

Genome-wide association studies of endometrial cancer: Latest developments and future directions.

Tracy A. O'Mara¹, Dylan M. Glubb¹, Pik Fang Kho¹, Deborah J. Thompson², Amanda B. Spurdle¹

1. Molecular Cancer Epidemiology Group, Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia
2. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Running Title: Genome-wide association studies of endometrial cancer

Keywords: Endometrial cancer; genome-wide association study; post-GWAS; genetics; Mendelian randomisation

Financial Support: TAO'M is supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship (APP1111246), PFK is supported by an Australian Government Research Training Program PhD Scholarship and QIMR Berghofer Postgraduate Top-Up Scholarship, ABS is supported by an NHMRC Senior Research Fellowship (APP1061779).

Corresponding Author:

Tracy O'Mara (tracy.omara@qimrberghofer.edu.au)

Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Brisbane, Queensland 4006, Australia. Phone: 61 7 3362 0389

Conflict of Interest Statement: The authors declare no potential conflicts of interest.

Word Count: 3,978

Number Figures: 2

Number Tables: 2

Abstract

Endometrial cancer, the most commonly diagnosed cancer of the female reproductive tract in developed countries, has a heritable component. To date, sixteen genetic risk regions have been robustly discovered by genome-wide association studies (GWAS) of endometrial cancer. Post-GWAS analyses including expression quantitative trait loci (eQTL) analysis and laboratory-based functional studies have been successful in identifying genes and pathways involved in endometrial carcinogenesis. Mendelian randomization analysis studies have confirmed factors causal for endometrial cancer risk, including increased body mass index and early onset of menarche. In this review, we summarize findings from GWAS and post-GWAS analyses of endometrial cancer. We discuss clinical implications of these findings, current knowledge gaps and future directions for the study of endometrial cancer genetics.

Introduction

Endometrial cancer, a malignancy of the lining of the uterus, is the most common gynaecological cancer diagnosed in developed countries, accounting for ~5% of all female cancers (<http://gco.iarc.fr/today/home>); age-standardized rates are steadily increasing (1). Endometrial carcinoma has traditionally been classified into two groups defined by histological subtype: endometrioid and non-endometrioid. Endometrioid endometrial cancers, which comprise the majority of endometrial cancers (80%), develop from glandular cells in the lining of endometrium and are generally associated with hyperplasia. These tumours are estrogen-dependent and tend to be low-grade with a favourable overall prognosis (2). Conversely, non-endometrioid endometrial cancers (commonly serous papillary or clear cell histology) are non-estrogen dependent and often exhibit a more aggressive clinical course that has a very poor prognosis (2).

Endometrial cancer treatment has remained almost static over the last four decades relying principally on surgery, with full hysterectomy the most common and effective treatment for early stage disease (reviewed in 3). However, the prognosis for advanced, recurrent or metastatic stage is still poor with high rates of recurrence and lower 3-year survival rates: endometrial cancer patients with local (vaginal) metastasis have a 3-year survival of 73%, pelvic metastasis 8% survival, and distant metastasis, 15% survival (4). Furthermore, the median survival time for advanced stage endometrial cancer is generally less than twelve months (5, 6).

Women with a family history of endometrial cancer have an approximately 2-fold increased risk of developing the disease (7, 8). While some of the association between family history and endometrial cancer risk is attributable to shared environmental and/or lifestyle risk factors, twin studies have estimated the heritability to be between 27-52% (9-12).

The currently known genetic architecture of endometrial cancer is displayed in **Figure 1** and is consistent with a polygenic model of inheritance. Rare germline pathogenic (i.e. high-risk) variants in cancer syndrome genes, i.e. DNA mismatch repair genes associated with Lynch Syndrome (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) or *PTEN* in the context of Cowden's disease explain ~3-5% of endometrial cancer cases at the population level (reviewed by 13). Evidence also supports the contribution of rare pathogenic variants in other DNA repair-

related genes to endometrial cancer risk, including *POLD1*, *POLE* and possibly *BRCA1* (reviewed by 13). Pathogenic variants in other genes are likely to exist, but we would expect the frequency of such variants to be low, and thus they could account for only a small proportion of endometrial cancer cases. In contrast, although each endometrial cancer predisposition variant (such as the sixteen risk variants identified to date in genome-wide association studies (GWAS)(14-17)) has only a modest effect on risk, together they are likely to explain far more of the familial relative risk of the disease.

This review will summarise the findings of recent endometrial cancer GWAS, and review post-GWAS analyses, including Mendelian randomization studies and the functional follow-up of endometrial cancer genetic risk regions to identify candidate target genes.

Genome-wide association studies of endometrial cancer risk

Since the late 2000s, genotyping arrays, consisting of hundreds of thousands of common genetic variants across the genome, have revolutionized the study of the genetic basis of complex traits. For many diseases, GWAS have been remarkably successful in unlocking the biology of disease and driving therapeutic development (18); early indications suggest that this will also be true for endometrial cancer.

In recent years, a series of GWAS in European populations have identified 16 endometrial cancer genetic risk regions (**Table 1**). The first endometrial cancer GWAS, reported in 2011, identified a novel genetic risk region at 17q12, intronic to *HNF1B* (19). This study involved a stage 1 GWAS of 1,265 endometrioid endometrial cancer cases and 5,190 controls, followed by stage 2 validation of 47 genetic variants in 3,957 cases and 6,886 controls. This finding was directionally concordant in a subsequent GWAS by the Epidemiology of Endometrial Cancer Consortium (E2C2), including 4,989 cases across two stages (20). A meta-analysis of these two GWAS studies, with a third GWAS from the National Study of Endometrial Cancer Genetics, totalling 4,907 cases, identified a new, intergenic, risk region at chromosome 6p22.3 (14).

The identification of additional endometrial cancer genetic risk regions has required collaboration and data sharing in order to achieve the sample sizes (and hence statistical power) required for the identification of variants with modest effect sizes. To this end, the

Endometrial Cancer Association Consortium (ECAC) was established to conduct large-scale meta-analyses of GWAS data. ECAC is an ongoing collaboration which currently includes twelve research groups, based in Europe, Australia and the USA, providing access to genetic and other information for more than 9,000 endometrial cancer cases of European ancestry.

Around the time that ECAC was established, a custom genotyping array (“iCOGS”) was designed to test genetic associations with the risk of hormone-related cancers. The content of the iCOGS ~200K SNP genotyping array (21) is enriched for genetic variants with some prior evidence of association with one or more hormonal cancer, including variants selected on the basis of the first endometrial cancer GWAS. The genotyping of 4,402 ECAC cases on the iCOGS genotyping array, combined with previous GWAS data, led to the identification of five novel endometrial cancer risk regions (Table 1) (15) and allowed for fine-mapping studies of the known signals at *HNF1B* (16) and *CYP19A1* (17) risk regions (the later had been previously identified in several candidate gene studies e.g. (22)).

Recently, ECAC genotyped an additional 2,689 endometrial cancer cases using the OncoArray array (23). Along with data from the iCOGS projects, E2C2, the first release of the UK Biobank and a new GWAS conducted by the Women’s Health Initiative, this resulted in the largest endometrial cancer meta-GWAS to date (12,906 cases), enabling us to identify a further nine genetic regions associated with endometrial cancer risk (24). One previously identified region, near *AKT1* on 14q32 (15), was not replicated by this analysis at a genome-wide level of significance, bringing the total number of established genetic risk regions for endometrial cancer to sixteen.

We have estimated that common genetic variants of the type that can be tagged by standard GWAS arrays potentially account for ~28% of the familial relative risk of endometrial cancer (24), and that the sixteen risk variants identified to date account for approximately one quarter of this figure, suggesting that many more genetic risk variants remain to be found.

One limitation of the endometrial cancer GWAS conducted to date is that they have been almost exclusively restricted to European-ancestry populations. An early GWAS of 832 Chinese endometrial cancer cases did not find any associations at the GWAS significance threshold, likely due to its small sample size and the necessity of using mostly European-ancestry cases in the replication stages (25) because of a lack of genotyped East-Asian

ancestry cases. The expansion of adequately powered endometrial cancer GWAS to wider populations is therefore a priority, as indeed it is for all types of cancer, in order to identify the risk variants which are most relevant to women with different ethnic backgrounds (26).

A second limitation is that the very small numbers of genotyped endometrial cancer cases with non-endometrioid histologies (e.g. only 434 of the 12,906 in the recent ECAC meta-analysis were of serous/mixed-serous histology, the most common non-endometrioid histology) have precluded meaningful subtype-specific analyses of the type which have proved fruitful in the study of ovarian cancer susceptibility (27). While there is currently no evidence for a difference in genetic architecture between endometrial cancer subtypes, the limited data currently available for non-endometrioid histologies do not allow for well-powered analyses of these subtypes. Thus, GWAS or sequencing studies using additional cases of rarer endometrial cancer subtypes are needed to increase statistical power, especially if subtypes can be analysed separately to provide cleaner histological phenotyping.

Mendelian randomization studies of endometrial cancer risk factors

Aside from genetic variants, numerous other factors have been reported as being associated with endometrial cancer risk, but observational studies alone are not always able to distinguish true, causal associations from artefactual associations caused by confounding or reverse causality. Mendelian randomization uses genetic variants known to be associated with a putative risk factor as ‘instruments’ in an instrumental variable analysis, thus testing the association of the risk factor in the absence of confounding (28, 29). The growing use of Mendelian randomization methods to examine risk factors for endometrial cancer (Table 2) has been facilitated by the success of GWAS in identifying the genetic variants associated with many of these proposed risk factors for endometrial cancer.

Obesity is the strongest risk factor for endometrial cancer, with observational studies observing up to eight-fold increased risk between obese women (body mass index; BMI ≥ 40 kg/m²) compared with lean women (BMI < 25 kg/m²). Mendelian randomization analysis has confirmed this relationship, finding strong evidence for a relationship between obesity, as measured by BMI, but not as measured by waist-hip ratio, and endometrial cancer risk (30). An earlier Mendelian randomization study of diabetic-related traits found a significant relationship between increased insulin levels and endometrial cancer risk (31), but did not

find type 2 diabetes or glucose levels to be associated with endometrial cancer, suggesting that observed associations between type 2 diabetes and endometrial cancer may be consequences of residual confounding (31).

Excessive endogenous and exogenous estrogen exposure, unopposed by progesterone, is a well-established risk factor for the development and progression of endometrial cancer (32). A Mendelian randomization analysis using the genetic variant most strongly associated with serum estradiol levels in postmenopausal women verified the relationship between postmenopausal estrogen levels and endometrial cancer (17). Further, each year of delay in menarche (which would be expected to reduce lifetime estrogen exposure), has been confirmed by Mendelian randomization as producing a ~12% reduction in endometrial cancer risk, even after adjusting for the effects of genetically-predicted BMI (24, 33).

The status of other hypothesized risk factors for endometrial cancer, including polycystic ovary syndrome (PCOS), endometriosis and uterine fibroids, remain unclear. The epidemiological associations reported for these risk factors (or lack thereof) may be confounded by co-existing conditions (e.g. PCOS with infertility/anovulation) or measurement limitations (e.g. under- and mis-diagnosis of endometriosis). Future Mendelian randomization studies to investigate these epidemiological risk factors, in conjunction with results from observational studies, will provide important information that can be used clinically to identify women at risk of endometrial cancer, and to inform prevention strategies.

Pleiotropy and cross-disease GWAS studies

One of the conclusions that can be drawn from the ever-growing catalogue of complex-trait GWAS results is that pleiotropy is very widespread (reviewed by 18). Endometrial cancer is no exception to this pattern; from the sixteen endometrial cancer genetic risk regions identified to date, fourteen are coincident with risk regions for other cancers (within 1Mb), twelve regions are associated with anthropometric traits, one region with endometriosis and six with traits associated with steroid hormone levels (as at May 2018; Table 1). While individual risk variants at these regions are not commonly shared between endometrial cancer and these traits, it is anticipated that the different trait-associated variants regulate the same target genes (see functional follow-up section below).

Some of the pleiotropy observed with endometrial cancer has been supported by LD Score regression analyses (34). Endometrial cancer GWAS summary statistics and GWAS data publicly available for 224 non-cancer traits have found several BMI-related traits to be significantly genetically correlated with endometrial cancer risk (24). A similar study performed using endometriosis GWAS summary statistics revealed a significant correlation between this disease and endometrial cancer risk ($r_g = 0.23$; $P=9.3\times 10^{-3}$)(35). LD Score regression analyses between endometrial cancer and other cancer types are in progress, and will likely yield intriguing results.

Given the widespread pleiotropy observed across the genome, it is not surprising that pleiotropic cross-disease GWAS meta-analysis has been used successfully to increase power and identify regions relevant to multiple diseases (36-38). A cross-cancer GWAS study of endometrial cancer with colorectal cancer identified a risk region at 12q24.12, where the most significant association is with a missense variant located in the *SH2B3* gene (39). Subsequent analyses in larger cohorts have found this variant to be independently associated with the risk of both of these cancers (24, 40). Future large-scale cross-cancer GWAS meta-analyses are planned to identify genetic risk regions important for carcinogenesis across multiple tissues. Cross-disease meta-analysis of GWAS data from endometrial cancer and endometriosis identified a risk region relevant to both diseases at 9p23 within the *PTPRD* gene (35). Further, a sub-genome-wide significant region in this cross-disease analysis (35) (12p12.1 locus rs2278868 $P = 5.5\times 10^{-6}$) was subsequently identified as a risk region for endometrial cancer in a larger cohort (24). Meta-analyses of GWAS data from other relevant traits or diseases (e.g. uterine fibroids) could also provide insights into pathways relevant for endometrial cancer aetiology.

Enrichment of endometrial cancer risk variants in functional elements

The correlation structure of common genetic variants means that the genetic variant with the most statistically significant association with disease at a particular locus in a GWAS is not necessarily causally associated with the disease – the apparent association may well be driven by a different causal variant(s) correlated through linkage disequilibrium. It is therefore conventional to think of the most significant variant at a locus as merely the ‘lead’ or ‘index’

variant for a wider set of correlated variants, any of which is a credible causal variant. Definitions vary, but these sets of credible causal variants are usually delineated according to the extent of the linkage disequilibrium with the lead variant, and the difference between the level of statistical significance of the lead variant and that of the candidate variants (41).

Of the credible causal variants identified from the endometrial cancer GWAS risk studies, only three identified to date are exonic. This observation is consistent with the distributions of the GWAS-identified variants for most other complex diseases (reviewed by 18). The sets of credible causal risk variants identified in GWASs of other diseases are enriched for localisation to active epigenetic marks, characteristic of regulatory elements, mapped from trait-relevant cells or tissues (42, 43). Correspondingly, credible causal variants from the most recently identified nine endometrial cancer genetic risk regions demonstrated a greater enrichment in active epigenetic marks from endometrial cancer cell lines and tissues, compared to credible causal variants for related (i.e. endometriosis) or unrelated (i.e. schizophrenia) diseases (24). Also, significantly more credible causal endometrial cancer risk variants localised to active epigenetic marks from estrogen-stimulated endometrial cancer cells in comparison to such marks from unstimulated cells (24). These findings thus support the use of these epigenetic marks in identifying functional (i.e. likely causal) variants at endometrial cancer risk regions and further highlight the role of estrogen in endometrial cancer development.

Functional follow-up studies of endometrial cancer risk GWAS

The identification of the target genes that mediate the effects of GWAS variants is an important step for clinical translation of findings e.g. the identification of opportunities for drug repositioning (44). Experimental and/or bioinformatic studies are required for target gene identification but are very often neglected (45). Several such functional studies have been performed to date for endometrial cancer risk GWAS and are discussed below.

The approaches taken in functional follow-up studies of GWAS are, in part, determined by the locations of the credible causal variants. For example, at the 17q12 endometrial cancer risk region, three candidate causal variants are located in an extended region of the *HNF1B* promoter and so reporter gene assays were performed to assess effects on promoter activity (16). In endometrial cancer cells, the risk alleles of two of the 17q12 credible causal variants

were associated with enhanced *HNF1B* promoter activity, and the effect of one of these variants (rs11263763) was supported by an association between the risk allele of rs11263763 and increased *HNF1B* expression in endometrial tumours (16). It is intriguing to note that rare *HNF1B* variants, reported to abrogate *HNF1B* expression, reduce secretion of insulin which Mendelian randomisation has shown to be an endometrial cancer risk factor (31). This finding is consistent with the upregulation of *HNF1B* promoter activity and transcription by the risk allele of rs11263763, potentially providing a mechanism for the effect of the endometrial cancer risk variation at this region through increased insulin secretion.

Due to haplotype structure, GWAS association signals often map to large genomic intervals, owing to extensive linkage disequilibrium (LD) between common variants, making identification of likely causal variants and their target genes extremely challenging. Moreover, it is not obvious which genes may be targeted for regulation; especially as functional elements such as enhancers can regulate genes through long-range chromatin looping interactions, up to two megabases away (reviewed in 46); **Figure 2**). Indeed, in general, GWAS variants regulate the nearest gene only one third of the time (47, 48) and likely target multiple genes through long-range chromatin looping (49). To address these issues, bioinformatic approaches that use correlations between gene expression and epigenomic features to identify enhancers and their corresponding target genes have been applied to endometrial cancer risk loci and revealed a number of candidate target genes (**Table 1**) (15). However, these candidate target genes still require validation by other means, as described below.

Long-range chromatin looping can be assessed experimentally to identify genes that may be targeted by credible causal variants through regulatory features such as enhancers. These approaches centre on the chromatin conformation capture (3C) technique which identifies interacting genomic regions (reviewed in 50). A 3C method was used at three endometrial cancer risk loci to identify looping between regions containing credible causal variants and the promoters of *MYC* (8q24.21 risk locus (51)), *KLF5* (13q22.1 risk locus (15)), and *AKT1* and *ZBTB42* (the non-replicated 14q32 risk locus (52)). Notably, the interactions between credible causal variants and these four genes had been bioinformatically predicted (**Table 1**) (15).

Looping interactions alone do not provide evidence of gene regulation. Therefore, reporter gene assays in endometrial cancer cells have been used to assess the effects of looping credible causal variants on promoter activity at endometrial cancer risk loci. The risk allele of a looping variant (rs9600103) at the 13q22.1 risk locus increased the activity of a minimal promoter (15). At the 14q32 risk locus, an allele of a looping variant (rs2494737) enhanced the activity of a canonical and alternative promoter of *AKT1*, but had no effect on *ZBTB42* promoter activity (52). No study has yet looked at the effect of looping candidate causal risk variants on *MYC* promoter activity at the 8q24 risk region.

Evidence from expression quantitative trait loci (eQTL) analyses has demonstrated that GWAS variants are enriched for variants that associate with gene expression (53, 54), indicating that these variants likely affect gene regulation and providing an approach to identify target genes at GWAS loci (54, 55). In the most recent endometrial cancer risk GWAS (24), eQTL analyses were performed using data from a variety of tissue sources to identify genes whose expression associated with risk variants from the newly identified loci (**Table 1**). Several of the identified eQTL genes were either tumour suppressors (*NF1* (56)), negative regulators of oncoproteins (*SH2B3* (57)) or oncogenes (*CDCA8* (58) and *WT1-AS* (59)). Consistent with these functions, risk alleles were associated with decreased expression of *NF1* and *SH2B3* and increased expression of *CDCA8* and *WT1-AS* (24).

Network analysis of candidate target genes of endometrial cancer risk variants

Analysis of the 25 candidate target genes identified to date (**Table 4**) has revealed a network which contains 18 of these genes (24). Major network hubs were either known oncoproteins or tumour suppressors, including the protein encoded by the candidate target gene *MYC* and proteins encoded by genes that are somatically mutated in endometrial cancer (*CCND1*, *CTNNB1* and *TP53* (60)). An enrichment of these and other network genes was observed in corresponding pathways such as cyclins and cell cycle regulation, Wnt/ β -catenin and P53 signalling. Interestingly, given the role of obesity in increasing endometrial cancer risk, there was also an enrichment of network genes in an adipogenesis pathway. As further candidate target genes are identified, this network will likely be refined and additional networks revealed. Furthermore, network genes may point to additional endometrial cancer genetic risk regions yet to be identified by GWAS.

Future functional follow-up approaches for endometrial cancer risk GWAS

Although candidate target genes have been identified at 11 of the 17 risk regions (**Table 1**), it is clear that other approaches are needed to identify further candidate target genes. The transcriptome-wide association study method (47) enables gene expression to be predicted in endometrial cancer GWAS datasets (from existing studies of tissue with genotype and gene expression data e.g. the Genotype Tissue Expression project (61) and The Cancer Genome Atlas (62), thus allowing the testing of associations between imputed gene expression and endometrial cancer risk. Additionally, a number of global 3C techniques are now available (50)) to systematically assess chromatin interactions across the entire genome, rather than being restricted to single loci, and could be used to assess all endometrial cancer genetic risk regions for candidate target genes. However, these approaches still do not provide definitive evidence that credible causal variants directly affect gene expression and require additional studies to determine variant function. The final steps to validate the functionality of credible causal variants and their target genes and determine their effects on cellular phenotypes will require systems such as CRISPR/Cas9 which could be used to generate isogenic cell lines that differ by the alleles of credible causal variants (reviewed in 63) or activate/inactivate chromatin encompassing these variants (reviewed in 64).

Conclusions and Future Directions

Plans by ECAC to conduct a significantly expanded meta-GWAS, including imputation to more dense reference panels, are likely to yield additional endometrial cancer risk loci. Sample sizes will also be increased by conducting cross-disease meta-analysis studies of endometrial cancer and related diseases: an additional benefit of such studies is the potential to identify novel pleiotropic risk loci, and hence insights into shared underlying biology. Looking further ahead, large-scale sequencing initiatives are required for the identification of rarer variants (i.e. not tagged by standard GWAS-type arrays) associated with modest to high risks of endometrial cancer.

The next phase of research will be to progress identified endometrial cancer risk variants to translational and clinical outcomes for patients. Results from Mendelian randomisation

analyses determining causality of epidemiological risk factors have potential to identify at-risk women for altered screening management. Another avenue for clinical translation will be the development of polygenic risk scores for endometrial cancer. Application of this approach in *BRCA1/2* pathogenic variant carriers is able to predict which individuals are most likely to develop breast or ovarian cancer (65). Similarly, a polygenic risk score could be developed to predict which women from Lynch Syndrome families are most likely to be diagnosed with endometrial cancer, and who would derive the most benefit from increased screening and/or risk-reducing steps. The development of polygenic risk scores reinforces the importance of performing GWAS in non-European populations. Given the differences in allele frequencies, effect sizes and LD patterns across ethnic groups, polygenic risk scores developed using primarily European sample sets may not be translatable to other populations. Therefore expansion of endometrial cancer GWAS to non-European populations is an essential research priority so that comparable polygenic risk scores will be available for all ethnic groups.

A key ongoing challenge is the identification of the causal genes that mediate the effects of endometrial cancer risk variants. Although post-GWAS studies of endometrial cancer risk have identified candidate target genes at individual risk regions, endometrial cancer functional genomic data needs to be generated in order to systematically assess all risk regions. Furthermore, it has not yet been experimentally shown that the regulation of candidate target genes contributes to endometrial cancer risk through effects on cellular phenotypes. Importantly, such experimental studies are necessary to spur the development of new therapies through the identification of new drug targets, or targets for which drugs already exist. Indeed, the use of a drug target with a genetic basis appears to improve the likelihood of successful drug development (66), highlighting the inherent potential for clinical translation from studies identifying the likely causal genes underlying endometrial cancer risk loci.

References

1. Lortet-Tieulent J, Ferlay J, Bray F, Jemal A. International Patterns and Trends in Endometrial Cancer Incidence, 1978–2013. *JNCI: Journal of the National Cancer Institute*. 2017:djx214-djx.
2. Webb PM. Environmental (nongenetic) factors in gynecological cancers: update and future perspectives. *Future Oncol*. 2015;11:295-307.
3. Lee WL, Lee FK, Su WH, Tsui KH, Kuo CD, Hsieh SL, et al. Hormone therapy for younger patients with endometrial cancer. *Taiwan J Obstet Gynecol*. 2012;51:495-505.

4. Creutzberg CL, van Putten WL, Koper PC, Lybeert ML, Jobsen JJ, Warlam-Rodenhuis CC, et al. Survival after relapse in patients with endometrial cancer: results from a randomized trial. *Gynecol Oncol.* 2003;89:201-9.
5. Rose PG, Brunetto VL, VanLe L, Bell J, Walker JL, Lee RB. A phase II trial of anastrozole in advanced recurrent or persistent endometrial carcinoma: a Gynecologic Oncology Group study. *Gynecol Oncol.* 2000;78:212-6.
6. Ray-Coquard I, Favier L, Weber B, Roemer-Becuwe C, Bougnoux P, Fabbro M, et al. Everolimus as second- or third-line treatment of advanced endometrial cancer: ENDORAD, a phase II trial of GINECO. *Br J Cancer.* 2013;108:1771-7.
7. Johnatty SE, Tan YY, Buchanan DD, Bowman M, Walters RJ, Obermair A, et al. Family history of cancer predicts endometrial cancer risk independently of Lynch Syndrome: Implications for genetic counselling. *Gynecol Oncol.* 2017;147:381-7.
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. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. 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.
10. Schildkraut JM, Risch N, Thompson WD. Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast/ovarian cancer relationship. *Am J Hum Genet.* 1989;45:521-9.
11. Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, et al. Most common 'sporadic' cancers have a significant germline genetic component. *Hum Mol Genet.* 2014;23:6112-8.
12. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA.* 2016;315:68-76.
13. Spurdle AB, Bowman MA, Shamsani J, Kirk J. Endometrial cancer gene panels: clinical diagnostic vs research germline DNA testing. *Mod Pathol.* 2017.
14. Chen MM, O'Mara TA, Thompson DJ, Painter JN, Australian National Endometrial Cancer Study G, Attia J, et al. GWAS meta-analysis of 16 852 women identifies new susceptibility locus for endometrial cancer. *Hum Mol Genet.* 2016;25:2612-20.
15. Cheng TH, Thompson DJ, O'Mara TA, Painter JN, Glubb DM, Flach S, et al. Five endometrial cancer risk loci identified through genome-wide association analysis. *Nat Genet.* 2016;48:667-74.
16. Painter JN, O'Mara TA, Batra J, Cheng T, Lose FA, Dennis J, et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum Mol Genet.* 2015;24:1478-92.
17. Thompson DJ, O'Mara TA, Glubb DM, Painter JN, Cheng T, Folkerd E, et al. CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. *Endocr Relat Cancer.* 2016;23:77-91.
18. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101:5-22.
19. Spurdle AB, Thompson DJ, Ahmed S, Ferguson K, Healey CS, O'Mara T, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat Genet.* 2011;43:451-4.
20. De Vivo I, Prescott J, Setiawan VW, Olson SH, Wentzensen N, Australian National Endometrial Cancer Study G, et al. Genome-wide association study of endometrial cancer in E2C2. *Hum Genet.* 2014;133:211-24.
21. Bahcall OG. iCOGS collection provides a collaborative model. Foreword. *Nat Genet.* 2013;45:343.

22. Setiawan VW, Doherty JA, Shu XO, Akbari MR, Chen C, De Vivo I, et al. Two estrogen-related variants in CYP19A1 and endometrial cancer risk: a pooled analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiol Biomarkers Prev.* 2009;18:242-7.
23. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev.* 2017;26:126-35.
24. O'Mara TA, Glubb DM, Amant F, Annibaldi D, Ashton K, Attia J, et al. Identification of nine new susceptibility loci for endometrial cancer. *Nat Commun.* 2018;9:3166.
25. Long J, Zheng W, Xiang YB, Lose F, Thompson D, Tomlinson I, et al. Genome-wide association study identifies a possible susceptibility locus for endometrial cancer. *Cancer Epidemiol Biomarkers Prev.* 2012;21:980-7.
26. Park SL, Cheng I, Haiman CA. Genome-Wide Association Studies of Cancer in Diverse Populations. *Cancer Epidemiol Biomarkers Prev.* 2018;27:405-17.
27. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet.* 2017;49:680-91.
28. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23:R89-98.
29. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008;27:1133-63.
30. Painter JN, O'Mara TA, Marquart L, Webb PM, Attia J, Medland SE, et al. Genetic Risk Score Mendelian Randomization Shows that Obesity Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial Cancer. *Cancer Epidemiol Biomarkers Prev.* 2016;25:1503-10.
31. Nead KT, Sharp SJ, Thompson DJ, Painter JN, Savage DB, Semple RK, et al. Evidence of a Causal Association Between Insulinemia and Endometrial Cancer: A Mendelian Randomization Analysis. *J Natl Cancer Inst.* 2015;107.
32. Kaaks R, Lukanova A, Kurzer MS. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiol Biomarkers Prev.* 2002;11:1531-43.
33. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet.* 2017;49:834-41.
34. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47:291-5.
35. Painter JN, O'Mara TA, Morris AP, Cheng THT, Gorman M, Martin L, et al. Genetic overlap between endometriosis and endometrial cancer: evidence from cross-disease genetic correlation and GWAS meta-analyses. *Cancer Med.* 2018;7:1978-87.
36. Kar SP, Beesley J, Amin Al Olama A, Michailidou K, Tyrer J, Kote-Jarai Z, et al. Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types. *Cancer Discov.* 2016;6:1052-67.
37. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet.* 2016;48:510-8.
38. Fehringer G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher FR, et al. Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. *Cancer Res.* 2016;76:5103-14.

39. Cheng TH, Thompson D, Painter J, O'Mara T, Gorman M, Martin L, et al. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci Rep*. 2015;5:17369.
40. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6:7138.
41. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93:779-97.
42. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551:92-4.
43. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317-30.
44. Pritchard JE, O'Mara TA, Glubb DM. Enhancing the Promise of Drug Repositioning through Genetics. *Front Pharmacol*. 2017;8:896.
45. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. *Am J Hum Genet*. 2018;102:717-30.
46. Krijger PH, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol*. 2016;17:771-82.
47. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48:245-52.
48. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48:481-7.
49. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet*. 2017;49:1602-12.
50. Davies JO, Oudelaar AM, Higgs DR, Hughes JR. How best to identify chromosomal interactions: a comparison of approaches. *Nature methods*. 2017;14:125-34.
51. Zhang X, Choi PS, Francis JM, Imielinski M, Watanabe H, Cherniack AD, et al. Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet*. 2016;48:176-82.
52. Painter JN, Kaufmann S, O'Mara TA, Hillman KM, Sivakumaran H, Darabi H, et al. A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates AKT1 through YY1 Binding. *Am J Hum Genet*. 2016;98:1159-69.
53. Kindt AS, Navarro P, Semple CA, Haley CS. The genomic signature of trait-associated variants. *BMC Genomics*. 2013;14:108.
54. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6:e1000888.
55. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16:197-212.
56. Rad E, Tee AR. Neurofibromatosis type 1: Fundamental insights into cell signalling and cancer. *Semin Cell Dev Biol*. 2016;52:39-46.
57. Devalliere J, Charreau B. The adaptor Lnk (SH2B3): an emerging regulator in vascular cells and a link between immune and inflammatory signaling. *Biochem Pharmacol*. 2011;82:1391-402.
58. Wrzeszczynski KO, Varadan V, Byrnes J, Lum E, Kamalakaran S, Levine DA, et al. Identification of Tumor Suppressors and Oncogenes from Genomic and Epigenetic Features in Ovarian Cancer. *PLOS ONE*. 2011;6:e28503.

59. Chiu HS, Somvanshi S, Patel E, Chen TW, Singh VP, Zorman B, et al. Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. *Cell reports*. 2018;23:297-312.e12.
60. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497:67-73.
61. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580-5.
62. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45:1113-20.
63. Bassett AR. Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mammalian Genome*. 2017;28:348-64.
64. Dominguez AA, Lim WA, Qi LS. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol*. 2016;17:5-15.
65. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst*. 2017;109.
66. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856-60.
67. Prescott J, Setiawan VW, Wentzensen N, Schumacher F, Yu H, Delahanty R, et al. Body Mass Index Genetic Risk Score and Endometrial Cancer Risk. *PLoS One*. 2015;10:e0143256.

Table 1. Genetic risk regions and candidate target genes identified by endometrial cancer genetic association studies to date.

Risk locus	Other relevant traits identified by GWAS at this locus*	Candidate target/Closest gene†	Evidence for targeting	Involvement in relevant pathways from network analysis	Studies reporting endometrial cancer risk association
1p34.3	Cancer: ovarian	<i>CDCA8</i>	eQTL (24)	-	(24)
2p16.1	Cancer: Hodgkin's lymphoma	<i>BCL11A</i>	-	-	(24)
6q22.3	Cancer: melanoma, neuroblastoma Anthropometric: height, body mass index Steroid hormone-related: bone mineral density	<i>SOX4, CASCI5</i>	-	-	(14, 24)
6q22.31	Cancer: bladder Anthropometric: height, hip circumference Steroid hormone-related: Age of menarche	<i>HEY2</i> <i>NCOA7</i>	Bioinformatic prediction (15) Bioinformatic prediction (15)	Notch signalling Endometrial cancer signalling	(14, 15, 24)
8q24.21	Cancer: acute lymphoblastic leukemia, breast, diffuse large B cell lymphoma, follicular lymphoma, glioma, Hodgkin's lymphoma, ovarian, pancreatic Anthropometric: height	<i>MIR1204</i> <i>MIR1205</i> <i>MIR1207</i> <i>MIR1208</i> <i>MYC</i>	Bioinformatic prediction (15) Bioinformatic prediction (15) Bioinformatic prediction (15) Bioinformatic prediction (15) Chromatin looping; bioinformatic prediction (15)	- - - - Molecular mechanisms of cancer Wnt/ β -catenin signalling Estrogen-mediated S-phase entry G1/S checkpoint regulation	(15, 24)
9p21.3	Cancer: acute lymphoblastic leukemia, breast, chronic lymphocytic leukemia, glioma, glioblastoma, lung, melanoma, multiple myeloma, nasopharyngeal, oral cavity, prostate Anthropometric: body mass index Endometriosis	<i>CDKN2A/B</i>	-	-	(24)
11p13	Anthropometric: body mass index, waist-hip ratio	<i>CCDC73</i> <i>EIF3M</i> <i>RCN1</i> <i>TCP11L1</i> <i>WT1-AS</i>	eQTL (24) eQTL (24) eQTL (24) eQTL (24) eQTL (24)	- EIF2 signalling - - -	(24)
12p12.1	Cancer: esophageal, renal cell Anthropometric: waist-hip ratio	<i>SSPN</i>	-	-	(24)
12q24.12	Cancer: breast, colorectal, esophageal Anthropometric: body mass index	<i>SH2B3</i>	eQTL (24)	-	(24)
12q24.21	Cancer: esophageal, laryngeal squamous cell, pancreatic, prostate Steroid hormone-related: mammographic density	<i>SNORA27</i>	-	-	(24)
13q22.1	Cancer: breast, chronic lymphocytic leukemia, pancreatic, prostate Steroid hormone-related: age of menarche	<i>KLF5</i>	Functional analyses; bioinformatic prediction (15)	Adipogenesis pathway	(14, 15, 24)
15q15.1	Cancer: Ewing sarcoma Anthropometric: body mass index, height Steroid hormone-related: age of menopause	<i>BMF</i> <i>GPR176</i> <i>SRP14-AS1</i> <i>SRP14</i>	Bioinformatic prediction (15) Bioinformatic prediction (15) Bioinformatic prediction (15) Bioinformatic prediction (15)	- - - -	(15, 24)
15q21.2	Anthropometric: body mass index, height Steroid hormone-related: bone mineral density, estradiol levels, follicle stimulating hormone	<i>CYP19A1</i>	eQTL and association with estradiol levels (17)	Estrogen-dependent breast cancer signalling	(15, 17, 24)
17q11.2	Cancer: breast, prostate Anthropometric: body mass index, height, hip circumference, waist circumference	<i>EVI2A</i> <i>NF1</i> <i>SUZ12</i>	eQTL (24) eQTL (24) eQTL (24)	- Molecular mechanisms of cancer -	(24)
17q12	Cancer: ovarian, pancreatic, prostate, testicular germ cell Anthropometric: body mass index, height	<i>HNF1B</i>	Functional analyses; eQTL (16)	Pigment epithelium-derived factor (PEDF) signaling	(14-16, 19, 24)
17q21.32	Cancer: ovarian Anthropometric: body fat percentage, body mass index, height, obesity	<i>SNX11</i>	eQTL (24)	-	(24)

*From GWAS catalog (<https://www.ebi.ac.uk/gwas/home>), accessed May 2018 †Bolded genes have evidence for being a candidate target gene, unbolded genes are closest genes

Table 2. Mendelian randomisation studies assessing the causal relationship between putative risk factors and endometrial cancer risk.

Trait assessed (no. variants used in risk score)	No. Cases	No. Controls	Association results	Comments	Reference
Type 2 Diabetes (49 variants)	1,287	8,273	0.91 (0.79-1.04); P=0.16	Endometrioid endometrial cancer cases only	(31)
Increased fasting glucose (36 variants)	1,287	8,273	1.00 (0.67-1.50); P=0.99	Endometrioid endometrial cancer cases only	(31)
Increased fasting insulin (18 variants)	1,287	8,273	2.34 (1.06-5.14); P=0.03	Endometrioid endometrial cancer cases only	(31)
Early insulin secretion (17 variants)	1,287	8,273	1.40 (1.12-1.76); P=0.003	Endometrioid endometrial cancer cases only	(31)
Body mass index (32 variants)	1,287	8,273	3.86 (2.24-6.64); P=1×10 ⁻⁶	Endometrioid endometrial cancer cases only	(31)
Body mass index (97 variants)	3,376	3,867	1.13 (1.04-1.22); P=0.002	Endometrioid endometrial cancer cases only. Association did not persist after adjustment for measured body mass index.	(67)
Body mass index (77 variants)	6,609	37,926	2.011 (1.94-2.28); P=3.4×10 ⁻¹⁷	All endometrial cancer cases. Remained significant after adjustment for measured body mass index (OR 1.23; P=5.3×10 ⁻⁴).	(30)
Waist-hip ratio (34 variants)	6,609	37,926	1.02 (0.99-1.04); P=0.09	All endometrial cancer cases. Waist-hip ratio variants are those that were associated with this trait amongst women.	(30)
Waist-hip ratio (47 variants)	6,609	37,926	0.97 (0.63-1.31); P=0.86	All endometrial cancer cases. Waist-hip ratio variants are those that were associated with this trait in men and women.	(30)
Serum estradiol level (1 variant)	6,608	37,925	1.15 (1.11-1.21); 4.8×10 ⁻¹¹	All endometrial cancer cases. CYP19A1 variant rs727478 used to predict serum estradiol level (10% increase per A-allele).	(17)
Menarche (age of onset) (237 variants)	6,609	37,926	0.78 (0.70-0.87); P=1.0×10 ⁻⁵	All endometrial cancer cases. Menarche variants were adjusted for genetically predicted body mass index	(33)
Adult Height (814 variants)	12,906	108,979	1.00 (0.95-1.06); P=0.90	All endometrial cancer cases.	(24)

Figures Legends

Figure 1. Schematic of the known genetic architecture of endometrial cancer. Low frequency genetic variants from *PTEN* and the mismatch repair genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*) are considered high risk variants (estimated risk >4-fold). Common variants (frequency > 1%) identified by GWAS are considered low risk variants (estimated risk <2-fold). No variants associated with moderate risk (estimated risk ~2 to 4-fold) are currently established. Although variants in other genes have been implicated in predisposition to endometrial cancer, including *STK11*, *POLD* and *POLE*, their clinical utility for the purpose of altering patient management is unclear, since risk estimates are imprecise, and/or based on a limited number of studies with potential for ascertainment bias (13).

Figure 2. Long-range gene regulation via chromatin looping events and functional consequences of genetic variants in epigenomic features. **Panel A** depicts chromatin looping bringing an enhancer (characterised by active histone marks) into proximity with a gene promoter, allowing transcription factor (TF) binding and gene transcription. **Panel B** depicts the same scenario with a variant allele (A to G change), resulting in further TF binding and increased gene expression.

Figure 1

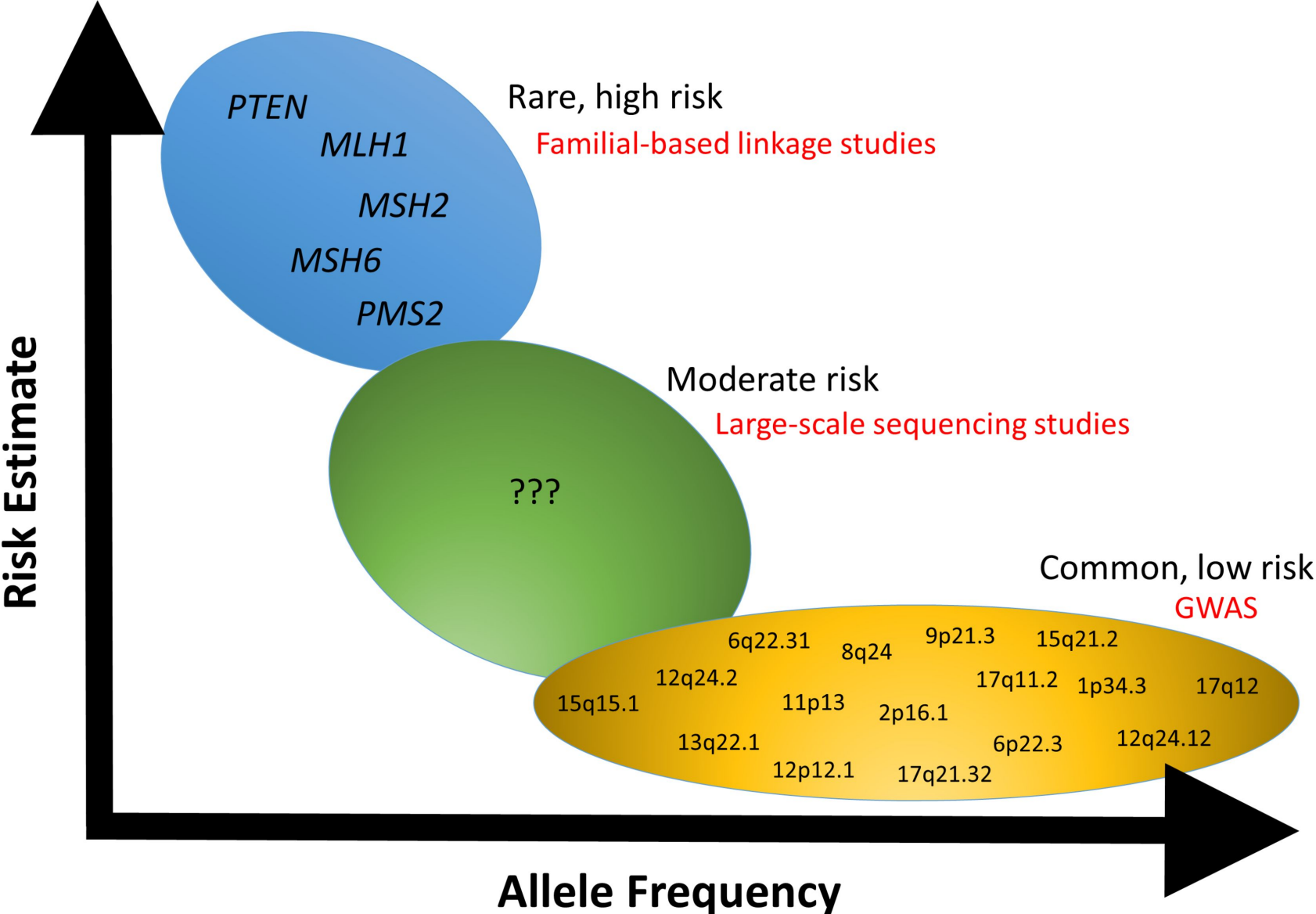


Figure 2

