Jussi Naukkarinen

Molecular Background of Common Dyslipidemias

Publications of the National Public Health Institute A 7/2008

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

Helsinki, Finland 2008

Jussi Naukkarinen

MOLECULAR BACKGROUND OF COMMON DYSLIPIDEMIAS

A C A D E M I C D I S S E R T A T I O N

To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in the Small Hall, University Main Building, on March 14th, at 12 noon.

National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2008

Publications of the National Public Health Institute
KTL A7 / 2008

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 978-951-740-773-1
ISSN 0359-3584
ISBN 978-951-740-774-8 (pdf)
ISSN 1458-6290 (pdf)

Yliopistopaino
Helsinki 2008

Supervised by

Academy Professor Leena Peltonen-Palotie M.D., Ph.D.
National Public Health Institute
Department of Molecular Medicine
and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland
and
The Broad Institute of MIT and Harvard
Boston, MA, USA
and
Welcome Trust Sanger Institute
Hinxton, Cambridge, UK

Reviewed by

Professor Matti J. Tikkanen M.D., Ph.D.
Helsinki University Central Hospital
Department of Medicine
University of Helsinki
Helsinki, Finland

Professor Olli Raitakari M.D., Ph.D.
University of Turku
Department of Clinical Physiology
Turku, Finland

Opponent

Professor Mark I. McCarthy M.D., Ph.D.
Oxford Centre for Diabetes, Endocrinology & Metabolism
Oxford, UK

“Nothing shocks me. I’m a scientist.”

-Indiana Jones

To my family.

Jussi Naukkarinen, Molecular Background of Common Dyslipidemias
Publications of the National Public Health Institute, A7/2008, 103 Pages
ISBN 951-740-773-1; 951-740-774-8 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

ABSTRACT

The leading cause of death in the Western world continues to be coronary heart disease (CHD). At the root of the disease process is dyslipidemia – an aberration in the relevant amounts of circulating blood lipids. Cholesterol builds up in the arterial wall and following rupture of these plaques, myocardial infarction or stroke can occur. Heart disease runs in families and a number of hereditary forms are known. The leading cause of adult dyslipidemia presently however is overweight and obesity. This thesis work presents an investigation of the molecular genetics of common, hereditary dyslipidemia and the tightly related condition of obesity.

Familial combined hyperlipidemia (FCHL) is the most common hereditary dyslipidemia in man with an estimated population prevalence of 1-6%. This complex disease is characterized by elevated levels of serum total cholesterol, triglycerides or both and is observed in about 20% of individuals with premature CHD. Our group identified the disease to be associated with genetic variation in the USF1 transcription factor gene. USF1 has a key role in regulating other genes that control lipid and glucose metabolism as well as the inflammatory response – all central processes in the progression of atherosclerosis and CHD. The first two works of this thesis aimed at understanding how these USF1 variants result in increased disease risk. Among the many, non-coding single-nucleotide polymorphisms (SNPs) that associated with the disease, one was found to have a functional effect. The risk-enhancing allele of this SNP seems to eradicate the ability of the important hormone insulin to induce the expression of USF1 in peripheral tissues. The resultant changes in the expression of numerous USF1 target genes over time probably enhance and accelerate the atherogenic processes.

High-density lipoprotein (HDL) is responsible for the transport of excess cholesterol from peripheral tissues to the liver for excretion in bile. This process is termed the reverse transport of cholesterol. An inverse correlation exists between the levels of HDL-C and the risk for CHD. High levels of HDL-C protect against atherosclerosis and this protection is decreased in patients with low HDL-C levels. The ABCA1 cholesterol transporter is instrumental in the initial transfer of cholesterol from cells to the nascent HDL particle for transport to the liver. In the third work of this thesis we investigated the early steps of the reverse transport of cholesterol in Finnish families with familial low-HDL-C to determine whether a defect there was involved

in the increased risk in these families for developing CHD. The role of ABCA1, critical to this process, was investigated in patients as well as healthy control subjects. Patients with low-HDL-C were found to carry rarer DNA variants (SNPs) significantly more often than healthy subjects and sets of these SNPs were related dose-dependently to HDL-C levels. While patient-derived macrophages were only marginally poorer at the efflux of cholesterol, they exhibited significantly higher ABCA1 expression levels. It follows, that the cholesterol efflux capability of patient-derived cells in proportion to the ABCA1 expression was markedly lower. These results suggest that genetic variation within the *ABCA1* gene and defective function of the protein are potential contributors to the impaired RCT process in Finnish families with low-HDL-C.

Dyslipidemias often represent an outcome of obesity and in the final work of this thesis we wanted to address the metabolic pathways related to acquired obesity. It is recognized that active processes in adipose tissue play an important role in the development of dyslipidemia, insulin resistance and other pathological conditions associated with obesity. To minimize the confounding effects of genetic differences present in most human studies, we investigated a rare collection of identical twins that differed significantly in the amount of body fat. In the obese, but otherwise healthy young adults, several notable changes were observed. In addition to chronic inflammation, the adipose tissue of the obese co-twins was characterized by a marked (47%) decrease in amount of mitochondrial DNA (mtDNA) – a change associated with mitochondrial dysfunction. The catabolism of branched chain amino acids (BCAAs) was identified as the most down-regulated process in the obese co-twins. A concordant increase in the serum level of these insulin secretagogues was identified. This hyperaminoacidemia may provide the feed-back signal from insulin resistant adipose tissue to the pancreas to ensure an appropriately augmented secretory response. The down regulation of BCAA catabolism correlated closely with liver fat accumulation and insulin. The single most up-regulated gene (5.9 fold) in the obese co-twins was osteopontin (*SPP1*) – a cytokine involved in macrophage recruitment to adipose tissue. SPP1 is here implicated as an important player in the development of insulin resistance.

These studies of exceptional study samples provide better understanding of the underlying pathology in common dyslipidemias and other obesity associated diseases –important for future improvement of intervention strategies and treatments to combat atherosclerosis and coronary heart disease.

Keywords: Complex disease genetics, dyslipidemia, cardiovascular disease, obesity, USF1, insulin resistance, FCHL, hypoalphalipoproteinemia

Jussi Naukkarinen, Molecular Background of Common Dyslipidemias
Kansanterveyslaitoksen julkaisuja, A7/2008, 103 sivua
ISBN 951-740-773-1; 951-740-774-8 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

TIIVISTELMÄ

Sydän- ja verisuonitaudit ovat edelleen suurin yksittäinen kuolinsyy länsimaissa. Keskeinen riskitekijä on elimistön häiriintynyt rasva-aineenvaihdunta ja veren epänormaali rasvapitoisuus, dyslipidemia. Kolesterolia kertyy verisuonien seinämiin muodostaen ahtauttavia plakkeja, jotka revetessään voivat aiheuttaa sydäninfarktin. Sydän- ja verisuonitaudit esiintyvät suvuittain, ja tunnetaankin useita periytyviä dyslipidemian muotoja. Yleisin aikuisiän dyslipidemian syy löytyy kuitenkin ylipainosta ja lihavuudesta. Tässä väitöskirjatutkimuksessa on selvitetty yleisten, perinnöllisten dyslipidemioiden molekyyligeneettistä taustaa, ja niihin vahvasti liittyvää lihavuutta.

Familiaalinen kombinoitu hyperlipidemia (FKH) on yleisin perinnöllisistä dyslipidemioista ja sen arvioitu esiintyvyys länsimaiden väestöissä on 1-6%. Tämän alunperin Suomessa kuvatun monitekijäisen taudin taustalla on lukuisten eri geenien ja ympäristön yhteisvaikutus. FKH:n liittyvät kohonneet veren kolesteroli- ja/tai triglyseridiarvot ja se löytyy jopa 20 %:lla varhaisen sydäninfarktin sairastaneista. Tutkimusryhmämme paikallisti taudille altistavan *USF1* geenin, jonka on sittemmin näytetty altistavan sydän- ja verisuonitaudille myös väestötasolla. USF1-geeni koodittaa muita geenejä säätelevää transkriptiotekijää ja sillä on tärkeä rooli elimistön rasva- ja sokeriaineenvaihdunnan sekä tulehdusreaktion säätelyssä. Väitöskirjatutkimuksen kahdessa ensimmäisessä osatyössä selvitettiin FKH:hon liittyvien geenimerkkien merkitystä USF1-geenin toiminnalle. Yhdellä näistä yleisistä emäsparin muutoksista paljastui olevan toiminnallinen merkitys. Riskialleelin kantajilla USF1-geenin normaali vaste elimistön insuliinihormonille vaikuttaa olevan puutteellinen ja tämä heijastuu USF1:n säätelemien geenien muuttuneessa ilmentymisessä. Ajan myötä näiden geenien puutteelliset vasteet todennäköisesti voimistavat ja kiihdyttävät ateroskleroosin kehittymistä.

HDL-hiukkaset kuljettavat ylimääräistä kolesterolia pois kehon soluista maksaan, josta se erittyy sapen mukana ulos. Tätä prosessia kutsutaan kolesterolin käänteiskuljetukseksi. Suurilla veren HDL-kolesterolipitoisuuksilla onkin sydän- ja verisuonitaudilta suojaava vaikutus, kun taas pieni HDL-kolesterolipitoisuus on itsenäinen riskitekijä sydän- ja verisuonitaudin kehittymiselle. Kolesterolin kuljettajaproteiini ABCA1 on tärkeässä asemassa siirtäessään kolestrolia HDL-hiukkaselle käänteiskuljetuksen ensivaiheissa. Kolmannessa osatyössä tutkittiin

tämän ABCA1-geenin ja sen toiminnan osallisuutta taudin kehittymiselle suomalaisissa pienen HDL-kolesterolipitoisuuden suvuissa. Potilaat kantoivat merkittävästi useammin harvinaisia *ABCA1* geenimuotoja kuin terveet verrokit. Potilaiden vaahtosolujen kyky siirtää ABCA1-proteiinin kautta kolesterolia ulos oli vain rajallisesti madaltunut, mutta samanaikaisesti *ABCA1* geenin ilmentyminen oli huomattavasti runsaampi. Täten kolesterolin siirtokyky *suhteessa ABCA1*-geenin ilmentymiseen oli potilailla merkittävästi madaltunut. Tulokset viittaavat siihen, että suomalaisissa pienen HDL-kolesterolipitoisuuden suvuissa häiriö ABCA1:n toiminnassa voi johtaa huonontuneeseen kolesterolin käänteiskuljetukseen ja täten altistaa sydän- ja verisuonitaudille.

Dyslipidemiat ovat usein seurausta lihavuudesta. Rasvakudosta ei enää mielletä vain staattiseksi varastoksi vaan sen tärkeillä endokriinisillä ja metabolisilla tehtävillä ymmärrettään nykyisin olevan vahva yhteys lihavuuden seurauksiin kuten dyslipidemiaan ja insuliiniresistenssiin. Neljännessä osatyössä tutkimmekin kuinka rasvakudoksen erinäiset metaboliset prosessit ovat häiriytyneet lihavuudessa. Geneettisten sekoittavien tekijöiden minimoimiseksi tutkimme harvinaisia, kehon rasvakoostumukseltaan huomattavasti eroavia, mutta perimältään samanlaisia identtisiä kaksosia. Rasvakudoksen geeniprofileissa paljastui useita patologisia muutoksia jo näissä nuorissa lihavissa, mutta muuten terveissä aikuisissa. Selvien tulehdusmerkkien lisäksi lihavien kaksosten rasvakudoksessa havaittiin merkittävä (47%) alentuma mitokondriaalisen DNA:n määrässä –muutos joka kuvastaa mitokondrioiden energiantuotannon aktiivista alasajoa. Suurin lihavuuteen liittyvä muutos oli haaraketjuisten aminohappojen katabolian aleneminen. Lihavien veressä havaittiin vastaavasti näiden insuliinineritystä lisäävien aminohappojen pitoisuuden nousu. Tämä saattaa edustaa insuliiniresistentin kudoksen palautesignaalia haimalle lisäerityksen aikaansaamiseksi. Kataboliareitin aktiivisuus korreloi vahvasti myös maksan rasvan määrän ja insuliiniresistenssin kanssa. Lisäksi osteopontiinin (*SPP1*) havaitiin olevan eniten (5,9 -kertaisesti) yli-ilmentyvä geeni lihavuudessa. Tämän rasvakudokseen makrofageja rekrytoivan sytokiinin havaittiin täten liittyvän tärkeällä tavalla insuliiniresistenssin syntyyn.

Väitöskirjatutkimusen tulokset syventävät ymmärrystä yleisten dyslipidemioiden ja muiden lihavuuteen liittyvien tautien molekyylitason mekanismeista luoden pohjaa ateroskleroosin sekä sydän- ja verisuonitaudin hoidon jatkuvalle kehitykselle.

Avainsanat: Kompleksitautigenetiikka, dyslipidemia, sydän- ja verisuonitauti, lihavuus, USF1, insuliiniresistenssi, FCHL, hypoalfalipoproteinemia

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I. **J. Naukkarinen**, M. Gentile, A. Soro-Paavonen, J. Saarela, H.A Koistinen, P. Pajukanta, MR. Taskinen, L. Peltonen. USF1 and Dyslipidemias: Converging Evidence for a Functional Intronic Variant. *Human Molecular Genetics*. 2005 Sep 1;14(17):2595-605.

II. **J. Naukkarinen**, E. Nilsson, H. A. Koistinen, V. Lyssenko, L.Groop, M-R. Taskinen and L. Peltonen. Transcript profiles from fat and muscle expose molecular mechanisms behind USF1-associated Dyslipidemia: Allelic imbalance and impaired insulin-reactivity. *Submitted.*

III. A. Soro-Paavonen*, **J. Naukkarinen***, M. Lee, H. Watanabe, E. Rantala, S. Söderlund, A. Hiukka, P. T.Kovanen, M. Jauhiainen, L. Peltonen, MR. Taskinen. Common ABCA1 variants, HDL levels and cellular cholesterol efflux in subjects with familial low-HDL. *Journal of Lipid Research*. 2007 Mar 19; [Epub ahead of print]

IV. K.H. Pietiläinen*, **J. Naukkarinen***, A. Rissanen, J. Saharinen, P. Ellonen, H. Keränen, A. Suomalainen, A. Götz, T. Suortti, H Yki-Järvinen, M. Oresic, J. Kaprio and L. Peltonen. Global transcript profiles of fat in monozygotic twins discordant for BMI: Pathways behind acquired obesity. *PLoS Medicine,* in press.

* The authors contributed equally to the study

ABBREVIATIONS

ABCA1	ATP-binding cassette transporter A1
ABCG1	ATP-binding cassette transporter G1
ABCG1	ATP-binding cassette transporter G1
ACACA	Acetyl-CoA carboxylase alpha
ACACB	Acetyl-CoA carboxylase beta
ADRB2	Beta-2-Adrenergic receptor
ADRB3	Beta-3-Adrenergic receptor
AGT	Angiotensinogen
APOA1	Apolipoprotein A1
APOA2	Apolipoprotein A2
APOA5	Apolipoprotein A5
APOB	Apolipoprotein B
APOC3	Apolipoprotein C3
APOE	Apolipoprotein E
BCAA	Branched chain amino acid
BG	Blood glucose
BMI	Body mass index
bp	Base pair
CAD	Coronary artery disease
CAPN10	Calpain 10
CD-CV	Common variant -common disease
cDNA	complementary deoxyribonucleic acid
CE	Cholesteryl ester
CFTR	Cystic fibrosis transmembrane conductance regulator
CHD	Coronary heart disease
CIDEA	Cell death-inducing DFFA-like effector a
CVD	Cardiovascular disease
DEXA	Dual energy X-ray absormetry
DNA	Deoxyribonucleic acid
DZ	Dizygotic
ELISA	Enzyme linked immunoSorbent assay
EMSA	Electrophoretic mobility shift assay
ENPP1	Ectonucleotide pyrophosphatase/phosphodiesterase 1
eQTL	Expression quantitative trait locus
EUFAM	European multicenter study on familial dyslipidemias with premature coronary heart disease
F11R/JAM1	Junction adhesion molecule
FASN	Fatty acid synthase
FC	Free cholesterol
FCHL	Familial combined hyperlipidemia

FFA	Free fatty acid
FH	Familial hypercholesterolemia
FTO	Fat mass and obesity associated
GCGR	Glucagon receptor
GCK	Glucokinase
GHRL	Ghrelin
GLUT4	Glucose transporter 4
GO	Gene Ontology
HDL	High density lipoprotein
HDL-C	High density lipoprotein cholesterol
HGP	Human Genome Project
HIV	Human immunodeficiency virus
HLA-E	Humam leukocyte antigen E
HNF4A	Hepatocyte nuclear factor 4-alpha
HOXB4	Homeobox B4
IBD	Identical by descent
IBS	Identical by state
IMT	Intima-media thickness
INSIG2	Insulin induced gene 2
IRS-1	Insulin receptor substrate 1
kb	Kilobase
KCNJ11	Inwardly rectifying potassium channel J11
LCAT	Lecithin:cholesterol acyltransferase
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LEPR	Leptin receptor
LIPE	Hormone sensitive lipase
LOD	Logarithm of odds
LPIN1	Lipin 1
L-PK	Liver-type pyruvate kinase
LPL	Lipoprotein lipase
MCP-1	Monocyte chemotactic protein 1
MCP-2	Monocyte chemotactic protein 2
MI	Myocardial infarction
MIP-4	Macrophage inflammatory protein 4
MIP-α-R	Macrophage inflammatory protein alpha
MIR-10	Macrophage immunoglobulin-like receptor 10
miRNA	MicroRNA
MODY	Maturity onset diabetes of the young
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
mtDNA	Mitochondrial DNA
mTOR	Mammalian target of rapamycin

MZ	Monozygotic
NIDDM	Non insulin-dependent diabetes mellitus
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase chain reaction
PPARα	Peroxisome proliferator-activated receptor alpha
PPARγ	Peroxisome proliferator-activated receptor gamma
PTPN6	Protein-tyrosine phosphatase, nonreceptor-type 6
QTL	Quantitative trait locus
REN	Renin
RNA	Ribonucleic acid
RT-PCR	Real-time polymerase chain reaction
SNP	Single nucleotide polymorphism
SPP1	Osteopontin
SR-B1	Scavenger receptor class B, member 1
T2D	Type 2 diabetes
TC	Total cholesterol
TCF7L2	Transcription factor 7-like 2
TG	Triglycerides
THRSP	Thyroid hormone-responsive spot-14
TNFRSF1B	Tumor necrosis factor receptor 1B
TXNIP	Thioredoxin-interacting protein
UCP2	Uncoupling protein 2
USF1	Upstream stimulatory factor 1
USF2	Upstream stimulatory factor 2
UTR	Untranslated region
VLDL	Very low density lipoprotein
WHO	World health organization

Gene symbols have been italicized in the text.

INTRODUCTION

Coronary heart disease (CHD) is the leading cause of death in the Western world and increasingly so in the developing world as it adopts the progressively more sedentary lifestyle and dietary habits of the West. Dyslipidemia refers to an abnormal amount of various lipids in the blood and leads to the development of atherosclerosis -the root cause of this CHD epidemic. A host of enzymes, carrier proteins and receptors are involved in the careful regulation of lipid metabolism and a perturbation of any of these parts can potentially lead to the development of dyslipidemia and CHD. A number of familial forms of dyslipidemia exist, but in adults the acquired dyslipidemia associated with obesity is most common.

Familial combined hyperlipidemia (FCHL) and familial low-HDL are two common hereditary dyslipidemias that both predispose to the development of premature (age <55 years for men, <65 years for women) CHD. Both are recognized to be complex diseases –that is the disease phenotype arises through the collective action of numerous genes and the environment. The presence of multiple predisposing genes, each increasing the disease risk only marginally, makes their identification challenging. Considerable efforts are currently being invested in the search for these genes and in characterizing the effects of the identified variants. Compared with the success at studying Mendelian, single-gene diseases, studies of complex disease are only recently being met with some degree of success. Complex diseases such as common dyslipidemias affect orders of magnitude more people globally than the rare, though generally more serious single-gene diseases. As such, these studies hold great potential for the improvement of human health –a selling point used to justify the enormous financial commitment that was the Human Genome Project. The first clinical applications of these findings are keenly awaited.

REVIEW OF THE LITERATURE

1. Genetics of complex/common disease

1.1 From descriptive genetics to functional genomics

Through the contributions of such scientists as Gregor Mendel, James Watson, Francis Crick, Rosalind Franklin and countless other investigators, the study of heredity has been transformed from a purely descriptive endeavor to an understanding of the phenotype at the molecular level. The discovery of the structure of DNA actually predated that of the correct number of human chromosomes[1], but it was the observation of chromosomal rearrangements that yielded the first described disease associated changes in the genetic material. Early work, employing the laborious positional cloning strategy in families with affected individuals, resulted in the discovery in 1989 of the first gene associated with a hereditary disease. Cystic fibrosis was shown to be caused by mutations in a gene coding for a chloride ion channel, aptly named the cystic fibrosis transmembrane conductance regulator (*CTFR*)[2,3].

Since this first disease gene discovery, the molecular understanding of traits that follow Mendelian rules of inheritance in families –and thus referred to as "Mendelian diseases" has proceeded with leaps and bounds –greatly aided by the tools provided by the human genome project. To date, the gene behind more than 2,200 Mendelian diseases has been identified (OMIM), whereas the progress in complex diseases that involve a number of genes, gene-gene and gene-environment interactions, has been much more modest (Figure 1). It is surely not for a lack of effort that so few reproducible complex disease gene findings have been reported, but rather it is the nature of the diseases and traits under study that make the hunt for these genes much more challenging. With more genes influencing the susceptibility

to disease, and this risk being modified by the environment, each genetic variant is likely to confer a smaller risk and thus be harder to detect. Complex diseases are also typically late-onset in nature, while monogenic diseases tend to be more severe and have an early onset (Figure 2). The difference between the two isn't always clear cut, as there exist examples of families that seem to be segregating a complex disease in a near-Mendelian fashion. Studies of such exceptional families have revealed genes, and importantly pathways, that also play a role in the development of complex disease[4].

Despite the sluggish beginnings, the past few years have been marked however with an increasing number of reproducible complex disease gene findings, representing the latest revolution in our understanding of human genetics (Figure 3). These new association studies are characterized by very large study samples of cases and controls (tens of thousands of individuals), the employment of new technologies that enable investigators to interrogate up to a million common genetic variants in a single experiment and the use of expression arrays that enable the precise measurement of the transcriptional activity of all genes. This has required the development of better statistical and analytical methods to mine the enormous amount of data generated. This thesis work has been carried out during this tumultuous time in the field of human genetics and includes a number of methods applying the modern technologies of molecular genetics.

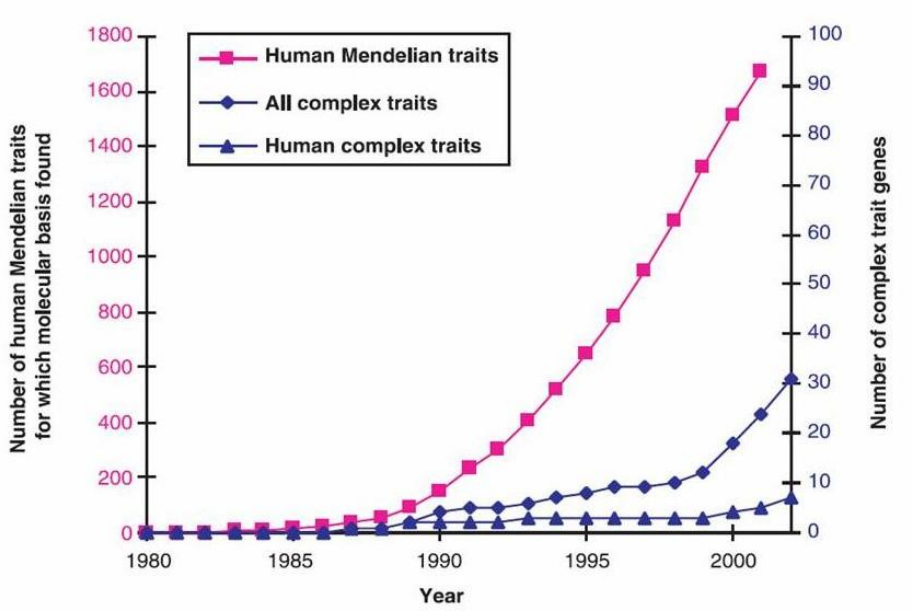


Figure 1. Accelerating pace of gene discovery. In comparison to the progress made with Mendelian, single-gene diseases, complex disease gene identification has been orders of magnitude more modest. Adapted from[5].

Perhaps telling of our growing understanding of human genetics is the fact that even the so-called single-gene diseases are no longer considered to be such clear cut cases: while mutations in single genes are known to cause disease, numerous other genes can modify the phenotype (as is the case with cystic fibrosis)[6], this raising the question whether any disease is truly a single-gene disease. What is being revealed is an enormous network of genetic interactions and pathways, orders of magnitude more complex than may have been assumed in the early days of genetics. It is the concern of functional genomics to ascribe meaning to the dynamic aspects of gene functions and interactions.

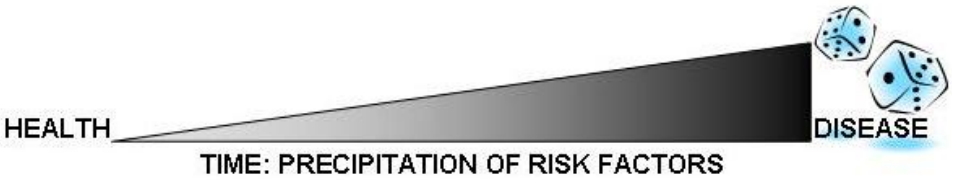


Figure 2. Development of a common/complex disease. The progression of a complex disease represents a continuum from health to pathology that develops over a long period of time. The diseases are thus typically late-onset and the outcomes (ex. myocardial infarct or stroke) are largely stochastic in nature.

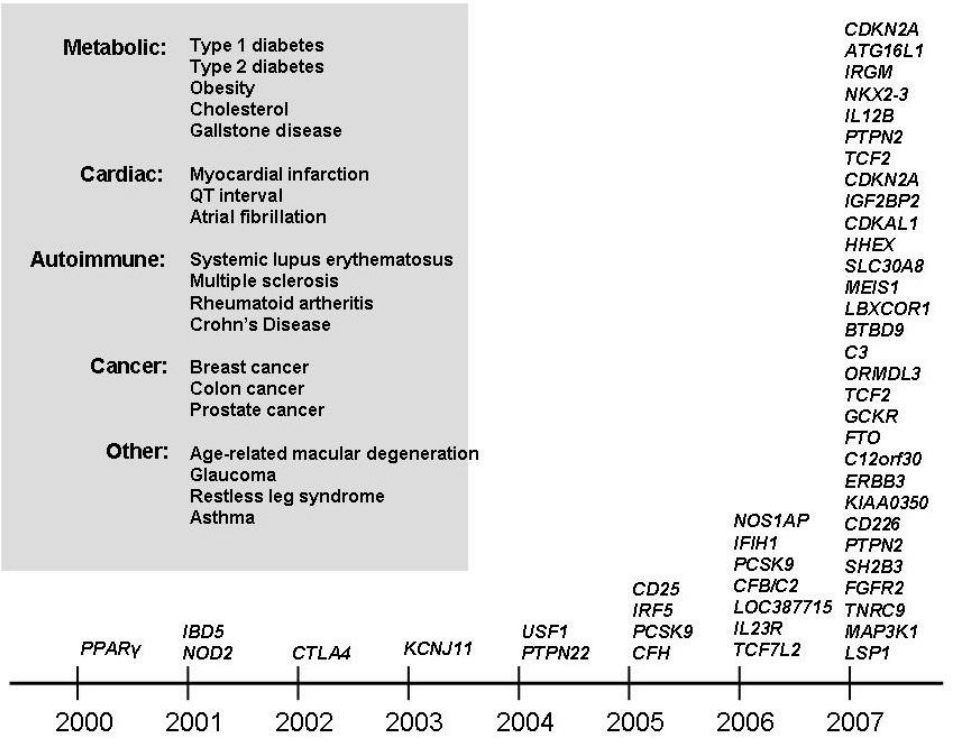


Figure 3. Pace of gene discovery for common diseases. With the advent of new technologies and truly massive study samples, the number of genes underlying common disease traits is beginning to gain momentum. Modified from a slide by Dr. David Altshuler.

1.2 The Human Genome Project

For the most part of the late 20th century, the identification of genes and their assignment onto the map of human chromosomes was a time-consuming task that would take for a committed group of researchers years to complete. It was recognized very early that it would be of paramount importance for the scientific community to have the complete sequence of the human genome available and to assign a position on this map for every human gene. The Human Genome Project (HGP) was formally initiated in October 1990 with the stated goal of identifying all the genes in human DNA, to determine the sequence of the 3 billion chemical base pairs that make up human DNA and to make this information publicly available in databases. The project was also to improve tools for data analysis and to address the ethical, legal and social issues that may arise. The project was completed faster than anticipated with the first "working draft" of the human genome announced in the year 2000 and an essentially complete sequence finished in 2003 –precisely 50 years after Watson and Crick published their seminal paper on the discovery of the structure of DNA[7] (Figure 4).

The HGP has succeeded in its goals and is the foundation of all research in the modern field of human genetics. Much like the space race of the 1950s -70s drove technological advancements in aerospace engineering and electronic communication, the HGP, along with the race between the public[8] and private[9] efforts as considerably improved and appended the genetic toolkit. The project has revealed the size of the human genome to be 3164.7 million nucleotides long and to contain an estimated 20,000-25,000 genes (HGP information web site). The true number of genes may yet drop some from these estimates as more research is done and our definitions of what constitutes a gene change. This is a considerably lower number of genes than previous estimates of about 100,000 genes, based on extrapolations from gene-rich regions of the genome.

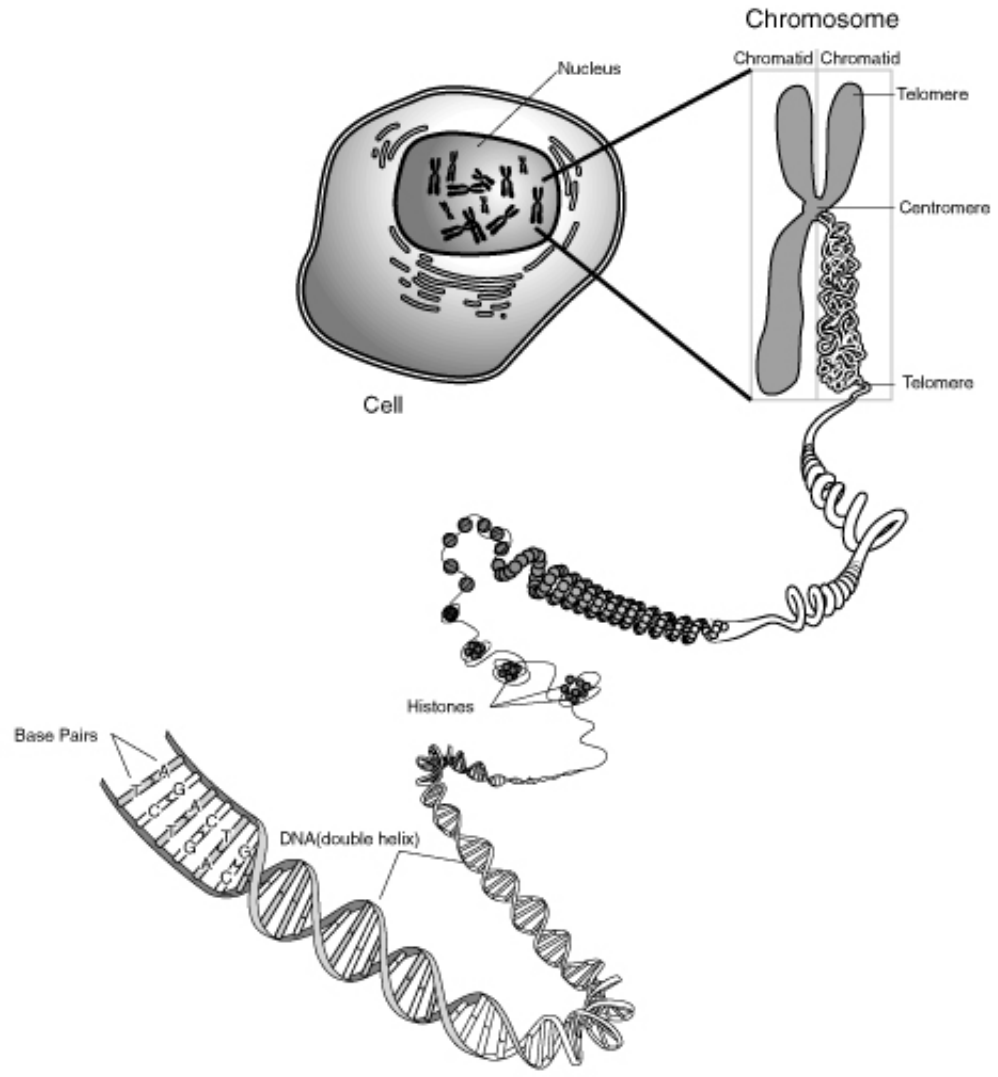


Figure 4. The structure of DNA and its packing into chromosomes in the cell nucleus. The entire length (about 3 meters) of DNA that makes up the human genome is tightly packed and organized into 23 pairs of chromosomes contained in the nucleus of the cell. (www.genome.gov/glossary.cfm)

Less than 2% of the genome codes for proteins, while the considerable majority is composed of non-coding, inter-genic- and repeating sequences. Some of these sequences, previously described as "junk-DNA", on closer inspection have been shown to be functionally relevant[10] and highly conserved –as has been learned from the comparison of sequence reads from the large number of other concurrent genome projects aimed at describing the full sequences of genomes of other species[11]. These comparisons have also revealed how closely related we are genetically to other species. These inter-species comparisons that can facilitate the identification of functional non-coding DNA elements in our genome are an essential tool in the study of complex disease, as a majority of the complex disease variants identified are located in these non-coding regions of our genome that presumably are involved in the regulation of gene expression[5].

1.3 Cataloguing the variation in the human genome

1.3.1 Types of sequence variation

With the exception of identical twins (excluding somatic mutations), no two individuals share the same precise sequence of nucleotides that makes up our genome. There exists variation in the sequence, and this variation is what in large part makes all individuals different. The genetic variation that exists in the human genome is present in several forms. The simplest and most abundant form of variation are changes in a single nucleotide of the genetic code. These single nucleotide polymorphisms (SNPs), can be roughly divided depending on their location in relation to genes to those that are found within the protein coding sequences of genes and those that are not. Further subdivision in the coding sequence SNPs can be made into those that either change an amino acid (non-synonymous SNPs) and those that cause no change (synonymous SNPs), thanks to the redundancy in the genetic code. A SNP may also introduce a STOP codon into

the coding sequence, thereby causing the production of a truncated protein that most often has lost its function. Most polymorphisms are selectively neutral, but some carry functional significance.

Another form of variation is simple deletion or insertion of a single (or a few) nucleotide(s). There can also be variation in the repeat number of a motif (mini- and micro-satellites). These motifs –composed of a repeating sequence of nucleotides, are prone to copying errors by the DNA replication machinery. This produces considerable variation in the number of repeats and this high degree of polymorphism is useful in genetic mapping. The vast number of polymorphisms identified, and subsequent accurate positioning of these markers along the chromosomal DNA strands have since the early years of the HGP provided us with a "handle" on individual segments of otherwise monotonously identical DNA between individuals. This ability, via these genetic markers, to discriminate between segments of DNA between maternal/paternal chromosomes and individuals enables the observation of their segregation in families and distribution in different populations –the basis for disease gene identification.

1.3.2 The HapMap project

Building on the work of the Human Genome Project that described the complete sequence of the human genome, the HapMap project was undertaken to catalogue all the common patterns of variation in the code. An estimated 10 million SNPs (with minor allele frequency >1%) exist in human populations. These SNPs are not inherited entirely independently of one another, but rather in more or less distinct blocks, with SNPs located in close proximity tending to be inherited together. This "non-random association of alleles at two or more loci" is termed linkage disequilibrium (LD). The pattern of LD is a result of the various forces of selection that have molded the human genome during the history of our species.

It was the proposition of the HapMap project that in order to describe the majority of the allelic diversity in a sample, it may not be necessary to genotype every variant. Rather, it may suffice to genotype only a subset of common SNPs (defined as having a frequency ≥5%) that "tag" a certain DNA segment carrying a number of markers in LD –a haplotype[12]. This greatly reduces the amount of genotyping needed to describe common variation in the genome, which directly translates to reduced cost of such studies[12].

The utility of the HapMap to aid in the identifying of disease variants rests on the assumptions behind the common disease common variant (CD-CV) hypothesis. In short the theory suggests that many common diseases, such as type 2 diabetes (T2D), asthma, obesity, Alzheimer's disease, psychiatric diseases and hypertension may be caused by a small number of common alleles. More correctly, that multiple common variants may confer *susceptibility* to these common diseases. Opponents of the CD-CV hypothesis argue however, that common diseases are caused not by a few common variants, but rather by multiple rare mutations, thus perhaps making the HapMap approach of only monitoring for common variants a futile affair[13,14]. With the first reports from massive genome-wide association studies (that rely on the HapMap as a reference) revealing multiple, replicable common disease gene findings, the starkest opponents may be losing ground. Here too the likely representation of true-biology probably lies somewhere between the two opposing views. Both common variants, as well as multiple rare variants contribute to the susceptibility to common diseases[15-17], with the HapMap approach proving to be powerful in detecting the prior, while the latter variants may only be detectable through large scale re-sequencing efforts. The rare variants as well as the "grey zone", the not so common, not so rare variants, behind common diseases and traits can potentially be identified by analyzing exceptional populations that due to their population history, exhibit an enrichment of rare alleles.

1.4 Genome-wide analyses of DNA

1.4.1 Linkage analysis

Most disease genes identified so far have been initially positioned in the genome via linkage analyses in exceptional families with multiple affected cases. The basis of linkage approach relies on the basic concept that in sexually reproducing animals such as humans, genetic diversity is generated in the process of meiosis through the formation of the male and female germ cells that recombine in fertilization to produce a new individual. Through the process of crossing over, a pair of homologous chromosomes exchanges segments, so that the resulting chromosomes contain segments from both of the original pairs. This crossing over takes place at least once on each chromosomal arm and the likelihood that two loci on the same chromosome are separated by a recombination event is related to their distance from each other. Loci on different chromosomes segregate independently and are thus said to have a recombination fraction (denoted by θ) of 0.50. Loci right next to one another may never be observed to be separated by a recombination and thus have a recombination fraction of $\theta = 0$. The recombination fraction thus ranges from zero (for complete linkage) to ½ (no linkage). It is the objective of linkage analysis to estimate this recombination fraction and to test whether this differs significantly from the null hypothesis of no linkage, i.e. $\theta = 0.50$.

The estimate of the recombination fraction is based on counting the number of observed recombinant individuals in a pedigree. However, in real-world pedigrees this may not always be possible to do in absolute terms and the test for linkage takes on a probabilistic expression –a likelihood ratio. The likelihood of linkage is compared to the likelihood of no linkage. By convention, this is typically expressed as a (base 10) logarithm of the ratio of these two likelihoods. This is called the logarithm of odds (LOD) score (Z).

$$Z(\theta) = \log_{10} \frac{L(\theta)}{L(\theta = 0.5)}$$

The calculation of this LOD score over a range of different values of θ and the value of θ at which the LOD score peaks is considered to represent the most likely recombination fraction, and thus the genetic distance between the two loci. In an attempt to identify disease genes, this test is done between genetic markers and disease status. Manual calculation of LOD scores can be done for simple pedigrees and disease states, but in large pedigrees and with diseases with unusual inheritance patterns this calculation must be done using one of several computer programs specially designed for these calculations. A LOD score of 3 translates to an odds ratio of 1000:1 and this has traditionally been considered a statistically significant demonstration of linkage, where as a LOD score of -2 is generally accepted to demonstrate the exclusion of linkage[18].

1.4.2 Association analysis

While linkage is a specific genetic relationship between loci in a pedigree, association in principle is merely a statistical statement about the co-occurrence of certain alleles with each other, or with a phenotype. Rarely however, do investigators have the luxury, or luck of directly testing the functional variant, but rather variants that are in proximity of the functional variant, and thus in linkage disequilibrium (LD) with it. This existence of LD between markers that are located close to each other is a demonstration of the fact that we are all in essence a part of an enormous pedigree of an unknown structure. Association studies that look for association between markers and disease status are in essence testing the bottom layer of this enormous pedigree where individuals with the same disease are assumed to also share nearby genetic markers, because they have inherited them from a distant, common ancestor (Figure 5).

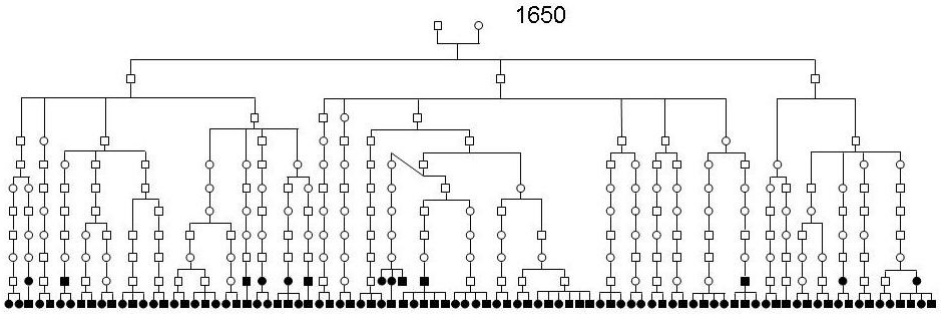


Figure 5. An extended Schitzophrenia pedigree from the Finnish isolate. Association studies in a population sample assume that affected individuals (black) share disease alleles passed down from a distant common ancestor. Only affected probands and ascendants needed to connect them in a pedigree are shown (other connections likely exist). Figure provided by Dr. Teppo Varilo.

Association analyses with a dense set of SNPs is the current gold standard when fine mapping a chromosomal region identified in a linkage scan utilizing hundreds of multiallelic microsatellite markers. Care must be exercised when designing and interpreting results from association studies, for not all observed associations are caused by linkage disequilibrium with a functional polymorphism. Spurious association can arise for several reasons, among them population stratification (the population under study consists of several genetically distinct subsets, each with their distinct allele frequencies), natural selection (two alleles may be associated because the are compatible with life only when inherited together) and simple statistical artifact[19] –an increasingly relevant worry, with the enormous number of markers, and thus tests performed, in the new genome-wide association studies. Some of the features of the etiology and study of complex disease are described in figure 6.

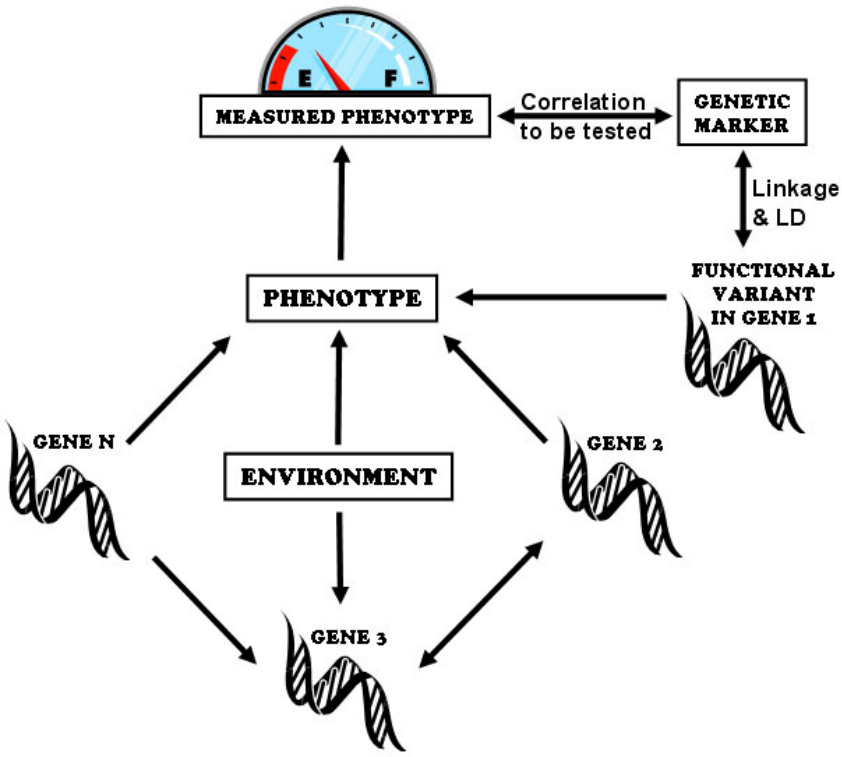


Figure 6. Schematic model of the genetic study of complex disease. This figure highlights the etiology of a complex trait, or disease, and some of the issues complicating its study via genetic means. An observed phenotype is shaped by the environment, gene-environment and gene-gene interactions. With more genes affecting the phenotype, each individual gene is likely to have a smaller effect size – making them more difficult to identify. Other confounding effects are introduced by our inept ability to precisely measure (and define) the phenotype, as well as the fact that more often than not, we are not testing the functional variant directly, but rather a marker in varying degrees of LD with it.

1.4.3 Combining the power of linkage and association

As alluded to above, linkage produces association within families, but not among unrelated people. The situation changes however, when two individuals sharing a certain trait or disease phenotype also share the same segment of DNA related to this trait -passed down from a distant common ancestor. In such cases the individuals are said to share the segment *identical by descent* (IBD) because it originated from the same allele in an earlier generation. This is in contrast to sharing an identical segment of DNA, without having inherited it from a common ancestor, but rather via a separate mutation –a condition known as *identical by state* (IBS). The length of the shared segment between individuals is related to the time (number of meioses) since the last common ancestor, as well as to the population history. A population which was founded by only a handful of individuals and has experienced a recent expansion in size is characterized by enrichment, by chance, of rare alleles (founder effect) as well as by LD blocks that are longer than observed in a more outbred population with a historically more constant population size (Figure 7).

Finland represents such a population isolate and consequently shows significantly less genetic diversity than most other Caucasian populations (Figure 8). In addition to relative genetic homogeneity, considerable environmental and cultural homogeneity reduce the confounding factors that hamper genetic studies in more outbred, mixed populations. The availability of good genealogical records and equal, high-quality national healthcare with detailed registries are another important reason genetic studies in the Finnish population have been so successful at identifying Mendelian disease genes. With the unique population history and genetic makeup, come a unique set of enriched diseases. A total of 36 rare, monogenic diseases have been enriched in the Finnish population and genetic studies have been able to locate the disease gene and mutation in 35 of these[20,21]. The genetic defect remains to be identified for PEHO, a form of infantile progressive encephalopathy[22].

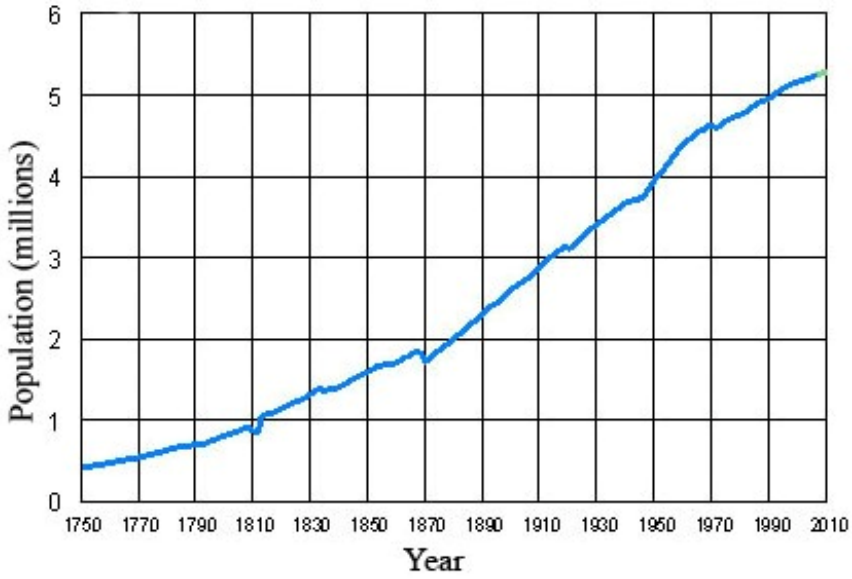


Figure 7. Development of Finnish population size. One quality of a population isolate that contributes toward an increased amount of LD is a relatively recent expansion of population size. Another characteristic is the presence of a number of population bottle-necks, small examples of which can be identified; Finland's war (1808-1809) and the Great Famine (1866-1868). Modified from Statistics Finland.

Studies in isolated populations have been very successful in mapping Mendelian disease-genes, but their utility in mapping complex-disease genes has been questioned. Several examples already exist however, where a disease gene identified in a population isolate among exceptional families with multiple affected individuals is upon further study revealed to be associated with an increased risk for disease in the general population as well as in more heterogenous populations[4]. Some informative examples include maturity onset diabetes of the young (MODY), familial combined hyperlipidemia (FCHL), familial Altzheimer's disease as well as rare forms of epilepsies and migraine[4].

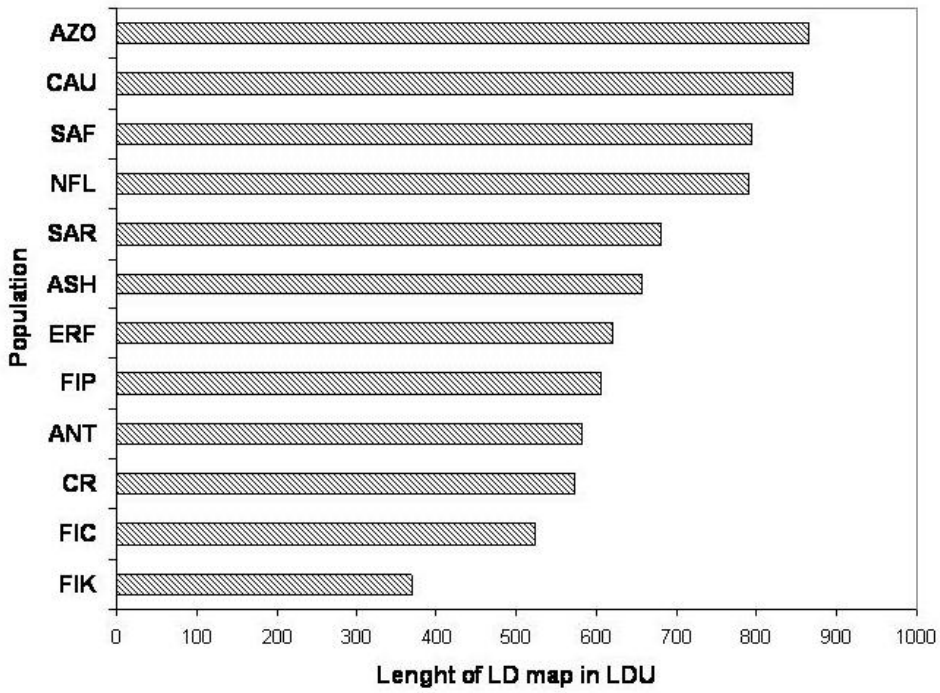


Figure 8. Length of LD map across various populations. Considerable differences in LD map length across populations. The length of the LD map in LDU (as defined in[23]) in 12 different population samples is depicted in order of decreasing map length. AZO: Azores; CAU: outbred European-derived sample; SAF: Afrikaner; NFL: Newfoundland; SAR: province of Nuoro in Sardinia; ASH: Ashkenazi; ERF: a village in southwestern Netherlands; FIP: Finland nationwide; ANT: Antioquia; CR: Central Colombia; FIC: early-settlement Finland; FIK: Finnish subisolate of Kuusamo. Adapted from[24].

1.5 Analyses of mRNA and the transcriptome

1.5.1 Expression analyses

The central dogma of molecular biology, as it is known, states that information flows from DNA to RNA to protein. Stated simply, DNA contains the encoded instructions for making the various proteins that largely make up our bodies and are

responsible for catalyzing the chemical reactions that sustain life. DNA specifies the synthesis of a messenger RNA (mRNA) that after being edited in the nucleus, takes this information into the cytoplasm of the cell where the code is read by the ribosomes and translated to a specific sequence of amino acids. These chains of amino acids fold in a unique manner to form functional proteins (Figure 9). This flow of information from DNA to RNA to protein is essentially a one-way process (with the exception of retro-viruses like HIV that are capable of turning their RNA genome to DNA that then incorporates into the host genome.)

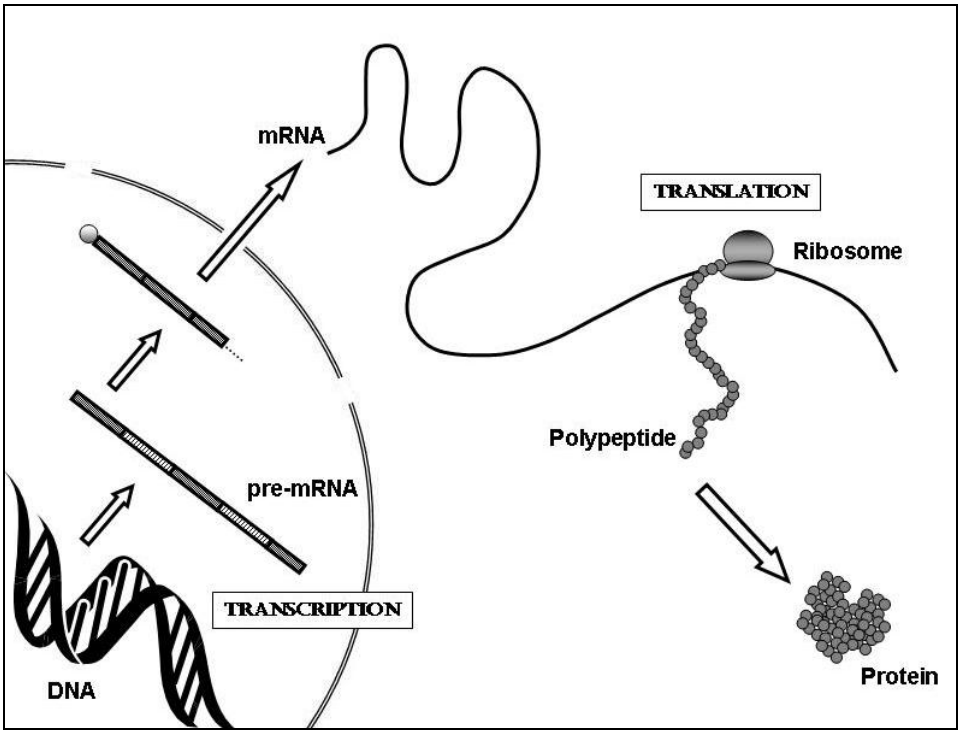


Figure 9. Simplified schematic of the process of protein synthesis. A strand of DNA encoding a protein is read and a messenger RNA (mRNA) is produced in the process of transcription. After splicing of introns and some additional modifications, the mRNA moves into the cytosol and the triplet code of nucleotides is read by the ribosomes. In this process of translation a polypeptide chain of amino acids is produced which then is folded in a unique manner to form a functional protein.

Of the estimated 20,000-25,000 genes that make up our genome, only a fraction are "turned on", or expressed in a given tissue at a given time. A quite different set of genes are expressed, and thus proteins made, in the different tissues that make up our bodies. Different diseased tissues also have their own expression "signature" compared to that of healthy tissues and detecting this signature can be important, because it can help to identify different subtypes of disease, such as different cancers, that respond to different treatments. Similarly, genetic diseases have their own specific expression profiles, that range from the total loss of a certain gene product (as is the case in many severe, early onset Mendelian diseases), to only a subtle change in the expression of a handful of genes, or a given biological pathway –as is suggested to be the case with most complex, late-onset diseases.

The temporal and spatial expression of genes is under complex control that takes place at multiple different levels: binding of tissue-specific transcription factors or hormones to cis-acting elements of DNA influence the transcription of the gene, as do the so-called epigenetic changes that are only beginning to be understood. As the name implies, epigenetic modification takes place outside, or "on top" of the DNA sequence, perhaps most well described in the form of methylation that can silence genes. Importantly, while these epigenetic modifications do not involve changes in DNA sequence, they present as mitotically and/or meiotically heritable changes in gene function[25]. These epigenetic changes seem to be in principle stable, but potentially affected by the environment. This raises the intriguing notion with an admittedly Lamarckian ring to it, that acquired changes may be in part passed down from parents to offspring[26]. Post-transcriptional regulation operates at the level of RNA processing, transport, stability and translation. In maintaining the proper expression of genes both temporally and spatially, the silencing of genes is just as important as the "turning on" of genes. New insights have relatively recently been gained to how this process takes place. MicroRNAs (miRNA) are short (21-23 nucleotides), non-coding single stranded RNA molecules that are partially complementary to specific mRNA molecules (their 3' UTRs) and can thus bind to

them and inhibit their translation into proteins[27]. This can take place through either direct inhibition of translation or by triggering the degradation of the mRNA transcript through a process similar to RNA interference. These miRNAs have raised considerable interest as their important role in controlling gene expression is beginning to be appreciated. Hundreds of miRNAs are predicted to be present in higher eukaryotes[28] and it is speculated that they may play a role in the pathogenesis, and in the future possibly in the treatment of numerous diseases[29].

Our ability to detect and accurately measure the global transcript profiles of biological samples has revolutionized genetics and provided a whole new " –ome" to investigate; namely the transcriptome. Active genes are actively transcribed and can be detected through their abundant mRNAs. The expression level of a gene represents the phenotype closest to the gene and is thus in a sense a purer phenotype than some of the classical ones such as measures of plasma triglycerides (TG) or cholesterol. There exist however, many levels of regulation beyond that of gene expression, so that conclusions made from transcript level changes alone must be made with the caveat that they may not be identically represented at the protein level. Expression profiles in essence provide investigators with a snap-shot of genes at the functional level.

The simplest method of measuring a gene's expression level is by reverse-transcriptase PCR (RT-PCR), whereby the total mRNA extracted from a tissue is first turned to complementary DNA (cDNA) by the enzyme reverse transcriptase. The abundance of cDNA is then measured with specifically designed primers in a machine that is capable of quantitating the amount of product produced in a PCR reaction. A larger amount of starting template cDNA results in a larger amount of amplified product and the amount of cDNA is directly related to the activity of the gene. This method is popular because of the relatively low cost associated with running the experiments, but is tedious because the expression of genes can only be measured one-by-one. New methods involving array technology with more than a million individuals probes printed onto special slides enable the detection and precise measurement of all

genes in a single experiment. This method produces an enormous amount of information, but is still relatively quite expensive to perform. The expression arrays used in this study are the Affymetrix U133 Plus 2.0 arrays that analyze the expression level of over 47,000 transcripts and variants, comprised of more than 54,000 probe sets and 1,300,000 distinct oligonucleotide features. While this platform enables the interrogation of all individual genes, it importantly enables the inspection of a collection of relevant inter-connected genes –a biological pathway.

The expression profiles of individuals exhibit large differences, not only because of differences in environmental exposure, but also because of their unique genetic make-up. Polymorphisms in the promoters of genes, enhancer elements, in transcription factors, methylation sites, miRNAs and their binding sites all have the potential to effect the expression of genes. The majority of polymorphisms that underlie the different susceptibility to complex disease are thought to be located in such sequences[30-32]. With the availability of genome-wide tools to analyze both all the common variation as well as the expression profiles of practically every gene in an individual, new and exciting possibilities have emerged from combining these two levels of genetic information. Using the expression levels of genes as quantitative phenotypes in linkage studies has recently enabled the identification of so-called expression quantitative trait loci (eQTLs) that contribute to phenotypic differences between individuals[33].

1.5.2 Extending the transcript analyses to biological pathways

No man is an island and no gene acts in isolation, but rather as a part of a biological pathway with numerous potential interactions with other genes. Especially in the case of complex disease where the measured changes in expression of a particular "disease gene" are likely to be only modest, the detection of these changes can be not only challenging methodologically and statistically, but also of only limited utility. The even subtle perturbation however, of an entire biological pathway in a

disease state can be of much higher utility, not only because of its likely greater effect on the individual over time, but as it uncovers a whole host of potential new candidate genes, as well as druggable targets.

The topography of a pathway can be viewed as a series of interconnected entities (genes, in this case) with central positions, or hubs, occupied by transcription factors and other proteins that have wide ranging effects on the expression of numerous other genes. Most familiar examples of biological pathways are the metabolic pathways, through which a given metabolite (for example glucose) is consecutively modified by a number of enzymes (for example the enzymes of glycolysis and the citric acid cycle), some of which are highly regulated and act as control points for the entire pathway. Other examples of pathways are signal transduction pathways and developmental pathways, all of which have different dynamics of operation and different regulatory characteristics.

Most current characterizations of biological pathways are incomplete approximations of the multi-dimensional nature of the true biology, but a consortium of scientists has attempted to describe gene products in terms of their associated cellular processes, molecular functions and cellular components. The Gene Ontology Consortium has assigned each gene in their database to these three categories in a tree-like arrangement[34]. This cataloguing of gene products to associated processes not only attempts to bring consistency to gene descriptions, but also facilitates the extension of expression analyses to the level of a whole biological pathway –an important tool in the study of complex diseases.

1.6 The concept of replication in the context of complex disease

One challenge that has come about with the advent of massive genome-wide association studies is to separate true findings of association with a phenotype from the slur of false positives that are sure to appear with the large amount of tests being performed in each study. Replication of original findings in independent collections

of samples is necessary for establishing the validity of an association. There exists however, disagreement over what constitutes a true replication in the context of complex disease findings. Unlike the replication of monogenic disease linkage findings, reports of which filled genetics journals until of late, for true replication in the case of complex disease it is not sufficient to merely replicate the same chromosomal locus as the original report. But what standards should one set for accepting a replication as valid? Is it sufficient to find a similar association with a different marker in the same gene? How strong should the LD be with the marker of the original report? Can that marker be located outside the gene and if so, how far? Should it be a requirement that the associated allele is the same in all studies? And concerning the phenotype; does the association have to be with the very same diagnostic endpoint as in the original finding, or can the association be with a component trait (or endophenotype) of the disease, or endpoint?

Many issues should be considered in the design, execution and reporting of a replication study in order for the scientific community to be able to draw conclusions of the validity of the findings. Issues involving the study population as well as phenotypes and markers analyzed, all are possible confounders. While a replication in a population of different ancestry should increase the confidence in the validity of the finding, failure to replicate should not necessarily be considered to invalidate the original finding, owing to differences both in the important environmental factors, as well as possibly different haplotype structure in the locus under study. Ideally, the replication should be with the same marker, or one in perfect LD (r^2=1.0) as the best result in the original report. However, since most initial reports do not claim to have identified the functional variant, testing and reporting on additional SNPs (even ones not tested initially) in the region should only aid in identifying the causal variant(s). Unfortunately, many reports of replication have been content at genotyping only the best associating variant(s) of the original study and thus the opportunity is lost for more finely dissecting the association signal. The importance of reporting negative results should also be

emphasized[35,36]. Publishing negative results is difficult, but disseminating such information may provide a crucial element in dissecting the effect of the proposed disease gene, as well as for aiding in making the decision of whether or not to pursue time consuming functional studies.

1.7 Identifying functional variants

In addition to competently performed replications of original findings, attaching functional significance to the identified polymorphisms holds great importance for the process of convincing the scientific community of the validity of a finding. Even more so in the case of complex disease genes, since the polymorphisms identified are generally more cryptic in nature and don't immediately present any obvious defect. This is in contrast to most mutations identified for the more serious, early onset monogenic diseases where the mutations usually result in serious damage to the protein structure. Our comprehension of the relationship between non-coding variants, gene expression and effects on phenotype is considerably more limited than our understanding of changes in coding sequence and resultant effects on protein function.

Functional variants for complex traits are more likely to be located in non-coding regions of the genome that affect transcription. Such variants that affect the transcription of the gene can be found in the promoter regions of the gene (that can be located several kilobases away, as is the case with the variant associated with adult type hypolactasia located nearly 14 kb upstream of the lactase gene)[37] and enhancer elements located in introns. Mutations or polymorphisms in the promoter and enhancer elements can change the affinity of transcription factors for their cognate sequences and thus affect the transcription of the gene. Sequence changes in the untranslated regions (UTRs) may affect the stability of the mRNA prior to translation into protein. These sequence changes, in addition to possible differences

in methylation/acetylation states of DNA, can have effects on the level of expression of the gene under different circumstances.

The presence of multiple, associated non-coding intragenic SNPs in high LD can make it difficult to prioritize SNPs for functional studies. Evaluating the inter-species conservation of the surrounding sequence can help in this task, as evolution conserves function. Special softwares have been designed to assist in this task[38]. They take use of multiple between-species comparisons as well as knowledge over the binding sequences of transcription factors and can predict which variants have the highest likelihood of being functionally relevant. In the end however, functionality must be tested using actual laboratory experiments. Some of the more commonly used assays include testing the sequence for promoter activity in a reporter-gene construct in cell lines, and testing for changes in DNA binding capability in electrophoretic mobility shift assays (EMSAs). The most conclusive evidence for the functionality of a variant is that the phenotype can be "rescued", or changed by replacing the variant allele. Proving this may be difficult in an in vitro test lacking the correct physiological environment, necessitating the use of experimental animals. Finally, given that the physiologies of humans and experimental species are not identical, the ultimate evidence for the functional significance often uses combined strategies of reaching from cellular experiments to population studies and analyses across species.

1.8 Extrapolation of complex disease findings

Geneticists use the terms "complex disease" and "common disease" interchangeably. By definition, they affect a considerably larger portion of the population than the much rarer monogenic diseases. The ultimate goal of the study of disease genetics should be to eventually translate the findings to benefit human health. Findings in monogenic diseases have firstly enabled the precise molecular

diagnosis of disease and given the tools needed for genetic counseling of families - exemplified by implementation of the molecular findings of the Finnish Disease Heritage (www.findis.org). Going from disease associated gene-identification to concrete health benefits in the context of common diseases is a much more challenging task, as the variants identified may only marginally increase the life-time risk of developing disease. The very use of the term "risk" exemplifies a fact about complex disease findings; their relevancy may only be obvious at the level of populations and that *individual* risk is essentially impossible to predict. This feature suggests that the benefits to human health are less likely to come in the form of prevention or prediction. The identification of complex disease-associated genes and importantly with them, biological pathways related to the disease process, may open up a new druggable system. Benefits to human health may thus be more likely to come in the form of new drugs. Perhaps the most prescribed class of type 2 diabetes (T2D) drugs are the Thiazolidinediones (ex. Rosiglitazone) that act to alleviate insulin resistance by binding to the transcription factor PPARγ[39] – a gene that has later been shown to associated strongly with T2D[40]. While in this case the drug discovery preceded the association of the gene with the disease, there is reasonable hope that similar gene discoveries will provide the new druggable targets that translate to considerable benefit to human health in the near future.

1.9 The thrifty gene hypothesis

The prevalence in the Western world of metabolic derangements such as obesity, type 2 diabetes and their sequelae has dramatically increased in the past decades to the point where medical professionals are using the term "epidemic" to describe the current situation. While it is generally accepted that obesity and T2D have a genetic component, our genes haven't changed during the time that this epidemic has come about. What *has* changed is our environment. In 1962, James V. Neel introduced the thrifty gene hypothesis, suggesting that what is behind the current increase in

obesity and related conditions is the mismatch that arises from having "stone-age genes in a space-age environment"[41]. The overwhelming majority of human evolution took place in the hunter-gatherer environment of the Paleolithic, and we are not only evolved from these Paleolithic hunter-gatherers, genetically speaking we still *are* them. The hypothesis posits that the existence of our ancestors consisted of cycles of famine and feast, physical activity and rest. In the current world, our existence is more characterized by sustained feast and physical inactivity –the cycle has stalled. In such an environment, the thrifty genes that conferred a reproductive advantage to those individuals who could efficiently store energy (in the form of fat and glycogen) are now the very genes that confer greater susceptibility to the so-called life-style diseases. The suggested mechanism may be seen at work among the geographically isolated Pacific Islanders where it is suggested that thrifty genotypes have been selected for even more strongly because of the long and difficult overseas journey required by their ancestors to reach the remote islands[42]. Following the so called "coca-colonization" of these communities, some groups where no recorded diabetes existed 50 years earlier, now have the highest rates of T2D in the world with more than 40% of the population afflicted[43].

Some opponents of the hypothesis question the ability of periods of famine to exert a selection pressure intense enough for thrifty genes to become common in a population[44]. It is also argued that the majority of deaths at the time of famine are not caused by starvation *per se*, but rather through infection and diarrhea –perhaps brought about indirectly by the lack of food, but not malnutrition itself. Those most likely to succumb to a famine are the very young and elderly, while those

individuals in their reproductive prime are most likely to survive –a fact that has been used to argue that no strong selective pressure can be asserted by famine. These arguments however do not address the other factor in the equation, i.e. the cycle of physical activity and rest. No clear cases of thrifty genes have been reported yet, but absence of proof is not proof of absence. With an increasing number of common obesity and T2D susceptibility variants being identified, the thrifty gene remains an intriguing hypothesis.

2. Common Dyslipidemias

2.1 Familial combined hyperlipidemia

2.1.1 History and diagnosis

In 1973, three groups in Helsinki[45], Seattle[46] and New York[47] independently described a new inherited disorder of lipid metabolism among young survivors of myocardial infarction. The disorder was named familial combined hyperlipidemia (FHCL) and the diagnosis consisted of a multiple types of dyslipidemia in the family: isolated high cholesterol (Fredrikson's lipid phenotype IIA), isolated high triglycerides (IV) and the combined phenotype (IIB). In addition, elevated levels of apoB and low levels of HDL are also often observed among affected individuals. Another feature of the condition is that affected individuals may change their lipid phenotype between visits to the clinic. While it was originally suggested that FCHL may be a case of a variable expression of a single autosomal gene[46] dominantly inherited, it is now clear that FCHL is a complex disease with multiple genes and environment affecting the phenotype. Clinical diagnosis relies on accurate population based, age- and sex-specific cut-off levels of lipid measurements. The presence of early coronary heart disease (CHD) in the family has been a requirement

for a diagnosis of FCHL in some studies. The cut-offs for plasma lipids have varied between different studies, resulting in variable diagnoses and encumbering the search for the molecular basis of the disease. Finnish studies use the 90th percentile in triglyceride and cholesterol measurements, appropriate for the population and individual, as the cut-off point to establish affection status (Figure 10). The population prevalence of FCHL is estimated to be 1-6%, making it the most common familial dyslipidemia predisposing to CHD[46,48]. The prevalence among patients with premature CHD is much higher, at approximately 20%[46].

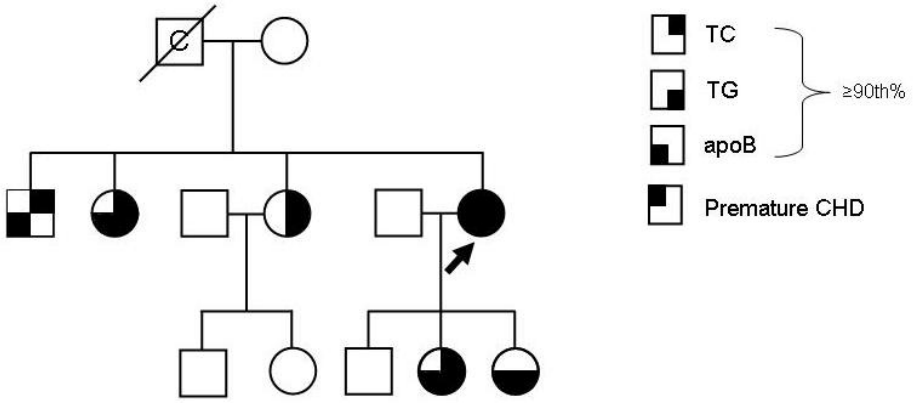


Figure 10. An example of a typical Finnish FCHL family. Modified from the thesis of Päivi Pajukanta[49].

2.1.2 Metabolic disturbances

The complex FCHL phenotype has been extensively dissected in an attempt to reveal the underlying metabolic disturbances. A primary defect in lipoprotein metabolism is the hypersecretion by the liver of apoB containing lipoproteins (VLDLs) as well as the delayed clearance from circulation of their atherogenic

remnants. This contributes to the elevated plasma triglycerides, apoB and total cholesterol. As would be expected of a multifactorial disease such as FCHL, involving a system as complex as lipid metabolism, it is difficult to discriminate between cause and effect in its etiology. The multiple different component traits of FHCL are listed in table 1. These quantitative traits are in themselves open to genetic analysis and may reveal causative- and modifying genes for FCHL.

Table 1. Additional FCHL component traits

Trait		Reference
HDL-C	↓	50
LPL activity	↓	51
Glucose tolerance	↓	52,53
Insulin	↑	54
ApoB containing lipoproteins	↑	55,56
Small dense LDL particles	↑	57,58
Abdominal obesity	↑	59,60
Free fatty acids	↑	61

Patients with FCHL are often overweight and to a variable degree insulin resistant[59]. These are also key features of the metabolic syndrome and accordingly, it has been suggested that the two diseases share etiological overlap. Obesity can mimic the FCHL lipid phenotypes and has been shown to be a confounder in making the FCHL diagnosis. Obesity has therefore been an exclusion criterion in some studies and can also be considered to modify the FCHL phenotype[59,62,63]. A fuller understanding of the physiological changes that occur as a result of acquired obesity are thus also important for disentangling the genetic, versus environmental contributions to the FCHL phenotype.

2.2 Dyslipidemias with overlapping phenotypes

2.2.1 Low-HDL

The FCHL phenotype is a mixed dyslipidemia with several component traits. It therefore shares phenotypic overlap with several other disorders with narrower definition of dyslipidemia. One such CVD predisposing disorder is familial low HDL. High density lipoprotein (HDL) is responsible for the reverse transport of cholesterol. Through this process, HDL transports cholesterol from extrahepatic tissues to the liver for excretion in bile (Figure 11). There exists a clear inverse relationship between the levels of HDL-C and the risk of developing premature coronary heart disease[64]. A low level of HDL is in fact the most common dyslipidemic finding in patients with premature coronary heart disease[65]. Aside from its role in the reverse transport of cholesterol, HDL has important anti-inflammatory effects that protect against CVD[66].

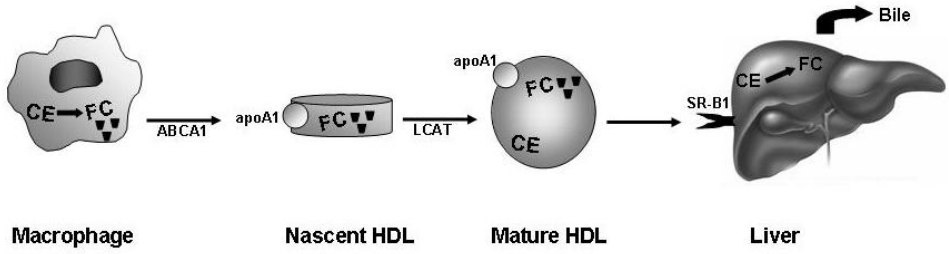


Figure 11. Reverse transport of cholesterol. CE: cholesteryl ester, FC: free cholesterol. (Modified from[67])

Mutations in several genes have been shown to cause rarer cases of familial low-HDL, most notably in the gene encoding for ATP-binding cassette transporter A1 (*ABCA1*) causing the rare Tangier's disease[68-70]. ABCA1 facilitates the efflux of cholesterol from peripheral cells to lipid poor apolipoprotein A1 (APOA1) –the main apolipoprotein of the HDL particle. The virtual absence of HDL in Tangier

patients is associated with severe, early CVD. Mutations causing low-HDL have also been identified in other genes, among them important components and modifiers of the HDL particle; *APOA1*[71], lecithin:cholesterol acyltransferase (*LCAT*)[72], and lipoprotein lipase (*LPL*)[73,74]. The most common forms of familial low-HDL however, are considered to be multifactorial and given the phenotypic overlap with FCHL, overlap in the genetic component to these two disorders most likely exists. In fact, variants in the gene encoding for apolipoprotein A5 (*APOA5*) have been associated with both low-HDL[75], as well as with FCHL[76-78].

2.2.2 The metabolic syndrome

The metabolic syndrome refers to a clustering of several metabolic risk factors in one person[79]. Not all physicians agree on the definition of the metabolic syndrome, perhaps appropriately also known as "syndrome-X." According to the International Diabetes Federation consensus published in 2005, the diagnostic criteria for the metabolic syndrome are (for Europeans) a waist circumference ≥94cm (♂), ≥80cm (♀) (indicating visceral fat accumulation) and at least two of the following:

Table 2. Diagnostic criteria for the metabolic syndrome.

Phenotype	**Diagnostic Criteria**
Serum triglycerides	≥1.70 mmol/l
High-density lipoprotein (HDL)	<1.03 mmol/l (♂), <1.29 mmol/l (♀)
Blood pressure	Systolic ≥130 mmHg, Diastolic ≥85 mmHg
Fasting blood glucose	≥5.6 mmol/l

The central defect (pun intended) in the metabolic syndrome is visceral obesity and the associated insulin resistance[80]. The risk for type 2 diabetes is larger among individuals with the metabolic syndrome, than could be predicted from the individual risk factors alone. Just as with type 2 diabetes, the prevalence of the

metabolic syndrome has risen dramatically over the last decades and presents a tremendous burden on the health care system.

2.2.3 Type II diabetes

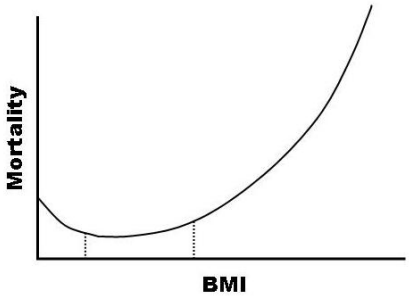
Also called diabetes mellitus type II and non insulin-dependent diabetes (NIDDM), type 2 diabetes (T2D) is a metabolic disorder very closely related to the metabolic syndrome. Both are characterized by insulin resistance and a relative inefficiency of the pancreas to secrete insulin. The dyslipidemia in T2D while similar to the metabolic syndrome, however, is more pronounced[81]. The age of onset for T2D is in general in adult age (although it is now also occurring in obese children) and the symptoms appear gradually, while the average age of diagnosis for type 1 diabetes is 14 years and the symptoms are more severe and the onset is more rapid. Type 1 diabetes results from the inability of the pancreatic β-cells to produce insulin, while in T2D insulin production is usually normal, or even increased.

Worldwide, T2D accounts for 90% of people with diabetes (WHO) and as the metabolic syndrome, is largely due to the increasingly sedentary lifestyle and obesity in the Western world. While environmental factors play the major role in the susceptibility to T2D, much research into the genetic component of the condition has also been done. Until now, not much progress has been made in uncovering these genes. Very few replicable findings have been made using the candidate gene, and positional cloning approaches, with exceptions being perhaps the genes *PPARγ*[40], *CAPN10*[82] and *KCNJ11*[83]. However, very recent findings from unbiased, genome-wide association studies have revealed replicable association to a novel T2D gene; *TCF7L2*[84]. This new finding represents some of the promise that is hoped from the study of complex disease genes; TCF7L2 is involved in the *wnt*-signalling pathway important for the development of the enteroendocrine system[85]. This pathway had not previously been considered to be potentially involved in T2D and

thus opens up a whole new, perhaps druggable, pathway and several interesting candidate genes. The best associating risk SNP variant in TCF7L2 confers a relative risk for developing T2D of 1.45 in heterozygotes and 2.41 in homozygotes –quite high for a complex disease variant[84]. Interestingly, the risk variant here is very common (with a frequency of 27.6% in the controls and 36.4% in the cases) perhaps suggesting that the variant had undergone positive selection in the past and may thus fit the thrifty gene hypothesis. In addition, variants in this gene have also been reported to be associated with the triglyceride component trait of FCHL[86].

2.3 Obesity and dyslipidemia

Considering the enormous quantity of food ingested, and energy expended in a year, it is an amazing feat by the systems maintaining homeostasis in the human body that body weight remains essentially constant. Most fundamentally, obesity is the result of a misbalance between excessive energy intake and insufficient expenditure. Extra energy in the time of plenty is stored by the body to be used at a time of scarcity (that in today's society seldom comes). Most of the energy storage is in the form of triglycerides in adipocytes and it is released from these depots as necessary. The amount of stored fat varies between individuals and a crude, but practical indicator of the amount of stored fat is the body mass index (BMI), which is the bodyweight in kilograms divided by the square of the height in meters. Values above 25 kg/m^2 are considered overweight and values above 30 kg/m^2 obese. Considerable morbidity and mortality is associated with the extremes of BMI, as can be observed from the J-shaped graph of mortality versus BMI (Figure 12, left).



Obesity is a problem because of the complications that are associated with it – among them atherosclerosis, hypertension and certain cancers. Important to the development of the morbidity associated with obesity is not only the amount of stored fat, but importantly *where* that fat is stored. Much more dangerous than subcutaneous fat, large amounts of visceral fat (that give some men the characteristic apple-shaped body) are associated with larger risk for disease[87]. Along with visceral obesity comes insulin resistance that can develop into type 2 diabetes. Dyslipidemic features that result from obesity are elevated levels of triglycerides, total cholesterol, free fatty acids and decreased levels of HDL –all atherogenic changes that predispose to coronary heart disease[79].

Adipose tissue is not merely the inert storage tissue that it was considered for a long time, but rather is now recognized as an important endocrine organ that actively produces hormones and cytokines that not only signal to the brain the amount of stored fat, but have profound impacts on the whole of human physiology. Leptin is the now classical adipose tissue derived hormone that acts as a lipostat, signaling to the satiety center in the hypothalamus to decrease food intake[88]. While considerable hope was initially placed on the potential of using exogenous leptin to treat obesity, no marked success has become of this approach. Rare cases of early-onset, very severe obesity are caused by inactivation of the leptin gene or its receptor[89], but it doesn't seem that common variation in the gene contributes to adiposity at the population level. Several other monogenic forms of obesity are known, but the search for common variants that affect BMI at the population level has only very recently reported success in identifying these common variants. Association of common variants in the FTO (fat mass and obesity associated) gene with obesity were recently reported by studies using the genome-wide association approach. Exemplifying the intimate relationship between obesity and T2D, the initial association was found for T2D, but upon closer inspection it was evident that the association with T2D was driven by the gene's effect on BMI[90]. The role of this

gene behind common human obesity has since been replicated in multiple studies[91-96] and such, *FTO* represents the first well replicated gene behind common forms of obesity.

It makes logical sense that some of the genes contributing to obesity would be expressed in the tissue most affected[97]. Genes involved in a number of processes taking place in adipocytes could potentially harbor genetic variation causing obesity. Some of these processes include adipogenesis, lipid turnover, endocrine/autocrine functions as well as mitochondrial respiratory functions[98] (Figure 13). The adipocyte mitochondria are only recently beginning to gain recognition as potentially very important factors in the development of obesity and insulin resistance[99].

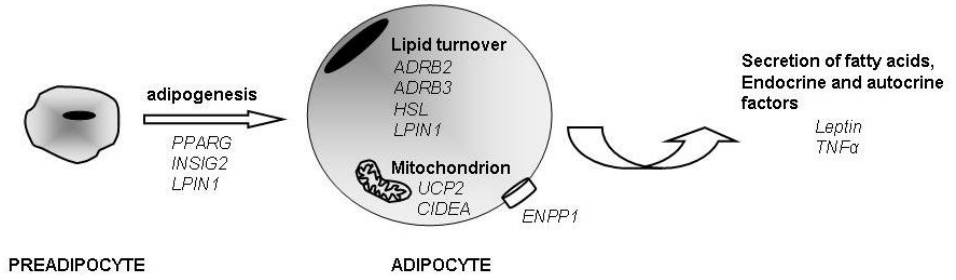


Figure 13. Genes involved in the regulation of adipocyte maturation, function and hormonal signaling may be involved in the predisposition to obesity. Modified from[98].

2.4 Relation to severe cardiovascular disease

Coronary heart disease is the major cause of death in the Western world today[100]. The significant role of cholesterol in the pathogenesis of atherosclerosis was first described by a Russian pathologist Nikolai Nikolajewitsch Anitschkow in a seminal paper in 1913. By feeding cholesterol to rabbits, he was able to induce vascular lesions closely resembling those found in humans. He described that these lesions contained cholesterol-laden foam-cells and speculated that the cholesterol was

entering the vessel wall from the blood. Nearly 100 years ago he famously stated that "there is no atherosclerosis without cholesterol", but it took more than 30 years before serious research following up his findings got underway. Considerable controversy ensued over the role of elevated levels of serum cholesterol in the development and progression of atherosclerosis. From the current vantage point, where nearly everyone is aware of the "good" and "bad" cholesterol and even know their own levels, it can be hard to accept that as recently as the late 1980s there were still vocal critics of the cholesterol-atherosclerosis idea. Since then an enormous body of epidemiological, pathological, genetic data and clinical observations has established dyslipidemias as independent risk factors for developing CHD[101].

The current understanding of the development of atherosclerotic lesions begins with excess LDL particles with their load of cholesterol, accumulating in the intimal layer of the artery wall. There they undergo chemical alterations and stimulate the endothelial wall to recruit monocytes from the circulation. These monocytes enter the vessel wall and mature there into macrophages that begin ingesting the excess LDL particles. Macrophages scavenging on the LDL particles become filled with these lipid droplets, taking on the appearance that gives them the name "foam-cell"[102]. Other immune cells are also recruited into the lesion, bringing about the inflammatory response –now understood to be a key component in the development of the plaque[103]. Chemical messages, cytokines, secreted by the inflammatory components of the lesion induce the smooth muscle cells of the media to migrate and form a fibrous cap over the lesion. As the lesion progresses adluminally, it can slowly occlude blood flow to the tissues that it serves. When this occurs in the coronary arteries, the individual may be symptomless at rest, but during physical exercise when the demand for oxygenated blood rises sharply, the resultant tissue hypoxia is experienced as angina pectoris. A large plaque may rupture and the resulting hematoma can very quickly completely occlude the vessel, resulting in an

infarct (Figure 14). This same process, when occurring in the blood vessels that supply the brain, results in stroke.

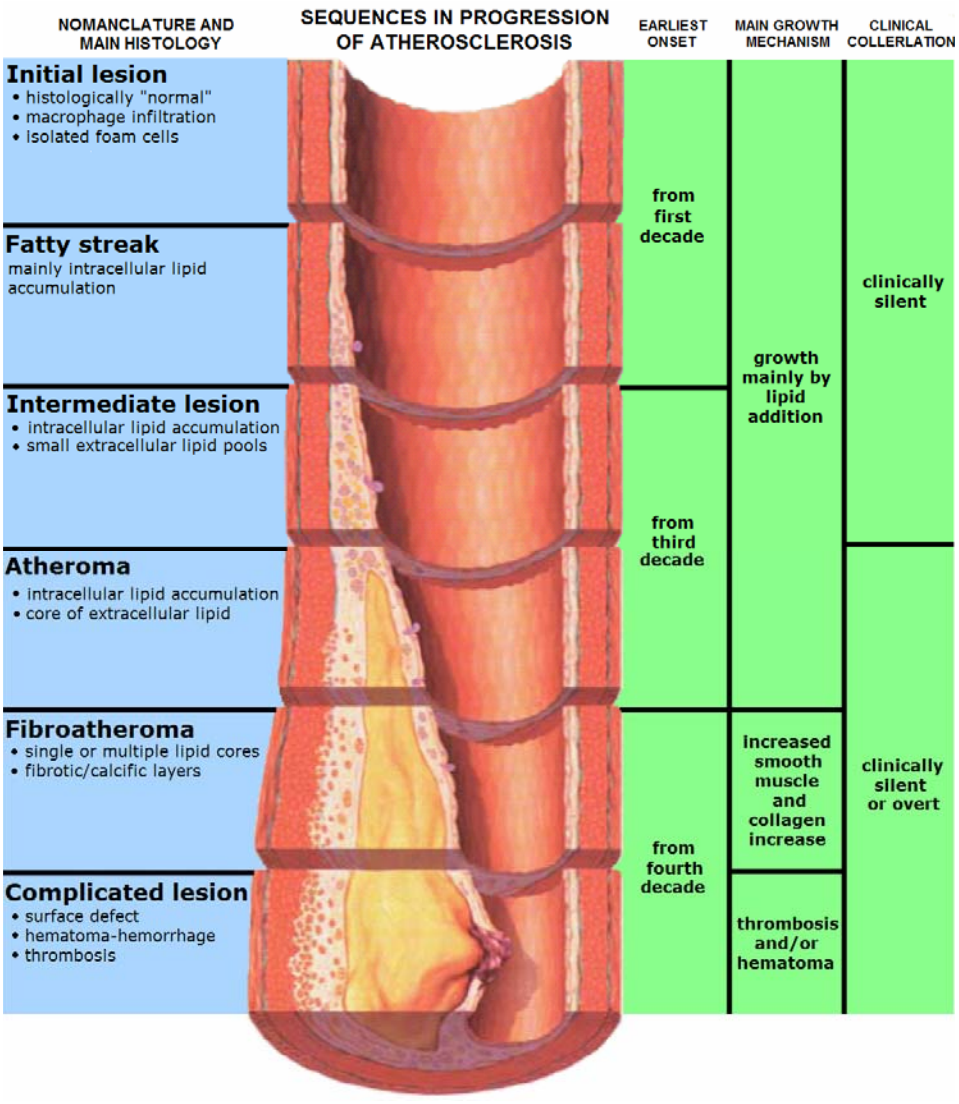


Figure 14. Progression of atherosclerosis. Accumulation of lipids and a recruitment of inflammatory cells into the intimal layer of arteries forms the lesions that are the hallmark of atherosclerosis. (Figure obtained from Wikimedia Commons / Atherosclerosis.)

3. Genetics of familial combined hyperlipidemia

3.1 Genetic loci identified and associated genes

A whole host of enzymes, transporters, channels and receptors are involved in the careful control of lipid metabolism and it is in theory possible that a change in any one of these components can result in the sorts of dyslipidemia as are observed in FCHL. To date, three genome-wide scans have been performed in a non-biased attempt to identify genes underlying the complex FCHL genotype. The scans performed in the Finnish[104], Dutch[105] and British[106] have identified several loci that likely harbor disease—associated genes. Three loci have been consistently replicated; 1q21-23, 11p and 16q22-24.1. In addition to genome-wide scans, multiple studies have evaluated the role of various candidate genes (Table 3).

Table 3. Genes with reported involvement with FCHL.

Gene	Chromosome	Type of Study	References
LPL	8p21.3	Family-based association	107
LCAT	16q22.1	Linkage	108
PPARA	22q13.31	Association	109
TNFRSF1B	1p36.22	Linkage and association	110
APOC3	11q23.3	Linkage, family-based association	111
APOA5	11q23.3	Association	76-78,112,113
HNF4A	20q13.12	Association	114
LEPR	1p31.3	Association	115
USF1	1q23.3	Linkage and association	116-120

(Modified from Naukkarinen, Ehnholm and Peltonen 2006)[121]

3.2 The 1q21-23 locus and dyslipidemia

The initial genome scan done for FCHL in the Finnish population identified a strong linkage signal in the 1q21-q23 locus[104]. Interestingly, the same issue of Nature

Genetics also reported that a region syntenic to the human 1q21-23 locus was linked in a mouse model of combined hyperlipidemia[122]. This raised considerable interest, for if the same mutation was behind the combined hyperlipidemia phenotype in both human and mouse, it would significantly simplify the identification of the gene. The linkage of FCHL to this 1q21 locus was since replicated in German and Chinese-[123], U.S. white-[124] and Mexican[125] FCHL families making it the best replicated FCHL locus. In addition, this region is also considered perhaps the best replicated T2D locus with convincing evidence of linkage in populations as diverse as Europeans[126-130], East-Asians[131,132], Native-Americans[133] and African-Americans (Steven C. Elbein, unpublished observation in[134]).

4. Upstream transcription factor 1 (USF1)

4.1 Association with FCHL

In 2004, sequencing of regional candidate genes and fine mapping with SNPs led to the 1q21 locus yielding the first FCHL associated gene. Multiple non-coding SNPs in the gene encoding upstream transcription factor 1 (*USF1*) were both linked and associated with FCHL and many of its component traits in 60 extended Finnish FCHL families[116]. It was also found that the *TXNIP* gene –the human orthologue of the *Hyplip1* gene shown to cause combined hyperlipidemia in the mouse, was not behind the disease in man. The strongest association was obtained with two SNPs in complete LD with eachother, located in the 3'UTR (rs3737787) and in intron 7 (rs2073658) (Figure 15). The association signal extended into the neighboring *F11R* (also known as *JAM1*) gene, especially among high triglyceride men. Early functional evidence for the involvement of the intronic SNP rs2073658 was also presented, as a 60bp intronic sequence containing this variant had an effect on transcription *in vitro* in the forward orientation. No effect of the SNP variants on the

transcription of the *USF1* gene itself could be identified in adipose tissue samples, but an overall effect on the transcript profiles could be identified.

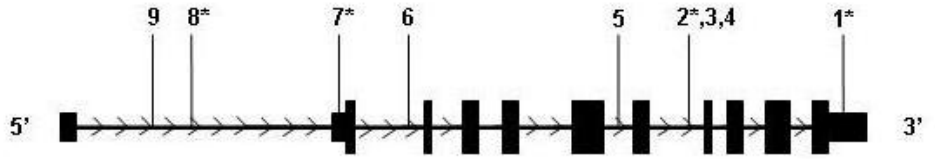


Figure 15. Schematic of the 6.7kb *USF1* gene. SNPs genotyped in the original study implicating USF1 in FCHL are marked with numbers. Associated SNPs are identified with asterisks.[116]

4.2 Central role in lipid- and glucose metabolism

The carefully regulated control of genes involved in the metabolism of lipids and glucose is essential in effectively responding to different nutritional conditions brought upon by fasting/refeeding. USF1 occupies a central position in the transcriptional regulation of a number of genes involved in lipid and glucose metabolism (Table 4), as well as components of the inflammatory- and stress-response. Together with a related transcription factor USF2 with which USF1 forms a functional heterodimer, they bind E-box sequences (CACGTG) in the promoters of their target genes and thus stimulate their transcription[135,136]. Perhaps the most important signal of changes in an individual's nutritional status is insulin secreted by the pancreas in response to a consumed meal. The ability of the USF transcription factor complex to stimulate the transcription of their target genes is influenced by this insulin signal and it makes possible the intricate control of important insulin

responsive genes such as fatty acid synthase (*FASN*)[137], acetyl-CoA carboxylase α (*ACACA*)[138] and apolipoprotein A5 (*APOA5*)[139].

USF1 is a ubiquitously expressed, evolutionarily conserved transcription factor of the basic helix-loop-helix leucine zipper family and has two known splicing variants [140]. The *USF1* gene spans 6.7kb, has 11 exons and codes for 310 amino acids. Knock-out experiments in mice have established that while either of the USF genes can be absent and still produce a viable animal, USF1/USF2 double knock-out mice display an embryonic-lethal phenotype –highlighting their important role in development[135]. In line with the known function of USF1, knock-out mice lacking this gene displayed a severely delayed induction of the FASN gene by refeeding a carbohydrate-rich diet[141]. Considering the ubiquitous expression, wide-ranging effects and evolutionary conservation (99% conservation between human and mouse/rat at the protein level) it is perhaps no surprise that no coding variants have been identified in the *USF1* gene.

Table 4. Selected USF1 target genes

Gene	Symbol
Apolipoprotein C3	*APOC3*
Apolipoprotein A2	*APOA2*
Apolipoprotein A5	*APOA5*
Apolipoprotein E	*APOE*
Hormone sensitive lipase	*LIPE*
Fatty acid synthase	*FASN*
Spot 14 protein	*Spot-14*
Acetyl-CoA carboxylase alpha	*ACACA*
Acetyl-CoA carboxylase beta	*ACACB*
ATP-binding cassette, subfamily A1	*ABCA1*
Osteopontin	*SPP1*
Renin	*REN*
Angiotensinogen	*AGT*
Glucokinase	*GCK*
Glucagon receptor	*GCGR*
Ghrelin	*GHRL*
Liver-type pyruvate kinase	*L-PK*

4.3 Role of USF1 in other lipid disorders and CVD

Following the first report of *USF1* association with FCHL and its component traits, a number of studies utilizing independently collected FCHL cohorts of different ethnic background attempted to replicate the association. Studies in Utah Caucasians[117], Mexicans[118] and Dutch[119,120] FCHL families all reported significant association of USF1 SNPs with the FCHL phenotype. Given the known function of USF1, it represents an attractive candidate gene not just for FCHL, but for many other dyslipidemias and related conditions. Several studies have thus also evaluated the role of *USF1* in data sets not ascertained for the FCHL phenotype[120,134,142-149]. It should be noted however, that two large scale studies ascertained for type 2 diabetes were not able to find evidence for the involvement of *USF1* in the disease[134,147] suggesting that the 1q21 locus harbors multiple disease susceptibility variants that underlie the repeated linkage of this region to type 2 diabetes and USF1 may not alone explain the linkage.

AIMS OF THE PRESENT STUDY

This thesis aimed at elucidating the molecular background of common dyslipidemias, such as FCHL and familial low-HDL by addressing the following specific aims:

I. Identify the dyslipidemia predisposing gene(s) underlying the multiple reports of linkage to the 1q21 locus.

II. Utilizing fat and muscle biopsies from patients, to elucidate the functional relevance of the variant(s) identified by correlating them with changes in global transcript profiles.

III. To investigate the role of the previously identified dyslipidemia gene, *ABCA1* in the pathogenesis of familial low-HDL by examining the reverse cholesterol transport process in patient and control samples.

IV. Using a special monozygotic twin resource to address biological pathways of acquired obesity –a strong modifier/confounder of the dyslipidemia phenotype.

MATERIALS AND METHODS

1. Study samples

1.1 Dyslipidemic family collection

The three first publications making up this thesis studied participants from large, multigenerational FCHL and low-HDL families collected as a part of the European Multicenter Study on Familial Dyslipidemias in patients with Premature Coronary Heart Disease (EUFAM)[150], begun in 1995. These families were recruited in the Helsinki, Turku and Kuopio University Central Hospitals and the collection consisted of three phases. First, the probands were selected from among patients undergoing elective coronary angiography or from a registry of patients with confirmed myocardial infarction (MI). In the second phase the probands and their 1st degree relatives were examined to identify low-HDL and FCHL families, followed in the third phase by examination of all accessible relatives and spouses. From among these families, a total of 77 fat biopsies were used in the first two studies of this thesis (19 in study I, 58 in Study II). Study III utilized a total of 72 individuals that included low-HDL family subjects (28 affecteds and 19 healthy relatives) and 25 healthy control subjects. All samples were collected in accordance with the Helsinki declaration and the ethics committees of the participating centers approved the study designs presented here.

1.2 FCHL families

The inclusion criteria for collection of the FCHL probands were as follows:

- Age 30-55 for males and 30-65 for females
- ≥ 50% stenosis in one or more coronary arteries, or verified MI
- Serum TC ≥ 90^{th} age- and sex specific Finnish population percentile, or TG ≥ 90^{th} percentile, or both.

Exclusion criteria:

- Type I diabetes mellitus
- hepatic or renal disease
- hypothyroidism
- FH

1.3 Low-HDL families

The inclusion criteria for collection of the low-HDL probands were as follows:

- Age 30-60 years
- ≥ 50% stenosis in one or more coronary arteries, or verified MI
- Serum HDL-C level below the 10^{th} age- and sex specific Finnish population percentile
- Additional lipid criteria: TC < 6.3 mmol/l in men and < 6.0 in women, TG < 2.3 mmol/l for both genders

Exclusion criteria:

- Type I or II diabetes mellitus
- hepatic or renal disease
- hypothyroidism
- FH
- BMI > 30 kg/m^2

The age- and sex specific Finnish population based cut-off values were based on the distribution of serum lipid values observed in the population survey FINRISK[151,152] and are listed below in table 5 for serum total cholesterol (TC) 90^{th} percentile, triglycerides (TG) 90^{th} percentile and high density lipoprotein (HDL) 10^{th} percentile. All serum measurements are reported in mmol/l.

Table 5. Population-based cut-off values for various serum lipids in mmol/l.

Age	TC Men	TC Women	TG Men	TG Women	HDL Men	HDL Women
5-11	5.9	6.2	1.6	1.4	1.0	1.1
12-14	6.0	5.8	1.7	1.6	1.0	1.1
15-17	5.3	5.6	1.9	1.5	0.9	1.0
18-20	5.9	6.0	2.0	1.7	0.9	1.1
21-24	5.7	5.9	2.3	1.6	0.8	1.1
25-29	6.2	6.1	2.4	1.7	0.9	1.1
30-34	6.6	6.2	2.7	1.7	0.9	1.1
35-39	7.0	6.4	2.9	1.8	0.9	1.1
40-44	7.2	6.6	3.2	1.8	0.9	1.1
45-49	7.4	6.8	3.4	1.9	0.9	1.1
50-54	7.5	7.1	3.5	2.1	0.9	1.1
55-59	7.5	7.3	3.5	2.3	0.9	1.1
60-	7.4	7.6	3.5	2.5	0.9	1.1

1.4 Twin samples

1.4.1 Monozygotic twins discordant for BMI

The participants in the fourth publication of this thesis work investigating the effects of acquired obesity were recruited from a population-based longitudinal study of five consecutive birth cohorts (1975-1979) of twins, their siblings and parents (N=2,453 families), identified through the national population registry of Finland[153]. Twin pairs included in the current study were enrolled based on their responses to questions on weight and height at age 23-27 years. From this cohort (Figure 16) the top 5% most obesity-discordant monozygotic twin pairs (one co-twin non-obese (BMI ~25 kg/m^2), and the other one obese (BMI ~30 kg/m^2)), with no significant height differences (<4 cm) were selected.

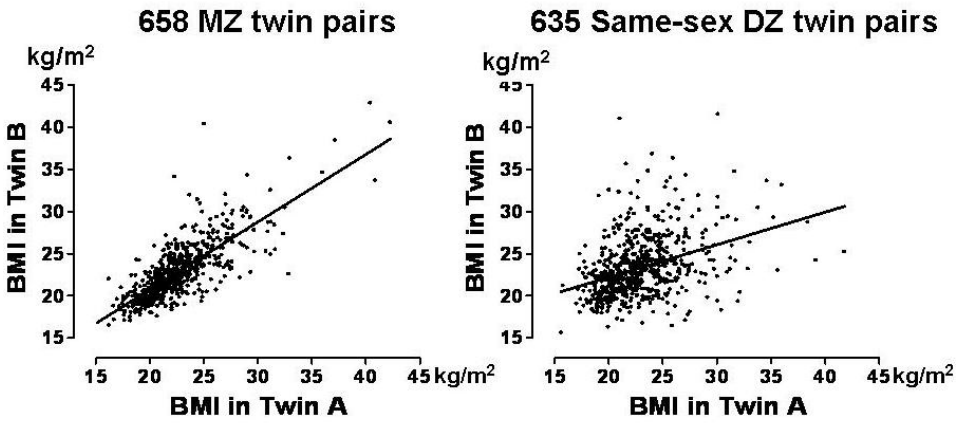


Figure 16. Between-twin correlation of BMI measurements. Figure provided by Dr. Kirsi Pietiläinen.

Screening all MZ twin pairs (N=658) resulted in identification of 18 pairs above the 95th percentile of BMI differences (3.1 kg/m^2). Fourteen of these pairs (eight male and six female pairs) were willing to participate and thirteen pairs (eight male and five female pairs, BMI differences 3.3-10.0 kg/m^2, height differences <3 cm) had adipose tissue samples available for the present study. All pairs were Caucasian, and their mean age was 25.8 (range 23.7-28.5) years. The subjects were healthy (based on medical history, clinical examination and structured psychiatric interview), normotensive and did not use any medications except contraceptives. Their weight had been stable for at least 3 months prior to the study. In addition, seven MZ twin pairs concordant for BMI (median BMI for Twin A 24.6 kg/m^2 vs Twin B 25.4 kg/m^2) recruited from the same cohort were studied as controls in the mitochondrial copy-number analyses.

1.4.2 Danish Twin Registry

In the second publication, subjects for the muscle biopsy were identified through The Danish Twin Register and selected as previously described[154,155]. A total of 98

young (aged 25-32 years) and elderly (aged 58-66 years) twin pairs were included in the clinical examination. Both genotype and skeletal muscle *USF1* mRNA expression levels were obtained from 71 of the twin pairs (58 younger monozygotic, 32 younger dizygotic, 22 elderly monozygotic and 30 elderly dizygotic twins).

2. Laboratory methods

2.1 Table of methods used in the study

Descriptions of methods used in this study can be found in the original publications accompanying this thesis, as described in Table 6.

Table 6. Materials and Methods used in this thesis study.

Material or Method	Original Publication
Study samples	
FCHL families	I, II
Low-HDL families	II, III
Finnish MZ twins	IV
Danish MZ and DZ twins	II
Phenotyping methods	
Serum lipid measurements	I-IV
Euglycemic hyperinsulinemic clamp	IV
Dual energy X-ray absormetry (DEXA)	IV
Magnetic resonance imaging (MRI)	IV
Carotid artery ultrasonography	III
Laboratory methods	
DNA extraction	I-IV
Polymerase chain reaction (PCR)	I-IV
Agarose gel electrophoresis	I-IV
Quantitative real-time PCR	II- IV
Genotyping	I-IV
Sequencing	II, IV
Fat biopsies and RNA extraction	I, II, IV
Muscle biopsies and RNA extraction	II
Electrophoretic mobility shift assay (EMSA)	I
Enzyme linked immunoSorbent assay (ELISA)	III, IV
Liquid chromatography	IV
Macrophage cholesterol efflux assay	III
Proton spectroscopy	IV
Statistical methods	
Marker selection	III
Haplotype construction	III
Association analyses	III
Regression analysis	III, IV
Expression analyses	I-IV
Pathway analysis	IV
Permutation	IV
Correlation analysis	I-IV
Transcription factor binding site prediction	II

RESULTS AND DISCUSSION

1. Scanning the *USF1* genomic landscape for a functional variant

A previous study by our group identified four non-coding, intergenic SNPs associated with FCHL and its component traits in the *USF1* gene located in chromosome 1q23.3. Typical of complex disease associated genes, none of the associated polymorphisms presented any obvious functional consequence that would explain the disease phenotype. Two variants in complete LD gave the strongest association signal and were thus considered the likeliest candidates for representing the functional polymorphism[116]. One is located in the 3'UTR (rs3737787 aka usf1s1) and the other in intron 7 (rs2073658 aka usf1s2). The complete LD made their prioritization by genetic means impossible and thus in study I, a comparative genetics approach was taken. The evolutionary conservation of all associating variants was evaluated by exploiting the sequence information of other species available at the time. Comparison of sequences surrounding the associating variants was made between human, chimp, dog, mouse, rat and chicken, revealing that the intron 7 SNP usf1s2 was located in a short patch of sequence conserved through all the species (Figure 17). Interestingly, the variant nucleotide itself was conserved and the ancestral allele represents the associated risk-allele. Sequence around the 3'UTR SNP usf1s1 was not itself conserved, despite being located in a transcribed region.

The possible functional role of the sequence around variant usf1s2 in intron 7 was suggested already in the original work by our group reporting the association with FCHL, as a 268bp segment (including this variant) enhanced the transcription of a reporter gene *in vitro*[116]. Based on this and the observed conservation, it was possible that that the sequence contained an enhancer element and therefore in study I, the ability of the sequence to bind protein was tested in an electrophoretic mobility

shift assay (EMSA). Using two partially overlapping probes it was shown that the sequence immediately surrounding variant usf1s2 bound nuclear protein(s).

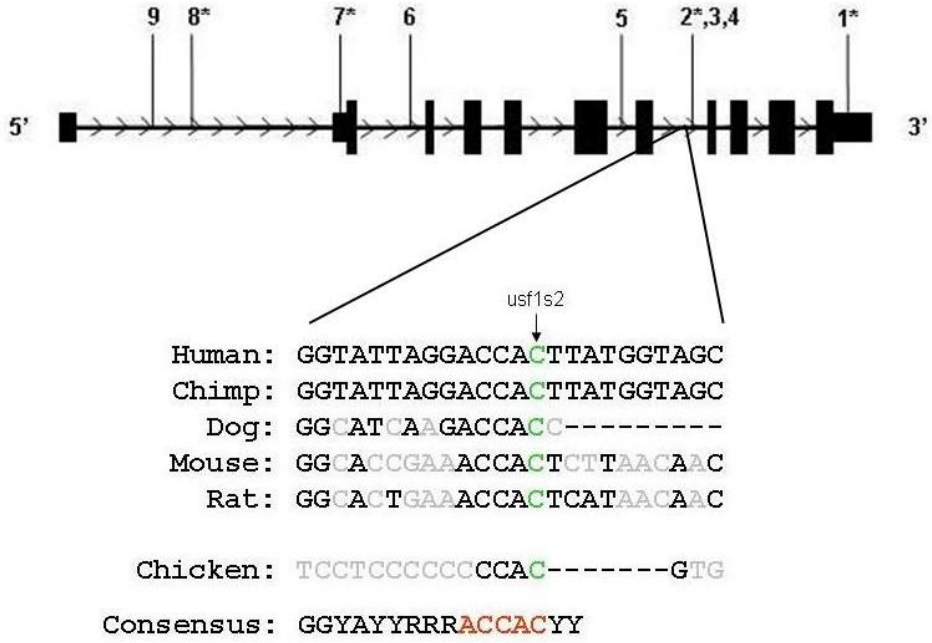


Figure 17. Conservation of sequence around variant usf1s2. Comparison of sequence available from multiple species shows the strongest associating variant usf1s2 to be located in a conserved patch of sequence in intron 7 of *USF1*. All identified variants are indicated by numbers with the associated variants marked by asterisks.

In the same issue of Nature Genetics as the original article by our group reporting the association of *USF1* with FCHL[116], a commentary by C. Shoulders suggested that an allele of the strongest associating SNP usf1s2 may activate an alternative promoter within *USF1*, resulting in the start of transcription from one of the two AUGs in exon 8 and thus a shorter protein product lacking the trans-activation domain[156]. This did not appear to explain the mechanism however, as in study I, *USF1* transcripts of equal length were produced in individuals carrying either allele.

2. Effects of *USF1* variants observed on numerous target genes

A qualitative or quantitative functional change in a transcription factor would be expected to be reflected on the expression pattern of its target genes under appropriate conditions. It is through the altered transcription of the numerous USF1 controlled lipid- and glucose metabolism related genes that the dyslipidemic phenotype associated with variants in USF1 is thought to arise. Our group also showed previously, first indications of this in a limited number of adipose tissue samples when an enrichment of lipid metabolism and immune response category genes differentially expressed in individuals carrying risk alleles (G) of usf1s2 was observed[116].

In order to determine the effect of different USF1 variants on the expression of known USF1 target genes, first in a collection of 19 fat biopsies (Study I), then in a larger set of 47 samples (Study II), known target genes were interrogated for differential expression in adipose tissue of individuals carrying different alleles of the disease associated SNP usf1s2. The two studies were carried out on different Affymetrix expression chips and thus had different number of genes detected. The second, larger study involving 47 fat biopsies were carried out on the newest U133 Plus 2.0 arrays and revealed that in a comparison of 24 carriers of usf1s2 risk allele (G) with 23 individuals homozygous for the non-risk allele (A), 10 of the 31 known USF1 target genes expressed in adipose tissue were significantly differentially regulated (Table 7). Of interest was the observation that those genes that were upregulated were mainly ones involved with the immune- or inflammatory response and that genes involved intimately in lipid metabolism predominated the list of downregulated genes –findings that agree with the dyslipidemic, pro-atherogenic state.

Table 7. Differentially expressed USF1 target genes in adipose tissue.

Gene	Fold Change (risk/protective)	P-value*
FCER1A	1.48	0.006
HLA-E	1.36	0.008
PTPN6	1.30	0.027
MAP2K1	1.24	0.007
HOXB4	0.84	0.002
THRSP	0.80	0.022
ABCA1	0.77	0.003
ACACB	0.68	1.6×10^{-4}
ACACA	0.53	0.001
AGT	0.54	0.002

*P-value for non-parametric Mann-Whitney U-test

While the list of known USF1 targets is growing, there are certainly many more yet to be identified. With the observed influence of the usf1s2 allele on the expression of known target genes, the *global* effect on gene expression in adipose tissue was evaluated in study II. To this end, first a list was generated of genes differentially expressed between the usf1s2 risk allele carriers and non-risk allele homozygotes. Of the 54,613 probes present, 30,155 passed initial filtering of the data set to exclude those genes with unreliably low signal intensity. Of this set, 4,847 probes (representing 3,693 unique gene products) were differentially expressed between the genotype groups (Welch t-test $P<0.05$). To test for the possible role of USF1 in controlling the expression of these genes, the oPOSSUM software was utilized[38]. This software combines a database of conserved transcription factor binding sites in human and mouse promoters with statistical methods in order to test whether the binding sites of certain transcription factors are over-represented in a given gene set. Of the 3,693 genes input, the software was able to analyze 2,897 and among this list of differentially expressed genes, binding sites for USF1 were significantly enriched ($P=2.8 \times 10^{-5}$). This implies a large-scale effect of the usf1s2 allele on the expression of numerous USF1 target genes that still remain to be identified.

3. Functional defect exemplifies an interaction with environment

Transcription factors control the expression of their target genes so that the functional proteins that they encode are appropriate in abundance, place and timing. When and where to activate transcription is in turn signaled by various environmental- and hormonal cues and for USF1 it has been shown that insulin is key in controlling the USFs post-transcriptionally[139]. Additional interest in the role of insulin in relation to USF1 was sparked by reports that carrier status of the usf1s2 risk allele had been associated with a decrease in antilipolytic insulin sensitivity[144] and a poorer response to oral glucose- and fat tolerance tests[149].

In study II, the potential effect of insulin on the expression of the *USF1* transcription factor itself was investigated in muscle biopsies from 142 Danish twins before and after a euglycemic hyperinsulinemic clamp. When analyzing all twins together, no change in the *USF1* transcript levels could be detected, however a clear difference could be observed when dividing the subjects based on their carrier status of the usf1s2 risk allele. While no differences in the expression level of *USF1* could be observed between the groups before the insulin administration, following the insulin challenge a dramatic increase in the transcript levels of *USF1* took place in the individuals homozygous for the non-risk allele (P=0.004). Those individuals carrying risk alleles of usf1s2 did not respond to the hyperinsulinemia (Figure 18). It seems that the risk-allele of *USF1* is unresponsive to insulin at the transcriptional level, while non-risk allele carriers respond by up-regulating their *USF1* expression. This provided the first allele-specific expression level difference in *USF1* and to our knowledge for the first time showed this transcription factor to be regulated by insulin also at the transcriptional level.

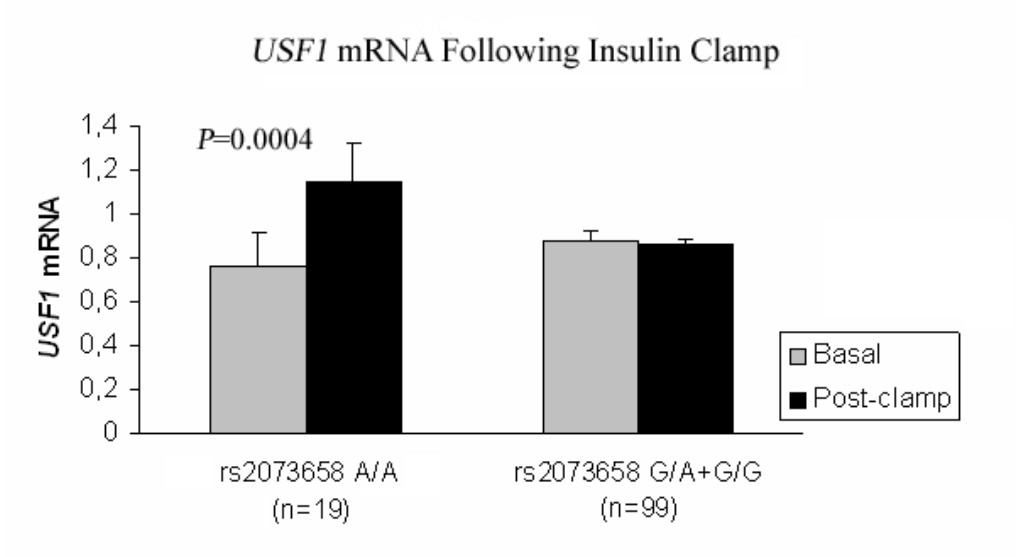


Figure 18. USF1 mRNA following insulin clamp. In muscle biopsies analyzed before and after a euglycemic hyperinsulinemic clamp, individuals carrying the risk allele of variant usf1s2 failed to upregulate their *USF1* transcript levels.

In addition, it could be observed that men displayed significantly higher USF1 mRNA levels compared to women (0.99 ± 0.06 [n = 71] vs. 0.72 ± 0.04 [n = 64], P_a= 0.002, and 1.0 ± 0.06 [n = 66] vs. 0.78 ± 0.05 [n = 59], P_a= 0.02, for basal and post-clamp *USF1* mRNA levels, respectively).

When one allele of a gene affects the expression of a gene, this may be detected in the heterozygous individual as allelic imbalance[157] (Figure 19). This requires the capability to accurately quantitate the two different alleles in a heterozygous sample and can be done by various different means, among them by quantitative sequencing[158]. Even relatively small differences in transcript levels can be detected, owing to the design of the experiment where in a heterozygous individual the other allele can serve as an internal control. In the absence of pre- and post- insulin clamp fat biopsies ascertained on the usf1s2 genotype, eleven adipose tissue samples from individuals heterozygous for usf1s2 were assayed for allelic imbalance of *USF1* by means of quantitative sequencing. Nine out of the eleven samples assayed exhibited

allelic imbalance of *USF1* with the risk allele of usf1s2 being under-expressed by ~20%. Importantly, there was a trend for higher imbalance (lower expression of the risk allele) with increasing levels of fasting serum insulin –as could be expected from the observation in muscle. Thus, in two relevant tissues the risk allele of *USF1* was expressed at a lower level compared to the non-risk allele and that this difference seemed to stem from an insensitivity of the risk allele to the inductive effect of insulin.

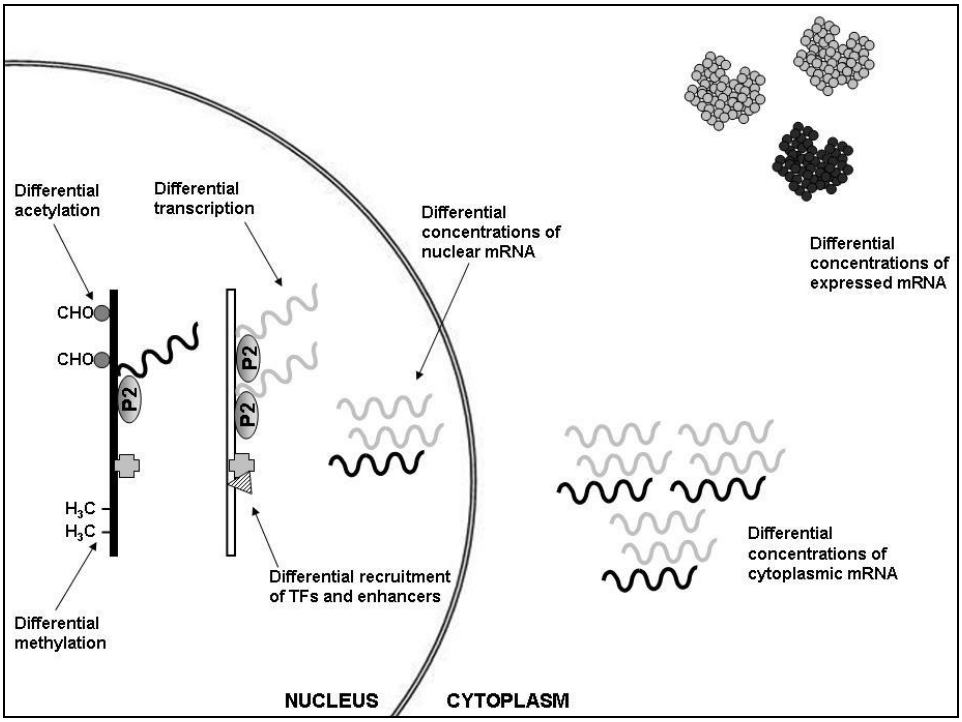


Figure 19. Mechanisms that can result in allelic imbalance. Certain polymorphisms can affect the acetylation/methylation pattern of genes, or the recruitment of transcription factors and enhancers and thereby result in differential transcription of the two alleles. In a heterozygous individual this phenomenon is termed allelic imbalance (AI) and can be detected using various techniques. Modified from[157].

The mounting data on the involvement of *USF1* variants in predisposition to dyslipidemia (particularly high triglycerides) and heart disease in multiple different populations has convincingly made it the most well-replicated dyslipidemia gene identified by the positional cloning method. While the genome-wide association studies performed on cardiovascular traits so far have not reported *USF1* among the top hits, its role in the pathogenesis of FCHL and related dyslipidemias especially in higher risk pedigrees is now well documented. The two strongest associating SNP variants of the original study have also been implicated in a majority of the subsequent reports, with the major allele of the variant rs2073658 associated increased susceptibility. Haplotypes carrying the minor allele of this variant however were deemed the risk variants in two population studies[134,146] perhaps highlighting the complex interplay of environmental and additional genetic variants and the difficulty this introduces in the absolute determination of risk alleles. Sequencing in the original- and subsequent studies has failed to identify any new variants within *USF1* that would present obvious functional potential and thus the results here described present the only functional evidence. While all the results point to the usf1s2 SNP being the functional polymorphism, the usf1s1 SNP located in the 3'UTR cannot conclusively be out-ruled, because of the complete LD between these two variants. It is also possible that polymorphisms within the neighboring *F11R* (a.k.a. *JAM1*) or close to it can contribute to the cardiovascular associations, given the fact that LD extends into this gene[116] and it too has recently been implicated in CVD, as a gene over-expressed in atherosclerotic plaques[159] and its levels in plasma associated with the severity of angiographically defined CAD[160]. Given the consistent linkage of this chromosomal region to T2D, yet the two negative reports on association[134,147] with *USF1*, it is likely that multiple disease predisposing variants will ultimately be found in this chromosomal region, in the vicinity of *USF1*. The role of USF1 in lipid metabolism and in the molecular background of FCHL however remains unchallenged. The possibly different role of disease predisposing variants of USF1 in men and women warrant further

investigation since a significant interaction between gender and *USF1* genotype was found in Dutch FCHL families and U.S. whites with CAD[120] as well as among women in the prospective population cohort of Finns (Silander *et al*., submitted). The observed difference in the transcript levels of *USF1* between the genders, as described here, may be involved. The overall incidence of CAD differs between men and women and surely there is a discordance in the battery of genes involved in the predisposition.

In conclusion, variation within the gene encoding transcription factor 1 (USF1) predisposes to FCHL, other related dyslipidemias and cardiovascular disease. The functional defect seems to be a variant within the binding site of an enhancer element within intron 7 that is involved in regulating the expression of *USF1* in response to insulin, as in a postprandial state. The transcript levels of *USF1* in carriers of the risk allele fail to up-regulate in the hyperinsulinemic state and this is in turn reflected in a delayed response of numerous relevant target genes. With time these numerous small perturbations lead to the permanent changes that we see in individuals who develop dyslipidemia and heart disease (Figure 20).

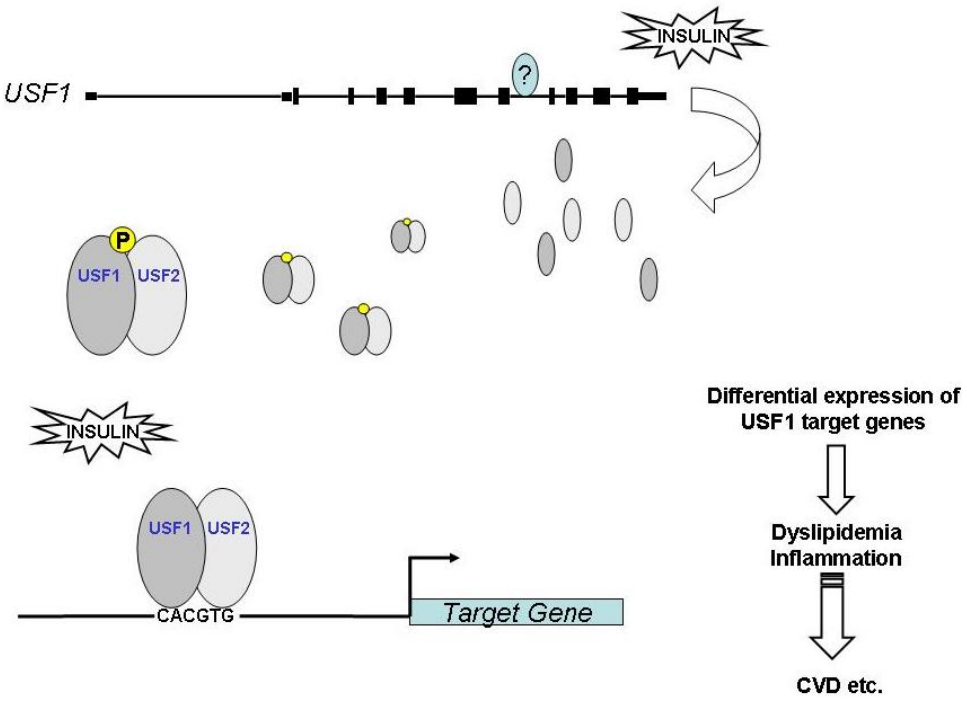


Figure 20. Schematic representation of the suggested mechanism through which genetic variation in the *USF1* gene predisposes to dyslipidemia and CVD.

4. Variation within the *ABCA1* gene and familial low-HDL

Rare monogenic forms of low-HDL have been identified where a defect in the gene encoding ATP-binding cassette, subfamily A1 (*ABCA1*) –another target gene of the USF1 transcription factor[161], causes a virtual absence of HDL-C in the plasma and a resultant accumulation of cholesteryl esters within cells[68-70]. In addition, it has been shown that up to 10% of people in the lowest HDL-C percentile are heterozygous for rare *ABCA1* mutations[162]. The possible role of common *ABCA1* variants in Finnish familial low-HDL was assessed in publication III of this thesis work. The role of *ABCA1* variants was examined at the level of the gene (genetic association), mRNA transcript (expression in patient derived macrophage foam-cells), functional

capacity (macrophage efflux capability to lipid-poor apoA-I) as well as clinical phenotype (correlation to several lipid measurements and carotid artery intima-media thickness). This study was carried out in a collection of affected family members and a control group including healthy volunteers and spouses with no CHD or lipid abnormalities.

Of 15 genotyped intragenic *ABCA1* SNPs, four variants displayed significantly different allele frequencies between the case and control groups with rare variants of three of them enriched in the low-HDL cases. Three of these variants were also directly associated with the levels of HDL-C (Table 8). Carrier status of the most common haplotype constructed of the three variants associated with HDL-C (L158, R219K and T1427) correlated with its levels in a dose-dependent manner: mean HDL-C level of individuals with two copies being 1.30 mmol/l, one copy 1.06 mmol/l, and no copies 0.85 mmol/l (P=0.025, Kruskal-Wallis test) (Figure 21). The likelihood of an individual belonging to the low HDL group when carrying either 0, 1 or 2 copies of this allelic haplotype were 80, 48, and 30%, respectively.

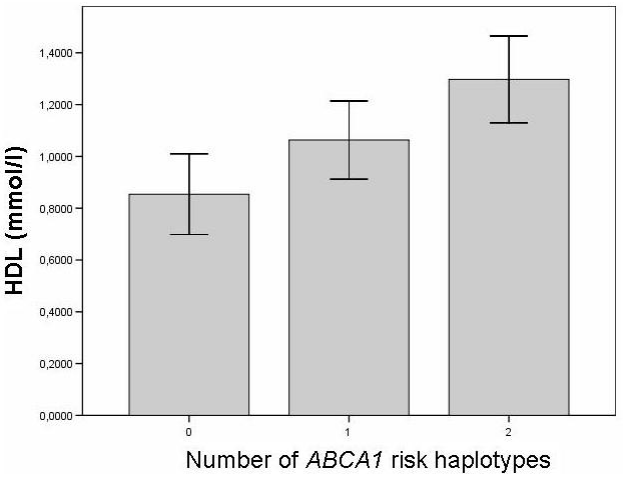


Figure 21. Dose dependent effect of the common *ABCA1* haplotype on HDL-C levels in serum. Error bars represent ±95th % confidence interval.

Table 8. ABCA1 allele frequencies between 47 low-HDL family members (19 unaffected and 28 affected subjects) and 25 control subjects, and their effect on cholesterol efflux, HDL-C and mean IMT in the pooled study sample (N=72).

SNP	AA residue	Allele frequencies and *P*-value in low HDL family members vs. controls	HDL-C (*P*-value)†	Mean IMT (*P*-value)‡
rs2472459	-	n.s.	n.s.	n.s.
rs2246293	-	n.s.	n.s.	n.s.
rs2515616	-	n.s.	n.s.	n.s.
rs1800978	-	n.s.	n.s.	n.s.
rs2740492	-	n.s.	n.s.	n.s.
rs3858075	-	n.s.	n.s.	n.s.
rs1929842	-	n.s.	n.s.	n.s.
rs2230805	L158	(0.45 vs 0.12) 0.004	0.009	n.s.
rs2230806	R219K	(0.40 vs 0.08) 0.005	0.007	n.s.
rs2487037	-	n.s.	n.s.	n.s.
rs2297409	-	(0.30 vs 0.04) 0.006	n.s.	n.s.
rs2066716	T1427	(0.11 vs 0.32) 0.053	0.003	n.s.
rs2230808	R1587K	n.s.	n.s.	n.s.
rs2066881	-	n.s.	n.s.	n.s.
rs4149341	-	n.s.	n.s.	0.025

Modified from[163].

In order to investigate the functional capacity of ABCA1 in patient samples, macrophages isolated from patients were loaded with acetyl-LDL to produce a foam cell model. These cells were then unloaded by incubating them with lipid-free apoA-1 as a primary cholesterol acceptor. Only a marginal, yet statistically significant difference could be observed in the cholesterol efflux capacity between cells derived from patients when compared with cells originating from healthy control subjects. Patient derived macrophages exhibiting a lower cholesterol efflux capacity, when normalized to macrophage protein content (dpm in medium/µg cell protein/18h) (Figure 22).

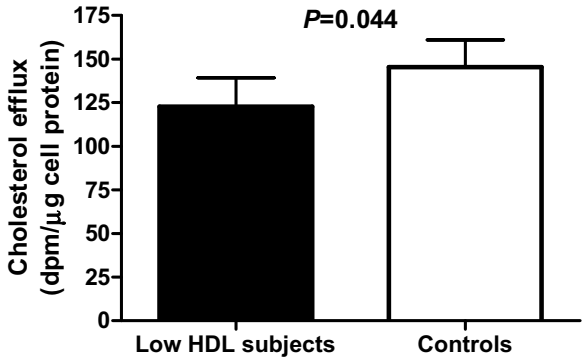


Figure 22. ApoA-1 mediated cellular cholesterol efflux in 22 low-HDL subjects and 21 controls. Error bars represent mean ± SEM. Modified from[163].

The transcript levels of the cholesterol transporters ABCA1 and ABCG1 are known to be up-regulated in response to cells becoming loaded with cholesterol and conversely to fall upon de-lipidation[164,165] and this was also observed in the Finnish samples. In the loaded macrophages, a distinct difference was observed in the transcript levels of *ABCA1*, with higher transcript levels somewhat surprisingly observed in the low-HDL group than in the control group ($P<0.001$). In light of this clear difference in *ABCA1* transcript levels, but only a marginal difference in efflux % to apoA-I, we next expressed efflux by cholesterol efflux to apoA-I (%)/relative *ABCA1* mRNA expression as a measure of the efflux efficiency of individual ABCA1 proteins. This ratio was significantly reduced in the low HDL group when compared to the controls (3.93±2.30 vs 7.85±2.14, $P<0.001$) (Figure 23).

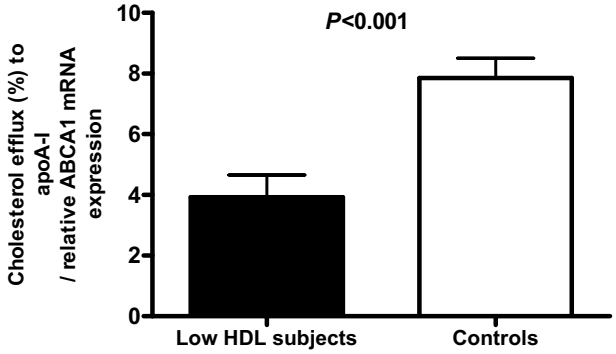


Figure 23. Cholesterol efflux efficiency in low-HDL subjects (N=10) compared with control subjects (N=11). Bars show mean ± 95% CI. Modified from[163].

The genotyped variants were also assessed for association with efflux efficiency of the patient macrophages. The L158 variant that had been associated with HDL-C levels also exhibited association with lower cholesterol efflux. Carriers of the rare alleles in L158 had decreased efflux (%) as compared to non-carriers in the whole study sample (4.6±1.0 vs 5.7±1.3, P=0.009). While evidently not the major defect behind low-HDL in these Finnish families, sub-optimal ABCA1 function contributes to the impaired reverse cholesterol transport process. The defect(s) resulting in low levels of HDL in these families may also be found in the acceptors of cholesterol and the several HDL modifying enzymes found in the circulation. As low-HDL is also considered a complex disease, the presence of multiple different defects, each with a small effect on the phenotype may be the likeliest scenario.

5. Consequences of acquired obesity –a confounding trait for dyslipidemias

As the single most common cause of dyslipidemia, understanding the physiological and molecular changes that occur in acquired obesity is of fundamental importance. Alongside the study of genetic predisposition to obesity and dyslipidemia,

describing these physiologic changes at the molecular level will aid in designing appropriate intervention strategies and treatments. In addition, the genetic- and physiologic studies complement each other: careful investigation of the acquired components of complex traits have the potential of exposing new pathways and candidate genes to evaluate in genetic studies and vise versa. Most studies on the effects of obesity have been hindered by the vast amount of environmental- and genetic variation inherent in human studies, so in study IV, a unique approach involving rare, monozygotic twins discordant for BMI was utilized to expose obesity induced changes in biological pathways in an identical genetic background. Fourteen pairs of monozygotic twins from the most discordant (5% extreme) in BMI in a national registry of 2,453 twin pairs were selected for analysis (Figure 10). In addition to well-documented changes in body morphology and dyslipidemia, a number of novel perturbations were observed in the obese co-twins.

A global analysis of adipose tissue gene expression pathways (as described by the GO-consortium) revealed that fat cells in an obese individual are not merely larger in size, but that a number of active, pathological processes are evident from the transcript profiles. An activation of inflammatory pathways predominated the list of up-regulated pathways (Table 9) underscoring at a transcriptional level the fact that obesity in large part is an inflammatory condition of fat tissue. An activation of both the adaptive- and innate immune systems could be observed as well as an increased transcription of several cytokines (*MCP-1, MCP-2, MIP-α-R, MIP-4, MIR-10*) that play roles in the recruitment of monocytes to sites of injury and infection. The single most over-expressed (5.9-fold) gene in the obese co-twins was osteopontin (*SPP1*), another cytokine involved in macrophage recruitment and stimulation of T cell proliferation during inflammation[166] and a known target gene of the *USF1* transcription factor[167]. In addition to the findings presented here, a very recent study on diet-induced obesity in mice also suggests that SPP1 may play a role in linking obesity to the development of insulin resistance by promoting inflammation and

accumulation of macrophages in adipose tissue[168]. Secretion of SPP1 was increased during obesity and unlike wild-type mice, animals lacking *SPP1* were protected from developing insulin resistance despite the diet-induced obesity. While the adipocytes themselves were expressing an increasing amount of SPP1 in obesity, the increase in *SPP1* mRNA in whole adipose tissues was in most part due to an increased of transcript levels in macrophages[168]. These results paint a picture where an increased degree of adipocity leads to secretion of SPP1 by the adipocyte, attracting macrophages from the circulation, which in turn begin secreting more SPP1 leading to an accelerating accumulation of adipose tissue macrophages. The presence of macrophages in adipose tissue has been shown to directly interfere with insulin signaling and insulin-stimulated glucose uptake in adipocytes by decreasing *GLUT4* and insulin receptor substrate 1 (*IRS-1*) expression[169]. Together with the concurrent report in the mouse model of diet-induced obesity, the novel observation in study IV of the dramatic increase of *SPP1* expression in obese humans for the first time links this cytokine as an important player in the development of insulin resistance. Over-expression of *PPARγ*, or treatment with a thiazolidinedione has been shown to suppress *SPP1* transcription[170] and *SPP1* deficient mice are protected from obesity associated insulin resistance[168], raising the possibility that specific targeting of this secreted cytokine could present a novel treatment strategy for type 2 diabetes.

Table 9. Top 10 most upregulated pathways in adipose tissue of obese-cotwins.

Pathway Name	Nominal *P*	Empirical *P**
Antigen processing and presentation of exogenous antigen	2.99E-11	0.0001
Inflammatory response	1.16E-10	0.0001
Humoral immune response	1.31E-10	0.0001
Locomotory behaviour	1.00E-09	0.0001
Immunoglobulin mediated immune response	1.06E-09	0.0001
Actin cytoskeleton	3.30E-09	0.0001
Antigen presentation	5.02E-09	0.0001
Chemotaxis	1.26E-08	0.0001
Taxis	1.26E-08	0.0001
Cellular defense response	5.88E-07	0.0001

Table 10. Most down-regulated pathways in adipose tissue of obese co-twins.

Pathway Name	Nominal *P*	Empirical *P**
Branched chain family amino acid metabolism	5.42E-08	0.0001
Negative regulation of transcription	4.81E-07	0.0001
Ligase activity, forming carbon-carbon bonds	2.21E-06	0.0001
Branched chain family amino acid catabolism	6.26E-06	0.0002
CoA carboxylase activity	7.11E-06	0.0002
Water-soluble vitamin metabolism	8.29E-06	0.0004
Muscle development	9.97E-06	0.0005

*Empirical P-values interpreted from the distribution of 10,000 permutation cycles.

Perhaps the most interesting findings came from interrogating the list of pathways down-regulated in the obeses co-twins (Table 10). The most significantly down-regulated pathway was involved in the catabolism of branched-chain amino acids (BCAAs), namely valine, leucine and isoleucine. Incidentally it is specifically these amino acids (especially leucine) that act as insulin secretagogues following a protein-rich meal[171] and are involved in the hypothalamic regulation of appetite through the mTOR signalling pathway (mammalian target of rapamycin)[172]. The role of BCAAs in obesity and the development of insulin resistance was first postulated in the 1960s[173] and their elevation in obese states has since been found in both rat-models of obesity[174,175] and in humans[176-179], but the related mechanism has not been settled. The possibility that in the obese, insulin resistant state the elevated BCAAs may serve as a feed-back signal to the pancreas to ensure an appropriately augmented secretory response was considered in the original report[173], but no mechanism was suggested. The observation here in study IV of a clearly reduced catabolism of these BCAAs may provide this mechanism (Figure 24).

Concordantly, along with the reduced catabolism of the BCAA's, the total plasma concentrations of BCAAs were elevated in the obese co-twins (P=0.025) and their levels correlated significantly with fS-insulin (r=0.50, P=0.028). Further, the activity (mean-centroid) of the BCAA catabolism pathway correlated negatively with all measures of body fat, i.e., with liver fat (r=-0.56, p=0.002), subcutaneous fat (r=-0.49, p=0.01) and intra-abdominal fat (r=-0.47, p=0.01). In a multiple regression

analysis including these fat depots, only liver fat remained significantly associated with the BCAA catabolism (R^2 0.29).

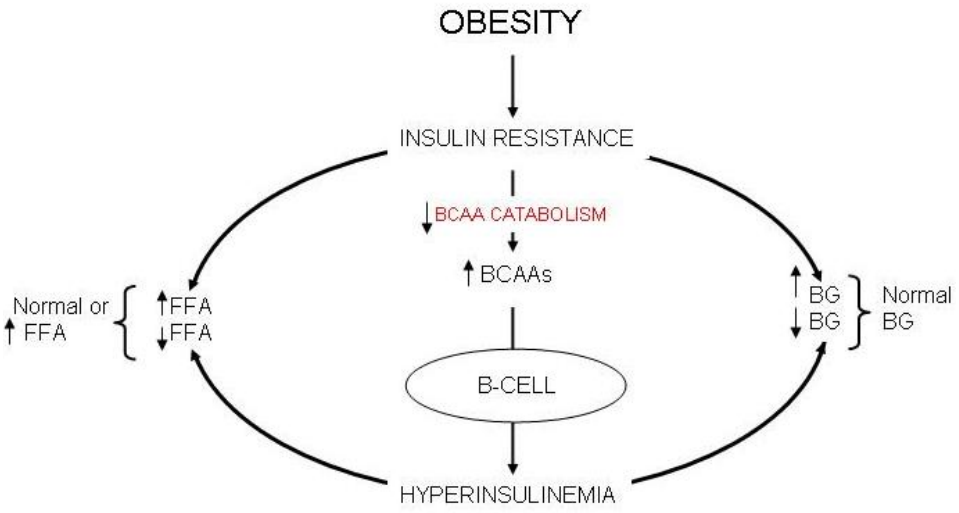


Figure 24. Postulated mechanism of the augmentation of insulin secretion in response to obesity associated insulin resistance. FFA: free fatty-acid, BG: blood glucose, BCAA: branched chain amino acid. Modified from[173].

The possible role of mitochondrial heteroplasmy behind the discordance in BMI was excluded through complete sequencing of mitochondrial DNA. No known obesity-predisposing mutations, or heteroplasmy could be detected between the co-twins. However, a considerable reduction in the mitochondrial DNA (mtDNA) copy number was observed in the obese co-twins –a novel finding in the study of human obesity. The mtDNA copy number in the obese co-twins was reduced to 53% of that in the non-obese co-twins (P=0.03). While reports exist of reduced mtDNA content in diabetic- and to a lesser extent in obese mice [180] and an impairment in the expression of mitochondrial oxidative phosphorylation genes in type 2 diabetic muscle[181] in humans, such a dramatic mtDNA loss in adipose tissue in human obesity is an entirely new observation. Interestingly the mitochondrial derangements including mtDNA loss

that was observed in the obese mice was only observed in adipose tissue, not in skeletal muscle or liver[180]. In study IV, no such differences in mtDNA content could be observed in leukocytes from the discordant twins, nor in the adipose tissue of 7 concordant MZ twins studied, strongly suggesting that the mtDNA depletion in these twins is weight-dependent. With the observed inflammatory state of the adipose tissue in the obese and the known ability of inflammation and oxidative stress to cause mitochondrial dysfunction, the mtDNA loss may well be yet another manifestation of the inflammation. The disturbed mitochondrial morphology and protein profile observed in a mouse model of obesity (*ob/ob*) could be reversed by treatment with a thiazolidinedione (rosiglitazone)[99] -the very same drug shown to downregulate *SPP1* expression[170] that was here connected with the development of inflammation in adipose tissue. This observation may provide the link between inflammation and mitochondrial mtDNA depletion in obesity.

A myriad of downstream effects of such a mitochondrial impairment could be hypothesized, among them a decreased maximal capacity for oxidative phosphorylation i.e. utilization of fat for energy production and a decreased catabolism of BCAAs –as this process takes place in mitochondria[182]. Decreased mitochondrial oxidative and phosphorylation activity may also result in the increase of intracellular fatty acid metabolites[183]. Such a dramatic decrease in the mtDNA content of the adipocytes of the obese, as here observed, surely has manifold sequelae that contribute to the multiple pathological features of the tissue, the precise consequences of which warrant further research.

As this fourth work of the thesis was purely cross-sectional in nature, direction of causation between the observed changes cannot be directly inferred. However, samples and measurements were obtained from one obese study participant that had gained a significant amount of weight upon a follow-up visit. These data, while in essence amount to a case-report, do show convincingly a further progression toward the pathological in the very same measures identified as differing between the obese

and lean twins. Despite the limited sample size, this, in addition to the trend for increased effects in twins with greater obesity discordance lends support to the notion that the changes are indeed sequelae, not proximal causes of obesity. A note worth considering is the fact that the pathological changes observed here were found in young and otherwise healthy, but obese individuals –a testament to the real health risks associated with obesity.

CONCLUSIONS OF THE STUDY

Shortly in reference to the aims listed for this thesis work on page 57, the following findings were presented:

I. While the linkage signal extends into the neighboring *F11R* gene, the known function of USF1 and the functional evidence for the specific associated *USF1* variant strongly indicate that *USF1* is the associated gene underlying the linkage with FCHL in chromosome 1q21.

II. The risk allele of the strongest associated SNP rs2073658 seems to eradicate the normal response of the *USF1* gene to the inducing effect of insulin in peripheral tissues. This change in the USF1 transcription factor results in the differential expression of its target genes in adipose tissue. It is through these downstream effects that the increased risk for dyslipidemia and the development of coronary heart disease is thought to arise.

III. Affected individuals from Finnish families with low-HDL carry rarer *ABCA1* variants more often then healthy control subjects, have higher expression of *ABCA1* in their macrophage foam-cells concurrently with only marginally diminished efflux capacity. Defects in the early steps of the reverse transport of cholesterol may predispose to the development of premature coronary heart disease in Finnish families with low-HDL.

IV. Numerous pathological changes can already be observed in the adipose tissue of young, obese but otherwise healthy individuals. Among the most prominent obesity associated changes were inflammation, a drastically diminished (63%) amount of mitochondrial DNA in adipocytes and a decreased catabolism of branched chain amino acids – a change that correlated strongly with liver fat and insulin resistance. The single most upregulated (5.9 –fold) gene in obesity was osteopontin (SPP1) – a cytokine involved in macrophage recruitment to adipose tissue and the development of insulin resistance.

CONCLUDING REMARKS

This thesis work took place during an exciting time in the field of human genetics. Between the beginning of the project in 2003 and the present time, we celebrated the 50th anniversary of the discovery of the structure of DNA, the essentially complete sequence of the human genome was revealed and the first genes behind complex disease were identified. The last few years have also seen the rise of genome-wide association studies that hold considerable promise for further discoveries and a deepened understanding of how genetic variation affects human health and disease on a global scale.

The *USF1* gene here investigated is one of the first examples of a gene identified underlying a complex human disease and it was discovered through the investigation of exceptional families segregating a multifaceted disease phenotype. The identification of *USF1* as the associated gene in familial combined hyperlipidemia (FCHL) provides an example of how utilizing well described, extended families from a population isolate such as Finland continues to be a relevant and powerful tool in disease gene identification in this new era of complex genetics. This work also exhibits many of the challenges associated with complex disease gene research, among them the difficulty in identifying the functional defect(s) that may manifest themselves as only moderate perturbations to "normal" physiology and only under certain conditions, as well as the intellectual frustration that can arise from the inability to conclusively exclude the possibility of additional functional variants. Complex diseases such as the common dyslipidemias here investigated represent a continuum in phenotype in which at the extreme end an individual can be considered affected. This is often however an arbitrary delineation and as in the case of FCHL where the lipid phenotype of a patient can switch classes, a definite demarcation of affected versus non-affected is not necessarily meaningful. This underscores the

reality of the complex disease geneticists' mantra that complex traits are a product of the interaction of a collection of genes with each other and the environment.

As noted, during this thesis project a number of new complex disease gene variants have been identified in genome-wide association studies, but to date no functionality has been assigned to the identified variants. The scientific community keenly awaits further confirmation of the findings and ultimately for an explanation at the molecular level of how the variants affect the disease process. This should be the ultimate goal of research of this sort if one wishes to proceed from mere reports of association to a full understanding of the genetics involved and the possibility of developing interventions and treatments to better human health.

The acquired component of complex disease is difficult to dissect in humans. To do this successfully requires an exceptional study sample and an unbiased global approach to analysis. The Finnish collection of monozygotic twins discordant for BMI was just such an exceptional sample that enabled the fine dissection of the molecular level events that take place in acquired obesity. In addition to confirming known events such as inflammation of adipose tissue, the study was able to shed light on completely novel aspects of the processes involved in the progression to insulin resistance and other pathological sequelae of obesity.

A common theme through this thesis work has been dyslipidemia, both in its familial and acquired forms. In addition, the USF1 transcription factor, somewhat serendipitously turns out to be a common thread in the four works comprising this thesis; *USF1* is the first gene associated with FCHL and as a transcription factor it is involved in the transcriptional regulation of *ABCA1* (here shown to be involved in the pathogenesis of low-HDL) as well as *SPP1* (here shown to be the most differentially expressed gene in obesity). In addition to highlighting the wide-ranging role of this transcription factor, this hints at the immense, interconnected network of genes and their interactions that is a key concept in beginning to understand the molecular background of complex disease.

ACKNOWLEDGEMENTS

This study was carried out in the Department of Molecular Medicine at the National Public Health Institute, Helsinki, Finland during the years 2003-2008. I wish to acknowledge the Director General Pekka Puska and the head of the Department of Molecular Medicine Adjunct Professor Anu Jalanko for providing exceptional research facilities and infrastructure.

Financial support was received from the Academy of Finland Centre of Excellence in Complex Disease Genetics, the Nordic Center of Excellence in Disease Genetics, National Institutes of Health grant 1RO1HL70150-01A1, the Finnish Foundation for Cardiovascular Research, the European Union Fifth Framework Program GenomEUtwin, the Biocentrum Helsinki Foundation, Helsinki University Central Hospital grants and the Finnish Medical Society Duodecim. All are gratefully acknowledged for their support. I also wish to thank all the participants of the studies here presented.

My deepest, heart-felt gratitude goes to my supervisor Academy Professor Leena Peltonen-Palotie. I never cease to admire the contagious enthusiasm she has for science and the fluency with which she is able to communicate complex concepts to both scientists and non-scientists alike. Leena's energy, dedication to her students and prodigious achievements are a source of inspiration for me. I want to thank her for challenging me as a scientist and for setting the bar so high as to nurture continued development. Working as a part of her team for these years has been a special privilege.

I wish to thank the members of my thesis committee, Professors Marja-Riitta Taskinen and Petri Kovanen from whom I have learned much and am also inspired by. The external reviewers of my thesis, Professors Matti J. Tikkanen and Olli Raitakari are thanked for their expertise, time and constructive comments to the manuscript. I am immensely grateful to Professor Mark McCarthy for accepting the role as opponent in my thesis. To have had such esteemed scientists review my work is truly an honor.

Numerous different scientists from literally around the world have contributed to the research presented here. Too numerous to mention them all, these talented people named in the accompanying publications are thanked for their invaluable contributions. I would like to acknowledge Assistant Professor Päivi Pajukanta for all her work on FCHL, for hosting me at UCLA and for the many valuable lessons. The secretaries Sari K., Tuija S., Sanna T., Mika K. and Sari M. are thanked for their help through everything. Jari R. is thanked for IT-support.

I have had the pleasure of learning from the numerous senior scientists in our laboratory: Kaisa Silander, Juha Saharinen, Massimiliano Gentile, Janna Saarela, Teppo Varilo, Marjo Kestilä, Iiris Hovatta, Ismo Ulmanen, Matti Jauhiainen, Vesa Olkkonen and Christian Ehnholm. Their insight into molecular genetic research and research-life in general has been invaluable. Juha and Massimiliano are especially

thanked for teaching me the complex world of global expression analyses, Matti and Vesa for their enthusiasm in lipid metabolism and Marjo for helping me navigate through the bureaucracy of getting this thesis published. I also want to thank my first mentors in science at the University of Washington: Professor Lawrence A. Loeb and Assistant Professor Manel Camps. I will always remember my time in the Loeb lab with warm thoughts. Along with Professors Mary-Claire King and Maynard Olson, you are very much the reason why I am in the field of human genetics.

I have had the pleasure of working with absolutely fantastic colleagues in the cardiovascular disease/dyslipidemia group that knows no real boundaries. Adjunct Professor Markus Perola is thanked for his guidance and good humor. Kati Kristiansson is thanked for open communication and fun company in the lab, as are Mervi and Kirsi. It has been a pleasure working with Krista for the past years and P-P is also thanked for joining the group. Good humor is the least expensive of all medicines.

The lab is filled with exceptional people that have made the work here so enjoyable. I want to thank all members of the lab (I'm running out of space!), especially Sampo, Johannes, Tero, Annina, Tintti, Outi, Juri, Pekka, Minttu, Minna, Olli, Nora, Marika, Emma N., Suvi, Annu, Markus, Annika, Jonna, Henna, Kaisu, Antti, Juho, Joe, Ida, Jaana, Ulrika, Tanja, Anu, Virpi, Heidi, Nabil, Karola, Riikka and past members Tony, Denis, Kismat, Aimee, Jenny, Suski, Emma P, Joni, Sarah, Tero and Aino. I know I have forgotten some names, but just know that you are all very much appreciated. A special thank you goes out to the superb laboratory technicians whose expertise and help has been invaluable: Anne and Lennu especially, working with you has been a delight.

"The better part of one's life consists of his friendships." I have been blessed in life with an abundance of good friends. I want to thank you all for your company, time, for lending me an ear and keeping me (relatively) sane through everything. Friends that are scattered around the world, the amazing people in my medical school class *Cursus Formosa et Fortia,* (esp. *Hypertrofia*, you know who you all are) and other people I have had the pleasure of meeting in school –too numerous to name, you all are beyond words.

Finally, I want to thank my parents Auvo and Anu, my brother Antti and sister Saara, as well as my grandparents for, well everything. For believing in me, encouraging me in everything I do and for your love and support I am forever grateful. I love you all.

Helsinki, February 14th, 2008

Jussi Naukkarinen

REFERENCES

1. Gartler, S.M. The chromosome number in humans: a brief history. *Nat Rev Genet* **7**, 655-60 (2006).
2. Riordan, J.R. et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066-73 (1989).
3. Kerem, B. et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073-80 (1989).
4. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lessons from studying monogenic disease for common disease. *Hum Mol Genet* **15 Spec No 1**, R67-74 (2006).
5. Glazier, A.M., Nadeau, J.H. & Aitman, T.J. Finding genes that underlie complex traits. *Science* **298**, 2345-9 (2002).
6. Buscher, R. & Grasemann, H. Disease modifying genes in cystic fibrosis: therapeutic option or one-way road? *Naunyn Schmiedebergs Arch Pharmacol* **374**, 65-77 (2006).
7. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
8. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
9. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
10. Lowe, C.B., Bejerano, G. & Haussler, D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A* **104**, 8005-10 (2007).
11. Birney, E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
12. The International HapMap Project. *Nature* **426**, 789-96 (2003).
13. Terwilliger, J.D. & Weiss, K.M. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* **9**, 578-94 (1998).
14. Weiss, K.M. & Terwilliger, J.D. How many diseases does it take to map a gene with SNPs? *Nat Genet* **26**, 151-7 (2000).
15. Cohen, J.C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72 (2004).
16. Cohen, J.C. et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A* **103**, 1810-5 (2006).
17. Romeo, S. et al. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* **39**, 513-6 (2007).

18. Ott, J. *Analysis of human genetic linkage*, xxii, 302 p. (Johns Hopkins University Press, Baltimore, 1991).
19. Schulze, T.G. & McMahon, F.J. Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. *Am J Med Genet* **114**, 1-11 (2002).
20. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nat Rev Genet* **1**, 182-90 (2000).
21. Peltonen, L., Jalanko, A. & Varilo, T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* **8**, 1913-23 (1999).
22. Salonen, R., Somer, M., Haltia, M., Lorentz, M. & Norio, R. Progressive encephalopathy with edema, hypsarrhythmia, and optic atrophy (PEHO syndrome). *Clin Genet* **39**, 287-93 (1991).
23. Maniatis, N. et al. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci U S A* **99**, 2228-33 (2002).
24. Service, S. et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* **38**, 556-60 (2006).
25. Russo, V.E.A., Martienssen, R.A. & Riggs, A.D. *Epigenetic mechanisms of gene regulation*, xii, 692 p. (Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1996).
26. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-8 (2007).
27. Lee, R.C., Feinbaum, R.L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54 (1993).
28. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. & Bartel, D.P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
29. Ross, J.S., Carlson, J.A. & Brock, G. miRNA: the new gene silencer. *Am J Clin Pathol* **128**, 830-6 (2007).
30. Mackay, T.F. Quantitative trait loci in Drosophila. *Nat Rev Genet* **2**, 11-20 (2001).
31. King, M.C. & Wilson, A.C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-16 (1975).
32. Korstanje, R. & Paigen, B. From QTL to gene: the harvest begins. *Nat Genet* **31**, 235-6 (2002).
33. Goring, H.H. et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* **39**, 1208-16 (2007).
34. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
35. McCarty, C.A. Refining the aetiology of myopia through negative results. *Br J Ophthalmol* **88**, 1 (2004).
36. O'Hara, B., Fowler, M.S. & Johnson, C.A. Why negatives should be viewed as positives. *Nature* **439**, 782 (2006).

37. Enattah, N.S. et al. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* **30**, 233-7 (2002).
38. Ho Sui, S.J. et al. oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* **33**, 3154-64 (2005).
39. Greene, D.A. Rosiglitazone: a new therapy for Type 2 diabetes. *Expert Opin Investig Drugs* **8**, 1709-1719 (1999).
40. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* **33**, 177-82 (2003).
41. Eaton, S.B., Eaton, S.B., 3rd, Sinclair, A.J., Cordain, L. & Mann, N.J. Dietary intake of long-chain polyunsaturated fatty acids during the paleolithic. *World Rev Nutr Diet* **83**, 12-23 (1998).
42. K©œbberling, J. & Tattersall, R. *The Genetics of diabetes mellitus*, xiii, 293 (Academic Press, London ; New York, 1982).
43. Eko©*, J.M., Zimmet, P. & Williams, D.R.R. *The epidemiology of diabetes mellitus : an international perspective*, xi, 437 (Wiley, New York, 2001).
44. Speakman, J.R. Thrifty genes for obesity and the metabolic syndrome--time to call off the search? *Diab Vasc Dis Res* **3**, 7-11 (2006).
45. Nikkila, E.A. & Aro, A. Family study of serum lipids and lipoproteins in coronary heart-disease. *Lancet* **1**, 954-9 (1973).
46. Goldstein, J.L., Schrott, H.G., Hazzard, W.R., Bierman, E.L. & Motulsky, A.G. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J Clin Invest* **52**, 1544-68 (1973).
47. Rose, H.G., Kranz, P., Weinstock, M., Juliano, J. & Haft, J.I. Inheritance of combined hyperlipoproteinemia: evidence for a new lipoprotein phenotype. *Am J Med* **54**, 148-60 (1973).
48. Hopkins, P.N. et al. Coronary artery disease risk in familial combined hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study. *Circulation* **108**, 519-23 (2003).
49. Pajukanta, P. *The search for familial combined hyperlipidemia susceptibility genes : by Pèaivi Pajukanta*, 59, [41] s. (National Public Health Institute, Helsinki, 1998).
50. de Graaf, J. & Stalenhoef, A.F. Defects of lipoprotein metabolism in familial combined hyperlipidaemia. *Curr Opin Lipidol* **9**, 189-96 (1998).
51. Babirak, S.P., Brown, B.G. & Brunzell, J.D. Familial combined hyperlipidemia and abnormal lipoprotein lipase. *Arterioscler Thromb* **12**, 1176-83 (1992).
52. Hunt, S.C. et al. Apolipoprotein, low density lipoprotein subfraction, and insulin associations with familial combined hyperlipidemia. Study of Utah

patients with familial dyslipidemic hypertension. *Arteriosclerosis* **9**, 335-44 (1989).
53. Aitman, T.J. et al. Defects of insulin action on fatty acid and carbohydrate metabolism in familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* **17**, 748-54 (1997).
54. Pihlajamaki, J., Karjalainen, L., Karhapaa, P., Vauhkonen, I. & Laakso, M. Impaired free fatty acid suppression during hyperinsulinemia is a characteristic finding in familial combined hyperlipidemia, but insulin resistance is observed only in hypertriglyceridemic patients. *Arterioscler Thromb Vasc Biol* **20**, 164-70 (2000).
55. Sniderman, A. et al. Association of coronary atherosclerosis with hyperapobetalipoproteinemia [increased protein but normal cholesterol levels in human plasma low density (beta) lipoproteins]. *Proc Natl Acad Sci U S A* **77**, 604-8 (1980).
56. Brunzell, J.D. et al. Plasma lipoproteins in familial combined hyperlipidemia and monogenic familial hypertriglyceridemia. *J Lipid Res* **24**, 147-55 (1983).
57. Austin, M.A., Brunzell, J.D., Fitch, W.L. & Krauss, R.M. Inheritance of low density lipoprotein subclass patterns in familial combined hyperlipidemia. *Arteriosclerosis* **10**, 520-30 (1990).
58. Hokanson, J.E., Austin, M.A., Zambon, A. & Brunzell, J.D. Plasma triglyceride and LDL heterogeneity in familial combined hyperlipidemia. *Arterioscler Thromb* **13**, 427-34 (1993).
59. Bredie, S.J., Tack, C.J., Smits, P. & Stalenhoef, A.F. Nonobese patients with familial combined hyperlipidemia are insulin resistant compared with their nonaffected relatives. *Arterioscler Thromb Vasc Biol* **17**, 1465-71 (1997).
60. Vakkilainen, J. et al. Glucose intolerance in familial combined hyperlipidaemia. EUFAM study group. *Eur J Clin Invest* **28**, 24-32 (1998).
61. Cabezas, M.C. et al. Impaired chylomicron remnant clearance in familial combined hyperlipidemia. *Arterioscler Thromb* **13**, 804-14 (1993).
62. Cullen, P., Farren, B., Scott, J. & Farrall, M. Complex segregation analysis provides evidence for a major gene acting on serum triglyceride levels in 55 British families with familial combined hyperlipidemia. *Arterioscler Thromb* **14**, 1233-49 (1994).
63. Bredie, S.J. et al. Segregation analysis of plasma apolipoprotein B levels in familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* **17**, 834-40 (1997).
64. Stein, O. & Stein, Y. Atheroprotective mechanisms of HDL. *Atherosclerosis* **144**, 285-301 (1999).
65. Genest, J.J., Jr. et al. Familial lipoprotein disorders in patients with premature coronary artery disease. *Circulation* **85**, 2025-33 (1992).
66. Barter, P.J. & Rye, K.A. High density lipoproteins and coronary heart disease. *Atherosclerosis* **121**, 1-12 (1996).

67. Pajukanta, P. Do DNA sequence variants in ABCA1 contribute to HDL cholesterol levels in the general population? *J Clin Invest* **114**, 1244-7 (2004).
68. Brooks-Wilson, A. et al. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* **22**, 336-45 (1999).
69. Bodzioch, M. et al. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat Genet* **22**, 347-51 (1999).
70. Rust, S. et al. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat Genet* **22**, 352-5 (1999).
71. Ordovas, J.M. et al. Apolipoprotein A-I gene polymorphism associated with premature coronary artery disease and familial hypoalphalipoproteinemia. *N Engl J Med* **314**, 671-7 (1986).
72. Kuivenhoven, J.A. et al. The molecular pathology of lecithin:cholesterol acyltransferase (LCAT) deficiency syndromes. *J Lipid Res* **38**, 191-205 (1997).
73. Heizmann, C. et al. DNA polymorphism haplotypes of the human lipoprotein lipase gene: possible association with high density lipoprotein levels. *Hum Genet* **86**, 578-84 (1991).
74. Wittrup, H.H. et al. A common substitution (Asn291Ser) in lipoprotein lipase is associated with increased risk of ischemic heart disease. *J Clin Invest* **99**, 1606-13 (1997).
75. Wang, Q.F. et al. Haplotypes in the APOA1-C3-A4-A5 gene cluster affect plasma lipids in both humans and baboons. *Hum Mol Genet* **13**, 1049-56 (2004).
76. Eichenbaum-Voline, S. et al. Linkage and association between distinct variants of the APOA1/C3/A4/A5 gene cluster and familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* **24**, 167-74 (2004).
77. Pennacchio, L.A. et al. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum Mol Genet* **11**, 3031-8 (2002).
78. Ribalta, J. et al. Newly identified apolipoprotein AV gene predisposes to high plasma triglycerides in familial combined hyperlipidemia. *Clin Chem* **48**, 1597-600 (2002).
79. Eckel, R.H., Grundy, S.M. & Zimmet, P.Z. The metabolic syndrome. *Lancet* **365**, 1415-28 (2005).
80. Frayn, K.N. Obesity and metabolic disease: is adipose tissue the culprit? *Proc Nutr Soc* **64**, 7-13 (2005).
81. Taskinen, M.R. LDL-cholesterol, HDL-cholesterol or triglycerides--which is the culprit? *Diabetes Res Clin Pract* **61 Suppl 1**, S19-26 (2003).
82. Horikawa, Y. et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **26**, 163-75 (2000).
83. Hani, E.H. et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia* **41**, 1511-5 (1998).

84. Grant, S.F. et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* **38**, 320-3 (2006).
85. Yi, F., Brubaker, P.L. & Jin, T. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. *J Biol Chem* **280**, 1457-64 (2005).
86. Huertas-Vazquez, A. et al. TCF7L2 is associated with high serum triacylglycerol and differentially expressed in adipose tissue in families with familial combined hyperlipidaemia. *Diabetologia* **51**, 62-9 (2008).
87. Janssen, I., Katzmarzyk, P.T. & Ross, R. Waist circumference and not body mass index explains obesity-related health risk. *Am J Clin Nutr* **79**, 379-84 (2004).
88. Zhang, Y. et al. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-32 (1994).
89. Karvonen, M.K. et al. Identification of new sequence variants in the leptin gene. *J Clin Endocrinol Metab* **83**, 3239-42 (1998).
90. Frayling, T.M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-94 (2007).
91. Hinney, A. et al. Genome Wide Association (GWA) Study for Early Onset Extreme Obesity Supports the Role of Fat Mass and Obesity Associated Gene (FTO) Variants. *PLoS ONE* **2**, e1361 (2007).
92. Peeters, A. et al. Variants in the FTO gene are associated with common obesity in the Belgian population. *Mol Genet Metab* (2007).
93. Wahlen, K., Sjolin, E. & Hoffstedt, J. The common rs9939609 gene variant of the fat mass and obesity associated gene (FTO) is related to fat cell lipolysis. *J Lipid Res* (2007).
94. Andreasen, C.H. et al. Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. *Diabetes* **57**, 95-101 (2008).
95. Scuteri, A. et al. Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet* **3**, e115 (2007).
96. Dina, C. et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* **39**, 724-6 (2007).
97. Arner, P. Hunting for human obesity genes? Look in the adipose tissue! *Int J Obes Relat Metab Disord* **24 Suppl 4**, S57-62 (2000).
98. Dahlman, I. & Arner, P. Obesity and polymorphisms in genes regulating human adipose tissue. *Int J Obes (Lond)* **31**, 1629-41 (2007).
99. Wilson-Fritch, L. et al. Mitochondrial remodeling in adipose tissue associated with obesity and treatment with rosiglitazone. *J Clin Invest* **114**, 1281-9 (2004).
100. Murray, C.J. & Lopez, A.D. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **349**, 1269-76 (1997).

101. Steinberg, D. Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy: part II: the early evidence linking hypercholesterolemia to coronary disease in humans. *J Lipid Res* **46**, 179-90 (2005).
102. Yla-Herttuala, S. Macrophages and oxidized low density lipoproteins in the pathogenesis of atherosclerosis. *Ann Med* **23**, 561-7 (1991).
103. Ross, R. Atherosclerosis--an inflammatory disease. *N Engl J Med* **340**, 115-26 (1999).
104. Pajukanta, P. et al. Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. *Nat Genet* **18**, 369-73 (1998).
105. Aouizerat, B.E. et al. A genome scan for familial combined hyperlipidemia reveals evidence of linkage with a locus on chromosome 11. *Am J Hum Genet* **65**, 397-412 (1999).
106. Naoumova, R.P. et al. Confirmed locus on chromosome 11p and candidate loci on 6q and 8p for the triglyceride and cholesterol traits of combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* **23**, 2070-7 (2003).
107. Hoffer, M.J. et al. The lipoprotein lipase (Asn291-->Ser) mutation is associated with elevated lipid levels in families with familial combined hyperlipidaemia. *Atherosclerosis* **119**, 159-67 (1996).
108. Aouizerat, B.E. et al. Linkage of a candidate gene locus to familial combined hyperlipidemia: lecithin:cholesterol acyltransferase on 16q. *Arterioscler Thromb Vasc Biol* **19**, 2730-6 (1999).
109. Eurlings, P.M., van der Kallen, C.J., Geurts, J.M., Flavell, D.M. & de Bruin, T.W. Identification of the PPARA locus on chromosome 22q13.3 as a modifier gene in familial combined hyperlipidemia. *Mol Genet Metab* **77**, 274-81 (2002).
110. Geurts, J.M. et al. Identification of TNFRSF1B as a novel modifier gene in familial combined hyperlipidemia. *Hum Mol Genet* **9**, 2067-74 (2000).
111. Gagnon, F. et al. Evidence of linkage of HDL level variation to APOC3 in two samples with different ascertainment. *Hum Genet* **113**, 522-33 (2003).
112. Evans, D., Buchwald, A. & Beil, F.U. The single nucleotide polymorphism -1131T>C in the apolipoprotein A5 (APOA5) gene is associated with elevated triglycerides in patients with hyperlipidemia. *J Mol Med* **81**, 645-54 (2003).
113. Talmud, P.J. et al. APOA5 gene variants, lipoprotein particle distribution, and progression of coronary heart disease: results from the LOCAT study. *J Lipid Res* **45**, 750-6 (2004).
114. Weissglas-Volkov, D. et al. Common hepatic nuclear factor-4alpha variants are associated with high serum lipid levels and the metabolic syndrome. *Diabetes* **55**, 1970-7 (2006).
115. van der Vleuten, G.M. et al. The Gln223Arg polymorphism in the leptin receptor is associated with familial combined hyperlipidemia. *Int J Obes (Lond)* **30**, 892-8 (2006).

116. Pajukanta, P. et al. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat Genet* **36**, 371-6 (2004).
117. Coon, H. et al. Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides. *Hum Genet* **117**, 444-51 (2005).
118. Huertas-Vazquez, A. et al. Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler Thromb Vasc Biol* **25**, 1985-91 (2005).
119. van der Vleuten, G.M. et al. The involvement of upstream stimulatory factor 1 in Dutch patients with familial combined hyperlipidemia. *J Lipid Res* **48**, 193-200 (2007).
120. Lee, J.C. et al. USF1 contributes to high serum lipid levels in Dutch FCHL families and U.S. whites with coronary artery disease. *Arterioscler Thromb Vasc Biol* **27**, 2222-7 (2007).
121. Naukkarinen, J., Ehnholm, C. & Peltonen, L. Genetics of familial combined hyperlipidemia. *Curr Opin Lipidol* **17**, 285-90 (2006).
122. Castellani, L.W. et al. Mapping a gene for combined hyperlipidaemia in a mutant mouse strain. *Nat Genet* **18**, 374-7 (1998).
123. Pei, W., Baron, H. & Muller-Myhsok, B. [Linkage of familial combined hyperlipidemia to chromosome 1q21-23 in Chinese and German families]. *Zhonghua Yi Xue Za Zhi* **80**, 25-7 (2000).
124. Coon, H. et al. Replication of linkage of familial combined hyperlipidemia to chromosome 1q with additional heterogeneous effect of apolipoprotein A-I/C-III/A-IV locus. The NHLBI Family Heart Study. *Arterioscler Thromb Vasc Biol* **20**, 2275-80 (2000).
125. Huertas-Vazquez, A. et al. Contribution of chromosome 1q21-q23 to familial combined hyperlipidemia in Mexican families. *Ann Hum Genet* **68**, 419-27 (2004).
126. Vionnet, N. et al. Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am J Hum Genet* **67**, 1470-80 (2000).
127. Wiltshire, S. et al. A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* **69**, 553-69 (2001).
128. Das, S.K., Hasstedt, S.J., Zhang, Z. & Elbein, S.C. Linkage and association mapping of a chromosome 1q21-q24 type 2 diabetes susceptibility locus in northern European Caucasians. *Diabetes* **53**, 492-9 (2004).
129. Ehm, M.G. et al. Genomewide search for type 2 diabetes susceptibility genes in four American populations. *Am J Hum Genet* **66**, 1871-81 (2000).

130. Hsueh, W.C. et al. Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* **52**, 550-7 (2003).
131. Ng, M.C. et al. Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* **53**, 1609-13 (2004).
132. Xiang, K. et al. Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. *Diabetes* **53**, 228-34 (2004).
133. Hanson, R.L. et al. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* **63**, 1130-8 (1998).
134. Zeggini, E. et al. Variation within the gene encoding the upstream stimulatory factor 1 does not influence susceptibility to type 2 diabetes in samples from populations with replicated evidence of linkage to chromosome 1q. *Diabetes* **55**, 2541-8 (2006).
135. Sirito, M., Lin, Q., Deng, J.M., Behringer, R.R. & Sawadogo, M. Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. *Proc Natl Acad Sci U S A* **95**, 3758-63 (1998).
136. Sirito, M. et al. Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. *Gene Expr* **2**, 231-40 (1992).
137. Griffin, M.J. & Sul, H.S. Insulin regulation of fatty acid synthase gene transcription: roles of USF and SREBP-1c. *IUBMB Life* **56**, 595-600 (2004).
138. Travers, M.T. et al. Promoter I of the ovine acetyl-CoA carboxylase-alpha gene: an E-box motif at -114 in the proximal promoter binds upstream stimulatory factor (USF)-1 and USF-2 and acts as an insulin-response sequence in differentiating adipocytes. *Biochem J* **359**, 273-84 (2001).
139. Nowak, M. et al. Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: role of upstream stimulatory factor. *Mol Cell Biol* **25**, 1537-48 (2005).
140. Gregor, P.D., Sawadogo, M. & Roeder, R.G. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev* **4**, 1730-40 (1990).
141. Casado, M., Vallet, V.S., Kahn, A. & Vaulont, S. Essential role in vivo of upstream stimulatory factors for a normal dietary response of the fatty acid synthase gene in the liver. *J Biol Chem* **274**, 2009-13 (1999).
142. Reiner, A.P. et al. USF1 gene variants, cardiovascular risk, and mortality in European Americans: analysis of two US cohort studies. *Arterioscler Thromb Vasc Biol* **27**, 2736-42 (2007).

143. Choquette, A.C. et al. Associations between USF1 gene variants and cardiovascular risk factors in the Quebec Family Study. *Clin Genet* **71**, 245-53 (2007).
144. Kantartzis, K. et al. Upstream transcription factor 1 gene polymorphisms are associated with high antilipolytic insulin sensitivity and show gene-gene interactions. *J Mol Med* **85**, 55-61 (2007).
145. Shibata, N. et al. Genetic association between USF 1 and USF 2 gene polymorphisms and Japanese Alzheimer's disease. *J Gerontol A Biol Sci Med Sci* **61**, 660-2 (2006).
146. Komulainen, K. et al. Risk alleles of USF1 gene predict cardiovascular disease of women in two prospective studies. *PLoS Genet* **2**, e69 (2006).
147. Gibson, F., Hercberg, S. & Froguel, P. Common polymorphisms in the USF1 gene are not associated with type 2 diabetes in French Caucasians. *Diabetes* **54**, 3040-2 (2005).
148. Hoffstedt, J., Ryden, M., Wahrenberg, H., van Harmelen, V. & Arner, P. Upstream transcription factor-1 gene polymorphism is associated with increased adipocyte lipolysis. *J Clin Endocrinol Metab* **90**, 5356-60 (2005).
149. Putt, W. et al. Variation in USF1 shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum Mol Genet* **13**, 1587-97 (2004).
150. Porkka, K.V. et al. Phenotype expression in familial combined hyperlipidemia. *Atherosclerosis* **133**, 245-53 (1997).
151. Vartiainen, E. et al. Twenty-year trends in coronary risk factors in north Karelia and in other areas of Finland. *Int J Epidemiol* **23**, 495-504 (1994).
152. Vartiainen, E. et al. Cardiovascular risk factor changes in Finland, 1972-1997. *Int J Epidemiol* **29**, 49-56 (2000).
153. Kaprio, J., Pulkkinen, L. & Rose, R.J. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res* **5**, 366-71 (2002).
154. Poulsen, P., Levin, K., Beck-Nielsen, H. & Vaag, A. Age-dependent impact of zygosity and birth weight on insulin secretion and insulin action in twins. *Diabetologia* **45**, 1649-57 (2002).
155. Poulsen, P. et al. Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins. *Diabetes* **54**, 275-83 (2005).
156. Shoulders, C.C. USF1 on trial. *Nat Genet* **36**, 322-3 (2004).
157. Pastinen, T. & Hudson, T.J. Cis-acting regulatory variation in the human genome. *Science* **306**, 647-50 (2004).
158. Ge, B. et al. Survey of allelic expression using EST mining. *Genome Res* **15**, 1584-91 (2005).
159. Babinska, A. et al. The F11 receptor (F11R/JAM-A) in atherothrombosis: overexpression of F11R in atherosclerotic plaques. *Thromb Haemost* **97**, 272-81 (2007).

160. Cavusoglu, E. et al. Association of plasma levels of F11 receptor/junctional adhesion molecule-A (F11R/JAM-A) with human atherosclerosis. *J Am Coll Cardiol* **50**, 1768-76 (2007).
161. Yang, X.P. et al. The E-box motif in the proximal ABCA1 promoter mediates transcriptional repression of the ABCA1 gene. *J Lipid Res* **43**, 297-306 (2002).
162. Frikke-Schmidt, R., Nordestgaard, B.G., Jensen, G.B. & Tybjaerg-Hansen, A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* **114**, 1343-53 (2004).
163. Soro-Paavonen, A. et al. Common ABCA1 variants, HDL levels, and cellular cholesterol efflux in subjects with familial low HDL. *J Lipid Res* **48**, 1409-16 (2007).
164. Klucken, J. et al. ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport. *Proc Natl Acad Sci U S A* **97**, 817-22 (2000).
165. Langmann, T. et al. Molecular cloning of the human ATP-binding cassette transporter 1 (hABC1): evidence for sterol-dependent regulation in macrophages. *Biochem Biophys Res Commun* **257**, 29-33 (1999).
166. Mazzali, M. et al. Osteopontin--a molecule for all seasons. *Qjm* **95**, 3-13 (2002).
167. Bidder, M. et al. Osteopontin transcription in aortic vascular smooth muscle cells is controlled by glucose-regulated upstream stimulatory factor and activator protein-1 activities. *J Biol Chem* **277**, 44485-96 (2002).
168. Nomiyama, T. et al. Osteopontin mediates obesity-induced adipose tissue macrophage infiltration and insulin resistance in mice. *J Clin Invest* **117**, 2877-88 (2007).
169. Lumeng, C.N., Deyoung, S.M. & Saltiel, A.R. Macrophages block insulin action in adipocytes by altering expression of signaling and glucose transport proteins. *Am J Physiol Endocrinol Metab* **292**, E166-74 (2007).
170. Oyama, Y., Akuzawa, N., Nagai, R. & Kurabayashi, M. PPARgamma ligand inhibits osteopontin gene expression through interference with binding of nuclear factors to A/T-rich sequence in THP-1 cells. *Circ Res* **90**, 348-55 (2002).
171. Hutson, S.M. The case for regulating indispensable amino acid metabolism: the branched-chain alpha-keto acid dehydrogenase kinase-knockout mouse. *Biochem J* **400**, e1-3 (2006).
172. Sipula, I.J., Brown, N.F. & Perdomo, G. Rapamycin-mediated inhibition of mammalian target of rapamycin in skeletal muscle cells reduces glucose utilization and increases fatty acid oxidation. *Metabolism* **55**, 1637-44 (2006).
173. Felig, P., Marliss, E. & Cahill, G.F., Jr. Plasma amino acid levels and insulin secretion in obesity. *N Engl J Med* **281**, 811-6 (1969).

174. Rafecas, I., Esteve, M., Remesar, X. & Alemany, M. Plasma amino acids of lean and obese Zucker rats subjected to a cafeteria diet after weaning. *Biochem Int* **25**, 797-806 (1991).
175. Wijekoon, E.P., Skinner, C., Brosnan, M.E. & Brosnan, J.T. Amino acid metabolism in the Zucker diabetic fatty rat: effects of insulin resistance and of type 2 diabetes. *Can J Physiol Pharmacol* **82**, 506-14 (2004).
176. Leclercq, B. & Seve, B. Influence of adiposity (genetic or hormonal) on the metabolism of amino acids and nutritional responses. *Reprod Nutr Dev* **34**, 569-82 (1994).
177. Marchesini, G., Bianchi, G., Rossi, B., Muggeo, M. & Bonora, E. Effects of hyperglycaemia and hyperinsulinaemia on plasma amino acid levels in obese subjects with normal glucose tolerance. *Int J Obes Relat Metab Disord* **24**, 552-8 (2000).
178. Schauder, P., Zavelberg, D., Langer, K. & Herbertz, L. Sex-specific differences in plasma branched-chain keto acid levels in obesity. *Am J Clin Nutr* **46**, 58-60 (1987).
179. Solini, A., Bonora, E., Bonadonna, R., Castellino, P. & DeFronzo, R.A. Protein metabolism in human obesity: relationship with glucose and lipid metabolism and with visceral adipose tissue. *J Clin Endocrinol Metab* **82**, 2552-8 (1997).
180. Choo, H.J. et al. Mitochondria are impaired in the adipocytes of type 2 diabetic mice. *Diabetologia* **49**, 784-91 (2006).
181. Mootha, V.K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
182. Hutson, S.M., Fenstermacher, D. & Mahar, C. Role of mitochondrial transamination in branched chain amino acid metabolism. *J Biol Chem* **263**, 3618-25 (1988).
183. Petersen, K.F. et al. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science* **300**, 1140-2 (2003).