From the Division of Gastroenterology and Hepatology,
Department of Medicine at Karolinska University Hospital Huddinge
Karolinska Institutet, Stockholm, Sweden

GENETIC AND ENVIRONMENTAL ASPECTS OF SYMPTOMATIC GALLSTONE DISEASE

Despina S. Katsika
M.D.

Stockholm 2009

ABSTRACT

Gallstone disease (GD) represents a major healthcare problem. Gallstones are likely to result from a complex interaction of the environment and the effects of multiple undetermined genes.

The aim of this study was: to evaluate the contribution of hereditary and environmental factors to the pathogenesis of gallstone disease by analyzing a large twin population; to examine the association between known environmental factors such as body mass index, alcohol and tobacco on symptomatic GD by analyzing parameter data from the Swedish Twin Registry; and to identify human candidate genes and polymorphisms for gallstone disease by selected sampling from the Swedish Twin Registry (STR).

In the first study we linked the STR with the Swedish inpatient-discharge and causes of death registries for symptomatic GD. Structural equation modelling techniques were used to estimate the contributions of genetic effects as well as shared and non-shared environmental factors to the development of symptomatic GD. In the second study we used the same screening procedure and evaluated those twins where weight, height, and data on use of alcohol and tobacco were provided by the STR and analyzed for possible associations by conditional logistic regression. In the third study we first identified the concordant monozygotic (MZ) and dizygotic (DZ) twins as well as discordant monozygotic (MZ) twins born between 1912 and 1958 alive in Stockholm County. We collected DNA, performed an abdominal ultrasound in case of undefined GD. For the *ABCG8* D19H polymorphism association analysis, we collected additional DNA from the nationwide TwinGene project identifying 20 MZ and 54 DZ cases as well as 109 MZ and 126 DZ controls.

A total of 43,141 twin pairs were screened in the first study. We found that concordances and correlations were higher in MZ compared with DZ twins, both for males and females. Genetic effects accounted for 25% (95% CI, 9%-40%), shared environmental effects for 13% (95% CI, 1%-25%), and unique environmental effects for 62% (95% CI, 56%- 68%) of the phenotypic variance among twins. In the second study we found that overweight and obesity were associated with significantly higher risk for GD in the whole study population (OR 1.86 and OR 3.38; CI: 1.52–2.28 and 2.28–5.02 respectively). High alcohol consumption was associated with a lower risk for GD in the whole population (OR 0.62; CI: 0.51–0.74) with no difference between discordant MZ and DZ twins (OR 1.08 and OR 0.96; CI: 0.82– 1.42 and 0.79–1.16). Smoking or smoke-free tobacco were not correlated with GD. Twenty-four (75%) out of 32 evaluable MZ twin pairs in Stockholm County were concordant for GD. Hetero- or homozygous 19H carriers were found in 5 concordant MZ twin pairs (20.8%), but only in 1 pair (12.5%) discordant for GD. Nationwide, we found 18.2% vs. 9.2% D19H carriers in MZ with and without GD, respectively, likewise 22.6% vs. 9.5% D19H in DZ. Overall D19H frequency was 20.8 % in cases compared to 9.4 % in controls. Association analysis showed that D19H allele significantly increased risk for GD (OR, 2.56; 95%CI, 1.28-5.15; p<0.01).

In conclusion, heritability is a major susceptibility factor for GD. There are positive associations between symptomatic GD and body mass index (BMI), and negative between GD and high alcohol consumption, whereas tobacco use has no impact. D91H was more common in cases than in controls and the association analysis found a significantly increased GD risk for twins carrying this *ABCG8* allele.

Key-words: Gallstone, Swedish Twin Registry, monozygotic, dizygotic, structural equation modeling, body mass index, alcohol, tobacco, *ABCG8* D19 H, *lith* genes, association study

LIST OF PUBLICATIONS

Genetic and environmental influences on symptomatic gallstone disease: a Swedish study of 43,141 twin pairs.
Katsika D, Grjibovski A, Einarsson C, Lammert F, Lichtenstein P, Marschall HU.
Hepatology 2005; 41(5): 1138-43
Copyright 2005 Hepatology; reprinted with permission.

Body mass index, alcohol, tobacco and symptomatic gallstone disease: a Swedish twin study.
Katsika D, Tuvblad C, Einarsson C, Lichtenstein P, Marschall HU.
J Intern Med. 2007; 262(5): 581-7
Copyright 2007 Journal of Internal Medicine; reprinted with permission.

Gallstone disease in Swedish twins is associated to ABCG8 D19H risk genotype.
Katsika D, Magnusson P, Krawczyk M, Grünhage F, Lichtenstein P, Einarsson C, Lammert F, Marschall HU. Submitted for publication

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Additive genetic effects
ABC	ATP-Binding Cassette
ABCB11	BSEP, Bile Salt Export Pump
ABCB4	MDR3, Multi Drug Resistant p-Glycoprotein 3
ABCG5/8	Hepatic cholesterol hemitransporters
AIC	Akaike Information Criterion
APOA/APOB	Apolipoprotein A/B
BMI	Body Mass Index
C	Common environmental effects
CI	Confidence Interval
CYP7A1	Cholesterol 7-Alpha-Hydroxylase
DNA	Deoxyribonucleic Acid
DZ	Dizygotic
E	Unique environmental effects
FXR/BAR	Farnesoid X Receptor/Bile Acid Receptor
GD	Gallstone Disease
GEE	Generalized Estimation Equation models
HDL	High Density Lipoprotein
HMG CoA	3-Hydroxy-3-metylglutaryl Coenzyme A
ICD	International Classification of Diseases
LDL	Low Density Lipoprotein
LXR	Liver X Receptor
MZ	Monozygotic
OR	Odds Ratio
OS	Opposite-Sexed
Ow/Ob	Overweight/Obesity
QTL	Quantitative Trait Locus
RR	Relative Risk
SEM	Structural Equation Modelling
SNP	Single-Nucleotide Polymorphism
STR	Swedish Twin Registry
UDCA	Ursodeoxycholic Acid
WHO	World Health Organization

1 INTRODUCTION

1.1 CHOLESTEROL GALLSTONE DISEASE

Cholesterol gallstone disease (GD) is one of the most common and health economically important gastrointestinal diseases. The disease represents a failure of biliary cholesterol homeostasis in which the physical-chemical balance of cholesterol solubility in bile is disturbed. The primary pathophysiologic defect is cholesterol supersaturation of gallbladder bile. The underlying defects are augmented intestinal cholesterol absorption, cholesterol synthesis, lipoprotein delivery, and hepatic cholesterol up-take, and disorders that uncouple phospholipid and/or cholesterol secretion. The molecular pathogenesis as well as the genetic susceptibility for gallstones is still obscured. Age, gender, race, obesity, diabetes, and parity have all been identified as significant risk factors for the development of gallstones (1-3). Gallstones are likely to result from a complex interaction of the environment and the effects of multiple undetermined genes. Recent progress in molecular biology and genetics indicated that the susceptibility for gallstones is based on improper function of proteins that regulate lipid synthesis and translocation. Studies in inbred mice revealed a number of candidate genes for gallstone disease (e.g., *Abcb4, Abcb11, Abcg5/8, Lrp2, Apoe, Cyp7a*), all encoding for proteins that are responsible for the synthesis and regulation of compounds involved in the metabolism and cellular transport of cholesterol and other biliary compounds (2-4).

1.1.1 Gallstone disease epidemiology and risk factors

Ultrasound studies indicate mean prevalence rates of 10–15% in adult Europeans, and of 3–5% in African and Asian populations (1, 5). In the US, the prevalence rates range from 5% for nonHispanic black men to 27% for Mexican-American women (6, 7), whereas in American Indians, gallstone disease is epidemic and found in 73% of adult female Pima Indians (8), and in 30% of male and 64% of female in other American Indians (7). Previous studies in Sweden have shown the overall prevalence to be approximately 15% (9). About 80% of all GD patients are asymptomatic but approximately 1-2% of patients per year will develop complications that will necessitate surgery (1, 10).

Female gender, fecundity, and a family history are strongly associated with the development of GD. Obesity and dyslipidemia, especially the combination of high triglycerides along with low HDL, hyperinsulinaemia and insuline-resistence, all part of the metabolic syndrome, are also associated with the formation of cholesterol gallstones as well as type 2 diabetes. Other factors known to increase the risk of gallstone formation include medications such as the use of oestrogen-replacement therapy or somatostatine analogues as well as conditions that promote gallbladder hypomotility such as prolonged fasting periods or total parenteral nutrition, and cervical spine injuries. Nutritional factors may promote or reduce the risk of gallstone formation. High caloric ("westernized") diets, especially high cholesterol and carbohydrate intake, increase the risk for gallstones whereas legumes, coffee and alcohol seem to have a protective effect as well as physical activity independent of weight-loss. A rapid weight loss may however increase the risk of gallstone formation (1, 2).

1.1.2　　　　Bile formation and cholesterol gallstone disease (Figure 1)

Intestinal cholesterol is transferred by the ABC-transporter ABCA1 to apoA1 particles that are taken up by the liver by high-density lipoprotein (HDL) receptor SRB1. Minor amounts of cholesterol derive from low-density lipoprotein LDL and chylomicron remnants and are taken up by LDL receptor LDLR and LDL-receptor-related protein LRP. Bile salts are mainly taken up by the liver via the sodium-dependent taurocholate transporter (NTCP) SLC10A1, and by organic anion transport proteins (OATPs) SLC21.

Hepatic de novo synthesis of cholesterol is under the control of hydroxymethyl-glutaryl- (HMG-) CoA-reductase. Part of cholesterol is esterified by acyl CoA:cholesterol acyltransferase (ACAT) and secreted as very low-density lipoprotein (VLDL) cholesterol or stored in the liver as cholesterol esters. Cholesterol may be metabolized into bile acids in the classical, neutral pathway via 7α- and 12α-hydroxylase (CYP7A1 and CYP8B1) reactions or in smaller amounts via the alternative, acidic pathway via an initial 27-hydroxylase (CYP27A1) reaction. The key regulatory enzyme in bile acid synthesis is CYP7A1.
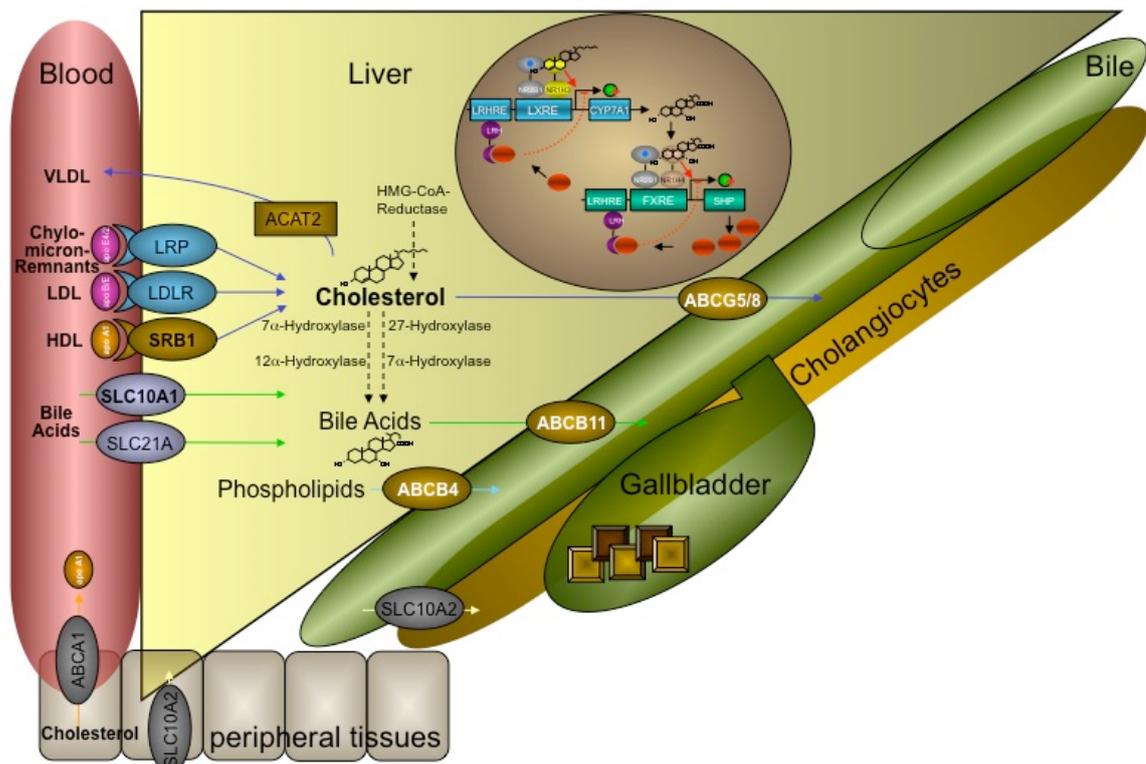
Figure 1: Uptake and excretion of biliary lipids, and major steps of cholesterol and bile acid synthesis, including nuclear receptor regulation. Adapted from (1).

Bile formation is essential for the removal of excess dietary cholesterol, either by direct excretion into bile or by conversion to bile salts. Bile is mainly an aqueous solution (90%) that contains three lipid species: cholesterol (4%), phospholipids (24%) and bile salts (72%) (1, 11). Hepatocytes express specific ATP-binding-cassette transport proteins — known as ABC transporters — for each of these three lipids at the canalicular membrane domain. The ABCB11 (BSEP) transporter is the bile salt export pump, ABCB4 (MDR3) is the transporter for the major biliary phospholipid phosphatidyl choline, and ABCG5 and ABCG8 form obligate heterodimers for biliary cholesterol

secretion (12). Gene expression levels of *ABCB4*, *ABCB11* and *ABCG5/8* are regulated by at least two nuclear receptors (NR) initially found to cholesterol and bile acid metabolism (13, 14). The farnesoid X or bile acid receptor FXR/BAR (official gene symbol *Nr1h4*) (15-18) regulates transcription of *ABCB4* (19) and *ABCB11* (20) while *ABCG5* and *ABCG8* are under control of the oxysterol or liver X receptors LXRα/β (21), perhaps mediated by FXR (22). Once secreted, hepatic bile is modified by bicarbonate- and chloride-rich secretions of cholangiocytes, accompanied with water influx through aquaporin channels (23). The chloride channel in cholangiocytes is the cystic fibrosis transmembrane conductance regulator (ABCC7, CFTR) that is mutated in cystic fibrosis (24, 25). The apical/ileal sodium-dependent bile salt transporter (ASBT/ISBT) SLC10A2 is expressed both in cholangiocytes and the intestine (Fig. 1.)

1.1.3   Pathophysiology

More than 80% of gallstones consist mainly of cholesterol and are formed within the gallbladder. Cholesterol crystals are embedded in a matrix of bile pigments, calcium salts and glycoproteins (26). Gallstones can be pure or mixed cholesterol gallstones as well as pure pigment stones. The latter can be brown or black pigment stones. Brown pigment stones are associated with infections of the biliary tract (bacterial and helminthic deconjugation of bilirubin glucuronides) and are more frequent in Asia. Black pigment stones mainly consist of calcium bilirubinate and are found in haemolytic anaemia or ineffective haematopoiesis and in patients with cystic fibrosis   Conditions associated with increased enterohepatic cycling of bilirubin such as terminal ileitis in Crohn's disease are also associated with black pigment stones although bile salt malabsorption in these patients may rather promote cholesterol gallstone formation (2, 3, 27-29).

The three key mechanisms that contribute to the formation of cholesterol gallbladder stones are cholesterol hypersaturation, of bile, gallbladder hypomotility and destabilization of bile by kinetic protein factors (1, 2).

1.1.4   Cholesterol hypersaturation

Cholesterol-supersaturated bile contains more cholesterol that can be solubilized by mixed micelles and multilamellar vesicles that fuse and aggregate to form solid cholesterol crystals. In principle, cholesterol hypersaturation of gallbladder bile may be the result of the hepatic hypersecretion of cholesterol or the hyposecretion of bile salts or lecithin (2).

The main cause of cholesterol hypersaturation is cholesterol hypersecretion that increases with age and may be caused by increased hepatic uptake or synthesis of cholesterol, decreased hepatic synthesis of bile salts, or decreased hepatic synthesis of cholesteryl esters. The major part of cholesterol is of dietary origin (80%), and *de novo* synthesis of cholesterol is only about 10%. Cholesterol hypersecretion is found only in patients with GD and not in healthy individuals although no single metabolic defect that can cause this hypersecretion has been identified in gallstone patients. Increased cholesterol saturation has also been associated with the excess in the bile acid pool of the secondary bile acid deoxycholic acid, formed from the primary cholic acid by 7-dehydroxylation, and presumably enriched in patients with prolonged intestinal transit (1, 2, 30).

3

1.1.5 Gallbladder motility

Whether gallbladder hypomotility is a primary defect or associated to cholesterol hypersaturation is debatable. However, impaired gallbladder motility is observed in conditions associated with GD such as diabetes, total parenteral nutrition and rapid weight loss. Cholecystokinin injections for patients receiving long-term total parenteral nutrition or small fatty meals during weight loss have been suggested in order to counteract gallbladder hypomotility in these patients (31-34).

1.1.6 Other factors

The nucleation of cholesterol microcrystals in bile is modulated by kinetic protein factors. Of a number of inhibitory or promoting proteins only mucine, a glycoprotein mixture secreted by biliary epithelial cells, has been consistently defined as a crystallization promoting protein in gallbladder sludge. It is suggested that the degradation of mucine by lysosomal enzymes is the major prokinetic mechanism rather than the correlation between cholesterol hypersaturation and the amount of mucine (35-41).

Intestinal *Helicobacter* species have also been suggested as a gallstone formation promoting factor in mice and they have been indeed identified in patients with GD but the prevalence of *Helicobacter* DNA in humans does not differ in gallstone patients and controls (42-50).

1.2 GENETIC FACTORS

Gallstone disease is likely to be the effect of the complex interaction of environmental factors and the effects of multiple, undetermined genes. Genetic factors that affect the susceptibility to gallstone formation are suggested by family studies that identified the prevalence of GD in first-degree relatives of patients with cholelithiasis as being two to four times higher than in stone-free controls. The high prevalence of GD among American Indians and Hispanics is also indicative of genetic factors (2).

Familial clustering of gallstones though, does not necessarily confirm the importance of genetic factors. Conclusive evidence was first provided by van der Linden in a Swedish study, where women married to and living together with gallstone patients did not have a significant increase in GD as compared to women married to the stone-free brother, suggesting that shared environment alone does not promote gallstone formation and genetic factors are involved (51).

In 1999, Duggirala et al. (52) used pedigree data to explore the genetic susceptibility to symptomatic GD in a Mexican-American population of 32 families and estimated a heritability (i.e., the proportion of the phenotypic variance of the trait that is due to genetic effects) of 44%. However, this association study did not control for shared environmental effects. A more recent family study from the United States performed a variance component analysis in 1,038 individuals taken from 358 families and calculated the heritability of symptomatic GD to be 29% (53).

As determined by cross-sectional ultrasound surveys, the prevalence of GD rates shows remarkable geographic variations. GD disease is common in most European countries as well as in North and South America. The prevalence still is low in Asia and Africa but increasing after the acquaintance to "westernized" diet. However, the role of high dietary cholesterol is unclear. This "lithogenic" diet seems to increase the risk of GD only in gallstone carriers but not in stone-free individuals, suggesting that intestinal cholesterol absorption and biliary secretion must be genetically determined (1, 2, 30).

1.2.1   Monogenic cholelithiasis

Only in specific groups of patients has monogenic predisposition for cholelithiasis been described. These include mutations in the genes that encode the ABC transporters for phosphatidylcholine (ABCB4) or bile salts (ABCB11), the rate-limiting enzyme for bile acid synthesis (CYP7A1), and the cholecystokinine type A receptor (CCK1R) (2, 54-59).

Rosmorduc et al. (56) provided the first evidence that a single gene defect causes the formation of gallstones. They identified point mutations in *ABCB4* in patients with "low phospholipid-associated cholelithiasis". This syndrome is characterized by cholesterol cholelithiasis before the age of forty, intrahepatic sludge and microlithiasis as well as recurrent biliary symptoms after cholecystectomy. Interestingly, variants of ABCB4 are also associated with intrahepatic cholestasis of pregnancy (ICP) that is also highly associated with GD (60). Symptomatic gallstone carriers with ABCB4 mutations as well as women with ICP seem to benefit from treatment with ursodeoxycholic acid (UDCA).

Mutations in *ABCG11* have also been associated with GD in a subgroup of patients with benign intrahepatic cholestasis (59) (BRIC-2, a minor form of progressive familial cholestasis type 2, PFIC-2), the majority of which will develop gallstones. In contrast, an association analysis in a German sample showed no evidence of association between the *lith* genes *ABCB11* and *LXRA* to gallstone susceptibility. The gallstone trait thus is not allelic to at the *ABCB11* locus for PFIC-3 (2).

Another association between a single-gene defect and gallstone formation has been suggested by Pullinger et al.(61) and confirmed later by a Chinese association study regarding a common single nucleotide polymorphism (SNP) within the CYP7A1 promoter, which is associated with increased LDL cholesterol levels and gallstone formation (62).

1.2.2   Genome analysis in inbred mice

During the last years, candidate genes for cholesterol gallstone disease have been identified in studies in inbred mouse strains that differ in the susceptibility for cholesterol gallstone formation when fed a lithogenic diet. Using the quantitative trait loci analysis (63, 64), a murine gallstone map was developed describing the chromosomal organization of candidate gene loci (65). Twenty-three candidate *lith* genes have been identified that are closely related to the regulation of synthesis, uptake and excretion of hepatobiliary lipids and proteins, e.g. genes that encode for sterol carrier protein, ABC transporters, nuclear receptors such as FXR, and mucine (12, 66-73). Likely candidate genes are *lith 1* (*Abcb11*), *lith 2* (*Abcc2*), *lith 7* (*Nr1h4*), *lith 9* (*Abcg5* and *Abcg8*), and *lith 13* (*Cckar*) (72). In addition, genes that encode for immune-related factors, e.g. Il4, have been postulated as *lith* genes (72-74).

1.2.3   Human candidate genes

Despite the large number of candidate genes identified in mice, human candidate gene studies have been sparse. Polymorphisms of genes encoding cholesterol transporting and metabolizing proteins apolipoprotein B and E (*APOB*, *APOE*), cholesteryl ester transport protein (*CETP*), and *CYP7A1* were found to be associated with cholesterol gallstone disease (2).

Recently, the first genome wide association study in a large cohort of gallstone patients from Germany by Buch et al. (75) and a linkage study in affected sib pairs by Grünhage et al. (76) identified a common variant (p. D19H) of the hepato-canalicular cholesterol transporter ABCG5/ABCG8 as genetic risk factor for gallstones. The p. D19H confers an increased risk of 2-3 and 7 for the heterozygous and homozygous carriers, respectively, and 8-11% of the total gallstone risk can be attributed to this variant (30, 75, 76). Carriers of this variant display lower levels of serum plant sterols and higher levels of cholesterol precursors, indicating decreased cholesterol absorption and thus increased cholesterol biosynthesis. The clinical relevance of this finding might be potential antilithogenic effects of HMG CoA reductase inhibitors for these carriers (30). It is obvious that more studies are needed to confirm these findings in larger cohorts and other ethnic groups.

1.3   DIAGNOSIS, CLINICAL COURSE, THERAPY AND PREVENTION

Although gallstone symptoms are not specific, postprandial abdominal pain with onset over an hour after food intake and radiation to the right upper back seems to be the most supportive of the diagnosis (77).

Diagnosis is often confirmed by abdominal ultrasonography which is accurate in >90% of the cases regarding gallbladder stones and intrahepatic stones (78) but may miss up to 50% of common bile duct stones (79). Magnetic Resonance Imaging or Cholangiography (MRC) as well as Endoscopic Ultrasound (EUS) and Endoscopic Retrograde Cholangio-Pancreatiography (ERCP) may display a higher sensitivity for duct stones. The latter bears a substantial procedural risk but offers therapeutic options such as sphincterotomy, stone extraction and biliary drainage (1). Today, the treatment of choice for symptomatic GD is laparoscopic cholecystectomy, which is associated with a shorter hospital stay, lower costs and the same complication frequency as open cholecystectomy. Mortality rates following cholecystectomy vary from 0.1 to 0.8%. Non-surgical approaches such as gallstone dissolution by chenodeoxycholic acid or UDCA and extracorporeal shock-wave lithotripsy (ESWL) have lost their impact during the years and are currently used only on a small number of selected symptomatic patients that usually do not qualify for surgery (1, 30, 80).

The natural history of GD is not well defined. Although the majority of patients with a single episode of biliary colic will develop repeated symptoms, only 1-3% of these symptomatic patients will develop complications within a year (30, 81-83). Consequently, approximately 50% of patients with biliary duct stones will develop complications, but 20% of these stones will pass spontaneously (30). Mild and moderate acute cholecystitis is preferably treated by early laparoscopic cholecystectomy though

patients may benefit by the use of intravenous antibiotics (30, 84). Asymptomatic cholecystolithiasis is generally not an indication for cholecystectomy (85).

Future perspectives include the regulation of cholesterol metabolism and secretion through stimulation (or even inhibition) of nuclear receptors as was shown for the prevention of GD by synthetic FXR agonists in mice (4, 86). In the future, individual risk profile assessment may allow distinguishing between different risk patients, by genetic counseling, appropriate medication and life-style changes (1).

1.4 TWIN STUDIES

1.4.1 Concordance and heritability

Most common and chronic human diseases belong to the group called "complex genetic diseases", which means that these diseases depend on complex interactions between several genetic variants and several environmental events. Twin studies have been a valuable source of information about the genetic basis of complex traits. They can be used to study the interaction of genotype with sex, age and lifestyle factors (87-93).

By facilitating comparison between MZ and DZ twins, twin registers represent some of the best resources for evaluating the importance of genetic variation in susceptibility to disease. Recent advantages on statistical modelling allow simultaneous analysis of many variables in relatives such as MZ and DZ twins. These advantages make it possible to carry out multivariate analyses by inclusion of two ore more dependent variables in one analysis, for example in estimating the genetic correlation of birth weight and blood pressure. It also allows the estimation of heritability that is the proportion of the total phenotypic variance in a given disease that can be attributed to genetic effects (87-89, 91).

The classical twin study compares the resemblance of MZ twins for a disease with the resemblance of DZ twins for the same disease (88). Because MZ twins share all their genes, whereas DZ twins share 50% of their segregated genes, this comparison allows the estimation of genes and environment (89). The concordance rate is defined as the occurrence of the same disease in both members of a pair of twins. Any heritable disease will be more concordant in MZ twins than in DZ twins. If MZ twins resemble each other more than DZ twins then the heritability of the phenotype can be estimated as twice the difference between MZ and DZ correlations. The proportion of the variance that is due to shared environment is the difference between the total twin correlation (r) and the part that is explained by heritability, that is:

$$r \, (MZ) \; - h^2 \qquad \text{in MZ twins and}$$
$$r \, (DZ) \; - h^2/2 \qquad \text{in DZ twins (88)}.$$

The advantages of using twins are many. The fact that a trait "runs in the family" is not sufficient evidence to assume that its aetiology is genetic. Families may share predisposing environment as well as genes. From a strictly genetic point of view there is no advantage of using DZ twins over sibling pairs. However, since most diseases vary with age different genes may influence a disease at different ages. DZ twins can remove

this confounder and they match MZ twins as closely as possible in gestation and rearing. In addition to this, twins are more cooperative, there are now large twin registries with data on a wide range of biometrical variables, and DZ twins are more likely to have the same father rather than other siblings (88).

It has been claimed that MZ twins share more similar post-natal environments than DZ twins and thus their similarity cannot be attributed to their genetic identity. However, most studies point out that the most similar treatment of MZ twins is rather the result of their genetic identity and the similar responses this elicits by the environment (89).

1.4.2 Structural Equation Modelling

Quantitative genetic methods can be used to investigate the relative importance of genetic and environmental influences on a phenotype. These methods are well developed for twin studies. Quantitative genetic methods include comparisons of concordances, as defined earlier, of intra class correlations and structural equation modelling. When MZ concordances are higher than DZ concordances, genetic influences are indicated. Intra class correlation is a statistical measure for the strength and the direction of resemblance between two variables or two family members (89, 94).

Structural equation modelling (SEM), also known as covariance modelling, is a method in which genotypic and environmental effects are modelled as the contribution of unmeasured (latent) variables to the potentially multivariate phenotypic differences between individuals (88). It estimates regression coefficients (parameters) between latent (unobserved) and observed variables. These estimates minimize the difference between the covariance structure of the observed data and that predicted by the model. Alternative models such as family resemblance due to shared genes versus shared environment can be compared by how well they fit the data (88). Information about shared genetic and shared environmental influences can be used to set up linear structural equations and fit models over all types of twins in order to best describe the causes of variation of the phenotype.

The total variance of a trait V(p) can be partitioned into genetic variance (A), shared environmental variance (C) and unique environmental variance (C). Thus, for twin 1 in a pair the equation can be written as (94):

$$V (p)1 = a \cdot A1 + c \cdot C1 + e \cdot E1$$

A similar equation can be written for the second twin in the pair.

The total covariance V(p) can be expressed by the equation:

$$V(p) = a^2 + c^2 + e^2$$

The covariances (Cov) for the MZ and DZ twins can be described as (94):

$$Cov (MZ) = a^2 + c^2$$
$$Cov (DZ) = 0.5 \cdot a^2 + c^2$$

The parameters a, c and e can be estimated by maximum likelihood methods. For this, several software programs such as LISREL http://sscicentral.com/lisrel/mainlis.htm or *Mx* http://www.vcu.edu/mx can be used.

SEM can accommodate the analysis of gender differences through the simultaneous analysis of data from female and male MZ and DZ twins. By analyzing data from DZ twins of opposite sexes (OS) it is possible to test whether the same genes are expressed in males and females. If the resemblance between twins of the opposite sexes is less than what is expected on the basis of heritability in males and females then this would indicate that different genes are operating in the two sexes (88, 95).

1.4.3 Co-twin analysis

The co-twin control method takes advantage of the fact that MZ and DZ twins share different degrees of genetic relatedness. These methods are used when the relationship between a putative risk factor and a disease is studied with control for genetic background and unmeasured early environmental factors shared by the twins. Both disease discordant and exposure discordant twins can be used (94).

In the co-twin control analysis with disease discordant twins, two control groups are used: External (non-related) controls and internal (co-twin) controls. First, the association between exposure and outcome is studied, where we compare twins diagnosed as cases with external controls (other twins not related to the cases). This evaluates the risk of a disease given an exposure, i.e., a risk factor, and is a classic case-control study. In the second step, the healthy co-twin in a pair (both MZ and DZ) is used as a control for the diseased twin. Because twins share the same intrauterine environment and are typically reared together, this controls for confounding from unmeasured early environmental factors in childhood or adolescent environment. If the associations between exposure and disease observed in the analysis with external controls remain the same in the co-twin analysis, it speaks in favour of a causal effect of the exposure on the disease. If on the contrary, the association is not as high in the co-twin analysis as it is in the analysis with external controls, this indicates that early environmental factors are responsible for this association between exposure and disease. In order to control for genetic background the healthy MZ twin is used as a control. If the disease is more often in MZ pairs with the exposure then this should indicate that the exposure contributes to the disease since MZ twins share the same genes. On the other hand, if an association exists in external controls among disease discordant DZ pairs but not between MZ pairs, then genetic effects have probably confounded the results (88, 94).

1.4.4 Genetic analysis, linkage and association studies

Human genetic disease can be classified into simple, Mendelian diseases in which each gene consists of a paternally and a maternally derived copy, which remains intact and distinct in the resulting gametes and where genes for different traits are independently inherited. Mendelian diseases though are typically rare, such as cystic fibrosis, Huntington's disease, Marfan syndrome and Duchenne's muscular dystrophy. The majority of genetic diseases are of a complex genetic architecture and are thus multifactorial, caused by one ore more genes in combination with environmental factors

that include gene-environment interactions (90, 96-101). The same categorisation exists for genes, the most common being susceptibility genes. Different states of a gene (alleles or polymorphisms) may confer different states of disease.

Susceptibility genes can be classified generally by two methods: Candidate-gene studies and genome screens. Candidate-genes studies examine the association between sequence variation in the candidate gene and the phenotype of interest. Variations in DNA are referred to as polymorphisms and alleles. These are examined for their relation with the phenotype, but first, the variation in the DNA sequence must be identified. The candidate-gene approach is successful when identifiable genes are known to be associated with variation in biologically relevant processes (96).

An alternative approach is the genome-scan approach in which locations, or markers, representing the entire genome are tested for associations between the genetic location and phenotype. Markers describe a DNA sequence whose location in the genome is known. Rapid and complete genome scans have been made possible by advantages in genetic technology. Whole-genome screens detect Quantitative Trait Loci (QTL) for disease. These are genetic loci or chromosomal regions that contribute to the variability of a complex quantitative trait. Quantitative traits are typically affected by several genes and the environment (88) and before one starts hunting for QTLs of a complex trait it is necessary to show that the trait is genetically influenced (90, 97, 98, 102).

SNPs are single base-pair changes in DNA that may represent functional differences in genes. The distance between the marker and the disease locus will disappear as SNPs within genes are identified and added to the marker map (92, 96).

Two complementary methods to candidate-gene analysis and genome screen analysis are association and linkage analyses. Association analysis identifies particular polymorphisms or alleles that increase the risk of disease. It examines the correlation of specific alleles at 2 loci that is a marker locus or candidate gene and disease phenotype. The shared genetic background will result in an association between near-by markers and the disease-associated polymorphism and ultimately in an association with the disease (96, 103).

Linkage analysis estimates the recombination fraction between 2 loci, that is a marker locus and a disease gene. The further apart these 2 loci are, the more likely they are to recombine during meiosis. If a marker and a gene are sufficiently close together on the same chromosome, then the original combination of paternal and maternal alleles is more likely to be inherited together, and the loci are said to be linked (96, 104).

Linkage analysis finds a region of linkage on a given chromosome, however for pinpointing a susceptibility gene for a complex disease, and identifying the disease-associated polymorphism in a larger region where linkage has already identified, association analysis has to be used (96, 105).

1.4.5   The Swedish Twin Registry

The Swedish Twin Registry (STR) was established in the late 1950s to study the importance of smoking and alcohol consumption on cancer and cardiovascular diseases

while controlling for genetic effects. Since then the Registry has expanded and focus has been broadened to different complex diseases. Extensive information on environmental and life-style factors has been collected through the last 35 years. Currently the register holds more then 170,000 twins and is constantly updated with information from numerous national databases (i.e., inpatient discharges, causes of death and birth condition). Additionally, specific surveys have been directed to subpopulations, generating plentiful of environmental, lifestyle and behavior data. The STR is continuously updated by cross-matching with national healthcare databases, especially registries with information on patients' discharges, cancer diagnoses, causes of death and conditions during birth (94, 106).

There are three cohorts in the registry: The cohort born in 1886-1925, where each potential pair of twins was manually followed until status in 1959 was established. Questionnaires were sent out to this cohort first in 1960-61 and later in 1963 and 1967 as well as in 1970 with mainly demographic, medical and life-style character information with main focus at cardiovascular and respiratory diseases. In 1970 a new cohort of twins born between 1926-1958 was compiled and questionnaires were mailed out to this cohort in 1972-1973. The third cohort consists of twins born 1959-1990, where only a small number of twins have been contacted (94).

The STR was recently updated on exposure information and symptoms from a large number of diseases through an extensive telephone interview of twins born in 1958 or earlier. In the Screening Across the Lifespan Twin-study (SALT-study), the subjects were interviewed regarding their occupation, education, and consumption of alcohol, tobacco and caffeine. Checklists of common illnesses, prescription and non-prescription use of medications were asked for (94).

During the last decade there has been considerable effort in detecting QTL for complex traits and diseases. Twin researchers have developed methodological analyses for linkage and association studies. MZ twins allow for measuring of gene-environment interactions and DZ twins are valuable in both linkage and association studies. DZ twins can be used in affected sib-pair analyses of linkage and concordantly affected MZ twins are valuable in association studies. Availability of data of basic biometric parameters in large twin samples makes it possible to select discordant and concordant twin pairs for quantitative genetic calculations (94, 106).

The TwinGene project
The STR has recently completed a study called TwinGene, where the primary objective was to collect blood samples from male and female twins for the establishment of a biobank containing DNA and serum and plasma. Between the years 2004 and 2008 population wide collection of blood on 13,600 twins born 1958 or earlier has been undertaken. The objectives were to identify genomic regions harboring genes influencing common, complex disorders, focusing on cardiovascular, metabolic and inflammatory disorders as well as to enable prospective cohort studies to assess the relative role of disease genes, biomarkers and functional for disease development.

The aim of the TwinGene project has been to systematically transform the oldest cohorts of the Swedish Twin Registry (STR) into a molecular-genetic resource. Together with this epidemiological information the biological samples gathered within TwinGene creates a platform equipped to investigate the genetic influences to a broad spectrum of health

related traits and common diseases as well as gene-environmental interactions in the etiology of common diseases.

Health status data from a self-reported questionnaire, a basic health check-up and routine biochemistry for have been included. Twins were contacted each month until the data collection was completed. Invitations to the study contained information of the study and its purpose. When the signed consent forms where returned, the subjects were sent blood-sampling equipment and asked to contact a local health facility for blood sampling. Subjects living in vicinity to the cities of Stockholm, Gothenburg, Malmö or Västerås were given the option of visiting hospital blood sampling facilities, in which case the health check-up were omitted

1.4.6 Previous twin studies and GD

A possible genetic basis for gallstone disease has been suggested by studies in first-degree relatives of patients relative to controls. Studies from Sweden (107, 108), Israel (109) and India (110) showed significantly higher ($\geq$2:1) prevalence rates of GD in the first degree relatives of the probands compared with controls, even in younger patients and by comparing with the spouses of probands (108). These studies confirmed GD by cholecystography, surgery or ultrasonography. A more recent study from the US estimated that genetic factors are at least responsible for 30% of symptomatic GD (53).

Several anecdotal reports of concordance of cholesterol gallstones in small numbers of MZ twins have been published (111-115). In a large Danish study (116), 1,900 unselected twin pairs born between 1870 and 1910 were sent a questionnaire, which requested that they describe "any and all admissions to hospital, where, when and for what condition". Diagnosis of GD was obtained from the chart descriptions of the hospitalized cases. In the entire cohort, the crude incidence of cholelithiasis was 2.6%. Of a total of 101 twin pairs with a hospital diagnosis of cholelithiasis, concordance for the disease was found among 14/25 monozygotic pairs compared with 6/40 (same sex) and 0/36 (different sex) DZ pairs. Zero concordance among white twins of different sex most likely reflects the different frequencies of symptomatic cholelithiasis between men and women. Further, if silent gallstone cases were ascertained, the actual frequency and concordance of gallstones might have been much more impressive (116). Kesaniemi et al. randomly selected male twins (17 MZ and 18 DZ pairs) from the Finnish Twin Cohort (111). The males were living apart but residing in the Helsinki area. Their ages ranged from 43 to 58 years (mean, 50 years) and mean body weight were 78 kg. By oral cholecystography and history of cholecystectomy, gallstones were ascertained in seven MZ and three DZ subjects, with two MZ and none of the DZ twin pairs being concordant for gallstones giving 40% pair-wise concordance for the former and 0% for the latter. These studies, albeit imperfect, constitute the best information on familial occurrence and monozygotic twin concordance for gallstones.

2 AIMS

Study I aimed to calculate the relative importance of genetic and environmental factors in symptomatic GD by conducting a quantitative genetic analysis on a large twin population and calculate the percentage of genetic, shared environmental and unique environmental effects on the total phenotypic variance.

Study II aimed (i) to examine the association between BMI, alcohol and tobacco consumption and GD, and (ii) to investigate whether potential associations are confounded by genetic and /or shared environmental factors.

Study III aimed to validate the contribution of the *ABCG8* D19H allele to GD by conducting an association analysis on a cohort of concordant MZ and concordant DZ twins as well as stone-free twin controls.

3 MATERIALS AND METHODS

Studies I and II were based on data from the STR as well as Swedish Hospital Discharge and Causes of Death Registries. Indexes were not contacted personally.

In the first part of Study III, patients were sent written information before considering participating along with a questionnaire. Informed consents from patients who agreed to participate where returned with the data we asked for. In the second part of Study III, DNA was extracted from the TwinGene Database at Karolinska Institutet after approval from the STR steering group as well as a new approval by the Ethics Committee. The subjects that had previously participated at the TwinGene project had already signed an informed consent for DNA analysis in future studies that are approved by the Ethics committee.

Studies I and II were approved by the data inspection authority as well as the PUL (Personuppgiftslagen) *ombudsman* who protects and assesses use of the Swedish personal security number. Study III was approved by the Ethics Committee and conducted in co-operation with the Biobank at Karolinska Institutet. The STR steering group as well as the local Ethics Committee at Karolinska Institutet Stockholm, Sweden, approved all of the studies.

3.1 SUBJECTS

3.1.1 Study I

In Study I we linked the first two collected cohorts (C1, twins born between 1900 and 1938; and C2, twins born between 1939 and 1958) of the Swedish Twin Registry to the Swedish Hospital Discharge and Causes of Death Registries. We then screened the registries for gallstone disease and gallstone surgery–related diagnoses codes (International Classification of Diseases [ICD] by the Word Health Organization [WHO]), according to the following search-criteria: ICD-8: 574, 575, 576; ICD-9: 574.0-574.5, 576.0-576.9; and ICD-10: K563, K800-K805, K808, JKA20, JKA21, JKB11, JKE00, JKE02, JKE06, JKE12, JKE15, and JKF10.

The total number of twin pairs screened was 43,141. Zygosity data were provided by the registry and were determined by a questionnaire that has been shown in validation studies to classify correctly more than 98% of pairs of twins. Fifteen twins with unknown zygosity were excluded.

3.1.2 Study II

In Study II the STR was linked to the Swedish Hospital Discharge and Causes of Death Registries for twins born between 1886 and 1958(117). The study population comprised all 58 402 twins born 1886–1958 in the STR, consisting of 19,950 MZ twins, 33,464 DZ

twins, and 4,988 twins of unknown zygosity; 27,692 were male and 30,710 female. In the separate analysis for each potential risk factor, we included those same-sexed twins in the STR who responded to a questionnaire in 1961 or1973 regarding the risk factors studied (94). Twins born between 1886 and 1925 (cohort C1) were evaluated in 1961 for smoking habits, in 1963 for smoking habits and BMI, and in 1967 and 1970 for smoking habits, alcohol and body mass index (BMI). If data on the same variable were available at different times, the most recent value was used. The 1973 questionnaire evaluated twins born between 1926 and 1958 (cohort C2) for alcohol, BMI, and smoking and smoke-free tobacco habits (94). The follow-up times were January 1, 1970 to December 31, 2002, for C1; and January 1, 1974 to December 31, 2002, for C2. To avoid bias through later lifestyle changes, we excluded twins that had answered the questionnaires after the diagnosis of GD was made.

3.1.3   Study III

Initially, we screened for MZ twins with GD, born between 1912 and 1956, and living in the greater Stockholm area, by linking the STR with the Swedish Hospital Discharge and Causes of Death Registries, as for Study I. We found 42 concordant and 146 discordant MZ twins that were asked in a letter to participate in the study. They were invited to donate blood for clinical-chemical and DNA analyses and to return a questionnaire about possible GD (surgery or X-ray/ultrasound evaluations). For comparison, we also invited a small number of dizygotic (DZ) twins with known GD. Twins who did not respond to the invitation letter were sent a reminder after two weeks. The GD-free twin in discordant pairs as determined by register or questionnaire data was invited for an ultrasound scan of the gallbladder, which was performed by the author at Karolinska University Hospital Huddinge. Blood samples were collected at a local primary health facility and were sent to the Biobank at the Karolinska Institute where plasma centrifugation and DNA extraction was performed. DNA was stored at –80 ºC. During this process, the samples were only identifiable by a barcode. An automatic report was generated and sent in encrypted mail.

Invitation letters, important information to the study subjects, informed consent, referrals for blood sample donation, ultrasound protocol as well as the questionnaire sent to the participating subjects were all approved by the STR steer group, the local Ethics committee, and the Biobank at Karolinska Institutet that also approved study logistics, sample collection, DNA extraction and storage procedures, as described in the Study Integration Plan (SIP; see Appendix).

Seventy-three twins donated blood and participated, if necessary, in the ultrasound investigation. This particular study population consisted of 65 MZ twins, 50 of them in pairs, and 8 DZ twins, 6 of them in pairs. Fifty-two (80%) and 5 (62.5%), respectively, of MZ and DZ twins were females. All 8 DZ twins (5 females, 3 males) had GD. After re-evaluation of GD status with ultrasonography in the apparently unaffected co-twins, we found 24 MZ twin pairs with but only 8 MZ twins without GD in the whole cohort. A DNA zygosity analysis at the Biobank at Karolinska Institute confirmed that all of the twins classified as MZ by the STR based on a questionnaire in fact were MZ.

Unexpectedly, the number of twins living in Stockholm County finally participating in the study represented only 39% of the screening population (73 out of 188 invited). Since in particular the number of GD-free control twins was low, the power was insufficient for an

association analysis. Therefore, we applied for a permission to utilize DNA from blood samples already collected for the TwinGene Database at the Biobank at Karolinska Institutet.

For TwinGene, blood samples were collected from male and female twins residing in different parts of Sweden for the establishment of a biobank containing DNA, serum and plasma. Between the years 2004 and 2008, population wide collection of blood from 12,600 twins born 1958 or earlier has been undertaken. Together with epidemiological information the biological samples gathered within TwinGene create a platform equipped to investigate the genetic influences to a broad spectrum of health related traits and common diseases as well as gene-environmental interactions in the etiology of common diseases.

Health status data from a self-reported questionnaire, a basic health check-up and routine biochemistry have been included. Invitations to the study contained information of the study and its purpose. When the signed consent forms where returned, the subjects were sent blood-sampling equipment and asked to make an appointment at their local health-care facility on Monday to Thursday but not the day before a national holiday, in order to ensure that the samples for DNA extraction and clinical-chemical analysis would reach the Biobank at Karolinska Institutet the following morning by over-night mail. The subjects were instructed to fast from 20:00 hours the previous night. By venipuncture a total of 50 ml of blood was drawn from each subject.

The study population of TwinGene was recruited among twins participating in the Screening Across the Lifespan Twin Study (SALT), which was a telephone interview study conducted in 1998-2002. Other inclusion criteria were that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study if they previously declined participation in future studies or if they had been enrolled in other STR DNA sampling projects. Average response-rate was 53%. Age-specific response-rates were 27%, 48%, 63%, 57% and 43% for subjects born before 1921, in the 1920's, in the 1930's, in the 1940's, and in the 1950's, respectively.

We screened the TwinGene database for GD as we did above. This search resulted in the identification of additional 20 concordant MZ pairs as well as 54 DZ twin individuals (26 twins from concordant DZ pairs, i.e., 13 pairs, as well as 28 twins from concordant DZ pairs where DNA was available only for one twin in each pair). Of the 20 MZ twins with GD, 2 were males and 18 females. Of the 54 DZ twins with GD, 17 were males and 37 females. From the TwinGene database 109 concordantly stone-free MZ twin pairs, 18 male and 91 female, as well as 126 non-related DZ twin individuals from 126 concordantly stone-free pairs (one from each pair) were selected as controls 44 males and 82 females. The controls were frequency matched by age, sex and zygosity to represent the distribution observed among the cases. Thus, the number of asymptomatic controls included was more than twice the size of the number of cases. The over-sampling of controls was done in order to compensate for the reduced power that may result from potential misclassification of controls due to lack of questionnaire or ultrasound verification of disease free status.

3.2 STATISTICS

3.2.1 Study I

In Study I we used SEM, also known as covariance modelling as a general approach for the analysis of variance and correlations. In SEM, genotypic and environmental effects are modeled as the contribution of unmeasured (latent) variables to the potentially multivariate phenotypic differences between individuals. The latent variables' contributions are estimated as regression coefficients in the linear regression of the observed variables on the latent variables by the maximum likelihood and weighted least squares (94). Data on all types of twins (male, female, MZ, DZ) are incorporated simultaneously and provide estimates of the variables. By including OS DZ twins one can compare phenotypic identity for symptomatic GD between twins of opposite sexes. To estimate the relative importance of genetic factors and to test whether these differ between men and women, models based on 2-by-2 contingency tables (twin A's status by twin B's status) on categorical data (dichotomous, i.e., disease or no disease) were constructed for MZ females, DZ females, MZ males, DZ females, and OS pairs. The software package used was *Mx* http://www.vcu.edu/mx (117)

The probandwise concordance (C) was calculated as the proportion of all persons with symptomatic GD whose twins had symptomatic GD metachronously (118, 119).

Concordance rate =

$$2 \cdot \text{concordant affected pairs} / (2 \cdot \text{concordant affected pairs} + \text{discordant pairs})$$

The 95% confidence intervals for C (CIc) were calculated as:

$$\text{CIc} = p \pm z \sqrt{(p(1-p)/n)}$$

where p, proportion of concordance; z =1.96, coefficient for a 95% confidence interval; and n, number of cases.

The relative risk for symptomatic GD for subjects whose twin had symptomatic GD compared with subjects whose twin did not was estimated as an odds ratio (OR) and was calculated as:

$$\text{OR} = a \cdot d / b \cdot c$$

where a, number of concordant pairs; b and c, each half the number of discordant pairs; and d, number of pairs without disease. The 95% confidence intervals for the risk (CI r) were estimated according to the Mantel-Haenszel method (120), using the SISA statistical program (SISA Binomial [database online]. Hilversum, The Netherlands: Uitenbroek; 2005. Available at: http://home.clara.net/sisa/binomial.htm).

In addition to concordance rates, tetrachoric correlations were calculated for MZ, DZ, and OS twin pairs. Tetrachoric correlations are calculated for two normally distributed phenotypic variables that are both expressed as a dichotomy (disease or no disease) and reflect the similarity of twin pairs. Thus, differences in correlations between various groups provide information about the presence of genetic effects. For example, if MZ twins display higher tetrachoric correlation coefficients than DZ twins, genetic effects are important.

The overall phenotypic variance (VP) is divided into

(1) one component due to inherited genetic factors
    (G = A · D; additive A or non-additive/dominant D),
(2) one component due to common environmental factors (C), and
(3) another component due to environmental factors unique for each twin (E).

Assuming heritability in the narrow sense (i.e., the absence of non-additive genetic variance [G = A]), the equation for variance (Vp) for one of the twins in a pair can be written as (94):

$$Vp = a \cdot A + c \cdot C + e \cdot E$$

Since heritability is not a universal factor but depends on the population, sex, and cohort being measured, we tested different models for males and females in two separate cohorts as well as both cohorts as a whole. An underlying normal distribution of susceptibility to the disease was assumed. A threshold value was defined as the sum of effects of many genetic and environmental factors that has to be exceeded for the disease to manifest itself. In the saturated model, the threshold value was calculated from the clinical prevalence. For model evaluation, a likelihood ratio test was used. The difference between twice the log-likelihood can be interpreted as a $\chi 2$ statistic. The principle of parsimony indicates that the model with fewer parameters to be estimated that still fits the data best is to be chosen.

The usual assumptions for a twin study were made, i.e., no random mating (since we just aimed to study the influence of the genotype), no gene/environment interaction, in that MZ twins share their entire genome whereas DZ share 50% of their segregated genes, equivalent environments (including prenatal) for MZ and DZ twins, known zygosity as well as the assumption that the twins are representative of the general population (94).

3.2.2   Study II

In Study II, we performed first a cohort study comparing cases to unaffected unrelated twins as well as a co-twin study comparing cases to unaffected co-twins. The covariates studied were BMI, alcohol consumption, smoking a well as the use of smoke-free tobacco (snuff) and were categorized as follows:

Body mass index [kg · m^2] data were categorized according to the WHO (121) as normal (18.5–24.9), overweight (25.0–29.9) and obese (≥30.0). Alcohol was stratified according

to the total amount consumed (122, 123) as nondrinkers (0 g / month; for both women and men), moderate consumers (>0–1800 g / month for women; >0–2400 g /month for men), and high consumers (>1800 g / month for women; >2400 g / month for men), i.e., statistically significantly associated with increased risk for liver disease (122, 123). Smoking was categorized as never, previous or current for all forms of smoking. The same principle was applied to smoke-free tobacco (in Swedish, *snuff*; an oral tobacco preparation popular in Sweden).

3.2.2.1 Cohort study comparing cases to unaffected unrelated twins

In the study population of 58,402 twins, we identified 1,666 twins with GD. In this cohort, logistic regression analysis for gender, age, zygosity, BMI and alcohol and tobacco habits was performed, including both concordant and discordant pairs, regardless of zygosity status. Dependence within twin pairs was accounted for by using Generalized Estimation Equation models (GEE) with SAS PROC GENMOD (124).

3.2.2.2 Co-twin study comparing cases to unaffected co-twins

Co-twin comparison of same-sexed twins was performed in 1,527 cases with GD and where the co-twin was without a history of GD. The analysis was subdivided in MZ and DZ twin pairs to investigate genetic or shared environmental factors. The co-twin analyses were performed with conditional logistic regression by the maximum likelihood method using SAS PROC PHREG (125),(126). Odds ratios (OR) and 95% confidence intervals (CI) were calculated.

By investigating the association within twin pairs discordant for GD, the influence of genetic and shared environmental factors is substantially reduced. Twins within the same pair share the same environment during infancy and childhood, so differences within MZ and DZ twin pairs should be independent of common environmental factors. In addition, within MZ twin pairs, differences are independent of genetic factors. If there were a causal effect of the risk factor on GD, we would expect the same association in GD discordant MZ and DZ twin pairs as in the whole study population that served as controls. On the other hand, if genetic effects were confounding the association we would expect the same association in discordant DZ twin pairs as well as in co-twins from the whole study population, but not in discordant MZ twin pairs. If shared environmental factors were confounding the association we would expect no difference between MZ and DZ twin pairs but a different association in co-twins from the whole study population. These conclusions are based on the assumption that any differences existing between MZ co-twins must necessarily originate from environmental influences including shared environmental factors. Differences between DZ co-twins could both be due to genetic and early environmental factors.

3.2.3 Study III

In Study III, allele and genotype frequencies were determined and tested for consistency with Hardy-Weinberg equilibrium using an exact test.

The association analysis was performed considering both MZ and DZ twins simultaneously and takes into account the dependence between twins in a pair (only one of the twins for each MZ pair was used while both members of 21 DZ pairs were used in the analysis). Allele and genotype frequencies were compared between cases and controls employing SAS PROC GENMOD (124).

3.3   GENOTYPING

Genomic DNA was isolated from EDTA anti-coagulated blood and transferred to the analyzing laboratory at Saarland University Hospital. DNA concentrations were determined photometrically (NanoDrop ND-1000 spectrophotometer, Peqlab Biotechnologie, Erlangen, Germany). For genotyping, we selected the functionally relevant non-synonymous coding SNP of the *ABCG8* gene rs11887534 (c.55G>C, p.D19H (76)). The SNP was genotyped using Taqman assays, as described (76). Since SNP rs11887534 is of the G ↔ C variant type, the identity of the minor allele is critical. The primer used identifies the C-alleles as minor alleles for rs11887534.

3.4   BIOCHEMISTRY

A biochemical analysis was performed from twins initially screened in Stockholm County. Blood sample donation was performed at the Karolinska University Hospital or at their local Health Care Center. Each study subjects' blood samples and referrals were packed in an envelope and sent to the Biobank at Karolinska Institutet for DNA extraction and storage as described in the Appendix. Prior to that, routine biochemistry analysis was performed at the Central Laboratory at Karolinska University Hospital. The samples were analyzed for: Hemoglobin, HbA1c, cholesterol (HDL- and LDL-cholesterol), triglycerides, ASAT, ALAT, ALP, γGT, bilirubin and CRP.

3.5   QUESTIONNAIRE

For the twins initially screened in Stockholm County, a questionnaire was sent to and answered by twins who consented to participating in the study. The questions addressed regarded diagnosis of GD (known diagnosis or not), surgery for GD (open or laparoscopic as well as age for surgery), family history defined as the number of first–degree relatives with GD, a history of diabetes (tablets or insulin), weight and height, tobacco and alcohol consumption stratified by grams of alcohol consumed during a month' period, as well as fecundity and hormonal replacement therapy for women).

4   RESULTS

4.1   STUDY I

From the total twin population of 43,141 pairs, 5,970 pairs of unknown zygosity consisting of 235 discordant, 40 concordant, and 5,677 healthy pairs were excluded from further calculations. The age range was 64 to 102 years in cohort 1 (C1) and 44 to 63 years in cohort 2 (C2). Among the 43,141 pairs evaluated in the whole cohort (C) we found a total of 4,394 individuals with symptomatic GD. The overall prevalence of symptomatic GD was 6.5% in C1 (7.3% and 6.8% for MZ and DZ females, 5.9% and 5.5% for MZ and DZ males, and 6.7% for OS twins, respectively), and 3.5% in C2 (5.2% and 4.7% for MZ and DZ females, 1.7% and 2.0% for MZ and DZ males, 3.5% for OS twins, respectively). Table 1 displays the probandwise concordance rates for symptomatic GD in MZ and DZ twins of both sexes as well as of OS pairs.

Concordance rates ranged from 6% for affected females in OS twin pairs to 24% for female MZ twins in C2. Concordance rates were higher for MZ compared with DZ twins, for both women and men. The differences between MZ and DZ twins were more pronounced in the younger cohort, C2. This result is also reflected by the odds ratios for symptomatic GD that ranged from 1.9 for OS twins in both cohorts to 17.6 for male MZ twins in the younger cohort 2 (Table 1).

The tetrachoric correlations (r) were estimated by *Mx* and are also shown in Table 1. Similar to concordances rates, MZ similarity exceeded DZ similarity in all cases indicating the presence of genetic effects. The correlations were generally higher in the younger cohort, although there was still overlapping CI compared with C1.

We found that MZ <2 rDZ in all cases, which implies a better fit of the ACE than the ADE model according to the algorithms used in SEM. In practice, it indicates that shared environmental effects are of more importance than dominant genetic effects. Significant sex differences were not found.

Cohort	Twin Type	Healthy Pairs	Discordant Pairs	Concordant Pairs	Probandwise Concordance Rate[a][b]	Odds Ratio[b]	Tetrachoric correlation[b]
FEMALES							
C1	MZ	3013	410	49	0.19 (0.16-0.23)	3.5 (2.5-4.9)	0.33 (0.23-0.42)
	DZ	5529	735	63	0.15 (0.12-0.17)	2.6 (1.9-3.4)	0.24 (0.16-0.32)
C2	MZ	2378	208	32	0.24 (0.19-0.29)	7.0 (4.5-10.9)	0.48 (0.36-0.58)
	DZ	3414	305	25	0.14 (0.11-0.18)	3.7 (2.3-5.8)	0.31 (0.19-0.42)
C	MZ	5391	618	81			0.39 (0.32-0.46)
	DZ	8943	1040	88			0.26 (0.20-0.33)
MALES							
C1	MZ	2498	280	24	0.15 (0.11-0.19)	3.1 (1.9-4.9)	0.28 (0.15-0.39)
	DZ	4096	456	25	0.10 (0.08-0.13)	2.0 (1.3-3.0)	0.16 (0.05-0.27)
C2	MZ	2116	58	7	0.19 (0.12-0.30)	17.6 (7.1-43.5)	0.56 (0.35-0.73)
	DZ	3403	124	8	0.11 (0.07-0.18)	7.1 (3.3-15.4)	0.39 (0.20-0.55)
C	MZ	4614	338	31			0.37 (0.26-0.47)
	DZ	7499	580	33			0.24 (0.15-0.33)
OPPOSITE-SEXED TWIN PAIRS							
C1	Female	5217	419	47	0.12 (0.10-0.14)	1.9 (1.4-2.7)	0.16 (0.04-0.28)
	Male		293				0.17 (0.05-0.28)
C2	Female	7380	389	17	0.06 (0.04-0.09)	1.9 (1.2-3.2)	0.12 (0.04-0.28)
	Male		123				0.14 (0.01-0.29)
C	Female	12597	808	64			0.31 (0.20-0.40)
	Male		416				0.09 (0.01-0.18)

[a]Probandwise concordance rate
= (number of affected twins in concordant pairs) / (total number of affected twins).
[b]95% confidence intervals in parentheses.

Table 1: Probandwise concordance rates

4.1.1 Model fitting

Different model assumptions were tested and the results are presented in Table 2. In a first attempt at model fitting, the whole cohort was analyzed. We found significant heritability both for females and males (25%; 95% CI, 9%-40%). The next step was to analyze the two cohorts separately with respect to females and males. When decomposing the phenotypic variance by SEM separately for both cohorts and both sexes, statistically significant effects of heritable factors were only observed for female twins in the younger cohort C2 (33%; 95% CI, 1%-58%). Heritability seems to be generally higher in the younger cohort, particularly for men in C2. Table 2 presents statistics on model fitting and the estimates of variance components based on the best-fitting models. Two different assumptions were tested: When different estimates for A, C, and E were assumed for the model, *Mx* returned no significant differences in the estimated parameters between sexes with a good fit to the model. This was confirmed by remodeling with identical estimates for both sexes from the beginning (ACE2 model in Table 2), supporting the assumption of no difference in estimates between sexes.

The ACE2 model represented the best fit according to Akaike's Information Criterion. Parameters in the best-fitting model ACE2 are estimated as follows: A, 25%; C, 13%; and E, 62%. We also tested the whole cohort for the ADE model, which was inferior to the ACE2 model (data not shown). CIs were generally large and overlapping in C1 and C2, suggesting calculations to be performed for the cohort as a whole and with the assumption of no sex differences, as reasoned above. Calculations for the whole cohort were important in terms of statistical power, since the number of males is relatively small compared with females. The model was run for 4 and 6 groups, respectively, excluding or including OS twin pairs, once again showing no differences in results, although there was an unsatisfactory model fit for the 6 group model. Thus, all data are presented for the best-fitting models ACE and ACE2 using 4 groups.

		Parameter Estimates				Fit of Model			
Model	Sex	A: Genetic Effects[a]	C: Shared Environmental Effects	E: Unique Environmental Effects	Cohort	X^2	Df	Probability	AIC
ACE	Women	0.17 (0-0.40)	0.16 (0-0.32)	0.67 (0.58-0.77)	C1	2.255	6	0.895	-9.745
ACE	Men	0.23 (0-0.39)	0.05 (0-0.27)	0.73 (0.62-0.85)	C1				
ACE2	Women	0.19 (0-0.37)	0.12 (0-0.27)	0.69 (0.62-0.77)	C1	4.196	8	0.839	-11.804
ACE2	Men	0.19 (0-0.37)	0.12 (0-0.27)	0.69 (0.62-0.77)	C1				
ACE	Women	0.33 (0.01-0.58)	0.14 (0-0.40)	0.53 (0.42-0.64)	C2	2.589	6	0.858	-9.411
ACE	Men	0.40 (0-0.74)	0.18 (0-0.54)	0.42 (0.26-0.63)	C2				
ACE2	Women	0.33 (0.06-0.58)	0.16 (0-0.38)	0.50 (0.41-0.60)	C2	3.877	8	0.868	-12.123
ACE2	Men	0.33 (0.06-0.58)	0.16 (0-0.38)	0.50 (0.41-0.60)	C2				
ACE	All	0.25 (0.09-0.40)	0.13 (0.01-0.25)	0.62 (0.56-0.68)	C	1.868	8	0.985	-14.132
AE	All	0.41 (0.36-0.46)	-	0.59 (0.54-0.64)	C	6.564	9	0.682	-11.436

C1, Cohort1; C2, Cohort 2; C, whole cohort; df, degrees of freedom;
AIC, Akaike Information Criterion (X^2 – 2 df); [a]95% confidence intervals in parentheses.
ACE2; different prevalence assumed, same estimates assumed for males and females.
Model in bold type the most suitable model according to the principle of parsimony

Table 2: Model fitting

4.2   STUDY II

The distributions of the covariates for twins with (cases) and without GD (controls) are presented in Table 3. There were more overweight twins with than without GD (32% vs. 23%). Moderate to high alcohol consumption was slightly more prevalent in twins without GD. The majority of the twins did not consume any tobacco products and the pattern of tobacco consumption was similar in the two groups. In the cohort study, we included all twins and calculated the risk for GD and evaluated each parameter per se in comparison with unrelated twins. The crude risks for BMI, smoking, smoke-free tobacco, and alcohol, respectively, and GD are presented in Table 4. Women had a significantly higher, more than doubled risk for GD. Overweight and obese twins had a higher risk for GD (OR, 1.86 and 3.38), which was statistically significant different from twins with normal BMI. Alcohol consumption at high levels had a statistically significant, negative association with GD (OR, 0.62). Previous or current smoking or use of smoke-free tobacco did not have a statistically significant impact for GD (Table 4). Results from the additional within-pair analyses, aiming to control for potential unmeasured confounding, are presented in Table 5. When we compared the cases with their healthy co-twins, we found a statistically significant decreased risk for GD when the healthy DZ co-twin was obese. No statistically significant difference was found in the association within discordant MZ and DZ twins regarding overweight or alcohol and GD. These results indicate that the positive association between overweight and GD, as well as the negative association of high alcohol consumption on GD observed in the cohort analyses probably have been confounded by shared environmental factors.

Variable	Type	Twins with Gallstone Disease N (%)	Twins without Gallstone Disease N (%)
Gender	female	1076 (65)	29634 (52)
	male	590 (35)	27102 (48)
	data missing	0	0
Zygosity	MZ	586 (35)	19364 (34)
	DZ	941 (56)	32523 (57)
	unknown	139 (9)	4849 (9)
BMI	normal	878 (63)	32547 (74)
	overweight	450 (32)	10150 (23)
	obese	70 (5)	1553 (4)
	data missing	268	12486
Alcohol	never	1153 (81)	33579 (73)
	moderate consumption	132 (9)	5383 (12)
	high consumption	139 (10)	7020 (15)
	data missing	242	10754
Smoking	never	832 (55)	25236 (51)
	previous	498 (33)	16853 (34)
	current	169 (12)	7108 (14)
	data missing	167	7539
Smoke-free tobacco	never	691 (96)	26024 (91)
	previous	20 (3)	1981 (7)
	current	7 (1)	524 (2)
	data missing	948	28207

Table 3: Demographics of twins with and without GD.

	Variable	OR	CI
Gender	Female	1.00 •	
	Male	0.47*	(0.39-0.56)
BMI	Normal	1.00 •	
	Overweight	1.86*	(1.52-2.28)
	Obese	3.38*	(2.28-5.02)
Alcohol	never	1.08	(0.96-1.20)
	moderate consumption	1.00 •	
	high consumption	0.62*	(0.51-0.74)
Smoking	Never	1.00 •	
	Previous	1.15	(0.95-1.39)
	Current	0.99	(0.78-1.27)
Smoke-free	Never	1.00 •	
tobacco	Previous	0.62	(0.37-1.04)
	Current	1.05	(0.49-2.23)

Significant differences marked by *. Reference category marked by •

Table 4: Cohort study. Crude odds ratios (OR) and 95% confidence intervals (CI) for gender, BMI and alcohol and tobacco use of twins with and without GD.

Variable	Zygosity	OR	CI
Overweight	MZ	1.16	(0.86-1.56)
	DZ	1.17	(0.97-1.42)
Obesity	MZ	0.79	(0.59-1.05)
	DZ	0.83*	(0.69-0.99)
No alcohol consumption	MZ	0.84	(0.64-1.11)
	DZ	0.99	(0.82-1.20)
High alcohol consumption	MZ	1.08	(0.82-1.42)
	DZ	0.96	(0.79-1.16)
Previous smoker	MZ	1.17	(0.93-1.47)
	DZ	1.08	(0.91-1.28)
Current smoker	MZ	0.85	(0.68-1.07)
	DZ	0.90	(0.76-1.06)
Previous use of	MZ	0.74	(0.32-1.75)
smoke-free tobacco	DZ	0.78	(0.42-1.45)
Current use of	MZ	1.20	(0.52-2.80)
smoke-free tobacco	DZ	1.26	(0.68-2.36)

Significant differences marked by *.

Table 5: Co-twin analysis. Odds ratios (OR) and 95% confidence intervals (CI) for symptomatic GD in same gender MZ and DZ twin pairs discordant for GD.

4.3 STUDY III

4.3.1 Genotyping

In Stockholm County, 42 GD-concordant (18 MZ and 24 DZ twins) and 146 GD-discordant MZ twins were screened. A total of 73 twins, 65 MZ and 8 DZ, consented and were included in the study. In the concordant cohort, we had 16 twins with GD (6 concordant MZ twins in pairs, 2 MZ twins from concordant pairs, 6 concordant DZ twins in pairs, and 2 DZ twins from concordant pairs). In the discordant cohort, we had 49 MZ twins with GD (27 MZ twins with GD according to the inpatient register, 13 MZ twins with GD according to the questionnaire, 9 MZ twins with ultrasonography verified GD and 8 MZ twins without GD. The sample information for all 73 twins genotyped is shown in Table 6, including analyses of both MZ twins from a pair. We thus identified only 8 GD negative twins.

The *ABCG8* D19H variant was successfully genotyped in all samples. The genotype distributions did not deviate from Hardy-Weinberg equilibrium (exact tests, all p>0.05). Overall frequency of the D19H minor allele among affected twins with unique MZ genomes (i.e., considering only one MZ twin from a pair) was 20.8%. In contrast, the overall allele frequency in Caucasian controls was expected to be less than 10% according to allele frequencies previously reported in the Entrez SNP database (http://www.ncbi.nih.gov/) and previous publications (127-132). Table 7 summarizes the allele and genotype distributions for the ABCG8 variant. Only one twin (DZ with GD) carried the CC allele. Considering unique genomes, we found that of 32 MZ twin pairs, 24 were concordant and 8 were discordant for GD (Table 7).

Hetero- or homozygous 19H carriers were observed in 6 MZ twin pairs, 5 of which were concordant for GD. All of the 8 DZ twins were concordant for GD (6 DZ twins in concordant pairs and 2 single DZ twins from concordant pairs). Of these, 4 carried the p. D19H allele (all females) and 4 did not. The small number of DZ twins as shown in the Table 7 was added to and evaluated within the DZ cohort of the nationwide screening. Discordant MZ twins were disregarded from further calculations as they cannot provide information in the association analysis (the same genotype contributes once to cases and once to controls).

ABCG8 D19H in Stockholm County twins

	No GD		GD
N	8 (11.0%)		65 (89.0%)
Males	0		16 (24.6%)
Females	8 (100%)		49 (75.4%)
Age (median, range)	61 (50-81)		59 (46-87)
allele/genotype			
GG	7 (87.5%)	49 (75.4%)	
GC	1 (12.5%)		15 (23.1%)
CC	0		1 (1.5%)

Table 6: Sample information and alleles and genotypes (count/frequency) for D19H in Stockholm County twins.

Unique genome analysis for ABCG8 D19H in Stockholm County twins

	MZ		DZ	
	No GD	GD	No GD	GD
N	8 (25.0%)	24 (75.0%)	0	8 (100%)
Males	0	6 (25.0 %)	0	3 (37.5%)
Females	8 (100%)	18 (75.0%)	0	5 (62.5%)
Age (median, range)	60 (54-86)	64 (51-78)		68 (51-79)
allele/genotype				
GG	7 (87.5%)	19 (79.2%)	0	4 (50%)
GC	1 (12.5%)	5 (20.8%)	0	3 (37.5%)
CC	0	0	0	1 (12.5%)

MZ cases, number of unique MZ genomes in each concordant MZ pair;
DZ cases, number of twins in concordant DZ pairs;
GD, no GD: number of unique MZ or DZ genomes with or without GD

<u>Table 7:</u> Sample information for unique MZ genomes and DZ genomes and alleles and genotypes (count/frequency) included for D19H analysis for twins screened in Stockholm County.

From the nationwide screening we included 20 MZ unique genomes from concordant MZ pairs with GD (the genotype from each pair only represented once), 54 DZ genomes from concordant DZ pairs with GD (one or both twins in each pair where available for genotyping) as well as 109 MZ unique genomes from stone-free MZ controls and 126 DZ genomes from stone-free non-related DZ controls. The total cohort, including the twins from Stockholm County, was used for the final analyses, as summarized in Table 8. GD-free twins from TwinGene served as controls, matched for age, zygosity and gender.

Unique genome analysis for ABCG8 D19H in Swedish twins

	MZ		DZ	
	No GD	GD	No GD	GD
N	109 (84.5%)	44 (27.3%)	126 (67.0%)	62 (33.0%)
Males	18 (16.5%)	8 (18.2%)	44 (34.9%)	20 (32.3%)
Females	91 (83.5%)	36 (81.8%)	82 (65.1%)	42 (67.7%)
Age (median, range)	68 (52-80)	65 (51-78)	70 (50-90)	70 (50-81)
allele/genotype				
GG	99 (90.8%)	36 (81.8%)	114 (90.5%)	48 (77.4%)
GC	9 (8.3%)	8 (18.2%)	11 (8.7%)	13 (21.0%)
CC	1 (0.9%)	0	1 (0.8%)	1 (1.6%)

MZ, unique genomes in concordant MZ pairs with GD and stone-free unique MZ genomes; DZ, twins in concordant DZ pairs with GD and non-related stone-free DZ twins.

<u>Table 8:</u> Sample information and alleles and genotypes (count/frequency) for ABCG8 D19H in all unique MZ genomes, DZ genomes and controls for Swedish twins included in a nationwide screening.

27

For the association analysis, as summarized in Table 9, we considered for power reasons the nationwide-screened MZ and DZ cases vs. controls, i.e., 44 MZ and 62 DZ cases vs. 109 MZ and 126 DZ controls. We observed that hetero- or homogeneous carriers of the D19H allele had a significantly increased risk for developing GD; the OR was 2.56 (95% CI, 1.28-5.15; p=0.008). There were 18.2% D19H carriers among the MZ cases compared to 9.2% in the MZ controls, and 22.6% D19H carriers among the DZ cases compared to 9.5% in the DZ controls. Overall, the D19H frequency was 20.8% (22 of 106) in cases compared to 9.4% (22/235) in controls.

Allele frequencies for D19H in Swedish twins with and without gallstones

	No GD	GD	Total
N	235 (68.9%)	106 (31.1%)	341 (100%)
GG	213 (71.7%)	84 (28.3%)	297 (100%)
GC	20 (48.8%)	21 (51.2%)	41 (100%)
CC	2 (66.7%)	1 (33.3%)	3 (100%)

Table 9: Sample information and alleles and genotypes (count/frequency) for ABCG8 D19H in cases and controls used in the association analysis for Swedish twins included in a nationwide screening.

4.3.2   Biochemical analysis

Serum cholesterol and triglycerids of twins screened in Stockholm county are shown in Table 10. Genotypes GC and CC are shown as one group due to the small number of CC positive twins. Twins hetero- or homozygous for the ABCG8 D19H variant (genotypes GC and CC) have lower total and LDL-cholesterol levels than twins with GD in general, as well as wild-type twins with GD. The opposite effect is observed for HDL-cholesterol. However, these differences were statistically not significant.

genotype	Gallstone disease			No gallstone disease		
	all	GG	GC / CC	all	GG	GC / CC
Total cholesterol	5.53	5.58	5.38	5.98	6.02	5.8
SD	1.04	1.07	0.97	0.89	0.99	0
HDL-cholesterol	1.50	1.48	1.56	1.53	1.42	2.1
SD	0.50	0.53	0.40	0.35	0.24	0
LDL-cholesterol	3.45	3.53	3.22	3.93	4.04	3.4
SD	0.90	0.91	0.85	0.72	0.75	0
Triglycerides	1.31	1.28	1.39	1.23	1.32	0.79
SD	0.56	0.59	0.45	0.36	0.33	0

Table 10: Plasma total, HDL- and LDL-cholesterol (mmol/L ± standard deviation), stratified according to D19H alleles in twins with and without gallstone disease screened in Stockholm County.

4.3.3 Questionnaire

The questionnaire results in the first part of Study III were: There were no statistically significant differences in diabetes mellitus, age at diagnosis, parity or hormonal replacement therapy in women though twins hetero- or homozygeous for the ABCG8 D19H genotype were generally more obese and drank less than wild type twins.

	All	GG	CG+CC
GD	67 (100%)	51 (76.1%)	16 (23.9%)
Surgery	52 (77.6%)	38 (56.7%)	14 (20.9%)
Open surgery	35 (52.2%)	24 (35.8%)	11 (16.4%)
Lap surgery	16 (23.9%)	13 (19.4%)	3 (4.5%)
Age at surgery ± SD	42.2 ± 12.0	41.3 ± 12.4	44.6 ± 10.9
1 degree relatives	1.6 ± 0.9	1.5 ± 0.9	1.8 ± 0.7
DM	6 (9.0%)	5 (7.5)	1 (1.5%)
DM tablets	5 (7.5%)	4 (6.0%)	1 (1.5%)
DM insulin	1 (1.5%)	1 (1.5%)	0
BMI ± SD	26.4 ± 4.9	25.9 ± 3.8	28.2 ± 7.3
Non-smokers	36 (53.7%)	28 (41.8%)	7 (10.5%)
Smokers	6 (9.0%)	6 (9.0%)	0
Previous smokers	25 (37.3%)	17 (25.4%)	8 (11.9%)
Alcohol (drinks) ± SD	2.8 ± 3.4	3.1 ± 3.6	1.6 ± 2.0
Parity (female)	1.3 ± 1.2	1.3 ± 1.2	1.3 ± 1.2
Hormone therapy (female)	10 (14.9%)	5 (7.5%)	5 (7.5%)

Table 11: Questionnaire data of twins screened in Stockholm County and included in the ABCG8 D19H genotype analysis.

5 GENERAL DISCUSSION

5.1 GENETIC VS. ENVIRONMENTAL EFFECTS

In the first study we investigated the impact of genetic and environmental factors on symptomatic GD by linking the STR to the Swedish Inpatients' Discharge Registry and the Causes of Death Registry. We found an increased risk for GD in co-twins of affected individuals among the MZ twins of both sexes regardless of the age cohort they belonged to. In all cases the correlations of MZ twins were greater than twice the correlations of DZ twins indicating the influence of additive genetic factors.

Concordance rates observed in previous twin studies with a total number of 148 twin pairs were 38% for MZ and 8% for DZ twins, as summarized in our previous study (117). Our data derived from an almost 30-fold larger cohort confirmed the higher concordance rate for MZ twins; however, only with a 2-fold increase, which points to the importance of environmental factors. Males had generally higher concordance rates than females, though results did not reach statistical significance. Testing different models including OS twins also supported the assumption that there is no difference GD in heritability between males and females.

We estimated the heritability to be 25% (95% CI, 9%-40%) for symptomatic GD in this very large sample. Nakeeb et al. (53) showed a heritability of 29% based on a self reported questionnaire whereas Duggirala et al. (52) estimated a heritability of 44%, though in a variance component linkage analysis models that did not discriminate between shared and non-shared environmental effects.

Our major knowledge on "lithogenic" genes comes from animal models. Previous human studies were scarce, including apolipoprotein E, the hepatic phospholipid transporter ABCB4 and the rate-limiting enzyme of bile acid synthesis CYP7A1. Recently though, human studies have identified the hepatic cholesterol hemitransporter ABCG8 as a risk factor for GD at p. D19.

It is important to keep in mind that this study has identified shared environmental factors to be of significance for GD, contribution by 13% (95% CI, 1%-25%). A gallstone-promoting diet during childhood, weight changes, biliary infections and recently even enteric *Helicobacter* species have been postulated as potential risk factors along with other unspecified environmental factors (1, 50). However, non-shared environment has been found to be the major contributing factor in symptomatic GD accounting for 62% (95% CI, 56%-68%) of the total variance. This is also supported by the fact that there are more discordant than concordant pairs in each group.

It is more likely that habits acquired later in life gain importance when the familial cluster is abandoned and thus, heritability is more obvious in the younger cohorts as is illustrated by the highest odds ratio of 17.2 for MZ males in C2. Even if the power of shared environmental factors is weak, which is often the case in twin studies on dichotomous phenotypes, simplified models did not allow disregarding this factor.

The prevalence of symptomatic GD in twins was 6.5% in the younger and 83% for the older cohort. Taking into account the fact that our data only include symptomatic GD,

most of which had to qualify for operability, one can extrapolate a prevalence of 1.6% in C1 and 13% in C2 considering that the other studies included approximately 50% of asymptomatic subjects. Heritability was shown to be higher for symptomatic than asymptomatic GD rendering symptomatic cohorts more suitable for genetic studies.

Family clustering in GD seems to be the effect of additive genetic factors, at least to some extent, though difficult to distinguish from shared-environmental factors even in such a large study. There may be different thresholds for the disease but there were no differences in heritability suggesting that the same genes operate in males and females. The age at diagnosis was not available and thus, the age-cohort was used as a rough indicator of age dependency.

5.2   BMI, ALCOHOL, TOBACCO AND SYMPTOMATIC GD

In the second study we investigated the impact of BMI, alcohol, smoking and smoke-free tobacco on symptomatic GD. In our first study we found that women were at a higher risk for GD. In this study we showed that this risk increased with increasing BMI, especially in overweight and obese twins (OR, 1.86 and 3.38, respectively).

In the co-twin analysis we could further study the aetiology of the association between BMI and GD- There were only minor increases in GD risks for discordant MZ and DZ overweight twins with no difference between zygosity groups (OR, 1.17 and 1.16, respectively). In contrast to the analysis with unrelated twin controls, the co-twin analysis suggested significantly lower risk for GD in DZ obese twins (OR 0.83). Obese MZ twins had also a lower, though not significantly lower, risk for GD (OR 0.79). One might speculate whether the obese twin of a DZ twin with GD restrained from further evaluation for GD as he or she was afraid of discomfort or surgery complications.

There were no differences between the zygosity groups in both overweight and obese twins, though there was a difference compared to the whole population suggesting shared environmental factors confounding this association. Possible such factors would be maternal diet or alcohol during pregnancy and childhood, which influence both BMI and GD.

The impact of alcohol consumption was more difficult to define. The risk was significantly decreased in twins with higher alcohol consumption (OR 0.62; CI 0.51-0.74) when compared to twins with moderate alcohol consumption, though these did not differ from twins abstaining from alcohol. Previous studies suggested that moderate alcohol consumption is associated with a lower risk for GD (133-138). In our study there was no difference within discordant DZ and MZ twins in the high alcohol consumption group (OR, 0.96 and 1.08, respectively; CI, 0.79-1.16 and 0.82-1.42), suggesting confounding by shared environmental factors.

Tobacco use was not found to have a significant association to GD, which is consistent with some (134, 139) but not all epidemiologic studies (133, 139-144). However, our data on smoking and smoke-free tobacco were limited.

This study of lifestyle-related environmental factors has its limitations. As to BMI, it was not able to define the proportion of GD that is due to obesity or heredity. For this purpose, one would have to conduct a bivariate quantitative analysis that includes twins without GD. However, data concerning BMI, alcohol and smoking were incomplete for the majority of twins without GD. In addition, our present information on these parameters was based on a self-reported questionnaire and it is possible that data on weight and alcohol consumption were underestimated. The total number of twins that could be evaluated thus was reduced to a modest 1,666. Furthermore, the question of smoke-free tobacco was only addressed to C1, thus leading to a large number of missing values. Despite the large number of twins screened, this study size might not have had sufficient power to reveal possible statistical differences. Twins were assumed to be representative of the whole population, which might not be the case.

The impact of co-morbidity was not evaluated in this cohort. High BMI or cardiopulmonary disease might increase the number of twins hospitalised but might on the other hand lead to a more conservative treatment on an outpatient basis. As the laparoscopic era progressed, many twins may have been operated on an ambulatory basis and were thus not included in the Hospital Discharge Registry, which was completed only from 1987 on; and although patients were included retrospectively, data may have been unavailable especially for the older cohort.

Despite all these limitations, our study has a long follow-up time of up to 40 years and there is no recall bias as these factors were registered prior to the diagnosis of GD.

In conclusion, this study confirmed the positive association between BMI and GD as well as the inverse association of high alcohol consumption and GD, though at levels associated with increased risk for liver disease. The co-twin analysis indicates that shared environmental factors rather than genetic factors have confounded these associations. Smoking was not shown to have an impact on GD.

5.3   GD AND THE *ABCG8* D19H GENOTYPE

In the third study we investigated the impact of the ABCG8 D19H genotype on symptomatic GD. Recently, Buch et al. identified the hepatocanalicular cholesterol transporter ABCG8 as the first common susceptibility factor for gallstones in humans (75). Grünhage et al. (76) demonstrated significant single-point linkage for the ABCG8 D19H variant in contrast to other ABCG5/G8 SNPs (76). Our present study in twins provides further evidence for the importance of the ABCG8 D19H variant. As in our previous studies we have investigated affected twin pairs, most of which were MZ sharing 100% of their segregated genes.

Although small in number, our study takes advantage of the fact that the cases are twins, most of them MZ. From the Stockholm County screening, of 32 MZ pairs, 24 were concordant and 8 were discordant for GD indicating a high impact of genetic factors in these genetically identical twins. Disease concordance in MZ twins is presently the best means for establishing the strength of the inherited genetic determinates of complex disease (145).

The number of concordant and discordant MZ twins alive in Stockholm County was low and the number of twins finally participating in all three parts of the study (questionnaire, ultrasound and donating blood samples) was further limited by the high age of the cohort addressed in this study. Co-morbidity, death during the evaluation period, dementia and difficulty to travel limited the total number of participating twins to a mere 39% (73 out of 188 invited). We thus did not reach the power needed for an association study. There was also an unexpected high number of GD in our sample. Notwithstanding, in our sample population 20.8% of all unique MZ GD patients carried the heterozygous D19H genotype, which is in accordance with the results by Grünhage et al. of 21.4% and 8.6%, respectively (76), indicating that the populations are comparable. Of note, the D19H variant was also replicated as a susceptibility factor for GD in Chilean Hispanics (75) and Chinese (146) where D19H increases the gallstone risk in patients younger than 50 years of age. However, in Chinese, other ABCG5/8 variants (*ABCG5* 604Q and *ABCG8* T400K) were found to be associated with GD as well (146, 147).

Overall, association studies are capable of identifying substantial genetic effects (i.e., OR >2.0) on disease phenotype with relatively small sample sizes (N about 200) and have high power to detect small effects (OR <2.0) but then require large sample sizes (about 1000) (145). Because of the small sample size of our study we did not perform an analysis of known confounding factors such as BMI (148) or plasma cholesterol. Of note, despite their generally high age, no twin within the first part of the study with GD and *ABCG8* D19H had elevated serum total or LDL-cholesterol. This is at least not in contrast to previous studies (128, 130, 131) that have shown significantly lower levels of total and LDL-cholesterol in 19H carriers. This findings lead to the assumption that D19H increases the ABCG5/G8 mediated transfer of cholesterol into bile, which conversely results in biliary cholesterol hypersaturation, the major pathogenetic defect in GD (1, 4). Of note, in non-obese Chinese gallstone patients, ABCG5/G8 expression increased significantly and correlated with the percentage of biliary cholesterol and cholesterol saturation index (149). In Chinese patients, the D19H polymorphism was also associated with greater insulin resistance, which promotes the formation of cholesterol gallstones (150). However, in contrast to European studies, Chinese D19H carriers displayed higher total and LDL cholesterol levels, which might be due to "over-compensation" of low cholesterol absorption rates in subjects consuming a diet rich in fibres and plant sterols (151).

Of all *ABCG8* D19H carriers among MZ twins screened in Stockholm County, 75% were concordant for GD. Recruitment of additional MZ and DZ twins as well as controls from the newly assembled DNA bank within the nationwide TwinGene project was performed for further genetic investigation. Considering the whole cohort for the MZ and DZ cases with matched controls from the TwinGene project, we found that the *ABCG8* D19H genotype was more common in both MZ and DZ twins with GD than in MZ and DZ controls (18.2% and 22.6% vs. 9.2% and 9.5%, respectively). Overall, the D19H frequency was significantly (p=0.008) higher in cases (20.8%) as compared to controls (9.4%).

A crucial step in the genetic study of a complex trait is the ability to confirm linkage or association results in independent samples. Despite its limitations, our study in MZ twins confirms the ABCG8 D19H variant as being a risk factor for GD, further providing evidence for the lithogenic effect of this risk allele.

33

6   CONCLUSIONS

• In Study I we established the different variance components that contribute to symptomatic GD to be:

Heritability	**25% (95% CI 0.09-0.40)**
Shared environmental factors	**13% (95% CI 0.01-0.25)**
Unique environmental factors	**62% (95% CI 0.56-0.68)**

There were gender differences for disease threshold but no gender differences for heritability. Family clustering in GD seems to be the result of additive genetic factors, at least to some extent, though difficult to distinguish from shared environmental factors.

• In Study II we found positive associations between increasing BMI and GD in the cohort study.

Associations were lower in the within-pair analyses with no difference between MZ and DZ twins indicating confounding by shared environmental factors. There were protective effects of high alcohol consumption with no differences between MZ and DZ in the within-pair analyses though with wide CI indicating confounding by shared environmental factors. Smoking or smoke-free tobacco do not have a significant impact on GD.

• In Study III we found that D19H genotype of the cholesterol hemitransporter ABCG8 was significantly ($p=0.008$) more prevalent in gallstone cases (20.8%) as compared to controls (9.4%), providing further evidence for the lithogenic effect of this risk allele.

7   ACKNOWLEDGEMENTS

times outside of work, the yearly skiing vacations to the Alps, the Midsummer parties, the dinner parties during the long winter evenings and making life more enjoyable. Special thanks to Daniel for the statistical help in study II.

My cousins Anna and Savvas. For always welcoming me back to Greece, sharing their time and friends with me and making me feel comfortable at my hometown. To Savvas for help with my Excel files in study II.

My good friends Maria and Roufia. For always being there for me. For the good times during all my years in Sweden, their non-judgemental support through work and personal matters, the long conversations, the fun trips as well as their unconditional friendship.

My childhood friends Danis and Sotiris for, almost a lifetime of friendship. For the summer vacations, the winter trips, always making me feel I have a second home in Athens, London or the Hague. For bringing joy in my life and always supporting me.

My brother Yannis for his love and support during all of our years. For always taking care of me. His wife Lily and my niece Eva for being my family in Greece.

My uncle Vassilis for always being there for me, supporting my decisions and encouraging me.

My mother Evi and my late father Kyros. For your endless support, always believing in me and making my dreams come true.

My husband Erik for your patience and understanding, your help with the statistical databases but most of all for your love and for sharing your life with me and tolerating my temper.

To our daughter with love.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Gallstenssjukdom är vanligt förekommande i Sverige och utgör ett stort hälsoproblem. Ålder, kön, etniskt ursprung, övervikt, antal barn hos kvinnor mm är kända riskfaktorer. Det är sannolikt ett komplext samspel mellan miljöfaktorer och icke-identifierade gener som bidrar till bildning av kolesterolgallstenar som utgör ca 90% av alla gallstenar. Skillnaden i gallstensförekomst hos olika etniska grupper och ökad förekomst inom vissa familjer antyder att det finns genetiska (ärftliga) faktorer som styr benägenheten att bilda gallstenar.

Tvillingstudier ger viktig information om genetiska faktorer och kan belysa samspelet mellan gener och kön, livsstil, ålder och deras betydelse för sjukdomsuppkomst. Min forskningsplan inleddes med att identifiera tvillingpar i hela Sverige där en eller båda tvillingar har känd gallstenssjukdom (samkörning av Tvillingregistret och Slutenvårdsregistret samt Dödsorsaksregistret för gallstensoperation och andra gallstensdiagnoser). Vi har identifierat totalt 43.141 tvillingpar födda före 1958. Våra resultat visar att sannolikheten att båda tvillingar i ett par har gallstenssjukdon är högre för enäggs- jämfört med tvåäggstvillingar, vilket talar för ärftliga faktorer. För att kartlägga betydelsen av ärftlighet vs miljö har vi använt oss utav statistiska modeller och visat att gallstenssjukdom beror till 25% på ärftliga faktorer, till 13% på faktorer i den gemmensamma tvillingmiljön och till 62% på unika miljöfaktorer.

För att studera hur en enskild mijöfaktor påverkar risken för gallstenssjukdom har vi tittat närmare på de tvillingpar där endast en tvilling i paret har gallstenssjukdom. Vi har med hjälp av Tvillingregistret samlat in data på body mass index (ett mått på övervikt), alkohol och rökning. Enäggstvillingar har gemensamma gener och skillnaden i sjukdomsförekomst mellan de två tvillingar i paret är oberoende av ärftliga faktorer. På samma sätt så kan skillnader i sjukdomsförekomst mellan tvåäggstvillingar i ett par bero på både ärftliga och miljöfaktorer. Studien visar att risken för gallstenssjukdom ökar med stigande body mass index och att hög alkoholkonsumption skyddar mot gallstenssjukdom men att dessa associationer beror på gemensamma miljöfaktorer och inte på ärftliga faktorer. Rökning och snusning har ingen effekt på gallstenssjukdom.

Nyligen har olika sk kandidatgener dvs gener som kan påverka gallstensbildning karakteriserats hos inavlade möss. Dessa gener styr bildningen av vissa proteiner som i sin tur styr omsättning och transport av kolesterol och andra gallprodukter mellan cellerna och har betydelse för bildning av kolesterolgallstenar. Närmare kartläggning av dessa gener hos människor med och utan gallsten borde kunna ge mycket viktig information om hur gallstenar bildas och hur man vid ärftlig benägenhet till gallstenssjukdom kan undvika sjukdomen. Vi har i den tredje studien tittat närmare på ABCG8 genen. Vi har identifierat de enäggstvillingar i Storstockholm där båda har gallstenssjukdom och bjudit in dem för att samla in DNA. Där gallstensdiagnosen var oklar gjorde vi ett ultraljud på gallblåsan. Av 32 enäggstvillingpar var 24 par konkordanta, dvs båda tvillingar i paret hade sjukdomen. Fem av 24 av dessa par bärare av ABCG8 D19H genen som förekommer oftare hos sjuka än friska tvillingar (20.8% mot 8.6%). Vi har vidare studerat samma gen i tvillingar från hela Sverige och hittat att de tvillingar som är bärare av ABCG8 D19H genen löper en 2.56 gånger större risk att utveckla gallstenssjukdom.

Sammantaget kan vi i denna avhandling visa att ärftlighet bidrar till gallstenssjukdom. Andra riskfaktorer är övervikt och kvinnligt kön men hög alkoholkonsumption skyddar mot sjukdom. Det finns ett antal gener som påverkar ärftligheten för sjukdomen varav ABCG8 D19H är associerad till ökad sjukdomsförekomst. Det finns dock fortfarande ett antal gener vars betydelse för gallstenssjukdom fortfarande är outforskad.

9 REFERENCES

1.   Marschall HU, Einarsson C. Gallstone disease. J Intern Med 2007;261:529-542.
2.   Lammert F, Sauerbruch T. Mechanisms of disease: the genetic epidemiology of gallbladder stones. Nat Clin Pract Gastroenterol Hepatol 2005;2:423-433.
3.   Wittenburg H, Lammert F. Genetic predisposition to gallbladder stones. Semin Liver Dis 2007;27:109-121.
4.   Portincasa P, Moschetta A, Palasciano G. Cholesterol gallstone disease. Lancet 2006;368:230-239.
5.   Kratzer W, Mason RA, Kachele V. Prevalence of gallstones in sonographic surveys worldwide. J Clin Ultrasound 1999;27:1-7.
6.   Everhart JE, Khare M, Hill M, Maurer KR. Prevalence and ethnic differences in gallbladder disease in the United States. Gastroenterology 1999;117:632-639.
7.   Everhart JE, Yeh F, Lee ET, Hill MC, Fabsitz R, Howard BV, Welty TK. Prevalence of gallbladder disease in American Indian populations: findings from the Strong Heart Study. Hepatology 2002;35:1507-1512.
8.   Sampliner RE, Bennett PH, Comess LJ, Rose FA, Burch TA. Gallbladder disease in pima indians. Demonstration of high prevalence and early onset by cholecystography. N Engl J Med 1970;283:1358-1364.
9.   Muhrbeck O, Ahlberg J. Prevalence of gallstone disease in a Swedish population. Scand J Gastroenterol 1995;30:1125-1128.
10.   Muhrbeck O. Symptoms of gallstone disease in a Swedish population. Eur J Gastroenterol Hepatol 1995;7:1209-1214.
11.   Portincasa P, Moschetta A, van Erpecum KJ, Calamita G, Margari A, vanBerge-Henegouwen GP, Palasciano G. Pathways of cholesterol crystallization in model bile and native bile. Dig Liver Dis 2003;35:118-126.
12.   Wittenburg H, Lyons MA, Li R, Churchill GA, Carey MC, Paigen B. FXR and ABCG5/ABCG8 as determinants of cholesterol gallstone formation from quantitative trait locus mapping in mice. Gastroenterology 2003;125:868-881.
13.   Lu TT, Repa JJ, Mangelsdorf DJ. Orphan nuclear receptors as eLiXiRs and FiXeRs of sterol metabolism. J Biol Chem 2001;276:37735-37738.
14.   Zollner G, Marschall HU, Wagner M, Trauner M. Role of nuclear receptors in the adaptive response to bile acids and cholestasis: pathogenetic and therapeutic considerations. Mol Pharm 2006;3:231-251.
15.   Makishima M, Okamoto AY, Repa JJ, Tu H, Learned RM, Luk A, Hull MV, et al. Identification of a nuclear receptor for bile acids. Science 1999;284:1362-1365.
16.   Parks DJ, Blanchard SG, Bledsoe RK, Chandra G, Consler TG, Kliewer SA, Stimmel JB, et al. Bile acids: natural ligands for an orphan nuclear receptor. Science 1999;284:1365-1368.
17.   Wang H, Chen J, Hollister K, Sowers LC, Forman BM. Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. Mol Cell 1999;3:543-553.
18.   Lu TT, Makishima M, Repa JJ, Schoonjans K, Kerr TA, Auwerx J, Mangelsdorf DJ. Molecular basis for feedback regulation of bile acid synthesis by nuclear receptors. Mol Cell 2000;6:507-515.

19.   Huang L, Zhao A, Lew JL, Zhang T, Hrywna Y, Thompson JR, de Pedro N, et al. Farnesoid X receptor activates transcription of the phospholipid pump MDR3. J Biol Chem 2003;278:51085-51090.

20.   Ananthanarayanan M, Balasubramanian N, Makishima M, Mangelsdorf DJ, Suchy FJ. Human bile salt export pump promoter is transactivated by the farnesoid X receptor/bile acid receptor. J Biol Chem 2001;276:28857-28865.

21.   Repa JJ, Berge KE, Pomajzl C, Richardson JA, Hobbs H, Mangelsdorf DJ. Regulation of ATP-binding cassette sterol transporters ABCG5 and ABCG8 by the liver X receptors alpha and beta. J Biol Chem 2002;277:18793-18800.

22.   Wang J, Einarsson C, Murphy C, Parini P, Bjorkhem I, Gafvels M, Eggertsen G. Studies on LXR- and FXR-mediated effects on cholesterol homeostasis in normal and cholic acid-depleted mice. J Lipid Res 2006;47:421-430.

23.   van Erpecum KJ. Biliary lipids, water and cholesterol gallstones. Biol Cell 2005;97:815-822.

24.   Hyde SC, Emsley P, Hartshorn MJ, Mimmack MM, Gileadi U, Pearce SR, Gallagher MP, et al. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature 1990;346:362-365.

25.   Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui LC, Antonarakis SE, Kazazian HH, Jr. A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. Nature 1990;346:366-369.

26.   Paumgartner G, Sauerbruch T. Gallstones: pathogenesis. Lancet 1991;338:1117-1121.

27.   Carey MC. Pathogenesis of gallstones. Am J Surg 1993;165:410-419.

28.   Brink MA, Slors JF, Keulemans YC, Mok KS, De Waart DR, Carey MC, Groen AK, et al. Enterohepatic cycling of bilirubin: a putative mechanism for pigment gallstone formation in ileal Crohn's disease. Gastroenterology 1999;116:1420-1427.

29.   Pereira SP, Bain IM, Kumar D, Dowling RH. Bile composition in inflammatory bowel disease: ileal disease and colectomy, but not colitis, induce lithogenic bile. Aliment Pharmacol Ther 2003;17:923-933.

30.   Lammert F, Miquel JF. Gallstone disease: from genes to evidence-based therapy. J Hepatol 2008;48 Suppl 1:S124-135.

31.   Stolk MF, Van Erpecum KJ, Peeters TL, Samsom M, Smout AJ, Akkermans LM, Vanberge-Henegouwen GP. Interdigestive gallbladder emptying, antroduodenal motility, and motilin release patterns are altered in cholesterol gallstone patients. Dig Dis Sci 2001;46:1328-1334.

32.   Stolk MF, van Erpecum KJ, Renooij W, Portincasa P, van de Heijning BJ, vanBerge-Henegouwen GP. Gallbladder emptying in vivo, bile composition, and nucleation of cholesterol crystals in patients with cholesterol gallstones. Gastroenterology 1995;108:1882-1888.

33.   van Erpecum KJ, Venneman NG, Portincasa P, Vanberge-Henegouwen GP. Review article: agents affecting gall-bladder motility--role in treatment and prevention of gallstones. Aliment Pharmacol Ther 2000;14 Suppl 2:66-70.

34.   Colecchia A, Sandri L, Bacchi-Reggiani ML, Portincasa P, Palasciano G, Mazzella G, Roda E, et al. Is it possible to predict the clinical course of gallstone disease? Usefulness of gallbladder motility evaluation in a clinical setting. Am J Gastroenterol 2006;101:2576-2581.

35.   Sahlin S, Glauman H, Danielsson A, Einarsson K. Lysosomal enzyme activities in gallbladder mucosa of gallstone-free subjects and patients with gallstones. J Hepatol 1996;25:895-899.

36.   Lee SP, LaMont JT, Carey MC. Role of gallbladder mucus hypersecretion in the evolution of cholesterol gallstones. J Clin Invest 1981;67:1712-1723.

37.   Lee SP, Nicholls JF. Nature and composition of biliary sludge. Gastroenterology 1986;90:677-686.

38.   Ko CW, Schulte SJ, Lee SP. Biliary sludge is formed by modification of hepatic bile by the gallbladder mucosa. Clin Gastroenterol Hepatol 2005;3:672-678.

39.   Holzbach RT, Kibe A, Thiel E, Howell JH, Marsh M, Hermann RE. Biliary proteins. Unique inhibitors of cholesterol crystal nucleation in human gallbladder bile. J Clin Invest 1984;73:35-45.

40.   Busch N, Lammert F, Marschall HU, Matern S. A new subgroup of lectin-bound biliary proteins binds to cholesterol crystals, modifies crystal morphology, and inhibits cholesterol crystallization. J Clin Invest 1995;96:3009-3015.

41.   Sahlin S, Ahlberg J, Einarsson K, Henriksson R, Danielsson A. Quantitative ultrastructural studies of gall bladder epithelium in gall stone free subjects and patients with gall stones. Gut 1990;31:100-105.

42.   Maurer KJ, Ihrig MM, Rogers AB, Ng V, Bouchard G, Leonard MR, Carey MC, et al. Identification of cholelithogenic enterohepatic helicobacter species and their role in murine cholesterol gallstone formation. Gastroenterology 2005;128:1023-1033.

43.   Maurer KJ, Rogers AB, Ge Z, Wiese AJ, Carey MC, Fox JG. Helicobacter pylori and cholesterol gallstone formation in C57L/J mice: a prospective study. Am J Physiol Gastrointest Liver Physiol 2006;290:G175-182.

44.   Fox JG, Dewhirst FE, Shen Z, Feng Y, Taylor NS, Paster BJ, Ericson RL, et al. Hepatic Helicobacter species identified in bile and gallbladder tissue from Chileans with chronic cholecystitis. Gastroenterology 1998;114:755-763.

45.   Fox JG, Yan LL, Dewhirst FE, Paster BJ, Shames B, Murphy JC, Hayward A, et al. Helicobacter bilis sp. nov., a novel Helicobacter species isolated from bile, livers, and intestines of aged, inbred mice. J Clin Microbiol 1995;33:445-454.

46.   Nilsson I, Shabo I, Svanvik J, Monstein HJ. Multiple displacement amplification of isolated DNA from human gallstones: molecular identification of Helicobacter DNA by means of 16S rDNA-based pyrosequencing analysis. Helicobacter 2005;10:592-600.

47.   Chen W, Li D, Cannan RJ, Stubbs RS. Common presence of Helicobacter DNA in the gallbladder of patients with gallstone diseases and controls. Dig Liver Dis 2003;35:237-243.

48.   Fallone CA, Tran S, Semret M, Discepola F, Behr M, Barkun AN. Helicobacter DNA in bile: correlation with hepato-biliary diseases. Aliment Pharmacol Ther 2003;17:453-458.

49.   Nilsson HO, Pietroiusti A, Gabrielli M, Zocco MA, Gasbarrini G, Gasbarrini A. Helicobacter pylori and extragastric diseases--other Helicobacters. Helicobacter 2005;10 Suppl 1:54-65.

50.   Hofmann AF. Helicobacter and cholesterol gallstones: do findings in the mouse apply to man? Gastroenterology 2005;128:1126-1129.

51.   van der Linden W. Genetic factors in gallstone disease. Clinical Gastroenterology 1973;2:603-614.

52.   Duggirala R, Mitchell BD, Blangero J, Stern MP. Genetic determinants of variation in gallbladder disease in the Mexican-American population. Genet Epidemiol 1999;16:191-204.

53.   Nakeeb A, Comuzzie AG, Martin L, Sonnenberg GE, Swartz-Basile D, Kissebah AH, Pitt HA. Gallstones: genetics versus environment. Ann Surg 2002;235:842-849.

54.   Paigen B. CMC: Gallstones. In: King R.A. RJI, Motulsky A.G.,, ed. The genetic basis of common diseases. 2nd edition ed. New York: Oxford University Press, 2002; 298-335.

55.   Lammert F, Carey MC, Paigen B. Chromosomal organization of candidate genes involved in cholesterol gallstone formation: a murine gallstone map. Gastroenterology 2001;120:221-238.

56.   Rosmorduc O, Hermelin B, Boelle PY, Parc R, Taboury J, Poupon R. ABCB4 gene mutation-associated cholelithiasis in adults. Gastroenterology 2003;125:452-459.

57.   Lammert F, Wang DQ, Hillebrandt S, Geier A, Fickert P, Trauner M, Matern S, et al. Spontaneous cholecysto- and hepatolithiasis in Mdr2-/- mice: a model for low phospholipid-associated cholelithiasis. Hepatology 2004;39:117-128.

58.   Fracchia M, Pellegrino S, Secreto P, Gallo L, Masoero G, Pera A, Galatola G. Biliary lipid composition in cholesterol microlithiasis. Gut 2001;48:702-706.

59.   van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, Bull LN, van den Berg IE, et al. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 2004;127:379-384.

60.   Glantz A, Marschall HU, Mattsson LA. Intrahepatic cholestasis of pregnancy: Relationships between bile acid levels and fetal complication rates. Hepatology 2004;40:467-474.

61.   Pullinger CR, Eng C, Salen G, Shefer S, Batta AK, Erickson SK, Verhagen A, et al. Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J Clin Invest 2002;110:109-117.

62.   Jiang ZY, Han TQ, Suo GJ, Feng DX, Chen S, Cai XX, Jiang ZH, et al. Polymorphisms at cholesterol 7alpha-hydroxylase, apolipoproteins B and E and low density lipoprotein receptor genes in patients with gallbladder stone disease. World J Gastroenterol 2004;10:1508-1512.

63.   Paigen B, Schork NJ, Svenson KL, Cheah YC, Mu JL, Lammert F, Wang DQ, et al. Quantitative trait loci mapping for cholesterol gallstones in AKR/J and C57L/J strains of mice. Physiol Genomics 2000;4:59-65.

64.   Wittenburg H, Lyons MA, Paigen B, Carey MC. Mapping cholesterol gallstone susceptibility (Lith) genes in inbred mice. Dig Liver Dis 2003;35 Suppl 3:S2-7.

65.   Yu L, Hammer RE, Li-Hawkins J, Von Bergmann K, Lutjohann D, Cohen JC, Hobbs HH. Disruption of Abcg5 and Abcg8 in mice reveals their crucial role in biliary cholesterol secretion. Proc Natl Acad Sci U S A 2002;99:16237-16242.

66.   Wang DQ, Lammert F, Paigen B, Carey MC. Phenotypic characterization of lith genes that determine susceptibility to cholesterol cholelithiasis in inbred mice. Pathophysiology Of biliary lipid secretion. J Lipid Res 1999;40:2066-2079.

67.   Wang DQ, Paigen B, Carey MC. Phenotypic characterization of Lith genes that determine susceptibility to cholesterol cholelithiasis in inbred mice: physical-chemistry of gallbladder bile. J Lipid Res 1997;38:1395-1411.

68.   Lammert F, Wang DQ, Paigen B, Carey MC. Phenotypic characterization of Lith genes that determine susceptibility to cholesterol cholelithiasis in inbred mice: integrated activities of hepatic lipid regulatory enzymes. J Lipid Res 1999;40:2080-2090.

69.   van Erpecum KJ, Wang DQ, Lammert F, Paigen B, Groen AK, Carey MC. Phenotypic characterization of Lith genes that determine susceptibility to cholesterol cholelithiasis in inbred mice: soluble pronucleating proteins in gallbladder and hepatic biles. J Hepatol 2001;35:444-451.

70.   Lammert F, Wang DQ, Wittenburg H, Bouchard G, Hillebrandt S, Taenzler B, Carey MC, et al. Lith genes control mucin accumulation, cholesterol crystallization, and gallstone formation in A/J and AKR/J inbred mice. Hepatology 2002;36:1145-1154.

71.   Lyons MA, Wittenburg H, Li R, Walsh KA, Leonard MR, Korstanje R, Churchill GA, et al. Lith6: a new QTL for cholesterol gallstones from an intercross of CAST/Ei and DBA/2J inbred mouse strains. J Lipid Res 2003;44:1763-1771.

72.   Lyons MA, Wittenburg H. Cholesterol Gallstone Susceptibility Loci: A Mouse Map, Candidate Gene Evaluation, and Guide to Human LITH Genes. Gastroenterology 2006.

73.   Lyons MA, Wittenburg H. Susceptibility to cholesterol gallstone formation: Evidence that LITH genes also encode immune-related factors. Biochim Biophys Acta 2006.

74.   Maurer KJ, Carey MC, Fox JG. Roles of infection, inflammation, and the immune system in cholesterol gallstone formation. Gastroenterology 2009;136:425-440.

75.   Buch S, Schafmayer C, Volzke H, Becker C, Franke A, von Eller-Eberstein H, Kluck C, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet 2007;39:995-999.

76.   Grunhage F, Acalovschi M, Tirziu S, Walier M, Wienker TF, Ciocan A, Mosteanu O, et al. Increased gallstone risk in humans conferred by common variant of hepatic ATP-binding cassette transporter for cholesterol. Hepatology 2007;46:793-801.

77.   Diehl AK. Epidemiology and natural history of gallstone disease. Gastroenterol Clin North Am 1991;20:1-19.

78.   Leopold GR, Amberg J, Gosink BB, Mittelstaedt C. Gray scale ultrasonic cholecystography: a comparison with conventional radiographic techniques. Radiology 1976;121:445-448.

79.   Portincasa P, Moschetta A, Petruzzelli M, Palasciano G, Di Ciaula A, Pezzolla A. Gallstone disease: Symptoms and diagnosis of gallbladder stones. Best Pract Res Clin Gastroenterol 2006;20:1017-1029.

80.   Keus F, de Jong JA, Gooszen HG, van Laarhoven CJ. Small-incision versus open cholecystectomy for patients with symptomatic cholecystolithiasis. Cochrane Database Syst Rev 2006:CD004788.

81.   Attili AF, De Santis A, Attili F, Roda E, Festi D, Carulli N. Prevalence of gallstone disease in first-degree relatives of patients with cholelithiasis. World J Gastroenterol 2005;11:6508-6511.

82.   Venneman NG, Besselink MG, Keulemans YC, Vanberge-Henegouwen GP, Boermeester MA, Broeders IA, Go PM, et al. Ursodeoxycholic acid exerts no beneficial effect in patients with symptomatic gallstones awaiting cholecystectomy. Hepatology 2006;43:1276-1283.

83.   Ransohoff DF, Gracie WA, Wolfenson LB, Neuhauser D. Prophylactic cholecystectomy or expectant management for silent gallstones. A decision analysis to assess survival. Ann Intern Med 1983;99:199-204.

84.   Gurusamy KS, Samraj K. Early versus delayed laparoscopic cholecystectomy for acute cholecystitis. Cochrane Database Syst Rev 2006:CD005440.

85.   Tewari M. Contribution of silent gallstones in gallbladder cancer. J Surg Oncol 2006;93:629-632.

86.   Moschetta A, Bookout AL, Mangelsdorf DJ. Prevention of cholesterol gallstone disease by FXR agonists in a mouse model. Nat Med 2004;10:1352-1358.

87.   Boomsma DI. Twin registers in Europe: an overview. Twin Res 1998;1:34-51.

88.   Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet 2002;3:872-882.
89.   Martin N, Boomsma D, Machin G. A twin-pronged attack on complex traits. Nat Genet 1997;17:387-392.
90.   Boomsma DI. Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. Behav Genet 1996;26:161-166.
91.   Boomsma DI, Molenaar PC. The genetic analysis of repeated measures. I. Simplex models. Behav Genet 1987;17:111-123.
92.   Burke W. Genetic testing. N Engl J Med 2002;347:1867-1875.
93.   Hall JG. Twinning. Lancet 2003;362:735-743.
94.   Lichtenstein P, De Faire U, Floderus B, Svartengren M, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. J Intern Med 2002;252:184-205.
95.   Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78-85.
96.   Hauser ER, Pericak-Vance MA. Genetic analysis for common complex disease. Am Heart J 2000;140:S36-44.
97.   Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241-247.
98.   Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994;265:2037-2048.
99.   Vogel F. [Significance of human genetics for a theory of disease]. Verh Dtsch Ges Pathol 1982;66:1-15.
100.   Monaco AP, Bertelson CJ, Middlesworth W, Colletti CA, Aldridge J, Fischbeck KH, Bartlett R, et al. Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. Nature 1985;316:842-845.
101.   Cordell HJ, Clayton DG. Genetic association studies. Lancet 2005;366:1121-1131.
102.   Risch N. Implications of multilocus inheritance for gene-disease association studies. Theor Popul Biol 2001;60:215-220.
103.   Martin ER, Gilbert JR, Lai EH, Riley J, Rogala AR, Slotterbeck BD, Sipe CA, et al. Analysis of association at single nucleotide polymorphisms in the APOE region. Genomics 2000;63:7-12.
104.   Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet 1955;7:277-318.
105.   Hoblinger A, Lammert F. Genetics of biliary tract diseases: new insights into gallstone disease and biliary tract cancers. Curr Opin Gastroenterol 2008;24:363-371.
106.   Lichtenstein P, Sullivan PF, Cnattingius S, Gatz M, Johansson S, Carlstrom E, Bjork C, et al. The Swedish Twin Registry in the third millennium: an update. Twin Res Hum Genet 2006;9:875-882.
107.   van der Linden W, Nakayama F. Gallstone disease in Sweden versus Japan. Clinical and etiologic aspects. Am J Surg 1973;125:267-272.
108.   van der Linden W, Westlin N. The familial occurrence of gallstone disease. II. Occurrence in husbands and wives. Acta Genet Stat Med 1966;16:377-382.
109.   Gilat T, Feldman C, Halpern Z, Dan M, Bar-Meir S. An increased familial frequency of gallstones. Gastroenterology 1983;84:242-246.

110.   Sarin SK, Negi VS, Dewan R, Sasan S, Saraya A. High familial prevalence of gallstones in the first-degree relatives of gallstone patients. Hepatology 1995;22:138-141.

111.   Kesaniemi YA, Koskenvuo M, Vuoristo M, Miettinen TA. Biliary lipid composition in monozygotic and dizygotic pairs of twins. Gut 1989;30:1750-1756.

112.   Duis BT. [Catamneses of twins; clinical findings in follow-up examination of a free-choice selection of a series of twins.]. Acta Gerontol (Milano) 1956;5:14-103.

113.   Doig RK. Illness in twins. III. Cholelithiasis. Med J Aust 1957;44:716-717.

114.   Koch G. [Results of a reexamination of the Berlin twin series after 20-25 years; preliminary results.]. Acta Genet Stat Med 1957;7:47-52.

115.   Koch G. Results of the 40-year, longitudinal study of a series of twins of Berlin (450 monozygotic and dizygotic twin pairs). Prog Clin Biol Res 1981;69 Pt C:201-209.

116.   Harvald B, Hauge M. A catamnestic investigation of Danish twins; a preliminary report. Dan Med Bull 1956;3:150-158.

117.   Katsika D, Grjibovski A, Einarsson C, Lammert F, Lichtenstein P, Marschall HU. Genetic and environmental influences on symptomatic gallstone disease: a Swedish study of 43,141 twin pairs. Hepatology 2005;41:1138-1143.

118.   Neale M, Cardon, LR. Methodology for Genetic Studies of Twins and Families. Dordrecht, The Netherlands: Kluwer Academic Publishers,. 1992.

119.   Neale M. Mx: statistical modelling, 5th ed. Richmond, VA: Virginia Commonwealth University. 1999.

120.   Kuritz SJ, Landis JR, Koch GG. A general overview of Mantel-Haenszel methods: applications and recent developments. Annu Rev Public Health 1988;9:123-160.

121.   Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. Bmj 2000;320:1240-1243.

122.   Becker U, Deis A, Sorensen TI, Gronbaek M, Borch-Johnsen K, Muller CF, Schnohr P, et al. Prediction of risk of liver disease by alcohol intake, sex, and age: a prospective population study. Hepatology 1996;23:1025-1029.

123.   Bellentani S, Saccoccio G, Costa G, Tiribelli C, Manenti F, Sodde M, Saveria Croce L, et al. Drinking habits as cofactors of risk for alcohol induced liver damage. The Dionysos Study Group. Gut 1997;41:845-850.

124.   Diggle P, Liang K, Zeger S. Analysis of Longitudinal Data. Oxford, UK: Oxford University Press. 1994.

125.   Freund R, Littell R. SAS System for Regression, 2nd edn Cary, NC: SAS Institute Inc. 1991.

126.   Stokes M, Davis C, Koch G. Categorical Data Analysis Using the SAS System. Cary, NC: SAS Institute Inc. 1995.

127.   Gylling H, Hallikainen M, Pihlajamaki J, Agren J, Laakso M, Rajaratnam RA, Rauramaa R, et al. Polymorphisms in the ABCG5 and ABCG8 genes associate with cholesterol absorption and insulin sensitivity. J Lipid Res 2004;45:1660-1665.

128.   Kajinami K, Brousseau ME, Nartsupha C, Ordovas JM, Schaefer EJ. ATP binding cassette transporter G5 and G8 genotypes and plasma lipoprotein levels before and after treatment with atorvastatin. J Lipid Res 2004;45:653-656.

129.   Acalovschi M, Ciocan A, Mostean O, Tirziu S, Chiorean E, Keppeler H, Schirin-Sokhan R, et al. Are plasma lipid levels related to ABCG5/ABCG8 polymorphisms? A preliminary study in siblings with gallstones. Eur J Intern Med 2006;17:490-494.

130. Hubacek JA, Berge KE, Cohen JC, Hobbs HH. Mutations in ATP-cassette binding proteins G5 (ABCG5) and G8 (ABCG8) causing sitosterolemia. Hum Mutat 2001;18:359-360.

131. Hubacek JA, Berge KE, Stefkova J, Pitha J, Skodova Z, Lanska V, Poledne R. Polymorphisms in ABCG5 and ABCG8 transporters and plasma cholesterol levels. Physiol Res 2004;53:395-401.

132. Lu K, Lee MH, Hazard S, Brooks-Wilson A, Hidaka H, Kojima H, Ose L, et al. Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively. Am J Hum Genet 2001;69:278-290.

133. Kono S, Eguchi H, Honjo S, Todoroki I, Oda T, Shinchi K, Ogawa S, et al. Cigarette smoking, alcohol use, and gallstone risk in Japanese men. Digestion 2002;65:177-183.

134. Kono S, Shinchi K, Ikeda N, Yanai F, Imanishi K. Prevalence of gallstone disease in relation to smoking, alcohol use, obesity, and glucose tolerance: a study of self-defense officials in Japan. Am J Epidemiol 1992;136:787-794.

135. La Vecchia C, Decarli A, Ferraroni M, Negri E. Alcohol drinking and prevalence of self-reported gallstone disease in the 1983 Italian National Health Survey. Epidemiology 1994;5:533-536.

136. Leitzmann MF, Giovannucci EL, Stampfer MJ, Spiegelman D, Colditz GA, Willett WC, Rimm EB. Prospective study of alcohol consumption patterns in relation to symptomatic gallstone disease in men. Alcohol Clin Exp Res 1999;23:835-841.

137. Leitzmann MF, Willett WC, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E. A prospective study of coffee consumption and the risk of symptomatic gallstone disease in men. Jama 1999;281:2106-2112.

138. Leitzmann MF, Tsai CJ, Stampfer MJ, Rimm EB, Colditz GA, Willett WC, Giovannucci EL. Alcohol consumption in relation to risk of cholecystectomy in women. Am J Clin Nutr 2003;78:339-347.

139. Kratzer W, Kachele V, Mason RA, Muche R, Hay B, Wiesneth M, Hill V, et al. Gallstone prevalence in relation to smoking, alcohol, coffee consumption, and nutrition. The Ulm Gallstone Study. Scand J Gastroenterol 1997;32:953-958.

140. Okamoto M, Yamagata Z, Takeda Y, Yoda Y, Kobayashi K, Fujino MA. The relationship between gallbladder disease and smoking and drinking habits in middle-aged Japanese. J Gastroenterol 2002;37:455-462.

141. Jorgensen T. Gall stones in a Danish population. Relation to weight, physical activity, smoking, coffee consumption, and diabetes mellitus. Gut 1989;30:528-534.

142. Jorgensen T, Jorgensen LM. Gallstones and diet in a Danish population. Scand J Gastroenterol 1989;24:821-826.

143. Rhodes M, Venables CW. Symptomatic gallstones--a disease of non-smokers? Digestion 1991;49:221-226.

144. McMichael AJ, Baghurst PA, Scragg RK. A case-control study of smoking and gallbladder disease: importance of examining time relations. Epidemiology 1992;3:519-522.

145. Juran BD, Lazaridis KN. Applying genomics to the study of complex disease. Semin Liver Dis 2007;27:3-12.

146. Kuo KK, Shin SJ, Chen ZC, Yang YH, Yang JF, Hsiao PJ. Significant association of ABCG5 604Q and ABCG8 D19H polymorphisms with gallstone disease. Br J Surg 2008;95:1005-1011.

147.	Wang Y, Jiang ZY, Fei J, Xin L, Cai Q, Jiang ZH, Zhu ZG, et al. ATP binding cassette G8 T400K polymorphism may affect the risk of gallstone disease among Chinese males. Clin Chim Acta 2007;384:80-85.

148.	Katsika D, Tuvblad C, Einarsson C, Lichtenstein P, Marschall HU. Body mass index, alcohol, tobacco and symptomatic gallstone disease: a Swedish twin study. J Intern Med 2007;262:581-587.

149.	Jiang ZY, Parini P, Eggertsen G, Davis MA, Hu H, Suo GJ, Zhang SD, et al. Increased expression of LXR alpha, ABCG5, ABCG8, and SR-BI in the liver from normolipidemic, nonobese Chinese gallstone patients. J Lipid Res 2008;49:464-472.

150.	Biddinger SB, Haas JT, Yu BB, Bezy O, Jing E, Zhang W, Unterman TG, et al. Hepatic insulin resistance directly promotes formation of cholesterol gallstones. Nat Med 2008;14:778-782.

151.	Chen ZC, Shin SJ, Kuo KK, Lin KD, Yu ML, Hsiao PJ. Significant association of ABCG8:D19H gene polymorphism with hypercholesterolemia and insulin resistance. J Hum Genet 2008;53:757-763.

APPENDIX

Study Integration Plan

Karolinska Institutet

GALLSTONE – KI Biobank Integration

Study Integration plan (SIP)

Author: Linda Lundström
Version: 1
Date: 2006-12-11

Document History

Version History

Date of this version: 2006-12-11

Version Number	Version Date	Summary of Changes	Author
1	2006-12-11	Initial draft	LL

Table of Contents

1. Background

1.1 Study Background

Cholesterol gallstone disease is one of the most common and health economically important gastrointestinal diseases. The disease represents a failure of biliary cholesterol homeostasis in which the physical-chemical balance of cholesterol solubility in bile is disturbed. The primary pathophysiologic defect is cholesterol supersaturation of gallbladder bile. The underlying defects are augmented intestinal cholesterol absorption, cholesterol synthesis, lipoprotein delivery, and hepatic cholesterol up-take, and disorders that uncouple phospholipid and/or cholesterol synthesis. The molecular pathogenesis as well as the genetic susceptibility for gallstones is still obscured.

Recent progress in molecular biology and genetics indicated that the susceptibility for gallstones are based on improper function of proteins that regulate lipid synthesis and translocation. Recent studies in inbred mice revealed candidate genes for gallstone disease. The genes identified (ABCB4, ABCB11, ABCG5/8, LRP2, APOE, CYP7A) are coding for proteins that are responsible for the synthesis and regulation of compounds involved in the metabolism and cellular transport of cholesterol and other biliary compounds.

By linking The Swedish Twin Registry with the Swedish inpatient-discharge and causes of death registries for symptomatic gallstone disease (GD) and gallstone surgery-related diagnoses in 43,141 twin pairs born between 1900 and 1958 we recently defined the contribution of genetic (25%) and shared (13%) or unique (62%) environmental factors for the development of symptomatic GD.

In order to study whether the candidate genes identified in inbred mice are also contributing to GD in humans, we aim to test for human candidate genes for gallstone disease by selected sampling from The Swedish Twin Registry for quantitative trait loci linkage and association studies. The Swedish Twin Registry is screened for a patient's history confirming gallbladder disease, e.g. by surgery, cholecystography or ultrasound. Based on previous Swedish data about the prevalence in gallstones, 42 concordant monozygotic and 146 discordant monozygotic twins will be investigated for environmental factors by the use of questionnaire and screened for gallstone disease by the use of ultrasonography. The twin brother or sister of an affected twin with unknown gallstone disease is invited for an ultrasound scan of the gallbladder. A biochemical analysis of blood lipids including lipoproteins and indicators of cholestasis and fatty liver disease and inflammation is included. Blood will be analysed for lipids and for polymorphisms of gallstone candidate genes (ABCB4, ABCB11, ABCG5/8, LRP2, APOE, CYP7A) and ABCG5/8 related to gallstone haplotypes.

The study is aimed to further identify persons at risk, to better understand the pathogenesis, and ultimately to develop novel strategies for prevention and treatment of GD.

Blood samples will be collected from 188 individuals and stored at KI Biobank. The collection is plan to start in February 2007 and will continue for two months. Six aliquots of plasma samples will be extracted from the blood samples.

1.2 Project Objectives

The Gallstone – KI Biobank Integration study is defined to achieve the following **primary objectives**:

1. Define information components that need to be transferred between the study organisation Gallstone and KI Biobank

2. Setup the IT communication infrastructure for transferring these information components between the organizations

2. Project logistics and scope

2.1 Concept Definitions

Several key concepts are used throughout this document, some of which are described in the table below.

Concept	Description
PNr	The Study donors personal identification number.
SampleID	Unique identifying number for every sample within the Gallstone study.
CDK	Customer Donor Key. The unique identifier of a study subject within the Gallstone study.
SDID	The referrals unique identifier of a study subject within the Gallstone study.
Consent desicion	The Study subjects have to give their consent to that their samples are stored in KI Biobank.
LIMS	Laboratory Information Management System. Stores information about the samples.
SDM	Study Donor Management. Stores information about the study subjects.

2.2 Project Logistics

This section will describe the major logistic steps in the study.

Involved organization	Responsibilities in Gallstone
Study Subject	Participates in the study.
KI Biobank	Responsible for processing and storing the biological samples as well as providing the fundamental IT logistics between the different parties involved in the study.
Gallstone	Responsible for contacting the Study Subjects as well as collecting consent decision and biological samples from them.

2.2.1 Logistics for Study Subject

The logistics for the Study Subject is described in figure 1.

Study Subject flowchart

	Karolinska Universitetssjukhuset or Health Centre	Study subject	Gallstone	Karolinska Universitetslaboratoriet	KI Biobank

1. A file is sent containing pnr, name, adress, CDK and consent decision

2. Add Study subject in SDM. Add samples in LIMS

3. Generate information for referrals

4. Print referrals containing pnr, name, adress, SDID, CDK and barcodes for the samples. Pack the referrals together with sample tubes and send to Study Subject

5. Receiving envelope with sample tubes and referrals.

6. Bring the envelope with sample tubes and referrals to the local health centre or to Karolinska Universitets- sjukhuset

7. Samples are collected and sent to KI Biobank.

8. Centrifugation of plasma samples

9. Receiving samples for storage and DNA extraction
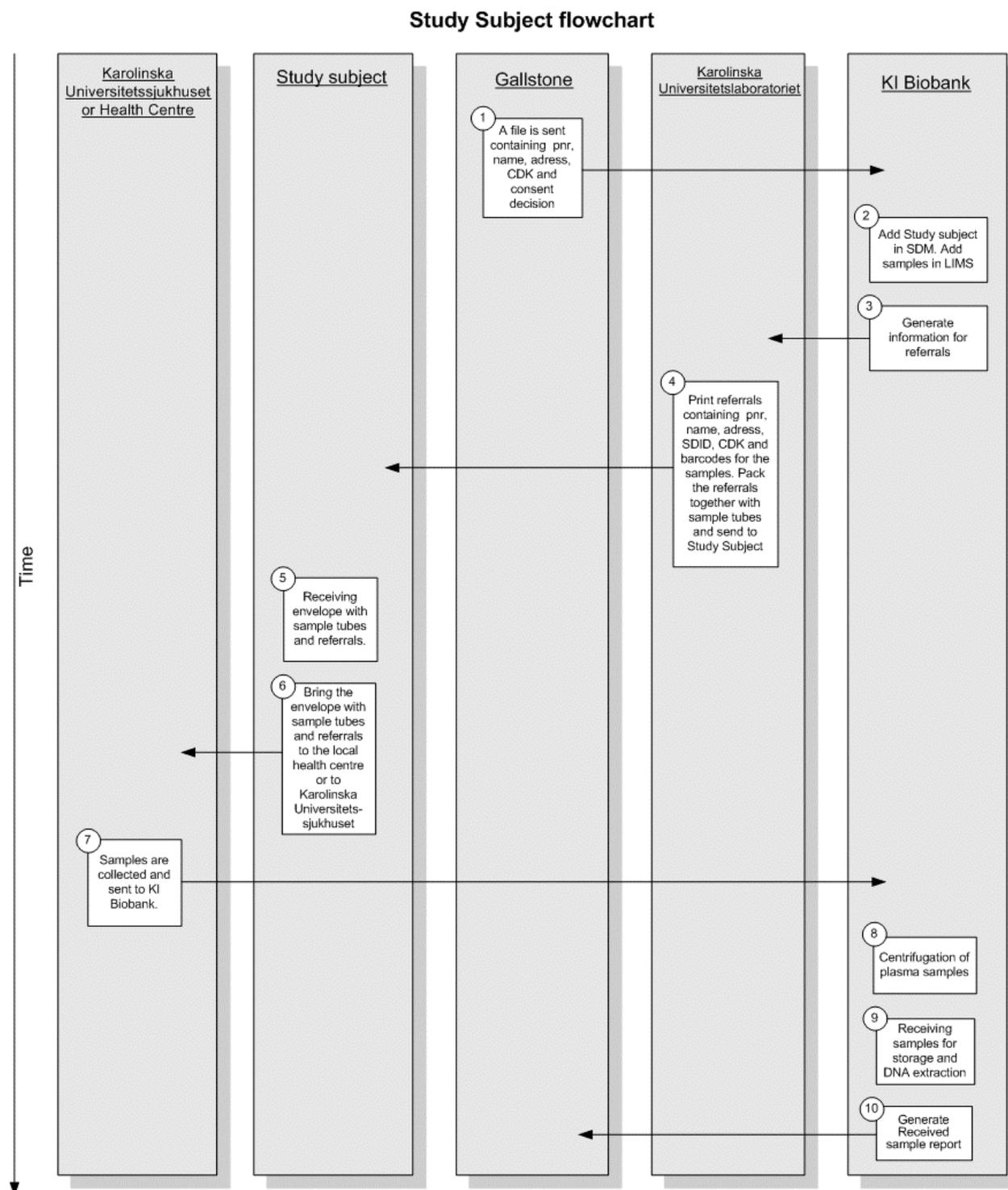
10. Generate Received sample report

Figure 1

A more detailed description of the stages described in the figure 1 is given here, together with information about scope.

Step	Description	Responsible party	In scope for this study
1	A file is sent containing pnr, name, address, CDK and consent decision	Gallstone	Yes
2	Add Study subject in SDM. Add samples in LIMS	KI Biobank	Yes
3	Generate information for referrals	Karolinska Universitetslaboratoriet	Yes
4	Print referrals containing pnr, name, adress, SDID, CDK and barcodes for the samples. Pack the referrals together with sample tubes and letters and send to Study Subject	KI Biobank	Yes
5	Receiving envelope with sample tubes and referrals.	Study subject	Yes
6	Bring the envelope with sample tubes and referrals to the local health centre or to Karolinska Universitets-sjukhuset	Study subject	Yes
7	Two blood samples are collected and sent to KI Biobank.	Karolinska Universitetssjukhuset	Yes
8	Centrifugation and aliquoting of plasma samples from the two blood samples.	KI Biobank	Yes
9	Receiving samples for storage and DNA extraction	KI Biobank	
10	Generate Received sample report	KI Biobank	

3. ID generation and Information transfer format

3.1 ID generation

One of the fundamental issues of the project is to handle various entity identities. Biological sample barcode identities and study subject identifiers must be properly designed and documented.

3.1.1 Sample barcodes

Each sample will be labelled with a unique identifier in the form of a sample barcode. The barcode contents will be generated by KI Biobank and will only be a simple sequence integer number, containing no other logic.

3.1.2 Study subject identifier

CDK is a number, which identifies a study subject within the Gallstone study.

3.1.3 Referrals

The referrals will be printed by Karolinska Universitetslaboratoriet and will contain personal identification number, SDID, CDK, name, address, two barcodes for blood sample and one extra barcode for each sample.

The referral will be sent to the Study subject.

3.2 Information transfer formats

This section defines the information transfer formats used for integrating the organizations involved in the electronic information flow in the Gallstone study. This includes clear definitions of the information that is mandatory and / or possible as well as the formats used for transferring this information.

3.2.1 Information transfer from Gallstone to KI Biobank

This information transfer corresponds to step 1 in section 2.2.1. The table below describes the Study Subject information variables of interest to transfer from Gallstone to KI Biobank whether it is mandatory or not. The information will be registered in an XSL-file.

Variable	Description	Example	Data format	Mandatory
pnr	Study subject personal identification number	ååååmmddxxxx	12-digit integer	Yes
name	Study subject name	Svea Sveasson	String	Yes
address	Study subject address	Sveavägen 2	String	Yes
postal code	Study subject postal code	171 25	String	Yes
city	Study Subject city	Stockholm	String	Yes
CDK	The studies unique identifier of a study subject within the Gallstone study.	1	String	Yes
Consent1	The answer of consent question number one. " Jag ger mitt godkännande till att de blodprover jag lämnar placeras i KI Biobank, att proverna används för forskning och att mina personuppgifter registreras enligt den information jag tagit del av"	1	1- digit Integer	Yes
Consent2	The answer of consent question number two "Jag ger mitt tillstånd till inhämtande av **journalkopior** eller att information baserad på min journal lämnas ut"	0	1- digit Integer	Yes
Consent3	The answer of consent question number three. "Jag ger mitt tillstånd till att proverna i KI Biobank används i **framtida forskning** som inte är beskriven här men som i förekommande fall kommer att granskas och godkännas av en etikprövningsnämnd"	1	1- digit Integer	Yes

3.2.1.1 Transfer format

Information about the study subject will be registered by Gallstone in a XSL-file and sent to KI Biobank.
An example of the file is shown below:

pnr	name	address	postal code	city	cdk	consent 1	consent2	consent3
197601011212	Svea Sveasson	Sveavägen 2	171 25	Stockholm	1	1	1	0
197602290412	Hans Svenssson	Brovägen 21A	371 42	Karlskrona	2	1	1	1

3.2.2 Referral information transfer from KI Biobank to Karolinska Universitetslaboratoriet

This information transfer corresponds to step 3 in section 2.2.1. The table below describes the study subject information variables of interest to transfer from KI Biobank to Karolinska Universitetslaboratoriet in order to print referrals.

Variable	File header	Example	Data format	Mandatory
Study subject personnummer	paco	197602290412	12-digit Integer	Yes
Study subject name	pana	Hans Svenssson	String	Yes
Study subject street address	paad1	Brovägen 21A	String	Yes
Study subject postal address and city	paad2	371 42 Karlskrona	String	Yes
KI Biobank-unique study subject identifier	sdid	0000008359	10-digit integer	Yes
The studies unique identifier of a study subject within the Gallstone study.	cdk	1	String	Yes
Blood sample barcode	edta1	00023901	8-digit integer	Yes
Blood sample barcode	edta2	00023902	8-digit integer	Yes

3.2.2.1 Transfer format

CSV files will be used for transferring the information from KI Biobank to Karolinska Universitetslaboratoriet . The files will be named with a timestamp name, such as 20060223105140.txt (meaning that the file was created on 2006-02-23 10:51:40). An example CSV file is shown below:

```
paco;pana;paad1;paad2;sdid;cdk;edta1;edta2;
192303137845;Liv Larsson;Tegelhöjden 19;14142 Huddinge;000008359;1;00023901;00023902
```

3.2.2 Received Sample information transfer from KI Biobank to Gallstone

This information transfer corresponds to step 10 in section 2.2.1. The table below describes the study subject information variables of interest to transfer from KI Biobank to Gallstone in order to notify the Gallstone study team of the samples that have been received.

Variable	Description	Example	Data format	Mandatory
Message sender	Who sent the message	KI Biobank	Free text string	Yes
Message receiver	Who received the message	Gallstone	Free text string	Yes
Message timestamp	When the message was sent	2006-11-21 8:04	Timestamp	Yes
Study subject CDK	The unique identifier of a study subject within the Gallstone study.	0045667	String	Yes
Sample type	The kind of biological sample	Blood	String	Yes
Number of samples	The number of samples of a specific type for a specific study subject		Integer	Yes
Received	The date when the samples were received at KI Biobank.	2006-11-12	String	Yes

3.2.2.1 Transfer format

A tab separated excel file will be used for transferring this information. An example of the file is shown below:

```
The following report has been automatically generated by the KI Biobank LIMS system.
For any questions or comments, please contact sdmbiobank@ki.se

Message sender       KI Biobank
Message receiver     GALLSTONE
Message timestamp    2006-11-21 8:04


CDK   Blood    Received
1     2        2006-11-12
2     2        2006-11-13
3     2        2006-11-15
4     2        2006-11-15
```

3.3   Schedule for information transfer

This section includes clear definitions of time scheduling for transferring of files, referrals and reports.

File with information about Study subject, see step 1 in section 2.2.1, will be sent each Monday and Wednesday information for referrals will be generated and sent to Karolinska Universitetslaboratoriet. The referrals will be printed within a day and sent to the Study Subject.

The Received sample report, see step 10 in section 2.2.1, will be sent regular every Friday evening.

3.4   Physical sample processing and storage

The general processing of the samples are given below:
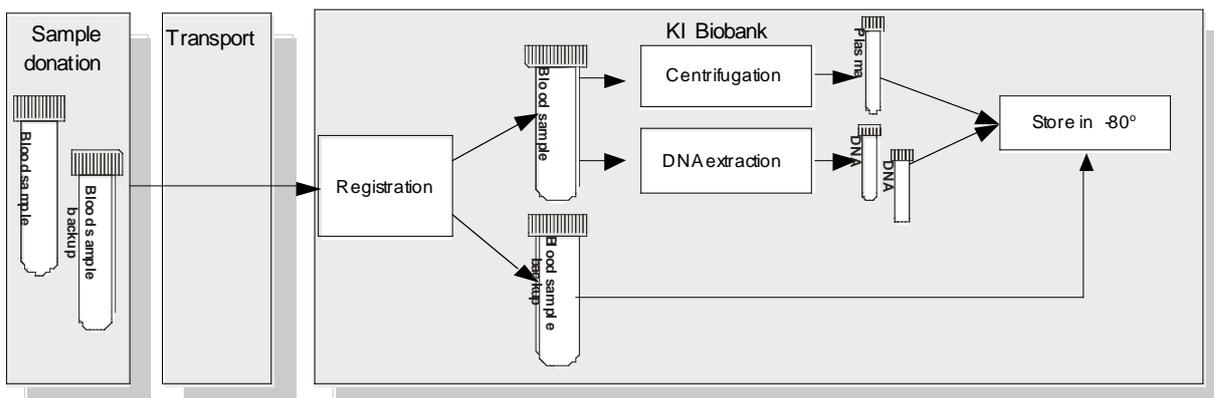
The stages in the sample processing at KI Biobank are explained in the table below.

Step	Description	Responsible party	In scope for this project
Sample Donation	Donation of two blood samples	Karolinska sjukhuset Solna or Healt Centre	No
Transport	Each study subjects samples and referral will be packed in one envelope and sent to KI Biobank	Karolinska sjukhuset Solna or Healt Centre	No
Registration	Registration of samples	KI Biobank	Yes
Centrifugation	Standardcentrifugation of the first blood sample which will be transferred into a new tube and later aliquoted into six ml tubes.	KI Biobank	Yes
DNA extraction	Robotized DNA extraction is performed on the remaining blood in the tube.	KI Biobank	Yes
Backup	The second blood sample is stored for backup.	KI Biobank	Yes
Storage	The samples are stored at -80°C.	KI Biobank	Yes

4. Project Organization

4.1 Gallstone Project Core Team

Following the identified roles and responsibilities for the Gallstone project core team members:

Name	Role	Responsibility
Hans-Ulrich Marschall	Project Manager	Project leadership
Despina Katsika	Project Manager	Project leadership ensuring delivery according to time plan and allocated project budget. Responsible for receiving and coordinate the referrals from KI Biobank. Responsible for sending the file with information about Study subject and consent to KI Biobank.

4.2 KI Biobank Project Core Team

Following the identified roles and responsibilities for the KI Biobank project core team members:

Name	Role	Responsibility
Linda Lundström	Project Manager	Project leadership ensuring delivery according to time plan and allocated project budget, SDM and LIMS integration
Lena Jakobsson	Coordinator	LIMS integration
Carita Rask	Coordinator	Lab
Gunnel Tybring	Project sponsor	Biobank Manager

4.3 Risk Management Plan

A proper risk analysis has not been performed, but some risk events have been identified and these are listed in the table below.

Rank	Risk event	Probability	Impact	Risks response
1	Samples arrive at KI Biobank without barcodes and paper referral form	Low	High	Nothing to do here, the samples has to be discarded. The GALLSTONE study administration gets notified.
2	Samples arrive at KI Biobank with barcodes but without paper referral form	Low	Low	The GALLSTONE study coordinator gets notified that a referral is missing
3	Samples arrive at KI Biobank with paper referral form but without barcodes.	Low	Low	Since the paper referral form arrives together with the samples and the samples is delivered per donor, it is not a problem to connect them to the donor. New barcodes is just printed to be fitted on the tubes. The Gallstone study administration gets notified.