssen M, van den Elzen C, Beltran-Valero de Bernabe D, Sabatelli P, Merlini L, Boon M, Scheffer H, Brockington M, Muntoni F, Huynen MA, Verrips A, Walsh CA, Barth PG, Brunner HG and van Bokhoven H (2005b). POMT2 mutations cause alpha-dystroglycan hypoglycosylation and Walker-Warburg syndrome. *J Med Genet* 42(12): 907-12.

Varon R, Gooding R, Steglich C, Marns L, Tang H, Angelicheva D, Yong KK, Ambrugger P, Reinhold A, Morar B, Baas F, Kwa M, Tournev I, Guerguelcheva V, Kremensky I, Lochmuller H, Mullner-

Eidenbock A, Merlini L, Neumann L, Burger J, Walter M, Swoboda K, Thomas PK, von Moers A, Risch N and Kalaydjieva L (2003). Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. *Nat Genet* 35(2): 185-9.

Vicart P, Caron A, Guicheney P, Li Z, Prevost MC, Faure A, Chateau D, Chapon F, Tome F, Dupret JM, Paulin D and Fardeau M (1998). A missense mutation in the alphaB-crystallin chaperone gene causes a desmin-related myopathy. *Nat Genet* 20(1): 92-5.

Vidal R, Revesz T, Rostagno A, Kim E, Holton JL, Bek T, Bojsen-Moller M, Braendgaard H, Plant G, Ghiso J and Frangione B (2000). A decamer duplication in the 3' region of the BRI gene originates an amyloid peptide that is associated with dementia in a Danish kindred. *Proc Natl Acad Sci U S A* 97(9): 4920-5.

Walker PD, Blitzer MG and Shapira E (1985). Marinesco-Sjogren syndrome: evidence for a lysosomal storage disorder. *Neurology* 35(3): 415-9.

Wang GS and Cooper TA (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8(10): 749-61.

Weinstock GM (2007). ENCODE: more genomic empowerment. *Genome Res* 17(6): 667-8.

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G and Lathrop M (1992). A second-generation linkage map of the human genome. *Nature* 359(6398): 794-801.

Weitzmann A, Volkmer J and Zimmermann R (2006). The nucleotide exchange factor activity of Grp170 may explain the non-lethal phenotype of loss of Sil1 function in man and mouse. *FEBS Lett* 580(22): 5237-40.

Williams TE, Buchhalter JR and Sussman MD (1996). Cerebellar dysplasia and unilateral cataract in Marinesco-Sjogren syndrome. *Pediatr Neurol* 14(2): 158-61.

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P and Altshuler D (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308(5718): 107-11.

Wolfsberg TG, Wetterstrand KA, Guyer MS, Collins FS and Baxevanis AD (2002). A user's guide to the human genome. *Nat Genet* 32 Suppl: 1-79.

Zhao L, Longo-Guess C, Harris BS, Lee JW and Ackerman SL (2005). Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nat Genet* 37(9): 974-9.

Zimmer C, Gosztonyi G, Cervos-Navarro J, von Moers A and Schroder JM (1992). Neuropathy with lysosomal changes in Marinesco-Sjogren syndrome: fine structural findings in skeletal muscle and conjunctiva. *Neuropediatrics* 23(6): 329-35.

Ziv Y, Frydman M, Lange E, Zelnik N, Rotman G, Julier C, Jaspers NG, Dagan Y, Abeliovicz D, Dar H, Borochowitz Z, Lathrop M, Gatti RA and Shiloh Y (1992). Ataxia-telangiectasia: linkage analysis in highly inbred Arab and Druze families and differentiation from an ataxia-microcephaly-cataract syndrome. *Hum Genet* 88(6): 619-26.