Independent replication and meta-analysis for endometriosis risk loci

Yadav Sapkota[1,†,*], Amelie Fassbender[2,3,†], Lisa Bowdler[1], Jenny N Fung[1], Daniëlle Peterse[2,3], Dorien O[2,3], Grant W Montgomery[1,§], Dale R Nyholt[1,4,§] and Thomas M D'Hooghe[2,3,5,§]

[1]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[2]KULeuven, Department of Development and Regeneration, Organ systems, Leuven, Belgium

[3]Department of Obstetrics and Gynaecology, Leuven University Fertility Centre, University Hospital Leuven, Leuven, Belgium

[4]Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia

[5]Division of Reproductive Biology, Institute of Primate Research, Karen, Nairobi, Kenya

[†]Equal contributions; [§]Equal contributions.

Correspondence:

Yadav Sapkota, PhD

Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston 4006 QLD, Australia

Tel. +61 7 3362 0228; Fax. +61 7 3362 0111

E-mail: Yadav.Sapkota@qimrberghofer.edu.au)

Running title: Replication study for endometriosis risk loci

Abstract

Endometriosis is a complex disease that affects 6-10% of women in their reproductive years and 20-50% of women with infertility. Genome-wide and candidate-gene association studies for endometriosis have identified ten independent risk loci, and of these, nine (rs7521902, rs13394619, rs4141819, rs6542095, rs1519761, rs7739264, rs12700667, rs1537377 and rs10859871) are polymorphic in European populations. Here, we investigate replication of the nine SNP loci in 998 laparoscopically- and histologically-confirmed endometriosis cases and 783 disease-free controls from Belgium. SNPs rs7521902, rs13394619 and rs6542095 show nominally significant ($P < 0.05$) associations with endometriosis, while the directions of effect for seven SNPs are consistent with the original reports. Association of rs6542095 at the *IL1A* locus with "All" ($P = 0.066$) and "Grade_B" ($P = 0.01$) endometriosis is noteworthy because this is the first successful replication in an independent population. Meta-analysis with the published results yields genome-wide significant evidence for rs7521902, rs13394619, rs6542095, rs12700667, rs7739264 and rs1537377. Notably, three coding variants in *GREB1* (near rs13394619) and *CDKN2B-AS1* (near rs1537377) also showed nominally significant associations with endometriosis. Overall, this study provides important replication in a uniquely characterised independent population and indicates that the majority of the original genome-wide association findings are not due to chance alone.

Keywords : endometriosis, genome-wide association study, replication, meta-analysis, IL1A

Endometriosis is the most common cause of pelvic pain that affects 6-10% of women in their reproductive years (Treloar et al., 1999) and 20-50% of women with infertility (Gao et al., 2006). It is primarily characterized by the presence of endometrium-like tissue in cavities other than the uterus. Besides severe pelvic pain, women with endometriosis may also suffer from heavy or irregular menstrual bleeding, pain during intercourse and exercise, infertility, lower abdominal and back pain, diarrhoea and/or constipation and chronic fatigue. Endometriosis has a complex aetiology, which results from an interplay of genetic and non-genetic risk factors and has an estimated heritability of approximately 51% (Treloar et al., 1999).

We and others have conducted genome-wide association (GWA) studies involving individuals of European and Japanese ancestries to identify genetic risk factors for endometriosis (Albertsen et al., 2013; Nyholt et al., 2012; Painter et al., 2011; Uno et al., 2010). Results from the four GWA studies strongly associated nine independent single nucleotide polymorphism (SNP) loci with endometriosis providing genome-wide significant ($P < 5 \times 10^{-8}$) evidence in at least one study. These risk loci include: rs10965235 in the *CDKN2BAS* gene at 9p21.3, rs12700667 at 7p15.2, rs7521902 near *WNT4* at 1p36.12, rs13391619 in the *GREB1* at 2p25.1, rs10859871 near *VEZT* at 12q22, rs4141819 at 2p14, rs7739264 near *ID4* at 6p22.3, rs1537377 near *CDKN2B-AS1* at 9p21.3, and rs1519761 at 2q23.3. The *GREB1* signal was implicated in a meta-analysis of the European (QIMRHCS+OX) and Japanese (BBJ) GWA data after combining published results for rs1339416 from Adachi et al. (2010) (Adachi et al., 2010) - a small GWA study comprising 696 endometriosis cases and 825 controls of Japanese descent. A recent follow-up study by ourselves further implicated *GREB1* locus in endometriosis risk, with stronger association signals within the region (Fung J.N. et al., 2015). Three risk loci (rs4141819 at 2p14, rs7739264 near *ID4* at 6p22.3 and rs1537377 near *CDKN2B-AS1* at 9p21.3) were implicated

when analysis was conducted after excluding endometriosis cases with minimal or mild [revised American Fertility Society (rAFS) (American Society for Reproductive Medicine, 1997) stage 1 or 2 disease] endometriosis (Nyholt et al., 2012). More recently, we conducted a candidate-gene association study (Sapkota et al., 2015) to investigate the potential role of the interleukin 1A (*IL1A*) variants reported by two small Japanese GWA studies for endometriosis (Adachi et al., 2010; Hata et al., 2013). The study results provided genome-wide significant evidence for association of rs6542095 in the *IL1A* with endometriosis. Taken together, results suggest a total of ten independent SNP loci for endometriosis at genome-wide significant level, and of these nine are polymorphic in populations of European ancestry.

The three replication studies (Pagliardini et al., 2015; Pagliardini et al., 2013; Sundqvist et al., 2013) performed to date, have only replicated association of rs7521902 ($P = 5.6 \times 10^{-3}$) (Pagliardini et al., 2013) and rs10859871 ($P = 6.9 \times 10^{-5}$) (Pagliardini et al., 2015) with endometriosis. Inherently hidden fine stratification among different populations of similar ancestry, which is often difficult to tease apart, power and possible variations in disease definition and/or classification may have contributed to these inconsistent results. Moreover, these replication studies have investigated only a fraction of the implicated SNP loci and hence additional replication studies are required to examine association of all the risk loci with endometriosis.

Here, we extended our previous study (Sundqvist et al., 2013) to evaluate all nine implicated SNP loci for endometriosis that are polymorphic in populations of European ancestry, utilizing GWA data for surgically confirmed 998 endometriosis cases and 783 disease-free controls from Belgium. Importantly, disease severity in the endometriosis cases have been prospectively graded using the rAFS classification system and hence are less likely to be misclassified as compared to the retrospective disease staging based on clinical records.

We also performed meta-analysis for the nine risk loci, after combining results from the current study with the relevant published results. Finally, we conducted GWA analysis of the observed data in the Belgian cohort to see if there were any novel loci associated with endometriosis risk identified in this sample.

Materials and Methods

Study participants

Cases (n = 1,077) and controls (n = 900) included in this study were recruited at the Leuven University Hospital, Belgium during 1993 - 2012, who had undergone laparoscopy for subfertility with or without pain. Presence of endometriosis in cases was laparoscopically- and histologically-confirmed based on electronic medical file records. The disease severity in women with endometriosis was prospectively graded according to the rAFS classification system. Endometriosis cases had either minimal (stage I; n = 380), mild (stage II, n = 229), moderate (stage III, n = 174), severe (stage IV; n = 284) or unknown (n = 10) disease. Absence of endometriosis in controls was confirmed laparoscopically. Both cases and controls were Caucasian in origin. All the study participants provided written informed consent and the study was approved by the Commission of Medical Ethics of the Leuven University Hospital, Belgium and QIMR Berghofer Human Ethics Research Committee, Australia.

DNA extraction and genotyping

DNA was purified from EDTA-stabilized whole blood collected for routine molecular diagnostic tests at the Centre for Human Genetics of University Hospitals, Leuven, Belgium. Following manufacturer's protocol, DNA was purified using Chemagic DNA blood special kit (Chemagen MSM I, PerkinElmer chemagen Technologies GmbH, Baesweiler, Germany)

based on the specific binding of DNA to paramagnetic beads, and Auto Pure LS Puregene chemistry (Qiagen, Venlo, The Netherlands) based on salting-out extraction and a manual salting-out procedure (home-brew). The choice of the extraction method was based on the available amount of blood and on the type of required molecular diagnostic test. DNA concentration was measured using Victor (PerkinElmer, Massachussetts, USA).

Whole genome genotyping of the DNA samples was performed using the Illumina HumanCoreExome 12v1.1 array at the Molecular Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, Australia, following manufacturer's standard protocol. For a quality control (QC) check, DNA concentrations of majority of the samples were re-measured on the BioTech Powerwave at the QIMR Molecular Epidemiology Laboratory, before genotyping. The Illumina HumanCoreExome genotyping arrays are the newer generations of Illumina GWA arrays, which consist of ~250,000 common tag SNPs ('Core') and ~250,000 predominantly rare coding variants ('exome'). The exome variants in the Illumina HumanCoreExome were carefully selected based on exome sequencing data in ~12,000 individuals.

Genotype calling and QC

Genotype data were called using a custom cluster file generated using ~2,000 good quality (<1% missing rate) samples and the GenCall algorithm within Illumina Genome Studio. Data were then further processed by zCall (Goldstein et al., 2012), a rare variant caller, to attempt to re-call missing genotypes. Following guidelines by the manufacturer and protocols developed for Exome chip data, quality control measures were applied to the Belgian GWA data. Briefly, samples with > 1% missing rates, outlying heterozygosity, non-European ancestries based on 1000 Genomes European populations, cryptic relatedness (pi-hat > 0.2) and gender discordances were excluded. Similarly, markers with poor separation of three

genotype clusters, excess heterozygosity, outlying mean theta and intensity values for heterozygote genotypes, > 1% missing rates, Hardy-Weinberg Equilibrium (HWE) $P < 10^{-6}$ in controls, and minor allele frequency (MAF) < 0.05% in cases and in controls were dropped.

Association analysis

Since rs1096523 in the *CDKN2BAS* gene at 9p21.3 is monomorphic in populations of European ancestry, we considered the remaining nine SNP loci (rs7521902, rs13394619, rs4141819, rs6542095, rs1519761, rs7739264, rs12700667, rs1537377 and rs10859871) for further analysis. To see if there were any novel association signals for endometriosis in the Belgian cohort, GWA analysis of the observed genotypes was performed using --assoc (for 'core' SNPs) and --fisher (for 'exome' variants) commands in Plink for data including all endometriosis cases ("All") and controls. Considering the relatively greater genetic loading of moderate-to-severe (rAFS stage III/IV or "Grade_B") endometriosis compared to mild or minimal (rAFS stage I/II or "Grade_A") endometriosis (Nyholt et al., 2012; Painter et al., 2011; Sapkota et al., 2015), additional analysis for "Grade_B" endometriosis cases versus controls was also performed. Strengths of association of SNPs with endometriosis are reported in terms of odds ratio (ORs) and confidence intervals (CIs).

Imputation and Meta-analysis

Of the nine SNPs, only six (rs7521902, rs13394619, rs4141819, rs6542095, rs7739264 and rs12700667) are assayed on Illumina HumanCoreExome 12v1.1 genotyping platform. Therefore, we imputed genotypes in chromosomes containing the nine SNP loci in the Belgian GWA data, using a reference panel of 1000 Genomes Project (March 2012 release). Imputation was carried out using SHAPEIT (Delaneau et al., 2012) and minimac programs (Li et al., 2009; Li et al., 2010) and following the two-step approach outlined in the online

Minimac: 1000 Genomes Imputation Cookbook (http://genome.sph.umich.edu/wiki/Minimac:_1000_Genomes_Imputation_Cookbook). Quality of the imputed genotypes was assessed by R^2 metric, which estimates the squared correlation between true and imputed genotypes. Poorly imputed SNPs indicated by $R^2 < 0.3$ were excluded from the downstream analyses. Association analyses of imputed genotype dosage scores of the nine SNP loci were conducted using Plink for "All" and "Grade_B" endometriosis cases separately.

After combining results of imputed dosage scores for the nine SNP loci from Belgian data with the published results from Nyholt *et al.* (2012) (Nyholt et al., 2012), Adachi *et al.* (2010) (Adachi et al., 2010), Albertsen *et al.* (2013) (Albertsen et al., 2013) and Sapkota *et al.* (2015) (Sapkota et al., 2015), we performed meta-analysis for "All" endometriosis cases and controls. A brief summary of the datasets used in this study is provided in **Table 1**. Results from Nyholt *et al.* (2012) (Nyholt et al., 2012) included summary statistics of the rs7521902, rs13394619, rs4141819, rs7739264, rs12700667, rs1537377 and rs10859871 obtained from the European (QIMRHCS+OX) and Japanese (BBJ) GWA data (**Table 1**). Similarly, we included results of the rs13394619 and rs6542095 from Adachi *et al.* (2010) (Adachi et al., 2010), obtained from combined analysis of 500K and 6.0 arrays in 696 cases and 825 controls of Japanese ancestry. Furthermore, results from Albertsen *et al.* (2013) (Albertsen et al., 2013) included summary statistics of the rs1519761 in their discovery and replication stages. Finally, we obtained imputed results of the rs6542095 in the QIMRHCS, OX and BBJ imputed data from Sapkota *et al.* (2015) (Sapkota et al., 2015).

Initial meta-analysis was conducted using a fixed-effects (inverse variance-weighted) model implemented in the GWAMA program (Magi & Morris, 2010). Heterogeneity of allelic associations was examined using the Cochran's Q statistic $P_{het} < 0.1$ (Cochran, 1954), as well as the I^2 index (Ioannidis et al., 2007), which indicates the proportion of variance

attributable to between-study heterogeneity. Meta-analysis of SNPs associated in fixed-effects model withevidence of heterogeneity ($P < 0.1$) was carried out using the Han Eskin random-effects model (RE2) (Han & Eskin, 2011) implemented in the METASOFT program. In contrast to the conventional random-effects model, the RE2 model increases power under heterogeneity. Furthermore, additional meta-analysis for the nine SNP loci was also performed by restricting to "Grade_B" endometriosis cases (wherever available) versus controls.

Results

Following the QC steps, a total of 998 endometriosis cases and 783 disease-free controls with 316,467 markers remained in the Belgian GWA data for downstream analysis. Of these, 246,071 were 'core' SNPs whereas 70,396 were 'exome' variants. The GWA analysis of observed genotypes of the 316,467 markers in the Belgian GWA study alone did not produce any genome-wide significant hits in either "All" or "Grade_B" analysis with few suggestive ($P < 1 \times 10^{-5}$) associations (data not shown). All nine SNPs were accurately imputed with $R^2 > 0.95$. Furthermore, we also compared imputed genotypes (dosage scores) of the six SNP loci (rs7521902, rs13394619, rs4141819, rs6542095, rs7739264 and rs12700667) for endometriosis with the observed true genotypes available in the Belgian data. Genotype concordances (as measured by the Pearson's correlation coefficient) between two sets of genotypes for the six SNPs were > 0.99 ($P < 2.2 \times 10^{-16}$).

Association analysis of the dosage scores of the nine implicated SNP loci in the Belgian data provided further insights into associations of these SNPs with endometriosis. Risk alleles and their frequencies of all nine SNPs were similar to the ones reported in the original studies (**Table 2**) (Albertsen et al., 2013; Nyholt et al., 2012; Painter et al., 2011; Sapkota et al., 2015) and their associations were stronger with "Grade_B" than "All"

endometriosis. Furthermore, effect directions of seven out of nine tested SNPs in either "All" or "Grade_B" endometriosis were in line with the published results. Three SNPs showed statistically significant association with endometriosis in either "All" or "Grade_B" disease at a nominal $P < 0.05$. SNP rs7521902 showed borderline marginal association (OR = 1.13; P = 0.12) with "All" endometriosis. As expected, its association was much stronger and statistically significant (OR = 1.30; P = 0.007) with "Grade_B" cases. A statistically significant association (OR = 1.14; P = 0.045) for rs13394619 was also observed for "All" endometriosis; however, the signal was slightly weaker (OR = 1.13; P = 0.164) in "Grade_B" cases. A borderline association (OR = 1.14; P = 0.06) with "All" endometriosis was observed for rs6542095, which was stronger and significant (OR = 1.26; P = 0.01) in "Grade_B" cases.

Meta-analysis including imputed data from Belgian cohort and the published results provide insights into SNP loci associated with endometriosis. Six SNP loci showed associations with either "All" or "Grade_B" endometriosis at genome-wide significance level ($P < 5 \times 10^{-8}$) and with similar directions of effect across all studies included in the analysis (**Table 3**). Of these, three SNPs were associated with both "All" and "Grade_B" endometriosis with a genome-wide significant evidence in the fixed-effects meta-analysis. These include: SNP rs7521902 near *WNT4* ["All" (OR = 1.17; 95% CI = 1.11-1.23; $P = 3.63 \times 10^{-8}$), "Grade_B" (OR = 1.25; 95% CI = 1.17-1.34; $P = 1.72 \times 10^{-10}$)], rs13394619 in *GREB1* ["All" (OR = 1.15; 95% CI = 1.10-1.20; $P = 9.13 \times 10^{-9}$), "Grade_B" (OR = 1.17; 95% CI = 1.11-1.24; $P = 3.02 \times 10^{-8}$)] and rs12700667 at 7p15.2 ["All" (OR = 1.19; 95% CI = 1.13-1.26; $P = 7.10 \times 10^{-10}$), "Grade_B" (OR = 1.29; 95% CI = 1.20-1.39; $P = 1.47 \times 10^{-11}$)]. The *IL1A* SNP (rs6542095) was genome-wide significantly associated with only "Grade_B" (OR = 1.22; 95% CI = 1.14-1.30; $P = 1.00 \times 10^{-9}$) endometriosis in fixed-effects meta-analysis, but after appropriately modelling for between-study heterogeneity in "All" endometriosis (P_{het} = 0.007) in the RE2 model, the association reached genome-wide

significance ($P = 3.35 \times 10^{-8}$). Statistical significance of association of the rs6542095 with "Grade_B" endometriosis also became stronger ($P = 4.90 \times 10^{-10}$) in the RE2 model, after accounting for between-study heterogeneity ($P_{het} = 0.01$). A strong association between rs7739264 near *ID4* and "All" endometriosis ($P = 1.93 \times 10^{-7}$) was observed, and the signal was further enriched in "Grade_B" endometriosis (OR = 1.20; 95% CI = 1.12-1.27; $P = 1.98 \times 10^{-8}$) achieving genome-wide significance. Similarly, near genome-wide significant evidence for association between rs1537377 near *CDKN2B-AS1* and "Grade_B" endometriosis (OR = 1.19; 95% CI = 1.12-1.27; $P = 9.27 \times 10^{-8}$) was observed, which was genome-wide significant ($P = 4.80 \times 10^{-8}$) after modelling for borderline between-study heterogeneity ($P_{het} = 0.10$). However, the effect of rs1537377 in "All" endometriosis was in opposite direction to the published results (**Table 2**).

While the remaining three SNPs (rs4141819, rs1519761 and rs10859871) did not produce genome-wide significant evidence for association with either "All" or "Grade_B" endometriosis in fixed-effects meta-analysis, they still showed strong associations with the disease ($P < 2.38 \times 10^{-5}$). Nonetheless, the effects of rs4141819 and rs1519761 in both "All" and "Grade_B" endometriosis, and rs10859871 in "All" endometriosis were in opposite directions to the published results (**Table 2**). SNP rs4141819 showed between-study heterogeneity ($P_{het} < 0.01$) in both "All" and "Grade_B" endometriosis and after accounting for this heterogeneity in the RE2 model, association of rs4141819 with "Grade_B" disease became stronger with near genome-wide significant evidence ($P = 3.63 \times 10^{-7}$). Significant between-study heterogeneity for rs1519761 was also observed in both "All" and "Grade_B" endometriosis, but its association with the disease ["All" ($P = 5.62 \times 10^{-6}$), "Grade_B" ($P = 1.94 \times 10^{-6}$)] slightly diluted in the RE2 model ["All" ($P = 1.07 \times 10^{-5}$), "Grade_B" ($P = 7.99 \times 10^{-6}$)]. A near genome-wide significant association between rs10859871 near *VEZT* and "All" endometriosis (OR = 1.16; 95% CI = 1.09-1.22; $P = 4.29 \times 10^{-7}$) was observed, with

slightly larger effect size (OR = 1.17; 95% CI = 1.10-1.25) in "Grade_B" disease although statistical significance of the signal was weaker ($P = 2.46 \times 10^{-6}$).

Discussion

Endometriosis is a complex disease and studies have shown that genetic risk factors substantially contribute to risk of endometriosis. Genetic studies, especially the GWA studies for endometriosis have identified ten SNP loci, of which nine are polymorphic in the populations of European origin (Albertsen et al., 2013; Nyholt et al., 2012; Painter et al., 2011; Sapkota et al., 2015). While much larger and well-powered GWA studies are needed to identify additional genetic risk factors involved in risk of endometriosis, replication studies are crucial to provide credibility that the initial genotype-phenotype associations are valid. Repeated observation of such associations in independent populations of similar ethnicity adds evidence that the associations are not due to chance alone. The previous two replication studies for endometriosis have investigated only a handful of the ten implicated SNP loci to date. Here, we report the most comprehensive replication study performed to date, in which we examine all nine implicated SNP risk loci for endometriosis that are polymorphic in populations of European ancestry, by utilizing GWA data in uniquely characterized 998 endometriosis cases and 783 controls from Belgium.

The risk alleles and their frequencies for all the nine SNPs in the Belgian replication cohort were comparable to the original studies (**Table 2**) (Albertsen et al., 2013; Nyholt et al., 2012; Painter et al., 2011; Sapkota et al., 2015). Moreover, direction of effects for seven of the nine SNPs for either "All" or "Grade_B" endometriosis were also consistent with the published results. Among these, we could successfully replicate associations of three SNPs (rs7521902, rs13394619 and rs6542095) with either "All" or "Grade_B" endometriosis at nominal $P < 0.05$, which is more often than by chance alone ($P = 0.008$; one-sided binomial

test). Significant association of rs6542095 at the *IL1A* locus with "All" (P = 0.066) and "Grade_B" (P = 0.01) endometriosis is noteworthy as this is the first successful replication in an independent population, providing further supporting evidence for a potential link between inflammation and endometriosis pathogenesis. More importantly, all the SNPs showed larger effect sizes with "Grade_B" than "All" endometriosis - an observation consistent with previous reports supporting greater genetic loading in moderate-to-severe disease (Nyholt et al., 2012; Painter et al., 2011; Sapkota et al., 2015).

Our meta-analysis including results from the current replication study and the published results produced genome-wide significant evidence for six (rs7521902 near *WNT4*, rs13394619 in *GREB1*, rs6542095 in *IL1A*, rs7739264 near *ID4*, rs12700667 at 7p15.2 and rs1537377 near *CDKN2B-AS1*) of the nine implicated SNPs in either "All" or "Grade_B" endometriosis, after accounting for between-study heterogeneity using the RE2 model, wherever appropriate (**Table 3**). With the exception of rs1519761 at 2q23.3 reported by Albertsen et al. (2013), the other two SNP loci (rs4141819 at 2p14 and rs10859871 near *VEZT*) also showed near genome-wide significance for "Grade_B" endometriosis in the RE2 model ($P = 3.63 \times 10^{-7}$) and for "All" endometriosis in the fixed-effects model ($P = 4.29 \times 10^{-7}$), respectively. The association signal for rs1519761 was the weakest ["All" ($P = 5.62 \times 10^{-6}$), "Grade_B" ($P = 1.94 \times 10^{-6}$)] amongst the nine risk loci and the signal was slightly diluted after accounting for observed between-study heterogeneity in the RE2 model ["All" ($P = 1.07 \times 10^{-5}$), "Grade_B" ($P = 7.99 \times 10^{-6}$)]. Association signal at this locus was also not replicated in a recent meta-analysis for endometriosis (Rahmioglu et al., 2014), suggesting further investigation is required to confirm a role for this locus in risk of endometriosis.

In our multi-ethnic GWA meta-analysis that strongly associated seven risk loci with endometriosis, we found stronger associations at six loci (rs56318008 at 1p36.12, rs77294520 at 2p25.1, rs2861694 at 2p14, rs6901079 at 6p22.3, rs7041895 at 9p21.3 and rs11107968 at

12q22), when we imputed genotypes in the region 2,500 kb upstream and downstream of the most significant genotyped SNP using the full reference panel from the 1000 Genomes Project Interim Phase 1 Haplotypes (2010-2011 data freeze). For the risk loci at 7p15.2, the genotyped SNP rs12700667 was the best signal. For the remaining six loci with stronger association signals ('best' SNPs) post-imputation than the genotyped SNP, we assessed for their replication in the Belgian cohort (**Table 4**). All six SNPs were accurately imputed in the current study with $R^2 > 0.85$. Association results were consistent with that of the original genotyped SNPs as shown in **Table 2**, in particular for the SNPs rs56318008 at 1p36 and rs77294520 at 2p25.1 that showed nominally significant associations with "All" ($P < 0.098$) and "Grade_B" ($P < 0.051$), providing further supporting evidence for implication of these risk loci in endometriosis.

As a first step to help identify causal variants at the nine SNP loci, we interrogated the ExomeChip data for putatively functional coding variants within genes harbouring or closest to the GWA SNPs. For the *GREB1* locus, we found another coding variant rs10929757 showing nominally significant association ($P = 0.015$) with endometriosis. The effect size of rs10929757 was similar (OR = 1.18) to the GWA SNP rs13394619 although they are poorly correlated ($r^2 = 0.25$). Similarly, we also observed nominally significant association ($P < 0.018$) for two coding variants (rs2383207 and rs4977574) in *CDKN2B-AS1* - the closest gene to the GWA SNP rs1537377 at 9p21.3. Despite lack of correlation ($r^2 = 0.011$ and 0.005, respectively) with rs1537377, the effect sizes for both variants were similar (ORs = 1.17 and 1.19, respectively). While these data may suggest independent association signals at *GREB1* and 9p21.3 loci, the three coding variants need to be further investigated in larger sample size for a more conclusive interpretation. We did not observe evidence of association for other coding variants at nominal $P < 0.05$ even though the effect sizes for some were comparable to the GWA SNPs at each risk loci (data not shown). We did not detect any rare

coding variants at the GWA loci in spite from the ExomeChip data. This may, in part, be due to the reduced power in the Belgian sample to detect such rare variants. Assuming a disease prevalence of 8%, our sample size only had 45% power to detect alleles of frequency 0.20 contributing to genotype relative risk of 1.15 (Purcell et al., 2003). As such, larger ExomeChip studies may be required to adequately investigate potential role of the coding/rare variants in risk of endometriosis and other complex traits. We cannot rule out the possibility of other types of rare functional variants at these loci - that are not adequately captured by either ExomeChip or current imputation methods - contributing to increased risk of endometriosis. These issues may be addressed by future studies utilizing larger sample sizes coupled with re-sequencing and further fine-mapping required to identify causal variants within the implicated GWA loci. Furthermore, controls used in this study are clinic-based endometriosis-free individuals who presented with symptoms of subfertility. As such, they may have different allele frequencies compared to 'population-based' controls used in the most GWA studies and hence caution should be used in interpreting these results.

Overall, results from the current replication study provide further supporting evidence for associations of the implicated SNP loci with endometriosis. Meta-analysis for these loci after including additional published results produced genome-wide significant evidence for the six loci, with similar magnitudes and directions of effect across studies and hence provided further evidence against any possibility of inflated genetic effects due to the winner's curse bias in the original study. More importantly, all the nine SNPs showed larger effect sizes with stage III/IV endometriosis than all cases, corroborating our previous observation for greater genetic loading in moderate-to-severe endometriosis.

Conflict of Interest

None.

Acknowledgements

We would like to all the participants in the endometriosis studies that are included in this analysis. We also thank many hospital directors and staff, gynaecologists, general practitioners and pathology services who provided assistance with confirmation of diagnoses. D.R.N. was supported by an NHMRC Fellowship (613674) and ARC Future Fellowship (FT0991022) schemes and G.W.M. was supported by the NHMRC Fellowships Scheme (339446, 619667).

References

Adachi, S., Tajima, A., Quan, J., Haino, K., Yoshihara, K., Masuzaki, H., . . . Tanaka, K. (2010). Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population. *J Hum Genet, 55*(12), 816-821. doi: 10.1038/jhg.2010.118

Albertsen, H. M., Chettier, R., Farrington, P., & Ward, K. (2013). Genome-wide association study link novel loci to endometriosis. *PLoS One, 8*(3), e58257. doi: 10.1371/journal.pone.0058257

American Society for Reproductive Medicine. (1997). Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil Steril, 67*(5), 817-821.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101-129.

Delaneau, O., Marchini, J., & Zagury, J. F. (2012). A linear complexity phasing method for thousands of genomes. *Nat Methods, 9*(2), 179-181. doi: 10.1038/nmeth.1785

Fung J.N., Holdsworth-Carson S.J., Sapkota Y., Zhao Z.Z., Jones L., Girling J.E., . . . G.W., M. (2015). Functional Evaluation of Genetic Variants Associated with Endometriosis near GREB1. *Hum Reprod. In Press.*

Gao, X., Outley, J., Botteman, M., Spalding, J., Simon, J. A., & Pashos, C. L. (2006). Economic burden of endometriosis. *Fertil Steril, 86*(6), 1561-1572. doi: 10.1016/j.fertnstert.2006.06.015

Goldstein, J. I., Crenshaw, A., Carey, J., Grant, G. B., Maguire, J., Fromer, M., . . . Neale, B. M. (2012). zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics, 28*(19), 2543-2545. doi: 10.1093/bioinformatics/bts479

Han, B., & Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet, 88*(5), 586-598. doi: 10.1016/j.ajhg.2011.04.014

Hata, Y., Nakaoka, H., Yoshihara, K., Adachi, S., Haino, K., Yamaguchi, M., . . . Tanaka, K. (2013). A nonsynonymous variant of IL1A is associated with endometriosis in Japanese population. *J Hum Genet, 58*(8), 517-520. doi: 10.1038/jhg.2013.32

Ioannidis, J. P., Patsopoulos, N. A., & Evangelou, E. (2007). Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One, 2*(9), e841. doi: 10.1371/journal.pone.0000841

Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype imputation. *Annu Rev Genomics Hum Genet, 10*, 387-406. doi: 10.1146/annurev.genom.9.081307.164242

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol, 34*(8), 816-834. doi: 10.1002/gepi.20533

Magi, R., & Morris, A. P. (2010). GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics, 11*, 288. doi: 10.1186/1471-2105-11-288

Nyholt, D. R., Low, S. K., Anderson, C. A., Painter, J. N., Uno, S., Morris, A. P., . . . Montgomery, G. W. (2012). Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet, 44*(12), 1355-1359. doi: 10.1038/ng.2445

Pagliardini, L., Gentilini, D., Sanchez, A. M., Candiani, M., Vigano, P., & Di Blasio, A. M. (2015). Replication and meta-analysis of previous genome-wide association studies confirm vezatin as the locus with the strongest evidence for association with endometriosis. *Hum Reprod*. doi: 10.1093/humrep/dev022

Pagliardini, L., Gentilini, D., Vigano, P., Panina-Bordignon, P., Busacca, M., Candiani, M., & Di Blasio, A. M. (2013). An Italian association study and meta-analysis with previous GWAS confirm WNT4, CDKN2BAS and FN1 as the first identified susceptibility loci for endometriosis. *J Med Genet, 50*(1), 43-46. doi: 10.1136/jmedgenet-2012-101257

Painter, J. N., Anderson, C. A., Nyholt, D. R., Macgregor, S., Lin, J., Lee, S. H., . . . Zondervan, K. T. (2011). Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet, 43*(1), 51-54. doi: 10.1038/ng.731

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics, 19*(1), 149-150.

Rahmioglu, N., Nyholt, D. R., Morris, A. P., Missmer, S. A., Montgomery, G. W., & Zondervan, K. T. (2014). Genetic variants underlying risk of endometriosis: insights from meta-analysis of

eight genome-wide association and replication datasets. *Hum Reprod Update, 20*(5), 702-716. doi: 10.1093/humupd/dmu015

Sapkota, Y., Attia, J., Gordon, S. D., Henders, A. K., Holliday, E. G., Rahmioglu, N., . . . Nyholt, D. R. (2015). Genetic burden associated with varying degrees of disease severity in endometriosis. *Mol Hum Reprod*. doi: 10.1093/molehr/gav021

Sapkota, Y., Low, S. K., Attia, J., Gordon, S. D., Henders, A. K., Holliday, E. G., . . . Nyholt, D. R. (2015). Association between endometriosis and the interleukin 1A (IL1A) locus. *Hum Reprod, 30*(1), 239-248. doi: 10.1093/humrep/deu267

Sundqvist, J., Xu, H., Vodolazkaia, A., Fassbender, A., Kyama, C., Bokor, A., . . . Falconer, H. (2013). Replication of endometriosis-associated single-nucleotide polymorphisms from genome-wide association studies in a Caucasian population. *Hum Reprod, 28*(3), 835-839. doi: 10.1093/humrep/des457

Treloar, S. A., O'Connor, D. T., O'Connor, V. M., & Martin, N. G. (1999). Genetic influences on endometriosis in an Australian twin sample. sueT@qimr.edu.au. *Fertil Steril, 71*(4), 701-710.

Uno, S., Zembutsu, H., Hirasawa, A., Takahashi, A., Kubo, M., Akahane, T., . . . Nakamura, Y. (2010). A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat Genet, 42*(8), 707-710. doi: 10.1038/ng.612