Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk


Published in:
Human Molecular Genetics

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright the authors 2014. This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk


See Supplementary Material for full list of members.

Received: April 26, 2014. Revised: October 13, 2014. Accepted: October 24, 2014.

© The Author 2014. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, 2Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation, and School of Biomedical Science and, 3Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, QLD, Australia, 4Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, 5Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care and 6Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, 7Department of Clinical Genetics, St George’s Hospital Medical School, London, UK, 8Grupo de Investigación Citogenética, Filogenia y Evolución de Poblaciones, Universidad del Tolima, Ibagué, Tolima, Colombia, 9Department of Biological Anthropology, College of Humanities and Educational Sciences, University of the Republic, Montevideo, Uruguay, 10Grupo de Medicina Xenómica, Fundación Galega de Medicina Xenómica (SERGAS) and CIBERER, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, 11Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, KSA, 12Genetic Predisposition to Colorectal Cancer Group, Gastrointestinal & Pancreatic Oncology Team, IDIBAPS/CIBERehd/Hospital Clinic, Centre Esther Koplowitz (CEK), Barcelona, Spain, 13Universidad Autónoma de Nuevo León, Pedro de Alba s/n, San Nicolás de Los Garza, Nuevo León, Mexico, 14Hospital A.C. Camargo, São Paulo, Brazil, 15Department of Genetics, Portuguese Oncology Institute, Porto, Portugal, 16Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal, 17Genome Center and Department of Biochemistry and Molecular Medicine, University of California, Davis, CA, USA, 18Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA, 19Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China, 20Hunter Medical Research Institute and, 21Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia, 22Centre for Information Based Medicine and School of Biomedical Science and Pharmacy, 23Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, 24Centre for Information Based Medicine and School of Medicine and Public Health, 25Faculty of Health, Centre for Information Based Medicine and the Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy and, 26Faculty of Health, School of Medicine and Public Health, University of Newcastle, NSW, Australia, 27Division of Hematology/Oncology, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA, 28University Hospital Erlangen, Friedrich–Alexander University Erlangen-Nuremberg, Erlangen, Germany, 29Gynaecology Research Unit and, 30Clinics of Gynaecology and Obstetrics, Hannover Medical School, Hannover, Germany, 31Dept. of Gynaecology, Friedrich Schiller University Jena, Jena, Germany, 32Vesalius Research Center, VIB, Leuven, Belgium, 33Department of Oncology, Laboratory for Translational Genetics, 34Division of Gynaecological Oncology, Department of Oncology, University Hospital Leuven, KU Leuven, Belgium, 35Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, 36Division of Epidemiology, Department of Health Science Research and, 37Departments of Laboratory Medicine and Pathology, and Health Science Research, Mayo Clinic, Rochester, MN, USA, 38Department of Clinical Science, Centre for Cancerbiomarkers, The University of Bergen, Norway, 39Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway, 40Department of Women’s and Children’s Health, 41Department of Molecular Medicine and Surgery and 42Department of Medical Epidemiology and Biostatistics, Karolinska
Abstract

Common variants in the hepatocyte nuclear factor 1 homeobox B (HNF1B) gene are associated with the risk of Type II diabetes and multiple cancers. Evidence to date indicates that cancer risk may be mediated via genetic or epigenetic effects on HNF1B gene expression. We previously found single-nucleotide polymorphisms (SNPs) at the HNF1B locus to be associated with endometrial cancer, and now report extensive fine-mapping and in silico and laboratory analyses of this locus. Analysis of 1184 genotyped and imputed SNPs in 6608 Caucasian cases and 37 925 controls, and 895 Asian cases and 1968 controls, revealed the best signal of association for SNP rs11263763 (P = 8.4 × 10^-14, odds ratio = 0.86, 95% confidence interval = 0.82–0.89), located within HNF1B intron 1. Haplotype analysis and conditional analyses provide no evidence of further independent endometrial cancer risk variants at this locus. SNP rs11263763 genotype was associated with HNF1B mRNA expression but not with HNF1B methylation in endometrial tumor samples from The Cancer Genome Atlas. Genetic analyses prioritized rs11263763 and four other SNPs in high-to-moderate linkage disequilibrium as the most likely causal SNPs. Three of these SNPs map to the extended HNF1B promoter based on chromatin marks extending from the minimal promoter region. Reporter assays demonstrated that this extended region reduces activity in combination with the minimal HNF1B promoter, and that the minor alleles of rs11263763 or rs8064454 are associated with decreased HNF1B promoter activity. Our findings provide evidence for a single signal associated with endometrial cancer risk at the HNF1B locus, and that risk is likely mediated via altered HNF1B gene expression.

Introduction

Endometrial cancer is the most common type of uterine cancer, and the fourth most diagnosed cancer in European and North American women (http://globocan.iarc.fr/). Traditionally, this cancer is divided into two etiological types (1): hormonally driven Type 1, endometrioid histology subtype with 'good' prognosis (~80% of cases), and Type 2, non-endometrioid largely...
serous or clear cell subtypes with poor prognosis. Recently, in-depth studies by The Cancer Genome Atlas (TCGA) have identified four distinct tumor categories with different prognostic characteristics, namely ‘copy number high’, ‘copy number low’, ‘POLEultramutated’ and ‘microsatellite instability hypermutated’ (2). We have previously identified single-nucleotide polymorphisms (SNPs) associated with endometrial cancer risk at the hepatocyte nuclear factor 1 homeobox B (HNF1B) locus using a genome-wide association study (GWAS) approach (3). The most significantly associated SNP was rs4430796 located in HNF1B intron 2, with the minor ‘G’ allele protective for endometrial cancer (3).

HNF1B is a member of the homeodomain-containing superfamily of transcription factors (TFs), and SNPs at this locus are already known to be associated with risk of Type II diabetes (4), prostate cancer (4–9) and two different ovarian cancer subtypes (10,11). However, fine-mapping studies have revealed a complex genetic architecture at the HNF1B locus, demonstrated by lead SNPs and the direction of genetic effects being inconsistent between cancer types (Table 1). For example, in prostate cancer, the signal is explained by a five-SNP haplotype that includes SNPs from two peaks of association (5) in HNF1B intron 2 (lead SNP rs4430796) and intron 4 (lead SNP rs4794758) (12). For ovarian cancer subtypes, SNP rs757210, in high linkage disequilibrium (LD) with rs4430796, was shown to be associated with decreased risk of clear cell ovarian cancer but increased risk of serous ovarian cancer (10,11). Signals were subsequently refined to rs11651755 in intron 1, with the minor ‘G’ allele protective for endometrial cancer (3).

Various analyses have been undertaken to assess the relationship between HNF1B locus cancer risk SNPs and altered regulation of HNF1B mRNA expression. Expression quantitative trait loci (eQTL) analysis indicates that rs4430796 is associated with altered HNF1B mRNA expression in lymphoblastoid cell lines generated from cord blood or circulating lymphocytes (3), and also in benign prostate tissue (13). However, SNP rs757210 in high LD with rs4430796 was not associated with HNF1B expression in normal ovarian tissue (10). Instead, this SNP was determined to be a methylation eQTL (mQTL), associated with HNF1B promoter methylation in serous ovarian tumors (10,11). In contrast, no such association is indicated for clear cell ovarian tumors, which mostly present with a CpG island methylator phenotype (CIMP) but are nevertheless unmethylated at the HNF1B promoter (11).

Here, we report the fine-scale mapping of the HNF1B locus incorporating data for 1184 genotyped and imputed SNPs in 6608 endometrial cancer cases and 37 925 controls of European ancestry, and analyses aimed at exploring the function of the most likely causal variants. Our results provide evidence for a single signal associated with endometrial cancer risk at the HNF1B locus, and that risk is likely mediated via altered HNF1B gene expression.

### Results

**Fine-mapping and association analysis reveals one independent signal for endometrial cancer**

Meta-analysis of the 1184 HNF1B region SNPs genotyped or imputed in the four Caucasian datasets (iCOGS fine-mapping, Australian National Endometrial Cancer Study (ANECS), Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) and National Study of the Genetics of Endometrial Cancer (NSECWGwas)) and passing our quality control measures identified 18 SNPs that reached genome-wide significance (P < 5.0 × 10⁻⁸) (Table 2; results for individual sample sets provided in Supplementary Material, Table S2). The best overall signal was observed for imputed SNP rs11263763 [P = 8.4 × 10⁻¹⁴, odds ratio (OR) = 0.90], located in HNF1B intron 1 (Fig. 1A; Supplementary Material, Table S3). All 17 additional SNPs reaching genome-wide significance were moderately to highly correlated (r² = 0.57–0.95) with rs11263763, including the original endometrial cancer GWAS SNP rs4430796 (r² to rs11263763 = 0.95, P = 9.7 × 10⁻¹², OR = 0.87), and the best SNP genotyped in all four datasets rs7501939 (r² to rs11263763 = 0.67, P = 3.7 × 10⁻⁹, OR = 0.88; Supplementary Material, Fig. S1). No SNP remained significant at P < 10⁻⁴ after analyses conditioning on rs11263763, indicating that there are no additional independent SNPs associated with endometrial cancer risk at this locus. Haplotype analysis in the iCOGS fine-mapping dataset (Table 3) confirmed that there was a single association signal arising from the set of SNPs in strong LD with genotyped SNPs rs11651755, rs8064454 and rs11651052; the three haplotypes containing the minor alleles of these SNPs were all similarly associated with endometrial cancer risk (P for the best haplotype = 8.1 × 10⁻⁶, OR = 0.88).

There was no significant heterogeneity in risk between studies for the best genotyped or imputed SNPs (Table 4; Fig. 1B). The OR for rs11263763 in the Asian SECCG dataset was non-significant (P = 5.7 × 10⁻¹⁰, OR = 0.96), although the power was low to detect an effect equivalent to that seen for the Caucasian datasets given the sample size (834 cases and 1936 controls) and lower minor allele frequency (MAF) (0.267) (see Materials and Methods). For both the Caucasian and Asian datasets high LD extends centromeric from rs11263763 to encompass part of intron 2, with a slightly larger LD block in the Asian dataset (7 versus 5 kb; Supplementary Material, Fig. S2): assuming the risk SNPs are the same in both populations, this indicates that the search for candidate causal SNPs should focus on the 5 kb region identified from analyses of Caucasian datasets. Meta-analysis of the five datasets (iCOGS

### Table 1. Existing evidence for HNF1B association, expression and methylation in prostate and ovarian cancers

<table>
<thead>
<tr>
<th>Disease</th>
<th>GWAS SNP</th>
<th>Location</th>
<th>Minor allele effect</th>
<th>Lead fine-mapping SNPs</th>
<th>r² to rs4430796</th>
<th>eQTL in normal/at risk tissue</th>
<th>mQTL in tumor tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate cancer</td>
<td>rs4430796 (4)</td>
<td>Intron 2</td>
<td>G ↓</td>
<td>rs7405696** (12)</td>
<td>0.71</td>
<td>↓ mRNA expression (13)</td>
<td>No information</td>
</tr>
<tr>
<td></td>
<td>rs11649743 (5)</td>
<td>Intron 4</td>
<td>A ↓</td>
<td>rs7497458** (12)</td>
<td>0.01</td>
<td>No information</td>
<td>No information</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>rs757210 (10)</td>
<td>Intron 2</td>
<td>A ↑</td>
<td>rs7405776 (11)</td>
<td>0.47</td>
<td>No change (10)</td>
<td>Tumor unmethylated—no reported association (11)</td>
</tr>
<tr>
<td></td>
<td>rs757210 (10)</td>
<td>Intron 2</td>
<td>A ↓</td>
<td>rs11651755 (11)</td>
<td>0.97</td>
<td>No change (10)</td>
<td></td>
</tr>
</tbody>
</table>

* r² to rs4430796 in the 1000 Genomes Pilot data; ** rs7405696 explains part of the risk at the HNF1B prostate cancer risk region 1, and rs7497458 explains all of the risk at risk region 2. Note that conditional analyses suggest a 5-SNP haplotype best captures the variation across this region, although not all of the prostate risk at the HNF1B locus is explained by this haplotype (12).*
<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Position</th>
<th>Alleles</th>
<th>iCOGS MAF</th>
<th>iCOGS Information</th>
<th>Caucasian only meta-analysis OR (95% CIs)</th>
<th>P-value</th>
<th>Caucasian/Asian meta-analysis OR (95% CIs)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11263763</td>
<td>36 103 565</td>
<td>A/G</td>
<td>0.47</td>
<td>0.96</td>
<td>0.86 (0.82, 0.89)</td>
<td>8.4 × 10⁻¹⁴</td>
<td>0.86 (0.83, 0.90)</td>
<td>2.7 × 10⁻¹²</td>
</tr>
<tr>
<td>rs11651052</td>
<td>36 102 381</td>
<td>G/A</td>
<td>0.47</td>
<td>1.00</td>
<td>0.86 (0.82, 0.89)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.87 (0.84, 0.91)</td>
<td>3.7 × 10⁻¹²</td>
</tr>
<tr>
<td>rs064454</td>
<td>36 101 586</td>
<td>C/A</td>
<td>0.47</td>
<td>1.00</td>
<td>0.86 (0.82, 0.89)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.87 (0.84, 0.91)</td>
<td>6.3 × 10⁻¹²</td>
</tr>
<tr>
<td>rs10908278</td>
<td>36 099 952</td>
<td>A/T</td>
<td>0.47</td>
<td>0.96</td>
<td>0.86 (0.83, 0.90)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.87 (0.84, 0.91)</td>
<td>6.3 × 10⁻¹²</td>
</tr>
<tr>
<td>rs11651755</td>
<td>36 099 840</td>
<td>T/C</td>
<td>0.48</td>
<td>1.00</td>
<td>0.86 (0.83, 0.90)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.87 (0.84, 0.91)</td>
<td>5.2 × 10⁻¹²</td>
</tr>
<tr>
<td>rs4430796</td>
<td>36 098 040</td>
<td>A/G</td>
<td>0.48</td>
<td>0.95</td>
<td>0.87 (0.83, 0.90)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.85, 0.92)</td>
<td>2.0 × 10⁻¹⁰</td>
</tr>
<tr>
<td>rs11263761</td>
<td>36 097 775</td>
<td>A/G</td>
<td>0.49</td>
<td>0.93</td>
<td>0.86 (0.83, 0.90)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.84, 0.91)</td>
<td>8.5 × 10⁻¹¹</td>
</tr>
<tr>
<td>rs7405696</td>
<td>36 102 035</td>
<td>C/G</td>
<td>0.43</td>
<td>1.00</td>
<td>0.87 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.85, 0.92)</td>
<td>4.0 × 10⁻¹⁰</td>
</tr>
<tr>
<td>rs12453443</td>
<td>36 103 121</td>
<td>G/C</td>
<td>0.43</td>
<td>0.96</td>
<td>0.87 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.85, 0.92)</td>
<td>6.8 × 10⁻¹⁰</td>
</tr>
<tr>
<td>rs75209</td>
<td>36 102 833</td>
<td>A/G</td>
<td>0.42</td>
<td>0.96</td>
<td>0.87 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.85, 0.92)</td>
<td>9.0 × 10⁻¹⁰</td>
</tr>
<tr>
<td>rs11263762</td>
<td>36 101 926</td>
<td>A/G</td>
<td>0.43</td>
<td>1.00</td>
<td>0.88 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.88 (0.85, 0.92)</td>
<td>6.1 × 10⁻¹⁰</td>
</tr>
<tr>
<td>rs2005705</td>
<td>36 096 300</td>
<td>G/A</td>
<td>0.45</td>
<td>1.00</td>
<td>0.88 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.85, 0.92)</td>
<td>3.0 × 10⁻⁹</td>
</tr>
<tr>
<td>rs12601991</td>
<td>36 101 633</td>
<td>T/G</td>
<td>0.42</td>
<td>1.00</td>
<td>0.88 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.85, 0.92)</td>
<td>1.5 × 10⁻⁹</td>
</tr>
<tr>
<td>rs9901746</td>
<td>36 103 149</td>
<td>A/G</td>
<td>0.43</td>
<td>0.96</td>
<td>0.88 (0.84, 0.91)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.85, 0.92)</td>
<td>2.8 × 10⁻⁹</td>
</tr>
<tr>
<td>rs11658063</td>
<td>36 103 872</td>
<td>G/C</td>
<td>0.39</td>
<td>0.96</td>
<td>0.88 (0.84, 0.92)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.85, 0.92)</td>
<td>7.4 × 10⁻⁹</td>
</tr>
<tr>
<td>rs11657964</td>
<td>36 100 767</td>
<td>G/A</td>
<td>0.40</td>
<td>1.00</td>
<td>0.88 (0.85, 0.92)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.86, 0.93)</td>
<td>3.6 × 10⁻⁸</td>
</tr>
<tr>
<td>rs7501939</td>
<td>36 101 156</td>
<td>C/T</td>
<td>0.40</td>
<td>1.00</td>
<td>0.88 (0.85, 0.92)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.86, 0.93)</td>
<td>3.8 × 10⁻⁸</td>
</tr>
<tr>
<td>rs4239217</td>
<td>36 098 987</td>
<td>A/G</td>
<td>0.40</td>
<td>1.00</td>
<td>0.88 (0.85, 0.92)</td>
<td>8.4 × 10⁻¹³</td>
<td>0.89 (0.86, 0.93)</td>
<td>5.0 × 10⁻⁸</td>
</tr>
</tbody>
</table>

*Build 37 position.

Major/minor alleles based on forward strand and MAF in Europeans.

MAF of iCOGS controls.

Average imputation information score for the fine-mapping iCOGS dataset, where SNPs with a score of ‘1’ are genotyped SNPs.

Caucasian-only case \( n = 6608 \), control \( n = 37,925 \): Caucasian/Asian dataset case \( N = 7442 \), control \( n = 39,861 \): Per allele OR for the minor allele relative to the major allele.

One-degree-of-freedom \( P \)-trend. The best imputed and best genotyped SNPs are noted in bold.

Figure 1. Regional association and forest plots for the HNF1B locus associated with endometrial cancer: (A) Locuszoom (14) plot of the log10 \( P \)-values for association between each SNP and endometrial cancer for the meta-analysis of the iCOGS fine-mapping dataset, and ANECS, SEARCH and NSECG GWAS datasets. The color of each point indicates the extent of LD with the top SNP rs11263763 (purple). Gene positions are given under the graph, and estimated recombination rates in cM/Mb are indicated by the blue line (right-hand scale). Genotyped SNPs are plotted as circles, and imputed SNPs as squares (info score \( \geq 0.7 \) for all plotted SNPs). The small peak of signal \( \sim 13 \) kb to the right of rs11263763 does not survive conditional analysis. (B) Forest plot of ORs for the GWAS and iCOGS fine-mapping datasets stratified by study and country for top SNP rs11263763 (study acronyms detailed in Supplementary Material, Table S1).

Downloaded from https://academic.oup.com/hmg/article-abstract/24/5/1478/794172
by Queen's University of Belfast user on 12 March 2018
Table 3. Association between haplotypes of genotyped SNPs in the iCOGS dataset and endometrial cancer risk

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Genotyped SNP</th>
<th>Haplotype Frequency</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base haplotype</td>
<td>11651506</td>
<td>0.406</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Haplotype 1</td>
<td>A C T A G G A</td>
<td>0.353</td>
<td>0.88 (0.83, 0.93)</td>
<td>8.1 × 10^{-8}</td>
</tr>
<tr>
<td>Haplotype 2</td>
<td>G C T A G G A</td>
<td>0.037</td>
<td>0.84 (0.73, 0.97)</td>
<td>1.6 × 10^{-4}</td>
</tr>
<tr>
<td>Haplotype 3</td>
<td>A T C C T A G</td>
<td>0.014</td>
<td>1.12 (0.91, 1.38)</td>
<td>2.8 × 10^{-4}</td>
</tr>
<tr>
<td>Haplotype 4</td>
<td>G T C C G G G</td>
<td>0.089</td>
<td>0.97 (0.89, 1.06)</td>
<td>5.7 × 10^{-3}</td>
</tr>
<tr>
<td>Haplotype 5</td>
<td>G T C C G A G</td>
<td>0.012</td>
<td>1.00 (0.80, 1.26)</td>
<td>9.8 × 10^{-4}</td>
</tr>
</tbody>
</table>

Simpliﬁed haplotypes are comprised of 7 of the 10 top genotyped SNPs in the iCOGS dataset: rs4239217 and rs11657964 cannot be separated from rs7501939, and rs7405696 cannot be separated from rs11263762. Our top SNP rs11263763 is in very high LD (r^2 = 0.87) with rs11651755, rs8064454 and rs11651052 (shown in bold).

The best associated haplotype is comprised of the minor alleles of the seven genotyped SNPs included in the simplified haplotype.

The minor alleles of risk-associated SNPs are associated with reduced HNF1B expression

Of the five prioritized SNPs, the top SNP rs11263763 and rs11651755 (r^2 = 0.91 to rs11263763 in TCGA dataset) were included in the Affymetrix 6.0 array used by TCGA to type their tumor samples. The remaining prioritized SNPs were well captured by rs11651755 (r^2 = 1.00 for rs11651052, rs8064464; r^2 = 0.96 for rs10908278). One additional SNP reaching genome-wide signiﬁcance for association with endometrial cancer risk (Table 2) was directly genotyped in the TCGA dataset (rs11658063, r^2 = 0.71 to rs11263763). There was evidence for association between genotype and HNF1B expression levels in endometrioid tumors for rs11263763 (P = 1.3 × 10^{-3}), rs11658063 (P = 5.0 × 10^{-3}) and marginally so for rs11651755 (P = 8.3 × 10^{-3}). We also tested the allelic effect of rs11263763 and rs11658063 on HNF1B expression in non-endometrioid tumors (total N = 52), and similarly identiﬁed eQTLs for both rs11263763 (P = 3.0 × 10^{-3}) and rs11658063 (P = 4.8 × 10^{-3}). However, these associations would not be considered statistically signiﬁcant after conservatively correcting for the total number of genes analyzed across the region, where P for signiﬁcance = 5.0 × 10^{-3} (0.05/10). In all instances, the minor allele was associated with decreased levels of HNF1B mRNA (rs11263763 Fig 2A, rs11658063 Fig 2B). SNP rs11658063 was well imputed in the Caucasian datasets (information score >0.94), but statistically is not a likely candidate causal SNP, with a likelihood over 13 000 times smaller than that of rs11263763 in the Caucasian meta-analysis (P = 1.41 × 10^{-5}). There was no evidence for association between genotypes of any of these three SNPs and expression of any of the other nine genes located within 1 Mb of HNF1B (data not shown).

Differential HNF1B isoform usage has been suggested to occur between benign and tumor prostate tissue (15). We also
investigated the association between rs11263763 genotype and HNF1B isoform expression and usage in the TCGA endometrial tumor sample. Three HNF1B isoforms were measured by TCGA, the presence of which was confirmed by our own mRNA analysis of endometrial cancer cell lines (Supplementary Material, Fig. S3): isoform A (uc010wdi.1), isoform B (uc002hok.3) and isoform C (uc010cve.1). Overall, there was no evidence for differential isoform usage by genotype ($P = 0.45$). The relationship between rs11263763 genotype and HNF1B expression level (decreased expression in ‘G’ allele carriers) was consistent across isoform A ($P = 2.2 \times 10^{-5}$) and isoform B ($P = 2.1 \times 10^{-7}$), but not with isoform C ($P = 5.8 \times 10^{-1}$), although this isoform was expressed at very low levels or absent in those samples assessed.

### Table 4. Best genotyped and imputed HNF1B SNPs associated with risk of endometrial cancer in four Caucasian and one Asian datasets

<table>
<thead>
<tr>
<th>Genotyped SNP: rs7501939</th>
<th>Position^b</th>
<th>Alleles^c</th>
<th>MAF Cases</th>
<th>MAF Controls</th>
<th>OR (95% CIs)^d</th>
<th>P-value^e</th>
<th>$P_{\text{heterogeneity}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12453443</td>
<td>36 103 156</td>
<td>A/G</td>
<td>0.37</td>
<td>0.40</td>
<td>0.90 (0.86-0.95)</td>
<td>5.4 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>rs8064454</td>
<td>36 101 586</td>
<td>G/A</td>
<td>0.35</td>
<td>0.39</td>
<td>0.83 (0.73-0.94)</td>
<td>3.3 x 10^{-3}</td>
<td></td>
</tr>
<tr>
<td>rs4430796</td>
<td>36 098 040</td>
<td>A/G</td>
<td>0.45</td>
<td>0.48</td>
<td>0.86 (0.78-0.94)</td>
<td>3.7 x 10^{-7}</td>
<td></td>
</tr>
</tbody>
</table>

^Best SNP genotyped in all four datasets.
^Build 37 position.
^Major/minor alleles based on forward strand and MAF in Europeans.
^Per allele OR for the minor allele relative to the major allele.
^One-degree-of-freedom $P_{\text{trend}}$.

### Table 5. Association signal for cases with endometrioid histology and non-endometrioid histology in the four Caucasian datasets

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Position^a</th>
<th>Alleles^b</th>
<th>iCOGS MAF^c</th>
<th>iCOGS Information^d</th>
<th>Endometrioid histology OR (95% CIs)^d</th>
<th>P-value^e</th>
<th>Non-endometrioid histology OR (95% CIs)^d</th>
<th>P-value^e</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11263763</td>
<td>36 103 565</td>
<td>A/G</td>
<td>0.47</td>
<td>0.96</td>
<td>0.86 (0.82, 0.89)</td>
<td>4.1 x 10^{-12}</td>
<td>0.85 (0.77, 0.95)</td>
<td>2.4 x 10^{-3}</td>
</tr>
<tr>
<td>rs11651052</td>
<td>36 102 381</td>
<td>G/A</td>
<td>0.47</td>
<td>1.00</td>
<td>0.86 (0.82, 0.90)</td>
<td>8.6 x 10^{-12}</td>
<td>0.85 (0.77, 0.94)</td>
<td>1.7 x 10^{-3}</td>
</tr>
<tr>
<td>rs8064454</td>
<td>36 098 040</td>
<td>G/A</td>
<td>0.49</td>
<td>0.93</td>
<td>0.87 (0.83, 0.91)</td>
<td>1.3 x 10^{-10}</td>
<td>0.85 (0.77, 0.94)</td>
<td>2.1 x 10^{-2}</td>
</tr>
</tbody>
</table>

^Build 37 position.
^Major/minor alleles based on forward strand and MAF in Europeans.
^MAF of iCOGS controls.
^Average imputation information score for the fine-mapping iCOGS dataset, where SNPs with a score of ‘1’ are genotyped SNPs.
^Endometrioid histology case $N = 5611$, Non-endometrioid histology case $N = 887$, control $N = 37 925$ for both analyses: Per allele OR for the minor allele relative to the major allele.
^One-degree-of-freedom $P_{\text{trend}}$. The top SNPs for each analysis are noted in bold.
No association between SNP rs11263763 and HNF1B methylation

There was no association between genotype and HNF1B CpG island methylation for rs11263763 (P = 0.42, Fig. 2C), or for rs11658063 (P = 0.42, Fig. 2D). Most of the TCGA samples (94%) were unmethylated (beta values < 0.2) at the 18 probes located within the HNF1B CpG region. We also assessed methylation of the mutL homolog 1 (MLH1) gene in tumor samples, as this is a marker of the CIMP-like phenotype in numerous cancers, including endometrial, colorectal and ovarian cancers (2,16). In the TCGA dataset of 196 tumors, there was no association between HNF1B expression and MLH1 methylation (P = 0.93) and no association between HNF1B genotype and MLH1 methylation (P = 0.58 for rs11263763, P = 0.22 for rs11658063). There was also no association between HNF1B genotype and MLH1 methylation in the independent sample of 182 ANECS endometrial cancer tumors (P = 0.91; assessed for rs4430796, r² = 0.95 to rs11263763). That is, endometrial tumors present with unmethylated HNF1B promoter status irrespective of CIMP phenotype, resembling the presentation observed for the clear cell ovarian cancer subtype (11).

The strongest candidate causal SNPs map to the extended HNF1B promoter region

The five SNPs most strongly associated with endometrial cancer cluster within a 5.5 kb region in HNF1B intron 1 (Fig. 3). Using Encyclopedia of DNA Elements (ENCODE) data, we show that three of these SNPs (rs11263763, rs11651052 and rs8064454) fall within the extended HNF1B promoter that is marked by H3K4Me3 and H3K4Me1, indicative of regulatory activity associated with promoters. This region also contains DNaseI hypersensitivity sites indicating open chromatin in multiple cell lines, including the endometrial cancer cell lines ECC1 and Ishikawa (Fig. 3). Furthermore, this region also covers a strong CpG island, and has a chromatin state in numerous ENCODE cell lines indicative of enhancer and promoter elements. While none of the 21 TFs tested to date in the ECC1 cell line bind in this region, several additional TFs do bind in other cancer and normal cell lines. All five SNPs are predicted to affect the ability of several TFs to bind DNA (Supplementary Material, Table S5). Notably, several of these TFs are implicated in endometrial cancer. This includes rs11263763 and rs11651755, both identified to be associated with HNF1B
expression in tumors, see above. Both SNPs are predicted to alter binding of p53, a prominent TF that plays a key role in response to DNA damage and other stress signals, and may have prognostic value in endometrial cancer (17). In addition, rs8064454 is predicted to create a binding site for zinc finger E-box-binding protein (ZEB) 1, a well-characterized transcriptional repressor (18) that has previously reported to be aberrantly expressed in aggressive endometrial cancers (19,20).

Two of the candidate causal SNPs reduce the extended HNF1B promoter activity

We used luciferase reporter assays to examine activity associated with the wild-type promoter region, and whether the risk-associated SNPs in the extended promoter region were associated with altered HNF1B promoter activity. Transfection of Ishikawa and EN-1078D cell lines showed that the minimal HNF1B promoter construct produced a significant increase in reporter gene activity above the empty pGL3 vector control (Fig. 4). However, the extended HNF1B promoter construct significantly reduced this basal promoter activity by 40–50%, suggesting the presence of a silencer element in the extended region. Notably, inclusion of the minor alleles of rs11263763 or rs8064454 were also generated. Cells were transiently transfected with each of these constructs and assayed for luciferase activity after 48 h. Error bars denote standard error of the mean (SEM) from three independent experiments. P-values were determined with a two-tailed t-test (*P<0.05, **P<0.01, ***P<0.001).

Discussion

Fine-mapping of the multi-cancer HNF1B locus on chromosome 17q12 has revealed the presence of one multivariant haplotype associated with the risk of endometrial cancer. The most significantly associated SNP rs11263763 is highly correlated with the original endometrial cancer hit at this locus, rs4430796 (3), an SNP also associated with the risk of prostate cancer and in high-to-moderate LD with risk SNPs for serous and clear cell ovarian cancers. Multiple independent HNF1B associations have now been reported for the lead SNPs in prostate cancer (in introns 1 and 4) (12), while associations are limited to a single peak in intron 3 for the serous ovarian cancer subtype, and a single peak in
intron 1 for the clear cell ovarian subtype (11). Our analyses refines the endometrial cancer association signal to a distinct peak in intron 1, and show that our top SNPs are associated with HNF1B expression in endometrial tumors, and are located within the extended HNF1B promoter that contains a negative regulatory element that inhibits gene expression.

HNF1B expression is altered in numerous cancers, with evidence to support a role as a tumor suppressor or oncogene depending on the tissue context. Down-regulation of HNF1B is associated with progression of hepatocellular carcinomas (21), and indicates poor prognosis of renal (22) and prostate (23) carcinomas. HNF1B expression has also been reported to be lower in primary serous ovarian tumors than in normal ovarian tissue (24). Epigenetic inactivation of HNF1B is seen in serous ovarian tumors, and has been detected in ovarian, colorectal, gastric and pancreatic cancer cell lines, suggesting that HNF1B promoter hypermethylation can be a feature of tumorigenesis (25).

Conversely, the HNF1B promoter is typically unmethylated and gene expression increased in clear cell ovarian tumors and cell lines compared with other ovarian cancer subtypes (11,26). HNF1B hypomethylation has recently been detected in additional clear cell histologies, including endometrial, cervical and renal clear cell cancers, suggesting HNF1B expression and promoter hypermethylation to be a general biomarker of cytoplasmic clearing (27). HNF1B over-expression in immortalized endometriosis epithelial cells (hypothesized cell of origin for clear cell ovarian cancer) led to altered morphology and multinucleation of cells (11), while siRNA knock-down of HNF1B led to the induction of apoptosis in clear cell ovarian cancer cells lines (26) and significantly inhibited the proliferation and anchorage-dependent colony formation in the prostate cancer cell lines LNCaP and RWPE1 (13). Additionally, a genome-wide screen of RNAi data generated for ~100 cell lines identified HNF1B as a major oncogene required for cancer cell survival (28).

Analyses by us and others indicate that HNF1B is the target gene for genetic risk associations with cancer in this region (3,10,11,13), although the mechanism of regulation mediated by risk SNPs is not necessarily the same between cancer subtypes. SNP rs4403796 is an eQTL (expression quantitative trait locus) associated with decreased HNF1B mRNA expression in benign prostate tissue (the at-risk tissue for prostate cancer) (3), while the serous ovarian cancer subtype lead risk SNP rs7405776 is an methylation quantitative trait locus (mQTL) associated with decreased expression in serous ovarian tumor tissue. At this point in time, neither eQTLs nor mQTLs have been reported for clear cell ovarian tumor tissue. TCGA datasets show the HNF1B promoter is unmethylated in both endometrioid endometrial and prostate tumors. For prostate cancer, no significant difference in HNF1B mRNA expression levels has been reported between malignant prostate tissue and between benign tissue (15), or observed from our analysis of tumor and normal prostate tissue from TCGA (data not shown). Further, although a shift in isoform usage was reported between benign tissue [predominantly isoform C, a transcriptional repressor (29)] and malignant tissue [predominantly isoform B, a transcriptional activator (29)] (15), this was not evident from our analysis of the larger prostate dataset from TCGA.

Our analyses of the TCGA and other data indicate that the effects of causal SNPs on endometrial cancer risk at this locus are more similar to those of prostate rather than ovarian cancer subtypes. There was no association between risk genotype and HNF1B promoter methylation as implicated for the serous cancer clear cell subtype. SNP rs11263763 is indicated as an eQTL in endometrial tumor tissue, with the minor (protective) allele associated with decreased HNF1B expression, although this SNP appears to have no effect on isoform usage as previously reported for prostate cancer. Importantly, our functional analysis showed that two of the three candidate causal SNPs located in the extended HNF1B promoter are associated with reduced promoter activity in vitro, suggesting that these SNPs are likely to be associated with reduced HNF1B expression in vivo. Further functional follow-up experiments focusing on the region encompassing this association peak, including additional SNPs belonging to our risk haplotype, will be required to confirm if any of the other prioritized likely causal SNP(s) exert additional effects on expression via alternative mechanisms (30). Such findings, once linked to genetic and regulatory data from multiple cancers, will provide a greater understanding of the mechanism by which the HNF1B genomic locus and the HNF1B protein mediate risks particularly of endometrial cancer, but also of different cancer subtypes. We also note the incomplete overlap between prioritized candidate causal SNPs identified as eQTLs in the TCGA dataset, and those shown to demonstrate altered function in vitro from our functional studies to date. It is likely that future fine-mapping studies that encompass direct genotyping of likely causal SNPs of interest in larger datasets of tumor and normal tissue will inform the role of eQTL data in the design of time-consuming functional analysis studies of candidate causal SNPs.

Building on recent findings reporting multiple shared cancer susceptibility loci (10,31–35), the knowledge that endometrial in addition to prostate, serous ovarian and clear cell ovarian cancer are associated with SNPs that influence HNF1B activity gives additional support for the concept of regulatory regions harboring multiple cancer risk SNPs that act in a tissue-specific manner. Further, these findings provide rationale for expansive multi-cancer studies of novel loci identified for any single cancer, including bioinformatically directed investigation of novel loci discovered for endometrial cancer in multiple other cancers. It will be relevant for such future genetic epidemiological studies to consider molecular stratification of all tumor types, since analyses documenting the genomic characteristics of endometrial and other solid tumors have shown that distinct molecular subgroups within endometrial cancer histological subtypes share genomic features with different subtypes of other hormonally related tumors (2). Together, such expansive cross-cancer studies may further our understanding of the different biological pathways that lead to cancer.

Materials and Methods

Fine-mapping dataset

The fine-mapping case dataset comprised 4402 women of European ancestry with a confirmed diagnosis of endometrial cancer (3535 with confirmed endometrioid histology), recruited via 11 separate studies in seven countries collectively called the Endometrial Cancer Association Consortium. The control dataset comprised 28,758 healthy female controls from the same countries, all participating in the Breast Cancer Association Consortium (BCAC) (31) or Ovarian Cancer Association Consortium (OCAC) (10) (see Supplementary Material, Information and Table S1). All cases and controls were genotyped at 211,155 SNPs using a custom Illumina Infinium iSelect array (iCOGS); arrays and control genotyping methods are summarized in (10,31–34), designed by the Collaborative Oncological Gene-environment Study (COGS). The iCOGS array includes 286 SNPs located 1 Mb upstream and downstream of the HNF1B (RefSeq NM_000458.2) gene, selected with the intention to carry out fine-mapping studies of this locus (34). See section entitled ‘HNF1B fine-mapping SNPs’ below for further information.
Caucasian GWAS datasets

**ANECS and SEARCH**

The results presented here are based on a re-analysis of our original GWAS dataset, including additional samples, all called using the Illuminus program (36). Cases comprised 1287 endometrioid subtype endometrial cancer cases from the ANECS (n = 606) and the UK SEARCH (n = 681) genotyped using Illumina 610K arrays (3). ANECS cases were compared with 3083 Australian controls recruited as part of the Brisbane Adolescent Twin Study (37,38) (n = 1846) and the Hunter Community Study (39) (n = 1237), also genotyped using Illumina Infinium 610k arrays. SEARCH cases were compared with 5190 individuals genotyped using Illumina Infinium 1.2M arrays as part of the Wellcome Trust Case Control Consortium (40).

**National Study of Endometrial Cancer Genetics Group**

In addition to the above samples we obtained genotype data from 919 endometrial cancer cases (795 with confirmed endometrioid histology) collected by the UK NSECG and genotyped using Illumina 660K arrays. These cases were compared with data generated for 895 controls drawn from the UK1/CORGI colorectal GWAS sample 504 515 SNPs. Regional imputation

As the aim of this study was to investigate the association signal around the HNF1B locus, we restricted our analyses to SNPs located within an ~1 Mb region surrounding HNF1B (Build37, chr17:35599377–36602919). To increase the number of SNPs in the analysis and provide identical coverage across the four Caucasian and one Asian datasets, we imputed genotypes for SNPs present in the 1000 Genomes dataset v3 (April 2012 release) which had not been genotyped in our studies using IMPUTE v2 (43) software. We allowed the IMPUTE software to select the most appropriate haplotypes from among the complete set of 1000 Genomes haplotypes (44). Imputation was conducted on inference panels based on the SNPs typed for each dataset (e.g. SNPs included on the iCOGS array, various Illumina arrays for the ANECS, SEARCH and NSECG GWAS and the Affymetrix 6.0 array for the SECGS GWAS). Imputation was conducted separately for the five datasets, and SNPs with imputation information score <0.7 and/or MAF <0.01 excluded prior to analysis. Following quality control 1184 genotyped and imputed SNPs were retained in all four Caucasian datasets. The most significant imputed SNP was individually genotyped in a subset of cases using standard protocols for the Fluidigm BioMark™ HD System (Fluidigm, South San Francisco, CA, USA) (Supplementary Material, Information) to confirm imputation accuracy, resulting in 99% concordance between the genotyped and imputed genotypes.

**Asian GWAS dataset**

**Shanghai Endometrial Cancer Genetic Study**

To assess LD structure of HNF1B SNPs in other populations, we analyzed data previously generated for a GWAS including 834 Asian endometrial cancer cases recruited to the Shanghai Endometrial Cancer Study (SECS) and 1936 controls who were recruited to the Shanghai Breast Cancer Study (SBCS; collectively termed SECGS here), genotyped using Affymetrix 6.0 arrays (42).

**HNF1B fine-mapping SNPs**

The SNPs included on the iCOGS chip for the fine-mapping of HNF1B were chosen by the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, to produce a set of 405 SNPs including: all known SNPs with MAF > 0.02 in Europeans (n = 255), SNPs with \( r^2 > 0.1 \) to two prostate cancer associated SNPs [rs11649743 and rs4430796 (also associated with endometrial cancer): \( r^2 = 0.36 \)], and a set of tagging SNPs for the LD blocks present across 150 kb of \( HNF1B \) around the locus, we restricted our analyses to SNPs which were present (i.e. a forward stepwise regression approach). Regional imputation

As the aim of this study was to investigate the association signal around the HNF1B locus, we restricted our analyses to SNPs located within an ~1 Mb region surrounding HNF1B (Build37, chr17:35599377–36602919). To increase the number of SNPs in the analysis and provide identical coverage across the four Caucasian and one Asian datasets, we imputed genotypes for SNPs present in the 1000 Genomes dataset v3 (April 2012 release) which had not been genotyped in our studies using IMPUTE v2 (43) software. We allowed the IMPUTE software to select the most appropriate haplotypes from among the complete set of 1000 Genomes haplotypes (44). Imputation was conducted on inference panels based on the SNPs typed for each dataset (e.g. SNPs included on the iCOGS array, various Illumina arrays for the ANECS, SEARCH and NSECG GWAS and the Affymetrix 6.0 array for the SECGS GWAS). Imputation was conducted separately for the five datasets, and SNPs with imputation information score <0.7 and/or MAF <0.01 excluded prior to analysis. Following quality control 1184 genotyped and imputed SNPs were retained in all four Caucasian datasets. The most significant imputed SNP was individually genotyped in a subset of cases using standard protocols for the Fluidigm BioMark™ HD System (Fluidigm, South San Francisco, CA, USA) (Supplementary Material, Information) to confirm imputation accuracy, resulting in 99% concordance between the genotyped and imputed genotypes.

**Association analysis**

The four imputed datasets were analyzed separately using unconditional logistic regression with a per-allele (1 degree of freedom) model using SNPTEST v2 (45). For the iCOGS dataset, analyses were performed adjusting for strata (six of the eight strata were defined by country, while the large UK dataset was divided into ‘SEARCH’ and ‘NSECG’) and including the first 10 principal components of the genomic kinship matrix, based on 37 000 uncorrelated iCOGS SNPs \( r^2 < 0.1 \), including \(~1000 selected as ancestry informative markers, using an in-house C++ program incorporating the Intel MKL libraries for eigenvectors (http://cge.medschl.cam.ac.uk/software/). One principal component was derived specifically for the Leuven (LES/LMBC) studies, for which there was substantial inflation not accounted for by the other principal components. The Caucasian GWAS datasets were analyzed as single stratum, with adjustment for the first two (ANECS and NSECG) and three (SEARCH) principal components.

Results (ORs) of the four studies were combined using standard fixed-effects meta-analyses. The I^2 statistic (46) was used to estimate the proportion of the variance due to between-study heterogeneity and the Q statistic to test for such heterogeneity. Analyses for all SNPs were repeated adjusting for the most significant SNP to assess whether multiple independent causal variants were present (i.e. a forward stepwise regression approach). The analyses were also repeated restricting the iCOGS and NSECG studies to those cases with endometrioid or non-endometrioid histology (the ANECS and SEARCH GWAS sample sets contained only endometrioid histology cases), and to iCOGS
cases and controls for whom BMI data were available. All statistical analyses used R software unless otherwise stated, and all statistical tests were two-sided. The association plot was produced using LocusZoom (14). LD between SNPs is reported as calculated for the HapMap3 (release 2) population (http://www.broadinstitute.org/mpg/snap/lidsearchpw.php). Haplotype analyses including the top genotyped SNPs in the iCOGS fine-mapping dataset were performed in Haplostats (http://www.mayo.edu/research/labs/statistical-genetics-genetic-epidemiology/software).

The power to detect an effect in the smaller Caucasian non-endometrioid tumor and Asian SECGS datasets, equivalent to that seen for the best SNP (rs11263763) in the main analysis including the four Caucasian datasets for all histologies, was calculated using QUANTO 1.1 (47). For the non-endometrioid dataset with an MAF of 0.47 in 887 cases and 37 925 controls, power to detect an equivalent effect was 87% at the 5% significance threshold and 22% at 10⁻³. For the SECGS dataset with an MAF of 0.27 in 834 cases and 1936 controls, power was 61% at the 5% significance threshold and 5% at 10⁻³.

Likelihood tests to select the most likely causal SNPs affecting endometrial cancer risk

To determine the most likely causative SNPs from among the top associated SNPs, the log-likelihoods of all tested SNPs were compared with that of the top SNP (rs11263763), using P-values from the overall (all-histologies) analysis in Caucasians. SNPs with log-likelihood ratios of <1 : 100 of being the top SNP were prioritized as potentially causal variants for follow-up in the bioinformatic and functional analyses (10,30,48).

Expression and methylation by genotype in endometrial tumors

To investigate in endometrial tumors the SNP effects previously demonstrated in benign prostate tissue (eQTL) and serous ovarian tumors (mQTL), we analyzed data from two different sources. TCGA: preprocessed SNP (Affymetrix 6.0 arrays), gene expression (RNA-Seq data generated using Illumina GAIIx and Illumina HiSeq platforms) and DNA methylation (Illumina Infinium HumanMethylation 450 Beadchips) data generated by TCGA for endometrial cancer tumor samples (2) were obtained through TCGA and the cBioPortal for Cancer Genomics (49,50)(Supplementary Material, Information). Analyses were restricted to samples of Caucasian ancestry with endometrioid subtype endometrial cancer, adjusting for copy number at the HNF1B locus. Associations between genotype and tumor HNF1B expression, HNF1B promoter methylation and tumor TCGA type were assessed by Kruskal–Wallis and Pearson correlation tests, with two-sided P-values <0.05 indicating a significant association.

ANECS: association between genotype at HNF1B SNP rs4430796 [generated through the original GWAS (3)] and tumor methylation at the MLH1 gene (a marker of the CIMP-like phenotype in endometrial cancer) (51) was assessed for 182 ANECS endometrial cancer cases for whom both data types were available.

Bioinformatic analysis to assess SNP functionality

Bioinformatic analyses to determine the most likely location and identity of putative causal SNPs that may influence the expression of HNF1B were conducted using a number of databases. Data produced by the ENCODE (52) project, indicating the location of open chromatin, DNA methylation, histone modification and TF binding in numerous cell lines including the endometrial cancer lines EEC1 and Ishikawa, were accessed through the UCSC Genome Browser (http://genome.ucsc.edu/ENCODE/). Multiple cell lines in addition to the endometrial cancer cell lines were included in the analysis to allow investigation of the range of possible potential regulatory mechanisms present across the HNF1B region. The is-rSNP software was used to predict which SNPs altered the ability of a TF to bind DNA (53). The is-rSNP program uses JASPAR and TRANSFAC databases to first determine if the two SNP alleles are predicted to localize in a potential TF binding site, based on binding scores computed using Position Weighted Matrices (PWM). For each potential TF, is-rSNP then calculates whether any of the two SNP alleles significantly alters the binding score.

Cell lines, plasmid construction and luciferase assays

Endometrial cancer cell lines Ishikawa and EN-1078D (kindly provided by Pamela Pollock, QUT, Brisbane) were grown in DMEM or DMEM:F12 medium, respectively, with 10% fetal calf serum and antibiotics. Cell lines were maintained under standard conditions routinely tested for Mycoplasma and short tandem repeat profiled. The HNF1B promoter-driven luciferase reporter constructs were generated by inserting a 908 bp (minimum promoter (Min prom), hg19; chr17:36104874–36105781) or 4651 bp fragment (extended promoter (Ext prom), hg19; chr17:36101131–36105781) with or without the minor alleles of rs11263763, rs11651052 or rs8064454 into the KpnI and HindIII sites of pGL3-basic. All HNF1B promoter sequences were commercially synthesized using GenScript (Life Research, Australia). Ishikawa and EN-1078D cells were transfected with equimolar amounts of luciferase reporter plasmids and 50 ng of pRLTK using Lipofectamine 2000. The total amount of transfected DNA was kept constant per experiment by adding carrier plasmid (pUC19). Luciferase activity was measured 48 h post-transfection using the Dual-Glo Luciferase Assay System on a Beckman-Coulter DTX-880 plate reader. To correct for any differences in transfection efficiency or cell lysate preparation, Firefly luciferase activity was normalized to Renilla luciferase. The activity of each test construct was calculated relative to an empty pGL3-basic construct, the activity of which was arbitrarily defined as 1.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

The authors thank the many individuals who participated in this study and the numerous institutions and their staff who supported recruitment, detailed in full in the Supplementary Material.

Conflict of Interest statement. None declared.

Funding

Fine-mapping analysis was supported by National Health and Medical Research Council (NHMRC) project grant (ID#1031333) to A.B.S., D.F.E. and A.M.D. Functional analysis was supported by NHMRC project grant (ID#1058415) to S.L.E., J.D.F. and A.M.D. A.B.S., P.M.W., G.W.M. and D.R.N. are supported by the NHMRC Fellowship scheme. D.F.E. is a Principal Research Fellow of Cancer Research UK. A.M.D. is supported by the Joseph Mitchell Trust. I.T. is supported by Cancer Research UK and the Oxford...
Comprehensive Biomedical Research Centre. Funding for the iCOGS infrastructure came from: the European Community’s Seventh Framework Programme under grant agreement no 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007 and C5047/A10692), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112—the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. ANECs recruitment was supported by project grants from the NHMR (ID#339435), the Cancer Council Queensland (ID#4196615) and Cancer Council Tasmania (ID#403091 and ID#457636). SEARCH recruitment was funded by a program grant from Cancer Research UK (C490/ A10124). Stage 1 and Stage 2 case genotyping was supported by the NHMRC (ID#552402, ID#1031339). Control data was generated by the Wellcome Trust Case Control Consortium (WTCCC), and a full list of the investigators who contributed to the generation of the data is available from the WTCCC website. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02—funding for this project was provided by the Wellcome Trust under award 085475. SNECC was supported by the EU FP7 CHIBCHA grant and CORGI by Cancer Research UK. Recruitment of the QMR Berghofer controls was supported by the NHMRC. The University of Newcastle, the Gladys M Brawn Senior Research Fellowship scheme, The Vincent Fairfax Family Foundation, the Hunter Medical Research Institute and the Hunter Area Pathology Service all contributed towards the costs of establishing the Hunter Community Study. The Bavarian Endometrial Cancer Study (BECS) was partly funded by the ELAN fund of the University of Erlangen. The Leuven Endometrium Study (LES) was supported by the Verelet Foundation for endometrial cancer. The Mayo Endometrial Cancer Study (MECS) and Mayo controls (MAY) were supported by grants from the National Cancer Institute of United States Public Health Service (R01 CA 092585 and R01 CA90899, R01 CA64277). The Breast Cancer Association Consortium (BCAC) is funded by Cancer Research UK (C1287/ A10118, C1287/A12014). The Ovarian Cancer Association Consortium (OCAC) is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PFD/RPCI.07), and the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge. Additional funding for individual control groups is detailed in the Supplementary Information.

References

22. Buchner, A., Castro, M., Hennig, A., Popp, T., Assmann, G.,
19. Singh, M., Spoelstra, N.S., Jean, A., Howe, E., Torkko, K.C.,
25. Terasawa, K., Toyota, M., Sagae, S., Ogi, K., Suzuki, H., Sonoda,
expression are associated with the development of prostate cancer.
BMC Cancer, 10, 315.
Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,
18. Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S.,
14. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S.,
16. Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,
expression are associated with the development of prostate cancer.
BMC Cancer, 10, 315.
Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,
18. Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S.,
14. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S.,
16. Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,
expression are associated with the development of prostate cancer.
BMC Cancer, 10, 315.
Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,
18. Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S.,
14. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S.,
16. Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit,