Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer

Tracy A O’Mara¹, Dylan M Glubb¹, Jodie N Painter¹, Timothy Cheng², Joe Dennis³, The Australian National Endometrial Cancer Study Group (ANECS)¹, John Attia⁴,⁵, Elizabeth G Holliday⁴,⁵, Mark McEvoy⁵, Rodney J Scott⁴,⁶,⁷,⁸, Katie Ashton⁴,⁷,⁸, Tony Pietro², Geoffrey Otton⁹, Mitul Shah¹⁰, Shahana Ahmed¹⁰, Catherine S Healey¹⁰, Maggie Gorman², Lynn Martin², National Study of Endometrial Cancer Genetics Group (NSECG)², Shirley Hodgson¹¹, Peter A Fasching¹²,¹³, Alexander Hein¹³, Matthias W Beckmann¹³, Arif B Ekici¹⁴, Per Hall¹⁵, Kamila Czene¹⁶, Hafed Darabi¹⁵, Jingmei Li¹⁵, Matthias Dürs¹⁶, Ingo Runnebaum¹⁶, Peter Hillemanns¹⁷, Thilo Dörk¹⁸, Diether Lambrechts¹⁹,²⁰, Jeroen Depreeuw¹⁹,²⁰,²¹, Daniela Annibali²¹, Frederic Amant²¹, Hui Zhao²²,²⁰, Ellen L Goode²²,²⁰, Sean C Dowdy²³,²⁰, Brooke L Fridley²⁴, Stacey J Winham²², Helga B Salvesen²⁵,²⁶, Tormund S Njølstad²⁵,²⁶, Jone Trovik²⁵,²⁶, Henrica M J Werner²⁵,²⁶, Emma Tham²⁷, Tao Liu²⁷, Miriam Mints²⁸, RENDOCAS²⁷,²⁸, Manjeet K Bolla³, Kyriaki Michaëllidou⁴, Jonathan P Tyrer¹⁰, Qin Wang³, John L Hopper²⁹, AOCS Group¹,³⁰, Julian Petö¹¹, Anthony J Swerdlow³¹,³²,³³, Barbara Burwinkel³⁴,³⁵, Hermann Brenner³⁶,³⁷,³⁸, Alfons Meindl³⁹, Hiltrud Brauch⁴⁰,⁴¹, Annika Lindblom²⁷, Jenny Chang-Claude³⁵, Fergus J Couch²²,²⁴, Graham G Giles²⁹,³⁴,³⁵, Vessela N Kristensen⁴⁵,⁴⁶,⁴⁷, Angela Cox⁴⁸, Paul D P Pharoah¹⁰, Alison M Dunning¹⁰, Ian Tomlinson², Douglas F Easton³,¹⁰, Deborah J Thompson³ and Amanda B Spurdle¹

¹Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, Brisbane, Queensland 4006, Australia
²Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
³Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK
⁴Hunter Medical Research Institute, John Hunter Hospital, Newcastle, New South Wales 2305, Australia
⁵School of Medicine and Public Health, Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, Newcastle, New South Wales 2305, Australia
⁶Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales 2305, Australia
⁷Centre for Information Based Medicine, ⁸School of Biomedical Sciences and Pharmacy, and ⁹School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales 2308, Australia
¹⁰Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK
¹¹Department of Clinical Genetics, St George’s, University of London, London SW17 0RE, UK
¹²Division of Hematology/Oncology, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA
¹³Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen 91054, Germany
¹⁴Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen 91054, Germany
¹⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden
¹⁶Department of Gynaecology, Jena University Hospital – Friedrich Schiller University, Jena 07743, Germany
¹⁷Hannover Medical School, Clinics of Gynaecology and Obstetrics, Hannover 30625, Germany
¹⁸Gynaecology Research Unit, Hannover Medical School, Hannover 30625, Germany
¹⁹Vesalius Research Center, Leuven 3000, Belgium
²⁰Laboratory for Translational Genetics, Department of Oncology, University Hospitals Leuven, Leuven 3000, Belgium
Abstract

Excessive exposure to estrogen is a well-established risk factor for endometrial cancer (EC), particularly for cancers of endometrioid histology. The physiological function of estrogen is primarily mediated by estrogen receptor alpha, encoded by ESR1. Consequently, several studies have investigated whether variation at the ESR1 locus is associated with risk of EC, with conflicting results. We performed comprehensive fine-mapping analyses of 3633 genotyped and imputed single nucleotide polymorphisms (SNPs) in 6607 EC cases and 37,925 controls. There was evidence of an EC risk signal located at a potential alternative promoter of the ESR1 gene (lead SNP rs79575945, \( P = 1.86 \times 10^{-5} \)), which was stronger for cancers of endometrioid subtype \( (P = 3.76 \times 10^{-6}) \). Bioinformatic analysis suggests that this risk signal is in a functionally important region targeting ESR1, and eQTL analysis found that rs79575945 was associated with expression of SYNE1, a neighbouring gene. In summary, we have identified a single EC risk signal located at ESR1, at study-wide significance. Given SNPs located at this locus have been associated with risk for breast cancer, also a hormonally driven cancer, this study adds weight to the rationale for performing informed candidate fine-scale genetic studies across cancer types.
Introduction

Endometrial cancer is the most commonly diagnosed gynaecological malignancy in developed countries (http://globocan.iarc.fr/). Excessive endogenous and exogenous estrogen exposure or estrogen exposure unopposed by progesterone is a well-established risk factor for the development and progression of endometrial cancer (Kaaks et al. 2002, Key & Pike 1988). Estrogen receptor alpha (encoded by ESR1) is the predominant receptor responsible for mediating the effects of estrogen in the endometrium.

A number of studies have previously been performed to investigate the hypothesis that variation at the ESR1 locus may be associated with predisposition to endometrial cancer (Weiderpass et al. 2000, Sasaki et al. 2002, Iwamoto et al. 2003, Einarsdottir et al. 2008, 2009, Wedren et al. 2008, Ashton et al. 2009, 2010, Sliwinski et al. 2010, Li et al. 2011), but results from these relatively underpowered studies (maximum sample size 713 cases and 1567 controls) have been conflicting. However, comprehensive candidate gene and genome-wide association studies of breast cancer, which shares many risk factors with endometrial cancer, have identified cancer-associated risk variants at the ESR1 locus (Dunning et al. 2009, Zheng et al. 2009, Turnbull et al. 2010, Hein et al. 2012). These findings indicate a need for similar large-scale and comprehensive genetic analysis of endometrial cancer to elucidate the role of ESR1 variants in the risk of endometrial cancer.

Here we present the results from fine-mapping of the ESR1 locus by dense SNP genotyping and imputation in 6607 endometrial cancer cases and 37,925 controls of European descent within the Endometrial Cancer Association Consortium.

Materials and methods

Datasets

Genotyping of the fine-mapping dataset was performed on a custom Illumina Infinium iSelect array (‘iCOGS’, designed by the Collaborative Oncological Gene-environment Study, details summarized in Bahcall (2013)). All studies have the relevant IRB approval in each country in accordance with the principles embodied in the Declaration of Helsinki, and informed consent was obtained from all participants. Details of iCOGS genotyping of endometrial cancer cases and control samples can be found in Supplementary Table 1, see section on supplementary data given at the end of this article and in Painter et al. (2014).

All cases and controls selected for analysis were of European ancestry, as defined by Identity-By-State (IBS) scores between study individuals and individuals in HapMap (http://hapmap.ncbi.nlm.nih.gov/). The final analysis of the iCOGS dataset included genotypes for 4401 women with a confirmed diagnosis of endometrial cancer and 28,758 healthy female controls genotyped by the Breast Cancer Association Consortium (BCAC) or the Ovarian Cancer Association Consortium (OCAC). Additionally, three Caucasian GWAS datasets (ANECS, SEARCH and NSECG) were as previously described, totalling 2206 cases and 9167 controls after quality control (Spurdle et al. 2011, Painter et al. 2014). Overall, there were 6607 endometrial cancer cases and 37,925 controls included in the meta-analysis of the four datasets (ANECS, SEARCH and NSECG GWAS datasets and the iCOGS dataset).

Fine-mapping

The study herein includes SNPs in a 1 Mb region including ESR1 (chr6: 151 600 000–152 650 000; NCBI build 37 assembly). SNPs with a minor allele frequency >2% using the 1000 Genomes Project (March 2010 Pilot version 60 CEU project data) were considered for inclusion for ESR1 fine-mapping on the iCOGS array by BCAC. In total, 975 SNPs were selected, comprising 277 SNPs correlated (r² >0.1) with three previously reported breast cancer associated SNPs (rs2046210, rs3757318 and rs3020314), and a 698 SNP set tagging all remaining SNPs in the region with r² >0.9.

Regional imputation

Genotypes for SNPs present in 1000 Genomes Phase 1 (April 2012 release) were imputed for the fine-mapping dataset and each GWAS dataset using IMPUTE V2.0 (Howie et al. 2009). Imputation was performed separately for each dataset. SNPs with an imputation information score >0.8 for all four datasets and minor allele frequency >0.01 were included in analysis. Following quality control, a total of 3633 genotyped and imputed SNPs were available across all four datasets (the three GWAS and iCOGS datasets).

Association analysis

Odds ratios for each SNP were estimated for the four imputed datasets separately, using unconditional logistic
regression with a per-allele (one degree-of-freedom) model, based on the expected genotyped dosages for the imputed SNPs. The GWAS datasets were each analysed as a single stratum, with adjustment for the first two (ANECS and NSECNG) and three (SEARCH) principal components. For the iCOGS dataset, analyses were performed adjusting for strata and for the first ten principal components, as previously described (Painter et al. 2014). The numbers of principal components included in the analyses were selected to adequately account for population stratification in each of the datasets. Results from the four studies were combined using standard fixed-effects meta-analysis, and between-study heterogeneity assessed by Q statistic were combined using standard fixed-effects meta-analysis, and between-study heterogeneity assessed by Q statistic.

To assess possible interaction with BMI group (≤30 kg/m² or >30 kg/m²) for lead SNP rs9341019, the significance of multiplicative interaction was assessed by the change in the likelihood ratio estimate after inclusion of a BMI-by-genotype interaction term to a simpler model without this term. Analyses were conducted using R, including the GenABEL (Aulchenko et al. 2007), meta packages (Schwarzer 2010) and SNPTESTv2 (Ferreira & Marchini 2011). All statistical tests were two-sided.

eQTL analysis

Data from endometrial tumours were accessed from The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network et al. 2013). Germline SNP genotypes (Affymetrix 6.0 arrays) were downloaded through the controlled access portal, while epidemiological data, normalized RNA-Seq data and copy-number information were downloaded through the public access TCGA portal. There were 290 TCGA patients (221 endometrioid histology) with complete genotype, RNA-Seq and copy-number data included in the analysis. Quality control was performed on the germline SNP genotypes as previously described (Carvajal-Carmona et al. 2015). To increase the number of SNPs in the analysis, we imputed genotypes for SNPs present in the 1000 Genomes dataset v3 in the ESR1 region (chr6: 150 125 000–152 650 000, April 2012 release) which were not genotyped by the Affymetrix 6.0 platform using minimac (Howie et al. 2012, Fuchsberger et al. 2015) Software. Haplotypes were phased using the MaCH program (Li et al. 2009, 2010) before running minimac for genotype imputation, using the recommended parameters (20 iterations of the Markov sampler and 200 states). SNPs imputed with a RSQR (quality measure) > 0.8 and minor allele frequency > 0.01 were included in the eQTL analysis. RNA-Seq expression for genes 500 kb upstream and downstream of ESR1 (SYNE1, ESR1, CCDC170, C6orf211, RMND1, ZBTB2, AKAPI2, MYCIT) were adjusted for somatic copy number variation, as previously described by Li et al. (2013). The associations between genotype and adjusted expression for each gene were evaluated using linear regression models by the mach2qtl program (Li et al. 2009, 2010). Associations were considered to be statistically significant after correction for the total number of genes analysed across the region (0.05/8 genes = 6.25 × 10⁻⁵).

Results

Meta-analysis performed on 3633 SNPs that passed quality control criteria in the four studies (iCOGS, ANECS, SEARCH and NSECNG) identified 401 SNPs associated with endometrial cancer risk with P < 0.05 (Supplementary Table 2, see section on supplementary data given at the end of this article), compared to 182 expected by chance. When analysis was restricted to endometrioid-only endometrial cancer, 411 mostly overlapping SNPs were identified to be associated with a P < 0.05 (Supplementary Table 2).

Imputed SNP rs79575945 displayed the strongest association for endometrial cancer risk (per A-allele OR 0.83 and 95% CI 0.77–0.90, P = 1.85 × 10⁻⁵; Fig. 1). The risk association was slightly stronger for endometrioid endometrial cancer (per A-allele OR 0.83 and 95% CI 0.77–0.90, P = 3.76 × 10⁻⁶; 5611 endometrioid cases and 37 926 controls). No other SNPs reached significance (P < 1.85 × 10⁻⁵) after conditioning on rs79575945, suggesting the presence of a single endometrial risk signal at this locus. Similar associations were observed for rs9341019 in the same linkage disequilibrium (LD) block as rs79575945, which was genotyped in all four datasets (rs9341019 OR 0.84 and 95% CI 0.76–0.92, P = 2.2 × 10⁻⁴; r² = 0.27 to rs79575945).

Supplementary Table 3, see section on supplementary data given at the end of this article lists the 47 SNPs most likely to be the causal variant underlying the risk associations with most significant ‘lead’ SNPs rs79575945. This SNP set was defined as the SNPs which were in LD (r² > 0.2) and had a likelihood of association.
with endometrial cancer <100:1 with the relevant lead SNP (Carvajal-Carmona et al. 2015, Glubb et al. 2015).

Given BMI is a major epidemiological risk factor for endometrial cancer, analyses were repeated adjusting for BMI in the subset of cases (n=4088) and controls (n=16 590) for whom BMI data were available, and also assessing the possible interaction of rs79575945 with BMI group (≤30 kg/m² or >30 kg/m²). There was no discernible difference in effect for rs79575945 (unadjusted OR=0.86, \(P=2.4 \times 10^{-3}\); adjusted OR=0.82, \(P=3.7 \times 10^{-4}\)), and no significant evidence of interaction of rs79575945 with BMI (\(P\)-interaction=0.15).

SNP rs79575945 was not significantly associated with risk of non-endometrioid endometrial cancer (OR 0.94 and 95% CI 0.80–1.13, \(P=0.54\)), although there was reduced power to detect association due to the smaller case sample size (iCOGS fine-mapping and NSECG GWAS datasets only, case n=887). No SNP reached study-wide significance for non-endometrioid endometrial cancer risk. Similarly, no significant associations were found in the case-only analysis, comparing endometrioid endometrial cancer patients to non-endometrioid patients (rs79575945 OR 1.08 and 95% CI 0.89–1.30, \(P=0.43\)).

None of the 47 potentially causal variants (Supplementary Table 3, see section on supplementary data given at the end of this article) showed evidence of an association with \(ESR1\) expression, using genotype and RNA-Seq data from TCGA. The strongest association observed for any SNP in this region with \(ESR1\) levels in endometrioid endometrial tumours was rs74575485 located upstream of the rs79575945 risk signal (\(r^2=0.001\)), but this SNP was not associated with risk (eQTL \(P=1.45 \times 10^{-3}\), risk \(P=0.77\)). We found evidence of an association between the top risk SNP rs79575945 and increased expression of \(SYNE1\) in endometrioid endometrial tumour (eQTL \(P=3.17 \times 10^{-5}\)). This association is considered to be statistically significant after correcting for the total number of genes analysed across the region (\(P\) for significance = \(6.25 \times 10^{-3}\)).

We integrated location of candidate causal SNPs with publicly available genomic data to assess likely functional relevance of SNPs. Candidate causal SNPs mapped to a potential regulatory element, which we defined by evidence of enhancer-specific histone modification (mono-methylation of H3 lysine 4 (H3K4Me1)), DNase hypersensitivity sites representative of open chromatin, and regions bound by transcription factors (Fig. 2). Super-enhancers annotated in the study by Hnisz et al. (2013) were also found to overlap with candidate causal SNPs (Fig. 2), indicating the functional importance of this region. Importantly, ENCODE data showed presence of DNase hypersensitivity sites and evidence for binding of transcription factors in Ishikawa endometrial cancer cells, indicating these regions may be active in endometrial tumours. The binding of these transcription factors were not found to be altered by the candidate causal SNPs, using two independent \textit{in silico} prediction algorithms (Supplementary Table 4, see section on supplementary data given at the end of this article). Candidate causal SNP rs9340770 was predicted to alter binding of p300 by HaploReg, and ENCODE data have shown p300 binding to occur at this region in Ishikawa cells (Encode Project Consortium et al. 2012).

**Discussion**

We have performed the largest and most comprehensive study assessing the association of SNPs across the \(ESR1\) gene with endometrial cancer risk. We provide evidence of a study-wide significant association between endometrial cancer risk and imputed SNP rs79575945. Our study implemented parameters to reduce imputation errors and minimize false-positive associations, including rigorous pre-imputation quality control, excluding rare SNPs (minor allele frequency <0.01) and using a high imputation quality score threshold (>0.8) for analyses (Marchini & Howie 2010). These measures, and the similar association observed for the best genotyped SNP in the same LD block as imputed lead SNP rs79575945, increase our confidence for the observed association. Given the strong prior evidence for association of this region with a hormonal cancer, as well as with other hormone-related phenotypes (Estrada et al. 2012, Perry et al. 2014),
we considered this a candidate-gene study. The consistency of SNP association with endometrial cancer risk between the four studies gives us confidence in this finding. Using tagger (de Bakker et al. 2005), 246 SNPs were calculated to be required to tag our region of interest by pairwise-tagging ($r^2 > 0.5$). The most strongly associated SNP had a $P$ value an order of magnitude smaller than the Bonferroni-adjusted significance threshold based on the number of independent SNPs at the locus ($P$ for significance = 0.05/246 = 2.0 × 10^{-4}). Notably, there was a more significant association for the endometrioid histology subtype which is well-established to be estrogen driven (Kaaks et al. 2002).

Neither SNP rs79575945, nor any other in the risk-associated SNP set, has been previously reported to be associated with endometrial cancer risk. Reported associated SNPs from smaller candidate studies investigating the effect of genetic variation at the ESR1 locus on endometrial cancer risk are not in LD ($r^2 < 0.2$) with rs79575945 and were not validated in our larger study (Table 1).

SNPs associated with multiple phenotypes have been mapped to the ESR1 locus, notably breast cancer (Zheng et al. 2009, Turnbull et al. 2010, Hein et al. 2012), which shares many risk factors with endometrial cancer, and age-of-menarche (Perry et al. 2014) and bone mineral density (Estrada et al. 2012), which are both associated with...
Table 1  Associations of ESR1 SNPs previously reported to be associated with endometrial cancer

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr:Position (b37)</th>
<th>Effect/ reference allele</th>
<th>Frequency of effect allele</th>
<th>Imputation information score</th>
<th>Genotyped/imputed</th>
<th>All cases</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>Endometrioid Only</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>Direction of effect relative to effect allele (reference reporting association with endometrial cancer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2234693</td>
<td>6: 152163335</td>
<td>T/C</td>
<td>0.46</td>
<td>1.00</td>
<td>Genotyped</td>
<td>0.98 (0.94–1.02)</td>
<td>0.35</td>
<td>0.98 (0.94–1.02)</td>
<td>0.32</td>
<td>0.12 (Y) Ashton et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9340799</td>
<td>6: 15216381</td>
<td>A/G</td>
<td>0.35</td>
<td>1.00</td>
<td>Genotyped</td>
<td>1.04 (1.0–1.08)</td>
<td>0.08</td>
<td>1.04 (1.00–1.09)</td>
<td>0.06</td>
<td>0.06 (Y) Einarsdottir et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3853250</td>
<td>6: 152159900</td>
<td>T/G</td>
<td>0.46</td>
<td>0.99</td>
<td>Imputed</td>
<td>0.98 (0.94–1.02)</td>
<td>0.35</td>
<td>0.98 (0.94–1.02)</td>
<td>0.32</td>
<td>0.12 (Y) Einarsdottir et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1709181</td>
<td>6: 152175180</td>
<td>T/C</td>
<td>0.26</td>
<td>0.98</td>
<td>Imputed</td>
<td>1.02 (0.98–1.07)</td>
<td>0.40</td>
<td>1.03 (0.99–1.08)</td>
<td>0.15</td>
<td>0.07 (Y) Einarsdottir et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4870057</td>
<td>6: 152171898</td>
<td>A/G</td>
<td>0.33</td>
<td>0.91</td>
<td>Imputed</td>
<td>1.04 (0.99–1.08)</td>
<td>0.11</td>
<td>1.04 (0.99–1.09)</td>
<td>0.09</td>
<td>0.05 (Y) Einarsdottir et al. (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2046210</td>
<td>6: 151948366</td>
<td>A/G</td>
<td>0.34</td>
<td>1.00</td>
<td>Genotyped</td>
<td>1.03 (0.99–1.07)</td>
<td>0.18</td>
<td>1.03 (0.99–1.08)</td>
<td>0.17</td>
<td>0.0005 (Y) Li et al. (2011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2077647</td>
<td>6: 152129077</td>
<td>T/C</td>
<td>0.48</td>
<td>1.00</td>
<td>Genotyped</td>
<td>0.99 (0.96–1.04)</td>
<td>0.78</td>
<td>1.00 (0.96–1.04)</td>
<td>0.85</td>
<td>0.11 (Y) Sasaki et al. (2002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1801132</td>
<td>6: 152265522</td>
<td>G/C</td>
<td>0.22</td>
<td>0.98</td>
<td>Imputed</td>
<td>1.03 (0.98–1.08)</td>
<td>0.27</td>
<td>1.01 (0.96–1.07)</td>
<td>0.65</td>
<td>0.01 (Y) Sliwinski et al. (2010)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SNPs previously reported as significantly associated with endometrial cancer risk were selected from a literature search. Significance thresholds were stated as $P<0.05$ in all publications. Sample sizes for studies were as follows: Ashton, 191 cases and 291 controls; Einarsdottir, 713 cases and 1567 controls; Iwamoto, 92 cases and 65 controls; Li, 953 cases and 947 controls; Sasaki, 113 cases and 200 controls; Sliwinski, 100 cases and 100 controls; Wedren, 702 cases and 1563 controls.
estrogen exposure. However, none of the SNPs reported by these studies are correlated with any of the variants found to be associated with endometrial cancer risk ($r^2 < 0.2$). The lack of overlap between risk variants for endometrial cancer, breast cancer and risk factors associated with estrogen exposure suggest that while these risks could be mediated through the same target gene, they are working via different regulatory mechanisms in different cell types.

Using log-likelihood ratios and LD, we have identified 47 candidate causal variants located at a potential alternative promoter of ESR1, represented by lead SNP rs79575945. Bioinformatics data provide evidence that these variants reside within a putative regulatory element for ESR1 and/or other genes in this region. By cross-referencing the catalogue created using 86 cell lines by Hnisz et al. (2013), we also provide evidence that candidate causal variants lie in a region encompassing super-enhancers that target ESR1. Super-enhancers consist of large clusters of transcriptional enhancers and are associated with genes that control and define cell identity (Loven et al. 2013, Whyte et al. 2013). The presence of super-enhancers overlapping the candidate causal variants indicates the functional importance of this region. Four candidate causal variants were predicted to alter transcription factor binding by two independent programs, is-rSNP (Macintyre et al. 2010) and HaploReg (Ward & Kellis 2012). However, none of these transcription factors identified have been examined by ENCODE. There was evidence of binding of transcription factors TAF1, NFIC, TCF12, p300, TEAD4 and FOXM1 overlapping candidate causal SNPs in Ishikawa cells by ENCODE. However, the binding of these transcription factors were not found to be altered by the candidate causal SNPs using is-rSNP and HaploReg. Given transcription factor binding frequently occurs in the absence of a known motif (Kheradpour & Kellis 2014), SNP effects may not have been correctly assessed in this analysis. Functional analysis would therefore be required to assess the impact of these SNPs on transcription factor binding. Using data from HaploReg alone, candidate causal SNP rs9340770 was predicted to alter binding of p300 and ENCODE data indicates that rs9340770 is in a region bound by p300 in Ishikawa cells. SNP rs9340770 is located upstream of an alternative transcript for ESR1, and the binding of p300 suggests this could be a putative promoter for these transcripts. Further functional work is required to uncover whether this SNP is affecting the expression of these alternative transcripts by disrupting p300 binding.

Although predicted to be the target gene bioinformatically, eQTL analysis using TCGA data did not find the candidate causal SNPs to be significantly associated with ESR1 expression. This is in line with previous fine-mapping studies performed for breast cancer, where candidate causal variants have not been found to act as eQTLs for predicted target genes in breast tissue samples (Ghoussaini et al. 2014, Glubb et al. 2015). The reason for this is unclear. It is possible that the effect of candidate SNPs on expression levels cannot always be detected in tumour tissue due to tissue-heterogeneity. Furthermore, eQTLs are context-dependent and might only be expressed in certain stages of cancer development, or only when under particular stimuli.

We did find candidate causal SNPs to be significantly associated with spectrin repeat containing, nuclear envelope 1 (SYNE1) expression in endometrioid endometrial cancer tissue. SYNE1 encodes Nesprin-1 which is reported to be involved in a variety of cellular processes, including Golgi and nucleus organization and cytokinesis (Zhang et al. 2001, Gough et al. 2003, Fan & Beck 2004). Genetic variation in SYNE1 has been reported to be associated with increased risk of invasive ovarian cancer (Doherty et al. 2010). SYNE1 is frequently methylated in lung adenocarcinoma and colorectal cancer (Schueler et al. 2007, Tessema et al. 2008) and mutations in SYNE1 have been reported in colorectal cancer (Sjoblom et al. 2006). Downregulation of an N-terminal isoform of Nesprin-1, Drop1, has been observed in cancers of the uterus, cervix, kidney, thyroid, pancreas and lung (Marne et al. 2008). Interestingly, a recent study has indicated a role for Nesprin-1 in the DNA damage response pathway, and identified Nesprin-1 as interacting with mismatch repair proteins MSH2 and MSH6 (Sur et al. 2014). Given that mismatch repair deficiency is observed in up to 30% of endometrial tumours (Kanaya et al. 2003), and the eQTL data from our study, the role of SYNE1 in endometrial cancer should be explored further.

In conclusion, we have identified a single endometrial cancer risk signal, at study-wide significance, located within a potential alternative promoter for ESR1. Lead SNP, rs79575945 is also reported to be associated with expression of SYNE1, adjacent to ESR1. Given SNPs at this locus have previously been identified as predisposing to breast cancer, also a hormonally driven cancer, this study adds weight to the rationale for performing informed candidate fine-scale genetic studies across cancer types (Carvajal-Carmona et al. 2015).

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/ERC-15-0319.
Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This work was supported by the National Health and Medical Research Council of Australia (ID#1031333 to A B Spurdle, DF, A M Dunning, ID#193943 to ANECS, ID#552402, QIMR Controls); National Health and Medical Research Council of Australia Fellowship Scheme (to A B Spurdle); Principal Research Fellow of Cancer Research UK (to D F Easton); Joseph Mitchell Trust (to A M Dunning); Oxford Comprehensive Biomedical Research Centre (to I Tomlinson); The European Community’s Seventh Framework Programme (grant agreement number 22175 (HEALTH-F2-2009-22175) (COGS); Cancer Research UK (C1281/A/10118 to COGS and BCAC, C1281/A/10170, C12922/A/11174, C1281/A/12014 to COGS and BCAC, C5047/A/15007, C5047/A/16692, C1897/A/16565, C490/A/10124 to SEARCH, CORGI - NSECG, to I Tomlinson); National Institutes of Health (CA128978, R01 CA122443 to MECS and MAY, P30 CA15083 to MECS, P50 CA136393 to MECS and MAY, CAHRES); Post-Cancer GWAS Initiative (1U19 CA148537, 1U19 CA148065, 1U19 CA148112 – the GAME-ON initiative); Department of Defence (W81XWH-10-1-0341); Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer; Komen Foundation for the Cure; Breast Cancer Research Foundation (MCBCS); VicHealth (MCCS); Cancer Council Victoria (MCCS); and NHS funding to the NIHR Biomedical Research Centre (UKBGS/ICR).

Author contribution statement

Acknowledgements
The authors thank the many individuals who participated in this study and the numerous institutions and their staff that supported recruitment, detailed in full in the Supplementary Text, see section on supplementary data given at the end of this article. Control data was generated by the Welcome 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. In addition, the results published here are based partly on data generated by TCGA, established by the NCI and the National Human Genome Research Institute. The authors also thank the specimen donors and relevant research groups associated with this project.

References
Bacalh O 2013 iCOGS collection provides a collaborative model. Foreword. Nature Genetics 45 343. (doi:10.1038/ng.2592)


