Department of Molecular Genetics
The Folkhälsan Institute of Genetics

and

Department of Medical Genetics
Haartman Institute
University of Helsinki
Finland

MOLECULAR GENETICS OF SELECTIVE INTESTINAL MALABSORPTION OF VITAMIN B12 THE GRÄSBECK-IMERSLUND DISEASE (MEGALOBLASTIC ANEMIA 1)

Maria Aminoff-Backlund

Academic dissertation

To be publicly discussed with the permission of the Faculty of Medicine, for the Department of Medical Genetics, University of Helsinki, in the large lecture hall of the Haartman Institute, Haartmaninkatu 3, Helsinki on June 28th, 2000, at 12 noon

Helsinki 2000

Supervised by

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

Reviewed by

Docent Erkki Elonen
Division of Haematology
Helsinki University Central Hospital

Docent Maija Wessman
Division of Genetics
Department of Biosciences
University of Helsinki

Official opponent

Docent Tom Pettersson
Division of Internal Medicine
Helsinki University Central Hospital

ISBN 952-91-2191-1
ISBN 952-91-2192-X (pdf version, http://ethesis.helsinki.fi/)

Gummerus Kirjapaino Oy
Saarijärvi 2000

To Anders

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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 In addition, some unpublished data are presented.

I Aminoff M., E. Tahvanainen, R. Gräsbeck, J. Weissenbach, H. Broch & A. de la Chapelle (1995). Selective intestinal malabsorption of vitamin B12 displays recessive Mendelian inheritance: assignment of a locus to chromosome 10 by linkage. Am J Hum Genet 57:824-831.

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

III Aminoff M., S. Brady, PJ. Verroust, SK. Moestrup & R. Krahe. The genomic structure of the human *CUBN* gene encoding, cubilin, the intrinsic factor-vitamin B12 receptor. *Submitted*.

IV Dugué B., M. Aminoff, I. Aimone-Gastin, E. Leppänen, R. Gräsbeck & J-L. Guéant (1998). A urinary radioisotope-binding assay to diagnose Gräsbeck-Imerslund disease. J Pediatr Gastr Nutr 26:21-25.

ABBREVIATIONS

BAC	bacterial artificial chromosome
bp	base pair
Cbl	cobalamin
cDNA	complementary DNA
CNCbl	cyanocobalamin
CEPH	Centre d'Etude du Polymorphisme Humain
cM	centiMorgan
cR	centiRays
CUB	abbreviation of Clr/s, Uegf and Bone morphogenic protein-1
CUBN	the cubilin gene
EGF	epidermal growth factor
EST	expressed sequence tag
FM1	Finnish mutation 1
FM2	Finnish mutation 2
FM3	Finnish mutation 3
HC	haptocorrin
HGP	Human Genome Project
IF	intrinsic factor
IF-B12	the intrinsic factor-vitamin B12 complex
IF-R	intrinsic factor receptor
kb	kilobase pairs
kDa	kilodalton
lod	logarithm of odds
LR-PCR	long range polymerase chain reaction
Mb	megabase pairs
MGA1	megaloblastic anemia 1
MIM	Mendelian Inheritance in Man
mRNA	messenger RNA
nd	not determined
p	short arm of a chromosome
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PIC	polymorphism information content
q	long arm of a chromosome
RH	radiation hybrid
RT-PCR	reverse transcriptase polymerase chain reaction
SNP	single-nucleotide polymorphism
STR	short tandem repeats
STS	sequence tagged site
TC II	transcobalamin II
YAC	yeast artificial chromosome

SUMMARY

This study was focused on determining the genetic background of Megaloblastic Anemia 1 (MGA1), also known as the Gräsbeck-Imerslund disease. MGA1 is an autosomal recessive disorder that belongs to the Finnish disease heritage. The disease is characterized by juvenile megaloblastic anemia, failure to thrive and neurological symptoms due to selective intestinal malabsorption of vitamin B12. The disease was originally described by Ralph Gräsbeck and collaborators in Finland and Olga Imerslund in Norway about 40 years ago and the majority of the Finnish and the Norwegian patients were identified during the following years. Since the early 1980's almost no new cases occurred in the two populations, leading to doubts concerning a Mendelian inheritance for the condition. The disease has generally been thought to be due to an error in the intrinsic factor receptor (IF-R) in the distal small intestine. For the present study patients with megaloblastic anemia due to cobalamin deficiency that fulfilled the diagnostic criteria were collected from both Finland and Norway. The patients from Finland were identified by perusing the hospital records.

Using six multiplex Finnish families the locus for MGA1 was mapped with linkage analysis to a 6-cM region on the short arm of chromosome 10. The linkage was confirmed with three Norwegian multiplex families and 11 additional Finnish families. Use of linkage disequilibrium (LD) and haplotype analysis in the Finnish families further narrowed the critical region to ~2 cM. A YAC contig was constructed over a distance of approximately 4 cM to positionally clone the MGA1 gene. Simultaneously, the obvious candidate gene for MGA1, the receptor for the intrinsic factor (IF)-B12 complex was identified by a functional approach by a Danish-French research team. The protein was named cubilin and the gene designated *CUBN*. *CUBN* was mapped to the same chromosomal region, by fluorescence in situ hybridization (FISH), radiation hybrid (RH) mapping and screening of YAC clones, as previously identified by linkage analysis in the Finnish and Norwegian MGA1 families. Screening of our YAC contig with *CUBN* intragenic markers further confirmed the gene as a functional positional candidate gene for MGA1.

The cubilin gene *CUBN* was screened for mutations in Finnish and Norwegian MGA1 families. Two mutations were identified in the Finnish population. The first Finnish mutation (FM1), found in the majority of the Finnish patients, was a 3916C->T missense mutation in CUB domain 8 changing proline to leucine (P1297L). The second mutation (FM2), seen in

only one affected, was a point mutation activating a cryptic splice site that results in the in frame insertion of multiple stop codons in the CUB domain 6 intron. No mutations have been identified in the Norwegian patients.

The genomic structure of the 36-domain cubilin protein was determined by LR-PCR and direct sequencing of our BAC contig covering the entire ~170 kb gene. A total of 67 exons and 66 introns were identified in addition to the putative promoter region.

The IF-R is expressed in the ileum and the kidney tubules and is found in urine. The urinary activity of the IF-R was measured in some of the Finnish MGA1 patients by using radioactive vitamin B12 (^{57}Co) cyanocobalamin (CNCbl) bound to intrinsic factor (IF) as a ligand. A markedly decreased and nearly undetectable binding activity of the IF-vitamin B12 complex was observed in the patients compared with their healthy relatives and the controls. The assay can therefore be used for initial diagnostic purposes. The characterization of the cubilin gene and the mutations responsible for MGA1 will in the future facilitate more exact diagnosis of new suspected cases at an earlier stage of the disease, which is important for an appropriate treatment.

1. INTRODUCTION

Megaloblastic anemias in children are mainly due to folate or vitamin B12 (cobalamin) deficiency (Chanarin 1987). Cobalamin (Cbl) deficiency is the most common cause of megaloblastic anemia in the Nordic countries while megaloblastic anemia due to folate deficiency is relatively rare (Gräsbeck 1984, Gräsbeck & Weber 1997). Megaloblastic anemia with neurological disturbances, recurrent infections, developmental delays and failure to thrive are characteristic symptoms of vitamin B12 deficiency in infancy (Visakorpi & Furuhjelm 1967, Campbell et al. 1981, Wulffraat et al. 1994). Megaloblastic anemia is a severe clinical condition that can be fatal if untreated.

All presently known inherited disorders in human cobalamin metabolism are single gene defects, inherited as autosomal recessive traits that can give rise to mental retardation and other severe neurological consequences (Linnell & Bhatt 1995). They affect either the absorption of cobalamin from the intestine, their transport in the blood or their intracellular metabolism. There are several different defects, including impaired function or expression of intrinsic factor (IF), the intrinsic factor receptor (IF-R), transcobalamin II (TC II) or the various reductases and synthases required for synthesis of adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl) (Linnell & Bhatt 1995).

Hereditary forms of vitamin B12 deficiency resulting in megaloblastic anemia are known to relate to low or absent secretion of intrinsic factor (IF) (Pernicious anemia, MIM 261000) (McNichol & Egan 1968, Katz et al. 1972), to decreased or absent synthesis of functional transcobalamin II (TC II deficiency, MIM 275350) (Hakami et al. 1971, Hitzig et al. 1974, Burman et al. 1979) or to a defect in the intestinal epithelium leading to decreased uptake of the IF-vitamin B12 complex and therefore to vitamin B12 deficiency (Gräsbeck-Imerslund disease, Megaloblastic anemia 1 (MGA1, MIM 261100) (Gräsbeck et al. 1960, Imerslund 1959, 1960).

At the time when this study was initiated, more than thirty years had elapsed since MGA1 was first independently described by Ralph Gräsbeck and co-workers in Finland (Gräsbeck et al. 1960) and Olga Imerslund in Norway (Imerslund 1959, 1960).
The goal of this study was to prove the existence of MGA1, its autosomal recessive mode of inheritance and define the gene responsible for MGA1.

2. REVIEW OF THE LITERATURE

2.1. Megaloblastic anemia

Megaloblastic anemia is a hematologic disorder characterized by the production in the bone marrow and increase in the peripheral blood of abnormally large nucleated cells, including immature erythrocytes, superlobulated polymorphic leukocytes and large platelets. There is usually a reduction in the total white cell and the red cell counts, and sometimes very low values are seen. Examination of the bone marrow is of great diagnostic importance, since in cobalamin deficiency megaloblasts are found, hence the name megaloblastic anemia. The condition is usually due to a deficiency of folate or vitamin B12. In these conditions, cells of other tissues, especially rapidly replicating, e.g. epithelial cells are also affected and the corresponding histology altered (Gräsbeck & Salonen 1976, Chanarin 1979). The best know condition is pernicious anemia caused by vitamin B12 deficiency resulting from lack of secretion of gastric intrinsic factor (IF). This, in turn, is usually due to atrophy of the gastric mucosa caused by an autoimmune process (Chanarin 1979).

Although the clinical symptoms for the different types of congenital megaloblastic anemia are similar, there is a major clinical difference in the age of onset. While neither one of the two congenital Cbl transport protein deficiencies, the more frequently occurring pernicious anemia or MGA1, manifests itself before the age of one year, most TC II deficient patients develop severe megaloblastic anemia as early as 1 to 3 months after birth (Gräsbeck & Salonen 1976, Burman et al. 1979, Hall 1992). In infants with congenital vitamin B12 malabsorption, megaloblastic anemia generally develops later between 12-18 months by which time the stored cobalamin received from the mother during the pregnancy is exhausted (Furuhjelm & Nevanlinna 1973, Linnell & Bhatt 1995).

Absorption tests using radioactive labeled vitamin B12, are useful techniques to determine vitamin B12 malabsorption even when there are no signs of vitamin B12 deficiency (Gräsbeck et al. 1956). The usual technique, Schilling's urinary excretion test, measures how much of the orally ingested radioactively labeled vitamin B12 is excreted in the urine following a flushing dose of intramuscular injected non-radioactive vitamin B12. It has been shown to be a useful measure of intestinal cobalamin absorption (Schilling 1953).

Another rare cause of megaloblastic anemia, previously more common in the eastern parts of Finland, where also MGA1 occurred more frequently, was fish tapeworm *Diphyllobothrium latum* infection (Nyberg et al. 1958, Gräsbeck et al. 1962). It was shown that *D. latum* impairs vitamin B12 absorption in its host and therefore frequently causes hematologic changes from mild macrocytosis to megaloblastic anemia. Tapeworm anemia very much resembles megaloblastic anemia caused by congenital selective vitamin B12 malabsorption (MGA1). It also occurred among very young children. Therefore it was of great importance to exclude the existence of *D. latum* in patients suffering from megaloblastic anemia (Nyberg et al. 1958, Gräsbeck et al. 1962, von Bonsdorff 1977). Today tapeworm anemia is very rare in Finland primarily as a result of changes in dietary habits, food preparation and population hygiene (Gräsbeck & Weber 1997).

2.1.1. Megaloblastic anemia 1 (MGA1)

Megaloblastic anemia 1, originally named Gräsbeck-Imerslund disease or Imerslund-Gräsbeck disease or syndrome, is a specific vitamin B12 malabsorption defect that is relatively easy to recognize on the basis of the frequent association of megaloblastic anemia and a benign proteinuria (Gräsbeck et al. 1960, Imerslund 1959, 1960). In the Gräsbeck-Imerslund disease the clinical signs are similar to those in juvenile pernicious anemia (lack of IF secretion), but the cause is defective uptake of the intrinsic factor-vitamin B12 (IF-B12) complex in the terminal ileum instead of impaired IF secretion (Gräsbeck 1972). The presence of proteinuria may indicate that the receptor facilitating IF-B12 uptake in the intestine is also important for kidney function (Gräsbeck 1997, Moestrup et al. 1998).

A recessively inherited form of megaloblastic anemia has also been detected in a family of giant schnauzer dogs. These dogs also have megaloblastic anemia as a result of selective vitamin B12 malabsorption and their phenotype greatly resembles MGA1 (Fyfe et al. 1989, 1991a).

The disease may escape attention since the first symptoms tend to be very unspecific, such as recurrent infections and failure to thrive (Gräsbeck & Kvist 1967). The disease is however usually diagnosed during the first 2-5 years of life. The diagnosis is based on clinical findings such as hematological tests revealing typical anemia, developmental delay and neurological lesions. The therapy is lifelong and consists of intramuscular injections of vitamin B12. When properly treated the disease is a fairly innocuous condition.

In megaloblastic anemia, hemoglobin decrease is usually less marked because of the increased size of the red cells leading to increased total hemoglobin content per cell. This results in a poor correlation between the cobalamin and the hemoglobin concentrations, making a hemoglobin determination alone an unreliable method in the diagnosis and therapy of cobalamin deficiency states. Since poor absorption of vitamin B12 is the usual cause of megaloblastic anemia, the serum cobalamin (Cbl) concentration is determined and Schilling absorption tests I (with Cbl only) and II (with IF-Cbl) are performed (Gräsbeck et al. 1960, Gräsbeck et al. 1962, Gräsbeck & Salonen 1976, Nexø et al. 1994, Linnell & Bhatt 1995).

2.1.2. The IF-receptor, a candidate gene for MGA1

Since the molecular background of the Gräsbeck-Imerslund disease has not been known there have been different hypotheses and speculations about the biochemical defect(s) underlying the disease. The most widely accepted theory, already suggested by Gräsbeck and co-workers (Gräsbeck et al. 1960), has been an abnormality or lack of the receptor for IF-B12 complex in the ileum. Lack of IF-receptor binding activity in the ileum has been demonstrated in the patients (Seetharam et al. 1981, Burman et al. 1985). A defective brush-border expression of the IF-receptor in the ileum has also been observed in the giant schnauzer dogs with the inherited intestinal malabsorption of vitamin B12 (Fyfe et al. 1991b). However, in other reported cases there seem to be no defect in the ileal receptors for the IF-B12 complex. Instead the defect may be in another of the links in the chain of reactions transferring cobalamin from the receptors in the ileum to transcobalamin II (TCII), which transports vitamin B12 in the blood (MacKenzie et al. 1972).

2.1.3. Vitamin B12

Vitamin B12 was first isolated as cyanocobalamin (CNCbl) in 1948 (Rickes et al. 1948, Smith & Parker 1948). It belongs to the corrin compounds, which are characterized by a corrin ring containing a central cobalt (Co) atom and various axial ligands (Gräsbeck & Salonen 1976).

Cobalamin, vitamin B12, is synthesized by bacteria and other microorganisms growing in soil and water and in the rumen or intestine of e.g. sheep and cattle (Allen 1975, Gräsbeck & Salonen 1976). Cobalamins are essential vitamins, which end up in higher animals via food chains. In man the cobalamins are obligatory nutrients. Rich dietary sources of cobalamin are liver, kidney, meat, seafood and dairy products (Faquharson & Adams 1976, Gräsbeck & Salonen 1976, Sandberg et al. 1981). The cobalamins play an important role in several

intracellular reactions in mammals, such as in the metabolism of e.g. protein, fats and carbohydrates, in blood formation and in neural functions (Gräsbeck & Nyberg 1957, Hansen & Nexø 1987). Vitamin B12 is needed for the synthesis of DNA, i.e. for supplying the methyl group to thymine. In vitamin B12 deficiency RNA and protein synthesis are not affected but DNA replication is, which results in large cells that do not divide (Gräsbeck & Salonen 1976).

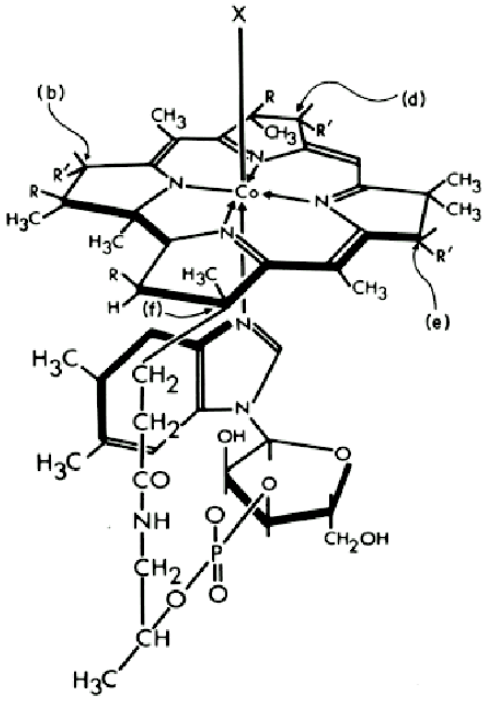

Figure 1.
The chemical structure of cobalamin. (From Fenton & Rosenberg 1978)

The intestinal absorption of vitamin B12 depends on its binding to specific transport proteins (Neale 1990). Cobalamin liberated by digestion is first bound to R-protein or haptocorrin (HC) (also called cobalophilin) contained in saliva, other secretions and leukocytes. Pancreatic enzymes break the complex and the vitamin is bound to intrinsic factor (IF), a protein secreted by the gastric mucosa. The IF-B12 complex is transported to the distal small intestine, where it attaches to the receptors on the enterocyte. For the attachment calcium ions and neutral pH are needed. Following the absorption of the IF-B12 complex to the ileal

receptor, according to the current view, the whole complex is internalized after which IF-B12 is segregated from the receptor and directed to the lysosomes for degradation of IF and the receptor is recycled to the membrane. TC II (also called transcobalamin: TC), present in the blood circulation and in various tissue fluids, is responsible for the essential delivery of cobalamin to most tissues (Allen 1975, Neale 1990, Linnell & Bhatt 1995). After oral intake there is a delay of several hours before the vitamin appears in the blood (Doscherholmen et al. 1957, Birn et al. 1997).

When there is an acute requirement for various metabolic functions in man there are large cobalamin stores available in liver and smaller ones in kidney, gut, lung endocrine glands and skeletal muscle. In cobalamin disorders all dividing cells in the body are affected although tissues with rapid cell formation such as bone marrow, blood and epithelia show the strongest signs (Gräsbeck & Salonen 1976, Linnell & Bhatt 1995).

2.2. The Finnish disease heritage

The concept of a "Finnish disease heritage" was introduced by Norio, Nevanlinna and Perheentupa in 1973 and consisted initially of 10 inherited diseases that were much more prevalent in Finland than in other populations (Norio et al. 1973). Today that concept includes some 30 diseases with a wide diversity of clinical phenotypes (de la Chapelle & Wright 1998, Peltonen et al. 1999) (see Table 1). The majority of the diseases are autosomal recessive with the exception of two autosomal dominant and two X-chromosomal recessive disorders. On the other hand, several recessive diseases that are common elsewhere are rare in the Finnish population, e.g. cystic fibrosis (Norio et al. 1973, Kere et al. 1989, de la Chapelle 1993, de la Chapelle & Wright 1998). Most disorders of the Finnish disease heritage are rare in the other Nordic countries. Nevertheless, progressive myoclonus epilepsy (EPM1) or Unverricht-Lundborg disease is an example of another disease in addition to MGA1 that also has been found in other Scandinavian countries (Norio 1981). Today the loci for the majority of the Finnish diseases have been assigned to a specific chromosome and in several cases the defective gene has been cloned and identified.

The Finns are a classic example of a genetically isolated population that is thought to descend mostly from a small number of original founders that existed about 2000 years ago. Due to linguistic, geographic and religious reasons the Finns have remained highly isolated from their Nordic and Slavic neighbors. The small isolated founder population that rapidly increased but with frequent “bottlenecks”, due to wars and severe epidemics and famines, allowed the founder effect and genetic drift to form the Finnish gene pool (Nevanlinna 1972).

Table 1. The Finnish Diseases

Disease Identified genes	Abbreviation	Locus	Gene	Defective protein	Reference
Aspartylglucosaminuria	AGU	4q	*AGA*	Aspartylglucosaminidase	Ikonen et al. 1991
Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy	APECED	21q	*AIRE*	Novel nuclear protein	The Finnish-German APECED consortium 1997, Nagamine at al. 1997
Choroidermia	CHM	Xq	*TCD*	Rab geranylgeranyl transferase	Sankila et al. 1992
Congenital chloride diarrhea	CLD	7q	*DRA*	Sulphate transporter	Höglund et al. 1996
Congenital nephrosis, Finnish type	CNF	19q	*NPHS1*	Nephrin	Kestilä et al. 1998
Cornea plana congenita	CNA2	12q	*KERA*	Keratocan	Pellegata et al. 2000
Diastrophic dysplasia	DTD	5q	*DTDST*	Sulphate transporter	Hästbacka et al. 1994
Familial amyloidosis, Finnish type	FAF	9q	*GSN*	Gelsolin	Maury et al. 1990
Gyrate atrophy of choroid and retina	GA, HOGA	10q	*OAT*	Omithine gamma-aminotransferase	Mitchell et al. 1988
Hereditary fructose intolerance	HFI, ALDOB	9q	*ALDOB*	Aldolase B	Cross et al. 1988
Hypergonadotropic ovarian dysgenesis	FSH-RO	2p	*FSHR*	Follicle stimulating hormone recep.	Aittomäki et al. 1995
Infantile neuronal ceroid lipofuscinosis	INCL	1p	*PPT*	Palmitoyl protein thioesterase	Vesa et al. 1995
Lysinuric protein intolerance	LPI	14q	*SLC7A7*	L amino acid transporter	Borsani et al. 1999, Torrents et al. 1999
Mulibrey nanism	MUL	17q	*MUL*	Novel RBCC protein	Avela et al. in press
Nonketotic hyperglycinemia	NKH	9p	*GLDC*	Glycine cleavage system; protein P	Kure at al. 1992
Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy	PLO-SL	19q			Paloneva et al. in press
Progressive epilepsy with mental retardation	EPMR	8p	*CLN8*	Novel transmembrane protein	Ranta et al. 1999
Progressive myoclonous epilepsy	EPM1	21q	*CSTB*	Cystatin B	Pennacchio et al. 1996
X-linked juvenile retinoschisis	RS	Xp	*RS1*	Retinoschisin	Sauer et al. 1997
Sialic acid storage disease (Salla disease)	SASD	6q	*SLC17A5*	Sialin; novel transporter	Verheijen et al. 1999
Selective intestinal malabsorption of vitamin B12	MGA1	10p	*CUBN*	Cubilin	**This study**
Variant late infantile ceroid lipofuscinosis, Finnish type	vLINCL	13q	*CLN5*	Novel transmembrane protein	Savukoski et al. 1998
Mapped loci					
Cartilage-hair hypoplasia	CHH	9p	nd		Sulisalo et al. 1993
Cohen syndrome	COH	8q	nd		Tahvanainen et al. 1994
Congenital lactase deficiency	CLD	2q	nd		Järvelä et al. 1998
Growth retardation with acidosis		2q	nd		Visapää et al. 1998
Hydrolethalus syndrome	HYDROLET	11q	nd		Visapää et al. 1999
Infantile-onset spinocerebellar ataxia	IOSCA	10q	nd		Nikali et al. 1995
Multiple contracture syndrome	LCCS	9q	nd		Mäkelä-Bengs et al. 1998
Meckel syndrome	MKS	17q	nd		Paavola et al. 1995
Muscle eye brain disease	MEB	1p	nd		Cormand et al. 1999
Tibial muscular dystrophy	TMD	2q	nd		Haravuori et al. 1998
Usher syndrome, type III	USH3	3q	nd		Sankila et al. 1995
Unmapped loci					
Infantile cerebellooptic atropy	PEHO	nd			Salonen et al. 1991
Rapadilino syndrome	RAPADILINO	nd			Kääriäinen et al. 1989

2.3. Identification of disease genes

There are different strategies for identifying human disease genes depending on how much is known about the pathogenesis of the disease and availability of already mapped and cloned putative candidate genes. The identification of disease causing genes can be accomplished either by functional or positional cloning (Collins 1992, Ballabio 1993). A functional cloning approach can be applied when the basic biochemical defect of the disease is known (Collins 1992). Both the hemophilia A gene (Gitschier et al. 1984) and the gene for phenylketonuria (Robson et al. 1982) were identified using this strategy. Another cloning strategy based on a similar but not that precise functional approach is the position-independent candidate gene approach. Using this procedure still some functional information about the disease gene is needed. There has to be a general idea of the molecular pathogenesis of closely related human or animal disease phenotypes (Collins 1995).

However, for the majority of inherited diseases the knowledge about the molecular background underlying the disease is usually limited and generally possible candidate genes have not yet been cloned and characterized. In such cases, mapping the disease gene to a specific subchromosomal localization by genetic linkage analysis followed by positional cloning makes cloning of novel genes possible. Pure positional cloning is, however, usually very time consuming without any factors limiting the critical candidate region, such as a strong linkage disequilibrium (for example the diastrophic dysplasia (DTD) gene, Hästbacka et al. 1994), usually only seen in genetically homogenous populations (Jorde 1995, Peltonen 2000), or disease associated visible cytogenetic rearrangements as in the dystrophin gene (Lindenbaum et al. 1979). Therefore the positional candidate gene approach, combining the knowledge of map position with the increasingly dense human transcript map, is today the most appealing and predominant method for cloning human disease genes (Boguski & Schuler 1995, Collins 1995).

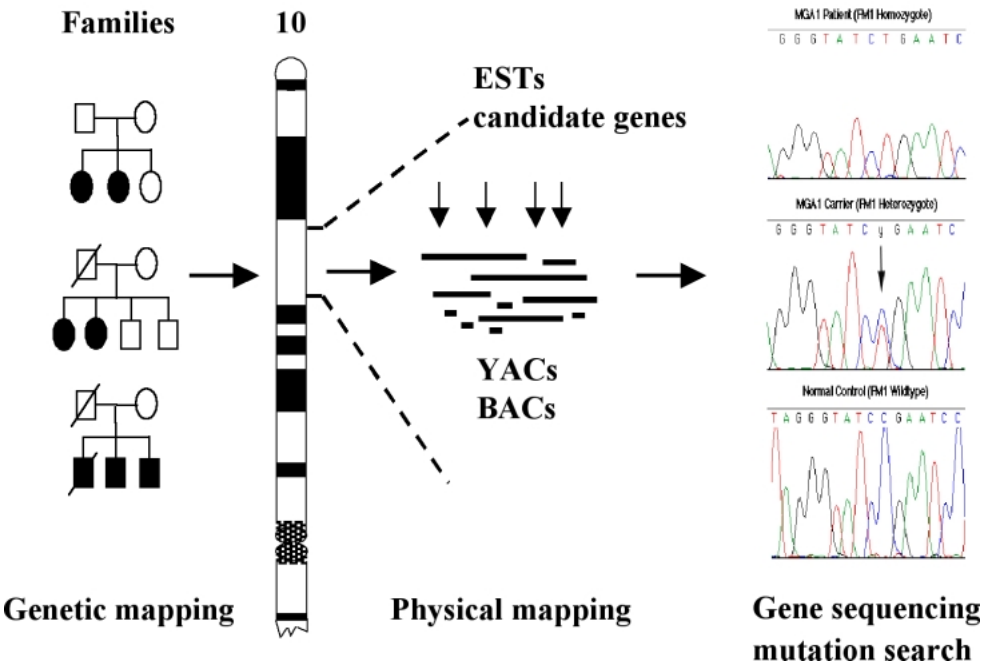

Figure 2. Schematic presentation of steps involved in identification of disease genes by positional cloning that starts with linkage analysis and ends up in identification of disease causing mutation(s).

2.3.1. Linkage and linkage disequilibrium analyses

Linkage analysis is often the first step towards localization and characterization of disease genes. Furthermore, since most inheritable diseases are known only by their phenotype and no obvious candidate gene generally exists, linkage analysis is the ultimate way to map novel disease genes. In order to perform linkage analysis, a sufficient number of multiplex families, with two or more affected children, have to be included in the study in order to confirm or exclude linkage. The mode of inheritance for the disease studied should be known. Mapping genes for hereditary diseases is based on the use of polymorphic markers spanning the genome, where cosegregation of alleles at the marker loci and a genetic trait in families are studied.

Linkage is observed when two loci located on the same chromosome are inherited together at a rate corresponding to the distance between them. The recombination fraction (θ) is used as a measure of the distance between two loci. Theoretically it ranges from (θ)=0, for loci close to each other, to (θ)=0.5 for loci far apart. Two loci are considered genetically linked when (θ)<0.5, i.e. recombination is observed in less than 50% of the meiosis. The likelihood of genetic linkage between loci is given in logarithm of odds, lod score (Z). At the maximum

total lod score (Zmax) of +3 or greater, linkage is considered proven while –2 or less is often evidence against linkage. The genetic distances between markers on a genetic map are given in centimorgan (cM), where two loci showing 1% recombination are 1 cM apart (Ott 1991, Terwilliger & Ott 1994).

On the other hand, to be able to identify and clone a specific gene, giving rise to the disorder, a more precise localization is necessary. Linkage disequilibrium is consequently a powerful statistical method that allows fine-scale mapping and identification of disease genes (Terwilliger 1995, Xiong & Guo 1997). Linkage disequilibrium or allelic association is a nonrandom association of alleles at linked loci and reflects the lack of historical recombinations between the marker and the disease locus. For disease gene mutations, disequilibrium can therefore be expected only if the majority of the patients have the same inherited mutation from a shared ancestor. The older the founder mutation is, the closer the marker has to be to indicate linkage disequilibrium.

In an isolated population, that originates from a small founder population and where the expansion of the population has occurred by growth rather than by immigration, the advantage of linkage disequilibrium in high-resolution mapping of disease genes can be utilized (de la Chapelle 1993, Peltonen 1997, de la Chapelle & Wright 1998, Peltonen et al. 1999). Identification of the genes for diastrophic dysplasia (Hästbacka et al. 1994), congenital chloride diarrhea (Höglund et al. 1996) and mulibrey nanism (Avela et al. in press) are good examples where highly informative linkage disequilibrium data facilitated the identification of the disease locus (Hästbacka et al. 1992, Höglund et al. 1995, Avela et al. 1997). Construction of chromosomal maps of genetically linked DNA markers has made almost the entire genome accessible to linkage studies in families where genetic traits are segregating (White et al. 1989).

2.3.2. DNA polymorphisms as markers

The initial construction of a genetic linkage map in human was based on the idea of using polymorphic restriction fragment length polymorphisms (RFLP) as markers (Botstein et al. 1980). The first comprehensive human genetic map was assembled by a combination of linkage analysis and physical localization of selected clones. Polymorphic loci were arranged into linkage groups estimated to be able to detect linkage to at least 95% of the human genome (Donis-Keller et al. 1987). Since the information obtained using RFLP is very

limited, the new genetic hypervariable minisatellite markers, a variable number of tandem repeats of short DNA sequence (VNTR), greatly improved linkage studies (Nakamura et al. 1987). The source of polymorphic markers has however increased. The standard tool in linkage analysis is nowadays the use of microsatellite markers, which are simple short tandem-repeats (STR). One of the most commonly used microsatellite markers have been the PCR-typeable $(AC)_n$ repeats (Weber & May 1989). The $(AC)_n$ dinucleotide repeats are highly tandemly repeated (~15-30 times) abundant DNA elements that have been found in eukaryotic genomes examined from yeast to human, indicating a high evolutionary conservation (Hamada et al. 1982, Weber & May 1989). Genetic linkage maps of the human genome have been constructed primarily based on these polymorphic $(AC)_n$ repeats (Weissenbach et al. 1992, Gyapay et al. 1994, Dib et al. 1996). A collection of tri- and tetra nucleotide short tandem repeat polymorphisms (STRP) are an example of other tandemly repeated polymorphic markers that are similarly used in constructing genome wide human linkage maps (Sheffield et al. 1995). However, the most common variations in the human genome are the frequently occurring, widely distributed single base pair differences, called single nucleotide polymorphisms (SNPs) (Collins et al. 1997, Wang et al. 1998). Although the SNPs are less informative, with only two alleles, than the other highly informative $(AC)_n$ and STRP markers, they are more abundant in the human genome and have a greater potential for future automated mapping (Wang et al. 1998). The maps with an ever-increasing number of genetic markers can be used to map any Mendelian trait, particularly monogenic human diseases (Gyapay et al. 1994).

2.3.3. Radiation hybrids

Radiation hybrid (RH) mapping is a very useful tool in refining the genetic localization of disease genes by physical mapping of linked DNA markers. Human RH maps are generated by a lethal irradiation of diploid human donor cells that are fused to a non-irradiated recipient rodent somatic cell line (Cox et al. 1990, Walter et al. 1994). In RH mapping the frequency of X-ray breakage between two markers is used as a statistical measure of the distance between markers and their order on the chromosome. Unordered DNA markers can be determined with a very high resolution. The resolution of RH maps depends on the dose of X-rays used to generate the hybrids. Hybrids generated with high doses, 8000-10,000 rad of X-rays, are very useful for ordering nearby DNA markers at a 500-kb level of resolution. The distances between markers are given in centiRays (cR), which are analogous to cM, where 1 cR corresponds to 1% breakage between two markers and is dependent on the radiation dose

(Cox et al. 1990). Hybrid panels with lower resolution have also been generated, which are more useful in ordering markers further apart (Walter et al. 1994). Since RH mapping is based on statistical likelihood, the RH map does not necessarily always represent the actual physical order or distances between markers on the chromosomes (Cox et al. 1990, Jones 1997).

A panel of RHs of the human genome is available and can be used to map polymorphic and non-polymorphic markers and for integrating already existing genetic and physical maps (Hudson et al. 1995, Gyapay et al. 1996, Stewart et al. 1997). Using the RH approach any human DNA sequence, that can be distinguished from rodent DNA background, can be mapped (Cox 1995). RH maps covering the whole genome are also available for mouse (McCarthy et al. 1997) and rat (Watanabe et al. 1999), allowing comparison and integration of maps from different species.

2.3.4. Physical mapping

The ability to physically localize and identify disease genes is greatly enhanced by integrating already existing genetic and physical maps. Unidentified Sequence Tagged Sites (STSs) and Expressed Sequence Tags (ESTs) can be mapped by PCR screening using either RH panels (Gyapay et al. 1996) or yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC) libraries (Hudson et al. 1995, Kim et al. 1996, Cai et al. 1998). A new updated physical map consisting of more than 40,000 STSs representing about 30,000 unique human genes was published in 1998 (Deloukas et al. 1998). This new gene map may consist of up to half of the estimated total number of 60,000-100,000 human genes (Antequera & Bird 1993, Fields et al. 1994), and is therefore of great help and a powerful tool in positional cloning of single and also more complex disease genes. The initial step in hunting for and identification of disease genes is the construction of a contig consisting of a set of physically overlapping cloned DNA fragments spanning the putative region of interest.

The first-generation physical map of the human genome was constructed by screening the YAC library from CEPH with more than 2,000 polymorphic STS markers distributed over 90% of the genome (Cohen et al. 1993). The physical map was far from complete with poor coverage for some of the chromosomes. A new updated YAC library was published a few years later covering about 75% of the human genome in 225 contigs (Chumakov et al. 1995). Because of their capability to contain large clones up to a megabase or more in size, YACs have provided a powerful tool for physical mapping and analysis of complex genomes

(Schlessinger 1990, Dausset et al. 1992). The YACs also played an important role in cloning of the Huntington disease gene (Zuo et al. 1992, The Huntington's Disease Collaborative Research Group 1993). Major disadvantages with using YAC libraries are, however, the remarkably high frequency of YACs that contain two or more unrelated pieces of DNA (chimeric YACs) and the instability of some regions (Green et al. 1991). Nevertheless, the development of the YAC cloning technology has directly enhanced the relationship among genetic, physical and functional mapping of genomes facilitating the identification of genes (Larin et al. 1997).

Therefore, for higher resolution physical mapping, overlapping BAC clones have proven to be more convenient to use than the YACs, mainly due to their smaller clone inserts (in average around 130-150 kb), clone stability and lower frequency of chimerism. BAC libraries serve to integrate genetic, STS and cytogenetic map information thus offering an enormous potential for identification of chromosomal rearrangement, mapping, genomic sequencing and functional studies (Ashworth et al. 1995, Kim et al. 1996, Cai et al. 1998, Korenberg et al. 1999).

2.3.5. Identification of candidate genes

There are a great variety of different methods for finding and isolating genes in cloned DNA. A well assembled genomic contig, with no gaps or unrelated sequences, provides a good starting point for both the search of novel coding sequences and fine mapping of previously identified candidate genes. Some of the commonly used methods for identifying coding sequences in cloned DNA are hybridization of the genomic candidate DNA clone against RNA or mRNA blots (Northern blotting) (Alwine et al. 1977), cDNA libraries (Bonaldo et al. 1994, Lovett 1994) or zoo-blots (Claudio et al. 1994) or the identification and cloning of CpG islands often associated with the 5' end of vertebrate genes (Bird 1987, Larsen et al. 1992, Cross & Bird 1995). Positional transcript mapping using these methods is both difficult and time consuming. A turning point for disease gene identification has been the Human Genome Project large-scale sequencing and physical mapping of an ever-increasing number of ESTs, providing candidate genes for many human genetic diseases (Berry et al. 1995, Schuler et al. 1996, Deloukas et al. 1998).

The Human Genome Project (HGP) is an international project with the major goal to produce whole-genome genetic maps, physical maps and a complete ~3,000 megabases (Mb) sequence map covering all the human chromosomes (Lander 1996, Collins et al. 1998). In the end of 1999 chromosome 22 was the first chromosome to be completely sequenced (Dunham et al. 1999). As early as in the spring of year 2000 at least 90% of the human DNA sequence might already be ready in a "working draft" form and it is projected that by 2003 the entire human genome sequence project will be completed (Human Genome Project: www.ornl.gov/hgmis/hg5yp/). This will tremendously influence the fields of genetics and will lead to a new understanding of genetic contributions to human diseases (Collins 1999, van Ommen et al. 1999, Wadman 1999).

2.3.6. Demonstration of mutations

The ultimate proof that a candidate gene indeed is the disease-causing gene is a demonstration of mutations in affected individuals. The search for mutations in genes can be very tedious and expensive (Dean 1995, Forrest et al. 1995). Once a candidate mutation has been observed, it also must be identified in other patients with the same phenotype. Furthermore, the mutation should segregate in the affected families according to the mode of inheritance. By screening a set of healthy unrelated controls from the same population, a sequence change due to a possible neutral polymorphism may be excluded.

Mutations can be detected e.g. by direct sequencing of the DNA segment (Forrest et al. 1995), by a gain or loss of a diagnostic restriction site visualized in an agarose gel (Prosser 1993), by altered banding patterns of single stranded DNA through non-denaturing gels (single strand conformation polymorphism, SSCP) (Sheffield et al. 1993) or by resolution of heteroduplex molecules by their instability in denaturing gradient gel electrophoresis (DGGE) (Cariello & Skopek 1993). After identifying the mutation(s) the question often still remains if the mutation actually causes the disease. Functional assays may be performed to prove the connection between the disease phenotype and the mutation(s), provided that the function of the protein is known (Aittomäki et al. 1995, Forrest et al. 1995).

2.3.7. Determination of the gene structure

Gene prediction and the recognition of the exon-intron structure of the coding region, in addition to the putative corresponding promoter region, has been improved by the contribution of available computer-assisted nucleotide sequence analysis. The primary computational approach to eukaryotic promoter recognition was by combining modules recognizing different individual binding sites in the sequence, and by using some kind of an overall description of how these sites should be spatially arranged. Because of the large number of putative transcription binding sites, it has been difficult to identify promoters correctly based only on the sequence information (Fickett & Hatzigeorgiou 1997, Knudsen 1999, Pedersen et al. 1999, Werner 1999). Since the knowledge of all the mechanisms involved in transcription, translation and alternative splicing are still far from complete, both exons and promoter sequences may in spite of all the available new software tools still be wrongly predicted and partitioned (Pedersen et al. 1999). However, a combinatorial approach of general promoter prediction with exon-intron predictions may markedly improve the accuracy of promoter recognition (Werner 1999). Successful and reliable computational analyzing programs for promoter recognition would also be useful in analyzing the sequence results from the Human Genome Project (Knudsen 1999).

3. AIMS OF THE STUDY

The principal purpose of this study was to:

1) To identify and collect patients from Finland and Norway, perform genealogical studies of the Finnish patients and to map the MGA1 gene with linkage analysis
2) To refine the localization of the gene by fine genetic mapping and physical mapping
3) To characterize the mutations in the MGA1 patients
4) To determine the genomic structure of the cubilin gene
5) To use an urinary radioisotope-binding assay to diagnose MGA1 patients

4. PATIENTS AND METHODS

4.1. Identification of patients

The majority of the Finnish MGA1 patients were diagnosed in the 1960's and are not currently under regular care by a physician. Because of the lack of a main responsible physician, the primary screening for patients included in this study was based on the study of hospital records from the 1940's to today. Nearly all children with severe megaloblastic anemia in Finland have customarily been referred to the Children's Hospital of the Helsinki University Central Hospital. Due to this almost all the MGA1 patients were found in the records of this hospital. Only sporadic new cases were identified via the National Social Insurance Institution (KELA-FPA). In Finland all individuals receiving vitamin B12 treatment orally or as parenteral injections are listed in the records of the Social Insurance Institution since the cost of the treatment is partly supported by them. The search for potential MGA1 patients was limited to the initiation of vitamin B12 treatment before the age of 15.

All the Norwegian patients participating in this study were originally diagnosed by Dr. Olga Imerslund in the 1950's (Imerslund, 1959, 1960), and the blood samples for this study were collected by Dr. Harald Broch. The clinical pictures of different types of megaloblastic anemia are very similar and it is consequently difficult to distinguish them from each other. Therefore the following selection criteria were used to identify possible MGA1 patients from other patients suffering from megaloblastic anemia (Gräsbeck et al. 1960, Imerslund & Bjørnstad 1963, Anttila & Salmi 1967, Broch et al. 1984):

1. Appearance of megaloblastic anemia within the first 5 years of life
2. Low serum vitamin B12 levels with good hematologic response to parenteral injections of vitamin B12
3. Serum folate not decreased
4. Schilling tests I and II showing malabsorption of labeled vitamin B12 even after the addition of exogenous intrinsic factor (IF)
5. Unhampered absorption of other nutrients when vitamin B12 stores are replenished (vitamin B12 deficiency causes secondary malabsorption)
6. Exclusion of severe malnutrition or a general malabsorption syndrome

7. Exclusion of fish tapeworm *Diphyllobothrium latum* infection

Professor Ralph Gräsbeck kindly scrutinized the patient data and in uncertain cases decided whether the diagnosis was acceptable or not. A total of 24 Finnish MGA1 patients in 17 families were chosen for this study. Altogether 33 potential Finnish MGA1 patients, in 24 families, were identified of which 27 patients in 19 families fulfilled the study criteria above. One of the patients lived abroad and was therefore not taking part in this study, while two other patients had recently died. Following a signed informed consent from the patients, parents and healthy siblings, venous blood and urine samples were collected from participating individuals. The study was approved by the Ethical Review Committee of the Department of Medical Genetics, University of Helsinki and the Finnish National Research and Development Centre for Welfare and Health, Ministry of Social Affairs and Health (STAKES).

4.2. MGA1 families and control individuals

A total of nine multiplex MGA1 families, six from Finland and three from Norway, with more than one affected child were included in the initial linkage study. After linkage was found the panel of individuals studied was expanded to include the rest of the Finnish MGA1 families. The Finnish and Norwegian pedigrees are shown in Figure 3.

DNA samples from 158 unrelated healthy Finnish individuals were used as controls and 138 of them originated from eastern and south-central Finland, where the disease is more prevalent. The remaining 20 samples were obtained from the Finnish Red Cross.

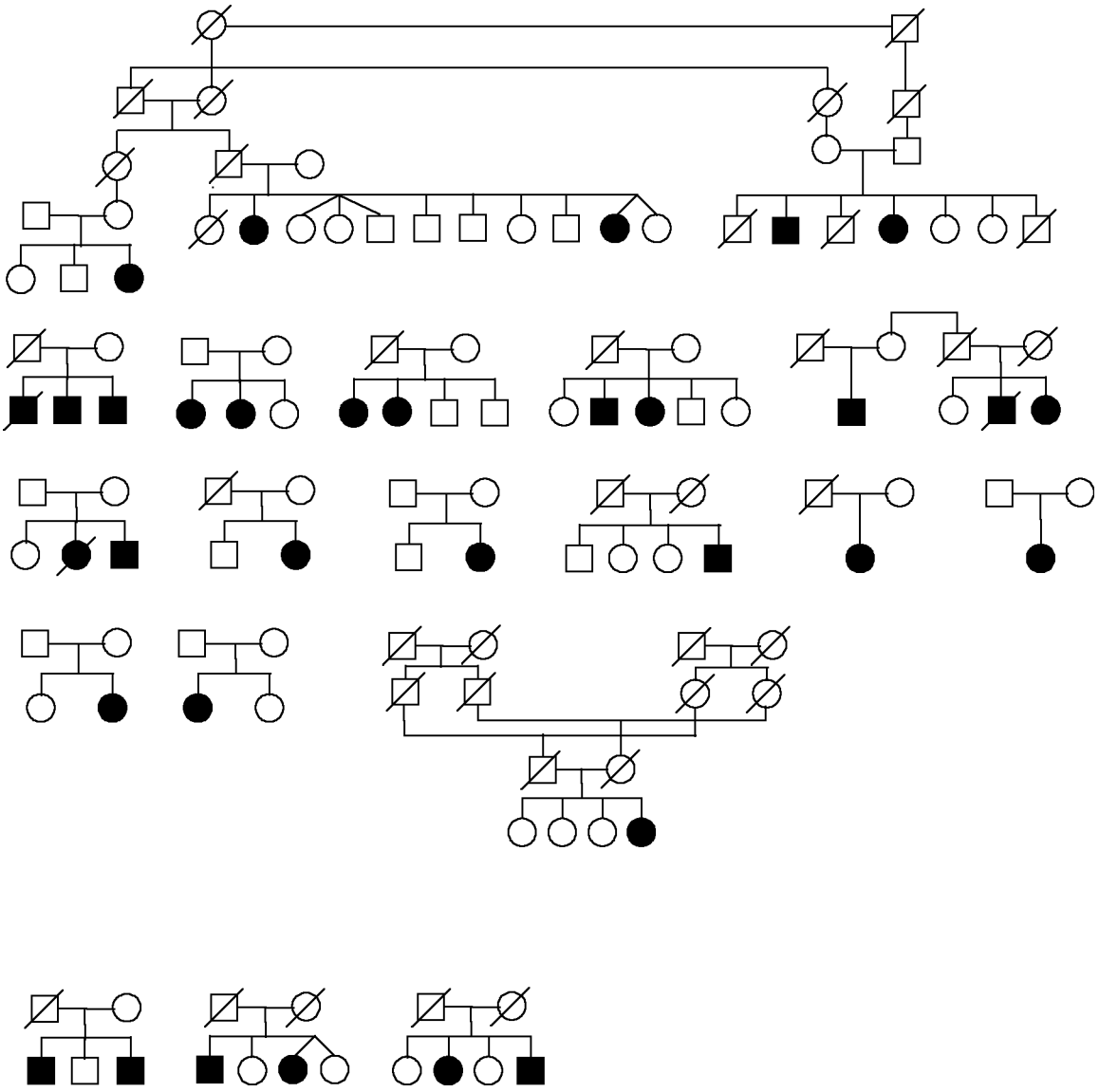

Figure 3. The pedigrees of the 17 Finnish and the three Norwegian (bottom) families used in linkage and haplotypes analyses.

4.3. Genealogical studies

A genealogical study was performed for the Finnish MGA1 patients identified for this study by examining the well established church parish record system. The church records include detailed information on birthplaces, deaths, marriages and movements for the majority of the population. The birthplaces for all the Finnish MGA1 patients' parents and grandparents and in some cases their great-grandparents could therefore be determined. Based on the information from these church registers the pedigrees of the Finnish MGA1 families were

drawn and the geographic distribution of birthplaces for the patients' grandparents was shown.

4.4. Molecular genetic studies

4.4.1. Genotyping

10-30 ml samples of venous blood were collected in EDTA or heparin from each consenting individual. DNA was extracted directly from blood leukocytes by standard methods. Most of the markers used in the initial random screening were highly polymorphic $(AC)_n$-repeats from the Généthon or Mansfield collections (Dib et al. 1996). A combination of two to six markers located ~ 20 cM apart were co-amplified in each PCR reaction using published protocols (Weber & May 1989). The α^{32}P dCTP radiolabelled co-amplified PCR fragments were separeted on 6% polyacrylamide gels and exposed to X-ray films for 1-7 days.

4.4.2. Linkage and linkage disequilibrium analyses

Linkage analyses were performed by computer programs from the LINKAGE package (Lathrop et al. 1984). The simulation program SLINK (Ott 1989, Weeks et al. 1990) was used to define a minimum number of individuals to be studied in the initial screening. Multipoint analysis was carried out using the program LINKMAP. All results were obtained under the assumption of autosomal recessive mode of inheritance, complete penetrance, with sex-average recombination fractions and allele frequencies obtained through the Genome Database (GDB) (http://www.gdb.org/) (Pearson 1991, Pearson et al. 1992, Cuticchia et al. 1993).

The haplotypes were constructed manually, assuming the least number of possible recombinations in each family. Allelic excess, the excess of an allele at a marker locus among the MGA1 chromosomes ($P_{affected}$) versus non-MGA1 chromosomes (P_{normal}) of parents or healthy siblings, was calculated according to the following formula:

$P_{excess} = (P_{affected}-P_{normal})/(1-P_{normal})$, where P denotes the allele frequency.

The P_{excess} values serve as a measure of the observed linkage disequilibrium.

4.4.3. Radiation hybrid analysis

The RH mapping panels Stanford G3 (Cox 1995) and Genebridge 4 (Gyapay et al. 1996) were used to determine the order of the microsatellite markers closest to the MGA1 locus. The PCR-amplified DNA was separated using electrophoresis in either a 1.5% agarose gel or in a 6% polyacrylamide gel. DNA was visualized with ethidium bromide or silver staining (Bassam et al. 1991).

4.4.4. Yeast artificial chromosomes

Prescreened YACs were obtained from the UK Human Genome Mapping Program Resource Center (http://hgmp.mrc.ac.uk/). Altogether 15 YACs containing any of the polymorphic markers D10S1653, D10S1763, D10S1661, D10S1477, D10S1476, D10S504, AFM234zf10, D10S1714 or D10S548 were initially identified. Total YAC DNA was prepared from colony purified YACs according to standard procedures (Hoffman & Winston 1987). The YAC contig, consisting of nine overlapping YACs between markers D10S1653 and D10S548, was used to determine the order of new STSs, ESTs and the *CUBN* intragenic markers.

4.4.5. Bacterial artificial chromosomes

Using the entire *CUBN* cDNA as a probe, 25 BAC clones were identified by screening the Roswell Park Cancer Institute (RPCI) human BAC library (http://bacpac.med.buffalo.edu). The colony purified BAC clones were grown in Luria Bertani (LB) media supplemented with chloramphenicol. BAC DNA was extracted using a modified version of the Plasmid Midiprep kit (Qiagen, Germany) protocol. A contig of 16 overlapping BACs spanning the region between the intragenic markers in the 5' and 3' region of the *CUBN* gene was assembled. The sizes of six BAC inserts used in direct sequencing were determined by the field inversion gel electrophoresis (FIGE) mapper system. BAC DNA was digested with the restriction enzymes *NotI* and *SacII*, the fragments were separeted in a 1% pulse field certified agarose gel, stained with ethidium bromide and visualized by UV fluorescence.

4.4.6. RNA extraction and cDNA synthesis

Total RNA was isolated from both fresh white blood cells and lymphoblastoid cell lines using the RNeasy RNA extraction kit or RNA Stat-60 (Chomczynski & Sacchi 1987). First strand cDNA was reverse transcribed from total cellular RNA using both random hexamer and oligo (dT) priming.

4.4.7. Sequencing and mutation analyses of the *CUBN* gene

For sequencing the 10.9 kb *CUBN* cDNA, multiple PCR primers were designed to cover the entire region in overlapping fragments. Both the 5' and 3' primers were tailed with M13-forward and M13-reverse tails, respectively, enabling sequencing in both directions. Sequencing was done using the BigDye-Terminator AmpliTaq FS Cycle sequencing kit (PE Applied Biosystems). The Sequence Navigator software (PE Applied Biosystems) was used to confirm heterozygous position and to align both directions of the respective ladders (Phelps et al. 1995). Sequence alterations were examined for restriction-site polymorphisms using the DNAStar software.

The Finnish mutation FM1 was initially identified by the sequencing of a RT-PCR amplified cDNA fragment. The screening of the FM1 mutation, C->T, in the rest of the MGA1 families and the control individuals was performed by direct sequencing of PCR amplified genomic DNA. Similarly, the second Finnish mutation FM2, a C->G transversion resulting in a splice-site mutation, was first identified from RT-PCR amplified cDNA fragments. The mutation was visualized in ethidium bromide stained agarose gels as two abnormal bands. The fragments were purified before sequencing. Screening of the FM2 mutation in the intra-CUB domain 6 intron was based on the loss of one of the two identified recognition sites for restriction enzyme *Dde1* visualized in an agarose gel.

4.4.8. Determining the exon-intron structure and a putative promoter region

Identification of the exon-intron structure of the *CUBN* gene was initiated by long-range (LR) PCR. LR-PCR was carried out using human DNA as template and Platinum Taq DNA Polymerase High Fidelity enzyme (Life Technologies, USA). Direct sequencing of the inserts of the BACs b7/b661M9, b11/b724P11, b12/b755F22, b15/b785G10, b16/b804N3 and b17/b830K8, was also performed to identify exon-intron boundaries.

A putative promoter region was identified by direct sequencing of the inserts of two BACs, b7/b661M9 and b11/b724P11. The sequence was analyzed for putative transcription factor elements using the following databases at World Wide Web: Genomatix (MatInspector professional: http://www.genomatix.gsf.de/cgi-bin/matinspector_prof/mat_fam.pl), GBF Bioinformatics (MatInspector V2.2: http://www.cbil.upenn.edu/cgi-bin/tess/Tess?_if=1&RQ=WECOME) (Quandt et al. 1995), TESS (BCM Search Launcher: Gene Feature Searches: http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html)

and BioInformatics and Molecular Analysis Section (BIMAS) (http://bimas.dcrt.nih.gov/molbio/signal). Genome-wide repeats in some of the introns and in the promoter region were identified using the National Center for Biotechnology Information Basic Local Alignment Search Tool (NCBI BLAST: http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al. 1997).

4.5. Functional studies

4.5.1. Western-blot analysis

To analyze the expression of cubilin protein, urine samples from Finnish MGA1 patients and healthy controls were collected and immediately frozen. The urine was dialyzed against water, lyophilized and resuspended in PBS. Concentrated urine samples were loaded on a non-reducing 4-16% SDS-gel and electroblotted onto Immobilon PVDF membrane. As primary and secondary antibodies a monoclonal mouse anti-human cubilin antibody (Sahali et al. 1992) and an alkaline phosphatase labeled anti mouse IgG were used, respectively.

4.5.2. Radioisotope binding assay

Urine specimens from 10 Finnish MGA1 patients from eight families, their healthy parents and siblings and 13 healthy controls were collected in the morning and kept at 4°C only for a few hours until analysis. The IF-receptor activity was measured by a radioisotope-binding assay. In the assay, free IF-B12 complexes were separated from the IF-B12 complexes bound to the receptor, by hydrophobic absorption of the receptor-IF-B12 complex to phenyl-Sepharose (Guéant at al. 1995). The urine samples were incubated with CN (^{57}Co) Cbl labeled IF before a suspension of phenyl-Sepharose was added. The amount of measured radioactivity corresponded to the urinary receptor activity for the labeled IF-B12 complex.

5. RESULTS AND DISCUSSION

5.1. Genetic assignment of the MGA1 locus

5.1.1. Linkage studies map the MGA1 locus to chromosome 10p (study I)

The 24 Finnish patients in 17 families that were accepted by Professor Ralph Gräsbeck as MGA1 patients, in addition to the three Norwegian families, were chosen for this study.

The linkage study was initially performed with six multiplex Finnish MGA1 families. After typing less than one hundred markers, a significant lod score value was first detected with marker D10S191 on chromosome 10p. At this point, also the three Norwegian multiplex families were included in the study. Additional markers were studied to confirm linkage and the highest lod scores were obtained with two markers centromeric to D10S191, D10S1476 and D10S1477. In addition a multi-point linkage analysis between the MGA1 locus and the seven closest marker loci localized the MGA1 gene closest to marker D10S1477 with a maximum lod score of 5.36. The use of multi-point linkage analysis did not provide any additional information suggesting that the marker D10S1477 and the MGA1 loci were not in strong linkage disequilibrium, which was also demonstrated by a low P_{excess} value (P_{excess} 0.53). Nonetheless, as a result from one of the Norwegian families, where the younger affected twin displayed a recombination, the gene was at this point placed more telomeric to marker D10S466. As the following step, the additional eleven Finnish MGA1 families with a single affected child were analyzed confirming linkage to chromosome 10p.

5.1.2. Linkage disequilibrium and haplotype analyses (studies I, II)

After analyzing an additional number of new polymorphic markers in the region, two of them AFM234zf10 and D10S504, showed a highly significant allelic association with MGA1 in the Finnish families. This result further refined the region to about 4 cM and once more switched the localization of the MGA1 gene centromeric to D10S1477. The area of strong linkage disequilibrium was very limited implying a quite old mutation for MGA1. As a result the younger Norwegian twin was considered not affected and her vitamin B12 treatment was discontinued under supervision of her physician. The most likely order and distance between the markers in the region according to the Whitehead institute/MIT Center for Genome Research (http://www-genome.wi.mit.edu/) and our haplotype and RH data are shown in Figure 4.

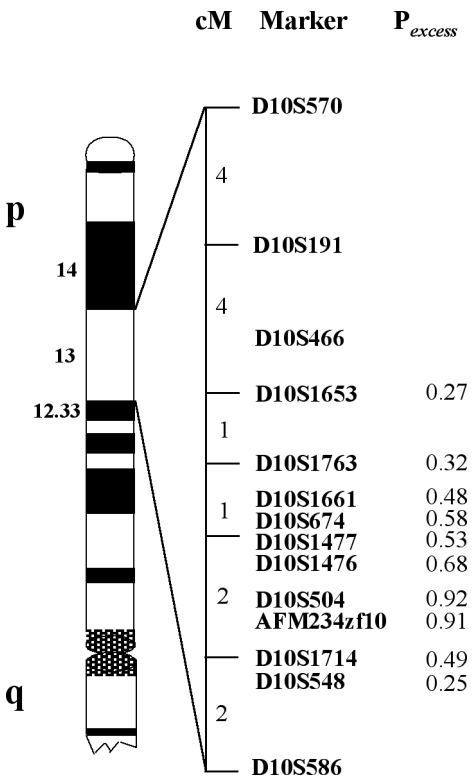

Figure 4. The genetic map of chromosome 10p is showing the relative position of the marker loci in the MGA1 region. P_{excess} values are also indicated for some of the markers closest to the MGA1 locus.

Initially, extended haplotypes over the MGA1 region were constructed with nine markers spanning ~14 cM between the markers D10S570 and D10S586 in the Finnish and Norwegian multiplex families. Later more markers were studied and added to the haplotypes. By genetic and linkage disequilibrium mapping, the critical region was limited to an interval less than 2 cM between markers D10S1476 and D10S548. The use of linkage disequilibrium has shown to be a powerful tool for high-resolution mapping of genes in isolated populations. The method has successfully been applied as guidance for the mapping and cloning of several autosomal recessively inherited disease genes in the Finnish population (de la Chapelle 1993, Hästbacka et al. 1994*,* Vesa et al. 1995, Höglund et al. 1996, de la Chapelle & Wright 1998, Peltonen 2000). Later a candidate gene encoding the intrinsic factor receptor (IF-R) precursor, cubilin was cloned by a functional approach and mapped to the same region by a Danish-French group (Kozyraki et al. 1998). After the gene for MGA1 was identified, five novel intragenic markers, four SNPs and a $(AC)_n$ repeat, were found when sequencing the gene and added to the haplotypes. The region with strong allelic association associated with FM1 was

very limited based on historical recombinations, implicating a very old Finnish mutation. This interpretation is consistent with our haplotype data, which suggested that most of the Finnish disease chromosomes carry the same ancestral mutation.

Table 2. Finnish haplotypes associated with the MGA1 gene.
The intragenic markers are FM2, FM1, CUB15P, CUB15IP and CUB27I(AC)n. CUB15P and CUB15IP are SNPs in CUB domain 15 and in intra CUB15 intron. Exonic nucleotides are indicated in capital letters and intronic nucleotides in small letters. CUB27I(AC)n is a highly polymorphic (AC)n repeat in the last intron of the cubilin gene.
(R: observed recombination events, NA: not applicable; ND: not determined)

Table 2a

		TEL		Markers		CEN
Mutation		D10S1476	**D10S504**	**AFM234zf10**	**D10S1714**	D10S548
p_{excess}	**N**	0.68	0.92	0.91	0.49	0.25
FM1	**6**	1	8	3	6	3
FM1	**1**	1	ND	3	6	3
FM1	**1**	1	8	3	**R**	3
FM1	**5**	1	8	3	6	1
FM1	**3**	1	8	3	8	4
FM1	**3**	1	8	3	14	1
FM1	**3**	2	8	3	6	1
FM1	**1**	1	8	3	3	3
FM1	**1**	5	8	3	3	3
FM1	**1**	5	8	3	7	1
FM1	**1**	2	8	3	8	2
FM1	**1**	2	8	3	8	4
FM1	**1**	1	9	3	5	1
FM1	**1**	1	4	6	6	4
FM1	**1**	5	8	6	2	2
FM1	**1**	2	ND	3	1	2
Total	**31**					
FM2	**2**	2	9	3	7	2
FM3	**1**	3	7	6	12	1

Table 2b Mutation		D10S504	AFM234zf10	FM2 (CUB6)	FM1 (CUB8)	CUB15P	CUB15IP	CUB27I(AC)n	D10S1714
p_{excess}	N	0.92	0.91	NA	NA	ND	ND	0.79	0.49
FM1	**12**	8	3	c	T	G	a	7	6
FM1	**1**	ND	3	c	T	G	a	7	6
FM1	**4**	8	3	c	T	G	a	7	8
FM1	**2**	8	3	c	T	G	a	7	14
FM1	**1**	8	3	c	T	G	a	7	7
FM1	**2**	8	3	c	T	G	a	7	3
FM1	**1**	9	3	c	T	G	a	7	5
FM1	**1**	4	6	c	T	G	a	7	6
FM1	**1**	8	6	c	T	G	a	2	2
FM1	**1**	8	3	c	T	G	c	7	14
FM1	**1**	8	3	c	T	G	c	8	6
FM1	**1**	8	3	c	T	G	c	8	8
FM1	**1**	8	3	c	T	C	c	7	6
FM1	**1**	8	3	c	T	C	c	8	6
FM1	**1**	ND	3	c	T	C	c	8	1
Total	**31**								
FM2	**2**	9	3	g	C	C	c	4	7
FM3	**1**	7	6	c	C	C	c	2	12

Table 3. Norwegian haplotypes associated with the MGA1 gene.
The intragenic markers are FM2, FM1, CUB15P, CUB15IP and CUB27I(AC)n. CUB15P and CUB15IP are SNPs in CUB domain 15 and in intra CUB15 intron. Exonic nucleotides are indicated in capital letters and intronic nucleotides in small letters. CUB27I(AC)n is a highly polymorphic (AC)n repeat in the last intron of the cubilin gene.
(R: observed recombination events, ND: not determined)

Table 3a					
Norway	D10S1476	**D10S504**	**AFM234zf10**	**D10S1714**	D10S548
FamA	2	ND	3	7	1
FamA	2	ND	3	12	2
FamD	5	ND	6	3	**R**
FamC	5	ND	6	6	4
FamC	2	ND	6	8	3
FamD	5	ND	13	7	1

Table 3b								
Norway	**D10S504**	**AFM234zf10**	FM2	FM1	CUB15P	CUB15IP	CUB27I(AC)n	**D10S1714**
FamD	ND	3	c	C	C	c	7	7
FamA	ND	3	c	C	G	c	7	12
FamC	ND	6	c	C	C	c	4	3
FamC	ND	6	c	C	C	c	4	6
FamA	ND	6	c	C	G	c	4	8
FamD	ND	13	c	C	G	c	4	7

Differences in the disease haplotypes between the Finnish and Norwegian suggested more than one common mutation. Our haplotype data suggested that most of the Finnish disease chromosomes carried the same ancestral mutation, while at least two Norwegian mutations were predicted based on their haplotypes.

5.2. Physical mapping

5.2.1. Yeast artificial chromosome contig (study II)

Since the region displaying linkage was quite large, the use of YAC clones was the only option to assemble a physical map for positional cloning. A YAC contig of nine YACs was constructed over an approximately 4-cM region between the flanking markers D10S1653 and D10S548. The adjacent and partially overlapping YACs were ordered by STS and EST content mapping. The STS content mapping identified four overlapping segments for the markers AFM234zf10 and D10S504 showing the highest allelic association with the MGA1 locus. The Mega YACs from CEPH ranged from 270 kb to 1780 kb in size and were all reported to be quite chimeric except for one (YAC-12). Because of the high degree of chimerism, the YAC contig was only used for refined mapping of new ESTs in the absence of any obvious candidate genes in the region.

5.2.2. Bacterial artificial chromosome contig (studies II, III)

In order to determine the genomic structure of the *CUBN* gene, a BAC contig was established covering the entire gene. The BACs are usually more stable than the YACs, due e.g. to smaller inserts, and are also very useful since they can be used for direct sequencing. The human BAC library was initially screened using 9.4 kb of the 10.9 kb *CUBN* cDNA as probe. Altogether 16 overlapping BAC clones were identified that were positive for the five intragenic markers in the *CUBN* gene. The BAC contig was then screened with the closest STS markers in the region. However, none of the BACs were positive for the STS markers flanking the gene, indicating that the distance between the markers and the gene was quite large.

5.3. A candidate gene

5.3.1. Cubilin - a functional and positional candidate gene for MGA1 (study II)

A functional candidate gene, *CUBN*, for MGA1 was identified by a functional approach and mapped to the same 6-cM chromosomal region by a Danish-French research group (Kozyraki et al. 1998) as we earlier identified by recombinant-based linkage analysis. The *CUBN* gene was encoding the receptor, cubilin (previously called gp280), present in the epithelium of intestine and kidney and the yolk sac (Moestrup et al. 1998). Previous studies had already indicated that the 460-kDa gp280 protein may act as a receptor for the IF-B12 complex and that gp280 and the IF- receptor in fact are identical proteins (Birn et al. 1997, Seetharam et al. 1997). The unique structure of cubilin showed a stretch of about 110 amino acids followed by a cluster of eight epidermal growth factor repeats (EGF) and a large cluster of 27 CUB domains (an abbreviation of Complement subcomponents Clr/Cls, Uegf and Bone morphogenic protein-1, Bork & Beckmann 1993) and lack of a typical N-terminal transmembrane segment. Due to the high amount of CUB domains the receptor was named cubilin (Moestrup et al. 1998). The assembling of cubilin has been studied in the bovine homologue, and it has been proposed that the receptor is organized into a trimer of identical subunits (Lindblom et al. 1999).

The CUB domain structure contains two layers of five-stranded antiparallel β-sheets connected with β-turns (Bork & Beckmann 1993, Romero et al. 1997). CUB domains are mainly seen in a few copies in developmental proteins, such as spermadhesins, talloid protein and bone morphogenetic protein-1 that comprise of multiple EGF and CUB binding domains (Bork & Beckmann 1993). The cubilin structure is although unique due to the very high number of 27 CUB domains.

Since cubilin has no classical transmembrane segment it might require a coreceptor. The colocalizing 600-kDa transmembrane transcobalamin-vitamin B12 receptor megalin (gp330) is suggested to carry out that function (Moestrup et al. 1996, Birn et al. 1997, Christensen et al. 1998, Moestrup et al. 1998). No binding of the IF-B12 complex direct to megalin has been shown (Birn et al. 1997). Megalin and cubilin are both expressed in the endocytic vesicles of the absorptive epithelia in the intestine, kidney and yolk sac (Birn et al. 1997, Moestrup et al. 1998).

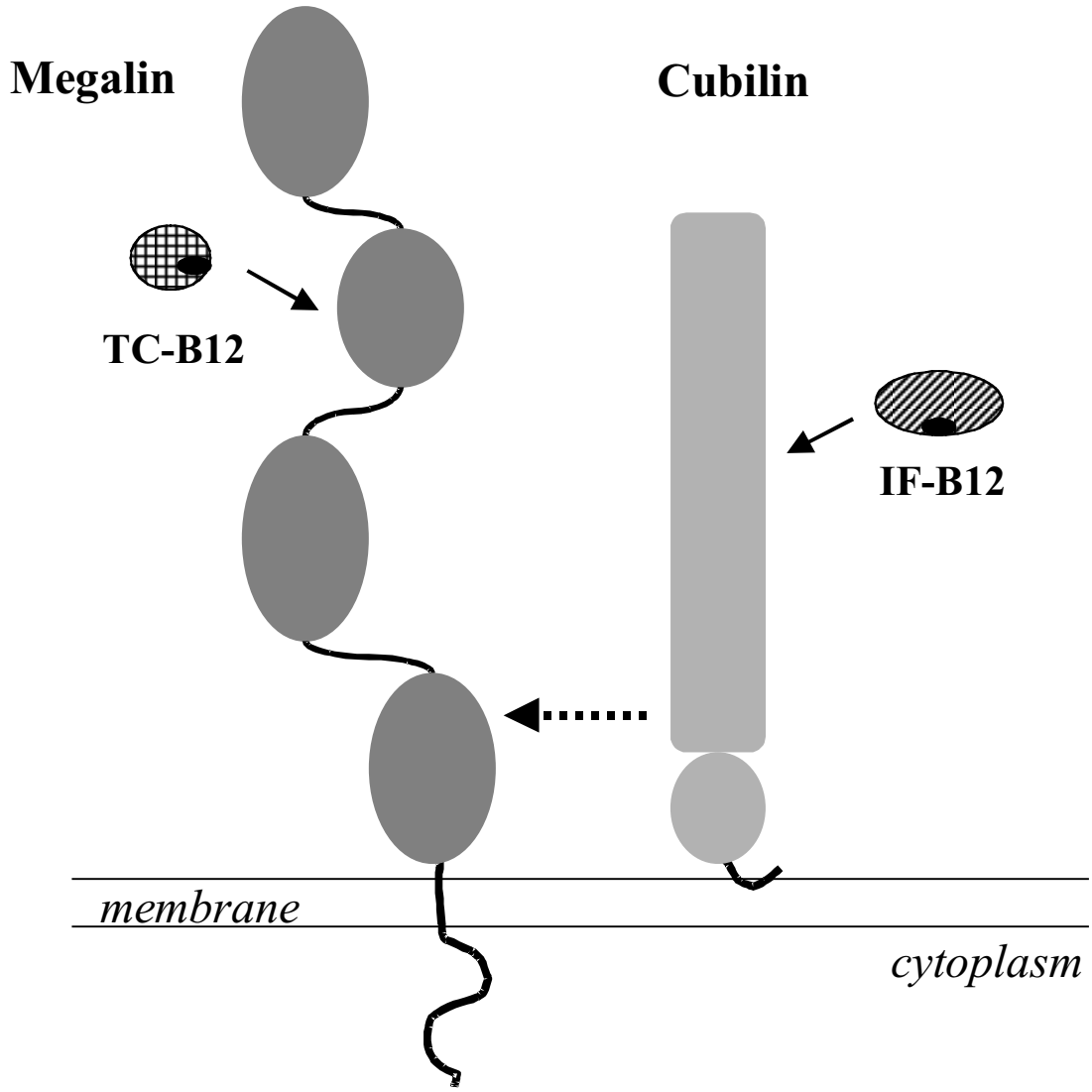

Figure 5. Epithelial receptors important for vitamin B12 uptake into the cell.
(Picture provided by Søren K. Moestrup)

The existence of 27 CUB domains in cubilin indicates a high number of putative sites for various protein interactions. Cubilin binds, as megalin does, the receptor-associated protein (RAP) (Birn et al. 1997, Kristiansen et al. 1999, Birn et al. 2000) and cubilin is suggested to play a role in the endocytosis and trafficking of immunoglobulin light chains in renal proximal tubule cells (Batuman et al. 1998). Additional studies have shown that cubilin also acts as a high-affinity apolipoprotein A-I receptor facilitating endocytosis of high-density lipoproteins (HDL) in the kidney (Hammad et al. 1999, Kozyraki et al. 1999).

Cubilin as many other CUB domain proteins also seems to play a role in embryonic development. Previous experimental studies performed in rats have shown that antibodies raised against cubilin induced fetal malformations including central nervous system defects (Sahali et al. 1992). Disruption of the gene for megalin in mice also causes severe central nervous system malformations that are associated with a decreased lipoprotein uptake in the early neuroectoderm and visceral yolk sac (Willnow et al. 1996). Other studies support a functionally significant association between cholesterol metabolism and the embryonic development of the central nervous system (Herz et al. 1997).

The *CUBN* gene was PCR screened with an intragenic marker against the established YAC contig and showed to be positive for the same YAC clones as the two markers, AFM234zf10 (P_{excess} 0.91) and D10S504 (P_{excess} 0.92), showing the strongest allelic association with MGA1. Consequently, the putative functional candidate gene also turned out to be a strong positional candidate gene for MGA1.

5.4. Mutation analyses

5.4.1. Mutational analyses of the *CUBN* gene (study II)

To test the assumption that mutations in *CUBN*, encoding the IF-receptor, cause hereditary megaloblastic anemia 1 (MGA1), RT-PCR and direct sequencing were performed with two MGA1 patients (one Finnish and one Norwegian), one of their children (heterozygous for MGA1) and one unrelated healthy control from both the Finnish and Norwegian population. Additionally, a Southern-blot analysis was carried out in the Finnish and Norwegian MGA1 families to exclude a possible chromosomal misalignment of the gene due to its highly repetitive nature.

Two different mutations in the *CUBN* gene were identified in the Finnish MGA1 families. The first Finnish mutation (FM1) identified was a point mutation, a C->T transition at nucleotide 3916 in CUB domain 8. This missense mutation causes an amino acid change from proline to leucine (P1297L), which causes an impaired recognition of the IF-B12 complex by cubilin (Kristiansen et al. in press). Patients from the majority of the Finnish families, 15 out of 17, were homozygous for this mutation, whereas one family was compound heterozygous and another was homozygous normal for FM1. In 316 control chromosomes, one FM1

heterozygote was identified indicating a 0.003 carrier frequency, which is in accordance with earlier estimations (Furuhjelm & Nevanlinna 1973). The proline residue is highly conserved in multiple CUB domains in the cubilin protein both in human and rat, indicating that it is functionally important (Bork & Beckmann 1993, Moestrup et al. 1998, Kozyraki et al. 1998). The cubilin receptor is otherwise also quite conserved among the two species showing a 69% identity (Kozyraki et al. 1998).

The second Finnish mutation, FM2, was a C->G transversion that resulted in a complex splice mutation by activating a cryptic splice donor site which results in an in-frame insertion of multiple stop codons in the intra CUB domain 6 intron. In contrast to the patients homozygous for FM1, no cubilin protein was detected with Western blot analysis in the urine sample from the patient homozygous for FM2. This suggested that FM2 might alter the stability of the mutant FM2 cubilin transcript. Additional studies have also shown that the FM2 patient had a high urinary apoA-I excretion level and an elevated fasting HDL-cholesterol concentration. In opposite, no apoA-I was detected in the urine of the two FM1 patients studied, indicating that these patients take up apoA-I in the kidney (Kozyraki et al. 1999). The FM2 mutation was identified in only one Finnish MGA1 family where the patient's parents were first cousins both paternally and maternally. No carriers were observed in either the Finnish control samples or the three Norwegian MGA1 families. Thus, it was concluded that FM2 constitutes a private mutation that most likely is unique and occurred relatively recently. Since one of the Finnish MGA1 was compound heterozygote for FM1, this indicates that there has to be a third Finnish mutation, FM3. Clinically, the FM1, FM2 and FM1/FM3 patients are not distinguishable.

Recognition of the binding site for IF-B12 to the CUB domains 5-8 in cubilin relates to the identification of the two mutations, FM1 and FM2. The two Finnish mutations are in the same critical region in cubilin that binds the IF-B12 complex (Kristiansen et al. 1999).

FM1 accounts for 31 of 34 (91%) disease chromosomes and constitutes the major Finnish MGA1 mutation. For many Finnish diseases for which the molecular defect has been identified, over 90% of disease alleles have shown to carry the same causative mutation (de la Chapelle & Wright 1998, Peltonen et al. 1999). In spite of sequencing the entire 10.9 kb *CUBN* coding region, as well as substantial portions of the intervening intronic sequence, in

three patients and one carrier from the Norwegian multiplex MGA1 families, no mutation has been found.

The genetic basis of selective vitamin B12 malabsorption with proteinuria has also been studied in a canine model (giant schnauzer dogs) closely resembling the human MGA1 phenotype (Fyfe et al. 1989, Fyfe et al. 1991a). The canine *CUBN* gene was shown to be highly identical with the human (83% identity) and the rat (70% identity) cubilin genes. Linkage to the canine *CUBN* gene was however rejected in the canine family indicating that the disorder resembling MGA1 is caused by a gene defect in a distinct gene product other than cubilin (Xu et al. 1999).

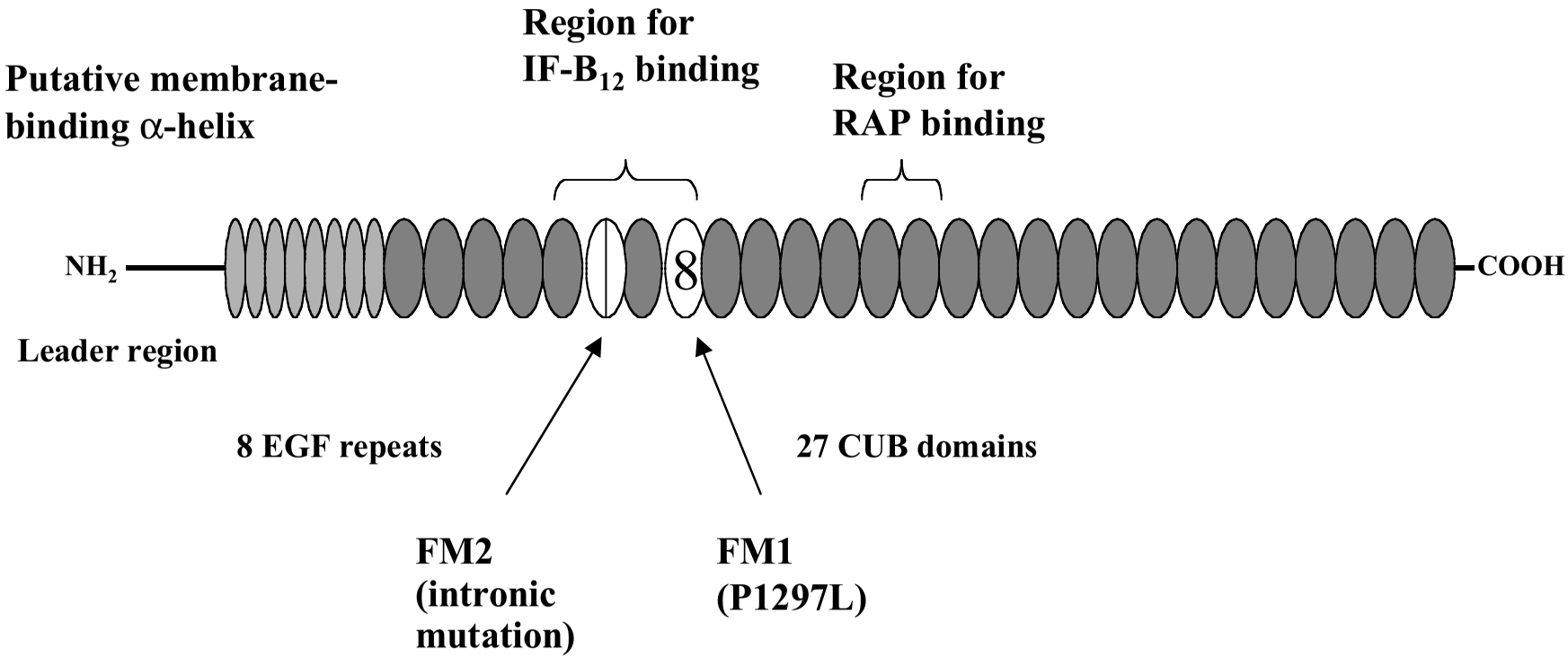

Figure 6. The cubilin protein with the two Finnish mutations FM1 in CUB domain 8 and FM2 in CUB domain 6 intron.
(Picture provided by Søren K. Moestrup)

5.5. **The genomic structure of the human cubilin gene** (study III)

The complete genomic structure of *CUBN* was determined by both LR-PCR and by direct sequencing of BAC clones that were positive for *CUBN* intragenic markers. The cubilin gene consists of 67 exons and 66 introns spanning about 170 kb. The leader (L) region is interrupted by three introns while the fourth intron is located in the junction between the leader region and the first EGF domain. The following EGF domains are all divided by introns except for EGF8. The next intron after the EGF repeats is located in the junction between EGF8 and the first CUB domain. The exon-intron pattern seems to be the same in almost every one of the following 27 CUB domains. All CUB domains are preceded by an intron and interrupted by an internal intron, except for CUB domain 11 that is interrupted by two internal introns (Figure 7). The size of the introns varied from 159 bp to about 11.3 kb.

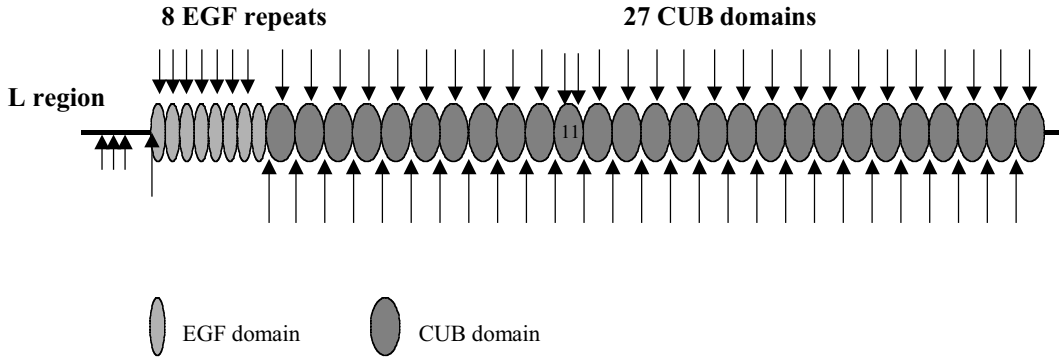

Figure 7. The genomic structure of the cubilin gene showing the distribution of the 66 introns (arrows), 3 introns in the leader (L) region, 1 intron before the first EGF repeat, 7 introns in the 8 EGF domains, 1 intron between the last EGF domain and the first CUB domain and 54 introns in the 27 CUB domains. (Picture provided by Søren K. Moestrup and modified)

No obvious TATA box was detected in the putative promoter sequence. Instead several other potential transcription-regulation sites, such as Sp1 binding sites, were identified. Moreover, binding sites for other common and more tissue specific were found. In addition to a possible CCAAT box and a CCAAT/enhancer binding protein site, the following transcription factor elements that are both tissue restricted and tissue specific were identified: different GATA factors, Lmo2/Tal-1/E2A complex and hepatocyte nuclear factors HNF-1 and HNF-4. The

human megalin/LRP-2 promoter has also recently been characterized (Knutson et al. 1998). Although cubilin and megalin are expressed in the same tissues, and presumably belong to the same family of receptors, their promoters only partly resemble each other. Both promoters have Sp1 binding sites, however the cubilin promoter lacks the atypical TATA element that seems to be essential for LRP-2 promoter activity (Knutson et al. 1998).

5.6. **Urinary assay of the IF-receptor** (study IV)

Previous studies have identified several different cobalamin-binding proteins in urine. These proteins include IF, haptocorrin (HC) and the IF-receptor (Gräsbeck et al. 1982, Wahlstedt & Gräsbeck 1985, Guéant et al. 1995). The IF-receptor activity in urine is measurable and correlates well with the receptor activity in ileal biopsies (Guéant et al. 1995, Safi et al. 1995), where similarly decreased or absent receptor activity has been observed in MGA1 patients (Burman et al. 1985).

In this study the receptor activity was measured in 10 Finnish MGA1 patients, 11 of their first-degree relatives and 13 healthy control individuals. The study was performed with the same patients used in the linkage analysis but before the cubilin gene had been cloned and the mutations identified. The receptor activity in the urine of the MGA1 patients was extremely low or almost undetectable compared with the healthy relatives and the controls. Similar results have also been observed by a French research team (Guéant et al. 1995, 1999). No distinction in receptor activity was observed between the FM2 and the FM1 patients. Differences in receptor activity were not observed among the heterozygote carriers and the healthy controls studied. Therefore this method might be reliable for diagnostic purposes but it is not able to detect possible carriers for the disease.

Further studies with the same Finnish MGA1 patients showed that they excrete significantly higher amounts of IF in the urine than the control individuals. The elevated values of IF in the urine may be caused by the absence of a functional IF-receptor in the kidney (Dugué et al. 1999). Patients from Kuwait have also been tested for receptor activity in urine but did not, in contrast to the Finnish patients, show any decrease of receptor activity and actually in one family there was increased excretion (Gräsbeck 1997, Dugué et al. 1999). Increased amounts

of an unstable IF-receptor detected in ileal biopsy specimens, has also been reported (Eaton et al. 1998).

5.7. **Genealogical studies** (unpublished data)

The MGA1 disease gene is predominantly clustered in the south-central and eastern parts of Finland. The distributions of the birthplaces for the grandparents of the MGA1 patients' participating in this study are shown in Figure 8. These two regions were settled in two expansions during the early settlement (old Finland) in which the population began to expand some 2,000-2,500 years ago and late settlement (new Finland) in which population expansion started mainly approximately 500 years ago in the 1500's (Norio et al. 1973). The presence of the MGA1 gene in both the old and new Finland indicates that the mutation is of a relatively old origin. In cases of autosomal recessive disorders where the mutations are geographically widespread, the founding has occurred a long time ago (de la Chapelle 1993). The major Finnish mutation FM1, most likely carried by the original settlers, is mainly seen in the parts of Finland that were inhabited later (new Finland) as a result of internal migration. In contrast to FM1, the two minor Finnish mutations FM2 and FM3 are observed in the old part of Finland along the coastline. These mutations may either have arisen from private mutations or arrived with new immigrants.

The church records have served as a population register for the majority of the population for about the last 10 generations (Norio et al. 1973) and made it possible to investigate for consanguinity between the MGA1 families. In an earlier genealogic study performed by Furuhjelm and Nevanlinna, 24% remotely consanguineous marriages between parents of MGA1 patients were identified (Furuhjelm & Nevanlinna 1973).

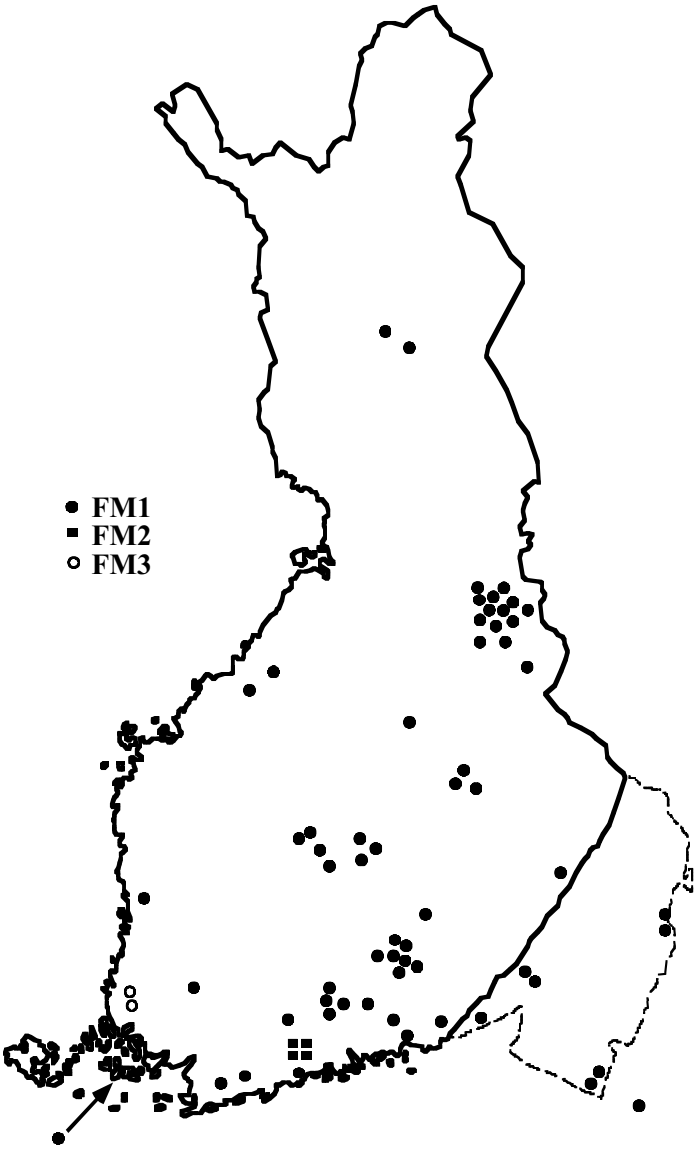

Figure 8. Geographic distribution of birthplaces for the MGA1 patients' grandparents.
Filled circles and squares indicate FM1 and FM2 grandparents, respectively. The grandparents for the postulated third mutation (FM3) are also indicated (open circles).

6. CONCLUDING REMARKS

At the beginning of this study the molecular background of MGA1 was not known, although evidence had been presented for the existence of an abnormal or a total lack of the intestinal receptor for the intrinsic factor-vitamin B12 complex. More than 30 years had passed since the disease was first described and the lack of new patients had even lead to speculations that the disease does not exist at all or that it does not show Mendelian inheritance. This study describes the identification of the Finnish MGA1 patients, genetic and physical mapping of the MGA1 gene and identification of the mutations causing MGA1.

The MGA1 locus was mapped by genetic linkage analysis to chromosome 10p in Finnish and Norwegian MGA1 families. A very strong candidate gene for MGA1, the one encoding the IF- receptor, cubilin, was cloned by a Danish-French research team and mapped to the MGA1 region. Two mutations were identified in the MGA1 gene in the Finnish population. The mutations include one missense mutation (FM1: P1297L) and one complex splice mutation (FM2). One of the MGA1 patients was compound heterozygote for FM1 suggesting a third Finnish mutation (FM3). However, no mutations were identified in the cubilin gene in the Norwegian MGA1 patients. Based on the mutation analysis of the cubilin gene in the Finnish and Norwegian patients it seems very possible that the same phenotype may result from one of several genetic errors. The Norwegian patients may have mutations in the promoter region which has not yet been analyzed, in some of the introns or in an another gene encoding a protein that is involved in the intracellular transport of vitamin B12 and is located in the near vicinity of the MGA1 gene. The presence of different haplotypes in the three Norwegian MGA1 families supports the hypothesis that another gene might be involved, although the Norwegian patients show linkage to the MGA1 locus and there is no evidence of an additional gene. Because of the complicated multistep nature of vitamin B12 absorption from the lumen of the intestine to the blood, genetic diversity in the disease pattern may very likely occur. All the steps involved in the transference of vitamin B12 across the enterocytes in the ileum are not completely understood and it is therefore very possible that more than one gene defect may cause MGA1.

The disease has also been frequently diagnosed in Turkey and several Arabian countries (Yetgin et al. 1983, Abdelaal & Ahmed 1991, Altay et al. 1995, Ismail et al. 1997). Those who studied the Saudi Arabian and Kuwaiti cases suggested an X-linked mode of inheritance

because of the abundance of affected males (Abdelaal & Ahmed 1991, Ismail et al. 1997). Considering the multistep nature of vitamin B12 absorption, heterogeneity in the genetic basis of the disease with a possible presence of an X-linked form would not be unexpected in the Arabian families (Ismail et al. 1997). A difference in penetrance between the Scandinavians and the Arabians may be another explanation why only males are affected, since in both materials the birth rate seems to be equal in males and females. Genetic heterogeneity may in part explain this syndrome. The differences in the expression levels of the IF-receptor in different populations support this assumption. While a decreased IF-receptor activity was observed in the urine from the Finnish MGA1 patients, no decrease and even a significant increase (in one family) of the urinary receptor activity was detected in the patients from Kuwait (Dugué et al. 1999). Overexpression of an unstable IF-receptor has also been detected in ileal biopsy specimens from two sisters suffering from MGA1 (Eaton et al. 1998).

Although the Finnish MGA1 mutations have been identified, the question still remains why almost no new patients have been diagnosed since the 1970's and 1980's in Finland and Norway. In contrast to the Arabian patients, where differences in the penetrance may occur, no evidence of nonpenetrance was observed among the Finnish and Norwegian families studied. In the Finnish MGA1 families none of the parents or the healthy siblings were homozygous for either the FM1 or the FM2 mutation thereby excluding nonpenetrance in these families. In the Norwegian families, on the other hand, where no mutation has been identified, nonpenetrance was excluded based on the haplotype analysis alone where none of the healthy relatives studied were homozygous for the affected haplotypes.

One explanation why the disease has "disappeared" may be a drop in the gene frequency due to migration to the bigger cities from isolated subpopulations, where most of the MGA1 patients originate. In the cities the gene frequencies would be lower due to a dilution of genes and homozygosity becomes very rare. In the rural high-incidence MGA1 areas in the northeastern parts of Finland the number of births has significantly dropped since the 1950's to present. In the eastern parts of Finland the numbers have dropped considerably from 19,610 to 8,754 births per year (~55%) between 1950 and 1970, while the decline of births in the whole country during that time was ~66% (from 98,065 to 64,559 births per year). Today little more than 6,000 children are born per year in eastern Finland and about 57,000 in the whole country. In the northeastern part of Finland, Kainuu-Kajanaland, which is a sparsely populated high-incidence MGA1 region, the number of births has decreased as well. While

the drop of births in the whole Finland was about 42% between 1950 and 1998, the number of births in Kainuu-Kajanaland dropped from 2,043 births per year to 873 births per year, (a 57% decline), during the same time period. During the corresponding time from the 1950s to today the number of births in southern Finland have, on the other hand, increased about 25% from 12,503 to 15,551 births per year (Official Statistics of Finland 1997, 1999). A decrease in the number of births in the rural high-incidence MGA1 areas could partly explain a decline in the number of new patients but not the present absence of patients. Another explanation would be that possible new patients suffering from megaloblastic anemia are treated with vitamin B12 without further examinations, which seems unlikely. As a result of the identification of the Finnish mutations, patients suffering from megaloblastic anemia can now relatively easily be tested for MGA1.

Drastic changes in environmental factors like the dietary habits among Finns and Norwegians during the last fifty years might have influenced the penetrance of the gene. This hypothesis could principally be very easily tested by population studies where healthy homozygotes for a MGA1 mutation should be found. However, a population study can be quite difficult to accomplish in practice since the amount of individuals studied has to be quite large due to the low frequency of the gene in both the Finnish and the Norwegian populations.

7. ACKNOWLEDGEMENTS

This study was carried out during 1994-2000 at the Folkhälsan Institute of Genetics, Department of Medical Genetics, University of Helsinki and at the Division of Human Cancer Genetics, Comprehensive Cancer Center, the Ohio State University. I wish to express my gratitude to the former and present heads of the Department of Medical Genetics, Helena Kääriäinen, Juha Kere, Leena Palotie and Pertti Aula for providing me with excellent research facilities.

I am thankful to the heads of Division of Genetics, Department of Biosciences, Olli Halkka and Jim Schröder and all my other teachers for introducing me to the interesting world of genetics.

I specially want to thank my supervisor Albert de la Chapelle for his endless enthusiasm for the research, for his support and encouraging guidance throughout this study and for teaching me never to give up, some things are just worth fighting for.

I also want to thank the other key person in this project, Ralph Gräsbeck who always knew that "his" disease really exists, for his valuable help in identifying the MGA1 patients and for his interest towards my work. Albert and Ralph were the only ones that right from the beginning believed in this project when I was searching for the MGA1 patients and were very supportive when many others were quite skeptical.

I would like to express my sincere gratitude to all my collaborators. Harald Broch for his enthusiasm to this study right from the beginning and for collecting the Norwegian MGA1 samples. Søren Moestrup for a very successful collaboration and Søren Moestrup and Pierre Verroust together with collaborators for identifying an excellent candidate gene for MGA1, it made my life much easier. Benoît Dugué for a good friendship that started already when I was doing my master thesis in Ralph Gräsbeck's group at the Minerva Foundation Institute for Medical Research.

I wish to thank Maija Wessman and Erkki Elonen for reviewing the manuscript and for their constructive criticism and helpful suggestion that helped me improve it.

I wish to thank Anna-Elina Lehesjoki for always being very positive, encouraging and confident that everything will turn out to the best in the end and for letting me use her room while she was in London.

My warmest thanks go to Sinikka Lindh for travelling around collecting blood and urine samples from the patients wherever they lived and for later being an indispensable link between the MGA1 patients and the research.

In autumn 1997 this project was transferred to the Division of Human Cancer Genetics at the Ohio State University where I was continuing my work in Ralf Krahe's research group. During the 18 months I spent in Ohio I had the opportunity to learn new things in the lab which I thank Ralf Krahe for. Furthermore I would like to thank Bob Chadwick and Cheryl Johnson and the whole sequencing lab for very beautiful sequences.

I also got many new friends during my stay and I would specially like to thank Suzanne Brady who made me feel a little bit like home in Ohio, Jody Carter who for a short period was involved in this project, and all the boys in the lab, Shanxiang Zhang, Jeff Palatini and Tim Wise, with whom I had many interesting discussions. The social life outside the lab was of course of great importance too. I have many nice memories from the ladies' evenings with both a lot of gossiping and good food in the nice company of Manisha Shah, Susan Whitman, Natalia Pellegata, Marilotta Turunen, Aranzazu de la Puente, Julie Eisel and the rest of the gang.

After my time in Ohio I moved to Annapolis, Maryland, where I was kindly provided with an office space at Kvaerner Masa Marine. I would like to thank John Avis for this opportunity and the rest of the people in the office for a nice working environment and for including me in the "sandwich gang".

I would like to thank my former roommates at Folkhälsan Maaret Ridanpää and Katarina Pelin for all the patience they have showed me every time I have been knocking on their door for help and advice.

I warmly thank all my "old" and "new" friends in the Folkhälsan lab. Ann-Liz Träskelin who was the first person whom I got to know in the Folkhälsan lab and who introduced me to the "real lab life". Tarja Joensuu and José Dieguez with whom I have been sharing many incredible dog-stories, for their encouragement and José Dieguez for his valuable comments to my fourth article. Everybody who are or have been in the same soup as me struggling with their thesis, Kristiina Avela, Susanna Ranta, Kati Donner, Kimmo Virtaneva, Laura Huopaniemi, Juha Kolehmainen, Kirsi Alakurtti, Riikka Hämäläinen and Charlotta Diesen, Anne Rantala, Riika Salmela, Johanna Tommiska and Henna Tyynismaa. Thank you for your friendship, support and all the fun, without you the days in the lab would be much more boring. Everything does not always have to be that serious although we in our daily working life many times are banging our heads against the wall.

I also want to thank Esa Tahvanainen for his helpfulness in the computer analyses of the linkage results.

Monica Virtaneva is acknowledged for reviewing the language of my thesis.

Aila Riikonen and Minna Maunula are acknowledged for their help in all practical matters.

I am grateful to my family and parents-in-law, especially my mother who has always been very supportive and understanding of almost everything I have decided to do and for taking me out with the dogs in the evenings when my brain cells have needed some refreshment.

Finally I want to thank my dear husband Anders for his endless love and support which have been very valuable during these years and for his real interest towards my work.

I am grateful to all the Gräsbeck-Imerslund patients and their families for their active participation in this study. Without their contribution this study would never have been possible and we would still not know the cause of the Gräsbeck-Imerslund disease.

This study has been financially supported by the Foundation of Pediatric Research, the Ulla Hjelt Fund and Medicinska Understödsföreningen Liv och Hälsa.

Helsinki, May 2000

Maria Aminoff-Backlund

8. REFERENCES

Abdelaal MA. & Ahmed AF. (1991). Imerslund-Gräsbeck syndrome in a Saudi family. Acta Paediatr Scand 80:1109-1112.

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

Allen RH. (1975). Human vitamin B12 transport proteins. Prog Hematol 9:57-84.

Altay C., Cetin M., Gumruk F., Irken G., Yetgin S. & Laleli Y. (1995). Familial selective vitamin B12 malabsorption (Imerslund-Gräsbeck syndrome) in a pool of Turkish patients. Pediatr Hemat Oncol 12:19-28.

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

Alwine JC., Kemp DJ. & Stark GR. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. P Natl Acad Sci USA 74:5350-5354.

Antequera F. & Bird A. (1993). Number of CpG islands and genes in human and mouse. P Natl Acad Sci USA 90:11995-11999.

Anttila R. & Salmi HA. (1967). Selective malabsorption of vitamin B12 with proteinuria in children. Acta Pediatr Scand 52:238-240.

Ashworth LK., Alegria-Hartman M., Burgin M., Devlin L., Carrano AV. & Batzer MA. (1995). Assembly of high-resolution bacterial artificial chromosome, P1-derived artificial chromosome, and cosmid contigs. Anal Biochem 224:564-571.

Avela K., Lipsanen-Nyman M., Perheentupa J., Wallgren-Pettersson C., Marchand S., Fauré S., Sistonen P., de la Chapelle A. & Lehesjoki A-E. (1997). Assignment of the mulibrey nanism gene to 17q by linkage and linkage-disequilibrium analysis. Am J Hum Genet 60:896-902.

Ballabio A. (1993). The rise and fall of positional cloning? Nat Genet 3:277-279.

Bassam BJ., Caetano-Anolles G. & Gresshoff PM. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80-83.

Batuman V., Verroust PJ., Navar GL., Kaysen JH., Goda FO., Campbell WC., Simon E., Pontillon F., Lyles M., Bruno J. & Hammond TG. (1998). Myeloma light chains are ligands for cubilin (gp280). Am J Physiol 275:F246-254.

Berry R., Stevens TJ., Walter NAR., Wilcox AS., Rubano T., Hopkins JA., Weber J., Goold R., Soares MB. & Sikela JM. (1995). Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. Nat Genet 10:415-423.

Bird AP. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342-346.

Birn H., Verroust PJ., Nexø E., Hager H., Jacobsen C., Christensen EI. & Moestrup SK. (1997). Characterization of an epithelial approximately 460-kDa protein that facilitates endocytosis of intrinsic factor-vitamin B12 and binds receptor-associated protein. J Biol Chem 272:26497-26504.

Birn H., Vorum H., Verroust PJ., Moestrup SK. & Christensen EI. (2000). Receptor-associated protein is important for normal processing of megalin in kidney proximal tubules. J Am Soc Nephrol 11:191-202.

Boguski MS. & Schuler GD. (1995). ESTablishing a human transcript map. Nat Genet 10:369-371.

Bonaldo MF., Yu MT., Jelenc P., Brown S., Su L., Lawton L., Deaven L., Efstratiadis A., Warburton D. & Soares MB. (1994). Selection of cDNAs using chromosome-specific genomic clones: application to human chromosome 13. Hum Mol Genet 3:1663-1673.

von Bonsdorff B. (1977). Diphyllobothriatis in Man. Academic Press London, New York, San Francisco. pp 94-97.

Bork P. & Beckmann G. (1993). The CUB domain. A widespread module in developmentally regulated proteins. J Mol Biol 231:539-545.

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

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

Broch H., Imerslund O., Monn E., Hovig T. & Seip M. (1984). Imerslund-Gräsbeck anemia: A long-term follow-up study. Acta Paediatr Scand 73:248-253.

Burman JF., Mollin DL., Sourial NA. & Sladden RA. (1979). Inherited lack of transcobalamin II in serum and megaloblastic anemia: a further patient. Brit J Haematol 43:27-38.

Burman JF., Jenkins WJ., Walker-Smith JA., Phillips AD., Sourial NA., Williams CB. & Mollin DL. (1985). Absent ileal uptake of IF-bound vitamin B12 in vivo in the Imerslund-Gräsbeck syndrome (familial vitamin B12 malabsorption with proteinuria). Gut 26:311-314.

Cai W., Jing J., Irvin B., Ohler L., Rose E., Shizuya H., Kim UJ., Simon M., Anantharaman T., Mishra B. & Schwartz DC. (1998). High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. P Natl Acad Sci USA 95:3390-3395.

Campbell AN., Inglis J. & Paynter AS. (1981). Failure to thrive associated with the Imerslund-Gräsbeck syndrome. Postgrad Med J 57:509-510.

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

Chanarin I. (1979). The megaloblastic anaemias. 2nd ed Blackwell Scientific, Oxford.

Chanarin I. (1987). Megaloblastic anaemia, cobalamin, and folate. J Clin Pathol 40:978-984.

Chomczynski P. & Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156-159.

Christensen EI., Birn H., Verroust P. & Moestrup SK. (1998). Megalin-mediated endocytosis in renal proximal tubule. Renal Failure 20:191-199.

Chumakov IM., Rigault P., Le Gall I., Bellanne-Chantleot C., Billault A., Guillou S., Soularue P., Guasconi G., Poullier E., Gros I., Belova M., Sambucy I-L., Susini L., Gervy P, Glibert F., Beaufils S., Bui H., Massart C., De Tand M-F., Dukasz F., Lecoulant S., Ougen P., Perrot V., Saumier M., Soravito C., Bahouyila R., Cohen-Akenine A., Barillot E., Bertrand S., Codani J-J., Caterina D., Georges I., Lacroix B., Lucotte G., Sahbatou M., Schmit C., Sangouard M., Tubacher E., Dib C., Fauré S., Fizames C., Gyapay G., Millasseau P., Nguyen S., Muselet D., Vignal A., Morissette J., Menninger J., Lieman J., Desai T., Banks A., Bray-Ward P., Ward D., Hudson T., Gerety S., Foote S., Stein L., Page DC., Lander ES., Weissenbach J., Le Paslier D. & Cohen D. (1995). A YAC contig map of the human genome. Nature Suppl 377:175-297.

Claudio JO., Marineau C. & Rouleau GA. (1994). The mouse homologue of the neurofibromatosis type 2 gene is highly conserved. Hum Mol Genet 3:185-190.

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

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

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

Collins FS. (1999). Shattuck lecture-medical and societal consequences of the Human Genome Project. New Engl J Med 341:28-37.

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

Collins FS., Patrinos A., Jordan E., Chakravarti A., Gesteland R. & Walters L. (1998). New Goals for the U.S. Human Genome Project: 1998-2003. Science 282:682-689.

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

Cox DR. (1995). Mapping with radiation hybrids. Genome Digest 2:14-15.

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

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

Cross SH. & Bird AP. (1995). CpG islands and genes. Curr Opin Genet Dev 5:309-314.

Cuticchia AJ., Fasman KH., Kingsbury DT., Robbins RJ. & Pearson PL. (1993). The GDB human genome data base anno 1993. Nucleic Acids Res 21:3003-3006.

Dausset J., Ougen P., Abderrahim H., Billault A., Sambucy JL., Cohen D. & Le Paslier D. (1992). The CEPH YAC library. Behring Inst Mitt 91:13-20.

Dean M. (1995). Resolving DNA mutations. Nat Genet 9:103-104.

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

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

Deloukas P., Schuler GD., Gyapay G., Beasley EM., Soderlund C., Rodriguez-Tome P., Hui L., Matise TC., McKusick KB., Beckmann JS., Bentolila S., Bihoreau M.-T., Birren BB., Browne J., Butler A., Castle AB., Chiannilkulchai N., Clee C., Day PJR., Dehejia A., Dibling T., Drouot N., Duprat S., Fizames C., Fox S., Gelling S., Green L., Harrison P., Hocking R., Holloway E., Hunt S., Keil S., Lijnzaad P., Louis-Dit-Sully CJ., Ma Mendis A., Miller J., Morissette J., Muselet D., Nusbaum HC., Peck A., Rozen S., Simon D., Slonim DK., Staples R., Stein LD., Stewart EA., Suchard MA., Thangarajah T., Vega-Czarny N., Webber C., Wu X., Hudson J., Auffray C., Nomura N., Sikela JM., Polymeropoulos MH., James MR., Lander ES., Hudson TJ., Myers RM., Cox DR., Weissenbach J., Boguski MS. & Bentley DR. (1998). A physical map of 30,000 human genes. Science 282:744-746.

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

Donis-Keller H., Green P., Helms C., Cartinhour S., Weiffenbach B., Stephens K., Keith TP., Bowden DW., Smith DR., Lander ES., Botstein D., Akots G., Rediker KS., Gravius T., Brown VA., Rising MB., Parker C., Powers JA., Watt ER., Kauffman ER., Bricker A., Phipps P., Muller-Kahle H., Fulton TR., Ng S., Schumm JW., Braman JC., Knowlton RG., Barker DF., Crooks SM., Lincoln SE., Daly MJ. & Abrahamson J. (1987). A genetic linkage map of the human genome. Cell 51:319-337.

Doscherholmen A., Hagen PS. & Liu M. (1957). Radioactive vitamin B12 absorption studies: results of direct measurement of radioactivity in the blood. Blood 12:336-346.

Dugué B., Ismail E., Sequeira F., Thakkar J. & Gräsbeck R. (1999). Urinary excretion of intrinsic factor and the receptor for its cobalamin complex in Gräsbeck-Imerslund patients: the disease may have subsets. J Pediatr Gastr Nutr 29:227-230.

Dunham I., Hunt AR., Collins JE., Bruskiewich R., Beare DM., Clamp M., Smink LJ., Ainscough R., Almeida JP., Babbage A., Bagguley C., Bailey J., Barlow K., Bates KN., Beasley O., Bird CP., Blakey S., Bridgeman AM., Buck D., Burgess J., Burrill WD., Burton J., Carder C., Carter NP., Chen Y., Clark G., Clegg SM., Cobley V., Cole CG., Collier RE., Connor RE., Conroy D., Corby N., Coville GJ., Cox AV., Davis J., Dawson E., Dhami PD., Dockree C., Dodsworth SJ., Durbin RM., Ellington A., Evans KL., Fey JM., Fleming K., French L., Garner AA., Gilbert JGR., Goward ME., Grafham D., Griffiths MN., Hall C., Hall R., Hall-Tamlyn G., Heathcott RW., Ho S., Holmes S., Hunt SE., Jones MC., Kershaw J., Kimberley A., King A., Laird GK., Langford CF., Leversha MA., Lloyd C., Lloyd DM., Martyn ID., Mashreghi-Mohammadi M., Matthews L., McCann OT., McClay J., McLaren S., McMurray AA., Milne SA., Mortimore BJ., Odell CN., Pavitt R., Pearce AV., Pearson D., Phillimore BJ., Phillips SH., Plumb RW., Ramsay H., Ramsey Y., Rogers L., Ross MT., Scott CE., Sehra HK., Skuce CD., Smalley S., Smith ML., Soderlund C., Spragon L., Steward CA., Sulston JE., Swann RM., Vaudin M., Wall M., Wallis JM., Whiteley MN., Willey D., Williams L., Williams S., Williamson H., Wilmer TE., Wilming L., Wright CL., Hubbard T., Bentley DR., Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S., Roe BA., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao HI., Lewis J., Lewis S., Lin S-P., Loh P., Malaj E., Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J., Korf I., Bedell JA., Hillier L., Mardis E., Waterston R., Wilson R., Emanuel BS., Shaikh T., Kurahashi H., Saitta S., Budarf ML., Mcdermid HE., Johnson A., Wong ACC., Morrow BE., Edelmann L., Kim UJ., Shizuya H., Simon MI., Dumanski JP., Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder CE. & O'Brien KP. (1999). The DNA sequence of human chromosome 22. Nature 402:489-495.

Eaton DM., Livingston JH., Seetharam B. & Puntis JW. (1998). Overexpression of an unstable intrinsic factor-cobalamin receptor in Imerslund-Gräsbeck syndrome. Gastroenterology 115:173-176.

Faquharson J. & Adams JF (1976). The forms of vitamin B12 in foods. Brit J Nutr 36:127-136.

Fenton WA. & Rosenberg LE. (1978). Genetic and biochemical analysis of human cobalamin mutants in cell culture. Annu Rev Genet 12:223-248.

Fickett JW. & Hatzigeorgiou AG. (1997). Eukaryotic promoter recognition. Genome Res 7:861-878.

Fields C., Adams MD., White O. & Venter C. (1994). How many genes in the human genome? Nat Genet 7:345-346.

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

Forrest S., Cotton R., Landegren U. & Southern E. (1995). How to find all those mutations. Nat Genet 10:375-376.

Furuhjelm U. & Nevanlinna HR. (1973). Inheritance of selective malabsorption of vitamin B12. Scand J Haematol 11:27-34.

Fyfe JC., Jezyk PF., Giger U. & Patterson DF. (1989). Inherited selective malabsorption of vitamin B12 in Giant Schnauzers. J Am Anim Hosp Assoc 25:533-539.

Fyfe JC., Giger U., Hall CA., Jezyk PF., Klumpp SA., Levine JS. & Patterson DF. (1991a). Inherited selective intestinal cobalamin malabsorption and cobalamin deficiency in dogs. Pediatr Res 29:24-31.

Fyfe JC., Ramanujam KS., Ramaswamy K., Patterson DF. & Seetharam B. (1991b). Defective brush-border expression of intrinsic factor-cobalamin receptor in canine inherited intestinal cobalamin malabsorption. J Biol Chem 266:4489-4494.

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

Green ED., Riethman HC., Dutchik JE. & Olson MV. (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. Genomics 11:658-669.

Gräsbeck R. (1972). Familial selective vitamin B 12 malabsorption. New Engl J Med 287:358.

Gräsbeck R. (1984). Biochemistry and clinical chemistry of vitamin B12 transport and the related diseases. Clin Biochem 17:99-107.

Gräsbeck R. (1997). Selective cobalamin malabsorption and the cobalamin-intrinsic factor receptor. Acta Biochim Pol 44:725-733.

Gräsbeck R., Nyberg W., Siurala M. (1956). The urinary excretion of radio-vitamin B12 ("Schilling test"). Nord Med 56:1656-1660.

Gräsbeck R. & Nyberg W. (1957). Medicinska och fysiologiska aspekter på vitamin B12. Nord Med 57:321-327.

Gräsbeck R., Gordin R., Kantero I. & Kuhlbäck B. (1960). Selective vitamin B12 malabsorption and proteinuria in young people: a syndrome. Acta Med Scand 167:289-296.

Gräsbeck R., Nyberg W., Saarni M. & von Bonsdorff B. (1962). Lognormal distribution of serum vitamin B12 levels and dependence of blood values on the B level in a large population heavily infected with *Diphyllobothrium latum*. J Lab Clin Med 59:419-429.

Gräsbeck R. & Kvist G. (1967). Congenital specific vitamin B12 malabsorption syndrome with proteinuria. Münch Med Wochenschr 109:1936-1944.

Gräsbeck R. & Salonen E-M. (1976). Vitamin B12. Prog Fd Nutr Sci 2:193-231.

Gräsbeck R., Wahlstedt V. & Kouvonen I. (1982). Radioimmunoassay of urinary intrinsic factor. A promising test for pernicious anaemia and gastric function. Lancet 1:1330-1332.

Gräsbeck R. & Weber T. (1997). In: Gahrton G. & Lundh B. (eds.) Blodsjukdomar, Lärobok i hematologi. Natur och Kultur, Stockholm. pp 92-97.

Guéant J-L., Saunier M., Gastin I., Safi A., Lamireau T., Duclos B., Bigard MA. & Gräsbeck R. (1995). Decreased activity of intestinal and urinary intrinsic factor receptor in Gräsbeck-Imerslund disease. Gastroenterology 108:1622-1628.

Guéant J-L., Chery C., Namour F., Aimone-Gastin I. & Wustinger M. (1999). Decreased affinity of urinary intrinsic factor-cobalamin receptor in a case of Gräsbeck-Imerslund syndrome. Gastroenterology 116:1274-1276.

Gyapay G., Morissette J., Vignal A., Dib C., Fizames C., Millasseau P., Marc S., Bernardi G., Lathrop M. & Weissenbach J. (1994). The 1993-94 Généthon human genetic linkage map. Nat Genet 7:246-339.

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

Hakami N., Neiman PE., Canellos GP. & Lazerson J. (1971). Neonatal megaloblastic anemia due to inherited transcobalamin II deficiency in two siblings. New Engl J Med 285:1163-1170.

Hall CA. (1992). The neurologic aspects of transcobalamin II deficiency. Brit J Haematol 80:117-120.

Hamada H., Petrino MG. & Kakunaga T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. P Natl Acad Sci USA 79:6465-6469.

Hammad SM., Stefansson S., Twal WO., Drake CJ., Fleming P., Remaley A., Brewer HB. Jr. & Argraves WS. (1999). Cubilin, the endocytic receptor for intrinsic factor-vitamin B(12) complex, mediates high-density lipoprotein holoparticle endocytosis. P Natl Acad Sci USA 96:10158-10163.

Hansen M. & Nexø E. (1987) The interaction of human transcobalamin isopeptides in cerebrospinal fluid and plasma with cobalamin and the cellular acceptor. Biochim Biophys Acta 926:359-364.

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

Herz J., Willnow TE. & Farese RV. Jr. (1997). Cholesterol, hedgehog and embryogenesis. Nat Genet 15:123-124.

Hitzig WH., Dohmann U., Pluss HJ. & Vischer D. (1974). Hereditary transcobalamin II deficiency: clinical findings in a new family. J Pediatr 85:622-628.

Hoffman CS. & Winston F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids from transformation of Escheria coli. Gene 57:267-272.

Hudson TJ., Stein LD., Gerety SS., Ma J., Castle AB., Silva J., Slonim DK., Baptista R., Kruglyak L., Xu S-H., Hu X., Colbert AME., Rosenberg C., Reeve-Daly MP., Rozen S., Hui L., Wu X., Vestergaard C., Wilson KM., Bae JS., Maitra S., Ganiatsas S., Evans CA., DeAngelis MM., Ingalls KA., Nahf RW., Horton LT. Jr, Oskin Anderson M., Collymore AJ., Ye W., Kouyoumjian V., Zemsteva IS., Tam J., Devine R., Courtney DF., Turner Renaud M., Nguyen H., O'Connor TJ., Fizames C., Fauré S., Gyapay G., Dib C., Morissette J., Orlin JB., Birren BW., Goodman N., Weissenbach J., Hawkins TL., Foote S., Page DC. & Lander ES. (1995). An STS-based map of the human genome. Science 270:1945-1954.

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

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

Höglund P., Sistonen P., Norio R., Holmberg C., Dimberg A., Gustavson K-H., de la Chapelle A. & Kere J. (1995). Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. Am J Hum Genet 57:95-102.

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

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

Imerslund O. (1959). Idiopathic chronic megaloblastic anemia in children. Oslo University Press, Boston, Oslo, London.

Imerslund O. (1960). Idiopathic chronic megaloblastic anemia in children. Acta Pediatr Suppl 49:3-65.

Imerslund O. & Bjørnstad P. (1963). Familial vitamin B12 malabsorption. Acta Haematol 30:1-7.

Ismail EA., Al Saleh Q., Sabry MA., Al Ghanim M. & Zaki M. (1997). Genotypic/phenotypic heterogeneity of selective vitamin B12 malabsorption (Gräsbeck-Imerslund syndrome) in two Bedouin families. Acta Paediatr 86:424-425.

Jones HB. (1997). Estimating physical distances from radiation hybrid mapping data. Genomics 43:258-266.

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

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

Katz M., Lee SK. & Cooper BA. (1972). Vitamin B(12) malabsorption due to biologically inert intrinsic factor. New Engl J Med 287:425-429.

Kere J., Norio R., Savilahti E., Estivill X. & de la Chapelle A. (1989). Cystic fibrosis in Finland: a molecular and genealogical study. Hum Genet 83:20-25.

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

Kim U-J., Birren BW., Slepak T., Mancino V., Boysen C., Kang H-L., Simon MI. & Shizuya H. (1996). Construction and characterization of a human bacterial artificial chromosome library. Genomics 34:213-218.

Knudsen S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics 15:356-361.

Knutson A., P. Hellman P., G. Akerstrom & G. Westin (1998). Characterization of the human megalin/LRP-2 promoter in vitro and in primary parathyroid cells. DNA Cell Biol 17:551-560.

Korenberg JR., Chen X-N., Sun Z., Shi Z-Y., Ma S., Vataru E., Yimlamai D., Weissenbach JS., Shizuya H., Simon MI., Gerety SS., Nguyen H., Zemsteva IS., Hui L., Silva J., Wu X., Birren BW. & Hudson TJ. (1999). Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and medicine. Genome Res 9:994-1001.

Kozyraki R., Kristiansen M., Silahtaroglu A., Hansen C., Jacobsen C., Tommerup N., Verroust PJ. & Moestrup SK. (1998). The human intrinsic factor-vitamin B12 receptor, cubilin: molecular characterization and chromosomal mapping of the gene to 10p within the autosomal recessive megaloblastic anemia (MGA1) region. Blood 91:3593-3600.

Kozyraki R., Fyfe J., Kristiansen M., Gerdes C., Jacobsen C., Cui S., Christensen EI., Aminoff M., de la Chapelle A., Krahe R., Verroust PJ. & Moestrup SK. (1999). The intrinsic factor-vitamin B12 receptor, cubilin, is a high-affinity apolipoprotein A-I receptor facilitating endocytosis of high-density lipoprotein. Nat Med 5:656-661.

Kristiansen M., Kozyraki R., Jacobsen C., Nexø E., Verroust PJ. & Moestrup SK. (1999). Molecular dissection of the intrinsic factor-vitamin B12 receptor, cubilin, discloses regions important for membrane association and ligand binding. J Biol Chem 274:20540-20544.

Kristiansen M., Aminoff M., Jacobsen C., de la Chapelle A., Krahe R., Verroust PJ. & Moestrup SK. The cubilin P1297L mutation associated with hereditary megaloblastic anemia 1 (MGA1) causes impaired recognition of intrinsic factor-vitamin B12 by cubilin. Blood (in press).

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

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

Lander ES. (1996). The new genomics: global views of biology. Science 274:536-539.

Larin Z., Monaco AP. & Lehrach H. (1997). Generation of large insert yeast artificial chromosome libraries. Mol Biotechnol 2:147-153.

Larsen F., Gundersen G., Lopez R. & Prydz H. (1992). CpG islands as gene markers in the human genome. Genomics 13:1095-1107.

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

Lindblom A., Quadt N., Marsh T., Aeschlimann D., Mörgelin M., Mann K., Maurer P. & Paulsson M. (1999). The intrinsic factor-vitamin B12 receptor, cubilin, is assembled into trimers via a coiled-coil α-helix. J Biol Chem 274:6374-6380.

Lindenbaum RH., Clarke G., Patel C., Moncrieff M. & Hughes JT. (1979). Muscular dystrophy in an X;1 translocation female suggests that Duchenne locus is on X chromosome short arm. J Med Genet 16:389-392.

Linnell JC. & Bhatt HR. (1995). Inherited errors of cobalamin metabolism and their management. Bailliere Clin Haem 8:567-601.

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

MacKenzie IL., Donaldson RM. Jr, Trier JS. & Mathan VI. (1972). Ileal mucosa in familial selective vitamin B12 malabsorption. New Engl J Med 286:1021-1025.

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

McCarthy LC., Terrett J., Davis ME., Knights CJ., Smith AL., Critcher R., Schmitt K., Hudson J., Spurr NK. & Goodfellow PN. (1997). A first-generation whole genome-radiation hybrid map spanning the mouse genome. Genome Res 7:1153-1161.

McNichol B. & Egan B. (1968). Congenital pernicious anemia: effects on growth, brain, and absorption of B12. Pediatr 42:149-156.

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

Moestrup SK., Birn H., Fischer PB., Petersen CM., Verroust PJ., Sim RB., Christensen EI. & Nexø E. (1996). Megalin-mediated endocytosis of transcobalamin-vitamin-B12 complexes suggests a role of the receptor in vitamin-B12 homeostasis. P Natl Acad Sci USA 93:8612-8617.

Moestrup SK., Kozyraki R., Kristiansen M., Kaysen JH., Holm Rasmussen H., Brault D., Pontillon F., Goda FO., Christensen EI., Hammond TG. & Verroust PJ. (1998). The intrinsic factor-vitamin B12 receptor and target of teratogenic antibodies is a megalin-binding peripheral membrane protein with homology to developmental proteins. J Biol Chem 273:5235-5242.

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

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

Nakamura Y., Leppert M., O'Connell P., Wolff R., Holm T., Culver M., Martin C., Fujimoto E., Hoff M., Kumlin E. & White R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616-1622.

Neale G. (1990). B12 binding proteins. Gut 31:59-63.

Nevanlinna HR. (1972) The Finnish population structure. A genetic and genealogical study. Hereditas 71:195-236.

Nexø E., Hansen M., Rasmussen K., Lindgren A. & Gräsbeck R. (1994). How to diagnose cobalamin deficiency. Scand J Clin Lab Inv Suppl 219:61-76.

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

Norio R. (1981). Diseases of Finland and Scandinavia. In: Rothschild H (ed.). Biocultural aspects of disease. Academic Press, New York. pp 359-415.

Norio R., Nevanlinna HR. & Perheentupa J. (1973). Hereditary diseases in Finland; rare flora in rare soul. Ann Clin Res 5:109-141.

Nyberg W., Gräsbeck R. & Sippola V. (1958). Urinary excretion of radiovitamin B12 in carriers of Diphyllobothrium Latum. New Engl J Med 259:216-219.

Official Statistics of Finland. (1997). Statistical yearbook of Finland 1997. Karisto, Hämeenlinna, Finland.

Official Statistics of Finland. (1999). Statistical yearbook of Finland 1999. Karisto, Hämeenlinna, Finland.

van Ommen GJ., Bakker E. & den Dunnen JT. (1999). The human genome project and the future of diagnostics, treatment, and prevention. Lancet 354:Suppl 1:SI5-10.

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

Ott J. (1991). Analysis of Human Linkage. Johns Hopkins University Press, Baltimore.

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

Paloneva J., Kestilä M., Wu J., Salminen A., Bühling T., Ruotsalainen V., Hakola P., Bakker A., Phillips J., Pekkarinen P., Lanier L., Timonen T. & Peltonen L. Loss of function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. Nat Genet (in press).

Pearson PL. (1991). The genome data base (GDB)-a human gene mapping repository. Nucleic Acids Res 19:Suppl, 2237-2239.

Pearson PL., Matheson NW., Flescher DC. & Robbins RJ. (1992). The GDB human genome data base anno 1992. Nucleic Acids Res 20:Suppl, 2201-2206.

Pedersen AG., Baldi P., Chauvin Y. & Brunak S. (1999). The biology of eukaryotic promoter prediction–a review. Comput Chem 23:191-207.

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

Peltonen L. (1997). Molecular background of the Finnish disease heritage. Ann Med 29:553-556.

Peltonen L. (2000). Positional cloning of disease genes: advantages of genetic isolates. Hum Hered 50:66-75.

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

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

Phelps RS., Chadwick RB., Conrad MP., Kronick MN. & Kamb A. (1995). Efficient, automatic detection of heterozygous bases during large-scale DNA sequence screening. Biotechniques 19:984-989.

Prosser J. (1993). Detecting single-base mutations. Trends Biotechnol 11:238-246.

Quandt K., Frech K., Karas H., Wingender, E. & Werner T. (1995). MatInd and MatInspector–New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res 23:4878-4884.

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

Rickes EL., Brink NG., Koniuszy FR., Wood TR. & Folkers K. (1948). Crystalline vitamin B12. Science 107:396-397.

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

Romero A., Romao MJ., Varela PF., Kolln I., Dias JM., Carvalho AL., Sanz L., Topfer-Petersen E. & Calvete JJ. (1997). The crystal structures of two spermadhesins reveal the CUB domain fold. Nat Struct Biol 4:783-788.

Safi A., Saunier M., Gastin I., Alibaba Y, Dugue B. & Guéant J-L. (1995). Intrinsic factor covalently bound to Sepharose as affinity medium for the purification of a soluble intrinsic factor receptor from human urine. J Chromatogr B 664:253-259.

Sahali D., Mulliez N., Chatelet F., Laurent-Winter C., Citadelle D., Roux C., Ronco P. & Verroust P. (1992). Co-expression in humans by kidney and fetal envelopes of a 280 kDa-coated pit-restricted protein. Similarity with the murine target of teratogenic antibodies. Am J Pathol 140:33-44.

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

Sandberg DP., Begley JA. & Hall CA. (1981). The content, binding, and forms of vitamin B12 in milk. Am J Clin Nutr 34:1717-1724.

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

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

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

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

Schilling RF. (1953). Intrinsic factor studies. II. The effect of gastric juice on the urinary excretion of radioactivity after the oral administration of radioactive vitamin B12. J Lab Clin Med 42:860-866.

Schlessinger D. (1990). Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. Trends Genet 6:255-258.

Schuler GD., Boguski MS., Stewart EA., Stein LD., Gyapay G., Rice K., White RE., Rodriguez-Tome P., Aggarwal A., Bajorek E., Bentolila S., Birren BB., Butler A., Castle AB., Chiannilkulchai N., Chu A., Clee C., Cowles S., Day PJR., Dibling T., Drouot N., Dunham I., Duprat S., East C., Edwards C., Fan JB., Fang N., Fizames C., Garrett C., Green L., Hadley D., Harris M., Harrison P., Brady S., Hicks A., Holloway E., Hui L., Hussain S., Louis-Dit-Sully C., J. Ma, MacGilvery A., Mader C., Maratukulam A., Matise TC., McKusick KB., Morissette J., Mungall A., Muselet D., Nusbaum HC., Page DC., Peck A., Perkins S., Piercy M., Qin F., Quackenbush J., Ranby S., Reif T., Rozen S., Sanders C., She X., Silva J., Slonim DK., Soderlund C., Sun WL., Tabar P., Thangarajah T., Vega-Czarny N., Vollrath D., Voyticky S., Wilmer T., Wu X., Adams MD., Auffray C., Walter NAR., Brandon R., Dehejia A., Goodfellow PN., Houlgatte R., Hudson JR. Jr., Ide SE., Iorio KR., Lee WY., Seki N.,

Nagase T., Ishikawa K., Nomura N., Phillips C., Polymeropoulos MH., Sandusky M., Schmitt K., Berry R., Swanson K., Torres R., Venter JC., Sikela JM., Beckmann JS., Weissenbach J., Myers RM., Cox DR., James MR., Bentley D., Deloukas P., Lander ES. & Hudson TJ. (1996). A gene map of the human genome. Science 274:540-546.

Seetharam B., Alpers DH. & Allen RH. (1981). Isolation and characterization of the ileal receptor for intrinsic factor-cobalamin. J Biol Chem 256:3785-3790.

Seetharam B., Christensen EI., Moestrup SK., Hammond TG. & Verroust PJ. (1997). Identification of rat yolk sac target protein of teratogenic antibodies, gp280, as intrinsic factor-cobalamin receptor. J Clin Invest 99:2317-2322.

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

Sheffield VC., Weber JL., Buetow KH., Murray JC., Even DA., Wiles K., Gastier JM., Pulido JC., Yandava C., Sunden SL, Mattes G., Businga T., McClain A., Beck J., Scherpier T., Gilliam J., Zhong J. & Duyk GM. (1995). A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837-1844.

Smith E. & Parker L. (1948). Purification of anti-pernicious anaemia factor. Biochem J 43:8.

Stewart EA., McKusick KB., Aggarwal A., Bajorek E., Brady S., Chu A., Fang F., Hadley D., Harris M., Hussain S., Lee R., Maratukulam A., O'Connor K., Perkins S., Piercy M., Qin F., Reif T., Sanders C., She X., Sun WL., Tabar P., Voyticky S., Cowles S., Fan JB., Mader C., Quackenbush J., Myers RM. & Cox DR. (1997). An STS-based radiation hybrid map of the human genome. Genome Res 7:422-433.

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

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

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

Terwilliger JD. & Ott J. (1994). Handbook of Human Genetic Linkage. London: The Johns Hopkins University Press. pp 105-134.

The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971-983.

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

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

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

Visakorpi JK. & Furuhjelm U. (1967). Selective malabsorption of vitamin B12. Mod Probl Pediat 11:150-160.

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

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

Wadman M. (1999). Human Genome Project aims to finish “working draft” next year. Nature 398:177.

Wahlstedt V. & Gräsbeck R. (1985). Cobalamin-binding proteins in human urine: identification and quantitation. J Lab Clin Med 106:439-446.

Walter MA., Spillett DJ., Thomas P., Weissenbach J. & Goodfellow PN. (1994). A method for constructing radiation hybrid maps of whole genomes. Nat Genet 7:22-28.

Wang DG., Fan JB., Siao CJ.; Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris MS., Shen N., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson TJ., Lipshutz R., Chee SM. & Lander ES. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077-1082.

Watanabe TK., Bihoreau M-L., McCarthy LC., Kiguwa SL., Hishigaki H., Tsuji A., Browne J., Yamasaki Y., Mizoguchi-Miyakita A., Oga K., Ono T., Okuno S., Kanemoto N., Takahashi E., Tomita K., Hayashi H., Adachi M., Webber C., Davis m., Kiel S., Knights C., Smith A., Critcher R., Miller J., Thangarajah T., Day PJR., Hudson JR. Jr, Irie Y., Takagi T., Nakamura Y., Goodfellow PN., Lathrop GM. & Tanigami A. (1999). A radiation hybrid map of the rat genome containing 5,255 markers. Nat Genet 22:27-36.

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

Weeks DE., Ott J. & Lathrop GM. (1990). SLINK: a general simulation program for linkage analysis. Am J Hum Genet Suppl 47:A204.

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

Werner T. (1999). Models for prediction and recognition of eukaryotic promoters. Mamm Genome 10:168-175.

White R., Lalouel J-M., Leppert M., Lathrop M., Nakamura Y. & O'Connell P. (1989). Linkage maps of human chromosomes. Genome 31:1066-1072.

Willnow TE., Hilpert J., Armstrong SA., Rohlmann A., Hammer RE., Burns DK. & Herz J. (1996). Defective forebrain development in mice lacking gp330/megalin. P Natl Acad Sci USA 93:8460-8464.

Wulffraat NM., De Schryver J., Bruin M., Pinxteren-Nagler E. & van Dijken PJ. (1994). Failure to thrive is an early symptom of the Imerslund Gräsbeck syndrome. Am J Pediatr Hematol Oncol 16:177-180.

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

Xu D., Kozyraki R., Newman TC. & Fyfe JC. (1999). Genetic evidence of an accessory activity required specifically for cubilin brush-border expression and intrinsic factor-cobalamin absorption. Blood 94:3604-3606.

Yetgin S., Özsoylu S. & Zamani VP. (1983). Imerslund-Gräsbeck syndrome and generalized malabsorption. Turk J Pediatr 25:193-196.

Zuo J., Robbins C., Taillon-Miller P., Cox DR. & Myers RM. (1992). Cloning of the Huntington disease region in yeast artificial chromosomes. Hum Mol Genet 1:149-159.

Internet addresses:

BCM Search Launcher: Gene Feature Searches:
http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html

BioInformatics & Molecular Analysis Section (BIMAS):
http://bimas.dcrt.nih.gov/molbio/signal/

Genome Database (GDB): http://www.gdb.org/

Human Genome Project (HGP): http://www.ornl.gov/hgmis/hg5yp/

MatInspector professional:
http://www.genomatix.gsf.de/cgi-bin/matinspector_prof/mat_fam.pl

MatInspector V2.2:
http://www.cbil.upenn.edu/cgi-bin/tess/Tess?_if=1&RQ=WECOME

National Center for Biotechnology Information (NCBI) BLAST:
http://www.ncbi.nlm.nih.gov/BLAST/

Roswell Park Cancer Institute (RPCI) human BAC library: http://bacpac.med.buffalo.edu

UK Human Genome Mapping Program Resource Center: http://hgmp.mrc.ac.uk/

Whitehead Institute/MIT Center for Genome Research:
http://www-genome.wi.mit.edu/