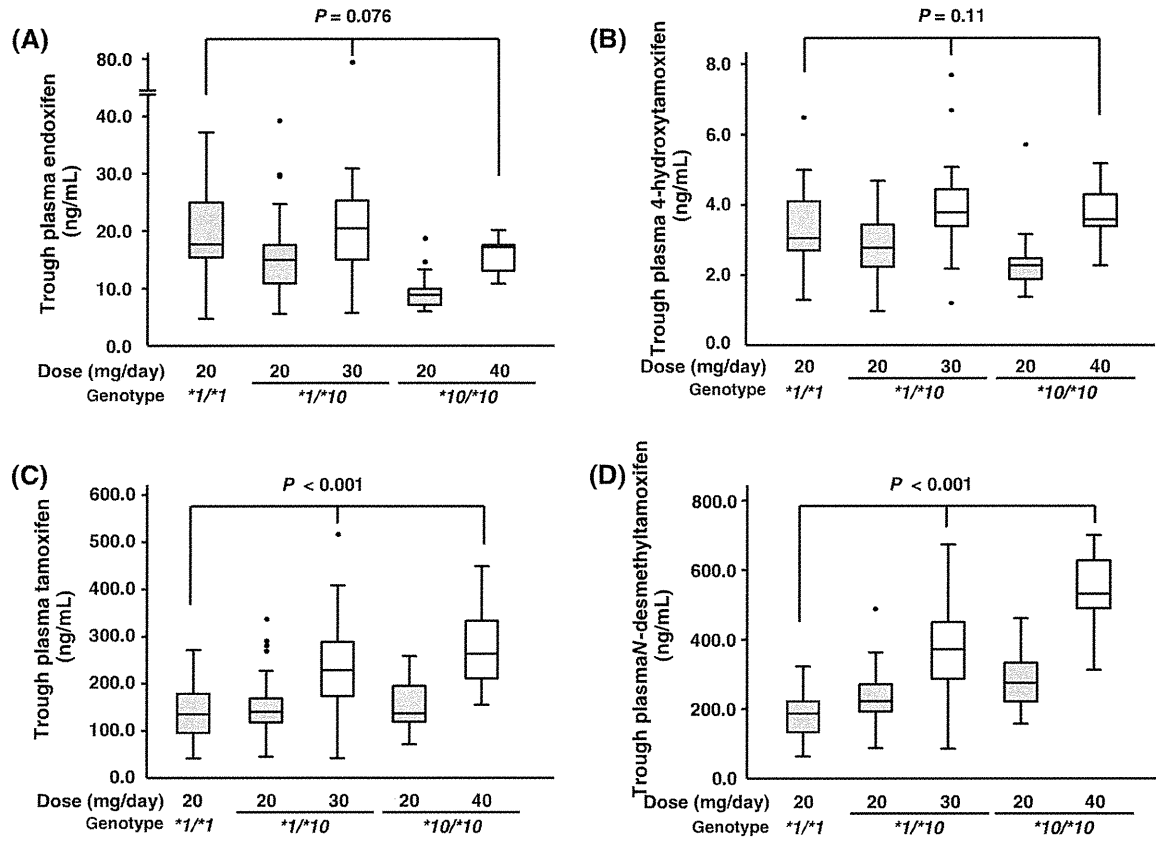

Fig. 3 Comparison of steady-state plasma concentrations of tamoxifen and its metabolites among the patients administrated with different dosages of tamoxifen. **a** Endoxifen, **b** 4-hydroxytamoxifen, **c** tamoxifen, and **d** *N*-desmethyltamoxifen. The *horizontal line* indicates the median concentration, the box covers the 25th–75th percentiles, and the maximum length of each whisker is 1.5× the interquartile range; *dots* outside the whiskers are outliers. The difference in plasma concentrations of tamoxifen and its metabolites among **1/*1* (20 mg/day), **1/*10* (30 mg/day), and **10/*10* (40 mg/day) genotypes for *CYP2D6* was evaluated by a one-way ANOVA test

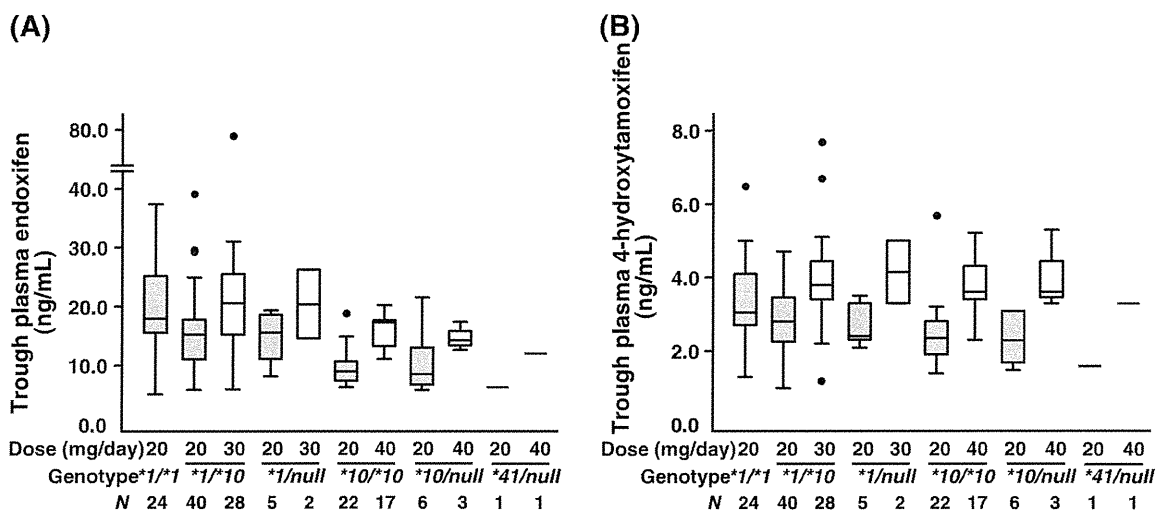

Fig. 4 Comparison of steady-state plasma concentrations of endoxifen (**a**) and 4-hydroxytamoxifen (**b**) among the patients administrated with different dosages of tamoxifen according to *CYP2D6* genotypes. The *horizontal line* indicates the median concentration, the box covers the 25th–75th percentiles, and the maximum length of each whisker is 1.5× the interquartile range; *dots* outside the whiskers are outliers. The difference in plasma concentrations of endoxifen and 4-hydroxytamoxifen between *CYP2D6*10* and null alleles was evaluated by the Student's *t* test. null: *CYP2D6*5, *21* and **36-*36*

Table 3 Association between tamoxifen dose and incidence of adverse events (all grades according to CTCAE v4.0)

Adverse events	*CYP2D6* genotype	Event/no event, no of patients (%)		After compared to before		After compared to **1/*1*	
		Before (20 mg/day)	After (30 or 40 mg/day)	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Hot flashes	**1/*1*	9/1 (90.0%)	–				
	**1/decreased and *1/null*	27/9 (75.0%)	18/9 (66.7%)	0.67 (0.22–2.00)	0.58	0.22 (0.02–2.04)	0.23
	Decreased/decreased and decreased/null	19/2 (90.5%)	19/3 (86.4%)	0.67 (0.10–4.45)	1.00	0.70 (0.06–7.74)	1.00
Hyperhidrosis	**1/*1*	9/1 (90.0%)	–				
	**1/decreased and *1/null*	21/15 (58.3%)	14/13 (51.9%)	0.77 (0.28–2.1)	0.62	0.12 (0.01–1.08)	0.056
	Decreased/decreased and decreased/null	13/8 (61.9%)	10/12 (45.5%)	0.51 (0.15–1.73)	0.36	0.09 (0.01–0.86)	0.024
Vaginal discharge	**1/*1*	7/3 (70.0%)	–				
	*1/decreased and *1/null*	30/6 (83.3%)	21/6 (77.8%)	0.70 (0.20–2.47)	0.75	1.50 (0.29–7.65)	0.68
	Decreased/decreased and decreased/null	12/9 (57.1%)	18/4 (81.8%)	3.38 (0.84–13.5)	0.10	1.93 (0.34–10.91)	0.65
Irregular menstruation	**1/*1*	1/9 (10.0%)	–				
	**1/decreased and *1/null*	3/33 (8.3%)	0/27 (0.0%)	0.17 (0.01–3.52)	0.25	0.12 (0.00–3.07)	0.27
	Decreased/decreased and decreased/null	2/19 (9.5%)	2/20 (9.1%)	0.95 (0.12–7.44)	1.00	0.90 (0.07–11.25)	1.00
Nausea or vomiting	**1/*1*	1/9 (10.0%)	–				
	**1/decreased and *1/null*	6/30 (16.7%)	3/24 (11.1%)	0.63 (0.14–2.76)	0.72	1.13 (0.1–12.27)	1.00
	Decreased/decreased and decreased/null	4/17 (19.0%)	3/19 (13.6%)	0.67 (0.13–3.44)	0.70	1.42 (0.13–15.64)	1.00
Eye disorders	**1/*1*	4/6 (40.0%)	–				
	**1/decreased and *1/null*	17/19 (47.2%)	12/15 (44.4%)	0.89 (0.33–2.44)	1.00	1.20 (0.27–5.25)	1.00
	Decreased/decreased and decreased/null	8/13 (38.1%)	11/11 (50.0%)	1.63 (0.48–5.47)	0.54	1.50 (0.33–6.83)	0.71
Malaise	**1/*1*	7/3 (70.0%)	–				
	**1/decreased and *1/null*	21/15 (58.3%)	12/15 (44.4%)	0.57 (0.21–1.57)	0.32	0.34 (0.07–1.62)	0.27
	Decreased/decreased and decreased/null	8/13 (38.1%)	7/15 (31.8%)	0.76 (0.22–2.67)	0.75	0.20 (0.04–1.01)	0.062
Reproductive system disorders-endometrial thickening	**1/*1*	0/24 (0.0%)	–				
	**1/decreased and *1/null*	2/43 (4.4%)	0/30 (0.0%)	0.28 (0.01–5.95)	0.87	–	–
	Decreased/decreased and decreased/null	0/29 (0.0%)	1/20 (4.8%)	4.12 (0.16–106.01)	0.84	3.59 (0.14–92.84)	0.84
Thromboembolic event	**1/*1*	1/23 (4.2%)	–				
	**1/decreased and *1/null*	1/44 (2.2%)	0/30 (0.0%)	0.47 (0.02–11.94)	1.00	0.26 (0.01–6.59)	0.85
	Decreased/decreased and decreased/null	0/29 (0.0%)	0/20 (0.0%)	–	–	0.36 (0.01–9.43)	1.00
Hepatobiliary disorders-exacerbation of hepatic steatosis	**1/*1*	0/24 (0.0%)	–				
	**1/decreased and *1/null*	0/45 (0.0%)	2/28 (6.7%)	7.71 (0.36–166.39)	0.74	4.30 (0.20-93.90)	0.85
	Decreased/decreased and decreased/null	0/29 (0.0%)	0/21 (0.0%)	–	–	–	–

CI confidence interval

Decreased: **10*, **41*; null: **5*, **21*, **36-*36*

increase of tamoxifen dose for the patients with *CYP2D6*1/*10*, **1*/null, **10/*10*, **10*/null, and **41*/null genotypes was an useful method to achieve the plasma levels of active metabolites of tamoxifen which was seen in the patients with *CYP2D6*1/*1* genotype.

Subjects who carry at least one decreased-function allele (*CYP2D6*10* or *CYP2D6*41*) or one null allele, remain to have a certain level of enzymatic activity although it is lower than the *CYP2D6*1/*1* genotype. Therefore, increased dose is an effective way to overcome the problem of reduced enzymatic activity and to increase the level of active metabolites for these populations. However, we could not evaluate the effects of increasing dose in the null/null patients because no null/null patient participated in this study. Recently, Irvin et al. reported that endoxifen concentration in PM patients, who were defined as homozygote for inactive alleles, was still lower after increasing tamoxifen dose to 40 mg/day (12.9 ng/ml) than that of patients classified as extensive metabolizers, who carry two alleles with normal activity (29.2 ng/ml) [28]. It should be noted that dose-adjustment strategy is useful for patients carrying at least one decreased-function allele or one null allele, while the postmenopausal patients with null/null genotype of *CYP2D6* might be more beneficial to take aromatase inhibitors instead of increased dose of tamoxifen, although further verification is required.

It has been well known that several adverse events were observed during tamoxifen therapy [29]. Hot flash is one of the most common adverse events, which was observed in up to 80% of patients prescribed with tamoxifen, and approximately 30% of them are relatively severe [29]. In this study, no significant difference was observed in the incidence of hot flash between the groups before and after increasing tamoxifen dose (Table 3). The incidence of hot flash has been suggested to be associated with the *CYP2D6* genotypes [16, 30], implying association with plasma levels of endoxifen and 4-hydrotamoxifen. The results from our preliminary investigation suggest that dose adjustment from 20 to 30 mg/day of tamoxifen for the patients with *CYP2D6*1/*10* and **1*/null and 40 mg/day for the patients with **10/*10*, **10*/null, and **41*/null may not affect the risk of adverse events, although tamoxifen and *N*-desmethyltamoxifen showed higher plasma concentrations in the patients receiving higher tamoxifen dose than those of *CYP2D6*1/*1* patients with 20 mg/day of tamoxifen. Further analysis using a larger number of patients is required to evaluate the influences of increase of tamoxifen dose on adverse events.

In conclusion, the dose-adjustment study based on the *CYP2D6* genotypes indicated that the increase of tamoxifen dose was able to increase the endoxifen plasma concentration, and expected to improve the prognosis of the tamoxifen-treated patients who show decreased CYP2D6 activity by genetic polymorphisms. A prospective large-scale study is required to confirm our dose-adjustment strategy for improvement of tamoxifen therapy in breast cancer patients.

Acknowledgments This study was mainly supported by a Grant-in-Aid for Leading Project of Ministry of Education, Culture, Sports, Science and Technology of Japan. This study was also supported in part by Grant-in-Aids for the Kobayashi Institute for Innovative Cancer Chemotherapy, the Japan Research Foundation for Clinical Pharmacology, the Takeda Science Foundation, and Young Scientists (B) (22790179) of Ministry of Education, Culture, Sports, Science and Technology of Japan. We express our heartfelt gratitude to all the study participants. We thank Yuka Kikuchi and Kumi Matsuda for technical assistance. We thank all other members and stuffs for their contribution to the sample collection and the completion of our study.

Disclosures None.

References

1. Early Breast Cancer Trialists' Collaborative Group (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 351:1451–1467
2. Early Breast Cancer Trialists' Collaborative Group (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717
3. Borgna JL, Rochefort H (1981) Hydroxylated metabolites of tamoxifen are formed in vivo and bound to estrogen receptor in target tissues. J Biol Chem 256:859–868
4. Lien EA, Solheim E, Lea OA, Lundgren S, Kvinnsland S, Ueland PM (1989) Distribution of 4-hydroxy-*N*-desmethyltamoxifen and other tamoxifen metabolites in human biological fluids during tamoxifen treatment. Cancer Res 49:2175–2183
5. Johnson MD, Zuo H, Lee KH, Trebley JP, Rae JM, Weatherman RV, Desta Z, Flockhart DA, Skaar TC (2004) Pharmacological characterization of 4-hydroxy-*N*-desmethyl tamoxifen, a novel active metabolite of tamoxifen. Breast Cancer Res Treat 85:151–159
6. Lim YC, Li L, Desta Z, Zhao Q, Rae JM, Flockhart DA, Skaar TC (2006) Endoxifen, a secondary metabolite of tamoxifen, and 4-OH-tamoxifen induce similar changes in global gene expression patterns in MCF-7 breast cancer cells. J Pharmacol Exp Ther 318:503–512
7. Lim HS, Lee JH, Lee SK, Lee SE, Jang IJ, Ro J (2007) Clinical implications of CYP2D6 genotypes predictive of tamoxifen pharmacokinetics in metastatic breast cancer. J Clin Oncol 25:3837–3845
8. Jin Y, Desta Z, Stearns V et al (2005) CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. J Natl Cancer Inst 97:30–39
9. Desta Z, Ward BA, Soukhova NV, Flockhart DA (2004) Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. J Pharmacol Exp Ther 310:1062–1075
10. Crewe HK, Notley LM, Wunsch RM, Lennard MS, Gillam EM (2002) Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: formation of the 4-hydroxy, 4'-hydroxy and *N*-desmethyl metabolites and isomerization of *trans*-4-hydroxytamoxifen. Drug Metab Dispos 30:869–874

11. Broly F, Gaedigk A, Heim M, Eichelbaum M, Morike K, Meyer UA (1991) Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. DNA Cell Biol 10:545–558
12. Sachse C, Brockmoller J, Hildebrand M, Muller K, Roots I (1998) Correctness of prediction of the CYP2D6 phenotype confirmed by genotyping 47 intermediate and poor metabolizers of debrisoquine. Pharmacogenetics 8:181–185
13. Nakamura K, Goto F, Ray WA, McAllister CB, Jacqz E, Wilkinson GR, Branch RA (1985) Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. Clin Pharmacol Ther 38:402–408
14. Yokota H, Tamura S, Furuya H, Kimura S, Watanabe M, Kanazawa I, Kondo I, Gonzalez FJ (1993) Evidence for a new variant *CYP2D6* allele *CYP2D6 J* in a Japanese population associated with lower in vivo rates of sparteine metabolism. Pharmacogenetics 3:256–263
15. Kiyotani K, Mushiroda T, Sasa M, Bando Y, Sumitomo I, Hosono N, Kubo M, Nakamura Y, Zembutsu H (2008) Impact of *CYP2D6*10* on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy. Cancer Sci 99:995–999
16. Goetz MP, Rae JM, Suman VJ et al (2005) Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. J Clin Oncol 23:9312–9318
17. Goetz MP, Knox SK, Suman VJ et al (2007) The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. Breast Cancer Res Treat 101:113–121
18. Schroth W, Antoniadou L, Fritz P, Schwab M, Muerdter T, Zanger UM, Simon W, Eichelbaum M, Brauch H (2007) Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. J Clin Oncol 25:5187–5193
19. Xu Y, Sun Y, Yao L et al (2008) Association between CYP2D6*10 genotype and survival of breast cancer patients receiving tamoxifen treatment. Ann Oncol 19:1423–1429
20. Schroth W, Goetz MP, Hamann U et al (2009) Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. JAMA 302:1429–1436
21. Kiyotani K, Mushiroda T, Imamura CK et al (2010) Significant effect of polymorphisms in *CYP2D6* and *ABCC2* on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. J Clin Oncol 28:1287–1293
22. Gjerde J, Geisler J, Lundgren S, Ekse D, Varhaug JE, Mellgren G, Steen VM, Lien EA (2010) Associations between tamoxifen, estrogens, and FSH serum levels during steady state tamoxifen treatment of postmenopausal women with breast cancer. BMC Cancer 10:313
23. Borges S, Desta Z, Li L et al (2006) Quantitative effect of *CYP2D6* genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. Clin Pharmacol Ther 80:61–74
24. Lim JS, Chen XA, Singh O, Yap YS, Ng RC, Wong NS, Wong M, Lee EJ, Chowbay B (2011) Impact of *CYP2D6*, *CYP3A5*, *CYP2C9* and *CYP2C19* polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. Br J Clin Pharmacol 71:737–750
25. Hosono N, Kato M, Kiyotani K et al (2009) *CYP2D6* genotyping for functional-gene dosage analysis by allele copy number detection. Clin Chem 55:1546–1554
26. Rae JM, Drury S, Hayes DF et al. (2010) Lack of correlation between gene variants in tamoxifen metabolizing enzymes with primary endopoints in the ATAC trial. San Antonio Breast Cancer Symposium, San Antonio, Abstract S1–7
27. Leyland-Jones B, Regan MM, Bouzyk M et al. for the BIG 1-98 Collaborative and International Breast Cancer Study Groups. (2010) Outcome according to CYP2D6 genotype among postmenopausal women with endocrine-responsive early invasive breast cancer randomized in the BIG 1-98 trial. San Antonio Breast Cancer Symposium, San Antonio, Abstract S1–8
28. Irvin WJ, Jr., Walko CM, Weck KE et al. (2011) Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: A Multicenter Study. J Clin Oncol 29(24):3232–3239
29. Day R (2001) Quality of life and tamoxifen in a breast cancer prevention trial: a summary of findings from the NSABP P-1 study. National surgical adjuvant breast and bowel project. Ann N Y Acad Sci 949:143–150
30. Henry NL, Rae JM, Li L et al (2009) Association between *CYP2D6* genotype and tamoxifen-induced hot flashes in a prospective cohort. Breast Cancer Res Treat 117:571–575

Published in final edited form as:
Nat Genet. 2012 October 28; 44(12): 1355–1359. doi:10.1038/ng.2445.

Genome-wide association meta-analysis identifies new endometriosis risk loci

Dale R. Nyholt[1,16], **Siew-Kee Low**[2,16], **Carl A. Anderson**[3], **Jodie N. Painter**[1], **Satoko Uno**[2,4], **Andrew P. Morris**[5], **Stuart MacGregor**[1], **Scott D. Gordon**[1], **Anjali K. Henders**[1], **Nicholas G. Martin**[1], **John Attia**[6,7], **Elizabeth G. Holliday**[6,7], **Mark McEvoy**[6,8,9], **Rodney J. Scott**[7,10,11], **Stephen H. Kennedy**[12], **Susan A. Treloar**[13], **Stacey A. Missmer**[14], **Sosuke Adachi**[15], **Kenichi Tanaka**[15], **Yusuke Nakamura**[2], **Krina T. Zondervan**[5,12,17], **Hitoshi Zembutsu**[2,17], and **Grant W. Montgomery**[1,17]

[1]Queensland Institute of Medical Research, Brisbane, QLD 4029, Australia.

[2]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan.

[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 2HH, UK.

[4]First Department of Surgery, Sapporo Medical University, School of Medicine, Sapporo, Japan.

[5]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.

Correspondence should be addressed to D.R.N (dale.nyholt@qimr.edu.au), K.T.Z. (krina.zondervan@well.ox.ac.uk), H.Z (zembutsh@ims.u-tokyo.ac.jp) or G.W.M (grant.montgomery@qimr.edu.au).; **Corresponding (submitting) author:** A/Prof Dale R Nyholt Queensland Institute of Medical Research Locked Bag 2000, Royal Brisbane Hospital, Herston, Qld 4029, Australia Tel: +61 7 3362 0258 Fax: +61 7 3362 0101 dale.nyholt@qimr.edu.au.

[16]These authors contributed equally to this work.

[17]These authors jointly directed this work.

Author contributions:

Manuscript preparation and final approval: D.R.N., S.-K.L., C.A.A., J.N.P., S.U., A.P M., S M., S D.G., A.K.H., N.G.M., J.A., E.G.H., M.M., R.J.S., S.H.K , S.A.T., S.A.M , S.A., K.T., Y.N., K.T Z., H.Z. & G.W.M.

Study conception and design: D R.N., S.M., Y.N., K.T.Z., H Z. & G.W.M.

GWAS data collection, sample preparation and clinical phenotyping: J N.P., S.U., A K.H., N.G.M., J.A., E.G.H., M.M., R.J.S., S.H.K., S.A.T., K.T Z., H.Z. & G.W.M.

Replication data collection, sample preparation and clinical phenotyping: S.A., K.T. & H.Z.

Replication genotyping: H.Z.

Data analysis: GWA analysis: D.R.N., C.A.A. & S.-K.L ; imputation and replication analysis: D.R N. & S.-K.L.; polygenic prediction, gene-based and meta-analysis: D.R.N.

Obtaining study funding: D.R.N., S.M., N.G.M., S.H.K., S.A.T., S.A.M., Y.N., K.T.Z. & G.W.M.

Competing financial interests The authors declare no competing financial interests.

URLs.

A Catalog of Published Genome-Wide Association Studies, www.genome.gov/gwastudies;
Gene Expression Omnibus (GEO) database, http://www.ncbi.nlm.nih.gov/gds/;
GENe Expression VARiation (Genevar) database, http://www.sanger.ac.uk/resources/software/genevar/;
GWAMA, http://www.well.ox.ac.uk/gwama/;
MaCH, http://www.sph.umich.edu/csg/abecasis/MaCH/;
Mammalian Gene Expression Uterus database (MGEx-Udb), http://resource.ibab.ac.in/cgi-bin/MGEXdb/microarray/scoring/interface/Homepage.pl;
METAL, http://genome.sph.umich edu/wiki/METAL_Program;
Metasoft, http://genetics.cs.ucla.edu/meta/index.html; **minimac**, http://genome.sph.umich.edu/wiki/Minimac;
minimac, http://genome.sph.umich.edu/wiki/Minimac;
Minimac: 1000 Genomes Imputation Cookbook, http://genome.sph.umich.edu/wiki/Minimac:_1000_Genomes_Imputation_Cookbook; 1000 **Genomes Project**, http://www.1000genomes.org/;
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/;
SNPSpD, http://genepi.qimr.edu.au/general/daleN/SNPSpD/;
Wellcome Trust Case-Control Consortium, http://www.wtccc.org.uk;

[6]Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Newcastle, NSW 2308, Australia.

[7]Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia.

[8]School of Medicine and Public Health, University of Newcastle, Newcastle, NSW 2308, Australia.

[9]Public Health Research Program, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia.

[10]School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW 2308, Australia.

[11]Division of Genetics, Hunter Area Pathology Service, Newcastle, NSW 2305, Australia.

[12]Nuffield Department of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.

[13]Centre for Military and Veterans' Health, University of Queensland, Mayne Medical School, 288 Herston Road, QLD 4006, Australia.

[14]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA.

[15]Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, 1-757 Asahimachi-dori, Niigata, Japan.

Abstract

We conducted a genome-wide association (GWA) meta-analysis of 4,604 endometriosis cases and 9,393 controls of Japanese[1] and European[2] ancestry. We show that rs12700667 on chromosome 7p15.2, previously found in Europeans, replicates in Japanese ($P = 3.6 \times 10^{-3}$), and confirm association of rs7521902 on 1p36.12 near *WNT4*. In addition, we establish association of rs13394619 in *GREB1* on 2p25.1 and identify a novel locus on 12q22 near *VEZT* (rs10859871). Excluding European cases with minimal or unknown severity, we identified additional novel loci on 2p14 (rs4141819), 6p22.3 (rs7739264) and 9p21.3 (rs1537377). All seven SNP effects were replicated in an independent cohort and produced $P < 5 \times 10^{-8}$ in a combined analysis. Finally, we found a significant overlap in polygenic risk for endometriosis between the European and Japanese GWA cohorts ($P = 8.8 \times 10^{-11}$), indicating that many weakly associated SNPs represent true endometriosis risk loci and risk prediction and future targeted disease therapy may be transferred across these populations.

Endometriosis (MIM131200) is a common gynecological disease associated with severe pelvic pain, affecting 6-10% of women in their reproductive years[3,4] and 20-50% of women with infertility[5]. Endometriosis risk is influenced by genetic factors and has an estimated heritability of around 51%[3].

Two large endometriosis GWA studies[1,2] have reported genome-wide significant associations. The first, in a Japanese sample of 1,423 cases and 1,318 controls obtained from the BioBank Japan (BBJ), with 484 cases and 3,974 controls for replication, implicated a SNP (rs10965235) in the *CDKN2BAS* gene on chromosome 9p21.3 (overall odds ratio (OR) = 1.44, 95% CI 1.30–1.59; $P = 5.57 \times 10^{-12}$)[1]. The second, by the International Endogene Consortium (IEC) in a sample of European ancestry from Australia (2,270 cases and 1,870 controls) and the UK (924 cases and 5,190 controls), with 2,392 cases and 2,271 controls from the US for replication, identified an intergenic SNP (rs12700667) on 7p15.2 (overall OR = 1.20, 95% CI 1.13–1.27; $P = 1.4 \times 10^{-9}$)[2]. These two studies did not report replication

of each other's top locus, partly because rs10965235 is monomorphic in Caucasian populations. The European study did find association with rs7521902 (OR = 1.16, 95% CI 1.08–1.25, $P = 9.0 \times 10^{-5}$) near the *WNT4* gene on 1p36.12, that was reported to be suggestively associated in the Japanese (OR = 1.20, 95% CI 1.11–1.29, $P = 2.2 \times 10^{-6}$).

Encouraged by the *WNT4* association and with accumulating evidence for many complex traits that the number of discovered variants is strongly correlated with experimental sample size[6], we sought to increase the ratio of controls to cases in the Australian GWA cohort and to perform a formal meta-analysis of the Australian (QIMR), UK (OX) and Japanese (BBJ) GWA data.

To increase the power of the Australian GWA dataset we matched the existing QIMR cases and controls[2] on ancestry to individuals from the Hunter Community Study (HCS)[7]. After stringent quality control (QC), the combined QIMRHCS GWA cohort consisted of 2,262 endometriosis cases and 2,924 controls, increasing the number of controls by 1,054 and the Australian effective sample size by 24%. We also performed more stringent QC incorporating the OX dataset, resulting in a revised OX GWA cohort of 919 endometriosis cases and 5,151 controls. All cases in the QIMRHCS and OX studies have surgically confirmed endometriosis and disease stage from surgical records using the rAFS classification system[8], subjects are grouped into stage A (stage I or II disease or some ovarian disease with a few adhesions; $n = 1{,}680$, 52.8%) or stage B (stage III or IV disease; $n = 1{,}357$, 42.7%), or unknown ($n = 144$, 4.5%). Details of the final GWA and independent replication case-control cohorts are summarized in Table 1 and a schematic of our study design is provided in Fig. 1.

Meta-analysis of all endometriosis 4,604 cases and 9,393 controls for the 407,632 SNPs overlapping in the QIMRHCS, OX and BBJ GWA data, showed that the A allele of rs12700667 at the European 7p15.2 locus (OR = 1.22, 95% CI 1.13–1.31, $P = 7.2 \times 10^{-8}$) also replicates in the Japanese GWA 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 GWA meta-analysis; we also confirmed association with allele A of rs7521902 at the 1p36.12 *WNT4* locus (OR = 1.18, 95% CI 1.11–1.25, $P = 4.6 \times 10^{-8}$) (Table 2).

The GWA meta-analysis identified a novel locus on 12q22 near the *VEZT* gene (allele C of rs10859871 OR = 1.18, 95% CI 1.12–1.25, $P = 5.5 \times 10^{-9}$). We also established association with allele G of rs13394619 in the *GREB1* gene on 2p25.1 (OR = 1.12, 95% CI 1.06–1.18, $P = 2.1 \times 10^{-5}$), previously reported (OR = 1.35, 95% CI 1.17–1.56, $P = 3.8 \times 10^{-5}$) in a small independent Japanese GWA study of 696 cases and 825 controls by Adachi et al (2010)[9]. The G allele of rs13394619 approached conventional genome-wide significance ($P \leq 5 \times 10^{-8}$) in combined analysis of the QIMRHCS, OX, BBJ, Adachi500K and Adachi6.0 GWA data (OR = 1.15, 95% CI 1.09–1.20, $P = 6.1 \times 10^{-8}$) (Table 2). In addition to the three genome-wide significant SNPs on chromosomes 1, 7 and 12 (rs7521902, rs12700667, rs10859871), the Manhattan plot of the all endometriosis GWA meta-analysis results (Supplementary Fig. 1) showed 34 SNPs reached genome-wide *suggestive* association ($P \leq 10^{-5}$).

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)[2], a secondary analysis was performed for the SNPs reaching genome-wide suggestive association, where the association results from QIMRHCS and OX Stage B cases versus controls, were meta-analyzed with the BBJ association results (stage information not available).

After excluding endometriosis cases with minimal (rAFS stage I-II) or unknown severity in the QIMRHCS and OX cohorts, GWA meta-analysis implicated novel loci on 2p14 (allele C of rs4141819 OR = 1.22, 95% CI 1.14–1.32, $P = 6.5 \times 10^{-8}$), 6p22.3 (allele T of rs7739264 OR = 1.21, 95% CI 1.13–1.30, $P = 5.8 \times 10^{-8}$) and 9p21.3 (allele C of rs1537377 OR = 1.22, 95% CI 1.14–1.30, $P = 1.0 \times 10^{-8}$) (Table 2, Supplementary Fig. 2, Supplementary Table 1-2 and Supplementary Note).

Annotated plots showing evidence for association in the combined QIMRHCS, OX and BBJ GWA data of genotyped SNPs across the seven implicated loci from the analysis of all cases and of stage B cases only are provided in Supplementary Figs. 3-9. Imputation up to the 1000 Genomes reference panel produced more significant *P* values and helped resolve the associated region at the 1p36.12 (rs56318008, $P_{all} = 1.3 \times 10^{-10}$), 2p25.1 (rs77294520, $P_{stageB} = 8.6 \times 10^{-8}$), 2p14 (rs2861694, $P_{stageB} = 7.9 \times 10^{-9}$), 6p22.3 (rs6901079, $P_{all} = 1.9 \times 10^{-8}$), 9p21.3 (rs7041895, $P_{stageB} = 5.1 \times 10^{-10}$) and 12q22 (rs11107968, $P_{all} = 3.9 \times 10^{-9}$) loci (Fig. 2 and Supplementary Figs. 10-16). Of particular note, the most significant imputed SNPs on 1p36.12, rs56318008 and rs3820282 ($P_{all} = 1.6 \times 10^{-10}$), are located 22 bp 5′ and *within* the *WNT4* gene, respectively.

Interestingly, the most associated genotyped SNP at 9p21.3 (rs1537377) is 55 kb centromeric to the genome-wide significant SNP reported in the original BBJ GWA[1] (rs10965235) located in the *CDKN2BAS* gene, and 49 kb 3′ to the transcription end site of *CDKN2BAS*. SNP rs10965235 is monomorphic in Caucasian populations and we investigated the independence of rs10965235 and rs1537377 in the BBJ GWA data. Firstly, in the BBJ GWA data, alleles of rs10965235 and rs1537377 are very weakly correlated, with linkage disequilibrium (LD) metrics of $r^2 = 0.028$ and $D' = 0.461$. Secondly, the allelic association *P* values for rs10965235 and rs1537377 are $P = 1.6 \times 10^{-4}$ and $P = 1.8 \times 10^{-2}$, respectively. After conditioning on rs10965235, weak residual association remains at rs1537377 ($P = 9.0 \times 10^{-2}$). Consequently, the data suggest there may be two independent genetic risk factors near the *CDKN2BAS* locus on 9p21.3. *CDKN2BAS* is a long non-coding RNA adjacent to and transcribed from the opposite strand to *CDKN2B* (p15), *CDKN2A* (p16) and *ARF* (p14). Loss of heterozygosity of *CDKN2A* and hypermethylation of the *CDKN2A* promoter have been reported in endometriosis[10,11].

To further validate the seven SNPs implicated by the meta-analysis, we carried out a replication study using a cohort of 1,044 cases and 4,017 controls obtained from the BioBank Japan independent of the BBJ GWA cohort. As shown in the forest plots of risk allele effects estimated using all cases versus controls (Fig. 3), the effects (ORs) were in the same direction for all seven implicated SNPs across the GWA and replication cohorts. With the exception of rs12700667, which was previously replicated ($P = 1.2 \times 10^{-3}$) in 2,392 cases and 2,271 controls from the US[2], and rs4141819 (with a marginal $P = 5.1 \times 10^{-2}$), all SNPs were replicated at the nominal $P < 0.05$ threshold (Table 2). All seven SNPs surpass the conventional genome-wide significant threshold of $P \leq 5 \times 10^{-8}$ after combined analysis of the GWA and replication cases and controls (Table 2). A conservative adjustment of the rs4141819 total *P* values ($P_{all} = 8.5 \times 10^{-8}$; $P_{stageB} = 4.1 \times 10^{-8}$) for performing two independent GWA studies (all and stage B endometriosis cases versus controls) would produce $P > 5 \times 10^{-8}$ ($P_{all\text{-}adjusted} = 1.7 \times 10^{-7}$; $P_{stageB\text{-}adjusted} = 8.2 \times 10^{-8}$). However, the accurately imputed (Rsq > 0.95) SNP rs2861694 ($P_{stageB} = 7.9 \times 10^{-9}$), in strong LD with rs4141819 ($r^2 = 0.981$, $D' = 1.0$; and $r^2 = 0.867$, $D' = 1.0$, in the 379 European and 286 Asian 1000 Genomes reference samples, respectively), would remain genome-wide significant ($P_{stageB\text{-}adjusted} = 1.6 \times 10^{-8}$).

The Q-Q plots for the QIMRHCS, OX and BBJ GWA data (Supplementary Fig. 17a-c) reflect our stringent quality control, while the GWA meta-analysis Q-Q plot (Supplementary

Fig. 17d), reveals a significant preponderance of small *P* values $<10^{-3}$, suggesting many of these nominally significant SNPs likely represent true signals[12]. To further examine the shared genetic risk across our European and Japanese populations we performed polygenic prediction analysis[13] to evaluate whether the aggregate effects of many variants of small effect in the BBJ GWA cohort, could predict affection status in the European GWA cohorts. The BBJ-derived risk scores significantly predicted affection status in the QIMRHCS (R^2 = 0.0064; $P = 6.9 \times 10^{-7}$), OX ($R^2 = 0.0057$; $P = 9.6 \times 10^{-6}$) and combined QIMRHCS+OX all endometriosis case-control sets ($R^2 = 0.0054$; $P = 8.8 \times 10^{-11}$). For the individual and combined QIMRHCS and OX case-control sets, the variance explained peaked in the SNP sets with BBJ GWA $P < 0.1$, using all GWA meta-analysis SNPs (Fig. 4a) and after excluding all SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1 (Fig. 4b). Analogously, performing the prediction in reverse, the QIMRHCS+OX-derived risk scores significantly predicted affection status in the BBJ case-control set ($R^2 = 0.0106$; $P = 3.3 \times 10^{-6}$) (Supplementary Fig. 18 and Supplementary Note).

A gene-based GWA analysis using VEGAS[14], which accounts for gene size and LD between SNPs, revealed 1,184 genes with a combined $P \leq 0.05$ and the top three ranked genes associated with endometriosis to be *WNT4* on 1p36.12 ($P = 5.0 \times 10^{-9}$), *VEZT* on 12q22 ($P = 5.7 \times 10^{-7}$) and *GREB1* on 2p25.1 ($P = 2.5 \times 10^{-5}$) (Supplementary Table 3). In addition to having genome-wide significant SNPs near these three genes, the *WNT4* and *VEZT* genes easily surpass our conservative gene-based significant association threshold of $P \leq 2.85 \times 10^{-6}$ (calculated as $P = 0.05 / 17{,}538$ independent genes). *WNT4* encodes for wingless-type MMTV integration site family, member 4 and is important for the development of the female reproductive tract[15] and steroidogenesis[16]. *VEZT* encodes vezatin, an adherens junction transmembrane protein that is down regulated in gastric cancer[17]. *GREB1* encodes growth regulation by estrogen in breast cancer 1, an early response gene in the estrogen regulation pathway involved in hormone dependent breast cancer cell growth[18]. For the four remaining implicated regions on 2p14, 6p22.3, 7p15.2 and 9p21.3, no genes were significant ($P \leq 1.3 \times 10^{-3}$) after adjusting VEGAS results for testing 37 genes across all seven regions, see Table 2, Supplementary Figs. 3-9 and Supplementary Table 4.

In conclusion, given their high gene-based ranking, proximity to genome-wide significant SNPs, known pathophysiology and reported gene expression (Supplementary Note and Supplementary Fig. 19), the *WNT4*, *VEZT* and *GREB1* genes are strong targets for further studies aimed at understanding the molecular pathogenesis of endometriosis. Our results also suggest that a considerable number of SNPs nominally implicated (e.g. $P < 0.1$) in the European and Japanese GWA cohorts represent true endometriosis risk loci. Moreover, the significant overlap in common polygenic risk for endometriosis indicates genetic risk prediction and future targeted disease therapy may be transferred across these populations.

ONLINE METHODS

GWA samples and phenotyping

Initially, 2,351 surgically-confirmed endometriosis cases were drawn from women recruited by The Queensland Institute of Medical Research (QIMR) study[19] and a further 1,030 cases were obtained from women recruited by the Oxford Endometriosis Gene (OXEGENE) study. Australian controls consisted of 1,870 individuals recruited by QIMR[2] and 1,244 individuals recruited by the Hunter Community Study (HCS)[7]. UK controls encompassed 6,000 individuals provided by the Wellcome Trust Case Control Consortium 2 (WTCCC2). Approval for the studies was obtained from the QIMR Human Ethics Research Committee, the University of Newcastle and Hunter New England Population Health Human Research

Ethics Committees, and the Oxford regional multi-centre and local research ethics committees. Informed consent was obtained from all participants prior to testing[2].

All Japanese GWA case and control samples were obtained from the BioBank Japan (BBJ) at the Institute of Medical Science, the University of Tokyo. A total of 1,423 cases were diagnosed with endometriosis by the following one or more examinations: multiple clinical symptoms, physical examinations, and laparoscopy or imaging tests. We utilized 1,318 female control samples consisting of healthy volunteers from Osaka-Midosuji Rotary Club, Osaka, Japan and women in the Biobank Japan who were registered to have no history of endometriosis. All participants provided written informed consent to this study. This study was approved by the ethical committees at the Institute of Medical Science, the University of Tokyo and Center for Genomic Medicine, RIKEN Yokohama Institute.

GWA genotyping and quality control (QC)

QIMR and OX cases, and QIMR controls were genotyped at deCODE Genetics on Illumina 670-Quad (cases) and 610-Quad (controls) BeadChips (Illumina Inc), respectively. HCS controls were genotyped at the University of Newcastle on 610-Quad BeadChips (Illumina Inc). The WTCCC2 controls were genotyped at the Wellcome Trust Sanger Institute using Illumina HumanHap1M BeadChips. Genotypes for QIMR cases and controls were called with the Illumina BeadStudio software. Standard quality control procedures were applied as outlined previously[20]. Briefly, individuals with call rates <0.95 then SNPs with a mean BeadStudio GenCall score < 0.7, call rates < 0.95, Hardy-Weinberg equilibrium $P < 10^{-6}$, and minor allele frequency (MAF) < 0.01 were excluded. Cryptic relatedness between individuals was identified through a full identity-by-state matrix. Ancestry outliers were identified using data from 11 populations of the HapMap 3 and five Northern European populations genotyped by the GenomeEUtwin consortium, using EIGENSOFT[21,22]. To increase the power of the Australian GWA dataset we ancestrally matched the existing QIMR cases and controls[2] to individuals from the Hunter Community Study (HCS)[7] genotyped on Illumina 610 chips. After stringent quality control, the resulting QIMRHCS GWA cohort consists of 2,262 endometriosis cases and 2,924 controls, increasing the Australian effective sample size by 24%.[2]

Quality control procedures for the OX genotype data resulted in the removal of SNPs with a genotype call rate < 0.99 and/or heterozygosity < 0.31 or > 0.33. Genome-wide IBS was estimated for each pair of individuals and one individual from each duplicate or related pair (IBS > 0.82) was removed. Genotype data were combined with CEU, CHB&JPT and YRI genotype data from HapMap 3 and individuals of non Northern European ancestry were identified using EIGENSOFT and subsequently removed. SNPs with a genotype call rate < 0.95 were removed, and this threshold was increased to 0.99 for SNPs with MAF < 0.05. In addition, SNPs showing a significant a) deviation from HWE ($P < 1 \times 10^{-6}$); b) difference in call rate between 58BC and NBS control groups ($P < 1 \times 10^{-4}$); c) difference in allele/genotype frequency between control groups ($P < 1 \times 10^{-4}$); d) difference in call rate between cases and controls ($P < 1 \times 10^{-4}$) and e) a MAF < 0.01 were removed.[2]

The BBJ cases and controls were genotyped using the Illumina HumanHap550v3 Genotyping BeadChip. Quality control included sample call rate ≥ 0.98, identity-by-state to exclude close relatedness samples and principal component analysis to exclude non-Asian samples. We also performed SNP quality control (call rate of ≥ 0.99 in both cases and controls and Hardy-Weinberg equilibrium test $P \geq 1.0 \times 10^{-6}$ in controls); 460,945 SNPs on all chromosomes passed the quality control filters and were further analyzed.[1]

GWA meta-analysis

For SNPs passing QC, tests of allelic association (--assoc) were performed using PLINK[23] in the separate QIMRHCS, OX and BBJ GWA datasets. The primary meta-analysis of all endometriosis cases versus controls in the QIMRHCS, OX and BBJ GWA data was performed using a fixed-effect (inverse variance-weighted) model, where the effect size estimates, or β-coefficients, are weighted by their estimated standard errors, utilizing the GWAMA software[24].

The threshold of 7.2×10^{-8} for GWA studies of dense SNPs and resequence data[25] proposed by Dudbridge and Gusnanto[26] was utilized to indicate genome-wide *significant* association, while SNPs with $P \leq 10^{-5}$ were considered to show a *suggestive* association [as used in the online 'Catalog of Published Genome-Wide Association Studies'].

Also, 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)[2], a secondary analysis was performed for suggestive SNPs ($P \leq 10^{-5}$); where the association results from QIMRHCS and OX stage B cases versus controls, were meta-analyzed with the BBJ association results. As previously demonstrated[2], the exclusion of minimal endometriosis cases has the potential to enrich true genetic risk effects, even taking into account the reduced sample size.

Consistency of allelic effects across studies was examined utilizing the *Cochran's Q* test[27]. Between-study (effect) heterogeneity was indicated by *Q statistic P* values < 0.1[28]. Meta-analysis of SNPs associated with fixed-effect $P \leq 10^{-5}$ and showing evidence of effect heterogeneity were also analyzed using the recently developed Han and Eskin's random effects model (RE2) implemented in the Metasoft software[29]. In contrast to the conventional DerSimonian-Laird random effects (RE) model[30], the RE2 model *increases* power under heterogeneity[29].

Genotype imputation analysis

In order to assess the impact of variants not present on the Illumina BeadChips and better define the associated regions, we imputed genotypes within ±2500 kb of the most significant genotyped SNP using the full reference panel from the 1000 Genomes project Interim Phase I Haplotypes (2010-11 data freeze, 2011-06 haplotypes). Imputation was performed separately for the QIMRHCS, OX and BBJ GWA datasets with only the overlapping genotyped SNPs within ±2500kb of the most significant genotyped SNP, using the MaCH and minimac programs[31,32] and following the two-step approach outlined in the online 'Minimac: 1000 Genomes Imputation Cookbook'. Analysis of imputed genotype dosage scores was performed using mach2dat[31,32] and PLINK. The quality of imputation was assessed by means of the Rsq statistic. Results for poorly imputed SNPs, defined as having an Rsq < 0.3, were subsequently removed. The results from association analysis of imputed data in the QIMRHCS, OX and BBJ datasets were then combined via meta-analysis of the β-coefficients weighted by their estimated standard errors using GWAMA.

Replication samples and genotyping

Independent of the BBJ GWA case-control cohort, a total of 1,044 cases and 4,017 controls were obtained from the BioBank Japan and utilized for replication. We note that 653 of these 1,044 cases were also utilized in a small GWA study (Adachi et al. 2010) of 696 cases and 825 controls[9]. To utilize all available association data for rs13394619 maximally, given there is incomplete overlap between the Adachi and our replication cases and zero overlap between the controls, we worked with the published results for rs13394619 in Adachi et al

(2010) and the results from comparing our non-overlapping 391 replication cases to our 4,017 replication controls.

The seven SNPs (rs7521902, rs13394619, rs4141819, rs7739264, rs12700667, rs1537377 and rs10859871) reaching genome-wide significance in the GWA meta-analysis were genotyped in the independent Japanese replication cohort using the multiplex PCR-based Invader assay (Third Wave Technologies), as previously described[1].

Replication and total association analyses

Tests of allelic association were performed using PLINK in the independent Japanese replication cohort. Because only the associations in the same direction would be considered as replicated, one-sided *P* values were obtained by halving the standard (two-sided) PLINK *P* values. To determine the total evidence for association, the one-sided replication *P* values were meta-analyzed with the QIMRHCS, OX, BBJ [and Adachi[9] 500K (290 cases and 262 controls) and 6.0 (406 cases and 563 controls) for rs13394619] GWA *P* values using METAL[33]. The *P* values observed in each case-control cohort were converted into a signed Z-score. Z-scores for each allele were combined across samples in a weighted sum, with weights proportional to the square-root of the sample size for each cohort[34]. Given that our cohorts have unequal numbers of cases and controls, we utilized the effective sample size, where $N_{eff} = 4 / (1 / N_{cases} + 1 / N_{controls})$[33]. We also performed meta-analysis of the β-coefficients weighted by their estimated standard errors using GWAMA to estimate the overall odds ratio and 95% CI for the genome-wide significant SNPs.

Polygenic prediction

The aim of the prediction analysis was to evaluate the aggregate effects of many variants of small effect. We summarized variation across nominally associated loci into quantitative scores and related the scores to disease state in independent samples. Although variants of small effect (e.g., genotype relative risk of 1.05) are unlikely to achieve even nominal significance, increasing proportions of "true" effects will be detected at increasingly liberal *P* value thresholds, e.g. $P < 0.1$ (i.e., ~10% of all SNPs), $P < 0.2$, etc. Using such thresholds, we defined large sets of "allele specific scores" in the "discovery" sample of the Japanese BioBank (BBJ) endometriosis case-control set (1,423 cases, 1,318 controls) to generate risk scores for individuals in the "target" sample of the QIMRHCS (2,262 cases, 2,924 controls), OX (919 cases, 5,151 controls) and combined European (QIMRHCS+OX) endometriosis case-control sets (3,181 cases, 8,075 controls). The term risk score is used instead of risk, as it is impossible to differentiate the minority of true risk alleles from the non-associated variants. In the discovery sample, we selected sets of allele specific scores for SNPs with the following levels of significance; $P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$, $P < 0.5$, $P < 0.6$, $P < 0.7$, $P < 0.8$, $P < 0.9$, $P < 1.0$. For each individual in the target sample, we calculated the number of score alleles that they possessed, each weighted by the log odds ratio from the discovery sample. To assess whether the aggregate scores reflect endometriosis risk, we tested for a higher mean score in cases compared to controls. Logistic regression was used to assess the relationship between target sample disease status and aggregate risk score. Nagelkerke's pseudo R^2 was used to assess the variance explained. Prediction was performed using all 407,632 SNPs overlapping the QIMRHCS, OX and BBJ GWA datasets, and after excluding the 6,163 SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1. We also performed the predictions in reverse, using QIMRHCS +OX-derived risk scores to predict affection status in the BBJ case-control set.

Gene-based association analysis

Gene-based approaches can be more powerful than traditional individual-SNP-based approaches in the presence of allelic heterogeneity. Therefore, utilizing the QIMRHCS, OX

and BBJ GWA data, we performed a genome-wide gene-based association study using VEGAS[14]. Briefly, for the 407,632 overlapping SNPs, the *P* values from the European GWA study (i.e., FE meta-analysis of QIMRHCS and OX GWA data) and the *P* values from the Japanese (BBJ) GWA study were analyzed separately using VEGAS. The VEGAS test incorporates evidence for association from all SNPs across a gene and accounts for gene size (number of SNPs) and LD between SNPs by using simulations from the multivariate normal distribution. The resulting European and Japanese gene-based *P* values were meta-analyzed using Stouffer's Z-score combined p-value method[34]. A total of 17,538 genes (including 50 kb 5′ and 3′ of their transcription start and end site, respectively[14]) contained association results for ≥1 SNP, so a Bonferroni adjusted significance threshold of $P \leq 2.85 \times 10^{-6}$ (0.05 / 17,538) was utilized to indicate genome-wide gene-based *significant* association.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge with appreciation all the women who participated in the QIMR, OX and BBJ studies. We thank Endometriosis Associations for supporting the study recruitment. We also thank the many hospital directors and staff, gynecologists, general practitioners and pathology services in Australia, the UK and the United States who provided assistance with confirmation of diagnoses. We thank Sullivan and Nicolaides Pathology and the Queensland Medical Laboratory Pathology for pro bono collection and delivery of blood samples and other pathology services for assistance with blood collection. The Hunter Community Study team thanks the men and women of the Hunter region who participated in the study.

The QIMR Study was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia (241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496610, 496739, 552485 and 552498), the Cooperative Research Centre for Discovery of Genes for Common Human Diseases (CRC), Cerylid Biosciences (Melbourne) and donations from N. Hawkins and S. Hawkins. D.R.N. was supported by the NHMRC Fellowship (339462 and 613674) and the ARC Future Fellowship (FT0991022) schemes. S.M. was supported by NHMRC Career Development Awards (496674, 613705). E.G.H. (631096) and G.W.M. (339446, 619667) were supported by the NHMRC Fellowships Scheme. We thank B. Haddon, D. Smyth, H. Beeby, O. Zheng, B. Chapman and S. Medland for project and database management, sample processing, genotyping and imputation. We thank Brisbane gynecologist D.T. O'Connor for his important role in initiating the early stages of the project and for confirmation of diagnosis and staging of disease from clinical records of many cases, including 251 in these analyses. We are grateful to the many research assistants and interviewers for assistance with the studies contributing to the QIMR collection. The Hunter Community Study was funded by the University of Newcastle, the Gladys M Brawn Fellowship scheme and the Vincent Fairfax Family Foundation in Australia.

The work presented here was supported by a grant from the Wellcome Trust (WT084766/Z/08/Z) and makes use of WTCCC2 control data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of these data is available from http://www.wtccc.org.uk. Funding for the WTCCC project was provided by the Wellcome Trust under awards 076113 and 085475. C.A.A. was supported by a grant from the Wellcome Trust (098051). A.P.M. was supported by a Wellcome Trust Senior Research Fellowship. S.H.K. is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health NIHR Biomedical Research Centres funding scheme. K.T.Z. is supported by a Wellcome Trust Research Career Development Fellowship (WT085235/Z/08/Z). We thank L. Cotton, L. Pope, G. Chalk and G. Farmer (University of Oxford). We also thank P. Koninckx (Leuven, Belgium), M. Sillem (Heidelberg, Germany), C. O'Herlihy and M. Wingfield (Dublin, Ireland), M. Moen (Trondheim, Norway), L. Adamyan (Moscow, Russia), E. McVeigh (Oxford, UK), C. Sutton (Guildford, UK), D. Adamson (Palo Alto, California, USA) and R. Batt (Buffalo, New York, USA) for providing diagnostic confirmation.

We thank the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

References

1. Uno S, et al. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. Nat Genet. 2010; 42:707–10. [PubMed: 20601957]

2. Painter JN, et al. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. Nat Genet. 2011; 43:51–4. [PubMed: 21151130]
3. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. Fertil Steril. 1999; 71:701–710. [PubMed: 10202882]
4. Montgomery GW, et al. The search for genes contributing to endometriosis risk. Hum Reprod Update. 2008; 14:447–57. [PubMed: 18535005]
5. Gao X, et al. Economic burden of endometriosis. Fertil Steril. 2006; 86:1561–72. [PubMed: 17056043]
6. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. Am J Hum Genet. 2012; 90:7–24. [PubMed: 22243964]
7. McEvoy M, et al. Cohort profile: The Hunter Community Study. Int J Epidemiol. 2010; 39:1452–63. [PubMed: 20056765]
8. American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril. 1997; 67:817–21. [PubMed: 9130884]
9. Adachi S, et al. Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. J Hum Genet. 2010; 55:816–21. [PubMed: 20844546]
10. Goumenou AG, Arvanitis DA, Matalliotakis IM, Koumantakis EE, Spandidos DA. Loss of heterozygosity in adenomyosis on hMSH2, hMLH1, p16Ink4 and GALT loci. Int J Mol Med. 2000; 6:667–71. [PubMed: 11078826]
11. Martini M, et al. Possible involvement of hMLH1, p16(INK4a) and PTEN in the malignant transformation of endometriosis. Int J Cancer. 2002; 102:398–406. [PubMed: 12402310]
12. Yang J, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011; 19:807–12. [PubMed: 21407268]
13. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–52. [PubMed: 19571811]
14. Liu JZ, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010; 87:139–45. [PubMed: 20598278]
15. Vainio S, Heikkila M, Kispert A, Chin N, McMahon AP. Female development in mammals is regulated by Wnt-4 signalling. Nature. 1999; 397:405–9. [PubMed: 9989404]
16. Guo X, et al. Down-regulation of VEZT gene expression in human gastric cancer involves promoter methylation and miR-43c. Biochem Biophys Res Commun. 2011; 404:622–7. [PubMed: 21156161]
17. Boyer A, et al. WNT4 is required for normal ovarian follicle development and female fertility. Faseb J. 2010; 24:3010–25. [PubMed: 20371632]
18. Rae JM, et al. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast Cancer Res Treat. 2005; 92:141–9. [PubMed: 15986123]
19. Treloar SA, 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; 77:365–376. [PubMed: 16080113]
20. Medland SE, et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet. 2009; 85:750–5. [PubMed: 19896111]
21. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
22. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
23. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
24. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC bioinformatics. 2010; 11:288. [PubMed: 20509871]
25. Bajpai AK, et al. MGEx-Udb: a mammalian uterus database for expression-based cataloguing of genes across conditions, including endometriosis and cervical cancer. PLoS One. 2012; 7:e36776. [PubMed: 22606288]

Europe PMC Funders Author Manuscripts

26. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008; 32:227–34. [PubMed: 18300295]
27. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954; 10:101–129.
28. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One. 2007; 2:e841. [PubMed: 17786212]
29. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011; 88:586–98. [PubMed: 21565292]
30. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–88. [PubMed: 3802833]
31. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]
32. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]
33. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
34. Stouffer, SA.; Suchman, EA.; DeVinney, LC.; Star, SA.; Williams, RMJ. Adjustment During Army Life. Princeton University Press; Princeton, NJ: 1949.

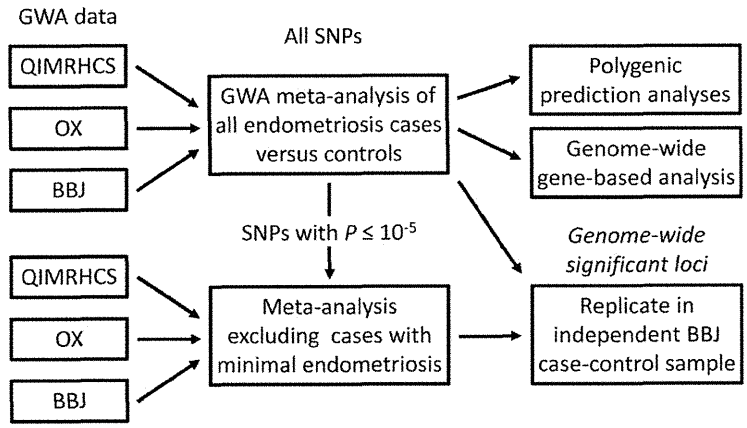

Figure 1.
Study design.

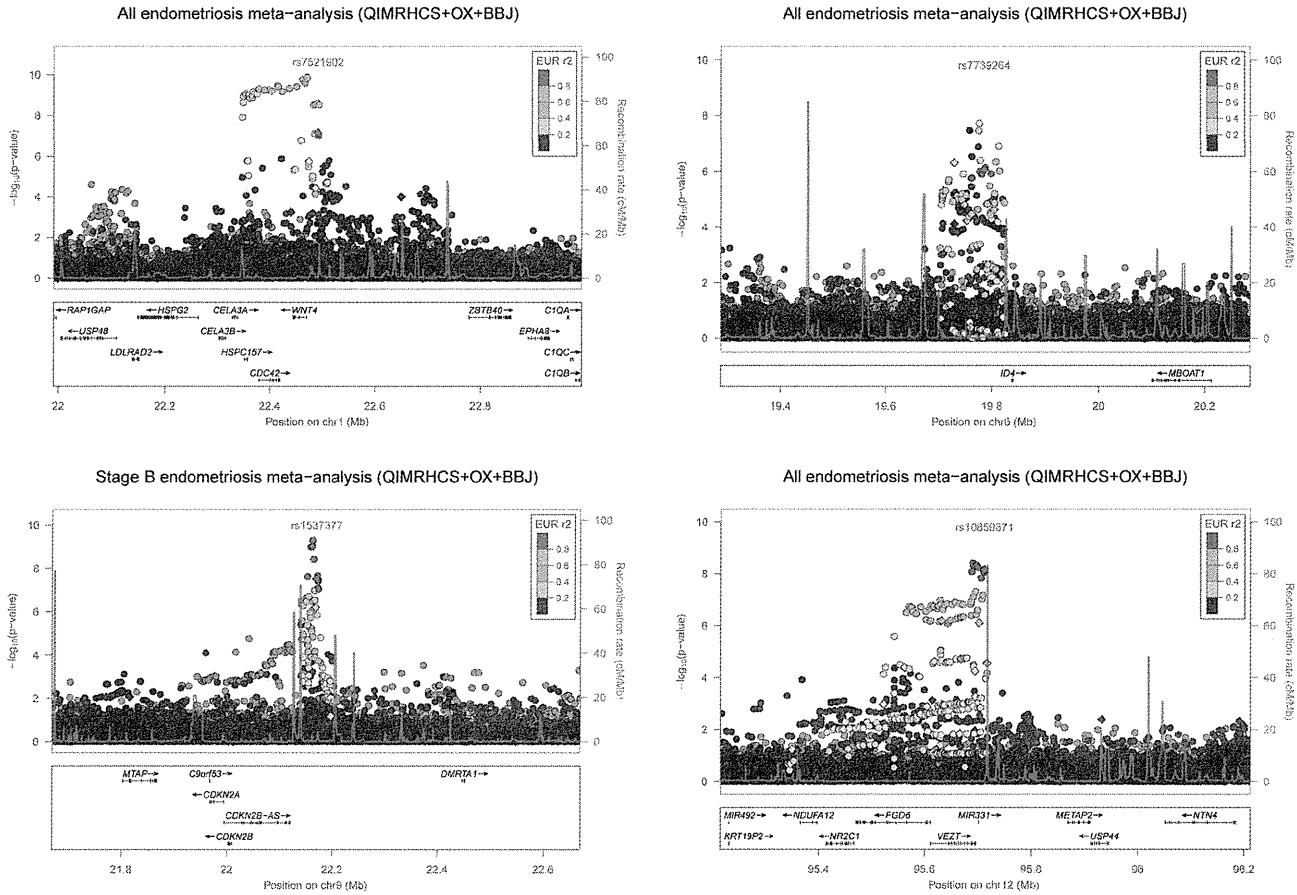

Figure 2.

Evidence for association with endometriosis from the QIMRHCS+OX+BBJ GWA meta-analysis across the 1p36.12 **(a)**, 6p22.3 **(b)**, 9p21.3 **(c)** and 12q22 **(d)** regions following imputation using the 1000 Genomes Project reference panel. Diamond and circle symbols represent genotyped and imputed SNPs, respectively. The most significant genotyped SNP is represented by a purple diamond. All other SNPs are color coded according to the strength of LD with the top genotyped SNP (as measured by r^2 in the European 1000 Genomes data).

Europe PMC Funders Author Manuscripts

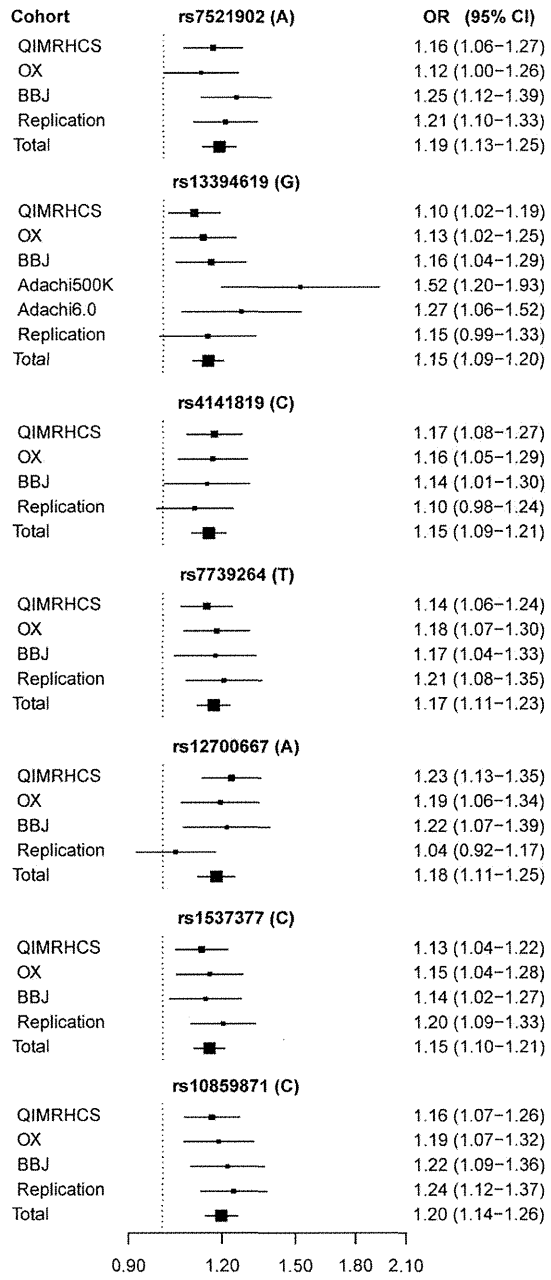

Figure 3.
Forest plots of risk allele effects for the seven genome-wide significant SNP loci in the individual and total endometriosis case-control cohorts.

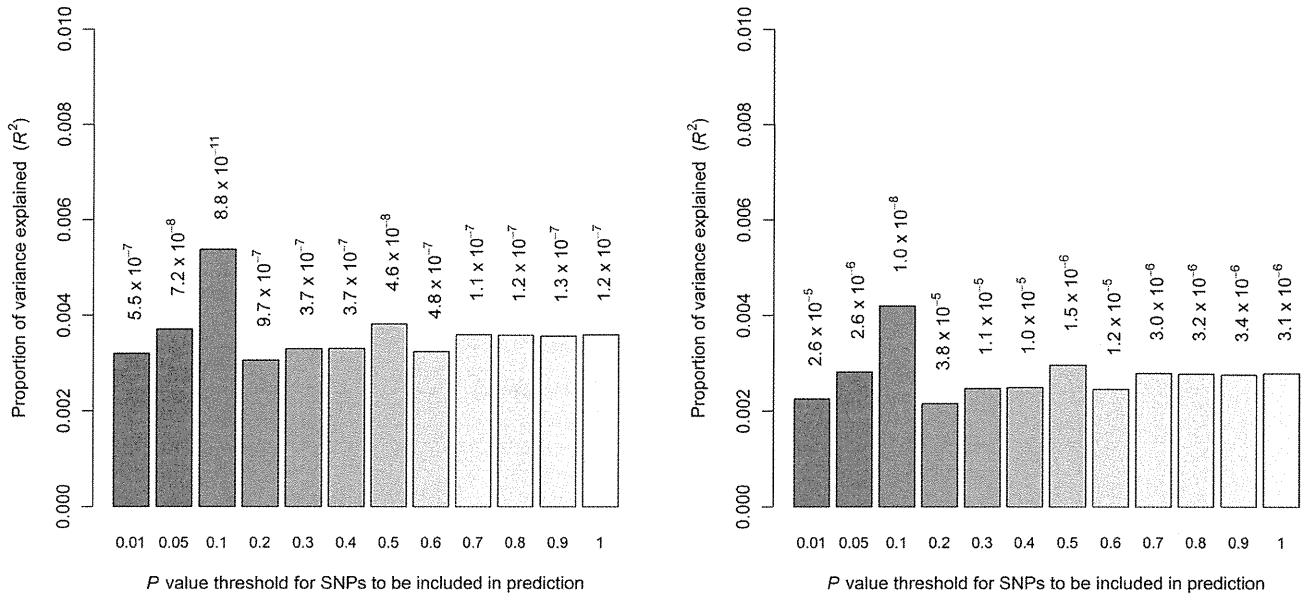

Figure 4.

Allele-specific score prediction for endometriosis, using the BBJ population as the discovery dataset and the QIMRHCS+OX population as the target dataset. The variance explained in the target dataset on the basis of allele-specific scores derived in the discovery dataset for twelve significance thresholds ($P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$, $P < 0.5$, $P < 0.6$, $P < 0.7$, $P < 0.8$, $P < 0.9$, $P < 1.0$, plotted left to right). The y-axis indicates Nagelkerke's pseudo R^2 representing the proportion of variance explained. The number above each bar is the P value for the target dataset prediction analysis (i.e. R^2 significance). Prediction was performed using all GWA meta-analysis SNPs **(a)** and after excluding all SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1 **(b)**. These figures show that the results were not driven by a few highly associated regions, indicating a substantial number of common variants underlie endometriosis risk.

Europe PMC Funders Author Manuscripts

Table 1

Summary of the endometriosis case-control cohorts

Cohort	Ancestry	No. of cases (stage B)	No. of controls
QIMRHCS GWA	European	2,262 (905)	2,924
OX GWA	European	919 (452)	5,151
BBJ GWA	Japanese	1,423	1,318
GWA meta-analysis		4,604	9,393
Replication	Japanese	1,044	4,017
Total		5,648	13,410