Review

Women's Health

Genetics of endometriosis

Endometriosis is a heritable complex disorder that is influenced by multiple genetic and environmental factors. Identification of these genetic factors will aid a better understanding of the underlying biology of the disease. In this article, we describe different methods of studying genetic variation of endometriosis, summarize results from genetic studies performed to date and provide recommendations for future studies to uncover additional factors contributing to the heritable component of endometriosis.

Keywords: association studies • candidate gene studies • common genetic variants • functional studies • genome-wide association studies • linkage studies • rare genetic variants • rhesus macaque • sub-phenotype • twin studies

Nilufer Rahmioglu*[,1], Grant W Montgomery[2] & Krina T Zondervan[1,3]

[1]Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

[2]Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia

[3]Nuffield Department of Obstetrics & Gynaecology, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DU, UK

*Author for correspondence: nilufer@well.ox.ac.uk

Endometriosis as a complex disease

Endometriosis is a complex disorder that is caused by combination of multiple genetic and environmental factors. This means that inherited genetic variants associated with endometriosis represent only part of the risk associated with the disease. Variants confer genetic susceptibility to develop the disease but environment and lifestyle factors also play an important role, both through interaction with genetic factors and independently [1].

The heritable component of endometriosis has been illustrated by many different studies [2–4]. Higher rates of endometriosis among relatives of endometriosis cases compared with controls was shown, with risk ratio compared with general population risk for sisters estimated at 5.2 and for cousins 1.6 in a population-based study in Iceland [3]. In addition, in a hospital-based study in the UK, consisting of 230 women with surgically confirmed endometriosis in 100 families, familial aggregation of endometriosis was shown [4]. However, estimates of familial aggregation in human populations are likely to be affected – to an unknown extent – by the fact that endometriosis is only reliably diagnosed through laparoscopy. The chance of being diagnosed with endometriosis may be influenced by having a family member already diagnosed with disease and it is difficult to get an accurate population-based estimate of disease risk [5]. Stronger evidence of heritability is provided by twin studies, that have shown higher concordance in monozygotic twins compared with dizygotic twins, a finding less likely to be affected by selection biases operating on diagnostic opportunity [6,7]. The largest twin study, among 3096 Australian female twins, estimated the heritable component of endometriosis at 51% [7]. In addition to the human studies, familial aggregation of endometriosis is shown in nonhuman primates that develop endometriosis spontaneously, such as the rhesus macaque evidence, which is also less likely to be affected by diagnostic bias [8].

The underlying biological mechanisms causing endometriosis remain poorly understood to date. One fundamental approach to uncover the underlying mechanisms and causes of endometriosis is through identification and understanding of the functions of the genetic variants responsible for the heritable component of the disease.

Future Medicine part of fsg

Discovery of underlying genetic risk variants

Approaches for the investigation of genetic variants underlying endometriosis divide into hypothesis-based and hypothesis-free methods. Hypothesis-based approaches rely on prior biological understanding of the disease through selection of functionally relevant genes/regions to test for association with the disease, whereas hypothesis-free approaches target the whole genome without preselecting any regions or genes.

Hypothesis-based candidate gene studies

Hypothesis-based candidate gene association studies have to date been the most common type of study in the investigation of genetic factors underlying complex diseases. Candidate genes studies can be either biological or positional. Biological candidate gene studies are based on prior selection of genes with a hypothesized function relevant to the disease. Variants in such genes are tested for association with disease risk. This type of biological candidate gene studies are inherently limited, as they assume a known underlying biological mechanism, a limitation that is particularly relevant in a disease such as endometriosis for which etiology is not well understood.

We previously published reviews of candidate genes studies of endometriosis, up to 1 April 2008 [9], and up to 1 April 2012 [10]. A systematic search of PubMed from 1 April 2012 to 1 December 2013 resulted in an additional 12 studies that have been published, testing 31 variants in 14 candidate genes (Table 1). Similar to other complex disease fields, candidate gene studies of endometriosis have been unsuccessful in providing replicable results [10]. For an identified association to be credible, the result needs to be replicated in an independent study in individuals from the same ethnic population. The reasons for failure of these biological candidate gene studies are now well understood. First, the tested biological hypothesis may not be true due to lack of understanding of the underlying biological mechanisms of the disease. Second, usually only a few selected variants are tested with incomplete coverage of the genetic variation in the candidate gene. Third, only the involvement of one or a few genes in a biological pathway is tested, missing potential other genes making up the pathway. Fourth, the cases and controls within and between these studies are not well defined or their definitions vary, which reduces the power to detect or replicate associations [11]. Last, the sample sizes are usually insufficient to detect genetic variants for complex traits.

In positional candidate gene studies, genes and variants are selected based on prior evidence of the likely genomic location of a disease risk variant from hypothesis-free whole genome linkage studies (described later in the article) combined with biological plausibility. This type of positional 'filtering' is a method that *a priori* has a greater likelihood of identifying genes harboring risk variants, although depending on the extent and knowledge of the region, the number of biologically interesting genes may still be large. Relatively few such studies have been performed in endometriosis [23–25].

Hypothesis-free studies

There are two types of hypothesis-free study designs: family-based linkage studies and genome-wide association studies (GWAS). The two designs are fundamentally different, but complementary, approaches in the identification of genetic risk variants across the whole allelic frequency spectrum. Linkage studies are aimed at identifying genetic variants that are rare in the general population, but are responsible for aggregation of disease in a family, whereas association studies are aimed at identifying common genetic variants in the general population related to disease risk.

Linkage studies are family/pedigree-based, and utilize the information of disease aggregation together with chromosomal recombination events, identifying chromosomal regions shared more commonly between multiple affected in a family than expected by the laws of Mendelian inheritance. They have been highly successful in identifying high-risk variants in monogenic diseases such as cystic fibrosis or Huntington's chorea [26,27]. In complex disease, however, due to underlying heterogeneity in (genetic) causality, linkage evidence tends to point at very large genomic regions that typically extend across 10–50 Mb, containing hundreds of genes. This is in contrast to association studies that can pinpoint a signal to 10–500 Kb depending on the local linkage disequilibrium (LD) structure.

A summary of genome-wide linkage studies of endometriosis to date is provided in Table 2.

Whole genome linkage studies

To conduct a linkage study of a complex disease such as endometriosis, many families with multiple affected women are required. The largest genome-wide linkage study for endometriosis was conducted by the IES, which included 1176 families from Australia (n = 931) and UK (n = 245) with at least two members – mainly affected sister pairs – with surgically diagnosed endometriosis (Table 2). This study identified a region of significant linkage on chromosome 10q26 and another region of suggestive linkage on chromosome 20p13 [33]. In a subsequent study by the IES, including a subset of 248 families with three or more affected members, an additional linkage region was identified on chromosome 7p13–15 (Figure 1) [34].

Table 1. Candidate gene association studies of endometriosis published from April 2012 to December 2013, and summarized evidence for the genes from all candidate gene association studies to date.

Candidate gene	Chr	Number of studies with association 2008–2013[†]		Significant/total up to 2013 (n/n)[‡]
		Significant (n cases; n controls)	Nonsignificant (n cases; n controls)	
ACP1	2p25	Italian (132; 359) [12]	–	1/1
CETP	16q21	Iranian (97; 107) [13]	–	1/1
CYP17A1	10q24	Polish (115; 197) [14]	–	2/10
CYP19A1	15q21	Polish (115; 197) [14]	–	5/7
DRD2	11q23	Brazilian (58; 49) [15]	–	1/1
FOXP3	Xp11.23	Han Chinese (314; 358) [16]	–	1/1
GSTM1	1p13	Brazilian (50; 46) [17]	–	8/14
GSTT1	22q11	Brazilian (50; 46) [17]	–	3/10
IL-10	1q31–32	–	Chinese (1080; 1134), Danish (100; 358)[§] [18]	5/7
TP53	17p13	Italian (132; 359) [12] Brazilian (50; 46) [17]	–	6/11
PTPN22	1q13	Italian (132; 359) [12]	–	4/5
TNFA	6p21	Iranian (135; 173) [19]	–	2/3
VEGF	6p21–12	Asian, Caucasian (3313; 3393)[¶] [20] Indian (302, 324) [21]	Iranian (135; 173) [19]	12/15
VEGFR-2	4q11–q12	Northern Chinese (571; 580) [22]	–	1/1

[†]When more than one polymorphism was investigated in a study, the result is indicated as significant if one or more of the variants were reported as significant by the authors.
[‡]Total number of candidate studies is combined with data from our review in 2012 [10].
[§]Literature-based meta-analysis of three polymorphisms in *IL-10* gene from eight case–control studies from seven Chinese and one Danish population [55].
[¶]Literature-based meta-analysis of five polymorphisms in *VEGF* gene from 14 Asian case–control studies [57].

To follow-up these linkage regions and identify genetic variants, 'fine mapping' studies have been performed. The first such study extensively genotyped the linkage region on chromosome 10q26 using 11,984 single nucleotide polymorphism (SNPs) in 1144 familial cases and 1190 controls (Figure 1). The study identified three independent signals with significant evidence of association to endometriosis including rs11592737 at 96.59 Mb, rs1253130 at 105.63 Mb and rs2250804 at 124.25 Mb [28]. However, only rs11592737 was robustly replicated ($p = 0.04$) in an independent cohort of 2079 cases and 7069 controls. rs11592737 is located in the *CYP2C19* gene, which is an important potential candidate gene for endometriosis. It is involved in drug metabolism as well as estrogen metabolism including conversion of E2 to estrone (E1) [29]. The association was unlikely to explain all of the linkage signal observed on chromosome 10q26, and further studies are needed to identify rare genetic variants in this region, through resequencing approaches. Functional studies are needed to identify the genes and determine the effects of the variants in underlying biological pathways.

A second follow-up study was performed for the chromosome 7p13–15 linkage region through sequencing of the coding regions and upstream regulatory regions of three strong candidate genes in the linkage region, namely *INHBA*, *SFRP4* and *HOXA10*, with roles in endometrial development (Figure 1) [30]. The sequencing was performed in 47 cases from 15 families, aimed at identifying rare variants that are not present on genotyping arrays. The study identified 11 variants, five of which were common (minor allele frequency [MAF] >0.05) and the remaining six variants were rare. However, none of these variants either individually or collectively was significantly associated with endometriosis risk, and rare variants in the coding regions or the regulatory regions of these three genes are unlikely to be

responsible for the linkage observed in these families. As only coding and regulatory regions were sequenced, it remains possible that the noncoding regions of these three genes (i.e., intronic regions) harbor rare variants associated with endometriosis, or indeed that other genes in this linkage region contain risk variants.

Genome-wide association studies

GWAS have been very successful in identifying common genetic risk variants for various complex diseases. GWAS became a reality after two major scientific and technological developments. Firstly, the International HapMap Project together with the International SNP Consortium mapped approximately 10 M common SNPs and the LD pattern between them in different ethnic populations [31], that allowed efficient genome-wide assays, through the genotyping of tagSNPs, of most of the common variation across the genome. Second, high-throughput technology generated genotyping arrays able to genotype 100,000s to millions of SNPs at ever decreasing costs. For a detailed description of the design of GWAS see Zondervan & Cardon [11]. All genetic variants identified for complex disorders and traits through GWAS are documented and regularly updated in the National Human Genome Research Institute (NHGRI) GWA Catalog [35].

To date, a total of five GWAS of endometriosis have been published in populations from European and Japanese ancestry (Table 2). The first study published was in 1432 cases (mixture of surgically diagnosed and clinically diagnosed women) and 1318 control patients from the Japanese population [32]. The discovery phase analysis included 460,945 SNPs after quality control criteria were applied. The second phase replication analysis included an independent set of 484 cases and 3974 control patients and the top 100 most significantly associated SNPs from the discovery phased were followed up. The study identified a significant association with rs10965235 located in *CDKN2B-AS1* gene on chromosome 9 ($p = 5.6 \times 10^{-12}$; odds ratio [OR]: 1.44 [1.30–1.59]) and a suggestive association with rs16826658 near *WNT4* on chromosome 1 ($p = 1.7 \times 10^{-6}$; OR: 1.20 [1.11–1.29]).

The second Japanese GWAS was on a smaller sample size of 696 cases (not surgically confirmed) and 825 controls including 282,828 SNPs tested after quality control [36]. This study did not discover any significant risk variants for endometriosis, which is likely to be due to a combination of the small sample size and suboptimal case definition.

The first GWAS on endometriosis among women of European ancestry was performed by an extension of the IES – the International Endogene Consortium (IEC), involving 3194 surgically confirmed endometriosis cases and 7060 controls from Australia and the UK, including 504,723 SNPs after quality control [37]. Disease severity of the cases was also categorized using rAFS classification system from retrospective surgical records, which resulted in two phenotypes: stage I and II or some ovarian disease with a few adhesions (mild disease) (n = 1686; 52.7%) and stage III and IV disease (n = 1364; 42.7%) or unknown (n = 144; 4.6%).

This study analyzed the effect of all SNPs combined showing a significantly higher 'genetic loading' among stage III and stage IV endometriosis cases (proportion of endometriosis variation explained by common SNPs: 0.34 [SD: 0.04]) compared with stage I and stage II cases (variation explained: 0.15 [SD: 0.15]; $p = 1.1 \times 10^{-3}$). Driven by this result, two GWA analyses were performed; including all endometriosis cases (n = 3194 cases), including only stage III and IV endometriosis cases (n = 1364 cases).

Table 2. Summary of the published genome-wide linkage and association studies of endometriosis.

Cohort	Ancestry	Cases (n)	Stage III/IV cases† (n)	Controls (n)	Ref.
Genome-wide linkage studies					
IES linkage analysis I	European	1176 affected sib pair families			[23]
IES linkage analysis II	European	248 families with 3 ≤ affected members			[24]
GWAS					
BBJ GWAS	Japanese	1423	–	1318	[28]
Adachi GWAS	Japanese	696	–	825	[29]
IEC GWAS	European	3194	1364	7060	[30]
BBJ and IEC meta-analysis	European and Japanese	4604	–	9393	[31]
Utah GWAS	European	2019	848	14,471	[32]

†The gene name at the single nucleotide polymorphism location or the nearest gene is given for each of the associated single nucleotide polymorphisms.
GWAS: Genome-wide association studies; IEC: International Endogene Consortium.

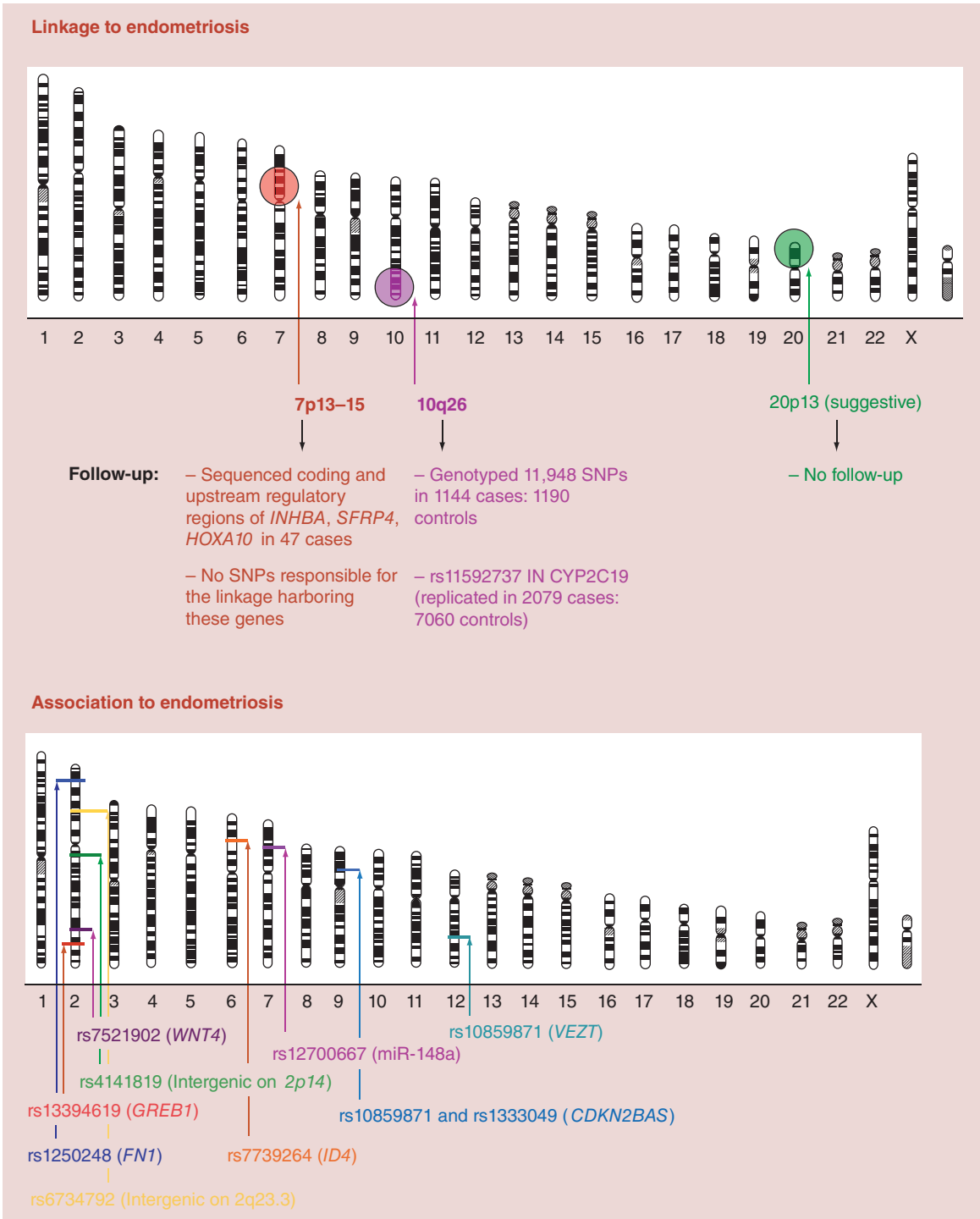

Figure 1. The linkage regions and association signals identified via whole genome linkage and whole genome association studies of endometriosis.
SNP: Single nucleotide polymorphism.

The 'all endometriosis' GWAS identified two genome-wide significant associations, rs12700667 in an intergenic region, which is 88 kb upstream of a microRNA (*miRNA-148a*) and 290 kb upstream of *NFEL2L3* on chromosome 7 ($p = 2.6 \times 10^{-7}$; OR: 1.22 [1.13–1.32]; rs1250248 in *FN1* locus on chromosome 2 ($p = 1.7 \times 10^{-5}$; OR: 1.17 [1.09–1.26]). Both these associations were stronger in the stage III and IV limited analysis ($p = 1.5 \times 10^{-9}$; OR: 1.38 [1.24–1.53]; $p = 3.2 \times 10^{-8}$ OR: 1.30 [1.19–1.43], respectively). In the rep-

lication phase, the top 70 most significantly associated SNPs from the discovery phase were followed up in an independent dataset of 2392 self-reported surgically confirmed cases and 2271 controls from the Nurses' Health Study I and II in the USA. The association of rs12700667 was replicated ($p = 1.2 \times 10^{-3}$, OR: 1.17 [1.06–1.28]) but rs12540248 showed no evidence of replication (p = 0.57). The combined analysis of Australian, UK and USA datasets further confirmed the association of rs12700667 with endometriosis (all endometriosis $p = 1.4 \times 10^{-9}$; OR: 1.20 [1.13–1.27]).

This study has also investigated the significant associations reported by the Japanese GWAS of endometriosis. Their results did not show an association with rs10965235 on chromosome 9 in the European women, since this SNP is monomorphic (not variable) in individuals of European ancestry. Furthermore, no SNPs in LD with rs10965235 in the European population showed any association with endometriosis, reflecting the different genetic backgrounds of the two studied populations. rs75121902 near *WNT4* locus, which is in strong LD with rs10965235 in the Japanese population (R^2 = 0.98), showed a strong association with stage III/IV endometriosis; a meta-analysis of all endometriosis cases (since severity of disease is not reported by the Japanese GWAS) from both GWAS datasets provided genome-wide significance of the variant ($p = 4.2 \times 10^{-8}$; OR: 1.19 [1.12–1.27]).

A formal genome-wide association meta-analysis of the IEC GWAS and the Japanese GWAS was subsequently performed, resulting in 4604 cases and 9393 controls from Japanese and European ancestry, and additional including an independent replication dataset of 1044 cases and 4017 controls from the BioBank Japan [38]. This meta-analysis confirmed the two associations published by the original papers and identified five additional genome-wide significant loci, namely rs13394619 in *GREB1* on chromosome 2 and rs10859871 near *VEZT* on chromosome 12 (OR: 1.20 [1.14–1.26]; $p = 5.1 \times 10^{-13}$), rs4141819 in an intergenic region on chromosome 2 (OR: 1.15 [1.09–1.21]; $p = 8.5 \times 10^{-8}$), rs7739264 near *ID4* on chromosome 6 (OR: 1.17 [1.11–1.23]; $p = 3.6 \times 10^{-10}$) and rs1537377 near *CDKN2B-AS1* on chromosome 9 (OR: 1.15 [1.10–1.21]; $p = 2.4 \times 10^{-9}$).

A fourth GWAS analysis was published in 2013, in 2019 cases and 14,471 controls of European ancestry from the USA [39]. This study reported two additional genome-wide significant associations, namely, rs2235529 near *WNT4* (in LD with rs7521902 $r^2 = 0.66$; $p = 8.65 \times 10^{-9}$) and rs1519761 in an intergenic region 280 kb upstream of *RND3* on chromosome 2 ($p = 4.70 \times 10^{-8}$). However, the study did not significantly replicate the rs12700667 signal on chromosome 7 (p = 0.12).

Recently, we have conducted a meta-analysis of all genome-wide significant results from all published GWAS and replication studies in endometriosis, amal-

Table 3. Eight genome-wide significant genetic variants robustly associated with endometriosis.

SNP	Chr	Position (HG19)	P_{meta} [†]	OR_{all} [‡] (95% CI)	$OR_{stage\ III/IV}$ [§] (95% CI)	Nearest gene (distance)
rs7521902	1	22490474	1.8×10^{-15}	1.18 (1.13–1.23)	1.25 (1.16–1.33)	*WNT4* (21 Kb)
rs13394619	2	11727257	2.9×10^{-8}	1.13 (1.07–1.20)	1.18 (1.11–1.24)	*GREB 1* (0)
rs4141819	2	67864425	8.8×10^{-6} (9.2×10^{-8})	1.08 (1.04–1.12)	1.16 (1.09–1.24)	Intergenic (*ETAA1*: 227 Kb)
rs1250248	2	216286843	1.1×10^{-4} (8.0×10^{-8})	1.11 (1.04–1.18)	1.26 (1.16–1.38)	*FN1* (0)
rs7739264	6	19785338	6.2×10^{-10}	1.11 (1.08–1.15)	1.20 (1.13–1.28)	*ID4* (52 Kb)
rs12700667	7	25901389	1.6×10^{-9}	1.13 (1.08–1.17)	1.22 (1.14–1.31)	Intergenic (*miR-148a*: 88 Kb, *NFE2L3*: 290 Kb)
rs1537377	9	22169450	1.0×10^{-8}	1.12 (1.08–1.17)	1.18 (1.11–1.26)	*CDKN2B-AS1* (48 Kb)
rs10859871	12	95711626	4.8×10^{-15}	1.18 (1.13–1.22)	1.19 (1.11–1.27)	*VEZT* (17 Kb)

[†]Meta-analysis p-value is from Rahmioglu *et al.* 2014 [40]. For rs4141819 and rs1250248 the meta-analysis p-values including only stage III/IV cases is given in parenthesis, which reach borderline significance.
[‡]Odds ratio including all endometriosis cases.
[§]Odds ratio including only stage III/IV cases.
SNP: Single nucleotide polymorphism.
Adapted from Rahmioglu *et al.* (2014) [40].

Table 4. Biological functions of potential genes identified for association with endometriosis.

Gene name	Known functions
WNT4	A protein-coding gene that is vital for development of the female reproductive organs. In knockout mice, the loss of WNT4 leads to complete absence of the Müllerian duct and its derivatives [41]. It has been also shown that WNT genes are expressed in the peritoneum and endometrium [42]
GREB1	An early response gene in the estrogen regulation pathway that is involved in hormone-dependent breast cancer cell growth. Its expression was increased in peritoneal eutopic endometrotic lesions when compared with eutopic endometrium, possibly with a role in oestrogen-dependent growth of endometriosis [43]
ETAA1	Encodes for a tumor-specific cell surface antigen in Ewing family of tumors, which is a group of cancers that form in bone or soft tissue that share common features as they develop from the same type of stem cell in the body [44]
FN1	Involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defense and metastasis [45]. It has been shown that *SOX2*, a gene encoding a transcription factor that targets FN1, is a key gene regulating cell migration in ovarian cancer [46]
ID4	An ovarian oncogene, which is also implicated in methylation-related regulatory pathways in breast cancinogenesis [47]. It is also overexpressed in ovarian, endometrial and breast cancer cell lines [48]
NFE2L3	A transcription factor suggested to be involved in cell differentiation, inflammation and carcinogenesis [49]. Its expression was increased in human breast cancer cells [50] and testicular carcinoma tissue samples [51]
miRNA_148a	A microRNA with a role in the Wnt/β-catenin signaling pathway that may have a role in endometriosis development through sex hormone homeostasis regulation [52] and fibrogenesis [53]
HOXA10	Belongs to homeobox A family of transcription factors, which plays essential roles in specifying regional differentiation of the Müllerian duct into oviduct, uterus, cervix and vagina [54]
CDKN2B-AS1	Involved in regulation of CDKN2B, CDKN2A and ARF tumor suppressor genes. CDKN2A inactivation has been reported in endometriosis and endometrial cancer [55]
VEZT	Encodes an adherens junction transmembrane protein. Also, it is a putative tumor suppressor gene, targeting cell migration and invasion genes, growth genes, cellular adhesion genes and a functionally validated cell cycle progression gene called *TCF19*. TCF19 is involved in maintaining immunological balance [56]

gamating the evidence for all genome-wide significantly reported loci across the studies [40]. Our study has shown robust evidence for six loci associated with endometriosis and two additional borderline associations with stage III/IV endometriosis. Furthermore, all eight loci showed stronger associations when the dataset is restricted to only more severe endometriosis cases (stage III/IV) (Table 3).

When the variants are located within genes, the biological consequences may act through these genes (e.g., for *GREB1* and *FN1*). Alternatively, SNPs within introns or in intergenic regions may act through *cis* regulation of the nearby genes, or may have a regulatory action on more distal genes, including trans effects on different chromosomes. A brief summary of the functions of the genes closest to the identified endometriosis risk variants are provided in Table 4. All regions need to be followed up in functional studies to better understand their roles in endometriosis disease etiology. For an extensive description of the potential biological roles of these genes with regard to endometriosis see Rahmioglu *et al.* 2014 [40].

Conclusion & future perspective

The progress from genetic studies of endometriosis, and that in other complex disease fields, indicate there are a number of areas, which we can work on to improve our understanding of the genetic origins of endometriosis.

The IES results demonstrated a greater genetic loading for stage III/IV cases compared with stage I/II cases [37], along with the stronger associations observed with all the confirmed endometriosis risk variants with the more severe stages of the disease (Table 3). These

results highlight the importance of collecting detailed subphenotype information on the endometriosis cases. Collection of detailed surgical data from the cases will allow assessment of genetic variants associated with different stages of the disease as well as specific subtypes of the disease such as deep infiltrating endometriosis, rectovaginal disease. Equally important is to collect the disease phenotypes in a standardized manner, which enables collaborative research and comparison of results between centers. The global WERF Endometriosis Biobanking and Phenome Harmonisation Project (EPHect) is developing freely available data collection tools to enable standardized data collection [57].

To date, the sample sizes of GWAS of endometriosis have been modest compared with other complex diseases, such as Breast cancer reaching sample size of 60,000 individuals and identifying up 68 genome-wide significant genetic risk variants [58]. Increased sample size of the GWAS meta-analysis from different populations will aid in identification of further common genetic risk variants of endometriosis [59]. This emphasizes the importance of collaboration between centers to reach larger numbers of well-defined endometriosis cases and controls.

The most commonly studied type of genetic variation is common SNPs through GWAS. However, there is a whole spectrum of rarer mutations (MAF < 0.05) both single sited and structural (e.g., copy number variations) that are likely to play a role in disease causation and explain part of the genetic heritability. These rare mutations are not captured by the genotyping arrays and are therefore missed by the GWAS. As described in the linkage studies section, there are three linkage regions identified for endometriosis, which are likely to harbor rare mutations. These regions need to be sequenced to investigate the rare mutations responsible.

It is important to note that the genetic variants identified through GWAS are unlikely to be the causal variants, but instead be in LD with the actual disease causing variants. Further sequencing or dense genotyping of these regions in large sample numbers are required to pin down the causal variants and understand their roles in downstream biological pathways. Once the genetic variants of endometriosis are identified, they need to be followed up in functional studies in relevant disease tissues such as eutopic and ectopic endometrium, to shed light on their roles in endometriosis. Functional studies include understanding the variation in different levels of the whole systems biology including, expression levels of genes (mRNA levels), protein levels and metabolite levels in different tissues and how these relate to changes in the DNA level. These will aid in understanding how the identified genetic variants cause the resulting disease phenotype.

Although mouse models of endometriosis are an important tool to better understand functional mechanisms involved in the disease [60], they are limited in that they represent an induced disease model. An important animal model in which the (genetic) origin of spontaneous disease can be studied is the rhesus macaque (*Macaca mulatta*). Endometriosis develops spontaneously in rhesus macaques, and the disease is morphologically and histologically identical to that seen in humans [61]. Captive rhesus macaques are kept in colonies, which creates a smaller gene pool and greater genetic homogeneity to investigate genetic risk variants of endometriosis [61]. Furthermore, they live in controlled environments, which make them a good resource to study retrospective environmental exposures and potentially to study gene–environment interactions.

So far, the identified genetic variants of endometriosis have small effects and only explain a small

Executive summary

- Endometriosis is a complex disorder with a considerable genetic component (heritability = 51%).
- Hypothesis-free genome-wide investigations of genetic variants causing endometriosis are more appropriate since the underlying mechanisms of endometriosis etiology are not well understood.
- To date, eight genome-wide significant robust common genetic variants (single nucleotide polymorphisms) have been associated with endometriosis. All these associations are more strongly associated with the more moderate/severe form of the disease (rAFS stage III/IV disease).
- To date, three linkage regions of endometriosis have been identified. These regions need to be followed up to identify the rare genetic variants involved that are expected to have larger effects on the disease risk.
- Future studies should include more detailed and standardized disease information on the cases to allow for subphenotype analysis to identify genetic variants associated with different subtypes or groups of the disease.
- Better coverage of the genomic variation in studies is crucial to investigate all of the genetic variation.
- Functional studies are needed to uncover the roles of the identified genetic variants on underlying biological pathways of endometriosis.
- The identified genetic variants are important in determining the underlying biology of endometriosis that will aid in developing personalized treatment options for patients.

proportion (3.4%) of the heritable component of the disease. Therefore, they are not suitable for use as clinical diagnostic markers. However, they provide very important starting points to investigate the underlying biological pathways contribution to endometriosis that will aid in understanding the etiology of disease. In the long term, it is hoped that they will aid in defining genetically heterogenous subtypes and allow stratification of patients for different treatments through the identification of differential pathways contributing to subtypes of the disease.

Financial & competing interests disclosure

GW Montgomery is supported by the NHMRC Fellowships Scheme (339446, 613667). KT Zondervan is supported by a Wellcome Trust Research Career Development Fellowship (WT085235/Z/08/Z). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Craig J. Complex diseases: research and applications. *Nat. Educ.* 1(1), 184 (2008).

2 Zondervan KT, Cardon LR, Kennedy SH. The genetic basis of endometriosis. *Curr. Opin. Obstet. Gynecol.* 13, 309–314 (2001).

3 Stefansson H, Geirsson RT, Steinthorsdottir V *et al.* Genetic factors contribute to the risk of developing endometriosis. *Hum. Reprod.* 17, 555–559 (2002).

4 Kennedy S, Mardon H, Barlow D. Familial endometriosis. *J. Assist. Reprod. Genet.* 12(1), 32–34 (1995).

5 Eskenazi B, Warner ML. Epidemiology of endometriosis. *Obstet. Gynecol. Clin. North Am.* 24(2), 235–58 (1997).

6 Hadfield RM, Mardon HJ, Barlow DH *et al.* Endometriosis in monozygotic twins. *Fertil. Steril.* 68, 941–942 (1997).

7 Treloar SA, O'Connor DT, O'Connor VM *et al.* Genetic influences on endometriosis in an Australian twin sample. *Fertil. Steril.* 71, 701–710 (1999).

8 Zondervan KT, Weeks DE, Colman R *et al.* Familial aggregation of endometriosis in a large pedigree of rhesus macaques. *Hum. Reprod.* 19, 448–455 (2004).

9 Montgomery GW, Nyholt DR, Zhao ZZ *et al.* The search for genes contributing to endometriosis risk. *Hum. Reprod. Update* 14(5), 447–457 (2008).

10 Rahmioglu N, Missmer SA, Montgomery GW, Zondervan KT. Insights into assessing the genetics of endometriosis. *Curr. Obstet. Gynecol. Rep.* 1, 124–137 (2012).

11 Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case–control association studies. *Nat. Protoc.* 2(10), 2492–2501 (2007).

12 Gloria-Bottini F, Ammendola M, Saccucci P, Pietropolli A, Magrini A, Bottini E. The association of *PTPN22* polymorphism with endometriosis: effect of genetic and clinical factors. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 169(1), 60–63 (2013).

13 Sahmani M, Ghaleh TD, Darabi M, Darabi M, Rashvand Z, Najafipour R. I405V polymorphism of *CETP* gene and lipid profile in women with endometriosis. *Gynecol. Endocrinol.* 29(7), 712–715 (2013).

14 Szczepańska M, Wirstlein P, Skrzypczak J, Jagodziński PP. Polymorphic variants of *CYP17* and *CYP19A* and risk of infertility in endometriosis. *Acta Obstet. Gynecol. Scand.* 92(10), 1188–1193 (2013)

15 Bilibio JP, Matte U, de Conto E, Genro VK, Souza CA, Cunha-Filho JS. Dopamine receptor D2 genotype (3438) is associated with moderate/severe endometriosis in infertile women in Brazil. *Fertil. Steril.* 99(5), 1340–1345 (2013).

16 Wu ZY, Wang WW, Wang T *et al.* Association of *FOXP3* gene polymorphism in Chinese women with endometriosis. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 30(1), 106–110 (2013).

17 Frare AB, Barbosa AM, Costa IR *et al. GSTM1* and *GSTT1* polymorphisms in endometriosis in women from Goiás, Brazil. *Genet. Mol. Res.* 12(3), 2764–2770 (2013).

18 Fan W, Li S, Chen Q, Huang Z, Ma Q, Xiao Z. Association between interleukin-10 promoter polymorphisms and endometriosis: a meta-analysis. *Gene* 515(1), 49–55 (2013)

19 Saliminejad K, Memariani T, Ardekani AM *et al.* Association study of the TNF-α -1031T/C and VEGF + 450G/C polymorphisms with susceptibility to endometriosis. *Gynecol. Endocrinol.* 29(11), 974–977 (2013).

20 Li YZ, Wang LJ, Li X *et al.* Vascular endothelial growth factor gene polymorphisms contribute to the risk of endometriosis: an updated systematic review and meta-analysis of 14 case–control studies. *Genet. Mol. Res.* 12(2), 1035–1044 (2013).

21 Vanaja MC, Rozati R, Nassaruddin K, Vishnupriya S. Association of *VEGF* +405G>C polymorphism with endometriosis. *Front. Biosci. (Elite Ed.)* 5, 748–754 (2013).

22 Kang S, Shi YY, Li Y *et al.* Association between genetic variants of the *VEGFR-2* gene and the risk of developing endometriosis in northern Chinese women. *Gynecol. Obstet. Invest.* 76(1), 32–37 (2013).

23 Lin J, Zong L, Kennedy SH, Zondervan KT. Coding regions of *INHBA*, *SFRP4* and *HOXA10* are not implicated in familial endometriosis linked to chromosome 7p13–15. *Mol. Hum. Reprod.* 17, 605–611 (2011).

24 Treloar SA, Zhao ZZ, Le L *et al.* Variants in *EMX2* and *PTEN* do not contribute to risk of endometriosis. *Mol. Hum. Reprod.* 13, 587–594 (2007).

25 Zhao ZZ, Nyholt DR, Le L, Treloar SA, Montgomery GW. Common variation in the *CYP17A1* and *IFIT1* genes on chromosome 10 does not contribute to the risk of endometriosis. *Open Reprod. Sci. J.* 1, 35–40 (2008).

26 Lander ES, Daly MJ. Multipoint linkage analysis of the cystic fibrosis region. *Prog. Clin. Biol. Res.* 329, 47–52 (1989).

27 Volkers WS, Went LN, Vegter-Van Der Vlis M, Harper PS, Caro A. Genetic linkage studies in Huntington's chorea. *Ann. Hum. Genet.* 44, 75–79 (1980).

28 Painter JN, Nyholt DR, Morris A *et al.* High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to *CYP2C19*. *Fertil. Steril.* 95, 2236–2240 (2011).

29 Cribb AE, Knight MJ, Dryer D *et al.* Role of polymorphic human CYP450 enzymes in estrone oxidation. *Cancer Epidemiol. Biomarkers Prev.* 15, 551–558 (2006).

30 Lin J, Zong L, Kennedy SH *et al.* Coding regions of *INHBA*, *SFRP4* and *HOXA10* are not implicated in familial endometriosis linked to chromosome 7p13–15. *Mol. Hum. Reprod.* 17, 605–611 (2011).

31 The International HapMap C. A haplotype map of the human genome. *Nature* 437, 1299–1320 (2005)

32 Uno S, Zembutsu H, Hirasawa A *et al.* A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat. Genet.* 42, 707–710 (2010).

33 Treloar SA, Wicks J, Nyholt DR *et al.* Genome-wide linkage study in 1,176 affected sister pair families identifies significant susceptibility locus for endometriosis on chromosome 20q26. *Am. J. Hum. Genet.* 77, 365–376 (2005).

34 Zondervan KT, Treloar SA, Lin J *et al.* Significant evidence of one or more susceptibility loci for endometriosis with near-Mendelian inheritance on chromosome 7p13–15. *Hum. Reprod.* 22, 717–728 (2007).

35 National Human Genome Research Institute. www.genome.gov/GWAStudies

36 Adachi S, Tajima A, Quan J *et al.* Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. *J. Hum. Genet.* 55, 816–821 (2010).

37 Painter JN, Anderson CA, Nyholt DR *et al.* Genome-wide association study identifies a locus at 7p15.2 associated with the development of endometriosis. *Nat. Genet.* 43, 51–54 (2011).

38 Nyholt DR, Low SK, Anderson CA *et al.* Genome-wide association meta-analysis identified new endometriosis risk loci. *Nat. Genet.* 44(12), 1355–1359 (2012).

39 Albertsen HM, Chettier R, Farrington P, Ward K. Genome-wide association study link novel loci to endometriosis. *PLoS ONE* 8, e58257 (2013).

40 Rahmioglu N, Nyholt DR, Morris AP, Missmer SA, Montgomery GW, Zondervan KT. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. *Hum. Reprod. Update* 20(5), 702–716 (2014).

41 Vainio S, Heikkila M, Kispert A, Chin N, McMahon AP. Female development in mammals is regulated by *Wnt-4* signalling. *Nature* 397, 405–409 (1999).

42 Gaetje R, Holtrich U, Engels K *et al.* Endometriosis may be generated by mimicking the ontogenetic development of the female genital tract. *Fertil. Steril.* 87, 651–656 (2007).

43 Pellegrini C, Gori I, Achtari C *et al.* The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. *Fertil. Steril.* 98, 1200–1208 (2012).

44 Borowski A, Dirksen U, Lixin L, Shi RL, Göbel U, Schneider EM. Structure and function of ETAA16: a novel cell surface antigen in Ewing's tumours. *Cancer Immunol. Immunother.* 55, 363–374 (2006).

45 Pankov R, Yamada KM. Fibronectin at a glance. *J. Cell Sci.* 115, 3861–3863 (2002).

46 Lou X, Han X, Jin C *et al. SOX2* targets fibronectin 1 to promote cell migration and invasion in ovarian cancer: new molecular leads for therapeutic intervention. *OMICS* 17(10), 510–518 (2013).

47 Verschuur-Maes AH, de Bruin PC, van Diest PJ. Epigenetic progression of columnar cell lesions of the breast to invasive breast cancer. *Breast Cancer Res. Treat.* 136, 705–715 (2012).

48 Ren Y, Cheung HW, von Maltzhan G *et al.* Targeted tumor-penetrating siRNA nanocomplexes for credentialing the ovarian cancer oncogene *ID4*. *Sci. Transl. Med.* 4, 147ra112 (2012).

49 Chevillard G, Blank V. NFE2L3 (NRF3): the Cinderella of the Cap'n'Collar transcription factors. *Cell. Mol. Life Sci.* 68, 3337–3348 (2011).

50 Rhee DK, Park SH, Jang YK. Molecular signatures associated with transformation and progression to breast cancer in the isogenic MCF10 model. *Genomics* 92, 419–428 (2008).

51 Almstrup K, Leffers H, Lothe RA *et al.* Improved gene expression signature of testicular carcinoma *in situ*. *Int. J. Androl.* 30, 292–302 (2007).

52 Wang Y, van der Zee M, Fodde R, Blok LJ. Wnt/Beta-catenin and sex hormone signaling in endometrial homeostasis and cancer. *Oncotarget* 1, 674–684 (2010).

53 Matsuzaki S, Darcha C. *In vitro* effects of a small-molecule antagonist of the Tcf/β-catenin complex on endometrial and endometriotic cells of patients with endometriosis. *PLoS ONE* 8, e61690 (2013).

54 Kobayashi A, Behringer RR. Developmental genetics of the female reproductive tract in mammals. *Nat. Rev. Genet.* 4, 969–980 (2003).

55 Martini M, Ciccarone M, Garganese G *et al.* Possible involvement of hMLH1, p16INK4a and PTEN in the malignant transformation of endometriosis. *Int. J. Cancer* 102, 398–406 (2002).

56 Miao R, Guo X, Zhi Q *et al.* VEZT, a novel putative tumor suppressor, suppresses the growth and tumorigenicity of gastric cancer. *PLoS ONE* 8, e74409 (2013).

57 WERF Endometriosis Phenome and Biobanking Harmonisation Project (EPHect). www.endometriosisfoundation.org/ephect

58 Michailidou K, Hall P, Gonzalez-Neira A *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* 45(4), 353–361 (2013).

59 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24 (2012).

60 Story L, Kennedy S. Animal studies in endometriosis: a review. *ILAR J.* 45(2), 132–138 (2004).

61 Zondervan K, Cardon L, Desrosiers R *et al.* The genetic epidemiology of spontaneous endometriosis in the rhesus monkey. *Ann. N. Y. Acad. Sci.* 955, 233–238 (2002).