

## ORIGINAL ARTICLE

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

Jodie N. Painter<sup>1</sup>, Tracy A. O'Mara<sup>1</sup>, Jyotsna Batra<sup>2</sup>, Timothy Cheng<sup>4</sup>, Felicity A. Lose<sup>1</sup>, Joe Dennis<sup>5</sup>, Kyriaki Michailidou<sup>5</sup>, Jonathan P. Tyrer<sup>6</sup>, Shahana Ahmed<sup>6</sup>, Kaltin Ferguson<sup>1</sup>, Catherine S. Healey<sup>6</sup>, Susanne Kaufmann<sup>1</sup>, Kristine M. Hillman<sup>1</sup>, Carina Walpole<sup>2</sup>, Leire Moya<sup>2</sup>, Pamela Pollock<sup>3</sup>, Angela Jones<sup>4</sup>, Kimberley Howarth<sup>4</sup>, Lynn Martin<sup>4</sup>, Maggie Gorman<sup>4</sup>, Shirley Hodgson<sup>7</sup>, National Study of Endometrial Cancer Genetics Group (NSECG)<sup>4,†</sup>, CHIBCHA Consortium<sup>4,†</sup>, Ma. Magdalena Echeverry De Polanco<sup>8</sup>, Monica Sans<sup>9</sup>, Angel Carracedo<sup>10,11</sup>, Sergi Castellvi-Bel<sup>12</sup>, Augusto Rojas-Martinez<sup>13</sup>, Erika Santos<sup>14</sup>, Manuel R. Teixeira<sup>15,16</sup>, Luis Carvajal-Carmona<sup>4,8,17</sup>, Xiao-Ou Shu<sup>18</sup>, Jirong Long<sup>18</sup>, Wei Zheng<sup>18</sup>, Yong-Bing Xiang<sup>19</sup>, The Australian National Endometrial Cancer Study Group (ANECS)<sup>1,†</sup>, Grant W. Montgomery<sup>1</sup>, Penelope M. Webb<sup>1</sup>, Rodney J. Scott<sup>20,21,22</sup>, Mark McEvoy<sup>23</sup>, John Attia<sup>20,23</sup>, Elizabeth Holliday<sup>20,24</sup>, Nicholas G. Martin<sup>1</sup>, Dale R. Nyholt<sup>1</sup>, Anjali K. Henders<sup>1</sup>, Peter A. Fasching<sup>27,28</sup>, Alexander Hein<sup>28</sup>, Matthias W. Beckmann<sup>28</sup>, Stefan P. Renner<sup>28</sup>, Thilo Dörk<sup>29</sup>, Peter Hillemanns<sup>30</sup>, Matthias Dürst<sup>31</sup>, Ingo Runnebaum<sup>31</sup>, Diether Lambrechts<sup>32,33</sup>, Lieve Coenegrachts<sup>34</sup>, Stefanie Schrauwen<sup>34</sup>, Frederic Amant<sup>34</sup>, Boris Winterhoff<sup>35</sup>, Sean C. Dowdy<sup>35</sup>, Ellen L. Goode<sup>36</sup>, Attila Teoman<sup>35</sup>, Helga B. Salvesen<sup>38,39</sup>, Jone Trovik<sup>38,39</sup>, Tormund S. Njolstad<sup>38,39</sup>, Henrica M.J. Werner<sup>38,39</sup>, Katie Ashton<sup>21,25</sup>, Tony Proietto<sup>26</sup>, Geoffrey Otton<sup>26</sup>, Gerasimos Tzortzatos<sup>40</sup>, Miriam Mints<sup>40</sup>, Emma Tham<sup>41</sup>, RENDOCAS<sup>41,†</sup>, Per Hall<sup>42</sup>, Kamila Czene<sup>42</sup>, Jianjun Liu<sup>43</sup>, Jingmei Li<sup>43</sup>, John L. Hopper<sup>44</sup>, Melissa C. Southey<sup>45</sup>, Australian Ovarian Cancer Study (AOCS)<sup>1,46,†</sup>, Arif B. Ekici<sup>47</sup>, Matthias Ruebner<sup>28</sup>, Nicola Johnson<sup>48</sup>, Julian Peto<sup>51</sup>,

<sup>†</sup> 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.

Barbara Burwinkel<sup>52,54</sup>, Frederik Marme<sup>52,53</sup>, Hermann Brenner<sup>55,58</sup>, Aida K. Dieffenbach<sup>55,58</sup>, Alfons Meindl<sup>59</sup>, Hiltrud Brauch<sup>60</sup>, The GENICA Network<sup>56,60,61,62,63,64,†</sup>, Annika Lindblom<sup>41</sup>, Jeroen Depreeuw<sup>32</sup>, Matthieu Moisse<sup>32</sup>, Jenny Chang-Claude<sup>57</sup>, Anja Rudolph<sup>65</sup>, Fergus J. Couch<sup>37</sup>, Janet E. Olson<sup>36</sup>, Graham G. Giles<sup>44,66,67</sup>, Fiona Bruinsma<sup>66</sup>, Julie M. Cunningham<sup>37</sup>, Brooke L. Fridley<sup>68</sup>, Anne-Lise Børresen-Dale<sup>69,70</sup>, Vessela N. Kristensen<sup>69,70,71</sup>, Angela Cox<sup>72</sup>, Anthony J. Swerdlow<sup>49,50</sup>, Nicholas Orr<sup>50</sup>, Manjeet K. Bolla<sup>5</sup>, Qin Wang<sup>5</sup>, Rachel Palmieri Weber<sup>73</sup>, Zhihua Chen<sup>74</sup>, Mitul Shah<sup>6</sup>, Juliet D. French<sup>1</sup>, Paul D.P. Pharoah<sup>6</sup>, Alison M. Dunning<sup>6</sup>, Ian Tomlinson<sup>4</sup>, Douglas F. Easton<sup>5,6</sup>, Stacey L. Edwards<sup>1</sup>, Deborah J. Thompson<sup>6</sup>, and Amanda B. Spurdle<sup>1,\*</sup>

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

Institutet, Karolinska University Hospital, Stockholm, Sweden, <sup>43</sup>Human Genetics, Genome Institute of Singapore, Singapore, <sup>44</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health and <sup>45</sup>Department of Pathology, Genetic Epidemiology Laboratory, The University of Melbourne, Melbourne, VIC, Australia, <sup>46</sup>Peter MacCullum Cancer Centre, Melbourne, VIC, Australia, <sup>47</sup>Institute of Human Genetics, University Hospital, Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany, <sup>48</sup>Breakthrough Breast Cancer Research Centre, <sup>49</sup>Division of Genetics and Epidemiology and, <sup>50</sup>Division of Breast Cancer Research, Institute of Cancer Research, London, UK, <sup>51</sup>London School of Hygiene and Tropical Medicine, London, UK, <sup>52</sup>Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, <sup>53</sup>National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany, <sup>54</sup>Molecular Epidemiology, C080, <sup>55</sup>Division of Clinical Epidemiology and Aging Research, <sup>56</sup>Molecular Genetics of Breast Cancer and <sup>57</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, <sup>58</sup>German Cancer Consortium (DKTK), Heidelberg, Germany, <sup>59</sup>Division of Tumor Genetics, Department of Obstetrics and Gynecology, Technical University of Munich, Munich, Germany, <sup>60</sup>Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology Stuttgart, University of Tuebingen, Germany, <sup>61</sup>Institute for Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany, <sup>62</sup>Department of Internal Medicine, Evangelische Kliniken Bonn GmbH, Johanniter Krankenhaus, Bonn, Germany, <sup>63</sup>Institute of Pathology, Medical Faculty of the University of Bonn, Bonn, Germany, <sup>64</sup>Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Bochum, Germany, <sup>65</sup>Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany, <sup>66</sup>Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, VIC, Australia, <sup>67</sup>Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia, <sup>68</sup>Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA, <sup>69</sup>Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway, <sup>70</sup>Faculty of Medicine, The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, University of Oslo, Oslo, Norway, <sup>71</sup>Division of Medicine, Department of Clinical Molecular Oncology, Akershus University Hospital, Ahus, Norway, <sup>72</sup>Department of Oncology, Sheffield Cancer Research Centre, University of Sheffield, Sheffield, UK, <sup>73</sup>Department of Community and Family Medicine, Duke University School of Medicine, Durham, NC, USA, and <sup>74</sup>Division of Population Sciences, Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA

\*To whom correspondence should be addressed at: QIMR Berghofer Medical Research Institute, 300 Herston Rd, Herston, 4006, Queensland, Australia. Tel: +61 733620379; Fax: +61 733620105; Email: amanda.spurdle@qimrberghofer.edu.au

## 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 ( $P = 8.4 \times 10^{-14}$ , 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 for the clear cell ovarian cancer subtype, and rs7405776 in intron 3 for the serous subtype (11).

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 (NSECG GWASs)] and passing our quality control measures identified 18 SNPs that reached genome-wide significance ( $P < 5.0 \times 10^{-8}$ ) (Table 2; results for individual sample sets provided in Supplementary Material, Table S2). The best overall signal was observed for imputed SNP rs11263763 [ $P = 8.4 \times 10^{-14}$ , odds ratio (OR) = 0.86], located in *HNF1B* intron 1 (Fig. 1A; Supplementary Material, Table S3). All 17 additional SNPs reaching genome-wide significance were moderately to highly correlated ( $r^2 = 0.57$ – $0.95$ ) with rs11263763, including the original endometrial cancer GWAS SNP rs4430796 ( $r^2$  to rs11263763 = 0.95,  $P = 9.7 \times 10^{-12}$ , OR = 0.87), and the best SNP genotyped in all four datasets rs7501939 ( $r^2$  to rs11263763 = 0.67,  $P = 3.7 \times 10^{-9}$ , 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 =  $8.1 \times 10^{-6}$ , 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 SECGS dataset was non-significant ( $P = 5.7 \times 10^{-1}$ , 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

Disease	GWAS SNP	Location	Minor allele effect	Lead fine-mapping SNPs	$r^2$ to rs4430796*	eQTL in normal/at risk tissue	mQTL in tumor tissue
Prostate cancer	rs4430796 (4)	Intron 2	G ↓	rs7405696** (12)	0.71	↓ mRNA expression (13)	No information
	rs11649743 (5)	Intron 4	A ↓	rs4794758** (12)	0.01	No information	No information
Ovarian cancer							
Serous	rs757210 (10)	Intron 2	A ↑	rs7405776 (11)	0.47	No change (10)	↑ methylation (11)
Clear cell	rs757210 (10)	Intron 2	A ↓	rs11651755 (11)	0.97	No change (10)	Tumor unmethylated—no reported association (11)

\* $r^2$  to rs4430796 in the 1000 Genomes Pilot data; \*\* rs7405696 explains part of the risk at the *HNF1B* prostate cancer risk region 1, and rs4794758 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 2.** Genome-wide significant signal for all-histology endometrial cancer following fine-mapping meta-analysis of the HNF1B locus in four Caucasian and one Asian datasets

SNP ID	Position <sup>a</sup>	Alleles <sup>b</sup>	iCOGS MAF <sup>c</sup>	iCOGS Information <sup>d</sup>	Caucasian only meta-analysis OR (95% CIs) <sup>e</sup>	P-value <sup>f</sup>	Caucasian/Asian meta-analysis OR (95% CIs) <sup>e</sup>	P-value <sup>f</sup>
<b>rs11263763</b>	<b>36 103 565</b>	<b>A/G</b>	<b>0.47</b>	<b>0.96</b>	<b>0.86 (0.82, 0.89)</b>	<b><math>8.4 \times 10^{-14}</math></b>	0.86 (0.83, 0.90)	$2.7 \times 10^{-13}$
rs11651052	36 102 381	G/A	0.47	1.00	0.86 (0.82, 0.89)	$1.3 \times 10^{-13}$	0.87 (0.84, 0.91)	$3.7 \times 10^{-12}$
rs8064454	36 101 586	C/A	0.47	1.00	0.86 (0.82, 0.89)	$2.4 \times 10^{-13}$	0.87 (0.84, 0.91)	$6.3 \times 10^{-12}$
rs10908278	36 099 952	A/T	0.47	0.96	0.86 (0.83, 0.90)	$8.6 \times 10^{-13}$	0.87 (0.84, 0.91)	$6.3 \times 10^{-12}$
rs11651755	36 099 840	T/C	0.48	1.00	0.86 (0.83, 0.90)	$2.3 \times 10^{-12}$	0.87 (0.84, 0.91)	$5.2 \times 10^{-12}$
rs4430796	36 098 040	A/G	0.48	0.95	0.87 (0.83, 0.90)	$9.7 \times 10^{-12}$	0.88 (0.85, 0.92)	$2.0 \times 10^{-10}$
rs11263761	36 097 775	A/G	0.49	0.93	0.86 (0.83, 0.90)	$5.5 \times 10^{-12}$	0.88 (0.84, 0.91)	$8.5 \times 10^{-11}$
rs7405696	36 102 035	C/G	0.43	1.00	0.87 (0.84, 0.91)	$1.1 \times 10^{-10}$	0.88 (0.85, 0.92)	$4.0 \times 10^{-10}$
rs12453443	36 104 121	G/C	0.43	0.96	0.87 (0.84, 0.91)	$1.9 \times 10^{-10}$	0.88 (0.85, 0.92)	$6.8 \times 10^{-10}$
rs757209	36 102 833	A/G	0.42	0.96	0.87 (0.84, 0.91)	$1.9 \times 10^{-10}$	0.88 (0.85, 0.92)	$9.0 \times 10^{-10}$
rs11263762	36 101 926	A/G	0.43	1.00	0.88 (0.84, 0.91)	$1.9 \times 10^{-10}$	0.88 (0.85, 0.92)	$6.1 \times 10^{-10}$
rs2005705	36 096 300	G/A	0.45	1.00	0.88 (0.84, 0.91)	$4.1 \times 10^{-10}$	0.89 (0.85, 0.92)	$3.0 \times 10^{-9}$
rs12601991	36 101 633	T/G	0.42	1.00	0.88 (0.84, 0.91)	$4.2 \times 10^{-10}$	0.88 (0.85, 0.92)	$1.5 \times 10^{-9}$
rs9901746	36 103 149	A/G	0.43	0.96	0.88 (0.84, 0.91)	$6.6 \times 10^{-10}$	0.89 (0.85, 0.92)	$2.8 \times 10^{-9}$
rs11658063	36 103 872	G/C	0.39	0.96	0.88 (0.84, 0.92)	$1.4 \times 10^{-9}$	0.89 (0.85, 0.92)	$7.4 \times 10^{-9}$
rs11657964	36 100 767	G/A	0.40	1.00	0.88 (0.85, 0.92)	$3.4 \times 10^{-9}$	0.89 (0.86, 0.93)	$3.6 \times 10^{-8}$
<b>rs7501939</b>	<b>36 101 156</b>	<b>C/T</b>	<b>0.40</b>	<b>1.00</b>	<b>0.88 (0.85, 0.92)</b>	<b><math>3.7 \times 10^{-9}</math></b>	0.89 (0.86, 0.93)	$3.8 \times 10^{-8}$
rs4239217	36 098 987	A/G	0.40	1.00	0.88 (0.85, 0.92)	$5.9 \times 10^{-9}$	0.89 (0.86, 0.93)	$5.0 \times 10^{-8}$

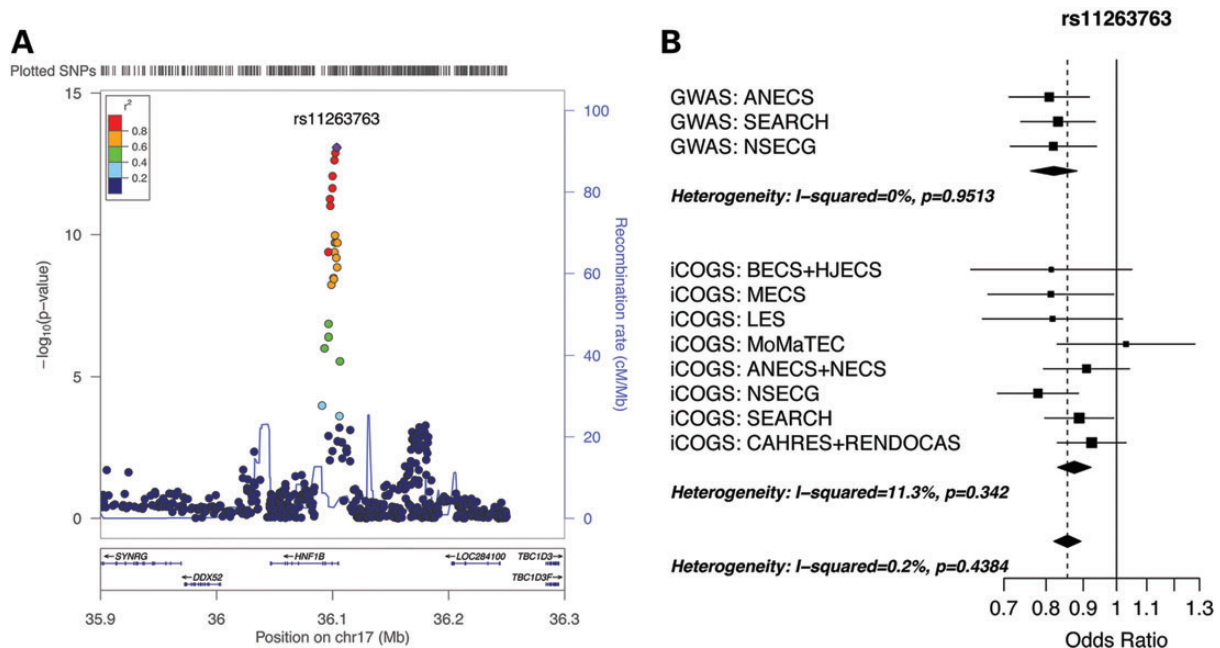
<sup>a</sup>Build 37 position.<sup>b</sup>Major/minor alleles based on forward strand and MAF in Europeans.<sup>c</sup>MAF of iCOGS controls.<sup>d</sup>Average imputation information score for the fine-mapping iCOGS dataset, where SNPs with a score of '1' are genotyped SNPs.<sup>e</sup>Caucasian-only 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.<sup>f</sup>One-degree-of-freedom  $P_{\text{trend}}$ . 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  $\log_{10}$  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 ~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).

Table 3. Association between haplotypes of genotyped SNPs in the iCOGS dataset and endometrial cancer risk

Haplotype	Genotyped SNPs <sup>a</sup>	rs2005705	rs11651755	rs7501939	rs8064454	rs12601991	rs11263762	rrs11651052	Haplotype frequency	OR (95% CI)	P-value
Base haplotype <sup>b</sup>		G	T	C	C	T	A	G	0.406	–	–
Haplotype 1 <sup>c</sup>		A	C	T	A	G	G	A	0.353	<b>0.88 (0.83, 0.93)</b>	<b>8.1 × 10<sup>-6</sup></b>
Haplotype 2		A	C	C	A	G	G	A	0.070	<b>0.84 (0.76, 0.93)</b>	<b>6.7 × 10<sup>-4</sup></b>
Haplotype 3		G	C	T	A	G	G	A	0.037	<b>0.84 (0.73, 0.97)</b>	<b>1.6 × 10<sup>-2</sup></b>
Haplotype 4		A	T	C	C	T	A	G	0.014	1.12 (0.91, 1.38)	2.8 × 10 <sup>-1</sup>
Haplotype 5		G	T	C	C	G	G	G	0.089	0.97 (0.89, 1.06)	5.7 × 10 <sup>-1</sup>
Haplotype 6		G	T	C	C	G	A	G	0.012	1.00 (0.80, 1.26)	9.8 × 10 <sup>-1</sup>

<sup>a</sup>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 rs11263762 (imputed) is in very high LD ( $r^2 = 0.87$ ) with rs11651755, rs8064454 and rs11651052 (shown in bold).

<sup>b</sup>Base haplotype comprised of the major alleles of the seven genotyped SNPs included in the simplified haplotype.

<sup>c</sup>The best associated haplotype is comprised of the minor alleles of the seven genotyped SNPs included in the simplified haplotype.

fine-mapping and ANECS, SEARCH, NSECG and SECGS GWASs) revealed an overall OR of 0.86 ( $P = 2.7 \times 10^{-13}$ ).

The association was similar for endometrioid-subtype cases only, with rs11263763 retaining the strongest association signal in the meta-analysis of the four Caucasian datasets ( $P = 4.1 \times 10^{-12}$ , OR = 0.86), and a genome-wide significant signal seen for the same 18 SNPs as above (Table 5; Supplementary Material, Table S2). Despite a reduction in power to detect an association in non-endometrioid cases due to the smaller case sample size (see Materials and Methods), these 18 SNPs also retained the best association signal for non-endometrioid cases (iCOGS fine-mapping and NSECG GWAS datasets). The top SNP in this analysis was rs10908278 ( $r^2$  to rs11263763 = 0.84;  $P = 1.3 \times 10^{-3}$ , OR = 0.85; signal for rs11263763  $P = 2.4 \times 10^{-3}$ , OR = 0.85) (Supplementary Material, Table S3).

Analyses were also performed adjusting for body mass index (BMI), a major epidemiological risk factor for endometrial cancer, in the subset of iCOGS cases ( $N = 2858$ ) and controls ( $N = 14\,098$ ) for whom BMI data was available. There was no attenuation in effect for our top SNPs [e.g. rs11263763\_unadjusted  $P = 4.5 \times 10^{-4}$ , OR = 0.89, 95% confidence intervals (CIs) 0.83–0.95; rs11263763\_adjusted  $P = 7.9 \times 10^{-4}$ , OR = 0.89, 95% CIs 0.83–0.95] (Supplementary Material, Table S4).

Log-likelihood tests based on the Caucasian datasets, comparing the all-histologies  $P$ -values of all tested SNPs against that of the top SNP rs11263763, prioritized five SNPs in HNF1B intron 1 for follow-up as potentially causal variants based on log-likelihood ratios < 1 : 100: rs11263763, rs11651052 ( $r^2$  to rs11263763 = 0.87), rs8064454 ( $r^2 = 0.87$ ), rs10908278 ( $r^2 = 0.83$ ) and rs11651755 ( $r^2 = 0.87$ ).

### 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 ( $P = 1.3 \times 10^{-2}$ ), rs11658063 ( $P = 5.0 \times 10^{-3}$ ) and marginally so for rs11651755 ( $P = 8.3 \times 10^{-2}$ ). 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 ( $P = 3.0 \times 10^{-2}$ ) and rs1165806 ( $P = 4.8 \times 10^{-2}$ ). 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 =  $5.0 \times 10^{-3}$  (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 ( $P = 1.41 \times 10^{-9}$ ). 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

**Table 4.** Best genotyped<sup>a</sup> and imputed HNF1B SNPs associated with risk of endometrial cancer in four Caucasian and one Asian datasets

	Position <sup>b</sup>	Alleles <sup>c</sup>	MAF Cases	MAF Controls	OR (95% CIs) <sup>d</sup>	P-value <sup>e</sup>	P <sub>heterogeneity</sub>
Genotyped SNP: rs7501939	36101156	G/A					
iCOGS			0.37	0.40	0.90 (0.86–0.95)	5.4 × 10 <sup>-5</sup>	
ANECs GWAS			0.35	0.39	0.83 (0.73–0.94)	3.3 × 10 <sup>-3</sup>	
SEARCH GWAS			0.36	0.40	0.84 (0.75–0.95)	4.2 × 10 <sup>-3</sup>	
NSECG GWAS			0.35	0.39	0.87 (0.76–0.99)	3.7 × 10 <sup>-2</sup>	
SECGS GWAS					1.04 (0.90–1.20)	6.2 × 10 <sup>-1</sup>	
Combined—Caucasian only					0.88 (0.85–0.92)	3.7 × 10 <sup>-9</sup>	5.0 × 10 <sup>-1</sup>
Combined—all 5 datasets					0.89 (0.86–0.93)	3.7 × 10 <sup>-8</sup>	5.0 × 10 <sup>-1</sup>
Imputed SNP: rs11263763	36103565	A/G					
iCOGS			0.43	0.47	0.87 (0.83–0.92)	6.8 × 10 <sup>-8</sup>	
ANECs GWAS			0.42	0.47	0.81 (0.71–0.92)	9.3 × 10 <sup>-4</sup>	
SEARCH GWAS			0.44	0.48	0.83 (0.74–0.93)	2.0 × 10 <sup>-3</sup>	
NSECG GWAS			0.42	0.47	0.82 (0.71–0.94)	3.9 × 10 <sup>-3</sup>	
SECGS GWAS					0.96 (0.84–1.10)	5.7 × 10 <sup>-1</sup>	
Combined—Caucasian only					0.86 (0.82–0.89)	8.4 × 10 <sup>-14</sup>	5.7 × 10 <sup>-1</sup>
Combined—all 5 datasets					0.86 (0.83–0.90)	2.7 × 10 <sup>-13</sup>	5.7 × 10 <sup>-1</sup>

<sup>a</sup>Best SNP genotyped in all four datasets.<sup>b</sup>Build 37 position.<sup>c</sup>Major/minor alleles based on forward strand and MAF in Europeans.<sup>d</sup>Per allele OR for the minor allele relative to the major allele.<sup>e</sup>One-degree-of-freedom P<sub>trend</sub>.**Table 5.** Association signal for cases with endometrioid histology and non-endometrioid histology in the four Caucasian datasets

SNP ID	Position <sup>a</sup>	Alleles <sup>b</sup>	iCOGS MAF <sup>c</sup>	iCOGS Information <sup>d</sup>	Endometrioid histology		Non-endometrioid histology	
					OR (95% CIs) <sup>d</sup>	P-value <sup>e</sup>	OR (95% CIs) <sup>d</sup>	P-value <sup>e</sup>
<b>rs11263763</b>	<b>36 103 565</b>	<b>A/G</b>	<b>0.47</b>	<b>0.96</b>	<b>0.86 (0.82, 0.89)</b>	<b>4.1 × 10<sup>-12</sup></b>	0.85 (0.77, 0.95)	2.4 × 10 <sup>-3</sup>
rs11651052	36 102 381	G/A	0.47	1.00	0.86 (0.82, 0.90)	8.6 × 10 <sup>-12</sup>	0.85 (0.77, 0.94)	1.7 × 10 <sup>-3</sup>
rs8064454	36 101 586	C/A	0.47	1.00	0.86 (0.82, 0.90)	1.4 × 10 <sup>-11</sup>	0.85 (0.77, 0.94)	1.7 × 10 <sup>-3</sup>
<b>rs10908278</b>	<b>36 099 952</b>	<b>A/T</b>	<b>0.47</b>	<b>0.96</b>	0.86 (0.83, 0.90)	6.4 × 10 <sup>-11</sup>	<b>0.85 (0.77, 0.94)</b>	<b>1.3 × 10<sup>-3</sup></b>
rs11651755	36 099 840	T/C	0.48	1.00	0.87 (0.83, 0.91)	1.3 × 10 <sup>-10</sup>	0.85 (0.77, 0.94)	2.1 × 10 <sup>-3</sup>
rs11263761	36 097 775	A/G	0.49	0.93	0.87 (0.83, 0.91)	2.1 × 10 <sup>-10</sup>	0.86 (0.78, 0.95)	3.9 × 10 <sup>-3</sup>
rs4430796	36 098 040	A/G	0.48	0.95	0.87 (0.83, 0.91)	2.7 × 10 <sup>-10</sup>	0.86 (0.78, 0.96)	5.0 × 10 <sup>-3</sup>
rs7405696	36 102 035	C/G	0.43	1.00	0.87 (0.84, 0.91)	1.1 × 10 <sup>-9</sup>	0.87 (0.79, 0.96)	5.6 × 10 <sup>-3</sup>
rs12453443	36 104 121	G/C	0.43	0.96	0.87 (0.83, 0.91)	1.4 × 10 <sup>-9</sup>	0.88 (0.79, 0.97)	9.5 × 10 <sup>-3</sup>
rs757209	36 102 833	A/G	0.42	0.96	0.87 (0.83, 0.91)	1.8 × 10 <sup>-9</sup>	0.87 (0.79, 0.96)	7.4 × 10 <sup>-3</sup>
rs11263762	36 101 926	A/G	0.43	1.00	0.88 (0.84, 0.91)	2.2 × 10 <sup>-9</sup>	0.87 (0.79, 0.96)	5.5 × 10 <sup>-3</sup>
rs12601991	36 101 633	T/G	0.42	1.00	0.88 (0.84, 0.92)	4.8 × 10 <sup>-9</sup>	0.87 (0.79, 0.96)	7.0 × 10 <sup>-3</sup>
rs11658063	36 103 872	G/C	0.39	0.96	0.87 (0.84, 0.91)	5.3 × 10 <sup>-9</sup>	0.89 (0.80, 0.99)	2.7 × 10 <sup>-2</sup>
rs9901746	36 103 149	A/G	0.43	0.96	0.88 (0.84, 0.92)	5.5 × 10 <sup>-9</sup>	0.88 (0.79, 0.97)	9.5 × 10 <sup>-3</sup>
rs2005705	36 096 300	G/A	0.45	1.00	0.88 (0.84, 0.92)	8.3 × 10 <sup>-9</sup>	0.88 (0.79, 0.97)	1.2 × 10 <sup>-2</sup>
rs4239217	36 098 987	A/G	0.40	1.00	0.88 (0.84, 0.92)	1.0 × 10 <sup>-8</sup>	0.90 (0.81, 1.00)	5.2 × 10 <sup>-2</sup>
rs11657964	36 100 767	G/A	0.40	1.00	0.88 (0.84, 0.92)	1.2 × 10 <sup>-8</sup>	0.89 (0.80, 0.99)	2.6 × 10 <sup>-2</sup>
rs7501939	36 101 156	C/T	0.40	1.00	0.88 (0.84, 0.92)	1.4 × 10 <sup>-8</sup>	0.89 (0.80, 0.99)	2.7 × 10 <sup>-2</sup>

<sup>a</sup>Build 37 position.<sup>b</sup>Major/minor alleles based on forward strand and MAF in Europeans.<sup>c</sup>MAF of iCOGS controls.<sup>d</sup>Average imputation information score for the fine-mapping iCOGS dataset, where SNPs with a score of '1' are genotyped SNPs.<sup>e</sup>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.<sup>f</sup>One-degree-of-freedom P<sub>trend</sub>. The top SNPs for each analysis are noted in bold.

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 ( $P = 2.2 \times 10^{-2}$ ) and isoform B ( $P = 2.1 \times 10^{-2}$ ), 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.

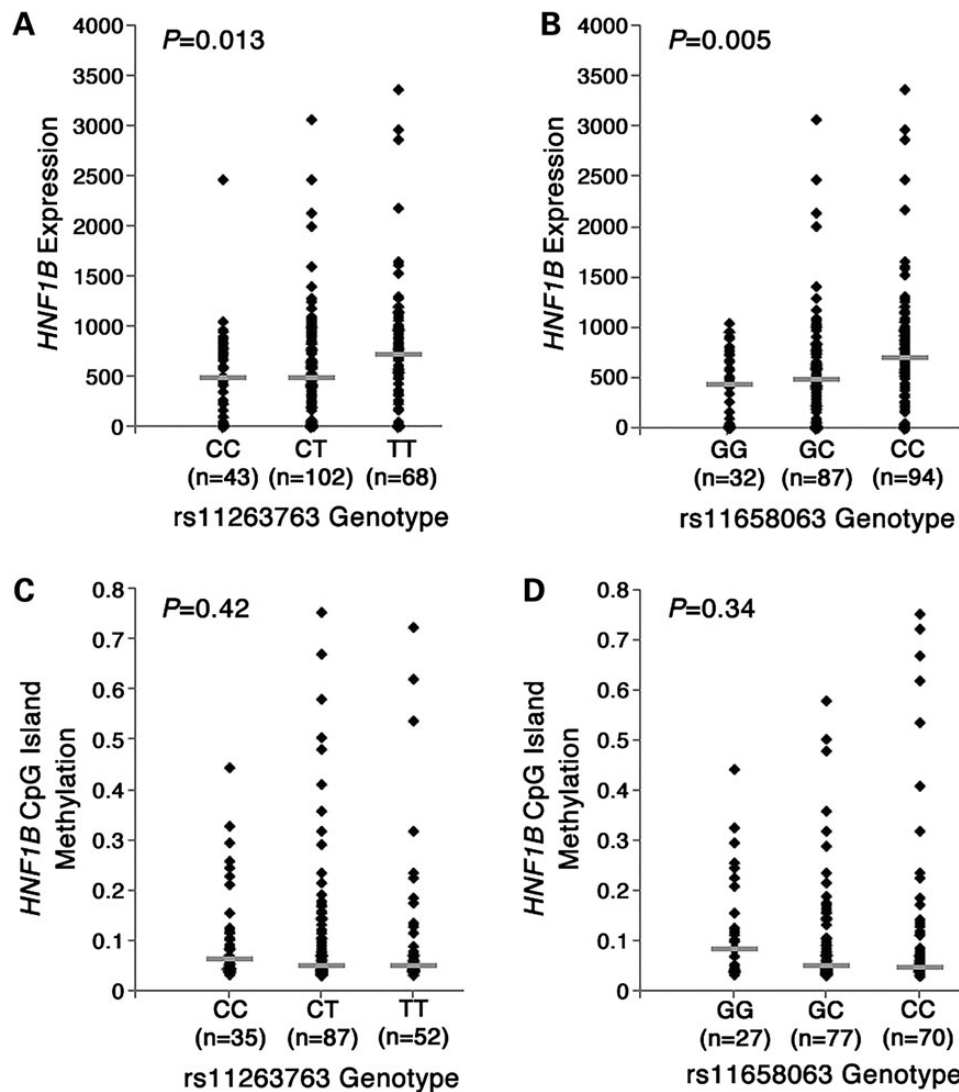


Figure 2. Association of genotypes with *HNF1B* expression as measured by RNA-Seq for rs11263763 (A) and rs11658063 (B), and with average *HNF1B* CpG island methylation for rs11263763 (C) and rs11658063 (D).

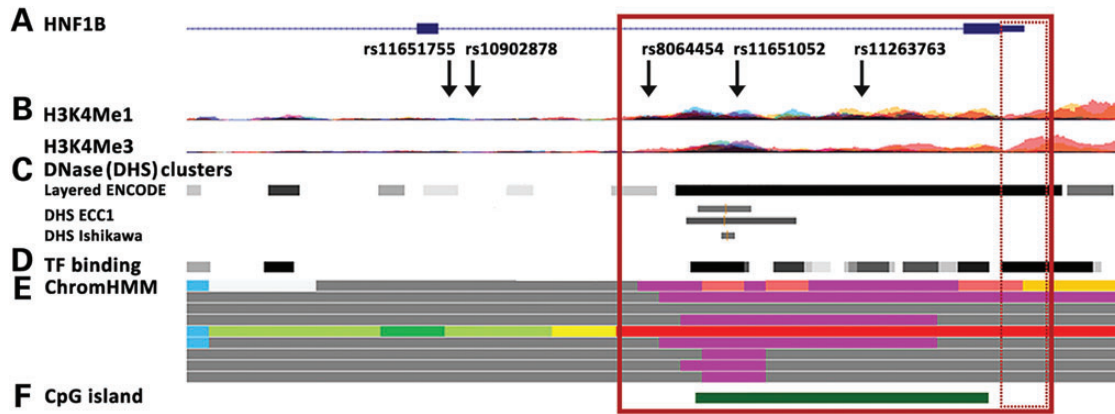
### No association between SNP rs11263763 and *HNF1B* methylation

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,  $r^2=0.95$  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).

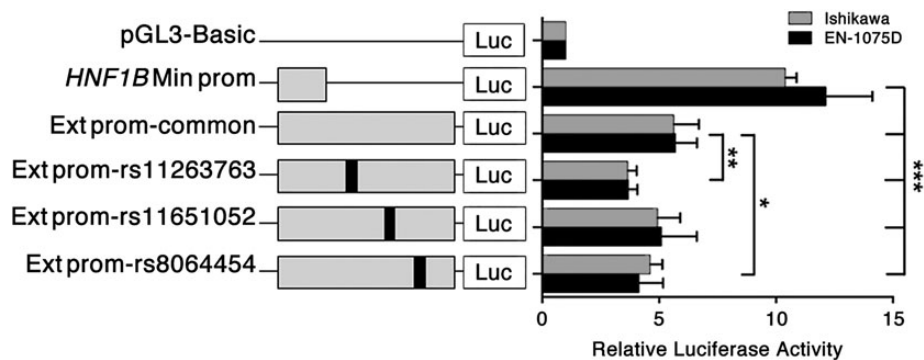
### The strongest candidate causal SNPs map to the extended *HNF1B* promoter region

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*





**Figure 3.** Genetic associations and epigenetic landscape at the *HNF1B* locus. (A) Enlarged image of the *HNF1B* intron 1 region, showing the epigenetic landscape in ENCODE cell lines. The top five likely causal SNPs are indicated in relation to marks of regulatory potential; (B) Histones H3K4Me1 (indicative of regulatory regions) and H3K4Me3 (indicative of promoters); (C) DNaseI hypersensitivity (DHS: indicative of open chromatin, with darker shading indicating stronger experimental signal) in 125 (layered) ENCODE cell lines and endometrial cancer ECC1 (DMSO and estradiol 10 m) and Ishikawa (4-OHTAM and estradiol 10 m) cell lines; (D) Transcription factor (TF) binding in 72 ENCODE cell lines; (E) Chromatin state in nine ENCODE cell lines, with the following color coding: bright red-active promoter; light red-weak promoter; purple-inactive/poised promoter; orange-strong enhancer; yellow-weak enhancer; blue-insulator; dark green-transcriptional transition; light green-weak transcribed; dark gray-repressed/heterochromatin; (F) *HNF1B* CpG island. The solid red box represents the extended promoter region, and the hatched box the minimal promoter region.



**Figure 4.** Luciferase reporter assays in endometrial cell lines demonstrate that SNPs rs11263763 and rs8064454 reduce the extended *HNF1B* promoter activity. The minimal *HNF1B* (Min prom) or extended *HNF1B* (Ext prom) promoters were cloned upstream of a luciferase reporter. An Ext prom construct containing either the wild-type haplotype or minor alleles of rs11263763, rs11651052 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 (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).

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 in the extended promoter constructs decreased relative wild-type *HNF1B* promoter activity by a further ~25% compared with the construct containing the major alleles (Fig. 4).

### 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 cancer 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 hypomethylation 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 eQTL 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 cancer sample set (41) previously genotyped using Illumina Hap550 arrays (Supplementary Material, Information).

### 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 (SBSC; 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):  $n = 45$ ], and a set of tagging SNPs for the LD blocks present across 150 kb of the *HNF1B* region (Build 37, 36025887–36175887:  $n = 105$ ). Additional SNPs within 1 Mb of *HNF1B*, utilized for association analyses and genotype imputation (see below), were chosen by the various COGS participants [PRACTICAL, OCAC, BCAC and The Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA)].

### Data quality control

Genotypes for the ANECS and SEARCH GWAS samples (cases and controls) were subjected to quality control as described previously (3). Genotypes for the iCOGS fine-mapping and NSECG GWAS samples were called using Illumina's proprietary GenCall algorithm (31), and subjected to quality control as follows. SNPs were excluded for call rate  $< 95\%$  ( $< 99\%$  for  $MAF < 5\%$ ),  $MAF < 0.1\%$  or deviations from Hardy-Weinberg equilibrium significant at  $10^{-7}$ . Samples were excluded for low overall call rate ( $< 95\%$ ), heterozygosity  $> 5$  standard deviations from the mean, non-female genotype (XO, XY or XXY) or  $< 85\%$  estimated European ancestry based on identity by state (IBS) scores between study individuals and individuals in HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) and multidimensional scaling. For cases, any

96-well plate containing  $\geq 5$  excluded samples was entirely excluded. For duplicate samples or those identified as close relatives by IBS probabilities  $> 0.85$ , the sample with the lower call rate was excluded, except for case-control relative pairs for which the case was retained. Following quality control, the iCOGs sample retained data for 197 627 SNPs, and the NSECG 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  $\sim 1$  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 GWASs 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 for the first 10 principal components of the genomic kinship matrix, based on 37 000 uncorrelated iCOGS SNPs ( $r^2 < 0.1$ ), including  $\sim 1000$  selected as ancestry informative markers, using an in-house C++ program incorporating the Intel MKL libraries for eigenvectors (<http://ccge.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/ldsearchpw.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  $10^{-4}$ . 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  $10^{-4}$ .

### 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 rs4430 796 [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 ECC1 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 NHMRC (ID#339435), the Cancer Council Queensland (ID#4196615) and Cancer Council Tasmania (ID#403031 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#1031333). 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. NSECG was supported by the EU FP7 CHIBCHA grant and CORGI by Cancer Research UK. Recruitment of the QIMR 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 Verelst 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 CA122443, U19 CA148112, P50 CA13 6393, and GAME-ON the NCI Cancer Post-GWAS Initiative U19 CA148112), the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. MoMaTEC received financial support from a Helse Vest Grant, the University of Bergen, Melzer Foundation, The Norwegian Cancer Society (Harald Andersens legat), The Research Council of Norway and Haukeland University Hospital. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, The NBN Children's Cancer Research Group, Ms Jennie Thomas and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet (numbers: 20110222, 20110483, 20110141 and DF 07015), The Swedish Labor Market Insurance (number 100069) and The Swedish Cancer Society (number 11 0439). The Cancer Hormone Replacement Epidemiology in Sweden Study (CAHRES, formerly called The Singapore and Swedish Breast/Endometrial Cancer Study; SASBAC) was supported by funding from the Agency for Science, Technology and Research of Singapore (A\*STAR), the US National Institutes of Health and the Susan G. Komen Breast Cancer Foundation. The Shanghai Endometrial Cancer Genetic Study (SECGS) was 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 (PPD/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

1. Tavassoli, F.A. and Stratton, M.R. (2003) *Tumors of the Breast and Female Genital Organs*. Lyon: IARC Press.
2. Cancer Genome Atlas Research Network. (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**, 67–73.
3. Spurdle, A.B., Thompson, D.J., Ahmed, S., Ferguson, K., Healey, C.S., O'Mara, T., Walker, L.C., Montgomery, S.B., Dermitzakis, E.T., Fahey, P. et al. (2011) Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat. Genet.*, **43**, 451–454.
4. Gudmundsson, J., Sulem, P., Steinthorsdottir, V., Bergthorsson, J.T., Thorleifsson, G., Manolescu, A., Rafnar, T., Gudbjartsson, D., Agnarsson, B.A., Baker, A. et al. (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.*, **39**, 977–983.
5. Sun, J., Zheng, S.L., Wiklund, F., Isaacs, S.D., Purcell, L.D., Gao, Z., Hsu, F.C., Kim, S.T., Liu, W., Zhu, Y. et al. (2008) Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat. Genet.*, **40**, 1153–1155.
6. Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J. et al. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, **40**, 316–321.
7. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, **40**, 310–315.
8. Takata, R., Akamatsu, S., Kubo, M., Takahashi, A., Hosono, N., Kawaguchi, T., Tsunoda, T., Inazawa, J., Kamatani, N., Ogawa, O. et al. (2010) Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.*, **42**, 751–754.
9. Liu, F., Hsing, A.W., Wang, X., Shao, Q., Qi, J., Ye, Y., Wang, Z., Chen, H., Gao, X., Wang, G. et al. (2011) Systematic confirmation study of reported prostate cancer risk-associated single nucleotide polymorphisms in Chinese men. *Cancer Sci.*, **102**, 1916–1920.
10. Pharoah, P.D., Tsai, Y.Y., Ramus, S.J., Phelan, C.M., Goode, E.L., Lawrenson, K., Buckley, M., Fridley, B.L., Tyrer, J.P., Shen, H. et al. (2013) GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.*, **45**, 362–370. 370e361-362.
11. Shen, H., Fridley, B.L., Song, H., Lawrenson, K., Cunningham, J.M., Ramus, S.J., Cicek, M.S., Tyrer, J., Stram, D., Larson, M.C. et al. (2013) Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. *Nat. Commun.*, **4**, 1628.
12. Berndt, S.I., Sampson, J., Yeager, M., Jacobs, K.B., Wang, Z., Hutchinson, A., Chung, C., Orr, N., Wacholder, S., Chatterjee,

- N. et al. (2011) Large-scale fine mapping of the HNF1B locus and prostate cancer risk. *Hum. Mol. Genet.*, **20**, 3322–3329.
13. Grisanzio, C., Werner, L., Takeda, D., Awoyemi, B.C., Pomerantz, M.M., Yamada, H., Sooriakumaran, P., Robinson, B.D., Leung, R., Schinzel, A.C. et al. (2012) Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis. *Proc. Natl Acad. Sci. USA*, **109**, 11252–11257.
  14. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
  15. Harries, L.W., Perry, J.R., McCullagh, P. and Crundwell, M. (2010) Alterations in LMTK2, MSMB and HNF1B gene expression are associated with the development of prostate cancer. *BMC Cancer*, **10**, 315.
  16. Hughes, L.A., Melotte, V., de Schrijver, J., de Maat, M., Smit, V.T., Bovee, J.V., French, P.J., van den Brandt, P.A., Schouten, L.J., de Meyer, T. et al. (2013) The CpG island methylator phenotype: what's in a name? *Cancer Res.*, **73**, 5858–5868.
  17. Garg, K., Leitao, M.M. Jr, Wynveen, C.A., Sica, G.L., Shia, J., Shi, W. and Soslow, R.A. (2010) p53 overexpression in morphologically ambiguous endometrial carcinomas correlates with adverse clinical outcomes. *Mod. Pathol.*, **23**, 80–92.
  18. Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H. and Foisner, R. (2005) DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene*, **24**, 2375–2385.
  19. Singh, M., Spoelstra, N.S., Jean, A., Howe, E., Torkko, K.C., Clark, H.R., Darling, D.S., Shroyer, K.R., Horwitz, K.B., Broadus, R.R. et al. (2008) ZEB1 expression in type I vs type II endometrial cancers: a marker of aggressive disease. *Mod. Pathol.*, **21**, 912–923.
  20. Spoelstra, N.S., Manning, N.G., Higashi, Y., Darling, D., Singh, M., Shroyer, K.R., Broadus, R.R., Horwitz, K.B. and Richer, J.K. (2006) The transcription factor ZEB1 is aberrantly expressed in aggressive uterine cancers. *Cancer Res.*, **66**, 3893–3902.
  21. Lazarevich, N.L., Cheremnova, O.A., Varga, E.V., Ovchinnikov, D.A., Kudrjavitseva, E.I., Morozova, O.V., Fleishman, D.I., Engelhardt, N.V. and Duncan, S.A. (2004) Progression of HCC in mice is associated with a downregulation in the expression of hepatocyte nuclear factors. *Hepatology*, **39**, 1038–1047.
  22. Buchner, A., Castro, M., Hennig, A., Popp, T., Assmann, G., Stief, C.G. and Zimmermann, W. (2010) Downregulation of HNF-1B in renal cell carcinoma is associated with tumor progression and poor prognosis. *Urology*, **76**, 507. e506–511.
  23. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. and Gerald, W.L. (2004) Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.*, **113**, 913–923.
  24. Cancer Genome Atlas Research, N. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
  25. Terasawa, K., Toyota, M., Sagae, S., Ogi, K., Suzuki, H., Sonoda, T., Akino, K., Maruyama, R., Nishikawa, N., Imai, K. et al. (2006) Epigenetic inactivation of TCF2 in ovarian cancer and various cancer cell lines. *Br. J. Cancer.*, **94**, 914–921.
  26. Tsuchiya, A., Sakamoto, M., Yasuda, J., Chuma, M., Ohta, T., Ohki, M., Yasugi, T., Taketani, Y. and Hirohashi, S. (2003) Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am. J. Pathol.*, **163**, 2503–2512.
  27. Cuff, J., Salari, K., Clarke, N., Esheba, G.E., Forster, A.D., Huang, S., West, R.B., Higgins, J.P., Longacre, T.A. and Pollack, J.R. (2013) Integrative bioinformatics links HNF1B with clear cell carcinoma and tumor-associated thrombosis. *PLoS ONE*, **8**, e74562.
  28. Shao, D.D., Tsherniak, A., Gopal, S., Weir, B.A., Tamayo, P., Stransky, N., Schumacher, S.E., Zack, T.I., Beroukhir, R., Garraway, L.A. et al. (2012) ATARIS: computational quantification of gene suppression phenotypes from multi-sample RNAi screens. *Genome. Res.*, **23**, 665–678.
  29. Bach, I. and Yaniv, M. (1993) More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *EMBO J.*, **12**, 4229–4242.
  30. Edwards, S.L., Beesley, J., French, J.D. and Dunning, A.M. (2013) Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.*, **93**, 779–797.
  31. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K. et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.*, **45**, 353–361. 361e351–352.
  32. Eeles, R.A., Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S. et al. (2013) Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.*, **45**, 385–391. 391e381–382.
  33. Garcia-Closas, M., Couch, F.J., Lindstrom, S., Michailidou, K., Schmidt, M.K., Brook, M.N., Orr, N., Rhie, S.K., Riboli, E., Feigelson, H.S. et al. (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.*, **45**, 392–398. 398e391–392.
  34. Sakoda, L.C., Jorgenson, E. and Witte, J.S. (2013) Turning of COGS moves forward findings for hormonally mediated cancers. *Nat. Genet.*, **45**, 345–348.
  35. Bahcall, O.G. (2013) iCOGS collection provides a collaborative model. (Foreword). *Nat. Genet.*, **45**, 343.
  36. Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P. and Clark, T.G. (2007) A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics*, **23**, 2741–2746.
  37. McGregor, B., Pfitzner, J., Zhu, G., Grace, M., Eldridge, A., Pearson, J., Mayne, C., Aitken, J.F., Green, A.C. and Martin, N.G. (1999) Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet. Epidemiol.*, **16**, 40–53.
  38. Painter, J.N., Anderson, C.A., Nyholt, D.R., Macgregor, S., Lin, J., Lee, S.H., Lambert, A., Zhao, Z.Z., Roseman, F., Guo, Q. et al. (2011) Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.*, **43**, 51–54.
  39. McEvoy, M., Smith, W., D'Este, C., Duke, J., Peel, R., Schofield, P., Scott, R., Byles, J., Henry, D., Ewald, B. et al. (2010) Cohort profile: the Hunter Community Study. *Int. J. Epidemiol.*, **39**, 1452–1463.
  40. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
  41. Houlston, R.S., Cheadle, J., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Spain, S.L., Broderick, P., Domingo, E., Farrington, S. et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, **42**, 973–977.
  42. Long, J., Zheng, W., Xiang, Y.B., Lose, F., Thompson, D., Tomlinson, I., Yu, H., Wentzensen, N., Lambrechts, D., Dork, T.

- et al. (2012) Genome-wide association study identifies a possible susceptibility locus for endometrial cancer. *Cancer Epidemiol. Biomarkers Prev.*, **21**, 980–987.
43. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
44. Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3*, **1**, 457–470.
45. Ferreira, T. and Marchini, J. (2011) Modeling interactions with known risk loci—a Bayesian model averaging approach. *Ann. Hum. Genet.*, **75**, 1–9.
46. Higgins, J.P. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, **21**, 1539–1558.
47. Gauderman, W.J. and Morrison, J. (2006) QUANTO 1.1: a computer program for power and sample size calculations for genetic-epidemiology studies. <http://biostats.usc.edu/software>.
48. Udler, M.S., Tyrer, J. and Easton, D.F. (2010) Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. *Genet. Epidemiol.*, **34**, 463–468.
49. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E. et al. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, **2**, 401–404.
50. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. et al. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, pl1.
51. Buchanan, D.D., Tan, Y.Y., Walsh, M.D., Clendenning, M., Metcalf, A.M., Ferguson, K., Arnold, S.T., Thompson, B.A., Lose, F.A., Parsons, M.T. et al. (2014) Tumor mismatch repair immunohistochemistry and DNA MLH1 methylation testing of patients with endometrial cancer diagnosed at age <60 years optimizes triage for population-level germline mismatch repair gene mutation testing. *J. Clin. Oncol.*, **32**, 90–100.
52. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R. et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
53. Macintyre, G., Bailey, J., Haviv, I. and Kowalczyk, A. (2010) is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics*, **26**, i524–i530.