Utilising Genomics for Risk Prediction and Targeted Prevention of Endometrial Cancer

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2021

CEMSEL BAFLIGIL

School of Medical Sciences
Division of Cancer Sciences

Contents

Final word count (including footnotes and endnotes): 56,172

List of Figures

List of Tables

Abbreviations

95% CI 95% confidence interval

AEH atypical endometrial hyperplasia
AKT/PKB protein kinase B
ANECS Australian National Endometrial Cancer Study
AUC area under the receiver-operator curve

BCAC Breast Cancer Association Consortium
BMI body mass index
BRC-Biobank Manchester Biomedical Research Centre Biobank
BRCA1 breast cancer type 1
BRCA1/2 breast cancer type 1/2
BRCA2 breast cancer type 2

CIGMR Centre for Integrated Genomic Medical Research
CS Cowden Syndrome
CYP19A1 cytochrome P450 19A1

D&C dilation and curettage
DETECT Developing tests for endometrial cancer detection
DNA deoxyribonucleic acid

E2C2 Epidemiology of Endometrial Cancer Consortium
EA effect allele
EAF effect allele frequency
ECAC Endometrial Cancer Association Consortium
EH endometrial hyperplasia
EIN endometrial intraepithelial neoplasia
EMBRACE Epidemiological study of Familial Breast Cancer
EMT epithelial-to-mesenchymal transition

ER oestrogen receptor

ESR1 oestrogen receptor 1

ESR2 oestrogen receptor 2

FH-Risk Family History Risk Study

FTO alpha-ketoglutarate dependent dioxygenase

GRCh37 Genome Reference Consortium human build 37

GWAS genome-wide association studies

HNF1B hepatocyte nuclear factor 1 β

HNPCC hereditary non-polyposis colorectal cancer

HRC Haplotype Reference Consortium

HRT hormone replacement therapy

HWE Hardy-Weinberg equilibrium

IBS identity-by-state

IGF insulin-like growth factor

IHC immunohistochemistry

IL interleukin

IQR interquartile range

IUS intrauterine system

KRAS kirsten rat sarcoma viral oncogene homolog

LD linkage disequilibrium

LRT likelihood ratio test

LS Lynch Syndrome

LVSI lymphovascular space invasion

MAF minor allele frequency

MC4R melanocortin 4 receptor

MDM2 murine double-minute 2 homolog

METFORMIN Metformin in non-diabetic women with endometrial cancer

MFT Manchester University NHS Foundation Trust

MGDL Manchester Genomic Diagnostic Laboratory

MIRENA Mirena for treatment of endometrial neoplastic abnormalities

MLH1 MutL homolog 1

MMR mismatch repair

MSH2 MutS homolog 2

MSI microsatellite instability

MSS microsatellite stable

NaOH sodium hydroxide

NGS next-generation sequencing

NICE National Institute for Health and Care Excellence

NSMP no specific molecular profile

NuHS Nurses' Health Study

OA other allele

OR odds ratio

PAGE Population Architecture Using Genomics and Epidemiology

PARP poly ADP-ribose polymerase

PC principal component

PCA principal component analysis

PCOS polycystic ovary syndrome

PCR polymerase chain reaction

PD1 programmed cell-death protein-1

PETALS Proportion of endometrial tumours associated with Lynch Syndrome

PGR progesterone receptor

PI3K phosphatidylinositol-3-kinase

PI3K/AKT/mTOR phosphatidylinositol-3-kinase/protein kinase B/mechanistic target of rapamycin

PIK3CA phosphatidylinositol-3-kinase catalytic subunit α

POLE DNA polymerase ϵ

PR-B progesterone receptor B

PR-A progesterone receptor A

PREMIUM Pre-surgical metformin for women with endometrial cancer: a randomised placebo controlled trial

PROCAS Predicting Risk of Breast Cancer at Screening

ProMisE Proactive Molecular Risk Classifier for Endometrial Cancer

PRS polygenic risk score

PTEN phosphatase and tensin homolog

Q-Q Quantile-Quantile

QC quality control

RB The impact of obesity and weight loss on the endometrium: a prospective cohort study

RBC red blood cell

RBP4 retinol-binding protein 4

REC Research Ethics Committee

RPM risk prediction model

RR relative risk

RRSO risk-reducing salpingo-oophorectomy

SD standard deviation

SE standard error

SGO Society of Gynecologic Oncology

SHBG sex hormone binding globulin

SIR standardised incidence ratio

SNP single nucleotide polymorphism

TCGA The Cancer Genome Atlas

TF transcription factor

TMEM18 transmembrane protein 18

TNF-α tumour necrosis factor α

TP53 tumour protein p53

Tris-HCl tris(hydroxymethyl)aminomethane hydrochloride

UKBB UK Biobank

UKCRN-ID UK Clinical Research Network identification number

WBC white blood cell

WHS Women's Health Study

Wnt/β-catenin wingless integrated homolog/beta catenin

WT1-AS WT1 Antisense RNA

Abstract

Background: Rates of endometrial cancer incidence and mortality are rising as a result of increasing rate of obesity and an aging population. Diagnosed early, endometrial cancer has an excellent prognosis but those with advanced disease have a very poor outlook. Epidemiological risk factors fail to accurately stratify women and there is no standardised screening programme for endometrial cancer. A growing body of evidence suggests that genetic factors contribute to the risk. The aim of this project was to investigate the potential benefit of genomic tools for predicting the risk of endometrial cancer to enable targeted and personalised prevention.

Methods: i) A systematic review of the literature was performed to identify a panel of robust single nucleotide polymorphisms (SNPs) contributing to endometrial cancer risk. ii) A genome-wide association study (GWAS) was conducted in a case-control study from the North West of England to identify and validate SNPs associated with endometrial cancer. iii) A polygenic risk score (PRS) was developed and refined in this cohort followed by validation in additional datasets. iv) The relationship between endometrial cancer risk and pathogenic *BRCA* variants was investigated by comparing the prevalence of cases identified in a *BRCA* carrier database to the general population. DNA sequencing was undertaken to examine the presence of any somatic pathogenic mutations.

Results: 24 SNPs most likely to influence endometrial cancer risk were identified through the systematic review. Six risk regions of suggestive significance were identified in the Manchester GWAS. 72 externally curated SNPs were investigated where a significant association was confirmed for ten SNPs. A refined PRS consisting of 40 SNPs was developed in our independent cohort resulting in 62% discriminatory ability. The PRS was applied to two other datasets with low to moderate success. No evidence for an increased risk of endometrial cancer in *BRCA* pathogenic carriers was observed. No pathogenic *BRCA* somatic mutations were identified in serous subtype tumours.

Conclusions: In this project, we report strong evidence in favour of SNPs influencing endometrial cancer risk and have independently validated the most robust SNPs. A PRS based on the most predictive SNPs was moderately capable of risk stratification at a level similar to published multivariable risk prediction models. If successfully validated, the PRS may be useful in improving the accuracy of risk prediction models in risk-based care. Risk-reduction measures and incorporation into risk prediction models are not warranted for carriers of pathogenic *BRCA* mutations.

Declaration

I hereby declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

This thesis is the output of three years of hard work and resilience which would not have been possible without the support of several people to whom I am forever grateful.

First and foremost, I would like to express my deepest gratitude and appreciation to my supervisory team: Prof Emma Crosbie, Prof Gareth Evans, Dr Miriam Smith and Dr Artitaya Lophatananon. Emma, special thanks to you for your optimism, enthusiasm and encouragement throughout my journey. I also would like to thank my advisor Dr Emma Woodward for her valuable insight along the way. Your guidance has been invaluable and I am grateful for having the opportunity to learn from all of you.

My heartfelt thanks to Dr Deborah Thompson and Prof Douglas Easton for welcoming me to their lab at University of Cambridge for three months. Deborah, I am most grateful to you for providing me with supervision and support during my placement. Your wealth of expertise and knowledge has no match. I would also like to thank Joe Dennis for tirelessly answering my bioinformatics-related questions.

I am incredibly grateful to Suzanne Carter for helping me find my precious DNA samples which mysteriously disappeared in transit.

I consider myself extremely fortunate to have a strong support network outside of work, all of whom emotionally supported me during this journey.

I am immensely grateful to my parents, Cemaliye and Ersel, whose undying belief in me has been the most precious gift. Thank you for your unconditional support for achieving my dreams. I cannot express how profoundly I appreciate my brother Ercem who is always there for me in times of need and lifts my spirits. I am most grateful to my fiancé, Visa, who kept me motivated and shared this journey with me. You are my rock, my calmer half and I consider myself incredibly lucky to have you by my side. My dearest friend and PhD buddy Cemre Su Osam. Your friendship has been a blessing and I am very much looking forward to sharing more joyful moments with you in the future.

Last but not least, this work would not have been possible without the financial support from the NIHR Manchester Biomedical Research Centre.

Preface and Thesis Structure

This thesis is submitted in the *alternative* format supported by the fact that the area of research studied is well-suited for publication of academic papers in high-impact, peer-reviewed journals. There are four results chapters presented in this thesis, which will or have resulted in publication of three substantial research papers where the author of this thesis is the first author on two and second author on one the publications. The author has led study designs, data collection and analysis, conducted laboratory and bioinformatics work, and drafted or, in the case of second authorship, contributed to the manuscripts.

Each of the results chapters presented in this thesis include a brief introduction to the chapter and the paper and a statement highlighting the contributions of each author. Two of the papers have already been published (Journal of Medical Genetics and European Journal of Cancer) and the other is currently being prepared for submission in a high-impact journal.

The paper detailing the systematic review presented in **Chapter 2**, is largely unaltered with the exception of correcting typing errors and editing the tables, figures and references to fit the format of the thesis. **Chapters 3** and **4** together form a single paper, however, in this thesis, the methods, results, discussion and any relevant table, figure or reference has been further extended to provide PhD level detail for the purpose of this thesis. The methods section of the final paper presented in **Chapter 5** has been expanded to demonstrate sufficient detail for better clarity.

This thesis is formed of eight chapters numbered 1 to 8, each with numbered (sub)sections (e.g. 1.1). **Chapter 1** introduces the existing literature on the growing burden of endometrial cancer, lack of standard screening and accurate prevention tools, the role of low-risk, common susceptibility variants, and gaps in our knowledge regarding genetic risk factors in this context. Lastly, this Chapter provides an overview of the research aims of this thesis.

Chapters 2-5 consist of the main thesis results. **Chapter 2** investigates the role of single nucleotide polymorphisms in modifying endometrial cancer risk by systematic review of literature and presents a panel of variants with the most robust evidence. **Chapter 3** presents a new genome-wide association study conducted on a unique cohort and investigates the association between common variants and endometrial cancer risk. **Chapter 4** presents the development and validation of an extended polygenic risk score for risk prediction purposes. **Chapter 5** investigates the assumed link between *BRCA* carrier status and increased endometrial cancer risk in a prospective cohort. Finally, **Chapter 6** summarises the key findings of this thesis, highlights the strengths and limitations, and discusses the findings and their contribution to the literature in the context of early detection and prevention of endometrial cancer. The published papers presented in **Chapters 2** and **5** are provided as published/unaltered form in the **Appendices**.

Research Output

Publications

1. **Bafligil, C**., Thompson, D. J., Lophatananon, A., Smith, M. J., Ryan, N. A., Naqvi, A., Evans, D. G. and Crosbie, E. J. (2020), 'Association between genetic polymorphisms and endometrial cancer risk: a systematic review', *Journal of Medical Genetics* **57**(9), 591-600.

   Contribution: I conducted this systematic review of literature, analysed the data and drafted the paper. This paper forms a key part of my project for providing an essential step for selection of single nucleotide polymorphisms for the polygenic risk score panel guided by the best evidence available.

2. Kitson, S., **Bafligil, C**., Ryan, N., Lalloo, F., Woodward, E., Clayton, R., Edmondson, R., Bolton, J., Crosbie, E., & Evans, D. G. (2020). '*BRCA1* and *BRCA2* pathogenic variant carriers and endometrial cancer risk-a cohort study', *European Journal of Cancer* **136**, 169-175.

   Contribution: I identified and prepared the DNA samples for the somatic *BRCA* mutation screenings for 15 high-grade serous endometrial cancers, interpreted the relevant data and contributed to the writing of this paper.

Presentations

1. **A polygenic risk score-based genetic stratification for endometrial cancer**.
   Cemsel Bafligil, Deborah J Thompson, Artitaya Lophatananon, Miriam Smith, Joe Dennis, D Gareth Evans, Emma J Crosbie
   Poster presentation at American Association for Cancer Research Annual Meeting, *June 2020*, Philadelphia, PA.

2. **Utilising single nucleotide polymorphisms to predict endometrial cancer**.
   Cemsel Bafligil

Oral presentation at The University of Manchester Division of Cancer Sciences Internal Seminar Series, *September 2020*, Manchester, UK.

3. **Constructing a panel of common predisposition variants to predict women that are at high-risk of developing endometrial cancer**.
   Cemsel Bafligil, Deborah Thompson, Miriam Smith, Artitaya Lophatananon, Neil Ryan, Anie Naqvi, Gareth D Evans, Emma J Crosbie
   Poster presentation at Society for Reproductive Investigation Annual Scientific Meeting, *March 2019*, Paris, France.

4. **Constructing a panel of common predisposition variants to predict women that are at high-risk of developing endometrial cancer**.
   Cemsel Bafligil, Deborah Thompson, Miriam Smith, Artitaya Lophatananon, Neil Ryan, Anie Naqvi, Gareth D Evans, Emma J Crosbie
   Two poster presentations at The University of Manchester Doctoral Academy Graduate Society PhD Conference and Postgraduate Summer Research Showcase, *June 2018*, Manchester, UK.

5. **Developing an endometrial cancer risk prediction model for targeted prevention strategies**.
   Cemsel Bafligil
   Oral presentation at NIHR Manchester BRC Cancer Prevention and Early Detection Showcase, *June 2018*, Manchester, UK.

Awards

1. University of Manchester Division of Cancer Sciences Internal Poster Session Prize Winner, *2020*.

2. NIHR Short Placement Award for Research Collaboration (SPARC) Round 5, *2019*.
   Placement title: "Utilising genomics data for early detection of endometrial cancer",

supervised by Prof. Douglas Easton and Dr. Deborah Thompson at Cancer Genetic Epidemiology, University of Cambridge.

3. NIHR Manchester BRC Cancer Prevention and Early Detection Theme Public Engagement Fund, *2018*.
   Public engagement video about endometrial cancer and brief introduction to PhD project (`https://youtu.be/2JBCyxAxQO8`).

1 Introduction

1.1 Epidemiology of Endometrial Cancer

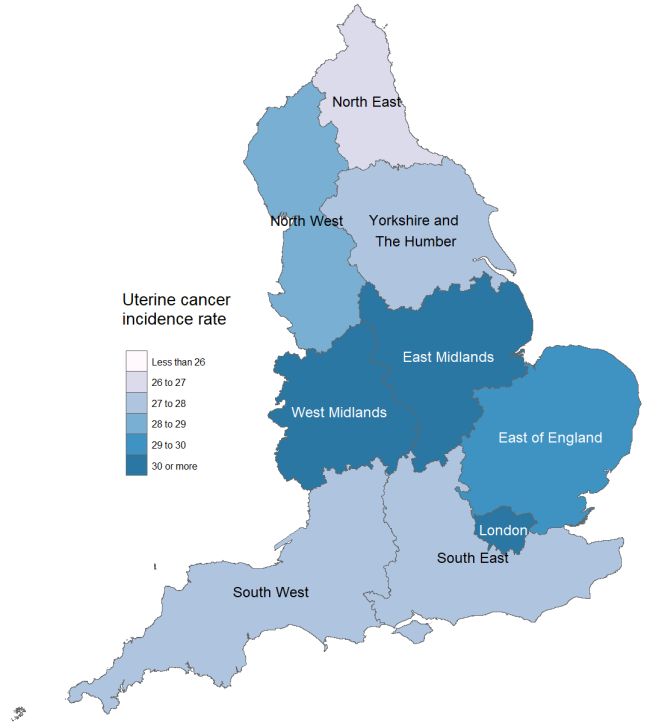
Endometrial cancer is the most common gynaecological cancer in developed countries, particularly in the United Kingdom and United States (Sundar *et al*., 2017; Siegel *et al*., 2019) (**Figure 1.1**). According to Cancer Research UK, roughly 9,500 new cases were diagnosed in 2017 alone, which accounts for 5% all new cancer diagnoses in females in the United Kingdom (CRUK, 2020). The incidence of endometrial cancer has more than doubled since the 1990s and has risen by roughly 15% within the last decade alone.

Despite increased complexity of management, mortality rates for endometrial cancer have also been rising with on average more than 2,300 women in the United Kingdom dying from endometrial cancer each year. Most women are diagnosed with endometrial cancer in its early stages when expected survival is highest (>90%); however, this falls dramatically to 15% when diagnosed at an advanced stage. More than half of new cases are diagnosed in women aged 65 and over, who often have other concurrent health concerns that may contribute to a less favourable prognosis.

1.2 Aetiology of Endometrial Cancer

Endometrial cancer arises from the abnormal growth of cells lining the uterus, called the endometrium. The endometrium is a heterogeneous tissue comprised of luminal and glandular epithelial cells embedded in a layer of stromal cells. It is a highly dynamic tissue that undergoes monthly cyclical changes induced by sex steroids during the childbearing years (Mihm *et al*., 2011). Oestrogen has a mitogenic effect on the endometrium whereas progesterone counterbalances this by inhibiting growth and inducing differentiation of glands and stroma (Cooke *et al*., 1997).

(a) (b)



(c)

Figure 1.1: *Incidence and projected mortality of endometrial cancer are on the rise.*

(a) Incidence and (b) mortality rates of endometrial cancer. (c) Age-standardised incidence rate (per 100,000) of endometrial cancer in England by region in 2017. b) Solid line represents the observed and dashed line represents the projected rates. Data obtained from Office for National Statistics (2019) and CRUK (2020). ASR, age-standardised rate.

Table 1.1: *Classification of endometrial cancer subtypes according to the dualistic model, adapted from Morice et al. (2016).*

	Type I	Type II
Associated Clinical Features	*Metabolic syndrome*: obesity, hyperlipidaemia, hyperglycaemia, and increased oestrogen concentrations	None
Grade	Low	High
Hormone Receptor Expression	Positive	Negative
Genomic Stability	Diploid, frequent MSI (40%)	Aneuploid
***TP53*mutation**	No	Yes
Prognosis	Good: overall 85% 5-year survival	Poor: overall 55% 5-year survival

MSI, microsatellite instability; *TP53*, tumour protein p53

1.3 Endometrial Cancer Subtypes

1.3.1 The Dualistic Model

Endometrial cancer is categorised into Type I and Type II tumours based on histomorphologic, clinical and endocrine characteristics (Table **1.1**) (Morice *et al*., 2016). The classification is based on a *dualistic model* first proposed by Bokhman in 1983 (Bokhman, 1983). Type I endometrial cancers are the most common subtype, accounting for roughly 85% of sporadic endometrial cancer cases and are of epithelial origin. These are generally defined as oestrogen-driven, may be preceded by atypical endometrial hyperplasia (AEH) and are often associated with low grade disease that has a favourable prognosis. On the other hand, Type II endometrial cancers account for 10-20% of endometrial cancers and are dominated by serous and clear cell carcinomas. Type II endometrial cancers are more aggressive and metastatic in nature, where oestrogen is not the driver of carcinogenesis. Although these tumours were thought to arise from atrophic endometrium, there is now growing evidence that they may arise from areas of dysplasia and hypertrophy (Fadare *et al*., 2006).

1.3.2 Endometrial Hyperplasia

Endometrial hyperplasia (EH) is a benign condition resulting in thickening of the endometrium, whereby endometrial stroma and glands are overgrown, with or without presence of atypia (Rosen *et al.*, 2019). EH occurs due to a hyperoestrogenic state of the endometrium such as in chronic anovulation or obesity. AEH, also known as endometrial intraepithelial neoplasia (EIN), is a monoclonal premalignant lesion developing from endometrial glands which has distinct histomorphologic features. AEH has a substantial risk of progressing to Type I endometrial cancer.

1.3.3 Molecular Classification

Although Bokhman's dualistic classification model is widely accepted, evidence from molecular studies point towards a more complicated and sub-structured model. Accordingly, Type I tumours often exhibit mutations in phosphatidylinositol-3-kinase/protein kinase B/mechanistic target of rapamycin (PI3K/AKT/mTOR) and wingless integrated homolog/beta catenin (Wnt/β-catenin) pathways as well as microsatellite instability (MSI) (Wilczynski *et al.*, 2016). Mutations in phosphatase and tensin homolog (*PTEN*), a tumour suppressor gene, have been linked to development of various forms of cancer including endometrial cancer. Another tumour suppressor gene, tumour protein p53 (*TP53*), is often found to be mutated in Type II endometrial cancers, particularly in serous carcinomas. However, it is also reported that these mutations are not mutually exclusive for different types of endometrial cancer (Talhouk and McAlpine, 2016). *PTEN* mutations may be seen in a small number of Type II endometrial cancers and, likewise, *TP53* mutations may be present in Type I endometrial cancers. Moreover, Setiawan and colleagues suggested that the aetiology of Type I and Type II endometrial cancers may not be as distinct as previously thought (Setiawan *et al.*, 2013). According to their pooled analysis from 10 cohort and 14 case-control studies, many of the risk factors analysed had comparable effects on both subtypes, with the exception of body mass index (BMI) which had a greater effect on Type I endometrial cancers.

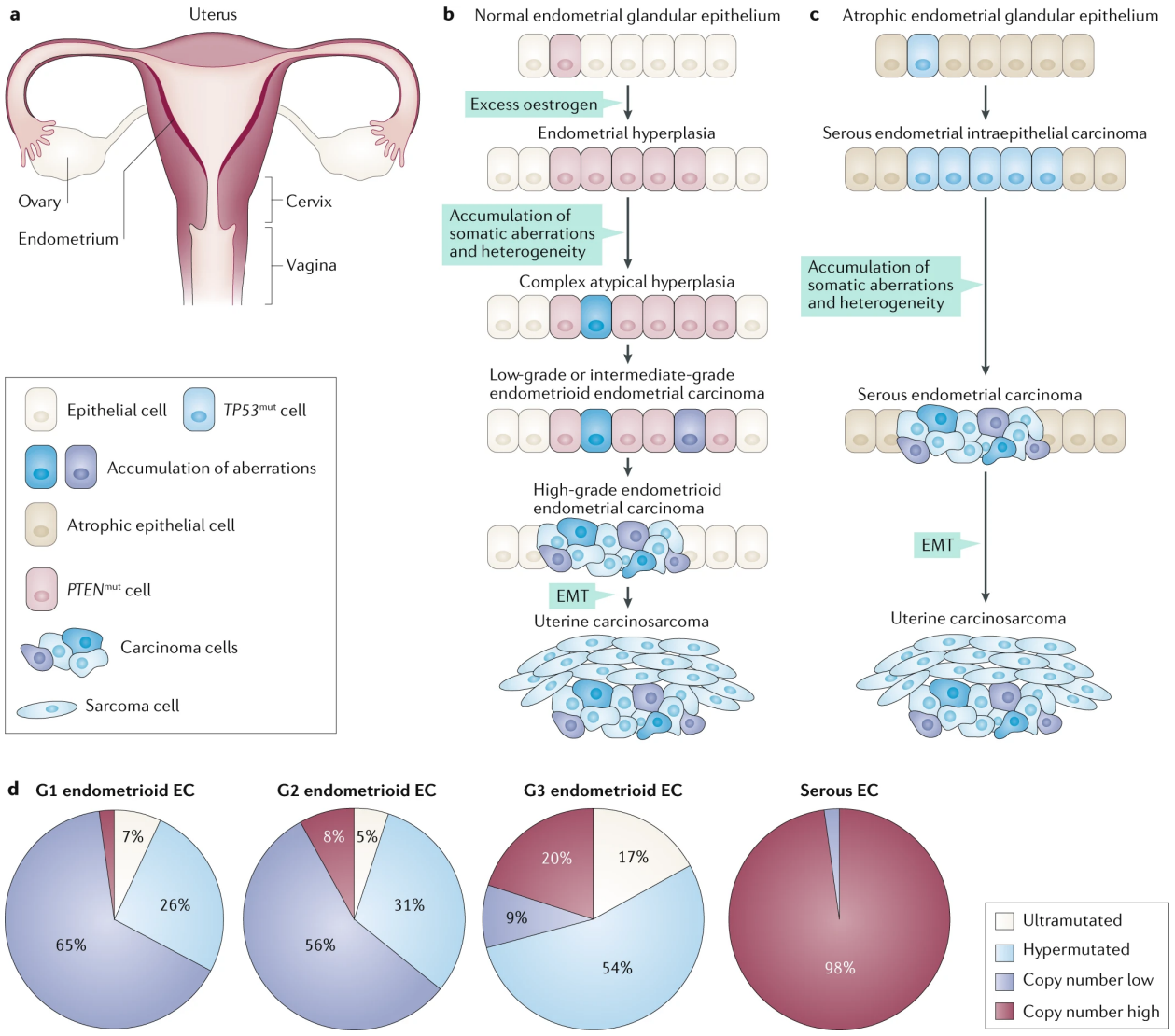


Figure 1.2: *Summary schematics of endometrial cancer aetiology and progression.*

a) Female reproductive system. **b** Illustration of how endometrioid endometrial cancer arises through precursor lesions; EH and AEH. **c)** Illustration of how serous endometrial cancer arise from the atrophic endometrium. Cells with acquired somatic mutations are coloured. If high grade carcinomas undergo an epithelial-to-mesenchymal transition (EMT), they give rise to uterine carcinosarcomas. **d)** Distribution of low- and intermediate-grade endometrioid, high-grade endometrioid and serous ECs among the Cancer Genome Atlas classification of molecular subgroups. Figure taken from Urick and Bell (2019).

Recognising the complexity of this disease, research efforts have been diverted to elucidate a more comprehensive classification method for endometrial cancer subtypes. A study by The Cancer Genome Atlas (TCGA) Research Network classified the endometrial cancer tumours into four major subtypes based on their distinct molecular profiles; i) DNA polymerase ϵ (*POLE*) (ultramutated) with an excessively high mutation rate and widespread DNA polymerase ϵ (*POLE*) mutations, ii) MSI (hypermutated) with high mutation rate and mismatch repair (MMR) deficiency, predominantly through MutL homolog 1 (*MLH1*) promoter methylation, iii) copy number-low (endometrioid) primarily comprising of microsatellite stable (MSS) endometrioid tumours, and iv) copy number-high (serous-like) consisting primarily of serous-like tumours with frequent *TP53* mutation and harbouring extensive somatic copy-number alterations (**Figure 1.3 a,b,d**) (Getz *et al.*, 2013). According to their results, the poorest survival observed was for the copy-number high (serous-like) subgroup whereas *POLE* (ultramutated) tumours showed the highest survival among all four subgroups (**Figure 1.3 c**).

A low-cost genomics-based molecular classification system, Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE), based on the TCGA study was developed by Talhouk and colleagues and later validated (Talhouk *et al.*, 2015, 2017; Kommoss *et al.*, 2018). In this system, MMR status and *TP53* mutations were investigated by immunohistochemistry (IHC) whereas *POLE* mutations were detected through sequencing. Here, *TP53* IHC was used as a cheaper alternative to act as a surrogate for the copy-number analysis conducted by TCGA. Accordingly, this system identified four molecular subgroups efficiently and reported similar survival rates to those described by the TCGA study (**Figure 1.6**).

Similar to the TCGA and ProMisE studies, the TransPORTEC international consortium also identified four distinct subgroups; p53 mutant, MSI, *POLE* mutant, and no specific molecular profile (NSMP) tumours (Stelloo *et al.*, 2015, 2016). Using PORTEC-1 and PORTEC-2 trials, the consortium evaluated the molecular subgroups in early-stage en-

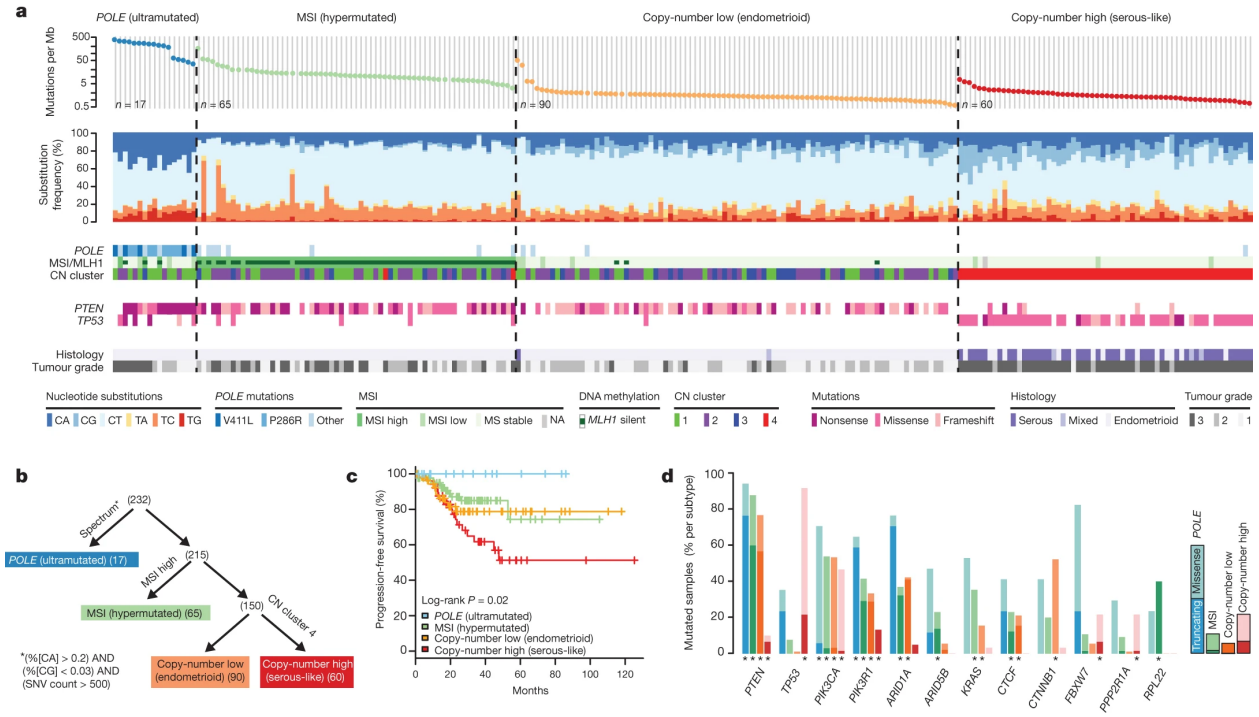


Figure 1.3: *Categorisation of endometrial cancer tumour subtypes according to molecular profiling by single nucleotide polymorphism-based copy-number analysis and exome sequencing by the TGCA.*

(a) Mutational landscape of the tumours. (b) TCGA classification system and subgroups identified. (c) Kaplan-Meier progression-free survival analysis of the four tumour subgroups. (d) Four subgroups harbour frequent mutations in different genes. Figure taken from Getz *et al*. (2013)

dometrioid endometrial cancers (Stelloo *et al.*, 2016). According to their results, p53 mutant subgroup was associated with high-grade tumours, hormone receptor loss and other mutations including >10% *L1CAM* (encodes a cell adhesion molecule). The latter was also associated with distant recurrence. Lymphovascular space invasion (LVSI) and abnormal *ARID1a* expression was common in MSI tumours. On the other hand, *POLE* mutant subgroup was more commonly seen in younger women with high-grade tumours and often co-existing with *PTEN* mutations. Finally, NSMP tumours were commonly low-grade and harboured mutations in *CTNNB1*. In this study, distant recurrence and endometrial cancer-associated mortality rates were comparable among the four subgroups.

Unlike the TCGA, the more recent TransPORTEC study using PORTEC-3 trial also included clear cell tumours in their analysis (Stelloo *et al.*, 2015). Here, the authors investigated the clinical outcome of the four molecular subgroups in terms of recurrence and distant metastasis. They reported that *POLE* mutant and MSI tumours showed no distant metastasis whereas this was observed in half and 39% of the p53 mutant and NSMP tumours, respectively (**Figure 1.4**). Five-year recurrence-free survival was over 93% for the *POLE* mutant and MSI tumours as opposed to 42% and 52% for the p53 mutant and NSMP tumours, respectively.

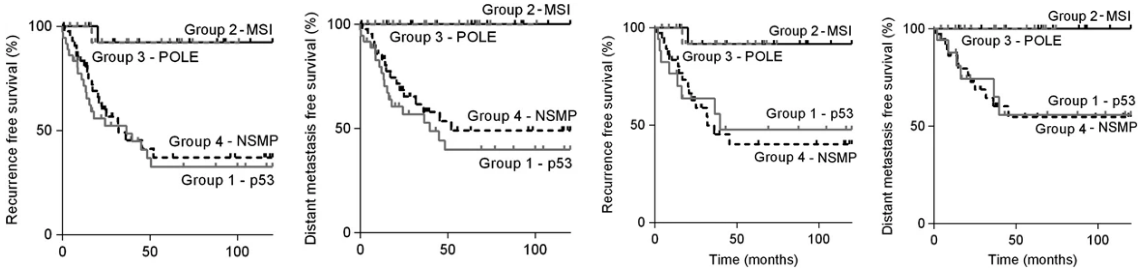


Figure 1.4: *Clinical outcome in correlation to the four molecular subgroups.*

Recurrence- and distant metastasis-free survival of (a) all and (b) endometrioid high-risk patients. Figure adapted from Stelloo *et al.* (2015).

Accurate molecular classification of endometrial cancer tumours is crucial as it offers

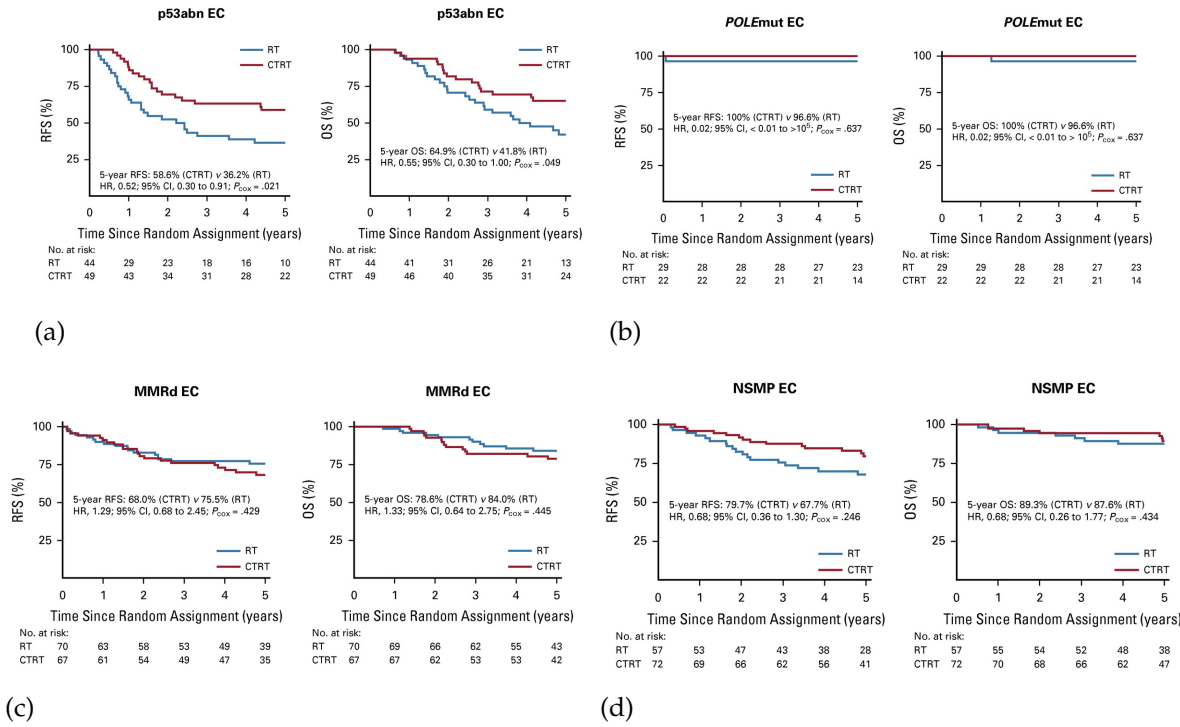


Figure 1.5: *Adjuvant treatment response in four molecular subgroups.*

Recurrence-free and overall survival of (a) p53 mutant, (b) *POLE* mutant, (c) MMR deficient and (d) NSMP ECs. CTRT, combined chemo/radio-therapy; RT, radiotherapy alone. Figure adapted from Leon-Castillo *et al*. (2020).

efficient prognostic risk assessment in terms of both clinical outcome and therapy response. Most recently, the molecular classification system was investigated in relation to adjuvant therapy in PORTEC-3 trial participants (Leon-Castillo *et al*., 2020). Accordingly, combined adjuvant chemo- and radiotherapy was most beneficial for the p53 mutant subgroup (**Figure 1.5** a) whereas radiotherapy alone was more beneficial to the MMR deficient subgroup (**Figure 1.5** c). Combined adjuvant therapy seemed to be more beneficial in NSMP group, though this association was not statistically significant (**Figure 1.5** d). Finally, *POLE* subgroup had excellent survival rate in both radiotherapy alone and combined therapy methods (**Figure 1.5** b). The former only achieved a slightly lower survival rate due to one patient experiencing recurrence after treatment. This is particularly important as the molecular classification of endometrial tumours may aid targeted therapy efforts to maximise treatment efficiency while balancing over- and under-treatment of different

tumour subtypes. Details and a comparative summary of the aforementioned studies is presented in **Figure 1.6**.

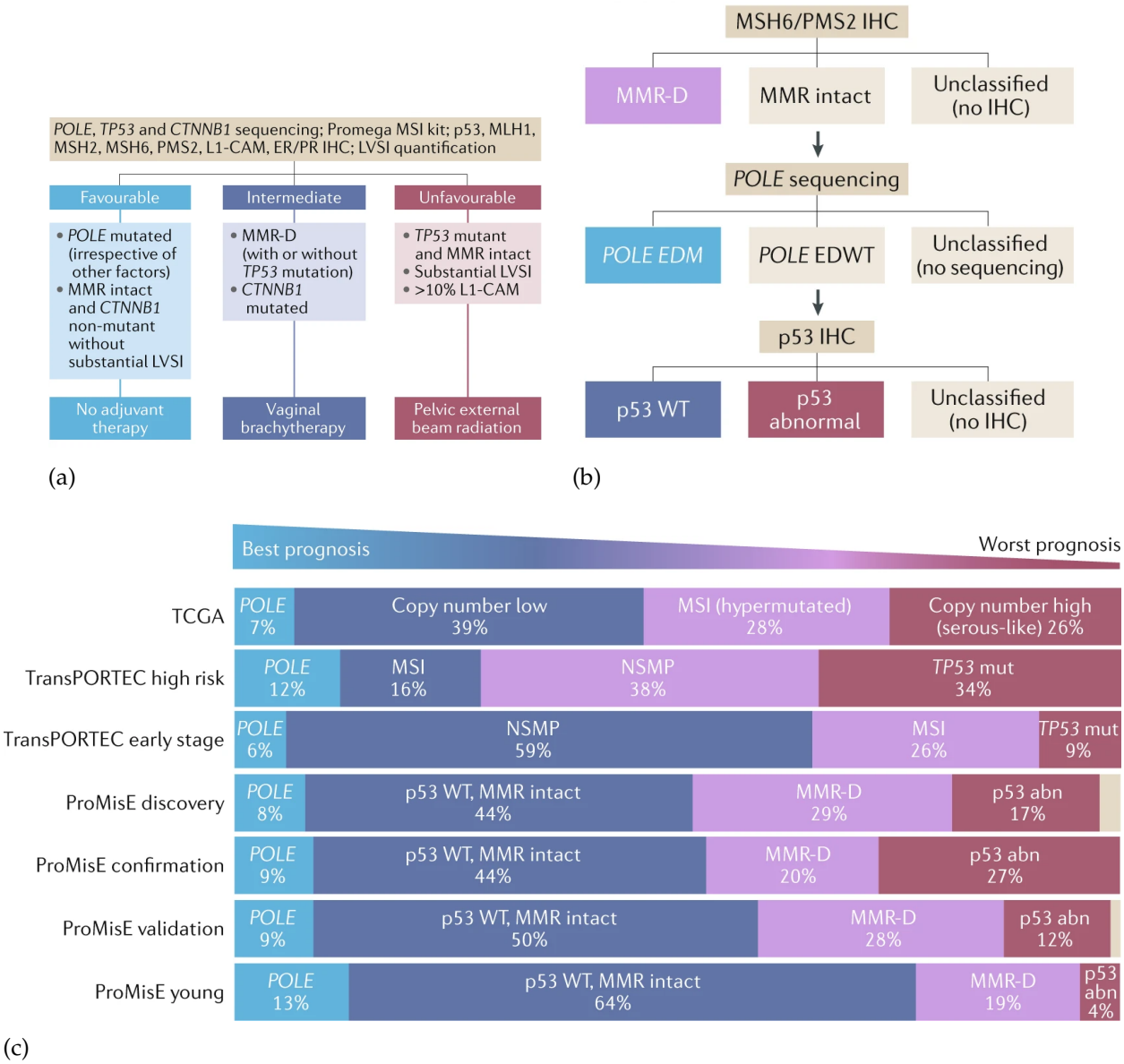


Figure 1.6: *Molecular subgroup classification systems for risk and therapy stratification.*

(a) TransPORTEC and (b) ProMisE molecular classification systems. (c) Prognostic comparison of the TCGA, TransPORTEC and ProMisE systems. Figure adapted from Urick and Bell (2019).

1.4 Clinical Features of Endometrial Cancer

Postmenopausal or abnormal bleeding is the most characteristic symptom of endometrial cancer (Sundar *et al*., 2017) while abnormal discharge or spotting may also occur to a less extent. Pelvic pain and weight loss may be experienced, most commonly in later stages. However, all these symptoms are also shared with benign gynaecological conditions. As such, only up to 10% of women presenting with abnormal bleeding receive an endometrial cancer diagnosis.

1.4.1 Diagnosis and Screening

Transvaginal ultrasound is often the first step in examining the uterus and nearby reproductive organs (Cooper *et al*., 2014; Schramm *et al*., 2017). Further tests include outpatient hysteroscopy and pathological examination of the endometrial biopsies. However, these procedures are expensive, invasive and/or intolerable for some patients (Urick and Bell, 2019). In the case of ultrasound, this procedure has poor specificity for diagnostic purposes particularly in asymptomatic women (Sundar *et al*., 2017).

Historically, dilation and curettage (D&C) was the preferred method for obtaining endometrial samples for pathological assessment. However, this method requires use of anaesthetics, is costly, and is associated with risks of infection and uterine perforation. Therefore, less invasive and cheaper alternatives have been developed for endometrial sampling such as the Pipelle or Tao brush. Although these are prone to missing the cancer and sufficient material may not be obtained for histopathologic analysis (van Hanegem *et al*., 2016), associated risks, duration and cost of these methods outweigh the drawbacks.

Detection of common endometrial cancer-associated mutations in the endometrium can also be achieved through minimally-invasive endometrial sampling, offering potential for an endometrial cancer screening tool. Two of the methods used for endometrial sampling are brush-assisted and uterine lavage sampling. These facilitate collection of endometrial cells from the lumen of the uterus which can then be tested for mutations and/or ploidy

status in a given set of genes. PapSEEK test using samples obtained from endometrial sampling can be used to test for the commonly mutated genes such as *POLE* and *TP53* (reviewed by Urick and Bell (2019)). However, the sensitivity and specificity of these are within the range of 93-100% and 46% to 100% depending on the sampling method used (Nair *et al*., 2016; Wang *et al*., 2018). Newer methods that are cost-effective and minimally-invasive are being trialled. One example is the use of blood-based infrared spectroscopy which can identify endometrial cancer with 87% sensitivity and 78% specificity (Paraskevaidi *et al*., 2020). This method investigates all the molecules, such as lipids, DNA and proteins, within the blood specimens.

Despite the increasing rates seen for endometrial cancer globally, no standard guidelines exist for screening. This may be partly due to the lack of sensitivity and specificity of some of the commonly available methods such as transvaginal ultrasound, or their costly and invasive nature such as the D&C and hysteroscopy. Although endometrial biopsy is the most accurate method, as discussed above, the majority of patients report considerable pain and access to the uterus may be difficult due to cervical or atrophy-related factors at the time of the procedure. There is also an acknowledged sampling error associated with this method which may result in missing the cancerous tissue. Thus, these methods are not preferred to use in the general population for screening purposes. A small number of guidelines exist from professional societies and are summarised in **Table 1.2** (Gentry-Maharaj and Karpinskyj, 2020). There is no evidence that screening improves survival outcomes from the disease. Therefore, utilising screening alongside effective risk prediction tools and prevention strategies will undoubtedly have a greater impact on survival and endometrial cancer-related mortality as well as reducing the alarming rise in the incidence.

Although there is a clear lack of consensus and standardised guidelines for diagnosis and screening, a recent testing guideline was published by the National Institute for Health and Care Excellence (NICE) in the United Kingdom in late 2020 (NICE, 2020).

Table 1.2: *Screening guidelines suggested by professional societies.*

Table adapted from Gentry-Maharaj and Karpinskyj (2020).

Risk	Definition	Mode of screening/prevention				Society
		No screening (Advise to visit GP/family physician if experience PMB, advise of increased risk after menopause)	Annual endometrial biopsy from age 35	Annual TVS from age 35	Hysterectomy (with salpingo-oophorectomy) from age 40	
Average	Population level, 3%	✓				ACS BGCS ESGO
Intermediate	↑ unopposed oestrogens, no family history (<10% risk)	✓				ACS BGCS ESGO
High	LS or family history (>10% risk)	✓				ACS
		✓	✓			BGCS
		✓	✓	✓		ESGO

ACS, American Cancer Society; BGCS, British Gynaecological Cancer Society; ESGO, European Society of Gynaecological Oncology; PMB, post-menopausal bleeding; TVS, transvaginal ultrasound.

The new guideline recommends testing for Lynch Syndrome (LS), an inherited genetic condition discussed in further detail in **Section 1.5.6.1, p 52**, for all women who receive an endometrial cancer diagnosis. As endometrial cancer is often the first cancer to develop in those with LS, the new guideline will likely result in more LS diagnoses, which will in turn result in requirement for targeted screening and risk-reducing strategies for other family members identified through cascade-testing.

1.4.2 Treatment

Different management options for endometrial cancer are employed depending on the type and extent of the disease as well as general health of the individual at the time of treatment (Sundar *et al.*, 2017). The standard treatment is total hysterectomy and bilateral salpingo-oophorectomy with or without adjuvant radiotherapy and/or chemotherapy. Hysterectomy may be accompanied by peritoneal lavage, omental biopsy and pelvic and para-aortic lymph node dissection depending on the histological subtype and likely stage of the disease. Main risks of surgery are bleeding, infection, and venous thromboem-

bolism; risks that are increased in morbidly obese women. Minimally invasive laparoscopic surgery has been reported to be less risky in terms of post-operative complications, however, overall survival remains the same as by traditional laparotomy (Walker *et al*., 2009; Janda *et al*., 2010, 2017). Following the operation, FIGO 2009 staging system is used to determine the stage of the cancer and direct adjuvant therapy. Chemotherapy and/or radiotherapy may be required in place of surgery for advanced or inoperable endometrial cancer.

1.4.2.1 Adjuvant Therapy

Adjuvant therapy is generally not required for women with early stage and early grade endometrioid tumours with no substantial myometrial invasion (Lu and Broaddus, 2020). Systemic chemotherapy (carboplatin and paclitaxel) and vaginal brachytherapy is offered to women with early stage but more aggressive tumours such as carcinosarcomas and serous tumours. PORTEC-3 trial has shown that combined chemo-and radiotherapy extends a greater survival benefit for patients with node-positive disease and molecular subtyping may aid in calibrating the regimen (Randall *et al*., 2019). For aggressive and recurrent disease, a combined chemotherapy regimen of paclitaxel and carboplatin is standard care whereas for serous tumours trastuzumab may also be added for a three-drug therapy (Fader *et al*., 2018). In contrast, for recurrent endometrioid tumours, hormonal agents may be used as a second- or third-line treatment (Carlson *et al*., 2014). For tumours with high MSI, an immune checkpoint inhibitor (pembrolizumab) may be an effective second-line treatment (Marabelle *et al*., 2019). High-grade tumours not exhibiting MSI may be treated with lenvatinib, a multi-tyrosine kinase inhibitor, and pembrolizumab (Makker *et al*., 2017).

Chemotherapy may be used for advanced endometrial cancer as a means to relieve symptoms and slow tumour growth and/or spread. Although chemotherapy is conventionally used to treat many forms of cancer, its many side effects are less than desirable and might lead to serious complications. Chemotherapy drugs are known to cause reduced immu-

nity, gastrointestinal problems, fatigue, and hair loss, bringing about emotional distress. Commonly, radiotherapy is used after surgery for most cases, and for recurrent and/or metastatic disease to treat an isolated pelvic recurrence with curative intent or relieve symptoms. Radiotherapy can either be external or internal, or both, depending on each individual case. Although internal radiotherapy has fewer side effects, it is only suitable for delivering radiotherapy to the vaginal vault, and both routes of delivery cause bowel, bladder and vaginal complications. Thus, newer and more convenient forms of treatments are needed to address the growing burden of cancer.

1.4.2.2 Newer Therapies

Detection of clinically actionable mutations will undeniably support the use of targeted therapy strategies. Though the potential of targeted therapies may be hindered by lack of comparable benefit on synonymous genetic mutations in different forms of cancer or different mutations of the same gene, improved outcomes have been reported for advanced cancer patients as a result of targeted therapies (Stockley *et al*., 2016; Rodon *et al*., 2018). A summary of the most frequently mutated genes in endometrial cancer which may be used for targeted therapies is presented in **Figure 1.7**.

Immunotherapy, including drugs such as pembrolizumab, a programmed cell-death protein-1 (PD1) signalling pathway inhibitor, holds great potential for endometrial cancer treatment. Pembrolizumab, for instance, is approved for the use of MMR deficient tumours and has been effective in a subset of endometrial cancer patients (Ott *et al*., 2017; Chan *et al*., 2020). Other immunotherapeutic approaches, which stimulate immune response against tumour cells, are categorised into two classes; passive and active therapy. Passive therapy involves introduction of external immunomodulatory agents such as checkpoint inhibitors, whereas active therapy seeks to enhance or stimulate an individual's immune system *in vivo* such as via anticancer vaccines (**Figure 1.8**). Ohno *et al*. evaluated the use of a vaccine against WT1 in a small subset of endometrial cancer patients and reported

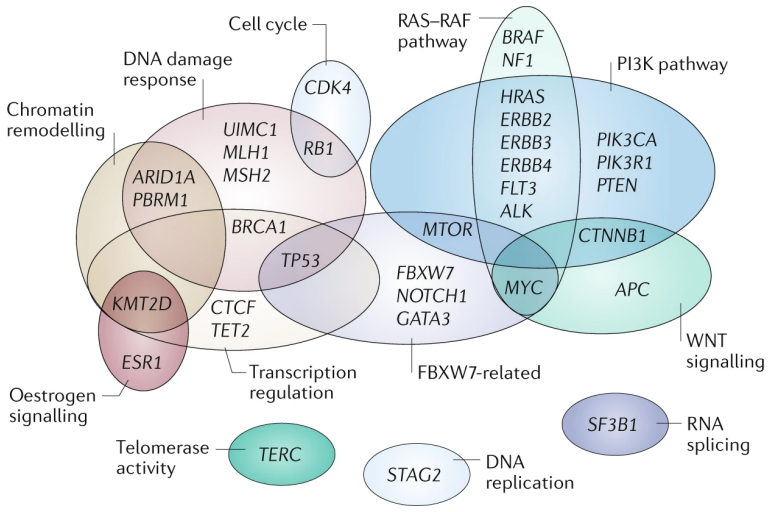


Figure 1.7: *Groups of genes seen with frequent mutations in metastatic endometrial cancer*.

Figure taken from Urick and Bell (2019).

modest disease control (Ohno *et al*., 2009). This particular vaccine may be important as one of the top single nucleotide polymorphisms (SNPs) reported to be associated with endometrial cancer risk was found near *WT1* and WT1 Antisense RNA (WT1-AS) (please refer to **Section 2, Table 2.1, p 84**) (O'Mara *et al*., 2018; Bafligil *et al*., 2020).

Other novel drugs including those that target angiogenesis and PI3K/AKT/mTOR pathway are also being trialled (Tran and Gehrig, 2017). Lenvatinib, a multikinase-inhibitor which has antiangiogenic activity, had moderate success in a small cohort of metastatic endometrial cancer patients when used together with pembrolizumab (Makker *et al*., 2017). Moreover, poly ADP-ribose polymerase (PARP) inhibitors are emerging anticancer therapeutics particularly in tumours with deoxyribonucleic acid (DNA) double-strand break repair deficiencies, also known as homologous recombination deficiency. This deficiency has been commonly associated with breast cancer type 1/2 (*BRCA1/2*) mutations but recently been noted to also be associated with non-endometrioid endometrial cancers and p53 mutant tumours (Ledermann *et al*., 2016; de Jonge *et al*., 2019a). Thus, PARP inhibitors may be useful in a small subset of endometrial cancer cases.

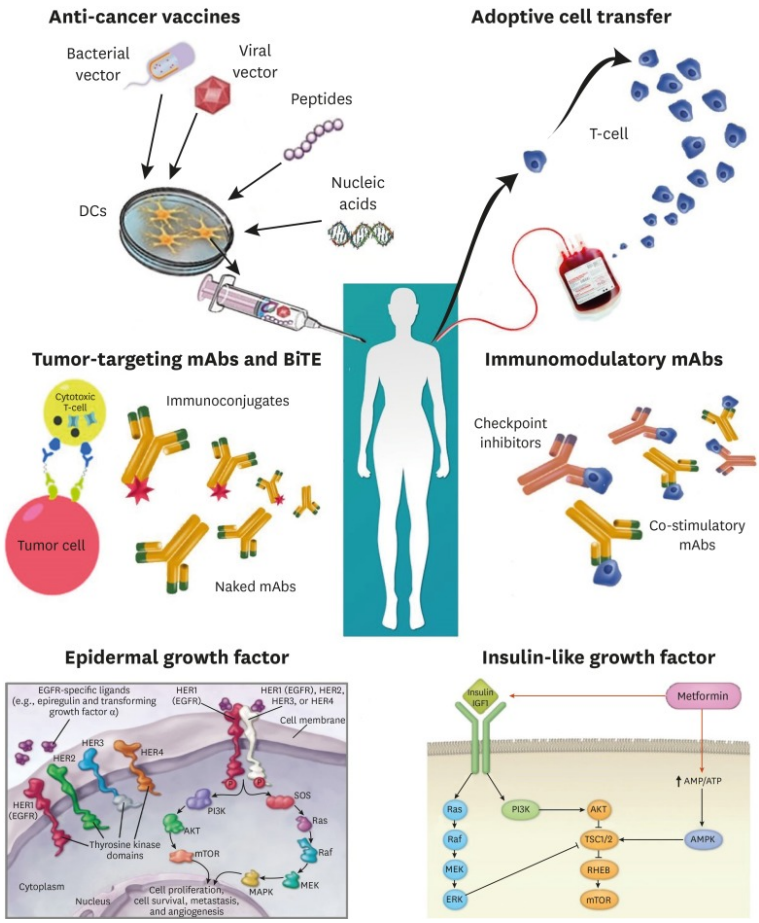


Figure 1.8: *Active and passive immunotherapy methods for cancer*.

BiTE, Bispecific T cell engager; DC, dendritic cell; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; HER, human epidermal growth factor receptor; IGF, insulin-like growth factor; mAb, monoclonal antibody; mTOR, mammalian target of rapamycin; PI3K, phosphatidylinositol-3 kinase. Figure taken from Di Tucci *et al*. (2019).

Based on the evidence supporting the molecular classification of endometrial cancer and support for newer therapeutic agents, four new trials are designed to refine adjuvant therapy based on each of the four molecular subgroups as part of the RAINBO platform (**Figure 1.9**) (TransPORTEC Consortium, 2020). p53 abnormal endometrial cancers will be allocated to the RED trial investigating adjuvant chemoradiotherapy proceeded with either niraparib (PARP inhibitor) or placebo treatment. MMR deficient cancers will be enrolled to receive radiotherapy with or without dostarlimab, a PD1 inhibitor while NSMP tumours will receive adjuvant radiotherapy with or without hormonal therapy. Lastly,

participants with *POLE* mutant tumours will be observed without any adjuvant therapy with the aim of investigating the benefit of descaling of adjuvant therapy.

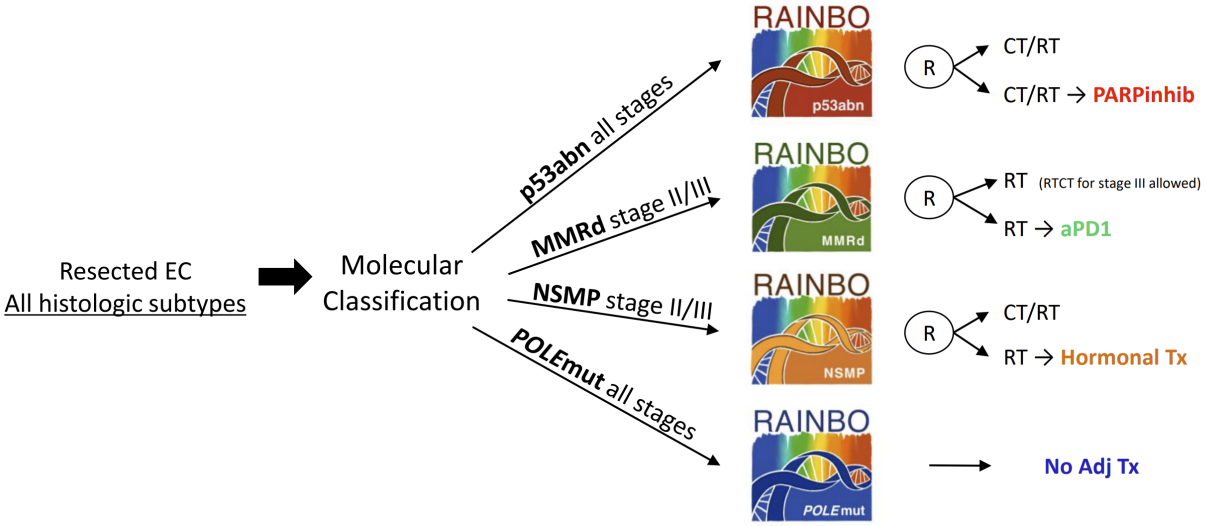


Figure 1.9: *Protocol design of the RAINBO umbrella program coordinated by the TransPORTEC consortium.*

aPD1, anti-PD1 agent, CT, chemotherapy; EC, endometrial cancer; PARPinhib, PARP inhibitor; RT, radiotherapy; Tx, treatment. Figure adapted from TransPORTEC Consortium (2020).

1.5 Risk Factors

Several lifestyle, genetic and reproductive factors have been attributed to modify the risk of endometrial cancer, as summarised in **Table 1.3**. Well-established non-genetic risk factors include age, obesity, insulin resistance and reproductive factors such as nulliparity, exposure to unopposed oestrogens and co-morbidity of polycystic ovary syndrome (PCOS) (Raglan *et al.*, 2019). These will be discussed in more detail in the following sections.

1.5.1 Age

Age is a strong risk factor for several cancers including endometrial cancer. The majority of the endometrial cancer cases are diagnosed in postmenopausal women (Constantine *et al.*, 2019). According to CRUK, the age group receiving most of the average yearly diagnoses is 65-69 years, while the highest incidence is seen in those aged 75 to 79 years old (**Figure**

Table 1.3: *Summary of risk factors associated with the risk of developing endometrial cancer.*

Risk Factor	Effect	Summary	Reference
Age	↑	Risk of developing endometrial cancer increases with age and according to CRUK, reaches peak at ∼65 years.	CRUK (2020)
Obesity	↑	Being overweight or obese increases risk of endometrial cancer. Excess adiposity leads to excess oestrogen bioavailability, insulin resistance and inflammation.	Renehan *et al*. (2008); Kyrgiou *et al*. (2017)
Diabetes	↑	Insulin and IGF1 promote proliferation and inhibit apoptosis. Hyperinsulinemia (type II diabetes) reduces SHGB and adiponectin levels, and increases bioavailability of oestrogen.	Friberg *et al*. (2007); Nead *et al*. (2015)
Adiponectin	↓	Anti-inflammatory, tumour suppressive (activation of p53) and promotes insulin sensitisation.	Gong *et al*. (2015)
Unopposed oestrogens	↑	Oestrogen stimulates endometrial growth. Examples include use of Tamoxifen, oestrogen-only HRT, nulliparity, long menstrual years (early menarche and late menopause).	Cooke *et al*. (1997); MacKintosh and Crosbie (2013)
Progestogens	↓	Progestogens counteract the prolifeative effect of oestrogen. Examples include COCP, IUD, parity.	Gompel (2020); Derbyshire *et al*. (2020)
PCOS	↑	Nearly 3-fold increase in lifetime risk of EC, related to insulin resistance, obesity and abnormal menstrual cycles.	Dumesic and Lobo (2013)
LS	↑	LS is caused by mutations in MMR genes. Cumulative risk of endometrial cancer with LS is up to 71% by the age of 70.	Ryan *et al*. (2017)
Cowden Syndrome (CS)	↑	CS often presents with *PTEN* mutations and risk of endometrial cancer with CS is up to 28%.	Constantinou and Tischkowitz (2017)
Family History	↑	Risk of endometrial cancer increases by 2 to 3-fold with first-or second- degree relatives with endometrial cancer.	Win *et al*. (2015)

1.10) (CRUK, 2020). This may be partly due to lack of progesterone in post-menopausal years which would normally balance the growth of endometrium by opposing oestrogen. The historic use of oestrogen only hormone replacement therapy (HRT) as well as tamoxifen for breast cancer prevention and treatment increases the risk of endometrial cancer through the same mechanism (explained further below). Another reason that may explain the increased risk of endometrial cancer later in life is the acquired oncogenic mutations over time (Moore *et al*., 2020). Analysing the genomic landscape of uterine lavage fluid, Nair *et al.* reported an age-related association of oncogenic mutations in genes including *PTEN* and phosphatidylinositol-3-kinase catalytic subunit α (*PIK3CA*) (Nair *et al*., 2016).

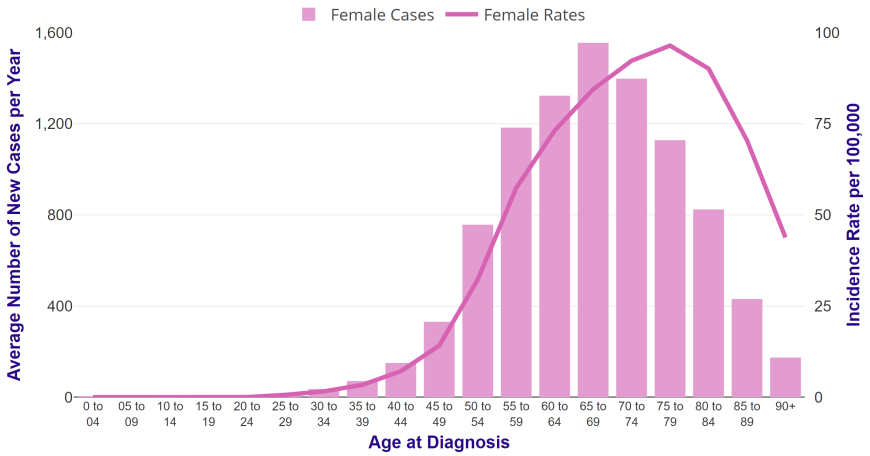


Figure 1.10: *Average number of new cases per annum and incidence rates per age group per 100,000 women in the UK.*

Figure adapted from CRUK (2020).

1.5.2 Obesity and Insulin Resistance

The strongest known risk factor for endometrial cancer is obesity, with an estimated 1.60 times increase in relative risk of endometrial cancer per each 5 kg/m^2 increase in BMI in a linear model (Renehan *et al.*, 2008; Kyrgiou *et al.*, 2017). In fact, obesity might be the prime cause of nearly 40% of endometrial cancer cases (Kaaks *et al.*, 2002). As global burden of obesity increases, endometrial cancer incidence and mortality is also projected to rise. Excess adipose tissue exerts a strong risk to endometrial cancer in terms of inflammatory and endocrine factors. Adipocytes are a source of oestrogen, which has a proliferative effect on the endometrium. In ovulatory premenopausal women, this effect is naturally antagonised by progesterone whereas in postmenopausal women, endogenous progesterone production is no longer active. This allows oestrogen-driven endometrial proliferation to proceed unchecked.

Previous research has shown that obese women have lower blood levels of sex hormone binding globulin (SHBG), and higher levels of insulin and insulin-like growth factors (IGFs) (Hernandez *et al.*, 2015). SHBG inactivates oestrogen and androgen upon binding thus regulating their levels in the body. Insulin and IGFs have a stimulatory effect on

endometrium that increases its proliferation rate. Insulin resistance is a common problem observed among obese women. In fact, according to recently published national statistics in England, 12% of women who are obese also have co-existing diabetes (NHS Digital, 2020b). It is a complicated process, though widely accepted to arise from impairment of glucose uptake due to increased free fatty acid production by excess adipose tissue. This increases the level of circulating insulin, which aside from stimulating endometrial growth, also inhibits production of SHBG, further increasing the bioavailability of oestrogen. However, a causal association between hyperinsulinemia and endometrial cancer, independent of BMI, was also reported using the Mendelian randomisation approach (Nead *et al.*, 2015).

Obesity also causes low-grade systemic inflammation which increases circulating pro-inflammatory cytokines (MacKintosh and Crosbie, 2013). Those secreted by adipose tissue are referred to as adipokines. Obese patients usually present with increased circulating tumour necrosis factor α (TNF-α), pro-inflammatory interleukins (ILs), in particular IL-6, and retinol-binding protein 4 (RBP4) in their blood. As hallmarks of chronic inflammation, these pro-inflammatory cytokines are known to promote angiogenesis, proliferation and DNA damage which collectively can lead to formation and persistence of cancerous tissue. Adiponectin, on the other hand, is an anti-inflammatory adipokine that inhibits proliferative and pro-angiogenic pathways and has a positive effect on insulin sensitivity (Gong *et al.*, 2015). Hence, elevated adiponectin levels are associated with lowered endometrial cancer risk.

Similarly, leptin is an adipokine that acts as a satiety hormone and regulates energy balance (Ahima, 2008). Obese individuals often develop leptin insensitivity, similar to insulin resistance. Reduced sensitivity to leptin results in inability to detect satiety which consequently causes excess storage of energy resources and weight gain. Interestingly, blood level of leptin is raised in obese individuals and those with endometrial cancer (Ma *et al.*, 2013). Therefore, excess oestrogen production, the establishment of a chronic inflammatory state, insulin resistance and weight gain all contribute to carcinogenesis in endometrial

tissue.

1.5.3 Unopposed Oestrogen Exposure

Obesity-related and other forms of excess and/or unopposed oestrogen exposure, are considered key risk factors for endometrial cancer (Kaaks *et al*., 2002; Kitson *et al*., 2017). Natural and exogenous oestrogens promote carcinogenesis by inducing DNA damage and increasing the risk of endometrial and breast cancer. Prolonged premenopausal exposure due to early menarche or late menopause i.e., increased numbers of menstrual cycles, increases the risk of developing gynaecological cancers. Similarly, in the case of exogenous oestrogen-only hormonal therapies, pro-carcinogenic effects of oestrogens would not be opposed by progesterone (similar to postmenopausal years) and hence, promoting excess growth of endometrium leading to endometrial thickness and hyperplasia (Yang *et al*., 2015). As mentioned before, AEH is often considered a precursor of Type I endometrial cancers.

Due to its protective effect on endometrium, progesterone-based treatments have attracted interest for prevention and reversal of carcinogenesis of the endometrium. Progestin-only contraceptives and progesterone-based intrauterine systems (IUSs) reverse hyperplasia and prevent endometrial cancer (Gompel, 2020). The PROTEC trial investigated the feasibility of using levonorgestrel-releasing IUS in class III obese women for the primary prevention of endometrial cancer (Derbyshire *et al*., 2020). The authors reported that this approach was acceptable to women with class III obesity for risk-reduction purposes, who may otherwise be unfit for other forms of treatment should the need arise.

1.5.3.1 Hormone Replacement Therapy

HRT is regularly prescribed to postmenopausal women experiencing severe symptoms or early menopause in order to minimize loss of bone density. However, there is a link between the use of oestrogen-alone HRT which increases not only breast cancer risk but

also relative endometrial cancer risk (Edey *et al.*, 2018). The risk of endometrial cancer is shown to be increased by continuous oestrogen-only HRT use but also to persist for at least five years after discontinuing its use (Grady *et al.*, 1995). However, most of these studies are conducted in postmenopausal women and it remains unclear whether the risk imposed by oestrogen alone hormone therapy mirrors this risk in premenopausal women. Current clinical guidelines are very clear that oestrogen-only HRT should not be used in women with an intact uterus due to the risks of endometrial cancer.

1.5.3.2 Tamoxifen Use

Tamoxifen is a selective oestrogen receptor (ER) modulator, often used for the treatment of ER positive breast cancer (Di Cristofano and Ellenson, 2007) and breast cancer prevention in high-risk women (NICE, 2019). It has long been suspected to increase the risk of endometrial cancer and research has shown that endometrial tissue selectively uptakes tamoxifen which in turn has a stimulatory effect on the endometrium. A large case-control study by Chen *et al.* compared the incidence of endometrial cancer in breast cancer patients that received tamoxifen treatment or not (Chen *et al.*, 2014a). According to their results, the incidence of endometrial cancer in women receiving tamoxifen increased by nearly 3-fold when tamoxifen use exceeded 3 years in duration. Furthermore, this risk was found to be more profound in women older than 35 years of age.

1.5.3.3 Polycystic Ovary Syndrome

PCOS is a common endocrine disorder that is characterised by elevated circulating androgens in women. PCOS causes abnormal periods, poor fertility and excess body and facial hair. Type II diabetes, obesity and endometrial cancer are all associated with PCOS. Previous studies indicated that PCOS increases the risk of endometrial cancer by nearly 3-fold, that accounts for up to 9% lifetime risk of endometrial cancer, compared to 3% in controls (Dumesic and Lobo, 2013). Prolonged exposure to unopposed oestrogen due to abnormal cycle lengths in women with PCOS increases the risk of hyperplasia, a precursor to Type I endometrial cancers (Tzur *et al.*, 2017). Moreover, insulin resistance, which is frequently

accompanied by obesity, often co-exists in women with PCOS and as explained earlier, affects the sex-steroid pathways leading to carcinogenesis in the endometrial tissue.

1.5.4 Race and Ethnicity

Evidence from a number of studies shows a variable distribution of endometrial cancer incidence across race and ethnicity (Jamison *et al.*, 2013; Edwards *et al.*, 2014; Cote *et al.*, 2015). Accordingly, after adjustment for hysterectomy prevalence, the incidence of endometrial cancer in non-Hispanic Black women is higher compared to non-Hispanic White women (Jamison *et al.*, 2013). Moreover, diagnoses of high-grade and more aggressive subtypes are more often received by non-Hispanic Black women than any other racial or ethnic group (Long *et al.*, 2013). Consequently, poorer survival rates are observed among this group (Cote *et al.*, 2015). This disparity between racial groups could be attributed to genetics and lifestyle factors, importantly considering the differing obesity rates between the developed and developing world. Interestingly, although lowest endometrial cancer incidence rates are seen among Asian women, the rates among American-born Asians in comparison to their Asian-born counterparts are higher, pointing out to a potential role of environmental or lifestyle factors in this disparity (Setiawan, 2016).

1.5.5 Family History

To date, a number of studies reported that individuals with first- or second-degree relative(s) diagnosed with endometrial cancer and, to a lesser extent, colorectal cancer are at higher risk of developing endometrial cancer. A meta-analysis by Win *et al.* estimated a pooled relative risk (95% confidence interval (95% CI)) of 1.82 (1.65-1.98) and 1.17 (1.03-1.31) associated with having first-degree relatives with endometrial and colorectal cancer, respectively (Win *et al.*, 2015). The cumulative risk of endometrial cancer up to 70 years of age was estimated to be just over 3% with a first-degree relative with a history of endometrial cancer. Family history in relation to inherited genetic conditions is discussed in more detail in the next section.

1.5.6 Genetic Mutations

A recent study by Moore and colleagues investigated the genomic landscape of normal and cancerous endometrial tissue (Moore *et al*., 2020). The authors showed that normal endometrial glandular cells harbour far less driver mutations than their cancerous counterparts (**Figure 1.11**), and that the mutational burden is proportional to increasing age and is inversely associated with parity. In keeping with previous research, *TP53* alterations were seen frequently in both endometrioid and non-endometrioid subtypes of endometrial cancer, though most commonly in high-grade serous carcinomas (90%). In endometrial cancer, *TP53* mutations are observed primarily in exons 5-8 (Schultheis *et al*., 2016).

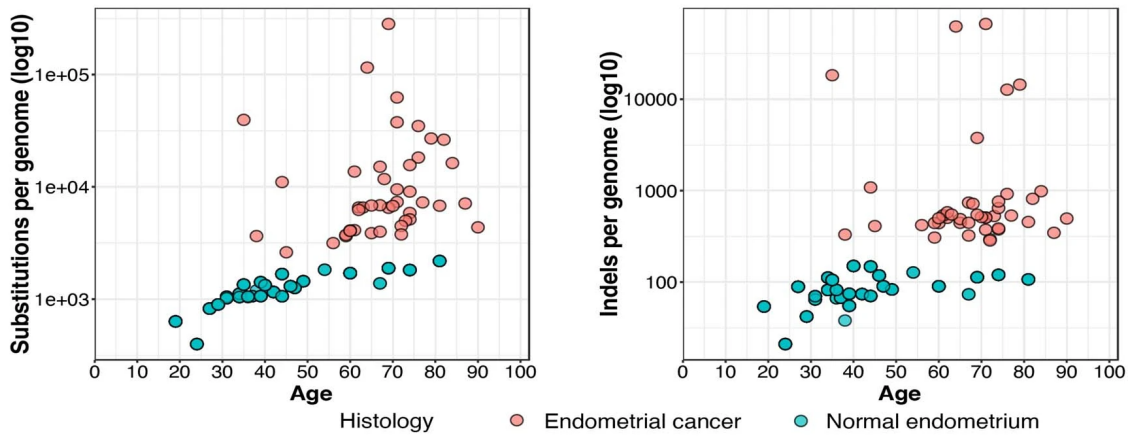
Furthermore, according to Moore *et al.* endometrioid tumours were seen to exhibit mutations in kirsten rat sarcoma viral oncogene homolog (*KRAS*), in line with previous reports of somatic *KRAS* mutations in up to 30% in this subgroup (Sideris *et al*., 2019; Moore *et al*., 2020). *KRAS*, a proto-oncogene, is involved in responding to extracellular signals that govern proliferation and differentiation such as growth factors. Somatic mutations in *KRAS* are found in approximately 10-30% of Type I endometrial cancers, as well as its precursor EH (Dobrzycka *et al*., 2009; Ring *et al*., 2017).

Mutations in the PI3K/AKT/mTOR pathway are also common in endometrial cancer, so that *PIK3CA* mutations may be found in approximately 28% of all endometrial cancer cases according to previous research, which was also replicated in the Moore article (Millis *et al*., 2016; Wilczynski *et al*., 2016; Moore *et al*., 2020). *PIK3CA* is an oncogene that is a part of the phosphatidylinositol-3-kinase (PI3K) signalling pathway. The PI3K pathway regulates proliferation and cell survival, thus, its deregulation is an important factor in tumour initiation and progression. Mutations in exons 9 and 20 have been notably implicated in endometrial carcinogenesis (Millis *et al*., 2016).

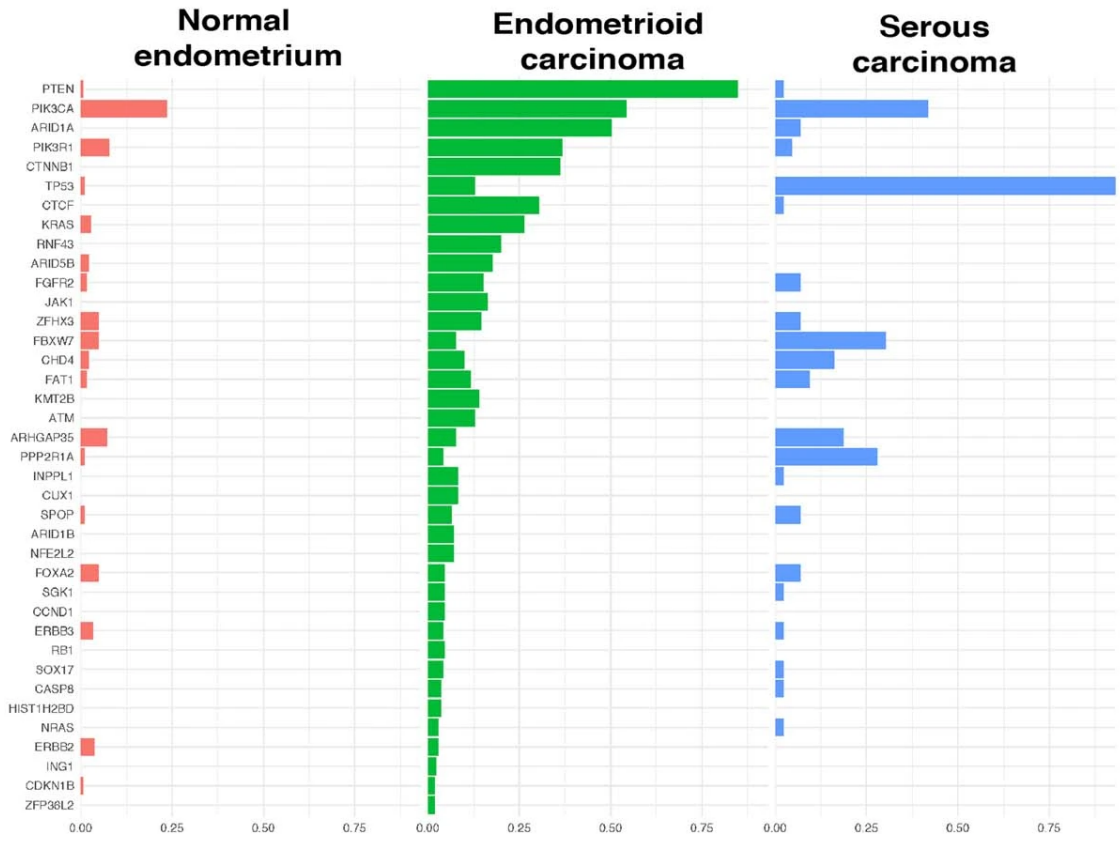
Lastly, *PTEN* is one of the most frequently mutated genes in cancer including endometrial cancer and these are more frequent in endometrioid tumours in comparison to serous or

serous-like tumours (Getz *et al*., 2013; Zhao *et al*., 2016; Moore *et al*., 2020). *PTEN* regulates cell cycle and is considered a tumour suppressor gene through its negative regulation of protein kinase B (AKT/PKB) signalling pathway. Frequent *PTEN* mutations are also found in CS, which is associated with increased risk of developing endometrial cancer (see **Section 1.5.6.2**).

Pathogenic variants in MMR genes and rare germline loss-of-function pathogenic variants in *PTEN* are known hallmarks of Lynch Syndrome (LS) and CS, respectively (Ryan *et al*., 2017; Burke and Gold, 2015). These are discussed in more detail below. It is clear that endometrial cancer can arise as a direct result of pathogenic mutations and that the mutational landscape of endometrial tumours is distinct from the normal endometrium, which is determined by age as well as parity. The latter is particularly important because endometrial cancer is age-dependent and is less common in women with children. Nevertheless, most endometrial cancers are thought to be sporadic in nature, where only 5-10% of the cases are indicated to be inherited.



(a)



(b)

Figure 1.11: *Genomic comparison between normal and cancerous endometria.*

(a) Normal endometrium has less substitutions and indels per genome in comparison to endometrial cancer, indicating less mutation burden. In both normal and, particularly, in cancer, both of these increase by age. (b) Driver mutations found in normal endometrial glands, endometrioid and serous ECs. Figure adapted from Moore *et al*. (2020).

1.5.6.1 Lynch Syndrome

LS, previously known as hereditary non-polyposis colorectal cancer (HNPCC), is a highly heritable condition caused by a defect affecting one of the four MMR genes (Burke and Gold, 2015) (**Table 1.4**). LS increases the risk of many types of cancer such as colorectal, endometrial and ovarian cancer. The lifetime risk of developing endometrial cancer in relation to mutations in MMR genes in LS have been reported to be up to 71% (Barrow *et al.*, 2009; Ryan *et al.*, 2017). Although LS is thought to affect up to one in 300 individuals, prevalence of LS-associated endometrial cancer accounts for up to 5% of all cases in the general population (Rossi *et al.*, 2017; Rosenblum *et al.*, 2020). Recently, a study by Ryan and colleagues found that just over 3% of women had LS in an unselected endometrial cancer population (Ryan *et al.*, 2020). Restricting the criteria by age <50 years for LS testing would have missed over half of the cases whereas testing only those with indicative family history would miss over 60%.

Table 1.4: *Mutation rates in MMR genes in overall LS and in EC.*

Table adapted from Tafe (2015).

Gene	Overall mutation rate in LS (%)	Rate in endometrial cancer (%)	Unique features	Typical IHC pattern with gene mutation or silencing	Comments
MLH1	50	24-40	5%-10% mutations are large deletions; promoter methylation	*MLH1* -/+, *PMS2* -, *MSH2* +, *MSH6* +	*MLH1* and *PMS2* form heterodimer
MSH2	40	50-66	17%-50% mutations are large deletions; promoter methylation	*MLH1* +, *PMS2* +, *MSH2* -, *MSH6* -	*MSH2* and *MSH6* form heterodimer
MSH6	7-10	10-13		*MLH1* +, *PMS2* +, *MSH2* +, *MSH6* -	Deletions are rare
PMS2	<5	<5	Highly homologous pseudogenes for exons 1-5, 9, and 11-15	*MLH1* +/-, *PMS2* -, *MSH2* +, *MSH6* +	Deletions are rare

The MMR genes function to recognise and repair erroneous DNA bases which may arise during replication, recombination, or DNA repair itself (**Figure 1.12**). Therefore, pathogenic mutations affecting the function of any of the MMR genes can lead to carcino-

genesis due to genomic instability. This can be measured by examining evidence for MSI. Furthermore, epigenetic silencing of the promoters for example through methylation can also impair the function of these genes. The most common mutations of MMR genes seen in endometrial cancer are in MutS homolog 2 (*MSH2*) and *MLH1* (Tafe, 2015) (**Table 1.4**).

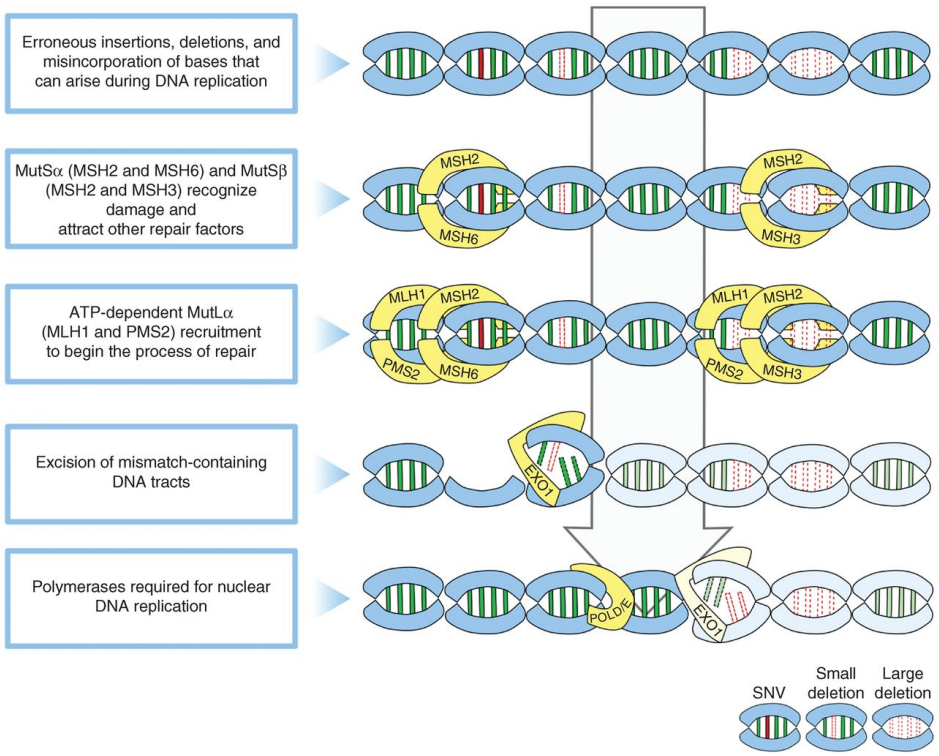


Figure 1.12: *Illustration of the DNA MMR mechanism.*

Figure taken from (Germano *et al.*, 2018).

1.5.6.2 Cowden Syndrome

CS is a rare autosomal dominant disorder with frequent germline loss-of-function mutations in *PTEN* and is associated with endometrial, breast and thyroid cancer risks (Burke and Gold, 2015). Estimates of CS risk in the population is likely to be inaccurate due to its rarity, however, it has been reported to be one in 200,000-250,000 prevalence (Nelen *et al.*, 1999; Pilarski, 2009). Although the risk of endometrial cancer with CS is a recently identified phenomenon, lifetime risk of endometrial cancer associated with CS is reported

to be up to 28% (Constantinou and Tischkowitz, 2017). However, part of this association may be explained by tamoxifen use for CS-associated breast cancer, as explained earlier. Mahdi *et al*. reported that CS-associated germline *PTEN* mutations in endometrial cancer patients can be predicted by 50 years or younger age, presence of macrocephaly, and/or prevalent or synchronous renal cell carcinoma (Mahdi *et al*., 2015). Finally, endometrioid tumours are the most prevalent in CS-related endometrial cancers, similar to the prevalent somatic *PTEN* mutations seen in this group.

1.5.6.3 Pathogenic Mutations in *BRCA* Genes

Breast cancer type 1/2 (*BRCA1/2*) genes play a critical role in DNA double-strand break repair. They were first identified in relation to breast cancer risk, and women who carry pathogenic germline mutations in the *BRCA1/2* genes are at severely heightened risk of hereditary breast and ovarian cancer. As such, although these account for only up to 10% of all breast cancer cases in women, having a pathogenic mutation in either of the *BRCA1/2* genes increases the lifetime risk of breast cancer by around 5-fold. For breast cancer type 1 (*BRCA1*) pathogenic mutation carriers, cumulative risk of developing breast and ovarian cancer by 80 years old is 72% and 44%, respectively (Kuchenbaecker *et al*., 2017a). The corresponding values for breast cancer type 2 (*BRCA2*) carriers are 69% and 17% for breast and ovarian cancer, respectively. In contrast, in the general population, one in eight (12%) of women are estimated to develop breast cancer whereas one in 78 (1.3%) will develop ovarian cancer (Pearce *et al*., 2015; Siegel *et al*., 2020).

Tamoxifen is traditionally used as an adjuvant therapy for breast cancer as well as a chemopreventative agent, particularly for the prevention of ER-positive breast cancer (Early Breast Cancer Trialists' Collaborative, 2011; Cuzick *et al*., 2013; Rugo *et al*., 2016). However, the use of tamoxifen is thought to increase the risk of developing endometrial cancer, though whether this significantly outweighs its benefits against breast cancer is debatable (Sestak and Cuzick, 2016). Prophylactic surgery such as mastectomy and salpingo-oophorectomy in women at high risk of developing breast or ovarian cancer,

such as due to pathogenic *BRCA1/2* carrier status, leads to substantial reduction in risk (Kauff *et al*., 2008; Heemskerk-Gerritsen *et al*., 2013; Boerner and Long Roche, 2020).

Some studies have suggested that *BRCA1/2* mutation carriers are at greater risk of developing endometrial cancer than the general population. Shu and colleagues claimed that women carrying pathogenic mutations in *BRCA1* and undergoing risk-reducing salpingo-oophorectomy (RRSO) without concurrent hysterectomy had a higher risk of developing serous or serous-like endometrial cancer (Shu *et al*., 2016). Further studies were published reporting an increased risk of endometrial cancer, in relation to *BRCA1* mutations in particular (Thompson and Easton, 2002; Segev *et al*., 2013). These observations may have been confounded by the use of tamoxifen for prior breast cancer which is a known risk factor for endometrial cancer (Beiner *et al*., 2007; Segev *et al*., 2013; Lee *et al*., 2017). Others have come forward with contradictory results (Breast Cancer Linkage Consortium, 1999; Levine *et al*., 2001) and international guidelines do not recommend prophylactic hysterectomy for women carrying pathogenic *BRCA1/2* mutations (Mary *et al*., 2020). Thus, the role of *BRCA1/2* in endometrial cancer predisposition remains inconclusive and warrants further investigation (Dullens *et al*., 2020).

1.6 Genetic Predisposition

With advancing DNA sequencing tools and better understanding of the role of inherent genetics in predisposition to disease, many genes have been linked to the onset and progression of various conditions. However, most diseases like cancer exhibit a complex genetic basis coupled with environmental or lifestyle factors that contribute to the risk. A recent review by Bianco *et al.* exploring the genetic basis of endometrial cancer highlighted its complexity and heterogeneity, and summarised a list of the most commonly studied genes in this context (Bianco *et al*., 2020).

Although genetic mutations in MMR genes show high penetrance in endometrial cancer, these are rare in the general population. Up to 95% of all endometrial cancer cases are

thought to be sporadic in nature which indicates low penetrance and likely polygenic input. Population studies have shown an effect of familial risk beyond presumed LS (Bermejo *et al*., 2004). In order to elucidate variant genetic alleles that define or contribute to endometrial cancer susceptibility, a number of candidate-gene studies and genome-wide association studies (GWAS) have been conducted particularly within the last decade.

1.6.1 Single Nucleotide Polymorphisms

There are an estimated 4 million SNPs per person distributed throughout the genome, at a rate of one every 300 nucleotides. While some SNPs exhibit a minor allele frequency (MAF) of 5% or higher, the vast majority have a MAF of 1%, which necessitates studying a substantially large sample size to identify genuine signals in association to a certain disease.

Most SNPs lie within the intronic regions of the DNA presumably having little or no effect on health. However, some SNPs have been strongly linked with response to a treatment and susceptibility to certain diseases or environmental factors. For instance, large-scale GWAS have identified many SNPs that are strongly linked to breast cancer (Evans *et al*., 2017).

1.6.2 Candidate Gene Association Studies

Candidate gene studies focus on pre-selection of putative candidate genes based on their relevance in the trait-of-interest being investigated (Patnala *et al*., 2013). Prior to the rise of GWAS, this method was the predominant choice to explore low-risk variants in a given disease due to being relatively fast and cheap. However, due to the nature of this approach, only a small number of variants are investigated. Hence, SNPs with an anticipated functional effect on their respective genes-of-interest are prioritised. This method can later be expanded to include tagging SNPs that exhibit high correlation i.e., in linkage disequilibrium (LD), that would cover the entire gene.

1.6.3 Genome-Wide Association Studies

High throughput genotyping methods have revolutionised the field of genetic epidemiology since the first large GWAS by Wellcome Trust Case Control Consortium (2007). A GWAS evaluates the association of certain traits with millions of SNPs without prior knowledge and selection of traits or SNPs, unlike candidate gene association studies.

In a typical GWAS subjects are genotyped for several hundred thousand to a million variants, commonly by using a commercial or custom-developed chip. The allele and genotype frequencies of these variants are investigated between the subject groups such as disease versus non-disease. In order to minimise the effects of the high false-positive rate seen in GWAS, population structure and relatedness are accounted for, statistical corrections for multiple testing are employed and a widely accepted genome-wide significance level of $P < 5E\text{-}8$ is used (Panagiotou *et al.*, 2012; Kaler and Purcell, 2019). However, large sample sizes are therefore required to detect associations with small P values.

1.7 Endometrial Cancer Susceptibility Variants

Many studies have been conducted, particularly using the candidate gene approach, to identify SNPs conferring susceptibility to endometrial cancer. Much of the research using the candidate gene approach focused on genes involved in oestrogen-related pathways, simply due to the hormone-related nature of endometrial cancer. One such study by Epidemiology of Endometrial Cancer Consortium (E2C2) reported two SNPs in the aromatase gene, cytochrome P450 19A1 (*CYP19A1*), to be associated with an increased risk of endometrial cancer (Setiawan *et al.*, 2009). This is the only instance, in the context of endometrial cancer, where an association of a locus identified through candidate-gene approach was then confirmed in a large-scale GWAS.

In contrast, fewer GWAS exist for endometrial cancer, most of which utilised overlapping sample sets, comprising mostly of endometrioid subtype and broad-European populations. Many of these cohorts have been pooled into large consortia such as the E2C2,

Endometrial Cancer Association Consortium (ECAC) and Population Architecture Using Genomics and Epidemiology (PAGE) (Olson *et al.*, 2009; Setiawan *et al.*, 2014). The latest and largest GWAS to look at endometrial cancer susceptibility identified nine new susceptibility loci and validated seven others reported previously (O'Mara *et al.*, 2018).

1.7.1 *HNF1B* Locus

Hepatocyte nuclear factor 1 β (*HNF1B*) encodes for a transcription factor (TF) and the resulting HNF1B protein is expressed in many organs and tissues including those of the reproductive system. There are three isoforms of HNF1B, two of which activate transcription while the third is a suppressor (Bach and Yaniv, 1993). *HNF1B* has been previously linked to multiple cancers including prostate, colorectal and ovarian cancer. A large-scale, multicentre GWAS conducted in 2011 identified, for the first time, three SNPs in *HNF1B* locus that are associated with decreased endometrial cancer risk in women of European descent (Spurdle *et al.*, 2011). G allele of rs4430796 showed the highest association, reaching significance in stage one (odds ratio (OR)=0.79, 95% CI 0.73-0.87, P = 3.06E-7) comprising 1,265 cases and 5,190 controls. In the second stage, this SNP was genotyped in an additional 3,957 cases and 6,886 controls and showed a weaker association (OR=0.87, 95% CI 0.81-0.94, P = 2E-4). The combined results of both stages reached P = 7.11E-10 with an OR (95% CI) of 0.84 (0.79-0.89), which was slightly stronger in the endometrioid-only subset (OR=0.82, 95% CI 0.77-0.87, P = 4.28E-11).

The other two *HNF1B* SNPs identified in the study were rs4239217 and rs7501939. The combined results of both stages were of similar magnitude to that of the lead SNP rs4430796 with ORs (95% CIs) 0.84 (0.80-0.90) and 0.86 (0.82-0.91) per G allele of rs4239217 and A allele of rs7501939, respectively. The combined P values for both SNPs reached P < 1E-7. For all three SNPs, the associations were restricted to endometrioid cases only, however, this is likely a direct result of the insufficient number of non-endometrioid cases in the study.

The same study reported that a surrogate SNP for rs4430796 (rs11651755), which was genotyped in 832 cases and 2,049 controls of Chinese ancestry, did not reach significance (P = 0.55). This could be attributed to different effect allele frequency (EAF) levels seen in different ethnicities such that G allele has a frequency of ∼0.48 in broad-European populations whereas in pan-Asian populations it is nearly halved to ∼0.28 (Sherry *et al*., 2001).

Another study reported an association between rs4430796 and rs7501939 and endometrial cancer in an ethnically mixed, but primarily European, cohort (Setiawan *et al*., 2012). Per allele ORs were reported as 0.82 (95% CI 0.75-0.89, P = 5.63E-6) for G allele of rs4430796 and 0.79 (95% CI 0.73-0.87, P = 3.77E-7) for A allele of rs7501939. Further fine-mapping of the *HNF1B* locus in a large multi-ethnic cohort revealed other SNPs, namely intronic rs11263763, that were significantly associated with altered *HNF1B* expression (Painter *et al*., 2015). A recent GWAS meta-analysis looking at cross-cancer shared risk regions between endometrial and ovarian cancers by Glubb *et al*. reported that rs11263763 is also significantly associated with clear cell ovarian cancer and validated the association of this SNP in overall and endometrioid subtype endometrial cancer (Glubb *et al*., 2020).

1.7.2 Hormonal Pathway Polymorphisms

1.7.2.1 *CYP19A1* Locus

The sex hormones (oestrogens, progestogens and androgens) are produced through the steroidogenesis pathway catalysed by a number of enzymes. A key steroidogenic enzyme, aromatase, encoded by *CYP19A1*, has received considerable attention for its potential role in endometrial carcinogenesis. Aromatase converts androstenedione and testosterone into oestradiol and oestrone, respectively.

A number of SNPs in the *CYP19A1* gene have been reported to date. In a Chinese candidate gene study, A/C and C/C genotypes of rs1870050, located in the promoter region, were reported to be associated with decreased risk of endometrial cancer with ORs of 0.81 (95% CI 0.68-0.97) and 0.58 (95% CI 0.42-0.80), respectively (Tao *et al*., 2007). Moreover,

the same study found that rs1065779 was associated with decreased risk of endometrial cancer in post-menopausal women.

Another study which investigated polymorphisms within the sex steroid pathway reported that rs4775936 of *CYP19A1* significantly increased endometrial cancer risk (OR=1.22, 95% CI 1.01-1.47, P = 0.04) per A allele (Lundin *et al*., 2012). When limited to the endometrioid subtype alone, this association was slightly lower. A meta-analysis on 4,998 cases and 8,285 controls from the E2C2 found that A alleles of both rs749292 and rs727479 of *CYP19A1* were significantly associated with increased risk of endometrial cancer; OR 1.1, 95% CI 1.09-1.21 (P = 7.1E-7) and OR 1.08, 95% CI 1.02-1.14 (P = 0.009), respectively (Setiawan *et al*., 2009). Interestingly, women aged 55 years or over and with higher BMI had a greater risk of endometrial cancer associated with these variants.

1.7.2.2 Oestrogen Receptor SNPs

Exposure to excess oestrogen is a well-established risk factor for endometrial cancer, particularly for the endometrioid subtype. Oestrogen functions through its receptors ER α and β, encoded by oestrogen receptor 1 (*ESR1*) and oestrogen receptor 2 (*ESR2*), respectively. Epithelial cells of the endometrium predominantly express ERα and thus, polymorphisms in *ESR1* may play a role in predisposition to endometrial cancer. Much of the previous work on *ESR1* SNPs yielded conflicting results and was limited by small sample sizes (Sasaki *et al*., 2002; Iwamoto *et al*., 2003; Wedren *et al*., 2008; Ashton *et al*., 2009a; Einarsdottir *et al*., 2009; Sliwinski *et al*., 2010; Li *et al*., 2011a).

Fine-mapping of the *ESR1* locus by O'Mara and colleagues, however, extensively genotyped and imputed SNPs of *ESR1* locus using 6,607 cases and 37,925 controls (O'Mara *et al*., 2015). According to their analysis an imputed SNP, rs79575945, showed the strongest association with an OR of 0.85 (95% CI 0.79-0.92) per A allele. While this association was strong in endometrioid subtype, it did not reach significance in the non-endometrioid subtype, similar to most other reported associations, possibly due to small sample sizes.

The team reported no other SNP in the *ESR1* locus that reached significance.

1.7.2.3 Progesterone Receptor SNPs

Progesterone counteracts the proliferative stimulation by oestrogen on the endometrium. The progesterone receptor is encoded by a single progesterone receptor (*PGR*) gene translated into two functional isoforms; progesterone receptor A (PR-A) and progesterone receptor B (PR-B) (Ellmann *et al.*, 2009). Genetic variants in *PGR* may affect biological function or expression of progesterone, hence, polymorphisms in *PGR* have been investigated in relation to endometrial carcinogenesis.

Genotyping 2,888 cases and 4,483 by candidate gene approach revealed a single SNP in the 3' region of *PGR* (rs11224561) that was associated with increased risk of endometrial cancer (OR=1.31, 95% CI 1.12-1.53, $P = 0.001$) (O'Mara *et al.*, 2011). Another SNP, rs608995, was reported to be associated with increased risk of endometrial cancer with an OR (95% CI) of 1.30 (1.06-1.59) per T allele ($P = 0.012$) (Lee *et al.*, 2010). A non-significant increase associated with T/T genotype of rs1042838 (OR=1.42, 95% CI 0.65-3.09, $P = 0.09$) was also noted.

A meta-analysis by Chen *et al.* pooled six studies totalling to 6,285 cases and 12,120 controls, and reported results for two SNPs, rs1042838 and rs10895068, in the *PGR* region in relation to endometrial cancer (Chen *et al.*, 2015). According to their results, rs1042838 was significantly associated with increased risk of endometrial cancer such that carrying T allele versus G allele moderately increased risk (OR 1.23, 95% CI 1.07-1.42, $P = 0.005$) while homozygous carriers of T/T versus G/G genotype had over 1.7-fold increase of risk (OR 1.72, 95% CI 1.12-2.65, $P = 0.013$). Moreover, the authors reported a modest increase in endometrial cancer risk with rs10895068 T allele versus C allele with an OR of 1.15 (95% CI 1.02-1.29, $P = 0.027$).

1.7.3 *MDM2* Locus

A key gene that has been extensively studied by candidate-gene studies in relation to its potential role in endometrial cancer is murine double-minute 2 homolog (*MDM2*), which acts as a negative regulator of the tumour suppressor gene *TP53*. A polymorphism commonly referred to as SNP309 (rs2279744), a T to G change, at the promoter region has been linked to increased *MDM2* expression as a result of increased affinity of its transcriptional activator SP1 (Bond *et al*., 2004). Aberrant *MDM2* expression subsequently leads to suppression of *TP53* activity, hence increasing the risk of tumourigenesis. ER has been also reported to bind to *MDM2* reporter and alter its activity in a gender-specific manner (Bond *et al*., 2006). Thus, it has been postulated that *MDM2* may be responsible for accelerated carcinogenesis in oestrogen-regulated endometrium.

In line with the role of *MDM2* in other cancers including breast cancer, Walsh *et al.* investigated the possible role of *MDM2* SNP309 in endometrial cancer in a small set of cases and controls (Walsh *et al*., 2007). They reported that homozygous G/G carriers had over 2.6-fold increase in endometrial cancer risk. A number of other studies and pooled meta-analyses reported a similar association, primarily in Asian populations, with a varying degree of association (Terry *et al*., 2008; Ueda *et al*., 2009; Li *et al*., 2011b; Peng *et al*., 2013; Zhao *et al*., 2014; Xue *et al*., 2016). However, some of these studies had methodological errors raising concerns about the validity of their results. Wang and colleagues pointed out that the studies by Ueda *et al*. (2009) and Nunobiki *et al*. (2009) had genotype distribution deviating from Hardy-Weinberg equilibrium (HWE) in controls which indicates heterogeneity in population with or without genotyping errors (Wang *et al*., 2014). Others reported contradicting results regarding this association (Ashton *et al*., 2009b; Yoneda *et al*., 2013).

Conversely, a few research teams claimed that a variant in codon 72 of *TP53* (rs1042522) and SNP309 of *MDM2* increase the endometrial cancer risk co-operatively, indicating that SNP309 does not increase the risk of endometrial cancer alone (Nunobiki *et al*., 2009;

Yoneda *et al*., 2013; Zajac *et al*., 2014). rs1042522 of *TP53* has also been studied individually and similar to *MDM2*, contradictory results were reported (Zubor *et al*., 2009; Gu *et al*., 2011; Zajac *et al*., 2012; Kafshdooz *et al*., 2014). Interestingly, according to Zajac *et al.* this SNP may be related to obesity and hypertension (Zajac *et al*., 2013).

Two other SNPs in *MDM2* promoter region have been described in relation to endometrial cancer risk in the literature. rs2870820 (C/T), referred to as SNP55, has been reported to affect expression of *MDM2*, which could potentially lead to carcinogenesis of the endometrium, despite having no direct association with endometrial cancer (Okamoto *et al*., 2015). On the other hand, rs117039649 (SNP285) has been described as a protective variant that antagonises binding of SP1 and decreases the risk of endometrial cancer (Knappskog *et al*., 2012a). This particular SNP appears to be distributed in a higher frequency among European populations (Knappskog *et al*., 2014).

1.7.4 Obesity-Related Polymorphisms

Obesity is a well-established risk factor for endometrial cancer. Excess oestrogen production by the adipose tissue and chronic inflammatory state of obesity are associated with endometrial cancer. Within the last decade, researchers attempted to further characterise the link between obesity and endometrial cancer.

Accordingly, SNPs related to obesity and BMI have been implicated in endometrial cancer (Lurie *et al*., 2011; Delahanty *et al*., 2011). Examining pooled data from E2C2, A allele of rs9939609, near alpha-ketoglutarate dependent dioxygenase (*FTO*) gene, was found to be associated with endometrial cancer in non-Hispanic Whites (Lurie *et al*., 2011), which was later confirmed in a Chinese population (Delahanty *et al*., 2011). C allele of rs6548238 located at transmembrane protein 18 (*TMEM18*) gene was shown to be strongly associated with BMI (P = 1E-18) and increase the risk of endometrial cancer (OR 1.28, 95% CI 1.04-1.59, P = 0.0215) (Lurie *et al*., 2011). Moreover, an association between C allele of rs17782313 at melanocortin 4 receptor (*MC4R*) gene and both BMI and endometrial cancer was reported

in Chinese women (OR 1.29, 95% CI 1.11-1.50, *P* = 7E-04) (Delahanty *et al*., 2011). However, this association was not observed in non-Hispanic White women (Lurie *et al*., 2011).

SNPs in adipokine genes such as *ADIPOQ*, encoding adiponectin, and *LEP*, encoding leptin, have also been reported. As such, in a study of 1,028 cases and 1,932 controls, A/A carriers of rs3774262 intronic to *ADIPOQ* had lower risk of endometrial cancer (OR 0.68, 95% CI 0.48-0.97, $P_{additive}$=0.025) (Chen *et al*., 2012). Similarly, homozygote carriers of the minor allele (C/C) of rs1063539 at the 3' untranslated region (OR 0.66, 95% CI 0.47-0.93, $P_{additive}$=0.017), and heterozygote carriers of (A/G) of rs12629945 at the 3' flanking region (OR 0.78, 95% CI 0.64-0.94, $P_{additive}$=0.026) had lower risk of endometrial cancer. A similar result was later obtained for rs1063539 by Aminimoghaddam *et al.* while Chen *et al.* found no association for this SNP (Chen *et al*., 2012; Aminimoghaddam *et al*., 2015).

Moreover, a reduced risk of endometrial cancer was also observed for T/T carriers of leptin SNP rs2071015 under recessive model (OR 0.70, 95% CI 0.54-0.90, $P_{recessive}$=0.006) (Chen *et al*., 2012). Another leptin SNP (rs12112075) was also reported to lower the endometrial cancer risk in homozygote A/G carriers (OR 0.51, 95% CI 0.27-0.99, *P* = 0.045) (Bienkiewicz *et al*., 2017).

In summary, higher energy intake, which may be mediated by SNPs in obesity-associated genes, increases the risk of endometrial cancer. Similarly, protective SNPs, such as those in adipokine genes, may reduce excess energy intake leading to a decreased risk of cancer.

1.7.5 SNPs in DNA Repair Pathways

DNA damage occurs in the genome regularly during replication of DNA and as a result of oxidative stress, reactive oxygen species surplus, radiation and pollutants. If left unrepaired, these can accumulate and result in aberrant transcription of key genes such as oncogenes. In fact, to date, various cancers including breast and prostate cancer have been linked to DNA damage. Depending on the type and cause of DNA damage, a number of

crucial DNA repair mechanisms are employed to repair the damage. An example is the MMR genes implicated in LS.

Aside from germline mutations in the MMR genes, a small number of studies also attempted to investigate polymorphisms that may be associated with endometrial cancer. In a small candidate-gene study, Poplawski and colleagues genotyped 100 cases and 100 controls to assess any association between endometrial cancer and rs4987188 in *MSH2* gene (Gly322Asp) and rs1800734 in *MLH1* gene (-94G>A) (Poplawski *et al*., 2015). The authors reported an increased risk of endometrial cancer with the G allele of rs1800734 (OR=2.71, 95% CI 1.81-4.08, $P < 0.001$). Carriers of both G/A genotype of rs1800734 and Gly/Gly genotype of rs4987188 showed the highest risk for endometrial cancer with an OR of 4.52 and 95% CI 2.41-8.49 ($P < 0.001$). It has been postulated that rs1800734 at *MLH1* plays a role in epigenetic silencing of its promoter, and thus, inactivating the gene (Chen *et al*., 2007). However, a recent meta-analysis by Russell *et al.* found no evidence for an association between neither rs1800734 (OR=1.06, 95% CI 0.8-1.33, $P = 0.60$) nor 126 other SNPs in the *MLH1* region and endometrial cancer with MSI in a total of 225 cases and 13,582 controls (Russell *et al*., 2020). The authors also reported that rs1800734 allelic status did not influence *MLH1* methylation and expression.

1.8 Risk Prediction and Stratification

Risk prediction enables tailored design of cancer prevention trials and informs healthcare professionals and at-risk individuals to make decisions about personalised prevention and treatment strategies. This is particularly important for endometrial cancer because the alarming rise in obesity, its largest risk factor, and reduction of hysterectomy rates for benign conditions has caused an increase in endometrial cancer incidence. Principally, the projected increase in incidence and associated mortality warrants immediate action to halt this trend from continuing any further.

Epidemiological and lifestyle factors such as BMI are largely modifiable but genetic predisposition is fixed throughout life. As shown by the GWAS studies, endometrial cancer predisposition is likely to be influenced by the polygenic genetic make-up of individuals. Thus, a polygenic risk score (PRS) may aid risk prediction efforts for endometrial cancer. Prior knowledge of the genetic predisposition for endometrial cancer, i.e., likelihood of developing the disease, will not only enable appropriate screening and prevention strategies to be employed but may also maximise the benefit of targeted treatments.

First, no screening guidelines for endometrial cancer currently exist at a national level. Employing risk prediction tools that use readily or easily accessible information such as BMI or SNPs will undoubtedly benefit both the individuals and the healthcare system by reducing cost and physical burden of endometrial cancer. Second, a personal risk score for an individual can be used to tailor both the preventive and screening efforts such that resources are directed at the individuals at high-risk as well as reducing the use of unnecessary invasive diagnostic and treatment modalities for individuals at low- to moderate-risk. As an example, regular surveillance may be offered to women at moderate-risk, who would have been otherwise offered a hysterectomy. Finally, prevention strategies can be more effectively exerted based on the multifaceted risk score of an individual. The personalised nature of this will allow healthcare professionals to offer the most appropriate guideline or treatment of choice, such as progesterone-based IUS as discussed in **Section 1.5.3, p 46** (Gompel, 2020; Derbyshire *et al*., 2020), weight loss interventions such as bariatric surgery, or changes in nutritional habits and physical activity (MacKintosh and Crosbie, 2013). Hysterectomy and RRSO, may be offered only to those at high-risk, such as women with LS. Evidence for and against prevention strategies has been reviewed in detail by MacKintosh and Crosbie (2018) and summarised in **Table 1.5**.

Aside from individual risk, familial risk needs addressing if a woman is found at high risk based on her risk score. As genetic make-up is shared between the members of a family, relatives of a genetically high-risk woman may be offered further testing to assess their

Table 1.5: *Summary of main risk factors for endometrial cancer and prevention methods.*

Table adapted from MacKintosh and Crosbie (2018).

Risk factor	Effect on endometrial cancer risk	Proposed mechanism	Proven methods of prevention	Potential methods of prevention
LS	Lifetime risk 70%, cf. 2-3% in general population	Mutations in DNA mismatch repair genes	Risk-reducing surgery	Aspirin
Tamoxifen	Postmenopausal RR 4.01 (95% CI 1.7-10.9)	Oestrogenic effects on endometrium	Low threshold to investigate abnormal bleeding	LNG-IUS
PCOS	Lifetime risk 9% OR 2.89	Insulin resistance Anovulatory cycles	Induce regular withdrawal bleeds	Weight reduction Metformin hormonal contraception
Obesity	RR 1.59 per 5 kg/m^2 increase in BMI	Activation of pro-proliferative pathways Anovulatory cycles	Bariatric surgery Physical activity	Non-surgical weight loss LNG-IUS
Diabetes	RR 1.42-4.1	Activation of pro-proliferative pathways	Bariatric surgery	Modulation of insulin resistance

LNG-IUS, levonorgestrel intrauterine system; RR, relative risk.

risks. The new NICE guidelines are an example of this approach wherein all women diagnosed with endometrial cancer are recommended to undergo testing for LS, which is then enhanced by cascade-testing within the family (NICE, 2020). The same can be applicable to non-genetic factors as environmental factors may also be shared by individuals living in the same household, for example a high-calorie diet. This will ensure that the endometrial cancer risk in families which may result either from shared genetic or environmental factors may be established and targeted ahead of time.

1.8.1 Risk Prediction Models

In order to be able to offer preventative and risk-reducing measures for a disease, such as prophylactic surgery, risk prediction models (RPMs) may be developed to predict health outcomes and aid clinical decision making. Based on patient characteristics and risk factors, RPMs calculate the probability of an individual experiencing a health condition in a given context. An RPM should be predictive for each patient in a cohort and ideally validated in a new cohort of patients. The sample size of development and validation cohorts should be sufficient to avoid overfitting of the model that could lead to inaccurate

prediction (Pavlou *et al*., 2015).

Three RPMs exist for identifying women at risk of endometrial cancer (**Table 1.6**). The most comprehensive endometrial cancer RPM to date, proposed by Kitson *et al*. combines obesity, insulin resistance, reproductive and a rudimentary genetic score to predict high-risk women (**Table 1.7**) (Kitson *et al*., 2017). Using a population-based cohort, Pfeiffer *et al*. and Husing *et al*. developed RPMs for predicting endometrial cancer. The former was validated externally, and the latter was later extended to include biomarkers (Pfeiffer *et al*., 2013; Fortner *et al*., 2017). These had moderate discriminatory power (0.62-0.77) (Pfeiffer *et al*., 2013; Husing *et al*., 2016; Fortner *et al*., 2017). Moreover, as demonstrated by Fortner *et al*., addition of serum biomarkers only slightly improves the power of the model (Fortner *et al*., 2017). This signifies the urgent need for a more detailed RPM to achieve better discrimination, such as by taking into account genetic risk factors alongside other factors.

Table 1.6: *Risk factors utilised in the developed, validated and proposed RPMs for endometrial cancer.*

Table adapted from Alblas *et al*. (2018). Additional data was obtained from Kitson *et al*. (2017).

Risk models	Age at				BMI	HRT[1]	OC[2]	Parity	Menopausal status	Smoking	Biomarkers[3]
	current	menopause	menarche	FFTP							
Kitson *et al*. (2017)[4]		✓	✓		✓		✓	✓			✓
Fortner *et al*. (2017)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Husing *et al*. (2016)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Pfeiffer *et al*. (2013)	✓	✓			✓	✓	✓	✓	✓	✓	

[1] Duration of use.
[2] Ever use.
[3] Biomarkers include adiponectin, oestrogens, IL1Ra, TNFα, testosterone, triglycerides.
[4] Additional factors include anovulation, C-peptide, ever use of tamoxifen, family history of EC, type 2 diabetes, PCOS, waist circumference, weight gain. FFTP, first full-term pregnancy.

Table 1.7: *The comprehensive RPM for endometrial cancer proposed by Kitson et al. (2017).*

Risk score	Risk factor	-2	-1	0	1	2	4	8
Obesity	BMI			<25 kg/m^2	25-30 kg/m^2	30-35 kg/m^2	35-40 kg/m^2	≥40 kg/m^2
	Waist circumference			<90 cm	90-100 cm	100-110 cm	>110 cm	
	Weight gain between 18-25 & 45-55 y			<5 kg	5-20 kg	>20 kg		
	Adiponectin	>5μg/mL						
Reproductive	Early menarche (<12 y) or late menopause(>55 y) OR Anovulation (≥6 mo, unrelated to pregnancy, breastfeeding or contraceptive use)			None	One or more			
	Parity	2+	1	0				
	COCP use	≥5 y	Never or <5 y					
	Ever use of tamoxifen			No		Yes		
	Free testosterone			≤17 pmol/L	>17 pmol/L			
Insulin	Type 2 diabetes			Absent		Present		
	PCOS			Absent			Present	
	C-peptide (non-fasting)			≤0.76 nmol/L		>0.76 nmol/L		
Genetic	Family history of EC			No first- or second-degree relatives affected			First-degree relative diagnosed at <50 y	Two or more first- or second-degree relatives diagnosed

Identifying individuals at high risk for LS by; Amsterdam (II), Bethesda (Revised), Society of Gynecologic Oncology (SGO) and Australian National Endometrial Cancer Study (ANECS) criteria employ family history and age (reviewed by Buchanan *et al*. (2014)). However, similar to endometrial cancer RPMs, these models do not incorporate genomic data in their assessment either. Therefore, it is crucial to assess whether adding a panel of genetic variants combined as a PRS into existing models can maximise risk prediction. This would allow development of personalised early detection and prevention tools.

1.8.2 Polygenic Risk Scores

As GWAS started to mount in the literature, so did the number of variants identified for various diseases. However, individually, the effects of the vast majority of these variants on predisposition to a disease are negligible. Thus, to harness the combinatorial effects of SNPs, a PRS can be devised. This is done by combining information from multiple SNPs to yield a number which reflects susceptibility of an individual to the disease of interest. It is important to note that the PRS explains only the relative risk of disease and does not reflect the absolute risk. In essence, the score compares the risk of individuals in one group, such as affected, with those in another group (unaffected). Therefore, it does not indicate the lifetime risk of a disease or causation but rather a correlation.

SNPs included in a PRS are independent of each other, i.e., not showing substantial evidence of LD, and are associated with the risk of the disease of interest as identified and/or validated by large GWAS. The number of risk alleles (0, 1 or 2) carried by the individual for a given SNP is multiplied by its effect size (OR or beta). All these are summed to create the PRS. The means of the PRS in cases and controls are expected to be different for the PRS to be informative of a risk. An example of a PRS construction and personalised risk assessment is illustrated in **Figure 1.13**.

The vast majority of the SNPs do not influence risk, and most of those reported to date are likely not the causal variant but may be located close to it, such as in a regulatory region.

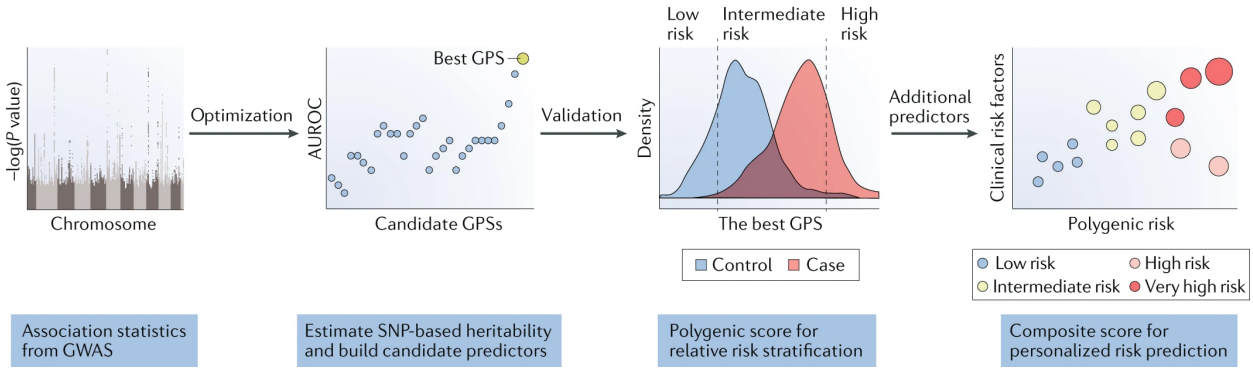


Figure 1.13: *Depiction of how a PRS and personalised risk assessment may be constructed.*

AUROC, area under receiver-operator curve; GPS, genome-wide polygenic risk score. Figure taken from Liu and Kiryluk (2018).

Nevertheless, because the risk associated with the risk alleles is often quite low, combining the effects into a PRS to indicate the overall risk conferred is a rational approach. This is particularly important to highlight, as most of the hereditary component of complex conditions such as cancer does not have an underlying monogenic cause but rather a polygenic or a multi-allelic association. It has been estimated that a PRS may contribute to risk as much as a monogenic pathogenic variant (Khera *et al.*, 2018).

A good example of the potential for PRS can be understood by looking at those devised for breast cancer. Despite accounting only for up to 5% of all breast cancer cases, SNPs have been shown to be effective in risk stratification for breast cancer. A 77-SNP PRS derived for breast cancer achieved mean PRS 0.69 for cases and 0.49 for controls, in 33,673 case and 33,381 control women of European descent (Mavaddat *et al.*, 2015). This was later validated in a larger dataset comprising 94,075 cases and 75,017 controls of European origin, and in an independent prospective set of 11,428 cases and 18,323 controls (Mavaddat *et al.*, 2019). The 77-SNP PRS achieved area under the receiver-operator curve (AUC) of 0.61 in the validation set and 0.60 in the prospective set. Another PRS including SNPs below genome-wide significance, totalling to 313, was also developed by the authors, that was shown to have a higher predictive value. The 313-SNP PRS achieved AUC of 0.64 in the validation set and 0.63 in the prospective set. This is important because these results indi-

cate the potential benefit of including suggestive SNPs in a PRS to improve predictivity.

PRSs also may be useful to estimate the penetrance or presentation of symptoms in individuals carrying high-risk genes. In breast and ovarian cancer, the absolute risk was shown to be predicted in *BRCA1* and *BRCA2* carriers, by the use of PRSs (Kuchenbaecker *et al*., 2017b). Moreover, SNPs can be useful for improving the predictive value of existing risk prediction models (RPMs), such as in the case of 18-SNP PRS used alongside mammographic density and other risk factors (van Veen *et al*., 2018; Evans *et al*., 2019). This indicates that even if not sufficiently predictive on its own, as may be expected from a single risk factor, a PRS can still be very useful when combined with other risk prediction tools.

For endometrial cancer, more precise and personalised risk prediction is urgently needed not only because of the lack of good-performing models, but also because of the risk posed by the growing burden of endometrial cancer in parallel to the obesity epidemic. Although PRSs may be moderately predictive, when incorporated into the RPMs, the improved models may provide more accurate risk prediction that allows targeted screening and prevention strategies to be designed. Furthermore, most of the risk factors used within the existing models are likely to be more applicable to the endometrioid tumours, such as obesity, for which early detection coupled with hysterectomy is highly curative. Predicting both of the major subtypes would indeed boost survival of patients who would have developed the more aggressive subtype tumours. Effective and personalised screening, diagnosis, management, and treatment modalities could be offered, should the risk prediction and stratification allow prediction of treatment response, type of tumour at-risk of being developed, and penetrance of high-risk pathogenic variants (Lewis and Vassos, 2020). However, currently, most PRSs for other conditions have low discriminatory power in the general population, and thus, a PRS for endometrial cancer should not be expected to perform well outside case-control studies (especially the enriched hospital-based cohorts) either.

1.9 Identifying the Research Question

1.9.1 Rationale

Endometrial cancer is the most commonly diagnosed gynaecological cancer among developed nations, with nearly 9,400 new cases diagnosed each year in the UK. The incidence of endometrial cancer has been increasing in parallel with the alarming obesity rise globally, making it a major health concern. Although Type I endometrial cancers have a favourable prognosis, overall endometrial cancer-related mortality is still over 20% and continues to rise. Moreover, currently available diagnostic measures and treatments are either too invasive, such as hysteroscopy and hysterectomy, or associated with significant undesirable consequences. There are no official endometrial cancer screening measures offered by the healthcare systems around the world. Epidemiological factors have been shown to have important roles in the risk of endometrial cancer, however these fail to accurately stratify all women.

An increased risk of endometrial cancer among those with affected first-degree relatives indicates the role of genetic predisposition. Within the past decade, studies on genetic variants predisposing to endometrial cancer have identified numerous SNPs, particularly through candidate-gene studies. However, those have not been validated in large GWAS, which may imply the widespread false-positives crowding the literature. In contrast, there have been far fewer but more robust GWAS which identified novel SNPs and validated others. Thus, we propose that a systematic review of the literature spanning the past decade will provide the best evidence for identifying a panel of SNPs robustly associated with endometrial cancer risk.

PRSs have been calculated in multiple diseases, including several cancers. Some of these have been useful in predicting high-risk versus low-risk individuals based on cumulative effects of multiple SNPs associated with the disease studied. Given the growing number of SNPs reported for influencing the risk of developing endometrial cancer, compiling the

effects of these SNPs in a PRS using the best evidence available to date may be useful for endometrial cancer risk assessment efforts.

Development and validation of a PRS in a clinically-rich dataset is important to elucidate the potential confounding effects of the clinical factors as well as to be able to account for the bias that could result from them. All large GWAS studies employing endometrial cancer cases and controls have made use of shared datasets. Firstly, this introduces bias by effectively looking at the same data by recycling the same studies, but also limits the validation efforts for any previously reported SNP or a potential PRS being developed. Therefore, independent datasets are essential to facilitate the validation of reported SNPs as well as development of a PRS.

Aside from PRSs, existing RPMs can be improved by looking at other genetic contributions, such as assessment of somatic or germline mutations. The role of germline mutations such as those that result in LS have been investigated by others. However, at the time of the project design, mutations in *BRCA1/2* genes, which have been postulated to be pathogenic, particularly for high-grade serous subset of endometrial cancer tumours, lack a conclusive evidence. Therefore, it is important to assess the risk conferred to individuals by *BRCA1/2* pathogenic carrier status. If indeed there is a considerable increased risk due to these mutations, it would be sensible to include them into an RPM to be able to adjust the overall risk accordingly.

1.9.2 Hypothesis

The overall hypothesis of this PhD was that genomics tools aid the prediction of the risk of developing endometrial cancer, exclusively or in combination with other risk factors. The first hypothesis was that SNPs influence the risk of endometrial cancer. The second hypothesis was that a PRS aids risk prediction of endometrial cancer. The third hypothesis was that the *BRCA1/2* genes increase the risk of developing endometrial cancer, particularly of the serous subtype.

1.9.3 Project Aims

In line with these specific hypotheses, this study aimed to:

1. systematically review the literature to identify a panel of top SNPs with the strongest evidence for endometrial cancer predisposition,

2. genotype prospectively collected endometrial cancer cases using the OncoArray chip, analyse the genotype data from these cases alongside OncoArray data obtained for independent controls, by conducting a GWAS for untargeted analysis and targeted analysis for the SNP panel,

3. derive a PRS using the most predictive SNPs and validate the resulting PRS in an independent dataset(s),
   and

4. investigate the association between *BRCA1/2* genes and endometrial cancer risk.

2 Association between genetic polymorphisms and endometrial cancer risk: a systematic review

This Chapter has been published in Journal of Medical Genetics (Bafligil *et al*., 2020) which can be found at **Appendix 8.1**. The article describes the methods and results of a systematic review of literature to identify the most robust SNPs that are associated with endometrial cancer predisposition. The SNP panel identified here was validated in our Manchester cohort (**Chapter 3, p 105**) and was used to derive a PRS described in **Chapter 4, p 150**.

2.1 Contributorship Statement

I performed the systematic review, data acquisition and synthesis, statistical analysis and wrote the manuscript. D.J.T and A.L provided statistical support for the analysis. N.A.J.R and A.N supported data acquisition.

2.2 Abstract

Introduction Endometrial cancer is one of the most commonly diagnosed cancers in women. Although there is a hereditary component to endometrial cancer, most cases are thought to be sporadic and lifestyle related. The aim of this study was to systematically review prospective and retrospective case-control studies, meta-analyses and genome-wide association studies to identify genomic variants that may be associated with endometrial cancer risk.

Methods We searched MEDLINE, Embase and CINAHL from 2007 to 2019 without restrictions. We followed PRISMA 2009 guidelines. The search yielded 3015 hits in total. Following duplicate exclusion, 2674 abstracts were screened and 453 full-texts evaluated based on our pre-defined screening criteria. 149 articles were eligible for inclusion.

Results We found that single nucleotide polymorphisms (SNPs) in *HNF1B*, *KLF*, *EIF2AK*, *CYP19A1*, *SOX4* and *MYC* were strongly associated with incident endometrial cancer. Nineteen variants were reported with genome-wide significance and a further five with suggestive significance. No convincing evidence was found for the widely studied *MDM2* variant rs2279744. Publication bias and false discovery rates were noted throughout the literature.

Conclusion Endometrial cancer risk may be influenced by SNPs in genes involved in cell survival, oestrogen metabolism and transcriptional control. Larger cohorts are needed to identify more variants with genome-wide significance.

2.3 Introduction

Endometrial cancer is the most common gynaecological malignancy in the developed world (Sundar *et al*., 2017). Its incidence has risen over the last two decades as a consequence of the ageing population, fewer hysterectomies for benign disease and the obesity epidemic. In the United States, it is estimated that women have a 1 in 35 lifetime risk of

endometrial cancer, and in contrast to cancers of most other sites, cancer-specific mortality has risen by approximately 2% every year since 2008 related to the rapidly rising incidence (Siegel *et al*., 2019).

Endometrial cancer has traditionally been classified into Type I and Type II based on morphology (Morice *et al*., 2016). The more common subtype, type I, is mostly comprised of endometrioid tumours and is oestrogen-driven, arises from a hyperplastic endometrium, presents at an early stage and has an excellent 5-year survival rate (Tzur *et al*., 2017). By contrast, type II includes non-endometrioid tumours, specifically serous, carcinosarcoma and clear cell subtypes, which are biologically aggressive tumours with a poor prognosis that are often diagnosed at an advanced stage (Clarke *et al*., 2019). Recent efforts have focused on a molecular classification system for more accurate categorization of endometrial tumours into four groups with distinct prognostic profiles (Getz *et al*., 2013; Stelloo *et al*., 2015).

The majority of endometrial cancers arise through the interplay of familial, genetic and lifestyle factors. Two inherited cancer predisposition syndromes, Lynch syndrome and the much rarer Cowden syndrome, substantially increase the lifetime risk of endometrial cancer, but these only account for around 3-5% of cases (Win *et al*., 2015; Constantinou and Tischkowitz, 2017; Ryan *et al*., 2019). Having first or second degree relative(s) with endometrial or colorectal cancer increases endometrial cancer risk, although a large European twin study failed to demonstrate a strong heritable link (Lichtenstein *et al*., 2000). The authors failed to show that there was greater concordance in monozygotic than dizygotic twins, but the study was based on relatively small numbers of endometrial cancers. Lu and colleagues reported an association between common single nucleotide polymorphisms (SNPs) and endometrial cancer risk, revealing the potential role of SNPs in explaining part of the risk in both the familial and general populations (Lu *et al*., 2014). Thus far, many SNPs have been reported to modify susceptibility to endometrial cancer; however, much of this work predated genome-wide association studies (GWAS) and is of variable

quality. Understanding genetic predisposition to endometrial cancer could facilitate personalized risk assessment with a view to targeted prevention and screening interventions (Kitson *et al*., 2017). This emerged as the most important unanswered research question in endometrial cancer according to patients, carers and healthcare professionals in our recently completed James Lind Womb Cancer Alliance Priority Setting Partnership (Wan *et al*., 2016). It would be particularly useful for non-endometrioid endometrial cancers, for which advancing age is so far the only predictor (Raglan *et al*., 2019).

We therefore conducted a comprehensive systematic review of the literature to provide an overview of the relationship between SNPs and endometrial cancer risk. We compiled a list of the most robust endometrial cancer-associated SNPs. We assessed the applicability of this panel of SNPs with a theoretical polygenic risk score (PRS) calculation. We also critically appraised the meta-analyses investigating the most frequently reported SNPs in *MDM2*. Finally, we described all SNPs reported within genes and pathways that are likely involved in endometrial carcinogenesis and metastasis.

2.4 Materials and Methods

Our systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) collaboration 2009 recommendations. The registered protocol is available through PROSPERO (CRD42018091907) (PROSPERO, 2020).

2.4.1 Search Strategy

We searched Embase, MEDLINE and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases via Healthcare Databases Advanced Search (HDAS) platform, from 2007 to 2018 to identify studies reporting associations between polymorphisms and endometrial cancer risk. Key words including MeSH (Medical Subject Heading) terms and free-text words were searched in both titles and abstracts. The following terms were used: "endomet*","uter*", "womb", "cancer(s)", "neoplasm(s)", "endometrium tumor", "carcinoma", "adenosarcoma", "clear cell carcinoma", "carcinosarcoma", "SNP", "single nu-

cleotide polymorphism", "GWAS", and "genome wide association study/ies". No other restrictions were applied. The search was repeated with time restrictions between 2018 and June 2019 to capture any recent publications.

2.4.2 Eligibility Criteria

Studies were selected for full-text evaluation if they were primary articles investigating a relationship between endometrial cancer and SNPs. Study outcome was either the increased or decreased risk of endometrial cancer relative to controls reported as an odds ratio (OR) with corresponding 95% confidence intervals (95% CIs).

2.4.3 Study Selection

Three independent reviewers screened all articles uploaded to a screening spread sheet developed by Helena VonVille (VonVille, H., 2015). Disagreements were resolved by discussion. Chronbach's alpha score was calculated between reviewers and indicated high consistency at 0.92. Case-control, prospective and retrospective studies, GWAS, and both discovery and validation studies were selected for full-text evaluation. Non-English articles, editorials, conference abstracts and proceedings, letters and correspondence, case reports and review articles were excluded.

Candidate-gene studies with at least 100 women and GWAS with at least 1,000 women in the case arm were selected to ensure reliability of the results, as explained by Spencer *et al.* 2009 (Spencer *et al.*, 2009). To construct a panel of up to 30 SNPs with strongest evidence of association, those with the strongest *P* values were selected. For the purpose of a SNP panel, articles utilizing broad European or multi-ethnic cohorts were selected. Where overlapping populations were identified, the most comprehensive study was included.

2.4.4 Data Extraction and Synthesis

For each study, the following data were extracted: SNP ID, nearby gene(s)/chromosome location, OR (95% CI), *P* value, minor or effect allele frequency (MAF/EAF), effect allele

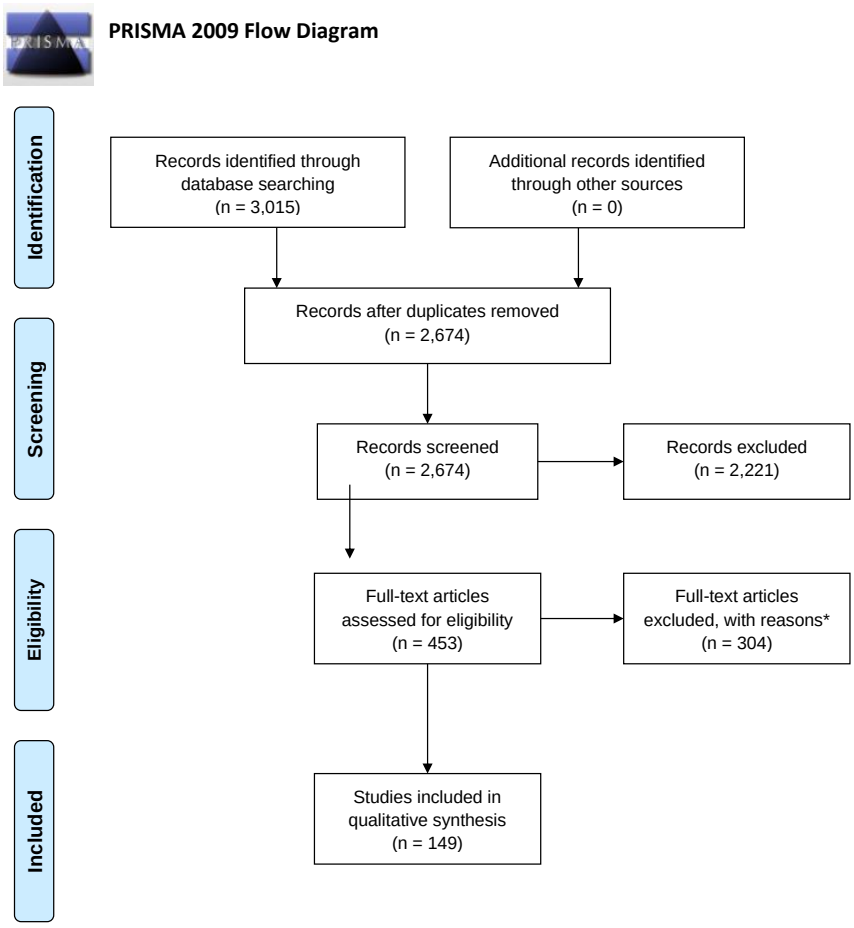
(EA) and other allele (OA), adjustment, ethnicity and ancestry, number of cases and controls, endometrial cancer type, and study type including discovery or validation study and meta-analysis. For risk estimates, a preference towards most adjusted results was applied. For candidate-gene studies, a standard *P* value of < 0.05 was applied and for GWAS a *P* value of < 5E-8, indicating genome-wide significance, was accepted as statistically significant. However, due to the limited number of SNPs with *P* values reaching genome-wide significance, this threshold was then lowered to < 1E-5, allowing for marginally significant SNPs to be included. As shown by Mavaddat *et al.* for breast cancer, SNPs that fall below genome-wide significance may still be useful for generating a PRS and improving the models (Mavaddat *et al*., 2019).

We estimated the potential value of a PRS constructed based on the most significant SNPs by comparing the predicted risk for a woman with a risk score in the top 1% of the distribution to the mean predicted risk. Per-allele ORs and MAFs were taken from the publications and standard errors (SEs) for the lnORs were derived from published 95% CIs. The PRS was assumed to have a Normal distribution, with mean $2\Sigma\beta_i p_l$ and SE, σ, equal to $\sqrt{2\Sigma\beta_i^2 p_l(1-p_i)}$, according to the Binomial distribution, where the summation is over all SNPs in the risk score. Hence the relative risk (RR) comparing the top 1% of the distribution to the mean is given by $exp(Z_{0.01}\sigma)$, where Z is the inverse of the standard normal cumulative distribution.

2.5 Results

The flow chart of study selection is illustrated in **Figure 2.1**. In total, 453-text articles were evaluated and of those, 149 articles met our inclusion criteria. One study was excluded from **Table 2.1**, for having an Asian-only population, as this would make it harder to compare with the rest of the results which were all either multi-ethnic or European cohorts, as stated in our inclusion criteria for the SNP panel (Geng *et al*., 2018). Any SNPs without 95% CIs were also excluded from any downstream analysis. Additionally, SNPs in linkage disequilibrium (r^2>0.2) with each other were examined, and of those in linkage

disequilibrium, the SNP with strongest association was reported. Per allele ORs were used unless stated otherwise.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 2.1: *Study selection flow diagram.*

* Reasons: Irrelevant articles, articles focusing on other conditions, non-GWAS/candidate-gene study related articles, technical and duplicate articles. GWAS, genome-wide association study. Adapted from: Moher *et al*. (2009)

2.5.1 Top SNPs Associated with Endometrial Cancer Risk

Following careful interpretation of the data, twenty-four independent SNPs with the lowest *P* values that showed the strongest association with endometrial cancer were obtained (**Table 2.1**) (Spurdle *et al*., 2011; Chen *et al*., 2014; O'Mara *et al*., 2015; Painter *et al*., 2016; O'Mara *et al*., 2018). These SNPs are located in or around genes coding for transcription factors, cell growth and apoptosis regulators, and enzymes involved in the steroidogenesis pathway. All the SNPs presented here were reported on the basis of a GWAS or in one case, an exome-wide association study, and hence no SNPs from candidate-gene studies made it to the list. This is partly due to the nature of larger GWAS providing more comprehensive and powered results as opposed to candidate gene studies. Additionally, a vast majority of SNPs reported by candidate-gene studies were later refuted by large-scale GWAS such as in the case of *TERT* and *MDM2* variants (Zajac *et al*., 2012; Cheng *et al*., 2015). The exception to this is the *CYP19* gene, where candidate-gene studies reported an association between variants in this gene with endometrial cancer in both Asian and broad European populations, and this association was more recently confirmed by large-scale GWAS (Tao *et al*., 2007; Lundin *et al*., 2012; Cheng *et al*., 2016; O'Mara *et al*., 2018). Moreover, a recent article authored by O'Mara and colleagues reviewed the GWAS that identified most of the currently known SNPs associated with endometrial cancer (O'Mara *et al*., 2019a).

Table 2.1: *List of top SNPs most likely to contribute to endometrial cancer risk identified through systematic review of recent literature.*

Reference	SNP ID	Nearby Gene(s)	Location	OR	LCI	UCI	*P*	EAF	EA	OA	Ethnicity	Cases (n)	Controls (n)	EC Type	Position	Datasets*
O'Mara *et al.* (2018)	rs11263761	HNF1B	17q12	1.15	1.12	1.19	3.20E-20	0.52	A	G	EUR	12906	108979	All	intronic	**NSECG**, UK1-CORGI, **SEARCH**, WTCCC, **ANECS**, QIMR, HCS, E2C2, BECS/HJECS, **iCOGS**‡, BBCC, BSUCH, ESTHER, GC-HBOC, GENICA, MARIE, **MoMaTEC**, NBCS, BBCS, SBCS, UKBGS, **NECS**, ABCFS, ABCTB, BCEES, MCCS, **LES**, LMBC, **BECS**, GESBC, HaBCS, CAHRES, RENDOCAS, MISS, pKARMA, SMC, CBR_STUDY98, MECS, MCBCS, MMHS, WHI, UKBB
	rs7981863	KLF5, KLF12	13q22.1	1.16	1.12	1.20	2.70E-17	0.72	C	T	EUR	12906	108979	All	intergenic	
	rs1740828	SOX4	6p22.3	1.15	1.11	1.19	4.20E-16	0.52	G	A	EUR	12906	108979	All	regulatory	
	rs17601876	CYP19A1	15q21.2	1.12	1.09	1.16	3.30E-14	0.48	G	A	EUR	12906	108979	All	intronic	
	rs4733613	MYC	8q24.21	1.18	1.13	1.24	7.50E-14	0.12	C	G	EUR	12906	108979	All	intergenic	
	rs3184504	SH2B3	12q24.11	1.10	1.07	1.14	1.10E-10	0.52	C	T	EUR	12906	108979	All	missense	
	rs2747716	HEY2, NCOA7, MYC	6q22.31	1.10	1.07	1.14	2.90E-10	0.57	A	G	EUR	12906	108979	All	intronic	
	rs9668337	SSPN	12p12.1	1.11	1.08	1.15	1.10E-09	0.74	A	G	EUR	12906	108979	All	non-coding exon	
	rs35286446	MYC	8q24.21	1.10	1.06	1.13	3.10E-09	0.58	GAT	G	EUR	12906	108979	All	intronic	
	rs10850382	LOC107984437	12q24.21	1.10	1.07	1.14	3.50E-09	0.31	T	C	EUR	12906	108979	All	regulatory	
	rs882380	SKAP, SMX11	17q21.32	1.10	1.06	1.13	4.70E-09	0.61	A	C	EUR	12906	108979	All	intronic	
	rs937213	EIF2AK4, BMF	15q15.1	1.09	1.06	1.13	5.10E-09	0.42	C	T	EUR	12906	108979	All	intronic	
	rs1679014	CDKN2A, CDKN2B	9p21.3	1.18	1.12	1.25	6.40E-09	0.07	T	C	EUR	12906	108979	All	intronic	
Painter *et al.* (2016)	rs2498794	AKT1	14q32.33	1.13	1.09	1.17	8.70E-09	0.48	G	A	EUR	7737	37144	All	intronic	**ANECS**, **SEARCH**, **NSECG**, **iCOGS**
O'Mara *et al.* (2018)	rs10835920	WT1, WT1-AS, EIF3M	11p13	1.09	1.06	1.13	1.30E-08	0.38	T	C	EUR	12906	108979	All	intergenic	**NSECG**, UK1-CORGI, **SEARCH**, WTCCC, **ANECS**, QIMR, HCS, E2C2, BECS/HJECS, **iCOGS**, BBCC, BSUCH, ESTHER, GC-HBOC, GENICA, MARIE, **MoMaTEC**, NBCS, BBCS, SBCS, UKBGS, **NECS**, ABCFS, ABCTB, BCEES, MCCS, **LES**, LMBC, **BECS**, GESBC, HaBCS, CAHRES, RENDOCAS, MISS, pKARMA, SMC, CBR_STUDY98, MECS, MCBCS, MMHS, WHI, UKBB
	rs139584729	MYC	8q24.21	1.40	1.25	1.58	2.40E-08	0.98	C	G	EUR	12906	108979	All	intergenic	
	rs148261157	BCL11A	2p16.1	1.26	1.16	1.36	3.40E-08	0.03	A	G	EUR	12906	108979	All	intergenic	
				1.25	1.14	1.38	4.70E-06	0.03				8758	46126	End		
				1.64	1.32	2.04	9.60E-06	0.03				1230	35447	NE		
	rs113998067	GNL2, RSPO1, CDCA8	1p34.3	1.23	1.14	1.32	3.60E-08	0.04	C	T	EUR	12906	108979	All	intergenic	
	rs1129506	EVI2A, NF1	17q11.2	1.10	1.06	1.13	4.30E-08	0.38	G	A	EUR	12906	108979	All	missense	
Spurdle *et al.* (2011)	rs673604	SFPQ	1p34	1.21	1.12	1.32	5.90E-06	0.08	G	A	EUR	1265	5190	All	regulatory	**ANECS**, **NECS**, **SEARCH**, WTCCC2, **BECS**, **LES**, **MoMaTEC**, **NSECG**, PECS, SASBAC, SECGS

Reference	SNP ID	Nearby Gene(s)	Location	OR	LCI	UCI	*P*	EAF	EA	OA	Ethnicity	Cases (n)	Controls (n)	EC Type	Position	Datasets*
O'Mara *et al.* (2015)	rs79575945	ESR1	6q25	1.20	1.11	1.30	3.76E-06	0.07	G	A	EUR	6607	37925	End	intronic	**iCOGS**, BCAC, OCAC, **ANECS**, **SEARCH**, **NSECG**
Chen *et al.* (2014)	rs1953358	LINC00520	14q22.3	1.36	1.20	1.53	4.76E-07	0.49	G	A	ME	1055	1778	End	intergenic	AHS, EDGE, FHCRC, MEC
	rs8178648	PROS1	chr3	1.71	1.37	2.12	1.53E-06	0.09	G	A	ME	1055	1778	End	intronic	
	rs9399840	N/A	6q16.3	1.33	1.18	1.49	3.01E-06	0.53	T	C	ME	1055	1778	End	intergenic	

The different studies listed here used overlapping datasets (in bold).

All locations were based on Genome Reference Consortium Human Build 37 (GRCh37). Variants located at 8q24.21 were obtained from a conditional model.

* NSECG: UK National Study of Endometrial Cancer Genetics, UK1-CORGI: UK Colorectal Tumour Gene Identification Consortium, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, WTCCC1/2: Wellcome Trust Case Control Consortium 1/2, ANECS: Australian National Endometrial Cancer Study, QIMR: Queensland Institute of Medical Research, HCS: Hunter Community Study, E2C2: NCI-supported international consortium of four US-based cohort studies, 2 US-based case-control studies and 1 Polish case-control study, BECS/HJECS: Bavarian Endometrial Cancer Study/Hannover-Jena Endometrial Cancer Study, BBCC: Bavarian Breast Cancer Cases and Controls, BSUCH: Breast Cancer Study of the University Clinic Heidelberg, ESTHER: ESTHER Breast Cancer Study, GC-HBOC: German Consortium for Hereditary Breast & Ovarian Cancer, GENICA: Gene Environment Interaction and Breast Cancer in Germany, MARIE: Mammary Carcinoma Risk Factor Investigation, MoMaTEC: Molecular Markers in Treatment of Endometrial Cancer, NBCS: Norwegian Breast Cancer Study, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, NSECG: National Study of Endometrial Cancer Genetics, BBCS: British Breast Cancer Study, SBCS: Sheffield Breast Cancer Study, UKBGS: UK Breakthrough Generations Study, ANECS: Australian National Endometrial Cancer Study, NECS: Newcastle Endometrial Cancer Study, ABCFS: Australian Breast Cancer Family Study, ABCTB: Australian Breast Cancer Tissue Bank, BCEES: Breast Cancer Employment and Environment Study, MCCS: Melbourne Collaborative Cohort Study, LES: Leuven Endometrial Cancer Study, LMBC: Leuven Multidisciplinary Breast Centre, BECS: Bavarian Endometrial Cancer Study, BBCC: Bavarian Breast Cancer Cases and Controls, BSUCH: Breast Cancer Study of the University Clinic Heidelberg, GENICA: Gene Environment Interaction and Breast Cancer in Germany, GESBC: Genetic Epidemiology Study of Breast Cancer by Age 50, HaBCS: Hannover Breast Cancer Study, MARIE: Mammary Carcinoma Risk Factor Investigation, CAHRES: Cancer Hormone Replacement Epidemiology, RENDOCAS: Registry of Endometrial Cancer in Sweden, MISS: Melanoma Inquiry of Southern Sweden, pKARMA: Karolinska Mammography Project for Risk Prediction of Breast Cancer, SMC: Swedish Mammography Cohort, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, NSECG: National Study of Endometrial Cancer Genetics, BBCS: British Breast Cancer Study, CBR_STUDY98: Cambridge BioResource, UKBGS: UK Breakthrough Generations Study, MECS: Mayo Endometrial Cancer Study, MCBCS: Mayo Clinic Breast Cancer Study, MMHS: Mayo Mammography Health Study, WHI: Women's Health Initiative, UKBB: UK BioBank, OCAC: Ovarian Cancer Association Consortium, AHS: Alberta Health Services, EDGE: Estrogen, Diet, Genetics and Endometrial Cancer, FHCRC: Fred Hutchinson Cancer Research Center, MEC: Multiethnic Cohort Study, PECS: Polish Endometrial Cancer Study, SASBAC: Singapore and Swedish Breast/Endometrial Cancer Study, SECGS: Shanghai Endometrial Cancer Genetic Study.

† iCOGS dataset breakdown was not indicated by the listed studies here other than O'Mara *et al.* (2018).

EA, effect allele; EAF, effect allele frequency; EC, endometrial cancer; EUR, European cohort; LCI, lower confidence intervals; ME, multiethnic; NE, non-endometrioid; OA, other allele; SNP, single nucleotide polymorphism; UCI, upper confidence interval.

Most of the studies represented in **Table 2.1** are GWAS and the majority of these involved broad European populations. Those having a multi-ethnic cohort also consisted primarily of broad European populations. Only 4 of the variants in **Table 2.1** are located in coding regions of a gene, or in regulatory flanking regions around the gene. Thus, most of these variants would not be expected to cause any functional effects on the gene or the resulting protein. An eQTL search using GTEx Portal showed that some of the SNPs are significantly associated ($P < 0.05$) with modified transcription levels of the respective genes in various tissues such as prostate (rs11263761), thyroid (rs9668337), pituitary (rs2747716), breast mammary (rs882380) and testicular (rs2498794) tissue, as summarised in **Table 2.2**.

The only variant for which there was an indication of a specific association with non-endometrioid endometrial cancer was rs148261157 near the *BCL11A* gene. The A allele of this SNP had a moderately higher association in the non-endometrioid arm (OR=1.64, 95% CI 1.32-2.04, P = 9.6E-6) compared to the endometrioid arm (OR=1.25, 95% CI 1.14-1.38, P = 4.7E-6) (O'Mara *et al.*, 2018).

Oestrogen receptors α and β encoded by *ESR1* and *ESR2*, respectively, have been extensively studied due to the assumed role of oestrogens in the development of endometrial cancer. O'Mara *et al.* reported a lead SNP (rs79575945) in the *ESR1* region that was associated with endometrial cancer (p = 1.86E-5) (O'Mara *et al.*, 2015). However, this SNP did not reach genome-wide significance in a more recent larger GWAS (O'Mara *et al.*, 2018). No statistically significant associations have been reported between endometrial cancer and SNPs in the *ESR2* gene region.

Table 2.2: *List of eQTL hits for the selected panel of SNPs.*

SNP ID	Significant eQTL for	*P*	Tissue	Other Gene(s)	Other Tissue(s)
rs17601876	GLDN CYP19A1 CYP19A1	1.20E-08 3.40E-07 0.0000058	Adipose - Subcutaneous Whole Blood Adipose - Subcutaneous	SPPL2A, DMXL2	Skin - Sun Exposed (Lower leg), Colon - Sigmoid, Cells - Cultured fibroblasts, Muscle - Skeletal, Spleen, Skin - Not Sun Exposed (Suprapubic), Nerve - Tibial
rs3184504	TMEM116 MAPKAPK5	0.00017 0.00026	Adipose - Subcutaneous Adipose - Subcutaneous	ALDH2, LINC01405, ADAM1B	Oesophagus - Mucosa, Skin - Not Sun Exposed (Suprapubic), Skin - Sun Exposed (Lower leg), Muscle - Skeletal, Artery - Aorta, Heart - Atrial Appendage, Artery - Tibial, Colon - Sigmoid, Brain - Nucleus accumbens (basal ganglia)
rs2747716	RP11-624M8.1 RP11-624M8.1 HEY2 HEY2 RP11-624M8.1 RP11-624M8.1	4.20E-11 8.20E-11 9.70E-10 2.10E-09 1.70E-07 0.0000013	Pituitary Adipose - Subcutaneous Testis Ovary Breast - Mammary Tissue Ovary	HDDC2	Artery - Tibial, Pancreas, Thyroid, Brain - Nucleus accumbens (basal ganglia), Brain - Substantia nigra, Oesophagus - Muscularis, Nerve - Tibial, Brain - Caudate (basal ganglia), Adipose - Visceral (Omentum), Brain - Spinal cord (cervical c-1), Artery - Aorta, Brain - Cortex, Brain - Hypothalamus, Muscle - Skeletal, Brain - Cerebellum, Heart - Left Ventricle, Brain - Putamen (basal ganglia), Brain - Frontal Cortex (BA9), Brain - Cerebellar Hemisphere
rs9668337	BHLHE41 SSPN	9.00E-17 0.00011	Thyroid Thyroid	RP11-283G6.3	Cells - Cultured fibroblasts
rs882380	SNX11 SNX11 SNX11 COPZ2 SKAP1 HOXB2	3.10E-25 1.00E-21 1.20E-13 9.30E-12 3.30E-08 0.000026	Adipose - Subcutaneous Whole Blood Breast - Mammary Tissue Testis Whole Blood Adipose - Subcutaneous	RP5-890E16.5, CBX1, LRRC46, MRPL10, RP11-6N17.4, CDK5RAP3, SP6, PRR15L, RP5-890E16.2, PNPO, RP11-6N17.3, HOXB1, HOXB-AS1, NFE2L1	Skin - Sun Exposed (Lower leg), Cells - Cultured fibroblasts, Adipose - Visceral (Omentum), Lung, Skin - Not Sun Exposed (Suprapubic), Pancreas, Spleen, Oesophagus - Muscularis, Artery - Aorta, Heart - Atrial Appendage, Liver, Colon - Transverse, Thyroid, Artery - Tibial, Colon - Sigmoid, Oesophagus - Gastroesophageal Junction, Stomach, Muscle - Skeletal, Small Intestine - Terminal Ileum, Prostate, Brain - Cerebellum, Brain - Cerebellar Hemisphere, Minor Salivary Gland, Adrenal Gland, Oesophagus - Mucosa
rs937213	EIF2AK4 EIF2AK4 RP11-521C20.5 RP11-521C20.5	4.70E-11 3.40E-08 5.40E-07 7.40E-07	Adipose - Visceral (Omentum) Breast - Mammary Tissue Testis Prostate	SRP14 NA NA NA	Thyroid, Oesophagus - Mucosa, Skin - Sun Exposed (Lower leg), Stomach, Oesophagus - Muscularis, Pancreas, Skin - Not Sun Exposed (Suprapubic), Colon - Transverse, Adipose - Subcutaneous, Lung, Colon - Sigmoid, Muscle - Skeletal, Nerve - Tibial, Whole Blood, Oesophagus - Gastroesophageal Junction, Artery - Tibial, Adrenal Gland, Spleen, Heart - Left Ventricle, Heart - Atrial Appendage
rs2498794	AKT1 ADSSL1 SIVA1 ADSSL1 SIVA1	1.70E-30 5.50E-25 1.80E-07 0.000026 0.000044	Thyroid Testis Adipose - Visceral (Omentum) Ovary Breast - Mammary Tissue	ZBTB42	Oesophagus - Mucosa, Artery - Tibial, Oesophagus - Muscularis, Skin - Sun Exposed (Lower leg), Skin - Not Sun Exposed (Suprapubic), Cells - Cultured fibroblasts, Artery - Aorta, Oesophagus - Gastroesophageal Junction, Adipose - Subcutaneous, Colon - Sigmoid, Colon - Transverse, Heart - Atrial Appendage
rs10835920	WT1-AS	0.0000055	Spleen	NA	Oesophagus - Muscularis
rs148261157	KIAA1841	0.000013	Oesophagus - Muscularis	NA	NA
rs113998067	RSPO1	2.70E-10	Artery - Tibial	EPHA10, FHL3, DNALI1	Nerve - Tibial, Artery - Aorta, Colon - Transverse
rs1129506	EVI2A NF1 NF1 NF1	4.30E-20 3.50E-09 2.20E-07 3.70E-07	Whole Blood Adipose - Subcutaneous Thyroid Testis	OMG, RAB11FIP4	Spleen, Oesophagus - Mucosa, Artery - Tibial, Lung, Artery - Aorta, Skin - Sun Exposed (Lower leg), Nerve - Tibial, Heart - Atrial Appendage, Adipose - Visceral (Omentum), Cells - Cultured fibroblasts, Liver, Stomach, Brain - Amygdala, Skin - Not Sun Exposed (Suprapubic), Brain - Caudate (basal ganglia), Muscle - Skeletal, Colon - Sigmoid

SNP ID	Significant eQTL for	*P*	Tissue	Other Gene(s)	Other Tissue(s)
rs673604	ZMYM1 MAP7D1	7.00E-07 0.00001	Adipose - Subcutaneous Whole Blood	RP4-665N4.8, ZMYM4, KIAA0319L, TFAP2E	Skin - Sun Exposed (Lower leg), Oesophagus - Muscularis, Cells - EBV-transformed lymphocytes, Oesophagus - Mucosa, Nerve - Tibial, Brain - Cerebellum
rs1953358	LINC00520	0.000015	Skin - Not Sun Exposed (Suprapubic)	NA	NA
rs8178648	PROS1	0.0003	Skin - Sun Exposed (Lower leg)	NA	NA

Top significant eQTL hits from different tissues are shown in the table. There were no significant hits reported for some SNPs which are hence not included in this table.

EBV, Epstein-Barr virus; SNP, single nucleotide polymorphism.

AKT is an oncogene linked to endometrial carcinogenesis. It is involved in the PI3K/AKT/-mTOR pro-proliferative signalling pathway to inactivate apoptosis and allow cell survival. The A allele of rs2494737 and G allele of rs2498796 were reported to be associated with increased and decreased risk of endometrial cancer in 2016, respectively (Cheng *et al*., 2016; Painter *et al*., 2016). However, this association was not replicated in a larger GWAS in 2018 (O'Mara *et al*., 2018). Nevertheless, given the previous strong indications, and biological basis that could explain endometrial carcinogenesis, we decided to include an *AKT1* variant (rs2498794) in our results.

PTEN is a multi-functional tumour suppressor gene that regulates the AKT/PKB signalling pathway and is commonly mutated in many cancers including endometrial cancer (Banno *et al*., 2014). Loss-of-function germline mutations in *PTEN* are responsible for Cowden syndrome, which exerts a lifetime risk of endometrial cancer of up to 28% (Constantinou and Tischkowitz, 2017). Lacey and colleagues studied SNPs in the *PTEN* gene region; however, none showed significant differences in frequency between 447 endometrial cancer cases and 439 controls of European ancestry (Lacey *et al*., 2011).

KRAS mutations are known to be present in endometrial cancer. These can be activated by high levels of KLF5 (transcriptional activator). Three SNPs have been identified in or around *KLF5* that are associated with endometrial cancer. The G allele of rs11841589 (OR=1.15, 95% CI 1.11-1.21, P = 4.83E-11), the A allele of rs9600103 (OR=1.23, 95% CI 1.16-1.30, P = 3.76E-12) and C allele of rs7981863 (OR=1.16, 95% CI 1.12-1.20, P = 2.70E-17) have all been found to be associated with an increased likelihood of endometrial cancer in large European cohorts (Chen *et al*., 2016; Cheng *et al*., 2016; O'Mara *et al*., 2018). It is worth noting that these SNPs are not independent, and hence they quite possibly tag the same causal variant.

The *MYC* family of proto-oncogenes encode transcription factors that regulate cell proliferation, which can contribute to cancer development if dysregulated. The recent GWAS

by O'Mara *et al*. reported three SNPs within the *MYC* region that reached genome-wide significance with conditional *P* values reaching at least 5E-8 (O'Mara *et al*., 2018).

To test the utility of these SNPs as predictive markers, we devised a theoretical PRS calculation using the log ORs and EAFs per SNP from the published data. The results were very encouraging with an RR of 3.16 for the top 1% versus the mean, using all the top SNPs presented in **Table 2.1** and 2.09 when using only the SNPs that reached genome-wide significance (including *AKT1*).

2.5.2 Controversy Surrounding *MDM2* Variant SNP309

MDM2 negatively regulates tumour suppressor gene *TP53*, and as such, has been extensively studied in relation to its potential role in predisposition to endometrial cancer. Our search identified six original studies of the association between *MDM2* SNP rs2279744 (also referred to as SNP309) and endometrial cancer, all of which found a statistically significant increased risk per copy of the G allele. Two more original studies were identified through our full-text evaluation; however, these were not included here as they did not meet our inclusion criteria; one due to small sample size, other due to studying rs2279744 status dependent on another SNP (Walsh *et al*., 2007; Gansmo *et al*., 2017). Even so, the two studies were described in multiple meta-analyses that are listed in **Table 2.3**. Different permutations of these eight original studies appear in at least eight published meta-analyses. However, even the largest meta-analysis contained <4000 cases (**Table 2.3**) (Zou *et al*., 2018).

In comparison, a GWAS including nearly 13,000 cases found no evidence of an association with OR and corresponding 95% CI of 1.00 (0.97 - 1.03) and a *P* value of 0.93 (personal communication) (O'Mara *et al*., 2018). Nevertheless, we cannot completely rule out a role for *MDM2* variants in endometrial cancer predisposition as the candidate-gene studies reported larger effects in Asians, whereas the GWAS primarily contained participants of European ancestry. There is also some suggestion that the SNP309 variant is in linkage

disequilibrium with another variant, SNP285, which confers an opposite effect.

It is worth noting that the SNP285C/SNP309G haplotype frequency was observed in up to 8% of Europeans, thus requiring correction for the confounding effect of SNP285C in European studies (Knappskog *et al*., 2014). However, aside from one study conducted by Knappskog *et al*., no other study including the meta-analyses corrected for the confounding effect of SNP285 (Knappskog *et al*., 2012). Among the studies presented in **Table 2.3**, Knappskog *et al*. (2012) reported that after correcting for SNP285, the OR for association of this haplotype with endometrial cancer was much lower, though still significant. Unfortunately, the meta-analyses which synthesised Knappskog *et al*. (2012) as part of their analysis, did not correct for SNP285C in the European-based studies they included (Peng *et al*., 2013; Wang *et al*., 2014; Xue *et al*., 2016). It is also concerning that two meta-analyses using the same primary articles failed to report the same result, in two instances (Wan *et al*., 2011; Wang *et al*., 2014; Xue *et al*., 2016; Li *et al*., 2011).

Table 2.3: *Characteristics of studies that examined MDM2 SNP rs2279744.*

Reference	OR (95% CI)	P	EAF	Ancestry	Cases (n)	Controls (n)	EC Type	Dataset(s)
Terry *et al.* (2008)	1.32 (1.11-1.56)	0.002	N/A	European	591	1543	N/A	Nurses' Health Study (NuHS), Women's Health Study (WHS)
Ashton *et al.* (2009)	1.37 (1.06-1.79)	N/A	0.56	Caucasian	191	291	All	Hospital based
Nunobiki *et al.* (2009)	2.28 (2.02-2.54)	0.03	0.49	Japanese	102	95	All	Hospital based
Ueda *et al.* (2009)	1.91 (1.5-3.47)	0.035	0.51	Japanese	119	108	All	Hospital based
Wan *et al.* (2011)	1.54 (1.21-1.94)	0.000	N/A	N/A	N/A	N/A	N/A	Walsh 2007*, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009
Li *et al.* (2011)	1.75 (1.16-2.63)	0.007	N/A	European, Asian	1001	1889	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009
Knappskog *et al.* (2012)	1.22 (1.03-1.44)	N/A	0.36	European	392	956	N/A	Hospital based
Zajac *et al.* (2012)	1.33 (1.12-1.58)	0.001	N/A	European	152	100	N/A	Hospital based
Yoneda *et al.* (2013)	1.64 (0.81-3.28)	0.45	0.45	Asian	125	200	All	Population based
Peng *et al.* (2013)	1.6 (1.21-2.13)	0.001	N/A	European, Asian	2069	4546	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Knappskog 2012, Yoneda 2013
	1.87 (1.29-2.73)	0.01	N/A	European	1842	4251	N/A	
Zhao *et al.* (2014)	1.41 (1.04-1.92)	0.03	N/A	European, Asian	1278	2189	N/A	Walsh 2007, Terry 2008, Ashton 2009, Ueda 2009, Zajac 2012, Yoneda 2013
	1.34 (1.07-1.69)	N/A	N/A	European	859	1707	N/A	
Wang *et al.* (2014)	1.32 (1.06-1.64)	0.01	N/A	European, Asian	1967	4460	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012, Knappskog 2012, Yoneda 2013
	1.14 (0.79-1.65)	0.49	N/A	European	1769	4172	N/A	
Xue *et al.* (2016)	1.46 (1.25-1.72)	N/A	N/A	European	1690	4151	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012, Knappskog 2012, Yoneda 2013
Zhang *et al.* (2018)	1.91 (1.5-3.47)	0.035	N/A	European, Asian	762	1041	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012
Zou *et al.* (2018)	1.23 (1.06-1.41)	0.005	N/A	European, Asian, Mixed	3535	6476	All	Walsh 2007, Terry 2008, Ashton 2009, Ueda 2009, Knappskog 2012, Zajac 2012, Yoneda 2013, Okamoto 2015, Gansmo 2017*

* Walsh *et al.* (2007) and Gansmo *et al.* (2017) did not meet eligibility criteria for us to include in our evaluation.
EAF, effect allele frequency; EC, endometrial cancer; SNP, single nucleotide polymorphism.

2.6 Discussion

This article represents the most comprehensive systematic review to date, regarding critical appraisal of the available evidence of common low-penetrance variants implicated in predisposition to endometrial cancer. We have identified the most robust SNPs in the context of endometrial cancer risk. Of those, only nineteen were significant at genome-wide level and a further five were considered marginally significant. The largest GWAS conducted in this field was the discovery- and meta-GWAS by O'Mara *et al.*, which utilized 12,906 cases and 108,979 controls (O'Mara *et al.*, 2018). Despite the inclusion of all published GWAS and around 5,000 newly genotyped cases, the total number did not reach anywhere near what is currently available for other common cancers such as breast cancer. For instance, BCAC (Breast Cancer Association Consortium), stands at well over 200,000 individuals with more than half being cases, and resulted in identification of 170 SNPs in relation to breast cancer (Hamdi *et al.*, 2016; Mavaddat *et al.*, 2019). A total of 313 SNPs including imputations were then used to derive a PRS for breast cancer (Mavaddat *et al.*, 2019). Therefore, further efforts should be directed to recruit more patients, with deep phenotypic clinical data to allow for relevant adjustments and subgroup analyses to be conducted for better precision.

A recent study by Zhang and colleagues examined the polygenicity and potential for SNP-based risk prediction for 14 common cancers, including endometrial cancer, using available summary-level data from European-ancestry datasets (Zhang *et al.*, 2020). They estimated that there are just over 1,000 independent endometrial cancer susceptibility SNPs, and that a PRS comprising all such SNPs would have an area under the receiver-operator curve of 0.64, similar to that predicted for ovarian cancer, but lower than that for the other cancers in the study. The modelling in the paper suggests that an endometrial cancer GWAS double the size of the current largest study would be able to identify susceptibility SNPs together explaining 40% of the genetic variance, but that in order to explain 75% of the genetic variance it would be necessary to have a GWAS comprising close to

150,000 cases and controls, far in excess of what is currently feasible.

We found that the literature consists mainly of candidate-gene studies with small sample sizes, meta-analyses reporting conflicting results despite using the same set of primary articles, and multiple reports of significant SNPs that have not been validated by any larger GWAS. The candidate-gene studies were indeed the most useful and cheaper technique available until the mid to late 2000s. However, a lack of reproducibility (particularly due to population stratification and reporting bias), uncertainty of reported associations, and considerably high false discovery rates make these studies much less appropriate in the post-GWAS era. Unlike the candidate-gene approach, GWAS do not require prior knowledge, selection of genes or SNPs, and provide vast amounts of data. Furthermore, both the genotyping process and data analysis phases have become cheaper, the latter particularly due to faster and open-access pre-phasing and imputation tools being made available.

It is clear from **Table 2.1** that some SNPs were reported with wide 95% CI, which can be directly attributed to small sample sizes particularly when restricting the cases to non-endometrioid histology only, low EAF or poor imputation quality. Thus, these should be interpreted with caution. Additionally, most of the SNPs reported by candidate-gene studies were not detected by the largest GWAS to date conducted by O'Mara *et al*. (2018). However, this does not necessarily mean that the possibility of those SNPs being relevant should be completely dismissed. Moreover, meta-analyses were attempted for other variants, however, these showed no statistically significant association and many presented with high heterogeneity between the respective studies (data not shown). Furthermore, as many studies utilized the same set of cases and/or controls, conducting a meta-analysis was not possible for a good number of SNPs. It is therefore unequivocal that the literature is crowded with numerous small candidate-gene studies and conflicting data. This makes it particularly hard to detect novel SNPs and conduct meaningful meta-analyses.

We found convincing evidence for nineteen variants that indicated the strongest associa-

tion with endometrial cancer, as shown in **Table 2.1**. The associations between endometrial cancer and variants in or around *HNF1B*, *CYP19A1*, *SOX4*, *MYC*, *KLF* and *EIF2AK* found in earlier GWAS were then replicated in the latest and largest GWAS. These SNPs showed promising potential in a theoretical PRS we devised based on published data. Using all 24 or genome-wide significant SNPs only, women with a PRS in the top 1% of the distribution would be predicted to have a risk of endometrial cancer 3.16 and 2.09 times higher than the mean risk, respectively.

However, the importance of these variants and relevance of the proximate genes in a functional or biological context is challenging to evaluate. Long distance promoter regulation by enhancers may disguise the genuine target gene. In addition, enhancers often do not loop to the nearest gene, further complicating the relevance of nearby gene(s) to a GWAS hit. In order to elucidate biologically relevant candidate target genes in endometrial cancer, O'Mara *et al*. looked into promoter-associated chromatin looping using a modern HiChIP approach (O'Mara *et al*., 2019b). The authors utilised normal and tumoural endometrial cell lines for this analysis which showed significant enrichment for endometrial cancer heritability with 103 candidate target genes identified across the 13 risk loci identified by the largest ECAC GWAS. Notable genes identified here were *CDKN2A* and *WT1*, and their antisense counterparts. The former was reported to be nearby of rs1679014 and the latter of rs10835920, as shown in **Table 2.1**. Moreover, 36 of the candidate target genes, 17 were found to be downregulated while 19 were upregulated in endometrial tumours.

The authors also investigated overlap between the 13 endometrial cancer risk loci and top eQTL variants for each target gene (O'Mara *et al*., 2019b). In whole blood, two particular lead SNPs; rs8822380 at 17q21.32 was a top eQTL for *SNX11* and *HOXB2*, whereas rs937213 at 15q15.1 was a top eQTL for *SRP14*. In endometrial tumour, rs7579014 at 2p16.1 was found to be a top eQTL for *BCL11A*. This is particularly interesting because *BCL11A* was the only nearby/candidate gene that had a GWAS association reported in both endometrioid and non-endometrioid subtypes. The study looked at protein-protein interac-

tions between endometrial cancer drivers and candidate target gene products. Significant interactions were observed with TP53 (most significant), AKT, PTEN, ESR1 and KRAS, among others. Finally, when 103 target candidate genes and 387 proteins were combined together, 462 pathways were found to be significantly enriched. Many of these are related to gene regulation, cancer, obesity, insulinaemia and oestrogen exposure. This study clearly showed a potential biological relevance for some of the SNPs reported by ECAC GWAS in 2018.

Most of the larger included studies used cohorts primarily composed of women of broad European descent. Hence, there are negligible data available for other ethnicities, particularly African women. This is compounded by the lack of reference genotype data available for comparative analysis, making it harder for research to be conducted in ethnicities other than Europeans. This poses a problem for developing risk prediction models that are equally valuable and predictive across populations. Thus, our results also are of limited applicability to non-European populations.

Furthermore, considering that non-endometrioid cases comprise a small proportion (20%) of all endometrial cancer cases, much larger cohort sizes are needed to detect any genuine signals for non-endometrioid tumours. Most of the evaluated studies looked at either overall/mixed endometrial cancer subtypes or endometrioid histology, and those that looked at variant associations with non-endometrioid histology were unlikely to have enough power to detect any signal with statistical significance. This is particularly concerning because non-endometrioid subtypes are biologically aggressive tumours with a much poorer prognosis that contribute disproportionately to mortality from endometrial cancer. It is particularly important that attempts to improve early detection and prevention of endometrial cancer focus primarily on improving outcomes from these subtypes. It is also worth noting that, despite the current shift towards a molecular classification of endometrial cancer, most studies used the over-arching classical Bokhman's classification system, type I versus type II, or no histological classification system at all. Therefore, it is

important to create and follow a standardized and comprehensive classification system for reporting tumour subtypes for future studies.

This study compiled and presented available information for an extensively studied, yet unproven in large datasets, SNP309 variant in *MDM2*. Currently, there is no convincing evidence for an association between this variant and endometrial cancer risk. Additionally, of all the studies, only one accounted for the opposing effect of a nearby variant SNP285 in their analyses. Thus, we conclude that until confirmed by a sufficiently large GWAS, this variant should not be considered significant in influencing the risk of endometrial cancer and therefore not included in a PRS. This is also true for the majority of the SNPs reported in candidate-gene studies, as the numbers fall far short of being able to detect genuine signals.

This systematic review presents the most up-to-date evidence for endometrial cancer susceptibility variants, emphasizing the need for further large-scale studies to identify more variants of importance, and validation of these associations. Until data from larger and more diverse cohorts are available, the top twenty-four SNPs presented here are the most robust common genetic variants that affect endometrial cancer risk. The multiplicative effects of these SNPs could be used in a PRS to allow personalised risk prediction models to be developed for targeted screening and prevention interventions for women at greatest risk of endometrial cancer.

2.7 References

Ashton, K. A., Proietto, A., Otton, G., *et al.* (2009). Polymorphisms in TP53 and MDM2 combined are associated with high grade endometrial cancer. *Gynecologic Oncology*, 113(1):109–114.

Banno, K., Yanokura, M., Iida, M., *et al.* (2014). Carcinogenic mechanisms of endometrial cancer: Involvement of genetics and epigenetics. *Journal of Obstetrics and Gynaecology Research*, 40(8):1957–1967.

Chen, M. M., Crous-Bou, M., Setiawan, V. W., *et al.* (2014). Exome-Wide Association Study of Endometrial Cancer in a Multiethnic Population. *Plos One*, 9(5):9.

Chen, M. M., O'Mara, T. A., Thompson, D. J., *et al.* (2016). GWAS meta-analysis of 16 852 women identifies new susceptibility locus for endometrial cancer. *Human Molecular Genetics*, 25(12):2612–2620.

Cheng, T. H. T., Thompson, D., Painter, J., *et al.* (2015). Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Scientific Reports*, 5:12.

Cheng, T. H. T., Thompson, D. J., O'Mara, T. A., *et al.* (2016). Five endometrial cancer risk loci identified through genome-wide association analysis. *Nature Genetics*, 48(6):667–674.

Clarke, M. A., Devesa, S. S., Harvey, S. V., and Wentzensen, N. (2019). Hysterectomy-Corrected Uterine Corpus Cancer Incidence Trends and Differences in Relative Survival Reveal Racial Disparities and Rising Rates of Nonendometrioid Cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, pages JCO1900151–JCO1900151.

Constantinou, P. and Tischkowitz, M. (2017). Genetics of gynaecological cancers. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 42:114–124.

Gansmo, L. B., Bjornslett, M., Halle, M. K., *et al.* (2017). MDM2 promoter polymorphism del1518 (rs3730485) and its impact on endometrial and ovarian cancer risk. *Bmc Cancer*, 17:6.

Geng, Y. H., Wang, Z. F., Jia, Y. M., *et al.* (2018). Genetic polymorphisms in CDH1 are associated with endometrial carcinoma susceptibility among Chinese Han women. *Oncology Letters*, 16(5):6868–6878.

Getz, G., Gabriel, S. B., Cibulskis, K., *et al.* (2013). Integrated genomic characterization of endometrial carcinoma. *Nature*, 497(7447):67–73.

Hamdi, Y., Soucy, P., Adoue, V., *et al.* (2016). Association of breast cancer risk with genetic variants showing differential allelic expression: Identification of a novel breast cancer susceptibility locus at 4q21. *Oncotarget*, 7(49):80140–80163.

Kitson, S. J., Evans, D. G., and Crosbie, E. J. (2017). Identifying High-Risk Women for Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model. *Cancer Prevention Research*, 10(1):1–13.

Knappskog, S., Gansmo, L. B., Dibirova, K., *et al.* (2014). Population distribution and ancestry of the cancer protective MDM2 SNP285 (rs117039649). *Oncotarget*, 5(18):8223–8234.

Knappskog, S., Trovik, J., Marcickiewicz, J., *et al.* (2012). *SNP285C modulates oestrogen receptor/Sp1 binding to the MDM2 promoter and reduces the risk of endometrial but not prostatic cancer*, volume 48, pages 1988–96. 2011 Elsevier Ltd, England.

Lacey, J. V., Yang, H. N., Gaudet, M. M., *et al.* (2011). Endometrial cancer and genetic variation in PTEN, PIK3CA, AKT1, MLH1, and MSH2 within a population-based case-control study. *Gynecologic Oncology*, 120(2):167–173.

Li, Y., Zhao, H. J., Sun, L., *et al.* (2011). MDM2 SNP309 is associated with endometrial cancer susceptibility: a meta-analysis. *Human Cell*, 24(2):57–64.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., *et al.* (2000). Environmental and heritable factors in the causation of cancer - Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343(2):78–85.

Lu, Y., Ek, W. E., Whiteman, D., *et al.* (2014). Most common 'sporadic' cancers have a significant germline genetic component. *Human Molecular Genetics*, 23(22):6112–6118.

Lundin, E., Wirgin, I., Lukanova, A., *et al.* (2012). Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiology*, 36(5):445–452.

Mavaddat, N., Michailidou, K., Dennis, J., *et al.* (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, 104(1):21–34.

Moher, D., Liberati, A., Tetzlaff, J., *et al.* (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7):1–6.

Morice, P., Leary, A., Creutzberg, C., *et al.* (2016). Endometrial cancer. *Lancet*, 387(10023):1094–1108.

Nunobiki, O., Ueda, M., Yamamoto, M., *et al.* (2009). Polymorphisms of p53 codon 72 and MDM2 promoter 309 and the risk of endometrial cancer. *Human Cell*, 22(4):101–106.

O'Mara, T. A., Glubb, D. M., Amant, F., *et al.* (2018). Identification of nine new susceptibility loci for endometrial cancer. *Nature Communications*, 9.

O'Mara, T. A., Glubb, D. M., Kho, P. F., *et al.* (2019a). Genome-wide association studies of endometrial cancer: Latest developments and future directions. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*.

O'Mara, T. A., Glubb, D. M., Painter, J. N., *et al.* (2015). Comprehensive genetic assessment

of the ESR1 locus identifies a risk region for endometrial cancer. *Endocrine-Related Cancer*, 22(5):851–861.

O'Mara, T. A., Spurdle, A. B., and Glubb, D. M. (2019b). Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. *Cancers (Basel)*, 11(10).

Painter, J. N., Kaufmann, S., O'Mara, T. A., *et al.* (2016). A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates AKT1 through YY1 Binding. *American Journal of Human Genetics*, 98(6):1159–1169.

Peng, Q. L., Mo, C. J., Qin, A. P., *et al.* (2013). MDM2 SNP309 polymorphism contributes to endometrial cancer susceptibility: evidence from a meta-analysis. *Journal of Experimental & Clinical Cancer Research*, 32:9.

PROSPERO (accessed: Oct 2020). Low-penetrance common predisposition variants in endometrial cancer risk: a systematic review. `https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018091907`.

Raglan, O., Kalliala, I., Markozannes, G., *et al.* (2019). Risk factors for endometrial cancer: An umbrella review of the literature. *Int J Cancer*, 145(7):1719–1730.

Ryan, N. A. J., Glaire, M. A., Blake, D., *et al.* (2019). The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genetics in Medicine*, 21(10):2167–2180.

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J Clin*, 69(1):7–34.

Spencer, C. C. A., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *Plos Genetics*, 5(5):13.

Spurdle, A. B., Thompson, D. J., Ahmed, S., *et al.* (2011). Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nature Genetics*, 43(5):451–454.

Stelloo, E., Bosse, T., Nout, R. A., *et al.* (2015). Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Modern Pathology*, 28(6):836–844.

Sundar, S., Balega, J., Crosbie, E., *et al.* (2017). BGCS uterine cancer guidelines: Recommendations for practice. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 213:71–97.

Tao, M. H., Cai, Q. Y., Zhang, Z. F., *et al.* (2007). Polymorphisms in the CYP19A1 (Aromatase) gene and endometrial cancer risk in Chinese women. *Cancer Epidemiology Biomarkers & Prevention*, 16(5):943–949.

Terry, K., McGrath, M., Lee, I. M., *et al.* (2008). MDM2 SNP309 is associated with endometrial cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 17(4):983–986.

Tzur, T., Kessous, R., and Weintraub, A. Y. (2017). Current strategies in the diagnosis of endometrial cancer. *Archives of Gynecology and Obstetrics*, 296(1):5–14.

Ueda, M., Yamamoto, M., Nunobiki, O., *et al.* (2009). Murine double-minute 2 homolog single nucleotide polymorphism 309 and the risk of gynecologic cancer. *Human Cell*, 22(2):49–54.

VonVille, H. (2015). Excel workbooks for systematic reviews. `https://showcase.dropbox.com/s/Excel-Workbooks-for-Systematic-Reviews-User-Guides-Kf4pYVTvFqSJdZR2Spzlu`.

Walsh, C. S., Miller, C. W., Karlan, B. Y., and Koeffler, H. P. (2007). Association between a functional single nucleotide polyrnorphism in the MDM2 gene and sporadic endometrial cancer risk. *Gynecologic Oncology*, 104(3):660–664.

Wan, Y., Wu, W., Yin, Z. H., *et al.* (2011). MDM2 SNP309, gene-gene interaction, and tumor susceptibility: an updated meta-analysis. *Bmc Cancer*, 11:9.

Wan, Y. L., Beverley-Stevenson, R., Carlisle, D., *et al.* (2016). Working together to shape the endometrial cancer research agenda: The top ten unanswered research questions. *Gynecologic Oncology*, 143(2):287–293.

Wang, L. H., Wang, X., Xu, W. T., and Hu, Y. L. (2014). MDM2 rs2279744 polymorphism and endometrial cancer: a meta-analysis. *Tumor Biology*, 35(4):3167–3170.

Win, A. K., Reece, J. C., and Ryan, S. (2015). Family History and Risk of Endometrial Cancer A Systematic Review and Meta-analysis. *Obstetrics and Gynecology*, 125(1):89–98.

Xue, Z. W., Zhu, X. L., and Teng, Y. C. (2016). Relationship Between Murine Double Minute 2 (MDM2) T309G Polymorphism and Endometrial Cancer Risk: A Meta-Analysis. *Medical Science Monitor*, 22:3186–3190.

Yoneda, T., Kuboyama, A., Kato, K., *et al.* (2013). Association of MDM2 SNP309 and TP53 Arg72Pro polymorphisms with risk of endometrial cancer. *Oncology Reports*, 30(1):25–34.

Zajac, A., Stachowiak, G., Pertynski, T., *et al.* (2012). Association between MDM2 SNP309 polymorphism and endometrial cancer risk in Polish women. *Polish Journal of Pathology*, 63(4):278–283.

Zhang, J. X., Zhang, Y., and Zhang, Z. Y. (2018). Association of rs2279744 and rs117039649 promoter polymorphism with the risk of gynecological cancer A meta-analysis of case-control studies. *Medicine*, 97(2):9.

Zhang, Y. D., Hurson, A. N., Zhang, H. Y., *et al.* (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature Communications*, 11(1):13.

Zhao, Y., Yang, X. E., Hao, X. J., *et al.* (2014). Common variant on MDM2 contributes to

endometrial cancer susceptibility: evidence based on 7 studies. *Tumor Biology*, 35(8):7555–7560.

Zou, X., Zhang, Y., Zhang, L., *et al.* (2018). Association between MDM2 SNP309 and endometrial cancer risk: A PRISMA-compliant meta-analysis. *Medicine*, 97(49):e13273–e13273.

3 Identification and validation of single nucleotide polymorphisms of genome-wide and suggestive significance associated with endometrial cancer

This chapter forms part of a research article that is being prepared for submission in a peer-reviewed journal. In brief, this chapter explains the generation of genotype data and a GWAS conducted on an independent set of clinically well-annotated cases and controls (Manchester study) to validate previously identified SNPs associated with endometrial cancer and identify further loci suggestive of an association. The results generated here were used to construct and validate some of the PRSs described in **Chapter 4**.

3.1 Contributorship Statement

I performed the laboratory work (DNA quantification, purification and normalisation, and some of the DNA extractions), genotype analyses, QC and GWAS, data extraction, statistical analysis, and drafted the manuscript. D.J.T and J.D supported data extraction, genotype analyses and GWAS. A.L, K.M, T.OM and E.J.C supported data extraction. Most of the DNA extractions were performed previously by either Manchester Genomic Diagnostic Laboratory (MGDL) or Hologic Ltd.

3.2 Abstract

Background Endometrial cancer is the most commonly diagnosed gynaecological malignancy in developed countries. It is estimated that nearly 30% of endometrial cancer familial risk is explained by common low-risk susceptibility variants. To date, multiple single nucleotide polymorphisms (SNPs) have been reported to influence the risk of endometrial cancer, however, most were identified by a single GWAS and many studies used overlapping datasets. Therefore, we aimed to validate the known SNPs and identify further regions in an independent cohort.

Methods We identified 612 endometrial cancer cases and 1,202 endometrial- and breast cancer-free controls from the North West of England. The samples were genotyped using the OncoArray chip and processed according to the OncoArray Consortium guidelines. Untyped genotypes were imputed using a two-step approach comprising of phasing and imputation, and GWAS was conducted by performing logistic regression for case-control status with or without adjustment for BMI and subtype-specific associations.

Results Six loci were identified at suggestive significance (P < 1E-5); lead SNPs rs12727038 (1q42.2), rs144065942 (2q14.2), rs9854980 (3p24.1), rs6580584 (5q32), rs74532550 (12p11.23) and rs111282723 (18q12.3). Of the 24 systematic review SNPs, 18 had same direction of OR, and seven were significant at $P < 0.05$. Importantly, the first identified endometrial cancer GWAS hit at *HNF1B* was also significant here. Twenty-nine of the 48 SNPs significant at suggestive threshold identified from the ECAC (Endometrial Cancer Association Consortium) GWAS had the same direction of association in our dataset, and three were significant at $P < 0.05$.

Conclusions Here we report independent validation of 72 SNPs identified through our recent systematic review or the latest ECAC GWAS. Moreover, we report six SNPs of suggestive significance identified in our GWAS. This solidifies the importance of SNPs, particularly of the previously reported SNPs, in endometrial cancer predisposition.

3.3 Introduction

Endometrial cancer is the most commonly diagnosed gynaecological malignancy in developed countries and is estimated to account for 7% of all new cancer diagnoses in women (Sundar *et al*., 2017; Siegel *et al*., 2020). In the last 50 years, cancer survival has improved for most cancers except for endometrial cancer. Traditionally, endometrial cancer has been classified into two subtypes based on morphology; endometrioid and non-endometrioid. The former is closely associated with obesity and is considered oestrogen-dependent. While the majority of endometrial tumours are of endometrioid histology (∼80%), non-endometrioid tumours have far worse prognosis and contribute to endometrial cancer-associated mortality more than their endometrioid counterparts (Clarke *et al*., 2019; Siegel *et al*., 2020).

Although hysterectomy is curative for early stage and/or low-grade disease, this is often not the case for late stage, metastatic or recurrent tumours, all of which are more commonly associated with the non-endometrioid subtype. The overall five-year survival rate currently stands at 80%, but this drops dramatically to 16% in women diagnosed with advanced stage disease (Morice *et al*., 2016; Siegel *et al*., 2020). Moreover, there is a marked discrepancy between survival rates in women of different ethnic origins. Among all disease stages, women of African heritage have lower survival rates compared to those of European heritage (Siegel *et al*., 2020).

Family history of endometrial cancer is well-established to increase the risk of disease by at least two-fold, but twin studies have shown variable heritability (Lichtenstein *et al*., 2000; Lu *et al*., 2014; Mucci *et al*., 2016; Johnatty *et al*., 2017). High-risk, rare variants such as those in the Lynch and Cowden syndrome-related genes (*PTEN* and the mismatch repair genes, respectively) only account for up to 5% of endometrial cancer cases at population level (Spurdle *et al*., 2017). Due to the rarity of these pathogenic variants, their contribution to the overall endometrial cancer prevalence is relatively small. O'Mara *et al.*

estimated that nearly 30% of familial relative risk of endometrial cancer may be explained by common low-risk susceptibility variants (O'Mara *et al*., 2018). To date, less than 20 genome-wide significant SNPs have been reported to influence endometrial cancer risk (O'Mara *et al*., 2018), and recently we have also compiled a panel of 24 most robust SNPs including suggestively significant SNPs (see **Chapter 2, p 76**) (Bafligil *et al*., 2020). The majority of these SNPs were identified by ECAC, which is the largest consortium investigating common risk variants in endometrial cancer predisposition and currently benefits from approximately 12,900 cases and 109,000 controls (O'Mara *et al*., 2018).

In contrast, over 100 SNPs have been reported to influence the risk of breast cancer, while 313 SNPs have recently been used to generate a polygenic risk score (PRS) to aid risk prediction efforts (Michailidou *et al*., 2017; Mavaddat *et al*., 2019; Rivandi *et al*., 2018). However, this is partly due to a far larger dataset comprised of ∼150,000 cases and equal number of controls for breast cancer through Breast Cancer Association Consortium (BCAC), which naturally holds greater power to detect an increased number of genome-wide significant risk loci. In the case of endometrial cancer, Zhang *et al.* predicted that over a thousand independent SNPs play a role in endometrial cancer predisposition and cohorts of similar magnitude to that of BCAC are required to be able to detect these signals (Zhang *et al*., 2020). Moreover, in our systematic review, we noted that the datasets used in GWAS for endometrial cancer often overlapped with others, which may hinder their ability to identify novel variants but also for meta-analyses to be conducted on these variants. Thus, it is essential to increase diversity and study size of the cohorts for the future GWAS efforts looking into endometrial cancer predisposition, and importantly, independently validate the reported SNPs.

The aim of this study was to conduct an independent GWAS using a clinically well-phenotyped, locally matched cohort of endometrial cancer cases and cancer-free controls from the North West of England. We sought to validate the 24 systematic review SNPs as well as SNPs of suggestive significance identified through ECAC. Moreover, we inves-

tigated SNPs of suggestive significance identified here, examined various comparisons including BMI adjustment, and conducted subtype-specific analyses.

3.4 Materials and Methods

3.4.1 Study Population

The Manchester study cohort comprising of geographically matched endometrial cancer cases and cancer-free controls was collated from patients recruited to multiple pre-existing, ethically approved studies across the faculty as detailed in **Table 3.1**.

Table 3.1: *Details of studies employed in the Manchester study.*

Study Name	REC Reference/UKCRN-ID	Recruitment Period
Proportion of endometrial tumours associated with Lynch Syndrome (PETALS)	15/NW/0733	2015-2018
The impact of obesity and weight loss on the endometrium: a prospective cohort study (RB)	12/NW/0050	2012-2016
Metformin in non-diabetic women with endometrial cancer (METFORMIN)	11/NW/0442	2012-2014
Pre-surgical metformin for women with endometrial cancer: a randomised placebo controlled trial (PREMIUM)	14/NW/1236	2015-2017
Mirena for treatment of endometrial neoplastic abnormalities (MIRENA)	14/NW/0056	2014-ongoing
Manchester Biomedical Research Centre Biobank (BRC-Biobank)	-	-
Developing tests for endometrial cancer detection (DETECT)	16/NW/0660	2016-ongoing
Predicting Risk of Breast Cancer at Screening (PROCAS)	09/H1008/81	2009-2015
Family History Risk Study (FH-Risk)	8611*	2010-2014
Clinical Genetics service	-	-
EVE Study	-	-
MEC Study	-	-

* UKCRN-ID
REC, Research Ethics Committee; UKCRN-ID, UK Clinical Research Network identification number

612 women from Greater Manchester who received a histological diagnosis of endometrial cancer were selected from participants consented to PETALS (n=214), RB (n=20), METFORMIN (n=23), PREMIUM (n=76), MIRENA (n=44), DETECT (n=54), PROCAS (n=20), EVE (n=16), and MEC (n=5) studies. A further 89 blood or DNA samples from endometrial cancer cases were obtained from the BRC-Biobank donated by eligible women. Finally, 51 DNA samples were obtained from the Clinical Genetics service and the FH-Risk study. All biopsy and hysterectomy specimens were reviewed by at least two specialist gynaeco-

logical pathologists as part of their routine clinical care at St Mary's Hospital, according to FIGO (2009) staging criteria.

Endometrial and breast cancer-free controls (n=1,202) without a history of hysterectomy were chosen from eligible women participating in the PROCAS study, which was a prospective case-control study recruiting women aged 46 to 73 for breast cancer risk prediction screening between 2009 and 2015. These have been genotyped previously as part of a BCAC study. Moreover, 20 women who were diagnosed with endometrial cancer after participating in the PROCAS study were included as cases in this study, three of which were already genotyped using the OncoArray chip.

All studies were approved by relevant ethics boards and written informed consents were obtained from all participants. Details of PETALS and PROCAS studies were published previously (Evans *et al.*, 2019; Ryan *et al.*, 2020).

3.4.2 Specimen Collection

Peripheral blood, up to 10 mL, from participants who gave written informed consent was collected into anticoagulant (K_2EDTA) containing vacutainers by the hospital clinical staff. The blood samples were then processed by the hospital Biobank staff. Of each blood specimen received, 0.5 mL was immediately frozen (whole blood) while the rest was spun at 1,500 x *g* to separate the plasma and blood cells. The cellular layer containing white blood cells (WBCs), approximately 0.5-1 mL, was collected and frozen. These are referred to as *buffycoat* samples. All blood samples were stored in -80°C freezers until DNA extraction.

3.4.3 DNA Extraction

For this study, a total of 609 genomic DNA samples were used for genotyping the cases (excluding the three previously genotyped cases). Of these, 442 (72.6%) were extracted from either whole blood or buffycoats by Hologic Ltd (Manchester, UK) using the Nu-

cleon extraction chemistry method (Cat no. SL8502, Gen-Probe Life Sciences Ltd). Another 133 (21.8%) DNA were obtained from the MGDL at the Manchester Centre for Genomic Medicine which were previously extracted for other studies and 20 of these were extracted for the PROCAS study using Oragene kit (DNA Genotek Inc., Ottawa, ON, Canada). The remaining 34 (5.6%) samples were extracted in house by C.B using a Gentra Puregene Blood Kit (Cat no. 158389, Qiagen). Details of the relevant DNA extraction methods are provided below.

3.4.3.1 Nucleon Chemistry Method

Compared to other methods, the Nucleon extraction method, used by Hologic Ltd, yields higher quality DNA which is purer and less fragmented. The resulting DNA often has $A_{260/280}$ ratios of 1.80-1.90, which is a widely accepted ideal purity rate for DNA. Moreover, this kit yields 35-40 μg DNA per mL of whole blood. Although, final DNA concentrations were not measured or normalised by Hologic Ltd, a representative number of samples were quantified using Picogreen, which is a highly accurate method for DNA quantification. Due to time restrictions and favourable yield using this method, this route was preferred for the majority of the samples.

3.4.3.2 Gentra Puregene Method

Gentra Puregene Blood Kit (Qiagen) typically yields a similar $A_{260/280}$ ratio between 1.7 and 1.9 and a similar final DNA yield of 35 μg per mL of blood to that of Nucleon kits. Additional genomic DNA from blood samples of recent recruits were extracted using this method according to manufacturer's protocol detailed below.

300 μL of blood (whole blood or buffycoat) was thawed quickly at 37°C and mixed with 900 μL of red blood cell (RBC) lysis solution. After a one-minute incubation at room temperature, the mixture was centrifuged at 13,000-16,000 x g for 20 seconds to pellet the WBCs. Then the supernatant was decanted and the remaining pellet with residual supernatant was vortexed vigorously to disperse the cells evenly. 300 μL cell lysis solution was

added to lyse the cells and vortexed vigorously for 10 seconds. 100 μL protein precipitation solution was added to the mixture and vigorously vortexed at high speed for 20 seconds. Following centrifugation at 13,000-16,000 x g for one minute, the supernatant was added to a clean tube containing 300 μL isopropanol. The resulting liquid was mixed by inverting 50 times until the DNA was precipitated and became visible.

DNA was then pelleted by centrifuging at 13,000-16,000 x g for one minute and then the supernatant was discarded. Then, 300 μL 70% ethanol was added to the pellet and inverted several times to wash, followed by one minute centrifugation at 13,000-16,000 x g. The supernatant was then discarded carefully and the pellet was left to air dry until the residual ethanol had evaporated. 100 μL hydration solution was added to the pellet and vortexed at medium speed for 5 seconds. As a final step, DNA was dissolved by incubating in a water bath pre-heated to 65°C for 1 hour and then incubated at room temperature overnight.

3.4.3.3 Oragene Method

Saliva specimens (approximately 2 mL) from the PROCAS participants were collected using Oragene saliva lysate tubes (DNA Genotek Inc., Ottawa, ON, Canada). Reagents within the tubes prevents degradation of the DNA and inactivate nucleases and bacteria, thus stabilising the samples. This collection method yields a higher total DNA per sample compared to other collections methods such as the buccal swabs. Genomic DNAs from the stabilised specimens were extracted using the Nucleon chemistry method by Hologic Ltd as explained earlier.

3.4.4 DNA Preparation

3.4.4.1 DNA Quantification

The DNA samples were quantified by NanoDrop Spectrophotometer (ThermoFisher). 2 μL of the elution buffer was used to blank the instrument. Then, 2 μL of each DNA sample was loaded onto the instrument for measurement.

10 μL of genomic DNA at a concentration of 100 ng/μL, if quantification was done using an absorbance method such as NanoDrop, was required by the genotyping service (Strangeways Research Laboratory, University of Cambridge). However, concentrations between 70 and 150 ng/μL were tolerated.

3.4.4.2 DNA Purification and Normalisation

Based on NanoDrop quantitation, any DNA sample that showed indication of a contamination, as measured by $A_{260/280}$ and $A_{260/230}$ ratios, and/or had a DNA concentration out of the 70 to 150 ng/μL range were treated with Zymo Genomic DNA Clean and Concentrator-10 purification columns (Cat no. D4011, Zymo Research), according to the manufacturer's instructions. Accordingly, the DNA was mixed with DNA binding buffer, two times its initial volume and vortexed briefly. After a brief centrifugation step to collect the liquid, the mixture was transferred to a binding column and spun at maximum speed for 30 seconds. Twice, 200 μL wash buffer (ethanol added) was added and spun at maximum speed for 30 seconds, each time. Flow-through was discarded carefully and columns were spun again (maximum speed for 30 seconds) to remove any residual wash buffer and ethanol. The column was transferred to a collection tube and 20 μL of 1xTE buffer, pre-heated to 65°C, was added onto the membrane. After 1 minute incubation at room temperature, the column was spun at maximum speed for one minute. The flow through, containing purified and concentrated DNA was then quantified again using NanoDrop. All consumables used for this process were free of endotoxin, RNase and DNase.

All DNA was plated into full-skirted ABGene 96-well SuperPlates (Cat no. AB-2800, Thermo Scientific) and stored at -20°C until genotyping.

3.4.5 Genotyping

High throughput genotyping of endometrial cancer cases was performed at Strangeways Research Laboratory, University of Cambridge using Infinium OncoArray-534K BeadChip

(Illumina). OncoArray is a custom chip designed by the OncoArray Consortium to investigate cancer predisposition and risk by looking into common SNPs related to common cancers including colorectal, breast and ovarian cancers but is also applicable to endometrial cancer due to wide SNP coverage (Amos *et al.*, 2017). Controls and three cases were previously genotyped at Cambridge with the same OncoArray chip as part of BCAC.

3.4.5.1 Infinium High Throughput Screening Protocol

Briefly, the DNA is denatured and amplified, which is then followed by fragmentation. The fragmented DNA is then precipitated and resuspended in hybridisation buffer. The DNA fragments will bind to silica beads coated with multiple oligonucleotide probes specific to each locus investigated. For loci where the allele change is *ambiguous* such as G to C or A to T, the probes bind to the DNA one base before the SNP locus of interest (n$\pm$1) (see illustration (**3.1**) below). For loci where the two alleles are unambiguous, the probes bind the DNA at the exact base (n), but at two separate known locations for each allele, allowing for distinguishing between the alleles.

$$5'..AGTT\underbrace{G}_{\text{n-1}}\overbrace{[G/T]}^{\text{n}}\underbrace{A}_{\text{n+1}}TCGA..3' \tag{3.1}$$

The green or red labelled nucleotides bind to the SNP locus depending on which variant is present and are then extended by DNA polymerase. The chips are scanned by HiScan (Illumina) or iScan (Illumina) readers which capture the colour and intensity of the specific signals upon excitation by lasers.

3.5 Data Analysis

Data analysis was conducted by the author, including time spent during an NIHR funded 3-month placement at Strangeways Research Laboratory, University of Cambridge, where the genotyping was done.

3.5.1 Study Power Calculation

A statistical power of at least 0.8 (80%) is expected for a study to be considered robust enough to detect genuine signals (Gupta *et al.*, 2016). It is generally accepted that a sufficiently powered GWAS requires approximately 1,200 unrelated cases when 500,000 SNPs are being analysed for an OR of 2 (Hong and Park, 2012). Indeed, the first endometrial cancer GWAS analysed 1,265 cases and reported a single locus reaching genome wide significance threshold (Spurdle *et al.*, 2011).

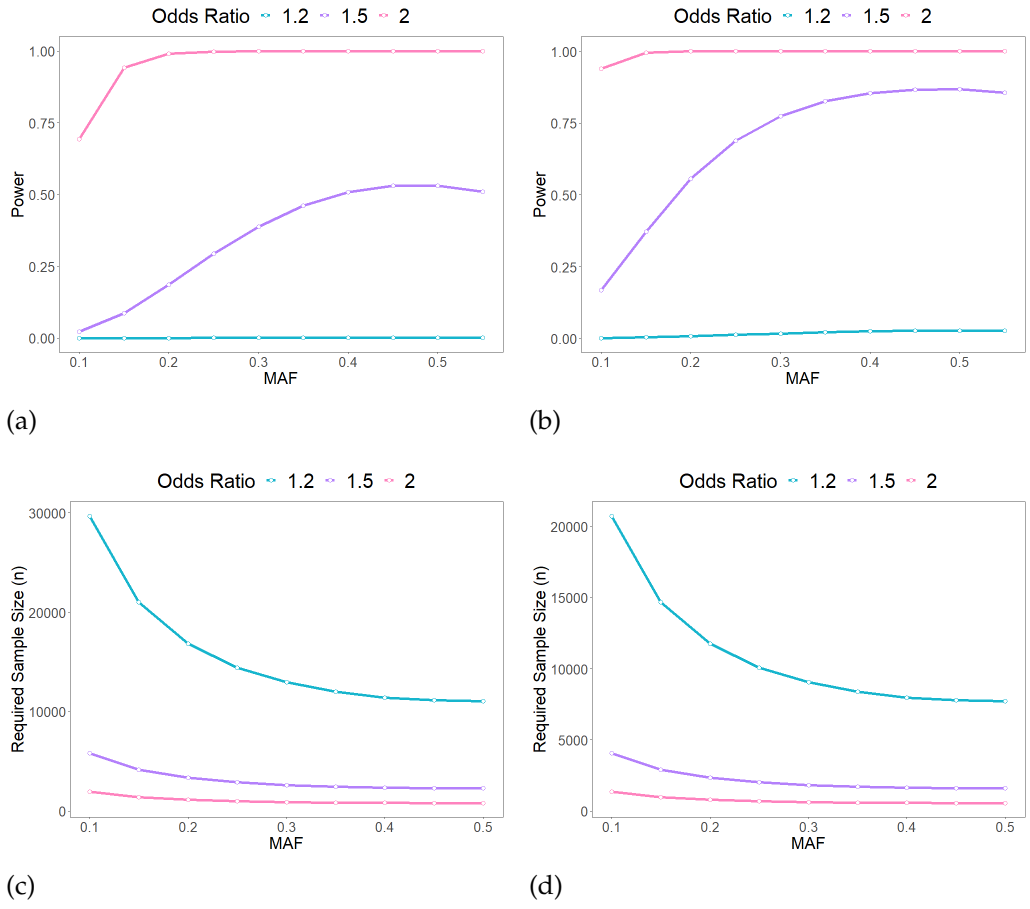


Figure 3.1: *Estimated study power (a,b) and sample size required (c,d) at alpha 5E-8 (a,c) and 1E-5 (b,d) when OR is assumed at 1.2, 1.5 or 2 and at MAF distributions ranging from 0.1 to 0.5.*

These estimations were calculated under the assumption of an additive model according to our sample size of 1,757 with a case rate of ∼0.32. The calculations were performed using genpwr package in R (Moore and Jacobson, 2021).

Statistical power of this one-stage GWAS study was computed for a series of MAFs (0.1-0.5) under the assumption of OR of 1.2, 1.5 and 2, a study power of 0.8 and *P* values of 5E-8 and 1E-5. According to the estimations calculated here, for discovery purposes, this study does not hold sufficient power at *P* value threshold of 5E-8 for most of the tested OR and MAF ranges. However, its power is estimated to be satisfactory at the suggestive *P* value threshold of 1E-5 and is expected to prove sufficient for nominal significance value of 0.05 for validation purposes. Therefore, keeping close attention to the power limitations of this study, we proceeded to interpret the results with caution where necessary.

3.5.2 Quality Control

Manifest of the chip (.bpm file) and raw data (green and red .idat files) were received from the genotyping lab (Strangeways Research Laboratory). These were uploaded onto GenomeStudio 2.0 (Illumina) where the sample sheet was matched with the .idat files using the respective chip barcodes and positions per each sample. The genotype clustering algorithm, GenTrain 3.0 by Illumina, was used to cluster calls into three clusters. Samples and SNPs were checked for call-rates and gender assignments. Although these were investigated for quality control (QC) purposes, no SNP or sample was excluded at this point. The genotype calls were then exported in PLINK format (.ped and .map), which is a format suited to perform QC using GenABEL package in R (GenABEL, 2013; R Core Team, 2020). Genotype analyses were performed in R version 3.6.0, using RStudio interface, unless stated otherwise (R Core Team, 2020; RStudio Team, 2020).

Data were loaded on R and the total sample number was counted as 1,814 (n_{cases}=612, $n_{controls}$ =1,202). There were 7 internal control samples attached for QC purposes. The total number of genotyped SNPs was 533,631. A list of SNPs was provided by the OncoArray Consortium guidelines due to having consistently low call-rates, low minor allele frequency (MAF) or poor genotyping quality, which were excluded. 19 non-endometrial primary cancers were identified by latest histopathology reports and were labelled for exclusion from the analyses.

3.5.2.1 Call-rate & Heterozygosity

For the remaining 494,763 SNPs, call-rate per each SNP was calculated by dividing the sum of number of samples with non-missing genotypes by total sample number. There were 578 SNPs with a call-rate below 95%, including those in Y chromosome. Call-rate per sample was also calculated and there were four samples with a total call-rate below 95%, excluding Y chromosome. These four samples and 295 SNPs (not inclusive of Y chromosome SNPs) were excluded. Heterozygosity, which is the fraction of heterozygous genotype calls of a sample, per plate was assessed. These showed no immediate evidence of a poor genotyping quality.

3.5.2.2 Sex Chromosome Checks

To identify individuals deviating from genetically female sex, a list of SNPs (n=92) in chromosome X with autosomal clusters as reference was merged with the data. Sample-wise summaries for X chromosome SNPs (n=13,962) were calculated after excluding any SNP with call-rate below 95% and a MAF smaller than 1%, leaving 10,384 SNPs. This would enable identification of genetically XO (Turner's Syndrome) individuals. After identifying samples with an X chromosome heterozygosity smaller than 0.2, one sample was flagged. However, the heterozygosity was not deviant enough to warrant exclusion. Therefore, this sample was kept with caution. A list of 300 Y chromosome SNPs was also merged with the data to evaluate Y chromosome SNPs which would indicate genetically male (XY) or XXY (Klinefelter Syndrome) individuals.

Similarly, after restricting the data to include those with X chromosome heterozygosity lower than 0.2 and calculating call-rate for the Y chromosome SNPs, the same sample was flagged, but was not excluded. This was because the Y chromosome call-rate and homozygosity for this individual was not high enough, which would be expected to be true for genetically male individuals.

3.5.2.3 Population Structure & Kinship

To identify individuals of major continental ancestry, a list of 33,661 uncorrelated SNPs and HapMap data from unrelated individuals were merged with the data. Identity-by-state (IBS) matrix was then calculated, which evaluates the relatedness between individuals based on shared alleles at a given locus. An ancestry map (**Figure 3.2**) was plotted with distinct continental ancestries (based on known HapMap samples) via multidimensional scaling through principal component analysis (PCA). The ancestry cut-off points were defined as ± 0.04 for the principal components (PCs) for this study, similar to large consortia. 21 individuals that clustered within or closer to Asian or African ancestries (n=21) were identified. Of those where self-reported ancestry data were available, all but one had matching genetic ancestry. The samples outside of the European genetic cluster were excluded due to lack of sufficient numbers for these ancestries to be able to impute the untyped genotypes separately.

Overall heterozygosity was checked again after excluding non-European individuals and accordingly, those that deviated beyond 5 standard deviation (SD) of mean heterozygosity were identified. Except for one sample, the remainder were borderline deviant therefore, only one sample was excluded at this stage. The kinship matrix, IBS, created earlier was loaded again without the HapMap samples, and used to identify pairs of individuals that were genetically identical e.g. monozygotic twins, unexpected duplicates or first-degree relatives. The threshold for first-degree relatedness was 0.85. In summary fifteen pairs of duplicates and one pair of first-degree relatives (one of which was excluded at an earlier stage) were identified within the sample set. One of each pair which had a lower call-rate was excluded.

Mitochondrial and Y chromosome SNPs were excluded. Then, deviation from Hardy-Weinberg equilibrium (HWE) was assessed both in the total dataset but also within the controls and cases, separately. HWE assesses the genotype distribution of a given SNP where in the absence of evolutionary factors would remain constant (see **Section 3.5.5,**

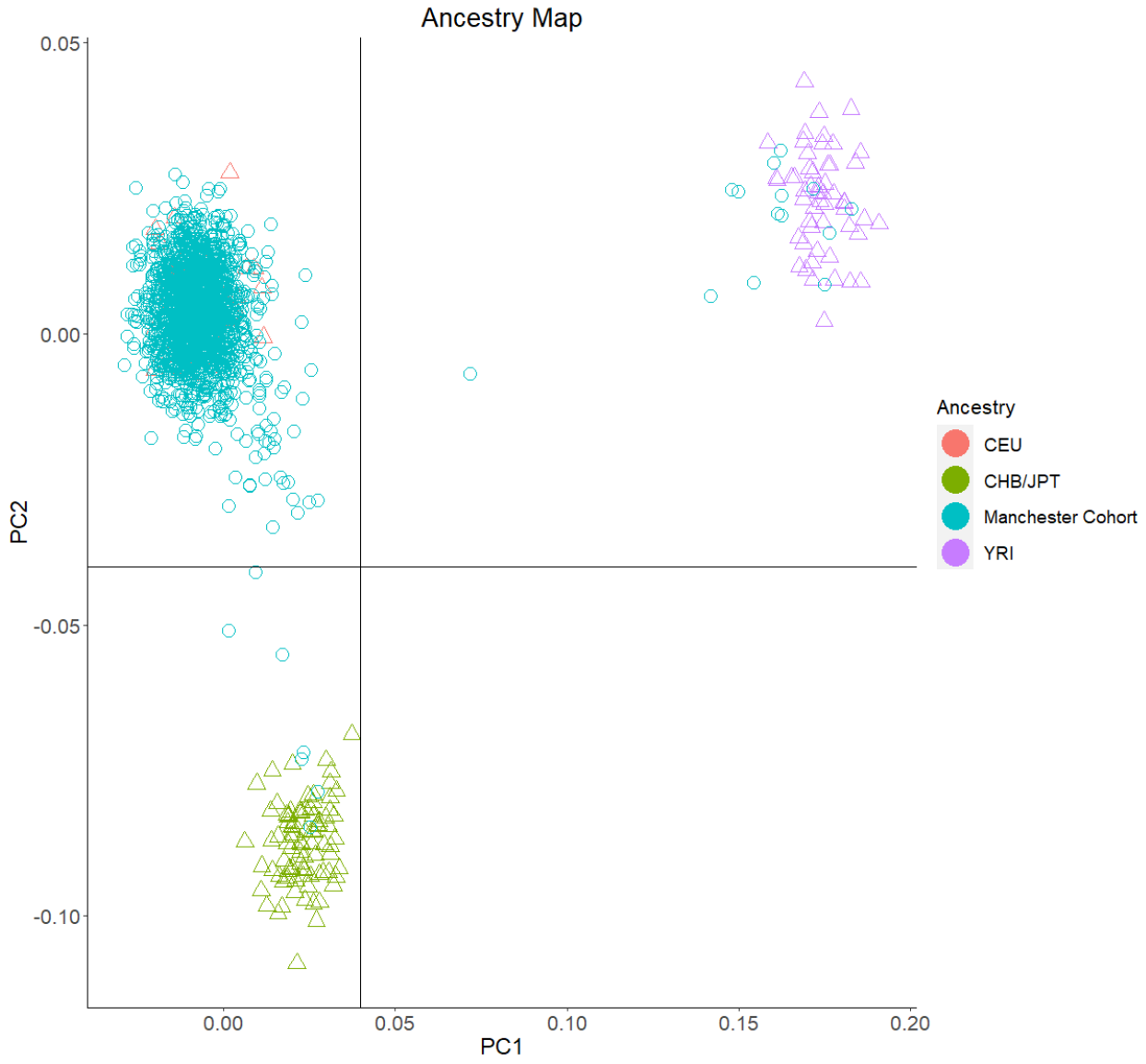


Figure 3.2: *Ancestral fractions of the Manchester samples were estimated using IBS matrix and plotted alongside known major continental ancestries.*

Manchester cohort (△) are represented here in comparison to the HapMap reference (○) for known European (CEU), Asian (CHB/JPT) and African (YRI) descent individuals.

Equation 3.2, p 121). Non-random mating, genetic drift, mutations and natural selection would disrupt the equilibrium. Accordingly, three SNPs were excluded for deviating from HWE at P < 1E-7 in controls and P < 1E-12 in cases.

Imputation input files (.gen, .sample and .strand) per each chromosome were then generated in R. The .gen files were updated to replace certain indel variant alleles (D/I) to match with 1000 Genomes alleles (A/T/G/C).

3.5.3 Phasing, Imputation & Logistic Regression

In order to expedite imputation, each chromosome was pre-phased which allows for estimating the haplotypes prior to imputation, thus reducing computing costs. The pre-phasing was run via SHAPEITv2 where each chromosome was referenced to Genome Reference Consortium human build 37 (GRCh37) assembly (Church *et al*., 2011; Delaneau *et al*., 2013; Zagury and Marchini, 2020). All samples were imputed with IMPUTEv2 using 1000 Genomes Project version 3 for non-overlapping 5 megabase intervals (Marchini *et al*., 2007; Howie *et al*., 2011; Auton *et al*., 2015; Howie and Marchini, 2020).

Logistic Regression (see **Section 3.5.5, Equation 3.7**) was performed on each chromosome segment by case-control status as well as after restricting to endometrioid or non-endometrioid histologies by using a purpose-written software (Tyrer, 2020). The output information included effect allele (EA), other allele (OA), effect allele frequency (EAF) in both cases and controls, imputation score (r^2), log odds ratio (OR), log standard error (SE), P value, Wald's statistic, and likelihood ratio test (LRT).

3.5.4 Post-Imputation QC

Output files from the logistic regression analyses were merged together. The data was then restricted to SNPs with EAF$\geq$ 0.1% in controls, r^2> 0.4, -1.1 <logOR< 1.1, and LRT< 73.5. Following this, multiple SNPs at the same genomic position in a given chromosome were identified. Only the first SNP on a given duplicated position was included.

To assess expected versus observed LRT values in our dataset, a Quantile-Quantile (Q-Q) plot was created by looking at the results of the list of known uncorrelated SNPs as before.

3.5.5 Statistical Analysis

Statistical analyses were carried out using R version 3.6.0, unless stated otherwise (R Core Team, 2020).

3.5.5.1 Hardy-Weinberg Equilibrium

Genotype distributions of selected SNPs were compared in endometrial cancer cases and controls and tested for HWE using:

$$p^2 + 2pq + q^2 = 1 \tag{3.2}$$

where p is the population frequency of "A" allele and q is the population frequency of "B" allele. Thus, genotypes of homozygous AA, heterozygous AB and homozygous BB are represented by p^2, $2pq$ and q_2 in the equation.

Deviation from HWE was assessed by chi-squared ($\tilde{\chi}^2$) test using:

$$\tilde{\chi}^2 = \sum_g \frac{(O_k - E_k)^2}{E_k} \tag{3.3}$$

for which $P < 0.05$ was accepted as significant. In this formula, O_k and E_k are the observed and expected values of the position k in a contingency table, respectively, and g represents the genotypes.

3.5.5.2 Genome-Wide Association Analysis

3.5.5.2.1 Odds Ratio

ORs for each genotype per SNP were calculated to assess the association and strength of genotypes with endometrial cancer risk. Formula for calculating OR is given below.

$$\text{OR} = \frac{\text{(n) exposed cases}/\text{(n) unexposed cases}}{\text{(n) exposed controls}/\text{(n) unexposed controls}} \tag{3.4}$$

95% confidence interval (95% CI) for a given OR is calculated by:

$$\exp^{\log(\text{OR})\pm 1.96\times SE} \tag{3.5}$$

where $SE = \sqrt{\sigma^2[\log(\text{OR})]}$.

3.5.5.2.2 Genome-Wide Association Threshold

A pre-determined genome-wide significance threshold is used here to discover SNPs with *P* value that survives the multiple testing correction inspired by Bonferroni. This threshold, *P* < 5E-8, is obtained by:

$$\text{P} = \frac{0.05}{1,000,000} \tag{3.6}$$

where 0.05 is the common significance threshold for P and 1 million is the estimated independent variants (Kuo, 2017).

3.5.5.2.3 Logistic Regression

To test for an association between a genotype of a SNP with a binary outcome, such as case-control status, logistic regression was employed for this GWAS. This statistical model assesses association of each SNP with the phenotype-of-interest and can be corrected for covariates such as BMI. The formula for a basic logistic regression is:

$$ln\frac{p}{1-p} = \beta_0 + \beta_1 G + \beta_2 X \tag{3.7}$$

where p is the probability of disease of interest, $\frac{p}{1-p}$ is the odds of this disease, G is the genotype, and X is other covariate. In this formula, β_1 is a regression coefficient measuring a change in $ln\frac{p}{1-p}$ per unit change in the genotype (G). OR can then be estimated by taking the exponentiation of β (e^{β}).

3.5.5.2.4 Genomic Inflation Factor

Differing allele frequencies due to differences between ancestries, particularly when cases and controls are not perfectly matched in this regard can lead to bias and inflation of the test statistics. The standard approach to detect inflation due to population substructure is to calculate the genomic inflation factor (λ). λ is the ratio of the median of the observed $\tilde{\chi}^2$ to the expected median of the $\tilde{\chi}^2$ distribution. Thus, λ can be calculated using:

$$\lambda = \frac{M_{\tilde{\chi}^2}}{0.45} \tag{3.8}$$

where $M_{\tilde{\chi}^2}$ is the median of the observed $\tilde{\chi}^2$ values, in this case the LRT, and 0.45 is the median of a $\tilde{\chi}^2$ distribution with one degree of freedom. Hence, a λ value of 1 indicates no genomic inflation.

λ_{1000} is a standardised estimate of genomic inflation regardless of sample size. This is calculated by:

$$\lambda_{1000} = \frac{1 + (500 * (\lambda - 1))}{1/(\frac{1}{n_{case}} + \frac{1}{n_{control}})} \tag{3.9}$$

where the estimated genomic inflation is expected to be 1 in the absence of bias.

3.5.5.3 Descriptive Statistics

Student's t-test was carried out to assess any significant differences for age and BMI between cases and controls, which uses the following formula:

$$t = \frac{(\overline{x} - \overline{y})}{\hat{\sigma}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{3.10}$$

where $\overline{x}$and $\overline{y}$ are the sample means ($\overline{x} = \frac{\sum x}{n}$) of groups x and y, n is the number of samples per group, while $\hat{\sigma}$ is the pooled variance estimate ($\hat{\sigma} = \frac{\sum x^2}{n} - \overline{x}^2$) .

3.6 Results

3.6.1 Characteristics of the Study Population

After exclusions, the Manchester cohort included 555 cases and 1,202 female controls (**Figure 3.3**). The cases included surgically or histopathologically-confirmed primary endometrial cancers whereas controls had no history of endometrial or breast cancer, and had not undergone hysterectomy. The majority of the cases were of endometrioid histological subtype (75.3%) and had median age and BMI of 64 years and 30.8 kg/m^2, respectively. The controls were slightly younger and substantially leaner with a median age and BMI of 59 years and 25.45 kg/m^2 (P_{age}=1.12E-12 and P_{BMI}=1.08E-36). Importantly, these were based on baseline measurements and although baseline median age of control participants was significantly smaller than the cases, the controls were followed-up for a median of 8.9 years. This indicates that the majority would have reached the age group most at-risk for endometrial cancer and not received a diagnosis.

3.6.2 Manchester Study GWAS

Previous studies have identified 19 genome-wide significant SNPs associated with endometrial cancer (O'Mara *et al.*, 2018; Bafligil *et al.*, 2020). Suggestive SNPs, which were shown to improve PRS performance by Mavaddat *et al.*, may reach genome-wide significance in future efforts with the expansion of the sample size of consortium-level GWAS. Thus, we obtained a list of suggestive SNPs from the largest endometrial cancer GWAS, ECAC. To validate these SNPs and further identify indicative loci, we conducted an independent GWAS using Manchester cases and controls after exclusions (**Figure 3.3**).

Several logistic regression analyses were performed to evaluate overall case-control (**Section 3.6.3, p 125**), BMI-adjusted case-control (**Section 3.6.3.1, p 128**), and subtype-restricted (endometrioid/non-endometrioid versus control) associations (**Sections 3.6.4, p 130** and **3.6.5, p 132**). Neither of the analyses showed evidence of a genomic inflation with λ and λ_{1000} calculated to be in the range of 1.02-1.03 and 1.03-1.07, respectively.

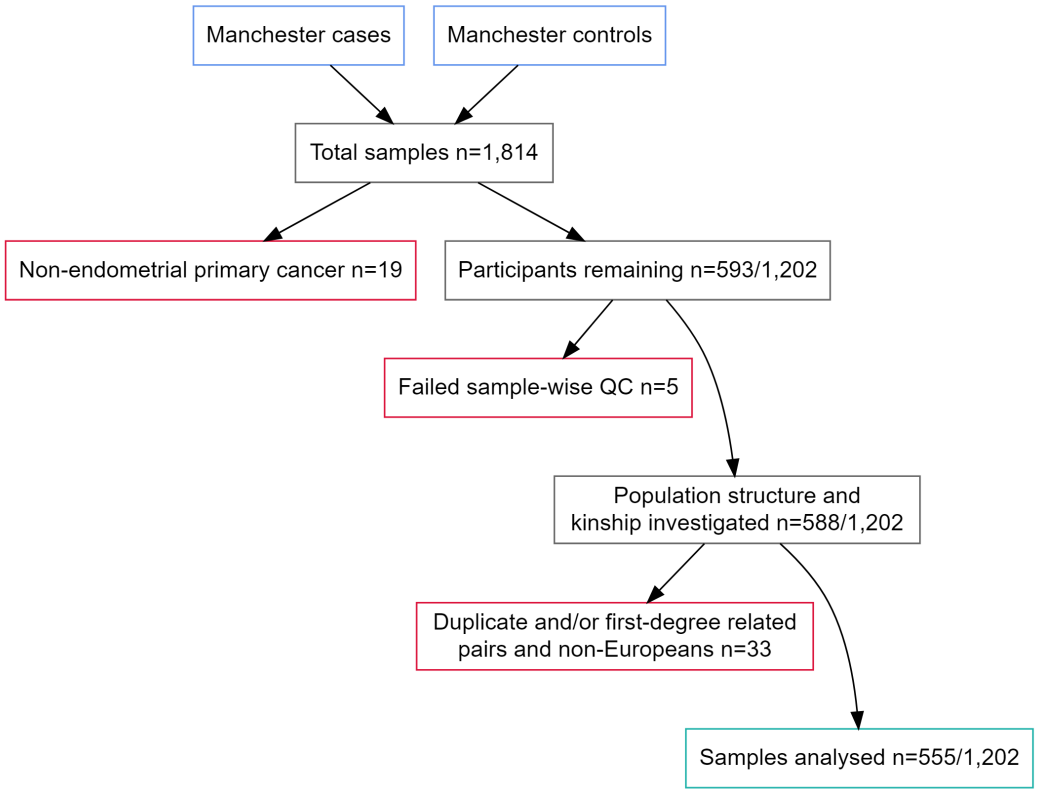


Figure 3.3: *Flowchart of sample selection and QC.*

Key: Exclusions | Studies | Samples at each stage | Final number of participants included in this study

3.6.3 Association of SNPs with Endometrial Cancer

Logistic regression was performed to assess the association of SNPs imputed with endometrial cancer risk based on disease status. The total number of SNPs that were analysed after restrictions described in **Section 3.5.3** and removal of duplicate probes at the same location (n=89,171) was 13,019,021.

Using the uncorrelated SNPs subset, both the λ and λ_{1000} were estimated to be an optimal level of 1.03. Then, using *qqman* and *snpStats* packages in R (Turner, 2017; Clayton, 2019), a Q-Q plot (**Figure 3.4**) and a Manhattan (**Figure 3.5**) plot were created to visualise the results. The majority of the SNPs observed a uniform distribution as seen in the Q-Q plot (**Figure 3.4**). Though no genome-wide significant SNPs were found due to sample

size, there were trails of SNPs seen in multiple chromosomes at the suggestive P value threshold of 1E-5 (**Figure 3.5**).

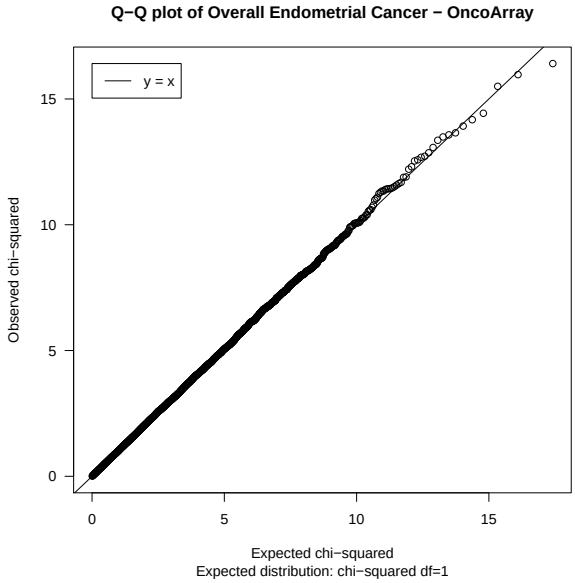


Figure 3.4: *A Q-Q plot comparing the expected and observed results for the uncorrelated SNPs after imputation.*

Nearly all SNPs follow the reference $\tilde{\chi}^2$ line indicating a ratio of 1, as expected. Each ∘ represents an individual SNP.

Two particular regions that were suggestive of an association with endometrial cancer were identified: 18q12.3 (rs111282723) and 3p24.1 (rs9854980). ORs of the lead SNPs located at the 18q12.3 and 3p24.1 loci were 2.61 (P = 3.86E-7) per A allele and 0.57 (P = 2.43E-6) per T allele, respectively. Other suggestive SNPs identified through this GWAS are presented in **Figure 3.5** and detailed below in **Section 3.6.3**. The SNPs identified here which were also included in the PRS analyses explained in **Section 4** were: rs12727038 (1q42.4), rs144065942 (2q14.2), rs9854980 (3p24.1), rs6580584 (5q32), rs74532550 (12p11.23) and rs111282723 (18q12.3).

One SNP identified here (rs5942128) located on X chromosome was not included in further analyses due to complete lack of data available for this chromosome within the ECAC

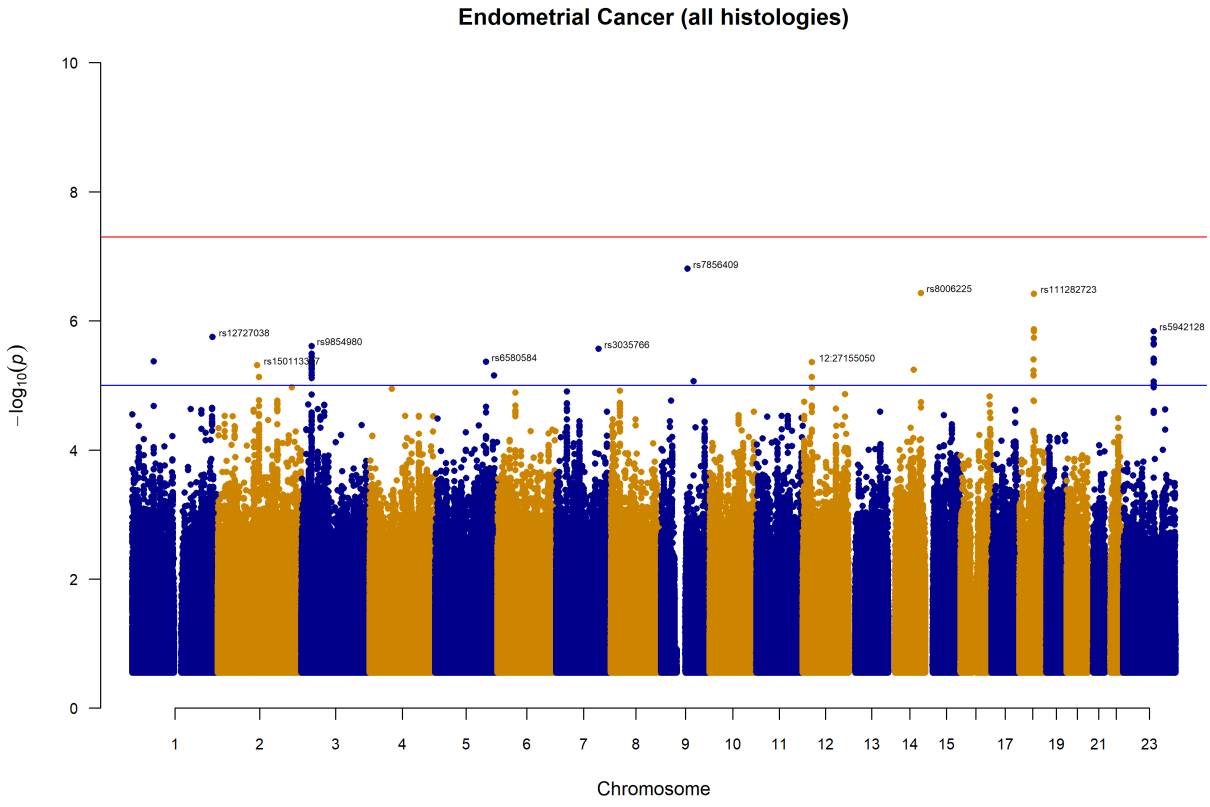


Figure 3.5: *A Manhattan plot of GWAS results from case versus control comparison within the Manchester cohort.*

Each dot represents a SNP with the $-\log_{10} P$ values (y-axis) for SNPs plotted against the genomic position (x-axis) across all 23 chromosomes. Horizontal lines indicate the P value of genome-wide significance at -log 5E-8 (red) and the suggestive significance limit at -log 1E-5 (blue). Top SNPs above the suggestive threshold per chromosome are annotated. 12:27155050 corresponds to rs74532550. Chromosome 23 represents the X chromosome.

GWAS. Therefore, although presented in the figures, this SNP is not included in the results table below (**Table 3.2**) or the PRSs.

3.6.3.1 BMI-Adjusted Analysis

BMI is an important covariate to consider due to its strong effect on endometrial cancer risk. Although, none of the SNPs identified through our systematic review were directly associated with BMI, we repeated the Manchester study association test additionally adjusting for BMI and compared the results to those of the unadjusted results above (Bafligil *et al.*, 2020). After exclusions including duplicate probes (n=87,289), there were 12,826,953 SNPs available for the analyses.

λ and λ_{1000} were estimated at 1.02 and 1.03, respectively. Q-Q and Manhattan plots are given below (**Figures 3.6** and **3.7**. A substantial trail of SNPs in chromosome 3 can be seen achieving strong P values, the strongest just short of genome-wide significance threshold.

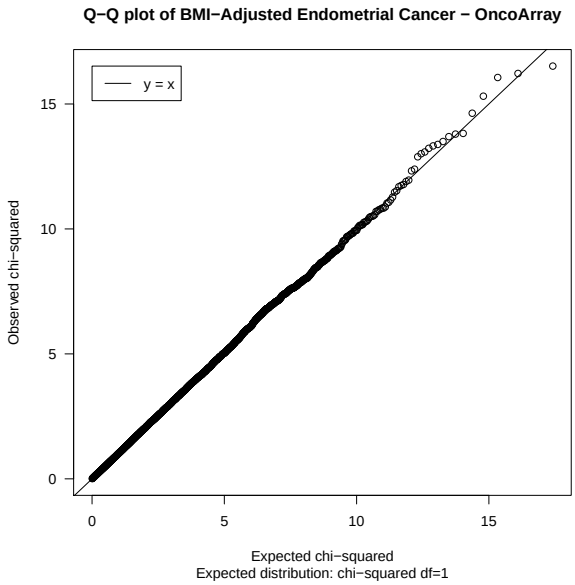


Figure 3.6: *A Q-Q plot comparing the expected and observed results for the uncorrelated SNPs of BMI-adjusted results.*

Nearly all SNPs follow the reference $\tilde{\chi}^2$ line indicating a ratio of 1, as expected. Each ○ represents an individual SNP.

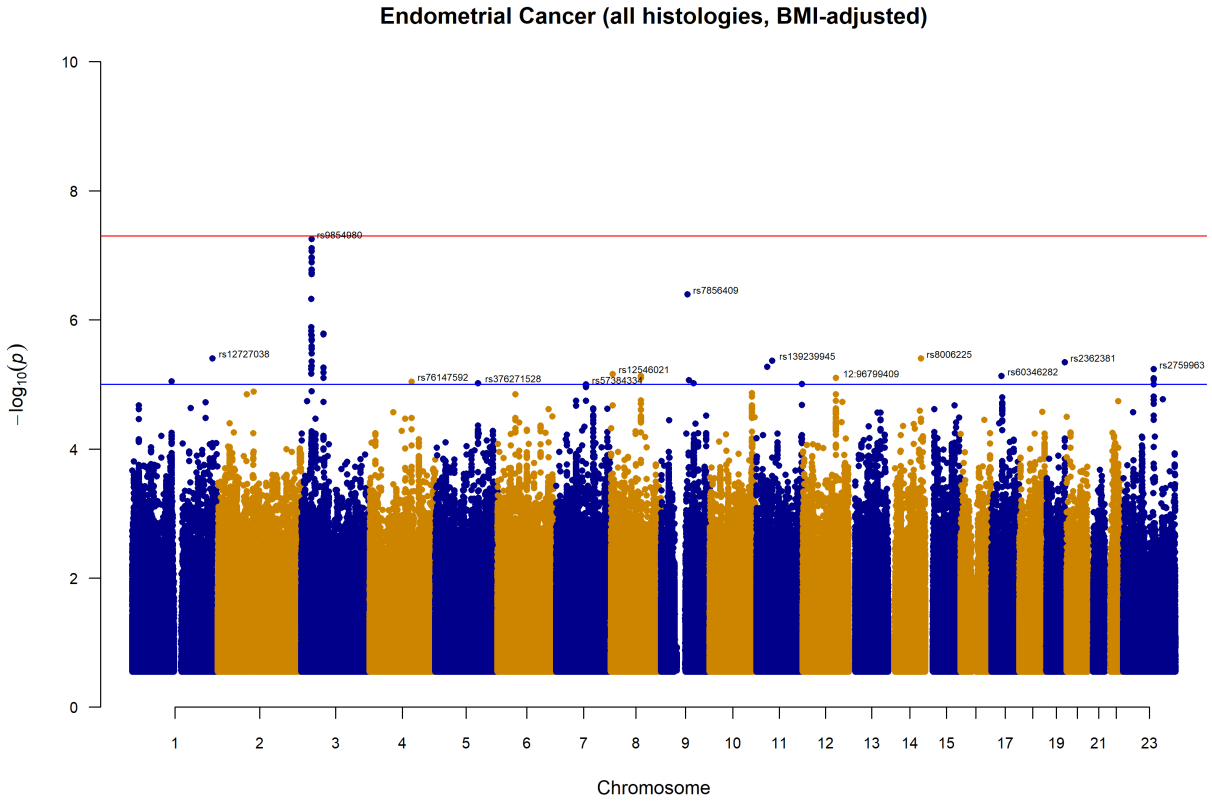


Figure 3.7: *A Manhattan plot of BMI-adjusted GWAS results from case versus control comparison within the Manchester cohort.*

Each dot represents a SNP with the $-\log_{10} P$ values (y-axis) for SNPs plotted against the genomic position (x-axis) across all 23 chromosomes. Horizontal lines indicate the P value of genome-wide significance at -log 5E-8 (red) and the suggestive significance limit at -log 1E-5 (blue). Top SNPs above the suggestive threshold per chromosome are annotated. Chromosome 23 represents the X chromosome.

3.6.4 Association of SNPs with Endometrioid Histotype

As reported in the systematic review of literature in **Chapter 2**, the vast majority of the studies to date have utilised datasets where the cases were primarily comprised of or restricted to endometrioid subtype tumours. Therefore, we also sought to investigate the results when the cases were restricted to endometrioid histology only. There were 415 endometrioid tumours available for the analyses. The total number of SNPs analysed after restrictions and removal of duplicate probes (n=87,242) were 12,818,561. Unlike overall analyses, there was a SNP that was significant at $P < 5E\text{-}8$ (**Figure 3.9**).

In this analysis, λ and λ_{1000} were estimated at 1.02 and 1.04, respectively, indicating no substantial genomic inflation. A Q-Q plot **Figure 3.8** and a Manhattan plot **Figure 3.9** were created to visualise the results.

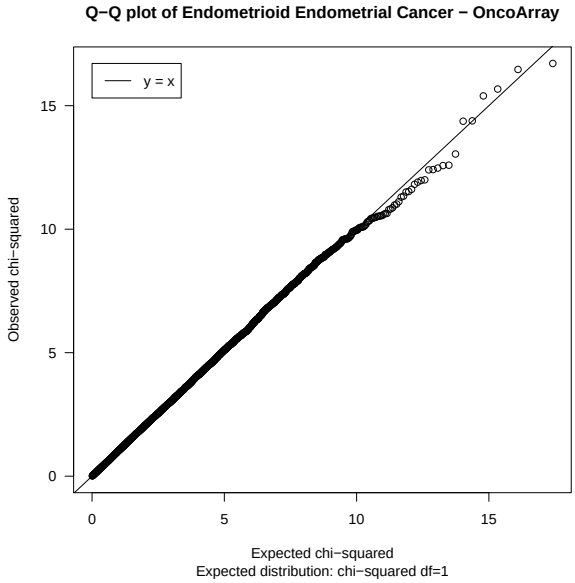


Figure 3.8: *A Q-Q plot comparing the expected and observed results for the uncorrelated SNPs for endometrioid-only versus control analyses after imputation.*

Nearly all SNPs follow the reference $\tilde{\chi}^2$ line indicating a ratio of 1, as expected. Each ∘ represents an individual SNP.

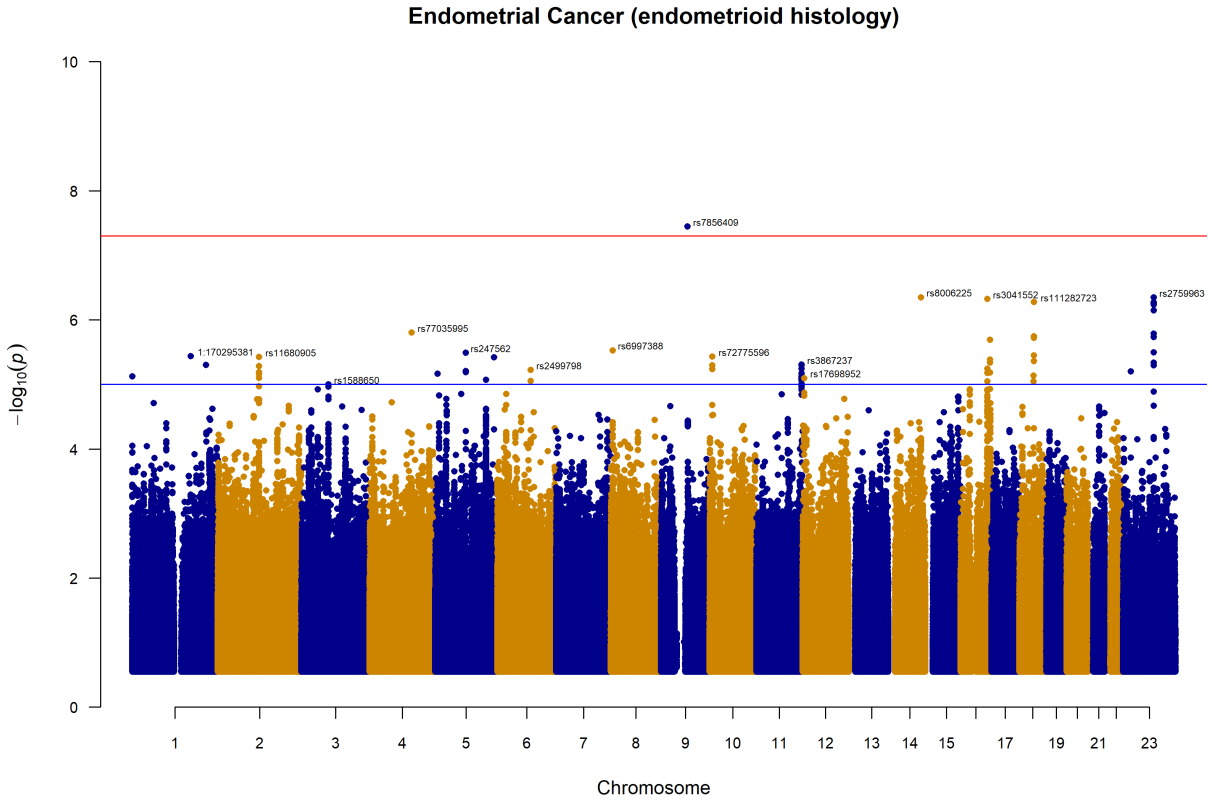


Figure 3.9: *A Manhattan plot of GWAS results from endometrioid subtype versus control comparison within the Manchester cohort.*

Each dot represents a SNP with the $-\log_{10} P$ values (y-axis) for SNPs plotted against the genomic position (x-axis) across all 23 chromosomes. Horizontal lines indicate the P value of genome-wide significance at -log 5E-8 (red) and the suggestive significance limit at -log 1E-5 (blue). Top SNPs above the suggestive threshold per chromosome are annotated. Chromosome 23 represents the X chromosome.

3.6.5 Association of SNPs with Non-endometrioid Histotype

To date, only one SNP (rs14826157, OR=1.64, 95% CI 1.32-2.04, P = 9.6E-6) has been shown to be suggestively associated with the non-endometrioid subset of endometrial tumours, while no SNP has been reported to be exclusively associated with this subtype (see **Chapter 2**) (O'Mara *et al*., 2018; Bafligil *et al*., 2020). In our Manchester cohort, there were 136 cases of non-endometrioid histology. Although the sample size is small, we nevertheless conducted association test while restricting the cases to non-endometrioid tumours only. After exclusions and removal of duplicate probes (n=76,897), there were 11,723,911 SNPs available for analyses.

In this analysis, λ and λ_{1000} were estimated at 1.02 and 1.07, respectively. Q-Q and Manhattan plots are illustrated in **Figures 3.10** and **3.11**.

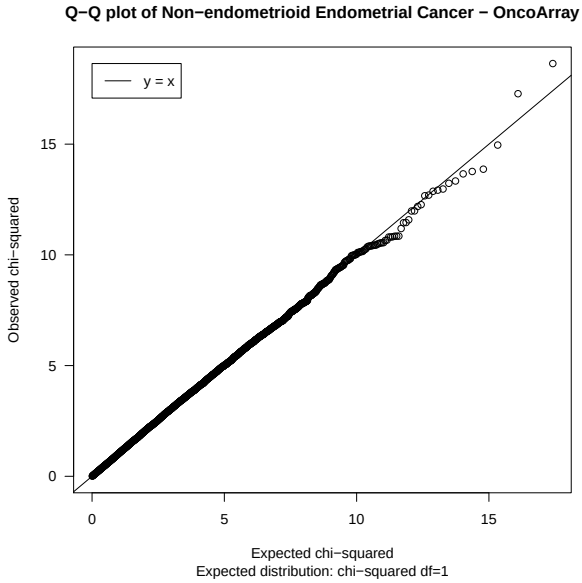


Figure 3.10: *A Q-Q plot comparing the expected and observed results for the uncorrelated SNPs for non-endometrioid subtype versus control analyses after imputation.*

Most SNPs follow the reference $\tilde{\chi}^2$ line indicating a ratio of 1, as expected. Each ∘ represents an individual SNP.

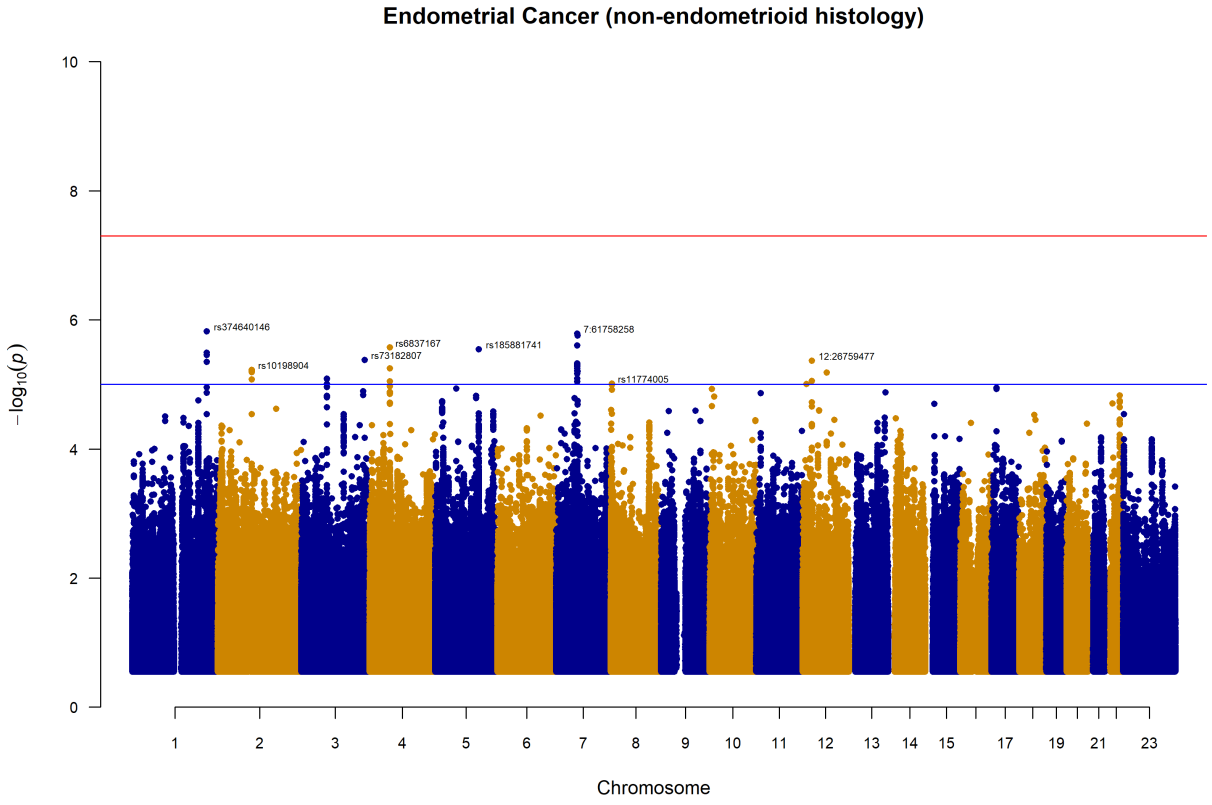


Figure 3.11: *A Manhattan plot of GWAS results from non-endometrioid versus control comparison within the Manchester cohort.*

Each dot represents a SNP with the $-\log_{10} P$ values (y-axis) for SNPs plotted against the genomic position (x-axis) across all 23 chromosomes. Horizontal lines indicate the P value of genome-wide significance at -log 5E-8 (red) and the suggestive significance limit at -log 1E-5 (blue). Top SNPs above the suggestive threshold per chromosome are annotated. Chromosome 23 represents the X chromosome.

3.6.6 Suggestive SNPs Identified Through Manchester GWAS

A number of suggestive SNPs were identified by the Manchester GWAS (overall case versus control analysis). Of those, only the loci with multiple SNPs in LD were selected to include in the PRS over single hits. One SNP, rs5942128 in chromosome 23, identified in the Manchester GWAS was not available in the ECAC dataset, and was excluded from further analyses. The details of the six SNPs identified here are given below in **Table 3.2**.

The lowest imputation score (r^2) among these SNPs was 0.77, and all were consistent between the different analyses assessed. Similarly, the EAFs and betas across the comparisons were similar. SEs, *P* values, LRT and Wald tests, on the other hand, were dissimilar, owing to the varied number of samples available for each of the analyses.

Table 3.2: *Details of the Manchester GWAS SNPs identified from overall (case-control) analysis.*

The SNP details presented here were obtained from overall, BMI-adjusted overall, endometrioid and non-endometrioid restricted analyses.

Analysis	rsID	Chr:Pos	EA	OA	EAF (con)	EAF (case)	r^2	Control (n)	Case (n)	beta	SE	Wald	LRT	*P*
Overall	rs12727038	1:233730259	C	T	0.27	0.35	0.85	1202	555	0.42	0.09	22.85	22.84	1.76E-06
	rs144065942	2:119953289	GGGAA	G	0.78	0.72	0.81			-0.43	0.10	20.22	20.10	7.34E-06
	rs9854980	3:29054809	T	C	0.14	0.10	1.00			-0.56	0.12	20.37	22.22	2.43E-06
	rs6580584	5:148220997	C	T	0.07	0.04	0.77			-0.93	0.22	18.06	21.13	4.29E-06
	rs74532550	12:27155050	A	G	0.11	0.17	0.98			0.49	0.11	21.46	21.10	4.36E-06
	rs111282723	18:41773405	A	T	0.02	0.06	0.99			0.96	0.19	25.48	25.80	3.78E-07
BMI-adjusted	rs12727038	1:233730259	C	T	0.27	0.35	0.85	1142	501	0.46	0.10	21.24	21.30	3.92E-06
	rs144065942	2:119953289	GGGAA	G	0.78	0.72	0.81			-0.38	0.11	12.22	12.11	5.02E-04
	rs9854980	3:29054809	T	C	0.14	0.09	1.00			-0.75	0.15	26.30	29.50	5.59E-08
	rs6580584	5:148220997	C	T	0.07	0.04	0.77			-0.80	0.24	10.83	12.27	4.60E-04
	rs74532550	12:27155050	A	G	0.11	0.16	0.98			0.46	0.12	14.47	14.21	1.64E-04
	rs111282723	18:41773405	A	T	0.02	0.06	0.99			0.81	0.22	14.20	14.10	1.73E-04
Endometrioid	rs12727038	1:233730259	C	T	0.27	0.35	0.85	1202	415	0.42	0.10	17.97	17.87	2.37E-05
	rs144065942	2:119953289	GGGAA	G	0.78	0.71	0.82			-0.47	0.11	19.64	19.41	1.06E-05
	rs9854980	3:29054809	T	C	0.14	0.10	1.00			-0.53	0.14	14.71	16.15	5.87E-05
	rs6580584	5:148220997	C	T	0.07	0.03	0.77			-1.02	0.25	16.22	19.84	8.42E-06
	rs74532550	12:27155050	A	G	0.11	0.15	0.98			0.40	0.12	11.46	11.11	8.57E-04
	rs111282723	18:41773405	A	T	0.02	0.06	0.99			1.04	0.21	25.82	25.16	5.27E-07
NE	rs12727038	1:233730259	C	T	0.27	0.36	0.85	1202	136	0.47	0.15	9.50	9.25	2.35E-03
	rs144065942	2:119953289	GGGAA	G	0.78	0.74	0.81			-0.30	0.17	3.22	3.12	7.73E-02
	rs9854980	3:29054809	T	C	0.14	0.09	1.00			-0.62	0.23	7.42	8.59	3.37E-03
	rs6580584	5:148220997	C	T	0.07	0.04	0.79			-0.84	0.39	4.61	5.66	1.74E-02
	rs74532550	12:27155050	A	G	0.11	0.20	0.97			0.74	0.17	18.18	16.67	4.45E-05
	rs111282723	18:41773405	A	T	0.02	0.06	0.99			0.77	0.30	6.54	5.65	1.75E-02

The positions are based on GRCh37 and r^2 indicates imputation score.
EA, effect allele; EAF, effect allele frequency; LRT, Likelihood Ratio Test; NE, non-endometrioid; OA, other allele; SE, Standard Error.
P values indicate $P < 1E\text{-}5$.

3.6.7 Validation of the SNPs Identified Through the Systematic Review

24 SNPs were identified through the systematic review detailed in **Chapter 2** (Bafligil *et al.*, 2020). These included 19 SNPs at genome-wide significance level (including the variant at *AKT1*), and the majority were identified by the largest endometrial cancer GWAS through ECAC (O'Mara *et al.*, 2018). We sought to validate these SNPs in our independent Manchester GWAS. The results are presented in **Table 3.3** and include overall and BMI adjusted analyses.

In both analyses, the lowest imputation score was 0.82, and for the majority of the SNPs, adjustment of BMI changed neither the direction nor the size of the effect of association (beta). Betas of two SNPs, rs139584729 and rs2498794, changed from -0.13 to 0.05 and 0.02 to -0.02, respectively. Moreover, the beta of rs79575945 saw a sizeable reduction from -0.12 to -0.02 when BMI was accounted for.

Of the 24 SNPs, 18 had betas in the same direction per EA in the overall analyses in comparison to the published results (**Tables 3.3** and **2.1**). Moreover, two SNPs, rs3184504 and rs10850382, had betas very close to 0.00 (equal to OR 1.00) in the overall analyses. On the other hand, 16 of the SNPs had same direction of effect sizes in the BMI-adjusted analyses. rs3184504 and rs882380 had betas very close to 0.00 in these analyses.

Seven and four of the 24 SNPs were significant at a P value threshold of < 0.05 in the overall and BMI-adjusted analyses, respectively (**Table 3.3**). All of these significant SNPs in both analyses had betas in the same direction as the published studies that they were identified in. Importantly, rs11263761 near the first identified endometrial cancer GWAS risk region *HNF1B* (17q12), which is the most robust SNP reported and validated to date, was also significant in both overall and BMI-adjusted analyses. Similarly, *CYP19A1* SNP rs17601876 (15q21.2), *EIF2AK4* SNP rs937213 (15q15.1), and *HEY2* SNP rs2747716 (6q22.31) were significant in both analyses. In addition to the previously mentioned, rs1740828 (6q22.3) of *SOX4*, rs35286446 (8q24) of *MYC* and rs9668337 (12p12.1) of *SSPN* were also

significant in the overall GWAS analysis.

Table 3.3: *Details of 24 systematic review SNPs obtained in overall and BMI-adjusted analyses of Manchester GWAS.*

Analysis	rsID	Chr:Pos	EA	OA	EAF (con)	EAF (case)	r^2	Control (n)	Case (n)	beta	SE	Wald	LRT	*P*
Overall	rs673604	1:35687815	C	T	0.08	0.09	1.00	1202	555	0.09	0.13	0.48	0.48	4.90E-01
	rs113998067	1:38073356	C	T	0.04	0.05	0.89			0.18	0.18	0.97	0.95	3.29E-01
	rs148261157	2:60897579	A	G	0.03	0.03	0.87			0.11	0.22	0.24	0.24	6.27E-01
	rs8178648	3:93605739	C	T	0.05	0.05	0.97			-0.01	0.17	0.01	0.01	9.34E-01
	rs9399840	6:104076463	T	C	0.52	0.53	1.00			0.05	0.07	0.48	0.48	4.90E-01
	rs2747716	6:126008372	G	A	0.44	0.39	1.00			-0.21	0.07	7.62	7.69	5.55E-03
	rs79575945	6:152158847	G	A	0.10	0.08	0.99			-0.12	0.13	0.81	0.83	3.63E-01
	rs1740828	6:21649085	A	G	0.47	0.43	1.00			-0.18	0.08	5.78	5.82	1.59E-02
	rs35286446	8:129445863	GAT	G	0.56	0.60	1.00			0.16	0.07	4.87	4.90	2.68E-02
	rs4733613	8:129599278	G	C	0.88	0.87	1.00			-0.15	0.11	1.85	1.83	1.77E-01
	rs139584729	8:129623902	G	C	0.03	0.02	0.97			-0.13	0.25	0.25	0.26	6.12E-01
	rs1679014	9:22207037	C	T	0.93	0.92	1.00			-0.17	0.14	1.45	1.43	2.31E-01
	rs10835920	11:32489664	T	C	0.39	0.42	1.00			0.08	0.07	1.29	1.29	2.57E-01
	rs3184504	12:111884608	C	T	0.52	0.54	1.00			0.00	0.07	0.00	0.00	9.83E-01
	rs10850382	12:115214548	T	C	0.33	0.33	1.00			0.01	0.08	0.01	0.01	9.12E-01
	rs9668337	12:26426338	A	G	0.74	0.76	1.00			0.21	0.09	5.58	5.68	1.72E-02
	rs7981863	13:73812141	T	C	0.30	0.29	1.00			-0.11	0.08	1.80	1.82	1.78E-01
	rs2498794	14:105245251	G	A	0.47	0.48	0.82			0.02	0.08	0.05	0.05	8.28E-01
	rs1953358	14:56295580	A	G	0.49	0.51	1.00			0.08	0.07	1.26	1.26	2.62E-01
	rs937213	15:40322124	C	T	0.41	0.45	1.00			0.17	0.08	4.96	4.96	2.59E-02
	rs17601876	15:51553909	G	A	0.47	0.53	0.99			0.23	0.07	9.87	9.93	1.63E-03
	rs1129506	17:29646032	A	G	0.61	0.63	1.00			0.11	0.08	1.82	1.83	1.76E-01
	rs11263761	17:36097775	A	G	0.51	0.57	0.98			0.25	0.08	10.54	10.64	1.11E-03
	rs882380	17:46294236	A	C	0.62	0.62	0.99			0.02	0.08	0.05	0.05	8.22E-01
BMI-adjusted	rs673604	1:35687815	C	T	0.08	0.10	1.00	1142	501	0.09	0.15	0.36	0.36	5.49E-01
	rs113998067	1:38073356	C	T	0.04	0.05	0.89			0.21	0.21	1.03	1.01	3.14E-01
	rs148261157	2:60897579	A	G	0.03	0.04	0.88			0.15	0.25	0.37	0.37	5.46E-01
	rs8178648	3:93605739	C	T	0.05	0.05	0.97			-0.03	0.20	0.03	0.03	8.68E-01
	rs9399840	6:104076463	T	C	0.52	0.53	1.00			0.09	0.08	1.28	1.28	2.58E-01
	rs2747716	6:126008372	G	A	0.44	0.39	1.00			-0.20	0.09	5.25	5.30	2.13E-02
	rs79575945	6:152158847	G	A	0.10	0.09	0.99			-0.02	0.14	0.01	0.01	9.11E-01
	rs1740828	6:21649085	A	G	0.46	0.43	1.00			-0.15	0.09	2.97	2.98	8.44E-02
	rs35286446	8:129445863	GAT	G	0.56	0.59	1.00			0.15	0.08	2.97	2.99	8.39E-02
	rs4733613	8:129599278	G	C	0.89	0.87	1.00			-0.15	0.13	1.30	1.29	2.57E-01
	rs139584729	8:129623902	G	C	0.03	0.02	0.97			0.05	0.28	0.03	0.03	8.66E-01
	rs1679014	9:22207037	C	T	0.93	0.92	1.00			-0.20	0.16	1.62	1.59	2.07E-01
	rs10835920	11:32489664	T	C	0.39	0.41	1.00			0.08	0.09	0.97	0.97	3.24E-01
	rs3184504	12:111884608	C	T	0.53	0.55	1.00			-0.01	0.08	0.01	0.01	9.31E-01
	rs10850382	12:115214548	T	C	0.33	0.33	1.00			0.08	0.09	0.74	0.74	3.89E-01
	rs9668337	12:26426338	A	G	0.73	0.76	1.00			0.18	0.10	3.32	3.37	6.65E-02
	rs7981863	13:73812141	T	C	0.31	0.29	1.00			-0.18	0.09	3.48	3.52	6.08E-02
	rs2498794	14:105245251	G	A	0.47	0.48	0.82			-0.02	0.09	0.06	0.06	8.09E-01
	rs1953358	14:56295580	A	G	0.49	0.51	1.00			0.10	0.08	1.42	1.43	2.32E-01
	rs937213	15:40322124	C	T	0.41	0.45	1.00			0.20	0.09	5.43	5.43	1.98E-02
	rs17601876	15:51553909	G	A	0.47	0.54	0.99			0.30	0.08	12.97	13.10	2.95E-04
	rs1129506	17:29646032	A	G	0.61	0.63	1.00			0.07	0.09	0.65	0.65	4.21E-01
	rs11263761	17:36097775	A	G	0.51	0.57	0.98			0.26	0.09	9.03	9.12	2.53E-03
	rs882380	17:46294236	A	C	0.61	0.61	0.99			0.00	0.09	0.00	0.00	9.88E-01

The positions are based on GRCh37 and r^2 indicates imputation score. rsIDs indicate same direction of betas obtained as published studies. P values indicate $P<0.05$.
EA, effect allele, EAF, effect allele frequency; LRT, Likelihood Ratio Test; OA, other allele; SE, standard error.

3.6.8 Validation of Suggestive SNPs Identified Through ECAC GWAS

A recent article by Mavaddat *et al.* showed that using suggestive SNPs in a breast cancer PRS increased the discriminatory performance of the PRS. Therefore, we extracted independent suggestive SNPs obtained from ECAC excluding UK Biobank (UKBB) data (explained in more detail in **Section 4.5.3, p 172**). Three SNPs were excluded due to failing QC in UKBB for the PRS analyses, as described in **Section 4.4.3.3, p 157**).

The results of the 48 suggestive ECAC SNPs obtained from our overall analysis is presented in **Table 3.4**. The lowest r^2 score was 0.63, and there were four SNPs with r^2 score of < 0.70 in total. Some SNPs had very low EAFs, less than 0.01. More than half of the SNPs (n=29) had the same direction of effect, while only three were significant at $P < 0.05$.

Table 3.4: *Results of the 48 suggestive ECAC SNPs obtained in Manchester GWAS analysis*.

rsID	Chr:Pos	EA	OA	EAF (con)	EAF (case)	r^2	Control (n)	Case (n)	beta	SE	Wald	LRT	*P*
rs11583244	1:225952474	T	C	0.40	0.39	0.95			-0.05	0.08	0.34	0.34	5.60E-01
rs7579014	2:60707894	A	G	0.63	0.63	0.98			0.08	0.08	0.97	0.98	3.23E-01
rs2920505	3:12335081	A	G	0.61	0.61	0.98			0.00	0.08	0.00	0.00	9.79E-01
rs73044842	3:13714562	A	G	0.01	0.01	0.89			-0.39	0.43	0.81	0.86	3.53E-01
rs2659685	3:128122396	A	G	0.25	0.21	1.00			-0.28	0.09	9.36	9.61	1.94E-03
rs72716510	4:136014605	T	A	0.06	0.06	0.90			0.05	0.17	0.09	0.09	7.64E-01
rs7713218	5:1283312	G	A	0.55	0.51	0.96			-0.14	0.08	3.23	3.23	7.22E-02
rs7729155	5:82058370	C	G	0.83	0.81	0.90			-0.11	0.10	1.26	1.25	2.63E-01
rs1357050	6:122391135	G	T	0.37	0.34	1.00			-0.11	0.08	2.19	2.20	1.38E-01
rs60515555	6:126000638	G	T	0.56	0.61	1.00			0.20	0.07	7.44	7.51	6.15E-03
rs76165228	6:150519175	G	A	0.87	0.84	0.90			-0.20	0.11	3.16	3.12	7.75E-02
rs2982732	6:152364273	A	G	0.31	0.35	0.99	1202	555	0.09	0.08	1.20	1.19	2.75E-01
rs149404333	6:29146703	C	A	0.12	0.10	1.00			-0.08	0.12	0.41	0.42	5.18E-01
rs9296422	6:43905816	C	G	0.25	0.23	0.98			-0.11	0.09	1.50	1.51	2.19E-01
rs149369224	7:103074830	C	G	0.01	0.00	0.94			-0.68	0.53	1.65	1.87	1.72E-01
rs117274813	7:68267626	T	G	0.01	0.01	0.65			-0.10	0.40	0.07	0.07	7.94E-01
rs117978821	7:99107775	C	T	0.03	0.02	0.99			-0.33	0.24	2.00	2.11	1.46E-01
rs1553183	8:3276592	C	T	0.34	0.33	0.99			-0.05	0.08	0.34	0.34	5.57E-01
rs10505508	8:129215924	T	C	0.28	0.29	1.00			0.07	0.08	0.65	0.65	4.21E-01

rsID	Chr:Pos	EA	OA	EAF (con)	EAF (case)	r^2	Control (n)	Case (n)	beta	SE	Wald	LRT	*P*
rs6470612	8:129218047	G	A	0.93	0.92	0.99			-0.08	0.14	0.30	0.29	5.88E-01
rs1356332	8:129525114	A	G	0.35	0.34	1.00			-0.06	0.08	0.53	0.53	4.65E-01
rs13250178	8:143997100	T	C	0.43	0.45	0.98			0.13	0.07	3.05	3.05	8.09E-02
rs6468088	8:32028139	T	A	0.01	0.02	0.83			0.01	0.34	0.00	0.00	9.68E-01
rs535955703	8:90386853	T	C	0.01	0.01	0.84			0.07	0.47	0.03	0.02	8.75E-01
rs1923357	9:10266786	T	C	0.81	0.81	0.99			0.05	0.09	0.26	0.26	6.09E-01
rs3808753	9:17616880	G	A	0.03	0.04	0.92			0.25	0.20	1.49	1.45	2.28E-01
rs138843373	10:58753125	T	C	0.01	0.01	0.65			-0.03	0.45	0.00	0.00	9.54E-01
rs4312007	10:87185828	T	G	0.48	0.50	1.00			0.05	0.07	0.54	0.54	4.63E-01
rs76762469	11:72549374	A	C	0.03	0.04	0.72			0.24	0.23	1.08	1.06	3.03E-01
rs558029	11:4265076	C	G	0.12	0.13	0.77			0.07	0.13	0.27	0.27	6.06E-01
rs7967338	12:109004950	C	T	0.36	0.34	0.99			-0.09	0.08	1.31	1.31	2.51E-01
rs117670121	12:28331672	A	G	0.02	0.01	0.77			-0.17	0.34	0.25	0.26	6.10E-01
rs1677893	12:78338386	T	A	0.56	0.55	0.99			-0.08	0.07	1.19	1.19	2.76E-01
rs11069840	13:110988807	C	G	0.08	0.09	0.69			0.16	0.16	1.08	1.06	3.03E-01
rs538150087	13:43925809	A	G	0.02	0.01	0.74			-0.40	0.34	1.41	1.49	2.22E-01
rs144076224	13:75668743	C	T	0.01	0.01	0.73	1202	555	-0.70	0.51	1.89	2.11	1.46E-01
rs139948912	14:29486916	G	A	0.01	0.01	0.75			-0.74	0.50	2.18	2.47	1.16E-01
rs4072776	14:66287974	C	T	0.47	0.48	0.75			0.02	0.09	0.08	0.08	7.74E-01
rs141265605	15:63822709	G	A	0.02	0.02	0.79			-0.51	0.32	2.59	2.81	9.34E-02
rs60310219	15:82382871	T	C	0.01	0.01	0.63			-0.03	0.52	0.00	0.00	9.57E-01
rs34676612	16:10446142	C	T	0.12	0.15	1.00			0.27	0.11	6.29	6.19	1.28E-02
rs1533495	17:36172155	T	C	0.72	0.72	0.98			0.02	0.08	0.04	0.05	8.32E-01
rs117338667	17:9337299	T	G	0.02	0.01	0.79			-0.34	0.36	0.88	0.92	3.37E-01
rs60856912	17:65892343	T	G	0.19	0.19	0.95			0.01	0.10	0.02	0.02	8.95E-01
rs150325239	17:75498474	T	C	0.01	0.01	0.77			-0.14	0.36	0.16	0.16	6.91E-01
rs4607001	20:22448537	T	C	0.21	0.23	0.99			0.12	0.09	1.69	1.68	1.95E-01
rs577034498	21:22378129	C	A	0.06	0.07	0.99			0.15	0.15	1.05	1.03	3.10E-01
rs9616483	22:49577020	A	T	0.12	0.13	0.98			0.10	0.11	0.77	0.76	3.83E-01

The positions are based on GRCh37 and r^2 indicates imputation score. rsIDs indicate same direction of betas obtained as ECAC (excluding UKBB. P values indicate $P < 0.05$.

EA, effect allele, EAF, effect allele frequency; LRT, Likelihood Ratio Test; OA, other allele; SE, standard error.

3.7 Discussion

Endometrial cancer remains one of the few cancers with an increasing incidence and population-level mortality rate (CRUK, 2020; Siegel *et al*., 2020). This can be attributed to the growing obesity epidemic, its largest risk factor, as well as an increase in non-endometrioid tumours which contribute to mortality far greater than their endometrioid counterparts (Crosbie and Morrison, 2014). While overall five-year survival rate is as high as 80%, this dramatically reduces to 16% with advanced stage disease (Morice *et al*., 2016; Siegel *et al*., 2020). Treatment for endometrial cancer carries substantial risks such as surgical complications, radio-/chemo-therapy side effects and infertility for women at childbearing ages. Therefore, it is imperative to have effective risk prediction and prevention strategies to address the growing burden of endometrial cancer.

Identifying high-risk women currently relies on epidemiological factors (see **Table 1.7**) (Pfeiffer *et al*., 2013; Kitson *et al*., 2017; Fortner *et al*., 2017; Alblas *et al*., 2018). While up to half of the endometrial cancers may be attributed to obesity alone, BMI lacks the necessary precision to accurately predict an endometrial cancer diagnosis. This is in part because BMI is an indirect measurement of adiposity and does not account for the muscle/fat mass ratio (Rothman, 2008). Moreover, BMI does not consider bone structure and fat distribution, which may further skew the accuracy. Another reason is that while BMI is associated with the risk of non-endometrioid subtype, this association is not as strong as it is with endometrioid tumours.

Using epidemiological factors, the three published risk prediction models (RPMs) for endometrial cancer show moderate discriminatory power (0.62-0.77) (Alblas *et al*., 2018). T could enhance the discriminatory performance of RPMs. These do not include the important genetic factors in their assessment. It is estimated that SNPs may explain approximately 28% of the familial relative risk of endometrial cancer (O'Mara *et al*., 2018), where with or without the absence of a pathogenic variant in a known endometrial cancer

risk gene, having a first-degree relative with endometrial cancer increases the risk by approximately two-fold (Win *et al.*, 2015).

Over recent years, growing evidence has been put forward in favour of SNPs in the context of endometrial cancer predisposition. The 24 most robust SNPs were identified by our recent systematic review (please refer to **Table 2.1**), where 19 were genome-wide significant and the majority were reported by the ECAC discovery- and meta-GWAS (O'Mara *et al.*, 2018; Bafligil *et al.*, 2020). We noted that the literature was crowded with small candidate-gene studies where all but one candidate SNP reported was later validated in a large GWAS. Moreover, there was a consistent lack of racial/ethnic diversity which limits the applicability of the reported SNP and associations to populations other than White Europeans. The studies also lacked histological subtype discrimination or were restricted to endometrioid subtype only, which limits the assessment of genetic predisposition to non-endometrioid endometrial cancer.

In light of these limitations to previous work, we (C.B) gathered a clinically well-annotated case-control cohort of histologically-confirmed endometrial cancer cases and geographically matched endometrial and breast cancer-free controls from the North West of England (Manchester study). We genotyped the cases using the custom OncoArray chip and obtained genotype data for controls (and 3 cases), which were previously genotyped with the same chip for BCAC. In our Manchester GWAS, we identified further suggestive loci as well as validating the 24 most robust SNPs from our systematic review (**Chapter 2, p 76**) and suggestive SNPs from the largest endometrial cancer consortium, ECAC.

Nearly two thirds of the low-risk susceptibility variants we sought to validate (n=72), comprised of SNPs with genome wide or suggestive significance from our systematic review and ECAC GWAS, showed the same effect size direction in the Manchester GWAS. Just over two fifths of the 24 SNPs from our systematic review were statistically significant

at $P < 0.05$. Though lower than the widely accepted genome-wide significance threshold, this threshold is appropriate for validation purposes and thus, our findings support the importance of these SNPs in endometrial cancer predisposition. Of note, *CYP19A1* SNP rs17601876, which was first reported by candidate-gene studies and later confirmed by large GWAS, was also significant in the Manchester study (P = 1.63E-3). Moreover, the most strongly associated hit reported to date, *HNF1B* SNP rs11263761, was also nominally significant here (P = 1.11E-03).

Further suggestive loci were identified through our Manchester GWAS. Of those, six were included in the PRS (please refer to **Chapter 4, p 160**). rs111282723 is an intergenic SNP near *SETBP1*, which encodes a DNA-binding protein and modulates transcription (Piazza *et al*., 2018). *SETBP1* has also been predicted as a target gene of GWAS hits for breast cancer (Fachal *et al*., 2020). rs9854980 is intronic to *RBMS3*, which encodes an RNA-binding protein belonging to the *c-myc* gene single-strand binding protein family (Wu *et al*., 2020). These genes are implicated in DNA replication, transcription, cell cycle progression and apoptosis. rs12727038 is an intergenic variant near *MAP3K21* and *PCNX2*, both expressed at moderate levels within glandular cells of endometrium (Uhlen *et al*., 2015; Shan *et al*., 2019; Human Protein Atlas, 2020; Uniprot, 2020). The former is a negative regulator of a toll-like receptor involved in activation of innate immune system, whereas the latter may be associated with microsatellite instability (MSI)-high colorectal tumours. rs144065942 is an intergenic variant close to *STEAP3* that encodes metalloreductase, important for mediating downstream responses to p53 (Yu *et al*., 2020). *STEAP3* is expressed at low and moderate levels in stromal and glandular compartments of the endometrium, respectively (Human Protein Atlas, 2020). rs6580584 is a regulatory region variant close to *SH3TC2*, involved in onset of a neurodegenerative disease, and *HTR4*, which encodes a serotonin receptor moderately expressed in glandular endometrial cells, and SNPs in this region were reported to be associated with type 2 diabetes (Jerath *et al*., 2018; Kwon *et al*., 2019; Human Protein Atlas, 2020).

The results explained above were obtained from overall case-control association analyses. Further associations were investigated within the Manchester study by adjusting for BMI in the case-control analyses and restricting the analyses to known endometrioid or non-endometrioid histologies. Due to incomplete data and smaller sample sizes available for these analyses, we have not proceeded to exclude any Manchester GWAS suggestive SNPs identified through the overall analysis. Nevertheless, the BMI-adjusted and subtype-specific analyses provided interesting results.

Accordingly, two SNPs remained suggestively significant after adjusting for BMI, rs12727038 and rs9854980, with the latter increasing in significance to P = 5.59E-8, just short of reaching genome-wide significance (**Table 3.2** and **Figures 3.5** and **3.7**). This is important because the loci this SNP is located at showed several SNPs that were suggestively associated with endometrial cancer, particularly increasing the confidence of this loci having a role in the predisposition to endometrial cancer, despite our relatively small sample size. However, it is important to approach these results with caution. None of these suggestive SNPs were found near genes that are known to be associated with BMI or obesity, and median BMI of our cohort was in line with national figures, and thus, representative of the local population (NHS Digital, 2020; National Cancer Intelligence Network, 2020). Moreover, despite the change in P values, the strength of the effect sizes of these SNPs remained similar with or without adjustment for BMI. Hence, the BMI adjusted results obtained here were used as an additional confirmation of the effect sizes and direction of these SNPs and not utilised for identification or exclusion of any SNP.

We also obtained results for overall and BMI-adjusted analyses for the systematic review SNPs for validation purposes (**Table 3.3**). Aside from two SNPs, the results between the two analyses were similar (rs139584729 and rs882380 had opposite effect size in BMI-adjusted analysis). This was an important step to show that BMI-adjustment did not drastically alter the magnitude of effect. Moreover, the opposite effect size of rs139584729 is likely due to a direct result of its very low EAF, which likely hindered the ability to

accurately assess its association, particularly in the smaller size imposed by restricting the data to those individuals with available BMI data. The other SNP, rs882380 had a quite low effect size of 0.02 (OR=1.02) in the overall analysis, which then completely levelled to 0.00 when adjusted for BMI. This may also be attributed to differing sample sizes between the two analyses, as slight shifts are expected with adjustments and sample size changes.

In the subtype-restricted analyses, two particular SNPs remained suggestively significant (rs6580584 and rs111282723) in the endometrioid-only analysis (**Table 3.2**). These SNPs had substantial effect sizes of -1.02 (OR=0.36) and 1.04 (OR=2.83) per allele, respectively. The small sample size can weaken or strengthen the effect size, especially in the case of particularly uncommon alleles, as may be the case for these SNPs which both have EAF $\leq$ 7%. Nevertheless, it is important that the Manchester GWAS suggestive SNPs showed similar effect sizes in the endometrioid-only results to those obtained from the overall analysis. As in the case of BMI adjusted results mentioned above, relatively smaller sample size, especially in the non-endometrioid subtype analyses, weakens the strength of association leading to higher P values obtained for most of the SNPs in these analyses. It is important to note that, although here we provided results for non-endometrioid versus control analysis, due to insufficient sample size of 136 cases, the results are not expected to reflect true associations and should be interpreted with caution.

The strength of this study is the use of geographically well-matched study population with available clinical data for the cases and a long follow-up data for the controls. Geographical matching of the cases and controls are particularly important because endometrial cancer incidence rates follow local sociodemographic factors such as deprivation and obesity (National Cancer Intelligence Network, 2020; NHS Digital, 2020). Thus, women in both arms of the study are expected to be exposed to similar exogenous risk factors in this locally representative population. Moreover, although the controls were leaner, as expected, and younger, any bias resulting from the latter is minimised with the use of long follow-up period of median >8 years. This is very important as it indicates

that the controls have most likely past the peak ages for receiving an endometrial cancer diagnosis. Those that were diagnosed with endometrial cancer after entry to the PROCAS study were identified and included in our study as cases. Finally, we were able to adjust for BMI and examine its potential confounding effect on our results. Previous GWAS often lacked associated clinical information and thus, were not able to assess the effect of BMI.

48 independent suggestive SNPs were obtained from ECAC (excluding UKBB data) for PRS purposes (see **Chapter 4, p 150**) and were evaluated in our Manchester study (**Table 3.4**). Three fifths of the 48 SNPs shared the same direction of effect with the ECAC data, while three were significant at $P < 0.05$. Some of the SNPs with opposite effect sizes had very uncommon EAFs, while others had an acceptable but relatively lower imputation quality scores. Additionally, these SNPs were suggestively associated in the much larger ECAC cohort comprising of over 12,000 cases, would not be expected to reach significance in our smaller cohort.

In summary, the results of this chapter show six SNPs at suggestive significance (P < 1E-5) for endometrial cancer in an independent and clinically well-phenotyped cohort from North West of England. Moreover, our results validated nearly two thirds of the genome-wide significant and suggestive SNPs identified in our systematic review (**Chapter 2, p 76**) and the largest endometrial cancer GWAS cohort (ECAC) (O'Mara *et al.*, 2018; Bafligil *et al.*, 2020). The genotype data generated from the Manchester GWAS were then used for PRS development for endometrial cancer risk prediction purposes, as detailed in the following chapter.

3.8 References

Alblas, M., Velt, K. B., Pashayan, N., *et al.* (2018). Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. *Critical Reviews in Oncology/Hematology*, 126:92–99.

Amos, C. I., Dennis, J., Wang, Z. M., *et al.* (2017). The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiology Biomarkers & Prevention*, 26(1):126–135.

Auton, A., Abecasis, G. R., Altshuler, D. M., *et al.* (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

Bafligil, C., Thompson, D. J., Lophatananon, A., *et al.* (2020). Association between genetic polymorphisms and endometrial cancer risk: a systematic review. *Journal of Medical Genetics*, 57(9):591–600.

Church, D. M., Schneider, V. A., Graves, T., *et al.* (2011). Modernizing reference genome assemblies. *PLoS Biol*, 9(7):e1001091.

Clarke, M. A., Devesa, S. S., Harvey, S. V., and Wentzensen, N. (2019). Hysterectomy-Corrected Uterine Corpus Cancer Incidence Trends and Differences in Relative Survival Reveal Racial Disparities and Rising Rates of Nonendometrioid Cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, pages JCO1900151–JCO1900151.

Crosbie, E. and Morrison, J. (2014). The emerging epidemic of endometrial cancer: time to take action. *Cochrane Database of Systematic Reviews*, (12).

CRUK (accessed: Oct 2020). Uterine cancer statistics. `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer`.

Delaneau, O., Howie, B., Cox, A. J., *et al.* (2013). Haplotype Estimation Using Sequencing Reads. *The American Journal of Human Genetics*, 93(4):687–696.

Evans, D. G. R., Harkness, E. F., Brentnall, A. R., *et al.* (2019). Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. *Breast Cancer Research and Treatment*, 176(1):141–148.

Fachal, L., Aschard, H., Beesley, J., *et al.* (2020). Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*, 52(1):56–73.

Fortner, R. T., Husing, A., Kuhn, T., *et al.* (2017). Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. *Int J Cancer*, 140(6):1317–1323.

Gupta, K. K., Attri, J. P., Singh, A., *et al.* (2016). Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi journal of anaesthesia*, 10(3):328–331.

Hong, E. P. and Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics Inform*, 10(2):117–122.

Howie, B., Marchini, J., and Stephens, M. (2011). Genotype Imputation with Thousands of Genomes. *G3: Genes | Genomes | Genetics*, 1(6):457.

Human Protein Atlas (Accessed: November 2020).

Jerath, N. U., Mankodi, A., Crawford, T. O., *et al.* (2018). Charcot-Marie-Tooth Disease type 4C: Novel mutations, clinical presentations, and diagnostic challenges. *Muscle & Nerve*, 57(5):749–755.

Johnatty, S. E., Tan, Y. Y., Buchanan, D. D., *et al.* (2017). Family history of cancer predicts endometrial cancer risk independently of Lynch Syndrome: Implications for genetic counselling. *Gynecol Oncol*, 147(2):381–387.

Kitson, S. J., Evans, D. G., and Crosbie, E. J. (2017). Identifying High-Risk Women for Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model. *Cancer Prevention Research*, 10(1):1–13.

Kuo, K. H. M. (2017). Multiple Testing in the Context of Gene Discovery in Sickle Cell Disease Using Genome-Wide Association Studies. *Genomics insights*, 10:1178631017721178–1178631017721178.

Kwon, Y.-J., Hong, K.-W., Park, B. J., and Jung, D.-H. (2019). Serotonin receptor 3B polymorphisms are associated with type 2 diabetes: The Korean Genome and Epidemiology Study. *Diabetes Research and Clinical Practice*, 153:76–85.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., *et al.* (2000). Environmental and heritable factors in the causation of cancer - Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343(2):78–85.

Lu, Y., Ek, W. E., Whiteman, D., *et al.* (2014). Most common 'sporadic' cancers have a significant germline genetic component. *Human Molecular Genetics*, 23(22):6112–6118.

Marchini, J., Howie, B., Myers, S., *et al.* (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39(7):906–13.

Mavaddat, N., Michailidou, K., Dennis, J., *et al.* (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, 104(1):21–34.

Michailidou, K., LindstrÃ¶m, S., Dennis, J., *et al.* (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678):92–94.

Morice, P., Leary, A., Creutzberg, C., *et al.* (2016). Endometrial cancer. *Lancet*, 387(10023):1094–1108.

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., *et al.* (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, 315(1):68–76.

National Cancer Intelligence Network (Accessed: December 2020). Outline of Uterine Cancer in the United Kingdom: Incidence, Mortality and Survival.

NHS Digital (Accessed: December 2020). Health Survey for England 2018: Adult and child overweight and obesity.

O'Mara, T. A., Glubb, D. M., Amant, F., *et al.* (2018). Identification of nine new susceptibility loci for endometrial cancer. *Nature Communications*, 9.

Pfeiffer, R. M., Park, Y., Kreimer, A. R., *et al.* (2013). Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies. *Plos Medicine*, 10(7):15.

Piazza, R., Magistroni, V., Redaelli, S., *et al.* (2018). SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nature Communications*, 9(1):2192.

Rivandi, M., Martens, J. W. M., and Hollestelle, A. (2018). Elucidating the Underlying Functional Mechanisms of Breast Cancer Susceptibility Through Post-GWAS Analyses. *Frontiers in genetics*, 9:280–280.

Rothman, K. J. (2008). BMI-related errors in the measurement of obesity. *International Journal of Obesity*, 32(3):S56–S59.

Ryan, N. A. J., McMahon, R., Tobi, S., *et al.* (2020). The proportion of endometrial tumours associated with Lynch syndrome (PETALS): A prospective cross-sectional study. *PLOS Medicine*, 17(9):e1003263.

Shan, J. X., Chouchane, A., Mokrab, Y., *et al.* (2019). Genetic Variation in CCL5 Signaling Genes and Triple Negative Breast Cancer: Susceptibility and Prognosis Implications. *Frontiers in Oncology*, 9:13.

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1):7–30.

Spurdle, A. B., Bowman, M. A., Shamsani, J., and Kirk, J. (2017). Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol*, 30(8):1048–1068.

Spurdle, A. B., Thompson, D. J., Ahmed, S., *et al.* (2011). Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nature Genetics*, 43(5):451–454.

Sundar, S., Balega, J., Crosbie, E., *et al.* (2017). BGCS uterine cancer guidelines: Recommendations for practice. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 213:71–97.

Uhlen, M., Fagerberg, L., Hallstrom, B. M., *et al.* (2015). Tissue-based map of the human proteome. *Science*, 347(6220):1260419.

Uniprot (Accessed: November 2020).

Win, A. K., Reece, J. C., and Ryan, S. (2015). Family History and Risk of Endometrial Cancer A Systematic Review and Meta-analysis. *Obstetrics and Gynecology*, 125(1):89–98.

Wu, Y. L., Meng, D. L., You, Y. X., *et al.* (2020). Increased expression of RBMS3 predicts a favorable prognosis in human gallbladder carcinoma. *Oncology Reports*, 44(1):55–68.

Yu, Z., Wang, H., Fang, Y., *et al.* (2020). Molecular chaperone HspB2 inhibited pancreatic cancer cell proliferation via activating p53 downstream gene RPRM, BAI1, and TSAP6. *Journal of Cellular Biochemistry*, 121(3):2318–2329.

Zhang, Y. D., Hurson, A. N., Zhang, H. Y., *et al.* (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature Communications*, 11(1):13.

4 Development and evaluation of polygenic risk scores for prediction of endometrial cancer risk in European women

This chapter forms part of a research article that is being prepared for submission in a high-impact peer-reviewed journal. Briefly, this chapter explains the development and evaluation of PRSs in the Manchester study, ECAC and UKBB datasets. The methods and results of the Manchester GWAS used in some parts are detailed in the previous chapter (**Chapter 3, p 105**).

4.1 Contributorship Statement

I performed the laboratory work (DNA quantification, purification and normalisation, and some of the DNA extractions), genotype analyses, QC and GWAS, data extraction, statistical analysis, PRS calculations and drafted the manuscript. D.J.T and J.D supported data extraction, genotype analyses and GWAS. A.L, K.M, T.OM and E.J.C supported data extraction. A.L and K.M supported data analysis. Most of the DNA extractions were performed previously by either MGDL or Hologic Ltd.

4.2 Abstract

Background Endometrial cancer is the most common gynaecological cancer in the UK. Whilst Lynch syndrome is the main heritable cause, multiple single nucleotide polymorphisms (SNPs) are known to influence genetic predisposition to endometrial cancer. We hypothesized that a polygenic risk score (PRS) derived from a panel of SNPs may be a useful component of an endometrial cancer risk prediction tool for targeted screening and prevention strategies.

Methods We developed four SNP panels derived from our recent systematic review of published studies and further suggestive SNPs from the Endometrial Cancer Association Consortium (ECAC) and an independent GWAS of surgically confirmed endometrial cancer cases and cancer-free controls from the North West of England. The PRS was then refined to only contain SNPs that showed favourable contribution to the PRS performance. All SNP panels were then evaluated in the North West, ECAC and UK Biobank (UKBB) cohorts.

Results The PRS was tested in 555 cases and 1,202 controls from the North West of England. The AUC from 19, 24, 72 and 78 SNP panels were 0.58, 0.55, 0.57 and 0.66, respectively. The refined PRS resulted in a 40 SNP-panel achieving an AUC of 0.62. In this PRS (PRS40), women in the third and second tertiles respectively had 2.4-fold and 1.3-fold increased risk of endometrial cancer compared to women in the first tertile (OR: 2.43, 95% CI 1.89-3.14; OR: 1.27, 95% CI 0.96-1.68; P_{trend} = 9.89E-13). The discriminatory performance of the SNP panels was lower in UKBB and similar in ECAC.

Conclusions In this study, we report an endometrial cancer PRS which was moderately capable of predicting endometrial cancer in a small case-control study, however, external validation is warranted. A PRS with a consistent and high predictive performance will support targeted screening and prevention interventions to predict those at greatest risk of endometrial cancer.

4.3 Introduction

Endometrial cancer is the most common gynaecological malignancy in the United Kingdom, with incidence and death rates rising steadily (Crosbie and Morrison, 2014). Despite an overall five-year survival rate approaching 80%, there is a marked discrepancy in survival between women diagnosed with early (>90%) and late stage (16%) disease (Morice *et al*., 2016; Siegel *et al*., 2020). Surgical treatment may be hazardous, particularly in elderly and obese women, and removes the opportunity for childbearing in younger women. There is therefore an urgent need for effective early detection and prevention strategies to tackle this growing disease burden (Kitson *et al*., 2017; O'Mara and Crosbie, 2020).

Identifying women at greatest risk of endometrial cancer will maximise the benefits and minimise the harms of targeted screening and prevention interventions. Whilst excess adiposity is a major risk factor for endometrial cancer (Crosbie *et al*., 2010), BMI alone is insufficient for accurate risk prediction; not all tumours are obesity-driven (Lu and Broaddus, 2020) and women with class III obesity (BMI >40kg/m^2) have a relatively modest lifetime risk of endometrial cancer of 10-15% (Kitson and Crosbie, 2019). It is clear that endometrial cancer risk is determined by both environmental and genetic influences (Raglan *et al*., 2019). The strongest genetic factors are pathogenic variants affecting the MMR genes, which cause Lynch Syndrome (LS) and a 40-60% lifetime risk of endometrial cancer (Ryan *et al*., 2017; Dominguez-Valentin *et al*., 2020). Rare pathogenic variants in other DNA repair-related genes, including *PTEN* (with or without CS), *POLE* and possibly *BRCA1/2* are also associated with increased risk (Constantinou and Tischkowitz, 2017; Spurdle *et al*., 2017). Where specific genetic variants are not identified, women with a first-degree relative with endometrial cancer have a two-fold higher risk of the disease (Win *et al*., 2015).

Indeed, it has been estimated that approximately 28% of the familial relative risk

of endometrial cancer is attributable to common SNPs (O'Mara *et al*., 2018). GWAS, including those by members of our group, have identified 16 common endometrial cancer susceptibility regions (O'Mara *et al*., 2018, 2019; Bafligil *et al*., 2020). Although the risk alleles for each SNP influence endometrial cancer risk by a small amount (9%-40% per allele), a PRS (i.e., the total number of risk alleles carried by an individual across all SNPs, weighted by each SNP's risk estimate) could be used to distinguish women at highest and lowest genetic risk of endometrial cancer. The potential of a PRS-based risk prediction model to rationalise screening algorithms and eligibility for chemoprevention strategies is already well established in breast cancer (Mavaddat *et al*., 2015).

The aim of this study was to develop and test a PRS for endometrial cancer risk prediction in European-descent populations, using SNPs identified from three sources: a new cohort of carefully phenotyped cases and controls from Manchester in the North West of England (Evans *et al*., 2019; Ryan *et al*., 2020), the ECAC GWAS (O'Mara *et al*., 2018), and a systematic review of literature (Bafligil *et al*., 2020).

4.4 Materials and Methods

4.4.1 Study Populations

4.4.1.1 Manchester Study

Women treated for endometrial cancer at Manchester University NHS Foundation Trust (MFT), who donated clinico-pathological data and a blood sample for future research, were the Manchester cases (see **Section 3.4.1**) (Ryan *et al*., 2020). Baseline clinical data included age, ethnicity, BMI (kg/m^2), histological subtype (endometrioid, serous, clear cell, carcinosarcoma), FIGO (2009) stage and grade. All pathology specimens were reviewed by at least two specialist gynaecological pathologists using confirmatory immunohistochemistry as necessary. We excluded cases where final pathology review indicated cancer of non-endometrial origin and women of non-European ancestry, due to small numbers (**Figure 3.3**).

Women of European descent participating in a local general population breast cancer screening study with (i) no personal history of endometrial or breast cancer, (ii) no endometrial or breast cancer diagnosis during follow up (median 8.86 years, interquartile range (IQR) 8.10-9.60, median age at censor 68.12, IQR=62.25-73.49), and (iii) an intact uterus, were the Manchester controls (Evans *et al.*, 2019) (**Figure 3.3**). Both studies were sponsored by the University of Manchester, approved by research ethics committees (**Table 3.1**) and conducted in accordance with Good Clinical Practice guidelines.

4.4.1.2 Endometrial Cancer Association Consortium

Details of ECAC study were published previously (O'Mara *et al.*, 2018). In summary, 40 datasets were compiled wherein all subjects were of European ancestry. This included 12,906 endometrial cancer cases and 108,979 controls, from both hospital- and population-based studies. For our study, a subset of ECAC without UKBB samples, and where individual level genotyping data were available (except Epidemiology of Endometrial Cancer Consortium and Womens' Health Initiative), was obtained. In total, 9,062 endometrial cancer cases and 41,461 controls were available for the analyses.

4.4.1.3 UK Biobank

The UKBB is an openly accessible prospective study that holds extensive genetic and phenotypic data on nearly half a million participants aged between 40 and 69, recruited from across the UK (Bycroft *et al.*, 2018). The participants are still being followed up to this day, allowing for extensive information to be collected over the years. The UKBB study was approved by the North West Centre for Research Ethics Committee (11/NW/0382). Details of sample collection were previously described by Bycroft *et al.* (2018).

4.4.2 Genotyping

Detailed explanations of the specimen preparation and genotyping processes for the Manchester study are given in **Chapter 3**. Briefly, genomic DNA was extracted from

peripheral blood and saliva for the Manchester cases and controls, respectively. DNA from case specimens was extracted using either Nucleon extraction kit (Cat no. SL8502, Gen-Probe Life Sciences Ltd) or Gentra Puregene Blood Kit (Cat no. 158389, Qiagen). DNA extracts were resuspended in standard TE buffer and quantified by NanoDrop Spectrophotometer (ThermoFisher). Samples with very high or low concentrations were treated with Genomic Clean and Concentrator-10 (Cat no. D4011, Zymo Research) according to manufacturer's protocol to normalise the concentrations to approximately 100 ng/μL. DNA from the controls was collected and extracted from saliva using Oragene kit (DNA Genotek Inc., Ottawa, ON, Canada) according to the manufacturer's protocols. All samples were stored at -80°C until genotyping. 10 μL of DNA per sample was then dispensed into full-skirted ABGene 96-well SuperPlates (Cat no. AB-2800, Thermo Scientific) prior to genotyping.

The genotyping platforms used in this study are summarised in **Table 4.1**. Manchester samples were genotyped using OncoArray 534K, custom designed by the OncoArray Consortium to include ∼250,000 GWAS backbone SNPs and ∼250,000 SNPs with previously known associations to five common cancers (breast, ovarian, prostate, colorectal and lung) (Amos *et al.*, 2017). Genotyping processes for both ECAC and UKBB were published previously (Bycroft *et al.*, 2017; O'Mara *et al.*, 2018). Briefly, individual studies in ECAC were genotyped using several genotyping arrays including OncoArray 534K and Illumina Human OmniExpress arrays.

Genotyping data for most UKBB participants is available to researchers. The majority of genotyping was performed using *UK Biobank Axiom Array* and around 50,000 samples were genotyped using *UK BiLEVE Axiom Array* (Affymetrix, currently a part of ThermoFisher Scientific) (Bycroft *et al.*, 2018). A total of 805,426 SNPs are available as genotyped data. The markers were positioned to Genome Reference Consortium human build 37 (GRCh37) assembly. QC was performed by UKBB team and the data was phased and imputed to a combination of Haplotype Reference Consortium (HRC), 1000 Genomes

Project v3 and UK10K haplotype panels SHAPEIT3 and IMPUTEv4 (Bycroft *et al.*, 2018). The imputation resulted in obtaining nearly 96 million SNPs.

Table 4.1: *Details of the three datasets used in the study.*

Study	Source of Samples	Case (n)	Control (n)	Genotyping Array	Imputation Panel(s)
Development Set	Manchester Cohort	555	1,202	Illumina OncoArray 534K	1000 Genomes v3
Validation Set I	ECAC[1]	9,062	41,461	Illumina660WQuads, Illumina Hap550, Illumina 610K, Illumina 1.2M, Illumina Infinium iSelect[2], Illumina OncoArray 534K	1000 Genomes v3, UK10K
Validation Set II	UKBB	1,676	116,960	Affymetrix UK BiLEVE Axiom, Affymetrix UK Biobank Axiom	1000 Genomes v3, UK10K, HRC

[1] ECAC data used here do not include UKBB cases or controls.
[2] iCOGS.
HRC, Haplotype Reference Consortium.

4.4.3 Data Extraction and Processing

4.4.3.1 Manchester Study

The Manchester GWAS study QC and analyses are detailed in the previous chapter (see **Sections 3.5.2** and **3.6.3**). In brief, genotype calling was performed at University of Cambridge for both Manchester cases and controls. Genotype analyses were carried out using GenABEL package (GenABEL, 2013) in R version 3.6.0 (R Core Team, 2020) according to OncoArray Consortium guidelines (Amos *et al.*, 2017). SNP-wise QC was conducted to exclude SNPs with call-rate <95% and deviating from HWE (P < 10E-7 in controls and P < 10E-12 in cases). Sample-wise QC was conducted to exclude samples with call-rate <95%, low or high heterozygosity and genetically XO, XXY and XY individuals. One pair of each duplicate samples or monozygotic twins and first-degree relatives were excluded using genomic kinship matrices, and non-European individuals were removed from further analyses based on known HapMap populations.

All samples were imputed using the v3 of 1000 Genomes Project reference panel (Auton *et al.*, 2015). Samples were phased with SHAPEITv2 (Zagury and Marchini, 2020) and genotypes were imputed with IMPUTEv2 (Howie and Marchini, 2020) for non-overlapping

5Mb intervals. The resulting data were restricted to SNPs with MAF≥0.1% and r^2>0.4 (imputation score), leaving ∼13 million markers. SNPs with LD r^2>0.2 were excluded from the PRS analyses. Three SNPs (rs9460655, rs113945442 and rs2305252) failed QC in the UKBB cohort and were excluded from the PRS (please refer to **Section 4.4.3.3**).

4.4.3.2 Endometrial Cancer Association Consortium

Individual level genotyping data were obtained from ECAC except Epidemiology of Endometrial Cancer Consortium and Womens' Health Initiative where the data were not available, and UKBB to avoid use of duplicated data (T O'Mara). GWAS results from ECAC (excluding UKBB samples) were also obtained through personal communication (T O'Mara). The SNPs reaching the suggestive threshold of P < 1E-5 were selected and any SNP that was already included in the SNP panel, or was in LD with other SNPs (r^2>0.2) were excluded. Data for chromosome X was not available. No SNP in this dataset had an overall imputation score less than 0.4, similar to Manchester GWAS.

4.4.3.3 UK Biobank

The full variable UKBB dataset including 502,536 individuals was obtained through the Centre for Integrated Genomic Medical Research (CIGMR). Endometrial cancer cases were selected using ICD10 (C55 and C54), ICD9 (179 and 182) and self-reported (1040) coding system (Bycroft *et al.*, 2017). This included 2,478 endometrial cancer cases, the majority of which were confirmed cases by the cancer registry. 217,422 female controls without any malignant cancers were also obtained.

All endometrial cancer cases were identified, and categorised as incident or prevalent cases (Groups 1 and 2) (**Table 4.2**). Healthy individuals without other malignancies were categorised into Group 3. Individuals with cancer diagnoses other than endometrial cancer were assigned to a Group 4. Endometrial cancer cases with missing information (prevalent or incident) were also noted (Group 5). All benign cases were assigned to Group 6. Then, self-reported cases were identified and assigned to appropriate groups

created earlier. Individuals with self-reported sex corresponding to males were removed from the whole dataset. Prior to genotype data acquisition, participants in the control arm who had hysterectomy at any point, diagnosed with any malignancy (excluding melanoma), and died were excluded for the purpose of this study, leaving 179,414 women. Assignment of cases and controls was performed using STATA 16 (StataCorp, 2019).

Table 4.2: *Distribution of endometrial cancer cases and controls within the UKBB dataset.*

The numbers represent the remaining set after exclusion of males.

Group	Definition	n
1	Incident endometrial cancer	1,065
2	Prevalent endometrial cancer	1,411
3	Controls without other malignant cancers	216,859
4	Controls with other malignant cancers	53,502
5	Endometrial cancer with missing data	2
6	Controls with neoplasms or benign tumours	563
	Total	273,402

A list of SNPs including the 24-SNP panel and a further suggestive SNPs from ECAC data (excluding UKBB) was created. The SNPs-of-interest were then extracted from the imputed UKBB .bgen files, hosted at CIGMR using PLINK2 (Purcell and Chang, 2020). Individuals who withdrew their consent (as of August 2020) were excluded and only females with matching self-reported and genetically inferred sex were included. In this cohort, the level of genetic data missingness tolerated was $\leq 2\%$, no chromosome aneuploidy was present and principal component analysis (PCA)-corrected heterozygosity was between ± 3 SD from the mean. Moreover, the individuals included in this study belonged to the genetic group of Europeans and had no relative within the cohort.

Following exclusions and matching with available genotype data, 1,676 cases and 116,960 controls were available for the analyses. Standard SNP-wise QC was performed using PLINK1.90 (Purcell and Chang, 2017) to assess deviation from HWE (P < 1E-10 in cases and P < 1E-6 in controls), SNP-wise missingness (>2%) and MAF.

Three SNPs were excluded from further analysis for failing QC; rs9460655 for deviating from HWE (P = 2.33E-78 in controls), high missingness at >57%, and , low imputation score at 0.58; rs113945442 for deviating from HWE (P = 2.9E-15 in controls); and rs2305252 for deviating from HWE (P = 2.14E-8 in controls). The lowest imputation score used for this study was 0.85.

4.4.4 Statistical Analyses

4.4.4.1 Population Structure and Post-Imputation QC

Population stratification within the ECAC GWAS was adjusted by using nine PCs; estimated from 33,661 uncorrelated SNPs with MAF>0.05 and pairwise r^2<0.1 using purpose-written software (`http://ccge.medschl.cam.ac.uk/software/pccalc`) (O'Mara *et al.*, 2018). The statistic inflation adjusted to a sample size of 1,000 cases and equal controls (λ_{1000}), using 33,301 uncorrelated SNPs (r^2<0.1), was assessed to be 1.004 in the latest ECAC GWAS, indicating no evidence of substantial population stratification. Same methods were followed for the Manchester GWAS, which are detailed in the previous chapter (**Chapter 3, p 125**), while the details for the other studies identified through the systematic review are summarised in **Table 2.1, p 84** (Bafligil *et al.*, 2020).

4.4.4.2 Descriptive Statistics

Descriptive statistics for participants' age and BMI were calculated within and between the three studies. For ECAC, only a partial assessment was carried out within its cases and controls due to limited availability of data. Student's t-test was carried out to assess any significant differences for age and BMI between cases and controls, which uses the following formula:

$$t = \frac{(\overline{x} - \overline{y})}{\hat{\sigma}\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{4.1}$$

where $\overline{x}$and $\overline{y}$ are the sample means ($\overline{x} = \frac{\sum x}{n}$) of groups x and y, n is the number of samples per group, while $\hat{\sigma}$ is the pooled variance estimate ($\hat{\sigma} = \frac{\sum x^2}{n} - \overline{x}^2$) .

Impact of BMI on PRS was assessed by comparing the means of controls in different BMI groups using one-way ANOVA and Tukey's tests.

Statistical analyses were conducted in R version 3.6.0 (R Core Team, 2020). All tests were two-sided and $P < 0.05$ was accepted as statistically significant, unless stated otherwise.

4.4.4.3 PRS Performance

In order to assess the performance of the PRS, an area under the receiver-operator curve (AUC) was plotted. This was done by plotting sensitivity (true positive rate); calculated by:

$$\text{Sensitivity} = n_{\text{True Positive}}/(n_{\text{True Positive}} + n_{\text{False Negative}}) \tag{4.2}$$

against false positive rate; calculated by

$$\text{False Positive Rate} = 1 - Specificity \tag{4.3}$$

where specificity is

$$n_{\text{True Negative}}/(n_{\text{True Negative}} + n_{\text{False Positive}}) \tag{4.4}$$

4.4.5 Development of the Extended PRS

SNPs used in the PRS were identified through three approaches: a) a systematic review of the literature (Bafligil *et al.*, 2020) and independent suggestive SNPs from b) ECAC GWAS (excluding UKBB-derived ECAC samples) and c) the Manchester GWAS. A suggestive SNP was defined as any SNP with $P < 1E\text{-}5$ but not genome-wide significant, that is $P < 5E\text{-}8$. For SNPs identified through the Manchester GWAS, suggestive SNPs from loci with multiple SNPs in LD were selected over single hits. The SNP selection from all sources involved exclusion of SNPs showing evidence of LD ($r^2>0.2$).

The PRS was derived using the formula:

$$\text{PRS} = \beta_{\text{k}}\chi_{\text{k}} + ... + \beta_{\text{n}}\chi_{\text{n}} \tag{4.5}$$

where β_{k} is the per-allele log OR for SNP k, χ_{k} is the allele dosage for SNP k, and n is the total number of SNPs included in the PRS.

Dosages for the whole SNP panel were obtained from imputation output files (Manchester and UKBB studies by C Bafligil, ECAC study by T O'Mara). Per-allele logORs (betas) used to calculate the PRS were obtained from the discovery studies, either through the systematic review (Bafligil *et al*., 2020) (n=6, excluding the ECAC SNPs), by personal communication (T O'Mara) from ECAC GWAS excluding UKBB data (n=66), or from Manchester GWAS (n=6) (**Chapter 3, p 125**). Details of the SNPs and the panels are summarised in **Supplementary Table S4.8, p187**.

A PRS was developed first using genome-wide significant SNPs (PRS19) from the systematic review as the base score, and then expanded by adding the remaining five SNPs identified from the literature (PRS24), and suggestive SNPs from ECAC (PRS72) and the Manchester GWAS (PRS78), 48 and six SNPs, respectively. The PRS was fitted as a continuous variable in a logistic regression to calculate AUC for assessment of the goodness of fit (Sing *et al*., 2005; Robin *et al*., 2011). We then divided the model into tertiles based on the PRS in controls and calculated ORs of the second (33.3>x<66.6) and third tertile (x>66.6), in comparison to first tertile (x<33.3). OR of having endometrial cancer in third or second tertiles versus first tertile was then calculated in R using *fmsb* package (Nakazawa, 2019). The PRS plots were created with *ggplot2* package in R (Wickham, 2016).

Initially, four PRSs were developed, two of which were based on our systematic review. PRS19 included the 19 genome-wide significant SNPs, including the *AKT1* SNP (discussed in **Chapter 2, p 89**); PRS24 comprised all 24 SNPs identified by the systematic review;

PRS72 included PRS24 and 48 suggestive SNPs from ECAC GWAS; and finally, PRS78 included PRS72 and the six suggestive SNPs identified in Manchester GWAS. Stepwise addition of SNPs was performed to investigate PRS performance based on effects of individual SNPs excluding the six Manchester SNPs. Finally, performance of the PRSs in Manchester cohort were investigated in known endometrioid and non-endometrioid subtype cases versus controls, individually.

4.4.6 Validation of the PRS

The four PRSs (PRS19, 24, 72 and 78) as well as the refined PRS determined by stepwise addition in the Manchester cohort was then applied to UKBB for independent, population-level validation. Due to inherent differences between study design and participant demographics between the Manchester and UKBB cohorts, we also sought to validate the PRSs in the largest dataset available, ECAC. Limitations for the use of both the UKBB and ECAC datasets will be discussed later.

4.5 Results

4.5.1 Development of the PRS in Manchester Cohort

4.5.1.1 Characteristics of the Study Population

After exclusions, the Manchester cohort included 555 cases and 1,202 female controls with a median age and BMI of 64 years and 30.8 kg/m^2 and 59 years and 25.45 kg/m^2, respectively **(Table 4.3)**. Most cases had low grade (64.5%), early stage (84.3%) endometrial cancer of endometrioid histological subtype (75.3%), consistent with national figures (CRUK, 2020). There was a statistically significant difference in age between cases and controls (P = 1.12E-12). Similarly, women with endometrial cancer had a higher BMI than control women with respective mean BMI of 32.7 kg/m^2 versus 26.5 kg/m^2 (P = 1.08E-36). More cases were also classified as overweight or obese in the cases than controls (78.8% versus 56.2, respectively).

Table 4.3: *Baseline characteristics for the women included in Manchester cohort.*

Missing data was primarily due to retrospectively identified cases from the FH-Risk study and Clinical Genetics service.

Characteristics	Category	Cases (n)	Controls (n)
BMI (kg/m^2)	<18.5	6	11
	18.5 <25	100	489
	25 <30	120	405
	30 <35	116	157
	35 <40	68	56
	$\geq$ 40	91	24
	Unknown	54	60
Age (years)	Mean	63.23	59
	<50	66	85
	50 <60	128	518
	60 <70	160	530
	$\geq$ 70	180	69
	Unknown	21	0
Tumour Grade	1	211	
	2	109	
	3	176	
	Unknown	59	
Stage	I	364	
	II	55	
	III	69	
	IV	9	
	Unknown	58	
Subtype	Endometrioid	415	
	Non-endometrioid	136	

4.5.1.2 Polygenic Risk Score

The base PRS (PRS19) achieved an AUC of 0.58 (95% CI 0.56-0.61), while the full 24-SNP systematic review panel (PRS24) achieved an AUC of 0.55 (95% CI 0.52-0.58) in the Manchester cohort (**Figures 4.1 and 4.2, Table 4.4**). PRS72 achieved a slightly better AUC (95% CI) of 0.57 (0.54-0.60) while PRS78 performed markedly better at 0.66 (95% CI 0.63-0.69). As not all 72 SNPs (excluding the six internally discovered suggestive SNPs) showed the same direction of effect between the discovery and Manchester cohorts, we tested stepwise addition of SNPs to the PRS to refine and improve the PRS performance. Using this approach, the best performing and refined PRS comprised 40 SNPs (PRS40) and achieved an AUC (95% CI) of 0.62 (0.59-0.65).

Table 4.4: *Details of the PRSs developed within the Manchester cohort.*

PRS Name	SNP Source	SNP *P* Value Cut-off	SNP (n)	Third vs First Tertile; OR (95% CI)	Second vs First Tertile; OR (95% CI)	P_{trend}	AUC (95% CI)
PRS19	Systematic review (genome-wide significant SNPs)	<5E-8	19	2.08 (1.61-2.68)	1.29 (0.99-1.70)	5.75E-9	0.58 (0.56-0.61)
PRS24	Systematic review (all)	<1E-5	24	1.51 (1.18-1.94)	1.24 (0.96-1.60)	1.11E-3	0.55 (0.52-0.58)
PRS40	Stepwise addition (excl. internal SNPs)	<1E-5	40	2.43 (1.89-3.14)	1.27 (0.96-1.68)	9.89E-13	0.62 (0.59-0.65)
PRS72	Systematic review and ECAC GWAS	<1E-5	72	1.67 (1.31-2.14)	1.11 (0.85-1.44)	2.55E-5	0.57 (0.54-0.60)
PRS78	Systematic review, ECAC and Manchester GWAS	<1E-5	78	3.39 (2.59-4.45)	1.65 (1.23-2.21)	2.22E-20	0.66 (0.63-0.69)

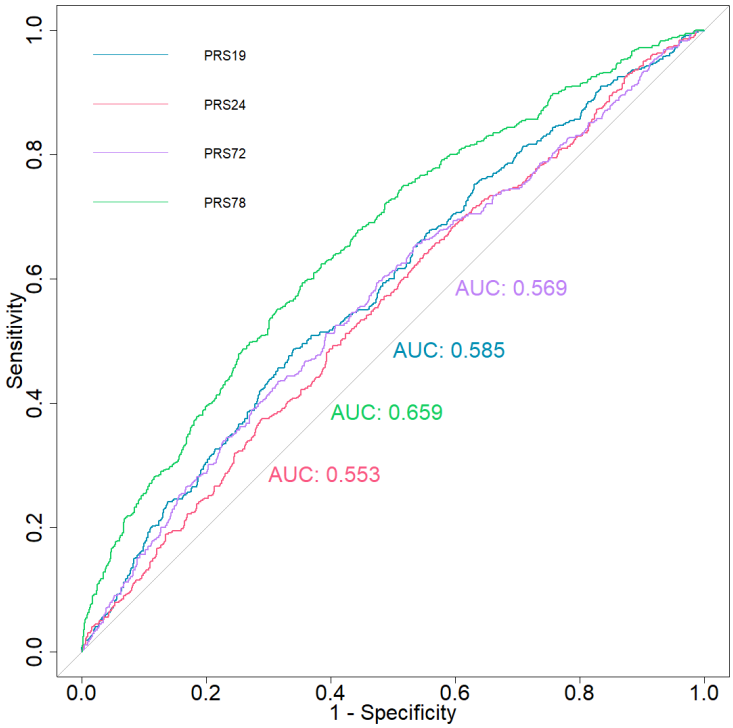


Figure 4.1: *AUCs obtained for PRS19, PRS24, PRS72 and PRS78 in Manchester cohort.*

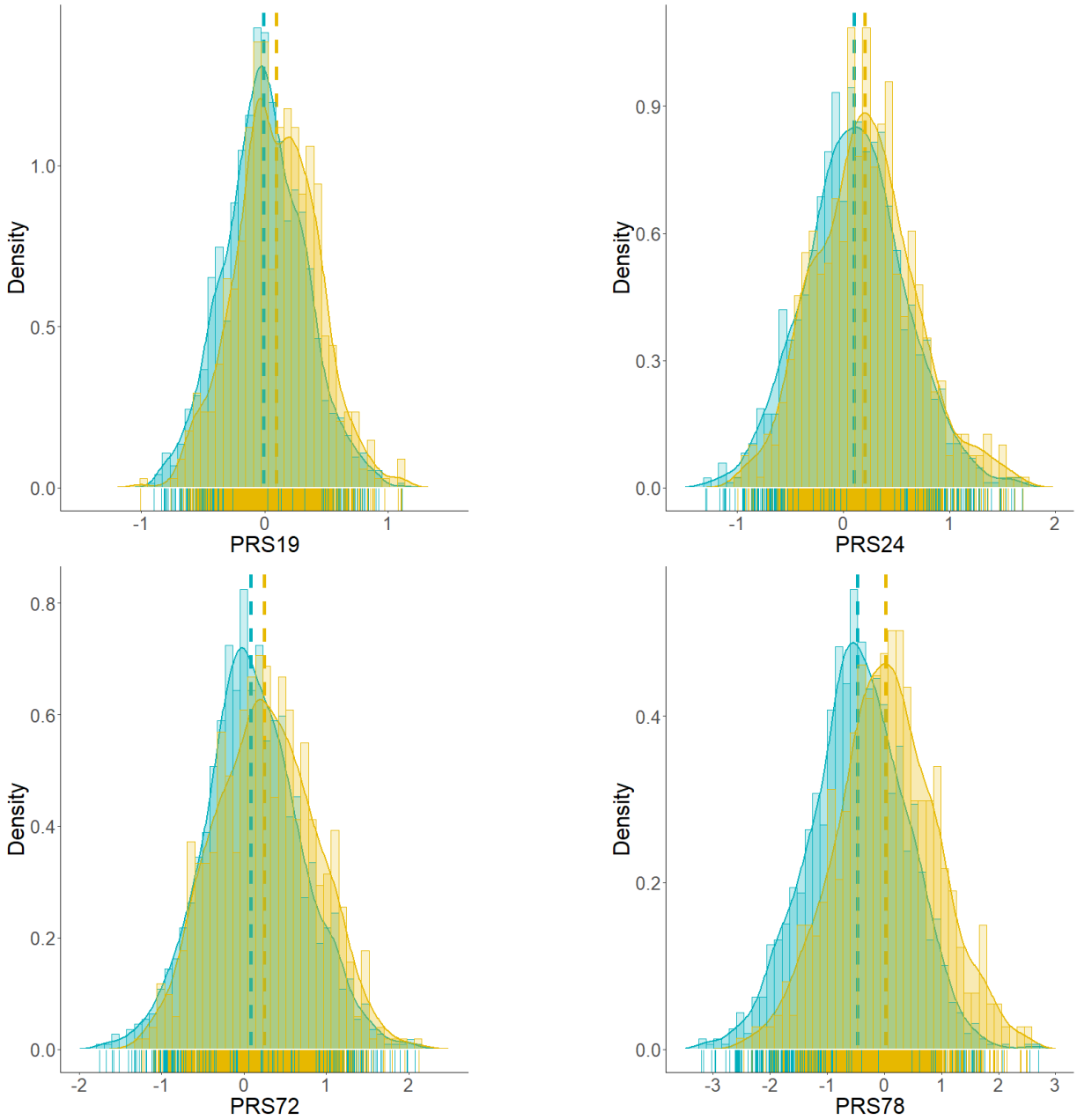
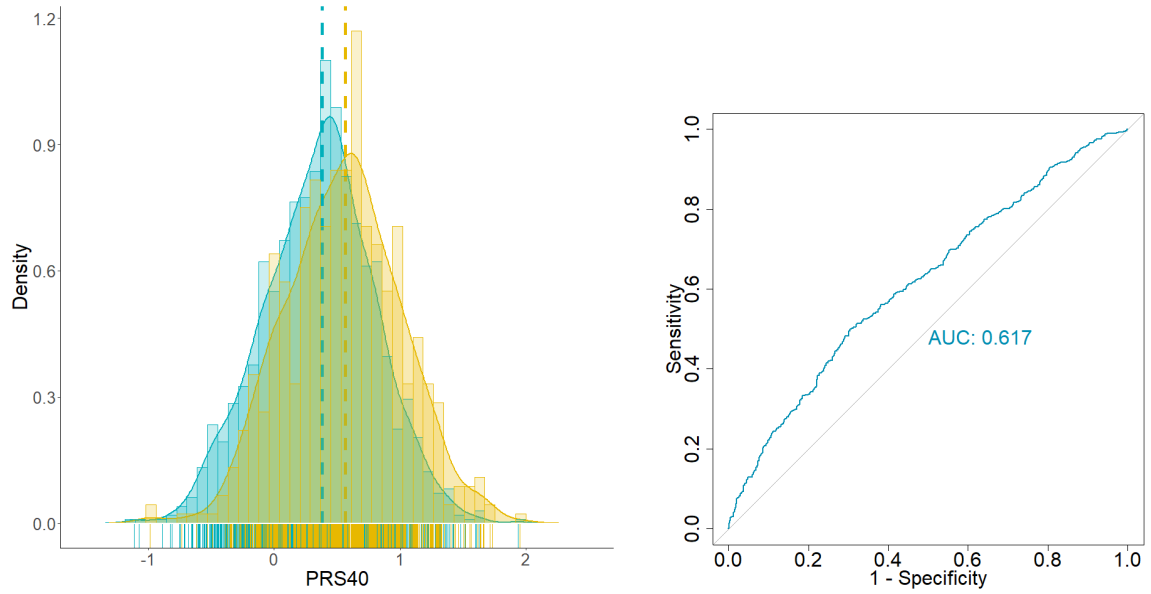


Figure 4.2: *PRS performances in Manchester cases and controls. Distribution of PRS19, PRS24, PRS72 and PRS78 is presented in each corresponding plot.*

The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.



(a) (b)

Figure 4.3: *The best performing PRS (PRS40) achieved in Manchester cohort after refining the SNP panel.*

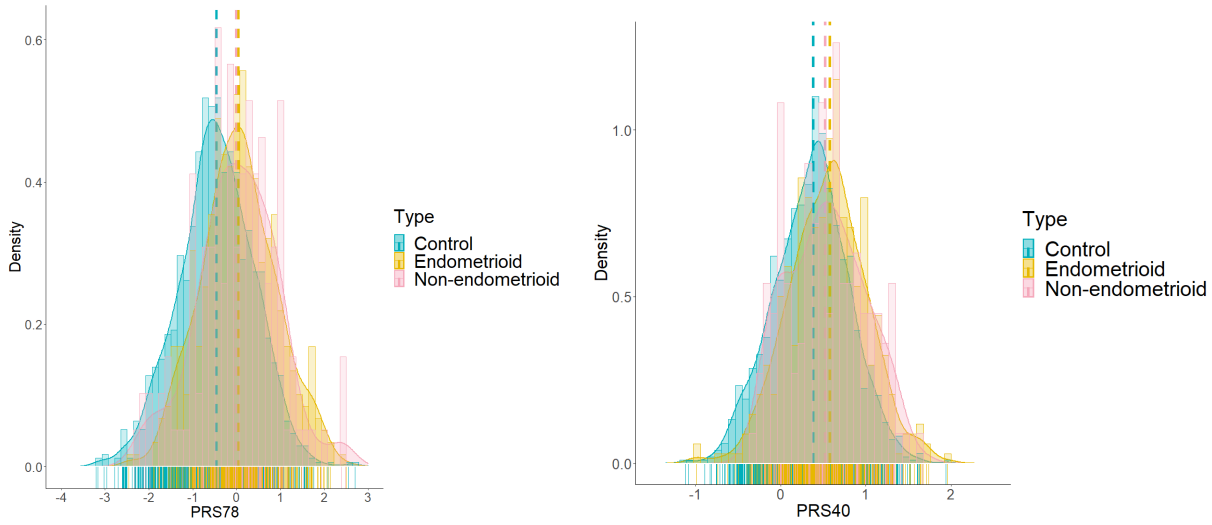
(a) Density and (b) AUC plots of PRS40. (a) The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.

According to the results, in all PRS constitutions, women in the third tertile of the PRS were at a significantly higher risk of endometrial cancer compared to those in the first tertile (**Table 4.4**). This risk was less pronounced in those ranking in the second tertile, though in PRS78 this was still statistically significant. The highest risk seen was with PRS78 and PRS40 where women in third tertile were at 3.39- and 2.43-fold increased risk of having endometrial cancer, respectively, compared to those in the first tertile.

Impact of BMI on the PRS score (PRS40) was investigated in the Manchester cohort. Women in the control group were assigned into one of the three BMI categories; i) underweight and normal (BMI up to 25 kg/m^2), ii) overweight (BMI between 25 and 30 kg/m^2), and iii) obese and extremely obese (BMI over 30 kg/m^2). Neither one-way ANOVA nor Tukey's test showed a statistically significant difference of the genetic risk between the three BMI categories ($P > 0.5$ for all comparisons).

4.5.1.3 Polygenic Risk Score Stratifies Risk in Endometrial Cancer Subtypes Similarly

Predicting the risk of both the two major subtypes of endometrial cancer, jointly or exclusively, particularly that of the more aggressive non-endometrioid histotype, is essential. Therefore, the performance of the PRSs in women with endometrioid (n=415) and non-endometrioid (n=136) tumours was examined separately. Among all PRS panels, the PRSs predicted the risk similarly in both subtypes, with the best discrimination obtained with PRS78 achieving AUCs (95% CIs) of 0.66 (0.63-0.69) and 0.65 (0.60-0.70) in endometrioid and non-endometrioid versus controls, respectively (**Figure 4.4 a**). The refined PRS, i.e., PRS40, also showed similar discriminatory ability in both of the histotypes (**Figure 4.4 b**). Accordingly, the AUC (95% CI) for endometrioid-only cases versus controls was 0.62 (0.59-0.65) while the corresponding value achieved for non-endometrioid-only cases versus controls was 0.61 (0.55-0.66).



(a) (b)

Figure 4.4: *Performance of (a) PRS78 and (b) PRS40 in Manchester endometrioid and non-endometrioid cases versus controls.*

The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis.

4.5.2 Population-Based Validation of the PRS in UK Biobank

4.5.2.1 Characteristics of the Study Population

The UKBB validation cohort comprised 1,676 cases and 116,960 female controls. The median age and BMI of UKBB cohort were 61 years and 28.3 kg/m^2 and 56 years and 25.8 kg/m^2 for cases and controls, respectively. The baseline characteristics of the study participants are given in **Table 4.5**.

Table 4.5: *Participant characteristics within the UKBB cohort used in the PRS analyses*.

Characteristics	Category	Cases (n)	Controls (n)
BMI (kg/m^2)	<18.5	7	897
	18.5 <25	425	48537
	25 <30	558	42047
	30 <35	354	16906
	35 <40	179	5728
	$\geq$ 40	142	2543
	Unknown	11	302
Age (years)	Mean	61	55
	<50	89	32239
	50-59	451	41607
	60-69	1118	42751
	$\geq$70	18	363
	Unknown	0	0

4.5.2.2 Polygenic Risk Score Validation

The AUC (95% CI) achieved for PRS19 and PRS24 was 0.56 (0.54-0.57) and 0.53 (0.52-0.54), respectively, in the UKBB cohort (**Figure 4.6**, **Table 4.6**). The corresponding values obtained for PRS72 and PRS78 were 0.54 (0.52-0.55) and 0.52 (0.51-0.54), respectively. Comparison of the AUC plots for the four PRSs is presented in **Figure 4.5**. The best performing PRS that was achieved in the Manchester cohort, as described above, was applied to the UKBB cases and controls as well for validation. The performance achieved in UKBB for PRS40 was rather low with an AUC of 0.54 (95% CI 0.53-0.55) (**Figure 4.7**).

Table 4.6: *Details of the PRSs validated within the UKBB cohort.*

PRS Name	SNP Source	SNP *P* Value Cut-off	SNP (n)	Third vs First Tertile; OR (95% CI)	Second vs First Tertile; OR (95% CI)	P_{trend}	AUC (95% CI)
PRS19	Systematic review (genome-wide significant SNPs)	<5E-8	19	1.55 (1.37-1.74)	1.20 (1.06-1.37)	5.29E-13	0.56 (0.54-0.57)
PRS24	Systematic review (all)	<1E-5	24	1.27 (1.13-1.43)	1.21 (1.07-1.36)	1.02E-4	0.53 (0.52-0.54)
PRS40	Best performing PRS in Manchester cohort	<1E-5	40	1.36 (1.20-1.53)	1.20 (1.06-1.36)	6.34E-7	0.54 (0.53-0.55)
PRS72	Systematic review and ECAC GWAS	<1E-5	72	1.32 (1.17-1.48)	1.16 (1.03-1.31)	6.54E-6	0.54 (0.52-0.55)
PRS78	Systematic review, ECAC and Manchester GWAS	<1E-5	78	1.18 (1.05-1.33)	1.03 (0.91-1.16)	5.28E-3	0.52 (0.51-0.54)

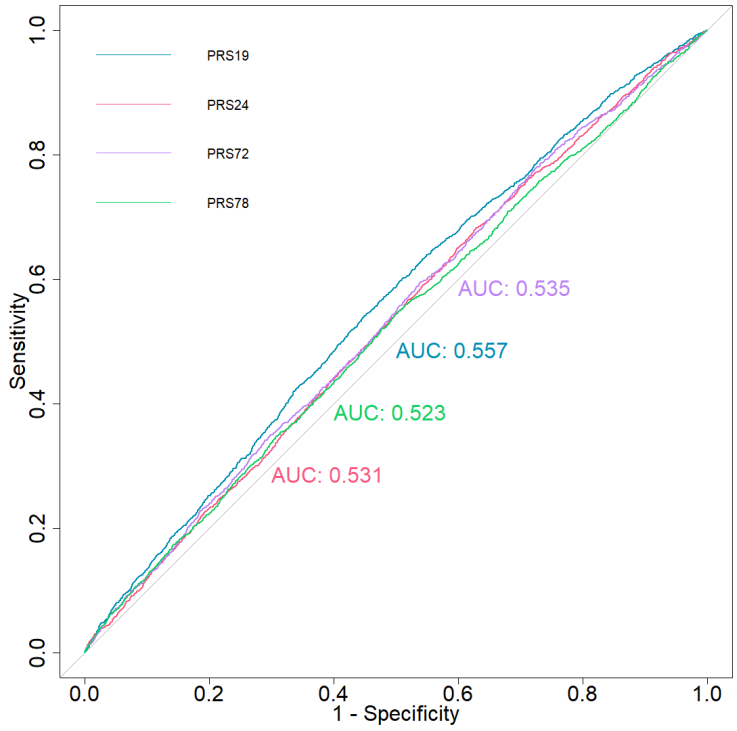


Figure 4.5: *AUC obtained for PRS19, PRS24, PRS72 and PRS78 in UKBB cohort.*

Although the discriminatory performance of the PRSs were low, a modestly increased risk was seen in all women in third versus first tertile among all PRSs, and the P_{trend} for the difference of risk between tertiles was significant for all combinations.

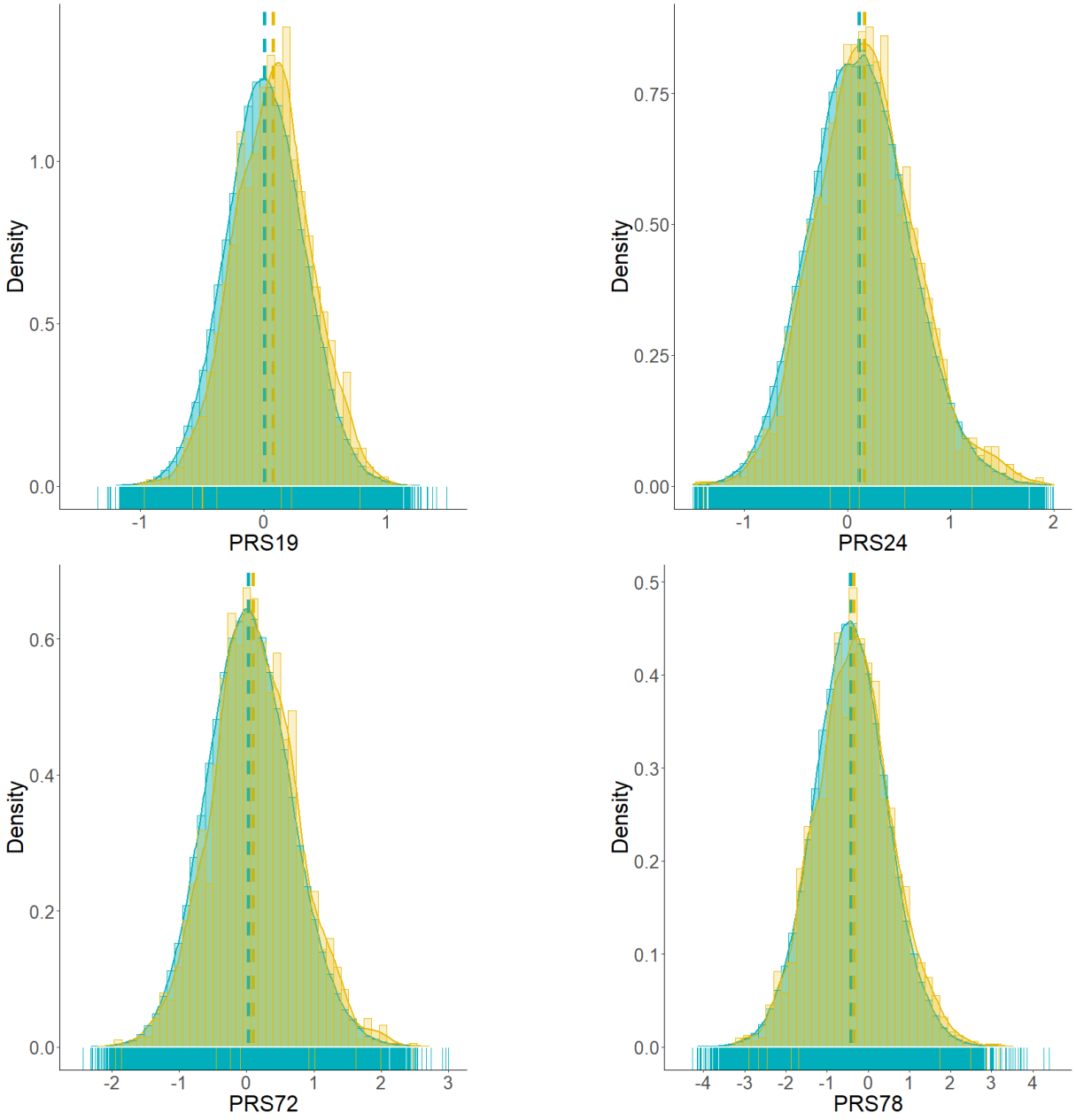
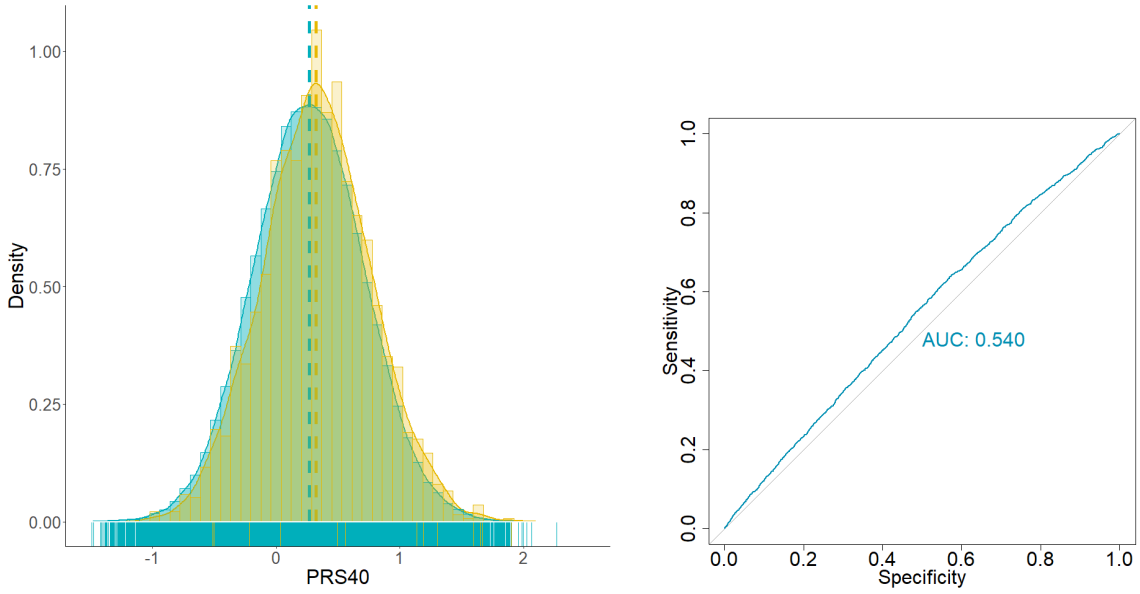


Figure 4.6: *PRS performances in UKBB cases and controls. Distribution of PRS19, PRS24, PRS72 and PRS78 is presented in each corresponding plot.*

The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.



(a) (b)

Figure 4.7: *Validation of the performance of PRS40 in UKBB cases and controls.*

(a) Density and (b) AUC plots of PRS40. (a) The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.

4.5.3 Cohort-Based Semi-Validation of the PRS in ECAC

4.5.3.1 Characteristics of the Study Population

The ECAC validation cohort comprised 9,062 cases and 41,461 female controls. Of the available data from ECAC, cases and controls had a median age and BMI of 63 years and 28.2 kg/m^2 and 55 years and 25 kg/m^2, respectively. Three quarters of the cases had the endometrial cancer histological subtype defined of which majority were endometrioid tumours (83.3%).

4.5.3.2 Polygenic Risk Score Validation

Though the majority of SNPs were discovered from ECAC study, the discriminatory performance of the extended SNP panel was tested in the ECAC dataset, excluding UKBB participants and unavailable data, for cohort-based *semi-validation* of the PRSs. In this cohort, PRS24 showed the lowest AUC among the four PRSs, 0.56 95% CI (0.55-0.56) (**Table 4.7**). The best AUC obtained was with PRS72 showing an AUC of 0.60 (95% CI 0.59-0.61). For PRS19 and PRS78, the corresponding values obtained were 0.59 (0.58-0.59) and 0.57 (0.56-0.58), respectively. The resulting PRS distribution and AUC plots are presented in **Figures 4.9** and **4.8** below. Validation of the refined PRS40 had a moderate performance in the ECAC cohort (AUC 0.59, 95% CI 0.58-0.60) (**Figure 4.10**).

Table 4.7: *Details of the PRSs validated within the ECAC cohort.*

PRS Name	SNP Source	SNP *P* Value Cut-off	SNP (n)	Third vs First Tertile; OR (95% CI)	Second vs First Tertile; OR (95% CI)	P_{trend}	AUC (95% CI)
PRS19	Systematic review (genome-wide significant SNPs)	<5E-8	19	1.94 (1.83-2.05)	1.38 (1.29-1.46)	1.22E-114	0.59 (0.58-0.59)
PRS24	Systematic review (all)	<1E-5	24	1.58 (1.49-1.67)	1.23 (1.16-1.31)	6.04E-57	0.56 (0.55-0.56)
PRS40	Best performing PRS in Manchester cohort	<1E-5	40	2.01 (1.89-2.13)	1.36 (1.28-1.45)	1.66E-128	0.59 (0.58-0.60)
PRS72	Systematic review and ECAC GWAS	<1E-5	72	2.16 (2.04-2.29)	1.41 (1.32-1.31)	1.75E-155	0.60 (0.59-0.61)
PRS78	Systematic review, ECAC and Manchester GWAS	<1E-5	78	1.69 (1.60-1.79)	1.22 (1.15-1.30)	2.25E-76	0.57 (0.56-0.58)

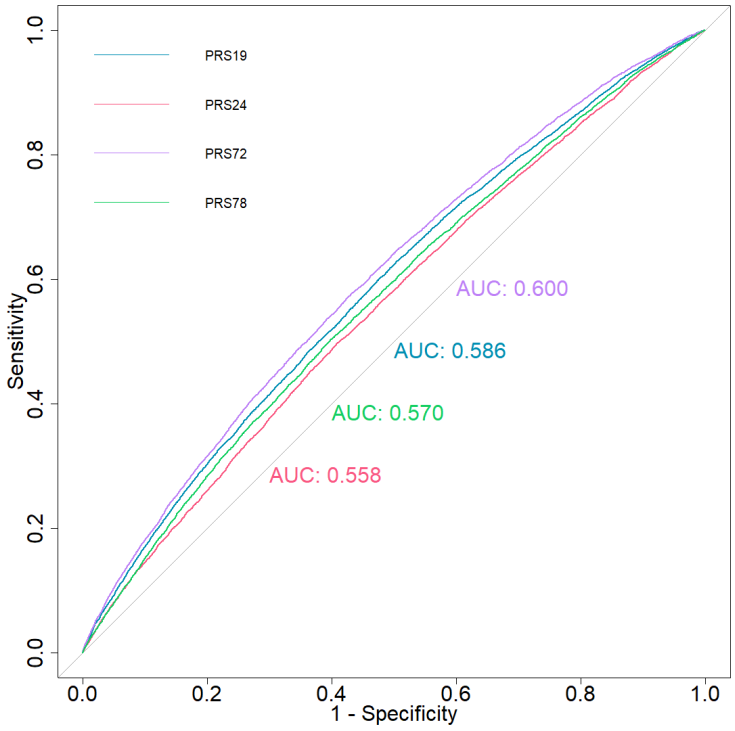


Figure 4.8: *AUC obtained for PRS19, PRS24, PRS72 and PRS78 in ECAC cohort.*

In this cohort, women in both the third and second tertiles had a statistically significantly increased risk of having endometrial cancer compared to the women in the first tertile, in all of the PRSs constructed. In particular, women ranking in the third tertile of the best performing PRS (PRS72) had 2.16-fold increased risk compared to those in the first tertile. For all PRSs, the P_{trend} for this difference in risk between tertiles was statistically significant.

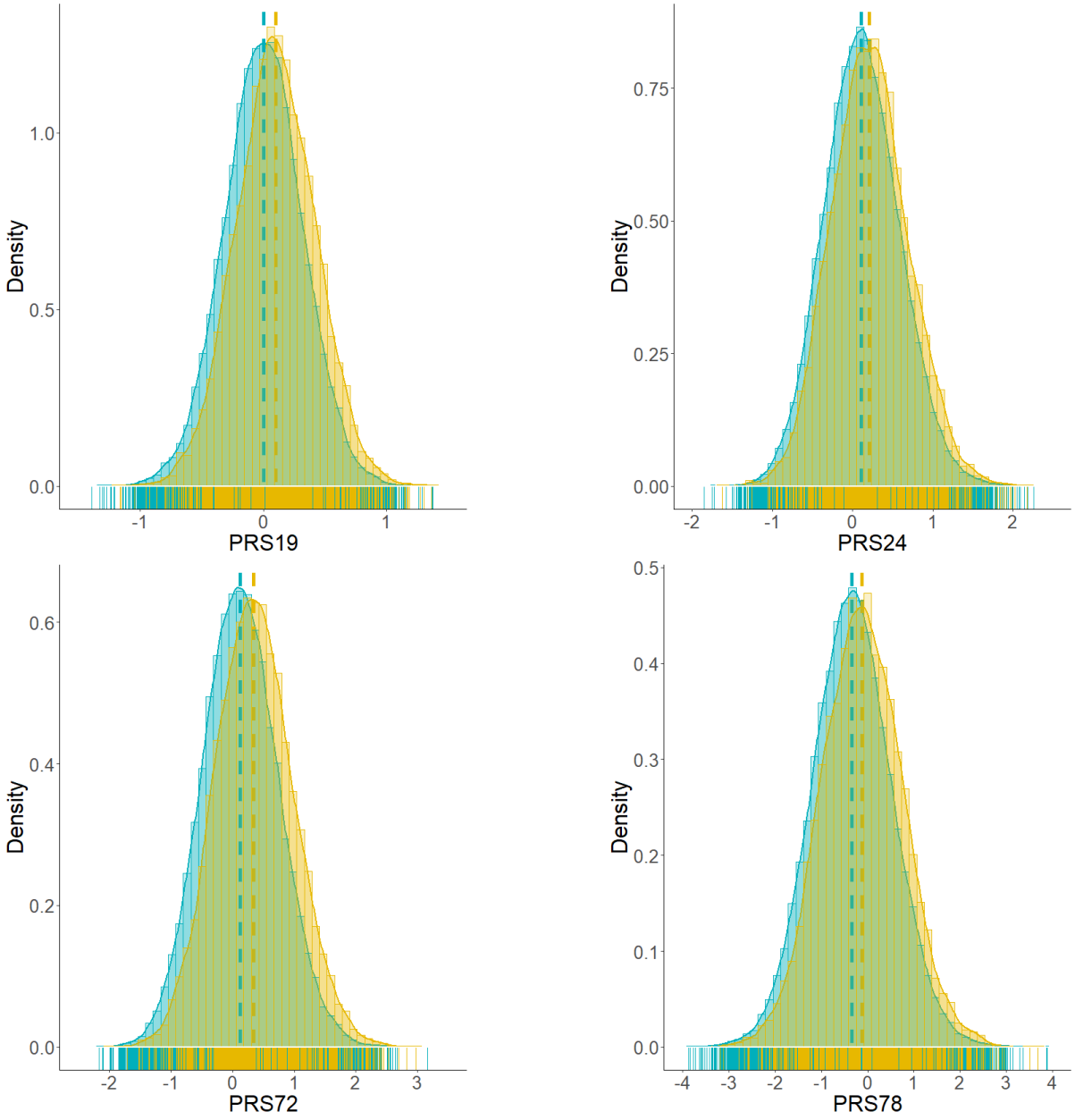
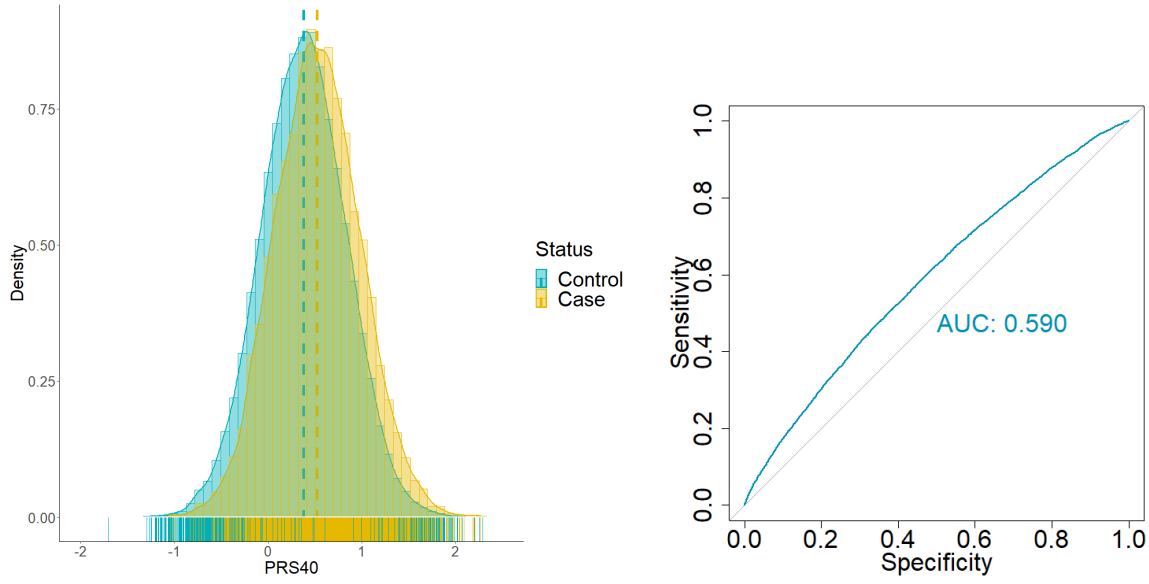


Figure 4.9: *PRS performances in ECAC cases and controls. Distribution of PRS19, PRS24, PRS72 and PRS78 is presented in each corresponding plot.*

The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.



(a) (b)

Figure 4.10: *Validation of the performance of PRS40 in available ECAC cases and controls.*

(a) Density and (b) AUC plots of PRS40. (a) The median PRS for endometrial cancer cases and controls are shown with the vertical dashed lines. Density of the PRS is also represented immediately above the x-axis, colour coded as above.

4.5.4 Differences Between Participants of the Three Cohorts

As the three datasets had participants recruited from different settings; hospital-based for Manchester cohort, population-based for UKBB cohort and mixed for ECAC cohort, we investigated whether the mean age and BMI were statistically different between Manchester and the other cohorts, where sufficient data were available (**Figure 4.11**).

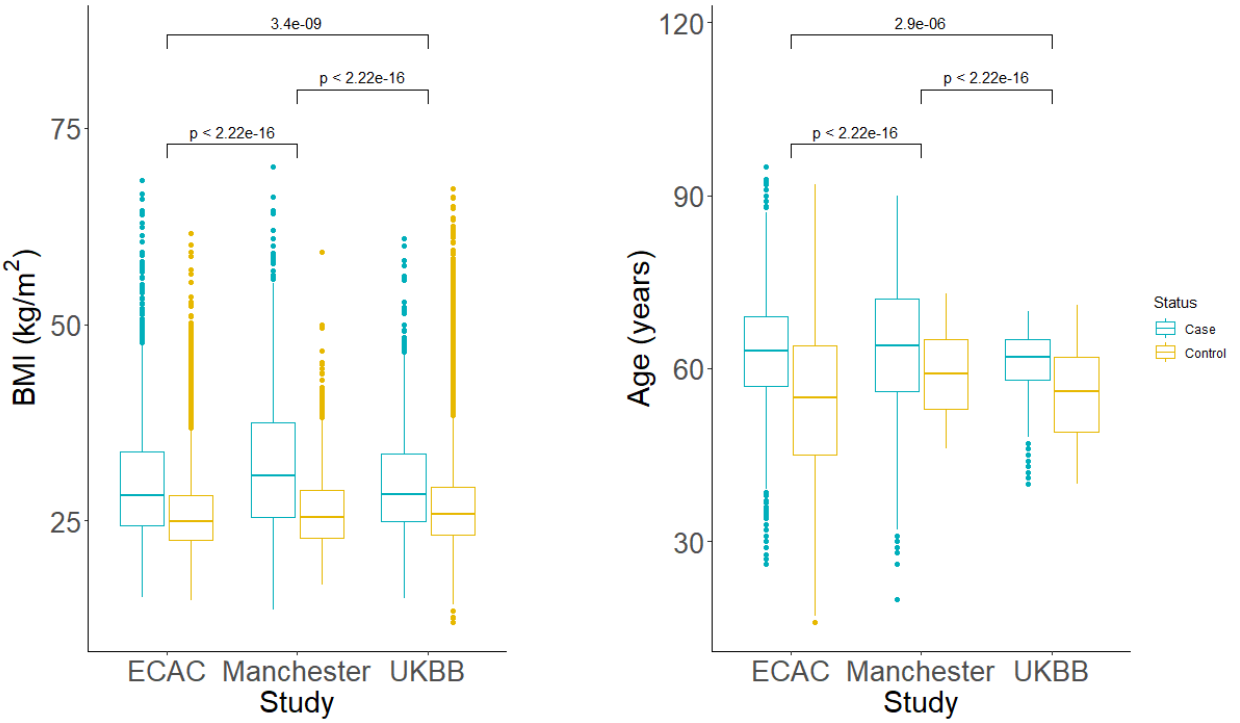


Figure 4.11: *Presentation of the differences in age and BMI between the three cohorts.*

Each study is represented by two box plots colour coded according to the disease status of the participants. The *P* values represent the statistical significance between overall age or BMI compared between each of the study pairs. *P* < 2.22E-16 represents the lowest *P* value that can be displayed on the plot by the computer.

Mean (SD) age at the time of enrolment was 63.2 (12.7) years for 534 women with endometrial cancer and 59 (6.9) years for 1,202 women without endometrial cancer, within the Manchester study. In UKBB, mean (SD) age for endometrial cancer cases was 61 (6) years and 55.4 (8) for healthy controls. For the limited available data in the ECAC cohort, mean (SD) age for endometrial cancer cases was 63.1 (9.1) years whereas this was 53.4 (13.4) years for the controls. In all three studies, there was a statistically significant

difference in age between cases and controls (Manchester study P = 1.12E-12, UKBB P = 5.98E-232, ECAC P = 0*). Similarly, BMI of cases and controls was statistically different within each of the three studies (P = 1.08E-36, 1.09E-66, 4.18E-168 for Manchester, UKBB and ECAC studies, respectively).

Moreover, overall age was statistically different between the three studies, though very similar means were observed between ECAC and UKBB (55.4 and 55 years, respectively). Importantly, controls of the Manchester cohort were followed up for a median of 8.9 years. Thus, while at baseline, the age between cases and controls were statistically significantly different, mean or median age between the cases and controls would have been similar at the time of this study. The median follow-up time for the UKBB cohort was 10 years (IQR 9.44-10.82 years), with a maximum of 13 years.

4.6 Discussion

In this study, we describe the development and validation of an endometrial cancer PRS in cases and controls of European descent. We show that genetic predisposition to endometrial cancer can be calculated from multiple SNPs found to influence endometrial cancer susceptibility in large GWAS. In the Manchester cohort, our best performing PRS had an AUC of 0.66 with (PRS78) inclusive of the six SNPs identified by Manchester GWAS. Using this PRS, women in the third and second tertiles have a nearly 3.4- and 1.7-fold increased risk of endometrial cancer compared to women in the first tertile, respectively. Then we refined the SNP panel to improve discriminatory ability and the corresponding values we obtained for the refined PRS (PRS40, excluding the Manchester SNPs) were 2.4-and 1.3-fold increased risk of endometrial cancer in third and second tertiles compared to the first, respectively. These data suggest that a PRS combining low-risk susceptibility variants can help identify women at greatest risk of endometrial

*The null value obtained here is a result of very low P value obtained for this test where the software is not able to display the exact value.

cancer for targeted screening and prevention interventions.

Endometrial cancer is strongly associated with obesity and the potential aetiological contribution of polygenic factors has been relatively poorly studied to date. Even so, genetic risk has largely been attributed to rare pathogenic variants affecting high risk genes, for example in LS. Over the past decade, GWAS have been employed to examine the influence of SNPs on endometrial cancer predisposition (O'Mara *et al*., 2019). Though limited in terms of ethnicity and pathological subtypes investigated, these provide important evidence that multiple independent loci are associated with endometrial cancer risk (Bafligil *et al*., 2020). Accordingly, polygenic methods may provide an opportunity to unpick the relative contribution of familial and general population risk of this complex disease (O'Mara and Crosbie, 2020).

In the current study, we tested the combinatorial effects of multiple SNPs shown to influence endometrial cancer predisposition in an independent cohort of well phenotyped cases and controls from the North West of England. Nearly two thirds of the low-risk susceptibility variants in externally sourced SNP panel, comprised of SNPs with genome wide or suggestive significance from our systematic review and ECAC GWAS, showed the same effect size direction in the Manchester GWAS. Around two fifths of the 24 SNPs from our systematic review were statistically significant at $P < 0.05$. Though lower than the widely accepted genome-wide significance threshold, these findings support the importance of these SNPs in endometrial cancer predisposition.

The Manchester cases and controls provided an independent cohort in which to develop and test our endometrial cancer PRS. They originated from the same geographical location and thus were carefully matched in terms of sociodemographic factors. Detailed clinico-pathological data ensured Manchester controls had an intact uterus, and histological subtype, grade and stage of endometrial cancer was available for the cases. Their median age, BMI and tumour characteristics were representative of national data (CRUK,

2020). Whilst controls were younger than cases at baseline, their prospective cancer-free follow up over a median 8.9 years is a major strength of the Manchester cohort. This meticulous characterisation of cases and appropriately matched controls may explain the better performance of the PRS in the Manchester dataset compared to UKBB. A limitation of the cohort is its small sample size, which precluded assessment of some of the lower frequency SNPs in our dataset.

Despite the addition of the six suggestive SNPs identified in Manchester GWAS greatly increasing the PRS performance in this cohort, the possibility of inflated effect sizes due to small GWAS sample size limits their applicability in other datasets. Importantly, these SNPs have not yet been validated in larger GWAS. Furthermore, using internally-discovered SNPs often leads to misleading estimations. Therefore, although seemingly very informative in our Manchester cohort, we do not promote their use in other datasets until they are validated in large GWAS.

Zhang *et al.* estimated that sample sizes equivalent to those available to the BCAC ($\sim$150,000 cases) would be necessary to explain three quarters of the genetic variance (Zhang *et al.*, 2020). The relatively small number of cases available ($\sim$13,000) to ECAC is therefore a major limitation of work in this area. It has been suggested that the genome-wide SNPs reported by ECAC to date can only explain $\sim$7% of the genetic variance (O'Mara *et al.*, 2018). ECAC is also limited by lack of age, BMI, hysterectomy status (controls) and detailed pathological data (cases). Validation of the discriminatory performance of our PRS in ECAC should be interpreted with caution, since most of the SNPs were discovered in this dataset. However, the relatively good performance of the PRS in Manchester cohort using SNPs primarily sourced from ECAC GWAS provides evidence for the robustness of these SNPs as well as their applicability in an independent and representative cohort.

The validation of the PRS in UKBB was disappointing, possibly due to the well-

documented healthy volunteer effect of this study, whose participants are generally younger, fitter, less obese and from higher socioeconomic backgrounds than their general population counterparts (Fry *et al*., 2017). The median follow up of 10 years in this cohort would be expected to bring the age of the controls, at the time of data acquisition, closer to the peak age period for endometrial cancer diagnosis. However, the latest cancer registry data update to UKBB data had taken place nearly 2.5 years prior. Moreover, all SNPs in our PRS were derived from European datasets and there were insufficient data to identify histological type-specific SNPs with any confidence, precluding an assessment of genetic predisposition to endometrioid versus non-endometrioid cancers.

Non-endometrioid endometrial cancers are associated with a less favourable prognosis and higher rate of mortality than their endometrioid counterparts (Lu and Broaddus, 2020). The small number of non-endometrioid tumours as well as lack of histopathological data available in the datasets so far has limited the discovery of genome-wide significant variants being identified for this subtype. Despite this, in our Manchester study, we showed that the best performing PRS constitution apply equally well in both subtypes, indicating their potential utility in identifying the group of patients with non-endometrioid tumours that are at a much higher risk of adverse outcomes.

To date, only two studies attempted to generate a PRS specific for predicting endometrial cancer risk. Choi *et al.* used 19 genome-wide significant SNPs from the latest ECAC GWAS and tested the PRS in UKBB (AUC 0.56) (Choi *et al*., 2020). On the other hand, Fritsche *et al.* used publicly available resources, namely the GWAS Catalogue and UKBB, and a more complex approach for selecting the SNPs (Fritsche *et al*., 2020). Nevertheless, using 20 SNPs from the GWAS Catalogue which presumably included the same SNPs we identified in our systematic review (Bafligil *et al*., 2020), mainly from the recent ECAC GWAS, the authors reported an AUC of 0.57 in UKBB as their best-performing PRS for endometrial cancer. Both of the studies, for their respective PRSs, reported similar performances to our findings in UKBB. Fritsche *et al.* also used a smaller cohort of 643

cases from Michigan Genomics Initiative, however, the mean age of the whole cohort was similar to that of UKBB (56.8 versus 56.9) and detailed information such as tumour histotype and BMI was not available. It is important to note that in both studies, the authors failed to remove the UKBB duplicates from the ECAC GWAS logORs or, in case of Choi *et al.*, validate their findings in an independent cohort. Thus, our extended PRS provides further evidence for the use of additional SNPs to achieve finer discriminatory PRS performance.

By contrast, multiple PRS have been developed for other cancer types. For example, using a PRS comprised of 313 SNPs, breast cancer can be predicted in women of European ancestry with an AUC of 0.63 (Mavaddat *et al.*, 2019). Furthermore, a panel of 18 SNPs improves the predictive performance of an established, clinically validated breast cancer risk prediction model (Evans *et al.*, 2019). Two PRS incorporating 30 and 22 SNPs for epithelial and high-grade serous ovarian cancer, respectively, have been described, the latter predicting substantial absolute risk differences for pathogenic variant carriers of *BRCA1* or *BRCA2* at PRS distribution extremes (Barnes *et al.*, 2020).

For endometrial cancer, three published risk prediction models have been proposed (Kitson *et al.*, 2017), developed (Husing *et al.*, 2016; Fortner *et al.*, 2017) or externally validated (Pfeiffer *et al.*, 2013) that integrate measures of obesity, reproductive risk, and/or insulin resistance, with moderate discriminatory power (0.62-0.77) (Alblas *et al.*, 2018). Zhang *et al.* estimated that over 1,000 independent SNPs could influence endometrial cancer risk and the best PRS achievable using these was estimated to have an AUC of 0.64 (Zhang *et al.*, 2020). Thus, our results support the notion that PRS performance for endometrial cancer risk is, expectedly, moderate. Moreover, although comprised of multiple SNPs, a PRS is nevertheless a single risk factor. Considering the multi-variable RPMs developed for predicting endometrial cancer showed moderate discriminatory ability, our extended PRS performance is rather satisfactory. Thus, incorporating an independent genetic score based on a panel of low-risk susceptibility variants to the RPMs

may substantially improve their overall accuracy and assist better and more personalised risk stratification.

Development of a personalised risk score for endometrial cancer was voted the most important research priority in endometrial cancer by clinicians, patients and carers in our James Lind Alliance Priority Setting Partnership (Wan *et al*., 2016). Effective risk assessment tools are crucial for implementing screening and prevention interventions at the population level. This is particularly important for endometrial cancer given the rapid rise in new diagnoses and deaths accompanying the current global obesity epidemic (Crosbie and Morrison, 2014). Models that risk stratify members of the general population will inform screening and prevention trial eligibility for high risk women while avoiding unnecessary interventions for those at low risk. Though it is unlikely that all will prove useful in a PRS, as shown by our PRS calculations, the 72 SNPs we present here provide the best evidence to construct a PRS for predicting the risk of endometrial cancer. In our Manchester cohort, the 40 SNP panel provided us with the most effective predictive value, however, we were unable to confirm this confidently in two other datasets. Thus, we propose our novel 40-SNP PRS for independent testing in large-scale datasets for validation.

4.7 References

Alblas, M., Velt, K. B., Pashayan, N., *et al.* (2018). Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. *Critical Reviews in Oncology/Hematology*, 126:92–99.

Amos, C. I., Dennis, J., Wang, Z. M., *et al.* (2017). The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiology Biomarkers & Prevention*, 26(1):126–135.

Auton, A., Abecasis, G. R., Altshuler, D. M., *et al.* (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

Bafligil, C., Thompson, D. J., Lophatananon, A., *et al.* (2020). Association between genetic polymorphisms and endometrial cancer risk: a systematic review. *Journal of Medical Genetics*, 57(9):591–600.

Barnes, D. R., Rookus, M. A., McGuffog, L., *et al.* (2020). Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*, 22(10):1653–1666.

Bycroft, C., Freeman, C., Petkova, D., *et al.* (2017). Genome-wide genetic data on 500,000 UK Biobank participants. *bioRxiv*, page 166298.

Bycroft, C., Freeman, C., Petkova, D., *et al.* (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209.

Choi, J., Jia, G., Wen, W., *et al.* (2020). Evaluating polygenic risk scores in assessing risk of nine solid and hematologic cancers in European descendants. *International Journal of Cancer*.

Constantinou, P. and Tischkowitz, M. (2017). Genetics of gynaecological cancers. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 42:114–124.

Crosbie, E. and Morrison, J. (2014). The emerging epidemic of endometrial cancer: time to take action. *Cochrane Database of Systematic Reviews*, (12).

Crosbie, E. J., Zwahlen, M., Kitchener, H. C., *et al.* (2010). Body Mass Index, Hormone Replacement Therapy, and Endometrial Cancer Risk: A Meta-Analysis. *Cancer Epidemiology Biomarkers & Prevention*, 19(12):3119–3130.

CRUK (accessed: Oct 2020). Uterine cancer statistics. `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer`.

Dominguez-Valentin, M., Sampson, J. R., SeppÃ¤lÃ¤, T. T., *et al.* (2020). Cancer risks by gene, age, and gender in 6350 carriers ofpathogenic mismatch repair variants: findings from the Prospective Lynch SyndromeDatabase. *Genetics in Medicine*, 22(1):15–25.

Evans, D. G. R., Harkness, E. F., Brentnall, A. R., *et al.* (2019). Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. *Breast Cancer Research and Treatment*, 176(1):141–148.

Fortner, R. T., Husing, A., Kuhn, T., *et al.* (2017). Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. *Int J Cancer*, 140(6):1317–1323.

Fritsche, L. G., Patil, S., Beesley, L. J., *et al.* (2020). Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. *The American Journal of Human Genetics*, 107(5):815–836.

Fry, A., Littlejohns, T. J., Sudlow, C., *et al.* (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9):1026–1034.

Husing, A., Dossus, L., Ferrari, P., *et al.* (2016). An epidemiological model for prediction of endometrial cancer risk in Europe. *European Journal of Epidemiology*, 31(1):51–60.

Kitson, S. J. and Crosbie, E. J. (2019). Endometrial cancer and obesity. *The Obstetrician & Gynaecologist*, 21(4):237–245.

Kitson, S. J., Evans, D. G., and Crosbie, E. J. (2017). Identifying High-Risk Women for Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model. *Cancer Prevention Research*, 10(1):1–13.

Lu, K. H. and Broaddus, R. R. (2020). Endometrial Cancer. *New England Journal of Medicine*, 383(21):2053–2064.

Mavaddat, N., Michailidou, K., Dennis, J., *et al.* (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, 104(1):21–34.

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., *et al.* (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*, 107(5):djv036.

Morice, P., Leary, A., Creutzberg, C., *et al.* (2016). Endometrial cancer. *Lancet*, 387(10023):1094–1108.

O'Mara, T. A. and Crosbie, E. J. (2020). Polygenic risk score opportunities for early detection and prevention strategies in endometrial cancer. *British Journal of Cancer*.

O'Mara, T. A., Glubb, D. M., Amant, F., *et al.* (2018). Identification of nine new susceptibility loci for endometrial cancer. *Nature Communications*, 9.

O'Mara, T. A., Glubb, D. M., Kho, P. F., *et al.* (2019). Genome-wide association studies of endometrial cancer: Latest developments and future directions. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*.

Pfeiffer, R. M., Park, Y., Kreimer, A. R., *et al.* (2013). Risk Prediction for Breast, Endometrial,

and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies. *Plos Medicine*, 10(7):15.

Raglan, O., Kalliala, I., Markozannes, G., *et al.* (2019). Risk factors for endometrial cancer: An umbrella review of the literature. *Int J Cancer*, 145(7):1719–1730.

Ryan, N. A. J., McMahon, R., Tobi, S., *et al.* (2020). The proportion of endometrial tumours associated with Lynch syndrome (PETALS): A prospective cross-sectional study. *PLOS Medicine*, 17(9):e1003263.

Ryan, N. A. J., Morris, J., Green, K., *et al.* (2017). Association of Mismatch Repair Mutation With Age at Cancer Onset in Lynch Syndrome: Implications for Stratified Surveillance Strategies. *JAMA Oncology*, 3(12):1702–1706.

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1):7–30.

Spurdle, A. B., Bowman, M. A., Shamsani, J., and Kirk, J. (2017). Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol*, 30(8):1048–1068.

Wan, Y. L., Beverley-Stevenson, R., Carlisle, D., *et al.* (2016). Working together to shape the endometrial cancer research agenda: The top ten unanswered research questions. *Gynecologic Oncology*, 143(2):287–293.

Win, A. K., Reece, J. C., and Ryan, S. (2015). Family History and Risk of Endometrial Cancer A Systematic Review and Meta-analysis. *Obstetrics and Gynecology*, 125(1):89–98.

Zhang, Y. D., Hurson, A. N., Zhang, H. Y., *et al.* (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature Communications*, 11(1):13.

4.8 Supplementary Data

Table S4.8: *Detailed list of SNPs used in the PRSs investigated in this study.*

rsID	OR	EA	Source	PRS19	PRS24	PRS72	PRS78	PRS40
rs10835920	1.09	T	SR	✓	✓	✓	✓	✓
rs10850382	1.10	T	SR	✓	✓	✓	✓	
rs2498794	1.13	G	SR	✓	✓	✓	✓	
rs17601876	1.12	G	SR	✓	✓	✓	✓	✓
rs1129506	0.91	A	SR	✓	✓	✓	✓	
rs11263761	1.15	A	SR	✓	✓	✓	✓	✓
rs882380	1.10	A	SR	✓	✓	✓	✓	✓
rs113998067	1.23	C	SR	✓	✓	✓	✓	✓
rs148261157	1.26	A	SR	✓	✓	✓	✓	
rs2747716	0.91	G	SR	✓	✓	✓	✓	✓
rs35286446	1.10	GAT	SR	✓	✓	✓	✓	✓
rs139584729	0.71	G	SR	✓	✓	✓	✓	
rs1679014	0.85	C	SR	✓	✓	✓	✓	
rs1740828	0.87	A	SR	✓	✓	✓	✓	✓
rs4733613	0.85	G	SR	✓	✓	✓	✓	✓
rs3184504	1.10	C	SR	✓	✓	✓	✓	✓
rs9668337	1.11	A	SR	✓	✓	✓	✓	✓
rs7981863	0.86	T	SR	✓	✓	✓	✓	
rs937213	1.09	C	SR	✓	✓	✓	✓	✓
rs1953358	0.74	A	SR		✓	✓	✓	
rs673604	1.21	G	SR		✓	✓	✓	✓
rs8178648	1.71	G	SR		✓	✓	✓	
rs9399840	1.33	T	SR		✓	✓	✓	✓

rsID	OR	EA	Source	PRS19	PRS24	PRS72	PRS78	PRS40
rs79575945	1.20	G	SR		✓	✓	✓	
rs138843373	1.57	T	ECAC GWAS			✓	✓	
rs4312007	1.08	T	ECAC GWAS			✓	✓	✓
rs76762469	1.18	A	ECAC GWAS			✓	✓	✓
rs558029	1.15	C	ECAC GWAS			✓	✓	✓
rs7967338	0.92	C	ECAC GWAS			✓	✓	✓
rs11583244	1.08	T	ECAC GWAS			✓	✓	
rs117670121	1.33	A	ECAC GWAS			✓	✓	
rs1677893	0.92	T	ECAC GWAS			✓	✓	✓
rs11069840	1.17	C	ECAC GWAS			✓	✓	✓
rs538150087	1.42	A	ECAC GWAS			✓	✓	
rs144076224	1.61	C	ECAC GWAS			✓	✓	
rs139948912	1.64	G	ECAC GWAS			✓	✓	
rs4072776	0.92	C	ECAC GWAS			✓	✓	
rs141265605	1.32	G	ECAC GWAS			✓	✓	
rs60310219	1.65	T	ECAC GWAS			✓	✓	
rs34676612	1.12	C	ECAC GWAS			✓	✓	✓
rs1533495	1.09	T	ECAC GWAS			✓	✓	
rs117338667	1.45	T	ECAC GWAS			✓	✓	
rs60856912	1.12	T	ECAC GWAS			✓	✓	✓
rs150325239	1.41	T	ECAC GWAS			✓	✓	
rs4607001	1.10	T	ECAC GWAS			✓	✓	✓
rs9616483	1.13	A	ECAC GWAS			✓	✓	✓
rs7579014	1.09	A	ECAC GWAS			✓	✓	
rs2920505	1.09	A	ECAC GWAS			✓	✓	✓
rs73044842	0.62	A	ECAC GWAS			✓	✓	✓
rs72716510	0.85	T	ECAC GWAS			✓	✓	

rsID	OR	EA	Source	PRS19	PRS24	PRS72	PRS78	PRS40
rs7713218	0.93	G	ECAC GWAS			✓	✓	✓
rs7729155	1.11	C	ECAC GWAS			✓	✓	
rs1357050	0.93	G	ECAC GWAS			✓	✓	✓
rs76165228	0.89	G	ECAC GWAS			✓	✓	✓
rs2982732	1.08	A	ECAC GWAS			✓	✓	✓
rs9296422	1.09	C	ECAC GWAS			✓	✓	
rs149369224	1.53	C	ECAC GWAS			✓	✓	
rs117274813	1.43	T	ECAC GWAS			✓	✓	
rs117978821	0.79	C	ECAC GWAS			✓	✓	✓
rs1553183	0.92	C	ECAC GWAS			✓	✓	✓
rs10505508	0.92	T	ECAC GWAS			✓	✓	
rs6470612	0.87	G	ECAC GWAS			✓	✓	✓
rs1356332	0.91	A	ECAC GWAS			✓	✓	✓
rs13250178	1.08	T	ECAC GWAS			✓	✓	✓
rs6468088	1.24	T	ECAC GWAS			✓	✓	
rs1923357	0.91	T	ECAC GWAS			✓	✓	
rs3808753	1.25	G	ECAC GWAS			✓	✓	✓
rs577034498	1.18	C	ECAC GWAS			✓	✓	✓
rs535955703	1.70	T	ECAC GWAS			✓	✓	
rs2659685	0.91	A	ECAC GWAS			✓	✓	✓
rs149404333	0.88	C	ECAC GWAS			✓	✓	✓
rs60515555	1.11	G	ECAC GWAS			✓	✓	✓
rs12727038	1.52	T	Mcr GWAS				✓	
rs144065942	0.65	G	Mcr GWAS				✓	
rs9854980	0.57	C	Mcr GWAS				✓	
rs6580584	0.40	T	Mcr GWAS				✓	
rs74532550	1.64	G	Mcr GWAS				✓	

rsID	OR	EA	Source	PRS19	PRS24	PRS72	PRS78	PRS40
rs111282723	2.61	T	Mcr GWAS				✓	

EA, effect allele; Mcr, Manchester; SR, systematic review.

5 *BRCA1* and *BRCA2* pathogenic variant carriers and endometrial cancer risk: A cohort study

This Chapter has been published at European Journal of Cancer (Kitson *et al.*, 2020) which can be found at **Appendix 8.2, p 273**. The article aimed to determine whether an increased risk for endometrial cancer exists in pathogenic *BRCA1/2* variant carriers. In this Chapter, the methods section describing the somatic tumour sequencing has been enriched for clarity of the thesis.

5.1 Contributorship Statement

I identified the high-grade serous endometrial cancer tumours for DNA sequencing, prepared the samples for the sequencing lab, analysed the sequencing data and contributed to writing the manuscript. N.A.J.R. and I contributed to data collection and analysis. F.L., E.R.W., R.D.C., R.J.E. and D.G.E. contributed in data collection. S.J.K. performed the analyses and drafted the manuscript. J.B. provided pathological review of tumour slides.

5.2 Abstract

Background An association between *BRCA* pathogenic variants and an increased endometrial cancer risk, specifically serous-like endometrial cancer, has been postulated but remains unproven, particularly for *BRCA2* carriers. Mechanistic evidence is lacking, and any link may be related to tamoxifen exposure or testing bias. Hysterectomy during risk-reducing bilateral salpingo-oophorectomy is, therefore, of uncertain benefit. Data from a large, prospective cohort will be informative.

Methods Data on UK *BRCA* pathogenic variant carriers were interrogated for endometrial cancer diagnoses. Standardised incidence ratio (SIR) were calculated in four distinct cohorts using national endometrial cancer rates; either from 1/1/1980 or age 20, prospectively from date of personal pathogenic variant report, date of family pathogenic variant report or date of risk-reducing salpingo-oophorectomy. Somatic *BRCA* sequencing of 15 serous endometrial cancers was performed to detect pathogenic variants.

Results Fourteen cases of endometrial cancer were identified in 2609 women (1350 *BRCA1* and 1259 *BRCA2*), of which two were prospectively diagnosed. No significant increase in either overall or serous-like endometrial cancer risk was identified in any of the cohorts examined (SIR = 1.70, 95% confidence interval = 0.74-3.33; no cases of serous endometrial cancer diagnosed). Results were unaffected by the *BRCA* gene affected, previous breast cancer or tamoxifen use. No *BRCA* pathogenic variants were detected in any of the serous endometrial cancers tested.

Conclusions Women with a BRCA pathogenic variant do not appear to have a significant increased risk of all-type or serous-like endometrial cancer compared with the general population. These data provide some reassurance that hysterectomy is unlikely to be of significant benefit if performed solely as a preventive measure.

5.3 Introduction

Since the publication of a number of reports describing diagnoses of serous endometrial cancer in *BRCA1* pathogenic variant carriers of Ashkenazi Jewish heritage (Hornreich *et al*., 1999; Lavie *et al*., 2000; Kaplan *et al*., 1998), there has been interest in a potential association between the *BRCA* pathogenic variants and an increased risk of endometrial cancer. A number of studies have sought to quantify the level of risk, although with conflicting results, with some finding evidence of an increased risk (Thompson and Easton, 2002; Lavie *et al*., 2010; Segev *et al*., 2013), particularly in *BRCA1* carriers, whilst others have found no association (Levine *et al*., 2001; Breast Cancer Linkage Consortium, 1999). Unfortunately, the absence of a suitable control group has prevented the results of these earlier studies being reconciled in a meta-analysis (de Jonge *et al*., 2017).

From a biological perspective, should a causative relationship exist between *BRCA* pathogenic variants and endometrial cancer, it would be anticipated that the increased risk would be restricted to the serous-like histological subtype, including p53 mutant uterine carcinosarcomas and mixed epithelial carcinomas (de Jonge *et al*., 2019). This has, however, not always been observed (Reitsma *et al*., 2013; Lee *et al*., 2017).

It has also been postulated that any observed association may be due to the use of tamoxifen for the prevention and treatment of breast cancer rather than a consequence of a *BRCA1* or *BRCA2* pathogenic variant *per se* (Segev *et al*., 2015; Shu *et al*., 2016). Whilst several prospective studies have examined the incidence of endometrial cancer after risk-reducing salpingo-oophorectomy (RRSO) in *BRCA1* and *BRCA2* pathogenic variant carriers compared with the general population, they have failed to consider the impact of the procedure on the rate of endometrial cancer within this specific population (Reitsma *et al*., 2013; Segev *et al*., 2015). Debate, therefore, continues within the scientific and medical communities as to whether risk-reducing hysterectomy should be offered to women with *BRCA1* and *BRCA2* pathogenic variants at the time of their

RRSO to reduce their subsequent endometrial cancer risk (Lee *et al*., 2017; Saule *et al*., 2018).

Limitations of the studies performed to date are that they have often been purely retrospective, have included only small numbers of *BRCA* pathogenic variant carriers, particularly those with *BRCA2* pathogenic variants, and have frequently omitted to undertake expert pathological review of the tumour tissue to ensure accurate subtyping. They have also often had short follow-up durations of only 5-6 years, which, when applied to a cohort with a median age of 40-50 years, means that they have limited power to detect endometrial cancer cases which predominately occur in older women.

This study, therefore, sought to determine whether *BRCA* pathogenic variants are associated with an increased risk of endometrial cancer compared with the general population using a large, well-described cohort of *BRCA1* and *BRCA2* pathogenic variant carriers with prospective follow-ups. It also aimed to determine whether there was a particular association between *BRCA* pathogenic variants and the serous histological subtype of endometrial cancer and the impact of RRSO on this risk.

5.4 Materials and Methods

5.4.1 Database and Study Population

A prospectively maintained database of *BRCA* pathogenic variant carriers at the Manchester Centre for Genomic Medicine was used to identify individuals aged >20 years for analysis. Data were collected on date of birth, personal and family pathogenic variant testing, salpingo-oophorectomy ± hysterectomy, breast, ovarian and endometrial cancer diagnosis, death and date of the last follow-up. Information on tamoxifen use was collected wherever possible. Women were eligible for the study if they had a *BRCA1* or *BRCA2* pathogenic variant identified between 01/01/1991 (the start date of the database) and 31/12/2017 and had not undergone a previous hysterectomy. Pathology reports were collated from affected individuals to determine endometrial cancer subtype, with slide review by an expert gynaecological pathologist (J.B.) where possible and TP53

immunohistochemistry in accordance with previously published protocols (Kitson *et al.*, 2019). Follow-up data were collected through medical record review and from the National Cancer Registration and Analysis Service, for women enrolled in the Epidemiological study of Familial Breast Cancer (EMBRACE) study, a national cohort study of *BRCA1/2* pathogenic variant carriers and non-affected family members (Centre for Cancer Genetic Epidemiology, 2017).

Women were considered in a number of distinct, but overlapping, cohorts; retrospectively assuming the follow-up started on 1/1/1980 (or age 20 years, whichever occurred later) and prospectively from the date of their family *BRCA* pathogenic variant identification (or age 20 years), from date of their personal *BRCA* pathogenic variant identification (or age 20 years) and from date of RRSO, where applicable. A nested case-control analysis was planned to evaluate the competing effect of RRSO on endometrial cancer incidence; however, no cases of endometrial cancer occurred in women who underwent RRSO. Women were censored at time of hysterectomy, diagnosis of cancer of the ovary, fallopian tube or peritoneum, death, the last follow-up or 31/12/17, whichever occurred first.

5.4.2 Somatic *BRCA* Sequencing

To establish whether serous endometrial cancers are associated with pathogenic variants in *BRCA*, we identified 15 serous endometrial cancers from the Manchester cases (see **Chapter 3, p 109**) treated at our institution and carried out somatic *BRCA* sequencing. DNA was previously extracted from formalin-fixed paraffin embedded blocks of tumour tissue, which had been obtained at the time of hysterectomy. The DNA samples were extracted using either EZ1 DNA tissue kit (Cat no: 953034, Qiagen) or COBAS DNA sample preparation kit (Cat no: 05985536190, Roche). COBAS method yields slightly more DNA product and is normally the preferred method for downstream *BRCA1/2* mutation detection by MGDL (Hu *et al.*, 2014). DNA samples were quantified using Qubit 2.0 Fluorometer (Thermo Fisher Scientific). This assay uses dyes which are selectively fluorescent when bound to nucleic acids or proteins and is highly accurate.

5.4.2.1 Sample Preparation

Next-generation sequencing (NGS) for *BRCA1/2* mutations was performed by the MGDL using their in-house developed protocol (Ellison *et al.*, 2015). All samples submitted to the MGDL had at least 4 ng/μL of DNA and 20 μL of total volume, corresponding to 80 ng of intact DNA. DNA was amplified using GeneRead DNAseq Targeted Exon Enrichment Breast Panel (Qiagen). First, DNA samples were diluted to 5 ng/μL using sterile Injection BP water. For the samples with less than 5 ng/μL DNA concentration, no dilution was necessary. A minimum of 40 μL DNA dilution was prepared for each sample. 4 polymerase chain reaction (PCR) master mixes (one per primer pool) were prepared, scaled according to number of samples being run. Each master mix included GeneRead PCR buffer, one of the primer pools, GeneRead Hotstart Taq DNA polymerase and DNase-free water. The master mixes and DNA samples were dispensed into a semi-skirted 96-well PCR plate via a Biomek NX automated liquid handler (Beckman Coulter). The plate was then sealed, pulse spun and run on a Veriti PCR machine with the following settings: 95°C for 10 minutes, 27 times at 95°C for 15 seconds followed by 60°C for 4 minutes, 72°C for 10 minutes and was left on a 4°C infinite cooling period.

The PCR product was confirmed by running the products on a 2% agarose gel. To do this, 2 μL of each PCR product mixed with 2 μL loading dye was loaded onto the gel and run at 120V for 20 minutes. After this, the 4 PCR products per individual were pooled using Biomek NX. Ampure XP beads, freshly prepared 80% ethanol and Injection water were mixed with the pooled samples using the Biomek NX, for purification of the PCR products. The quality and quantity of purified DNA samples were measured on the TapeStation using D1000 High Sensitivity kit (Agilent). Briefly, the high sensitivity D1000 ladder and sample buffer were vortexed and pulse spun. Into each well, 2 μL of sample buffer and 2 μL of 1 in 10 diluted sample was aliquoted, and 2 μL of ladder was loaded into a separate well. After sealing the plate, it was vortexed at 2,000 rpm for 1 minute and pulse spun prior to loading on the TapeStation. The PCR products were expected to be approximately

150-180 base pairs in size and 25-125 ng in quantity.

5.4.2.2 Library Preparation

Based on the TapeStation results, an automated worksheet calculated the volume of purified DNA and EB buffer required for TruSeq DNA PCR-free library preparation. The appropriate volumes were added into a fresh 4titude Framestar 96-well plate. Next, 40 μL of End Repair Mix 2 was added to each well and mixed by pipetting. The plate was sealed, briefly vortexed and centrifuged. The following program was run in a thermal cycler (lid preheated to 100°C); 30°C for 30 minutes and a 4°C infinite cooling period. DNA size was selected for by automated TruSeq purification. After adding the TruSeq purification beads, samples were stored at -20°C overnight.

The ends of the PCR products were adenylated by adding 12.5 μL of A-Tailing Mix to each of the wells containing 17.5 μL of eluate. After sealing, vortexing and centrifuging the plate briefly, the following program was run in a thermal cycler (lid preheated to 100°C); 37°C for 30 minutes, 70°C for 5 minutes, 4°C for 5 minutes and keeping at this temperature for infinite hold. 2.5 μL of Buffer EB was added to each well, followed by 2.5 μL of Ligation Mix 2, and 2.5 μL of the DNA adapter. The plate was sealed, vortexed and centrifuged briefly. Then, the plate was placed in a thermal cycler with a preheated lid at 100°C and run at 30°C for 10 minutes with an infinite 4°C cooling period. The plate was immediately removed from the thermal cycler and 5 μL of Stop Ligation Buffer was added to each well.

For purification of the final library, TruSeq purification beads were added to each samplevia the Biomek NX automated liquid handler. Each library was then quantified using SYBR FAST qPCR kit (Kappa Biosystems). Each sample was diluted to 1:80 by adding 2 μL of sample to 158 μL of EB buffer. In a fresh plate, 4 μL of the 1:80 dilution was added to 196 μL EB buffer to create 1:4000 dilution. Each final dilution, size standards 1-6 and negative controls containing EB buffer were prepared in triplicate. Mastermix

containing 12 μL Kapa MM, 4 μL nuclease-free water per 1 reaction was prepared in a microbiological safety cabinet. 16 μL of the mastermix was added to each well including standards and negative control. 4 μL of the diluted samples and standards were added to each corresponding well. The qPCR was set up to run on the following program; 95°C for 5 minutes to denature the DNA, and 35 cycles of 95°C for 30 seconds followed by 60°C for 45 seconds. The qPCR data was then analysed using an in-house generated calculations template. Using the TruSeq final library plate, the samples were pooled into a single tube and stored in -20°C until required.

5.4.2.3 Pooled Library Preparation for Sequencing

Fresh dilutions of 0.2N sodium hydroxide (NaOH) and 0.2N tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) were prepared. 62.5 μL of 20 ρM PhiX control library was mixed with 37.5 μL HT1 hybridisation buffer to achieve 12.5 ρM PhiX. Based on starting molarity, appropriate volume of pooled amplicon library and 0.2N NaOH were mixed together to denature the DNA. After a 5 minute incubation at room temperature, appropriate volume of 0.2 N Tris-HCl was added to neutralise NaOH. Appropriate volume of HT1 buffer was then added to achieve a final concentration of 12.5 ρM. Finally, 6 ρL of 12.5 ρM PhiX was added to the mix and the tube was kept on ice. DNA sequencing was performed on MiSeq V.2 sequencer (Illumina).

5.4.2.4 Data Analysis

The sequencing data was processed by MGDL as described by Ellison *et al*. (2015), using an analysis and variant calling pipeline based on the bc-bio nextgen framework (bcbio, 2015). The genomic positions were based on GRCh37 assembly. The filtered data was then examined by C.B and searching any variant-of-interest on ClinVar to determine the clinical significance of the variants (Landrum *et al*., 2018).

5.4.3 Statistical Analysis

Expected endometrial cancer incidence rates were calculated using age-standardised UK-specific data available from the Office for National Statistics (CRUK, 2020) in 5 year intervals and were adjusted for local hysterectomy rates, as calculated using data from the Predicting Risk of Breast Cancer at Screening (PROCAS) study (Evans *et al.*, 2019). This was a large risk assessment study conducted in the Greater Manchester area developing breast cancer risk algorithms. The risk of endometrial cancer relative to the general population was evaluated with SIRs, calculated as the observed number of endometrial cancer cases divided by the expected number of cases.

Subgroup analyses were performed based on *BRCA1/2* pathogenic variant status, history of breast cancer and tamoxifen use and endometrial cancer histological subtype. Serous-like endometrial cancers included serous endometrial cancer, uterine carcinosarcomas with a serous epithelial component and mixed serous epithelial tumours in keeping with the findings of de Jonge *et al*. (2019). The expected number of serous-like endometrial cancer cases was calculated assuming 10% of all endometrial cancers were of the serous-like histotype (Hamilton *et al*., 2006; Sagae *et al*., 2014). The Byar's approximation of the exact Poisson distribution was used to calculate the 95% confidence limits using the methodology of Breslow and Day (Breslow and Day, 1987). Statistical analysis was performed using MS Excel by S.J.K (Microsoft Corporation, 2016).

5.5 Results

Of 2,609 women, 1,350 (51.7%) had a *BRCA1* pathogenic variant and 1,259 (48.3%) had a *BRCA2* pathogenic variant. The median age at baseline, the last follow-up and the length of the follow-up varied according to the cohort examined (Table **5.1**).

There were 14 cases of endometrial cancer identified; 12 occurred before the confirmation of a personal *BRCA* pathogenic variant (i.e., were identified retrospectively) and two cases were identified prospectively. The clinical characteristics of the endometrial cancer cases

Table 5.1: *Demographic data and follow-up duration for cohorts examined.*

Cohort	Size (n)	Median age at baseline, yrs (IQR)	Median age at last follow-up, yrs (IQR)	Total follow-up, women years at risk (median)	*BRCA* pathogenic variant status (%)	Prior history of BC (%)	Prior history of tamoxifen use (%)
Retrospective cohort (1/1/1980-31/12/17)	2609	20.0 (20.0-31.6)	48.8 (40.5-57.9)	59199 (23.8)	*BRCA1 1350 (51.7%)* *BRCA2 1259 (48.3%)*	Yes 1259 (48.2%) No 1350 (51.7%)	Yes 311 (11.9%) No 557 (21.3%) Unknown 1741 (66.6%)
Date of family pathogenic variant report	1811	44.1 (34.9-54.6)	49.1 (40.2-59.1)	9412 (3.5)	*BRCA1 906 (50.0%)* *BRCA2 905 (50.0%)*	Yes 907 (50.1%) No 901 (49.8%) Unknown 3 (0.2%)	Yes 260 (14.4%) No 490 (27.1%) Unknown 1061 (58.6%)
Date of personal pathogenic variant report	1617	45.1 (37.0-55.1)	49.2 (40.9-58.8)	6375 (2.4)	*BRCA1 817 (50.5%)* *BRCA2 800 (49.5%)*	Yes 847 (52.4%) No 768 (47.5%) Unknown 2 (0.1%)	Yes 258 (16.0%) No 478 (29.6%) Unknown 881 (54.4.%)
Date of RRSO	546	45.8 (40.4-52.6)	51.3 (45.4-58.9)	2865 (2.9)	*BRCA1 274 (50.2%)* *BRCA2 272 (49.8%)*	Yes 283 (51.8%) No 263 (48.2%)	Yes 103 (18.9%) No 178 (32.6%) Unknown 265 (48.5%)

BC, breast cancer; IQR, interquartile range; RRSO, risk-reducing salpingo-oophorectomy.

are given in **Supplementary Table S5.3**. Pathology review was possible for six of the 14 endometrial cancer cases identified, with TP53 immunohistochemistry performed in three cases to aid diagnosis. Most cases were of endometrioid subtype, with one retrospectively identified case of a mixed serous and endometrioid tumour and two cases of endometrial carcinosarcoma. Only the mixed serous tumour demonstrated diffuse p53 staining, in keeping with a mutant-like pattern. Both prospectively diagnosed endometrial cancer cases were of endometrioid subtype and occurred in index cases. There were two cases of proven synchronous ovarian and endometrial cancers, one identified prospectively and the other retrospectively, and a further suspected case within the retrospective cohort, which could not be confirmed as the original slides were not available for review. There were no cases of endometrial cancer in women who underwent RRSO.

The overall risk of endometrial cancer was not significantly increased in any of the four cohorts studied (from 1/1/1980 adjusted SIR = 1.70, 95% confidence interval (95% CI) = 0.74-3.33; date of family pathogenic variant report adjusted SIR = 0.89, 95% CI = 0.12-3.02; date of personal pathogenic variant report adjusted SIR = 1.21, 95% CI= 0.09-4.48; date of RRSO adjusted SIR incalculable, Table **5.2**).

Table 5.2: *Observed and expected endometrial cancer rates in BRCA pathogenic variant carriers.*

Year	Expected	Observed	SIR	CI lower 95%	CI upper 95%
1/1/1980-31/12/2017					
1980-1984	0.28	0	0.00	0.00	0
1985-1989	0.4	0	0.00	0.00	0
1990-1994	0.61	1	1.64	0.00	13.11
1995-1999	0.82	4	4.90	0.01	31.45
2000-2004	1.08	3	2.77	0.05	14.56
2005-2009	1.35	4	2.97	0.13	13.6
2010-2014	1.3	1	0.77	0.03	3.59
2015-2017	0.44	1	2.27	0.00	23.4
Total	6.27	14	2.23	0.84	4.78
Adjusted*	8.22	14	1.70	0.74	3.33
Serous-like EC	0.82	3	3.66	0.01	23.41
BRCA1 only	3.68	7	1.9	0.47	5.05
Date of family pathogenic variant mutation report					
1990-1994	0	0	0	0	0
1995-1999	0.05	1	19.39	0	1420.26
2000-2004	0.2	1	5.01	0	102.28
2005-2009	0.49	0	0	0	0
2010-2014	0.65	0	0	0	0
2015-2017	0.33	0	0	0	0
Total	1.71	2	1.17	0.1	4.6
Adjusted*	2.25	2	0.89	0.12	3.02
Serous-like EC	0.23	0	0	0	0
BRCA1 only	0.98	1	1.02	0.01	5.73
Date of personal pathogenic variant mutation report					

1990-1994	0	0	0	0	0
1995-1999	0.03	1	32.35	0	3906.24
2000-2004	0.14	1	7.33	0	212.49
2005-2009	0.37	0	0	0	0
2010-2014	0.5	0	0	0	0
2015-2017	0.22	0	0	0	0
Total	1.26	2	1.59	0.06	7.55
Adjusted*	1.65	2	1.21	0.09	4.88
Serous-like EC	0.17	0	0	0	0
BRCA1 only	0.73	1	1.36	0	9.46
Date of RRSO[3]					
1990-1994	0.02	0	0	0	0
1995-1999	0.04	0	0	0	0
2000-2004	0.09	0	0	0	0
2005-2009	0.19	0	0	0	0
2010-2014	0.29	0	0	0	0
2015-2017	0.13	0	0	0	0
Total	0.76	0	0	0	0
Adjusted*	0.99	0	0	0	0
Serous-like EC	0.10	0	0	0	0
BRCA1 only	0.47	0	0	0	0

* Expected data adjusted for hysterectomy prevalence data.

CI, confidence interval; EC, endometrial cancer; RRSO, risk-reducing salpingo-oophorectomy; SIR, standardised incidence ratio.

Subgroup analyses failed to find any difference in endometrial cancer risk between women with a *BRCA1* or *BRCA2* pathogenic variant, a history of breast cancer or tamoxifen use. Neither was there a specific increase in the risk of serous-like endometrial

cancer (cohort from 1/1/1980 SIR = 3.66, 95% CI = 0.01-23.41, SIR incalculable in the prospective cohorts as no cases of serous endometrial cancer diagnosed).

Furthermore, we assessed the presence of a pathogenic variant in *BRCA1/2* in first-degree relatives, of a proven carrier, who had developed endometrial cancer without previous breast or synchronous ovarian cancer. Five of seven (71%) did not carry the family variant. If endometrial cancer was associated, then more than 50% should have carried the pathogenic variant.

Of the 15 serous endometrial cancers analysed, none contained *BRCA1/2* pathogenic variants. The majority of the variants obtained were common single nucleotide polymorphisms or indels. Two variants located at Chr17: 41245400 of *BRCA1* Chr13:32931911 of *BRCA2*, both of which appeared relatively rare, were described on ClinVar as of uncertain significance and likely benign, respectively. There was no difference in depth of sequencing coverage between the samples extracted using either EZ1 and COBAS methods.

5.6 Discussion

This study did not find a significant increase in the incidence of endometrial cancer in women with a pathogenic variant in either the *BRCA1* or *BRCA2* genes. This finding was unaffected by the *BRCA* gene affected, a personal history of breast cancer or tamoxifen use. The study was unable to address whether RRSO reduces the risk of endometrial cancer specifically in this population, due to the lack of endometrial cancer cases in women who underwent RRSO. No specific association between *BRCA1/2* pathogenic variants and serous endometrial cancer was detected; there was neither an increased risk of serous endometrial cancer in *BRCA1/2* pathogenic variant carriers nor were pathogenic variants detected in the *BRCA1/2* genes within the tumour tissue from 15 unselected serous endometrial cancers.

These reassuring findings are consistent with those of Levine *et al*. (2001) , who described

a relative risk of endometrial cancer of 0.75 (95% CI = 0.24-2.34, *P* = 0.6) in 199 Ashkenazi Jews with *BRCA1/2*, and of Lee *et al*. (2017), who failed to find an increase in serous or endometrioid endometrial cancer in their moderately sized Australasian population (*BRCA1* SIR = 2.87, 95% CI = 0.59-8.43, *P* = 0.18, *BRCA2* SIR = 2.01, 95% CI = 0.24-7.30, *P* = 0.52). The largest study to date, conducted across 11 different countries, however, yielded contradictory results, noting a significantly increased risk of endometrial cancer in *BRCA1* pathogenic variant carriers and those exposed to tamoxifen (Segev *et al*., 2013). Of the 4,893 women studied, 3,536 were *BRCA1* pathogenic variant carriers, explaining why there was no statistically significant increase in endometrial cancer risk in the *BRCA2* group, despite similar SIRs (*BRCA1* SIR = 1.91, 95% CI = 1.06-3.19, *P* = 0.03, *BRCA2* SIR = 1.75, 95% CI = 0.55-4.23, *P* = 0.2). The association between tamoxifen use and an increase in endometrial cancer incidence in *BRCA* pathogenic variant carriers has been confirmed in a subsequent case-control study undertaken by the same group, which found a 6.21-fold increase in risk compared with non-users (95% CI = 2.21-17.5, *P* = 0.0005), which the authors suggested could provide an explanation for the observed association (Segev *et al*., 2015). These findings were not, however, replicated in the present study.

Whilst the same authors also described a lower incidence of endometrial cancer in women who underwent oophorectomy for any reason, this has not been confirmed in other cohorts of women who have undergone specific RRSO, that is, in the absence of any tubo-ovarian disease (Shu *et al*., 2016). A beneficial effect of RRSO may have been anticipated if serous endometrial cancers originated in the fallopian tube. The fact that in our cohort only one case of mixed serous and endometrioid endometrial cancer was diagnosed means that we are unable to provide any robust data to confirm or refute this hypothesis, except to state that this case occurred in a woman who had not undergone RRSO and no cases of endometrial cancer were diagnosed in the RRSO cohort.

The number of endometrial cancer cases observed in each of the cohort studies, including our own, has been small (2-17) and could well explain the difference in statistical

significance of SIRs that all approximate to a value of 2. Indeed, only two cases of endometrial cancer were diagnosed in our prospective cohorts, which may indicate a testing bias in those with endometrial cancer in previously reported studies. An unbiased assessment of testing first-degree relatives with only endometrial cancer supports our premise that there is unlikely to be any substantial increase in endometrial cancer risk. It is, however, arguable that even if an SIR of 2 is validated (none of our upper confidence limits exclude this), the level of risk is insufficient to recommend hysterectomy at the time of bilateral salpingo-oophorectomy as a risk-reducing measure. Given the increased potential morbidity associated with more extensive surgery, evidence of benefit is certainly warranted to outweigh these additional risks. Whether there is a clear benefit in specific subgroups of *BRCA* pathogenic variant carriers is currently unknown; neither ours nor previously published studies have included data on body mass index (BMI) and hence have been unable to adjust for this in analyses.

The strengths of this study include the confirmation of *BRCA* diagnoses with histological reports and contemporaneous expert pathological review of slides, although unfortunately this was not universally achievable due to the lack of availability of tumour tissue for assessment. The study also included the largest number of *BRCA2* pathogenic variant carriers to date, increasing our understanding of endometrial cancer risk in this specific population. As with previous studies, robust methodology has been employed to compare observed with national expected endometrial cancer rates, with adjustment made for local hysterectomy rates. We were able to include data on *BRCA1/2* pathogenic variant status of first-degree relatives of women who developed endometrial cancer and of unselected serous endometrial cancer cases to corroborate our findings. Our somatic *BRCA* testing has been shown to have high sensitivity in high-grade serous ovarian cancer (Kotsopoulos *et al*., 2018).

The potential lack of power in our study is a limitation and one that we have attempted to address by contacting the EMBRACE study (Easton D) to ensure endometrial cancer cases

have not been missed. It does mean that our ability to detect differences in endometrial cancer risk in any specific subgroups or subtype of endometrial cancer is limited. Whilst the length of the follow-up, particularly for the retrospective cohort, is a clear advantage of this work, the median age at censoring remains younger than 50 years, well below the average age of endometrial cancer diagnoses in the UK (CRUK, 2020). Re-analysis of the data at a later date will be performed to increase the duration of the follow-up and potentially the number of endometrial cancer diagnoses. Subgroup analyses based upon tamoxifen use was limited due to the fact that two thirds of women in the database did not have data collected on their exposure to the drug, although the vast majority of women with a pathogenic variant in a *BRCA* gene without a history of breast cancer were not known to have taken tamoxifen. The low prevalence of tamoxifen use within the cohort may well explain why no association was observed between tamoxifen use and an increase in endometrial cancer risk. Additional efforts to reduce the amount of missing data within our data set will be made to address this.

Data were unfortunately not routinely collected on hormone replacement therapy (HRT) use by women who underwent RRSO and so the impact of this on subsequent endometrial cancer risk could not be assessed. Whilst oestrogen-only HRT may be associated with a lower rate of subsequent breast cancer (Kotsopoulos *et al*., 2018; Marchetti *et al*., 2018), this must be balanced against the impact this could have on the risk of malignant changes within the endometrium. Whilst every attempt was made to undertake histological review of all endometrial cancer cases occurring within the cohort, this was unfortunately not possible for eight cases where slides were unavailable. Four of these cases were also operated on at another hospital and, as it was not possible to retrieve the pathology reports for these tumours, this may impact upon the results of our subgroup analysis of serous-like endometrial cancers.

5.7 Conclusion

In conclusion, *BRCA1* and *BRCA2* pathogenic variant carriers do not appear to be at a significant increased risk of endometrial cancer compared with the general population. Neither does there appear to be a specific association between *BRCA1/2* pathogenic variants and serous endometrial cancer. Women and clinicians should be reassured that hysterectomy at the time of bilateral salpingo-oophorectomy is unlikely to be of benefit if performed solely for the purpose of trying to reduce subsequent endometrial cancer risk.

5.8 References

Breast Cancer Linkage Consortium (1999). Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst*, 91(15):1310–6.

Breslow, N. E. and Day, N. E. (1987). Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci Publ*, (82):1–406.

Centre for Cancer Genetic Epidemiology (2017). EMBRACE Study.

CRUK (accessed: Oct 2020). Uterine cancer statistics. `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer`.

de Jonge, M. M., Mooyaart, A. L., Vreeswijk, M. P., *et al.* (2017). Linking uterine serous carcinoma to BRCA1/2-associated cancer syndrome: A meta-analysis and case report. *Eur J Cancer*, 72:215–225.

de Jonge, M. M., Ritterhouse, L. L., de Kroon, C. D., *et al.* (2019). Germline BRCA-Associated Endometrial Carcinoma Is a Distinct Clinicopathologic Entity. *Clin Cancer Res*, 25(24):7517–7526.

Ellison, G., Huang, S., Carr, H., *et al.* (2015). A reliable method for the detection of BRCA1 and BRCA2 mutations in fixed tumour tissue utilising multiplex PCR-based targeted next generation sequencing. *BMC Clin Pathol*, 15:5.

Evans, D. G. R., Harkness, E. F., Brentnall, A. R., *et al.* (2019). Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. *Breast Cancer Research and Treatment*, 176(1):141–148.

Hamilton, C. A., Cheung, M. K., Osann, K., *et al.* (2006). Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br J Cancer*, 94(5):642–6.

Hornreich, G., Beller, U., Lavie, O., *et al.* (1999). Is uterine serous papillary carcinoma a BRCA1-related disease? Case report and review of the literature. *Gynecol Oncol*, 75(2):300–4.

Hu, Y. C., Zhang, Q., Huang, Y. H., *et al.* (2014). Comparison of two methods to extract DNA from formalin-fixed, paraffin-embedded tissues and their impact on EGFR mutation detection in non-small cell lung carcinoma. *Asian Pac J Cancer Prev*, 15(6):2733–7.

Kaplan, E. J., Caputo, T. A., Shen, P. U., *et al.* (1998). Familial papillary serous carcinoma of the cervix, peritoneum, and ovary: a report of the first case. *Gynecol Oncol*, 70(2):289–94.

Kitson, S. J., Maskell, Z., Sivalingam, V. N., *et al.* (2019). PRE-surgical Metformin In Uterine Malignancy (PREMIUM): a Multi-Center, Randomized Double-Blind, Placebo-Controlled Phase III Trial. *Clin Cancer Res*, 25(8):2424–2432.

Kotsopoulos, J., Gronwald, J., Karlan, B. Y., *et al.* (2018). Hormone Replacement Therapy After Oophorectomy and Breast Cancer Risk Among BRCA1 Mutation Carriers. *JAMA Oncol*, 4(8):1059–1065.

Landrum, M. J., Lee, J. M., Benson, M., *et al.* (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, 46(D1):D1062–d1067.

Lavie, O., Ben-Arie, A., Segev, Y., *et al.* (2010). BRCA germline mutations in women with uterine serous carcinoma–still a debate. *Int J Gynecol Cancer*, 20(9):1531–4.

Lavie, O., Hornreich, G., Ben Arie, A., *et al.* (2000). BRCA1 germline mutations in women with uterine serous papillary carcinoma. *Obstet Gynecol*, 96(1):28–32.

Lee, Y. C., Milne, R. L., Lheureux, S., *et al.* (2017). Risk of uterine cancer for BRCA1 and BRCA2 mutation carriers. *Eur J Cancer*, 84:114–120.

Levine, D. A., Lin, O., Barakat, R. R., *et al.* (2001). Risk of endometrial carcinoma associated with BRCA mutation. *Gynecol Oncol*, 80(3):395–8.

Marchetti, C., De Felice, F., Boccia, S., *et al.* (2018). Hormone replacement therapy after prophylactic risk-reducing salpingo-oophorectomy and breast cancer risk in BRCA1 and BRCA2 mutation carriers: A meta-analysis. *Crit Rev Oncol Hematol*, 132:111–115.

Reitsma, W., Mourits, M. J., de Bock, G. H., and Hollema, H. (2013). Endometrium is not the primary site of origin of pelvic high-grade serous carcinoma in BRCA1 or BRCA2 mutation carriers. *Mod Pathol*, 26(4):572–8.

Sagae, S., Susumu, N., Viswanathan, A. N., *et al.* (2014). Gynecologic Cancer InterGroup (GCIG) consensus review for uterine serous carcinoma. *Int J Gynecol Cancer*, 24(9 Suppl 3):S83–9.

Saule, C., Mouret-Fourme, E., Briaux, A., *et al.* (2018). Risk of Serous Endometrial Carcinoma in Women With Pathogenic BRCA1/2 Variant After Risk-Reducing Salpingo-Oophorectomy. *J Natl Cancer Inst*, 110(2).

Segev, Y., Iqbal, J., Lubinski, J., *et al.* (2013). The incidence of endometrial cancer in women with BRCA1 and BRCA2 mutations: an international prospective cohort study. *Gynecol Oncol*, 130(1):127–31.

Segev, Y., Rosen, B., Lubinski, J., *et al.* (2015). Risk factors for endometrial cancer among women with a BRCA1 or BRCA2 mutation: a case control study. *Fam Cancer*, 14(3):383–91.

Shu, C. A., Pike, M. C., Jotwani, A. R., *et al.* (2016). Uterine Cancer After Risk-Reducing Salpingo-oophorectomy Without Hysterectomy in Women With BRCA Mutations. *JAMA Oncol*, 2(11):1434–1440.

Thompson, D. and Easton, D. F. (2002). Cancer Incidence in BRCA1 mutation carriers. *J Natl Cancer Inst*, 94(18):1358–65.

5.9 Supplementary Data

Table S5.3: *Characteristics of patients diagnosed with endometrial cancer.*

Pt	Prospective/ retrospective EC diagnosis	Age at start of follow-up	*BRCA1/2*	History of BC?	History of prior tamoxifen use?	Previous RRSO?	Time*	Histology	FIGO (2009) stage	Status	Pathology review?	Additional details
1	Prospective	54.5	*BRCA1*	Yes	No	No	2.3	Endometrioid	1b	Dead	Yes	Synchronous diagnosis of stage 3c ovarian cancer. Death related to ovarian cancer
2	Prospective	50.2	*BRCA2*	Yes	No	No	3.0	Endometrioid	1b	Alive	No	
3	Retrospective	42.0	*BRCA1*	No	No	No	13.4	Adenocarcinoma	3	Dead	Yes	Morphology in keeping with endometrial origin. No subtype information available
4	Retrospective	40.4	*BRCA1*	Yes	Unknown	No	25.3	Unknown	Unknown	Dead	No	Unable to locate histology report
5	Retrospective	36.5	*BRCA2*	Yes	Yes	No	16.4	Endometrioid	1a	Alive	No	
6	Retrospective	37.8	*BRCA2*	No	No	No	17.7	Endometrioid	2	Alive	No	
7	Retrospective	41.9	*BRCA1*	Yes	Unknown	No	16.0	Endometrioid	1a	Alive	Yes	
8	Retrospective	41.5	*BRCA2*	Yes	Yes	No	27.7	Endometrioid	Unknown	Alive	No	Unable to confirm stage
9	Retrospective	39.9	*BRCA2*	Yes	Yes	No	33.3	Mixed serous and endometrioid	1a	Alive	Yes	
10	Retrospective	40.5	*BRCA1*	No	No	No	29.9	Unknown	Unknown	Deceased	No	Unable to locate histology report. Simultaneous ovarian cancer diagnosis. Death related to ovarian cancer
11	Retrospective	52.7	*BRCA2*	Yes	Unknown	No	23.6	Carcinosarcoma	3c	Deceased	Yes	Death related to breast cancer
12	Retrospective	37.0	*BRCA2*	No	No	No	23.4	Unknown	Unknown	Alive	No	Unable to locate histology report
13	Retrospective	23.6	*BRCA1*	No	No	No	26.0	Endometrioid	Unknown	Deceased	Yes	Tubal tumour deposit reported on original report, but slides unavailable for review
14	Retrospective	20.0	*BRCA1*	Yes	Unknown	No	31.6	Carcinosarcoma	Unknown	Alive	No	Unable to locate slides for expert review

* Time from start of follow-up to diagnosis (yrs).
EC, endometrial cancer; Pt, patient; RRSO, risk-reducing salpingo-oophorectomy.

6 Thesis Discussion and Conclusion

6.1 Thesis Overview

Development of a personalised risk score for endometrial cancer emerged as the most important research priority from the James Lind Alliance Priority Setting Partnership survey (Wan *et al*., 2016). Genetic risk factors, aside from family history, have largely been left out from the existing endometrial cancer RPMs. Better precision of the risk prediction tools can be achieved by expanding the scope of the model with the inclusion of genetic scores, such as a PRS. This will allow personalised or targeted screening to be established which consequently would enable efficient use of our resources, reduce the burden of cancer, avoid over- and undertreatment, and implement effective prevention strategies to high-risk women. Therefore, this PhD study aimed at investigating the utilisation of genomics tools to aid accurate risk prediction of endometrial cancer.

Firstly, evidence for the role of SNPs in endometrial cancer predisposition and the most robust SNPs from only the high-quality data were collated through a systematic review of literature. This was undertaken to construct a panel of SNPs which would form the basis of a future PRS and to highlight potential gaps in current knowledge and research on the polygenic basis regarding endometrial cancer. The use of a systematic review allowed for retrieval of a much larger number of studies which otherwise might have been missed. Then, hundreds of papers identified were examined meticulously and included in our analysis provided they met certain pre-determined criteria, ensuring that only the SNPs with the best evidence were considered for our panel.

Next, using a carefully matched cases and controls, we genotyped a previously untested, enriched cohort to provide us a basis for; i) validation of the SNP panel (targeted analysis), ii) identify new risk regions (untargeted analysis), and iii) to develop a novel, extended PRS to aid the much-needed risk prediction for endometrial cancer. For this body of work,

we used the OncoArray Consortium custom chip, which comprised of nearly 534,000 typed SNPs, and followed a validated, harmonised, and robust GWAS QC and analysis workflow developed by the consortium (Amos *et al.*, 2017). The wide coverage of the OncoArray chip enabled us not only to directly acquire nearly half a million SNPs, but also to impute millions more across the entire genome. By using these vast data, coupled with extensive clinical information available for the participants of our study, we were able to construct an extended PRS in a representative population. We then applied the PRSs that we developed to external datasets. To do this, we used data from the April 2019 release of UKBB and latest ECAC GWAS (excluding UKBB participants) published in 2018.

Lastly, pathogenic *BRCA1/2* mutations have been investigated by others in relation to endometrial cancer risk, particularly of the serous subtype. Despite these efforts, contradicting results were presented, all of which were utilising retrospectively collected datasets. Due to this uncertain relationship, clinical guidelines do not encourage hysterectomy at the time of RRSO in *BRCA1/2* pathogenic carriers for risk reduction. Thus, we investigated this phenomenon in a large, prospectively maintained cohort by comparing the endometrial cancer prevalence to the national rates. We investigated the effect of HRT and tamoxifen use on the results. We also sequenced the two *BRCA* genes in available serous tumours identified from our well-phenotyped Manchester cases.

In the next section, each chapter is discussed with respect to its novel contribution to the literature in this field and strengths. Overall limitations of this thesis are then discussed in a separate section followed by possible future research directions based on the body of work presented in this thesis.

6.2 Discussion

6.2.1 Systematic Review of Literature

Though several systematic reviews and meta-analyses were previously published, often focusing on a single or a small number of SNP(s) or gene(s), our *blanket* systematic review

of literature spanning the past decade was a first of its kind in the field of endometrial cancer. Our systematic review provided the most comprehensive evidence for endometrial cancer susceptibility variants by considering any SNP reported by all studies meeting our inclusion criteria. Furthermore, the systematic review critically appraised the strengths and weaknesses of the relevant studies and emphasized the need for large-scale, clinically-rich studies to identify more variants of importance and validate these associations. Finally, the systematic review concluded and summarised the most robust 24 SNPs that affect endometrial cancer risk and provided theoretical evidence supporting their potential use in a PRS.

6.2.2 Manchester GWAS

GWAS is a preferred method over candidate-gene association studies and has a great potential to identify SNPs associated with the risk of a given disease. The GWAS approach is largely responsible for discoveries of genetic risk factors for polygenic or complex diseases such as cancer. However, in regard to endometrial cancer, the lack of detailed participant information in GWAS to date and their use of overlapping datasets have been highlighted by our systematic review. Thus, our Manchester GWAS comprising previously untested and sociodemographically matched cases and controls with generally complete clinical data provided a valuable source to validate the previously reported SNPs, identify new associations and an independent cohort to develop a novel PRS for endometrial cancer risk prediction. In summary, six suggestive regions were identified by our GWAS, yet validation of these in larger GWAS is imperative. Moreover, we validated a substantial number of SNPs identified by our systematic review. Validation of further SNPs of suggestive significance identified from the latest ECAC GWAS was also investigated here. In total, of the 72 SNPs we sought to validate, nearly two thirds had matching effect direction and ten were nominally significant in our study. Development of an independently validated PRSs would not have been possible without our new Manchester case-control study.

It is worth noting that our Manchester study control participants were followed-up for a

median 8.9 years, which enabled us to identify participants who developed endometrial or breast cancer during follow up. Any individual who later developed endometrial cancer and had available DNA was added to our cases. Similarly, to avoid introduction of allele bias, all participants who later developed breast cancer were excluded from our dataset. This is particularly important because unlike previous GWAS, we minimised any potential bias by capturing cases and controls more accurately. This has a great impact on SNP EAF distribution between cases and controls, consequently affecting the direction and magnitude of the effect of association.

6.2.3 Development and Validation of PRS

As genotype is determined and fixed at conception, a risk factor based on multiple low-risk alleles combined in a PRS provides a reliable indication of overall risk. Unlike most of the widely used risk factors, such as BMI which fluctuates throughout life, a well-calibrated and validated PRS has the potential to be used in any age group at any time of life. Given the reduction in costs associated with genotyping over the recent years a stable risk score that stays constant through life and is personalised, is very much needed to halt and reverse the growing burden of endometrial cancer across nations.

Using our deeply phenotyped cohort, we developed and evaluated several PRSs, which we then attempted to validate in two other cohorts. Firstly, in Manchester cohort, we developed four PRSs, which performed moderately. We then refined the SNP panel by stepwise addition to the PRS and assessing individual benefit. The refined PRS, PRS40, performed the best with an AUC of 0.62 and women in the top 33% of this PRS were at over 2.4-fold increased risk of endometrial cancer compared to women in the bottom 33%. The AUC achieved here as a single risk factor was similar to some of those reported by the multi-variable RPMs. To avoid overestimation, we did not include the six Manchester GWAS-identified SNPs in this approach. Therefore, we hypothesise that our new PRS, curated with externally sourced SNPs, will aid risk prediction efforts by improving the accuracy of RPMs.

By contrast, our population level validation of the various PRSs in UKBB showed far lower discriminatory ability. This is likely a result of the well documented healthy volunteer effect within the UKBB (Fry *et al*., 2017). Moreover, PRSs are often reported to perform worse in population settings (Lewis and Vassos, 2020). The results we obtained with UKBB were therefore to be expected. Finally, we attempted to validate the PRSs in the largest available GWAS cohort, ECAC, which performed relatively well (the limitations regarding this are discussed in the next section). Despite the publication of two recent articles reporting an endometrial cancer PRS each, both were limited in terms of their methodological approach and the cohorts they utilised (e.g. UKBB) (Choi *et al*., 2020; Fritsche *et al*., 2020). Hence, our extended SNP panel comprised of 78 SNPs is the first time this comprehensive panel was used to develop, refine, and validate an endometrial cancer PRS.

6.2.4 Pathogenic *BRCA* Mutations and Endometrial Cancer Risk

Pathogenic mutations in the DNA repair-related *BRCA* genes have attracted significant interest from the research community in relation to the elevated hereditary breast and ovarian cancer risk conferred to carriers. An increased risk of endometrial cancer, particularly of the serous or serous-like subtype, in pathogenic *BRCA1/2* carriers have been postulated by others (Thompson and Easton, 2002; Lavie *et al*., 2010; Segev *et al*., 2013), often reaching contradictory conclusions (Levine *et al*., 2001; Breast Cancer Linkage Consortium, 1999) in retrospective cohorts. Our study on the other hand, using a prospectively maintained database, found no evidence for a significantly increased risk of all or serous-like endometrial cancer compared to the general population.

We compared the endometrial cancer diagnoses in *BRCA* pathogenic variant carriers to SIR calculated in four distinct cohorts using national endometrial cancer rates. We found that, regardless of the *BRCA* gene studied, the risk of endometrial cancer cases

was no higher than it is at the population level. Moreover, previous diagnosis of breast cancer and use of tamoxifen, which may have contributed to the positive association seen in other studies, did not alter the results. Finally, we sequenced the *BRCA* genes in 15 serous endometrial cancer tumours to detect pathogenic variants. The somatic tumour sequencing was undertaken as a pragmatic strategy to overcome the lack of ethics coverage for germline sequencing in these participants. Absence of pathogenic variants in tumour tissue was considered a proxy for absence of pathogenic germline mutations in *BRCA1/2*. Indeed, we found no pathogenic variants or variants of clinical significance. If pathogenic variants were to be detected, this would provide sufficient evidence for seeking extension of the ethics approval to allow for germline sequencing of *BRCA* genes in this cohort of patients.

If we were to observe an increased risk of endometrial cancer in pathogenic *BRCA* variant carriers, this information could have been useful not only in terms of risk-reduction, where this patient group would have been offered regular screening or elective hysterectomy, but also in terms of risk prediction. The latter would have been possible by the addition of *BRCA* mutational status testing to the RPM. Considering that the presumed risk with *BRCA* variants related to serous or serous-like endometrial tumours, the model could potentially be used to predict the risk of developing this particular aggressive tumour subtype for which currently age is the only predictor. Thus, our study provides valuable insight into the relationship between *BRCA* genes and endometrial cancer risk, concluding that there seems to be no apparent risk and hysterectomy at the time of RRSO is unlikely of any benefit to the pathogenic variant carriers.

6.3 Limitations

Despite our best attempts to strengthen our methods and use the best data, approach and/or resources available, our studies were still susceptible to methodological and theoretical limitations.

First, while we identified over a hundred studies that were analysed for the systematic review, due to overlapping and high heterogeneity between the datasets, we were not able to produce further SNPs by compiling the smaller studies into a meta-analysis. This resulted in a relatively small number of SNPs being identified by our analysis, and only 19 of those were genome-wide significant, including the *AKT1* variant for which the association was not replicated in the latest ECAC GWAS. Moreover, due to a paucity of large-scale GWAS in this field, most of the SNPs we identified were reported by the same study (ECAC GWAS). Similarly, although we considered candidate-gene studies in our systematic review, due to lack of validation of their results in large GWAS, they were not included in our final list of SNPs. Finally, because of the limited results we initially obtained, we chose to lower our SNP inclusion threshold to the suggestive significance level of $P < 1E\text{-}5$. However, despite this limitation, there is evidence that SNPs of suggestive significance may still be useful in a PRS (Mavaddat *et al*., 2019). Our theoretical PRS calculations comparing the RR between the 19 genome-wide significant SNP panel and the 24-SNP panel inclusive of the suggestive SNPs, supported this notion by showing an improvement in the RR with 24-SNP panel.

Second, while we conducted a GWAS using robust and validated methods in a rather homogeneous and carefully collated cohort, our small size was one of the main limitations of our GWAS. Due to this, we were unable to identify any genome-wide significant SNPs and the suggestive SNPs we identified here have to be interpreted with caution until validation by larger GWAS. Moreover, despite our best efforts to conduct subtype-specific analyses, even smaller sample sizes for each of these, 415 and 136 per subtype, further reduced the study power required for meaningful interpretation of the results. Finally, possibly for the same reason, we also were unable to validate all SNPs identified by our systematic review, irrespective of their discovery GWAS significance.

Regarding our PRS study, the performances we obtained were likely a direct result of different allele frequency distributions and, thus, effect size and/or direction observed

between the discovery and Manchester GWAS. The opposite effect direction is a result of reversed genotype distribution between the cases and controls. We expect this to affect the lower PRS performances we have seen in Manchester and UKBB cohorts, compared to ECAC, where most of the SNPs were identified from. Although it is unlikely that all SNPs we used, particularly those with suggestive significance, would be informative in a PRS, we believe that this limitation may have attenuated the discriminatory ability of 72-SNP PRS in our Manchester cohort.

The performance of our different PRS constitutions in UKBB were rather disappointing. This is highly likely due to the well-documented healthy volunteer bias in UKBB. Moreover, we showed that the UKBB cases were younger and leaner than our Manchester cases. There is a possibility of further bias introduced through the use of different genotype chips and imputation reference panels by the UKBB researchers to generate the data we obtained. However, this is purely a speculation at this point, though, in my opinion warrants further scrutiny. In terms of our validation efforts, because of the vast differences between our Manchester and UKBB cohorts, we used available ECAC data to validate our PRSs. However, the majority of the SNPs we used here were identified from ECAC GWAS and this limits the validity of our ECAC PRS results. Moreover, unlike our Manchester cohort, we did not have any information on the hysterectomy history of the controls in ECAC. Therefore, in spite of the success of our PRS in the Manchester cohort, further validation in a similarly enriched cohort is needed to confirm our results. I believe that while our results are nevertheless encouraging, genetic risk prediction efforts for endometrial cancer still require substantial research before achieving clinical utility.

In general, the aim of the PRSs generated is to aid clinical decision making by personalising preventive and/or treatment measures. In the case of endometrial cancer, the aim in the first instance would be to utilise the PRS alongside other risk factors to accurately predict women who are at high-risk of developing the disease. High-risk women could then be offered interventions such as weight loss or management options for those

presenting with elevated weight, regular endometrial cancer screening, hormonal therapy such as IUS, and hysterectomy where necessary. However, adoption of PRSs in healthcare settings thus far has not been widespread, largely owing to a lack of immediate benefit for certain diseases and discrepant predictive accuracy across ethnicities.

Clinical implementation of our endometrial cancer PRS is currently unattainable, as is the case for the vast majority of existing PRSs for other diseases. Most importantly, in our Manchester cases and controls, the 19-SNP PRS provided 8% advantage over if it were due to chance only (i.e. 58% vs 50%). Therefore, this PRS would not be ready to be used in clinical practice until improved by further addition of risk-influencing SNPs to improve prediction capacity not only in Europeans but also in other ancestries and validated in other independent datasets. Even so, in the first instance, clinical utility may only be achieved when the PRS is successfully incorporated into an RPM, rather than a stand-alone prediction tool and in risk-based or hospital settings rather than in the general population.

As mentioned, a major disadvantage for any PRS would be the limited applicability of the PRSs in the general population. As many of the nations across the modern world are of multi-ethnic societies, the existing PRSs which have largely been developed and tested in European ancestries would be comparatively less effective in persons of other ethnicities. Given that the variants which we utilised in our PRS were discovered in European populations, the performance of the PRS cannot be expected to perform equally in non-European ancestries and thus, its applicability at the population level is unknown.

PRSs for other diseases such as breast cancer have been also useful in disease subtyping. While our endometrial cancer PRS showed similar performance in both of the traditional endometrial cancer subtypes, due to our inability to validate this outside of our Manchester dataset, it is unclear whether this phenomenon would be replicated in other studies. Therefore, given the rudimentary state of ours and published PRSs for endometrial cancer,

I highly doubt we will see the NHS or indeed other healthcare organisations adopting them for prediction or prevention purposes in the near future and certainly not until the predictive performance is replicated consistently by us and/or others.

Lastly, our *BRCA* study investigating the relationship between pathogenic variants in *BRCA1/2* genes and increased risk of endometrial cancer did not find an elevated risk in a prospective cohort. It is important to note that BMI was not adjusted for in the analyses of this and previous studies. This hinders our ability to draw meaningful conclusions on whether hysterectomy could be of any benefit in a specific group of patients such as those with increased BMI. Another limitation of our study is the lack of power, which we attempted to minimise by confirming the completeness of data. Due to low study power, we may not have been able to detect endometrial cancer risk differences in any patient or tumour subgroups. Moreover, the median age at censoring in our study is well below the national average of endometrial cancer diagnoses. Thus, analysis of our prospectively maintained database may need to be repeated in future to lengthen the follow-up duration which may increase the number of endometrial cancers identified.

Subgroup analysis looking at tamoxifen use was possible, though this data was not complete. However, we expect this to prove a minimal limitation due to the fact that chemoprevention in *BRCA* carriers is not a standard protocol thus we do not expect the majority of the women without a history of breast cancer to have taken tamoxifen. Similarly, our data was incomplete in regard to the use of HRT by women who underwent RRSO. Therefore, our ability to examine the risk conferred by HRT use was not possible. Lastly, despite our best attempts, we were not able to acquire pathology reports or histology slides for some of the endometrial cancer cases, limiting our ability to determine the tumour subtype for all cases.

6.4 Future Work and Conclusion

Further research can be proposed and undertaken based on the body of work presented in this thesis. Based on our systematic review, we made it clear that at present, the number of genome-wide significant SNPs associated with endometrial cancer risk is rather limited as is the availability of large-scale, independent, diverse and clinically-detailed cohorts to conduct high-quality GWAS. Yet, others estimate that over a thousand undiscovered SNPs are associated with endometrial cancer (Zhang *et al.*, 2020), and this provides insight into the potential increase we will be seeing in the number of SNPs reported in the future. Appropriately, the continual expansion of ECAC, such as incorporation of our representative and well-matched Manchester GWAS into their next meta-GWAS, will undoubtedly result in identification of further SNPs reaching the genome-wide significance threshold. The vast data available to us, and others, as a result of the wide-coverage genotype chips available as well as the use of efficient imputation tools, make it possible to extract further genotype data to expand the PRS in the future, should there be sufficient evidence of an association.

Due to small sample sizes, unavailability of data on tumour subtype and lack of harmonised classification systems used for tumour histotypes between studies, no genome-wide significant SNP has been identified for non-endometrioid subtype of endometrial cancer. Utilisation of data such as ours as part of ECAC, for instance, may provide sufficient numbers for the subtype specific analyses to gain sufficient power to detect signals specific for this subtype at the genome-wide significance level. It would be interesting to construct the PRS separately for each subtype and examine its performance by incorporating subtype-specific SNPs in each. Although we have shown here that our proposed PRS works well in both of the major subtypes of endometrial cancer, it would be interesting to see whether the slightly lower discriminatory performance seen in non-endometrioid cases were due to the smaller size of the subset of patients, or non-specificity of the SNPs comprising the PRS.

Women of African heritage are disproportionately affected with aggressive and high-grade endometrial cancer (Francies *et al*., 2020). However, large-scale endometrial cancer GWAS have primarily focused on European cohorts, and thus it is unclear if and which of the genome-wide significant SNPs reported to date are associated with endometrial cancer risk in other ancestries. Further research investigating the association between both with SNPs with known associations, but also ancestry-specific SNPs is of critical importance. All SNPs used in our PRS were identified in Europeans. For that reason, whether the PRS would be applicable to other ethnicities is currently unknown. Therefore, it is crucial to extend the efforts across to different ancestries to be able to tackle the risk of endometrial cancer effectively in all women in any population.

Research groups studying various diseases including other cancers have started to incorporate PRSs into RPMs to improve risk prediction efforts (Lakeman *et al*., 2020; Elliott *et al*., 2020). We have shown that a PRS comprised of 40 SNPs curated from the best available evidence at the time of study conception has moderate risk stratification ability in a hospital-based and locally representative cohort. If this performance were to be validated independently, the next logical step, in my opinion, would be to test the utility of the PRS as part of an RPM to examine whether the accuracy of the model will be improved with the addition of the PRS. Importantly, this will require careful adjustment and calibration of the model so that the weight of the PRS is not overestimated.

For our PRS, we relied on known genome-wide or suggestive SNPs identified primarily of a single source (ECAC). As mentioned before, current predictions estimate over a thousand SNPs to be of importance in endometrial cancer risk. Therefore, I think that using more complex and comprehensive tools such as LDpred may be of important use to determine the most predictive SNPs and expand and/or refine the PRS in a more reliable manner (Marquez-Luna *et al*., 2020; Prive *et al*., 2020).

Unfortunately, we did not have the data for family history of endometrial cancer or self-diagnosis or family history of LS available for our controls. In the future, where this data is available for the cohort studied, the correlation between these and PRS performance should be investigated. This would be of particular importance to determine whether the PRS is of any benefit in these certain groups, and whether adjustment of the RPM is necessary when both risk factors are present.

Future research efforts could also be directed at devising polygenic hazard scores, which aim to identify individuals with genetic risk in an age-specific manner. Identifying older women at risk of developing endometrial cancer is of particular clinical importance because endometrial cancer is far more common in older women than it is in young ones. Similar to endometrial cancer, prostate cancer is more commonly diagnosed in older men and polygenic hazard scores for prostate cancer have been effective (Huynh-Le *et al.*, 2020; Karunamuni *et al.*, 2021). Therefore, I expect this approach to be of benefit in the context of endometrial cancer risk prediction as well as guiding targeted screening strategies.

Whilst our PRS requires further validation and refinement, it nevertheless presents a promising start for which improvements can be introduced upon. As mentioned above, these could and, ideally, should include expansion of the SNP portfolio, validation in larger, multi-ethnic and independent datasets, investigating and correcting the SNP panel and effect sizes to match various ethnicity requirements, and implementing the PRS into an RPM for obtaining a more accurate prediction tool. Based on the final form of the PRS, there could be two main choices for obtaining the PRS. A routine blood draw, or even a self-collected saliva sample, could be used for the purpose of genotyping. Then, genotyping could be achieved by either using a scalable, smaller-scale and targeted method such as the iPLEX MassARRAY system (Agena Bioscience), or a commercially available, wide-coverage chip such as the cancer-specific OncoArray (Illumina) used in this study. The former is cheaper and, thus, more appropriate for a PRS containing only a small number of SNPs such as PRS19. The latter, on the other hand, while overall has

a higher cost per chip, could be just as cost-effective considering its wide applicability for generating multiple PRSs for several cancers and other diseases by using the same chip where all consumables, staff salaries and computational costs of the downstream analyses up to the point of specific PRS generation would be shared. Furthermore, using a wide-coverage array would allow later modifications to the PRS should the need arise. Finally, newer genotyping methods, such as genotyping-by-sequencing which is considered more cost-effective than array genotyping, can detect novel genotypes and is more reliable in rare variant genotyping, may be a more appropriate method of choice at a later time should a more comprehensive PRS for endometrial cancer become ready for use in the clinical practice.

According to Pennington *et al*., endometrial cancer associated costs per person within two years post-diagnosis totals over £9,000 on average in the United Kingdom, for all stages including AEH (Pennington *et al*., 2016). When considering the more aggressive disease stages III and IV, this figure rises above £18,600 and reaches nearly £17,000, respectively. According to the authors, diagnosis and surgery alone costs approximately £5,700 per person. This poses a critical financial problem for the NHS, in view of the projected continued increase in endometrial cancer incidence and mortality rates in the coming years. In comparison, ever-continuing progress in the reduction of costs associated with genotyping and indeed other forms of genetic screening methods, and the interusability of these methods to generate multiple PRSs or genetic scores from the same material will prove highly advantageous. Therefore, although currently the results presented here and published by others do not provide sufficient support for immediate clinical implementation of a PRS, in the future, its investment value as part of actionable prediction and prevention efforts will be beneficial in levelling down clinical costs associated with endometrial cancer diagnosis and treatment by reducing the number of blanket surgeries or diagnostic procedures that would otherwise be required.

Finally, because we had incomplete data collected for the use of tamoxifen or HRT in

our *BRCA* database, it would be sensible to repeat these analyses in another dataset where there is complete or near-complete data available to be able to draw meaningful conclusions. Furthermore, our date of censoring was below the national average and at the time of writing, this would have limited our ability to identify endometrial cancer cases that will develop later due to the age-dependent nature of endometrial cancer. Thus, reanalysing the data at a later point in time will no doubt provide us with a more robust dataset to reaffirm our conclusions.

In summary, this body of work highlights the promise of genomic tools and resources to support our understanding of endometrial cancer aetiology and augment risk prediction strategies. It has shown the potential of low-risk susceptibility variants, combined into a PRS as presented in this thesis, that can be used to identify women at high-risk of developing endometrial cancer. Importantly, the PRS seems to be applicable to both endometrioid and non-endometrioid cases. Unfortunately, complete validation of the PRS was not possible, however, the data presented here also indicate that the PRS, when validated and/or improved, may be more useful in risk-based clinical settings than at the population level. Lastly, it has shown that carrying pathogenic *BRCA* mutations increases the risk of developing neither the overall nor the serous subtype endometrial cancer and thus is of no benefit for risk prediction and/or reduction purposes.

7 References

Ahima, R. S. (2008). Revisiting leptin's role in obesity and weight loss. *Journal of Clinical Investigation*, 118(7):2380–2383.

Alblas, M., Velt, K. B., Pashayan, N., *et al.* (2018). Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. *Critical Reviews in Oncology/Hematology*, 126:92–99.

Aminimoghaddam, S., Shahrabi-Farahani, M., Mohajeri-Tehrani, M., *et al.* (2015). Epistatic interaction between adiponectin and survivin gene polymorphisms in endometrial carcinoma. *Pathology Research and Practice*, 211(4):293–297.

Amos, C. I., Dennis, J., Wang, Z. M., *et al.* (2017). The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiology Biomarkers & Prevention*, 26(1):126–135.

Ashton, K. A., Proietto, A., Otton, G., *et al.* (2009a). Estrogen receptor polymorphisms and the risk of endometrial cancer. *Bjog-an International Journal of Obstetrics and Gynaecology*, 116(8):1053–1061.

Ashton, K. A., Proietto, A., Otton, G., *et al.* (2009b). Polymorphisms in TP53 and MDM2 combined are associated with high grade endometrial cancer. *Gynecologic Oncology*, 113(1):109–114.

Auton, A., Abecasis, G. R., Altshuler, D. M., *et al.* (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

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(11):4229–42.

Bafligil, C., Thompson, D. J., Lophatananon, A., *et al.* (2020). Association between genetic polymorphisms and endometrial cancer risk: a systematic review. *Journal of Medical Genetics*, 57(9):591–600.

Banno, K., Yanokura, M., Iida, M., *et al.* (2014). Carcinogenic mechanisms of endometrial cancer: Involvement of genetics and epigenetics. *Journal of Obstetrics and Gynaecology Research*, 40(8):1957–1967.

Barnes, D. R., Rookus, M. A., McGuffog, L., *et al.* (2020). Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genetics in Medicine*, 22(10):1653–1666.

Barrow, E., Robinson, L., Alduaij, W., *et al.* (2009). Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: a report of 121 families with proven mutations. *Clinical Genetics*, 75(2):141–149.

Beiner, M. E., Finch, A., Rosen, B., *et al.* (2007). The risk of endometrial cancer in women with BRCA1 and BRCA2 mutations. A prospective study. *Gynecologic Oncology*, 104(1):7–10.

Bermejo, J. L., Buchner, F. L., and Hemminki, K. (2004). Familial risk of endometrial cancer after exclusion of families that fulfilled Amsterdam, Japanese or Bethesda criteria for HNPCC. *Annals of Oncology*, 15(4):598–604.

Bianco, B., Barbosa, C. P., Trevisan, C. M., *et al.* (2020). Endometrial cancer: a genetic point of view. *Translational Cancer Research*, 9(12):7706–7715.

Bienkiewicz, J., Romanowicz, H., Malinowski, A., and Smolarz, B. (2017). Association of Single Nucleotide Polymorphism -2548 G/A (rs12112075) of leptin gene with endometrial cancer and uterine leiomyomas. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 218(Supplement C):113–118.

Boerner, T. and Long Roche, K. (2020). Salpingectomy for the Risk Reduction of Ovarian Cancer: Is It Time for a Salpingectomy-alone Approach? *Journal of Minimally Invasive Gynecology*.

Bokhman, J. V. (1983). 2 PATHOGENETIC TYPES OF ENDOMETRIAL CARCINOMA. *Gynecologic Oncology*, 15(1):10–17.

Bond, G. L., Hirshfield, K. M., Kirchhoff, T., *et al.* (2006). MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. *Cancer Research*, 66(10):5104–5110.

Bond, G. L., Hu, W., Bond, E. E., *et al.* (2004). A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell*, 119(5):591–602.

Breast Cancer Linkage Consortium (1999). Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst*, 91(15):1310–6.

Breslow, N. E. and Day, N. E. (1987). Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci Publ*, (82):1–406.

Buchanan, D. D., Tan, Y. Y., Walsh, M. D., *et al.* (2014). Tumor Mismatch Repair Immunohistochemistry and DNA MLH1 Methylation Testing of Patients With Endometrial Cancer Diagnosed at Age Younger Than 60 Years Optimizes Triage for Population-Level Germline Mismatch Repair Gene Mutation Testing. *Journal of Clinical Oncology*, 32(2):90–100.

Burke, W. M. and Gold, M. A. (2015). Endometrial Cancer. *Obstetrics and Gynecology*, 125(4):1006–1026.

Bycroft, C., Freeman, C., Petkova, D., *et al.* (2017). Genome-wide genetic data on 500,000 UK Biobank participants. *bioRxiv*, page 166298.

Bycroft, C., Freeman, C., Petkova, D., *et al.* (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209.

Carlson, M. J., Thiel, K. W., and Leslie, K. K. (2014). Past, present, and future of hormonal therapy in recurrent endometrial cancer. *International journal of women's health*, 6:429–435.

Centre for Cancer Genetic Epidemiology (2017). EMBRACE Study.

Chan, J. K., Lakomy, D. S., McDonald, Y., and Kapp, D. S. (2020). Long-term durable responses after pembrolizumab immunotherapy for recurrent, resistant endometrial cancer. *Gynecologic Oncology Reports*, 33:100581.

Chen, H. P., Taylor, N. P., Sotamaa, K. M., *et al.* (2007). Evidence for heritable predisposition to epigenetic silencing of MLH1. *International Journal of Cancer*, 120(8):1684–1688.

Chen, J.-Y., Kuo, S.-J., Liaw, Y.-P., *et al.* (2014a). Endometrial Cancer Incidence in Breast Cancer Patients Correlating with Age and Duration of Tamoxifen Use: a Population Based Study. *Journal of Cancer*, 5(2):151–155.

Chen, M. M., Crous-Bou, M., Setiawan, V. W., *et al.* (2014b). Exome-Wide Association Study of Endometrial Cancer in a Multiethnic Population. *Plos One*, 9(5):9.

Chen, M. M., O'Mara, T. A., Thompson, D. J., *et al.* (2016). GWAS meta-analysis of 16 852 women identifies new susceptibility locus for endometrial cancer. *Human Molecular Genetics*, 25(12):2612–2620.

Chen, X. L., Xiang, Y. B., Long, J. R., *et al.* (2012). Genetic polymorphisms in obesity-related genes and endometrial cancer risk. *Cancer*, 118(13):3356–3364.

Chen, Z., Wang, S. H., Zhou, J., *et al.* (2015). Contribution of PGR genetic polymorphisms to the pathogenesis of endometrial cancer: A meta-analysis. *Journal of Cancer Research and Therapeutics*, 11(4):810–817.

Cheng, T. H. T., Thompson, D., Painter, J., *et al.* (2015). Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Scientific Reports*, 5:12.

Cheng, T. H. T., Thompson, D. J., O'Mara, T. A., *et al.* (2016). Five endometrial cancer risk loci identified through genome-wide association analysis. *Nature Genetics*, 48(6):667–674.

Choi, J., Jia, G., Wen, W., *et al.* (2020). Evaluating polygenic risk scores in assessing risk of nine solid and hematologic cancers in European descendants. *International Journal of Cancer*.

Church, D. M., Schneider, V. A., Graves, T., *et al.* (2011). Modernizing reference genome assemblies. *PLoS Biol*, 9(7):e1001091.

Clarke, M. A., Devesa, S. S., Harvey, S. V., and Wentzensen, N. (2019). Hysterectomy-Corrected Uterine Corpus Cancer Incidence Trends and Differences in Relative Survival Reveal Racial Disparities and Rising Rates of Nonendometrioid Cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, pages JCO1900151–JCO1900151.

Constantine, G. D., Kessler, G., Graham, S., and Goldstein, S. R. (2019). Increased Incidence of Endometrial Cancer Following the Women's Health Initiative: An Assessment of Risk Factors. *Journal of Womens Health*, 28(2):237–243.

Constantinou, P. and Tischkowitz, M. (2017). Genetics of gynaecological cancers. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 42:114–124.

Cooke, P. S., Buchanan, D. L., Young, P., *et al.* (1997). Stromal estrogen receptors mediate mitogenic effects of estradiol on uterine epithelium. *Proceedings of the National Academy of Sciences of the United States of America*, 94(12):6535–6540.

Cooper, N. A. M., Barton, P. M., Breijer, M., *et al.* (2014). Cost-effectiveness of diagnostic strategies for the management of abnormal uterine bleeding (heavy menstrual bleed-

ing and post-menopausal bleeding): a decision analysis. *Health technology assessment (Winchester, England)*, 18(24):1–vi.

Cote, M. L., Ruterbusch, J. J., Olson, S. H., *et al.* (2015). The Growing Burden of Endometrial Cancer: A Major Racial Disparity Affecting Black Women. *Cancer Epidemiology Biomarkers & Prevention*, 24(9):1407–1415.

Crosbie, E. and Morrison, J. (2014). The emerging epidemic of endometrial cancer: time to take action. *Cochrane Database of Systematic Reviews*, (12).

Crosbie, E. J., Zwahlen, M., Kitchener, H. C., *et al.* (2010). Body Mass Index, Hormone Replacement Therapy, and Endometrial Cancer Risk: A Meta-Analysis. *Cancer Epidemiology Biomarkers & Prevention*, 19(12):3119–3130.

CRUK (accessed: Oct 2020). Uterine cancer statistics. `http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer`.

Cuzick, J., Sestak, I., Bonanni, B., *et al.* (2013). Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *The Lancet*, 381(9880):1827–1834.

de Jonge, M. M., Auguste, A., van Wijk, L. M., *et al.* (2019a). Frequent Homologous Recombination Deficiency in High-grade Endometrial Carcinomas. *Clin Cancer Res*, 25(3):1087–1097.

de Jonge, M. M., Mooyaart, A. L., Vreeswijk, M. P., *et al.* (2017). Linking uterine serous carcinoma to BRCA1/2-associated cancer syndrome: A meta-analysis and case report. *Eur J Cancer*, 72:215–225.

de Jonge, M. M., Ritterhouse, L. L., de Kroon, C. D., *et al.* (2019b). Germline BRCA-Associated Endometrial Carcinoma Is a Distinct Clinicopathologic Entity. *Clin Cancer Res*, 25(24):7517–7526.

Delahanty, R. J., Beeghly-Fadiel, A., Xiang, Y. B., *et al.* (2011). Association of Obesity-related Genetic Variants With Endometrial Cancer Risk: A Report From the Shanghai Endometrial Cancer Genetics Study. *American Journal of Epidemiology*, 174(10):1115–1126.

Delaneau, O., Howie, B., Cox, A. J., *et al.* (2013). Haplotype Estimation Using Sequencing Reads. *The American Journal of Human Genetics*, 93(4):687–696.

Derbyshire, A. E., Allen, J. L., Gittins, M., *et al.* (2020). PROgesterone Therapy for Endometrial Cancer prevention in obese women (PROTEC) trial: a feasibility study. *Cancer Prev Res (Phila)*.

Di Cristofano, A. and Ellenson, L. H. (2007). Endometrial Carcinoma. *Annual Review of Pathology: Mechanisms of Disease*, 2(1):57–85.

Di Tucci, C., Capone, C., Galati, G., *et al.* (2019). Immunotherapy in endometrial cancer: new scenarios on the horizon. *Journal of gynecologic oncology*, 30(3):e46–e46.

Dobrzycka, B., Terlikowski, S. J., Mazurek, A., *et al.* (2009). Mutations of the KRAS oncogene in endometrial hyperplasia and carcinoma. *Folia Histochemica Et Cytobiologica*, 47(1):65–68.

Dominguez-Valentin, M., Sampson, J. R., SeppÃ¤lÃ¤, T. T., *et al.* (2020). Cancer risks by gene, age, and gender in 6350 carriers ofpathogenic mismatch repair variants: findings from the Prospective Lynch SyndromeDatabase. *Genetics in Medicine*, 22(1):15–25.

Dullens, B., de Putter, R., Lambertini, M., *et al.* (2020). Cancer Surveillance in Healthy Carriers of Germline Pathogenic Variants inBRCA1/2: A Review of Secondary Prevention Guidelines. *Journal of Oncology*, 2020:13.

Dumesic, D. A. and Lobo, R. A. (2013). Cancer risk and PCOS. *Steroids*, 78(8):782–785.

Early Breast Cancer Trialists' Collaborative, G. (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *The Lancet*, 378(9793):771–784.

Edey, K. A., Rundle, S., and Hickey, M. (2018). Hormone replacement therapy for women previously treated for endometrial cancer. *Cochrane Database Syst Rev*, 5(5):Cd008830.

Edwards, B. K., Noone, A. M., Mariotto, A. B., *et al.* (2014). Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer*, 120(9):1290–1314.

Einarsdottir, K., Darabi, H., Czene, K., *et al.* (2009). Common genetic variability in ESR1 and EGF in relation to endometrial cancer risk and survival. *British Journal of Cancer*, 100(8):1358–1364.

Elliott, J., Bodinier, B., Bond, T. A., *et al.* (2020). Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA*, 323(7):636–645.

Ellison, G., Huang, S., Carr, H., *et al.* (2015). A reliable method for the detection of BRCA1 and BRCA2 mutations in fixed tumour tissue utilising multiplex PCR-based targeted next generation sequencing. *BMC Clin Pathol*, 15:5.

Ellmann, S., Sticht, H., Thiel, F., *et al.* (2009). Estrogen and progesterone receptors: from molecular structures to clinical targets. *Cellular and Molecular Life Sciences*, 66(15):2405–2426.

Evans, D. G., Brentnall, A., Byers, H., *et al.* (2017). The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. *Journal of Medical Genetics*, 54(2):104–110.

Evans, D. G. R., Harkness, E. F., Brentnall, A. R., *et al.* (2019). Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. *Breast Cancer Research and Treatment*, 176(1):141–148.

Fachal, L., Aschard, H., Beesley, J., *et al.* (2020). Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*, 52(1):56–73.

Fadare, O., Liang, S. X., Ulukus, E. C., *et al.* (2006). Precursors of endometrial clear cell carcinoma. *American Journal of Surgical Pathology*, 30(12):1519–1530.

Fader, A. N., Roque, D. M., Siegel, E., *et al.* (2018). Randomized Phase II Trial of Carboplatin-Paclitaxel Versus Carboplatin-Paclitaxel-Trastuzumab in Uterine Serous Carcinomas That Overexpress Human Epidermal Growth Factor Receptor 2/neu. *J Clin Oncol*, 36(20):2044–2051.

Fortner, R. T., Husing, A., Kuhn, T., *et al.* (2017). Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. *Int J Cancer*, 140(6):1317–1323.

Francies, F. Z., Marima, R., Hull, R., *et al.* (2020). Genomics and splicing events of type II endometrial cancers in the black population: racial disparity, socioeconomic and geographical differences. *American journal of cancer research*, 10(10):3061–3082.

Friberg, E., Orsini, N., Mantzoros, C. S., and Wolk, A. (2007). Diabetes mellitus and risk of endometrial cancer: a meta-analysis. *Diabetologia*, 50(7):1365–74.

Fritsche, L. G., Patil, S., Beesley, L. J., *et al.* (2020). Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. *The American Journal of Human Genetics*, 107(5):815–836.

Fry, A., Littlejohns, T. J., Sudlow, C., *et al.* (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9):1026–1034.

Gansmo, L. B., Bjornslett, M., Halle, M. K., *et al.* (2017). MDM2 promoter polymorphism del1518 (rs3730485) and its impact on endometrial and ovarian cancer risk. *Bmc Cancer*, 17:6.

Geng, Y. H., Wang, Z. F., Jia, Y. M., *et al.* (2018). Genetic polymorphisms in CDH1 are associated with endometrial carcinoma susceptibility among Chinese Han women. *Oncology Letters*, 16(5):6868–6878.

Gentry-Maharaj, A. and Karpinskyj, C. (2020). Current and future approaches to screening for endometrial cancer. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 65:79–97.

Germano, G., Amirouchene-Angelozzi, N., Rospo, G., and Bardelli, A. (2018). The Clinical Impact of the Genomic Landscape of Mismatch Repair-Deficient Cancers. *Cancer Discovery*, 8(12):1518.

Getz, G., Gabriel, S. B., Cibulskis, K., *et al.* (2013). Integrated genomic characterization of endometrial carcinoma. *Nature*, 497(7447):67–73.

Glubb, D. M., Thompson, D. J., Aben, K. K. H., *et al.* (2020). Cross-cancer genome-wide association study of endometrial cancer and epithelial ovarian cancer identifies genetic risk regions associated with risk of both cancers. *Cancer Epidemiology, Biomarkers & Prevention*.

Gompel, A. (2020). Progesterone and endometrial cancer. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 69:95–107.

Gong, T. T., Wu, Q. J., Wang, Y. L., and Ma, X. X. (2015). Circulating adiponectin, leptin and adiponectin-leptin ratio and endometrial cancer risk: Evidence from a meta-analysis of epidemiologic studies. *International Journal of Cancer*, 137(8):1967–1978.

Grady, D., Gebretsadik, T., Kerlikowske, K., *et al.* (1995). HORMONE REPLACEMENT THERAPY AND ENDOMETRIAL CANCER RISK - A METAANALYSIS. *Obstetrics and Gynecology*, 85(2):304–313.

Gu, Y., Zhou, X., and Zhang, S. L. (2011). Meta-analysis of an association of codon 72 polymorphisms of the p53 gene with increased endometrial cancer risk. *Genet Mol Res*, 10(4):3609–19.

Gupta, K. K., Attri, J. P., Singh, A., *et al.* (2016). Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi journal of anaesthesia*, 10(3):328–331.

Hamdi, Y., Soucy, P., Adoue, V., *et al.* (2016). Association of breast cancer risk with genetic variants showing differential allelic expression: Identification of a novel breast cancer susceptibility locus at 4q21. *Oncotarget*, 7(49):80140–80163.

Hamilton, C. A., Cheung, M. K., Osann, K., *et al.* (2006). Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *Br J Cancer*, 94(5):642–6.

Heemskerk-Gerritsen, B. A. M., Menke-Pluijmers, M. B. E., Jager, A., *et al.* (2013). Substantial breast cancer risk reduction and potential survival benefit after bilateral mastectomy when compared with surveillance in healthy BRCA1 and BRCA2 mutation carriers: a prospective analysis. *Annals of Oncology*, 24(8):2029–2035.

Hernandez, A. V., Pasupuleti, V., Benites-Zapata, V. A., *et al.* (2015). Insulin resistance and endometrial cancer risk: A systematic review and meta-analysis. *European Journal of Cancer*, 51(18):2747–2758.

Hong, E. P. and Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics Inform*, 10(2):117–122.

Hornreich, G., Beller, U., Lavie, O., *et al.* (1999). Is uterine serous papillary carcinoma a BRCA1-related disease? Case report and review of the literature. *Gynecol Oncol*, 75(2):300–4.

Howie, B., Marchini, J., and Stephens, M. (2011). Genotype Imputation with Thousands of Genomes. *G3: Genes | Genomes | Genetics*, 1(6):457.

Hu, Y. C., Zhang, Q., Huang, Y. H., *et al.* (2014). Comparison of two methods to extract DNA from formalin-fixed, paraffin-embedded tissues and their impact on EGFR mutation detection in non-small cell lung carcinoma. *Asian Pac J Cancer Prev*, 15(6):2733–7.

Human Protein Atlas (Accessed: November 2020).

Husing, A., Dossus, L., Ferrari, P., *et al.* (2016). An epidemiological model for prediction of endometrial cancer risk in Europe. *European Journal of Epidemiology*, 31(1):51–60.

Huynh-Le, M.-P., Fan, C. C., Karunamuni, R., *et al.* (2020). Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. *medRxiv*, page 19012237.

Iwamoto, I., Fujino, T., Douchi, T., and Nagata, Y. (2003). Association of estrogen receptor alpha and beta(3)-adrenergic receptor polymorphisms with endometrial cancer. *Obstetrics and Gynecology*, 102(3):506–511.

Jamison, P. M., Noone, A. M., Ries, L. A. G., *et al.* (2013). Trends in Endometrial Cancer Incidence by Race and Histology with a Correction for the Prevalence of Hysterectomy, SEER 1992 to 2008. *Cancer Epidemiology Biomarkers & Prevention*, 22(2):233–241.

Janda, M., Gebski, V., Brand, A., *et al.* (2010). Quality of life after total laparoscopic hysterectomy versus total abdominal hysterectomy for stage I endometrial cancer (LACE): a randomised trial. *The Lancet Oncology*, 11(8):772–780.

Janda, M., Gebski, V., Davies, L. C., *et al.* (2017). Effect of Total Laparoscopic Hysterectomy vs Total Abdominal Hysterectomy on Disease-Free Survival Among Women With Stage I Endometrial Cancer: A Randomized Clinical Trial. *JAMA*, 317(12):1224–1233.

Jerath, N. U., Mankodi, A., Crawford, T. O., *et al.* (2018). Charcot-Marie-Tooth Disease type 4C: Novel mutations, clinical presentations, and diagnostic challenges. *Muscle & Nerve*, 57(5):749–755.

Johnatty, S. E., Tan, Y. Y., Buchanan, D. D., *et al.* (2017). Family history of cancer predicts endometrial cancer risk independently of Lynch Syndrome: Implications for genetic counselling. *Gynecol Oncol*, 147(2):381–387.

Kaaks, R., Lukanova, A., and Kurzer, M. S. (2002). Obesity, endogenous hormones, and

endometrial cancer risk: A synthetic review. *Cancer Epidemiology Biomarkers & Prevention*, 11(12):1531–1543.

Kafshdooz, T., Tabrizi, A. D., Ardabili, S. M. M., *et al.* (2014). Polymorphism of p53 Gene Codon 72 in Endometrial Cancer: Correlation with Tumor Grade and Histological Type. *Asian Pacific Journal of Cancer Prevention*, 15(22):9603–9606.

Kaler, A. S. and Purcell, L. C. (2019). Estimation of a significance threshold for genome-wide association studies. *BMC Genomics*, 20(1):618.

Kaplan, E. J., Caputo, T. A., Shen, P. U., *et al.* (1998). Familial papillary serous carcinoma of the cervix, peritoneum, and ovary: a report of the first case. *Gynecol Oncol*, 70(2):289–94.

Karunamuni, R. A., Huynh-Le, M.-P., Fan, C. C., *et al.* (2021). Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. *Prostate Cancer and Prostatic Diseases*.

Kauff, N. D., Domchek, S. M., Friebel, T. M., *et al.* (2008). Risk-reducing salpingo-oophorectomy for the prevention of BRCA1- and BRCA2-associated breast and gynecologic cancer: a multicenter, prospective study. *J Clin Oncol*, 26(8):1331–7.

Khera, A. V., Chaffin, M., Aragam, K. G., *et al.* (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9):1219–1224.

Kitson, S. J., Bafligil, C., Ryan, N. A. J., *et al.* (2020). BRCA1 and BRCA2 pathogenic variant carriers and endometrial cancer risk: A cohort study. *European journal of cancer (Oxford, England : 1990)*, 136:169–175.

Kitson, S. J. and Crosbie, E. J. (2019). Endometrial cancer and obesity. *The Obstetrician & Gynaecologist*, 21(4):237–245.

Kitson, S. J., Evans, D. G., and Crosbie, E. J. (2017). Identifying High-Risk Women for

Endometrial Cancer Prevention Strategies: Proposal of an Endometrial Cancer Risk Prediction Model. *Cancer Prevention Research*, 10(1):1–13.

Kitson, S. J., Maskell, Z., Sivalingam, V. N., *et al.* (2019). PRE-surgical Metformin In Uterine Malignancy (PREMIUM): a Multi-Center, Randomized Double-Blind, Placebo-Controlled Phase III Trial. *Clin Cancer Res*, 25(8):2424–2432.

Knappskog, S., Gansmo, L. B., Dibirova, K., *et al.* (2014). Population distribution and ancestry of the cancer protective MDM2 SNP285 (rs117039649). *Oncotarget*, 5(18):8223–8234.

Knappskog, S., Trovik, J., Marcickiewicz, J., *et al.* (2012a). *SNP285C modulates oestrogen receptor/Sp1 binding to the MDM2 promoter and reduces the risk of endometrial but not prostatic cancer*, volume 48, pages 1988–96. 2011 Elsevier Ltd, England.

Knappskog, S., Trovik, J., Marcickiewicz, J., *et al.* (2012b). *SNP285C modulates oestrogen receptor/Sp1 binding to the MDM2 promoter and reduces the risk of endometrial but not prostatic cancer*, volume 48, pages 1988–96. 2011 Elsevier Ltd, England.

Kommoss, S., McConechy, M. K., Kommoss, F., *et al.* (2018). Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series. *Ann Oncol*, 29(5):1180–1188.

Kotsopoulos, J., Gronwald, J., Karlan, B. Y., *et al.* (2018). Hormone Replacement Therapy After Oophorectomy and Breast Cancer Risk Among BRCA1 Mutation Carriers. *JAMA Oncol*, 4(8):1059–1065.

Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., *et al.* (2017a). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama*, 317(23):2402–2416.

Kuchenbaecker, K. B., McGuffog, L., Barrowdale, D., *et al.* (2017b). Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *JNCI: Journal of the National Cancer Institute*, 109(7).

Kuo, K. H. M. (2017). Multiple Testing in the Context of Gene Discovery in Sickle Cell Disease Using Genome-Wide Association Studies. *Genomics insights*, 10:1178631017721178–1178631017721178.

Kwon, Y.-J., Hong, K.-W., Park, B. J., and Jung, D.-H. (2019). Serotonin receptor 3B polymorphisms are associated with type 2 diabetes: The Korean Genome and Epidemiology Study. *Diabetes Research and Clinical Practice*, 153:76–85.

Kyrgiou, M., Kalliala, I., Markozannes, G., *et al.* (2017). Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ*, 356:j477.

Lacey, J. V., Yang, H. N., Gaudet, M. M., *et al.* (2011). Endometrial cancer and genetic variation in PTEN, PIK3CA, AKT1, MLH1, and MSH2 within a population-based case-control study. *Gynecologic Oncology*, 120(2):167–173.

Lakeman, I. M. M., Rodriguez-Girondo, M., Lee, A., *et al.* (2020). Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. *Genetics in Medicine*, page 9.

Landrum, M. J., Lee, J. M., Benson, M., *et al.* (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, 46(D1):D1062–d1067.

Lavie, O., Ben-Arie, A., Segev, Y., *et al.* (2010). BRCA germline mutations in women with uterine serous carcinoma–still a debate. *Int J Gynecol Cancer*, 20(9):1531–4.

Lavie, O., Hornreich, G., Ben Arie, A., *et al.* (2000). BRCA1 germline mutations in women with uterine serous papillary carcinoma. *Obstet Gynecol*, 96(1):28–32.

Ledermann, J. A., Drew, Y., and Kristeleit, R. S. (2016). Homologous recombination deficiency and ovarian cancer. *Eur J Cancer*, 60:49–58.

Lee, E., Hsu, C., Haiman, C. A., *et al.* (2010). Genetic variation in the progesterone receptor gene and risk of endometrial cancer: a haplotype-based approach. *Carcinogenesis*, 31(8):1392–1399.

Lee, Y. C., Milne, R. L., Lheureux, S., *et al.* (2017). Risk of uterine cancer for BRCA1 and BRCA2 mutation carriers. *Eur J Cancer*, 84:114–120.

Leon-Castillo, A., de Boer, S. M., Powell, M. E., *et al.* (2020). Molecular Classification of the PORTEC-3 Trial for High-Risk Endometrial Cancer: Impact on Prognosis and Benefit From Adjuvant Therapy. *Journal of Clinical Oncology*, 38(29):3388–3397.

Levine, D. A., Lin, O., Barakat, R. R., *et al.* (2001). Risk of endometrial carcinoma associated with BRCA mutation. *Gynecol Oncol*, 80(3):395–8.

Lewis, C. M. and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, 12(1):44.

Li, G., Xiang, Y. B., Courtney, R., *et al.* (2011a). Association of a single nucleotide polymorphism at 6q25.1,rs2046210, with endometrial cancer risk among Chinese women. *Chinese journal of cancer*, 30(2):138–43.

Li, Y., Zhao, H. J., Sun, L., *et al.* (2011b). MDM2 SNP309 is associated with endometrial cancer susceptibility: a meta-analysis. *Human Cell*, 24(2):57–64.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., *et al.* (2000). Environmental and heritable factors in the causation of cancer - Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343(2):78–85.

Liu, L. and Kiryluk, K. (2018). Genome-wide polygenic risk predictors for kidney disease. *Nature Reviews Nephrology*, 14(12):723–724.

Long, B., Liu, F. W., and Bristow, R. E. (2013). Disparities in uterine cancer epidemiology, treatment, and survival among African Americans in the United States. *Gynecologic Oncology*, 130(3):652–659.

Lu, K. H. and Broaddus, R. R. (2020). Endometrial Cancer. *New England Journal of Medicine*, 383(21):2053–2064.

Lu, Y., Ek, W. E., Whiteman, D., *et al.* (2014). Most common 'sporadic' cancers have a significant germline genetic component. *Human Molecular Genetics*, 23(22):6112–6118.

Lundin, E., Wirgin, I., Lukanova, A., *et al.* (2012). Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiology*, 36(5):445–452.

Lurie, G., Gaudet, M. M., Spurdle, A. B., *et al.* (2011). The Obesity-Associated Polymorphisms FTO rs9939609 and MC4R rs17782313 and Endometrial Cancer Risk in Non-Hispanic White Women. *Plos One*, 6(2):8.

Ma, Y., Liu, Z. W., Zhang, Y., and Lu, B. J. (2013). Serum leptin, adiponectin and endometrial cancer risk in Chinese women. *Journal of Gynecologic Oncology*, 24(4):336–341.

MacKintosh, M. L. and Crosbie, E. J. (2013). Obesity-driven endometrial cancer: is weight loss the answer? *Bjog-an International Journal of Obstetrics and Gynaecology*, 120(7):791–794.

MacKintosh, M. L. and Crosbie, E. J. (2018). Prevention Strategies in Endometrial Carcinoma. *Current Oncology Reports*, 20(12):101.

Mahdi, H., Mester, J. L., Nizialek, E. A., *et al.* (2015). Germline PTEN, SDHB-D, and KLLN alterations in endometrial cancer patients with Cowden and Cowden-like syndromes: An international, multicenter, prospective study. *Cancer*, 121(5):688–696.

Makker, V., Rasco, D. W., Dutcus, C. E., *et al.* (2017). A phase Ib/II trial of lenvatinib (LEN) plus pembrolizumab (Pembro) in patients (Pts) with endometrial carcinoma. *Journal of Clinical Oncology*, 35(15-suppl):5598–5598.

Marabelle, A., Le, D. T., Ascierto, P. A., *et al.* (2019). Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *Journal of Clinical Oncology*, 38(1):1–10.

Marchetti, C., De Felice, F., Boccia, S., *et al.* (2018). Hormone replacement therapy after prophylactic risk-reducing salpingo-oophorectomy and breast cancer risk in BRCA1 and BRCA2 mutation carriers: A meta-analysis. *Crit Rev Oncol Hematol*, 132:111–115.

Marchini, J., Howie, B., Myers, S., *et al.* (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39(7):906–13.

Marquez-Luna, C., Gazal, S., Loh, P.-R., *et al.* (2020). LDpred-funct: incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *bioRxiv*.

Mary, B. D., Robert, P., Matthew, B. Y., *et al.* (2020). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic, Version 1.2020. *Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw*, 18(4):380–391.

Mavaddat, N., Michailidou, K., Dennis, J., *et al.* (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, 104(1):21–34.

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., *et al.* (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute*, 107(5):djv036.

Michailidou, K., LindstrÃ¶m, S., Dennis, J., *et al.* (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678):92–94.

Mihm, M., Gangooly, S., and Muttukrishna, S. (2011). The normal menstrual cycle in women. *Animal Reproduction Science*, 124(3-4):229–236.

Millis, S. Z., Ikeda, S., Reddy, S., *et al.* (2016). Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19 784 Diverse Solid Tumors. *JAMA Oncol*, 2(12):1565–1573.

Moher, D., Liberati, A., Tetzlaff, J., *et al.* (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7):1–6.

Moore, L., Leongamornlert, D., Coorens, T. H. H., *et al.* (2020). The mutational landscape of normal human endometrial epithelium. *Nature*, 580(7805):640–646.

Morice, P., Leary, A., Creutzberg, C., *et al.* (2016). Endometrial cancer. *Lancet*, 387(10023):1094–1108.

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., *et al.* (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, 315(1):68–76.

Nair, N., Camacho-Vanegas, O., Rykunov, D., *et al.* (2016). Genomic Analysis of Uterine Lavage Fluid Detects Early Endometrial Cancers and Reveals a Prevalent Landscape of Driver Mutations in Women without Histopathologic Evidence of Cancer: A Prospective Cross-Sectional Study. *PLoS Med*, 13(12):e1002206.

National Cancer Intelligence Network (Accessed: December 2020). Outline of Uterine Cancer in the United Kingdom: Incidence, Mortality and Survival.

Nead, K. T., Sharp, S. J., Thompson, D. J., *et al.* (2015). Evidence of a Causal Association Between Insulinemia and Endometrial Cancer: A Mendelian Randomization Analysis. *JNCI: Journal of the National Cancer Institute*, 107(9).

Nelen, M. R., Kremer, H., Konings, I. B., *et al.* (1999). Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. *Eur J Hum Genet*, 7(3):267–73.

NHS Digital (Accessed: December 2020a). Health Survey for England 2018: Adult and child overweight and obesity.

NHS Digital (Accessed: December 2020b). Statistics on Obesity, Physical Activity and Diet, England, 2020: Adult overweight and obesity.

NICE (2019). Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer (NICE Clinical Guidelines, No. 164).

NICE (2020). Testing strategies for Lynch syndrome in people with endometrial cancer: Diagnostics guidance [DG42].

Nunobiki, O., Ueda, M., Yamamoto, M., *et al.* (2009). Polymorphisms of p53 codon 72 and MDM2 promoter 309 and the risk of endometrial cancer. *Human Cell*, 22(4):101–106.

Office for National Statistics (2019). Cancer registration statistics, England: 2017.

Ohno, S., Kyo, S., Myojo, S., *et al.* (2009). Wilms' tumor 1 (WT1) peptide immunotherapy for gynecological malignancy. *Anticancer Res*, 29(11):4779–84.

Okamoto, K., Tsunematsu, R., Tahira, T., *et al.* (2015). SNP55, a new functional polymorphism of MDM2-P2 promoter, contributes to allele-specific expression of MDM2 in endometrial cancers. *Bmc Medical Genetics*, 16:15.

Olson, S. H., Chen, C., De Vivo, I., *et al.* (2009). Maximizing resources to study an uncommon cancer: E2C2-Epidemiology of Endometrial Cancer Consortium. *Cancer Causes & Control*, 20(4):491–496.

O'Mara, T. A. and Crosbie, E. J. (2020). Polygenic risk score opportunities for early detection and prevention strategies in endometrial cancer. *British Journal of Cancer*.

O'Mara, T. A., Fahey, P., Ferguson, K., *et al.* (2011). Progesterone receptor gene variants and risk of endometrial cancer. *Carcinogenesis*, 32(3):331–335.

O'Mara, T. A., Glubb, D. M., Amant, F., *et al.* (2018). Identification of nine new susceptibility loci for endometrial cancer. *Nature Communications*, 9.

O'Mara, T. A., Glubb, D. M., Kho, P. F., *et al.* (2019a). Genome-wide association studies of endometrial cancer: Latest developments and future directions. *Cancer epidemiol-*

ogy, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.

O'Mara, T. A., Glubb, D. M., Painter, J. N., *et al.* (2015). Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocrine-Related Cancer*, 22(5):851–861.

O'Mara, T. A., Spurdle, A. B., and Glubb, D. M. (2019b). Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. *Cancers (Basel)*, 11(10).

Ott, P. A., Bang, Y. J., Berton-Rigaud, D., *et al.* (2017). Safety and Antitumor Activity of Pembrolizumab in Advanced Programmed Death Ligand 1-Positive Endometrial Cancer: Results From the KEYNOTE-028 Study. *J Clin Oncol*, 35(22):2535–2541.

Painter, J. N., Kaufmann, S., O'Mara, T. A., *et al.* (2016). A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates AKT1 through YY1 Binding. *American Journal of Human Genetics*, 98(6):1159–1169.

Painter, J. N., O'Mara, T. A., Batra, J., *et al.* (2015). Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Human Molecular Genetics*, 24(5):1478–1492.

Panagiotou, O. A., Ioannidis, J. P. A., and for the Genome-Wide Significance, P. (2012). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *International Journal of Epidemiology*, 41(1):273–286.

Paraskevaidi, M., Morais, C. L. M., Ashton, K. M., *et al.* (2020). Detecting Endometrial Cancer by Blood Spectroscopy: A Diagnostic Cross-Sectional Study. *Cancers*, 12(5).

Patnala, R., Clements, J., and Batra, J. (2013). Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC genetics*, 14:39–39.

Pavlou, M., Ambler, G., Seaman, S. R., *et al.* (2015). How to develop a more accurate risk prediction model when there are few events. *BMJ*, 351.

Pearce, C. L., Stram, D. O., Ness, R. B., *et al.* (2015). Population distribution of lifetime risk of ovarian cancer in the United States. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 24(4):671–676.

Peng, Q. L., Mo, C. J., Qin, A. P., *et al.* (2013). MDM2 SNP309 polymorphism contributes to endometrial cancer susceptibility: evidence from a meta-analysis. *Journal of Experimental & Clinical Cancer Research*, 32:9.

Pennington, M., Gentry-Maharaj, A., Karpinskyj, C., *et al.* (2016). Long-Term Secondary Care Costs of Endometrial Cancer: A Prospective Cohort Study Nested within the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *PloS one*, 11(11):e0165539–e0165539.

Pfeiffer, R. M., Park, Y., Kreimer, A. R., *et al.* (2013). Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies. *Plos Medicine*, 10(7):15.

Piazza, R., Magistroni, V., Redaelli, S., *et al.* (2018). SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nature Communications*, 9(1):2192.

Pilarski, R. (2009). Cowden Syndrome: A Critical Review of the Clinical Literature. *Journal of Genetic Counseling*, 18(1):13–27.

Poplawski, T., Sobczuk, A., Sarnik, J., *et al.* (2015). Polymorphism of DNA mismatch repair genes in endometrial cancer. *Exp Oncol*, 37(1):44–7.

Prive, F., Arbel, J., and Vilhjalmsson, B. J. (2020). LDpred2: better, faster, stronger. *bioRxiv*.

PROSPERO (accessed: Oct 2020). Low-penetrance common predisposition variants in endometrial cancer risk: a systematic review. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018091907.

Raglan, O., Kalliala, I., Markozannes, G., *et al.* (2019). Risk factors for endometrial cancer: An umbrella review of the literature. *Int J Cancer*, 145(7):1719–1730.

Randall, M. E., Filiaci, V., McMeekin, D. S., *et al.* (2019). Phase III Trial: Adjuvant Pelvic Radiation Therapy Versus Vaginal Brachytherapy Plus Paclitaxel/Carboplatin in High-Intermediate and High-Risk Early-Stage Endometrial Cancer. *Journal of Clinical Oncology*, 37(21):1810–1818.

Reitsma, W., Mourits, M. J., de Bock, G. H., and Hollema, H. (2013). Endometrium is not the primary site of origin of pelvic high-grade serous carcinoma in BRCA1 or BRCA2 mutation carriers. *Mod Pathol*, 26(4):572–8.

Renehan, A. G., Tyson, M., Egger, M., *et al.* (2008). Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*, 371(9612):569–78.

Ring, K. L., Yates, M. S., Schmandt, R., *et al.* (2017). Endometrial Cancers With Activating KRas Mutations Have Activated Estrogen Signaling and Paradoxical Response to MEK Inhibition. *International Journal of Gynecological Cancer*, 27(5):854–862.

Rivandi, M., Martens, J. W. M., and Hollestelle, A. (2018). Elucidating the Underlying Functional Mechanisms of Breast Cancer Susceptibility Through Post-GWAS Analyses. *Frontiers in genetics*, 9:280–280.

Rodon, J., Soria, J.-C., Berger, R., *et al.* (2018). WINTHER: An international WIN Consortium precision medicine trial using genomic and transcriptomic analysis in patients with advanced malignancies. *Journal of Clinical Oncology*, 36(15-suppl):12011–12011.

Rosen, M. W., Tasset, J., Kobernik, E. K., *et al.* (2019). Risk Factors for Endometrial Cancer or Hyperplasia in Adolescents and Women 25 Years Old or Younger. *Journal of Pediatric and Adolescent Gynecology*, 32(5):546–549.

Rosenblum, R. E., Ang, C., Suckiel, S. A., *et al.* (2020). Lynch Syndrome-Associated Variants and Cancer Rates in an Ancestrally Diverse Biobank. *JCO Precision Oncology*, (4):1429–1444.

Rossi, L., Le Frere-Belda, M. A., Laurent-Puig, P., *et al.* (2017). Clinicopathologic Characteristics of Endometrial Cancer in Lynch Syndrome A French Multicenter Study. *International Journal of Gynecological Cancer*, 27(5):953–960.

Rothman, K. J. (2008). BMI-related errors in the measurement of obesity. *International Journal of Obesity*, 32(3):S56–S59.

Rugo, H. S., Rumble, R. B., Macrae, E., *et al.* (2016). Endocrine Therapy for Hormone Receptor-Positive Metastatic Breast Cancer: American Society of Clinical Oncology Guideline. *Journal of Clinical Oncology*, 34(25):3069–3103.

Russell, H., Kedzierska, K., Buchanan, D. D., *et al.* (2020). The MLH1 polymorphism rs1800734 and risk of endometrial cancer with microsatellite instability. *Clinical Epigenetics*, 12(1).

Ryan, N. A. J., Glaire, M. A., Blake, D., *et al.* (2019). The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genetics in Medicine*, 21(10):2167–2180.

Ryan, N. A. J., McMahon, R., Tobi, S., *et al.* (2020). The proportion of endometrial tumours associated with Lynch syndrome (PETALS): A prospective cross-sectional study. *PLOS Medicine*, 17(9):e1003263.

Ryan, N. A. J., Morris, J., Green, K., *et al.* (2017). Association of Mismatch Repair Mutation With Age at Cancer Onset in Lynch Syndrome Implications for Stratified Surveillance Strategies. *Jama Oncology*, 3(12):1702–1706.

Sagae, S., Susumu, N., Viswanathan, A. N., *et al.* (2014). Gynecologic Cancer InterGroup (GCIG) consensus review for uterine serous carcinoma. *Int J Gynecol Cancer*, 24(9 Suppl 3):S83–9.

Sasaki, M., Tanaka, Y., Kaneuchi, M., *et al.* (2002). Polymorphisms of estrogen receptor alpha gene in endometrial cancer. *Biochemical and Biophysical Research Communications*, 297(3):558–564.

Saule, C., Mouret-Fourme, E., Briaux, A., *et al.* (2018). Risk of Serous Endometrial Carcinoma in Women With Pathogenic BRCA1/2 Variant After Risk-Reducing Salpingo-Oophorectomy. *J Natl Cancer Inst*, 110(2).

Schramm, A., Ebner, F., Bauer, E., *et al.* (2017). Value of endometrial thickness assessed by transvaginal ultrasound for the prediction of endometrial cancer in patients with postmenopausal bleeding. *Arch Gynecol Obstet*, 296(2):319–326.

Schultheis, A. M., Martelotto, L. G., De Filippo, M. R., *et al.* (2016). TP53 Mutational Spectrum in Endometrioid and Serous Endometrial Cancers. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists*, 35(4):289–300.

Segev, Y., Iqbal, J., Lubinski, J., *et al.* (2013). The incidence of endometrial cancer in women with BRCA1 and BRCA2 mutations: an international prospective cohort study. *Gynecol Oncol*, 130(1):127–31.

Segev, Y., Rosen, B., Lubinski, J., *et al.* (2015). Risk factors for endometrial cancer among women with a BRCA1 or BRCA2 mutation: a case control study. *Fam Cancer*, 14(3):383–91.

Sestak, I. and Cuzick, J. (2016). Endometrial cancer risk in postmenopausal breast cancer patients treated with tamoxifen or aromatase inhibitors. *Expert Rev Endocrinol Metab*, 11(5):425–432.

Setiawan, V. W. (2016). *Endometrial Cancer Among Asian Americans*, pages 219–231. Springer International Publishing, Cham.

Setiawan, V. W., Doherty, J. A., Shu, X. O., *et al.* (2009). Two Estrogen-Related Variants in CYP19A1 and Endometrial Cancer Risk: A Pooled Analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiology Biomarkers & Prevention*, 18(1):242–247.

Setiawan, V. W., Haessler, J., Schumacher, F., *et al.* (2012). HNF1B and Endometrial Cancer Risk: Results from the PAGE study. *Plos One*, 7(1):6.

Setiawan, V. W., Schumacher, F., Prescott, J., *et al.* (2014). Cross-cancer pleiotropic analysis of endometrial cancer: PAGE and E2C2 consortia. *Carcinogenesis*, 35(9):2068–2073.

Setiawan, V. W., Yang, H. P., Pike, M. C., *et al.* (2013). Type I and II Endometrial Cancers: Have They Different Risk Factors? *Journal of Clinical Oncology*, 31(20):2607–2618.

Shan, J. X., Chouchane, A., Mokrab, Y., *et al.* (2019). Genetic Variation in CCL5 Signaling Genes and Triple Negative Breast Cancer: Susceptibility and Prognosis Implications. *Frontiers in Oncology*, 9:13.

Sherry, S., Ward, M., Kholodov, M., *et al.* (2001). dbSNP: the NCBI database of genetic variation. `https://www.ncbi.nlm.nih.gov/snp/`.

Shu, C. A., Pike, M. C., Jotwani, A. R., *et al.* (2016). Uterine Cancer After Risk-Reducing Salpingo-oophorectomy Without Hysterectomy in Women With BRCA Mutations. *JAMA Oncol*, 2(11):1434–1440.

Sideris, M., Emin, E. I., Abdullah, Z., *et al.* (2019). The Role of KRAS in Endometrial Cancer: A Mini-Review. *Anticancer Research*, 39(2):533–539.

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J Clin*, 69(1):7–34.

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1):7–30.

Sliwinski, T., Sitarek, P., Stetkiewicz, T., *et al.* (2010). Polymorphism of the ER alpha and CYP1B1 genes in endometrial cancer in a Polish subpopulation. *Journal of Obstetrics and Gynaecology Research*, 36(2):311–317.

Spencer, C. C. A., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *Plos Genetics*, 5(5):13.

Spurdle, A. B., Bowman, M. A., Shamsani, J., and Kirk, J. (2017). Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol*, 30(8):1048–1068.

Spurdle, A. B., Thompson, D. J., Ahmed, S., *et al.* (2011). Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nature Genetics*, 43(5):451–454.

Stelloo, E., Bosse, T., Nout, R. A., *et al.* (2015). Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Modern Pathology*, 28(6):836–844.

Stelloo, E., Nout, R. A., Osse, E. M., *et al.* (2016). Improved Risk Assessment by Integrating Molecular and Clinicopathological Factors in Early-stage Endometrial Cancer-Combined Analysis of the PORTEC Cohorts. *Clinical Cancer Research*, 22(16):4215.

Stockley, T. L., Oza, A. M., Berman, H. K., *et al.* (2016). Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med*, 8(1):109.

Sundar, S., Balega, J., Crosbie, E., *et al.* (2017). BGCS uterine cancer guidelines: Recommendations for practice. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 213:71–97.

Tafe, L. J. (2015). Targeted Next-Generation Sequencing for Hereditary Cancer Syndromes A Focus on Lynch Syndrome and Associated Endometrial Cancer. *Journal of Molecular Diagnostics*, 17(5):472–482.

Talhouk, A. and McAlpine, J. N. (2016). New classification of endometrial cancers: the development and potential applications of genomic-based classification in research and clinical care. *Gynecologic Oncology Research and Practice*, 3(1):14.

Talhouk, A., McConechy, M. K., Leung, S., *et al.* (2015). A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer*, 113(2):299–310.

Talhouk, A., McConechy, M. K., Leung, S., *et al.* (2017). Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer*, 123(5):802–813.

Tao, M. H., Cai, Q. Y., Zhang, Z. F., *et al.* (2007). Polymorphisms in the CYP19A1 (Aromatase) gene and endometrial cancer risk in Chinese women. *Cancer Epidemiology Biomarkers & Prevention*, 16(5):943–949.

Terry, K., McGrath, M., Lee, I. M., *et al.* (2008). MDM2 SNP309 is associated with endometrial cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 17(4):983–986.

Thompson, D. and Easton, D. F. (2002). Cancer Incidence in BRCA1 mutation carriers. *J Natl Cancer Inst*, 94(18):1358–65.

Tran, A. Q. and Gehrig, P. (2017). Recent Advances in Endometrial Cancer. *F1000Res*, 6:81.

TransPORTEC Consortium (2020). RAINBO: Refining Adjuvant treatment IN endometrial cancer Based On molecular profile.

Tzur, T., Kessous, R., and Weintraub, A. Y. (2017). Current strategies in the diagnosis of endometrial cancer. *Archives of Gynecology and Obstetrics*, 296(1):5–14.

Ueda, M., Yamamoto, M., Nunobiki, O., *et al.* (2009). Murine double-minute 2 homolog single nucleotide polymorphism 309 and the risk of gynecologic cancer. *Human Cell*, 22(2):49–54.

Uhlen, M., Fagerberg, L., Hallstrom, B. M., *et al.* (2015). Tissue-based map of the human proteome. *Science*, 347(6220):1260419.

Uniprot (Accessed: November 2020).

Urick, M. E. and Bell, D. W. (2019). Clinical actionability of molecular targets in endometrial cancer. *Nature Reviews Cancer*, 19(9):510–521.

van Hanegem, N., Prins, M. M., Bongers, M. Y., *et al.* (2016). The accuracy of endometrial sampling in women with postmenopausal bleeding: a systematic review and meta-analysis. *Eur J Obstet Gynecol Reprod Biol*, 197:147–55.

van Veen, E. M., Brentnall, A. R., Byers, H., *et al.* (2018). Use of Single-Nucleotide Polymorphisms and Mammographic Density Plus Classic Risk Factors for Breast Cancer Risk Prediction. *Jama Oncology*, 4(4):476–482.

VonVille, H. (2015). Excel workbooks for systematic reviews. `https://showcase.dropbox.com/s/Excel-Workbooks-for-Systematic-Reviews-User-Guides-Kf4pYVTvFqSJdZR2Spzlu`.

Walker, J. L., Piedmonte, M. R., Spirtos, N. M., *et al.* (2009). Laparoscopy Compared With Laparotomy for Comprehensive Surgical Staging of Uterine Cancer: Gynecologic Oncology Group Study LAP2. *Journal of Clinical Oncology*, 27(32):5331–5336.

Walsh, C. S., Miller, C. W., Karlan, B. Y., and Koeffler, H. P. (2007). Association between a functional single nucleotide polyrnorphism in the MDM2 gene and sporadic endometrial cancer risk. *Gynecologic Oncology*, 104(3):660–664.

Wan, Y., Wu, W., Yin, Z. H., *et al.* (2011). MDM2 SNP309, gene-gene interaction, and tumor susceptibility: an updated meta-analysis. *Bmc Cancer*, 11:9.

Wan, Y. L., Beverley-Stevenson, R., Carlisle, D., *et al.* (2016). Working together to shape the endometrial cancer research agenda: The top ten unanswered research questions. *Gynecologic Oncology*, 143(2):287–293.

Wang, L. H., Wang, X., Xu, W. T., and Hu, Y. L. (2014). MDM2 rs2279744 polymorphism and endometrial cancer: a meta-analysis. *Tumor Biology*, 35(4):3167–3170.

Wang, Y., Li, L., Douville, C., *et al.* (2018). Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci Transl Med*, 10(433).

Wedren, S., Lovmar, L., Humphreys, K., *et al.* (2008). Estrogen receptor alpha gene polymorphism and endometrial cancer risk - a case-control study. *Bmc Cancer*, 8:11.

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(7145):661–678.

Wilczynski, M., Danielska, J., and Wilczynski, J. (2016). An update of the classical Bokhman's dualistic model of endometrial cancer. *Menopause Review-Przeglad Menopauzalny*, 15(2):63–68.

Win, A. K., Reece, J. C., and Ryan, S. (2015). Family History and Risk of Endometrial Cancer A Systematic Review and Meta-analysis. *Obstetrics and Gynecology*, 125(1):89–98.

Wu, Y. L., Meng, D. L., You, Y. X., *et al.* (2020). Increased expression of RBMS3 predicts a favorable prognosis in human gallbladder carcinoma. *Oncology Reports*, 44(1):55–68.

Xue, Z. W., Zhu, X. L., and Teng, Y. C. (2016). Relationship Between Murine Double Minute 2 (MDM2) T309G Polymorphism and Endometrial Cancer Risk: A Meta-Analysis. *Medical Science Monitor*, 22:3186–3190.

Yang, C. H., Almomen, A., Wee, Y. S., *et al.* (2015). An estrogen-induced endometrial hyperplasia mouse model recapitulating human disease progression and genetic aberrations. *Cancer Med*, 4(7):1039–50.

Yoneda, T., Kuboyama, A., Kato, K., *et al.* (2013). Association of MDM2 SNP309 and TP53 Arg72Pro polymorphisms with risk of endometrial cancer. *Oncology Reports*, 30(1):25–34.

Yu, Z., Wang, H., Fang, Y., *et al.* (2020). Molecular chaperone HspB2 inhibited pancreatic cancer cell proliferation via activating p53 downstream gene RPRM, BAI1, and TSAP6. *Journal of Cellular Biochemistry*, 121(3):2318–2329.

Zajac, A., Smolarz, B., Stachowiak, G., and Wilczynski, J. R. (2014). TP53 and MDM2 polymorphisms and the risk of endometrial cancer in postmenopausal women. *Medical Oncology*, 31(11):6.

Zajac, A., Stachowiak, G., Pertynski, T., *et al.* (2012). Association between MDM2 SNP309 polymorphism and endometrial cancer risk in Polish women. *Polish Journal of Pathology*, 63(4):278–283.

Zajac, A., Stachowiak, G., Smolarz, B., and Wilczynski, J. R. (2013). Polymorphisms of codon 72 of the TP53 gene in endometrial carcinoma of postmenopausal women. *Postepy Higieny I Medycyny Doswiadczalnej*, 67:1312–1318.

Zhang, J. X., Zhang, Y., and Zhang, Z. Y. (2018). Association of rs2279744 and rs117039649 promoter polymorphism with the risk of gynecological cancer A meta-analysis of case-control studies. *Medicine*, 97(2):9.

Zhang, Y. D., Hurson, A. N., Zhang, H. Y., *et al.* (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature Communications*, 11(1):13.

Zhao, M., Liu, Y., and O'Mara, T. A. (2016). ECGene: A Literature-Based Knowledgebase of Endometrial Cancer Genes. *Human Mutation*, 37(4):337–343.

Zhao, Y., Yang, X. E., Hao, X. J., *et al.* (2014). Common variant on MDM2 contributes to endometrial cancer susceptibility: evidence based on 7 studies. *Tumor Biology*, 35(8):7555–7560.

Zou, X., Zhang, Y., Zhang, L., *et al.* (2018). Association between MDM2 SNP309 and endometrial cancer risk: A PRISMA-compliant meta-analysis. *Medicine*, 97(49):e13273–e13273.

Zubor, P., Stanclova, A., Kajo, K., *et al.* (2009). The p53 Codon 72 Exon 4 BstUI Polymorphism and Endometrial Cancer in Caucasian Women. *Oncology*, 76(3):173–183.

7.1 Packages and Software

bcbio (2015). bcbio-nextgen.

Clayton, D. (2019). *snpStats: SnpMatrix and XSnpMatrix classes and methods*. R package version 1.34.0.

GenABEL (2013). *GenABEL: genome-wide SNP association analysis*. R package version 1.8-0.

Howie, B. and Marchini, J. (2020). IMPUTE [2].

Microsoft Corporation (2016). MS Excel 2016.

Moore, C. and Jacobson, S. (2021). *genpwr: Power Calculations Under Genetic Model Misspecification*. R package version 1.0.4.

Nakazawa, M. (2019). *fmsb: Functions for Medical Statistics Book with some Demographic Data*. R package version 0.7.0.

Purcell, S. and Chang, C. (2017). PLINK [1.9].

Purcell, S. and Chang, C. (2020). PLINK [2].

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and MÃ¼ller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20):7881.

StataCorp (2019). Stata Statistical Software: Release 16.

Turner, S. (2017). *qqman: Q-Q and Manhattan Plots for GWAS Data*. R package version 0.1.4.

Tyrer, J. (2020). Mlogit.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Zagury, J. and Marchini, J. (2020). SHAPEIT [2].

8 Appendices

8.1 Appendix 1

Publication: **Bafligil, C**., Thompson, D. J., Lophatananon, A., Smith, M. J., Ryan, N. A., Naqvi, A., Evans, D. G. and Crosbie, E. J. (2020), 'Association between genetic polymorphisms and endometrial cancer risk: a systematic review', *Journal of Medical Genetics* **57**(9), 591-600.

OPEN ACCESS

ORIGINAL RESEARCH

Association between genetic polymorphisms and endometrial cancer risk: a systematic review

Cemsel Bafligil ,[1] Deborah J Thompson,[2] Artitaya Lophatananon,[3] Miriam J Smith ,[4] Neil AJ Ryan,[1] Anie Naqvi,[5] D Gareth Evans ,[4] Emma J Crosbie [1,6]

[1]Division of Cancer Sciences, University of Manchester, Manchester, UK
[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[3]Division of Population Health, University of Manchester, Manchester, UK
[4]Division of Evolution and Genomic Sciences, University of Manchester, Manchester, UK
[5]University of Manchester Medical School, Manchester, UK
[6]Department of Obstetrics and Gynaecology, Manchester University NHS Foundation Trust, Manchester, UK

Correspondence to
Dr Emma J Crosbie, Division of Cancer Sciences, University of Manchester, Manchester, UK; emma.crosbie@manchester.ac.uk

Received 26 August 2019
Revised 24 November 2019
Accepted 15 December 2019

To cite: Bafligil C, Thompson DJ, Lophatananon A, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2019-106529

ABSTRACT

Introduction Endometrial cancer is one of the most commonly diagnosed cancers in women. Although there is a hereditary component to endometrial cancer, most cases are thought to be sporadic and lifestyle related. The aim of this study was to systematically review prospective and retrospective case–control studies, meta-analyses and genome-wide association studies to identify genomic variants that may be associated with endometrial cancer risk.

Methods We searched MEDLINE, Embase and CINAHL from 2007 to 2019 without restrictions. We followed PRISMA 2009 guidelines. The search yielded 3015 hits in total. Following duplicate exclusion, 2674 abstracts were screened and 453 full-texts evaluated based on our pre-defined screening criteria. 149 articles were eligible for inclusion.

Results We found that single nucleotide polymorphisms (SNPs) in *HNF1B*, *KLF*, *EIF2AK*, *CYP19A1*, *SOX4* and *MYC* were strongly associated with incident endometrial cancer. Nineteen variants were reported with genome-wide significance and a further five with suggestive significance. No convincing evidence was found for the widely studied *MDM2* variant rs2279744. Publication bias and false discovery rates were noted throughout the literature.

Conclusion Endometrial cancer risk may be influenced by SNPs in genes involved in cell survival, oestrogen metabolism and transcriptional control. Larger cohorts are needed to identify more variants with genome-wide significance.

INTRODUCTION

Endometrial cancer is the most common gynaecological malignancy in the developed world.[1] Its incidence has risen over the last two decades as a consequence of the ageing population, fewer hysterectomies for benign disease and the obesity epidemic. In the USA, it is estimated that women have a 1 in 35 lifetime risk of endometrial cancer, and in contrast to cancers of most other sites, cancer-specific mortality has risen by approximately 2% every year since 2008 related to the rapidly rising incidence.[2]

Endometrial cancer has traditionally been classified into type I and type II based on morphology.[3] The more common subtype, type I, is mostly comprised of endometrioid tumours and is oestrogen-driven, arises from a hyperplastic endometrium, presents at an early stage and has an excellent 5 year survival rate.[4] By contrast, type II includes non-endometrioid tumours, specifically serous, carcinosarcoma and clear cell subtypes, which are biologically aggressive tumours with a poor prognosis that are often diagnosed at an advanced stage.[5] Recent efforts have focused on a molecular classification system for more accurate categorisation of endometrial tumours into four groups with distinct prognostic profiles.[6 7]

The majority of endometrial cancers arise through the interplay of familial, genetic and lifestyle factors. Two inherited cancer predisposition syndromes, Lynch syndrome and the much rarer Cowden syndrome, substantially increase the lifetime risk of endometrial cancer, but these only account for around 3–5% of cases.[8–10] Having first or second degree relative(s) with endometrial or colorectal cancer increases endometrial cancer risk, although a large European twin study failed to demonstrate a strong heritable link.[11] The authors failed to show that there was greater concordance in monozygotic than dizygotic twins, but the study was based on relatively small numbers of endometrial cancers. Lu and colleagues reported an association between common single nucleotide polymorphisms (SNPs) and endometrial cancer risk, revealing the potential role of SNPs in explaining part of the risk in both the familial and general populations.[12] Thus far, many SNPs have been reported to modify susceptibility to endometrial cancer; however, much of this work predated genome wide association studies and is of variable quality. Understanding genetic predisposition to endometrial cancer could facilitate personalised risk assessment with a view to targeted prevention and screening interventions.[13] This emerged as the most important unanswered research question in endometrial cancer according to patients, carers and healthcare professionals in our recently completed James Lind Womb Cancer Alliance Priority Setting Partnership.[14] It would be particularly useful for non-endometrioid endometrial cancers, for which advancing age is so far the only predictor.[15]

We therefore conducted a comprehensive systematic review of the literature to provide an overview of the relationship between SNPs and endometrial cancer risk. We compiled a list of the most robust endometrial cancer-associated SNPs. We assessed the applicability of this panel of SNPs with a theoretical polygenic risk score (PRS) calculation. We also critically appraised the meta-analyses investigating the

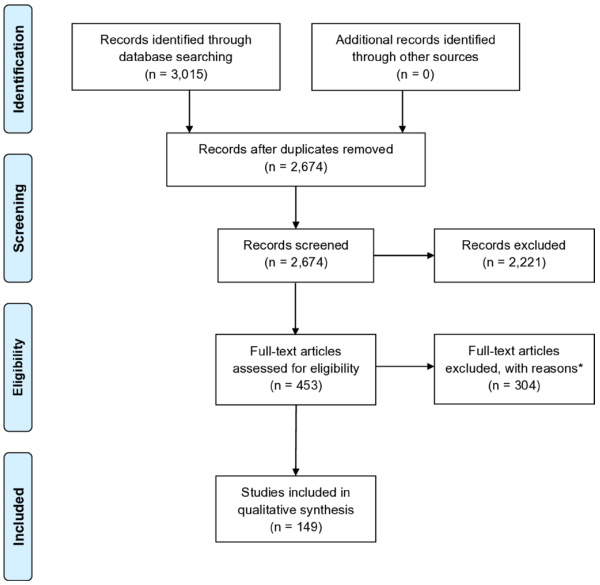


Figure 1 Study selection flow diagram. *Reasons: irrelevant articles, articles focusing on other conditions, non-GWAS/candidate-gene study related articles, technical and duplicate articles. GWAS, genome-wide association study. Adapted from: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(6): e1000097. doi:10.1371/journal.pmed1000097.

most frequently reported SNPs in *MDM2*. Finally, we described all SNPs reported within genes and pathways that are likely involved in endometrial carcinogenesis and metastasis.

METHODS

Our systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) collaboration 2009 recommendations. The registered protocol is available through PROSPERO (CRD42018091907).[16]

Search strategy

We searched Embase, MEDLINE and Cumulative Index to Nursing and Allied Health Literature (CINAHL) databases via the Healthcare Databases Advanced Search (HDAS) platform, from 2007 to 2018, to identify studies reporting associations between polymorphisms and endometrial cancer risk. Key words including MeSH (Medical Subject Heading) terms and free-text words were searched in both titles and abstracts. The following terms were used: "endomet*","uter*", "womb", "cancer(s)", "neoplasm(s)", "endometrium tumour", "carcinoma", "adenosarcoma", "clear cell carcinoma", "carcinosarcoma", "SNP", "single nucleotide polymorphism", "GWAS", and "genome-wide association study/ies". No other restrictions were applied. The search was repeated with time restrictions between 2018 and June 2019 to capture any recent publications.

Eligibility criteria

Studies were selected for full-text evaluation if they were primary articles investigating a relationship between endometrial cancer and SNPs. Study outcome was either the increased or decreased risk of endometrial cancer relative to controls reported as an odds ratio (OR) with corresponding 95% confidence intervals (95% CIs).

Study selection

Three independent reviewers screened all articles uploaded to a screening spreadsheet developed by Helena VonVille.[17] Disagreements were resolved by discussion. Chronbach's α score was calculated between reviewers and indicated high consistency at 0.92. Case–control, prospective and retrospective studies, genome-wide association studies (GWAS), and both discovery and validation studies were selected for full-text evaluation. Non-English articles, editorials, conference abstracts and proceedings, letters and correspondence, case reports and review articles were excluded.

Candidate-gene studies with at least 100 women and GWAS with at least 1000 women in the case arm were selected to ensure reliability of the results, as explained by Spencer *et al*.[18] To construct a panel of up to 30 SNPs with the strongest evidence of association, those with the strongest p values were selected. For the purpose of an SNP panel, articles utilising broad European or multi-ethnic cohorts were selected. Where overlapping populations were identified, the most comprehensive study was included.

Data extraction and synthesis

For each study, the following data were extracted: SNP ID, nearby gene(s)/chromosome location, OR (95% CI), p value, minor or effect allele frequency (MAF/EAF), EA (effect allele) and OA (other allele), adjustment, ethnicity and ancestry, number of cases and controls, endometrial cancer type, and study type including discovery or validation study and meta-analysis. For risk estimates, a preference towards most adjusted results was applied. For candidate-gene studies, a standard p value of <0.05 was applied and for GWAS a p value of $<5\times10^{-8}$, indicating genome-wide significance, was accepted as statistically significant. However, due to the limited number of SNPs with p values reaching genome-wide significance, this threshold was then lowered to $<1\times10^{-5}$, allowing for marginally significant SNPs to be included. As shown by Mavaddat *et al*, for breast cancer, SNPs that fall below genome-wide significance may still be useful for generating a PRS and improving the models.[19]

We estimated the potential value of a PRS based on the most significant SNPs by comparing the predicted risk for a woman with a risk score in the top 1% of the distribution to the mean predicted risk. Per-allele ORs and MAFs were taken from the publications and standard errors (SEs) for the lnORs were derived from published 95% CIs. The PRS was assumed to have a Normal distribution, with mean $2\sum\beta_i p_i$ and SE, σ, equal to $\sqrt{2\sum\beta_i^2 p_i(1-p_i)}$, according to the binomial distribution, where the summation is over all SNPs in the risk score. Hence the relative risk (RR) comparing the top 1% of the distribution to the mean is given by $\exp(Z_{0.01}\sigma)$, where Z is the inverse of the standard normal cumulative distribution.

RESULTS

The flow chart of study selection is illustrated in figure 1. In total, 453 text articles were evaluated and, of those, 149 articles met our inclusion criteria. One study was excluded from table 1, for having an Asian-only population, as this would make it harder to compare with the rest of the results which were all either multi-ethnic or Caucasian cohorts, as stated in our inclusion criteria for the SNP panel.[20] Any SNPs without 95% CIs were also excluded from any downstream analysis. Additionally,

Table 1 List of top SNPs most likely to contribute to endometrial cancer risk identified through systematic review of recent literature[21–25]

Reference	SNP ID	Nearby gene(s)	Location	OR	LCI	UCI	P	EAF	EA	OA	Ethnicity	Cases (n)	Controls (n)	EC type	Position	Datasets*
O'Mara *et al*, 2018[21]	rs11263761	HNF1B	17q12	1.15	1.12	1.19	3.20e-20	0.52	A	G	EUR	12906	108979	All	Intronic	**NSECG**, UK1-CORGI, **SEARCH**, WTCCC, **ANECS**, QIMR, HCS, E2C2, BECS/ HJECS, **iCOGS**† (BBCC, BSUCH, ESTHER, GC-HBOC, GENICA, MARIE, **MoMaTEC**, NBCS, BBCS, SBCS, UKBGS), **NECS**, ABCFS, ABCTB, BCEES, MCCS, **LES**, LMBC, **BECS**, GESBC, HaBCS, CAHRES, RENDOCAS, MISS, pKARMA, SMC, CBR_STUDY98, MECS, MCBCS, MMHS, WHI, UKBB
	rs7981863	KLF5, KLF12	13q22.1	1.16	1.12	1.20	2.70e-17	0.72	C	T	EUR	12906	108979	All	Intergenic	
	rs1740828	SOX4	6p22.3	1.15	1.11	1.19	4.20e-16	0.52	G	A	EUR	12906	108979	All	Regulatory region	
	rs17601876	CYP19A1	15q21.2	1.12	1.09	1.16	3.30e-14	0.48	G	A	EUR	12906	108979	All	Intronic	
	rs4733613	MYC	8q24.21	1.18	1.13	1.24	7.50e-14	0.12	C	G	EUR	12906	108979	All	Intergenic	
	rs3184504	SH2B3	12q24.11	1.10	1.07	1.14	1.10e-10	0.52	C	T	EUR	12906	108979	All	Missense	
	rs2747716	HEY2, NCOA7, MYC	6q22.31	1.10	1.07	1.14	2.90e-10	0.57	A	G	EUR	12906	108979	All	Intronic	
	rs9668337	SSPN	12p12.1	1.11	1.08	1.15	1.10e-09	0.74	A	G	EUR	12906	108979	All	Non-coding exon	
	rs35286446	MYC	8q24.21	1.10	1.06	1.13	3.10e-09	0.58	GAT	G	EUR	12906	108979	All	Intronic	
	rs10850382	LOC107984437	12q24.21	1.10	1.07	1.14	3.50e-09	0.31	T	C	EUR	12906	108979	All	Regulatory region	
	rs882380	SKAP, SNX11	17q21.32	1.10	1.06	1.13	4.70e-09	0.61	A	C	EUR	12906	108979	All	Intronic	
	rs937213	EIF2AK4, BMF	15q15.1	1.09	1.06	1.13	5.10e-09	0.42	C	T	EUR	12906	108979	All	Intronic	
	rs1679014	CDKN2A, CDKN2B	9p21.3	1.18	1.12	1.25	6.40e-09	0.07	T	C	EUR	12906	108979	All	Intronic	
Painter *et al*, 2016[22]	rs2498794	AKT1	14q32.33	1.13	1.09	1.17	8.70e-09	0.48	G	A	EUR	7737	37144	All	Intronic	**ANECS**, **SEARCH**, **NSECG**, **iCOGS**
O'Mara *et al*, 2018[21]	rs10835920	WT1, WT1-AS, EIF3M	11p13	1.09	1.06	1.13	1.30e-08	0.38	T	C	EUR	12906	108979	All	Intergenic	**NSECG**, UK1-CORGI, **SEARCH**, WTCCC, **ANECS**, QIMR, HCS, E2C2, BECS/ HJECS, **iCOGS**, BBCC, BSUCH, ESTHER, GC-HBOC, GENICA, MARIE, **MoMaTEC**, NBCS, BBCS, SBCS, UKBGS, **NECS**, ABCFS, ABCTB, BCEES, MCCS, **LES**, LMBC, **BECS**, GESBC, HaBCS, CAHRES, RENDOCAS, MISS, pKARMA, SMC, CBR_STUDY98, MECS, MCBCS, MMHS, WHI, UKBB
	rs1395847729	MYC	8q24.21	1.40	1.25	1.58	2.40e-08	0.98	C	G	EUR	12906	108979	All	Intergenic	
	rs1482611157	BCL11A	2p16.1	1.26	1.16	1.36	3.40e-08	0.03	A	G	EUR	12906	108979	All	Intergenic	
				1.25	1.14	1.38	4.70e-06	0.03				8758	46126	Endometrioid		
				1.64	1.32	2.04	9.60e-06	0.03				1230	35447	Non-endometrioid		
	rs113998067	GNL2, RSPO1, CDCA8	1p34.3	1.23	1.14	1.32	3.60e-08	0.04	C	T	EUR	12906	108979	All	Intergenic	
	rs1129506	EVI2A, NF1	17q11.2	1.10	1.06	1.13	4.30e-08	0.38	G	A	EUR	12906	108979	All	Missense	
Spurdle *et al*, 2011[23]	rs673604	SFPQ	1p34	1.21	1.12	1.32	5.90e-06	0.08	G	A	EUR	1265	5190	All	Regulatory region	**ANECS**, **NECS**, **SEARCH**, WTCCC2, **BECS**, **LES**, **MoMaTEC**, **NSECG**, PECS, SASBAC, SECGS
O'Mara *et al*, 2015[24]	rs79575945	ESR1	6q25	1.20	1.11	1.30	3.76e-06	0.07	G	A	EUR	6607	37925	Endometrioid	Intronic	**iCOGS**, BCAC, OCAC, **ANECS**, **SEARCH**, **NSECG**
Chen *et al*, 2014[25]	rs1953358	LINC00520	14q22.3	1.36	1.20	1.53	4.76e-07	0.49	G	A	ME	1055	1778	Endometrioid	Intergenic	AHS, EDGE, FHCRC, MEC
	rs8178648	PROS1	chr3	1.71	1.37	2.12	1.53e-06	0.09	G	A	ME	1055	1778	Endometrioid	Intronic	
	rs9399840	N/A	6q16.3	1.33	1.18	1.49	3.01e-06	0.53	T	C	ME	1055	1778	Endometrioid	Intergenic	

Continued

Table 1 Continued

Reference	SNP ID	Nearby gene(s)	Location	OR	LCI	UCI	P	EAF	EA	OA	Ethnicity	Cases (n)	Controls (n)	EC type	Position	Datasets*

The different studies listed here used overlapping datasets (in bold).
All locations were based on Genome Reference Consortium Human Build 37 (GRCh37). Variants located at 8q24.21 were obtained from a conditional model
*NSECG: UK National Study of Endometrial Cancer Genetics, UK1-CORGI: UK Colorectal Tumour Gene Identification Consortium, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, WTCCC1/2: Wellcome Trust Case Control Consortium 1/2, ANECS: Australian National Endometrial Cancer Study, QIMR: Queensland Institute of Medical Research, HCS: Hunter Community Study, E2C2: NCI-supported international consortium of four US-based cohort studies, 2 US-based case-control studies and 1 Polish case-control study, BECS/HJECS: Bavarian Endometrial Cancer Study/Hannover-Jena Endometrial Cancer Study, BBCC: Bavarian Breast Cancer Cases and Controls, BSUCH: Breast Cancer Study of the University Clinic Heidelberg, ESTHER: ESTHER Breast Cancer Study, GC-HBOC: German Consortium for Hereditary Breast & Ovarian Cancer, GENICA: Gene Environment Interaction and Breast Cancer in Germany, MARIE: Mammary Carcinoma Risk Factor Investigation, MoMaTEC: Molecular Markers in Treatment of Endometrial Cancer, NBCS: Norwegian Breast Cancer Study, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, NSECG: National Study of Endometrial Cancer Genetics, BBCS: British Breast Cancer Study, SBCS: Sheffield Breast Cancer Study, UKBGS: UK Breakthrough Generations Study, ANECS: Australian National Endometrial Cancer Study, NECS: Newcastle Endometrial Cancer Study, ABCFS: Australian Breast Cancer Family Study, ABCTB: Australian Breast Cancer Tissue Bank, BCEES: Breast Cancer Employment and Environment Study, MCCS: Melbourne Collaborative Cohort Study, LES: Leuven Endometrial Cancer Study, LMBC: Leuven Multidisciplinary Breast Centre, BECS: Bavarian Endometrial Cancer Study, BBCC: Bavarian Breast Cancer Cases and Controls, BSUCH: Breast Cancer Study of the University Clinic Heidelberg, GENICA: Gene Environment Interaction and Breast Cancer in Germany, GESBC: Genetic Epidemiology Study of Breast Cancer by Age 50, HaBCS: Hannover Breast Cancer Study, MARIE: Mammary Carcinoma Risk Factor Investigation, CAHRES: Cancer Hormone Replacement Epidemiology, RENDOCAS: Registry of Endometrial Cancer in Sweden, MISS: Melanoma Inquiry of Southern Sweden, pKARMA: Karolinska Mammography Project for Risk Prediction of Breast Cancer, SMC: Swedish Mammography Cohort, SEARCH: UK Studies of Epidemiology and Risk factors in Cancer Heredity, NSECG: National Study of Endometrial Cancer Genetics, BBCS: British Breast Cancer Study, CBR_STUDY98: Cambridge BioResource, UKBGS: UK Breakthrough Generations Study, MECS: Mayo Endometrial Cancer Study, MCBCS: Mayo Clinic Breast Cancer Study, MMHS: Mayo Mammography Health Study, WHI: Women's Health Initiative, UKBB: UK BioBank, OCAC: Ovarian Cancer Association Consortium, AHS: Alberta Health Services, EDGE: Oestrogen, Diet, Genetics and Endometrial Cancer, FHCRC: Fred Hutchinson Cancer Research Centre, MEC: Multiethnic Cohort Study, PECS: Polish Endometrial Cancer Study, SASBAC: Singapore and Swedish Breast/Endometrial Cancer Study, SECGS: Shanghai Endometrial Cancer Genetic Study.
†iCOGS dataset breakdown was not indicated by the listed studies here other than O'Mara *et al*., 2018.[21]
EA, effect allele; EAF, effect allele frequency; EC, endometrial cancer; EUR, European cohort; LCI, lower confidence interval; ME, multiethnic; OA, other allele; SNP, single nucleotide polymorphism; UCI, upper confidence interval.

SNPs in linkage disequilibrium (r^2 >0.2) with each other were examined, and of those in linkage disequilibrium, the SNP with strongest association was reported. Per allele ORs were used unless stated otherwise.

Top SNPs associated with endometrial cancer risk

Following careful interpretation of the data, 24 independent SNPs with the lowest p values that showed the strongest association with endometrial cancer were obtained (table 1).[21–25] These SNPs are located in or around genes coding for transcription factors, cell growth and apoptosis regulators, and enzymes involved in the steroidogenesis pathway. All the SNPs presented here were reported on the basis of a GWAS or in one case, an exome-wide association study, and hence no SNPs from candidate-gene studies made it to the list. This is partly due to the nature of larger GWAS providing more comprehensive and powered results as opposed to candidate gene studies. Additionally, a vast majority of SNPs reported by candidate-gene studies were later refuted by large-scale GWAS such as in the case of *TERT* and *MDM2* variants.[26 27] The exception to this is the *CYP19* gene, where candidate-gene studies reported an association between variants in this gene with endometrial cancer in both Asian and broad European populations, and this association was more recently confirmed by large-scale GWAS.[21 28–30] Moreover, a recent article authored by O'Mara and colleagues reviewed the GWAS that identified most of the currently known SNPs associated with endometrial cancer.[31]

Most of the studies represented in table 1 are GWAS and the majority of these involved broad European populations. Those having a multi-ethnic cohort also consisted primarily of broad European populations. Only four of the variants in table 1 are located in coding regions of a gene, or in regulatory flanking regions around the gene. Thus, most of these variants would not be expected to cause any functional effects on the gene or the resulting protein. An eQTL search using GTEx Portal showed that some of the SNPs are significantly associated ($p<0.05$) with modified transcription levels of the respective genes in various tissues such as prostate (rs11263761), thyroid (rs9668337), pituitary (rs2747716), breast mammary (rs882380) and testicular (rs2498794) tissue, as summarised in table 2.

The only variant for which there was an indication of a specific association with non-endometrioid endometrial cancer was rs148261157 near the *BCL11A* gene. The A allele of this SNP had a moderately higher association in the non-endometrioid arm (OR 1.64, 95% CI 1.32 to 2.04; $p=9.6\times10^{-6}$) compared with the endometrioid arm (OR 1.25, 95% CI 1.14 to 1.38; $p=4.7\times10^{-6}$).[21]

Oestrogen receptors α and β encoded by *ESR1* and *ESR2*, respectively, have been extensively studied due to the assumed role of oestrogens in the development of endometrial cancer. O'Mara *et al* reported a lead SNP (rs79575945) in the *ESR1* region that was associated with endometrial cancer ($p=1.86\times10^{-5}$).[24] However, this SNP did not reach genome-wide significance in a more recent larger GWAS.[21] No statistically significant associations have been reported between endometrial cancer and SNPs in the *ESR2* gene region.

AKT is an oncogene linked to endometrial carcinogenesis. It is involved in the PI3K/AKT/mTOR pro-proliferative signalling pathway to inactivate apoptosis and allow cell survival. The A allele of rs2494737 and G allele of rs2498796 were reported to be associated with increased and decreased risk of endometrial cancer in 2016, respectively.[22 30] However, this association was not replicated in a larger GWAS in 2018.[21] Nevertheless, given

Table 2 List of eQTL hits for the selected panel of SNPs

SNP ID	Significant eQTL for	P	Tissue	Other gene(s)	Other tissue(s)
rs17601876	GLDN	1.2e-08	Adipose – subcutaneous	SPPL2A, DMXL2	Skin – sun exposed (lower leg); colon – sigmoid; cells – cultured fibroblasts; muscle – skeletal; spleen; skin – not sun exposed (suprapubic); nerve – tibial
	CYP19A1	3.4e-07	Whole blood		
	CYP19A1	5.8e-06	Adipose – subcutaneous		
rs3184504	TMEM116	1.7e-04	Adipose – subcutaneous	ALDH2, LINC01405, ADAM1B	Oesophagus – mucosa; skin – not sun exposed (suprapubic); skin – sun exposed (lower leg); muscle – skeletal; artery – aorta; heart – atrial appendage; artery – tibial; colon – sigmoid; brain – nucleus accumbens (basal ganglia)
	MAPKAPK5	2.6e-04	Adipose – subcutaneous		
rs2747716	RP11-624M8.1	4.2e-11	Pituitary	HDDC2	Artery – tibial; pancreas; thyroid; brain – nucleus accumbens (basal ganglia); brain – substantia nigra; oesophagus – muscularis; nerve – tibial; Brain – caudate (basal ganglia); adipose – visceral (omentum); brain – spinal cord (cervical c-1); artery – aorta; brain – cortex; brain – hypothalamus; muscle – skeletal; brain – cerebellum; heart – left ventricle; brain – putamen (basal ganglia); brain – frontal cortex (BA9); brain – cerebellar hemisphere
	RP11-624M8.1	8.2e-11	Adipose – subcutaneous		
	HEY2	9.7e-10	Testis		
	HEY2	2.1e-09	Ovary		
	RP11-624M8.1	1.7e-07	Breast – mammary tissue		
	RP11-624M8.1	1.3e-06	Ovary		
rs9668337	BHLHE41	9.0e-17	Thyroid	RP11-283G6.3	Cells – cultured fibroblasts
	SSPN	1.1e-04	Thyroid		
rs882380	SNX11	3.1e-25	Adipose – subcutaneous	RP5-890E16.5, CBX1, LRRC46, MRPL10, RP11-6N17.4, CDK5RAP3, SP6, PRR15L, RP5-890E16.2, PNPO, RP11-6N17.3, HOXB1, HOXB-AS1, NFE2L1	Skin – sun exposed (lower leg); cells – cultured fibroblasts; adipose – visceral (omentum); lung; skin – not sun exposed (suprapubic); pancreas; spleen; oesophagus – muscularis; artery – aorta; heart – atrial appendage; liver; colon – transverse; thyroid; artery – tibial; colon – sigmoid; oesophagus – gastro-oesophageal junction; stomach; muscle – skeletal; small intestine – terminal Ileum; prostate; brain – cerebellum; brain – cerebellar hemisphere; minor salivary gland; adrenal gland; oesophagus – mucosa
	SNX11	1.0e-21	Whole blood		
	SNX11	1.2e-13	Breast – mammary tissue		
	COPZ2	9.3e-12	Testis		
	SKAP1	3.3e-08	Whole blood		
	HOXB2	2.6e-05	Adipose – subcutaneous		
rs937213	EIF2AK4	4.7e-11	Adipose – visceral (omentum)	SRP14	Thyroid; oesophagus – mucosa; skin – sun exposed (lower leg); stomach; oesophagus – muscularis; pancreas; skin – not sun exposed (suprapubic); colon – transverse; adipose – subcutaneous; lung; colon – sigmoid; muscle – skeletal; nerve – tibial; whole blood; oesophagus – gastro-oesophageal junction; artery – tibial; adrenal gland; spleen; heart – left ventricle; heart – atrial appendage
	EIF2AK4	3.4e-08	Breast – mammary tissue	N/A	
	RP11-521C20.5	5.4e-07	Testis	N/A	
	RP11-521C20.5	7.4e-07	Prostate	N/A	
rs2498794	AKT1	1.7e-30	Thyroid	ZBTB42	Oesophagus – mucosa; artery – tibial; oesophagus – muscularis; skin – sun exposed (lower leg); skin – not sun exposed (suprapubic); cells – cultured fibroblasts; artery – aorta; oesophagus – gastro-oesophageal junction; adipose – subcutaneous; colon – sigmoid; colon – transverse; heart – atrial appendage
	ADSSL1	5.5e-25	Testis		
	SIVA1	1.8e-07	Adipose – visceral (omentum)		
	ADSSL1	2.6e-05	Ovary		
	SIVA1	4.4e-05	Breast – mammary tissue		
rs10835920	WT1-AS	5.5e-06	Spleen	N/A	Oesophagus –- muscularis
rs148261157	KIAA1841	1.3e-05	Oesophagus – muscularis	N/A	N/A
rs113998067	RSPO1	2.7e-10	Artery – tibial	EPHA10, FHL3, DNALI1	Nerve – tibial; artery – aorta; colon – transverse
rs1129506	EVI2A	4.3e-20	Whole blood	OMG, RAB11FIP4	Spleen; oesophagus – mucosa; artery – tibial; lung; artery – aorta; skin – sun exposed (lower leg); nerve – tibial; heart – atrial appendage; adipose – visceral (omentum); cells – cultured fibroblasts; liver; stomach; brain – amygdala; skin – not sun exposed (suprapubic); brain – caudate (basal ganglia); muscle – skeletal; colon – sigmoid
	NF1	3.5e-09	Adipose – subcutaneous		
	NF1	2.2e-07	Thyroid		
	NF1	3.7e-07	Testis		
rs673604	ZMYM1	7.0e-07	Adipose – subcutaneous	RP4-665N4.8, ZMYM4, KIAA0319L, TFAP2E	Skin – sun exposed (lower leg); oesophagus – muscularis; cells – EBV-transformed lymphocytes; oesophagus – mucosa; nerve – tibial; brain – cerebellum
	MAP7D1	1.0e-05	Whole blood		
rs1953358	LINC00520	1.5e-05	Skin – not sun exposed (suprapubic)	N/A	N/A
rs8178648	PROS1	3.0e-04	Skin – sun exposed (lower leg)	N/A	N/A

Top significant eQTL hits from different tissues are shown in the table. There were no significant hits reported for some SNPs which are hence not included in this table.
EBV, Epstein-Barr virus; SNP, single nucleotide polymorphism.

the previous strong indications, and biological basis that could explain endometrial carcinogenesis, we decided to include an *AKT1* variant (rs2498794) in our results.

PTEN is a multi-functional tumour suppressor gene that regulates the AKT/PKB signalling pathway and is commonly mutated in many cancers including endometrial cancer.[32] Loss-of-function germline mutations in *PTEN* are responsible for Cowden syndrome, which exerts a lifetime risk of endometrial cancer of up to 28%.[9] Lacey and colleagues studied SNPs in the *PTEN* gene region; however, none showed significant differences in frequency between 447 endometrial cancer cases and 439 controls of European ancestry.[33]

KRAS mutations are known to be present in endometrial cancer. These can be activated by high levels of KLF5

(transcriptional activator). Three SNPs have been identified in or around *KLF5* that are associated with endometrial cancer. The G allele of rs11841589 (OR 1.15, 95% CI 1.11 to 1.21; $p=4.83\times10^{-11}$), the A allele of rs9600103 (OR 1.23, 95% CI 1.16 to 1.30; $p=3.76\times10^{-12}$) and C allele of rs7981863 (OR 1.16, 95% CI 1.12 to 1.20; $p=2.70\times10^{-17}$) have all been found to be associated with an increased likelihood of endometrial cancer in large European cohorts.[21 30 34] It is worth noting that these SNPs are not independent, and hence they quite possibly tag the same causal variant.

The *MYC* family of proto-oncogenes encode transcription factors that regulate cell proliferation, which can contribute to cancer development if dysregulated. The recent GWAS by O'Mara *et al* reported three SNPs within the *MYC* region that reached genome-wide significance with conditional p values reaching at least 5×10^{-8}.[35]

To test the utility of these SNPs as predictive markers, we devised a theoretical PRS calculation using the log ORs and EAFs per SNP from the published data. The results were very encouraging with an RR of 3.16 for the top 1% versus the mean, using all the top SNPs presented in table 1 and 2.09 when using only the SNPs that reached genome-wide significance (including *AKT1*).

Controversy surrounding *MDM2* variant SNP309

MDM2 negatively regulates tumour suppressor gene *TP53*, and as such, has been extensively studied in relation to its potential role in predisposition to endometrial cancer. Our search identified six original studies of the association between *MDM2* SNP rs2279744 (also referred to as SNP309) and endometrial cancer, all of which found a statistically significant increased risk per copy of the G allele. Two more original studies were identified through our full-text evaluation; however, these were not included here as they did not meet our inclusion criteria—one due to small sample size, the other due to studying rs2279744 status dependent on another SNP.[36 37] Even so, the two studies were described in multiple meta-analyses that are listed in table 3. Different permutations of these eight original studies appear in at least eight published meta-analyses. However, even the largest meta-analysis contained <2000 cases (table 3)[38]

In comparison, a GWAS including nearly 13 000 cases found no evidence of an association with OR and corresponding 95% CI of 1.00 (0.97 to 1.03) and a p value of 0.93 (personal communication).[21] Nevertheless, we cannot completely rule out a role for *MDM2* variants in endometrial cancer predisposition as the candidate-gene studies reported larger effects in Asians, whereas the GWAS primarily contained participants of European ancestry. There is also some suggestion that the SNP309 variant is in linkage disequilibrium with another variant, SNP285, which confers an opposite effect.

It is worth noting that the SNP285C/SNP309G haplotype frequency was observed in up to 8% of Europeans, thus requiring correction for the confounding effect of SNP285C in European studies.[39] However, aside from one study conducted by Knappskog *et al*, no other study including the meta-analyses corrected for the confounding effect of SNP285.[40] Among the studies presented in table 3, Knappskog *et al* (2012) reported that after correcting for SNP285, the OR for association of this haplotype with endometrial cancer was much lower, though still significant. Unfortunately, the meta-analyses which synthesised Knappskog *et al* (2012), as part of their analysis, did not correct for SNP285C in the European-based studies they included.[38 41 42] It is also concerning that two meta-analyses using the same primary articles failed to report the same result, in two instances.[38 42–44]

Table 3 Characteristics of studies that examined *MDM2* SNP rs2279744

Reference	OR (95% CI)	P values	EAF	Ancestry	Cases (n)	Controls (n)	EC type	Dataset(s)
Terry 2008[48]	1.32 (1.11 to 1.56)	0.002	N/A	European	591	1543	N/A	NHS (Nurses' Health Study), WHS (Women's Health Study)
Ashton 2009[49]	1.37 (1.06 to 1.79)	N/A	0.56	Caucasian	191	291	All	Hospital based
Nunobiki 2009[50]	2.28 (2.02 to 2.54)	0.030	0.49	Japanese	102	95	All	Hospital based
Ueda 2009[51]	1.91 (1.5 to 3.47)	0.035	0.51	Japanese	119	108	All	Hospital based
Wan 2011[43]	1.54 (1.21 to 1.94)	0.000	N/A	N/A	N/A	N/A	N/A	Walsh 2007,[36] Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009
Li 2011[44]	1.75 (1.16 to 2.63)	0.007	N/A	European, Asian	1001	1889	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009
Knappskog 2012[40]	1.22 (1.03 to 1.44)	N/A	0.36	European	392	956	N/A	Hospital based
Zajac 2012[27]	1.33 (1.12 to 1.58)	0.001	N/A	European	152	100	N/A	Hospital based
Yoneda 2013[52]	1.64 (0.81 to 3.28)	0.450	0.45	Asian	125	200	All	Population based
Peng 2013[41]	1.6 (1.21 to 2.13)	0.001	N/A	European, Asian	2069	4546	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Knappskog 2012, Yoneda 2013
	1.87 (1.29 to 2.73)	0.010	N/A	European	1842	4251	N/A	
Zhao 2014[53]	1.41 (1.04 to 1.92)	0.030	N/A	European, Asian	1278	2189	N/A	Walsh 2007, Terry 2008, Ashton 2009, Ueda 2009, Zajac 2012, Yoneda 2013
	1.34 (1.07 to 1.69)	N/A	N/A	European	859	1707	N/A	
Wang 2014[38]	1.32 (1.06 to 1.64)	0.010	N/A	European, Asian	1967	4460	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012, Knappskog 2012, Yoneda 2013
	1.14 (0.79 to 1.65)	0.490	N/A	European	1769	4172	N/A	
Xue 2016[42]	1.46 (1.25 to 1.72)	N/A	N/A	European	1690	4151	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012, Knappskog 2012, Yoneda 2013
Zhang 2018[54]	1.91 (1.5 to 3.47)	0.035	N/A	European, Asian	762	1041	N/A	Walsh 2007, Terry 2008, Ashton 2009, Nunobiki 2009, Ueda 2009, Zajac 2012
Zou 2018[55]	1.23 (1.06 to 1.41)	0.005	N/A	European, Asian, mixed	3535	6476	All	Walsh 2007, Terry 2008, Ashton 2009, Ueda 2009, Knappskog 2012, Zajac 2012, Yoneda 2013, Okamoto 2015, Gansmo 2017[37]

*Walsh *et al* 2007 and Gansmo *et al* 2017 did not meet eligibility criteria for us to include in our evaluation.
EAF, effect allele frequency; EC, endometrial cancer; SNP, single nucleotide polymorphism.

DISCUSSION

This article represents the most comprehensive systematic review to date, regarding critical appraisal of the available evidence of common low-penetrance variants implicated in predisposition to endometrial cancer. We have identified the most robust SNPs in the context of endometrial cancer risk. Of those, only 19 were significant at genome-wide level and a further five were considered marginally significant. The largest GWAS conducted in this field was the discovery- and meta-GWAS by O'Mara *et al*, which utilised 12 096 cases and 108 979 controls.[21] Despite the inclusion of all published GWAS and around 5000 newly genotyped cases, the total number did not reach anywhere near what is currently available for other common cancers such as breast cancer. For instance, BCAC (Breast Cancer Association Consortium) stands at well over 200 000 individuals with more than half being cases, and resulted in identification of ~170 SNPs in relation to breast cancer.[19 45] A total of 313 SNPs including imputations were then used to derive a PRS for breast cancer.[19] Therefore, further efforts should be directed to recruit more patients, with deep phenotypic clinical data to allow for relevant adjustments and subgroup analyses to be conducted for better precision.

A recent pre-print study by Zhang and colleagues examined the polygenicity and potential for SNP-based risk prediction for 14 common cancers, including endometrial cancer, using available summary-level data from European-ancestry datasets.[46] They estimated that there are just over 1000 independent endometrial cancer susceptibility SNPs, and that a PRS comprising all such SNPs would have an area under the receiver-operator curve of 0.64, similar to that predicted for ovarian cancer, but lower than that for the other cancers in the study. The modelling in the paper suggests that an endometrial cancer GWAS double the size of the current largest study would be able to identify susceptibility SNPs together explaining 40% of the genetic variance, but that in order to explain 75% of the genetic variance it would be necessary to have a GWAS comprising close to 150 000 cases and controls, far in excess of what is currently feasible.

We found that the literature consists mainly of candidate-gene studies with small sample sizes, meta-analyses reporting conflicting results despite using the same set of primary articles, and multiple reports of significant SNPs that have not been validated by any larger GWAS. The candidate-gene studies were indeed the most useful and cheaper technique available until the mid to late 2000s. However, a lack of reproducibility (particularly due to population stratification and reporting bias), uncertainty of reported associations, and considerably high false discovery rates make these studies much less appropriate in the post-GWAS era. Unlike the candidate-gene approach, GWAS do not require prior knowledge, selection of genes or SNPs, and provide vast amounts of data. Furthermore, both the genotyping process and data analysis phases have become cheaper, the latter particularly due to faster and open-access pre-phasing and imputation tools being made available.

It is clear from table 1 that some SNPs were reported with wide 95% CI, which can be directly attributed to small sample sizes particularly when restricting the cases to non-endometrioid histology only, low EAF or poor imputation quality. Thus, these should be interpreted with caution. Additionally, most of the SNPs reported by candidate-gene studies were not detected by the largest GWAS to date conducted by O'Mara *et al*.[21] However, this does not necessarily mean that the possibility of those SNPs being relevant should be completely dismissed. Moreover, meta-analyses were attempted for other variants; however, these showed no statistically significant association and many presented with high heterogeneity between the respective studies (data not shown). Furthermore, as many studies utilised the same set of cases and/or controls, conducting a meta-analysis was not possible for a good number of SNPs. It is therefore unequivocal that the literature is crowded with numerous small candidate-gene studies and conflicting data. This makes it particularly hard to detect novel SNPs and conduct meaningful meta-analyses.

We found convincing evidence for 19 variants that indicated the strongest association with endometrial cancer, as shown in table 1. The associations between endometrial cancer and variants in or around *HNF1B*, *CYP19A1*, *SOX4*, *MYC*, *KLF* and *EIF2AK* found in earlier GWAS were then replicated in the latest and largest GWAS. These SNPs showed promising potential in a theoretical PRS we devised based on published data. Using all 24 or genome-wide significant SNPs only, women with a PRS in the top 1% of the distribution would be predicted to have a risk of endometrial cancer 3.16 and 2.09 times higher than the mean risk, respectively.

However, the importance of these variants and relevance of the proximate genes in a functional or biological context is challenging to evaluate. Long distance promoter regulation by enhancers may disguise the genuine target gene. In addition, enhancers often do not loop to the nearest gene, further complicating the relevance of nearby gene(s) to a GWAS hit. In order to elucidate biologically relevant candidate target genes in endometrial cancer, O'Mara *et al* looked into promoter-associated chromatin looping using a modern HiChIP approach.[47] The authors utilised normal and tumoural endometrial cell lines for this analysis which showed significant enrichment for endometrial cancer heritability, with 103 candidate target genes identified across the 13 risk loci identified by the largest ECAC GWAS. Notable genes identified here were *CDKN2A* and *WT1*, and their antisense counterparts. The former was reported to be nearby of rs1679014 and the latter of rs10835920, as shown in table 1. Moreover, of the 36 candidate target genes, 17 were found to be downregulated while 19 were upregulated in endometrial tumours.

The authors also investigated overlap between the 13 endometrial cancer risk loci and top eQTL variants for each target gene.[47] In whole blood, of the two particular lead SNPs, rs8822380 at 17q21.32 was a top eQTL for *SNX11* and *HOXB2*, whereas rs937213 at 15q15.1 was a top eQTL for *SRP14*. In endometrial tumour, rs7579014 at 2p16.1 was found to be a top eQTL for *BCL11A*. This is particularly interesting because *BCL11A* was the only nearby/candidate gene that had a GWAS association reported in both endometrioid and non-endometrioid subtypes. The study looked at protein–protein interactions between endometrial cancer drivers and candidate target gene products. Significant interactions were observed with TP53 (most significant), AKT, PTEN, ESR1 and KRAS, among others. Finally, when 103 target candidate genes and 387 proteins were combined together, 462 pathways were found to be significantly enriched. Many of these are related to gene regulation, cancer, obesity, insulinaemia and oestrogen exposure. This study clearly showed a potential biological relevance for some of the SNPs reported by ECAC GWAS in 2018.

Most of the larger included studies used cohorts primarily composed of women of broad European descent. Hence, there are negligible data available for other ethnicities, particularly African women. This is compounded by the lack of reference genotype data available for comparative analysis, making it harder for research to be conducted in ethnicities other than Europeans. This poses a problem for developing risk prediction

models that are equally valuable and predictive across populations. Thus, our results also are of limited applicability to non-European populations.

Furthermore, considering that non-endometrioid cases comprise a small proportion (~20%) of all endometrial cancer cases, much larger cohort sizes are needed to detect any genuine signals for non-endometrioid tumours. Most of the evaluated studies looked at either overall/mixed endometrial cancer subtypes or endometrioid histology, and those that looked at variant associations with non-endometrioid histology were unlikely to have enough power to detect any signal with statistical significance. This is particularly concerning because non-endometrioid subtypes are biologically aggressive tumours with a much poorer prognosis that contribute disproportionately to mortality from endometrial cancer. It is particularly important that attempts to improve early detection and prevention of endometrial cancer focus primarily on improving outcomes from these subtypes. It is also worth noting that, despite the current shift towards a molecular classification of endometrial cancer, most studies used the overarching classical Bokhman's classification system, type I versus type II, or no histological classification system at all. Therefore, it is important to create and follow a standardised and comprehensive classification system for reporting tumour subtypes for future studies.

This study compiled and presented available information for an extensively studied, yet unproven in large datasets, SNP309 variant in *MDM2*. Currently, there is no convincing evidence for an association between this variant and endometrial cancer risk. Additionally, of all the studies, only one accounted for the opposing effect of a nearby variant SNP285 in their analyses. Thus, we conclude that until confirmed by a sufficiently large GWAS, this variant should not be considered significant in influencing the risk of endometrial cancer and therefore not included in a PRS. This is also true for the majority of the SNPs reported in candidate-gene studies, as the numbers fall far short of being able to detect genuine signals.

This systematic review presents the most up-to-date evidence for endometrial cancer susceptibility variants, emphasising the need for further large-scale studies to identify more variants of importance, and validation of these associations. Until data from larger and more diverse cohorts are available, the top 24 SNPs presented here are the most robust common genetic variants that affect endometrial cancer risk. The multiplicative effects of these SNPs could be used in a PRS to allow personalised risk prediction models to be developed for targeted screening and prevention interventions for women at greatest risk of endometrial cancer.

Twitter Emma J Crosbie @DrEmmaCrosbie

Contributors CB planned the study, did the systematic review, analysed the data and wrote the manuscript. DJT and AL supervised the study and provided statistical support for the analysis. MJS supervised the study. NAJR and AN supported data acquisition. DGE and EJC designed and planned the study, provided supervision and wrote the manuscript. EJC provided funding for the study. All authors reviewed and approved the final manuscript.

Funding CB, DGE, AL, MJS and EJC are all supported by the National Institute for Health Research (NIHR) Manchester Biomedical Research Centre (IS-BRC-1215-20007). NAJR was supported through a Medical Research Council (MRC) Doctoral Research Fellowship (MR/M018431/1), EJC through an NIHR Clinician Scientist Fellowship (NIHR-CS-012-009) and DGE is an NIHR Senior Investigator (NF-SI-0513–10076). This research was funded by the NIHR Manchester BRC.

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The protocol for this systematic review was published at PROSPERO and the data that inform this manuscript are available upon reasonable request from the corresponding author.

ORCID iDs
Cemsel Bafligil https://orcid.org/0000-0002-2365-1194
Miriam J Smith http://orcid.org/0000-0002-3184-0817
D Gareth Evans http://orcid.org/0000-0002-8482-5784
Emma J Crosbie http://orcid.org/0000-0003-0284-8630

REFERENCES

1 Sundar S, Balega J, Crosbie E, Drake A, Edmondson R, Fotopoulou C, Gallos I, Ganesan R, Gupta J, Johnson N, Kitson S, Mackintosh M, Martin-Hirsch P, Miles T, Rafii S, Reed N, Rolland P, Singh K, Sivalingam V, Walther A. BGCS uterine cancer guidelines: recommendations for practice. *Eur J Obstet Gynecol Reprod Biol* 2017;213:71–97.
2 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7–34.
3 Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. *Lancet* 2016;387:1094–108.
4 Tzur T, Kessous R, Weintraub AY. Current strategies in the diagnosis of endometrial cancer. *Arch Gynecol Obstet* 2017;296:5–14.
5 Clarke MA, Devesa SS, Harvey SV, Wentzensen N. Hysterectomy-corrected uterine corpus cancer incidence trends and differences in relative survival reveal racial disparities and rising rates of nonendometrioid cancers. *J Clin Oncol* 2019;37:1895–908.
6 Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DAGetz G, Gabriel SB, Cibulskis K, Lander E, Sivachenko A, Sougnez C, Lawrence M, Dooling D, Fulton R, Fulton L, Kalicki-Veizer J, McLellan MD, O'Laughlin M, Schmidt H, Wilson RK, Ye K, Ally A, Balasundaram M, Birol I, Butterfield YSN, Carlsen R, Carter C, Chu A, Chuah E, Chun HJE, Dhalla N, Guin R, Hirst C, Holt RA, Jones SJM, Lee D, HI L, Marra MA, Mayo M, Moore RA, Mungall AJ, Plettner P, Schein JE, Sipahimalani P, Tam A, Varhol RJ, Saksena G, Onofrio RC, Schumacher SE, Tabak B, Carter SL, Hernandez B, Gentry J, Salvesen HB, Ardlie K, Winckler W, Beroukhim R, Meyerson M, Hadjipanayis A, Lee S, Mahadeshwar HS, Park P, Protopopov A, Ren XJ, Seth S, Song XZ, Tang JB, RB X, Yang LX, Zeng D, Chin L, Zhang JH, Auman JT, Balu S, Bodenheimer T, Buda E, Hayes DN, Hoyle AP, Jefferys SR, Jones CD, Meng SW, Mieczkowski PA, Mose LE, Parker JS, Perou CM, Roach J, Shi Y, Simons JV, Soloway MG, Tan DH, Topal MD, Waring S, JY W, Hoadley KA, Baylin SB, Bootwalla MS, Lai PH, Triche TJ, Van Den Berg D, Weisenberger DJ, Cho J, DiCara D, Frazer S, Heiman D, Jing R, Lin P, Mallard W, Stojanov P, Voet D, Zhang HL, Zou LH, Noble M, Reynolds SM, Shmulevich I, Aksoy BA, Antipin Y, Ciriello G, Dresdner G, Gao JJ, Gross B, Jacobsen A, Ladanyi M, Reva B, Sander C, Sinha R, Sumer SO, Taylor BS, Cerami E, Weinhold N, Benz S, Goldstein T, Haussler D, Ng S, Szeto C, Stuart J, Annala M, Broom BM, Casasent TD, ZL J, Liang H, YL L, Unruh AK, Wakefield C, Weinstein JN, Zhang NX, Liu YX, Broaddus R, Adams C, Barr T, Black AD, Bowen J, Deardurff J, Frick J, Gastier-Foster JM, Grossman T, Harper HA, Hart-Kothari M, Helsel C, Hobensack A, Kuck H, Kneile K, Leraas K, Lichtenberg TM, McAllister C, Pyatt RE, Ramirez NC, Tabler TR, Vanhoose N, White P, Wise L, Zmuda E, Barnabas N, Berry-Green C, Blanc V, Boice L, Button M, Farkas A, Green A, MacKenzie J, Nicholson D, Kalloger SE, Gilks CB, Karlan BY, Lester J, Orsulic S, Borowsky M, Cadungog M, Czerwinski C, Huelsenbeck-Dill L, Iacocca M, Petrelli N, Witkin G, Nemirovich-Danchenko E, Potapova O, Rotin D, Berchuck A, Birrer M, DiSaia P, Monovich L, Curley E, Gardner J, Mallery D, Penny R, Dowdy SC, Winterhoff B, Dao L, Gostout B, Meuter A, Teoman A, Dao F, Olvera N, Bogomolniy F, Garg K, Soslow RA, Abramov M, Bartlett JMS, Kodeeswaran S, Parfitt J, Moiseenko F, Clarke BA, Goodman MT, Carney ME, Matsuno RK, Fisher J, Huang M, Rathmell WK, Thorne L, Van Le L DR, Edwards R, Elishaev E, Zorn K, Goodfellow PJ, Mutch D, Kahn AB, Bell DW, Pollock PM, Wang C, Wheeler DA, Shinbrot E, Ayala B, Chu AL, Jensen MA, Kothiyal P, Pihl TD, Pontius J, Pot DA, Snyder EE, Srinivasan D, Shaw KRM, Sheth M, Davidsen T, Eley G, Ferguson ML, Demchok JA, Yang LM, Guyer MS, Ozenberger BA, Sofia HJ, Shen R, Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013;497:67–73.
7 Stelloo E, Bosse T, Nout RA, MacKay HJ, Church DN, Nijman HW, Leary A, Edmondson RJ, Powell ME, Crosbie EJ, Kitchener HC, Mileshkin L, Pollock PM, Smit VT, Creutzberg CL. Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol* 2015;28:836–44.
8 Win AK, Reece JC, Ryan S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet Gynecol* 2015;125:89–98.

9 Constantinou P, Tischkowitz M. Genetics of gynaecological cancers. *Best Pract Res Clin Obstet Gynaecol* 2017;42:114–24.
10 Ryan NAJ, Glaire MA, Blake D, Cabrera-Dandy M, Evans DG, Crosbie EJ. The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genet Med* 2019.
11 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
12 Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, Pharoah PD, Thompson DJ, Dunning AM, Hayward NK, Chenevix-Trench G, Macgregor S, Q-MEGA and AMFS Investigators, ANECS-SEARCH, UKOPS-SEARCH, BEACON Consortium. Most common 'sporadic' cancers have a significant germline genetic component. *Hum Mol Genet* 2014;23:6112–8.
13 Kitson SJ, Evans DG, Crosbie EJ. Identifying high-risk women for endometrial cancer prevention strategies: proposal of an endometrial cancer risk prediction model. *Cancer Prev Res* 2017;10:1–13.
14 Wan YL, Beverley-Stevenson R, Carlisle D, Clarke S, Edmondson RJ, Glover S, Holland J, Hughes C, Kitchener HC, Kitson S, Miles T, Morley R, Morrison J, Nelson L, Powell M, Sadler L, Tomlinson A, Tylko-Hill K, Whitcombe J, Crosbie EJ. Working together to shape the endometrial cancer research agenda: the top ten unanswered research questions. *Gynecol Oncol* 2016;143:287–93.
15 Raglan O, Kalliala I, Markozannes G, Cividini S, Gunter MJ, Nautiyal J, Gabra H, Paraskevaidis E, Martin-Hirsch P, Tsilidis KK, Kyrgiou M. Risk factors for endometrial cancer: an umbrella review of the literature. *Int J Cancer* 2019;145:1719–30.
16 PROSPERO. *Low-penetrance common predisposition variants in endometrial cancer risk: a systematic review*, 2018.
17 VonVille H. Excel workbooks for systematic reviews, 2015. Available: https://showcase.dropbox.com/s/Excel-Workbooks-for-Systematic-Reviews-User-Guides-Kf4pYVTvFqSJdZR2Spzlu [Accessed 17 Oct 2017].
18 Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping CHIP. *PLoS Genet* 2009;5:13.
19 Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen T-H, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomäki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barrdahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale A-L, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collée JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Försti A, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W, John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarek K, Kataja V, Keeman R, Kerin MJ, Khusnutdinova E, Kiiski JI, Knight JA, Ko Y-D, Kosma V-M, Koutros S, Kristensen VN, Krüger U, Kühl T, Lambrechts D, Le Marchand L, Lee E, Lejbkowicz F, Lilyquist J, Lindblom A, Lindström S, Lissowska J, Lo W-Y, Loibl S, Long J, Lubiński J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Muñoz-Garzon VM, Neuhausen SL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon T-W, Perez JIA, Pérez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pylkäs K, Rack B, Radice P, Rau-Murthy R, Rennert G, Rennert HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu X-O, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thöne K, Tollenaar RAEM, Tomlinson I, Truong T, Tzardi M, Ulmer H-U, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Ziogas A, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, García-Closas M, Simard J, Easton DF, Freeman LEB, Alnaes GIG, YD K, WY L, Kostovska IM, Rennert HS, Clarke C, Balleine R, Baxter R, Braye S, Carpenter J, Dahlstrom J, Forbes J, Lee CS, Marsh D, Morey A, Pathmanathan N, Scott R, Simpson P, Spigelman A, Wilcken N, Yip D, Zeps N, Sexton A, Dobrovic A, Christian A, Trainer A, Fellows A, Shelling A, De Fazio A, Blackburn A, Crook A, Meiser B, Patterson B, Saunders C, Hunt C, Amor D, Ortega DG, Edkins E, Salisbury E, Haan E, Macrea F, Farshid G, Lindeman G, Trench G, Mann G, Giles G, Gill G, Thorne H, Campbell I, Hickie I, Caldon L, Winship I, Cui J, Flanagan J, Kollias J, Visvader J, Taylor J, Burke J, Saunus J, Hopper J, Beesley J, Kirk J, French J, Tucker K, Wu K, Phillips K, Forrest L, Lipton L, Andrews L, Lobb L, Walker L, Kentwell M, Spurdle M, Cummings M, Gleeson M, Harris M, Jenkins M, Young MA, Delatycki M, Wallis M, Burgess M, Brown M, Southey M, Bogwitz M, Field M, Friedlander M, Gattas M, Saleh M, Aghmesheh M, Hayward N, Pachter N, Cohen P, Duijf P, James P, Fong P, Butow P, Williams R, Kefford R, Dawson SJ, Lok S, O'Connell S, Greening S, Nightingale S, Edwards S, Fox S, McLachlan SA, Lakhani S, Dudding T, Antill Y, Sahlberg KK, Ottestad L, Karesen R, Schlichting E, Holmen MM, Sauer T, Haakensen V, Engebraten O, Naume B, Fossa A, Kiserud CE, Reinertsen KV, Helland A, Riis M, Geisler J, ABCTB Investigators, kConFab/AOCS Investigators, NBCS Collaborators. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet* 2019;104:21–34.
20 Geng YH, Wang ZF, Jia YM, Zheng LY, Chen L, Liu DG, XH L, Tian XX, Fang WG. Genetic polymorphisms in CDH1 are associated with endometrial carcinoma susceptibility among Chinese Han women. *Oncol Lett* 2018;16:6868–78.
21 O'Mara TA, Glubb DM, Amant F, Annibali D, Ashton K, Attia J, Auer PL, Beckmann MW, Black A, Bolla MK, Brauch H, Brenner H, Brinton L, Buchanan DD, Burwinkel B, Chang-Claude J, Chanock SJ, Chen C, Chen MM, Cheng THT, Clarke CL, Clendenning M, Cook LS, Couch FJ, Cox A, Crous-Bous M, Czene K, Day F, Dennis J, Depreeuw J, Doherty JA, Dörk T, Dowdy SC, Dürst M, Ekici AB, Fasching PA, Fridley BL, Friedenreich CM, Fritschi L, Fung J, García-Closas M, Gaudet MM, Giles GG, Goode EL, Gorman M, Haiman CA, Hall P, Hankison SE, Healey CS, Hein A, Hillemanns P, Hodgson S, Hoivik EA, Holliday EG, Hopper JL, Hunter DJ, Jones A, Krakstad C, Kristensen VN, Lambrechts D, Marchand LL, Liang X, Lindblom A, Lissowska J, Long J, Lu L, Magliocco AM, Martin L, McEvoy M, Meindl A, Michailidou K, Milne RL, Mints M, Montgomery GW, Nassir R, Olsson H, Orlow I, Otton G, Palles C, Perry JRB, Peto J, Pooler L, Prescott J, Proietto T, Rebbeck TR, Risch HA, Rogers PAW, Rübner M, Runnebaum I, Sacerdote C, Sarto GE, Schumacher F, Scott RJ, Setiawan VW, Shah M, Sheng X, Shu X-O, Southey MC, Swerdlow AJ, Tham E, Trovik J, Turman C, Tyrer JP, Vachon C, VanDen Berg D, Vanderstichele A, Wang Z, Webb PM, Wentzensen N, Werner HMJ, Winham SJ, Wolk A, Xia L, Xiang Y-B, Yang HP, Yu H, Zheng W, Pharoah PDP, Dunning AM, Kraft P, De Vivo I, Tomlinson I, Easton DF, Spurdle AB, Thompson DJ. Identification of nine new susceptibility loci for endometrial cancer. *Nat Commun* 2018;9.
22 Painter JN, Kaufmann S, O'Mara TA, Hillman KM, Sivakumaran H, Darabi H, Cheng THT, Pearson J, Kazakoff S, Waddell N, Hoivik EA, Goode EL, Scott RJ, Tomlinson I, Dunning AM, Easton DF, French JD, Salvesen HB, Pollock PM, Thompson DJ, Spurdle AB, Edwards SL. A common variant at the 14q32 endometrial cancer risk locus activates Akt1 through YY1 binding. *Am J Hum Genet* 2016;98:1159–69.
23 Spurdle AB, Thompson DJ, Ahmed S, Ferguson K, Healey CS, O'Mara T, Walker LC, Montgomery SB, Dermitzakis ET, Fahey P, Montgomery GW, Webb PM, Fasching PA, Beckmann MW, Ekici AB, Hein A, Lambrechts D, Coenegrachts L, Vergote I, Amant F, Salvesen HB, Trovik J, Njolstad TS, Helland H, Scott RJ, Ashton K, Proietto T, Otton G, Tomlinson I, Gorman M, Howarth K, Hodgson S, Garcia-Closas M, Wentzensen N, Yang H, Chanock S, Hall P, Czene K, Liu J, Li J, Shu X-O, Zheng W, Long J, Xiang Y-B, Shah M, Morrison J, Michailidou K, Pharoah PD, Dunning AM, Easton DF, Australian National Endometrial Cancer Study Group, National Study of Endometrial Cancer Genetics Group. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet* 2011;43:451–4.
24 O'Mara TA, Glubb DM, Painter JN, Cheng T, Dennis J, Attia J, Holliday EG, McEvoy M, Scott RJ, Ashton K, Proietto T, Otton G, Shah M, Ahmed S, Healey CS, Gorman M, Martin L, Hodgson S, Fasching PA, Hein A, Beckmann MW, Ekici AB, Hall P, Czene K, Darabi H, Li J, Dürst M, Runnebaum I, Hillemanns P, Dörk T, Lambrechts D, Depreeuw J, Annibali D, Amant F, Zhao H, Goode EL, Dowdy SC, Fridley BL, Winham SJ, Salvesen HB, Njølstad TS, Trovik J, Werner HMJ, Tham E, Liu T, Mints M, Bolla MK, Michailidou K, Tyrer JP, Wang Q, Hopper JL, Peto J, Swerdlow AJ, Burwinkel B, Brenner H, Meindl A, Brauch H, Lindblom A, Chang-Claude J, Couch FJ, Giles GG, Kristensen VN, Cox A, Pharoah PDP, Dunning AM, Tomlinson I, Easton DF, Thompson DJ, Spurdle AB, Australian National Endometrial Cancer Study Group (ANECS), National Study of Endometrial Cancer Genetics Group (NSECG), RENDOCAS, AOCS Group. Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocr Relat Cancer* 2015;22:851–61.
25 Chen MM, Crous-Bou M, Setiawan VW, Prescott J, Olson SH, Wentzensen N, Black A, Brinton L, Chen C, Chen C, Cook LS, Doherty J, Friedenreich CM, Hankinson SE, Hartge P, Henderson BE, Hunter DJ, Le Marchand L, Liang X, Lissowska J, Lu L, Orlow I, Petruzella S, Polidoro S, Pooler L, Rebbeck TR, Risch H, Sacerdote C, Schumacher F, Sheng X, Shu X-ou, Weiss NS, Xia L, Van Den Berg D, Yang HP, Yu H, Chanock S, Haiman C, Kraft P, De Vivo I, Vivo D I. Exome-wide association study of endometrial cancer in a multiethnic population. *PLoS One* 2014;9:9.
26 Cheng THT, Thompson D, Painter J, O'Mara T, Gorman M, Martin L, Palles C, Jones A, Buchanan DD, Win AK, Hopper J, Jenkins M, Lindor NM, Newcomb PA, Gallinger S, Conti D, Schumacher F, Casey G, Giles GG, Pharoah P, Peto J, Cox A, Swerdlow A, Couch F, Cunningham JM, Goode EL, Winham SJ, Lambrechts D, Fasching P, Burwinkel B, Brenner H, Brauch H, Chang-Claude J, Salvesen HB, Kristensen V, Darabi H, Li J, Liu T, Lindblom A, Hall P, de Polanco ME, Sans M, Carracedo A, Castellvi-Bel S, Rojas-Martinez A, Aguiar Jnr S, Teixeira MR, Dunning AM, Dennis J, Otton G, Proietto T, Holliday E, Attia J, Ashton K, Scott RJ, McEvoy M, Dowdy SC, Fridley BL, Werner HMJ, Trovik J, Njolstad TS, Tham E, Mints M, Runnebaum I, Hillemanns P, Dörk T, Amant F, Schrauwen S, Hein A, Beckmann MW, Ekici A, Czene K, Meindl A, Bolla MK,

Michailidou K, Tyrer JP, Wang Q, Ahmed S, Healey CS, Shah M, Annibali D, Depreeuw J, Al-Tassan NA, Harris R, Meyer BF, Whiffin N, Hosking FJ, Kinnersley B, Farrington SM, Timofeeva M, Tenesa A, Campbell H, Haile RW, Hodgson S, Carvajal-Carmona L, Cheadle JP, Easton D, Dunlop M, Houlston R, Spurdle A, Tomlinson I. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci Rep* 2015;5:12.
27 Zając A, Stachowiak G, Pertyński T, Romanowicz H, Wilczyński J, Smolarz B. Association between MDM2 SNP309 polymorphism and endometrial cancer risk in Polish women. *Pjp* 2012;4:278–83.
28 Lundin E, Wirgin I, Lukanova A, Afanasyeva Y, Krogh V, Axelsson T, Hemminki K, Clendenen TV, Arslan AA, Ohlson N, Sieri S, Roy N, Koenig KL, Idahl A, Berrino F, Toniolo P, Hallmans G, Försti A, Muti P, Lenner P, Shore RE, Zeleniuch-Jacquotte A. Selected polymorphisms in sex hormone-related genes, circulating sex hormones and risk of endometrial cancer. *Cancer Epidemiol* 2012;36:445–52.
29 Tao MH, Cai Q, Zhang Z-F, Xu W-H, Kataoka N, Wen W, Xiang Y-B, Zheng W, Shu XO. Polymorphisms in the CYP19A1 (aromatase) gene and endometrial cancer risk in Chinese women. *Cancer Epidemiol Biomarkers Prev* 2007;16:943–9.
30 Cheng TH, Thompson DJ, O'Mara TA, Painter JN, Glubb DM, Flach S, Lewis A, French JD, Freeman-Mills L, Church D, Gorman M, Martin L, Hodgson S, Webb PM, Attia J, Holliday EG, McEvoy M, Scott RJ, Henders AK, Martin NG, Montgomery GW, Nyholt DR, Ahmed S, Healey CS, Shah M, Dennis J, Fasching PA, Beckmann MW, Hein A, Ekici AB, Hall P, Czene K, Darabi H, Li J, Dörk T, Dürst M, Hillemanns P, Runnebaum I, Amant F, Schrauwen S, Zhao H, Lambrechts D, Depreeuw J, Dowdy SC, Goode EL, Fridley BL, Winham SJ, Njølstad TS, Salvesen HB, Trovik J, Werner HM, Ashton K, Otton G, Proietto T, Liu T, Mints M, Tham E, Consortium C, Jun Li M, Yip SH, Wang J, Bolla MK, Michailidou K, Wang Q, Tyrer JP, Dunlop M, Houlston R, Palles C, Hopper JL, Peto J, Swerdlow AJ, Burwinkel B, Brenner H, Meindl A, Brauch H, Lindblom A, Chang-Claude J, Couch FJ, Giles GG, Kristensen VN, Cox A, Cunningham JM, Pharoah PDP, Dunning AM, Edwards SL, Easton DF, Tomlinson I, Spurdle AB, National Study of Endometrial Cancer Genetics Group (NSECG), Australian National Endometrial Cancer Study Group (ANECS), RENDOCAS, AOCS Group. Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat Genet* 2016;48:667–74.
31 O'Mara TA, Glubb DM, Kho PF, Thompson DJ, Spurdle AB. Genome-wide association studies of endometrial cancer: latest developments and future directions. *Cancer Epidemiol Biomarkers Prev* 2019;28:1095–102.
32 Banno K, Yanokura M, Iida M, Masuda K, Aoki D. Carcinogenic mechanisms of endometrial cancer: involvement of genetics and epigenetics. *J Obstet Gynaecol Res* 2014;40:1957–67.
33 Lacey JV, Yang H, Gaudet MM, Dunning A, Lissowska J, Sherman ME, Peplonska B, Brinton LA, Healey CS, Ahmed S, Pharoah P, Easton D, Chanock S, Garcia-Closas M. Endometrial cancer and genetic variation in PTEN, PIK3CA, Akt1, MLH1, and MSH2 within a population-based case-control study. *Gynecol Oncol* 2011;120:167–73.
34 Chen MM, O'Mara TA, Thompson DJ, Painter JN, Attia J, Black A, Brinton L, Chanock S, Chen C, Cheng TH, Cook LS, Crous-Bou M, Doherty J, Friedenreich CM, Garcia-Closas M, Gaudet MM, Gorman M, Haiman C, Hankinson SE, Hartge P, Henderson BE, Hodgson S, Holliday EG, Horn-Ross PL, Hunter DJ, Le Marchand L, Liang X, Lissowska J, Long J, Lu L, Magliocco AM, Martin L, McEvoy M, Olson SH, Orlow I, Pooler L, Prescott J, Rastogi R, Rebbeck TR, Risch H, Sacerdote C, Schumacher F, Wendy Setiawan V, Scott RJ, Sheng X, Shu X-O, Turman C, Van Den Berg D, Wang Z, Weiss NS, Wentzensen N, Xia L, Xiang Y-B, Yang HP, Yu H, Zheng W, Pharoah PDP, Dunning AM, Tomlinson I, Easton DF, Kraft P, Spurdle AB, De Vivo I, LG L, Setiawan VW, Vivo D I, Australian National Endometrial Cancer Study Group (ANECS), National Study Of Endometrial Cancer Genetics Group (NSECG). GWAS meta-analysis of 16 852 women identifies new susceptibility locus for endometrial cancer. *Hum Mol Genet* 2016;25:ddw092–20.
35 Setiawan VW, Haessler J, Schumacher F, Cote ML, Deelman E, Fesinmeyer MD, Henderson BE, Jackson RD, Vöckler J-S, Wilkens LR, Yasmeen S, Haiman CA, Peters U, Le Marchand L, Kooperberg C. Hnf1B and endometrial cancer risk: results from the page study. *PLoS One* 2012;7:6.
36 Walsh CS, Miller CW, Karlan BY, Koeffler HP. Association between a functional single nucleotide polymorphism in the MDM2 gene and sporadic endometrial cancer risk. *Gynecol Oncol* 2007;104:660–4.
37 Gansmo LB, Bjørnslett M, Halle MK, Salvesen HB, Romundstad P, Hveem K, Vatten L, Dørum A, Lønning PE, Knappskog S. Mdm2 promoter polymorphism del1518 (rs3730485) and its impact on endometrial and ovarian cancer risk. *BMC Cancer* 2017;17:6.
38 Wang L-H, Wang X, Xu W-T, Hu Y-L. Mdm2 rs2279744 polymorphism and endometrial cancer: a meta-analysis. *Tumor Biol* 2014;35:3167–70.
39 Knappskog S, Gansmo LB, Dibirova K, Metspalu A, Cybulski C, Peterlongo P, Aaltonen L, Vatten L, Romundstad P, Hveem K, Devilee P, Evans GD, Lin D, Van Camp G, Manolopoulos VG, Osorio A, Milani L, Ozcelik T, Zalloua P, Mouzaya F, Bliznetz E, Balanovska E, Pocheshkova E, Kučinskas V, Atramentova L, Nymadawa P, Titov K, Lavryashina M, Yusupov Y, Bogdanova N, Koshel S, Zamora J, Wedge DC, Charlesworth D, Dörk T, Balanovsky O, Lønning PE. Population distribution and ancestry of the cancer protective MDM2 SNP285 (rs117039649). *Oncotarget* 2014;5:8223–34.
40 Knappskog S, Trovik J, Marcickiewicz J, Tingulstad S, Staff AC, Romundstad P, Hveem K, Vatten L, Salvesen HB, Lønning PE. SNP285C modulates oestrogen receptor/Sp1 binding to the MDM2 promoter and reduces the risk of endometrial but not prostatic cancer. *Eur J Cancer* 2012;48:1988–96.
41 Peng Q, Mo C, Qin A, Lao X, Chen Z, Sui J, Wu J, Zhai L, Yang S, Qin X, Li S. Mdm2 SNP309 polymorphism contributes to endometrial cancer susceptibility: evidence from a meta-analysis. *J Exp Clin Cancer Res* 2013;32.
42 Xue Z, Zhu X, Teng Y. Relationship between murine double minute 2 (MDM2) T309G polymorphism and endometrial cancer risk: a meta-analysis. *Med Sci Monit* 2016;22:3186–90.
43 Wan Y, Wu W, Yin Z, Guan P, Zhou B. Mdm2 SNP309, gene-gene interaction, and tumor susceptibility: an updated meta-analysis. *BMC Cancer* 2011;11:9.
44 Li Y, Zhao H, Sun L, Huang L, Yang Q, Kong B. Mdm2 SNP309 is associated with endometrial cancer susceptibility: a meta-analysis. *Hum Cell* 2011;24:57–64.
45 Hamdi Y, Soucy P, Adoue V, Michailidou K, Canisius S, Lemaçon A, Droit A, Andrulis IL, Anton-Culver H, Arndt V, Baynes C, Blomqvist C, Bogdanova NV, Bojesen SE, Bolla MK, Bonanni B, Borresen-Dale A-L, Brand JS, Brauch H, Brenner H, Broeks A, Burwinkel B, Chang-Claude J, Couch FJ, Cox A, Cross SS, Czene K, Darabi H, Dennis J, Devilee P, Dörk T, Dos-Santos-Silva I, Eriksson M, Fasching PA, Figueroa J, Flyger H, García-Closas M, Giles GG, Goldberg MS, González-Neira A, Grenaker-Alnæs G, Guénel P, Haeberle L, Haiman CA, Hamann U, Hallberg E, Hooning MJ, Hopper JL, Jakubowska A, Jones M, Kabisch M, Kataja V, Lambrechts D, Le Marchand L, Lindblom A, Lubinski J, Mannermaa A, Maranian M, Margolin S, Marme F, Milne RL, Neuhausen SL, Nevanlinna H, Neven P, Olswold C, Peto J, Plaseska-Karanfilska D, Pylkäs K, Radice P, Rudolph A, Sawyer EJ, Schmidt MK, Shu X-O, Southey MC, Swerdlow A, Tollenaar RAEM, Tomlinson I, Torres D, Truong T, Vachon C, Van Den Ouweland AMW, Wang Q, Winqvist R, Zheng W, Benitez J, Chenevix-Trench G, Dunning AM, Pharoah PDP, Kristensen V, Hall P, Easton DF, Pastinen T, Nord S, Simard J, Collaborators N, Truong Th, Ouweland AMWVD, Easton A, Pastinen T, Nord S, Simard J, Couch FJ, NBCS Collaborators, kConFab/AOCS Investigators. Association of breast cancer risk with genetic variants showing differential allelic expression: identification of a novel breast cancer susceptibility locus at 4q21. *Oncotarget* 2016;7:80140–63.
46 Zhang Y, Wilcox A, Zhang H, Choudhury PP, Easton D, Milne R, Simard J, Hall P, Michailidou K, Dennis J, Schmidt M, Chang-Claude J, Gharahkhani P, Whiteman D, Campbell P, Hoffmeister M, Jenkins M, Peters U, Hsu L, Gruber S, Casey G, Schmit S, O'Mara T, Spurdle A, Thompson D, Tomlinson I, Vivo D I, Landi MT, Law M, Iles M, Demenais F, Kumar R, MacGregor S, Bishop T, Ward S, Bondy M, Houlston R, Wiencke J, Melin B, Barnholtz-Sloan J, Kinnersley B, Wrensch M, Amos C, Hung R, Brennan P, McKay J, Caporaso N, Berndt S, Birmann B, Camp N, Kraft P, Rothman N, Slager S, Berchuck A, Pharoah PDP, Sellers T, Gayther S, Pearce C, Goode E, Schildkraut J, Moysich K, Amundadottir L, Jacobs E, Klein A, Petersen G, Risch H, Stolzenberg-Solomon R, Wolpin B, Li D, Eeles R, Haiman C, Kote-Jarai Z, Schumacher F, Olama AAA, Purdue M, Scelo G, Dalgaard M, Greene M, Grotmol T, Kanetsky P, McGlynn K, Nathanson K, Turnbull C, Wiklund F, Chanock S, Chatterjee N, Garcia-Closas M, Bcac B, Ccfr C, Ecac G, GenoMel G, Ilcco I, InterLymph O, Oral Cancer G, Panc P. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *bioRxiv* 2019.
47 O'Mara TA, Spurdle AB, Glubb DM, Endometrial Cancer Association Consortium Endometrial Cancer Association Consortium. Analysis of promoter-associated chromatin interactions reveals biologically relevant candidate target genes at endometrial cancer risk loci. *Cancers* 2019;11:1440.
48 Terry K, McGrath M, Lee I-M, Buring J, De Vivo I, Vivo D I. Mdm2 SNP309 is associated with endometrial cancer risk. *Cancer Epidemiol Biomarkers Prev* 2008;17:983–6.
49 Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, Attia J, Gilbert M, Hamann U, Scott RJ. Polymorphisms in TP53 and MDM2 combined are associated with high grade endometrial cancer. *Gynecol Oncol* 2009;113:109–14.
50 Nunobiki O, Ueda M, Yamamoto M, Toji E, Sato N, Izuma S, Okamoto Y, Torii K, Noda S. Polymorphisms of p53 codon 72 and MDM2 promoter 309 and the risk of endometrial cancer. *Hum Cell* 2009;22:101–6.
51 Ueda M, Yamamoto M, Nunobiki O, Toji E, Sato N, Izuma S, Okamoto Y, Torii K, Noda S. Murine double-minute 2 homolog single nucleotide polymorphism 309 and the risk of gynecologic cancer. *Hum Cell* 2009;22:49–54.
52 Yoneda T, Kuboyama A, Kato K, Ohgami T, Okamoto K, Saito T, Wake N. Association of MDM2 SNP309 and TP53 Arg72Pro polymorphisms with risk of endometrial cancer. *Oncol Rep* 2013;30:25–34.
53 Zhao Y, Yang X, Hao X, Pan X, Zhao B, Ma J, Fang J, Zhao M. Common variant on MDM2 contributes to endometrial cancer susceptibility: evidence based on 7 studies. *Tumor Biol* 2014;35:7555–60.
54 Zhang JX, Zhang Y, Zhang ZY. Association of rs2279744 and rs117039649 promoter polymorphism with the risk of gynecological cancer a meta-analysis of case-control studies. *Medicine* 2018;97:9.
55 Zou XW, Zhang Y, Zhang L, JX L, Zhu CJ, Cheng QH, Zhou JH, Chen YG. Association between MDM2 SNP309 and endometrial cancer risk a PRISMA-compliant meta-analysis. *Medicine* 2018;97:11.

8.2 Appendix 2

Publication: Kitson, S., **Bafligil, C**., Ryan, N., Lalloo, F., Woodward, E., Clayton, R., Edmondson, R., Bolton, J., Crosbie, E., & Evans, D. G. (2020). 'BRCA1 and BRCA2 pathogenic variant carriers and endometrial cancer risk-a cohort study', *European Journal of Cancer* **136**, 169-175.

European Journal of Cancer 136 (2020) 169–175

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

BRCA1 and BRCA2 pathogenic variant carriers and endometrial cancer risk: A cohort study

Sarah J. Kitson [a], Cemsel Bafligil [a], Neil A.J. Ryan [a], Fiona Lalloo [b], Emma R. Woodward [b,e], Richard D. Clayton [c], Richard J. Edmondson [a,c], James Bolton [d], Emma J. Crosbie [a,c,1], D. Gareth Evans [b,e,1,*]

[a] *Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, St Mary's Hospital, Manchester, UK*
[b] *Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK*
[c] *Department of Obstetrics and Gynaecology, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK*
[d] *Department of Histopathology, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK*
[e] *Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK*

Received 2 March 2020; received in revised form 11 May 2020; accepted 24 May 2020
Available online 19 July 2020

KEYWORDS
Endometrial cancer;
BRCA1;
BRCA2;
Risk;
Serous endometrial cancer

Abstract ***Background:*** An association between *BRCA* pathogenic variants and an increased endometrial cancer risk, specifically serous-like endometrial cancer, has been postulated but remains unproven, particularly for *BRCA2* carriers. Mechanistic evidence is lacking, and any link may be related to tamoxifen exposure or testing bias. Hysterectomy during risk-reducing bilateral salpingo-oophorectomy is, therefore, of uncertain benefit. Data from a large, prospective cohort will be informative.
Methods: Data on UK *BRCA* pathogenic variant carriers were interrogated for endometrial cancer diagnoses. Standardised incidence ratios (SIRs) were calculated in four distinct cohorts using national endometrial cancer rates; either from 1/1/1980 or age 20, prospectively from date of personal pathogenic variant report, date of family pathogenic variant report or date of risk-reducing salpingo-oophorectomy. Somatic *BRCA* sequencing of 15 serous endometrial cancers was performed to detect pathogenic variants.
Results: Fourteen cases of endometrial cancer were identified in 2609 women (1350 *BRCA1* and 1259 *BRCA2*), of which two were prospectively diagnosed. No significant increase in

* *Corresponding author*: Manchester Centre for Genomic Medicine, St Mary's Hospital, Oxford Road, Manchester, M13 9WL, UK.
E-mail address: Gareth.evans@mft.nhs.uk (D.G. Evans).
[1] Denotes equal contribution.

https://doi.org/10.1016/j.ejca.2020.05.030

either overall or serous-like endometrial cancer risk was identified in any of the cohorts examined (SIR = 1.70, 95% confidence interval = 0.74–3.33; no cases of serous endometrial cancer diagnosed). Results were unaffected by the *BRCA* gene affected, previous breast cancer or tamoxifen use. No *BRCA* pathogenic variants were detected in any of the serous endometrial cancers tested.

Conclusions: Women with a *BRCA* pathogenic variant do not appear to have a significant increased risk of all-type or serous-like endometrial cancer compared with the general population. These data provide some reassurance that hysterectomy is unlikely to be of significant benefit if performed solely as a preventive measure.

1. Introduction

Since the publication of a number of reports describing diagnoses of serous endometrial cancer in *BRCA1* pathogenic variant carriers of Ashkenazi Jewish heritage [1–3], there has been interest in a potential association between the *BRCA* pathogenic variants and an increased risk of endometrial cancer. A number of studies have sought to quantify the level of risk, although with conflicting results, with some finding evidence of an increased risk [4–6], particularly in *BRCA1* carriers, whilst others have found no association [7,8]. Unfortunately, the absence of a suitable control group has prevented the results of these earlier studies being reconciled in a meta-analysis [9]. From a biological perspective, should a causative relationship exist between *BRCA* pathogenic variants and endometrial cancer, it would be anticipated that the increased risk would be restricted to the serous-like histological subtype, including p53 mutant uterine carcinosaromas and mixed epithelial carcinomas [10]. This has, however, not always been observed [11,12]. It has also been postulated that any observed association may be due to the use of tamoxifen for the prevention and treatment of breast cancer rather than a consequence of a *BRCA1* or *BRCA2* pathogenic variant *per se* [13,14]. Whilst several prospective studies have examined the incidence of endometrial cancer after risk-reducing salpingo-oophorectomy (RRSO) in *BRCA1* and *BRCA2* pathogenic variant carriers compared with the general population, they have failed to consider the impact of the procedure on the rate of endometrial cancer within this specific population [12,13]. Debate, therefore, continues within the scientific and medical communities as to whether risk-reducing hysterectomy should be offered to women with *BRCA1* and *BRCA2* pathogenic variants at the time of their RRSO to reduce their subsequent endometrial cancer risk[11,15]. Limitations of the studies performed to date are that they have often been purely retrospective, have included only small numbers of *BRCA* pathogenic variant carriers, particularly those with *BRCA2* pathogenic variants, and have frequently omitted to undertake expert pathological review of the tumour tissue to ensure accurate subtyping. They have also often had short follow-up durations of only 5–6 years, which, when applied to a cohort with a median age of 40–50 years, means that they have limited power to detect endometrial cancer cases which predominately occur in older women. This study, therefore, sought to determine whether *BRCA* pathogenic variants are associated with an increased risk of endometrial cancer compared with the general population using a large, well-described cohort of *BRCA1* and *BRCA2* pathogenic variant carriers with prospective follow-ups. It also aimed to determine whether there was a particular association between *BRCA* pathogenic variants and the serous histological subtype of endometrial cancer and the impact of RRSO on this risk.

2. Materials and methods

A prospectively maintained database of BRCA pathogenic variant carriers at the Manchester Centre for Genomic Medicine was used to identify individuals aged >20 years for analysis. Data were collected on date of birth, personal and family pathogenic variant testing, salpingo-oophorectomy ± hysterectomy, breast, ovarian and endometrial cancer diagnosis, death and date of the last follow-up. Information on tamoxifen use was collected wherever possible. Women were eligible for the study if they had a *BRCA1* or *BRCA2* pathogenic variant identified between 01/01/1991 (the start date of the database) and 31/12/2017 and had not undergone a previous hysterectomy. Pathology reports were collated from affected individuals to determine endometrial cancer subtype, with slide review by an expert gynaecological pathologist (J.B.) where possible and TP53 immunohistochemistry in accordance with previously published protocols [16]. Follow-up data were collected through medical record review and from the National Cancer Registration and Analysis Service, for women enrolled in the Epidemiological study of Familial Breast Cancer (EMBRACE) study, a national cohort study of *BRCA1* and *2* pathogenic variant carriers and non-

affected family members [17]. Women were considered in a number of distinct, but overlapping, cohorts; retrospectively assuming the follow-up started on 1/1/1980 (or age 20 years, whichever occurred later) and prospectively from the date of their family *BRCA* pathogenic variant identification (or age 20 years), from date of their personal *BRCA* pathogenic variant identification (or age 20 years) and from date of RRSO, where applicable. A nested case-control analysis was planned to evaluate the competing effect of RRSO on endometrial cancer incidence; however, no cases of endometrial cancer occurred in women who underwent RRSO. Women were censored at time of hysterectomy, diagnosis of cancer of the ovary, fallopian tube or peritoneum, death, the last follow-up or 31/12/17, whichever occurred first.

To establish whether serous endometrial cancers are associated with pathogenic variants in BRCA, we identified 15 serous endometrial cancers treated at our institution and carried out somatic BRCA sequencing. DNA was extracted from formalin-fixed paraffin embedded blocks of tumour tissue, which had been obtained at the time of hysterectomy. DNA extraction was performed using either COBAS (Cat no: 05985536190, Roche) or EZ1 (Cat no: 953034, Qiagen) extraction kits. The DNA was quantified using Qubit broad range assay and reagents. Sanger DNA sequencing for *BRCA1/2* mutations was undertaken by the Manchester Genomics Diagnostics Laboratory using their in-house developed protocol; details of which have been published elsewhere [18]. In brief, 80 ng of intact DNA was amplified using GeneRead DNAseq Targeted Exon Enrichment Breast Panel (Qiagen). PCR products were purified using Ampure XP beads and quantified on the 2200 TapeStation using D1000 High Sensitivity kit (Agilent). These were then adenylated, and adaptors were ligated using TruSeq PCR-free Library Preparation Kit according to manufacturer's protocol (Illumina). The resulting libraries were cleaned and selected for size using GeneRead (Qiagen) size selection columns, before undergoing quantification using the KAPA Library Quantification Kit (Kapa Biosystems). Each library was normalised to 0.5 nM with EB buffer (Qiagen). Samples were pooled, denatured with 0.2 N NaOH of equal volume, neutralised with 200 mM Tris of equal volume and diluted with HT1 solution (Illumina) to achieve a final 12.5pM library concentration. For sequencing, 594uL of the pooled library mix and 6 pL of 12.5pM PhiX control library were mixed and loaded on to MiSeq V.2 (Illumina). Data were processed as previously described.

Expected endometrial cancer incidence rates were calculated using age-standardised UK-specific data available from the Office for National Statistics [19] in 5 year intervals and were adjusted for local hysterectomy rates, as calculated using data from the Predicting the Risk of Cancer At Screening (PROCAS) study [20]. This was a large risk assessment study conducted in the Greater Manchester area developing breast cancer risk algorithms. The risk of endometrial cancer relative to the general population was evaluated with standardised incidence ratios (SIR), calculated as the observed number of endometrial cancer cases divided by the expected number of cases. Subgroup analyses were performed based on *BRCA1/2* pathogenic variant status, history of breast cancer and tamoxifen use and endometrial cancer histological subtype. Serous-like endometrial cancers included serous endometrial cancer, uterine carcinosarcomas with a serous epithelial component and mixed serous epithelial tumours in keeping with the findings of de Jonge et al [10]. The expected number of serous-like endometrial cancer cases was calculated assuming 10% of all endometrial cancers were of the serous-like histotype [21,22] The Byar's approximation of the exact Poisson distribution was used to calculate the 95% confidence limits using the methodology of Breslow and Day [23]. Statistical analysis was performed using MS Excel (2016).

3. Results

Of 2609 women, 1350 (51.7%) had a *BRCA1* pathogenic variant and 1259 (48.3%) had a *BRCA2* pathogenic variant. The median age at baseline, the last follow-up and the length of the follow-up varied according to the cohort examined (Table 1).

There were 14 cases of endometrial cancer identified; 12 occurred before the confirmation of a personal BRCA pathogenic variant (i.e. were identified retrospectively) and two cases were identified prospectively. The clinical characteristics of the endometrial cancer cases are given in Supplementary Table 1. Pathology review was possible for six of the 14 endometrial cancer cases identified, with TP53 immunohistochemistry performed in three cases to aid diagnosis. Most cases were of endometrioid subtype, with one retrospectively identified case of a mixed serous and endometrioid tumour and two cases of endometrial carcinosarcoma. Only the mixed serous tumour demonstrated diffuse p53 staining, in keeping with a mutant-like pattern. Both prospectively diagnosed endometrial cancer cases were of endometrioid subtype and occurred in index cases. There were two cases of proven synchronous ovarian and endometrial cancers, one identified prospectively and the other retrospectively, and a further suspected case within the retrospective cohort, which could not be confirmed as the original slides were not available for review. There were no cases of endometrial cancer in women who underwent RRSO.

The overall risk of endometrial cancer was not significantly increased in any of the four cohorts studied (from 1/1/1980 adjusted SIR = 1.70, 95% confidence interval [CI] = 0.74–3.33; date of family pathogenic

Table 1
Demographic data and follow-up duration for cohorts examined.

Cohort	Size (n)	Median age at baseline, yrs (IQR)	Median age at last follow-up, yrs (IQR)	Total follow-up, women years at risk (median)	*BRCA* pathogenic variant status (%)	Prior history of breast cancer (%)	Prior history of tamoxifen use (%)
Retrospective cohort (1/1/1980–31/12/17)	2609	20.0 (20.0–31.6)	48.8 (40.5–57.9)	59199 (23.8)	*BRCA1* 1350 (51.7%) *BRCA2* 1259 (48.3%)	Yes 1259 (48.2%) No 1350 (51.7%)	Yes 311 (11.9%) No 557 (21.3%) Unknown 1741 (66.6%)
Date of family pathogenic variant report	1811	44.1 (34.9–54.6)	49.1 (40.2–59.1)	9412 (3.5)	*BRCA1* 906 (50.0%) *BRCA2* 905 (50.0%)	Yes 907 (50.1%) No 901 (49.8%) Unknown 3 (0.2%)	Yes 260 (14.4%) No 490 (27.1%) Unknown 1061 (58.6%)
Date of personal pathogenic variant report	1617	45.1 (37.0–55.1)	49.2 (40.9–58.8)	6375 (2.4)	*BRCA1* 817 (50.5%) *BRCA2* 800 (49.5%)	Yes 847 (52.4%) No 768 (47.5%) Unknown 2 (0.1%)	Yes 258 (16.0%) No 478 (29.6%) Unknown 881 (54.4.%)
Date of RRSO	546	45.8 (40.4–52.6)	51.3 (45.4–58.9)	2865 (2.9)	*BRCA1* 274 (50.2%) *BRCA2* 272 (49.8%)	Yes 283 (51.8%) No 263 (48.2%)	Yes 103 (18.9%) No 178 (32.6%) Unknown 265 (48.5%)

IQR, interquartile range.

variant report adjusted SIR = 0.89, 95% CI = 0.12–3.02; date of personal pathogenic variant report adjusted SIR = 1.21, 95% CI = 0.09–4.48; date of RRSO adjusted SIR incalculable, Table 2).

Subgroup analyses failed to find any difference in endometrial cancer risk between women with a *BRCA1* or *BRCA2* pathogenic variant, a history of breast cancer or tamoxifen use. Neither was there a specific increase in the risk of serous-like endometrial cancer (cohort from 1/1/1980 SIR = 3.66, 95% CI = 0.01–23.41, SIR incalculable in the prospective cohorts as no cases of serous endometrial cancer diagnosed).

Furthermore, we assessed the presence of a pathogenic variant in *BRCA1/2* in first-degree relatives, of a proven carrier, who had developed endometrial cancer without previous breast or synchronous ovarian cancer. Five of seven (71%) did not carry the family variant. If endometrial cancer was associated, then more than 50% should have carried the pathogenic variant.

Of the 15 serous endometrial cancers analysed, none contained *BRCA1/2* pathogenic variants.

4. Discussion

This study did not find a significant increase in the incidence of endometrial cancer in women with a pathogenic variant in either the *BRCA1* or *BRCA2* genes. This finding was unaffected by the *BRCA* gene affected, a personal history of breast cancer or tamoxifen use. The study was unable to address whether RRSO reduces the risk of endometrial cancer specifically in this population, due to the lack of endometrial cancer cases in women who underwent RRSO. No specific association between *BRCA1/2* pathogenic variants and serous endometrial cancer was detected; there was neither an increased risk of serous endometrial cancer in *BRCA1/2* pathogenic variant carriers nor were pathogenic variants detected in the *BRCA1/2* genes within the tumour tissue from 15 unselected serous endometrial cancers.

These reassuring findings are consistent with those of Levine *et al.* [7], who described a relative risk of endometrial cancer of 0.75 (95% CI = 0.24–2.34, p = 0.6) in 199 Ashkenazi Jews with *BRCA1/2*, and of Lee *et al.* [11], who failed to find an increase in serous or endometrioid endometrial cancer in their moderately sized Australasian population (*BRCA1* SIR = 2.87, 95% CI = 0.59–8.43, p = 0.18, *BRCA2* SIR = 2.01, 95% CI = 0.24–7.30, p = 0.52). The largest study to date, conducted across 11 different countries, however, yielded contradictory results, noting a significantly increased risk of endometrial cancer in *BRCA1* pathogenic variant carriers and those exposed to tamoxifen [5]. Of the 4893 women studied, 3536 were *BRCA1* pathogenic variant carriers, explaining why there was no statistically significant increase in endometrial cancer risk in the *BRCA2* group, despite similar SIRs (*BRCA1* SIR = 1.91, 95% CI = 1.06–3.19, p = 0.03, *BRCA2* SIR = 1.75, 95% CI = 0.55–4.23, p = 0.2). The association between tamoxifen use and an increase in endometrial cancer incidence in BRCA pathogenic variant carriers has been confirmed in a subsequent case-control study undertaken by the same group, which found a 6.21-fold increase in risk compared with non-users (95% CI = 2.21–17.5, p = 0.0005), which the authors suggested could provide an explanation for the observed association [14]. These findings were not, however, replicated in the present study. Whilst the same authors also described a lower incidence of endometrial cancer in women who underwent oophorectomy for any reason, this has not been confirmed in other cohorts of women who have undergone specific RRSO, that is, in the

Table 2
Observed and expected endometrial cancer rates in BRCA pathogenic variant carriers.

Year	Expected	Observed	SIR	CI lower 95%	CI upper 95%
1/1/1980–31/12/2017					
1980–1984	0.28	0	0.00	0.00	0
1985–1989	0.4	0	0.00	0.00	0
1990–1994	0.61	1	1.64	0.00	13.11
1995–1999	0.82	4	4.90	0.01	31.45
2000–2004	1.08	3	2.77	0.05	14.56
2005–2009	1.35	4	2.97	0.13	13.6
2010–2014	1.3	1	0.77	0.03	3.59
2015–2017	0.44	1	2.27	0.00	23.4
Total	6.27	14	2.23	0.84	4.78
Adjusted[a]	8.22	14	1.70	0.74	3.33
Serous-like endometrial cancer	0.82	3	3.66	0.01	23.41
BRCA1 only	3.68	7	1.9	0.47	5.05
Date of family pathogenic variant mutation report					
1990–1994	0	0	0	0	0
1995–1999	0.05	1	19.39	0	1420.26
2000–2004	0.2	1	5.01	0	102.28
2005–2009	0.49	0	0	0	0
2010–2014	0.65	0	0	0	0
2015–2017	0.33	0	0	0	0
Total	1.71	2	1.17	0.1	4.6
djusted[a]	2.25	2	0.89	0.12	3.02
Serous-like endometrial cancer	0.23	0	0	0	0
BRCA1 only	0.98	1	1.02	0.01	5.73
Date of personal pathogenic variant mutation report					
1990–1994	0	0	0	0	0
1995–1999	0.03	1	32.35	0	3906.24
2000–2004	0.14	1	7.33	0	212.49
2005–2009	0.37	0	0	0	0
2010–2014	0.5	0	0	0	0
2015–2017	0.22	0	0	0	0
Total	1.26	2	1.59	0.06	7.55
Adjusted[a]	1.65	2	1.21	0.09	4.88
Serous-like endometrial cancer	0.17	0	0	0	0
BRCA1 only	0.73	1	1.36	0	9.46
Date of RRSO					
1990–1994	0.02	0	0	0	0
1995–1999	0.04	0	0	0	0
2000–2004	0.09	0	0	0	0
2005–2009	0.19	0	0	0	0
2010–2014	0.29	0	0	0	0
2015–2017	0.13	0	0	0	0
Total	0.76	0	0	0	0
Adjusted[a]	0.99	0	0	0	0
Serous-like endometrial cancer	0.10	0	0	0	0
BRCA1 only	0.47	0	0	0	0

CI, confidence interval; RRSO, risk-reducing salpingo-oophorectomy; SIR, standardised incidence ratio.
[a] Expected data adjusted for hysterectomy prevalence data.

absence of any tubo-ovarian disease [13]. A beneficial effect of RRSO may have been anticipated if serous endometrial cancers originate in the fallopian tube. The fact that in our cohort only one case of mixed serous and endometrioid endometrial cancer was diagnosed means that we are unable to provide any robust data to confirm or refute this hypothesis, except to state that this case occurred in a woman who had not undergone RRSO and no cases of endometrial cancer were diagnosed in the RRSO cohort.

The number of endometrial cancer cases observed in each of the cohort studies, including our own, has been small (2–17) and could well explain the difference in statistical significance of SIRs that all approximate to a value of 2. Indeed, only two cases of endometrial cancer were diagnosed in our prospective cohorts, which may indicate a testing bias in those with endometrial cancer in previously reported studies. An unbiased assessment of testing first-degree relatives with only endometrial cancer supports our premise that there is unlikely to be any substantial increase in endometrial cancer risk. It is, however, arguable that even if an SIR of 2 is validated (none of our upper confidence limits exclude this), the level of risk is insufficient to recommend hysterectomy at the time of bilateral salpingo-oophorectomy as a risk-reducing measure. Given the increased potential morbidity associated with more extensive surgery, evidence of benefit is certainly warranted to outweigh these additional risks. Whether there is a clear benefit in specific subgroups of *BRCA* pathogenic variant carriers is currently unknown; neither ours nor previously published studies have included data on body mass index (BMI) and hence have been unable to adjust for this in analyses.

The strengths of this study include the confirmation of endometrial cancer diagnoses with histological reports and contemporaneous expert pathological review of slides, although unfortunately this was not universally achievable due to the lack of availability of tumour tissue for assessment. The study also included the largest number of *BRCA2* pathogenic variant carriers to date, increasing our understanding of endometrial cancer risk in this specific population. As with previous studies, robust methodology has been employed to compare observed with national expected endometrial cancer rates, with adjustment made for local hysterectomy rates. We were able to include data on *BRCA1/2* pathogenic variant status of first-degree relatives of women who developed endometrial cancer and of unselected serous endometrial cancer cases to corroborate our findings. Our somatic BRCA testing has been shown to have high sensitivity in high-grade serous ovarian cancer [25].

The potential lack of power in our study is a limitation and one that we have attempted to address by contacting the EMBRACE study (Easton D) to ensure endometrial cancer cases have not been missed. It does

mean that our ability to detect differences in endometrial cancer risk in any specific subgroups or subtype of endometrial cancer is limited. Whilst the length of the follow-up, particularly for the retrospective cohort, is a clear advantage of this work, the median age at censoring remains younger than 50 years, well below the average age of endometrial cancer diagnoses in the UK [24]. Re-analysis of the data at a later date will be performed to increase the duration of the follow-up and potentially the number of endometrial cancer diagnoses. Subgroup analyses based upon tamoxifen use was limited due to the fact that two thirds of women in the database did not have data collected on their exposure to the drug, although the vast majority of women with a pathogenic variant in a *BRCA* gene without a history of breast cancer were not known to have taken tamoxifen. The low prevalence of tamoxifen use within the cohort may well explain why no association was observed between tamoxifen use and an increase in endometrial cancer risk. Additional efforts to reduce the amount of missing data within our data set will be made to address this. Data were unfortunately not routinely collected on hormone replacement therapy (HRT) use by women who underwent RRSO and so the impact of this on subsequent endometrial cancer risk could not be assessed. Whilst oestrogen-only HRT may be associated with a lower rate of subsequent breast cancer [25,26], this must be balanced against the impact this could have on the risk of malignant changes within the endometrium. Whilst every attempt was made to undertake histological review of all endometrial cancer cases occurring within the cohort, this was unfortunately not possible for eight cases where slides were unavailable. Four of these cases were also operated on at another hospital and, as it was not possible to retrieve the pathology reports for these tumours, this may impact upon the results of our subgroup analysis of serous-like endometrial cancers.

5. Conclusion

In conclusion, *BRCA1* and *BRCA2* pathogenic variant carriers do not appear to be at a significant increased risk of endometrial cancer compared with the general population. Neither does there appear to be a specific association between *BRCA1/2* pathogenic variants and serous endometrial cancer. Women and clinicians should be reassured that hysterectomy at the time of bilateral salpingo-oophorectomy is unlikely to be of benefit if performed solely for the purpose of trying to reduce subsequent endometrial cancer risk.

Ethics approval and consent for publication

All women consented to the inclusion of their personal information in a prospective database held at the Manchester Centre for Genomic Medicine (The study was approved by the Central Manchester Research Ethics Committee (10/H1008/24 and 11/H1003/3), the review of their clinical records and cancer registry data for information on subsequent cancer diagnoses and the publication of any studies undertaken using their anonymised data. The endometrial tumour tissue used for *BRCA1/2* sequencing was donated by women enrolled into other, unrelated, studies at our institute [27] (Proportion of Endometrial Tumours Associated Lynch Syndrome (PETALS) study North West Research Ethics Committee reference 15/NW/0733, Cancer Research UK clinical trial database, ref-13595). The study was performed in accordance with the Declaration of Helsinki.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Funding

S.J.K. reports being a National Institute for Health Research (NIHR) Academic Clinical Lecturer. N.A.J.R. reports being a Medical Research Council Doctoral Research Fellow (MR/M018431/1). E.J.C., E.R.W., C.B. and D.G.E. are supported through the NIHR Manchester Biomedical Research Centre (IS-BRC1215-20007), and D.G.E. reports being an NIHR Senior Investigator (NF–SI-0513-10076).

Role of the funding source

The funding source had no role in the design of this study, its execution, analysis, interpretation of the results or decision to submit the results.

Author contributions

E.J.C. and D.G.E. contributed in designing the study and supervising its execution. F.L., E.R.W., R.D.C., R.J.E. and D.G.E. contributed in data collection. S.J.K. performed the analyses and drafted the manuscript. C.B. and N.A.J.R. contributed to data collection and analysis. J.B. provided pathological review of tumour slides. All authors contributed to the final manuscript.

Conflict of interest statement

The authors report no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2020.05.030.

References

[1] Hornreich G, Beller U, Lavie O, Renbaum P, Cohen Y, Levy-Lahad E. Is uterine serous papillary carcinoma a BRCA1-related disease? Case report and review of the literature. Gynecol Oncol 1999;75:300–4.

[2] Lavie O, Hornreich G, Ben Arie A, Renbaum P, Levy-Lahad E, Beller U. BRCA1 germline mutations in women with uterine serous papillary carcinoma. Obstet Gynecol 2000;96:28–32.

[3] Kaplan EJ, Caputo TA, Shen PU, Sassoon RI, Soslow RA. Familial papillary serous carcinoma of the cervix, peritoneum, and ovary: a report of the first case. Gynecol Oncol 1998;70:289–94.

[4] Thompson D, Easton DF, Breast Cancer Linkage C. Cancer Incidence in BRCA1 mutation carriers. J Natl Canc Inst 2002;94:1358–65.

[5] Segev Y, Iqbal J, Lubinski J, Gronwald J, Lynch HT, Moller P, et al. The incidence of endometrial cancer in women with BRCA1 and BRCA2 mutations: an international prospective cohort study. Gynecol Oncol 2013;130:127–31.

[6] Lavie O, Ben-Arie A, Segev Y, Faro J, Barak F, Haya N, et al. BRCA germline mutations in women with uterine serous carcinoma–still a debate. Int J Gynecol Canc Off J Int Gynecolog Canc Soc 2010;20:1531–4.

[7] Levine DA, Lin O, Barakat RR, Robson ME, McDermott D, Cohen L, et al. Risk of endometrial carcinoma associated with BRCA mutation. Gynecol Oncol 2001;80:395–8.

[8] Breast Cancer Linkage C. Cancer risks in BRCA2 mutation carriers. J Natl Canc Inst 1999;91:1310–6.

[9] de Jonge MM, Mooyaart AL, Vreeswijk MP, de Kroon CD, van Wezel T, van Asperen CJ, et al. Linking uterine serous carcinoma to BRCA1/2-associated cancer syndrome: a meta-analysis and case report. Eur J Canc 2017;72:215–25.

[10] de Jonge MM, Ritterhouse LL, de Kroon CD, Vreeswijk MPG, Segal JP, Puranik R, et al. Germline BRCA-associated endometrial carcinoma is a distinct clinicopathologic entity. Clin Canc Res Off J Am Assoc Canc Res 2019;25(24):7517–26.

[11] Lee YC, Milne RL, Lheureux S, Friedlander M, McLachlan SA, Martin KL, et al. Risk of uterine cancer for BRCA1 and BRCA2 mutation carriers. Eur J Canc 2017;84:114–20.

[12] Reitsma W, Mourits MJ, de Bock GH, Hollema H. Endometrium is not the primary site of origin of pelvic high-grade serous carcinoma in BRCA1 or BRCA2 mutation carriers. Mod Pathol Off J U. S. Canad Acad Pathol Inc 2013;26:572–8.

[13] Shu CA, Pike MC, Jotwani AR, Friebel TM, Soslow RA, Levine DA, et al. Uterine cancer after risk-reducing salpingo-oophorectomy without hysterectomy in women with BRCA mutations. JAMA Oncol 2016;2:1434–40.

[14] Segev Y, Rosen B, Lubinski J, Gronwald J, Lynch HT, Moller P, et al. Risk factors for endometrial cancer among women with a BRCA1 or BRCA2 mutation: a case control study. Fam Cancer 2015;14:383–91.

[15] Saule C, Mouret-Fourme E, Briaux A, Becette V, Rouzier R, Houdayer C, et al. Risk of serous endometrial carcinoma in women with pathogenic BRCA1/2 variant after risk-reducing salpingo-oophorectomy. J Natl Canc Inst 2018:110.

[16] Kitson S, Maskell Z, Sivalingam VN, Allen JL, Ali S, Burns S, et al. PRE-surgical Metformin in Uterine Malignancy (PREMIUM): a multi-center, randomized double-blind, placebo-controlled phase 3 trial. Clin Canc Res Off J Am Assoc Canc Res 2018;25(8):2424–32.

[17] Centre for Cancer Genetic Epidemiology UoC. EMBRACE study. 2017.

[18] Ellison G, Huang S, Carr H, Wallace A, Ahdesmaki M, Bhaskar S, et al. A reliable method for the detection of BRCA1 and BRCA2 mutations in fixed tumour tissue utilising multiplex PCR-based targeted next generation sequencing. BMC Clin Pathol 2015;15:5.

[19] Cancer Research UK. Uterine cancer incidence statistics. 2019.

[20] Evans DGR, Harkness EF, Brentnall AR, van Veen EM, Astley SM, Byers H, et al. Breast cancer pathology and stage are better predicted by risk stratification models that include mammographic density and common genetic variants. Breast Canc Res Treat 2019;176:141–8.

[21] Sagae S, Susumu N, Viswanathan AN, Aoki D, Backes FJ, Provencher DM, et al. Gynecologic Cancer InterGroup (GCIG) consensus review for uterine serous carcinoma. Int J Gynecol Canc Off J Int Gynecolog Canc Soc 2014;24:S83–9.

[22] Hamilton CA, Cheung MK, Osann K, Chen L, Teng NN, Longacre TA, et al. Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. Br J Canc 2006;94:642–6.

[23] Breslow NE, Day NE. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. IARC Sci Publ 1987:1–406.

[24] Cancer Research UK. Uterine cancer statistics. 2016.

[25] Kotsopoulos J, Gronwald J, Karlan BY, Huzarski T, Tung N, Moller P, et al. Hormone replacement therapy after oophorectomy and breast cancer risk among BRCA1 mutation carriers. JAMA Oncol 2018;4:1059–65.

[26] Marchetti C, De Felice F, Boccia S, Sassu C, Di Donato V, Perniola G, et al. Hormone replacement therapy after prophylactic risk-reducing salpingo-oophorectomy and breast cancer risk in BRCA1 and BRCA2 mutation carriers: a meta-analysis. Crit Rev Oncol-Hematol 2018;132:111–5.

[27] Sivalingam VN, Kitson S, McVey R, Roberts C, Pemberton P, Gilmour K, et al. Measuring the biological effect of presurgical metformin treatment in endometrial cancer. Br J Canc 2016;114(3):281–9.