View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Helsingin yliopiston digitaalinen arkisto

Mikko Kuokkanen

Molecular Genetics of Lactase Deficiencies

Publications of the National Public Health Institute A 20/2006

Department of Molecular Medicine
National Public Health Institute Helsinki, Finland
and
Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2006

Mikko Kuokkanen

MOLECULAR GENETICS OF LACTASE DEFICIENCIES

ACADEMIC DISSERTATION

To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in Lecture Hall 3 of Biomedicum Helsinki, Haartmaninkatu 8,on December 5th, at 12 noon.

Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2006

Publications of the National Public Health Institute
KTL A20 / 2006

Copyright National Public Health Institute

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-666-5
ISSN 0359-3584
ISBN 951-740-667-3 (pdf)
ISSN 1458-6290 (pdf)

Kannen kuva - cover graphic: Kirsi Kuokkanen, Nursing baby, aquarelle 2006

Edita Prima Oy
Helsinki 2006

Supervised by

Docent Irma Järvelä
Helsinki University Hospital
Laboratory of Molecular Genetics and
University of Helsinki
Department of Medical Genetics
Helsinki, Finland

Reviewed by

Professor Mikko Hallman
University of Oulu
Department of Pediatrics and
Biocenter Oulu
Oulu, Finland

Docent Johanna Schleutker
University of Tampere
Institute of Medical Technology
Laboratory of Cancer Genetics and
Tampere University Hospital
Tampere, Finland

Opponent

Professor Helena Kääriäinen
University of Turku
Department of Medical Genetics
Turku, Finland

The only solution.

Isn't it amazing?

Jim Morrison, Shaman’s Blues (1969)

To my family

Mikko Kuokkanen, Molecular Genetics of Lactase Deficiencies
Publications of the National Public Health Institute, A20/2006, 117 Pages
ISBN 951-740-666-5; 951-740-667-3 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/suomi/julkaisut/julkaisusarjat/kansanterveyslaitoksen_julkaisuja_a/

ABSTRACT

Congenital lactase deficiency (CLD) (MIM 223000) is a rare autosomal recessive gastrointestinal disorder characterized by watery diarrhea in infants fed with breast milk or other lactose-containing formulas. The CLD locus was previously assigned by linkage and linkage disequilibrium analyses on 2q21 in 19 Finnish families. In this study, the molecular background of this disorder is reported. The CLD locus was refined in 32 CLD patients in 24 families by using microsatellite and single nucleotide polymorphism (SNP) haplotypes. Mutation analyses were performed by direct sequencing. We identified 5 distinct mutations in the lactase (*LCT*) gene, encoding the enzyme that hydrolyzes lactose in the intestinal lumen. Twenty-seven patients out of thirty-two (84%) were homozygous for a nonsense mutation, c.4170T→A (Y1390X, designated as Fin_{major}). In addition, four rare mutations were detected two of which, a four-nucleotide deletion (c.4998_5001delTGAG) and a two-nucleotide deletion (c.653_654delCT), predicted a frameshift and protein truncation at S1666fsX1722 and S218fsX224 of the 1927 amino acid polypeptide, respectively. Two point mutations, c.804G→C and c.4087G→A, would result in amino acid substitutions Q268H and G1363S, respectively. Five patients were compound heterozygous carrying the Fin_{major} mutation and one of the four minor mutations. These findings facilitate genetic testing of CLD in clinical practice and enable genetic counseling. The present data also provide the basis for detailed characterization of the molecular pathogenesis of this disorder.

Adult-type hypolactasia (MIM 223100) (lactase non-persistence, lactose intolerance) is an autosomal recessive gastrointestinal condition that is a result of a decline in the activity of lactase in the intestinal lumen after weaning. Adult-type hypolactasia is considered to be a normal phenomenon among mammals and symptoms are remarkably milder than experienced in CLD. Recently, a variant C/T_{-13910} was shown to associate with the adult-type hypolactasia trait, locating 13.9 kb upstream of the *LCT* gene (Enattah et al. 2002). In this study, the functional significance of the C/T_{-13910} variant was determined by studying the *LCT* mRNA levels in intestinal biopsy samples in children and adults with different genotypes. Intestinal biopsy samples were taken from 15 children or adolescents and from 52 adults with abdominal complaints. The expression of *LCT* mRNA was demonstrated in patients

heterozygous for the C/T$_{-13910}$ variant and an informative expressed SNP located in the coding region of *LCT*. RT-PCR followed by solid-phase minisequencing was applied to determine the relative expression levels of the *LCT* alleles using an informative SNP located in exon 1. In children, the C$_{-13910}$ allele was observed to be downregulated after five years of age in parallel with lactase enzyme activity. The expression of the *LCT* mRNA in the intestinal mucosa in individuals with the T$_{-13910}$ A$_{-22018}$ alleles was 11.5 times higher than that found in individuals with the C$_{-13910}$, G$_{-22018}$ alleles. These findings suggest that the C/T$_{-13910}$ associated with adult-type hypolactasia is associated with the transcriptional regulation of the *LCT* gene. The presence of the T$_{-13910}$ A$_{-22018}$ allele also showed significant elevation lactase activity.

Galactose, the hydrolysing product of the milk sugar lactose, has been hypothesized to be poisonous to ovarian epithelial cells. Hence, consumption of dairy products and lactase persistence has been proposed to be a risk factor for ovarian carcinoma. To investigate whether lactase persistence is related to the risk of ovarian carcinoma the C/T$_{-13910}$ genotype was determined in a cohort of 782 women with ovarian carcinoma. The C/T$_{-13910}$ genotype was determined by the solid-phase minisequencing method from 327 Finnish, 303 Polish, 152 Swedish patients and 938 Finnish, 296 Polish and 97 Swedish healthy subjects serving as controls. Lactase persistence did not associate significantly with the risk for ovarian carcinoma in the Finnish (OR 0.77, 95% CI, 0.57-1.05, p=0.097), in the Polish (OR 0.95, 95% CI, 0.68-1.33, p=0.75), or in the Swedish populations (OR 1.63, 95% CI, 0.65-4.08, p=0.29). The findings do not support the hypothesis that lactase persistence increases the risk for ovarian carcinoma.

Keywords: *LCT*, CLD, adult-type hypolactasia, lactase persistence/non-persistence, ovarian carcinoma, C/T$_{-13910}$, single nucleotide polymorphism, solid-phase minisequencing

Mikko Kuokkanen, Molecular Genetics of Lactase Deficiencies
Kansanterveyslaitoksen julkaisuja, A20/2006, 117 sivua
ISBN 951-740-666-5; 951-740-667-3 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/suomi/julkaisut/julkaisusarjat/kansanterveyslaitoksen_julkaisuja_a/

TIIVISTELMÄ

Synnynnäinen laktaasin puutos (congenital lactase deficiency, CLD [MIM 223000]) on vakava peittyvästi periytyvä suolistosairaus. Potilailla on tyypillistä runsas vetinen ripuli, joka ilmenee maitoruokinnan alettua. Ripuli johtaa ravintoaineiden imeytymishäiriöön ja kasvun hidastumiseen. Potilaiden laktaasiaktiivisuus on havaittu ohutsuolessa hyvin alhaiseksi. Vaikea ripuli johtuu nimenomaan hydrolysoimattoman laktoosin kertymisestä suolistoon aiheuttaen osmoosin, vaikean kuivumistilan, asidoosin ja painon menetyksen syntymäpainon alapuolelle. Laktoositon ruokavalio poistaa oireet ja palauttaa kehityksen normaaliksi. CLD-vauvoja syntyy Suomessa yksi vuodessa (1:60000). Se kuuluu yhdessä 36 muun harvinaisen Suomessa esiintyvän sairauden kanssa ns. suomalaiseen tautiperintöön.

Tutkimuksen tarkoituksena oli tunnistaa CLD:lle altistavat geneettiset tekijät geenien kartoitus- ja sekvensointimenetelmien avulla. Tutkimuksessa analysoitiin 32 CLD-potilasta 24 eri perheestä. Löysimme viisi CLD:lle altistavaa mutaatiota laktaasigeenistä (*LCT*), joista Y1390X-mutaation havaittiin olevan yleisin ns. $\text{Fin}_{\text{major}}$-mutaatio. Löydökset mahdollistavat sairauden perinnöllisen testauksen ja neuvonnan.

Lapsuuden jälkeen tai nuorella aikuisiällä havaittu primääri maitosokerin imeytymishäiriö (MIM 223100) (tunnettu myös laktoosi-intoleranssina) on yleinen laktaasin aktiivisuuden alenemisesta johtuva ilmiö. Laktaasiaktiivisuus alenee suolistossa 5-10% lapsuudesta havaitusta ja laktoosi ruokavaliossa aiheuttaa suolistoperäisiä oireita. Laktaasin puutos on nisäkkäille luonnollinen ilmiö, näin lapsi vierotetaan rinnasta käyttämään kiinteätä ravintoa. Kuitenkin joillain ihmisillä korkea laktaasiaktivisuus säilyy koko elinajan. Kyseiseen ilmiasuun on havaittu

liittyvän yhden emäksen muutos $C \rightarrow T_{-13910}$ noin 14 kiloemästä laktaasigeenistä ylävirtaan. Muutos sijaitsee *MCM6*- (minichromosome maintenance deficient 6) geenin intronissa 13 kromosomissa 2q21. T_{-13910}-muutoksen perineet henkilöt säilyttävät laktaasiaktiivisuuden ja kyvyn pilkkoa laktoosia.

Tämän tutkimuksen tarkoituksena oli määrittää C/T_{-13910} emäsmuutoksen vaikutus laktaasigeenin ilmentymiseen lähetti-RNA tasolla lapsilla ja aikuisilla. Laktaasigeenin ilmentymistä tutkittiin geenispesifisellä minisekvensointimenetelmällä pohjukaissuolesta otetuista näytteistä. Havaitsimme, että laktaasia säädellään transkriptiotasolla, C_{-13910}-muutos sallii laktaasigeenin lähetti-RNA:n laskun. Ilmiö oli havaittavissa lapsilla viiden ikävuoden jälkeen. Aikuisilla T_{-13910}-emäsmuutos vastasi 92% havaitusta laktaasigeenin lähetti-RNA:n ilmentymisestä, joka näkyi myös korkeana laktaasiaktiivisuutena.

Galaktoosi on glukoosin ohella laktoosin hajoamistuote. Galaktoosin on havaittu eläinkokeissa olevan myrkyllinen munasarjojen epiteelisoluille ja sen on epäilty aiheuttavan munasarjasyöpää. Tästä syystä korkean laktaasiaktiivisuuden ja maitotuotteiden runsaan nauttimisen on oletettu nostavan munasarjasyövän riskiä. Tässä tutkimuksessa määritettiin 782 suomalaisen, puolalaisen ja ruotsalaisen munasarjasyöpäpotilaan ja 1331 verrokkinäytteen C/T_{-13910}-genotyyppi tutkiaksemme onko korkealla laktaasiaktiivisuudella yhteyttä syövän syntyyn. Genotyypillä ei havaittu olevan vaikutusta syöpäriskiin tutkituissa väestöryhmissä. Jatkotutkimuksella pyritään selvittämään maitotuotteiden kulutuksen merkitystä munasarjasyöpään eri C/T_{-13910} genotyypeillä.

Avainsanat: synnynnäinen laktaasin puutos, geenikartoitus, suomalainen tautiperintö, laktoosin imeytymishäiriö, C/T_{-13910}, minisekvensointi, munasarjasyöpä, galaktoosi

CONTENTS

Abbreviations **12**

List of original publications **14**

1 Introduction **15**

2 Review of the literature **17**

2.1 THE FINNISH DISEASE HERITAGE 17

2.1.1 Causes of the unique gene pool 20

2.1.1.1 The settlement history of Finland 20

2.1.1.2 The internal migration in the 16th century 22

2.2 LACTASE 24

2.2.1 Structure 25

2.2.2 Biosynthesis 26

2.2.3 Regulation 28

2.2.3.1 Proximal LCT promoter 28

2.2.3.2 Distal regulatory elements 30

2.3 CONGENITAL LACTASE DEFICIENCY 32

2.3.1 Clinical features 33

2.3.2 Assignment of the locus to 2q21 34

2.4 ADULT-TYPE HYPOLACTASIA 35

2.4.1 Diagnosis 36

2.4.2 Onset and prevalence 37

2.4.3 Genetics 38

2.4.3.1 Evolutionary favour of lactase persistence 38

2.4.3.2 Mechanism of downregulation 39

2.4.3.3 Discovery of the responsible variant 40

2.4.3.3.1 Functional evidence 42

2.4.3.4 T_{-13910}, the only variant ruling lactase persistence for humanity? 45

2.5 LACTASE PERSISTENCE, NON-PERSISTENCE AND DAIRY CONSUMPTION RISK FACTORS FOR OTHER DISEASES? 46

2.5.1 Ovarian carcinoma 49

2.5.1.1 Dairy consumption and ovarian carcinoma risk 50

2.6 IDENTIFICATION OF DISEASE GENES 53
2.6.1 The Human Genome Project 53
2.6.2 The Human Genome.. 53
2.6.3 Positional cloning 54
2.6.4 Genetic mapping... 54
2.6.4.1 Genetic markers 55
2.6.4.2 Linkage analysis 56
2.6.4.3 Linkage disequilibrium 56
2.6.4.4 Haplotype analysis..59
2.6.4.5 Towards causative mutations 59
2.7 GENE EXPRESSION STUDIES 59
2.7.1 Techniques for mRNA detection 60
3 Aims of the study 63
4 Materials and methods 64
4.1 STUDY SUBJECTS 64
4.1.1 Finnish CLD families and controls (I).. 64
4.1.2 *LCT* mRNA study in children (II) 64
4.1.3 *LCT* mRNA study in adults (III)... 64
4.1.4 Lactase persistence and ovarian carcinoma (IV) 65
4.1.5 Ethical considerations... 65
4.1.6 Methods not described in the publications 66
4.1.6.1 Sequencing..66
4.1.6.2 Microsatellite genotyping 67
4.1.6.3 Statistical analyses 67
5 Results and discussion 68
5.1 RESULTS OF CONGENITAL LACTASE DEFICIENCY (I)................................68
5.1.1 Linkage, linkage disequilibrium and haplotype analyses 68
5.1.2 CLD mutations in *LCT* 70
5.1.2.1 The impact of Fin_{major} on *LCT* transcription.............................72
5.1.3 Genealogy and carrier frequencies of the CLD mutations in Finland... 72
5.2 DISCUSSION OF CONGENITAL LACTASE DEFICIENCY 73
5.3 RESULTS OF *LCT* MRNA STUDIES 77

5.3.1 *LCT* expression and quantitation in children (II) 77
5.3.2 *LCT* expression in adults (III) 78
5.3.2.1 Quantitation of *LCT* mRNA levels 79
5.4 DISCUSSION OF *LCT* MRNA STUDIES (II, III) 80
5.5 RESULTS OF LACTASE PERSISTENCE AND OVARIAN CARCINOMA (IV)...... 85
5.6 DISCUSSION OF LACTASE PERSISTENCE AND OVARIAN CARCINOMA 86
6 Concluding remarks and future prospects 88
7 Acknowledgements 90
8 Electronic database information 93
9 References.. 94

ABBREVIATIONS

χ^2	chi-square test
λ	proportion of excess of alleles in chromosomes carrying the disease allele
θ	recombination fraction
Caco-2 cell	human intestinal adenocarcinoma cell
cDNA	complementary DNA
Cdx	caudal-related protein
CEPH	Centre d'Etude du Polymorphisme Humain
CLD	congenital lactase deficiency
cM	centiMorgan
cSNP	coding single nucleotide polymorphism
DARS	aspartyl-tRNA synthetase
DNA	deoxyribonucleic acid
dNTP	deoxynucleosidetriphosphate
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridization
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
HNF	hepatocyte nuclear factor
HRR-LRT	haplotype relative risk likelihood-ratio
in situ	on the spot
in vitro	in test tube
L	likelihood
LCT	lactase

LD	linkage disequilibrium
LOD, Z	logarithm of odds
L/S	ratio of lactase to sucrase
LTTE	lactose tolerance test with ethanol
MCM6	minichromosome maintenance deficient 6
MIM	Mendelian Inheritance in Man
mRNA	messenger RNA
NMD	Non-sense mediated mRNA decay
Oct	octamer-binding transcription factor
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
PTH	parathyroid hormone
q	long arm of chromosome
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RT	reverse transcriptase
SCCA	squamous cell carcinoma antigen
SNP	single nucleotide polymorphism
STR	short tandem repeat
tRNA	transfer RNA

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by their Roman numerals:

I Kuokkanen M, Kokkonen J, Enattah NS, Ylisaukko-oja T, Komu H, Varilo T, Peltonen L, Savilahti E, & Järvelä I. (2006) Mutations in the Translated Region of the Lactase Gene (*LCT*) Underlie Congenital Lactase Deficiency. Am J Hum Genet. 78:339-44

II Rasinperä H, Kuokkanen M, Enattah NS, Kolho KL, Savilahti E, Orpana A & Järvelä I. (2005): Transcriptional downregulation of the lactase (*LCT*) gene during childhood. Gut. 54:1660-1

III Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana A & Järvelä I (2003): Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. Gut 52:647-652.

IV Kuokkanen M, Butzow R, Rasinperä H, Medrek K, Nilbert M, Malander S, Lubinski J & Järvelä I. (2005): Lactase persistence and ovarian carcinoma risk in Finland, Poland and Sweden. Int J Cancer. 117:90-94

These articles are reproduced with the kind permission of their copyright holders.

Publications II and III has previously appeared in theses by Heli Rasinperä (2006) and Nabil Sabri Enattah (2005), respectively.

1 INTRODUCTION

Congenital lactase deficiency (CLD) is a severe gastrointestinal disorder in newborns. It is characterized by watery diarrhea during the first days of life in infants who are fed lactose-containing milk. Diarrhea is caused by diminished lactase activity in the intestinal mucosa. CLD is a rare autosomal recessively inherited disorder and belongs to the group of 36 rare diseases found in Finland (Norio et al. 1973; Savilahti et al. 1983; Norio 2000; Norio 2003c). Holzel and colleagues (1959) discovered the first patients in 1959. Adult-type hypolactasia (also known as lactase non-persistence, primary lactose malabsorption) is a common gastrointestinal condition whose recessive inheritance was shown in 1973 (Sahi et al. 1973). It is caused by developmental downregulation of lactase (*LCT*) resulting in unhydrolysed lactose in the intestinal mucosa. Clinical symptoms (lactose intolerance) such as diarrhea are mainly due to osmosis (Auricchio et al. 1963; Dahlqvist et al. 1963). Adult-type hypolactasia represents a normal physiological condition after weaning (Simoons 1970). However, the majority of northern Europeans and some nomadic population in Africa and Arabia have the ability to maintain lactase activity and digest lactose throughout life (lactase persistence) (Swallow and Hollox 2000). Adult-type hypolactasia has shown to associate with the C/T_{-13910} variant, which is located 13.9 kilobases (kb) upstream of the *LCT* gene. The C/C_{-13910} genotype is associated with lactase non-persistence whereas C/T_{-13910} and T/T_{-13910} genotypes are associated with lactase persistence.

Galactose, the hydrolyzing product of the milk sugar lactose, has been hypothesized to be toxic to ovarian epithelial cells and consumption of dairy products with lactase persistence has been suggested to be a risk factor for ovarian carcinoma. Ovarian carcinoma is the fourth most common cause of cancer death in women. The cause and pathogenesis of this disease has remained obscure.

In this thesis, linkage disequilibrium (LD) and haplotype mapping was applied to refine the CLD locus and direct sequencing to identify mutations underlying CLD. In addition, the functional effect of the variant C/T_{-13910} was studied on the mRNA level of the *LCT* gene in children and adults using the solid-phase minisequencing method. In order to assess whether lactase persistence is a risk factor for ovarian carcinoma we have determined the C/T_{-13910} variant in Finnish, Polish and Swedish women with ovarian carcinoma and defined its relation to their corresponding control populations.

2 REVIEW OF THE LITERATURE

2.1 The Finnish disease heritage

The exceptional pattern of rare hereditary disorders in Finland, the Finnish disease heritage, was introduced by Doctors Norio, Nevanlinna and Perheentupa in 1973 (Norio et al. 1973). The very first of these distinctive disorders the Finnish type congenital nephrotic syndrome was described in 1956 (Hallman et al. 1956). Its mode of inheritance was traced back ten years later from the very carefully maintained records of church books reaching back to the 16th century (Norio 1966). Records of church books and land tax registers have been the keys to exploring Finnish hereditary disorders. There are 36 disorders which belong to the Finnish disease heritage, 32 of which are autosomal recessive, two X-chromosomal and two dominantly inherited (Table 1). The incidence of these disorders is in the range of 1:10000-1:100000. The higher prevalence of certain disorders in Finland and on the other hand the lack of disorders that are prevalent elsewhere in the world have inspired researchers to investigate the origin of Finns and their genes (Markkanen et al. 1987; Norio 2000; Norio 2003a). Disorders of the Finnish disease heritage are introduced in Table 1.

Table 1. Disorders of the Finnish disease heritage and their defective proteins or loci in alphabetical order

Disease [OMIM number]	Defective gene/protein or locus	Reference
Amyloidosis V [105120] **(dominant)**	Gelsonin (GSN)	Levy et al. 1990, Maury et al. 1990
Aspartylglucosaminuria (AGU) [208400]	Aspartylglucosaminidase (AGA)	Ikonen at al. 1991
Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) [240300]	Autoimmune regulator (AIRE)	Nagamine et al. 1997, The Finnish-German APECED Consortium 1997
Batten disease [204200]	CLN3	The International Batten Disease Consortium 1995
Cartilage-hair hypoplasia (CHH) [250250]	RNA component of Rnase MRP (RMRP)	Ridanpää et al. 2001
Choroidemia (CHM) [303100] **(X-linked)**	Rab escort protein 1 (REP1)	Sankila et al. 1992
Cohen syndrome (COH1) [216550]	COH1	Kolehmainen et al. 2003
Congenital chloride diarrhea (CCD) [214700]	Solute carrier family 26, member 3 (SLC26A3)	Höglund et al. 1996
Congenital lactase deficiency (CLD) [223000]	LCT	Study I
Congenital nephrosis (CNF) [256300]	Nephrin	Kestilä et al. 1998
Cornea plana congenita (CNA2) [217300]	Keratocan (KERA)	Pellegata et al. 2000
Diastrophic dysplasia (DTD) [222600]	Solute carrier family 26, member 2 (SLC26A2)	Hästbacka et al. 1994

Free sialic acid storage disease (Salla disease) [604369]	Solute carrier family 17, member 5 (SLC17A5)	Verheijen et al. 1999
Growth retardation, aminoaciduria, cholestasis, iron overload lactacidosis and early death (GRACILE) [603358]	BCS1-like (BCS1L)	Visapää et al. 200
Gyrate atropy (GA) [258870]	Ornithine aminotransferase (OAT)	Mitchell et al. 1989
Hydrolethalus syndrome (HLS) [236680]	HYLS 1	Mee et al. 2005
Hypergonadotrophic ovarian dysgenesis (ODG1) [233300]	FSH receptor (FSHR)	Aittomäki et al. 1995
Infantile neuronal ceroid-lipofuscinosis (INCL) [256730]	Palmitoyl protein thioesterase 1	Vesa et al. 1995
Infantile onset spinocerebellar ataxia (IOSCA) [271245]	Twinkle and Twinky	Nikali et al. 2005
Lethal arthrogyposis with anterior horn cell disease	unknown	Clinical characterization by Vuopala et al. 1995
Lethal congenital contracture syndrome (LCCS) [253310]	9q34	Mäkelä-Bengs et al. 1998
Lysinuric protein intolerance (LPI) [222700]	Solute carrier family 7, member 7 (SLC7A7)	Borsani et al. 1999, Torrents et al. 1999
Meckel syndrome (MKS1) [249000]	MKS1	Kyttälä et al. 2006
Megaloblastic anemia 1 [261100]	Cubilin (CUBN)	Aminoff et al. 1999
Mulibrey nanism [253250]	TRIM37	Avela et al. 2000
Muscle-eye-brain disease (MEB) [253280]	POMGnT1	Diesen et al. 2004
Nonketotic hyperglycinemia (NKH) [605899]	Glycine decarboxylase (GLDC)	Kure et al. 1992

Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (PLO-SL) [221770]	TYRO protein tyrosine kinase-binding protein (TYROBP) and TREM2	Paloneva et al. 2000, Paloneva et al. 2002
Progressive encephalopathy with edema, hypsarrhythmia and optic atrophy (PEHO syndrome) [260565]	unknown	Clinical characterization by Salonen et al. 1991
Progressive epilepsy with mental retardation (EPMR) [600143]	CLN8	Ranta et al. 1999
Progressive myoclonus epilepsy (EPM1) [254800]	Cystatin B (CSTB)	Pennacchio et al. 1996; Virtaneva et al. 1997
RAPADILINO syndrome [266280]	RECQL4	Siitonen et al. 2003
Retinoschisis [312700] **(X-linked)**	Retinoschisin	The Retinoschisis Consortium 1998
Tibial muscle dystrophy (TMD) [600334] **(dominant)**	Titin (TTN)	Hackman et al. 2002
Usher syndrome type III [276902]	USH3A	Joensuu et al. 2001
Variant late infantile neuronal ceroid-lipofuscinosis (vLINCL) [256731]	CLN5	Savukoski et al. 1998

Thirty-two disorders are recessively inherited, two are dominantly inherited and two are X-linked traits (in **bold** text). The table is based on Norio (2000) and Norio (2003c).

2.1.1 Causes of the unique gene pool

2.1.1.1 The settlement history of Finland

The predominant hypothesis is that Finland has been inhabited continuously from the last glacial period approximately 10000 years ago (Nunez 1987; Norio 2003b). The oldest archaeological discovery from southern Finland is circa 9000 years old

whereas the first signs of settlement date from 8000 years ago (Huurre 1992; Jutikkala and Pirinen 1996). It cannot be known for sure who the first settlers were. They could have been inhabitants, despite the arctic circumstances, living on the coast of Scandinavia from the glacial period or/and immigrants who returned from the south as the ice melted. However, some of these first settlers are probably the ancestors of the Saami and some of them adopted agriculture and merged with the growing southern settler groups. In prehistoric Scandinavia times the warm Gulf stream offered bearable living conditions for a small population whose source of livelihood was based on hunting, fishing, and gathering (Norio 2000; Norio 2003b). Nowadays, there are only 60000 Saami living in the northern regions of Finland, Sweden, Norway and the Russian Kola Peninsula of these 6000 individuals live in Finland. The climate and source of livelihood were not conducive to expansion of population, the size has been estimated to have been constant throughout history (Lehtivirta and Seurujärvi-Kari 1991). Still warming climate forced the Saami to retreat gradually to the north away from growing settlers groups who had adopted agriculture (Eriksson 1973; Lehtivirta and Seurujärvi-Kari 1991). These settler groups carried the genes of today's Finns. Where did the farmers come from? The dual theory of Finland's inhabitation by Eriksson (1973) and Norio (1981) suggests there was an early migration of eastern Uralic speakers ~4000 years ago and a later migration from the south ~2000 years ago (Figure 1). Kittles and his colleagues's (1998) Y chromosome haplotype study supported this theory as they observed two different male lineages. Mitochondrial and nuclear DNA analyses show that the majority of Finnish genes are of central European origin (Cavalli-Sforza and Piazza 1993; Sajantila et al. 1995; Sajantila and Pääbo 1995; Lahermo et al. 1996; Laan and Pääbo 1997; Torroni et al. 1998). The Saami and other Europeans including Finns have been suggested to have a divergent population history (Cavalli-Sforza and Piazza 1993; Sajantila et al. 1995; Sajantila and Pääbo 1995; Lahermo et al. 1996). A recent mitochondrial and Y chromosomal study suggests that most likely the Saami have European origin after all. The large genetic difference between the

Saami and other Europeans is explained by the Saami representing a narrow subset of Europeans (Tambets et al. 2004). Since the last glacial period all migrations from south, east and also small ones from west shaped the Finnish population genetically, linguistically and culturally but kept as a relatively small founder population.

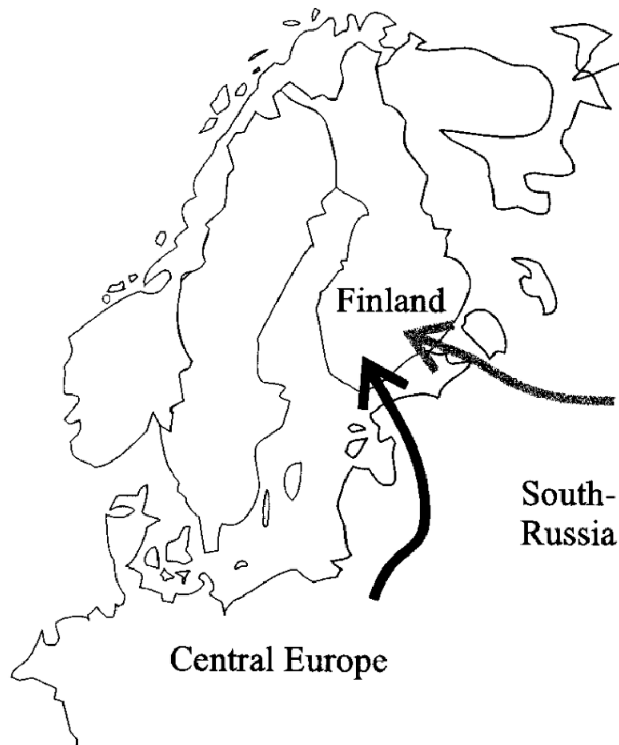


Figure 1. The main two migration waves came to the area of Finland from south and east. Adapted from Varilo (1999).

2.1.1.2 The internal migration in the 16th century

In the 16th century the Finnish population had grown to 250 000 and inhabitation who were focused in the coastal regions (Figure 2). Sweden acquired Denmark after the Liberation War in 1521-1523 and Gustavus of Vasa (1523-1560) was crowned

monarch of a new independent state Sweden-Finland. King Gustavus of Vasa was interested in his kingdom and made administrative reforms, which included increasing taxation and district court sessions. The superpower politics of King Gustavus of Vasa favoured the inhabitation of wildernesses to enlarge the country's borders. Farming were encouraged by lower taxation for farmers; also the growth of the population led to pressure to cultivate more land. The majority of inhabitants originated from the area of southeastern Finland called South Savo (Figure 2). Within two centuries nearly the whole area of Finland was inhabited and the population had increased to 400000. There was a great famine in 1696-1698 which decreased the population to 250000 but from the 1700s the Finnish population has grown pretty quickly to its present figure of 5300000 e.g. (Jokipii 1992; Jokipii and Rikkinen 1992; Jutikkala and Pirinen 1996; Varilo 1999; Norio 2003a).

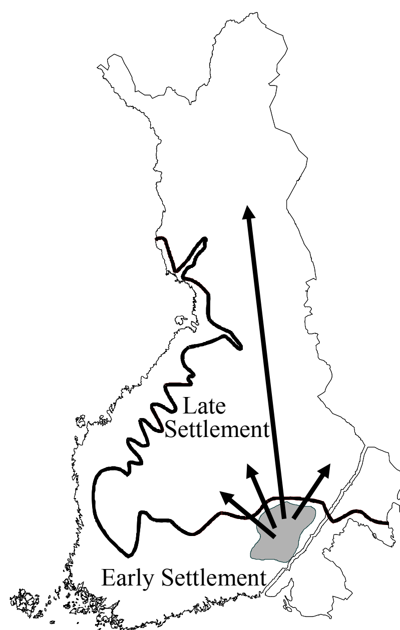


Figure 2. The internal migration movement led to regional subisolates in the Finland area in the 16th century. The map is modified from Varilo (1999).

The small number of original ancestors both due to genetic drift and regional subisolates, a rapid increase in population and linguistic and geographical isolation have strongly shaped the Finnish population and created our unique gene pool. Our carefully maintained church books, land registers and history books have given us tools to study our origin and rare diseases (Norio et al. 1973; Norio 2000).

2.2 Lactase

Lactase (*LCT*), also known as lactase-phlorizin hydrolase, belongs to the β-galactosidase family having both lactase (EC 3.2.1.108) and phlorizin hydrolase activity (EC 3.2.1.62). *LCT* is exclusively expressed in the small intestine and it is responsible for hydrolysing lactose (Figure 3) to glucose and galactose. Lactose is a major disaccharide found in milk synthesized by lactose synthetase in the mammary gland. Human milk has the highest lactose content (7%) e.g. cow milk contains 4.8% of lactose. Phlorizin hydrolase is responsible for hydrolysing aryl- and alkyl-β-glycosides such as phlorizin and galactosyl- and glycosyl-β-ceramides (Sahi 1994b). Lactase activity is delimited in mammals whereas phlorizin hydrolase activity has been detected in vertebrates. The natural substrate of phlorizin hydrolase has been thought to be glycosyl ceramides that inhere in the diet of most vertebrates and are also found in milk (Leese and Semenza 1973; Skovbjerg et al. 1981). Both catalytic activities are located on a single polypeptide chain (Wacker et al. 1992).

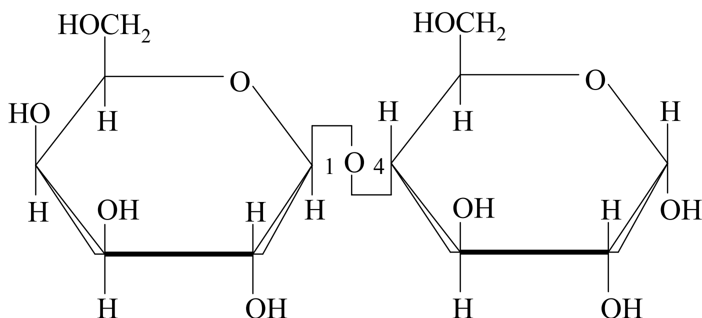


Figure 3. Lactose the milk sugar. In lactose, a glucose unit and a galactose unit are joined by an β-1, 4 glycosidic linkage.

2.2.1 Structure

The *LCT* gene reaches a genomic size of 50 kb and it is composed of 17 exons. There is one kb promotor region preceding *LCT* (Boll et al. 1991). The size of messenger RNA (mRNA) is 6274 bases and the primary translation product (pre-pro-LCT) is 1927 amino acids (Figure 3). Pre-pro-LCT consists of an N-terminal signal sequence, pre-domain (19 amino acids), a pro-domain size of 849 amino acids, an extracellular domain of 1014 amino acids, a hydrophobic transmembrane anchor domain (19 amino acids), and a short C-terminal cytosolic domain of 26 amino acids. Pro-LCT contains four internal repeats of which the pro-domain includes repeats I-II and an extracellular domain includes repeats III-IV (Figure 4). Phlorizin hydrolase activity locates in repeat III and lactase activity locates in repeat IV (Neele et al. 1995; Zecca et al. 1998; Arribas et al. 2000). These four repeats contain 38-55% identical amino acids. Mantei and colleagues (1988) suggest that repeats III-IV are results of duplication of the I-II repeats of one ancestor gene. Sequence similarities of the I-IV regions to β-glycosidases of archaebacteria, eubacteria, and fungi support this hypothesis and thus, lactase belongs to the β-glucosidase and β-galactosidase superfamily (Naim 2001).

2.2.2 Biosynthesis

Mature lactase is anchored to the intestinal membrane by a hydrophobic region near its carboxy terminus and the catalytic sites of the enzyme are located in the lumen of the intestine (Mantei et al. 1988). How is a lactase polypeptide processed to its active form? Intestinal epithelial cells synthesize lactase as a single chain pre-pro-LCT precursor polypeptide that is translocated over the endoplasmic reticulum (ER) guided by a signal sequence (Figure 4). The signal sequence is removed during the process resulting in pro-LCT with a molecular weight of 215 kiloDalton (kDa) (Naim and Naim 1996). Even if the pro-region of LCT (LCTα) does not have enzymatic activity (Naim 1995) it is has been shown to function as an intramolecular chaperone essential for the folding of pro-LCT (Oberholzer et al. 1993; Naim et al. 1994; Jacob et al. 2002). In ER, pro-LCT is glycosylated by mannose rich N-linked oligosaccharides. Glycosylated pro-LCT forms a homodimer that is further transferred to the Golgi apparatus. The transmembrane anchor domain and the cytosolic domain participate directly in the dimerization and apical sorting of lactase. The O-linked sugars of the pro-LCT dimer are glycosylated and N-linked sugars are further processed in cis-Golgi resulting in a glycoprotein with a molecular weight of 230 kDa. Glycosylation and dimerization of pro-LCT is required for efficient intracellular transport and enzyme activity (Naim et al. 1991; Naim and Lentze 1992; Naim 1994; Naim and Naim 1996; Panzer et al. 1998; Jacob et al. 2000). Pro-LCT undergoes two proteolytic cleavage steps before it finally becomes mature lactase. The first cleavage takes place intracellulary and removes the large LCTα at R734/L735 resulting in a membrane bound $LCT\beta_{initial}$ (L735-Y1927). $LCT\beta_{initial}$ is targeted to the intestine brush border membrane where it is cleaved by trypsin at R868/A869 leading to a 160 kDa mature lactase, $LCT\beta_{final}$ (Figure 4) (Jacob et al. 1996; Wüthrich et al. 1996).

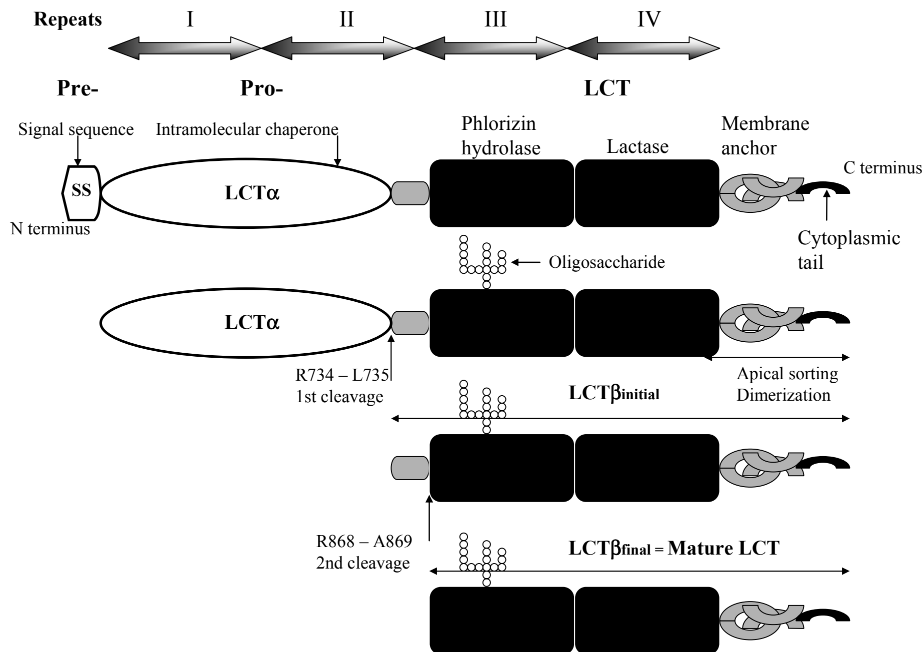


Figure 4. The structure and biosynthesis of lactase. Lactase is synthesized as a pre-pro-LCT precursor. The pre-pro-LCT contains a cleavable signal sequence that translocates the polypeptide over the endoplasmic reticulum (ER) (von Heijne 1986; Mantei et al. 1988). The pro-lactase region consists of four homologous repeats (I-IV). Repeats I and II are located in the pro-domain while repeats III-IV are found in mature lactase (Mantei et al. 1988). Later in the ER, the signal sequence is cut off yielding pro-LCT that is N-glycosylated and forms homodimers (Naim et al. 1991; Naim and Lentze 1992; Naim 1994; Naim and Naim 1996; Panzer et al. 1998; Jacob et al. 2000). The region involved in apical sorting and dimerization is indicated. The pro-region, LCTα, is shown to function as an intramolecular chaperone for folding of pro-LCT (Oberholzer et al. 1993; Naim et al. 1994; Jacob et al. 2002). The pro-LCT is transferred to the Golgi apparatus and O-glycosylated. The pro-LCT is further processed by two proteolytic cleavages: an intracellular cleavage that occurs between R734 and L735 produces LCT$\beta_{initial}$ and a cleavage in the intestinal lumen between R868-A869 generates LCTβ_{final} the mature enzyme (Jacob et al. 1996; Wüthrich et al. 1996). The figure is based on Naim (2001) and Troelsen (2005).

2.2.3 Regulation

Milk is the major nutrient consumed by newborn mammals. As lactose is the main carbohydrate and the most important source of energy in milk, it is necessary for mammals to be able to hydrolyse lactose after birth. *LCT* is expressed only by small intestinal enterocytes located at the crypt-villus junction and on the villus. Although, lactase activity is required from birth in mammals the timing of high *LCT* expression differs between rodents and humans. Humans have high *LCT* expression from birth whereas in mouse lactase expression reaches its maximum level 3 days after birth. This has been thought to be due to the intestinal immaturity of rodents. Fully developed crypts are not detected until 2-3 days after birth in mice. This finding has made *LCT* an ideal marker for fully differentiated enterocytes (Klein 1989; Freeman et al. 1993; Freund et al. 1995; Troelsen 2005).

2.2.3.1 Proximal LCT promoter

The proximal LCT promoter, a 150-basepair fragment upstream of the transcription initiation site (-150), is conserved in mouse, rat, rabbit, pig and human. Promoter analyses have revealed three transcription factor binding sites (CE1a, CE2c and GATA-site) so called cis-elements in this fragment (Figure 5) (Troelsen et al. 1992; Fitzgerald et al. 1998; Spodsberg et al. 1999). The functional role of transcription factors in these binding sites has been studied in transfection studies of promoter-reporter gene constructs and in DNA-protein and protein-protein interaction analyses using human intestinal adenocarcinoma (Caco-2) cells.

It has been shown that the caudal-related homeobox protein (Cdx-2) and the homeobox protein HOXC11 bind to CE1a (Troelsen et al. 1997; Mitchelmore et al. 1998; van Wering et al. 2002b). Cdx-2 is a significant factor for intestinal transcription and maintenance (Chawengsaksophak et al. 1997; Freund et al. 1998) and it is capable of activating pig (Troelsen et al. 1997) rat (Fang et al. 2000;

Krasinski et al. 2001) and human LCT promoters (Krasinski et al. 2001). Cdx-2 was also detected to bind to a site overlapping the *LCT* TATA-box. However, it remained unknown whether Cdx-2 binds directly to the site or through protein-protein interactions with a transcription initiation complex (van Wering et al. 2002b). The significance of this observation is to be clarified. HOXC11 is expressed in human fetal intestines but it could not be detected after birth suggesting a role in early intestinal development (Mitchelmore et al. 1998). Additionally, there are unidentified factors which repress *LCT* expression through the CE1a-site (Troelsen et al. 1997; van Wering et al. 2002b). Repression of *LCT* transcription by those factors was suggested to take place in cells that do not express Cdx-2 (van Wering et al. 2002b). This would indicate a tissue-specific LCT promoter activation by Cdx-2.

CE2c has been shown to be crucial for the LCT promoter activity. Mutations of the CE2c-site remarkably decrease the LCT promoter activity in pigs (Spodsberg et al. 1999) and in human (van Wering et al. 2002a). Hepatocyte nuclear factor-1α (HNF-1α) has been shown to bind to the CE2c-site (Mitchelmore et al. 1998; Spodsberg et al. 1999; Mitchelmore et al. 2000; Krasinski et al. 2001; van Wering et al. 2002a). HNF-1α has been shown to activate pig (Spodsberg et al. 1999) rat and human LCT promoters (Krasinski et al. 2001). Furthermore, HNF-1α and Cdx-2 has been shown to physically interact with and cooperatively activate pig (Mitchelmore et al. 2000) and human LCT promoters (Krasinski et al. 2001). HNF-1α was shown to be essential for *LCT* expression *in vivo*. *LCT* mRNA was reduced 95% in HNF-1α(-/-) mice compared with wild types (Bosse et al. 2006). HNF-1 transcription factors are expressed in a number of tissues such as intestine, liver, kidney, pancreas, stomach (Ott et al. 1991; Cereghini et al. 1992) and they have been shown to be involved in the activation of a number of intestinal genes (Bosse et al. 2006), genes of developmental appearance and embryonic regulation, as reviewed in Weber et al. (1996), and regulating the expression of genes that are expressed in the liver, kidney, and pancreas, reviewed in Pontoglio (2000).

GATA-4, -5 and -6 zinc finger transcription factor has been demonstrated to be capable of binding and activating individually the human (Fitzgerald et al. 1998; Krasinski et al. 2001), rat (Fang et al. 2001; Krasinski et al. 2001) LCT promoters. A study showed that GATA-4 is the key GATA-factor regulating lactase expression in mice (van Wering et al. 2004). GATA-4/HNF-1α and GATA-5/HNF-1α has been shown to physically interact with and synergistically activate mouse, rat and human LCT promoters (Krasinski et al. 2001; van Wering et al. 2002a; van Wering et al. 2004). GATA factors are expressed in a number of tissues. A subfamily of GATA-1, -2, and -3 is expressed in hematopoietic stem cells regulating differentiation-specific gene expression in T-lymphocytes, erythroid cells, and megakaryocytes, reviewed in Orkin (1998). A subfamily of GATA-4, -5, and -6 are expressed in a variety of mesoderm- and endoderm-derived tissues such as heart, liver, lung, gonad and intestine where they regulate tissue-specific gene expression, reviewed in Molkentin (2000).

The pancreatic duodenal homeobox-1 protein (Pdx-1) has been demonstrated to repress endogenous LCT promoter activity when Pdx-1 was overexpressed in Caco-2 cells. Pdx-1 is expressed the anterior duodenal region of the intestine (Guz et al. 1995) where *LCT* expression is repressed in adult mammals. A Pdx-1 candidate binding site TAAT was identified in the rat promoter but mutation analysis of the binding site failed to recover *LCT* expression suggesting a more complex regulation pattern (Wang et al. 2004). Pdx-1 binding sites have not been identified in human or pig LCT promoters (Troelsen 2005).

2.2.3.2 Distal regulatory elements

Even though, there is no doubt of the importance of the proximal LCT promoter in the regulation of the *LCT* gene it has a relatively weak effect on transcription in Caco-2 cells (Troelsen et al. 1992). There is evidence that distal regulatory elements, enhancers, are needed to complete the proximal promoter in order to obtain full *LCT* expression (Figure 5). This type of enhancer sequences was identified in the pig

locating around -850 of the lactase gene, this was necessary for high expression in Caco-2 cells (Spodsberg et al. 1999; Troelsen et al. 2003a). This enhancer region includes three binding sites CE2a, nt20 and CE2b of which CE2a is known to bind HNF-1α, the other two in binding factors are unidentified (Simon et al. 1995; Spodsberg et al. 1999; Troelsen et al. 2003a). The enhancer element is not conserved in human nor rat. Furthermore, a negative regulatory region CE3 has been identified in the pig and human promoter (Spodsberg et al. 1999). The Forkhead box (Fox) factors FREAC-2 and -3 have been shown to bind to this element in the pig, (Spodsberg et al. 1999) in human the same element binds intestinal nuclear factors (Hollox et al. 1999).

Experiments in transgenic mice have shown that distal regulatory sequences located at -2038-(+15) of the rat LCT promoter are important for tissue specificity, correct spatial and developmental expression (Krasinski et al. 1997; Lee et al. 2002; Wang et al. 2006). DNAse footprint analyses have revealed several Fox and a leucine zipper factor C/EBP in the rat LCT promoter (Figure 5). HNF-3 has been shown to activate the promoter whereas a negative regulatory region similar to the pig and human promoter repress the activity (Verhave et al. 2004). Even if distal regulatory regions are not conserved in pig, rat and human the same type of pattern does exist.

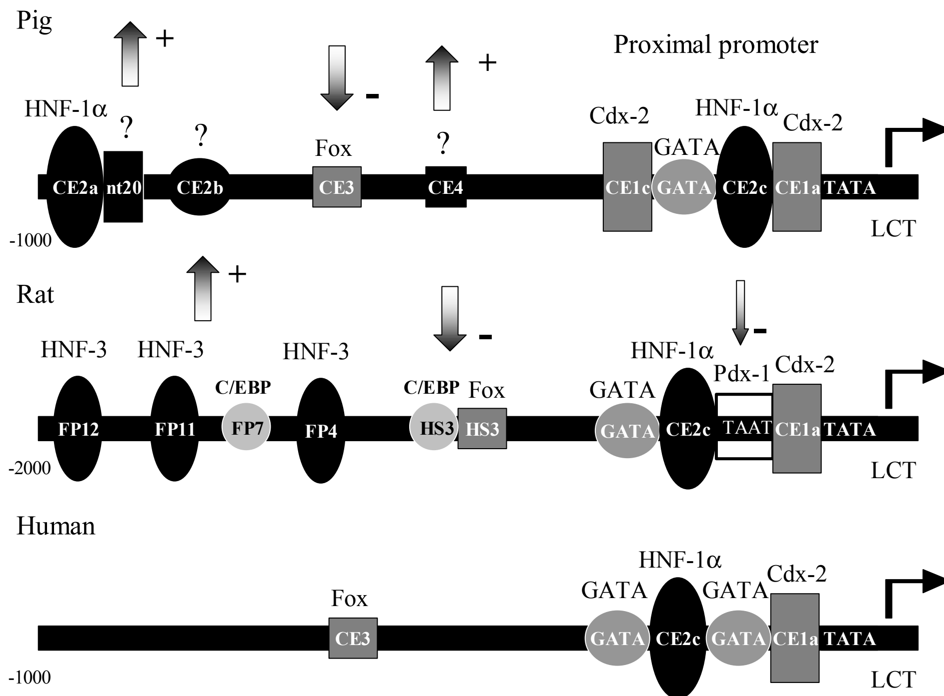


Figure 5. Proximal and distal regulatory elements of pig, rat and human LCT promoters. Arrows with + or – indicate the regulatory effect of cis-elements via proximal promoter on *LCT* transcription *in vitro*. Cis-elements are named as follows: CE (cis-element), HS (hypersensitive site) and FP (footprint). Transcription factors that bind the cis-elements are shown, question marks indicate unknown factors. The Figure is based on studies referred to on pages 26-29 and Troelsen (2005).

2.3 Congenital lactase deficiency

Congenital lactase deficiency (CLD [MIM 223000]) is an autosomal recessively inherited severe gastrointestinal disorder. Holzel et al. (1959) described the first patients in 1959. Launiala, Kuitunen, and Visakorpi (1966) discovered the absence of lactase activity in duodenal specimens of infants with explosive diarrhea after breastfeeding. In a clinical study on 16 patients Savilahti and colleagues (1983) gave the first evidence for a recessive mode of inheritance for CLD. So far, 50 patients in

42 families have been diagnosed in Finland (Savilahti et al. 1983; Study I) and several cases have been reported elsewhere in the world (Holzel 1967). CLD has an incidence of 1:60000 in the Finnish population (Savilahti E personal communication). The birthplaces of great-grandparents of 31 Finnish CLD families demonstrate that CLD mutation is emphasized in central Finland (Figure 6) (Järvelä et al. 1998). CLD is one of the 36 rare monogenic disorders enriched in Finland (Norio et al. 1973; Norio 2000; Norio 2003c).

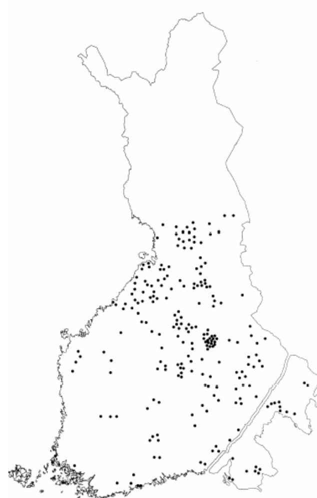


Figure 6. The birthplaces of the great-grandparents of 31 CLD families in Finland. The figure is adapted from Järvelä et al. (1998).

2.3.1 Clinical features

CLD is characterized by a pure watery diarrhea that is due to osmosis developed by unhydrolysed lactose. Feces are acidic (pH 4.5-5.5) and contain large quantities of lactose as a result of undeveloped bacterial flora. Later, at the age of 1.7 years or older lactose is absent, probably fermented by colonic bacterial flora creating flatulence and abdominal pains. The severe diarrhea followed by dehydration,

metabolic acidosis, and weight loss is usually diagnosed during the first weeks or months of life (Savilahti et al. 1983). Some patients have hypercalcemia and medullar nephrocalcinosis (Saarela et al. 1995). Serum cholesterol is reduced but the levels of triglycerides are normal. Regardless of the symptoms, CLD infants are lively and they have a good appetite (Savilahti et al. 1983). Lactase activity determined in duodenal biopsy specimen is very low (0-10 U/g protein). The morphology of duodenum and sucrase, maltase and isomaltase activities are observed to be normal. Normal psychomotor development and growth is achieved and the symptoms disappear when the patients are put on a lactose-free diet (Savilahti et al. 1983).

2.3.2 Assignment of the locus to 2q21

As lactase is responsible for the hydrolysis of lactose into galactose and glucose in the intestinal lumen it was an obvious candidate for CLD. *LCT* was localized to chromosome 2q and subsequently 2q21 by Kruse et al. (1988) and Harvey et al. (1993), respectively (Figure 7). Using this information Järvelä and colleagues (1998) found linkage for CLD on 2q21 in the close proximity of the *LCT* gene in 19 Finnish families. However, extended haplotype analysis by seven polymorphic microsatellite markers seemed to refine the CLD locus in the 350 kb area between D2S314 and D2S2385 from two centiMorgans (cM) telomeric from *LCT* (Figure 7). These microsatellites gave the highest LD in the disease alleles. This finding supported the hypothesis of one major mutation in most CLD patients in Finland. Based on strong LD spanning an area of ~8 cM at the CLD locus, the mutation was estimated to have been enriched by approximately 30 generations in the Finnish population (Järvelä et al. 1998). Sequence analysis of *LCT* and its promoter region failed to reveal disease-causing mutations in a Finnish CLD patient (Poggi and Sebastio 1991). This finding and the haplotype analyses of Finnish CLD patients encouraged the authors to suggest that there is a novel gene underlying CLD (Järvelä et al. 1998).

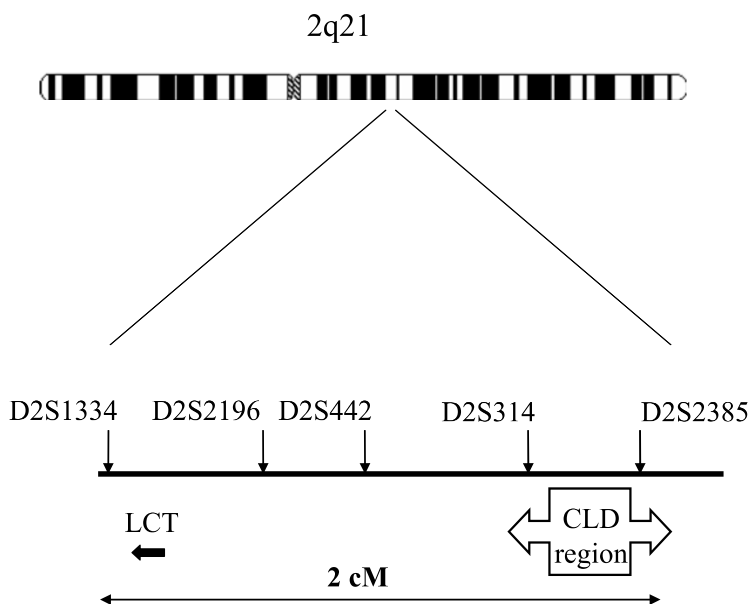


Figure 7. Physical map of the CLD region. According Järvelä et al. (1998) the critical CLD region located 2 cM telomeric from the *LCT* gene.

2.4 Adult-type hypolactasia

Adult-type hypolactasia (MIM 223100) (lactase non-persistence, lactose intolerance) is an autosomal recessive gastrointestinal condition that is the result of a decline in the activity of lactase in the intestinal lumen after weaning. Downregulation of lactase is considered as a normal phenomenon among mammals, and symptoms are remarkably milder than experienced in CLD. Milk intolerance was described in 1901 when carbohydrate ingestion was noted to be linked with pathogenesis of diarrhea (Montgomery et al. 1991). Röhmann and Nagano (1903) demonstrated the pathophysiologic mechanism of this condition in dogs as early as 1903 when they observed unhydrolysed lactose molecules in the intestinal lumen, reviewed in Sahi (1994a). The lactase activity was detected in the duodenum but the activity was highest in the jejunum and decreased towards the ileum. Absorption was observed in

the jejunum (Borgström et al. 1957). In 1963, Auricchio et al. (1963) and Dahlqvist et al. (1963) reported a decrease in lactase activity from birth in humans. Interestingly, some humans have the ability to maintain lactase activity and digest lactose throughout their lives (lactase persistence). The reasons for these phenomena have been speculated on for a quite some time. It was noted that lactase activity could not be induced by intake of lactose (Keusch et al. 1969b). Genetic factors were observed to play a role when evidence for the autosomal recessive inheritance was published in 1973 (Sahi et al. 1973). These fascinating characteristics have inspired researchers to study the molecular background of the lactase non-persistence and persistence phenotypes.

2.4.1 Diagnosis

Diagnosis of adult-type hypolactasia is commonly assessed by the lactose tolerance test (LTT). The method is indirect and is based on a serial of measurements of glucose in blood after oral lactose ingestion (50 g). A small rise in blood glucose concentration (<1.1 mmol/l) after a 20, 40, 60 and 90 minutes of lactose load indicate hypolactasia. A rise in blood glucose of more than 1.7 mmol/l indicates lactase persistence. In addition, clinical symptoms, such as possible stomach pains, distension, cramps, flatulence, nausea and diarrhea are registered (Sahi and Launiala 1978; Arola 1994). Lactose tolerance test with ethanol (LTTE) was a successful step towards a more specific test. Galactose in blood is measured instead of glucose and ethanol is used to inhibit the conversion of galactose to glucose in the liver (Jussila 1969; Isokoski et al. 1972; Arola 1994). One measurement at 40 min after lactose and ethanol ingestion is sufficient for a diagnosis. Blood galactose concentrations less than 0.3 mmol/l indicate hypolactasia (Isokoski et al. 1972). In addition, there are some applications of LTTE, which rely on urinary galactose determinations, reviewed in detail in Arola (1994). Tolerance tests for glucose and galactose are needed for exclusion of secondary malabsorption (Jussila 1969). Diagnosis by the

breath hydrogen test (BHT) after lactose ingestion is based on the definition of hydrogen from exhaled air. Hydrogen concentration is measured by gas chromatography from a set of samples taken for 2 to 6 hours. Usually hydrogen concentration >20 ppm indicates hypolactasia (Metz et al. 1975; Metz et al. 1976; Arola 1994). However, a definitive diagnosis of adult-type hypolactasia can be made by measuring lactase together with sucrase and maltase activities directly from intestinal biopsy specimen and exploring the histology of the mucosa (Dahlqvist 1984). Lactase activity less than 10 U/g protein and a ratio of lactase to sucrase activities less than 0.30 indicate hypolactasia (Jussila 1969; Dahlqvist 1984). Histological evaluations are important to study possible secondary reasons for a low lactase activity. Infections and inflammations have been found to decrease activities of several intestinal enzymes (Phillips et al. 1988). The comparison of indirect LTT, different LTTEs and BHT with direct mucosal disaccharidases shows that their sensitivities and specificities vary between 69 to 100% (Isokoski et al. 1972).

2.4.2 Onset and prevalence

In a number of populations lactase activity starts to decline a few years after weaning. The age at downregulation varies between populations. For example in Thais downregulation takes place between 1-2 years of age (Keusch et al. 1969a). In African children, the first signs of decline of lactase activity was detected at 3 years of age but in Finnish children somewhat later at 5-12 years of age (Rasinperä et al. 2004). A British study of different ethnic groups demonstrated that downregulation of *LCT* expression was detected from the second year of life, although the extent and onset was not constant (Wang et al. 1998b). The reasons for these timing variations are unknown.

The prevalence of adult-type hypolactasia varies between populations. The prevalence of adult-type hypolactasia is highest in Asia, for example 90-100% of

Thais, Chinese, and Japanese suffer from a decline of lactase activity (Flatz and Saengudom 1969). A total of 81-91% of the blacks of Africa cannot digest lactose (Cook and Kajubi 1966; Olatunbosun and Kwaku Adadevoh 1971). Lactase non-persistence is a frequent phenotype among native population of Australia and America reviewed in Swallow and Hollox (2000). In contrast, a low prevalence of hypolactasia 1-9.6% have been observed in Sweden and Denmark (Gudmand-Höyer et al. 1969; Dahlqvist and Lindquist 1971; Nilsson TK personal communication). Also, some nomadic populations in Africa and Arabia such as the Beja, Beduin, Fulbe and Tuareg can tolerate lactose efficiently, only 0-24% of them develop hypolactasia, reviewed in Swallow and Hollox (2000). Among the Finns the prevalence is about 17-18% while the prevalence in Finnish Swedish-speaking population is only 8% (Jussila et al. 1970; Sahi 1974; Enattah et al. 2002).

2.4.3 Genetics

2.4.3.1 Evolutionary favour of lactase persistence

Lactase non-persistence is considered to be the ancestral phenotype, lactase persistence has been regarded to be evolutionary advantageous when milk from domestic cows became available as a source of nutrition (Hollox 2005). It has been rationalized that this gene-culture coevolution created selective pressure for individuals who could efficiently use milk for nutrition in adulthood. This has been supported by observations as dairy farming and lactase persistence coincide, by the epidemiologic data as there are a wide distribution of the prevalence in lactase persistence between populations and by the fact that lactase persistence is genetically determined (Simoons 1969; McCracken 1970; Simoons 1970; McCracken 1971; Sahi et al. 1973; Sahi 1994a; Enattah et al. 2002). The gene-culture coevolution hypothesis was strongly supported when a high diversity in cattle milk protein genes and lactase persistence was demonstrated to coincide in

Europe (Beja-Pereira et al. 2003). Furthermore, convincing lines of evidence for the selective pressure of lactase persistence were obtained when genetic data nearby *LCT* was evaluated in northern European or derived populations. The lactase persistence locus was observed to contain an exceptionally long haplotype, LD was detected up to 1 Mb (Hollox et al. 2001; Enattah et al. 2002; Poulter et al. 2003; Bersaglieri et al. 2004) which was demonstrated to be longer and more common (77% of northern Europeans) than expected only by chance (Bersaglieri et al. 2004). The strong selection was estimated to have been taken place during past the 10000 years, 400 generations coinciding with a dairy culture (Bersaglieri et al. 2004; Coelho et al. 2005; Myles et al. 2005; Enattah et al. unpublished). The selection power was calculated to be 1.4-15% which is in agreement with previous estimations of 1-7% (Cavalli-Sforza 1973; Heston and Gottesman 1973; Flatz and Rotthauwe 1977; Aoki 1986; Flatz 1987). These findings indicate that lactase persistence has undergone the strongest positive selection seen in the human genome (Bersaglieri et al. 2004). Indeed, genetic, cultural and epidemiologic signs indicate that the ability to use milk efficiently for nutrition improved the survival opportunities of early farmers.

2.4.3.2 Mechanism of downregulation

Mutation analysis of *LCT* and its one kb sized promoter area have not shown the sequence variation associated with adult-type hypolactasia (Boll et al. 1991; Lloyd et al. 1992). In addition, a number of single nucleotide polymorphisms (SNPs) have been identified in one Mb fragment of LD at *LCT* but none of them could explain the lactase persistence phenotype (Harvey et al. 1995; Harvey et al. 1998; Hollox et al. 2001; Poulter et al. 2003). Many studies have been conducted at the cellular level to explain the phenotype differences and as a result several factors have been reported to influence the decline of lactase after childhood. At first, posttranscriptional regulation was suggested when *LCT* mRNA levels could not be

seen to correlate with lactase activity (Sebastio et al. 1989). Both slow processing of the lactase protein or/and reduction of pro-LCT synthesis were observed in metabolic labelling studies in lactase non-persistent individuals (Sterchi et al. 1990; Witte et al. 1990; Lloyd et al. 1992). Biosynthesis of pro-LCT has been detected to correlate with *LCT* mRNA levels but not lactase activity. A comparison of *LCT* mRNA levels and lactase activity/*LCT* mRNA level ratios indicated a heterogeneous pattern of regulation in both hypolactasic and lactase persistent individuals (Rossi et al. 1997). Even, mosaic regulation of lactase was observed from individuals with adult-type hypolactasia (Maiuri et al. 1994). However in the majority of cases, the *LCT* mRNA level have been shown to correlate with lactase activity or the ratio of lactase to sucrase (L/S) activities indicating that the decline of lactase activity is regulated at the transcriptional level (Escher et al. 1992; Lloyd et al. 1992; Fajardo et al. 1994). Analogous results have been observed in studies on animals (Krasinski et al. 1994; Lacey et al. 1994). Later, the lactase persistence/non-persistence was demonstrated to be controlled by a cis-acting element (Wang et al. 1995). In that study, *LCT* steady state mRNA levels were studied utilizing SNPs in the coding region of *LCT*. Certain types of lactase persistent were detected to have asymmetric allelic mRNA expression suggesting that expression of the *LCT* alleles could be independently regulated. Authors concluded that a developmentally regulated trans-acting DNA-binding protein could bind to only one kind of lactase allele and influence transcription and/or mRNA stability (Wang et al. 1995; Wang et al. 1998b).

2.4.3.3 Discovery of the responsible variant

Enattah and colleagues (2002) used the candidate gene approach and restricted the locus of adult-type hypolactasia to a 47 kb area at the 5'-end of the *LCT* gene on 2q21 using LD and haplotype analysis of nine extended Finnish families. Diagnoses for adult-type hypolactasia were assessed by LTTE. Sequence analysis of the 47 kb region exposed two SNPs C/T_{-13910} (rs4988235) and G/A_{-22018} (rs182549), 14 kb and

22 kb upstream from the initiation codon of *LCT*, respectively. The SNPs C/T$_{-13910}$ and G/A$_{-22018}$ are located in introns 13 and 9 of the minichromosome maintenance deficient 6 (*MCM6*) gene, respectively (Figure 8). These SNPs cosegregated completely with the adult-type hypolactasia trait in Finnish families. All hypolactasic individuals showed homozygosity for C$_{-13910}$ and G$_{-22018}$. Furthermore, in an independent sample set of 236 individuals, with biochemically verified disaccharidase activities from intestinal biopsy specimens from four different populations, the C/T$_{-13910}$ SNP was completely found to be associated with the trait and the SNP G/A$_{-22018}$ associated in 229 of 236 cases (Enattah et al. 2002). The frequency of the C/C$_{-13910}$ in a sample set of 1047 individuals was in agreement with the reported prevalence of adult-type hypolactasia in Finnish, French, African American and South Korean populations and North American Caucasians (Sahi 1974; Simoons 1978; Cuddenec et al. 1982). As early as this time, the lactase persistence allele, T$_{-13910}$, was anticipated to be very old as it was found in distally related populations (Enattah et al. 2002).

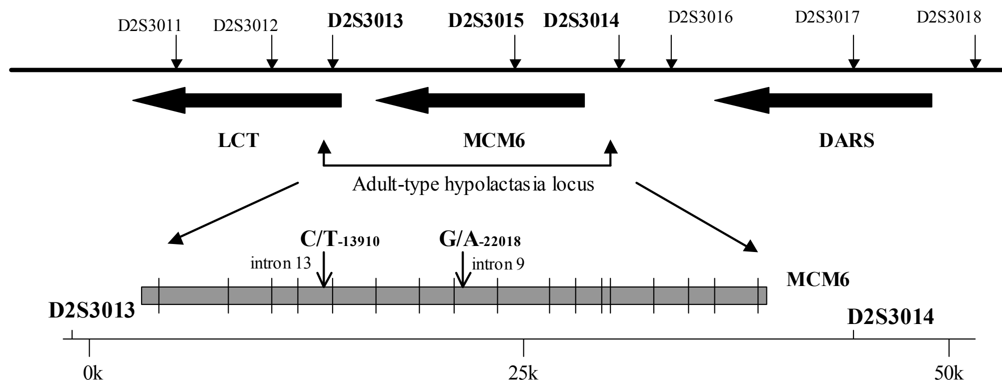


Figure 8. The physical map of the adult-type hypolactasia locus. The adult-type hypolactasia locus is located between markers D2S3013 and D2S3014. The horizontal arrows indicate the positions of lactase (*LCT*), minichromosome maintenance deficient 6 (*MCM6*) and aspartyl-tRNA synthetase (*DARS*). SNPs C/T$_{-13910}$ (rs4988235) and G/A$_{-22018}$ (rs182549) are located in intron 13 and 9 of the *MCM6* gene, 14 and 22 kb from *LCT*, respectively. Modified from Enattah et al (2002).

2.4.3.3.1 Functional evidence

Both the SNPs, C/T_{-13910} and G/A_{-22018}, were found in introns of *MCM6*, which are located in close proximity of, only 2.5 kb upstream, *LCT*. MCM6 is a mammalian homologue of mis5 of yeast, it functions as a cell cycle factor (Takahashi et al. 1994) and its mammalian version was initially identified in intestinal crypt cells of the rat (Sykes and Weiser 1995). *MCM6* is expressed in various human tissues including intestine during development. However, intestinal expressions of *MCM6* and *LCT* were not observed to match with lactase persistence or non-persistence individuals suggesting that these genes are independently regulated (Harvey et al. 1996). Thus, it seems that the influence of the *MCM6* gene on adult-type hypolactasia is only structural. The next evident step was to define the significance of C/T_{-13910} and G/A_{-22018} on the regulation of lactase activity. The answer for the dilemma was sought in the present study (III, see discussion) but later also by others using human (Troelsen et al. 2003b) and rat (Olds and Sibley 2003) LCT promoter-reporter gene construct analyses in Caco-2 cells. Both C_{-13910} and T_{-13910} variants were observed to enhance LCT promoter activity. The region containing T_{-13910} was demonstrated to increase 25%-75% LCT promoter activity compared to the results with the region cloned with C_{-13910} in undifferentiated Caco-2 (Olds and Sibley 2003; Troelsen et al. 2003b). Interestingly, the enhancer activity of C_{-13910} and T_{-13910} was detected to be several times higher in differentiated Caco-2 cells. The differentiated Caco-2 cells are commonly considered as a better model as they endogenously express *LCT*. Also, the difference of enhancer activities between C_{-13910} and T_{-13910} was found to be clearer, the T_{-13910} variant had a 3-6-fold increase in transcription compared with the C_{-13910} variant. Furthermore, electrophoretic mobility shift assays (EMSAs) showed a strong interaction between the T_{-13910} variant and a nuclear factor in nuclear extracts of HeLa and differentiated Caco-2 cells. On the contrary, the C_{-13910} variant had a weak interaction suggesting that the observed difference of LCT promoter activities are due to the binding capacity of C_{-13910} and T_{-13910} of a nuclear factor (Troelsen et al. 2003b). Analogous analyses of

the G/A_{-22018} variant with human and rat LCT promoters did not reveal any significant functional role for phenotype differences of adult-type hypolactasia (Olds and Sibley 2003; Troelsen et al. 2003b). The G/A_{-22018} variants were detected to possess a modest silencer effect on the C/T_{-13910} enhancer activity but there were no differences where one of the variants G_{22018} or A_{-22018} were present (Troelsen et al. 2003b). These results suggest that only C/T_{-13910} has a functional significance for enhancing *LCT in vitro* (Olds and Sibley 2003; Troelsen et al. 2003b). Although, G/A_{-22018} was observed to closely associate with the adult-type hypolactasia trait (Enattah et al. 2002) it probably is in LD locating only 8 kb from C/T_{-13910}.

The hunt for the nuclear factor, which binds to the C/T_{-13910} enhancer, was started from nuclear extracts of HeLa and CaCo-2 cells. DNA affinity purification analyses using the DNA region containing the T_{-13910} variant demonstrated co-purification of Octamer-binding transcription factor-1 (Oct-1) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Super shift assays showed that Oct-1 binds directly to the T_{-13910} variant. Furthermore, the binding was detected to be stronger to the T_{-13910} variant than the C_{-13910} variant (Lewinsky et al. 2005). Earlier, GAPDH has been shown to directly interact with Oct-1 (Zheng et al. 2003) which could explain its observed co-purification with Oct-1 (Lewinsky et al. 2005). In addition, DNase footprint and supershift analyses identified binding motifs for Cdx-2, GATA-6, Fox and HNF-4α nearby C/T_{-13910} (Figure 9). Mutation analyses of the Oct-1 binding site of T_{-13910} enhancer led a 5-fold decrease of expression using LCT promoter-reporter gene construct in Caco-2 cells. Also, mutation analyses of GATA, Fox and HNF-4α sites remarkably diminished the T_{-13910} enhancer activity suggesting they are important factors belonging to the enhancer structure; in contrast mutations in the Cdx-2 site had no significant effect on enhancer activity. Co-transfection analyses of Oct-1, GATA-6, HNF-4α, Cdx-2 and HNF-1α and HNF-1β with the T_{-13910} enhancer uncovered the fact that Oct-1 triggers reporter gene expression (133-fold) only with HNF-1α. The HNF-1α motif has not been found within the C/T_{-13910}

enhancer but it probably acts via a proximal promoter, which has a fine-characterized HNF-1α binding site. Co-transfection of Oct-1 and HNF-1α using the C_{-13910} enhancer led to a 112-fold increase in enhancer activity. Over-expression of Oct-1 could not alone initiate the T_{-13910} enhancer suggesting Oct-1 is not the limiting factor in Caco-2 cells. Over-expression of HNF-1α, GATA-6, HNF-4α and Cdx-2 was observed to induce the C/T_{-13910} enhancer activity but only HNF-1α was observed to maintain the difference in activity between the T_{-13910} and C_{-13910} enhancers. In addition, the C/T_{-13910} enhancer had no influence on MCM6 promoter activity. The results suggest that the C/T_{-13910} variant locates in the middle of a sophisticated enhancer structure consisting of at least Oct-1, GATA, Fox and HNF-4α binding sites (Figure 9). Oct-1 with C/T_{-13910} directs specifically the LCT promoter activity and enhances reporter gene expression with the T_{-13910} variant significantly over the C_{-13910} variant. These *in vitro* observations offer a convincing model for T_{-13910}-induced lactase persistence in humans (Lewinsky et al. 2005).

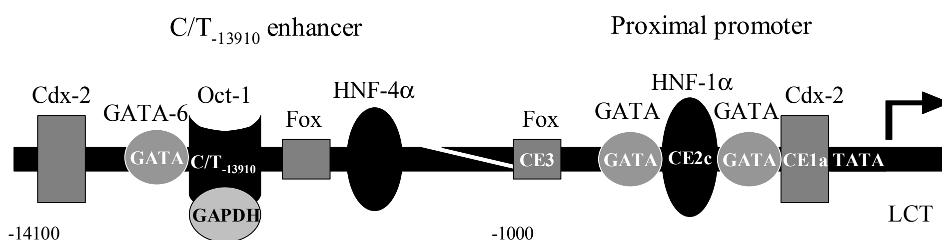


Figure 9. Structure of the C/T_{-13910} enhancer and the proximal promoter. Analysis of region of the C/T_{-13910} enhancer revealed the binding sites for Cdx-2, GATA-6, Oct-1, Fox and HNF-4α from 14045-13825 bp upstream of the *LCT* gene. Oct-1 interacts directly with the C/T_{-13910} site and has a stronger affinity to the T_{-13910} than the C_{-13910} variant. Oct-1 was observed to interact with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) that possibly functions as a co-activator for Oct-1. Mutation analyses of the binding sites demonstrated that the GATA-6, Fox and HNF-4α sites belong to the enhancer structure together with the Oct-1 site eliminating the enhancer activity. Disrupt of the Cdx-2 binding site had no effect on the enhancer activity. In expression studies of the C/T_{-13910} enhancer and the proximal LCT promoter construct, Oct-1 was observed to enhance reporter gene expression only when co-expressed with HNF-1α. This is assumed to occur through interactions with Oct-1 and HNF-1α that is bound to the CE2c site at the proximal promoter (Lewinsky et al. 2005).

2.4.3.4 T_{-13910}, the only variant ruling lactase persistence for humanity?

Originally, the T_{-13910} variant was cloned in the Finnish population, which is a population that has been proven to be invaluable for tracing mutations underlying rare monogenic diseases (Enattah et al. 2002). Interestingly, the variant has also been found in other European Caucasians or their descendants. The prevalence of the T_{-13910} allele has been observed to be in agreement with the epidemiological observations of lactase persistence in Austria, (Obermayer-Pietsch et al. 2004; Bodlaj et al. 2006) Germany, (Büning et al. 2003) France, US, (Enattah et al. 2002) UK, (Rasinperä et al. 2005) Ireland (Mulcare et al. 2004) Poland and Sweden (Study IV). C/T_{-13910} has also been shown to match, with few exceptions, with sets of individuals whose ability to hydrolyse lactose was determined by duodenal enzyme assays in the UK (Poulter et al. 2003) and by LTT in Sweden (Nilsson and Johansson 2004; Ridefelt and Håkansson 2005). In Austria, BHT diagnosed lactase status correlated with the C/C_{-13910} genotype, however the test indicated lactase deficiency in 14% of individuals with C/T_{-13910} and T/T_{-13910} (Hogenauer et al. 2005). The results are acceptable if it is taken into account that the accuracy of BHT is not very high (Arola 1994). Outside of Europe the T_{-13910} allele was strongly represented in Pakistan and Algeria, lower frequencies were observed in Middle Eastern populations (Bersaglieri et al. 2004). In Pakistan, five out of eight population groups had the T_{-13910} allele frequency $\geq$0.30 (Bersaglieri et al. 2004) that is in scale with the phenotyped data (0.20-0.33) in this region (Abbas and Ahmad 1983; Ahmad and Flatz 1984). There is no published data of lactase persistence available for Algerians. The T_{-13910} allele was suggested to have been spread between populations along dairy farming thus, increasing traces of T_{-13910} is seen in neighbouring European populations (Bersaglieri et al. 2004). There are certain nomadic populations with known pastoralism and high lactase persistence in Africa and the Arabian Peninsula. To shed some light on lactase persistence in Africans, different populations were analysed for the T_{-13910} variant. Statistically, T_{-13910} could

not match with five out of seven population groups with the published lactase persistence incidences. Four of those five populations came from Sudan one came from Senegal. On the contrary, in two population groups from Cameroon T_{-13910} associated with the epidemiological observations. Based on these results, these five groups may have a divergent genetic background for lactase persistence. A variant that could explain lactase persistence in these groups was not identified (Mulcare et al. 2004). The study has its limitations including the small number of analysed subjects and carefully characterized phenotypes are required to confirm the results. As there may be other variants apart from the T_{-13910} one influencing lactase persistence, at least in Africans, it will be interesting to see whether any new variant found shares the same enhancer element or if its mechanism of up-regulation is more complex. Also, it is to be determined what variant nomadic populations of Arabia possess to allow lactase expression in adulthood. Be that as it may, the case is not closed yet. There is still work to be done to uncover causes or to confirm the present one for lactase persistence globally. Thus far, the T_{-13910} variant is the only one that has been published to be causative for the lactase persistence phenotype.

There have been discussions of the diagnostic value of the C/T_{-13910} variant for adult-type hypolactasia (Swallow 2006). Based on studies in children (Rasinperä et al. 2004) and the population and functional studies reviewed here, C/T_{-13910} can be regarded as a suitable first-stage screening test for adult-type hypolactasia in adults of European ancestry.

2.5 Lactase persistence, non-persistence and dairy consumption risk factors for other diseases?

Milk contains a large quantity of calcium, lactose and many other components such as vitamin D and various vitamins of B. Milk is a versatile nutrient, and its consumption and effects on health have been studied with a passion. Obviously high

lactase activity facilitates the use of milk as nutrition and this has been noted as an increasing milk consumption in western countries (Obermayer-Pietsch et al. 2004; Rasinperä et al. 2004; Rasinperä et al. 2006). Hence, lactase persistence or non-persistence has been suggested to influence indirectly the pathogenesis of certain diseases. Perhaps one of the most studied is the link between the milk consumption and the development of osteoporosis. Calcium is the main mineral involved in bone formation and it is needed throughout life. The recommended daily intake of calcium is about 900 mg/day for adults, 1200 mg/day for adolescents and the elderly (National Nutrition Council 1999). As milk is an excellent source of bioavailable calcium, it has a positive effect on calcium balance and therefore on bone formation. Especially, the consumption of dairy products in childhood and adolescence is seen as an increasing bone density in adulthood and an ability to protect women against osteoporosis (Gueguen and Pointillart 2000). In addition, the C/C_{-13910} genotype defining lactase non-persistence associated with bone fractures in elderly people in Austria and Finland (Obermayer-Pietsch et al. 2004; Enattah et al. 2005b) but in young Finnish men the genotype had no effect (Enattah et al. 2004a). With this background lactase persistence could help maintain the milk drinking habit from childhood to adolescence and hence decrease the risk of osteoporosis. However, studies focussed on this issue are controversial and there is not consensus among researchers. Even, a case control study using genetically defined lactase status by the C/T_{-13910} variant could not find any link between lactase non-persistence and osteoporosis in Finnish postmenopausal women (Enattah et al. 2005a). Lactose has also been suggested to enhance calcium absorption in the intestine but studies on subjects with lactase persistence and non-persistence have shown no real proof of this. In general, it has been rationalized that diminished calcium availability in non-persistence subjects might explain the more efficient calcium absorption, reviewed in Gueguen and Pointillart (2000), Prentice (2004). Osteoporosis is a complex disease, which is caused by genetic and environmental factors. It seems that dietary calcium for its part during the peak of bone formation reduces bone fracture risk and

osteoporosis in adulthood. Yet, the effect of lactase non-persistence is more arguable (Gueguen and Pointillart 2000; Prentice 2004). The composition of Finnish 1.5% milk is represented in Table 2 (Fineli, National Public Health Institute, Finland).

Table 2. Milk (1.5%) content

Milk component	Quantity	Milk component	Quantity
Energy	191 kJ	Iodine	16.0 µg
Fat	1.5 g	Selenium	2.8 µg
Protein	3.0 g	Vitamine A	12.9 µg
Lactose	4.8 g	Vitamine D	0.5 µg
Cholesterol	6.4 mg	Vitamine E	<0.1 mg
Sodium	41.0 mg	Vitamine K	0.33 µg
NaCl	104.5 mg	Vitamine C	1.1 mg
Potassium	150.0 mg	Folate	4.2 µg
Magnesium	11.0 mg	Niasine	0.8 mg
Calcium	120.0 mg	Vitamine B2	0.19 mg
Phosphorus	90.0 mg	Vitamine B1	0.04 mg
Iron	< 0.1 mg	Vitamine B12	0.4 µg
Zinc	0.4 mg	Carotenoid	8.7 µg

Milk consumption has been proposed to have a modest protective effect against colorectal cancer (World Cancer Research Fund 1997). Primarily, the protective effect has been suggested to be mediated by calcium and vitamin D. However, only large cohort studies have been able to confirm the relation, most case control studies have failed to find supportive evidences for the hypothesis (Norat and Riboli 2003). Lactase persistence variant, T_{-13910}, was found to have a modest protective effect against colorectal cancer in the Finnish population. However, there was no significant risk effect observed in the British or Spanish populations (Rasinperä et al. 2005). This might be related to the fact that Finnish milk consumption is the highest in the world (The World Dairy Situation 2003). The findings require more detailed studies concerning the possible mechanism between components of milk and colorectal carcinogenesis. Therefore, another association study would not seem to be the most reasonable alternative.

Lactase non-persistence has been suggested to associate also with Crohn's disease (Mishkin et al. 1997; von Tirpitz et al. 2002) and lactase persistence with type I and II diabetes (Meloni et al. 2001). Association analysis of C/T_{-13910} with patients with Crohn's disease (Büning et al. 2003) and type I and II diabetes (Enattah et al. 2004b) did not support those earlier findings. These and other's findings (Park et al. 1990) suggest that the hypotheses might be closer to a myth than a real cause and effect relation.

2.5.1 Ovarian carcinoma

Ovarian cancer is the sixth most common cause of cancer among women but it is one of the deathliest disease of the female genital tract in the world (Parkin et al. 2001; Schulz et al. 2004). Ovarian cancer occurrence rates are observed to be highest in Europe and North America and lowest in Asia (World Cancer Research Fund 1997). Epithelial ovarian cancer (ovarian carcinoma) consists of 90% of ovarian cancers; germ cell and stromal cell ovarian cancers are responsible for 10% of cases. Ovarian carcinoma can be divided into four major histological subtypes: serous, endometrioid, mucinous and clear cells tumours of which the serous subtype is the most common one consisting of about 40% of ovarian carcinoma tumours. Papillary serous, Brenner cell, undifferentiated adenocarcinomas and sarcomas represent a minority of the histological subtypes. The majority of ovarian cancer patients are postmenopausal women in their late 50's, reviewed by Schulz et al. (2004). About 90% of the cases of all types of ovarian cancers are sporadic and hereditary factors are proposed to describe 5-10% of cases (Runnebaum and Stickeler 2001; Schulz et al. 2004). The aetiology of ovarian carcinoma is poorly known but both genetic and environmental factors are involved. For example, in sporadic ovarian carcinomas inactivation of the *p53* tumour suppressor gene has been observed (Wen et al. 1999; Christie and Oehler 2006). The *BRCA1* and *BRCA2* genes are occasionally mutated in hereditary ovarian carcinomas (Menkiszak et al.

2003; Christie and Oehler 2006). Numerous "incessant" ovulation and hormonal factors have also been hypothesized to be involved in its pathogenesis (Risch 1998). Thus, factors that reduce ovulation cycles such as multiparity, lactation, tubal ligation and hysterectomy or use of oral contraceptives are associated with a decreased risk of ovarian carcinoma (Whittemore et al. 1992; Hankinson et al. 1993). Also, nutritional features have been suggested as contributory factors (Haenszel and Kurihara 1968; Schulz et al. 2004).

2.5.1.1 Dairy consumption and ovarian carcinoma risk

Dairy products have been hypothesized to play a role in the tumour genesis of ovarian carcinoma. Milk, as known, has a high content of lactose and it is further metabolized by lactase to glucose and galactose. Galactose is converted to glucose in the liver by the Leloir pathway (Xu et al. 1989; Reichardt and Woo 1991). The hypothesis was born in the 1980's when animal studies showed that high quantities of dietary galactose is poisonous to oocytes (Chen et al. 1981; Schwartz and Storozuk 1988). In addition, young women with galactosemia were observed to develop ovarian failure (Kaufman et al. 1981). Galactosemia is an autosomal recessive disorder with an almost total absence of galactose-1-phosphate uridyltransferase (GALT) activity leading metabolites of galactose to accumulate in the ovary and other tissues (Xu et al. 1989; Reichardt and Woo 1991). Hence, the premature ovarian failure was suggested to be due to the direct toxicity of galactose, its metabolites or elevated gonadotropin levels (Cramer and Welch 1983; Harlow et al. 1991; Leslie 2003). The hypothesis truly got off the ground when milk consumption and lactase persistence were observed to correlate with ovarian carcinoma incidence (Cramer 1989). Since, then a number of studies have been conducted to define the significance of milk or its components and/or lactase persistence/non-persistence to ovarian carcinogenesis. Once again the results are controversial. Lactase persistence and lactose consumption were shown to associate

with increased ovarian carcinoma risk in Sardinia, Italy (Meloni et al. 1999) and lactose intolerance was suggested to indicate a protective effect against ovarian carcinoma at an early age in eastern U.S. (Cramer et al. 2000) but in studies with subjects from Canada (Risch et al. 1994) and with Caucasian from the U.S. (Herrinton et al. 1995) no such association was found. Diagnosis of lactase phenotypes has been often obtained using LTTs and self-reports of clinical symptoms that naturally increase the distribution. In multiple studies, milk or lactose consumption was found to have a protective effect on ovarian carcinoma (Goodman et al. 2002a; Goodman et al. 2002b; Salazar-Martinez et al. 2002; Yen et al. 2003). Two cohort studies from the U.S. (Kushi et al. 1999; Fairfield et al. 2004) and one from Sweden (Larsson et al. 2004) found a correlation between lactose consumption and increased ovarian carcinoma risk, especially that of serous subtype (Fairfield et al. 2004; Larsson et al. 2004). In contrast, cohorts from the Netherlands (Mommers et al. 2006) and another from the U.S. (Kiani et al. 2006) did not show any association whatsoever. Several case control studies have failed to find a clear relation of milk or lactose consumption to ovarian tumour genesis (Mettlin and Piver 1990; Engle et al. 1991; Risch et al. 1994; Herrinton et al. 1995; Webb et al. 1998; Britton et al. 2000; Bertone et al. 2001; Pan et al. 2004; Parazzini et al. 2004). Meta-analyses showed that only cohort studies found a link between milk/dairy/lactose consumption and an increased risk for ovarian carcinogenesis, but not case control studies (Larsson et al. 2006), nor the studies combined (Qin et al. 2005; Larsson et al. 2006). However, a recent pooled analysis of 12 cohort studies showed no statistically significant association between lactose consumption and the risk of ovarian carcinoma (Genkinger et al. 2006). Dietary vitamin D has been shown to protect from the development of ovarian carcinoma (Bidoli et al. 2001; Salazar-Martinez et al. 2002). In addition, calcium consumption has been observed to decrease ovarian carcinoma risk (Bidoli et al. 2001; Goodman et al. 2002b). Calcium, dietary vitamin D and endogenous vitamin D formation has been shown to downregulate the synthesis of parathyroid hormone (PTH) (Takeuchi et al. 1995;

Malabanan et al. 1998). PTH is involved in dietary calcium absorption increasing 1,25 dihydroxyvitamin D synthesis. It has been suggested that PTH is involved in cancer stimulating IGF-1 formation or acting as a co-mitogen in preneoplastic lesion expressing PTH receptors (McCarty 2000). However, all studies have not supported the potential protective effect of vitamin D and/or calcium (Tzonou et al. 1993; Kushi et al. 1999; Goodman et al. 2002b; Salazar-Martinez et al. 2002; Pan et al. 2004). The results about the dairy consumption as a risk factor for ovarian carcinogenesis are contentious. Milk consists of a huge number of various components and there is no clear consensus of the effect of any of them. However, it must be taken into account and they are difficult analyse individually.

The influence of GALT activity as directly involving the galactose hypothesis has been studied in the development of ovarian carcinoma. Some studies have used the N314D polymorphism of GALT as a molecular marker as it has been shown to diminish the GALT activity to 74-43% from normal (Elsas et al. 1994; Langley et al. 1997). Decreased activity of GALT together with lactose consumption was detected to associate with ovarian carcinoma (Cramer et al. 1989). However, later studies could not find any significant difference in GALT activity (Herrinton et al. 1995; Webb et al. 1998; Cramer et al. 2000; Goodman et al. 2002a) or in the frequency of N314D (Goodman et al. 2002a) between ovarian carcinoma patients and controls. The frequency of N314D was found to be higher in women with serous ovarian carcinoma but not endometroid or clear cell subtypes (Morland et al. 1998). On the contrary, another study found that N314D associated with endometroid and clear cell ovarian carcinomas but not with the serous subtype (Cramer et al. 2000). In addition, the N314D polymorphism and lactose consumption associated with borderline ovarian carcinoma (Cozen et al. 2002).

2.6 Identification of disease genes

2.6.1 The Human Genome Project

The Human Genome Project (HGP) had the ambitious goal to determine the structure of the human DNA sequence. The HGP was initiated in various National Institutes of Health and the U.S. Department of Energy in the United States in 1990. In 1998 a private company, Celera Genomics, announced a three-year project aimed at revealing the human genome nucleotide sequence. In February 2001 they and the public International Human Genome Sequencing Consortium published a nearly complete sequence of the euchromatic portion of the human genome (Lander et al. 2001; Venter et al. 2001).

2.6.2 The Human Genome

The genetic map length of the human genome (unit of Morgan) between two loci defines the average number of genetic recombinations (crossing-over) occurring in a gamete during meiosis. The recombination frequency fluctuates with chromosomal regions and it is different for males and females. The genetic length of the whole genome is 4301 cM in females and 2849 cM in males (Collins et al. 1996). One centiMorgan is estimated to be roughly equal to a physical length of 1 Mb (Ott 1999).

The physical length of the euchromatic portion of the haploid human genome is ~2.91 billion bp. Sequence analysis showed circa 24500-26000 protein encoding transcripts and 6500-12000 additional genes obtained by gene prediction programs with weak similarities to other species. Exons cover only 1.1-1.5% of the genome, 24-33% is covered by introns, and 66-75% of the genome consists of intragenic DNA without any known function. The organization of the human genome primary

structure provides scientists with a tool to understand human biology (Lander et al. 2001; Venter et al. 2001).

2.6.3 Positional cloning

In positional cloning a disease gene is isolated based on its chromosomal location without any knowledge of its function or pathogenesis. A chromosomal location is defined by genetic mapping using family material adequate for linkage analyses. Locus can be further refined by LD and haplotype mapping (Collins 1992; Collins 1995). Today in the post genomic era, genes of interest are listed in databases, thus regional candidate genes can be carefully chosen and/or systematically screened for mutation.

2.6.4 Genetic mapping

Genetic mapping requires family material with a precisely diagnosed inherited disease. Only reliable diagnosis by experienced clinician(s) can lead to useful results using the genetic mapping approach. For linkage analysis two or more affected individuals are required. The collected family material is analysed with polymorphic microsatellite markers positioned at about 4-10 cM intervals throughout the genome. Until linkage is reached, the study can be further carried on by fine mapping with more frequent microsatellite maps to get as close as possible to the mutation responsible for the disease. Botstein et al. (1980) introduced the principles of genetic mapping with DNA markers in 1980 and Gusella et al. (1983) were the first to identify a human disease gene locus (Huntington disease) using this approach in 1983.

2.6.4.1 Genetic markers

There are a large number of sequence variations in the human genome between individuals. Most of them are located in the non-coding regions but such harmless alterations in the coding regions are not unusual. The segregation of these variations to offspring within families is the foundation for genetic mapping. The sequence variations are called polymorphisms when its most common allele has a population frequency of less than 95-99% (Botstein et al. 1980; Ott 1999). The usefulness of a marker for linkage analysis depends on the number of alleles and their gene frequencies (degrees of polymorphism). Short tandem repeats (STR) are highly polymorphic usually consisting of 10-50 copies of di-, tri-, tetra-, or pentanucleotide repeat elements. The most common repeat $(CA)_n$ is distributed approximately in 30 kb intervals in the human genome (Hearne et al. 1992). Today, thousands of microsatellite markers and several genetic maps are available in public databases (e.g. Marshfield Medical Genetic Foundation, Genome Database). High-resolution recombination maps have also been created and they are available as a supplementary file (Kong et al. 2002). The use of single nucleotide polymorphisms (SNPs) as biallelic markers have led to seriously new opportunities again not seen since the times of biallelic restriction fragment length polymorphism (RFLPs) detection. SNPs are observed about once in every kilobase, they are also found in the coding regions of the genes (cSNP). They are more stable than STRs but at the same time carry less information (Li and Sadler 1991; Wang et al. 1998a). SNPs can be easily and automatically identified using primer extension techniques and detected by scintillation, fluorometry or mass spectrometry based methods. They are suitable for genetic mapping and gene expression studies using cSNP as a marker (Syvänen et al. 1990; Syvänen et al. 1993; Lander 1996; Pastinen et al. 1996; Braun et al. 1997; Haff and Smirnov 1997; Pastinen et al. 1997; Ross et al. 1998; Syvänen 1998; Tang et al. 1999). Wang DG and colleagues (1998) published the first genetic SNP map in 1998.

2.6.4.2 Linkage analysis

The purpose of linkage analysis is to assign a disease locus its specific chromosomal region. It is based on comparing the inheritance of alleles of polymorphic markers with the disease phenotype. If a polymorphic marker is so close to a disease locus that they cannot be separated by recombination (or it is extremely rare) they are said to be genetically linked. The techniques of the statistical significance of linkage testing are based on maximum likelihood estimation and likelihood ratio testing. The most used approach is the logarithm of odds (LOD) score method. The likelihood (L) that two loci, e.g. a marker and a disease locus, are linked with a recombination fraction (θ) is compared the likelihood that they are not linked (θ=½). The LOD score Z(θ) is obtained when the ratio of these likelihoods are converted into $\log_{10}$,

$$Z(\theta) = \log_{10}\left[\frac{L(\theta)}{L(½)}\right].$$

A LOD score value ≥3, which means that two loci are 1000 times more likely to be linked than not linked, is considered to be significant evidence for linkage. On the other hand, a LOD score value ≤-2 is evidence against linkage (probably a 100 times likely not to be linked) (Morton 1955; Ott 1999). Linkage analyses are performed by comparing a disease locus with each marker locus as a two-point linkage analysis and analysis of numerous linked loci, multipoint analysis. Multipoint analysis uses the information of adjacent marker loci thus, the order of loci matters.

2.6.4.3 Linkage disequilibrium

Linkage disequilibrium is a phenomenon of non-random association of alleles at linked loci. The association can be statistically measured by the traditional chi-square (χ^2) test, the Fisher's exact test, (Armitage and Berry 1987) r^2, (Ardlie et al.

2002) and the likelihood based lambda (λ) test (Terwilliger 1995). The chi-square and a Fisher's exact tests evaluate independence of a 2x2 table. The chi-square test is adjusted for large sample sets.

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})}{E_{ij}},$$

where O_{ij} is the observed frequency of the *i*th row and the *j*th column, E is the expected frequency of the *i*th row and the *j*th column, R is a number of rows and C is a number of columns (Armitage and Berry 1987). Lambda informs about the proportion of excess of a certain allele in disease chromosomes, relative to its population frequency.

$$\lambda_i = \alpha(1-\theta)^n ,$$

Where λ_i is proportion of excess of allele *i* in chromosomes carrying the disease allele, α_i is proportion of disease alleles originally in association with allele *i*, θ is recombination fraction, n is number of generation since introduction of the founder disease allele into the population (Terwilliger 1995).

LD is used for fine mapping of disease genes and it is powerful in situations where chromosomes are carrying a mutation originating from one common ancestor. An efficient LD mapping also requires a low number of the original settlers, rapid expansion of the population without major immigration like the Finns or an ancient small population which has been stable for a long time like the Saami, and a large enough number of patients (Edwards 1980; de la Chapelle 1993; de la Chapelle and Wright 1998; Terwilliger et al. 1998; Peltonen et al. 1999). The power of LD mapping was showed in a rare Finnish monogenic disease infantile-onset

spinocerebellar ataxia (IOSCA) where initial gene localization was obtained using four affected individuals from two consanguineous pedigrees (Nikali et al. 1995). LD mapping has also been successfully used in narrowing the disease locus in other rare Finnish diseases such as diastrophic dysplasia (DTD) (Hästbacka et al. 1992) or autoimmune polyendocrinopathy-candidosis-ectodermal dystrophy (APECED) (Aaltonen et al. 1994). The length of LD can be utilized to estimate the age of mutations. Meiotic recombinations which may have occurred over time decrease the length of the observed LD generation after generation thus, indicating the older mutation (Varilo 1999). Figure 10 represents the principle used for LD mapping.

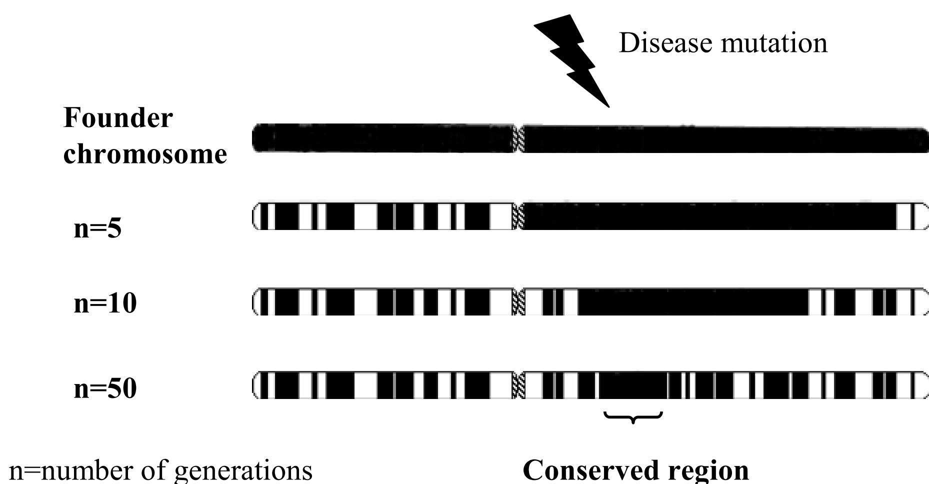


Figure 10. The basis of linkage disequilibrium (LD) mapping. The conserved region of a founder chromosome, exposed to disease mutation, shorten after generations due to recombinations in meiosis. In the course of time, only a short chromosomal segment is preserved from the founder chromosome. This event is detected as LD and can be used in a similar manner in haplotype analyses and for evaluating the age of mutation (de la Chapelle et al. 1994; Ott 1999; Varilo 1999; Norio 2000).

2.6.4.4 Haplotype analysis

Haplotype analysis, like LD mapping, is used for further restriction of disease locus in the isolated populations. Haplotype analysis is the identification of shared ancestral disease alleles of informative marker loci in a critical region. Recombinations events can dramatically reduce the shared region. Haplotype analysis extends beyond recombinations events by revealing old recombinations and thus further restricts the critical region needed to identify a disease gene, exemplified in many Finnish diseases such as progressive myoclonus epilepsy (EPM1) from about 7 cM to 0.30 cM (Lehesjoki et al. 1993).

2.6.4.5 Towards causative mutations

When a disease locus has been defined to a certain chromosomal region by genetic mapping, the virtually completed human genome project gives a possibility to use bioinformatics to locate ESTs, protein homologies, known genes or predicted genes. Also, genome projects of other organisms help to identify conserved gene structures and regulatory elements. Modern capillary techniques enable a fast mutation screening by cycle sequencing. The validation of found sequence variations can be performed by carrier screening, functional analyses *in vitro* and *in vivo*, and model organisms.

2.7 Gene expression studies

A fundamental part of disease gene characterization, as well as the basic gene studies, is to examine gene expression. Gene expression can be studied by a number of techniques for example, histochemically and immunohistochemically evaluations at the protein level. Cell specific gene expression can be detected using *in situ* hybridization at the mRNA level (Bleeker et al. 2000; Ropponen et al. 2001; van

Noorden 2002). Furthermore, more accurate sensitive expression detection techniques have been developed.

2.7.1 Techniques for mRNA detection

The monitoring of mRNAs has been found to be an important approach to studying gene function. Alwine et al. (1977) modified the southern blot technique (Southern 1975) and created the specific northern blot hybridization assay for the detection of mRNA. Northern blotting requires electrophoretic separation of total RNA or mRNA on an agarose gel, transferring to a nitrocellulose sheet, and hybridization with a ^{32}P-labeled DNA probe. The results are visualized by autoradiography.

Since Mullis et al's (1986) development of the PCR method, RT-PCR has been utilized efficiently to study the expression of genes. RT-PCR is a sensitive technique that enables the quantitative monitoring of changes of mRNA expression at the cellular level. Real-time PCR offers the opportunity to follow the accumulation of PCR products during amplification. It is based on specific hybridization of the fluorescent probes. Quantitation can be obtained using internal standard competitively during amplification and detection is based on decrease in fluorescein quenching by rhodamine after Taq DNA polymerase's exonuclease cleavage of dual-labelled probes during the exponential phase of amplification (Livak et al. 1995), see also (Holland et al. 1991). Alternatively, fluorescence can be monitored by the resonance energy transfer of fluorescein to Cy5 between adjacently hybridized probes (Wittwer et al. 1997).

A number of competitive endpoint RT-PCR based techniques have been developed. The quantitation is achieved by coamplification of *in vitro*-generated internal standard RNA with target mRNA. For example, in the early quantitation methods the PCR products of the internal standard and target mRNA had to differ in length (Wang et al. 1989) or the size difference were generated by restriction endonuclease

digestion (Becker-Andre and Hahlbrock 1989). The detection needed electrophoretic separation followed by northern blot hybridisation. The initial copy number of the target mRNA is estimated from the ratio to the internal standard. However, more sensitive methods of endpoint quantitation have been introduced. The versatile enzyme-linked immunosorbent (ELISA) PCR assay relies on the hybridization of a solid-support captured amplification product with a digoxigenin-labeled internal oligonucleotide probe. Detection is performed by optical reading after an anti-digoxigenin alkaline phosphatase coupled antibody treatment (Alard et al. 1993). Solid-phase minisequencing enables quantitation of highly homologous sequences that differ from each other only by one single nucleotide (Figure 11). It utilizes DNA polymerase to a single nucleotide primer extension. PCR is performed by a biotinylated primer which is immobilized to a streptavidine coated solid support and denaturated. A detection primer is designed to anneal adjacent to the single nucleotide polymorphism and minisequencing is performed separately with the expected ^{3}H-labeled dNTPs. The ratio of two incorporated labels as a measure of the two sequences is determined by a liquid scintillation counter (Syvänen et al. 1990; Syvänen et al. 1992a; Syvänen et al. 1992b). Solid-phase minisequencing has been used for quantitation of rare transcripts (Ikonen et al. 1991) and e.g. the detection method using fluorescent labels in multiplex analyses (Pastinen et al. 1996; Pastinen et al. 1997). Furthermore, solid-phase minisequencing was used for accurate determination of the relative mRNA levels of two closely related mRNAs (squamous cell carcinoma antigen; SCCA1 and 2). Quantitation was based on the ratio of the SCCA2 to SCCA1 genes in the same cell as internal standard for each other (Stenman et al. 1999), see also (Karttunen et al. 1996). Multiple genes or whole genomes can be simultaneously studied by the complementary DNA microarray technology. DNA microarray expression analysis involves reverse transcription (RT) of isolated mRNA with fluorescently labelled dNTPs. A mixture of equivalent amounts of both labelled cDNA sample are hybridized to the

microarray and the ratio of labels are measured by laser confocal fluorescent scanning (Schena et al. 1995; Lockhart et al. 1996; Brown and Botstein 1999).

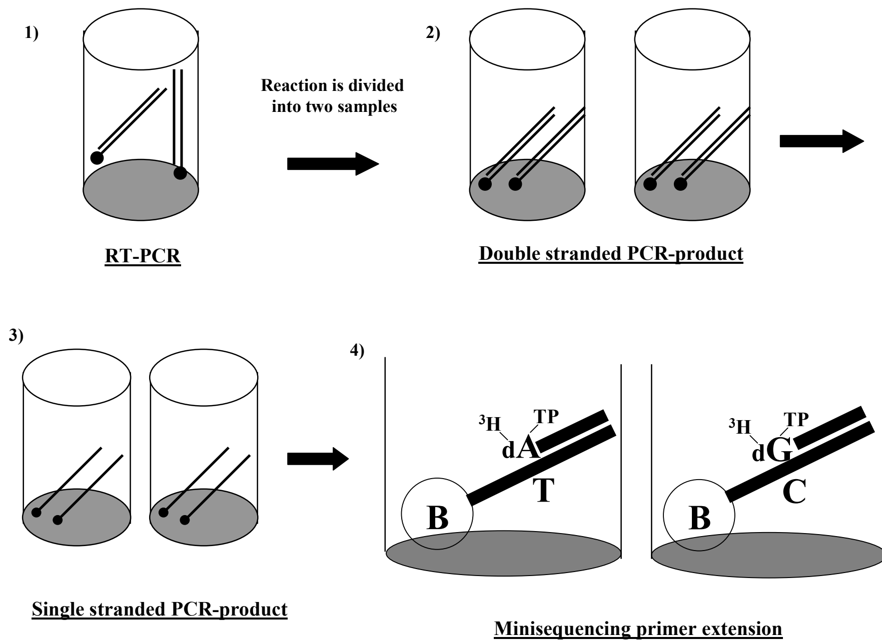


Figure 11. The principle of the solid-phase minisequencing. The initiation step is PCR or RT-PCR with a biotinylated primer. 2. Then dividing the reaction into two samples and immobilizing streptavidine-coated microplates. 3. Denaturation of the complementary strand. 4. A primer-guided tritium labelled nucleotide incorporation assay. As a result, the ratios of two nucleotides as a measure of two alleles are calculated (Syvänen et al. 1990; Syvänen et al. 1992a; Syvänen et al. 1992b). The Figure is modified from Study III.

3 AIMS OF THE STUDY

Aims of the present study were:

1. To identify causative mutations for congenital lactase deficiency (I).

2. To determine the functional consequences of the C/T$_{-13910}$ variant associated with adult-type hypolactasia both in children (II) and in adults (III).

4. To study a potential relationship between lactase persistence and ovarian carcinoma (IV).

4 MATERIALS AND METHODS

4.1 Study subjects

4.1.1 Finnish CLD families and controls (I)

Initially, family material consisted of 19 Finnish families, 27 patients and 64 healthy family members (Järvelä et al. 1998) at the end of the study a total number of 32 patients, 76 healthy family members from 24 families were included. The CLD diagnoses were based on clinical symptoms as described elsewhere (Savilahti et al. 1983). The carrier frequencies of the CLD mutations were determined from 556 anonymous blood donors obtained from the Finnish Red Cross Blood Transfusion Service.

4.1.2 *LCT* mRNA study in children (II)

Fifteen Finnish children and adolescents aged 10 months to 23 years from the Hospital for Children and Adolescents were included in the study.

4.1.3 *LCT* mRNA study in adults (III)

A total of 52 patients having unspecific abdominal complaints from Herttoniemi Hospital participated in the study. The diagnosis for adult-type hypolactasia was based on lactase and sucrase activities. An L/S ratio <0.30 indicated hypolactasia. Lactase, sucrase and maltase activities were determined from a duodenal biopsy specimen based on (Dahlqvist 1984). To exclude secondary lactase deficiencies all biopsies underwent histological examinations. Enzyme activities were determined at the Laboratory of the Hospital for Children and Adolescents.

4.1.4 Lactase persistence and ovarian carcinoma (IV)

The Finnish sample set consisted of 327 patients with ovarian carcinoma treated at the Department of Obstetrics and Gynaecology, Helsinki University Central Hospital throughout the years 1989-1998. The clinical diagnosis of ovarian carcinoma in the Finnish patients was made at the mean age of 56 years ± 14. The analysis of the Finnish control material was reported earlier (Enattah et al. 2002).

Polish samples consisted of 303 patients with ovarian carcinoma treated at the Department of Chemotherapy, Regional oncology Centre, Szczecin, during the years 2001-2004. The mean age of diagnosis was 54 (range 21-80) years. Control material consisted of 296 unselected adults.

The 152 Swedish samples were obtained from the southern Swedish health care region and were diagnosed with ovarian carcinoma during 1998-2000. The mean age of diagnosis was 59 (range 22-82) years. This patient material has previously been examined for germline mutations in *BRCA1* and *BRCA2*. The control samples were obtained from healthy individuals and from patients without known malignancy operated on at the Department of Orthopedics at the University Hospital in Lund.

4.1.5 Ethical considerations

The Ethics Committee of Helsinki University Hospital approved the lactase studies with a decision numbered 523/E7/2001. Written informed consent was obtained from the patients and controls. The ovarian carcinoma project has been approved by the Ethics Committee of the Department of Obstetrics and Gynecology, HUCH (040/95) and National Authority for Medicolegal Affairs (1129/32/300/03). As regards fresh tissue and blood samples informed consent has been obtained from each patient. The study was done in accordance with the Helsinki declaration.

Methods used in this study have been described in the original articles below and listed in Table 3.

Table 3. Methods used in the study.

Methods	Original publication
Bioinformatics used	I
DNA extraction with a strip method	III
RNA extraction	III
Polymerase chain reaction	III
Genotyping by Solid-phase minisequencing	IV
Quantitation by Solid-phase minisequencing	III
RT-PCR	III
Electrophoresis:	
Agarose gel electrophoresis	III
ABI377	III

4.1.6 Methods not described in the publications

4.1.6.1 Sequencing

The PCR products (~5 µl) were cycle sequenced in a total volume of 10 µl on both strands with 4 µl BigDye™ Terminator RR mix, (Applied Biosystems, Perkin Elmer, Foster City, CA, USA) and 3.5 pmol of primers. The sequencing reaction included a 30-second denaturation step at +96°C, a 15-second primer-annealing step at +55°C and a 4-minute extension step at +60°C. The sequencing reactions were cycled 25 times and stored at +4°C. Sequenced products were electrophoresed on an ABI377 or ABI3730 DNA Sequencher (Perkin Elmer).

4.1.6.2 Microsatellite genotyping

One of the two PCR primers of each reaction included a fluorescent label (6-FAM, TET, HEX) at its 5' end. PCR products were pooled and electrophoresed with TAMRA 500 size marker (Applera Corporation, Norwalk, CT, USA) on an ABI 377 (Applied Biosystems) using a 6% polyacrylamide gel. The GENESCAN 3.0 program together with GENOTYPER 2.0 (Applera Corporation) was used for determining the genotypes. The genotypes were validated by two individuals, and organized using LINKBASE 0.952 (Suomalainen 1998 unpublished). The pedigree inconsistencies were checked by PEDCHECK 1.1 (O'Connell and Weeks 1998).

4.1.6.3 Statistical analyses

Parametric two-point linkage analyses were performed with the recessive mode of inheritance assuming complete penetrance and 0.001 disease allele frequency. The analyses were carried out with the MLINK program of the linkage package (Lathrop and Lalouel 1984; Lathrop et al. 1984) using the ANALYZE 2.1 program (Hiekkalinna et al. 2005). Association analyses were carried out using the haplotype relative risk likelihood-ratio algorithm based test (HRR-LRT). In addition, LD was measured by a likelihood based lambda (λ) test which gives information about the proportion excess of a certain allele in disease chromosomes, relative to its population frequency (Terwilliger 1995). Haplotypes were constructed using the GENEHUNTER 2.1 program (Kruglyak et al. 1996) and the CLD founder haplotype was generated assuming that there were a minimum number of recombinations.

5 RESULTS AND DISCUSSION

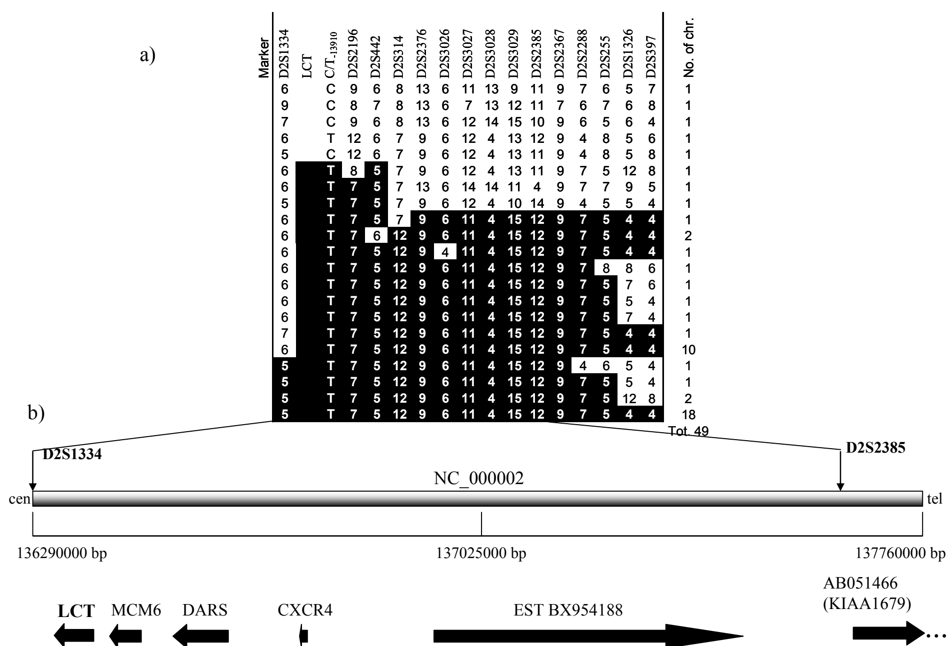
5.1 Results of congenital lactase deficiency (I)

Previously, the CLD locus was assigned by linkage analysis on 2q21. LD and haplotype analysis appeared to restrict the critical locus between microsatellites D2S314 and D2S2385 (Järvelä et al. 1998). Hence, the locus excluded *LCT* as a causative gene. The observation was supported by the sequence analysis of the coding region of *LCT* that not succeeded in finding disease mutations in a Finnish CLD patient (Poggi and Sebastio 1991). However, sequence analyses of transcripts between D2S314 and D2S2385 could not identify disease-causing mutations (data not shown). In this study, the CLD locus was further analysed using a denser microsatellite map. In addition, five novel CLD families were included in the analyses.

5.1.1 Linkage, linkage disequilibrium and haplotype analyses

Fifteen microsatellite markers and C/T_{-13910} were genotyped covering 5.88 cM on 2q21-2q22 in 24 families with a total of 32 affected children. Twenty-one different haplotypes were detected in 48 disease chromosomes (Figure 12 a). The highest parametric two-point LOD score of 7.24, with a θ value of 0, was achieved with marker D2S2385. Strong evidence for linkage with LOD scores $\geq$2.95 was obtained with all the included markers. LD was evaluated by calculation of the allelic association using HRR-LRT and lambda (λ) methods. The HRR-method gave statistically the most significant p-values of 1×10^{-10} and 9×10^{-10} with markers D2S314 and D2S2385, respectively. A significant LD was observed with all markers except D2S1334 and D2S2288, of which p-values were higher than 0.01. In contrast, a λ-value 0.84 gave the strongest evidence for LD with marker D2S3026 between the previous markers, decreasing to the lowest value of 0.30 at marker

D2S1334. The founder haplotype, cen-5-T_{-13910}-7-5-12-9-6-11-4-15-12-9-7-5-4-4-tel, was present in 18 disease chromosomes from which 15 of the remaining 20 different haplotypes have been derived (Figure 12 b). Five disease chromosomes each of which had a diverged haplotype and could not linked to the major haplotype. Haplotypes was detected to share a region that includes *LCT* and we carried on to perform sequence analysis of this gene. The haplotype and transcript map of the critical DNA region for CLD is shown in Figure 12.



Marker D2S1334	LCT	C/T_{-13910}	D2S2196	D2S442	D2S314	D2S2376	D2S3026	D2S3027	D2S3028	D2S3029	D2S2385	D2S2367	D2S2288	D2S255	D2S1326	D2S397	No. of chr.
6		C	9	6	8	13	6	11	13	9	11	9	7	6	5	7	1
9		C	8	7	8	13	6	7	13	12	11	7	6	7	6	8	1
7		C	9	6	8	13	6	12	14	15	10	9	6	5	6	4	1
6		T	12	6	7	9	6	12	4	13	12	9	4	8	5	6	1
5		C	12	6	7	9	6	12	4	13	11	9	4	8	5	8	1
6		T	8	5	7	9	6	12	4	13	11	9	7	5	12	8	1
6		T	7	5	7	13	6	14	14	11	4	9	7	7	9	5	1
5		T	7	5	7	9	6	12	4	10	14	9	4	5	5	4	1
6		T	7	5	7	9	6	11	4	15	12	9	7	5	4	4	1
6		T	7	6	12	9	6	11	4	15	12	9	7	5	4	4	2
6		T	7	5	12	9	4	11	4	15	12	9	7	5	4	4	1
6		T	7	5	12	9	6	11	4	15	12	9	7	8	8	6	1
6		T	7	5	12	9	6	11	4	15	12	9	7	5	7	6	1
6		T	7	5	12	9	6	11	4	15	12	9	7	5	5	4	1
6		T	7	5	12	9	6	11	4	15	12	9	7	5	7	4	1
7		T	7	5	12	9	6	11	4	15	12	9	7	5	4	4	1
6		T	7	5	12	9	6	11	4	15	12	9	7	5	4	4	10
5		T	7	5	12	9	6	11	4	15	12	9	4	6	5	4	1
5		T	7	5	12	9	6	11	4	15	12	9	7	5	5	4	1
5		T	7	5	12	9	6	11	4	15	12	9	7	5	12	8	2
5		T	7	5	12	9	6	11	4	15	12	9	7	5	4	4	18
																	Tot. 49

Figure 12. a) Haplotypes of 24 Finnish CLD families on 2q21. Haplotypes were constructed using 15 microsatellites and C/T_{-13910}. Haplotypes derived from one major haplotypes are highlighted.

b) Transcript map of the critical DNA region for CLD. Lactase (*LCT*), minichromosome maintenance deficient 6 (*MCM6*), aspartyl-tRNA synthetase (*DARS*), chemokine receptor type 4 (*CXCR4*) and KIAA1679 are the known genes in the region (indicated by horizontal arrows). The centromere is on the left.

5.1.2 CLD mutations in *LCT*

The *LCT* gene was sequenced in all 32 CLD patients in 24 families and a total of five different mutations were identified. Twenty-seven patients out of thirty-two were homozygous for a substitution of T→A at c.4170 that was designated as the $\text{Fin}_{\text{major}}$ mutation (Table 4, Figure 13). All their parents analysed (five were not available) and five patients were c.4170T→A heterozygotes. The c.4170T→A mutation was predicted to result in a premature stop codon Y1390X in exon nine, thus leading to a truncation of the eight remaining *LCT* exons. Two patients out of those five $\text{Fin}_{\text{major}}$ heterozygotes had a paternally transmitted deletion of four nucleotides, c.4998_5001delTGAG, in their other disease allele in exon 14. This mutation was anticipated to end in a frameshift and an early stop codon after 55 changed amino acids (S1666fsX1722). One of those five patients had a deletion of two nucleotides c.653_654delCT in exon two predicting a frameshift and truncation at S218fsX224. The patient's father carried the mutation. The last two mutations, each of which was found in a different patient, were substitutions of c.804G→C and c.4087G→A. They are expected to result in amino acid substitutions of histidine for glutamine at Q268H and serine for glycine at G1363S in exon 3 and 14, respectively. Two patients were heterozygous for Q268H or G1363S and $\text{Fin}_{\text{major}}$. Both mutations were detected to be as a maternal origin. In addition, the c.4087G→A mutation was carried by a healthy sibling in the family. The CLD mutations and corresponding lactase activities of duodenal biopsy specimens are shown in Table 4.

Table 4. Identified CLD mutations in the *LCT* gene and detected lactase activities of duodenal biopsy specimens.

Mutation	Consequence in protein	Exon	Mean lactase activities U/g protein (range)
c.4170T→A	Y1390X	9	2.4 (0-7)
c.4998_5001delTGAG	S1666fsX1722	14	5 (3-7)
c.653_654delCT	S218fsX224	2	2
c.804G→C	Q268H	3	5
c.4087G→A	G1363S	9	3

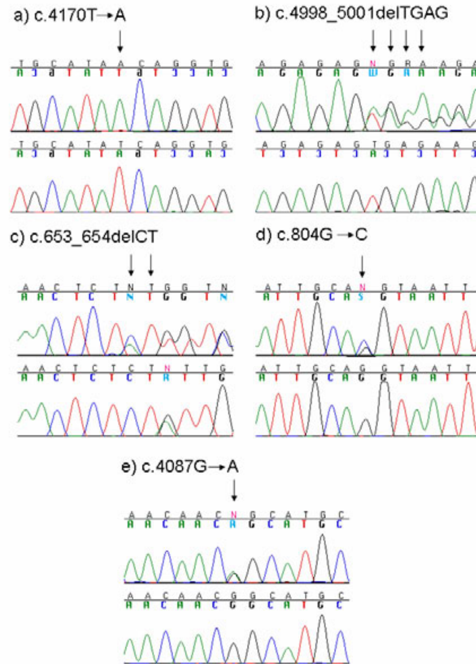


Figure 13. CLD mutations in the *LCT* gene. DNA sequence chromatograms of five identified CLD mutations, a-e. Mutation Y1390X (a) is homozygote, whereas the rest of mutations are heterozygotes. The first row shows affected and the second row wild type sequences.

5.1.2.1 The impact of Fin_{major} on *LCT* transcription

To explore the consequence of the Fin_{major} mutation at the mRNA level, *LCT* transcripts were analyzed using total RNA extracted from duodenal biopsies of a five-year old patient and five control subjects aged 6.7-17 years (Table 1, Study II). The patient was heterozygous for Fin_{major} and C/T_{-13910} that defines the lactase non-persistence/persistence phenotype. The patient's duodenal lactase activity was 6 U/g of protein and the L/S ratio was 0.08. All control subjects had the C/T_{-13910} genotype and lactase activity (determined from four subjects) ranged between 21-29 U/g of protein and L/S ratio between 0.28-0.62. RT-PCR followed by solid-phase minisequencing was employed to quantitate mRNA levels using an informative SNP G/A_{+593} in *LCT*. The T_{-13910} allele of the control subjects was expressed about five times higher than the downregulated C_{-13910} allele. On the contrary, the T_{-13910} allele carrying the Fin_{major} was not able to maintain the *LCT* expression. The phenomenon is probably due to nonsense-mediated mRNA decay (NMD). This mechanism is harnessed to eliminate the production of potentially harmful truncated proteins degrading the nascent mRNA (Maquat 2005).

5.1.3 Genealogy and carrier frequencies of the CLD mutations in Finland

On the basis of the genealogical studies Fin_{major} and S1666fsX1722 mutations were tracked to central and northern Finland, respectively. Mutations S218fsX224, Q268H and G1363S were traced in eastern Finland. The carrier frequency for the CLD mutations was determined among 556 anonymous donors representing both the early and the late settlement regions of Finland (Figure 14). Carriers of the Fin_{major} were found in all towns analyzed locating in later settlement regions. The highest carrier frequency of was found in Nilsiä 1:35 (4/140) predicting an incidence of ~1:4800 for CLD in this particular region. In this region, a cluster of the ancestors

was also trailed using church book records (Järvelä et al. 1998). No carriers for the four more rare CLD mutations were observed in any subpopulation screened.

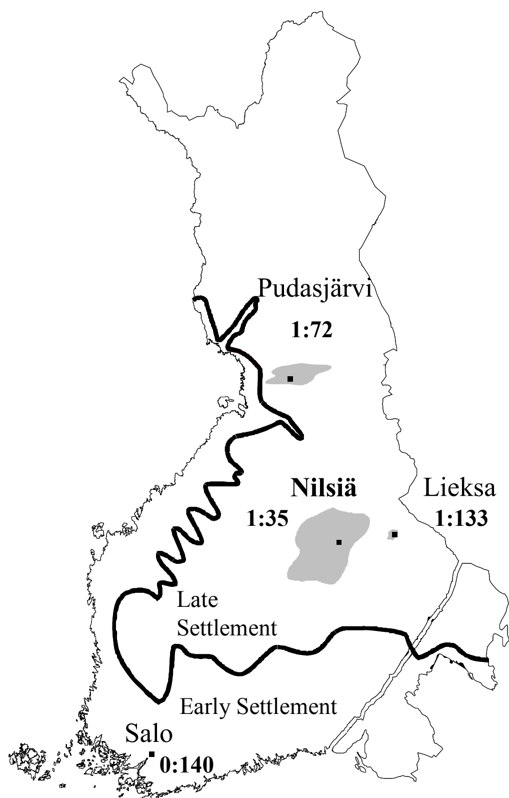


Figure 14. Carrier frequency of the Y1390X mutation in four subpopulations of Finland. Modified from Study I.

5.2 Discussion of congenital lactase deficiency

CLD patients suffer from deficiency of lactase activity in the intestinal lumen. Thus, infants, fed breast or other lactose containing milks, are exposed to substantial watery diarrhea that risks normal growth and development. Breast milk is, indisputably, an essential source of nutrition postnatally. For example, antibodies of

mother's milk stimulate infants' immune systems to protect them from fatal infections. The importance of lactose is indicated by the fact that breast milk contains 6.5% of lactose that represents 37% of its energy resource (Fineli, National Public Health Institute, Finland). Furthermore, acidic lactose facilitates the development of the valuable colonic bacterial floor of newborns.

It took nearly a half of the century to unravel the molecular background of CLD as the very first patients were described already in 1959 (Holzel et al. 1959). In this study, five distinctive mutations in the *LCT* gene underlying this disease are finally reported. The mutation Y1390X was found to be the prevailing one as 90% (43/48) of disease chromosomes carried the mutation. The observation is considered to reflect the founder effect of the Finnish population. The rest of the mutations were rare being specific to one or two families only.

As described in chapter 2.2.3 lactase is synthesized as a precursor of 1927 amino acids. The lactase precursor includes a pro-domain (repeats I and II) and an active domain (repeats III and IV), which possesses the catalytically active sites for phlorizin hydrolase and lactase (Figure 15). The pro-domain functions as an intermolecular chaperone assisting in the formation of the functional enzyme complex. It is cleaved off in the Golgi complex yielding the mature enzyme. Mutations G1363S, Y1390X (Fin_{major}) and S1666fsX1722 hit directly in the active domain while S218fsX224 and Q268H were found in the pro-domain. Fin_{major} and S1666fsX1722 truncate lactase polypeptide at repeat IV and S218fsX224 at repeat I. Fin_{major} was observed not be able to retain *LCT* expression probably affecting the stability of *LCT* mRNA directing the mutated transcript to NMD. NMD is across eukaryotes a conserved mRNA surveillance mechanism that verifies that mRNA has the appropriate quality for translation. NMD prevents the production of potentially fatal truncated proteins created by premature termination codons (Wagner and Lykke-Andersen 2002; Maquat 2005). However, premature termination codons closer than 50 nucleotides upstream of the last intron escape from NMD. This allows

accumulation of mutant protein resulting in a dominant negative effect as observed in dominant forms of β-thalassemia and Marfan syndrome (Baserga and Benz 1988; Frischmeyer and Dietz 1999). Although, NMD has been observed to link with nonsense mutations they hardly had enough evolutionary pressure to maintain the mechanism. Indeed, NMD has been observed to regulate normal cellular processes and mute genomic noise (Carter et al. 1995; Mendell et al. 2004). Lactase activity and its mobility on polyacrylamide gels were studied in duodenal biopsies of three of our CLD patients with Fin_{major}. In this preliminary study, brush-border lactase protein and activity was barely detectable, which supports undisputedly the NMD hypothesis (Freiburghaus et al. 1976). Alleles with frameshift mutations, S218fsX224 and S1666fsX1722, resulting eventually in stop codons may also be guided to NMD. Q268H substitutes the alcaline histidine for an uncharged glutamine and was found in LCTα the intramolecular chaperone. Q268H may affect the structure of LCTα disturbing the maturation process of lactase polypeptide. G1363S substitutes a small glycine for a reactive serine. The mutation locates close to catalytically active sites possibly changing the three-dimensional structure of the LCT polypeptide hence, inactivating the enzyme. The importance of the Q268 and G1363 residues are indicated by the fact that they are conserved across the species. The schematic structure of lactase polypeptide and the location of the five CLD mutations are illustrated in Figure 15.

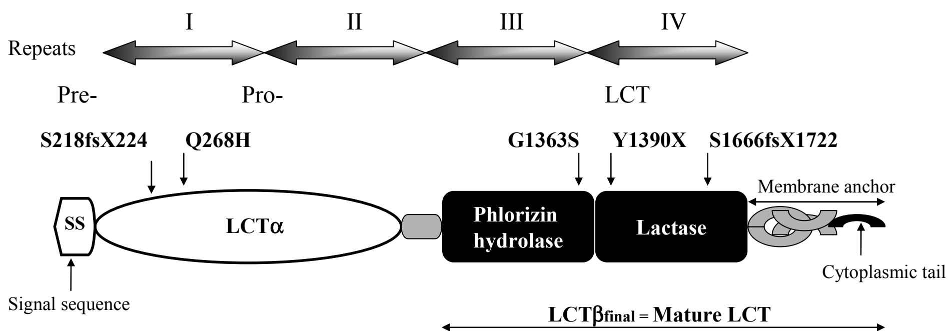


Figure 15. Structure of the translated lactase polypeptide and the location of the identified CLD mutations. The primary translation product (pre-pro-LCT) contains 1927 amino acids. The pro-LCT consists of four homologous repeats (I-IV). Repeats I and II are located in the pro-domain while repeats III-IV are found in the mature lactase (Mantei et al. 1988). Mutations S218fsX224 and Q268H were identified in LCTα that functions as an intramolecular chaperone. Mutations G1363S, Y1390X and S1666fsX1722 were found in mature lactase.

Regional clusters of CLD in the late settlement region of Finland, the rareness of disease, and the detected ~6 cM interval for LD indicate a fairly late introduction of the disease alleles into the population. The timing is expected to coincidence with many of the diseases in the Finnish disease heritage like IOSCA whose disease alleles were estimated to have been introduced before the major internal population movement to the inland in the 1500s some 30 generations, 750 years ago (Varilo et al. 1996).

Contrary to the rare CLD, lactase non-persistence is a common enzyme deficiency worldwide. Lactase activity in adult-type hypolactasia declines to a mean value of 5.9 U/g protein (Study III) differing significantly from the CLD patients' mean value of 2.4 U/g protein [$t(36) = 5.0$; $p = 0.0001$]. Interestingly, the CLD Fin_{major} mutation was found in the background of the lactase persistent allele, T_{-13910}, in Finns. As showed in this study, some Fin_{major} carriers with the C/T_{-13910} genotype may suffer from lactose-related symptoms after downregulation of the C_{-13910} allele.

5.3 Results of *LCT* mRNA studies

5.3.1 *LCT* expression and quantitation in children (II)

Fifteen subjects of forty-five (33%) were observed to be heterozygous for the A/G_{+593} polymorphism residing in exon 1 of the *LCT* gene and, thus their intestinal biopsies could be used for further analyses. Twelve of them were heterozygous for the lactase persistent C/T_{-13910} and three were homozygous for the lactase non-persistent C/C_{-13910} genotype (Table 5). The subjects with the C/T_{-13910} genotype (age range: 10 months to 17 years) had high lactase activity, ranging from 21 to 113 U/g of protein.

Table 5. Lactase activities, L/S and allelic ratios in children.

Age (years)	Genotype (C/T_{-13910})	C_{-13910} expressed (%)	Lactase activity (U/g protein)	L/S ratio
1.1	C/C	52	ND	ND
5.0	C/C	51	24	0.28
22.8	C/C	49	6	0.08
0.8	C/T	48	85	1.11
1.1	C/T	52	113	1.02
4.0	C/T	48	31	0.49
4.3	C/T	42	53	0.48
4.7	C/T	40	40	0.62
4.9	C/T	42	ND	ND
6.7	C/T	18	22	0.28
7.6	C/T	16	ND	ND
11.1	C/T	13	29	0.54
14.9	C/T	17	21	0.40
17.0	C/T	24	29	0.62

ND not determined

A standard curve was prepared for the G/A_{+593} of the *LCT* gene. An equal level of lactase mRNA was transcribed from both C_{-13910} and T_{-13910} allele in children younger than 5 years. In those aged between 4 and 5, a decline in the lactase mRNA transcribed from the C_{-13910} allele started to appear. In the children aged four to five years, the relative expression from the C-allele was 48% in two cases and, 40-42% in another three. In children aged six and seven years relative *LCT* mRNA expression was reduced to 18% and 16%. In children >6 years of age the relative expression was reduced to 13-24 % of the expression from that of the T_{-13910}-allele (see Figure 16, and Table 5). In contrast, in children with the C/C_{-13910} genotype the relative expression from both alleles was equal as expected.

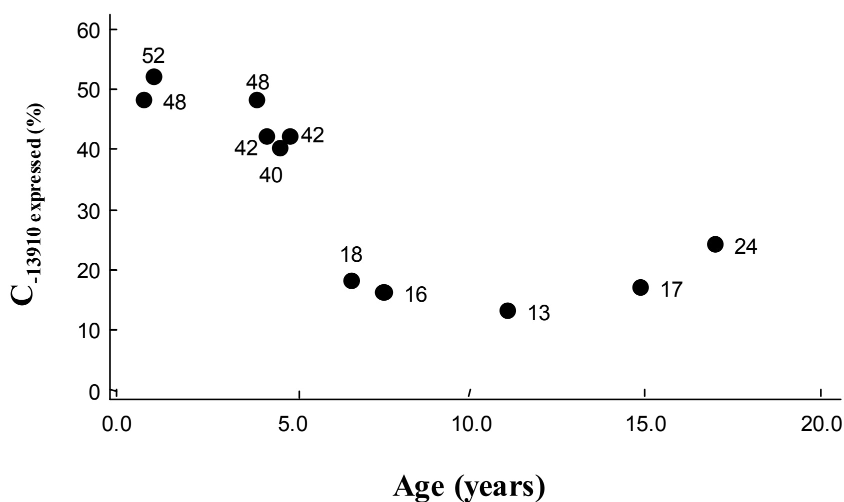


Figure 16. Relative proportion of *LCT* expressed by the C_{-13910} allele.

5.3.2 *LCT* expression in adults (III)

Relative allelic expression of *LCT* transcripts was studied using 52 patients. Eleven of the 52 patients (21%) were observed to be hypolactasic with lactase activities (4-9 U/g), L/S ratios (0.05-0.18), and genotypes C/C_{-13910} and G/G_{-22018}. Twenty-three (44%) patients' lactase activities ranged between 13-49 U/g, their L/S ratios from

0.35 to 0.69 and they were heterozygous for C/T$_{-13910}$ and G/A$_{-22018}$. Eighteen of the 52 (35%) patients were homozygous for T/T$_{-13910}$ and A/A$_{-22018}$ and their lactase activities ranged from 18 to 87 U/g and their L/S ratios from 0.49-1.05 (Figure 17). Using Fisher's exact test, the statistically significant difference between each genotype and their corresponding L/S ratios was found to have a p-value <0.0001.

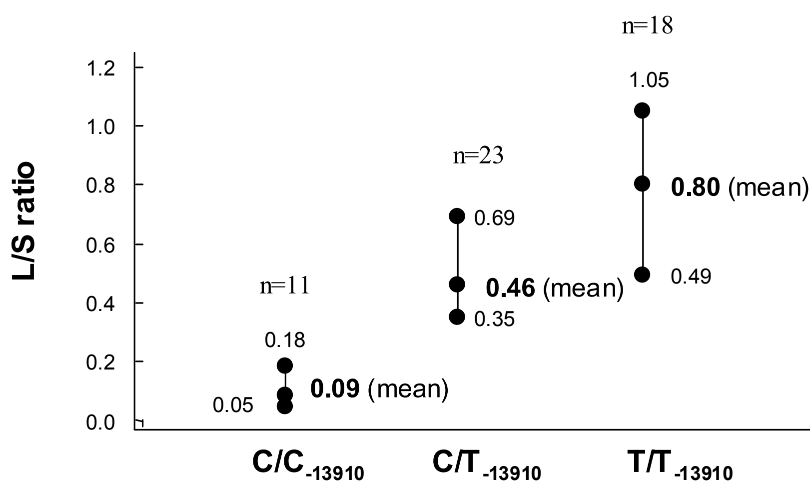


Figure 17. Correlation of *LCT* genotypes with L/S ratios. The effect of the T$_{-13910}$ variant was observed as elevation of L/S ratio, ($p<0.0001$).

5.3.2.1 Quantitation of *LCT* mRNA levels

The most informative single nucleotide polymorphisms (cSNPs) in the coding region of the *LCT* gene are located in exon 1 (c.593G→A) and exon 17 (c.5578C→T) in the Finnish families. These cSNPs were primarily chosen to quantify the relative mRNA levels. Twenty-two of fifty-two (42%) were found to be heterozygous for one or both of these cSNPs. Novel single nucleotide polymorphisms were searched for in the 30 uninformative patients by direct sequencing of exons 2, 6, 10, and/or 13. Six patients were found to be informative at least for one of the sites. Seventeen of the remaining 24 patients, were homozygous for T/T$_{-13910}$, A/A$_{-22018}$, 7 of those 24 were heterozygous for C/T$_{-13910}$, G/A$_{-22018}$.

Thus, 11 hypolactasic patients, 16 lactase persistent heterozygous and one lactase persistent homozygous could be further analysed for relative allelic expression.

To relate the actual difference in the expression of the *LCT* alleles in the minisequencing assay a standard curve was prepared for the G/A_{+593} cSNP in exon 1. The relative quantitation of the *LCT* alleles showed that the persistent allele represented 92% +/- 6% (mean +/- standard error of mean, range 78-99%, n=14) of the total *LCT* mRNA quantity. A clear increase in the R-value, suggesting allelic asymmetry of *LCT* mRNA expression, was also seen independently when quantitation was performed for any of the cSNPs located in exons 1, 2, 6, 10, 13, or 17 (Table 6). Hypolactasic patients with the C/C_{-13910} and G/G_{-22018} genotypes had equal expression of both alleles (47% +/-1%, n=7). Only one subject homozygous for lactase persistent T_{-13910}, A_{-22018} had an informative SNP of the coding region of the *LCT* gene. This individual equally expressed both *LCT* alleles (Table 6).

Table 6. The genotypes of each patient and their expressed allelic ratio of the *LCT* gene.

		Genotype	Cpm ratio of the cSNPs of the LCT gene				
Diagnosis	N	C/T_{-13910}/ G/A_{-22018}	Exon1 A/G	Exon6 G/T	Exon10 G/A	Exon13 C/T	Exon17 C/T
Hypolactasic	11/11	CC/GG	0.58-0.70	-	1.1	1.3-1.4	1.5-4.7
Lactase persistent	16/23	CT/GA	3.3-37	-	5.5-7.6	-	13-64
Lactase persistent	1/18	TT/AA	-	1.6	-	-	-

5.4 Discussion of *LCT* mRNA studies (II, III)

The decline of lactase activity after childhood, adult-type hypolactasia, is probably the best-known enzyme deficiency in humans. Hence, it has inspired scientists to study its occurrence for a century. The development and creation of methods in the

natural science over the last two decades has provided opportunities to approach this dilemma at the molecular level. As discussed in section 2.4.3.2 ‘Mechanism of downregulation’, the downregulation of lactase activity has been suggested to operate at the transcriptional and/or posttranscriptional level. This enigma climaxed in Enattah and his colleagues’ (2002) linkage disequilibrium and haplotype studies in Finnish families, which exposed two variants associated with adult-type hypolactasia supporting the transcriptional regulation theory.

In this study, those variants affecting the developmental changes in *LCT* expression was evaluated in biopsy specimens by determining the biological activities of lactase and sucrase and quantitating the mRNA levels by solid-phase minisequencing in Finnish children and adults. In children the C_{-13910} allele, determining lactase non-persistence, declined by the age of six years (Figure 18). In children aged six years or more the C_{-13910} allele contributed only 13-24% to the produced *LCT* mRNA. The age of *LCT* mRNA was observed to decline is pretty well to scale with that of the age of downregulation of lactase activity in Finnish children (Sahi 1994a; Rasinperä et al. 2004). However, the age when downregulation takes place varies between populations. In Thai children the downregulation event was observed between one and two years of age (Keusch et al. 1969a). In African children, the decline of lactase activity was detected from three years of age (Rasinperä et al. 2004). A British study of different ethnic groups demonstrated that the decline of *LCT* mRNA expression was detected from the second year of life (Wang et al. 1998b). The spatial and temporal gene regulation involved in the timing of *LCT* decline is so far unsolved. Timing variables should be taken into account when the usefulness of the genetic testing of C/T_{-13910} for adult-type hypolactasia is estimated. Tests before the age of downregulation occurs can lead to diminished milk consumption too early. Thus, positive effects of milk may be remained unexploited. The determination of C/T_{-13910} after 8-12 of years is considered to reflect the lactase phenotypes reliably (Rasinperä et al. 2004). In adults, the effect of a persistent allele T_{-13910} A_{-22018} was seen in elevation of lactase activity that was furthermore observed in significant

correlations in lactase activities between all the three genotypes C/C_{-13910} G/G_{-22018}, C/T_{-13910} G/A_{-22018}, and T/T_{-13910} A/A_{-22018}. The quantitation of relative *LCT* mRNA levels showed a clear difference between patients with lactase persistent and non-persistent genotypes. The lactase persistent allele, T_{-13910} A_{-22018}, had an average 11.5-fold higher expression than the non-persistent allele, C_{-13910} G_{-22018} (Figure 18). Lactase non-persistence patients having the C/C_{-13910} G/G_{-22018} genotype, and patients homozygous for lactase persistent with the T/T_{-13910} A/A_{-22018} genotype evenly expressed both alleles. These observations unambiguously show that the variants C/T_{-13910}, G/A_{-22018}, associated with adult-type hypolactasia, are also associated with the transcriptional regulation of the *LCT* gene. These results agree with previous findings that indicated that the decline of lactase activity is regulated at the transcriptional level (Escher et al. 1992; Lloyd et al. 1992; Fajardo et al. 1994) operating by a cis-acting element (Wang et al. 1995; Wang et al. 1998b).

Both the SNPs, C/T_{-13910} and G/A_{-22018}, are located in introns of *MCM6*, which locate at 5'-end of *LCT*. MCM6 is a mammalian homologue of mis5 of yeast, it functions as a cell cycle factor (Takahashi et al. 1994). *MCM6* is expressed in a variety of human tissues including intestine. Intestinal expressions of *MCM6* and *LCT* were not observed to correlate with subjects of lactase persistence or non-persistence suggesting that these genes are independently regulated (Harvey et al. 1996). So, obviously the influence of the *MCM6* gene on adult-type hypolactasia is only structural. The observation was confirmed later when the C/T_{-13910} construct was not able to activate the MCM6 promoter (Lewinsky et al. 2005).

Recently, studies have been focussed to clarify the functional role of the C/T_{-13910} G/A_{-22018} variants. Two studies using human (Troelsen et al. 2003b) and rat (Olds and Sibley 2003) LCT promoter-reporter gene construct analyses with C/T_{-13910} region in Caco-2 cells demonstrated independently differential regulation of LCT promoter activity. Furthermore, EMSA showed that the observed difference was a consequence of the binding capacity of the nuclear factors for the C_{-13910} and T_{-13910}

allele constructs (Troelsen et al. 2003b). The T_{-13910} variant had a stronger interaction than C_{-13910} suggesting a model for the T_{-13910} as a strong transcriptional enhancer. Similar analyses of the G/A_{-22018} variant did not reveal any significant functional role for phenotype differences of lactase persistence/non-persistence (Olds and Sibley 2003; Troelsen et al. 2003b). Later, DNA affinity purification analyses showed that Oct-1 binds directly to the C/T_{-13910} variant co-purifying with GAPDH and furthermore, the binding was detected to be stronger to the T_{-13910} variant than the C_{-13910} variant. DNase footprint and supershift analyses identified binding motifs for Cdx-2, GATA-6, Fox and HNF-4α in the close proximity of C/T_{-13910}. In mutation analyses, GATA-6, Fox and HNF-4α motifs were detected to be necessary for the pursuit of the C/T_{-13910} enhancer. Co-transfection analyses with C/T_{-13910} demonstrated that Oct-1 was able to activate LCT promoter only with HNF-1α. Interestingly, a HNF-1α site has not been found within the C/T_{-13910} enhancer but it probably is acting via a proximal promoter, which has a HNF-1α binding site. The results are convincing. The C/T_{-13910} variant seems to locate in the heart of an enhancer structure consisting of Oct-1, GATA, Fox and HNF-4α binding sites (Figure 18). Oct-1 with the C/T_{-13910} sequence directs specifically the LCT promoter activity and enhances reporter gene expression with the T_{-13910} variant significantly over the C_{-13910} variant. These discoveries offer a plausible model for the T_{-13910}-induced lactase persistence phenotype in humans (Lewinsky et al. 2005).

Although, we have learned a great deal about the molecular background of the lactase persistence phenotype the virtual cascade that leads to activation of the *LCT* gene is still obscure. Interestingly, Oct-1 has been reported to recruit co-factors for chromatin structure enhancing or silencing genes depending on the tissue type or promoter structure (Zabel et al. 2002; Remenyi et al. 2004). A similar mechanism was also considered as a potential one for the lactase persistence phenotype (Lewinsky et al. 2005). The T_{-13910} allele has been demonstrated to match the functional, phenotypic and epidemiological data of European and European derived

populations (reviewed in section 2.4.3 ‘Genetics’). It has been reported that statistically the T_{-13910} allele could not explain the lactase persistence incidence in some African populations (Mulcare et al. 2004). If carefully characterized phenotypes can confirm the results, there are still unidentified variants to be found, which probably will provide their own interesting characteristics to explain the lactase persistence mechanism.

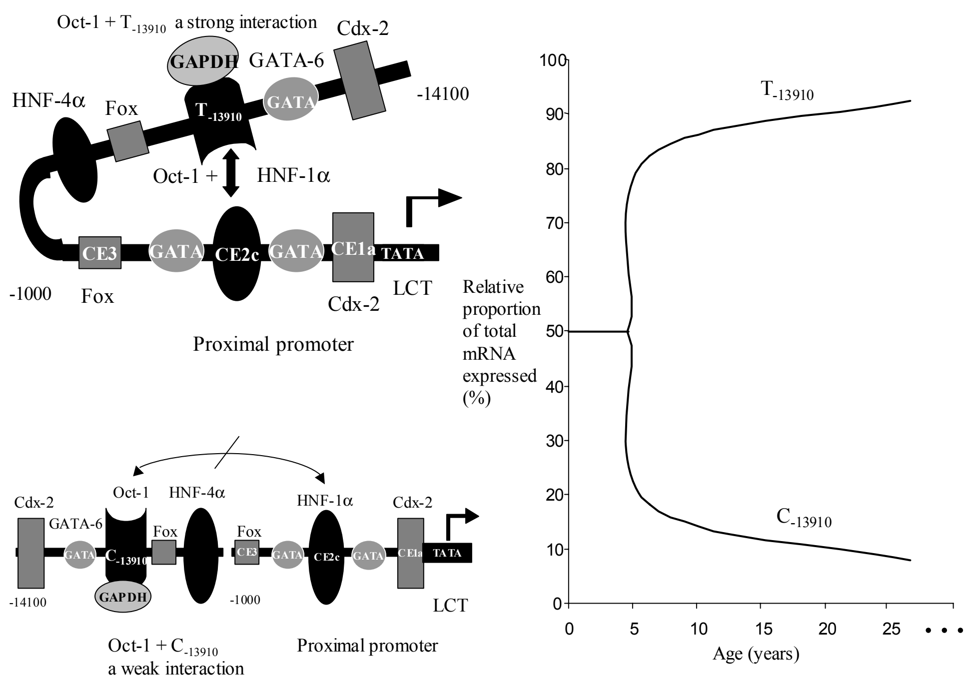


Figure 18. Regulation of the *LCT* gene – an overview. Regardless of the genotype of the C/T_{-13910} enhancer, *LCT* is fully expressed up to five years of age. The C_{-13910} variant has weaker interaction with Oct-1 than the T_{-13910} variant (Lewinsky et al. 2005), hence yet unknown mechanism allows the downregulation of the C_{-13910} allele at the age of six (Study II). In contrast, the T_{-13910} allele maintains *LCT* expression and is responsible for an average 92% of expressed *LCT* mRNA (Study III).

5.5 Results of lactase persistence and ovarian carcinoma (IV)

Lactase persistence had no statistically significant effect on the development of ovarian carcinoma in the Finnish, Polish or Swedish populations. Combined analysis of all three populations showed no significant risk for ovarian cancer (Figure 19). In addition, dissecting ovarian carcinoma tumours by histological subtypes for analyses did not affect the outcome (Table 1, Study IV).

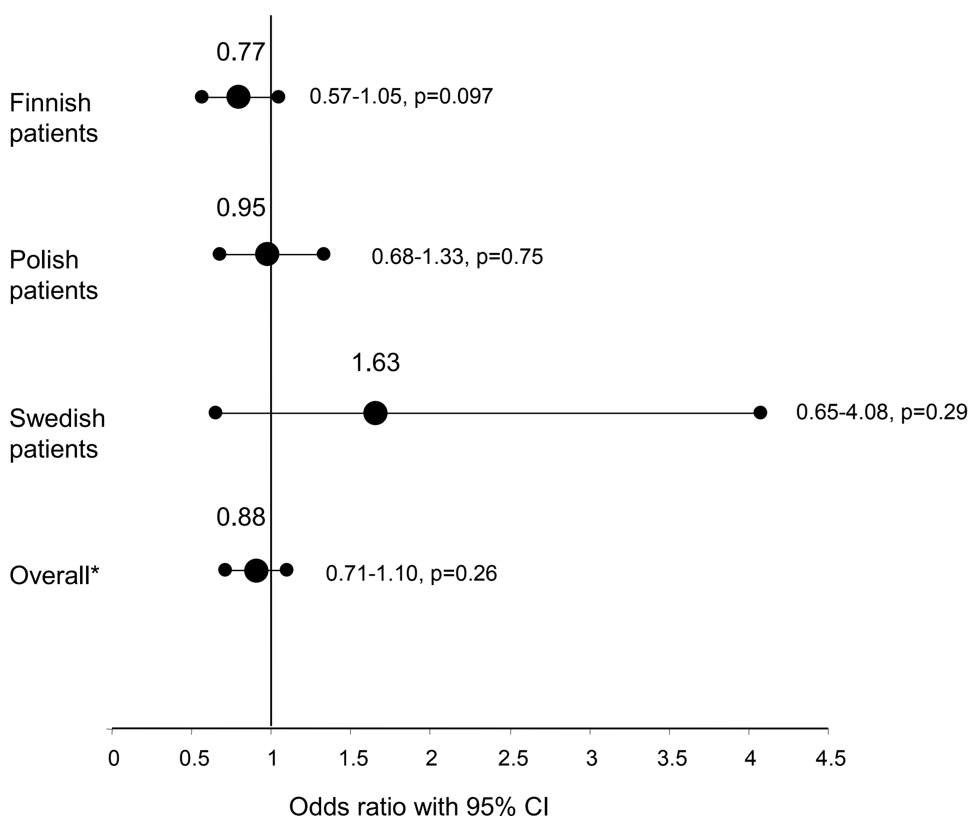


Figure 19. Molecularly defined lactase persistence and its risk to ovarian carcinoma pathogenesis in the Finnish, Polish and Swedish populations. *Mantel-Haenszel common odds ratio estimates, (CI) confidence interval.

The frequency of the C/C$_{-13910}$ genotype was detected as 33.1% in the Polish controls. This was a bit lower than the earlier reported incidence of lactase non-persistence (37.5%) diagnosed by BHT in the Polish population (Socha et al. 1984) In Swedish controls, the C/C$_{-13910}$ frequency was observed to be 10.3% which is in concordance with the recently observed occurrence of C/C$_{-13910}$ (9.6%) in the Swedish population (Nilsson TK personal communication).

5.6 Discussion of lactase persistence and ovarian carcinoma

Ovarian cancer is a deadly disease among women with a high incidence in western industrialized countries (World Cancer Research Fund 1997). Epithelial ovarian cancer (ovarian carcinoma) is the major cancer type explaining 90% of ovarian cancer cases. The aetiology is still for the most part unravelled but both genetic and environmental factors affect the pathogenesis. The pathogenesis has been studied intensively tracing frequent ovulation cycles and hormonal issues as contributory factors to the disease. Dietary factors particularly a high consumption of milk (lactose) and lactase persistence which is closely related its consumption have been put forward as risk factors for ovarian tumour genesis (Cramer 1989). Galactose, the hydrolysing product of lactose, was observed to be toxic to oocytes in animals studies (Chen et al. 1981; Schwartz and Storozuk 1988). Thus, galactose has been hypothesized to promote the development of ovarian carcinoma. In order to test the hypothesis, the C/T$_{-13910}$ variant that specifies lactase phenotypes, was defined in ovarian carcinoma patients and compared with healthy controls in the Finnish, Polish and Swedish populations. We observed that lactase persistence had a modest reduced effect (OR 0.77) on ovarian carcinoma risk but the trend was not statistically significant. Lactase persistence did not associate with the ovarian carcinoma risk either in the Polish or Swedish populations. The results might be related to the fact that Finnish milk consumption is the highest in the world (The

World Dairy Situation 2003). Lactase persistence facilitate milk's use as a nutrition which has been noted as an increasing milk consumption (Obermayer-Pietsch et al. 2004; Rasinperä et al. 2004; Rasinperä et al. 2006). However, these observations cannot replace the dairy consumption data of subjects of this study. As reviewed in chapter 2.5.1.1 'Dairy consumption and ovarian carcinoma risk' many studies have tried to shed light on the risks or benefits of milk and its components for ovarian tumour genesis. But there is not been any breakthroughs concerning the issue. Meta-analyses have shown that only large prospective cohort studies show a modest increased risk between high milk consumption and ovarian carcinoma (Genkinger et al. 2006; Larsson et al. 2006). Interestingly, the same trend was observed in a meta-analysis between milk consumption and colorectal cancer (Norat and Riboli 2003). Thus, large sample sizes probably play a role in these kinds of studies.

Milk is a versatile nutrient and an important source of calcium and many other traces and vitamins (Table 2). For example, dietary calcium has been observed to decrease bone fracture risk and to prevent osteoporosis in adulthood (Gueguen and Pointillart 2000; Prentice 2004). Milk has also been regarded as benefit for early farmers since the lactase persistence mutation arouse 5000-10000 years ago. The selection power of lactase persistence is the strongest seen in the human genome (Cavalli-Sforza 1973; Heston and Gottesman 1973; Flatz and Rotthauwe 1977; Aoki 1986; Flatz 1987; Bersaglieri et al. 2004). Perhaps, an abstemious consumption brings the best out of milk also today. The results of our study do not support the hypothesis that lactase persistence promotes the development of ovarian carcinoma.

6 CONCLUDING REMARKS AND FUTURE PROSPECTS

There are two different types of lactase deficiencies in humans; a rare recessively inherited CLD and a common so-called adult-type hypolactasia with milder features. CLD belongs to the Finnish disease heritage and is caused by mutations detected in the translated region of lactase, which is responsible for hydrolysing lactose in the intestine. In this study, five unique mutations were found in 32 patients in 24 families. They all were expected to inactivate mucosal lactase activities. Mutation Y1390X, Fin_{major}, is anticipated to direct *LCT* transcripts to NMD that is typically the consequence of premature stop codons. A detailed analysis of impact of other four mutations on the development of CLD should be conducted in the future. Especially, Q268H and G1363S mutations can bring new insights for the mechanism of lactase maturation and for crucial sites of activities of lactase and phlorizin hydrolase. From the clinical viewpoint, this study has created a direct DNA-based diagnosis and a carrier identification of CLD that has thus far been relied on clinical symptoms completed by lactase activity analysis from duodenal biopsy specimens.

The development of adult-type hypolactasia, lactase non-persistence, is dependent on a variant C/T_{-13910} locating 13.9 kb upstream from *LCT* (Enattah et al. 2002). In this study, the C_{-13910} allele was detected to allow the transcriptional downregulation of *LCT*. The timing of the downregulation was detected at the age of six. In adults, the T_{-13910} allele represented a mean 92% of the expressed *LCT* mRNA when compared with the C_{-13910} allele. The C/T_{-13910} variant locates in the middle of enhancer element consisting of Oct-1, GATA, Fox and HNF-4α binding sites. Oct-1 was demonstrated to bind directly to the C/T_{-13910} site. It has been speculated that Oct-1 may induce chromatin changes and activate *LCT* (Lewinsky et al. 2005). The major goal of future functional studies will be to clarify which genes are directly involved in transcriptional downregulation of *LCT*. For example, evidence of

activated or inactivated genes can be traced from duodenal biopsies by an Affymetrix array. It remains also unknown which variants and/or mechanisms are at the background of lactase persistence in some African and Arabia nomads.

Milk is a significant part of the Northern European and American diet. Milk's lactose has been suggested to be a risk factor for the development of ovarian carcinomas. As lactase persistence is associated with increased milk consumption, C/T_{-13910} was tested in ovarian carcinoma patients and in the Finnish, Polish and Swedish populations. Lactase persistence did not associate with ovarian carcinoma risk in any of the populations tested. Multiple studies have been conducted in an attempt to shed light on dietary milk and its components in relation to ovarian tumour genesis (section 2.5.1.1). However, the results have remained controversial possibly reflecting the fact that milk has no cancer promoting effect on the ovary. On the other hand, components of milk are hard to dissect in this type of studies. Cell line studies are a more exact approach to study promoting or inhibiting factors of tumour growth. For example, high concentrations of 1,25 dihydroxyvitamin D has been detected to inhibit human ovarian cancer cell lines (Ahonen et al. 2000; Miettinen et al. 2004). However, only two case-control studies have shown that dietary vitamin D diminish ovarian carcinoma risk (Bidoli et al. 2001; Salazar-Martinez et al. 2002). Ovarian cancer is a deadly disease, about 110000 women per year perishes from the disease in the world (Parkin et al. 2001). Even if some genes have found to be causal for the disease, the background of the most of cases has remained obscure. It is probable that eventually efforts will recompense those involved in the research and their findings pave the way for risk factors and mechanisms underlying ovarian carcinoma. It is hoped that new diagnostic markers would permit more accurate and specific diagnosis and offer new tools for drug development for this lethal cancer.

7 ACKNOWLEDGEMENTS

This research was carried out in the Department of Molecular Medicine at the National Public Health Institute in Helsinki and in the Department of Medical Genetics at the University of Helsinki. I want to thank the former and the present heads of the National Public Health Institute Jussi Huttunen and Pekka Puska for providing excellent research facilities.

The pride and joy of our laboratory my supervisor Docent Irma Järvelä is thanked for giving me the opportunity to work in this project. I am grateful to her for introducing me the world of molecular genetics. I want to thank her for educative, endless and amusing scientific and non-scientific talks during my time here. I wish to thank Academy Professor Leena Peltonen-Palotie for her interest toward my work and for supporting me with her great knowledge in science. I want to thank her for the educative, exciting and inspiring discussions that have given me a drive for my scientific career.

I would like to thank Professor Helena Kääriäinen for accepting the role as Opponent in my thesis defence.

Professor Mikko Hallman and Docent Johanna Schleutker are thanked for reviewing my thesis. Their constructive comments cannot be praised enough. Bachelor of Science M.A. Donald J.M. Smart is acknowledged for language revision of this thesis.

I wish to thank excellent clinicians Professor Erkki Savilahti, Docent Jorma Kokkonen, Aino Oksanen MD, PhD and Docents Ralf Bützow and Kaija-Leena Kolho. Their contributions to my work have been priceless and discussions fruitful. Docent Arto Orpana is thanked for the brilliant guidance of the mRNA quantitation and for pieces of advice concerning research in general. I would like to thank Docent Teppo Varilo for the genealogical studies of CLD and for his cheerful company in the laboratory.

I wish to thank Docent Marjo Kestilä for her many, many helpful discussions and guidance and interest toward my research. You always had a time for me! Senior Docents Ismo Ulmanen and Anu Jalanko are warmly thanked help in various matters in the lab. Iski is also thanked for organizing numerous unforgettable excursions.

I would like to thank Professor Jan Lubinski, and Doctors Krzysztof Medrek, Susanne Malander and Mef Nilbert for the collaboration in the ovarian carcinoma study (IV).

I wish to thank the members of the lactase group Nabil Enattah and Heli Rasinperä for the help in the lab and for interesting discussions during these years. Especially, issues concerning the right colour of stockings and the lost of a missile in the desert have been memorable.

Tero Ylisaukko-oja is thanked for his friendship, support, brilliant and not so brilliant (if possible) conversations with a wide perspective. I think we have shared a few laughs and theories. Knowing you has been a pleasure. Thanks also for badminton and tennis matches but do not worry as Niklas stated in his thesis “se kasvattaa luonnetta”.

Hanna Komu is acknowledged for her skilful handling of lab tasks. Karola Rehnström is thanked for her cheery personality, beautiful smile and introducing the blond perspective in the lab. Elli Kempas is thanked for the help and guidance of various methods. The newcomer of the lactase group Suvi Torniainen and former and present lab members of Irma: Anne Koivisto, Susanna Lemmelä, Ilona Nummela, Sanna Seitsonen, Ulla Sundström, Hanna Nurmi, Sini Penttilä, Kristiina Pulli, Selina Mäkinen and Mari Rossi are acknowledged for their joyful company.

I would like to thank Teemu Perheentupa and Jari Raikko for solving computer-related problems and good conversations. Teemu is thanked for introducing me crisis meetings on Fridays.

The sequencing lab run by Pekka (Charlie) and the Angels Heli, Susanna and Tiina thanked for their swift and accurate analyses and happy company. Niklas is thanked

for his tennis lessons. I would like to thank past and present members of the department. Iiris, Olli, Iita, Heli, Mira, Anna, Anne, Kati, Maria, Jesper, Henna, Jani, Kaitsu, Juha I, Juha P, Nina, Tarja, Outi, Tintti, Annina, Juho, Juha S, Minna, Ritva, Arja, Päivi, Sisko, Jussi, Heidi, Elina, Tanja, Markus P, Markus, Ville H, Kaisu N, Kaisu L, Katri, Laura, Nora, Ville K, Tero H, Juri, Kaisa, Johanna S, Mervi, Janna, Minttu, Anu, William, Joe, Emma, Marika, Joni, Tiina P, Anne N and many others are thanked for cheerful company in the lab. Secretaries Sari Kivikko, Sari Mustala, Tuija Svahnbäck are thanked for the help whenever I needed. I wish to thank Jouni, Tuomo and Katja, Teemu, Kortsu, Junnu, Isto and Marika, Vesa and Tarja, Tomi, Janne, Jani, Olli, Kimmo, Terho and Jenni, Tauno and Seija for their friendship and activities in real life.

I owe my gratitude to my parents Jorma and Pirkko and my sisters Riitta and Virpi, without their support and care this study would have been impossible.

I owe my deepest gratitude to my wife Kirsi, her patience, love and care have been precious.

I wish to thank the patients and families for participating in the study.

This study has been financially supported by grants from the Finnish Cultural Foundation, Biomedicum Helsinki Foundation, the Sigrid Jusélius Foundation, Helsinki University Hospital Research Funding and the Center of Excellence in Disease Genetics of the Academy of Finland.

Helsinki, 16th of November 2006

Mikko Kuokkanen

8 ELECTRONIC DATABASE INFORMATION

Basic Local Alignment Search Tool, http://www.ncbi.nlm.nih.gov/blast/

Baylor College of Medicine, http://searchlauncher.bcm.tmc.edu/

Expressed Sequence Tags Database, http://www.ncbi.nlm.nih.gov/dbEST/

Finnish Disease Database, http://www.findis.org./

Fineli, Finnish Food Composition Database, http://www.fineli.fi/index.php?lang=en

GenBank, http://www.ncbi.nlm.nih.gov/

GENSCAN, http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.cgi

Genome Database, http://gdbwww.gdb.org/

Marshfield Medical Genetic Foundation, http://research.marshfieldclinic.org/

National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/

Online Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/Omim/

Reference Sequence Database, http://www.ncbi.nlm.nih.gov/RefSeq/

SWISS-PROT, http://www.ebi.ac.uk/swissprot/

University of California, Santa Cruz, http://genome.ucsc.edu/

9 REFERENCES

Aaltonen J, Björses P, Sandkuijl L, Perheentupa J, Peltonen L (1994) An autosomal locus causing autoimmune disease: autoimmune polyglandular disease type I assigned to chromosome 21. Nat Genet 8:83-7.

Abbas H, Ahmad M (1983) Persistence of high intestinal lactase activity in Pakistan. Hum Genet 64:277-8

Ahmad M, Flatz G (1984) Prevalence of primary adult lactose malabsorption in Pakistan. Hum Hered 34:69-75

Ahonen MH, Zhuang YH, Aine R, Ylikomi T, Tuohimaa P (2000) Androgen receptor and vitamin D receptor in human ovarian cancer: growth stimulation and inhibition by ligands. Int J Cancer 86:40-6

Aittomäki K, Lucena JL, Pakarinen P, Sistonen P, Tapanainen J, Gromoll J, Kaskikari R, Sankila EM, Lehvaslaiho H, Engel AR, et al. (1995) Mutation in the follicle-stimulating hormone receptor gene causes hereditary hypergonadotropic ovarian failure. Cell 82:959-68.

Alard P, Lantz O, Sebagh M, Calvo CF, Weill D, Chavanel G, Senik A, Charpentier B (1993) A versatile ELISA-PCR assay for mRNA quantitation from a few cells. Biotechniques 15:730-7.

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. Proc Natl Acad Sci U S A 74:5350-4.

Aminoff M, Carter JE, Chadwick RB, Johnson C, Grasbeck R, Abdelaal MA, Broch H, Jenner LB, Verroust PJ, Moestrup SK, de la Chapelle A, Krahe R (1999) Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. Nat Genet 21:309-13.

Aoki K (1986) A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. Proc Natl Acad Sci U S A 83:2929-33

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299-309

Armitage P, Berry G (1987) Statistical Methods in Medical Research. Blackwell Scientific Publications, Oxford

Arola H (1994) Diagnosis of hypolactasia and lactose malabsorption. Scand J Gastroenterol Suppl 202:26-35

Arribas JC, Herrero AG, Martin-Lomas M, Canada FJ, He S, Withers SG (2000) Differential mechanism-based labeling and unequivocal activity assignment of the two active sites of intestinal lactase/phlorizin hydrolase. Eur J Biochem 267:6996-7005

Auricchio S, Rubino A, Tosi R, Semenza G, Landolt M, Kistler HJ, Prader A (1963) Die quantitative Disaccharidasen-Aktivität des menschlichen Dunndarms und der erworbene Lactasemangel des Erwachsenen. Helv Med Acta 30:690-2.

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

Baserga SJ, Benz EJ (1988) Nonsense mutations in the human beta-globin gene affect mRNA metabolism. Proc Natl Acad Sci U S A. 85:2056-60

Becker-Andre M, Hahlbrock K (1989) Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic Acids Res 17:9437-46.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Chamberlain AT, Nunes TP, Metodiev S, Ferrand N, Erhardt G (2003) Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nat Genet 35:311-3

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111-20

Bertone ER, Hankinson SE, Newcomb PA, Rosner B, Willet WC, Stampfer MJ, Egan KM (2001) A population-based case-control study of carotenoid and vitamin A intake and ovarian cancer (United States). Cancer Causes Control 12:83-90

Bidoli E, La Vecchia C, Talamini R, Negri E, Parpinel M, Conti E, Montella M, Carbone MA, Franceschi S (2001) Micronutrients and ovarian cancer: a case-control study in Italy. Ann Oncol 12:1589-93

Bleeker FE, Hazen LG, Kohler A, Van Noorden CJ (2000) Direct comparison of the sensitivity of enzyme histochemical and immunohistochemical methods: cathepsin B expression in human colorectal mucosa. Acta Histochem 102:247-57.

Bodlaj G, Stöcher M, Hufnagl P, Hubmann R, Biesenbach G, Stekel H, Berg J (2006) Genotyping of the lactase-phlorizin hydrolase -13910 polymorphism by LightCycler PCR and implications for the diagnosis of lactose intolerance. Clin Chem 52:148-51

Boll W, Wagner P, Mantei N (1991) Structure of the chromosomal gene and cDNAs coding for lactase- phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. Am J Hum Genet 48:889-902.

Borgström BA, Dahlqvist A, Lundh G, Sjövall J (1957) Studies of digestion and absorption in the human. J Clin Invest 36:1521-1536

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.

Bosse T, van Wering HM, Gielen M, Dowling LN, Fialkovich JJ, Piaseckyj CM, Gonzalez FJ, Akiyama TE, Montgomery RK, Grand RJ, Krasinski SD (2006) Hepatocyte nuclear factor-1alpha is required for expression but dispensable for histone acetylation of the lactase-phlorizin hydrolase gene in vivo. Am J Physiol Gastrointest Liver Physiol 290:G1016-24

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-31.

Braun A, Little DP, Köster H (1997) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 43:1151-8

Britton JA, Westhoff C, Howe GR, Gammon MD (2000) Lactose and benign ovarian tumours in a case-control study. Br J Cancer 83:1552-5

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33-7.

Büning C, Ockenga J, Krüger S, Jurga J, Baier P, Dignass A, Vogel A, Strassburg C, Weltrich R, Genschel J, Lochs H, Schmidt H (2003) The C/C(-13910) and G/G(-22018) genotypes for adult-type hypolactasia are not associated with inflammatory bowel disease. Scand J Gastroenterol 38:538-42

Carter MS, Doskow J, Morris P, Li S, Nhim RP, Sandstedt S, Wilkinson MF (1995) A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. J Biol Chem. 270:28995-9003

Cavalli-Sforza LL (1973) Analytic review: some current problems of human population genetics. Am J Hum Genet 25:82-104

Cavalli-Sforza LL, Piazza A (1993) Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet 1:3-18

Cereghini S, Ott MO, Power S, Maury M (1992) Expression patterns of vHNF1 and HNF1 homeoproteins in early postimplantation embryos suggest distinct and sequential developmental roles. Development 116:783-97

Chawengsaksophak K, James R, Hammond VE, Kontgen F, Beck F (1997) Homeosis and intestinal tumours in Cdx2 mutant mice. Nature 386:84-7

Chen YT, Mattison DR, Feigenbaum L, Fukui H, Schulman JD (1981) Reduction in oocyte number following prenatal exposure to a diet high in galactose. Science 214:1145-7

Christie M, Oehler MK (2006) Molecular pathology of epithelial ovarian cancer. J Br Menopause Soc 12:57-63

Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G, Rocha J (2005) Microsatellite variation and evolution of human lactase persistence. Hum Genet 117:329-39

Collins A, Frezal J, Teague J, Morton NE (1996) A metric map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci U S A 93:14771-5.

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 9:347-50.

Cook GC, Kajubi SK (1966) Tribal incidence of lactase deficiency in Uganda. Lancet 1:725-9.

Cozen W, Peters R, Reichardt JK, Ng W, Felix JC, Wan P, Pike MC (2002) Galactose-1-phosphate uridyl transferase (GALT) genotype and phenotype, galactose consumption, and the risk of borderline and invasive ovarian cancer (United States). Cancer Causes Control 13:113-20

Cramer DW (1989) Lactase persistence and milk consumption as determinants of ovarian cancer risk. Am J Epidemiol 130:904-10

Cramer DW, Greenberg ER, Titus-Ernstoff L, Liberman RF, Welch WR, Li E, Ng WG (2000) A case-control study of galactose consumption and metabolism in relation to ovarian cancer. Cancer Epidemiol Biomarkers Prev 9:95-101

Cramer DW, Harlow BL, Willett WC, Welch WR, Bell DA, Scully RE, Ng WG, Knapp RC (1989) Galactose consumption and metabolism in relation to the risk of ovarian cancer. Lancet 2:66-71

Cramer DW, Welch WR (1983) Determinants of ovarian cancer risk. II. Inferences regarding pathogenesis. J Natl Cancer Inst 71:717-21

Cuddenec Y, Delbruck H, Flatz G (1982) Distribution of the adult lactase phenotypes--lactose absorber and malabsorber--in a group of 131 army recruits. Gastroenterol Clin Biol 6:776-9.

Dahlqvist A (1984) Assay of intestinal disaccharidases. Scand J Clin Lab Invest 44:169-72

Dahlqvist A, Hammond JB, Crane RK, Dumphy HJV, Littman A (1963) Intestinal lactase deficiency and lactose intolerance in adults. Gastroenterology 45:488-491

Dahlqvist A, Lindquist B (1971) Lactose intolerance and protein malnutrition. Acta Paediatr Scand 60:488-94.

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857-65

de la Chapelle A, Hästbacka J, Lehesjoki AE, Sulisalo T, Kere J, Tahvanainen E, Sistonen P (1994) Kytkentä ja kytkentäepätasapaino suomalaisessa tautiperinnössä. Duodecim 110:654-64

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

Diesen C, Saarinen A, Pihko H, Rosenlew C, Cormand B, Dobyns WB, Dieguez J, Valanne L, Joensuu T, Lehesjoki AE (2004) POMGnT1 mutation and phenotypic spectrum in muscle-eye-brain disease. J Med Genet 41:e115

Edwards JH (1980) Allelic association in man. In: Ericksson AW, Forsius HR, Nevanlinna HR, Workman PL, Norio R (eds) Population Structure and Genetic Disorders. Academic Press, New York, pp 239-255

Elsas LJ, Dembure PP, Langley S, Paulk EM, Hjelm LN, Fridovich-Keil J (1994) A common mutation associated with the Duarte galactosemia allele. Am J Hum Genet 54:1030-6

Enattah N, Pekkarinen T, Välimäki MJ, Löyttyniemi E, Järvelä I (2005a) Genetically defined adult-type hypolactasia and self-reported lactose intolerance as risk factors of osteoporosis in Finnish postmenopausal women. Eur J Clin Nutr 59:1105-11

Enattah N, Välimäki VV, Välimäki MJ, Löyttyniemi E, Sahi T, Järvelä I (2004a) Molecularly defined lactose malabsorption, peak bone mass and bone turnover rate in young finnish men. Calcif Tissue Int 75:488-93

Enattah NS, Forsblom C, Rasinperä H, Tuomi T, Groop PH, Järvelä I (2004b) The genetic variant of lactase persistence C (-13910) T as a risk factor for type I and II diabetes in the Finnish population. Eur J Clin Nutr 58:1319-22

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30:233-7.

Enattah NS, Sulkava R, Halonen P, Kontula K, Järvelä I (2005b) Genetic variant of lactase-persistent C/T-13910 is associated with bone fractures in very old age. J Am Geriatr Soc 53:79-82

Engle A, Muscat JE, Harris RE (1991) Nutritional risk factors and ovarian cancer. Nutr Cancer 15:239-47

Eriksson AW (1973) Genetic polymorphisms in Finno-Ugrian populations. Finns, Lapps and Maris. Isr J Med Sci 9:1156-70.

Escher JC, de Koning ND, van Engen CG, Arora S, Buller HA, Montgomery RK, Grand RJ (1992) Molecular basis of lactase levels in adult humans. J Clin Invest 89:480-3.

Fairfield KM, Hunter DJ, Colditz GA, Fuchs CS, Cramer DW, Speizer FE, Willett WC, Hankinson SE (2004) A prospective study of dietary lactose and ovarian cancer. Int J Cancer 110:271-7

Fajardo O, Naim HY, Lacey SW (1994) The polymorphic expression of lactase in adults is regulated at the messenger RNA level. Gastroenterology 106:1233-41.

Fang R, Olds LC, Santiago NA, Sibley E (2001) GATA family transcription factors activate lactase gene promoter in intestinal Caco-2 cells. Am J Physiol Gastrointest Liver Physiol 280:G58-67.

Fang R, Santiago NA, Olds LC, Sibley E (2000) The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. Gastroenterology 118:115-27.

Fitzgerald K, Bazar L, Avigan MI (1998) GATA-6 stimulates a cell line-specific activation element in the human lactase promoter. Am J Physiol 274:G314-24.

Flatz G (1987) Genetics of lactose digestion in humans. Adv Hum Genet 16:1-77

Flatz G, Rotthauwe HW (1977) The human lactase polymorphism: physiology and genetics of lactose absorption and malabsorption. Prog Med Genet 2:205-49

Flatz G, Saengudom C (1969) Lactose tolerance in Asians: a family study. Nature 224:915-6.

Freeman TC, Collins AJ, Heavens RP, Tivey DR (1993) Genetic regulation of enterocyte function: a quantitative in situ hybridisation study of lactase-phlorizin hydrolase and Na(+)-glucose cotransporter mRNAs in rabbit small intestine. Pflugers Arch 422:570-6

Freiburghaus AU, Schmitz J, Schindler M, Rotthauwe HW, Kuitunen P, Launiala K, Hadorn B (1976) Protein patterns of brush-border fragments in congenital lactose malabsorption and in specific hypolactasia of the adult. N Engl J Med 294:1030-2

Freund JN, Domon-Dell C, Kedinger M, Duluc I (1998) The Cdx-1 and Cdx-2 homeobox genes in the intestine. Biochem Cell Biol 76:957-69

Freund JN, Jost B, Duluc I, Morel G (1995) Ultrastructural study of intestinal lactase gene expression. Biol Cell 83:211-7

Frischmeyer PA, Dietz HC (1999) Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet. 8:1893-900

Genkinger JM, Hunter DJ, Spiegelman D, Anderson KE, Arslan A, Beeson WL, Buring JE, et al. (2006) Dairy products and ovarian cancer: a pooled analysis of 12 cohort studies. Cancer Epidemiol Biomarkers Prev 15:364-72

Goodman MT, Wu AH, Tung KH, McDuffie K, Cramer DW, Wilkens LR, Terada K, Reichardt JK, Ng WG (2002a) Association of galactose-1-phosphate uridyltransferase activity and N314D genotype with the risk of ovarian cancer. Am J Epidemiol 156:693-701

Goodman MT, Wu AH, Tung KH, McDuffie K, Kolonel LN, Nomura AM, Terada K, Wilkens LR, Murphy S, Hankin JH (2002b) Association of dairy products, lactose, and calcium with the risk of ovarian cancer. Am J Epidemiol 156:148-57

Gudmand-Höyer E, Dahlqvist A, Jarnum S (1969) Specific small-intestinal lactase deficiency in adults. Scand J Gastroenterol 4:377-86

Gueguen L, Pointillart A (2000) The bioavailability of dietary calcium. J Am Coll Nutr 19:119S-136S

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, et al. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234-8.

Guz Y, Montminy MR, Stein R, Leonard J, Gamer LW, Wright CV, Teitelman G (1995) Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. Development 121:11-8

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

Haenszel W, Kurihara M (1968) Studies of Japanese migrants. I. Mortality from cancer and other diseases among Japanese in the United States. J Natl Cancer Inst 40:43-68

Haff LA, Smirnov IP (1997) Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res 7:378-88

Hallman N, Hjelt L, Ahvenainen EK (1956) Nephrotic syndrome in newborns and young infants. Ann Paediatr Fenn 2:227-241

Hankinson SE, Hunter DJ, Colditz GA, Willett WC, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE (1993) Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. Jama 270:2813-8

Harlow BL, Cramer DW, Geller J, Willett WC, Bell DA, Welch WR (1991) The influence of lactose consumption on the association of oral contraceptive use and ovarian cancer risk. Am J Epidemiol 134:445-53

Harvey CB, Fox MF, Jeggo PA, Mantei N, Povey S, Swallow DM (1993) Regional localization of the lactase-phlorizin hydrolase gene, LCT, to chromosome 2q21. Ann Hum Genet 57:179-85.

Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH, Cooper BT, Barton R, Sarner M, Korpela R, Swallow DM (1998) Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. Ann Hum Genet 62:215-23.

Harvey CB, Pratt WS, Islam I, Whitehouse DB, Swallow DM (1995) DNA polymorphisms in the lactase gene. Linkage disequilibrium across the 70-kb region. Eur J Hum Genet 3:27-41

Harvey CB, Wang Y, Darmoul D, Phillips A, Mantei N, Swallow DM (1996) Characterisation of a human homologue of a yeast cell division cycle gene, MCM6, located adjacent to the 5' end of the lactase gene on chromosome 2q21. FEBS Lett 398:135-40.

Hearne CM, Ghosh S, Todd JA (1992) Microsatellites for linkage analysis of genetic traits. Trends Genet 8:288-94.

Herrinton LJ, Weiss NS, Beresford SA, Stanford JL, Wolfla DM, Feng Z, Scott CR (1995) Lactose and galactose intake and metabolism in relation to the risk of epithelial ovarian cancer. Am J Epidemiol 141:407-16

Heston LL, Gottesman II (1973) The evolution of lactose tolerance. In: Gottesman II, Heston LL (eds) Summary of the conference on lactose and milk intolerance. DHEW Publication No. (OCD), Washington D.C., pp 73-79

Hiekkalinna T, Terwilliger JD, Sammalisto S, Peltonen L, Perola M (2005) AUTOGSCAN: powerful tools for automated genome-wide linkage and linkage disequilibrium analysis. Twin Res Hum Genet 8:16-21

Hogenauer C, Hammer HF, Mellitzer K, Renner W, Krejs GJ, Toplak H (2005) Evaluation of a new DNA test compared with the lactose hydrogen breath test for the diagnosis of lactase non-persistence. Eur J Gastroenterol Hepatol 17:371-6

Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A 88:7276-80.

Hollox E (2005) Evolutionary genetics: genetics of lactase persistence--fresh lessons in the history of milk drinking. Eur J Hum Genet 13:267-9

Hollox EJ, Poulter M, Wang Y, Krause A, Swallow DM (1999) Common polymorphism in a highly variable region upstream of the human lactase gene affects DNA-protein interactions. Eur J Hum Genet 7:791-800

Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM (2001) Lactase haplotype diversity in the Old World. Am J Hum Genet 68:160-172.

Holzel A (1967) Sugar malabsorption due to deficiencies of disaccharidase activities and of monosaccharide transport. Arch Dis Child 42:341-52

Holzel A, Schwarz V, Sutcliffe KW (1959) Defective lactose absorption causing malnutrition in infancy. Lancet i:1126-1128

Huurre M (1992) Folio 153 Settlement. In: Atlas of Finland. National Board of Survey and Geographical Society of Finland. Map Centre, Helsinki, pp 2-5

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 2:204-11.

Hästbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, et al. (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073-87.

Höglund P, Haila S, Socha J, Tomaszewski L, Saarialho-Kere U, Karjalainen-Lindsberg ML, Airola K, Holmberg C, de la Chapelle A, Kere J (1996) Mutations of the Down-regulated in adenoma (DRA) gene cause congenital chloride diarrhoea. Nat Genet 14:316-9.

Ikonen E, Baumann M, Gron K, Syvänen AC, Enomaa N, Halila R, Aula P, Peltonen L (1991) Aspartylglucosaminuria: cDNA encoding human aspartylglucosaminidase and the missense mutation causing the disease. Embo J 10:51-8.

Isokoski M, Jussila J, Sarna S (1972) A simple screening method for lactose malabsorption. Gastroenterology 62:28-32

Jacob R, Peters K, Naim HY (2002) The prosequence of human lactase-phlorizin hydrolase modulates the folding of the mature enzyme. J Biol Chem 277:8217-25

Jacob R, Radebach I, Wüthrich M, Grünberg J, Sterchi EE, Naim HY (1996) Maturation of human intestinal lactase-phlorizin hydrolase: generation of the brush border form of the enzyme involves at least two proteolytic cleavage steps. Eur J Biochem 236:789-95.

Jacob R, Weiner JR, Stadge S, Naim HY (2000) Additional N-glycosylation and its impact on the folding of intestinal lactase-phlorizin hydrolase. J Biol Chem 275:10630-7.

Joensuu T, Hämäläinen R, Yuan B, Johnson C, Tegelberg S, Gasparini P, Zelante L, Pirvola U, Pakarinen L, Lehesjoki AE, de la Chapelle A, Sankila EM (2001) Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3. Am J Hum Genet 69:673-84.

Jokipii M (1992) Folio 153 Settlement. In: Atlas of Finland. National Board of Survey and Geographical Society of Finland. Map Centre, Helsinki, pp 5-9

Jokipii M, Rikkinen K (1992) Folio 153 Settlement. In: Atlas of Finland. National Board of Survey and Geographical Society of Finland. Map Centre, Helsinki, pp 9-11

Jussila J (1969) Diagnosis of lactose malabsorption by the lactose tolerance test with peroral ethanol administration. Scand J Gastroenterol 4:361-8

Jussila J, Isokoski M, Launiala K (1970) Prevalence of lactose malabsorption in a Finnish rural population. Scand J Gastroenterol 5:49-56

Jutikkala E, Pirinen K (1996) History of Finland. WSOY, Juva

Järvelä I, Enattah NS, 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 63:1078-85.

Karttunen L, Lönnqvist L, Godfrey M, Peltonen L, Syvänen AC (1996) An accurate method for comparing transcript levels of two alleles or highly homologous genes: application to fibrillin transcripts in Marfan patients' fibroblasts. Genome Res 6:392-403.

Kaufman FR, Kogut MD, Donnell GN, Goebelsmann U, March C, Koch R (1981) Hypergonadotropic hypogonadism in female patients with galactosemia. N Engl J Med 304:994-8

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 1:575-82.

Keusch GT, Troncale FJ, Miller LH, Promadhat V, Anderson PR (1969a) Acquired lactose malabsorption in Thai children. Pediatrics 43:540-5.

Keusch GT, Troncale FJ, Thavaramara B, Prinyanont P, Anderson PR, Bhamarapravathi N (1969b) Lactase deficiency in Thailand: effect of prolonged lactose feeding. Am J Clin Nutr 22:638-41

Kiani F, Knutsen S, Singh P, Ursin G, Fraser G (2006) Dietary risk factors for ovarian cancer: the Adventist Health Study (United States). Cancer Causes Control 17:137-46

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171-9.

Klein RM (1989) Small intestinal cell proliferation during development. In: Lebenthal E (ed) Human Gastrointestinal Development. Raven Press, New York, pp 367-392.

Kolehmainen J, Black GC, Saarinen A, Chandler K, Clayton-Smith J, Traskelin AL, Perveen R, Kivitie-Kallio S, Norio R, Warburg M, Fryns JP, de la Chapelle A, Lehesjoki AE (2003) Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. Am J Hum Genet 72:1359-69

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

Krasinski SD, Estrada G, Yeh KY, Yeh M, Traber PG, Rings EH, Buller HA, Verhave M, Montgomery RK, Grand RJ (1994) Transcriptional regulation of intestinal hydrolase biosynthesis during postnatal development in rats. Am J Physiol 267:G584-94

Krasinski SD, Upchurch BH, Irons SJ, June RM, Mishra K, Grand RJ, Verhave M (1997) Rat lactase-phlorizin hydrolase/human growth hormone transgene is expressed on small intestinal villi in transgenic mice. Gastroenterology 113:844-55

Krasinski SD, Van Wering HM, Tannemaat MR, Grand RJ (2001) Differential activation of intestinal gene promoters: functional interactions between GATA-5 and HNF-1 alpha. Am J Physiol Gastrointest Liver Physiol 281:G69-84.

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

Kruse TA, Bolund L, Grzeschik KH, Ropers HH, Sjöström H, Norén O, Mantei N, Semenza G (1988) The human lactase-phlorizin hydrolase gene is located on chromosome 2. FEBS Lett 240:123-6.

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 90:160-4.

Kushi LH, Mink PJ, Folsom AR, Anderson KE, Zheng W, Lazovich D, Sellers TA (1999) Prospective study of diet and ovarian cancer. Am J Epidemiol 149:21-31

Kyttälä M, Tallila J, Salonen R, Kopra O, Kohlschmidt N, Paavola-Sakki P, Peltonen L, Kestilä M (2006) MKS1, encoding a component of the flagellar apparatus basal body proteome, is mutated in Meckel syndrome. Nat Genet 38:155-7

Laan M, Pääbo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435-8.

Lacey SW, Naim HY, Magness RR, Gething MJ, Sambrook JF (1994) Expression of lactase-phlorizin hydrolase in sheep is regulated at the RNA level. Biochem J 302 (Pt 3):929-35

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus ML (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58:1309-22.

Lander ES (1996) The new genomics: global views of biology. Science 274:536-9.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921.

Langley SD, Lai K, Dembure PP, Hjelm LN, Elsas LJ (1997) Molecular basis for Duarte and Los Angeles variant galactosemia. Am J Hum Genet 60:366-72

Larsson SC, Bergkvist L, Wolk A (2004) Milk and lactose intakes and ovarian cancer risk in the Swedish Mammography Cohort. Am J Clin Nutr 80:1353-7

Larsson SC, Orsini N, Wolk A (2006) Milk, milk products and lactose intake and ovarian cancer risk: a meta-analysis of epidemiological studies. Int J Cancer 118:431-41

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

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci U S A 81:3443-6.

Launiala K, Kuitunen P, Visakorpi JK (1966) Disaccharidases and histology of duodenal mucosa in congenital lactose malabsorption. Acta Paediatr Scand 55:257-63.

Lee SY, Wang Z, Lin CK, Contag CH, Olds LC, Cooper AD, Sibley E (2002) Regulation of intestine-specific spatiotemporal expression by the rat lactase promoter. J Biol Chem 277:13099-105

Leese HJ, Semenza G (1973) On the identity between the small intestinal enzymes phlorizin hydrolase and glycosylceramidase. J Biol Chem 248:8170-3

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

Lehtivirta J, Seurujärvi-Kari I (1991) Saamelaiset. In: Laakso J (ed) Uralilaiset kansat: tietoa suomen sukukielistä ja niiden puhujista. WSOY, Helsinki, pp 123-155

Leslie ND (2003) Insights into the pathogenesis of galactosemia. Annu Rev Nutr 23:59-80

Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, Troelsen JT (2005) T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum Mol Genet 14:3945-53

Levy E, Haltia M, Fernandez-Madrid I, Koivunen O, Ghiso J, Prelli F, Frangione B (1990) Mutation in gelsolin gene in Finnish hereditary amyloidosis. J Exp Med 172:1865-7.

Li WH, Sadler LA (1991) Low nucleotide diversity in man. Genetics 129:513-23.

Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4:357-62.

Lloyd M, Mevissen G, Fischer M, Olsen W, Goodspeed D, Genini M, Boll W, Semenza G, Mantei N (1992) Regulation of intestinal lactase in adult hypolactasia. J Clin Invest 89:524-9.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 14:1675-80.

Maiuri L, Rossi M, Raia V, Garipoli V, Hughes LA, Swallow D, Noren O, Sjostrom H, Auricchio S (1994) Mosaic regulation of lactase in human adult-type hypolactasia. Gastroenterology 107:54-60

Malabanan A, Veronikis IE, Holick MF (1998) Redefining vitamin D insufficiency. Lancet 351:805-6

Mantei N, Villa M, Enzler T, Wacker H, Boll W, James P, Hunziker W, Semenza G (1988) Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. Embo J 7:2705-13.

Maquat LE (2005) Nonsense-mediated mRNA decay in mammals. J Cell Sci 118:1773-6

Markkanen M, Isaksson R, Norio R (1987) Erikoista suvussamme - Suomalainen tautiperintö lyhyesti. Duodecim 103:1434-1463

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 276:75-7.

McCarty MF (2000) Parathyroid hormone may be a cancer promoter - an explanation for the decrease in cancer risk associated with ultraviolet light, calcium, and vitamin D. Med Hypotheses 54:475-82

McCracken RD (1970) Adult lactose tolerance. Jama 213:2257-60.

McCracken RD (1971) Origins and implications of the distribution of adult lactase deficiency in human populations. J Trop Pediatr Environ Child Health 17:7-10.

Mee L, Honkala H, Kopra O, Vesa J, Finnila S, Visapaa I, Sang TK, Jackson GR, Salonen R, Kestila M, Peltonen L (2005) Hydrolethalus syndrome is caused by a missense mutation in a novel gene HYLS1. Hum Mol Genet 14:1475-88

Meloni GF, Colombo C, La Vecchia C, Pacifico A, Tomasi P, Ogana A, Marinaro AM, Meloni T (2001) High prevalence of lactose absorbers in Northern Sardinian patients with type 1 and type 2 diabetes mellitus. Am J Clin Nutr 73:582-5

Meloni GF, Colombo C, La Vecchia C, Ruggiu G, Mannazzu MC, Ambrosini G, Cherchi PL (1999) Lactose absorption in patients with ovarian cancer. Am J Epidemiol 150:183-6

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet. 2004 36:1073-8

Menkiszak J, Gronwald J, Gorski B, Jakubowska A, Huzarski T, Byrski T, Foszczynska-Kloda M, et al. (2003) Hereditary ovarian cancer in Poland. Int J Cancer 106:942-5

Mettlin CJ, Piver MS (1990) A case-control study of milk-drinking and ovarian cancer risk. Am J Epidemiol 132:871-6

Metz G, Blendis LM, Jenkins DJ (1976) Letter: H2 breath test for lactase deficiency. N Engl J Med 294:730

Metz G, Jenkins DJ, Peters TJ, Newman A, Blendis LM (1975) Breath hydrogen as a diagnostic method for hypolactasia. Lancet 1:1155-7

Miettinen S, Ahonen MH, Lou YR, Manninen T, Tuohimaa P, Syvälä H, Ylikomi T (2004) Role of 24-hydroxylase in vitamin D3 growth response of OVCAR-3 ovarian cancer cells. Int J Cancer 108:367-73

Mishkin B, Yalovsky M, Mishkin S (1997) Increased prevalence of lactose malabsorption in Crohn's disease patients at low risk for lactose malabsorption based on ethnic origin. Am J Gastroenterol 92:1148-53

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

Mitchelmore C, Troelsen JT, Sjostrom H, Noren O (1998) The HOXC11 homeodomain protein interacts with the lactase-phlorizin hydrolase promoter and stimulates HNF1alpha-dependent transcription. J Biol Chem 273:13297-306

Mitchelmore C, Troelsen JT, Spodsberg N, Sjöström H, Norén O (2000) Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. Biochem J 346 Pt 2:529-35.

Molkentin JD (2000) The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. J Biol Chem 275:38949-52

Mommers M, Schouten LJ, Goldbohm RA, van den Brandt PA (2006) Dairy consumption and ovarian cancer risk in the Netherlands Cohort Study on Diet and Cancer. Br J Cancer 94:165-70

Montgomery RK, Buller HA, Rings EH, Grand RJ (1991) Lactose intolerance and the genetic regulation of intestinal lactase-phlorizin hydrolase. Faseb J 5:2824-32

Morland SJ, Jiang X, Hitchcock A, Thomas EJ, Campbell IG (1998) Mutation of galactose-1-phosphate uridyl transferase and its association with ovarian cancer and endometriosis. Int J Cancer 77:825-7

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

Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DM, Bradman N, Thomas MG (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. Am J Hum Genet 74:1102-10

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

Myles S, Bouzekri N, Haverfield E, Cherkaoui M, Dugoujon JM, Ward R (2005) Genetic evidence in support of a shared Eurasian-North African dairying origin. Hum Genet 117:34-42

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 63:506-16.

Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, Krohn KJ, Lalioti MD, Mullis PE, Antonarakis SE, Kawasaki K, Asakawa S, Ito F, Shimizu N (1997) Positional cloning of the APECED gene. Nat Genet 17:393-8.

Naim HY (1994) Processing and transport of human small intestinal lactase-phlorizin hydrolase (LPH). Role of N-linked oligosaccharide modification. FEBS Lett 342:302-7

Naim HY (1995) The pro-region of human intestinal lactase-phlorizin hydrolase is enzymatically inactive towards lactose. Biol Chem Hoppe Seyler 376:255-8

Naim HY (2001) Molecular and cellular aspects and regulation of intestinal lactase- phlorizin hydrolase. Histol Histopathol 16:553-61.

Naim HY, Jacob R, Naim H, Sambrook JF, Gething MJ (1994) The pro region of human intestinal lactase-phlorizin hydrolase. J Biol Chem 269:26933-43

Naim HY, Lacey SW, Sambrook JF, Gething MJ (1991) Expression of a full-length cDNA coding for human intestinal lactase- phlorizin hydrolase reveals an uncleaved, enzymatically active, and transport-competent protein. J Biol Chem 266:12313-20.

Naim HY, Lentze MJ (1992) Impact of O-glycosylation on the function of human intestinal lactase-phlorizin hydrolase. Characterization of glycoforms varying in enzyme activity and localization of O-glycoside addition. J Biol Chem 267:25494-504

Naim HY, Naim H (1996) Dimerization of lactase-phlorizin hydrolase occurs in the endoplasmic reticulum, involves the putative membrane spanning domain and is required for an efficient transport of the enzyme to the cell surface. Eur J Cell Biol 70:198-208.

National Nutrition Council (1999) Finnish Nutrition Recommendations.32

Neele AM, Einerhand AW, Dekker J, Buller HA, Freund JN, Verhave M, Grand RJ, Montgomery RK (1995) Verification of the lactase site of rat lactase-phlorizin hydrolase by site-directed mutagenesis. Gastroenterology 109:1234-40

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

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 56:1088-95.

Nilsson TK, Johansson CA (2004) A novel method for diagnosis of adult hypolactasia by genotyping of the -13910 C/T polymorphism with Pyrosequencing technology. Scand J Gastroenterol 39:287-90

Norat T, Riboli E (2003) Dairy products and colorectal cancer. A review of possible mechanisms and epidemiological evidence. Eur J Clin Nutr 57:1-17

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

Norio R (1981) Diseases of Finland and Scandinavia. In: Rothschild HR (ed) Biocultural aspects of disease. Academic Press, New York, pp 359-414

Norio R (2000) Suomi-neidon geenit - tautiperinnön takana juurillemme johtamassa. Otava, Keuruu

Norio R (2003a) Finnish Disease Heritage I: characteristics, causes, background. Hum Genet

Norio R (2003b) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet

Norio R (2003c) The Finnish disease heritage III: the individual diseases. Hum Genet

Norio R, Nevanlinna HR, Perheentupa J (1973) Hereditary diseases in Finland; rare flora in rare soul. Ann Clin Res 5:109-41.

Nunez MG (1987) A Model for the early settlement of Finland. Fennoscandia archaeologica IV:3-18

Oberholzer T, Mantei N, Semenza G (1993) The pro sequence of lactase-phlorizin hydrolase is required for the enzyme to reach the plasma membrane. An intramolecular chaperone? FEBS Lett 333:127-31

Obermayer-Pietsch BM, Bonelli CM, Walter DE, Kuhn RJ, Fahrleitner-Pammer A, Berghold A, Goessler W, Stepan V, Dobnig H, Leb G, Renner W (2004) Genetic predisposition for adult lactose intolerance and relation to diet, bone density, and bone fractures. J Bone Miner Res 19:42-7

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259-66

Olatunbosun DA, Kwaku Adadevoh B (1971) Lactase deficiency in Nigerians. Am J Dig Dis 16:909-14.

Olds LC, Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. Hum Mol Genet 12:2333-40

Orkin SH (1998) Embryonic stem cells and transgenic mice in the study of hematopoiesis. Int J Dev Biol 42:927-34

Ott J (1999) Analysis of Human Genetic Linkage. Johns Hopkins University Press, Baltimore

Ott MO, Rey-Campos J, Cereghini S, Yaniv M (1991) vHNF1 is expressed in epithelial cells of distinct embryonic origin during development and precedes HNF1 expression. Mech Dev 36:47-58

Paloneva J, Kestilä M, Wu J, Salminen A, Bohling T, Ruotsalainen V, Hakola P, Bakker AB, Phillips JH, Pekkarinen P, Lanier LL, Timonen T, Peltonen L (2000) Loss-of-function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. Nat Genet 25:357-61.

Paloneva J, Manninen T, Christman G, Hovanes K, Mandelin J, Adolfsson R, Bianchin M, Bird T, Miranda R, Salmaggi A, Tranebjaerg L, Konttinen Y, Peltonen L (2002) Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. Am J Hum Genet 71:656-62.

Pan SY, Ugnat AM, Mao Y, Wen SW, Johnson KC (2004) A case-control study of diet and the risk of ovarian cancer. Cancer Epidemiol Biomarkers Prev 13:1521-7

Panzer P, Preuss U, Joberty G, Naim HY (1998) Protein domains implicated in intracellular transport and sorting of lactase-phlorizin hydrolase. J Biol Chem 273:13861-9

Parazzini F, Chiaffarino F, Negri E, Surace M, Benzi G, Franceschi S, Fedele L, La Vecchia C (2004) Risk factors for different histological types of ovarian cancer. Int J Gynecol Cancer 14:431-6

Park RH, Duncan A, Russell RI (1990) Hypolactasia and Crohn's disease: a myth. Am J Gastroenterol 85:708-10

Parkin DM, Bray F, Ferlay J, Pisani P (2001) Estimating the world cancer burden: Globocan 2000. Int J Cancer 94:153-6

Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvänen AC (1997) Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Res 7:606-14.

Pastinen T, Partanen J, Syvänen AC (1996) Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. Clin Chem 42:1391-7.

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 25:91-5.

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

Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, D'Amato E, Ramirez L, Faham M, Koskiniemi M, Warrington JA, Norio R, de la Chapelle A, Cox DR, Myers RM (1996) Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). Science 271:1731-4.

Phillips AD, Smith MW, Walker-Smith JA (1988) Selective alteration of brush-border hydrolases in intestinal diseases in childhood. Clin Sci (Lond) 74:193-200

Poggi V, Sebastio G (1991) Molecular analysis of the lactase gene in the congenital lactase deficiency (abstract). Am J Hum Genet 49:105

Pontoglio M (2000) Hepatocyte nuclear factor 1, a transcription factor at the crossroads of glucose homeostasis. J Am Soc Nephrol 11 Suppl 16:S140-3

Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, Kajander K, Sarner M, Korpela R, Swallow DM (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. Ann Hum Genet 67:298-311

Prentice A (2004) Diet, nutrition and the prevention of osteoporosis. Public Health Nutr 7:227-43

Qin LQ, Xu JY, Wang PY, Hashi A, Hoshi K, Sato A (2005) Milk/dairy products consumption, galactose metabolism and ovarian cancer: meta-analysis of epidemiological studies. Eur J Cancer Prev 14:13-9

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 AE (1999) The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. Nat Genet 23:233-6.

Rasinperä H, Forsblom C, Enattah NS, Halonen P, Salo K, Victorzon M, Mecklin JP, Järvinen H, Enholm S, Sellick G, Alazzouzi H, Houlston R, Robinson J, Groop PH, Tomlinson I, Schwartz S, Jr., Aaltonen LA, Järvelä I (2005) The C/C-13910 genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in the Finnish population. Gut 54:643-7

Rasinperä H, Saarinen K, Pelkonen A, Järvelä I, Savilahti E, Kolho KL (2006) Molecularly defined adult-type hypolactasia in school-aged children with a previous history of cow's milk allergy. World J Gastroenterol 12:2264-8

Rasinperä H, Savilahti E, Enattah NS, Kuokkanen M, Tötterman N, Lindahl H, Järvelä I, Kolho KL (2004) A genetic test which can be used to diagnose adult-type hypolactasia in children. Gut 53:1571-6

Reichardt JK, Woo SL (1991) Molecular basis of galactosemia: mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase. Proc Natl Acad Sci U S A 88:2633-7

Remenyi A, Scholer HR, Wilmanns M (2004) Combinatorial control of gene expression. Nat Struct Mol Biol 11:812-5

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

Ridefelt P, Håkansson LD (2005) Lactose intolerance: lactose tolerance test versus genotyping. Scand J Gastroenterol 40:822-6

Risch HA (1998) Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. J Natl Cancer Inst 90:1774-86

Risch HA, Jain M, Marrett LD, Howe GR (1994) Dietary lactose intake, lactose intolerance, and the risk of epithelial ovarian cancer in southern Ontario (Canada). Cancer Causes Control 5:540-8

Ropponen KM, Kellokoski JK, Pirinen RT, Moisio KI, Eskelinen MJ, Alhava EM, Kosma VM (2001) Expression of transcription factor AP-2 in colorectal adenomas and adenocarcinomas; comparison of immunohistochemistry and in situ hybridisation. J Clin Pathol 54:533-8.

Ross P, Hall L, Smirnov I, Haff L (1998) High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol 16:1347-51

Rossi M, Maiuri L, Fusco MI, Salvati VM, Fuccio A, Auricchio S, Mantei N, Zecca L, Gloor SM, Semenza G (1997) Lactase persistence versus decline in human adults: multifactorial events are involved in down-regulation after weaning. Gastroenterology 112:1506-14

Runnebaum IB, Stickeler E (2001) Epidemiological and molecular aspects of ovarian cancer risk. J Cancer Res Clin Oncol 127:73-9

Saarela T, Simila S, Koivisto M (1995) Hypercalcemia and nephrocalcinosis in patients with congenital lactase deficiency. J Pediatr 127:920-3

Sahi T (1974) Lactose malabsorption in Finnish-speaking and Swedish-speaking populations in Finland. Scand J Gastroenterol 9:303-8

Sahi T (1994a) Genetics and epidemiology of adult-type hypolactasia. Scand J Gastroenterol Suppl 202:7-20

Sahi T (1994b) Hypolactasia and lactase persistence. Historical review and the terminology. Scand J Gastroenterol Suppl 202:1-6

Sahi T, Isokoski M, Jussila J, Launiala K, Pyörälä K (1973) Recessive inheritance of adult-type lactose malabsorption. Lancet 2:823-6.

Sahi T, Launiala K (1978) Manifestation and occurrence of selective adult-type lactose malabsorption in Finnish teenagers. A follow-up study. Am J Dig Dis 23:699-704.

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

Sajantila A, Pääbo S (1995) Language replacement in Scandinavia. Nat Genet 11:359-60.

Salazar-Martinez E, Lazcano-Ponce EC, Gonzalez Lira-Lira G, Escudero-De los Rios P, Hernandez-Avila M (2002) Nutritional determinants of epithelial ovarian cancer risk: a case-control study in Mexico. Oncology 63:151-7

Salonen R, Somer M, Haltia M, Lorentz M, Norio R (1991) Progressive encephalopathy with edema, hypsarrhythmia, and optic atrophy (PEHO syndrome). Clin Genet 39:287-93.

Sankila EM, 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 1:109-13.

Savilahti E, Launiala K, Kuitunen P (1983) Congenital lactase deficiency. A clinical study on 16 patients. Arch Dis Child 58:246-52.

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 19:286-8.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-70.

Schulz M, Lahmann PH, Riboli E, Boeing H (2004) Dietary determinants of epithelial ovarian cancer: a review of the epidemiologic literature. Nutr Cancer 50:120-40

Schwartz MZ, Storozuk RB (1988) Influence of epidermal growth factor on intestinal function in the rat: comparison of systemic infusion versus luminal perfusion. Am J Surg 155:18-22

Sebastio G, Villa M, Sartorio R, Guzzetta V, Poggi V, Auricchio S, Boll W, Mantei N, Semenza G (1989) Control of lactase in human adult-type hypolactasia and in weaning rabbits and rats. Am J Hum Genet 45:489-97.

Siitonen HA, Kopra O, Kääriainen H, Haravuori H, Winter RM, Saamanen AM, Peltonen L, Kestilä M (2003) Molecular defect of RAPADILINO syndrome expands the phenotype spectrum of RECQL diseases. Hum Mol Genet 12:2837-44

Simon TC, Roberts LJ, Gordon JI (1995) A 20-nucleotide element in the intestinal fatty acid binding protein gene modulates its cell lineage-specific, differentiation-dependent, and cephalocaudal patterns of expression in transgenic mice. Proc Natl Acad Sci U S A 92:8685-9

Simoons FJ (1969) Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. I. Review of the medical research. Am J Dig Dis 14:819-36.

Simoons FJ (1970) Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. II. A culture historical hypothesis. Am J Dig Dis 15:695-710.

Simoons FJ (1978) The geographic hypothesis and lactose malabsorption. A weighing of the evidence. Am J Dig Dis 23:963-80.

Skovbjerg H, Sjostrom H, Noren O (1981) Purification and characterisation of amphiphilic lactase/phlorizin hydrolase from human small intestine. Eur J Biochem 114:653-61

Socha J, Ksiazyk J, Flatz G, Flatz SD (1984) Prevalence of primary adult lactose malabsorption in Poland. Ann Hum Biol 11:311-6

Spodsberg N, Troelsen JT, Carlsson P, Enerbäck S, Sjöström H, Norén O (1999) Transcriptional regulation of pig lactase-phlorizin hydrolase: involvement of HNF-1 and FREACs. Gastroenterology 116:842-54.

Stenman J, Finne P, Stahls A, Grenman R, Stenman UH, Palotie A, Orpana A (1999) Accurate determination of relative messenger RNA levels by RT-PCR. Nat Biotechnol 17:720-2.

Sterchi EE, Mills PR, Fransen JA, Hauri HP, Lentze MJ, Naim HY, Ginsel L, Bond J (1990) Biogenesis of intestinal lactase-phlorizin hydrolase in adults with lactose intolerance. Evidence for reduced biosynthesis and slowed-down maturation in enterocytes. J Clin Invest 86:1329-37.

Swallow DM (2006) DNA test for hypolactasia premature. Gut 55:131; author reply 131-2

Swallow DM, Hollox EJ (2000) The genetic polymorphism of intestinal lactase activity in adult humans. In: Scriver CR, Beaudet AL, Sly SS, Valle D (eds) The Metabolic and Molecular Basis of Inherited Disease. McGraw-Hill, pp 1562-1651.

Sykes DE, Weiser MM (1995) Rat intestinal crypt-cell replication factor with homology to early S- phase proteins required for cell division. Gene 163:243-7.

Syvänen AC (1998) Solid-phase minisequencing as a tool to detect DNA polymorphism. Methods Mol Biol 98:291-8

Syvänen AC, Aalto-Setälä K, Harju L, Kontula K, Söderlund H (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 8:684-92.

Syvänen AC, Ikonen E, Manninen T, Bengtström M, Söderlund H, Aula P, Peltonen L (1992a) Convenient and quantitative determination of the frequency of a mutant allele using solid-phase minisequencing: application to aspartylglucosaminuria in Finland. Genomics 12:590-5.

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

Syvänen AC, Söderlund H, Laaksonen E, Bengtström M, Turunen M, Palotie A (1992b) N-ras gene mutations in acute myeloid leukemia: accurate detection by solid-phase minisequencing. Int J Cancer 50:713-8.

Takahashi K, Yamada H, Yanagida M (1994) Fission yeast minichromosome loss mutants mis cause lethal aneuploidy and replication abnormality. Mol Biol Cell 5:1145-58

Takeuchi A, Okano T, Ishida Y, Kobayashi T (1995) Effects of dietary vitamin D intake on plasma levels of parathyroid hormone and vitamin D metabolites in healthy Japanese. Miner Electrolyte Metab 21:217-22

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, Tolk HV, et al. (2004) The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am J Hum Genet 74:661-82

Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Köster H (1999) Chip-based genotyping by mass spectrometry. Proc Natl Acad Sci U S A 96:10016-20

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 56:777-87.

Terwilliger JD, Zollner S, Laan M, Pääbo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. Hum Hered 48:138-54.

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

The International Batten Disease Consortium (1995) Isolation of a novel gene underlying Batten disease, CLN3. Cell 82:949-57.

The Retinoschisis Consortium (1998) Functional implications of the spectrum of mutations found in 234 cases with X-linked juvenile retinoschisis. Hum Mol Genet 7:1185-92.

The World Dairy Situation (2003) Bulletin of the International Dairy Federation 384/2003.

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 21:293-6.

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonne-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137-52.

Troelsen JT (2005) Adult-type hypolactasia and regulation of lactase expression. Biochim Biophys Acta 1723:19-32

Troelsen JT, Mitchelmore C, Olsen J (2003a) An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. Gene 305:101-11

Troelsen JT, Mitchelmore C, Spodsberg N, Jensen AM, Norén O, Sjöström H (1997) Regulation of lactase-phlorizin hydrolase gene expression by the caudal- related homoeodomain protein Cdx-2. Biochem J 322:833-8.

Troelsen JT, Olsen J, Moller J, Sjostrom H (2003b) An upstream polymorphism associated with lactase persistence has increased enhancer activity. Gastroenterology 125:1686-94

Troelsen JT, Olsen J, Noren O, Sjostrom H (1992) A novel intestinal trans-factor (NF-LPH1) interacts with the lactase-phlorizin hydrolase promoter and co-varies with the enzymatic activity. J Biol Chem 267:20407-11

Tzonou A, Hsieh CC, Polychronopoulou A, Kaprinis G, Toupadaki N, Trichopoulou A, Karakatsani A, Trichopoulos D (1993) Diet and ovarian cancer: a case-control study in Greece. Int J Cancer 55:411-4

Wacker H, Keller P, Falchetto R, Legler G, Semenza G (1992) Location of the two catalytic sites in intestinal lactase-phlorizin hydrolase. Comparison with sucrase-isomaltase and with other glycosidases, the membrane anchor of lactase-phlorizin hydrolase. J Biol Chem 267:18744-52

Wagner E, Lykke-Andersen J (2002) mRNA surveillance: the perfect persist. J Cell Sci. 115:3033-8

van Noorden CJ (2002) Direct comparison of enzyme histochemical and immunohistochemical methods to localize an enzyme. Mar Environ Res 54:575-7.

van Wering HM, Bosse T, Musters A, de Jong E, de Jong N, Hogen Esch CE, Boudreau F, Swain GP, Dowling LN, Montgomery RK, Grand RJ, Krasinski SD (2004) Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. Am J Physiol Gastrointest Liver Physiol 287:G899-909

van Wering HM, Huibregtse IL, van der Zwan SM, de Bie MS, Dowling LN, Boudreau F, Rings EH, Grand RJ, Krasinski SD (2002a) Physical interaction between GATA-5 and hepatocyte nuclear factor-1alpha results in synergistic activation of the human lactase-phlorizin hydrolase promoter. J Biol Chem 277:27659-67

van Wering HM, Moyer L, Grand RJ, Krasinski SD (2002b) Novel interaction at the Cdx-2 binding sites of the lactase-phlorizin hydrolase promoter. Biochem Biophys Res Commun 299:587-93

Wang AM, Doyle MV, Mark DF (1989) Quantitation of mRNA by the polymerase chain reaction. Proc Natl Acad Sci U S A 86:9717-21.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, et al. (1998a) Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. Science 280:1077-82.

Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, Clay P, Walker-Smith JA, Swallow DM (1998b) The genetically programmed down-regulation of lactase in children. Gastroenterology 114:1230-6.

Wang Y, Harvey CB, Pratt WS, Sams VR, Sarner M, Rossi M, Auricchio S, Swallow DM (1995) The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. Hum Mol Genet 4:657-62.

Wang Z, Fang R, Olds LC, Sibley E (2004) Transcriptional regulation of the lactase-phlorizin hydrolase promoter by PDX-1. Am J Physiol Gastrointest Liver Physiol 287:G555-61

Wang Z, Maravelias C, Sibley E (2006) Lactase gene promoter fragments mediate differential spatial and temporal expression patterns in transgenic mice. DNA Cell Biol 25:215-22

Varilo T (1999) The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. Thesis, University of Helsinki, Dep. of Medical Genetics and National Public Health Institute, Dep. of Human Molecular Genetics, Helsinki

Varilo T, Nikali K, Suomalainen A, Lönnqvist T, Peltonen L (1996) Tracing an ancestral mutation: genealogical and haplotype analysis of the infantile onset spinocerebellar ataxia locus. Genome Res 6:870-5

Webb PM, Bain CJ, Purdie DM, Harvey PW, Green A (1998) Milk consumption, galactose metabolism and ovarian cancer (Australia). Cancer Causes Control 9:637-44

Weber H, Strandmann EP, Holewa B, Bartkowski S, Zapp D, Zoidl C, Ryffel GU (1996) Regulation and function of the tissue-specific transcription factor HNF1 alpha (LFB1) during Xenopus development. Int J Dev Biol 40:297-304

Wen WH, Reles A, Runnebaum IB, Sullivan-Halley J, Bernstein L, Jones LA, Felix JC, Kreienberg R, el-Naggar A, Press MF (1999) p53 mutations and expression in ovarian cancers: correlation with overall survival. Int J Gynecol Pathol 18:29-41

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. Science 291:1304-51.

Verhave M, Krasinski SD, Christian SI, Van Schaik S, Van Den Brink GR, Doting EM, Maas SM, Wolthers KC, Grand RJ, Montgomery RK (2004) Regulatory regions in the rat lactase-phlorizin hydrolase gene that control cell-specific expression. J Pediatr Gastroenterol Nutr 39:275-85

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 23:462-5.

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 376:584-7.

Whittemore AS, Harris R, Itnyre J (1992) Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. IV. The pathogenesis of epithelial ovarian cancer. Collaborative Ovarian Cancer Group. Am J Epidemiol 136:1212-20

Virtaneva K, D'Amato E, Miao J, Koskiniemi M, Norio R, Avanzini G, Franceschetti S, Michelucci R, Tassinari CA, Omer S, Pennacchio LA, Myers RM, Dieguez-Lucena JL, Krahe R, de la Chapelle A, Lehesjoki AE (1997) Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. Nat Genet 15:393-6.

Visapää I, Fellman V, Vesa J, Dasvarma A, Hutton JL, Kumar V, Payne GS, Makarow M, Van Coster R, Taylor RW, Turnbull DM, Suomalainen A, Peltonen L (2002) GRACILE syndrome, a lethal metabolic disorder with iron overload, is caused by a point mutation in BCS1L. Am J Hum Genet 71:863-76.

Witte J, Lloyd M, Lorenzsonn V, Korsmo H, Olsen W (1990) The biosynthetic basis of adult lactase deficiency. J Clin Invest 86:1338-42.

Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22:130-1, 134-8.

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683-90.

von Tirpitz C, Kohn C, Steinkamp M, Geerling I, Maier V, Moller P, Adler G, Reinshagen M (2002) Lactose intolerance in active Crohn's disease: clinical value of duodenal lactase analysis. J Clin Gastroenterol 34:49-53

World Cancer Research Fund (1997) American Institute for Cancer Research. Ovary, food , nutrition and the prevention of cancer: a global perspective. Washington DC. American Institute for Cancer Research:288-293

Vuopala K, Ignatius J, Herva R (1995) Lethal arthrogryposis with anterior horn cell disease. Hum Pathol 26:12-9.

Wüthrich M, Grünberg J, Hahn D, Jacob R, Radebach I, Naim HY, Sterchi EE (1996) Proteolytic processing of human lactase-phlorizin hydrolase is a two- step event: identification of the cleavage sites. Arch Biochem Biophys 336:27-34.

Xu YK, Ng WG, Kaufman FR, Lobo RA, Donnell GN (1989) Galactose metabolism in human ovarian tissue. Pediatr Res 25:151-5

Yen ML, Yen BL, Bai CH, Lin RS (2003) Risk factors for ovarian cancer in Taiwan: a case-control study in a low-incidence population. Gynecol Oncol 89:318-24

Zabel MD, Wheeler W, Weis JJ, Weis JH (2002) Yin Yang 1, Oct1, and NFAT-4 form repeating, cyclosporin-sensitive regulatory modules within the murine CD21 intronic control region. J Immunol 168:3341-50

Zecca L, Mesonero JE, Stutz A, Poiree JC, Giudicelli J, Cursio R, Gloor SM, Semenza G (1998) Intestinal lactase-phlorizin hydrolase (LPH): the two catalytic sites; the role of the pancreas in pro-LPH maturation. FEBS Lett 435:225-8

Zheng L, Roeder RG, Luo Y (2003) S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. Cell 114:255-66