Author(s): Coviello, Andrea D; Haring, Robin; Wellons, Melissa, et al
Title: A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation.
Year: 2012
Journal Title: Plos Genetics
Vol and number: 8 : 7
Pages: 1-12
ISSN: 1553-7390
Discipline: Biomedicine
School/Other Unit: School of Medicine
Item Type: Journal Article
Language: en
DOI: http://dx.doi.org/doi:10.1371/journal.pgen.1002805
URN: URN:NBN:fi:uta-201210181058
URL: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002805

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.
A Genome-Wide Association Meta-Analysis of Circulating Sex Hormone–Binding Globulin Reveals Multiple Loci Implicated in Sex Steroid Hormone Regulation


1 Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts, United States of America, 2 Section of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts, United States of America, 3 National Heart, Lung, and Blood Institute’s The Framingham Heart Study, Framingham, Massachusetts, United States of America, 4 Institute for Clinical Chemistry and Laboratory Medicine, University of Münster, Münster, Germany, 5 Department of Medicine and Department of Obstetrics and Gynecology, The University of Alabama at Birmingham, Birmingham, Alabama, United States of America, 6 Department of Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America, 7 Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital and University of Tampere School of Medicine, Tampere, Finland, 8 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, 9 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 10 Department of Public Health, Indiana University School of Medicine, Indianapolis, Indiana, United States of America, 11 Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana, United States of America, 12 University of Texas Health Science Center at Houston, Houston, Texas, United States of America, 13 Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Oxford, United Kingdom, 14 Department of Twin Research and Genetic Epidemiology, King’s College London, London, London, United Kingdom, 15 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, 16 Program on Genomics and Nutrition and the Center for Metabolic Disease Prevention, School of Public Health, University of California Los Angeles, Los Angeles, California, United States of America, 17 Center for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, 18 Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland, United States of America, 19 Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 20 Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, North Carolina, United States of America, 21 Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland, United States of America, 22 Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, 23 Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, 24 Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital, and Harvard Medical School, Boston, Massachusetts, United States of America, 25 Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands, 26 Netherlands Consortium of Healthy Aging, Rotterdam, The Netherlands, 27 Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter, United Kingdom, 28 Institute of Diagnostics, University of Oulu, Oulu, Finland, 29 Department of Obstetrics and Gynecology, University Hospital of Oulu, Oulu, Finland, 30 National Institute for Health and Welfare and Institute of Health Sciences, University of Oulu, Oulu, Finland, 31 Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy, 32 Department of Prosthetic Dentistry, Gerostomatology, and Dental Materials, University of Greifswald, Greifswald, Germany, 33 Experimental and Clinical Endocrinology, University of Lübeck, Lübeck, Germany, 34 Cancer Research Center Lyon, INSERM U1052, Lyon, France, 35 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom, 36 California Pacific Medical Center, San Francisco, California, United States of America, 37 Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, United States of America, 38 Division of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Massachusetts, United States of America, 39 Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, 40 Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands, 41 Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany, 42 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, 43 Hannover Unified Biobank, Hannover Medical School, Hannover, Germany, 44 Hebrew SeniorLife Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, United States of America, 45 Clinical and Molecular Osteoporosis Research Unit, Department of Clinical Sciences and Department of Orthopaedics, Lund University, Malmö, Sweden, 46 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, 47 Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland.
Abstract

Sex hormone-binding globulin (SHBG) is a glycoprotein responsible for the transport and biologic availability of sex steroid hormones, primarily testosterone and estradiol. SHBG has been associated with chronic diseases including type 2 diabetes (T2D) and with hormone-sensitive cancers such as breast and prostate cancer. We performed a genome-wide association study (GWAS) meta-analysis of 21,791 individuals from 10 epidemiologic studies and validated these findings in 7,046 individuals in an additional six studies. We identified twelve genomic regions (SNPs) associated with circulating SHBG concentrations. Loci near the identified SNPs included SHBG (rs12150660, 17p13.1, p = 1.8 × 10−13), PRMT6 (rs17496332, 1p13.3, p = 1.4 × 10−12), GCKR (rs700933, 2p23.3, p = 2.2 × 10−10), ZBTB10 (rs440837, 8q21.13, p = 3.4 × 10−9), MBD1C (rs57910927, 10q21.3, p = 6.1 × 10−7), SLCO1B1 (rs4194956, 12p12.1, p = 1.9 × 10−6), NRZ2 (rs8023580, 15q26.2, p = 8.3 × 10−5), ZNF652 (rs2411984, 17q21.32, p = 3.5 × 10−3), TDGF3 (rs1573036, 1q22.3, p = 4.1 × 10−4), LHCB1 (rs10454142, 2p16.3, p = 1.3 × 10−4), BAAP2L1 (rs3779195, 2q21.3, p = 2.7 × 10−3), and UGT2B15 (rs293428, 4q13.2, p = 5.5 × 10−3). These genes encompass multiple biologic pathways, including hepatic function, lipid metabolism, carbohydrate metabolism and T2D, androgen and estrogen receptor function, epigenetic effects, and the biology of sex steroid hormone-responsive cancers including breast and prostate cancer. We found evidence of sex-differentiated genetic influences on SHBG. In a sex-specific GWAS, the loci 4q13.2-6p21 and Xq22.3-TDG3 were significant in men only (men p = 2.5 × 10−6, women p = 0.66, heterogeneity p = 0.003). Additionally, three loci showed strong sex-differentiated effects: 17p13.1-SHBG and Xq22.3-TDG3 were stronger in men, whereas 8q21.12-ZBTB10 was stronger in women. Conditional analyses identified additional signals at the SHBG gene that together almost double the proportion of variance explained at the same loci. Using sex-specific GWAS, the loci 4q13.2-6p21 and Xq22.3-TDG3 were stronger in women, respectively. The evidence for sex-differentiated associations and allelic heterogeneity highlight the importance of considering these features when estimating complex trait variance.


Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received February 20, 2012; Accepted May 19, 2012; Published July 19, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This meta-analysis is a collaborative effort involving data from many individual studies and many sources of funding. The details of funding sources for each study are detailed in Text S1 as well as below. Framingham Heart Study (PHS): The phenotype-genotype association analyses were funded through grants from the NIA R21 AG032598 (JM Murabito, KL Lunetta), R01 HL094755 (AD Coviello, RS Vasan, S Bandinelli), and R01 AG13026 (RS Vasan, S Bandinelli), R01 AR/AG 41398 (DP Kiel). This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARE) project. This work was partially supported by the National Heart, Lung, and Blood Institute’s Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinCaA), funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Gothenburg Osteoporosis and Obesity Determinants (GOOD) Study: Financial support was received from the Swedish Research Council (K2010-54X-09894-19-3, 2006-3832, and K2010-52X-20229-05-3), the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg’s Foundation, Petrus and Augusta Hedlund’s Foundation, the Våra GötaLand Foundation, the Göteborg Medical Society, the Novo Nordisk Foundation, and the European Commission grant HEALTH-F2-2008-201865-GEFOS. We would like to acknowledge Maria Netheran at the genomics core facility at University of Gothenburg for statistical analyses. We would also like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technologie under grants #01AK03 A-H and #01 IG 07015 G for access to their grid resources. We would also like to thank Karel Estrada, Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands, for advice regarding
These authors also contributed equally to this work.

* E-mail: John.perry@pms.ac.uk (JRB Perry); murabito@bu.edu (JM Murabito)

Competing Interests: The current research is supported by CA87969, CA49449, CA40356, CA128034, and U01-CA98233 from the National Cancer Institute. We acknowledge the study as part of the Cancer Genetic Markers of Susceptibility (CGEMS) initiative of the NCI. We particularly acknowledge the contributions of R. Hoover, A.Hutchinson, K. Jacobs and was received from the Swedish Research Council (2006-3832), the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg HL087114. European Prospective Investigation into Cancer and Nutrition (Prospect-EPIC): The Prospect-EPIC study was funded by ''Europe against Cancer'' Programme of the European Commission (Health-F2-2008-201865-GEFOS). Nurses’ Health Study (NHS): The NHS breast cancer GWAS was performed by the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping led by Leena Peltonen and thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping led by Leena Peltonen and thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping led by Leena Peltonen and thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping led by Leena Peltonen. The authors have declared that no competing interests exist.
Author Summary

Sex hormone-binding globulin (SHBG) is the key protein responsible for binding and transporting the sex steroid hormones, testosterone and estradiol, in the circulatory system. SHBG regulates their bioavailability and therefore their effects in the body. SHBG has been linked to chronic diseases including type 2 diabetes and to hormone-sensitive cancers such as breast and prostate cancer. SHBG concentrations are approximately 50% heritable in family studies, suggesting SHBG concentrations are under significant genetic control; yet, little is known about the specific genes that influence SHBG. We conducted a large study of the association of SHBG concentrations with markers in the human genome in ~22,000 white men and women to determine which loci influence SHBG concentrations. Genes near the identified genomic markers in addition to the SHBG protein coding gene included PRMT6, GCKR, ZBTB10, JMJ1C, SLCO1B1, NR2F2, ZNF652, TDRF3, LHCGR, BAUAP2L1, and UGT2B15. These genes represent a wide range of biologic pathways that may relate to SHBG function and sex steroid hormone biology, including liver function, lipid metabolism, carbohydrate metabolism and type 2 diabetes, and the development and progression of sex steroid hormone-responsive cancers.

Introduction

Sex hormone-binding globulin (SHBG) is a protein secreted mainly by the liver that binds to the sex steroids, testosterone, dihydrotestosterone, and estradiol, transports them in the circulation, and influences their action in target tissues by regulating their bioavailability. SHBG thereby influences the expression of sex hormone sensitive phenotypes including sexual characteristics and reproductive function in men and women [1]. In addition to regulating sex steroid hormone effects, SHBG may exert independent effects through its own receptor [2]. Variation in SHBG concentration has also been associated with various chronic diseases including cancers [3], polycystic ovary syndrome (PCOS) [4,5] and type 2 diabetes (T2D) [6,7]. Although SHBG is estimated to have a heritable component (~50%) [8], little is known about the genetic regulation of SHBG. Polymorphisms at the **SHBG** gene locus have been associated with SHBG concentrations [9,10], but much remains unknown about specific genetic variants that may determine circulating SHBG concentrations. Identifying genetic factors that influence SHBG may provide insights into the biology of sex steroid hormone regulation, metabolism and tissue effects that underlie their relationship with chronic diseases such as T2D as well as hormone-sensitive cancers such as breast and prostate cancer.

Results

We identified nine loci associated with SHBG concentrations at the genome-wide significance threshold of p = 5 \times 10^{-8} (Table 1 and Figure 1) in a genome-wide association study (GWAS) meta-analysis of circulating SHBG concentrations in 21,791 men and women from 10 studies (Table S1). All nine lead SNPs at these loci had effects in the same direction (seven with p<0.05) in the validation dataset of 7,046 men and women from six additional studies (Table S2). The strongest association was within the **SHBG** locus (rs12150666, p = 2 \times 10^{-10}). Together, these nine lead SNPs explained 7.2% of the genetic variance (assuming 50% heritability) in SHBG concentrations.

We next performed a series of additional analyses to explain more of the phenotypic variance (Figure 2). First, we hypothesized that genetic effects may be different in men and women, as SHBG concentrations are >50% higher in females than males, and may be differentially regulated between sexes. In a sex stratified analysis, three of the nine loci showed evidence of sex-differentiated effects at p<0.02 when we would not expect any signals to have reached this level of significance by chance. The associations at the 17p13.1-SHBG and Xq22.3 loci were stronger in males whereas the association at the 8q21.13 locus was stronger in females. To investigate the apparent differential sex effect for the X chromosome further we ran a recessive regression model for the X chromosome SNP rs1573036 in women in the Framingham Heart Study and found no association with SHBG suggesting the sex-differentiated effect is not the result of a recessive inheritance pattern. Sex stratified GWAS identified one novel signal in men, which showed no association in women (4q13.2: men p = 2.5 \times 10^{-6}, women p = 0.66, heterogeneity p = 0.003).

A series of conditional analyses were performed to identify statistically independent signals. At the **SHBG** locus, three apparently independently additional signals separate from the main index SNP were observed, based on low (r^2<0.05) pairwise correlations in HapMap (rs6258 p = 2.7 \times 10^{-46}, rs1625895 p = 1.2 \times 10^{-14} and rs5333894 p = 2.5 \times 10^{-11}). A series of iterative conditional analyses (Table 2) involving SNPs at the **SHBG** locus generated a final regression model including six statistically independent **SHBG** SNPs. Four of these SNPs (#1–4 Table 2) retained GWS when conditioned against the other five, and two were nominally associated (SNP#5 p = 0.0001, SNP#6 p = 0.01). Re-running the GWAS meta-analysis adjusting for these six SNPs revealed evidence for three additional statistically independent (pairwise HapMap r^2<0.01) signals at the **SHBG** locus (SNP#7 p = 1.5 \times 10^{-7}, SNP#8 p = 1.6 \times 10^{-5}, SNP#9 p = 9.9 \times 10^{-6}) (Figure 3). There were also two additional trans signals located at 2p16.3 and 7q21.3 (Table 1). Although the 2p16.3 signal dropped below GWS when combined with follow-up samples (p = 1 \times 10^{-7}), the index SNP at 2p16.3 is ~300 kb away from a strong candidate gene, the luteinizing hormone receptor gene (**LHCGR**).

The majority of pair-wise correlations for the nine **SHBG** locus SNPs highlighted by our conditional analyses showed very low HapMap r^2 values. However, the pairwise D’ values are often high (Table S3) indicating that no or few recombination events have occurred between some SNPs, and that combinations of SNPs may be tagging un-typed variants on a common haplotype. To investigate this possibility, we performed more extensive analyses in a single study (NFBC1966, n = 4467). We used a denser set of SNPs imputed from the June 2011 version of the 1000 Genomes data and performed model selection analyses. Model selection identifies a set of SNPs that best explain phenotypic variation, while simultaneously penalizing each SNP included in this set, and therefore correlated SNPs tend to be excluded from the final model. These analyses consistently included at least seven SNPs in the model, although it is hard to estimate the false-negative rate of using the reduced sample size. While we are underpowered to accurately pinpoint the exact number of independent signals, these analyses support the results of the conditional analysis and suggest that multiple variants at the **SHBG** locus are independently associated with SHBG concentrations.

Data from an independent study, the InCHIANTI study, was used to calculate the proportion of genetic variance in SHBG concentrations explained when accounting for sex specific effects, the multiple signals of association at the **SHBG** locus, and the additional trans signals identified post conditional analysis. In men and women we explained ~15.6% and ~8.4% of the heritable...
### Table 1. SNPs representing loci associated with circulating SHBG concentrations.

<table>
<thead>
<tr>
<th>Index SNP</th>
<th>Analysis</th>
<th>Region</th>
<th>Nr Gene</th>
<th>Chr</th>
<th>Position</th>
<th>Effect Allele</th>
<th>Other Allele</th>
<th>EAF</th>
<th>Discovery Samples</th>
<th>Discovery+Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17496332</td>
<td>Main</td>
<td>1p13.3</td>
<td>PRMT6</td>
<td>1</td>
<td>107347898</td>
<td>a</td>
<td>g</td>
<td>0.67</td>
<td>-0.026</td>
<td>0.0066</td>
</tr>
<tr>
<td>rs780093</td>
<td>Main</td>
<td>2p23.3</td>
<td>GCKR</td>
<td>2</td>
<td>27596107</td>
<td>t</td>
<td>c</td>
<td>0.40</td>
<td>-0.033</td>
<td>0.0043</td>
</tr>
<tr>
<td>rs440837</td>
<td>Main</td>
<td>8q21.13</td>
<td>ZBTB10</td>
<td>8</td>
<td>81624529</td>
<td>a</td>
<td>g</td>
<td>0.78</td>
<td>-0.030</td>
<td>0.0052</td>
</tr>
<tr>
<td>rs7910927</td>
<td>Main</td>
<td>10q21.3</td>
<td>JMJD1C</td>
<td>10</td>
<td>64808916</td>
<td>t</td>
<td>g</td>
<td>0.51</td>
<td>-0.044</td>
<td>0.0043</td>
</tr>
<tr>
<td>rs4149056</td>
<td>Main</td>
<td>12p12.1</td>
<td>SLCO1B1</td>
<td>12</td>
<td>21222816</td>
<td>t</td>
<td>c</td>
<td>0.82</td>
<td>0.032</td>
<td>0.0057</td>
</tr>
<tr>
<td>rs8023580</td>
<td>Main</td>
<td>15q26.2</td>
<td>NR2F2</td>
<td>15</td>
<td>94509295</td>
<td>t</td>
<td>c</td>
<td>0.72</td>
<td>-0.029</td>
<td>0.0049</td>
</tr>
<tr>
<td>rs12150660</td>
<td>Main</td>
<td>17p13.1</td>
<td>SHBG</td>
<td>17</td>
<td>7462640</td>
<td>t</td>
<td>g</td>
<td>0.24</td>
<td>0.100</td>
<td>0.0053</td>
</tr>
<tr>
<td>rs2411984</td>
<td>Main</td>
<td>17q21.32</td>
<td>ZNF652</td>
<td>17</td>
<td>44800750</td>
<td>a</td>
<td>g</td>
<td>0.28</td>
<td>0.034</td>
<td>0.0049</td>
</tr>
<tr>
<td>rs157036</td>
<td>Main</td>
<td>Xq22.3</td>
<td>TDGFI3</td>
<td>23</td>
<td>109706724</td>
<td>t</td>
<td>c</td>
<td>0.39</td>
<td>0.031</td>
<td>0.0043</td>
</tr>
<tr>
<td>rs10454142</td>
<td>Conditional</td>
<td>2p16.3</td>
<td>LHCGR</td>
<td>2</td>
<td>48499903</td>
<td>t</td>
<td>c</td>
<td>0.69</td>
<td>0.026</td>
<td>0.0047</td>
</tr>
<tr>
<td>rs3779195</td>
<td>Conditional</td>
<td>7q21.3</td>
<td>BAIA2P2L1</td>
<td>7</td>
<td>97831289</td>
<td>a</td>
<td>t</td>
<td>0.17</td>
<td>-0.033</td>
<td>0.0057</td>
</tr>
<tr>
<td>rs293428</td>
<td>Sex-specific</td>
<td>4q13.2</td>
<td>UGT2B15</td>
<td>4</td>
<td>69626371</td>
<td>a</td>
<td>g</td>
<td>0.69</td>
<td>-0.023</td>
<td>0.0047</td>
</tr>
</tbody>
</table>

**All SNPs are on the + strand and positions are based on build 36. EAF = ‘effect allele frequency’. Beta units are per-allele effect estimates in natural log transformed nmol/L. Sex column gives the sex with the largest per-allele beta estimate. Missing values for conditional SNPs as sex-specific conditional analysis was not performed.**

doi:10.1371/journal.pgen.1002805.t001
component respectively. The SHBG locus accounted for $\sim$10% and $\sim$6.6% of the genetic variation in men and women respectively with the lead SNP in isolation accounting for $\sim$7.8% and $\sim$3.3% of the variation in men and women, respectively.

We identified genes near the associated SNPs and explored their biologic relevance to SHBG. The genes associated with identified SNPs included the SHBG locus (rs12150660, 17p13.1, $p = 1.8 \times 10^{-106}$), PRMT6 (rs17496332, 1p13.3, $p = 1.4 \times 10^{-11}$), GCKR (rs780093, 2p23.3, $p = 2.2 \times 10^{-16}$), ZBTB10 (rs408377, 8q21.3, $p = 3.4 \times 10^{-10}$), JMJD1C (rs7910927, 10q21.3, $p = 6.1 \times 10^{-35}$), SLCO1B1 (rs4149056, 12p12.1, $p = 1.9 \times 10^{-49}$), NR2F2 (rs8023580, 15q26.2, $p = 8.3 \times 10^{-12}$), ZNF652 (rs2411984, 17q21.32, $p = 3.5 \times 10^{-14}$), TDGF3 (rs1573036, Xq22.3, $p = 4.1 \times 10^{-18}$), LHCGR (rs10454142, 2p16.3, $p = 1.5 \times 10^{-47}$), BAIAP2L1 (rs779195, 7q21.3, $p = 2.7 \times 10^{-46}$), and UGT2B15 (rs293428, 4q13.2, $p = 5.5 \times 10^{-36}$) (Figure 1).

We used the online tool STRING (www.string-db.org) to perform pathway analyses to explore possible interactions between the SHBG gene and the proteins encoded by the 11 most plausible genes nearest the 11 SNPs listed above. There was an interaction noted between GCKR and JMJD1C which were associated with the lipoprotein fractions VLDL and HDL, respectively [11]. In an expanded analysis, we assessed protein interactions among SHBG and 67 genes within 500 kb of our 11 identified SNPs and uncovered additional protein interaction pathways. An interaction between two proteins encoded by GTF2A1L and STON1 was found; these proteins are co-expressed in testicular germ cells in the mouse [12]. An interaction between LHCGR and BR3 encoded proteins that are associated with the G-protein coupled receptor complex in the human luteinizing hormone receptor was also identified [13]. Finally, an interaction between LHCGR and LAPP (amylin) proteins which are components of a ligand/G-protein receptor/G-protein alpha subunit complex was found (database: www.reactome.com).

Targeted analysis of two strong candidate genes, hepatocyte nuclear factor-4α (HNF4α) and peroxisome-proliferating receptor γ (PPARγ) did not identify any SNPs at HNF4α but did identify one SNP, rs2920502, at PPARγ that reached statistical significance ($p = 9.9 \times 10^{-5}$) and a second SNP at PPARγ, rs13081389, that reached nominal significance ($p = 0.01$).

**Discussion**

In total, we identified 12 genomic regions associated with circulating SHBG concentrations, including extensive allelic heterogeneity at the SHBG locus itself. Conditional meta-analyses carried out at the SHBG locus, identified nine genome-wide significant SNPs with low correlation ($r^2 < 0.01$) between them. Two of these signals (rs6258 [10] and rs6259) are missense variants and two are low frequency variants (MAF < 2%). Furthermore, rs12150660 is highly correlated ($r^2 > 0.95$) [10] with a pentanu-

---

**Figure 1.** Manhattan plot of the autosomal SNPs identified in the GWA meta-analysis. The Manhattan plot depicts the SNPs identified in the GWA analysis labeled with the nearest gene on the plot. The SNP identified on the X chromosome, rs1573036, at Xq22.3, is not included in this figure.

doi:10.1371/journal.pgen.1002805.g001
A tandem repeat, which affects SHBG expression \textit{in-vitro} [14]. To our knowledge, the magnitude of secondary signals observed at this locus are the largest seen for any complex trait.

The proportion of genetic variance in SHBG serum concentrations explained when accounting for sex specific effects, the multiple signals of association at the SHBG locus, and the

**Table 2. Statistically independent signals at the SHBG gene locus.**

<table>
<thead>
<tr>
<th>SNP #</th>
<th>Model</th>
<th>Conditioned On SNP #</th>
<th>SNP</th>
<th>Position</th>
<th>Effect allele</th>
<th>Other allele</th>
<th>EAF</th>
<th>Beta</th>
<th>SE</th>
<th>p-value</th>
<th>Discovery p-value</th>
<th>Discovery Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Full model</td>
<td>2–6</td>
<td>rs12150660</td>
<td>7462640</td>
<td>t</td>
<td>g</td>
<td>0.24</td>
<td>0.082</td>
<td>0.005</td>
<td>1.89E-55</td>
<td>1.19E-79</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>Full model</td>
<td>1,3–6</td>
<td>rs6258</td>
<td>7475403</td>
<td>t</td>
<td>c</td>
<td>0.02</td>
<td>-0.272</td>
<td>0.017</td>
<td>1.03E-60</td>
<td>2.69E-46</td>
<td>-0.2613</td>
</tr>
<tr>
<td>3</td>
<td>Full model</td>
<td>1–2,4–6</td>
<td>rs1641537</td>
<td>7486446</td>
<td>t</td>
<td>c</td>
<td>0.14</td>
<td>-0.064</td>
<td>0.006</td>
<td>1.20E-64</td>
<td>8.19E-39</td>
<td>-0.0814</td>
</tr>
<tr>
<td>4</td>
<td>Full model</td>
<td>1–3,5–6</td>
<td>rs1625895</td>
<td>7518840</td>
<td>t</td>
<td>c</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.006</td>
<td>1.75E-60</td>
<td>1.17E-14</td>
<td>-0.052</td>
</tr>
<tr>
<td>5</td>
<td>Full model</td>
<td>1–4,6</td>
<td>rs6259</td>
<td>7477252</td>
<td>a</td>
<td>g</td>
<td>0.11</td>
<td>0.026</td>
<td>0.007</td>
<td>0.0001</td>
<td>1.46E-07</td>
<td>0.0372</td>
</tr>
<tr>
<td>6</td>
<td>Full model</td>
<td>1–5</td>
<td>rs10432029</td>
<td>7331393</td>
<td>a</td>
<td>g</td>
<td>0.79</td>
<td>0.0136</td>
<td>0.006</td>
<td>0.01</td>
<td>7.52E-16</td>
<td>0.0446</td>
</tr>
<tr>
<td>7</td>
<td>Conditional</td>
<td>1–6</td>
<td>rs9016175</td>
<td>7425536</td>
<td>a</td>
<td>g</td>
<td>0.05</td>
<td>-0.057</td>
<td>0.05</td>
<td>1.46E-07</td>
<td>5.2E-12</td>
<td>-0.10</td>
</tr>
<tr>
<td>8</td>
<td>Conditional</td>
<td>1–6</td>
<td>rs8077824</td>
<td>7588951</td>
<td>a</td>
<td>g</td>
<td>0.02</td>
<td>0.075</td>
<td>0.018</td>
<td>4.58E-05</td>
<td>0.01</td>
<td>0.0451</td>
</tr>
<tr>
<td>9</td>
<td>Conditional</td>
<td>1–6</td>
<td>rs9303218</td>
<td>7339386</td>
<td>t</td>
<td>c</td>
<td>0.77</td>
<td>0.026</td>
<td>0.006</td>
<td>9.89E-06</td>
<td>1.21E-11</td>
<td>0.0344</td>
</tr>
</tbody>
</table>

All SNPs are on the+strand and positions are based on build 36. EAF = 'effect allele frequency'. Beta units are per-allele effect estimates in natural log transformed nmol/L. Full model' SNPs were all included in a single regression model, where the effect estimates for each SNP are adjusted for the effect of the others in the model. 'Conditional' SNPs are SNPs with low pair-wise LD (HapMap r$^2$<0.01) that were identified after conditioning on the full model SNPs.

doi:10.1371/journal.pgen.1002805.t002
additional trans signals identified post conditional analysis was
15.6% in men and 8.4% in women. The SHBG locus accounted for
10% and 6.6% of the genetic variance in men and women, respectively, with the lead SNP explaining most of the genetic variation at 7.8% for men and 3.3% for women. Thus additional signals at the SHBG locus identified through conditional analyses approximately doubled the variance of the trait explained. While we provide evidence for multiple variants associated with SHBG concentrations, further studies are needed to pinpoint the causal loci and functional variants. For the 11 regions outside the SHBG locus, most have biologically plausible related genes within 300 kb.

Biology of Plausible Genes near Identified SNPs
Several genes near the identified SNPs regulate sex steroid production and function. The NR2F2 locus (15q26.2) encodes a nuclear receptor important in testicular Leydig cell function, the primary source of gonadal testosterone production [15], and has been linked to male infertility [16]. NR2F2 has also been associated with estrogen receptor alpha (ERα) signaling and may influence hormone responsivity in breast cancer [17]. PRMT6 (1p13.3) also encodes a nuclear receptor regulatory protein that mediates estrogen signaling as a co-activator of the estrogen receptor [18]. LHCGR (2p16.3) encodes the luteinizing hormone receptor which was associated with polycystic ovary syndrome (PCOS) in a recent GWAS [19,20]. PCOS is both a reproductive and metabolic disorder characterized by higher testosterone serum concentrations as well as an increased prevalence of obesity, insulin resistance, and T2D in women. Inappropriate secretion of luteinizing hormone leads to increased ovarian production of testosterone. Coincident lower SHBG concentrations contribute to increased bioavailable testosterone concentrations and the expression of both reproductive and metabolic phenotypes in PCOS [21,22,23].

The SLCO1B1 locus encodes a liver-specific transporter of thyroid hormone as well as estrogens which impact liver production of SHBG [24]. JMJD1C (10q21.3), also known as TRIP 8 (thyroid hormone receptor interactor protein 8 [25]), may impact SHBG concentrations via thyroid hormone effects on liver protein production. Thyroid hormone may alter SHBG production through effects on T3 levels which is known to regulate SHBG transcription [26,27].

Many of the genes identified are involved in carbohydrate and lipid metabolism and liver function. The GCKR locus (2p23.3) encodes a protein that regulates glucokinase activity and has been associated with T2D in several ethnic populations [28,29,30,31]. GCKR has been associated with metabolic and inflammatory traits including triglyceride concentrations and other lipid fractions [30,32], fasting plasma glucose [33,34], insulin concentrations, uric acid, c-reactive protein (CRP), and non-alcoholic fatty liver disease which are all characteristic of the metabolic syndrome and T2D [28,35,36,37,38,39,40,41,42]. The SLCO1B1 locus (12p12.1) codes for a protein, hepatocyte protein anion-transporting polypeptide 1B1, involved in liver metabolism of both endogenous and exogenous compounds [43]. Consistent with SLCO1B1’s role in liver metabolism, the same SNP (rs4149056) has been associated with circulating bilirubin concentrations in previous GWAS [44]. BAIAP2L1 (7q21.3) encodes a protein important in cytoskeleton organization [45] that has been associated with the inflammatory marker CRP in patients with arthritis [46]. BAIAP2L1 is also known as TRK5 [insulin receptor tyrosine kinase substrate] which

Figure 3. Allelic heterogeneity at the SHBG gene locus. There was significant allelic heterogeneity at the SHBG gene locus. The nine independent signals identified in the SHBG gene are shown in relation to their position within the gene. All positions based on build 36. Not all genes are shown.
doi:10.1371/journal.pgen.1002805.g003
is involved in insulin receptor signaling [47] and may relate to insulin resistant states including obesity and T2D [48,49, 50,51,52,53,54]. We conducted a targeted analysis of PPARδ, a gene that influences SHBG gene expression in the liver [1,55] and is associated with T2D [56,57]. Our analysis identified one significant SNP (rs2920502, p = 9.9 × 10^-5) and a second nominally significant SNP (rs13081389, p = 0.01) at PPARδ. Some of the identified genes involved in hepatic metabolism of lipids and carbohydrates may be affect SHBG concentrations indirectly through effects on the SHBG transcription regulator HNF4α although HNF4α itself was not identified in this meta-analyses [27,58,59,60].

The UGT2B15 locus (4q13.2) was significantly associated with SHBG concentrations in men but not women in this meta-analysis. UGT2B15 belongs to a family of genes (the UGT2B gene family) that code for enzymes involved in the metabolism of sex hormones through glucuronidation which allows for excretion of sex steroids through the kidney and the gut via bile excretion [61,62], primary clearance mechanisms for sex steroids [63]. UGT2B15 is involved in the conjugation and inactivation of testosterone [64]. An association between rs293428 in the UGT2B15 locus and circulating SHBG concentrations in men is supported by a previous study demonstrating that a non-synonymous SNP in UGT2B15 (rs1902023; D85Y) is associated with serum SHBG concentrations in younger adult men [65]. UGT2B15 is thought to play a significant role in local tissue inactivation of androgens in androgen dependent prostate cancer [66,67]. The mechanism behind the influence of genetic variants in UGT2B15 on SHBG concentrations is unknown, but one may speculate that UGT2B15 affects the local androgenic environment in selected tissues, which in turn results in regulation of SHBG concentrations.

In addition to UGT2B15, three other genes near the identified SNPs are associated with carcinogenesis, particularly in the prostate and breast. ZBTB10 (8q21.13) has been linked to breast cancer [68]. In breast cancer cell lines ZBTB10 is suppressed by ROS-microRNA27a thereby enhancing ERα expression and mediating estrogen effects [17]. The ZNF652 (17q21.32) locus codes for a DNA binding protein thought to act as a tumor suppressor gene in breast cancer [69,70,71] that is also co-expressed with the androgen receptor in prostate cancer [72]. TDGF3, teratocarcinoma derived growth factor 3, is the only significant region identified on the X chromosome (Xq22.3). TDGF3 is a pseudogene of TDGF1 located on chromosome 3p23-p21 that has been associated with testicular germ cell tumors [73].

Strengths and Limitations

This GWAS meta-analysis incorporated data from approximately 22,000 men and women from 16 epidemiologic cohorts. The overall size of the study yields power but the meta-analysis of data from different epidemiologic studies requires the inclusion of different laboratory methods. The different studies used a variety of assays to measure serum SHBG concentrations although the vast majority were immunoassays (Tables S1 and S2, Text S1) with similar methodologies. Variation introduced by the use of different SHBG assays would result in loss of statistical power and likely bias toward the null. Additionally, the majority of women were post-menopausal as ascertained by self-report in all studies (Table S1). SHBG concentrations, like testosterone, decline only slightly across the menopause [74] so adjustment for menopause status is not necessary. SHBG may also increase with ovulation and be slightly higher in the luteal versus the follicular phase of the menstrual cycle in premenopausal women, but most studies did not collect data on menstrual phase at the time of SHBG measurement so adjustment for menstrual phase was not possible [75]. Finally, individuals were not excluded based on health status, therefore some individuals with chronic conditions that may affect hepatic production of or clearance of proteins including SHBG such as liver disease, renal disease, or severe malnutrition, may have been included in this analysis.

Conclusion

SHBG synthesis in the liver is known to be affected directly or indirectly by estrogens, androgens and thyroid hormones and has been observed to be inversely associated with the higher insulin concentrations characteristic of insulin resistant states such as T2D [1,6]. In summary, the results of this GWAS reflect these influences. Three regions map to proteins related to hepatic function [12p12.1-SLC01B1 [76], 2p23.3-GCKR [77] and 10q21.3-JMJD1C [77]]. In addition, 2p23.3-GCKR and 7q21.3-BAIP2L1 (alias insulin receptor tyrosine kinase substrate [IRTKS]) are involved in susceptibility to T2D [48] and insulin signaling [47], respectively. Two signals also mapped to loci involved in thyroid hormone regulation (10q21.3-JMJD1C and 12p12.1-SLC01B1). One signal mapped to the receptor for luteinizing hormone 2p16.3-LHGR [29], the hormone that stimulates testosterone production. Five regions mapped to genes previously implicated in androgen and estrogen signaling [1p13.3-PRAMT6 [18], 8q21.13-ZBTB10 [17], 12p12.1-SLC01B1 [76], 15q26.2-NR2F2 [70], 4q13.2-UGT2B15 [63]].

We have combined a conventional GWAS approach with detailed additional analyses, including sex stratification, conditional analysis and imputation from 1000 Genomes. Our results demonstrate that these approaches can lead to an appreciable gain in heritable variance explained. It does however highlight the complexity of elucidating individual variant causality through statistical approaches. In addition to the extensive allelic heterogeneity at the SHBG locus, our data identify loci with a role in sex steroid hormone metabolism, which may help elucidate the role of sex steroid hormones in disease, particularly T2D and hormone-sensitive cancers.

Methods

We performed a genome wide association study (GWAS) meta-analysis of 21,791 individuals (Table S1: 9,390 women, 12,401 men) from ten observational studies. Data from an additional six studies totaling 7,046 individuals (Table S2: 4,509 women; 2,537 men) were used for validation. The proportion of variance explained was estimated in an independent study (InCHIANTI, n = 1,129). The individual study protocols were approved by their respective institution’s ethics committee/institutional review board and all participants provided informed consent prior to participation. Individuals known to be taking hormonal contraceptives or hormone replacement therapy at time of SHBG measurement were excluded from analysis. Age, sex and body mass index (BMI) were included as covariates. After applying standard quality control measures, imputed genotypes were available for approximately 2.5 M SNPs. See Figure 2 for an overview of the analytic plan and the Text S1 for further information for individual studies included in this meta-analysis.

GWAS Conditional Meta-Analysis Steps

Conditional analysis #1. The initial starting point for the conditional analysis was the four SHBG locus SNPs that all showed low Hapmap LD (r^2<0.05) with each other: rs12150660 (lead SNP Table 1), rs6258 p = 2.7 × 10^-16, rs1625895 p = 1.2 × 10^-14 and rs3653894 p = 2.5 × 10^-11. Each cohort fitted a single
regression model, fitting SHBG concentrations against these four genome-wide significant SHBG locus SNPs (rs12150660, rs6258, rs1625895 and rs3853894), in addition to age, sex and BMI. After meta-analyzing the results from all cohorts, three of the SNPs retained genome wide significance when regressed against each other, with the fourth SNP narrowly missing that threshold (rs3853894, p = 4.1 × 10^{-6}).

**Conditional GWAS #1 (Table 1, conditional analysis).** We next performed a conditional GWAS meta-analysis, where each study included, as additional covariates to the original analysis plan, the ten genome-wide significant autosomal SNPs (the eight ‘Main’ signals from Table 1 and the two unique SHBG locus signals described above in addition to the lead SNP rs12150660: rs6258 and rs1625895). Three additional signals (independence based on HapMap r²<0.05) at the SHBG locus reached genome-wide significance (rs1641537 p = 7.8 × 10^{-12}, rs6259 p = 1.5 × 10^{-12} and rs10432029 p = 3 × 10^{-8}), giving a total of six independent signals in this gene region. In addition, two novel signals reached genome-wide significance in the conditional analysis, at 7q21.3 (rs3779195 p = 1 × 10^{-8}) and 2p16.3 (rs10454142 p = 3 × 10^{-8}). After replication, only rs3779195 at the BAAP2L1 locus retained genome-wide significance.

**Conditional analysis #2 (Table 2, full model).** Given the six signals observed at the SHBG locus (three through conditional analysis #1 rs12150660, rs6258, rs1625895, three through LD estimates from conditional GWAS #1: rs1641537, rs6259, rs10432029), we sought to confirm which of these six were truly independent by a second round of conditional analysis. All discovery and replication cohorts fitted a single regression model of the six SNPs (SNPs # 1–6, Table 2) against SHBG concentrations, using the same parameters and covariates as conditional analysis #1. Four of the six SNPs (#1–4: rs12150660, rs6258, rs1641537, and rs1625895) retained genome-wide significance when conditioned against each other, with two showing nominal evidence of association (SNP #5 rs6259, p = 0.0001; SNP #6 rs10432029, p = 0.01).

**Conditional GWAS #2 (Table 2, conditional model).** Finally, we performed a second conditional GWAS analysis, adjusting for the six SHBG locus SNPs which had evidence of association from conditional analysis #2. All the discovery cohorts were used in this analysis, in addition to three replication cohorts (total sample size 24,354). This analysis revealed evidence for a further three independent signals at the SHBG locus (based on HapMap r²<0.01), SNP #7 rs9901675 p = 1.5 × 10^{-7}, SNP #8 rs8077824 p = 4.6 × 10^{-5}, and SNP #9 rs9393218 p = 9.9 × 10^{-8}.

**Sensitivity Analysis—Allelic Heterogeneity at the SHBG Locus**

We performed a sensitivity analysis using samples from the 1966 Northern Finland Birth Cohort (NFBC1966) study to further investigate allelic heterogeneity at the SHBG locus (Text S1). The conditional meta-analysis showed evidence for up to nine signals at the SHBG locus, but it is possible that these signals could be explaining a much smaller number of causal variants in the region. Since 1000 Genomes imputation allows us to assess the genetic variation associated with a phenotype across a much denser set of markers, it increases our power to detect allelic heterogeneity within a region. Therefore, 1000 Genomes imputation was carried out on all the samples in the NFBC1966 study and forward selection was used to identify the set of SNPs that best explain the variation in the SHBG phenotype. 1000 Genomes imputation was carried out using IMPUTE2. The mean genotype probabilities for each SNP were calculated and used in the model selection step. Only SNPs 250 kb upstream and 250 kb downstream from the SHBG locus (7283453–7786700 bp) were used in the analysis. All SNPs with MAF <0.1% or an imputation quality score less than 0.4 were excluded from the analysis. In total, 1978 SHBG region SNPs measured or imputed in 4467 samples from the NFBC1966 study were used in the sensitivity analysis. Forward selection was implemented in R (version 2.13.0) using the stepAIC package to estimate the Akaike Information Criterion (AIC), an inclusion parameter. Given the high degree of correlation between the SNPs in this region, we increased the penalty |k| on the number of terms included in the model to 12 (where it is usually two), to minimize possible over fitting. The final model included seven SNPs, adjusted for sex and BMI.

**Pathway Analysis**

We examined potential interactions among the proteins encoded by the SHBG locus and the proteins encoded by the 11 genes (ZBTB10, TDFG1, ZNF632, PRMT6, JMJDC1, GCKR, BAAP2L1, LHCGR, SLC01B1, UGT2B15, NR2F2) closest to the 11 identified SNPs using pathway analysis with Search Tool for the Retrieval of Interaction Genes/Proteins (STRING) Pathways Analysis (www.string-db.org). The interactions explored by STRING include direct (physical) and indirect (functional) associations. We then expanded the analysis to examine protein interactions among the SHBG gene and the proteins encoded by 67 genes within 500 kb of the 11 identified SNPs.

**Targeted Candidate Gene Analysis**

We conducted targeted analysis of two strong candidate genes, hepatocyte nuclear factor-4α (HNF4α) and peroxisome-proliferating receptor γ (PPARγ). Statistical significance thresholds were set correcting for the number of SNPs tested in each gene region (±100 kb).

**Supporting Information**

**Table S1** Characteristics of 21,791 individuals from 10 discovery cohorts included in the meta-analysis. (DOC)

**Table S2** Characteristics of 8,175 individuals from the six cohorts included in the validation analysis (WHI, CARDIA, Prospect-EPIC, MrOs, NHS, YFS) and the independent cohort used to estimate the proportion of genetic variance explained by the identified SNPs (IntChianti). (DOC)

**Table S3** Hapmap (release 22) linkage disequilibrium estimates for the nine SHBG locus single nucleotide polymorphisms. (DOC)

**Text S1** Supplementary Methods with Specific Cohort Information. (DOC)

**Acknowledgments**

We are grateful to all study participants and staff in our participating studies. Full study acknowledgements are available in Text S1.

**Author Contributions**

Conceived and designed the experiments: FH de Jong, O Raitakari, A Teumer, C Olsson, JM Maruthi, JRB Perry. Analyzed the data: JRB Perry, AD Covelli, R Haring, M Wellons, D Vaidya, T Lehtimäki, S Keidson, KL Lunetta, C He, M-R Jarvelin. Writes the paper: AD Covelli, R Haring, M Wellons, D Vaidya, T Lehtimäki, S Keidson, KL Lunetta, C He, TM Frayling, A Murray, S Franks, M-R Jarvelin, FH de
References


