Genetic and Functional Identification of the Likely Causative Variant for Cholesterol Gallstone Disease at the ABCG5/8 Lithogenic Locus

Oliver von Kampen,1,* Stephan Buch,1* Michael Nothnagel,2 Lorena Azocar,3 Hector Molina,3 Mario Brosch,1 Wiebke Erhart,1 Witigo von Schönfels,3 Jan Egberts,4 Marcus Seeger,1 Alexander Arlt,1 Tobias Balschun,5 Andre Franke,5 Markus M. Lerch,6 Julia Mayerle,7 Wolfgang Kratzer,7 Bernhard O. Boehm,8 Klaus Huse,9 Bodo Schniewind,4 Katharina Tiemann,10 Zhao-Yan Jiang,11 Tian-Quan Han,11 Balraj Mittal,12 Anshika Srivastava,12 Mogens Fenger,13 Torben Jørgensen,14 Ramin Schirin-Sokhan,18 Anke Tönjes,15 Henning Wittenburg,16 Michael Stumvoll,15 Holger Kalthoff,17 Frank Lammert,18 Jürgen Tepel,19 Klaus Puschel,21 Thomas Becker,4 Stefan Schreiber,1 Matthias Platzer,9 Henry Völzke,20 Michael Krawczak,2 Juan Francisco Miquel,3 Clemens Schafmayer,4 and Jochen Hampe1

The sterolin locus (ABCG5/ABCG8) confers susceptibility for cholesterol gallstone disease in humans. Both the responsible variant and the molecular mechanism causing an increased incidence of gallstones in these patients have as yet not been identified. Genetic mapping utilized patient samples from Germany (2,808 cases, 2,089 controls), Chile (680 cases, 442 controls), Denmark (366 cases, 766 controls), India (247 cases, 224 controls), and China (280 cases, 244 controls). Analysis of allelic imbalance in complementary DNA (cDNA) samples from human liver (n = 22) was performed using pyrosequencing. Transiently transfected HEK293 cells were used for [3H]-cholesterol export assays, analysis of protein expression, and localization of allelic constructs. Through fine mapping in German and Chilean samples, an ~250 kB disease-associated interval could be defined for this locus. Lack of allelic imbalance or allelic splicing of the ABCG5 and ABCG8 transcripts in human liver limited the search to coding single nucleotide polymorphisms. Subsequent mutation detection and genotyping yielded two disease-associated variants: ABCG5-R50C (P = 4.94 × 10−9) and ABCG8-D19H (P = 1.74 × 10−10) in high

Abbreviations: GFC, gallstone-free controls; GWAS, genome-wide association study; LD, linkage disequilibrium; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SNG, single nucleotide polymorphism.

From the 1Department of Internal Medicine I, Hospital Schleswig-Holstein, Kiel, Germany; 2Institute for Medical Informatics and Statistics, Hospital Schleswig-Holstein, Kiel, Germany; 3Institute of Medical Informatics and Statistics, Hospital Schleswig-Holstein, Kiel, Germany; 4Department of Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India; 5Department of Clinical Biochemistry, Heidekrone University Hospital, Heidekrone, Denmark; 6Department of General Internal Medicine, Faculty of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile; 7Department of General Internal Medicine I, University of Ulm Medical Centre, Ulm, Germany; 8Division of Endocrinology and Diabetes, University of Ulm Medical Centre, Ulm, Germany; 9Leibniz Institute for Aging Research, Fritz Lipmann Institute, Jena, Germany; 10Institute for Hematopathology Hamburg, Hamburg, Germany; 11Department of Surgery, Shanghai Institute of Digestive Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 12Department of Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India; 13Department of Clinical Biochemistry, Heidekrone University Hospital, Heidekrone, Denmark; 14Research Centre for Prevention and Health, Glsuntrup University Hospital, Glsuntrup, Denmark; 15University of Leipzig, Department of Medicine, Division of Endocrinology and Diabetes, and IFB Adiposity Disease, University Hospital Leipzig, Leipzig, Germany; 16Institute for Experimental Cancer Research, University Hospital Schleswig-Holstein, Kiel, Germany; 17Department of Internal Medicine II, University Hospital Hamburg, Germany; 18Department of Surgery, Hospital Onabrück, Onabrück, Germany; 19Institute of Community Medicine, University of Greifswald, Greifswald, Germany; 20Department of Family Medicine, Faculty of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile.

Received May 11, 2012; accepted July 23, 2012.

Supported by the Deutsche Forschungsgemeinschaft (DFG Ha 3091/3-1), the German Ministry of Education and Research (BMBF) and the Federal State of Schleswig-Holstein through the pogen biobank. WK and BOB received grant support from the Federal State of Baden-Württemberg (Centre of Excellence “Metabolic disorders”) and the Deutsche Forschungsgemeinschaft (DFG GrK 1041). The Study of Health in Pomerania (SHIP) is conducted by the Research Network of Community Medicine of the University Medicine Greifswald. SHIP is funded by the BMBF and the Ministries of Cultural Affairs and of Social Affairs Ministry of the Federal State of Mecklenburg-West Pomerania. Projects in Chile were funded by Fondo Nacional Cientifico y Tecnologico (FONDECYT, grant no. 1080325) and the FONDAP Center for Genome Regulation (grant no. 15000007) to J.E.M., by a joint BMBF/Conicyt project (grant no. 170-2009 to J.H. and J.F.M.), by a joint DFG/Conicyt project (444 CHL-1113/15/0-1 to F.L., J.F.M., H.W, and J.H.), and the Alexander von Humboldt Foundation (to J.E.M.).

*These authors contributed equally to this work.
pairwise linkage disequilibrium ($r^2 = 0.95$). [3H]-cholesterol export assays of allelic constructs harboring these genetic candidate variants demonstrated increased transport activity (3.2-fold, $P = 0.003$) only for the ABCG8-19H variant, which was also superior in nested logistic regression models in German ($P = 0.018$), Chilean ($P = 0.030$), and Chinese ($P = 0.040$) patient samples. Conclusion: This variant thus provides a molecular basis for biliary cholesterol hypersecretion as the mechanism for cholesterol gallstone formation, thereby drawing a link between “postgenomic” and “pregenomic” pathophysiological knowledge about this common complex disorder. (HEPATOLOGY 2013;57:2407-2417)

Gallstone disease is a frequent and economically relevant health problem worldwide.1-3 In Western countries the prevalence of cholelithiasis is ~10%-20%.1,4,5 Moreover, between 20% and 40% of gallstone patients become symptomatic or develop complications,9 and the annual number of cholecystectomies exceeds 700,000 in the U.S. and 170,000 in Germany.7,8 Economically, gallstone disease has been identified as the second most costly disorder of the digestive tract.8 A genetic component in the susceptibility to cholesterol gallstones has been recognized as early as 1937.9,10 There is compelling evidence for familial clustering of the disease, and an increased concordance rate has been observed in monozygotic as compared to dizygotic twins.11

Variation at the ABCG5/8 locus was initially identified as a susceptibility factor for gallstone disease in a genome-wide association study (GWAS), and this association has since been validated in independent studies of German, Romanian, Swedish, Danish, Chilean, and Chinese cases and controls.12-17 The findings on ABCG8 are also consistent with an earlier identification, in QTL mapping studies, of its murine ortholog at the mouse lithogenic locus Lith9.18 However, although coding variant ABCG8-D19H has often been used as the major tagging single nucleotide polymorphism (SNP) in both the discovery and the replication studies in humans, its causative role is by no means clear. In fact, identification of the disease-causing variant(s) at ABCG5/8 is still lacking, and no functional studies clarifying the likely disease mechanism have been reported so far.

Here we performed a series of experiments, combining genetic mapping in different populations and functional assays, with a view to identify the causative mechanisms of gallstone formation at the ABCG5/8 locus (Fig. 1). Our results draw a link between the classical “pregenomic” pathophysiological knowledge of the disease, namely, biliary cholesterol hypersecretion, and a disease-associated genetic variant in this common complex disorder.19,20

Materials and Methods

Patient Recruitment and Characterization. All patients with gallstone disease had undergone cholecystectomy or were diagnosed with cholecystolithiasis using B-mode ultrasonography. The gallstone-free controls (GSF) were confirmed to be gallstone-free by ultrasonography or computed tomography. Details about recruitment and clinical characterization has been reported previously for the German,12,21,22 Danish,23 Chilean,2 Indian,24 and Chinese25 patients. All patients and controls gave written informed consent prior to the study, and all study protocols were approved by the Ethics Committee of the Kiel Medical Faculty (Ethikkommission der Medizischen Fakultät der Christian-Albrechts-Universität Kiel, #A156/03).

SNP Selection, Genotyping, and Data Analysis. A set of tagging SNPs covering the extended ABCG8/ABCG5 gene region was selected from the CEU HAPMAP dataset (release 28) using Haploview26 with the following parameters: minor allele frequency ≥3%, pairwise $r^2 ≥ 0.8$ between SNPs, $P > 0.05$ for the HWE test in controls, including some redundant SNPs for more robust coverage. In addition, all coding SNPs, splice site SNPs, 5′ untranslated region (UTR)
SNPs, and SNPs located in the intergenic region of ABCG5 and ABCG8 that had a minor allele frequency >0.001 in Caucasians, as reported in dbSNP (release 132) or in the scientific literature, or that were detected by Sanger sequencing in our own mutation search were included in the genotyping. Genotyping was performed using either the Sequenom or the Taqman platform as described27,28 (see Supporting Methods for further details). All markers were tested for a possible deviation from Hardy-Weinberg equilibrium in the controls before inclusion in the subsequent statistical analysis. Single marker association tests were performed using Haploview26 and PLINK,29 comprising \( \chi^2 \) or Fisher's exact tests for contingency tables, as appropriate. Logistic regression analysis was performed using R (www.r-project.org) and SPSS (PASW Statistics 18).

**Mutation Detection at the ABCG5 and ABCG8 Loci.** All coding sequence and adjacent splice sites located in the disease-associated region were investigated for potential disease mutations by Sanger sequencing. Primer sequences are provided in the Supporting Methods.

**Tissue Samples, Pyrosequencing, and Reverse-Transcription Polymerase Chain Reaction (RT-PCR).** Human liver samples were obtained either by surgical or percutaneous biopsy. Total RNA from human liver tissue was isolated using the RNeasy kit from Qiagen (Hilden, Germany) and subsequently reverse-transcribed (Advantage RT-for-PCR kit, Clontech Laboratories, Palo Alto, CA). Matching DNA was isolated from peripheral blood samples. All patients provided written informed consent prior to the study and the sampling protocol was approved by the Ethics Committee of the Kiel Medical Faculty (Ethikkommission der Medizinischen Fakultät der Christian-Albrechts-Universität Kiel, #D425/07). Pyrosequencing was performed as described30 using primers as reported in the Supporting Methods.

**Vectors and Transfection.** Expression vectors for wildtype (WT) ABCG5 and ABCG8 were constructed by subcloning of the coding sequences of both proteins into plasmids pCEV-Cplus and pEF-HA-neo. The different versions of both transporters, defined by the respective alleles of R50C and D19H, were introduced by site-directed mutagenesis (QuickChange Lightning Site-Directed Mutagenesis Kit, Agilent Technologies, Santa Clara, CA) according to the manufacturer’s protocol. HEK cells were transfected using Effectene Transfection Reagent (Qiagen) according to the manufacturer’s protocol.

**Cholesterol Efflux Assay.** Cholesterol efflux was measured as described by Vrins et al.31 In brief, HEK cells transiently expressing different allelic variants of ABCG5-R50C and ABCG8-D19H were incubated in 24-well cell culture plates with 0.5 mL Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10 mM HEPES, pH 7.4, 30 mg/mL cholesterol, 0.5 mCi/mL \[^{3}H\]-cholesterol and 0.2% fatty acid free bovine serum albumin (BSA) for 24 hours. After removal of loading medium, cells were washed four times with DMEM, supplemented with 0.2% fatty acid free BSA, and efflux was initiated by addition of 0.5 mL DMEM supplemented with 10 mM taurocholic acid and 0.5 mM phosphatidylcholine. After 2 hours medium was collected and centrifuged (10,000 g, 5 minutes). The remaining cell associated \[^{3}H\]-cholesterol was determined after extraction for 30 minutes with 0.5 mL isopropanol. Samples of loading medium, efflux medium, and the isopropanol extract were mixed with 4 mL Aquasafe 500 Plus scintillation cocktail (Zinsser Analytic, Frankfurt, Germany) and analyzed with a beta-scintillation counter LS 6500 (Beckman Coulter, Brea, CA). The radioactivity released to the efflux medium was expressed as a fraction of the total radioactive cholesterol. Cholesterol efflux was analyzed compared to the cholesterol efflux of cells coexpressing WT ABCG5 and ABCG8 after baseline correction from cells transfected with empty expression vectors only.

**Plasma Membrane Detection of ABCG5, G8, and Allelic Variants and Western Blot Analysis.** HEK cells were grown in 10 mL culture dishes for 48 hours to 80% confluency. Biotinylation of cell surface proteins was performed with a Cell Surface Protein Isolation Kit (Pierce, Rockford, IL) according to a modified protocol (see Supporting Methods). Fifteen \( \mu \text{g} \) of each
membrane fraction and total protein lysate were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with 12% PA-Gels. Proteins were blotted onto 0.45 µm polyvinylidene difluoride membranes (Millipore, Billerica, MA) and probed with the following primary antibodies from Sigma-Aldrich: monoclonal anti-FLAG M2 antibody, polyclonal anti-HA, and monoclonal anti-β-actin. As secondary antibodies, mouse horseradish peroxidase (HRP)-linked antimouse and antirabbit conjugates (GE Healthcare, Little Chalfont, UK) were used.

Immunofluorescence Microscopy. Cells were grown on 60 mm plates to 80% confluence, transfected as described above with 2 µg total plasmid complementary DNA (cDNA). Cells were fixed with cold acetone 36 hours after transfection and subsequently permeabilized with 0.1% Triton X-100 for 15 minutes, blocked in 10% goat serum in phosphate-buffered saline (PBS), and incubated overnight with ABCG8 monoclonal antibody (NBP71706, Novus Biologicals, Littleton, CO). After washing with PBS 5 times, cells were incubated with a secondary antimouse antibody (Alexa 568) for 1 hour, washed with PBS 1×, and mounted in Fluoromount G (Electron Microscopy Sciences, Hatfield, PA). Labeled sections were examined and captured with a confocal microscope (Olympus BX51, Olympus, Tokyo, Japan).

Results

Linkage Disequilibrium (LD)-Based Definition of the Disease Interval in the ABCG5/8 Locus. The overall study workflow is depicted in Fig. 1. A total of 1,266 German cholelithiasis patients and 1,000 gallowstone-free controls ("mapping panel," Table 1) were used to analyze the disease association and the local LD structure using 57 known tagging SNPs at the ABCG5/8 locus (Fig. 2A) selected CEU HAPMAP variants. In a single point analysis (Supporting Table 1), the most significant disease associations were observed in an allele-based test for variants rs11887534 (ABCG8-D19H, \( P = 1.7 \times 10^{-10} \)), rs72875462 (ABCG8 intronic, \( P = 2.1 \times 10^{-10} \)), and rs56132765 (ABCG8-V151V \( P = 3.9 \times 10^{-10} \)). SNP rs6756629 (ABCG5-R50C) yielded the second most significant association (\( P = 4.9 \times 10^{-8} \)) of all previously known, nonsynonymous variants in ABCG5/8. The same set of 57 tagging SNPs was also analyzed in a Chilean case control sample, where it yielded a similar pattern of disease association and intermarker LD (Supporting Table 1, Fig. 2B). Based on pairwise \( r^2 > 0.5 \) to any of the four variants mentioned above, a disease-associated genomic region was defined using the mapping panel, comprising map positions 44.05 Mb to 44.08 Mb (hg19/GRCh37 genome assembly) on chromosome 2. Owing to the opposite orientation of the two ABCG genes, this region contained the five 5′ exons of ABCG5 and the six 5′ exons of ABCG8.

Disease-Associated Variants Are Not Associated with Allelic Splicing or Allelic Expression. Because formal definition of the disease-associated genomic region on chromosome 2 confined the location of potentially causal variants to the ABCG5 and ABCG8 genes, their transcripts were next evaluated for allelic expression differences. To this end, pyrosequencing assays for exonic variants rs6720173, rs6756629 in ABCG5 and rs11887534, rs56132765, and rs4148217 in ABCG8 were applied to 22 cDNA samples from human liver, including at least four heterozygotes for each variant. No allelic imbalance was detected for any of the five SNPs, including the most significantly disease-associated variant rs11887534 (Fig. 3A). This lack of allelic imbalance was also indicative of an absence of allele-dependent splicing. In order to systematically exclude the presence of variants affecting transcript structure, however, alternative splicing was examined by PCR across both transcripts. To this end, cDNAs from homozygous carriers of the major allele and

Table 1. Overview of Patient Samples

<table>
<thead>
<tr>
<th>Designation</th>
<th>Origin</th>
<th>N&lt;sub&gt;cases&lt;/sub&gt;</th>
<th>Median Age</th>
<th>% Male</th>
<th>BMI</th>
<th>N&lt;sub&gt;controls&lt;/sub&gt;</th>
<th>Median age</th>
<th>% Male</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Germany 1</td>
<td>1,266</td>
<td>54</td>
<td>36</td>
<td>27.2</td>
<td>1,000</td>
<td>64</td>
<td>36</td>
<td>26.8</td>
</tr>
<tr>
<td>Mapping</td>
<td>Chile 1</td>
<td>167</td>
<td>55</td>
<td>0</td>
<td>29.5</td>
<td>81</td>
<td>55</td>
<td>0</td>
<td>29.9</td>
</tr>
<tr>
<td>Differentiation</td>
<td>Germany 2</td>
<td>1,542</td>
<td>59</td>
<td>28</td>
<td>27.4</td>
<td>1,089</td>
<td>61</td>
<td>28</td>
<td>27.6</td>
</tr>
<tr>
<td>Differentiation</td>
<td>Chile all</td>
<td>680</td>
<td>55</td>
<td>10</td>
<td>29.1</td>
<td>442</td>
<td>55</td>
<td>10</td>
<td>28.8</td>
</tr>
<tr>
<td>Differentiation</td>
<td>Denmark</td>
<td>321</td>
<td>60</td>
<td>38</td>
<td>25.5</td>
<td>816</td>
<td>60</td>
<td>38</td>
<td>26.2</td>
</tr>
<tr>
<td>Differentiation</td>
<td>India</td>
<td>247</td>
<td>48</td>
<td>32</td>
<td>na</td>
<td>224</td>
<td>44</td>
<td>40</td>
<td>Na</td>
</tr>
<tr>
<td>Differentiation</td>
<td>China</td>
<td>280</td>
<td>54</td>
<td>36</td>
<td>24.6</td>
<td>244</td>
<td>57</td>
<td>36</td>
<td>22.9*</td>
</tr>
</tbody>
</table>

*Mapping samples* were utilized for the definition of the disease interval. *Differentiation* cohorts were genotyped for the assignment of the likely causative mutations among the significant functional variants in the disease interval. The "Chile all" sample included the individuals from the Chilean mapping sample.

*BMI information was only available for 41% of the control individuals."
heterozygous individuals for the two most strongly disease-associated variants rs11887534 and rs6756629 were studied in an exon walk. However, no evidence of allele-dependent splicing was found (Fig. 3B).

**ABCG8-D19H and ABCG5-R50C Are the Only Functional Candidates in the Disease Interval.** Owing to the absence of allelic imbalance and allele-dependent splicing, disease mechanisms such as differential
transcription efficiency (caused by promoter polymorphisms or variable intronic enhancers) as well as differential transcript structure or stability could be safely ruled out. Therefore, we next investigated, by Sanger sequencing, all protein-coding DNA sequence in the disease-associated region for potential disease mutations, using samples from 53 cases and 43 controls including 33 carriers of the risk allele of rs11887534 (Supporting Table 2). In addition to known coding variants (according to dbSNP release 132), only one new variant in exon11 of ABCG8 (Ala548Ala) was detected, namely, in a gallstone patient lacking the risk
allele of rs11887534. All known coding variants (dbSNP release 132) and a selection of synonymous variants from dbSNP were subsequently genotyped in the mapping panel (single point association results in Supporting Table 1). Again, SNP rs11887534 was found to be the most significantly disease associated variant.

**ABCG8-D19H, but Not ABCG5-R50C, Leads to Increased Cholesterol Transport.** Of all coding SNPs investigated in the disease-associated region, only rs6756629 (ABCG5-R50C; odds ratio [OR] allelic $\hat{\text{OR}} = 1.96$, 95% confidence interval [CI]: 1.56-2.47) and rs11887534 (ABCG8-D19H; $\hat{\text{OR}} = 2.07$, 95% CI: 1.65-2.60) were found to be significantly associated with cholelithiasis in our mapping panel. The significance of their disease associations was approximately seven orders of magnitude higher than that of the next “best” SNP (rs6720173, ABCG5-Q604E, $\hat{\text{OR}} = 0.0024$) which, moreover, is located well outside the disease-associated region defined above. The ORs of the two variants of interest were also found to be very similar and the variants showed high pairwise LD in the German mapping sample ($r^2 = 0.95$), thereby indicating that assessment of their possible causative role by genetic means alone would be difficult.

Because ABCG5 and ABCG8 form a heterodimeric biliary sterol transporter, expression constructs using all four possible allele combinations of the two SNPs were therefore transiently transfected into HEK293 cells. In line with previous observations, each transporter constituent was glycosylated only upon coexpression of the other constituent, irrespective of their allelic status, thereby indicating correct shuttling into the cell membrane (Fig. 4; Supporting Fig. 1). In line with these observations, no transporter was detected in the cell membrane fractions when expressed alone. Membrane integration was also confirmed by immunofluorescence assays using antibodies that target ABCG8 (Fig. 5). Thus, expression of ABCG8-19D or ABCG8-19H alone yielded only intracellular fluorescence, whereas coexpression with either ABCG5-50R or ABCG5-50C resulted in a membrane-associated fluorescence signal. Because ABCG5 is retained in the endoplasmic reticulum (ER) unless ABCG8 is expressed in the same cell, the membrane associated fluorescence in cells expressing the ABCG5 dimerization partner (i.e., ABCG8-19D or -19H) is an indirect proof of the correct assembly of the two transporter constituents. Signals for ABCG5 and ABCG8 were similar for all four allelic combinations, both in a western blot analysis using monoclonal tag antibodies (Fig. 4) or antibodies directed against human ABCG5 and ABCG8 protein (Supporting Fig. 1), thereby suggesting that differential protein expression is unlikely to represent the primary disease mechanism.

Next, cholesterol efflux assays using [3H]-labeled cholesterol were performed and revealed a 3.2-fold higher transport efficiency when the ABCG8-19H allele was included in the expression construct (two-sided $t$ test, $P = 0.0032$ and $P = 0.0036$ for the coexpression of either ABCG5-50R or ABCG5-50C, respectively; Fig. 6). Whether ABCG5-50R or ABCG5-50C was present in the construct did not influence cholesterol efflux, given the respective ABCG8-D19H allele. Thus, among the two genetically linked candidate SNPs ($r^2 = 0.95$), we may conclude that an increased sterol transport efficiency mediated by ABCG8-19H alone is a likely functional effect of the two SNPs.

**ABCG8-D19H But Not ABCG5-R50C Captures the Genetic Risk Across Populations.** In addition to the functional experiments, nested logistic regression analyses were performed to statistically evaluate the causative role of individual variants in the disease-associated region. In the German mapping panel, none of the coding SNPs listed in Supporting Table 1 significantly improved the allelic risk model over the inclusion of ABCG8-D19H alone. When ABCG5-R50C
was included as a mandatory explanatory variable, ABCG8-D19H significantly improved the model fit ($P = 0.0175$), thereby suggesting again that ABCG8-D19H is the major causative variant in the region. To confirm this result, rs6756629 (ABCG5-R50C) and rs11887534 (ABCG8-D19H) were also genotyped in additional case-controls samples of German, Chilean, Danish, Indian, and Chinese origin (Table 2). Sample size and haplotype structure allowed differentiation between the effects of the two SNPs in the Chilean ($P = 0.030$) and Chinese substudy ($P = 0.040$), in the latter owing to the lower level of LD between the two variants in this population ($r^2 = 0.73$) despite the limited sample size. In each of the samples, ABCG8-D19H consistently showed an equal or higher allelic OR as compared to ABCG5-R50C, thereby supporting the above conclusion. Moreover, logistic regression analysis containing ABCG8-D19H and other tagging SNPs (Supporting Table 1) were performed in the German and Chilean mapping panels also revealing no significant improvement over ABCG8-D19H alone (all $P > 0.1$). Thus, a prominent role of hitherto undetected rare variants tagged by one of the other SNPs at the locus of interest is also unlikely.

**Discussion**

Biliary cholesterol hypersaturation has long been established, through bile analysis in patients and through animal models, as one of the key mechanisms of gallstone formation.33-37 With the introduction of Fourier Transform Infrared Spectroscopy (FTIR) as a means of analyzing gallstone composition in the 1970s,38,39 cholesterol has been confirmed as the predominant compound of gallstones.40 Indeed, a recent analysis of over 1,000 samples from Germany revealed cholesterol as the most abundant substance in over 93% of gallstones.41 ABCG5 and ABCG8 form obligate heterodimers for biliary sterol secretion and export...
cholesterol from the canalicular membrane of hepatocytes into the biliary lumen. Loss of function mutations of ABCG5/8 cause a rare autosomal recessive disorder, known as sitosterolemia, that is characterized by increased fractional absorption and decreased biliary secretion of neutral sterols. In contrast, hepatic overexpression of ABCG5/8 increases hepatic sterol transport and biliary sterol content. In the present study we were able to link a statistically inferred candidate disease mutation at the ABCG5/8 locus to this pathomechanism. More specifically, we could show that the histidine-encoding allele of rs11887534 (ABCG8-D19H) is associated with an increased allelic cholesterol transport efficiency of the ABCG5/8 heterodimer, thereby contributing to biliary cholesterol hypersecretion and supersaturation and, most likely, to a predisposition to gallstone formation. This increased biliary cholesterol transport may also contribute to the lower serum cholesterol and phytosterol levels observed previously in carriers of ABCG8-D19H.

Further searches for the disease-causing variant(s) could thus be confined to the coding sequence of the two genes. After extensive genotyping, only two candidate variants remained: ABCG5-R50C and ABCG8-D19H, both with similar allelic odds ratios (Supporting Table 1; Table 2). Because the two SNPs were in high LD, formal genetic differentiation of the two variants was achieved only in three out of the six investigated patient samples from Germany, Denmark, Chile, India, and China, implicating rs11887534 (ABCG8-D19H) as the likely risk factor. Differentiation in the Chinese patient sample was aided by the markedly different LD structure in this population. In line with these statistical results, only ABCG8-D19H was found to yield a functional effect in in vitro coexpression experiments.

Despite our systematic and intensive mutation search, it cannot be ruled out that hitherto undiscovered rare variants at the ABCG5/8 locus also play a causative role in cholelithiasis. However, in a logistic regression analysis none of the known common tagging SNPs improved the model fit significantly over rs11887534 (ABCG8-D19H), which implies that there is probably not much “missing heritability” left at this locus.

The role of the ABCG5/8 heterodimer as a biliary cholesterol transporter and cholesterol supersaturation as a key pathomechanism of gallstone formation on the functional side and the association of the ABCG5/8 locus with gallstone risk using rs11887534 as the tagging variant were known previously. In this report, we genetically establish rs11887534 (ABCG8-D19H) through thorough transcript mapping, mutation detection, and association analysis in ethnically different populations as the likely causative variant for gallstone susceptibility. Further, we show that rs11887534 (ABCG8-D19H) increases the transport efficiency of cholesterol of the sterolin heterodimer, thereby drawing a link between the known basic pathomechanism

Table 2. Results of Single Marker Association Tests (Denoted “Single”) and Nested Logistic Regression Analysis of ABCG5-R50C and ABCG8-D19H

<table>
<thead>
<tr>
<th>Cases Control</th>
<th>N</th>
<th>Fca</th>
<th>Fco</th>
<th>ORallelic [CI:95%]</th>
<th>D19H</th>
<th>Fca</th>
<th>Fco</th>
<th>ORallelic [CI:95%]</th>
<th>Nested Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany 1</td>
<td>1,266</td>
<td>1,000</td>
<td>0.104</td>
<td>0.056</td>
<td>1.96 [1.56-2.47]</td>
<td>0.110</td>
<td>0.056</td>
<td>2.07 [1.65-2.60]</td>
<td>0.079</td>
</tr>
<tr>
<td>Germany 2</td>
<td>1,542</td>
<td>1,089</td>
<td>0.102</td>
<td>0.049</td>
<td>2.23 [1.78-2.80]</td>
<td>0.107</td>
<td>0.051</td>
<td>2.23 [1.79-2.79]</td>
<td>0.092</td>
</tr>
<tr>
<td>Denmark</td>
<td>336</td>
<td>766</td>
<td>0.089</td>
<td>0.053</td>
<td>1.73 [1.22-2.45]</td>
<td>0.092</td>
<td>0.055</td>
<td>1.75 [1.24-2.47]</td>
<td>0.376</td>
</tr>
<tr>
<td>Chile all</td>
<td>680</td>
<td>442</td>
<td>0.124</td>
<td>0.075</td>
<td>1.75 [1.29-2.35]</td>
<td>0.123</td>
<td>0.070</td>
<td>1.85 [1.37-2.51]</td>
<td>0.963</td>
</tr>
<tr>
<td>India</td>
<td>247</td>
<td>224</td>
<td>0.047</td>
<td>0.022</td>
<td>2.14 [1.01-4.54]</td>
<td>0.047</td>
<td>0.022</td>
<td>2.14 [1.01-4.54]</td>
<td>0.963</td>
</tr>
<tr>
<td>China</td>
<td>280</td>
<td>244</td>
<td>0.009</td>
<td>0.012</td>
<td>0.72 [0.21-2.38]</td>
<td>0.011</td>
<td>0.008</td>
<td>1.31 [0.37-4.67]</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Fca: allele frequency in cases, Fco: allele frequency in controls, R50C | D19H denotes R50C introduced to the model given D19H is contained in the model. The allelic odds ratios of D19H are consistently equal or greater than for R50C in all patient samples. The P-value of patient samples, where successful differentiation of the effect between D19H and R50C was possible is denoted in bold.
of cholesterol gallstone formation to this specific genetic variant.

References