

No part of this digital document may be reproduced, stored in a retrieval system or transmitted commercially in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

Chapter 13

Molecular Biology of Endometriosis

*Sosuke Adachi and Takayuki Enomoto**

Department of Obstetrics and Gynecology,
Niigata University Faculty of Medicine, Niigata Japan

Abstract

Endometriosis is considered a common yet complex disease, with many factors, including genetic and environmental, involved in its development. Evidence for an underlying genetic cause comes from the observation that the incidence in relatives of affected women is more common than in women without a family history of endometriosis. Although many genetic approaches, such as linkage analysis and candidate-gene-based case-control studies have been performed looking for this genetic link, none have yet been successful in identifying a replicable genetic factor for endometriosis. Like other common complex diseases, the hazard ratio of genetic factors contributing to endometriosis is estimated to be low. Recently, high throughput genome-wide association (GWA) analysis using high-density single nucleotide polymorphism (SNP) arrays has been performed to identify genetic susceptibility to a number of complex diseases, leading to many genetic markers for those diseases being identified. In this chapter, we review the many genetic factors contributing to endometriosis risk based on the findings of the very latest reports, and present an outline summary of GWA analyses conducted for endometriosis.

Endometriosis is an estrogen-dependent disorder observed in 7–10% of women of reproductive age and in up to 50% of such women suffering with infertility[1]. It is characterized by the presence of endometrial-like glands and stroma outside of the uterine cavity, primarily on the pelvic peritoneum and ovaries[2]. The main pathological processes associated with the disease are peritoneal inflammation and fibrosis, and the formation of pelvic adhesions and ovarian cysts[3]. The common complaint symptoms include dysmenorrhea, dyspareunia, noncyclic pelvic pain and infertility[4,5].

* Corresponding author: enomoto@med.niigata-u.ac.jp.

Endometriosis is a complex disease, arising from the interplay between multiple genetic and environmental factors[3,6]. The involvement of genetic factors in its development is strongly supported by different types of studies[7-11]. The incidence of endometriosis is approximately seven times higher in relatives of women with endometriosis than in those without a family history[12]. In addition, evidence of genetic risk of endometriosis has been suggested by studies of twins showing increased concordance rates of the development of endometriosis in monozygotic twins[1,13,14] compared with dizygotic twins. The heritability of human endometriosis is estimated to be 51% [2,10,11,13,14].

Numerous hypothesis-based candidate-gene studies have reported an association for markers in candidate genes and endometriosis susceptibility. In these studies, variants were tested from candidate genes involved in sex steroid biosynthesis and signaling pathways, estradiol metabolism, adhesion molecules and matrix enzymes, immunological and inflammatory pathways, growth factor systems, cell cycle regulation, oncogenes, apoptosis, and angiogenic factors[3,6,15]. For example, Yoshida et al. investigated in Japanese women the association between endometriosis and polymorphisms of the E-cadherin gene, a central player in epithelial-to-mesenchymal transition. Allele frequency analysis indicated that there was a marginally higher frequency of the rs4783689 C allele in women with endometriosis compared with controls (corrected $P=0.007$; OR=1.37; 95% CI, 1.14–1.64)[4,5,16]. However, the results of these hypothesis-based candidate gene studies have generally not been reproducible in follow-up replication studies[3,6]. This lack of success is likely due to a number of reasons, including differences in disease definitions, ethnic populations and the lack of power in small-scale studies, which detect only variants with large effects and are prone to detection of false-positive results[6-11,17]. The strength of the true effect is often overestimated in discovery datasets in an effect known as the ‘winner’s curse’[12,18]. Replication studies require a much larger sample size to detect a real association than do first studies.

The genetic components underlying endometriosis have been examined directly by a genome-wide linkage study with a large set of families. A total of 1,176 affected sib-pairs were recruited from Australian and United Kingdom families, giving the study sufficient statistical power (80%) to detect a locus with a sibling relative risk of 1.3 using nonparametric linkage analysis. The analysis of the families demonstrated a significant linkage to chromosome 10q26[19]. Taking a more stringent criteria of endometriosis families (248 families with 3 or more affected members), nonparametric and parametric linkage analysis identified a susceptibility region on 7p13-15 with near-Mendelian patterns of inheritance[20]. Follow-up analyses and a fine-mapping association study across the reported linkage peak on chromosome 10q26 region was conducted using 11,984 SNPs in 1,144 familial cases and 1,190 control patients[21]. Linkage analyses on families grouped by endometriosis symptoms (primarily subfertility) provided increased evidence for linkage (LOD score=3.62) near a previously reported linkage peak. This study identified three independent association signals at 96.59 Mb (rs11592737), 105.63 Mb (rs1253130), and 124.25 Mb (rs2250804). However, only linkage to rs11592737 (in the cytochrome P450 subfamily C (*CYP2C19*) gene) was replicated in an independent sample study of 2,079 cases and 7,060 population controls.

Two genome-wide association studies (GWAS)s of endometriosis have been published in women of Japanese ancestry[22,23] and one in women of European ancestry[24], and there has been one meta-analysis report of these GWASs[25]. The first Japanese GWAS included 1,423 endometriosis cases and 1,318 controls in the discovery datasets[22]. After the

application of SNP quality control, 460,945 SNPs were included in the first stage analysis. A replication analysis was conducted using an independent set of 484 cases and 3,974 controls. Then the top 100 SNPs that had most significant P values from the first set were genotyped. A significant association for one SNP was confirmed, that of rs10965235 in the *CDKN2BAS* gene on chromosome 9p21 ($P=6.79\times 10^{-6}$, OR=1.56, 95% CI, 1.29–1.89, Bonferroni-adjusted $P=4.89\times 10^{-4}$). A combined result of the discovery study and the replication study strongly supported the association of the locus on chromosome 9p21 (combined $P=5.57\times 10^{-12}$, OR=1.44, 95% CI 1.30–1.59). *CDKN2BAS* encodes a cyclin-dependent kinase inhibitor 2B antisense RNA. By fine mapping, the SNP showing the strongest association was located in intron 16 of *CDKN2BAS* and was implicated in regulating the expression of the p15, p16 and p14 tumor suppressor genes. In addition, a SNP, rs16826658, in the LD block including *WNT4* on chromosome 1p36, revealed a possible association with endometriosis ($P=1.66\times 10^{-6}$, OR=1.20, 95% CI 1.11–1.29)[22]. *WNT4* encodes the wntless-type MMTV integration site family member 4, which is considered to play an important role in the development of the female genital tract[26].

A second Japanese GWAS was a meta-analysis of two GWAS on two case-control datasets[23]. After quality control, 282,828 SNPs were tested for association in 696 patients with endometriosis and 825 controls. The meta-analysis revealed that a common susceptibility locus conferring a large effect on the disease risk was unlikely. On the other hand, an excess of SNPs with P -values $<10^{-4}$ was observed in the meta-analysis. Of note, four of the top five SNPs with p values $<10^{-5}$ were located in and around *IL1A* (interleukin 1a) on 2q13. IL-1a is a member of the interleukin-1 cytokine family, which binds to the IL-1 receptor 1, acts as an agonist for the receptor, and regulates production of other proinflammatory cytokines and chemokines that drive further inflammation[27]. In this second GWA analysis, SNP rs17761446, which is in perfect linkage disequilibrium with rs10965235 ($D'=1$, $r^2=1$, HapMap JPT population) reported by the first GWAS[22], showed a P -value= 8.9×10^{-3} , with a summary OR=1.29 (95% CI, 1.10–1.48), suggesting the existence of an endometriosis-susceptibility locus at 9p21.

The third GWAS (European) included 3,194 endometriosis cases and 7,060 controls from Australia and the UK[24]. They assessed disease stage from surgical records using the rAFS classification system and grouped the subjects into two phenotypes: stage A (stage I or II disease or some ovarian disease with a few adhesions; $n=1,686$, 52.7%) or stage B (stage III or IV disease; $n=1,364$, 42.7%), or unknown ($n=144$, 4.6%). After the SNP quality control, 504,723 SNPs were included in the analyses. They showed a significantly increased genetic loading among endometriosis cases with stage B compared to cases with stage A. The estimate for stage B endometriosis (0.34, SD=0.04) was significantly higher than that for stage A endometriosis (0.15, SD=0.04).

Subsequently, two GWAS were performed, using (i) 3,194 “all” endometriosis cases and (ii) 1,364 stage B cases. For “all” endometriosis, the strongest signal observed was rs12700667 in an intergenic region on chromosome 7p15.2 ($P=2.6\times 10^{-7}$, OR=1.22, 95% CI, 1.13–1.32). The 7p15.2 signal for stage B cases was considerably stronger, producing $P=1.5\times 10^{-9}$, OR=1.38, 95% CI, 1.24–1.53 for rs12700667 and $P=6.0\times 10^{-8}$, OR=1.34, 95% CI, 1.21–1.49 for the nearby SNP rs7798431 ($r^2=0.87$). A second strong association was found for rs1250248 (2q35) within *FNI* ($P=3.2\times 10^{-8}$). In the replication analysis, they genotyped 70 SNPs that produced nominal evidence of association with “all” ($P < 1.0\times 10^{-4}$) or stage B ($P < 1.0\times 10^{-4}$ in stage B and $P < 1.0\times 10^{-3}$ in “all” endometriosis cases) in an

independent dataset comprising 2,392 cases and 2,271 controls. Association with ‘all’ endometriosis for the two SNPs on 7p15.2 was replicated, with $P=1.2\times 10^{-3}$, OR=1.17, 95% CI, 1.06–1.28 for rs12700667 and $P=1.6\times 10^{-3}$, OR=1.17, 95% CI, 1.06–1.28 for rs7798431. There was no evidence (nominal $P\leq 0.05$) for replication of rs12540248 (*FNI*) or association with the remaining 70 SNPs. Analysis of all 5,586 cases and 9,331 controls from combined cohorts further confirmed association between “all” endometriosis and 7p15.2, producing $P=1.4\times 10^{-9}$, OR=1.20, 95% CI, 1.13–1.27 for rs12700667 and $P=1.1\times 10^{-7}$, OR=1.18, 95% CI, 1.11–1.25 for rs7798431.

The estimated percentage of “all” endometriosis variance explained by rs12700667 was 0.36, or 0.69% of the estimated 51% heritability of endometriosis[10]. The associated SNPs are located in a ~924-kb intergenic region containing at least one noncoding RNA (AK057379), predicted transcripts and regulatory elements, and a miRNA (hsa-mir-148a) ~88 kb upstream of rs12700667. They found no evidence for association with rs10965235 on chromosome 9p21 (rs10965235 is monomorphic in individuals of European descent, reflecting the different ancestral genetic backgrounds between the studies), nor with any SNPs in LD with rs10965235 reported by Japanese GWAS[22]. However, there was evidence for replication of rs7521902 on 1p36, close to *WNT4*, for both “all” endometriosis ($P=9.0\times 10^{-5}$, OR=1.16, 95% CI, 1.08–1.25) and stage B cases ($P=7.5\times 10^{-6}$, OR=1.25, 95% CI, 1.13–1.38). A meta-analysis of the “all” endometriosis OR with the reported Japanese OR of 1.25 (95% CI, 1.12–1.39) for rs7521902 produced a genome-wide significant P value of 4.2×10^{-8} (OR=1.19, 95% CI, 1.12–1.27).

Recently, a meta-analysis of two previously reported GWASs for endometriosis[22,24] was reported[25]. This analysis included 4,604 endometriosis cases and 9,393 controls of Japanese[22] and European[24] ancestry for 407,632 SNPs. An allele of rs12700667 at the 7p15.2 locus in individuals of European ancestry (OR=1.22, 95% CI, 1.13–1.31; $P=7.2\times 10^{-8}$) also replicates in the Japanese GWAS data (OR=1.22, 95% CI, 1.07–1.39; $P=3.6\times 10^{-3}$), producing an overall OR of 1.22 (95% CI, 1.14–1.30) and $P=9.3\times 10^{-10}$ in the GWAS meta-analysis.

They confirmed the association with allele A of rs7521902 at the 1p36.12 *WNT4* locus (OR=1.18, 95% CI, 1.11–1.25; $P=4.6\times 10^{-8}$). They confirmed association of rs13394619 in *GREB1* on 2p25.1 with endometriosis (OR=1.12, 95% CI=1.06–1.18; $P=2.1\times 10^{-5}$), previously reported in a second Japanese GWAS²³ (OR=1.35, 95% CI=1.17–1.56; $P=3.8\times 10^{-5}$). The association of rs13394619 approached conventional genome-wide significance ($P\leq 5\times 10^{-8}$) in combined analysis of the three GWASs datasets[22-24] (OR=1.15, 95% CI, 1.09–1.20; $P=6.1\times 10^{-8}$).

GREB1 encodes growth regulation by estrogen in *breast cancer 1*, an early response gene in the estrogen regulation pathway that is involved in hormone-dependent breast cancer cell growth²⁸. The GWAS meta-analysis identified a novel locus at 12q22 near *VEZT* (rs10859871: OR=1.18, 95% CI, 1.12–1.25; $P=5.5\times 10^{-9}$). *VEZT* encodes vezatin, an adherens junctions transmembrane protein that is downregulated in gastric cancer[29]. Given the substantially greater genetic loading of moderate-to-severe (stage B) endometriosis (rAFS stage III or IV disease) compared to minimal (stage A) endometriosis (rAFS stage I or II disease)[24], a secondary analysis was performed. Excluding endometriosis cases with minimal or unknown severity, GWAS meta-analysis implicated additional novel loci at 2p14 (rs4141819: OR=1.22, 95% CI, 1.14–1.32; $P=6.5\times 10^{-8}$), 6p22.3 (rs7739264 : OR=1.21, 95% CI, 1.13–1.30; $P=5.8\times 10^{-8}$) and 9p21.3 (rs1537377: OR=1.22, 95% CI, 1.14–1.30; $P=1.0\times 10^{-7}$

⁸). To further validate the seven SNPs (rs7521902, rs13394619, rs4141819, rs7739264, rs12700667, rs1537377, rs10859871) implicated by the meta-analysis, replication study was performed, including 1,044 cases and 4,017 controls obtained from a Japanese population independent of the first GWAS cohort[22]. All seven SNPs were replicated in an independent cohort and associated at $P < 5 \times 10^{-8}$ in a combined analysis. Polygenic prediction analysis revealed that the aggregate effects of many variants of small effect in the Japanese GWAS cohort could predict affected status in the European GWAS cohort ($P=8.8 \times 10^{-11}$). A gene-based genome-wide association analysis using the VEGAS (versatile gene-based association study) program[30] identified 1,184 genes with combined P of ≤ 0.05 and determined that the top 3 ranked genes associated with endometriosis were *WNT4*, at 1p36.12 ($P=5.0 \times 10^{-9}$), *VEZT*, at 12q22 ($P=5.7 \times 10^{-7}$), and *GREB1*, at 2p25.1 ($P=2.5 \times 10^{-5}$).

In conclusion, the meta-analysis of GWASs for endometriosis shows that *WNT4*, *VEZT* and *GREB1* may be strong susceptibility genes for endometriosis. These results also indicate that a considerable number of weakly-associated SNPs (for example, at $P < 0.1$) represent true endometriosis-risk loci in both Japanese and European populations. ORs for replicated, genome-wide significant signals are all lower than 1.5, suggesting that there are no common variants with large effects for endometriosis. Although the susceptible variants themselves have only small effects and explain only a very small proportion of the pathogenesis of endometriosis, they do provide important information for specialized diagnostic options (i.e. underlying pathways contributing to the disease) and suggest important approaches for the future treatment of endometriosis.

References

- [1] Giudice, L. C. (2010). Clinical practice. Endometriosis. *N Engl J Med*, 362, 2389–2398.
- [2] Bulun, S. E. Endometriosis. (2009). *N Engl J Med*, 360, 268–279.
- [3] Montgomery, G. W. (2008). *et al.* The search for genes contributing to endometriosis risk. *Hum. Reprod. Update*, 14, 447–457.
- [4] Giudice, L. C. & Kao, L. C. (2004). Endometriosis. *Lancet* 364, 1789–1799.
- [5] Berkley, K. J., Rapkin, A. J. & Papka, R. E. (2005). The pains of endometriosis. *Science*, 308, 1587–1589.
- [6] Rahmioglu, N., Missmer, S. A., Montgomery, G. W. & Zondervan, K. T. (2012). Insights into Assessing the Genetics of Endometriosis. *Curr Obstet Gynecol Rep*, 1, 124–137.
- [7] Kennedy, S., Mardon, H. & Barlow, D. (1995). Familial endometriosis. *J Assist Reprod Genet*, 12, 32–34.
- [8] Simpson, J. L. & Bischoff, F. Z. (2002). Heritability and molecular genetic studies of endometriosis. *Ann NY Acad Sci*, 955, 239–51; discussion, 293–5, 396–406.
- [9] Stefansson, H. *et al.* (2002). Genetic factors contribute to the risk of developing endometriosis. *Hum Reprod*, 17, 555–559.
- [10] Treloar, S. A., O'Connor, D. T., O'Connor, V. M. & Martin, N. G. (1999). Genetic influences on endometriosis in an Australian twin sample. sueT@qimr.edu.au. *Fertil Steril*, 71, 701–710.

-
- [11] Treloar, S. *et al.* (2002). The International Endogene Study: a collection of families for genetic research in endometriosis. *Fertil Steril*, 78, 679–685.
- [12] Simpson, J. L., Elias, S., Malinak, L. R. & Buttram, V. C. (1980). Heritable aspects of endometriosis. I. Genetic studies. *Am J Obstet Gynecol*, 137, 327–331 (1980).
- [13] Hadfield, R. M., Mardon, H. J., Barlow, D. H. & Kennedy, S. H. (1980). Endometriosis in monozygotic twins. *Fertil Steril*, 68, 941–942.
- [14] Moen, M. H. (1994). Endometriosis in monozygotic twins. *Acta Obstet Gynecol Scand*, 73, 59–62.
- [15] Falconer, H., D'Hooghe, T. & Fried, G. (2007). Endometriosis and genetic polymorphisms. *Obstet Gynecol Surv*, 62, 616–628.
- [16] Yoshida, K. *et al.* (2012). Possible involvement of the E-cadherin gene in genetic susceptibility to endometriosis. *Hum Reprod*, 27, 1685–1689.
- [17] Zondervan, K. T. & Cardon, L. R. (2007). Designing candidate gene and genome-wide case-control association studies. *Nat Protoc*, 2, 2492–2501.
- [18] Zollner, S. & Pritchard, J. K. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, 80, 605–615.
- [19] Treloar, S. A. *et al.* (2005). Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. *Am J Hum Genet*, 77, 365–376.
- [20] Zondervan, K. T. *et al.* (2007). Significant evidence of one or more susceptibility loci for endometriosis with near-Mendelian inheritance on chromosome 7p13-15. *Hum Reprod*, 22, 717–728.
- [21] Painter, J. N. *et al.* (2011). High-density fine-mapping of a chromosome 10q26 linkage peak suggests association between endometriosis and variants close to CYP2C19. *Fertil Steril*, 95, 2236–2240.
- [22] Uno, S. *et al.* (2010). A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat Genet*, 42, 707–710.
- [23] Adachi, S. *et al.* (2010). Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. *J Hum Genet*, 55, 816–821.
- [24] Painter, J. N. *et al.* (2011). Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet*, 43, 51–54.
- [25] Nyholt, D. R. *et al.* (2012). Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet*, 44, 1355–1359.
- [26] Gaetje, R., *et al.* (2007). Endometriosis may be generated by mimicking the ontogenetic development of the female genital tract. *Fertil Steril* 87, 651–656.
- [27] Fitzgerald, K. A. (2009). Integrating IL-1 alpha in antiviral host defenses. *Immunity*, 31, 7–9.
- [28] Rae, J. M. *et al.* (2005). GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res. Treat.*, 92, 141–149.
- [29] Boyer, A. *et al.* (2010). WNT4 is required for normal ovarian follicle development and female fertility. *FASEB J.*, 24, 3010–3025.
- [30] Liu, J. Z. *et al.* (2010). A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*, 87, 139–145.