

ORIGINAL ARTICLE

Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population

Sosuke Adachi¹, Atsushi Tajima², Jinhua Quan¹, Kazufumi Haino¹, Kosuke Yoshihara¹, Hideaki Masuzaki³, Hidetaka Katabuchi⁴, Kenichiro Ikuma^{5,12}, Hiroshi Suginami^{6,13}, Nao Nishida⁷, Ryoza Kuwano⁸, Yuji Okazaki⁹, Yoshiya Kawamura¹⁰, Tsukasa Sasaki¹¹, Katsushi Tokunaga⁷, Ituro Inoue² and Kenichi Tanaka¹

To identify susceptibility genes for endometriosis in Japanese women, genome-wide association (GWA) analysis was performed using two case–control cohorts genotyped with the Affymetrix Mapping 500K Array or Genome-Wide Human SNP Array 6.0. In each of the two array cohorts, stringent quality control (QC) filters were applied to newly obtained genotype data, together with previously analyzed data from the Japanese Integrated Database Project. After QC-based filtering of samples and single nucleotide polymorphisms (SNPs) in each cohort, 282 838 SNPs in both genotyping platforms were tested for association with endometriosis using a meta-analysis of the two GWA studies with 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 $< 10^{-4}$ (36 vs 28 SNPs expected by chance) 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 1 α), which might be a functional candidate gene for endometriosis. Further studies with larger case–control cohorts will be necessary to elucidate the genetic risk factors.

Journal of Human Genetics (2010) 55, 816–821; doi:10.1038/jhg.2010.118; published online 16 September 2010

Keywords: endometriosis; genome-wide association study; Japanese; meta-analysis; single nucleotide polymorphism

INTRODUCTION

Endometriosis is an estrogen-dependent disorder observed in 5–10% of women of reproductive age¹ and in 20–50% of women with infertility.² It is characterized by the presence of endometrial glands and stroma outside the uterine cavity, primarily on the pelvic peritoneum and ovaries.³ The main pathological processes associated with the disease are peritoneal inflammation and fibrosis, and the formation of pelvic adhesions and ovarian cysts.⁴ The common symptoms include dysmenorrhea, dyspareunia, noncyclic pelvic pain and infertility.^{1,5}

Endometriosis is a common complex disease caused by genetic and environmental factors including their interactions.⁴ There is ample

evidence that substantial components of genetic factors have an important role in the disease risk.^{6–10} Its incidence in relatives of affected women in USA is up to seven times the incidence in women without family histories of endometriosis.¹¹ The relative risk of endometriosis in female siblings of the case subjects is estimated to be 5.7 in Japanese.¹² In addition, evidence of genetic risk of endometriosis has been suggested by twin studies showing increased concordance rates of the development of endometriosis in monozygotic twins^{13,14} compared with dizygotic twins. The heritability of human endometriosis was estimated at 51%.^{9,10,13,14}

Genetic components underlying endometriosis have been examined directly by a genome-wide linkage study with a large set of families. A

¹Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; ²Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan; ³Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; ⁴Department of Reproductive Medicine and Surgery/Gynecology, Faculty of Medical and Pharmaceutical Science, Kumamoto University, Kumamoto, Japan; ⁵Department of Obstetrics and Gynecology, Takarazuka City Hospital, Takarazuka, Japan; ⁶Department of Obstetrics and Gynecology, National Hospital Organization Kyoto Medical Center, Kyoto, Japan; ⁷Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; ⁸Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; ⁹Department of Neurology, Tokyo Metropolitan Matsuzawa Hospital, Tokyo, Japan; ¹⁰Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan and ¹¹Office for Mental Health Support and Graduate School of Education, The University of Tokyo, Tokyo, Japan

¹²Present address: Department of Gynecology, Osaka Central Hospital, Osaka, Japan.

¹³Present address: Department of Obstetrics and Gynecology, Takano-hara Central Hospital, Nara, Japan.

Correspondence: Professor K Tanaka, Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, 1-757 Asahimachi-dori, chuo-ku, Niigata 951-8510, Japan.

E-mail: tanaken@med.niigata-u.ac.jp

or Professor I Inoue, Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan.

E-mail: ituro@is.icc.u-tokai.ac.jp

Received 6 July 2010; revised 3 August 2010; accepted 4 August 2010; published online 16 September 2010

total of 1176 affected sib-pairs were recruited in the Australian and the UK groups having sufficient statistical power (80%) to detect a locus with a sibling relative risk of 1.3 using nonparametric linkage analysis. The analysis of the Australian and the UK families demonstrated a significant linkage to chromosome 10. Taking more stringent criteria of endometriosis families (three or more affected individuals in a family), nonparametric and parametric linkage analysis identified 7p13–15 as a susceptibility locus with near-Mendelian inheritance.¹⁵ However, identification of the causal genes from these loci is not thus far successful.^{3,4,16}

An alternative approach to identify susceptibility genes of common complex diseases such as endometriosis is genome-wide association (GWA) analysis using high-density single nucleotide polymorphism (SNP) arrays that have recently been exemplified by a number of groups.¹⁷ In this study, we carried out a meta-analysis of two GWA studies to search for novel genetic contributors underlying endometriosis in Japanese population.

MATERIALS AND METHODS

Study subjects

All women with endometriosis were registered at the Niigata University Hospital, the Nagasaki University Hospital, the Kumamoto University Hospital, the Takarazuka City Hospital, the National Hospital Organization Kyoto Medical Center and several hospitals in Niigata, Toyama and Yamagata Prefecture (for details, see Acknowledgements). The affected subjects comprised 728 Japanese women who fulfilled any of the following diagnostic criteria: (i) women who underwent laparotomy or laparoscopic surgery for diseases other than endometriosis, each of which procedure was also used to provide biopsy-proven evidence of endometriosis, (ii) women who were verified to have endometriosis in diagnostic laparoscopy and (iii) women who were diagnosed to have ovarian cysts by imaging diagnostics. The mean ages for 500K and 6.0

array cohorts were 31.1 ± 5.8 years and 34.9 ± 7.9 years, respectively. The clinical characteristics of all the cases studied are summarized in Table 1.

Control samples for association analysis comprised 834 Japanese women from various resources as follows: (i) 96 fertile women or those with benign gynecological tumors, with no history of endometriosis diagnosed at the Niigata University Hospital (age, 37.1 ± 7.4 years, mean \pm s.d.), (ii) 241 Affymetrix CEL files (including per-probe intensity values) from the Japanese Integrated Database Project,¹⁸ each of which had passed 93% call-rate threshold (GeneChip Human Mapping 500K array), or the 86% threshold (Genome-Wide Human SNP array 6.0) with the Dynamic Model algorithm to generate an initial quality control (QC) call rate (for 500K array cohort, panic disorder control cohort ($n=81$ females),¹⁹ multiple system atrophy control cohort ($n=77$ females), control database cohort ($n=17$ females) and for 6.0 array cohort, control database cohort ($n=66$ females)²⁰) and (iii) Genotype count data for 906 703 SNPs genotyped with 6.0 arrays from the Japanese Integrated Database Project (for 6.0 array cohort, late-onset Alzheimer's disease control cohort ($n=497$ females)) (Table 2).

Genomic DNA was extracted from peripheral blood lymphocytes per sample using a QIAamp DNA Blood Maxi Kit (QIAGEN, Tokyo, Japan) according to the manufacturer's protocol. DNA samples used in this study were derived from individuals after they had given written informed consent. The study protocol was approved by the ethical committees of the University of Niigata and the affiliated hospitals.

Genotyping

Genotyping was conducted on two types of the Affymetrix platforms, GeneChip Human Mapping 500K array or Genome-Wide Human SNP array 6.0 (Affymetrix, Santa Clara, CA, USA), according to the manufacturer's instruction. In this study, 411 subjects (315 cases and 96 controls with fertility or no endometriosis) were newly genotyped with 500K arrays, and 413 subjects (413 cases) were genotyped with 6.0 arrays (Table 2). Genotype calls were determined with the Bayesian Robust Linear Model using Mahalanobis distance classifier (BRLMM) algorithm²¹ for 500K arrays or the Birdseed v2 algorithm for 6.0 arrays, embedded in Affymetrix Genotyping Console 3.0.1 (Affymetrix).

Sample QC

The following QC filters were applied to exclude samples in each of array cohorts (Table 2). We excluded from analysis samples that showed >5% missing genotypes (per-individual call rate <95%) and were outliers with respect to genome-wide heterozygosity (<21% or >30% heterozygous SNP rate), for which excess heterozygosity could be attributable to DNA contamination.²²

Samples showing cryptic relatedness were excluded. Cryptic relatedness was examined by pairwise identity-by-state analysis with gPLINK²³ using nearly independent SNPs, among which no pair were correlated with $r^2 > 0.2$. We estimated the degree of relatedness for each pair of samples, and found duplicates, first- and second-degree pairs based on pairwise identity-by-descent

Table 1 Characteristics of women with endometriosis

Characteristic	500K array cohort ($n=315$)	6.0 array cohort ($n=413$)
Mean age at diagnosis (years (s.d.))	31.1 (5.8)	34.9 (7.9)
<i>Diagnostic procedures (n (%))</i>		
Pathological examination	250 (79.3)	270 (65.3)
Diagnostic laparoscopy	21 (6.7)	11 (2.7)
Imaging	44 (14.0)	132 (32.0)

Table 2 Sample quality control process

Genotyping platform	Cohort type	Initial size ^a	Excluded subjects				Population outliers ^c	Additional subjects Genotype count data	Final size
			Per-individual call rate: <95% ^b	Heterozygous SNP rate: >30% or <21%	Cryptic relatedness ^c				
500K array	Case	315	5	0	14	6	—	290	
	Control	271 ^d	6	1 ^e	2	0	—	262	
6.0 array	Case	413	3	0	2	2	—	406	
	Control	66 ^f	0	0	0	0	497 ^g	563	

^aNumber of CEL files that passed 93% call-rate threshold (500K array) or the 86% threshold (6.0 array) with the Dynamic Model algorithm.

^bGenotype calls were determined using the BRLMM algorithm (500K array) or the Birdseed v2 algorithm (6.0 array).

^cCryptic relatedness and population outliers were examined by analyses of pairwise identity-by-state similarities using gPLINK.

^dThe CEL files were from 96 women with fertility or no endometriosis and 175 women in the Japanese Integrated Database Project.

^eSample with genome-wide deficiency of heterozygosity was removed.

^fAll the CEL files were from the Japanese Integrated Database Project.

^gGenotype counts for 906 703 SNPs in 497 women were from the Japanese Integrated Database Project.

Table 3 SNP quality control (QC) process

QC metric	500K array		6.0 array	
	Remaining genotyped SNPs	Genomic inflation factor	Remaining genotyped SNPs	Genomic inflation factor
Initial data	499 264	—	905 013	—
MAF \geq 1% in cases and controls	390 595	—	686 407	—
MR $<$ 4% in cases and controls (500K array)	343 430	—	—	—
MR $<$ 2% in cases and controls (6.0 array)	—	—	588 448	—
Differential missingness ^a	332 560	—	558 228	—
Exact HWE $P \geq 10^{-5}$ in cases and controls	330 389	1.022	557 299	1.048
Remaining SNPs common to both platforms (genomic inflation factor)	282 838 (1.031) ^b			

Abbreviations: GWA, genome-wide association; HWE, Hardy–Weinberg equilibrium; MR, per-SNP missing rate; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

^aWe excluded SNPs that showed P -values of the MR differences between cases and controls $<$ 0.05 in each array cohort.

^bSNPs that passed QC filters in both genotyping platforms were used for meta-analysis of the current GWA studies.

sharing. We kept only one subject from the duplicate or relative samples inferred. Samples showing population outliers were also excluded. Because it is known that there is a population substructure in Japanese population,²⁴ we performed multidimensional scaling analysis with the use of gPLINK for our cohort with 90 HapMap samples (45 JPT and 45 CHB samples) as references to provide a two-dimensional projection of the data based on pairwise identity-by-state distances. After removal of samples that were clearly separate from the main cluster of our cohorts and HapMap JPT samples, we further performed similar multidimensional scaling analysis only using the present cohort samples. In the secondary multidimensional scaling analysis, samples showing $>$ 4 s.d. from the mean along any of the two-dimensional scaling axes were regarded as population outliers to be excluded.

After the above sample QC, genotype count data for 497 women, which had been independently quality controlled in late-onset Alzheimer's disease cohort (R. Kuwano, personal communication), were added to perform the following SNP QC (Table 2).

SNP QC

Data underwent QC in each of array cohorts, and only SNPs fulfilling the following criteria were included: (i) minor allele frequency \geq 0.01 in cases and controls, (ii) per-SNP missing rate $<$ 4% (for 500K array) or $<$ 2% (for 6.0 array) in cases and controls, (iii) P -value of exact test of the missing rate differences between cases and controls \geq 0.05 and (iv) P -value of exact test of Hardy–Weinberg equilibrium $\geq 10^{-5}$ in cases and controls (Table 3). For these calculations, we used R statistical environment version 2.9.0 (<http://www.r-project.org>).

Statistical analysis

In cohort-wise association analysis, we tested for single-locus association between each QC-passed SNP (Table 3) and endometriosis in an allele frequency model (1 degree of freedom) using the χ^2 -test. We then assessed the association by combining cohort-wise odds ratios (ORs) in each array cohort based on a meta-analysis of the two GWA studies. To examine between-cohort heterogeneity of the effect size of SNP association, we conducted the Cochran's Q -test²⁵ and used P -value $<$ 0.01 as evidence of significant heterogeneity. When the between-cohort heterogeneity was not significant, a fixed-effects model of meta-analysis was conducted using the Mantel–Haenszel methods.²⁶ We applied a random-effects model (the DerSimonian and Laird method²⁷) of meta-analysis to SNPs with significant between-cohort heterogeneity for combining the results from the two cohorts. From the meta-

analysis, we estimated summary log-ORs of risk alleles with standard errors, and obtained P -values from the score test (two-tailed) for 282 838 SNPs common to two genotyping platforms. To evaluate the degree of overdispersion of test statistics, the genomic inflation factor was calculated by the ratio of the median of the observed test statistics to that of the expected χ^2 -values. For general statistical analysis, we used R statistical environment version 2.9.0 (especially, the R-package metafor (<http://www.wvbauer.com/>)). We used Haploview 4.2 software²⁸ to generate Manhattan plot of genome-wide significance (P -values) and to draw linkage disequilibrium map on 2q13, based on HapMap JPT data from the HapMap database (<http://hapmap.ncbi.nlm.nih.gov/>). We calculated the statistical power of the analysis using Genetic Power Calculator,²⁹ assuming that type I error rate was 5×10^{-7} , the prevalence of endometriosis was 10%, genotype relative risk (GRR) was 1.5 or 2.0 and unselected controls (random population samples) were used as controls in the case–control association analysis.

RESULTS

Cohort-wise analysis

A total of 586 female subjects (315 cases and 271 controls) were genotyped with the Affymetrix Mapping 500K arrays, and 976 women (413 cases and 563 controls) were genotyped with the Affymetrix SNP 6.0 arrays (Table 2). The newly and previously genotyped data were subjected to well established sample QC filters. After sample QC, the remaining case–control data consisted of 290 cases and 262 controls in 500K array cohort, and 406 cases and 563 controls in 6.0 array cohort (Table 2).

We further performed data cleaning for autosomal and X-chromosomal SNPs (500K array, 499 264 SNPs and 6.0 array, 905 013 SNPs) to extract genotype data that passed SNP QC filters. From the pre-cleaned SNPs, we selected 330 389 and 557 299 SNPs through the SNP QC filters in 500K and 6.0 array cohorts, respectively (Table 3).

We tested allelic association between each QC-passed SNP and endometriosis in each array cohort. The genomic inflation factors were 1.022 and 1.048 for 500K array and 6.0 array cohorts, respectively (Table 3), indicating that systematic inflation of genetic association due to population stratification or undetected genotyping error was unlikely in each of the array cohorts.

Meta-analysis of two GWA studies

Among QC-passed SNPs in each array cohort, 282 838 SNPs on autosomes and X-chromosome were overlapped between 500K and 6.0 arrays, and used for a meta-analysis of two GWA studies. When combining the results from the two array cohorts, we assessed between-cohort heterogeneity in genetic effects of the respective SNPs using the Cochran's Q -test.²⁵ According to the absence or presence of between-cohort heterogeneity in cohort-wise ORs for each SNP, we selected an appropriate model (fixed-effects or random-effects model, respectively) of meta-analysis for calculating a summary OR. Figure 1 shows quantile–quantile (Q - Q) plot of P -values ($-\log_{10}$ scale) for association with endometriosis from the present genome-wide meta-analysis. The genomic inflation factor of 1.031 was similar to those from the two array cohorts (Table 3).

From the GWA meta-analysis, P -values across autosomes and X-chromosomes for association with endometriosis are shown in Figure 2. In our meta-analysis, we found that none of the SNPs analyzed surpassed a genome-wide significant threshold of 5×10^{-7} .²² This meta-analysis had 80% power to detect a common risk allele (risk allele frequency, 0.1–0.8) with GRR of 2.0 (Supplementary Figure 1), suggesting that common risk variants with large effect sizes (GRRs $>$ 2.0) are unlikely to have roles in the development of endometriosis. On the other hand, we found an excess of SNPs with P -values $<$ 10^{-4}

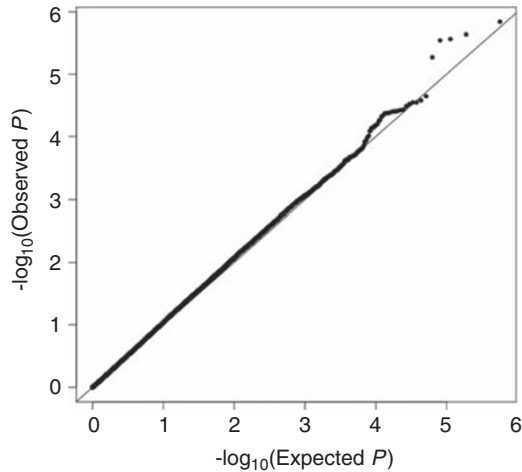


Figure 1 Log quantile-quantile (Q-Q) plot of P -values from genome-wide meta-analysis of 282 838 SNPs on autosomes and X-chromosome. The genetic inflation factor was 1.031.

(36 SNPs observed vs 28 SNPs expected by chance). Some of these may be demonstrated to be true risk variants, as additional cohorts would be evaluated. For replication study, we show five SNPs with P -values $< 10^{-5}$ for association with endometriosis from the meta-analysis in Table 4. Four of the five SNPs with P -values $< 10^{-5}$ were located in and around *IL1A* (interleukin 1 α) on 2q13, and were in high linkage disequilibrium with each other (Supplementary Figure 2).

As additional data, we provide 31 SNPs with P -values between 10^{-5} and 10^{-4} from the meta-analysis in Supplementary Table 1. Furthermore, five SNPs showing association with endometriosis at P -values $< 10^{-6}$ only in each cohort are shown in Supplementary Table 2. These lists of SNPs may also aid the search for susceptibility loci associated with endometriosis.

DISCUSSION

Endometriosis is a condition that millions of women suffer worldwide in terms of pelvic pain and infertility, and its etiology remains to be largely uncovered. Despite much evidence showing the involvement of

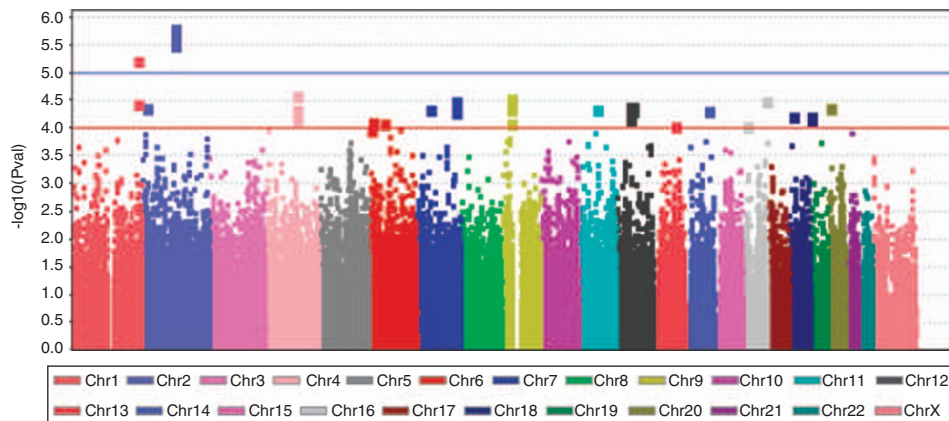


Figure 2 Summary of results from meta-analysis of two GWA studies by chromosome. Manhattan plot for single-locus association (y-axis, shown as $-\log_{10} P$ -values) of 282 838 SNPs with endometriosis in the meta-analysis by SNP position along chromosome (x-axis).

Table 4 SNPs showing association with endometriosis at P -values $< 10^{-5}$ from meta-analysis of two GWA studies

SNP	Chr	Position ^a	Risk allele	Cohort	Risk allele frequency		OR (95% CI)	P-value	P_{het}^b	Nearest gene
					Cases	Controls				
rs801112	1q42.13	227 063 245	C	500K	0.167	0.099	1.82 (1.27–2.61)	1.1×10^{-3}	0.49	<i>RHOU</i> , downstream
				6.0	0.154	0.105	1.55 (1.19–2.04)	1.3×10^{-3}		
				Combined ^c			1.65 (1.33–2.05)	5.4×10^{-6}		
rs6542095	2q13	113 245 654	C	500K	0.803	0.748	1.38 (1.04–1.83)	2.8×10^{-2}	0.45	<i>IL1A</i> , downstream
				6.0	0.796	0.712	1.58 (1.27–1.96)	2.9×10^{-5}		
				Combined			1.50 (1.27–1.78)	2.9×10^{-6}		
rs11677416	2q13	113 245 711	T	500K	0.946	0.909	1.75 (1.10–2.81)	1.9×10^{-2}	0.50	<i>IL1A</i> , downstream
				6.0	0.948	0.894	2.15 (1.49–3.10)	3.9×10^{-5}		
				Combined			2.00 (1.50–2.67)	2.3×10^{-6}		
rs3783550	2q13	113 249 356	C	500K	0.803	0.747	1.38 (1.04–1.84)	2.5×10^{-2}	0.47	<i>IL1A</i> , intron 6
				6.0	0.797	0.713	1.58 (1.27–1.95)	3.1×10^{-5}		
				Combined			1.51 (1.27–1.79)	2.7×10^{-6}		
rs3783525	2q13	113 258 290	A	500K	0.803	0.744	1.40 (1.05–1.86)	2.1×10^{-2}	0.47	<i>IL1A</i> , intron 1
				6.0	0.798	0.712	1.60 (1.29–1.98)	1.9×10^{-5}		
				Combined			1.52 (1.28–1.81)	1.4×10^{-6}		

Abbreviations: Chr, chromosomes; CI, confidence intervals; GWA, genome-wide association; OR, odds ratio; SNP, single nucleotide polymorphism.

^aNCBI genome build 36.3.

^bCochran's Q -statistic was used to test for between-study heterogeneity.

^cCohort-wise ORs were combined using a fixed-effects model (Mantel-Haenszel) of meta-analysis for five SNPs.

genetic components in the disease risk, both linkage and candidate gene-based case-control studies have not been successful to identify replicable genetic susceptibility to endometriosis thus far. We performed a meta-analysis of GWA studies to identify genetic susceptibility underlying endometriosis in Japanese population using two independent cohorts comprised of 696 cases and 825 controls. Population stratification is the first concern, particularly in the current GWA studies, because the case-control samples were recruited by multiple centers distributed in Honshu and Kyushu islands in Japan, where the population structure is genetically evident.²⁴ After removal of samples and SNPs through the sequential QC processes in each array cohort, we observed genomic inflation factors of 1.022 and 1.048 in 500K array and 6.0 array cohorts, respectively. Thus, the inflation of false-positive rates on genetic association would be within an acceptable level (genomic inflation factor <1.1)²⁴ for a case-control association study, indicating that the current QC processes are successful.

We assessed single-locus association of 282 838 SNPs with endometriosis using a genome-wide meta-analysis of the results from two array cohorts. Despite using QC-passed samples of Japanese ethnicity, there might be between-cohort heterogeneity in genetic effects because of sampling bias in each case-control cohort and/or chip-to-chip difference by using two types of genotyping platforms. To accommodate the possible between-cohort heterogeneity, we combined the association results using a random-effects model (the DerSimonian and Laird method) of meta-analysis, a more conservative approach,^{25,27} only when the between-cohort heterogeneity was very significant (P -value of the Cochran's Q -test statistic <0.01); otherwise, we combined the results using a fixed-effects model (the Mantel-Haenszel method) of meta-analysis. This meta-analysis approach caused no marked under- or overdispersion of test statistics for genetic association, because genomic inflation factor of 1.031 in combined data set was almost equivalent to those from the two array cohorts. This suggests that the current meta-analysis could provide a genome-wide summary of genetic association with endometriosis.

In the GWA meta-analysis, there were no association signals reaching the genome-wide significance (P -values < 5×10^{-7}). This GWA study is sufficiently powered to detect common risk alleles with $GRR \geq 2.0$ (Supplementary Figure 1) so that there is less likely to be common and large-effect susceptibility loci contributing to the risk of endometriosis across the genome. On the other hand, common variants with modest and/or small effects might confer the disease risk, as an excess of associated SNPs with P -values < 10^{-4} (36 SNPs observed vs 28 SNPs expected by chance) is observed in the GWA meta-analysis. To examine this possibility using GWA approaches, independent sets of larger case-control cohorts will be required.

Recently, Uno *et al.*³⁰ have reported that rs10965325, which is located in *CDKN2BAS* on chromosome 9p21, is significantly associated with endometriosis in the Japanese population using a GWA study. In our GWA analysis, SNP rs17761446, which is in perfect linkage disequilibrium with rs10965235 ($D' = 1$, $r^2 = 1$, HapMap JPT population (Phase III, Rel no. 2)), showed P -value = 8.9×10^{-3} with summary OR = 1.29 (95% CI, 1.10–1.48), suggesting the existence of an endometriosis susceptibility locus at 9p21.

For replication and GWA analysis with other cohorts, we provide a set of SNPs that are associated with endometriosis in the current meta-analysis at the significant level of P -value < 10^{-4} (Table 4 and Supplementary Table 1). We also give another set of SNPs showing association at P -values < 10^{-6} only in each cohort (Supplementary Table 2), although the SNPs have no significant summary ORs from the meta-analysis owing to large between-cohort heterogeneity. These

SNPs might be regarded as candidates for future genetic studies. Among the SNPs listed, four of the top five SNPs with P -values < 10^{-5} (Table 4) are mapped in and around *IL1A* on 2q13. The *IL-1 α* encoded by *IL1A* is a member of the interleukin 1 cytokine family. This cytokine 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.³¹ There is a general agreement that endometriosis is essentially a pelvic inflammatory process and these cytokines, including *IL-1 α* , could be important for the development and progression of endometriosis. Kondera-Anasz *et al.*³² have found that *IL-1 α* levels in peritoneal fluid and serum are higher in women with endometriosis than in women without the disease, and therefore this gene might be further regarded as a functional candidate. In any case, we hope that the SNP lists will be useful in an extensive search of susceptibility loci for endometriosis.

In summary, this is the GWA study to endometriosis in Japanese population. Our data may be still preliminary because it is a single meta-analysis GWA study with limited sample size. It will be necessary to conduct more detailed GWA investigations using independent sets of larger endometriosis cohorts in Japan and other countries with higher density platforms for genotyping common and rare alleles. Further genetic studies will expand our findings and provide specialized diagnostic options and approaches for the treatment of endometriosis.

ACKNOWLEDGEMENTS

All women with endometriosis were registered at the Niigata University Hospital, the Nagasaki University Hospital, the Kumamoto University Hospital, the Takarazuka City Hospital, the National Hospital Organization Kyoto Medical Center, the Nagaoka Red Cross Hospital, the Kido Hospital, the Niigata Cancer Center Hospital, the Suibarago Hospital, the Joetsu General hospital, the Kaetsu Hospital, the Niigata Medical Center, the Muikamachi Hospital, the Toyama Red Cross Hospital and the Syonai Hospital. We thank Katsunori Kashima, Nobumichi Nishikawa, Naoki Fukui and Masayuki Yamaguchi for the management of DNA samples. We thank China Seki, Akemi Yukawa, Hiromi Kamura and Eriko Tokubo for their technical assistance. The data used for this research are partly available at the website of 'DB Development for Medical Application from Disease Analysis (<https://gwas.lifesciencedb.jp/index.html>)', which is a part of the 'Integrated Database Project' funded by the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was supported, in part, by 2008 Research and Study Program of Tokai University Educational System General Research Organization (AT).

- Giudice, L. C. & Kao, L. C. Endometriosis. *Lancet* **364**, 1789–1799 (2004).
- Gao, X., Outley, J., Botteman, M., Spalding, J., Simon, J. A. & Pashos, C. L. Economic burden of endometriosis. *Fertil. Steril.* **86**, 1561–1572 (2006).
- Bulun, S. E. Endometriosis. *N. Engl. J. Med.* **360**, 268–279 (2009).
- Montgomery, G. W., Nyholt, D. R., Zhao, Z. Z., Treloar, S. A., Painter, J. N., Missmer, S. A. *et al.* The search for genes contributing to endometriosis risk. *Hum. Reprod. Update* **14**, 447–457 (2008).
- Berkley, K. J., Rapkin, A. J. & Papka, R. E. The pains of endometriosis. *Science* **308**, 1587–1589 (2005).
- Kennedy, S., Mardon, H. & Barlow, D. Familial endometriosis. *J. Assist. Reprod. Genet.* **12**, 32–34 (1995).
- Simpson, J. L. & Bischoff, F. Z. Heritability and molecular genetic studies of endometriosis. *Ann. NY Acad. Sci.* **955**, 239–251; discussion 293–295, 396–406 (2002).
- Stefansson, H., Geirsson, R. T., Steinthorsdottir, V., Jonsson, H., Manolescu, A., Kong, A. *et al.* Genetic factors contribute to the risk of developing endometriosis. *Hum. Reprod.* **17**, 555–559 (2002).
- Treloar, S., Hadfield, R., Montgomery, G., Lambert, A., Wicks, J., Barlow, D. H. *et al.* The International Endogene Study: a collection of families for genetic research in endometriosis. *Fertil. Steril.* **78**, 679–685 (2002).
- Treloar, S. A., O'Connor, D. T., O'Connor, V. M. & Martin, N. G. Genetic influences on endometriosis in an Australian twin sample. *sueT@qimr.edu.au. Fertil. Steril.* **71**, 701–710 (1999).

- 11 Simpson, J. L., Elias, S., Malinak, L. R. & Buttram, V. C. Jr Heritable aspects of endometriosis. I. Genetic studies. *Am. J. Obstet. Gynecol.* **137**, 327–331 (1980).
- 12 Kashima, K., Ishimaru, T., Okamura, H., Suginami, H., Ikuma, K., Murakami, T. *et al.* Familial risk among Japanese patients with endometriosis. *Int. J. Gynaecol. Obstet.* **84**, 61–64 (2004).
- 13 Hadfield, R. M., Mardon, H. J., Barlow, D. H. & Kennedy, S. H. Endometriosis in monozygotic twins. *Fertil. Steril.* **68**, 941–942 (1997).
- 14 Moen, M. H. Endometriosis in monozygotic twins. *Acta. Obstet. Gynecol. Scand.* **73**, 59–62 (1994).
- 15 Zondervan, K. T., Treloar, S. A., Lin, J., Weeks, D. E., Nyholt, D. R., Mangion, 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).
- 16 Treloar, S. A., Zhao, Z. Z., Le, L., Zondervan, K. T., Martin, N. G., Kennedy, S. *et al.* Variants in EMX2 and PTEN do not contribute to risk of endometriosis. *Mol. Hum. Reprod.* **13**, 587–594 (2007).
- 17 Pearson, T. A. & Manolio, T. A. How to interpret a genome-wide association study. *JAMA* **299**, 1335–1344 (2008).
- 18 Koike, A., Nishida, N., Inoue, I., Tsuji, S. & Tokunaga, K. Genome-wide association database developed in the Japanese Integrated Database Project. *J. Hum. Genet.* **54**, 543–546 (2009).
- 19 Otowa, T., Tani, H., Sugaya, N., Yoshida, E., Inoue, K., Yasuda, S. *et al.* Genome-wide association study of panic disorder in the Japanese population. *J. Hum. Genet.* **54**, 122–126 (2009).
- 20 Nishida, N., Koike, A., Tajima, A., Ogasawara, Y., Ishibashi, Y., Uehara, Y. *et al.* Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics* **9**, 431 (2008).
- 21 Rabbee, N. A genotype calling algorithm for Affymetrix SNP arrays. *Bioinformatics* **22**, 7–12 (2006).
- 22 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* **447**, 661–678 (2007).
- 23 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 24 Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
- 25 Nakaoka, H. & Inoue, I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *J. Hum. Genet.* **54**, 615–623 (2009).
- 26 Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
- 27 Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One*, **2**, e841 (2007).
- 28 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- 29 Purcell, S., Cherny, S. S. & Sham, P. C. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149–150 (2003).
- 30 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.* **42**, 707–710 (2010).
- 31 Fitzgerald, K. A. Integrating IL-1 alpha in antiviral host defenses. *Immunity* **31**, 7–9 (2009).
- 32 Kondera-Anasz, Z., Sikora, J., Mielczarek-Palacz, A. & Jonca, M. Concentrations of interleukin (IL)-1alpha, IL-1 soluble receptor type II (IL-1 sRII) and IL-1 receptor antagonist (IL-1 Ra) in the peritoneal fluid and serum of infertile women with endometriosis. *Eur. J. Obstet. Gynecol. Reprod Biol.* **123**, 198–203 (2005).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)