

# Genetic Biomarkers for Endometriosis

Sang Hong Lee, Yadav Sapkota, Jenny Fung, and Grant W. Montgomery

**Abstract** GWAS studies identified seven genomic regions with robust evidence for genome-wide significant association with endometriosis risk. One important question that arises is whether these genetic markers can be used to predict risk of developing endometriosis for individual women. As with most complex diseases, the effect sizes for genetic markers linked to endometriosis risk are small with odds ratios less than 1.3. If we combine information from all seven markers, we explain only 1.85% of the total phenotypic variance on the liability scale (assuming a population prevalence of endometriosis of 8%) with no predictive power for individual risk.

To explore the ability of all common genetic markers to predict endometriosis risk in individuals, we conducted simulations to quantify how useful endometriosis risk prediction is given current parameters. Applying our estimate of heritability ( $h^2 = 0.26$ ) from all common SNPs and assuming data were available from ~30,000 endometriosis cases, the proportion of variance explained by the risk predictor is still only ~0.08. To improve this prediction would require a far greater sample size. Current data may be useful for population-based stratification into risk categories. This can have applications in some cases such as improved efficiency of screening in breast cancer. In the future, risk prediction for endometriosis might be improved through combining genetic risk scores with clinical data, estimates of environmental effects such as DNA methylation signals, and/or better understanding of disease subtypes.

**Keywords** Endometriosis • Genetic biomarkers • Risk • Prediction • Simulation

---

S.H. Lee

School of Environmental and Rural Science, The University of New England,  
Armidale, NSW 2351, Australia

Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, The  
University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

Y. Sapkota

Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital,  
Memphis, TN 38105, USA

J. Fung • G.W. Montgomery (✉)

The Institute for Molecular Bioscience, University of Queensland,  
Brisbane, QLD 4072, Australia

e-mail: [g.montgomery1@uq.edu.au](mailto:g.montgomery1@uq.edu.au)

## Introduction

Endometriosis is a complex disorder influenced by multiple genetic and environmental factors. The genetic contributions to endometriosis risk are well documented, and several studies show that disease risk is higher among the relatives of endometriosis cases compared to controls in both hospital- [1, 2] and population-based [3] samples. This is further supported by twin studies showing an increased concordance in monozygotic when compared to dizygotic twins [4, 5] with the strongest evidence for genes influencing endometriosis from large-scale studies in twins [6] and in the Icelandic population [3].

Once the role for genetic variation was established, research efforts were directed toward identifying the genetic factors responsible. A large number of “candidate gene” studies have been published looking for association between endometriosis risk and genetic markers within biologically plausible candidate genes. In general, these studies were not successful with few results replicated or supported by later genome-wide association studies (GWAS) [7, 8]. Possible reasons for this have been reviewed elsewhere [7, 9]. Genetic linkage studies have also reported genomic regions that might harbor genetic variants increasing risk for endometriosis [10, 11], but no genes within these regions show significant association with disease risk.

In the last 5 years, the focus has shifted to large GWAS projects employing high throughput methods to genotype many thousand representative common genetic variants. This approach has revolutionized gene discovery for a wide spectrum of complex traits. Several GWAS studies for endometriosis have been published and results reviewed recently [8]. Seven genomic regions show genome-wide association with endometriosis with robust evidence across different populations and ethnic groups.

Given robust association between genetic markers and endometriosis risk, can these genetic markers be used to predict risk of developing endometriosis for individual women? Unfortunately, as with most complex diseases, the effect sizes for these genetic markers linked to increased endometriosis risk are small with odds ratios less than 1.3. Consequently, markers with robust evidence for association provide little power to predict a woman’s risk of disease. More recently, methods have been developed to use genome-wide SNP genotype data for prediction. In this chapter we discuss genomic regions associated with endometriosis identified from GWAS studies and discuss prediction of individual risk from the associated markers and from genome-wide SNP prediction.

## Genomic Regions Associated with Endometriosis Risk

Genome-wide association results have been published from four studies [12–15] and a meta-analysis [16] of summary data from the International Endogene Consortium study and the larger Japanese study. In addition, replication studies for

some of the key SNPs identified in the GWAS studies have been published [17–19]. There is excellent agreement across all studies for the major regions implicated in endometriosis risk [8]. Six regions showed evidence for genome-wide significant association in all cases, severe cases or both groups, and results for the region around fibronectin 1 (*FNI*) are close to genome-wide significance with strongest evidence in severe cases [8]. Recently, a meta-analysis of imputed data from Nyholt et al. [16] and the published results from Adachi et al. [12] confirmed association between endometriosis risk and SNPs in the region of interleukin 1A (*IL1A*) on chromosome 2 [20] adding a further important region for follow-up.

## Heritability

Heritability ( $h^2$ ) is an estimate of the proportion of variation in disease risk due to genetic factors. Traditionally this was estimated from similarities and differences in risk for relatives. One widely used design is to consider disease risks between pairs of identical and nonidentical twins. If the risk is higher for identical twins, this is evidence for a genetic contribution to disease risk since identical twins share 100% of their genome while nonidentical twins share only 50%. Using the classical twin design in a large sample of Australian twins, heritability for endometriosis risk was estimated at ~50% [6]. The remaining 50% of risk is due to other factors including environmental influences.

Studies trying to dissect the genetic and nongenetic causes using familial aggregation studies based on phenotypic observations alone must make explicit assumptions about shared environmental influences that are difficult to exclude entirely. More recently, whole-genome genotyping through GWAS provides an alternative method to estimate genetic contributions to disease risk independent of assumptions about shared environment necessary in family-based designs. We have used this method to estimate the genetic contribution to endometriosis from common genetic markers, sometimes called the SNP-heritability. After standard QC, the number of samples and SNPs used for estimating the genetic variance was 10,135 individuals (3154 cases and 6981 controls) with ~500,000 common SNPs [21]. We estimated that SNP-heritability on the liability scale was 0.26 (SE 0.04) assuming the population prevalence is 0.08 [21].

The difference between the SNP-heritability and heritability estimated from twin studies is likely due to several factors including uncertainty in the heritability from twin studies, the 500,000 common variants do not capture all the contributions from many rare variants, gene x environmental effects were not properly modelled, and possible heterogeneity from combining all endometriosis cases with different levels of severity and presentation. This difference between twin estimates of heritability and SNP-heritability is a general phenomenon for complex diseases. Many studies are now investigating those factors that can explain the discrepancy between the SNP-heritability and heritability. For the present study, the estimate of SNP-heritability

sets the practical upper limit for the ability of genetic markers to predict risk of disease. For endometriosis, this is ~25% of the variation and requires using data from genome-wide genotyping.

## Genetic Architecture of Endometriosis

The genetic architecture for a disease or trait is defined as the number of loci affecting the trait, the distribution of effect sizes, interactions between the genes or loci, and interactions with the environment [22]. GWAS results provide strong evidence for genomic regions associated with endometriosis risk. However, association results must pass stringent thresholds for significance and be replicated in independent studies before risk variants are accepted as contributing to disease risk. Only a few of the “top hits” meet these criteria in most genome-wide studies. Many other variants lie just below the threshold. A proportion of these markers will be “truly” associated with disease, but cannot be distinguished from the other false positive signals.

Larger studies help to discover more of the risk variants, but the application of multivariate statistical approaches to the entire marker data set can also be used in other important ways to understand the nature of genetic contributions to disease risk. Genetic risk prediction (GRP) methods make use of the aggregate effects of many genetic variants where one data set serves as discovery sample, with association tested in a target set [23]. Variants of small effect (e.g., with genotype relative risk of 1.05) are unlikely to achieve even nominal significance in a GWAS analysis; however, increasing proportions of true effects will be detected at increasingly liberal  $p$ -value thresholds. In the discovery sample, sets of allele-specific scores are selected for SNPs with the different levels of significance (e.g.,  $P < 0.01, 0.05, 0.1, 0.2$ , etc.). Genetic risk scores for individuals in the target set are then calculated as the sum of the copies of risk alleles for that individual in the target set weighted by the allelic effects (log odds ratio) estimated from the discovery set. The term risk score is used instead of risk, as it is impossible to differentiate the minority of true risk alleles from the nonassociated markers.

## Applications of GWAS Data Beyond the Top Hits

Genetic profiles can be used in important ways to investigate the genetic architecture of endometriosis. Our results show that analyses of all SNPs in the endometriosis GWAS data sets provide powerful approaches to investigate subgroups of endometriosis and understand shared genetic contributions across studies [14, 16, 24].

It is often difficult to determine the relationship between disease classes with strongly overlapping symptoms. In genetic studies of endometriosis, the Revised American Fertility Society (rAFS) classification system is commonly used to stage disease severity and assigns patients to one of four stages (I–IV) on the basis of the

extent of the disease and the associated adhesions present [7, 8]. Other classification systems have been proposed including ovarian vs. peritoneal disease and deep infiltrating vs. superficial disease. Whether these subclasses represent the natural history of one disorder, or are in fact different disease subtypes, is an important consideration in endometriosis research. Analysis of genome-wide marker data can assess the genetic contribution to individual disease subclasses and also the shared genetic contribution to each subclass providing new insights into the different disease presentations [24].

We have applied genetic risk prediction methods to show a stronger genetic contribution to severe disease compared with minimal/mild cases of endometriosis [14]. Further analysis of different disease classes [24] confirms the stronger genetic association with severe disease. In addition, mild forms of the disease in the discovery sample predict milder forms of disease in the target sample, but not more severe forms. Larger samples will be needed to confirm this result, but the data suggest distinct genetic contributions to mild forms of the disease. Similar methods also show strong genetic overlap for endometriosis cases in both European and Japanese populations [16].

Taken together, results from the GWAS, estimates of SNP-heritability, and polygenic prediction methods demonstrate that genetic contributions to endometriosis risk are due to a large number of common variants each with small effects. No common variants with large effects have been detected. Genome-wide significant “hits” all have small effects (odds ratios  $<1.3$ ), and many more genetic variants affecting disease risk remain to be discovered.

## Risk Prediction for Endometriosis

As noted above, the individual risks conferred by markers showing genome-wide significant association with endometriosis are low and do not help with prediction of individual risk. Even if we combine information from loci discovered from GWAS, they still have poor predictive power to discriminate individual risk. We combined results for the seven genome-wide significant loci from data on the reference allele frequencies and odds ratio from meta-analysis of the combined Australian and UK samples including 3181 case and 8075 controls [16]. Using a liability threshold model [25], the variance explained by the seven SNPs was 1.85% of the total phenotypic variance estimated on the liability scale assuming a population prevalence of 8%.

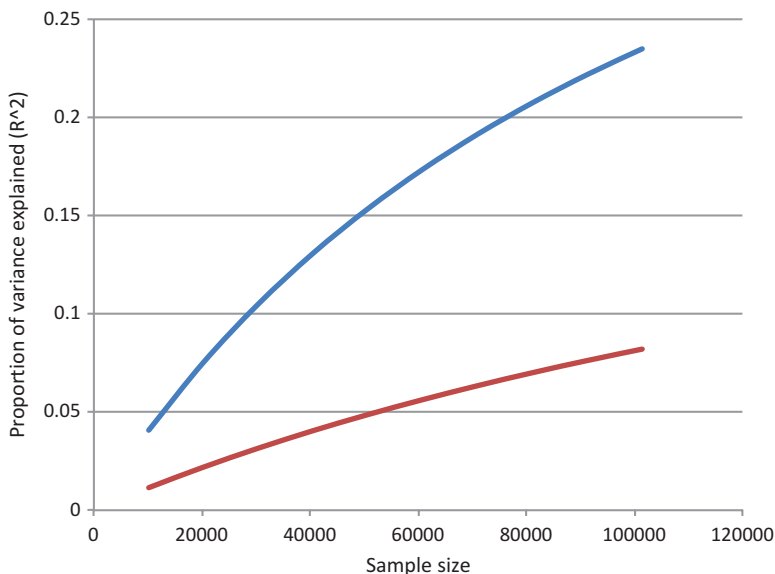
To explore the ability of all common genetic markers to predict endometriosis risk in individuals, we conducted simulations to quantify how useful endometriosis risk prediction is given current parameters. In this case, data from a large discovery sample are used to rank markers positively associated with endometriosis risk and develop a marker set which, when the markers are genotyped in an individual, would provide some prediction of disease risk. The accuracy of the prediction depends on a number of parameters and is strongly influenced by the size of the discovery sample.

## Prediction Accuracy and Sample Size

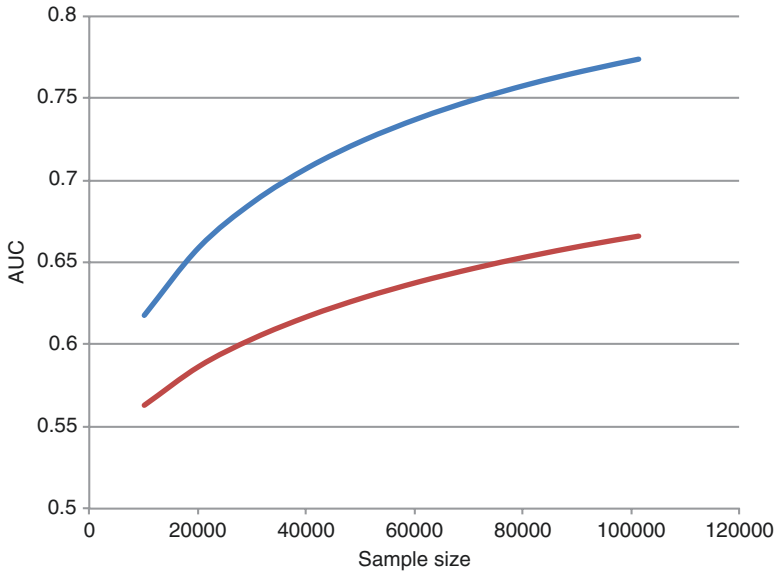
Using recent results on prediction accuracy of polygenic scores derived from quantitative genetic theory [26, 27], we quantified the relationship between sample size of the discovery sample and prediction accuracy. We assumed that endometriosis was polygenic [14, 16], the population prevalence was 0.08, and heritability on the liability scale was either 0.26 [21] based on SNP-heritability or 0.5 [6] from twin studies. We further assumed that the proportion of cases in the discovery sample was ~30% (~twice the number of controls compared with the number of cases) and 8% for validation set (i.e., population sample). The effective number of SNPs was assumed to be 50,000 [28].

Results show that when the heritability is  $h^2 = 0.26$  (Fig. 1), the proportion of variance explained by the risk predictor is ~0.08 even with 101,350 individuals (31,540 cases and 69,810 controls). However, the same proportion of variance can be achieved with ~20,270 individuals when heritability is  $h^2 = 0.5$ . A similar pattern is observed for the area under the curve (AUC; Fig. 2). An AUC of 0.65 requires 101,350 individuals with  $h^2 = 0.26$ , but requires only 20,270 individuals with  $h^2 = 0.5$ .

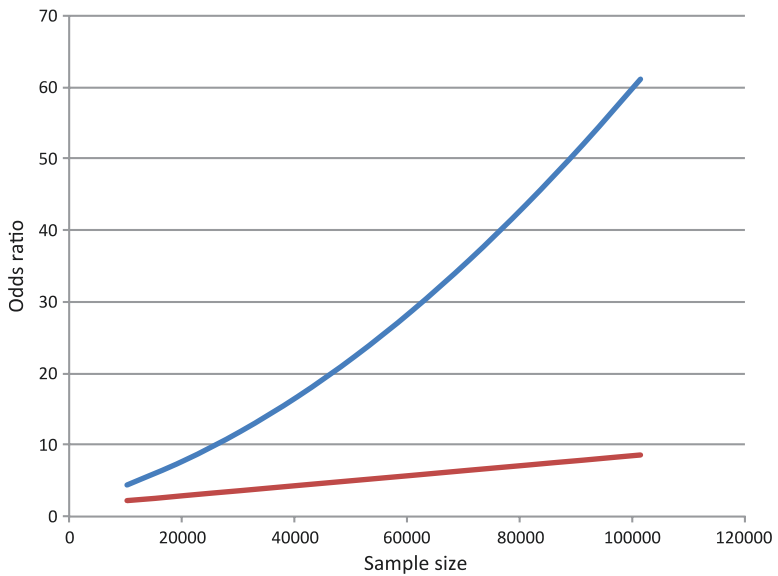
Following a common epidemiological approach to assess a continuous risk factor [23, 29], individuals were stratified into deciles according to the ranked values of the genetic risk predictors. We quantify the odds ratio of case-control status by contrasting the top decile to the lowest decile (Fig. 3). This approach is powerful even with a relatively small discovery sample, indicating this may be a valuable tool to stratify a heterogeneous population into groups.



**Fig. 1** The proportion of variance in endometriosis risk explained ( $R^2$ ) is plotted against sample size for the discovery sample. The *red line* assumes  $h^2 = 0.26$ , and the *blue line* assumes  $h^2 = 0.5$



**Fig. 2** Area under ROC (receiver operating characteristic) curve (AUC) plotted against sample size. The red line assumes  $h^2 = 0.26$ , and the blue line assumes  $h^2 = 0.5$



**Fig. 3** Odds ratios of individuals stratified into deciles based on genetic risk predictors in validation data set, using the decile with the lowest risk as the baseline, plotted against sample size. Red line with  $h^2 = 0.26$  and blue line with  $h^2 = 0.5$

## Summary and Future Directions

Current genome-wide significant “hits” provide no power for prediction of endometriosis risk for individual women. Studies of the genetic architecture of endometriosis and comparison with other complex diseases show that the genetic contribution to endometriosis is due to a large number of genetic variants each with small effects. The genome-wide significant “hits” represent only those markers that pass the stringent threshold required to account for the multiple testing required in GWAS analyses. Many other variants will be associated with endometriosis risk among the top SNPs that do not exceed the threshold and can provide useful information for prediction.

As we show in our simulations, the precision of genetic risk predictors constructed from a discovery sample depends on the size of the discovery sample and heritability of the disease. A very large discovery sample will be necessary to develop genetic risk scores with any accuracy for prediction. The meta-analysis of the International Endogene Consortium and Japanese GWAS studies analysed 4604 endometriosis cases. A new consortium of international groups is assembling data for ~17,000 cases and a large number of controls. Even with a discovery sample of this size, genetic risk predictors will still only explain a small proportion of variance in disease risk at SNP-heritability of 0.26.

Future developments may improve the prospects for including genetic markers in predictive tests for endometriosis risk. Risk prediction is an active area of research and a number of groups are working on ways to improve prediction estimates [30]. Although genetic markers do not provide accurate estimates for individual risk prediction, we show that current estimates may still be useful for population-based stratification into risk categories. This approach is being considered in breast cancer screening where including risk scores could change the current recommendations based on age [31]. Inclusion of risk scores could allow younger women with equivalent absolute risk to benefit from screening while decreasing by ~25% the proportion of women in current age groups where screening is considered useful [31].

Risk estimates from genetic marker data could be combined with clinical information to improve prediction. In breast cancer, addition of risk estimates from marker data for seven loci gave a small improvement in risk prediction based on family history, reproductive information, environment, and lifestyle factors [30]. Another consideration is that the current GWAS “hits” are unlikely to be the functional variants [32]. Identifying the true functional variants at each locus may improve the accuracy of risk estimates.

Risk prediction may vary across different disease subtypes. Current GWAS studies in endometriosis include cases of clinically diagnosed and self-reported disease and are combined across all disease stages. We have shown that the genetic architecture may differ between mild and severe forms of the disease [24]. If this is the case, separation of cases into meaningful subtypes may improve the precision of risk predictors within subtypes. However, very large studies will be necessary to achieve appropriate power for the different subtypes.



Endometriosis is influenced by genetic variation and also by environmental factors. One promising approach being used in other complex traits is the study of genome-wide methylation signals [33, 34]. Methylation signals are themselves influenced by genetic variation [32], but they also capture past and present environmental effects [34]. As we have seen, the accuracy of risk prediction depends on the disease heritability. Risk prediction in endometriosis would be improved if the heritability explained by genetic markers was nearer to the estimate ( $h^2 = 0.5$ ) from twin studies [6]. Even if we can account for all of the genetic variation, this still leaves half of the variance in endometriosis risk unexplained.

One approach we are following up is whether genome-wide methylation can capture some of the environmental influence and be used to improve disease prediction. Using similar approaches we have evaluated combining genetic risk scores and methylation risk scores for prediction in studies on body mass index (BMI) and height [35]. BMI has modest heritability and is influenced by environment, while height has very high heritability. Combining risk scores from GWAS and methylation substantially increases prediction for BMI but does not improve prediction for height.

In conclusion, genetic variants associated with endometriosis risk do not provide useful markers to predict individual risk for endometriosis, whether restricting the markers to genome-wide significant results or combining data into polygenic risk scores. Much larger genetic studies will be required to approach useful prediction. There are promising developments to improve prediction through combining genetic data with other data. This includes clinical data and predictors from genome-wide methylation signals. Further studies will be required to determine if these approaches are useful for endometriosis risk prediction.

**Acknowledgments** GWM is an NHMRC Principal Research Fellow (1078399), SHL is an ARC Future Fellow (FT160100229), and the work was supported by NHMRC project grants (1026033, 1049472, 1080157).

## References

1. Kennedy S, Mardon H, Barlow D. Familial endometriosis. *J Assist Reprod Genet.* 1995;12(1):32–4.
2. Simpson JL, Bischoff FZ. Heritability and molecular genetic studies of endometriosis. *Ann N Y Acad Sci.* 2002;955:239–51.
3. Stefansson H, Geirsson RT, Steinthorsdottir V, Jonsson H, Manolescu A, Kong A, et al. Genetic factors contribute to the risk of developing endometriosis. *Hum Reprod.* 2002;17(3):555–9.
4. Hadfield RM, Mardon HJ, Barlow DH, Kennedy SH. Endometriosis in monozygotic twins. *Fertil Steril.* 1997;68(5):941–2.
5. Moen MH. Endometriosis in monozygotic twins. *Acta Obstet Gynecol Scand.* 1994;73(1):59–62.
6. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. *Fertil Steril.* 1999;71(4):701–10.
7. Montgomery GW, Nyholt DR, Zhao ZZ, Treloar SA, Painter JN, Missmer SA, et al. The search for genes contributing to endometriosis risk. *Hum Reprod Update.* 2008;14:447–57.

8. 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*. 2014;20(5):702–16.
9. Zondervan KT, Cardon LR, Kennedy SH. What makes a good case-control study? Design issues for complex traits such as endometriosis. *Hum Reprod*. 2002;17(6):1415–23.
10. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, et al. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. *Am J Hum Genet*. 2005;77(3):365–76.
11. Zondervan KT, Treloar SA, Lin J, Weeks DE, Nyholt DR, Mangion J, et al. Significant evidence of one or more susceptibility loci for endometriosis with near-Mendelian autosomal inheritance on chromosome 7p13-15. *Hum Reprod*. 2007;22:717–28.
12. Adachi S, Tajima A, Quan J, Haino K, Yoshihara K, Masuzaki H, et al. Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. *J Hum Genet*. 2010;55(12):816–21.
13. Albertsen HM, Chettier R, Farrington P, Ward K. Genome-wide association study link novel loci to endometriosis. *PLoS One*. 2013;8(3):e58257.
14. Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, Lee SH, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet*. 2011;43(1):51–4.
15. Uno S, Zembutsu H, Hirasawa A, Takahashi A, Kubo M, Akahane T, et al. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat Genet*. 2010;42(8):707–10.
16. Nyholt DR, Low SK, Anderson CA, Painter JN, Uno S, Morris AP, et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet*. 2012;44(12):1355–9.
17. Pagliardini L, Gentilini D, Sanchez AM, Candiani M, Vigano P, Di Blasio AM. Replication and meta-analysis of previous genome-wide association studies confirm vezatin as the locus with the strongest evidence for association with endometriosis. *Hum Reprod*. 2015;30(4):987–93.
18. Pagliardini L, Gentilini D, Vigano P, Panina-Bordignon P, Busacca M, Candiani M, et al. An Italian association study and meta-analysis with previous GWAS confirm WNT4, CDKN2BAS and FN1 as the first identified susceptibility loci for endometriosis. *J Med Genet*. 2013;50(1):43–6.
19. Sundqvist J, Xu H, Vodolazkaia A, Fassbender A, Kyama C, Bokor A, et al. Replication of endometriosis-associated single-nucleotide polymorphisms from genome-wide association studies in a Caucasian population. *Hum Reprod*. 2013;28(3):835–9.
20. Sapkota Y, Low S-K, Adachi S, Attia J, Gordon SD, Henders AK, et al. Confirmed association between endometriosis and the interleukin 1A (IL1A) locus. *Hum Reprod*. 2015;30(1):239–48.
21. Lee SH, Harold D, Nyholt DR, ANZGene Consortium, International Endogene Consortium, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet*. 2013;22(4):832–41.
22. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011;187(2):367–83.
23. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–7.
24. Sapkota Y, Attia J, Gordon SD, Henders AK, Holliday EG, Rahmioglu N, et al. Genetic burden associated with varying degrees of disease severity in endometriosis. *Mol Hum Reprod*. 2015;21(7):594–602.
25. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol*. 2011;35(5):310–7.
26. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9(3):e1003348.
27. Lee SH, Wray NR. Novel genetic analysis for case-control genome-wide association studies: quantification of power and genomic prediction accuracy. *PLoS One*. 2013;8(8):e71494.

28. Visscher PM, Hemani G, Vinkhuyzen AAE, Chen G-B, Lee SH, Wray NR, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 2014;10(4):e1004269.
29. Maier R, Moser G, Chen GB, Ripke S, Cross-Disorder Working Group of the Psychiatric Genomics C, Coryell W, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet.* 2015;96(2):283–94.
30. Lee SH, Weerasinghe WM, Wray NR, Goddard ME, van der Werf JH. Using information of relatives in genomic prediction to apply effective stratified medicine. *Sci Rep.* 2017;7:42091.
31. Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr Opin Genet Dev.* 2015;30C:32–41.
32. Fung JN, Rogers PA, Montgomery GW. Identifying the biological basis of GWAS hits for endometriosis. *Biol Reprod.* 2015;92(4):87.
33. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* 2014;15(5):R73.
34. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014;24(11):1725–33.
35. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, Harris SE, Liewald D, Henders AK, Mendelson MM, Liu C, Joehanes R, Liang L, Levy D, Martin NG, Starr JM, Wijmenga C, Wray NR, Yang J, Montgomery GW, Franke L, Deary IJ, Visscher PM, Heijmans BT, 't Hoen PA, van Meurs J, Isaacs A, Jansen R, Franke L, Boomsma DI, Pool R, van Dongen J, Hottenga JJ, van Greevenbroek MM, Stehouwer CD, van der Kallen CJ, Schalkwijk CG, Wijmenga C, Zhernakova S, Tigchelaar EF, Slagboom PE, Beekman M, Deelen J, van Heemst D, Veldink JH, van den Berg LH, van Duijn CM, Hofman BA, Uitterlinden AG, Jhamai PM, Verbiest M, Suchiman HE, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Mei H, van Iterson M, van Galen M, Bot J, van 't Hof P, Deelen P, Nooren I, Moed M, Vermaat M, Zhernakova DV, Luijk R, Bonder MJ, van Dijk F, Arindrarto W, Kielbasa SM, Swertz MA, van Zwet EW. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet.* 2015;97(1):75–85.