RBMO

ELSEVIER

ARTICLE

Candidate genes for age at menarche are associated with endometriosis

BIOGRAPHY

Professor Mikhail Churnosov is Head of the Department of Biomedical Disciplines of the Belgorod State University. His areas of research activity include the study of the role of polymorphisms of candidate genes, gene–gene and gene–environmental interactions in the formation of frequently occurring multifactorial diseases among the Russian population.

Irina Ponomarenko[1], Evgeny Reshetnikov[1,*], Alexey Polonikov[2], Irina Verzilina[1], Inna Sorokina[1], Elizaveta E Elgaeva[3,4], Yakov A Tsepilov[3,4], Anna Yermachenko[5,6], Volodymyr Dvornyk[7], Mikhail Churnosov[1]

KEY MESSAGE

Sixteen candidate genes for age at menarche were reported to be associated with endometriosis. The observed pleiotropy of these loci suggests a shared genetic architecture for the two phenotypes.

ABSTRACT

Research question: Are the candidate genes for age at menarche associated with a risk of endometriosis?

Design: Fifty-two candidate single nucleotide polymorphisms (SNP) for age at menarche, their gene–gene and gene–environment interactions were analysed for possible association with endometriosis in a sample of 395 patients and 981 controls. Association of the polymorphisms was analysed using logistic regression according to three main genetic models (additive, recessive and dominant). The gene–gene and gene–environment interactions were analysed for the second-, third- and fourth-order models with adjustment for covariates and multiple comparisons with subsequent cross-validation.

Results: Sixteen SNP for age at menarche out of the 52 studied were associated with endometriosis. Polymorphism rs6589964 *BSX* was associated with endometriosis according to the additive and recessive models (OR 1.27–1.47, $P_{perm} \leq 0.006$). Fourteen SNP were associated with the disease within 12 most significant models of gene–gene interactions ($P_{perm} \leq 0.008$). Twelve SNP involved in 10 most significant models of SNP–induced abortion interactions are associated with endometriosis. Fourteen of the 16 polymorphisms associated with endometriosis demonstrated pleiotropic effects: they were also associated with either age at menarche (7 SNP) or height and/or body mass index (10 SNP) in the studied sample. The 16 SNP associated with endometriosis and 316 SNP linked to them have regulatory and expression quantitative trait locus significance for 28 genes contributing to the G alpha signal pathway (fold enrichment 31.09, $P_{FDR} = 0.001$) and responses to endogenous stimuli (fold enrichment 16.01, $P_{FDR} = 0.027$).

Conclusions: Sixteen SNP for age at menarche out of the 52 studied were associated with endometriosis.

[1] Department of Medical Biological Disciplines, Belgorod State University, Belgorod 308015, Russia
[2] Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk 305041, Russia
[3] Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University Novosibirsk, Russia
[4] Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics Novosibirsk, Russia
[5] Department of Social Epidemiology, Pierre Louis Institute of Epidemiology and Public Health, Paris 75571, France
[6] Sorbonne Université, Paris 75320, France
[7] Department of Life Sciences, College of Science and General Studies, Alfaisal University, Riyadh 11533, Saudi Arabia

KEYWORDS
Age at menarche
Association study
Endometriosis
Gene–environment interactions
Gene–gene interactions

*Corresponding author. E-mail address: reshetnikov@bsu.edu.ru (E Reshetnikov). https://doi.org/10.1016/j.rbmo.2020.04.016 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

INTRODUCTION

Endometriosis is classically defined as the presence of endometrial glands and stroma in ectopic locations, primarily the pelvic peritoneum, ovaries and rectovaginal septum (*Burney and Giudice, 2012*). Endometriosis occurs in 6–10% of women of reproductive age (*Treloar, 1999*). It is diagnosed in 20–50% of infertile women (*Gao, 2006*). Definitive diagnosis can only be established through surgical (laparoscopic) visualization of the lesions, ideally with histological verification, resulting in reported average diagnostic delays worldwide of 7–10 years from onset of symptoms (*Nnoaham et al., 2011*). The financial burden of endometriosis on the healthcare system is substantial, with the direct and indirect annual costs estimated at US$12,419 (approx. €9579) per affected woman (*Simoens et al., 2012*).

Hereditary factors are important in the development of endometriosis (*Stefansson et al., 2002*). Based on twin studies, their contribution was estimated at around 51% (*Treloar, 1999*). Recent estimates come from genome-wide association studies (GWAS). To date, the National Genome Research Institute's catalogue of GWAS (https://www.ebi.ac.uk/gwas) includes 10 papers on the study of endometriosis (*Adachi et al., 2010*; *Albertsen et al., 2013*; *Borghese et al., 2015*a; *Nyholt et al., 2012*; *Painter et al., 2011*; *Sapkota et al., 2017a*; *Sobalska-Kwapis et al., 2017*; *Uimari et al., 2017*; *Uno et al., 2010*; *Wang et al., 2017*). More than 20 significant loci associated with the development of endometriosis were identified in these studies. According to *Sapkota et al. (2017a)*, 19 GWAS-identified significant single nucleotide polymorphisms (SNP) determine 5.19% of the variability of endometriosis. However, only a few of the GWAS-identified significant SNP were replicated between studies (*Pagliardini et al., 2013*, *2015*; *Rahmioglu et al., 2014*; *Sapkota et al., 2015*, *2017b*; *Sundqvist et al., 2013*).

One of the risk factors for the development of endometriosis is age at menarche (*Nnoaham et al., 2012*; *Zondervan et al., 2016*). Early menarche increases 'exposure to menstruation' in the course of a woman's life which, in turn, may increase the risk of endometriosis (*Nnoaham et al., 2012*; *Zondervan et al., 2016*). Research data about the association of age at menarche and endometriosis are controversial. In some studies, early menarche was associated with an increased risk of endometriosis (*Candiani et al., 1991*; *Treloar et al., 2010*), whereas in others, either age of menarche was not associated with the disease (*Ashrafi et al., 2016*; *Peterson et al., 2013*) or early menarche was a protective factor (*Buck Louis et al., 2007*). A meta-analysis of 18 studies of the associations between age at menarche and endometriosis (3805 patients and 9526 controls) showed a slight increase in the risk of endometriosis with early menarche (*Nnoaham et al., 2012*).

The significantly different heritability estimates for endometriosis [51% according to twin studies (*Treloar, 1999*), 26% according to common SNP-based heritability (*Lee et al., 2012*), and only slightly above 5% according to GWAS data for significant SNP (*Sapkota et al., 2017a*)] raise a problem of so-called 'missing heritability'. This problem may be addressed by studies of candidate genes associated with particular risk factors for endometriosis (e.g. age at menarche, some anthropometric characteristics, etc.). For example, several loci located within or near genes *KIFAP3*, *CAB39L*, *WNT4* and *GRB14* were associated with both endometriosis and the ratio of waist and hip circumference (*Rahmioglu et al., 2015*). More than 350 GWAS-identified polymorphic loci have been reported as associated with menarcheal age (*Day et al., 2017*). However, no studies concerning the possible association of the menarche candidate genes with the risk of endometriosis have so far been conducted. The work presented here is dedicated to filling this gap.

The purpose of this study was to analyse the association of 52 candidate loci for age at menarche with endometriosis. A previous study examined the association of these loci with age at menarche and some anthropometric characteristics (height, body mass index [BMI]) in the same sample (*Ponomarenko et al., 2019*). This makes comparison and interpretation of the results more meaningful and less biased thanks to the elimination of possible between-sample heterogeneity.

MATERIALS AND METHODS

Study subjects

The study sample included 1435 women: 415 patients with endometriosis and 1020 control subjects. Recruitment of the participants was carried out through the Perinatal Centre of the Belgorod Regional Clinical Hospital of St. Joasaph during 2008–2013. Endometriosis was diagnosed by certified doctors from the Department of Gynecology at the above centre using laparotomy/laparoscopy and subsequent histological confirmation (n = 415, 100%). Hysterectomy followed by morphological study of the surgical material was performed in 228 patients (54.94%). All patients underwent hysteroscopy with targeted therapeutic and diagnostic curettage of the uterus and subsequent histological examination of the obtained material. Among 415 patients with endometriosis, stage I disease was detected in 35.90% (n = 149), stage II in 53.98% (n = 224) and stage III and IV in 10.12% (n = 42). Staging was determined according to the revised American Society for Reproductive Medicine (rASRM) classification (*American Society for Reproductive Medicine 1997*). The control group included women without clinical (an asymptomatic women without chronic pelvic pain, etc.) or ultrasound signs of benign proliferative diseases of the reproductive organs. Participants in the control sample were recruited during regular medical examinations at the above perinatal centre. The ages of endometriosis patients (39.75 ± 9.01 years) and the control group (40.73 ± 8.60 years) were similar (P = 0.10).

The following exclusion criteria were adopted: non-Russian descent (self-declared), a birthplace outside of Central Russia (*Sorokina et al., 2018*), malignant tumours of the small pelvis and breast, chronic severe diseases of the vital organs (heart, respiratory or renal failure) and severe autoimmune diseases. The study was approved by the Regional Ethics Committee of Belgorod State University on 10 April 2008. All participants signed an informed consent form prior to enrolment in this study.

For each participant, the following data were collected: anthropometric characteristics (height, weight, BMI) and age at menarche (as described elsewhere, *Ponomarenko et al., 2019*), details of the menstrual cycle (menstrual cycle length, duration of menses), reproductive characteristics (age at first birth, number of pregnancies and childbirths, spontaneous and induced abortions, infertility), family history

of endometriosis, marital status, use of oral contraceptives, smoking and alcohol use and history of surgery to the pelvic organs. All participants were also comprehensively examined for the presence of gynaecological diseases (uterine leiomyoma, endometrial hyperplasia, adenomyosis).

Blood sample collection and DNA handling

Blood (5 ml) was drawn by a certified nurse from the ulnar vein of each participant into a plastic vial (Vacutainer®) containing 0.5 mol/l EDTA solution (pH = 8.0). Total genomic DNA was isolated from the buffy coat using the standard phenol–chloroform method and then assessed for quality using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Inc., USA). Only samples with A260/A280 = 1.7–2.0 were included in the analysis. The isolated DNA was stored at –80°C.

SNP selection

In total, 52 candidate loci for age at menarche were included in the study. The following selection criteria were applied (*Ponomarenko, 2018*, *2019*): (i) previously reported associations with age at menarche or traits sharing common biological pathways with menarche (anthropometric characteristics, obesity, vitamin D metabolism, etc.), (ii) regulatory potential (regSNP), (iii) effect on gene expression (eSNP), (iv) tag value (tagSNP) and (v) minor allele frequency (MAF) >5%.

The regulatory potential of the SNP and effect on gene expression were estimated using the following online tools: HaploReg (v4.1) (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php), RegulomeDB (https://regulomedb.org/regulome-search/), rSNPBase (http://rsnp.psych.ac.cn/index.do), SNPinfo Web Server – SNP Function Prediction (FuncPred) (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html), Blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser) and GTExPortal (http://www.gtexportal.org).

Information about the selected SNP is given in Supplementary Table 1. All SNP appear to have a significant regulatory potential (Supplementary Table 2), 43 of them (82.69%) are eSNP (Supplementary Table 3), 29 are tagSNP, and 17 are associated with various anthropometric characteristics (Supplementary Table 4).

Out of the 52 selected SNP, 14 were associated with age at menarche according to the GWAS results and 28 according to the candidate gene association studies (Supplementary Table 4). In addition, 10 SNP, which did not demonstrate significant association with age at menarche, were either associated or tag with the traits related to menarche (e.g. vitamin D metabolism, polycystic ovary syndrome development, anthropometric characters, etc., Supplementary Table 4). These SNP included rs1884051 *ESR1*, rs3020394 *ESR1*, rs12324955 *FTO*, rs4633 *COMT*, rs222020 *GC*, rs222003 *GC*, rs1544410 *VDR*, rs3756261 *EGF*, rs7766109 *F13A1* and rs2252673 *INSR*. All these SNP have a significant regulatory potential; nine of them are eSNP and eight are tagSNP.

Several of these 52 loci were previously reported to have an association with age at menarche (14 SNP), height (16 SNP) and BMI (15 SNP) in the same sample of Russian women (*Ponomarenko et al., 2019*).

SNP genotyping

DNA samples were genotyped using the Sequenom MassARRAY® iPLEX platform, which is based on MALDI-TOF (matrix-assisted laser desorption/ionization time-of-flight) mass spectrometry at the Centre of Genomic Sciences at the University of Hong Kong. The analysed DNA samples had concentrations of 10–15 ng/ml. Assay Design Suite 1.0 (http://agenabio.com/assay-design-suite-10-software) was used to create a single-well iPLEX SNP genotyping assay. For this purpose, the 52 SNP of interest were retrieved from dbSNP of NCBI and imported according to their ID to Assay Design Suite 1.0. After completing the consecutive automatic steps, the genotyping assay was successfully generated and tested for cross-amplification.

Data quality control

The quality of the genotypic data was assessed according to the missing call rate, defined as the fraction of missing calls per SNP for all samples. MassARRAY® Typer 4.0 software was used for cluster analysis of the genotype calls. All samples successfully passed the quality control with the following parameters: SNP with call rate >95%, success rate of duplicate check >99.5%, and success rate of the blank check >90%. Individuals with a proportion of determined genotypes <95% out of the maximum possible number were excluded from the analysis (n = 59).

The final sample consisted of 1376 women: 395 patients with endometriosis and 981 individuals in the control group. The proportion of the determined genotypes for the 52 SNP was 98.82%. The general characteristics of the study participants are given in TABLE 1.

Statistical analysis

Logistic regression was used to assess association of clinical and clinical–anamnestic risk factors with the development of the disease. The calculations were performed using the *epicalc* package in the R software environment [version 3.4.0 (2017-04-21)].

All polymorphisms were checked for their correspondence to the Hardy–Weinberg equilibrium (HWE) using the chi-squared test. The associations of the 52 loci with endometriosis were analysed using logistic regression according to three basic genetic models: additive (i.e. comparison of all genotypes, e.g. TT versus TC versus CC), dominant (CC/TC versus TT, where C is a minor allele) and recessive (CC versus TC/TT, where C is a minor allele) with adjustment for covariates. The following covariates were applied as qualitative variables (yes/no): the history of pelvic surgery, history of infertility, the presence of induced abortions in the anamnesis; and as quantitative variables (value of the trait): menstrual cycle length, parity, the number of induced abortions in the anamnesis (TABLE 1). The adaptive permutation test was applied to adjust for multiple comparisons (*Che et al., 2014*). The significance level was set at $P_{perm} < 0.01$ (after the Bonferroni correction based on the numbers of genetic models studied).

The size of the studied sample (395 patients with endometriosis and 981 individuals of the control group) allowed determination of the differences in the distribution of genetic polymorphisms between patients and controls at the level of OR = 1.26–1.64 for the additive model, OR = 1.40–1.68 for the dominant model and OR = 1.46–5.82 for the recessive model (at 80% power, alpha = 0.05 for a two-sided test).

The haplotype blocks were determined using the 'confidence intervals' algorithm

TABLE 1 CHARACTERISTICS OF PARTICIPANTS FROM THE CASE AND CONTROL GROUPS

Parameters	Cases (*n* = 395)	Controls (*n* = 981)	*P*-value
Age (years)	39.75 ± 9.01	40.73 ± 8.60	0.10
Height (m)	1.65 ± 0.06	1.65±0.06	0.39
Weight (kg)	72.65 ± 14.38	72.49 ± 13.37	0.82
BMI (kg/m^2)	26.63 ± 5.31	26.66 ± 4.61	0.84
Proportion of participants by relative BMI:			
Underweight (<18.50)	4.30 (17)	1.12 (11)	
Normal weight (18.50–24.99)	37.72 (149)	42.41 (416)	0.97
Overweight (25.00–29.99)	31.65 (125)	30.48 (299)	
Obese (>30.00)	26.33 (104)	25.99 (255)	
Family history of endometriosis (yes)	6.08 (24)	1.94 (19)	0.0007
Married	82.53 (326)	85.93 (843)	0.13
Smoking (yes)	18.23 (72)	17.33 (170)	0.75
Drinking alcohol (≥7 drinks per week)	4.05 (16)	3.06 (30)	0.45
History of pelvic surgery (laparoscopy and/or laparotomy)	15.19 (60)	9.99 (98)	0.009
Oral contraceptive use	8.10 (32)	10.09 (99)	0.30
Age at menarche and menstrual cycle			
Age at menarche, years	13.29 ± 1.27	13.27 ± 1.25	0.97
Proportion of participants by relative age at menarche			
Early (<12 years)	6.33 (25)	6.42 (63)	0.92
Average (12–14 years)	80.76 (319)	79.51 (780)	
Late (>14 years)	12.91 (51)	14.07 (138)	
Duration of menstrual bleeding (days)	5.13 ± 1.56	4.94 ± 0.94	0.17
Menstrual cycle length (days)	27.66 ± 2.28	28.15 ± 2.24	<0.0001
Reproductive characteristics			
Age at first birth (years)	21.25 ± 3.04	21.71 ± 3.49	0.07
Gravidity	2.60 ± 2.31	2.45 ± 1.55	0.58
No. of births	1.07 ± 0.97	1.51 ± 0.67	<0.0001
No. of spontaneous abortions	0.21 ± 0.61	0.24 ± 0.51	0.08
No. of induced abortions	1.25 ± 1.61	0.67 ± 0.99	0.009
No. of induced abortions:			
0	46.58 (184)	58.92 (578)	
1	17.22 (68)	23.75 (233)	
2	19.24 (76)	10.40 (102)	<0.0001
3	8.61 (34)	5.40 (53)	
≥4	8.35 (33)	1.53 (15)	
History of infertility	33.42 (132)	5.20 (51)	<0.0001
Gynaecological pathologies			
Uterine leiomyoma	52.41 (207)	–	–
Endometrial hyperplasia	46.33 (183)	–	–
Adenomyosis	43.04 (170)	–	–

Data are presented as mean ± SD or % (*n*).

at D′ > 0.8 as implemented in HaploView v.4.2 (https://www.broadinstitute.org/haploview/haploview). Statistical power for each SNP was computed using Quanto 1.2.4. The association analyses were conducted using the PLINK v.2.050 software (available at http://zzz.bwh.harvard.edu/plink).

The gene–gene interactions were analysed for the two-, three- and four-locus models using MB-MDR (model-based multifactor dimensionality reduction) (*Calle et al., 2008*, *2010*) as implemented in the namesake software (v.2.6) for the R programming environment. MB-MDR is a modification of MDR and makes it possible to analyse

gene–gene interactions with adjustment for covariates and validation by the permutation test (*Mahachie et al., 2012*). The permutation test was demonstrated to be efficient for analysis of large massifs of GWAS data without a reduction of the power (Che R. *et al.*, 2014). For the permutation test, the following threshold *P*-values (after the Bonferroni correction based on the numbers of combinations studied for 52 SNP) were adopted for models of gene–gene interactions: $P < 3.8 \times 10^{-5}$ (<0.05/1326) for two-locus models, $P < 2.3 \times 10^{-6}$ (<0.05/22,100) for three-locus models, and $P < 1.8 \times 10^{-7}$ (<0.05/270,725) for four-locus models. The significance level was set at $P_{perm} < 0.01$.

The gene–environment interactions of the candidate genes for age at menarche with induced abortions in terms of their possible contribution to endometriosis were analysed. Induced abortions were included in the analysis of gene–environment interactions because (i) they were a risk factor for endometriosis in the study sample (TABLE 1) and (ii) their high prevalence as a birth control method is a specific feature of the reproductive behaviour of Russian women in comparison with other countries of the world (*David et al., 2007*; *Douglas et al., 2014*; *Sedgh et al., 2007*). The study was conducted using MB-MDR with adjustment for covariates (family history of endometriosis, history of pelvic surgery, history of infertility, menstrual cycle length, parity) and multiple comparisons (1000 permutations) as indicated above. The permutation test was applied to the selected best models of gene–environment interactions with the significance level of $P < 1 \times 10^{-22}$. The significance level was set at $P_{perm} < 0.01$.

The cross-validation of the most significant models of gene–gene and gene–environmental interactions associated with endometriosis was performed by GMDR (generalized multifactor dimensionality reduction) (*Chen et al., 2011a*; *Lou et al., 2007*) (http://www.ssg.uab.edu/gmdr) as implemented in the GMDR software (Beta 0.9) (http://sourceforge.net/projects/gmdr). Cross-validation consistency (CVC), testing balanced accuracy, sensitivity and specificity models were computed with adjustment for covariates. Correction for multiple comparisons was performed using the permutation test. One thousand permutations were performed with 10-fold cross-validation, which provides a level of statistical significance of at least $P_{perm} < 0.001$ for a validated model.

The gene–gene and gene–environment interactions and their relative contribution to the total variance of the trait within the second-, third- and fourth-order models were visualized using the MDR method (http://www.multifactordimensionalityreduction.org), as implemented in MDR v.3.0.2 (http://sourceforge.net/projects/mdr).

Evaluation of casual relationships between age at menarche and endometriosis

Casual relationships between age at menarche and endometriosis were analysed by two approaches: two-sample fixed-effect inverse variance weighted Mendelian randomization (IVW MR) and genetic causality proportion (GCP) analysis.

Two-sample fixed-effect inverse variance weighted Mendelian randomization

To achieve the biggest power in IVW MR, this study used SNP associated with age at menarche from a recent meta-analysis (*Day et al., 2017*) as instrumental variable and the biggest publicly available full GWAS summary statistics on endometriosis as the outcome. The GWAS data for endometriosis (245,494 females of European ancestry, 4252 cases, 241,242 controls) were obtained from GeneATLAS (http://geneatlas.roslin.ed.ac.uk/trait/?trait = 503). The list of instrumental variables was compiled according to the following procedure. First, SNP associated with menarcheal age were obtained from the recent GWAS (*Day et al., 2017*). From the 389 SNP reported in this study, 275 replicated SNP with *P*-values of the meta-analysis for both discovery and replication cohorts lower than those reported for the discovery only were selected. Then those 275 SNP were clumped and pruned for independence in PLINK v.1.90b6.15 (https://www.cog-genomics.org/plink) by retaining only one SNP within a 10,000 kb window among SNP correlated at $r^2 > 0.001$. Summary statistics of the resulting 176 SNP were used as instrumental variables in IVW MR. After subsequent harmonization with outcome data and filtering by MAF (<0.05), 139 SNP were retained for analysis. Seven SNP (rs1040070, rs10897450, rs2542420, rs4780885, rs8087304, rs9422857 and rs9522262) were excluded, because they were palindromic and had intermediate allele frequencies. Thus, 132 instrumental variables were finally selected for the analysis. Harmonization and IVW MR analysis were performed using the R package TwoSampleMR v.0.5.0 (*Hemani et al., 2018*). The other methods of two-sample MR (MR-Egger, weighted median, simple mode, weighted mode) were applied as sensitivity analyses.

Horizontal pleiotropy was controlled using MR-PRESSO method v1.0 (*Verbanck et al., 2018*), which identified rs11031040, rs6185 and rs6933660 as outliers. These three SNP were excluded from the list of instrumental variables to rerun the Mendelian randomization analysis. The full list of the instrumental variables and its association statistics with age at menarche and endometriosis are given in Supplementary Table 5.

Genetic causality proportion

To measure the GCP, GWAS summary statistics for endometriosis as described above were used, and the largest available full GWAS summary statistics on age of menarche (176,008 females of European ancestry) from the Neale lab database (http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank). The latent causal variable (LCV) model implemented in R (*O'Connor et al., 2018*) was used. Genetic correlations between these two traits were calculated using pre-calculated linkage disequilibrium scores estimated using linkage disequilibrium score regression and European ancestry samples from 1000 Genomes Project data (*Bulik-Sullivan et al., 2015*). Data reformatting was performed in the R package GenomicSEM v.0.0.2 (https://github.com/MichelNivard/GenomicSEM).

IVW Mendelian randomization of age at menarche and endometriosis using SNP from this study

In the very last step, IVW Mendelian randomization with available summary statistics for association with endometriosis was performed. A search was carried out for an overlap of 52 SNP (TABLE 2) with the data of the largest available GWAS for age at menarche (described above), performing clumping with above-described parameters, and filtered by the $P < 5 \times 10^{-8}$ for

association with age at menarche. This analysis yielded 14 SNP that were used as instrumental variables for IVW Mendelian randomization (see Supplementary Table 6 for the full list of instrumental variables).

Functional SNP

The SNP that were endometriosis-associated or strongly linked were analysed for their functional significance (non-synonymous SNP, regulatory potential and eQTL). The SNP in strong linkage disequilibrium ($r^2 \geq 0.8$) with those associated with endometriosis were determined using the online version of HaploReg (v4.1) (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php). The linkage disequilibrium was estimated using the data of the European population from the 1000 Genomes Project.

Non-synonymous SNP and their predictive potential were analysed using SIFT (https://sift.bii.a-star.edu.sg/www/SIFT_dbSNP.html).

Regulatory effects

The *in silico* analysis of the regulatory potential of the candidate SNP for endometriosis was conducted using HaploReg (v4.1), RegulomeDB (Version 1.1) (https://regulomedb.org/regulome-search/), rSNPBase (http://rsnp.psych.ac.cn/index.do) and SNP Function Prediction (FuncPred) (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html).

The relationship of a polymorphic locus (reference and alternative alleles) with the affinity of the regulatory DNA motif to transcription factors was determined by the difference between the LOD scores of the alternative (alt) and reference (ref) alleles (*Ward et al., 2012*): ΔLOD = LOD (alt) – LOD (ref). A negative value of this indicator suggests an increase in the affinity of this motif by the reference allele; in contrast, a positive value demonstrates a connection between the alternative allele and an increase in the affinity of the analysed DNA motif.

The possible regulatory effects of polymorphisms in strong linkage disequilibrium ($r^2 \geq 0.8$) with the endometriosis-associated SNP were analysed using HaploReg (v4.1).

Expression QTL

The effect of the candidate SNP for endometriosis on gene expression level (*cis*- and *trans*-eQTL) was estimated in peripheral blood using the data from the Blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser), and in other organs and tissues using the GTExPortal data (http://www.gtexportal.org) as of 10.12.2017 (Release V7 updated on 09/05/2017) (dbGaP Accession phs000424.v7.p2). To determine significant eQTL, a false discovery rate (FDR) ≤0.05 was applied. Likewise, eQTL values of SNP in strong linkage disequilibrium ($r^2 \geq 0.8$) with the endometriosis-associated polymorphisms were estimated.

Pathway analyses

The functional significance of the genes associated with endometriosis in the various biological pathways was studied using the Gene Ontology portal tools (PANTHER Overrepresentation Test from 13.04.2017; PANTHER version 12.0 from 10.07.2017, available at http://geneontology.org). The following databases were utilized: Gene Ontology molecular function, Gene Ontology biological process, PANTHER protein class, PANTHER pathway, PANTHER molecular function, PANTHER biological process and Reactome pathway. The results of multiple comparisons were adjusted with the FDR test. The gene interaction networks were constructed using GeneMANIA (version 3.5.0, accessed on 13 March 2017, available at http://genemania.org) and the automatic weighting for the network.

RESULTS

Study participant characteristics

The groups of patients with endometriosis (n = 395) and controls (n = 981) did not differ by age or BMI ($P > 0.05$) (TABLE 1). Patients with endometriosis had higher rates of family history of the disease (OR 3.27; 95% CI 1.70–6.31; $P < 0.001$), history of pelvic surgery (laparoscopy and/or laparotomy) (OR 1.61; 95% CI 1.12–2.31; $P < 0.01$), history of infertility (OR 9.15; 95% CI 6.35–13.20; $P < 0.001$), presence (OR 1.64; 95% CI 1.29–2.09; $P < 0.001$) and number of induced abortions in the anamnesis (OR 1.32; 95% CI 1.17–1.48; $P < 0.001$), shorter menstrual cycle length (OR 0.91; 95% CI 0.86–0.96; $P < 0.001$) and lower parity (OR 0.78; 95% CI 0.71–0.87; $P < 0.001$) (TABLE 1). These risk factors for endometriosis were used as covariates in the association analyses.

SNP and haplotype association analysis

Data about the SNP studied are provided in Supplementary Tables 7 and 8. All polymorphisms had MAF >5% and were in the HWE ($P_{Bonf} < 0.001$).

The rs6589964 *BSX* polymorphism was associated with endometriosis according to the additive (OR 1.27; 95% CI 1.07–1.51; $P = 0.006$, $P_{perm} = 0.006$, power 81.28%) and recessive (OR 1.47; 95% CI 1.11–1.93; $P = 0.006$; $P_{perm} = 0.005$, power 80.73%) models (TABLE 2). None of the haplotypes showed significant association with endometriosis (Supplementary Table 9).

SNP × SNP interactions

The 12 most significant 2, 3 and 4-locus models of SNP × SNP interactions associated with endometriosis were determined ($P_{perm} \leq 0.008$, CVC = 10/10, testing balanced accuracy 49.61–56.29%, sensitivity and specificity of the best models were 89.37% and 79.10%, respectively) (TABLE 3 and Supplementary Table 10). Out of the 14 SNP involved in the models, loci rs6589964 *BSX* and rs7579411 *LHCGR* contributed to the largest number of models (seven and five, respectively). Locus rs6589964 *BSX* was involved in the most significant models of intergenic interactions at all the levels considered. The most significant associations with endometriosis were determined for the combinations of genotypes rs6589964 AA *BSX* × rs10441737 TT *ZNF483* (beta = 0.64, $P = 0.0004$) and rs6589964 AA *BSX* × rs10441737 TT *ZNF483* × rs7138686 *RBJ* (beta = 0.79, $P = 0.0007$) (Supplementary Table 11).

The graph of gene–gene interactions of 14 SNP that make up the 12 best models associated with endometriosis (Supplementary Figure 1) suggests the synergistic nature of these interactions; the highest percentage of entropy is determined by interactions rs10441737 *ZNF483* × rs11031010 *FSHB* (0.58%), rs7759938 *LIN28B* × rs1782507 *FSHB* (0.53%) and polymorphism rs6589964 *BSX* (0.42%).

Gene–environment interactions

Second-, third- and fourth-order models of interactions of 12 loci with induced abortions were analysed. The 10 best models associated with the development of endometriosis were determined ($P < 0.001$, CVC = 10/10,

TABLE 2 ASSOCIATIONS OF THE 52 SNP WITH ENDOMETRIOSIS

chr	SNP	*n*	Additive model				Dominant model				Recessive model			
			OR	95% CI		*P*-value	OR	95% CI		*P*-value	OR	95% CI		*P*-value
				L95	U95			L95	U95			L95	U95	
1	rs1514175	1373	0.92	0.77	1.09	0.343	0.78	0.61	1.00	0.051	1.15	0.82	1.60	0.417
1	rs466639	1372	0.93	0.71	1.22	0.604	0.98	0.73	1.31	0.885	0.33	0.07	1.44	0.139
1	rs7538038	1373	1.17	0.95	1.44	0.131	1.21	0.95	1.55	0.130	1.21	0.69	2.13	0.512
2	rs713586	1372	0.94	0.79	1.12	0.504	1.02	0.78	1.32	0.892	0.80	0.58	1.10	0.174
2	rs2164808	1373	0.96	0.80	1.14	0.630	0.96	0.73	1.26	0.783	0.92	0.68	1.25	0.606
2	rs7589318	1370	0.94	0.78	1.14	0.532	0.92	0.73	1.18	0.528	0.93	0.61	1.44	0.759
2	rs4374421	1321	1.01	0.84	1.22	0.909	0.89	0.69	1.13	0.333	1.46	0.98	2.17	0.060
2	rs7579411	1362	0.96	0.81	1.15	0.680	0.83	0.64	1.08	0.158	1.15	0.85	1.56	0.362
2	rs6729809	1327	1.02	0.85	1.24	0.805	1.00	0.78	1.28	0.986	1.12	0.75	1.68	0.570
2	rs4953616	1368	1.10	0.90	1.33	0.357	1.05	0.82	1.33	0.724	1.42	0.91	2.23	0.124
2	rs6732220	1372	1.03	0.84	1.26	0.768	1.06	0.83	1.35	0.652	0.95	0.57	1.59	0.851
2	rs4953655	1374	0.98	0.80	1.19	0.829	1.02	0.80	1.30	0.878	0.79	0.46	1.35	0.382
2	rs887912	1315	0.79	0.64	0.97	0.023	0.75	0.58	0.97	0.026	0.72	0.41	1.26	0.247
2	rs12617311	1369	0.95	0.79	1.14	0.595	0.92	0.72	1.17	0.505	0.99	0.66	1.46	0.946
3	rs6438424	1363	1.13	0.95	1.34	0.160	1.21	0.91	1.59	0.187	1.15	0.87	1.52	0.326
4	rs2013573	1372	0.77	0.60	0.97	0.030	0.74	0.57	0.96	0.024	0.82	0.35	1.95	0.654
4	rs13111134	1371	0.84	0.68	1.04	0.111	0.78	0.61	1.01	0.055	1.04	0.57	1.90	0.889
4	rs222003	1373	0.97	0.68	1.37	0.856	0.99	0.69	1.41	0.952	0.00	0.00	inf	0.999
4	rs222020	1373	1.18	0.91	1.53	0.214	1.17	0.88	1.56	0.273	1.63	0.59	4.51	0.343
4	rs3756261	1359	0.93	0.66	1.31	0.665	0.93	0.65	1.32	0.673	0.81	0.06	11.01	0.874
5	rs757647	1363	0.97	0.79	1.19	0.751	0.98	0.77	1.25	0.861	0.88	0.50	1.53	0.646
6	rs7766109	1372	1.05	0.88	1.25	0.617	1.10	0.84	1.45	0.481	1.01	0.75	1.37	0.943
6	rs4946651	1373	0.91	0.77	1.09	0.315	0.92	0.71	1.18	0.505	0.84	0.61	1.17	0.312
6	rs7759938	1372	0.89	0.73	1.08	0.248	0.87	0.68	1.11	0.265	0.86	0.53	1.38	0.529
6	rs314280	1334	0.96	0.80	1.15	0.652	0.95	0.74	1.23	0.707	0.94	0.66	1.34	0.726
6	rs314276	1342	0.89	0.73	1.08	0.225	0.88	0.69	1.13	0.310	0.81	0.52	1.25	0.338
6	rs3020394	1374	0.95	0.79	1.14	0.560	0.96	0.75	1.22	0.739	0.85	0.56	1.28	0.461
6	rs1884051	1374	0.95	0.79	1.15	0.599	0.97	0.77	1.24	0.834	0.83	0.54	1.28	0.409
6	rs7753051	1374	0.89	0.73	1.08	0.221	0.79	0.62	1.00	0.051	1.80	0.76	1.80	0.472
7	rs1079866	1372	1.08	0.86	1.36	0.491	1.09	0.84	1.41	0.521	1.16	0.57	2.35	0.684
8	rs2288696	1371	0.97	0.78	1.21	0.792	1.00	0.78	1.29	0.981	0.70	0.33	1.51	0.367
9	rs2090409	1306	1.04	0.86	1.24	0.713	1.16	0.90	1.50	0.246	0.84	0.58	1.22	0.356
9	rs10980926	1369	0.86	0.71	1.05	0.137	0.87	0.68	1.11	0.250	0.73	0.45	1.16	0.181
9	rs10441737	1283	0.82	0.67	1.00	0.050	0.82	0.64	1.05	0.119	0.67	0.42	1.07	0.097
11	rs10769908	1356	0.99	0.83	1.18	0.938	0.94	0.72	1.24	0.676	1.05	0.78	1.41	0.749
11	rs555621	1370	0.98	0.82	1.17	0.809	0.98	0.76	1.27	0.897	0.95	0.68	1.33	0.775
11	rs11031010	1359	0.79	0.61	1.04	0.094	0.77	0.57	1.04	0.089	0.76	0.27	2.13	0.607
11	rs1782507	1368	1.09	0.91	1.30	0.375	1.05	0.82	1.35	0.680	1.25	0.87	1.80	0.231
11	rs6589964	1372	**1.27**	**1.07**	**1.51**	**0.006**	1.30	0.98	1.71	0.068	**1.47**	**1.11**	**1.93**	**0.006**
12	rs1544410	1373	1.01	0.84	1.20	0.931	0.96	0.75	1.22	0.724	1.13	0.80	1.59	0.500
14	rs999460	1374	0.83	0.69	1.00	0.050	0.83	0.65	1.05	0.122	0.70	0.46	1.06	0.092
14	rs4986938	1370	0.96	0.80	1.16	0.667	0.86	0.67	1.10	0.223	1.22	0.84	1.78	0.302
15	rs2241423	1365	0.98	0.79	1.23	0.877	1.04	0.80	1.34	0.792	0.66	0.32	1.36	0.258

(continued on next page)

Table 2 – (continued)

chr	SNP	*n*	Additive model				Dominant model				Recessive model			
			OR	95% CI		*P*-value	OR	95% CI		*P*-value	OR	95% CI		*P*-value
				L95	U95			L95	U95			L95	U95	
16	rs12444979	1370	0.91	0.71	1.16	0.443	0.95	0.72	1.25	0.696	0.52	0.21	1.30	0.162
16	rs9939609	1372	1.01	0.85	1.20	0.889	0.93	0.72	1.20	0.572	1.15	0.85	1.56	0.359
16	rs12324955	1372	0.97	0.80	1.17	0.757	1.05	0.82	1.33	0.721	0.73	0.46	1.15	0.177
18	rs1398217	1365	1.15	0.97	1.38	0.115	1.22	0.93	1.59	0.145	1.19	0.87	1.62	0.287
19	rs2252673	1369	0.98	0.79	1.21	0.849	0.92	0.72	1.19	0.541	1.31	0.74	2.33	0.358
20	rs1073768	1370	1.06	0.89	1.26	0.497	1.07	0.81	1.40	0.654	1.10	0.83	1.46	0.508
22	rs4633	1374	0.99	0.83	1.17	0.866	1.00	0.76	1.31	0.989	0.96	0.73	1.27	0.790
23	rs5930973	1355	1.20	0.85	1.70	0.306	NA	NA	NA	NA	NA	NA	NA	NA
23	rs3092921	1371	0.88	0.63	1.22	0.445	NA	NA	NA	NA	NA	NA	NA	NA

All results were obtained after adjustment for covariates.

Statistically significant values after adjustment by the adaptive permutation test (1000 permutations).

CI = confidence interval; inf = infinity; n = number of participants; NA = not applicable (loci are located on X chromosome and therefore parameters of the dominant and recessive models are not applicable); OR = odds ratio.

testing balanced accuracy 54.71–58.97%, sensitivity and specificity of the best models 72.91% and 75.94%, respectively) (TABLE 4 and Supplementary Table 12). The polymorphisms of the *LHCGR* and *FSHB* genes contributed to the largest number of the best models (eight and four, respectively) (TABLE 4). Polymorphisms rs6438424 *3q13.32* and rs12324955 *FTO* were significantly associated with endometriosis only when interacting with induced abortions. The following combinations of genotypes and induced abortions manifested the most significant associations with endometriosis:

TABLE 3 THE MOST SIGNIFICANT MODELS OF SNP × SNP INTERACTIONS ASSOCIATED WITH ENDOMETRIOSIS

n	Models of SNP × SNP interactions	NH	*beta* H	WH	NL	*beta* L	WL	P_{perm}
Two-locus models ($P < 3 \times 10^{-4}$)								
1	rs10441737 *ZNF483* × rs4374421 *LHCGR*	2	0.559	17.40	1	−0.327	3.95	0.001
2	rs6589964 *BSX* × rs1544410 *VDR*	2	0.660	15.88	1	−0.288	3.34	0.003
3	rs11031010 *FSHB* × rs10441737 *ZNF483*	1	0.256	4.11	2	−0.722	15.43	<0.001
4	rs10980926 *ZNF483* × rs4374421 *LHCGR*	2	0.463	12.69	1	−0.426	7.24	0.008
5	rs6589964 *BSX* × rs10441737 *ZNF483*	1	0.645	12.55	1	−0.651	3.06	0.007
Three-locus models ($P < 5 \times 10^{-8}$)								
1	rs6589964 *BSX* × rs10441737 *ZNF483* × rs713586 *RBJ*	4	0.915	31.36	4	−0.817	14.34	<0.001
2	rs6589964 *BSX* × rs1073768 *GHRH* × rs1544410 *VDR*	4	1.140	30.19	1	−0.848	6.50	<0.001
3	rs6589964 *BSX* × rs7759938 *LIN28B* × rs7579411 *LHCGR*	5	0.754	29.61	2	−0.559	8.30	<0.001
Four-locus models ($P < 8 \times 10^{-13}$)								
1	rs6589964 *BSX* × rs1782507 *FSHB* × rs7579411 *LHCGR* × rs1514175 *TNNI3K*	12	1.418	64.89	3	−1.281	18.22	<0.001
2	rs6589964 *BSX* × rs713586 *RBJ* × rs7579411 *LHCGR* × rs1514175 *TNNI3K*	10	1.424	57.54	2	−0.702	7.66	<0.001
3	rs2090409 *TMEM38B* × rs7759938 *LIN28B* × rs rs555621 *FSHB* × rs7579411 *LHCGR*	11	1.103	57.05	1	−1.227	3.99	<0.001
4	rs1782507 *FSHB* × rs7759938 *LIN28B* × rs7579411 *LHCGR* × rs1514175 *TNNI3K*	9	1.291	51.43	1	−0.978	4.86	<0.001

Results were obtained using the MB-MDR method with adjustment for covariates.

NH = number of significant high-risk genotypic combinations associated with endometriosis.

beta H = coefficient of the logistic regression for significant high-risk genotypic combinations associated with endometriosis.

WH = Wald test value for significant high-risk genotypic combinations associated with endometriosis.

NL = number of significant low-risk genotypic combinations associated with endometriosis.

beta L = coefficient of the logistic regression for significant low-risk genotypic combinations associated with endometriosis.

WL = Wald test value for significant low-risk genotypic combinations associated with endometriosis.

P_{perm} = *P*-value for the permutation test (1000 permutations).

TABLE 4 THE MOST SIGNIFICANT MODELS OF GENE–ENVIRONMENT INTERACTIONS ASSOCIATED WITH ENDOMETRIOSIS

n	Models of gene–environment interactions	NH	beta H	WH	NL	beta L	WL	P_{perm}
Two-order interaction models								
1	rs4374421 *LHCGR* × abortions	2	0.662	26.13	2	−0.504	17.41	<0.001
2	rs12324955 *FTO* × abortions	2	0.592	24.07	2	−0.521	18.78	<0.001
3	rs11031010 *FSHB* × abortions	1	0.582	22.30	1	−0.384	9.86	<0.001
Three-order interaction models								
1	rs12324955 *FTO* × rs4374421 *LHCGR* × abortions	3	0.818	34.51	2	−0.478	11.00	<0.001
2	rs7579411 *LHCGR* × rs1514175 *TNNI3K* × abortions	3	0.581	13.07	5	−0.754	33.57	<0.001
Four-order interaction models								
1	rs10980926 *ZNF483* × rs4374421 *LHCGR* × rs555621 *FSHB* × abortions	9	1.028	57.15	2	−0.866	7.37	<0.001
2	rs6438424 *3q13.32* × rs10441737 *ZNF483* × rs4374421 *LHCGR* × abortions	6	1.242	54.19	3	−0.676	9.65	<0.001
3	rs10441737 *ZNF483* × rs4374421 *LHCGR* × rs555621 *FSHB* × abortions	7	1.121	53.67	3	−0.709	11.98	<0.001
4	rs6589964 *BSX* × rs7759938 *LIN28B* × rs4374421 *LHCGR* × abortions	6	1.201	51.50	3	−0.652	10.07	<0.001
5	rs1782507 *FSHB* × rs7579411 *LHCGR* × rs1514175 *TNNI3K* × abortions	10	1.311	51.39	4	−0.892	17.81	<0.001

The results were obtained using the MB-MDR method with adjustment for covariates.

abortions = induced abortions; beta H, coefficient of the logistic regression for significant high-risk genotypic and abortions combinations associated with endometriosis; beta L, coefficient of the logistic regression for significant low-risk genotypic and abortions combinations associated with endometriosis; NH = number of significant high-risk genotypic and abortions combinations associated with endometriosis; NL = number of significant low-risk genotypic and abortions combinations associated with endometriosis; P_{perm} = *P*-value for the permutation test (1000 permutations); WH = the Wald test value for significant high-risk genotypic and abortions combinations associated with endometriosis; WL = the Wald test value for significant low-risk genotypic and abortions combinations associated with endometriosis.

rs4374421 TT *LHCGR* × abortion (beta = 0.58, $P < 0.0001$), rs11031010 CC *FSHB* × abortion (beta = 0.55, $P < 0.0001$), rs6438424 AC *3q13.32* × rs10441737 TT *ZNF483* × rs4374421 TT *LHCGR* × abortion (beta = 1.20, $P < 0.0001$) (Supplementary Table 13).

The graph of the interactions between the SNP and induced abortions within the best models mentioned above suggests that the abortions made the greatest contribution (0.91%) to the entropy of the trait (endometriosis) (Supplementary Figure 2). Except for rs12324955 *FTO*, no other SNP manifested significant pairwise interactions with induced abortions.

SNP associated with endometriosis are also associated with age at menarche, height and BMI in adults

This study also analysed whether the SNP associated with endometriosis are associated with menarche age, height and BMI of adults in the studied group of women (the respective data were reported elsewhere; *Ponomarenko et al., 2019*). Out of the 16 SNP associated with endometriosis either individually (1 SNP) or through SNP × SNP (14 SNP) and gene–environment (12 SNP) interactions, 14 SNP (87.5%) are also associated with menarcheal age (7 SNP), height and/or BMI (10 SNP) (Supplementary Table 14). Of these, the rs1073768 *GHRH* polymorphism was associated with all the phenotypes under consideration, and four SNP associated with endometriosis were also associated with two of the three phenotypes analysed (i.e. age of menarche and/or height and/or BMI).

Evaluation of causal relationships between age at menarche and endometriosis

Two-sample fixed-effect inverse variance weighted Mendelian randomization (IVW MR)

None of the methods (MR-Egger, weighted median, inverse variance weighted, simple mode, weighted mode) revealed significant results either in primary (full results presented in Supplementary Note 1) or follow-up Mendelian randomization analysis after excluding the outliers (full results presented in Supplementary Note 2). The Cochran's Q test in the primary analysis for both MR-Egger and inverse variance weighted methods showed significant heterogeneity among the instrumental variables (Q = 182.1556, Q *P*-value = 0.0017410 and Q = 182.3320, Q *P*-value = 0.0020406, respectively). In contrast, no heterogeneity was detected after removing SNP with horizontal pleiotropy.

Genetic causality proportion

The LCV approach did not detect significant non-zero genetic correlation (estimated value −0.11, SE 0.07) between age at menarche and endometriosis (GCP −0.21, SE 0.29, $P = 0.50$).

IVW MR of age at menarche and endometriosis using SNP from this study

No significant signal was detected using the 14 independent SNP reported in this study (Supplementary Note 3). However, the results for all used methods except the simple mode suggested a large negative effect of age at menarche on the risk of endometriosis (IVW MR beta = −1.03). This is supportive of possible relationships between these two traits.

Functional SNP

Non-synonymous SNP

None of the 16 SNP associated with endometriosis was missense. However, one of them, rs713586 (2p23.3), was in linkage disequilibrium with rs11676272 ($r^2 = 0.93$), which results in the Ser107Pro replacement in the ADCY3 polypeptide. This amino acid substitution has a predictor value of 'TOLERATED' in the SIFT (Sorting Tolerant from Intolerant) database with a SIFT score = 0.42, which does not exceed the threshold value (≤0.05).

Regulatory effects

The data on regulatory effects of the SNP associated with endometriosis are presented in Supplementary Table 15. The most significant regulatory potential was determined for rs10980926 and rs10441737 *ZNF483*. According to the HaploReg (v4.1), allele A of rs6589964 *BSX* increases affinity to transcription factors CDP_1 (ΔLOD = 3.8), Foxa_known1 (ΔLOD = 4.7) and Foxa_known2 (ΔLOD = 2.8).

The regulatory potential of the 16 SNP associated with endometriosis and of 316 SNP strongly linked ($r^2 \geq 0.8$) to those associated with endometriosis was studied (Supplementary Table 16). Out of these, three SNP (one non-synonymous and two synonymous) were located in exons of the studied genes, four polymorphisms were found in 3′-UTR, 78 were located in introns and 247 in intergenic regions. There were 11 SNP located in the evolutionarily conservative regions (Supplementary Table 16).

Among the above 316 SNP, more than 280 had regulatory significance. The most pronounced regulatory effects were determined for SNP that were in linkage disequilibrium ($r^2 \geq 0.8$) with loci rs713586 *RBJ* (4 SNP), rs555621, rs11031010, rs1782507 of the *FSHB* gene (4 SNP), rs10441737 *ZNF483* (1 SNP), rs7759938 *LIN28B* (2 SNP), rs2090409 *TMEM38B* (1 SNP) and rs1544410 *VDR* (2 SNP) (Supplementary Table 16).

For example, rs6749422, which is strongly linked to rs713586 *RBJ* ($r^2 = 0.92$), is located in an evolutionarily conservative region that has promoter and enhancer histone marks in 17 and 18 tissues, respectively, and is a region of hypersensitivity to DNase-1 in 27 tissues, a binding region of three regulatory proteins, and a region of eight regulatory motifs.

Polymorphism rs1222218 is linked to three loci (rs555621, rs11031010 and rs1782507) of the *FSHB* gene ($r^2 = 0.28, 0.21, 0.82$, respectively) associated with endometriosis. This SNP is located in the region of histones marking promoters in 24 tissues, the DNase hypersensitivity region in 49 tissues, a region binding to regulatory proteins in 19 tissues and a region of 11 regulatory motifs. Similarly, rs7941939 linked to the above three loci ($r^2 = 0.26$, 0.20, 0.80, respectively) is located in the region of 32 regulatory DNA motifs.

Importantly, the loci that are in linkage disequilibrium with the endometriosis-associated SNP manifest their regulatory effects in organs and tissues that are related to the development of endometriosis, i.e. various brain regions (hypothalamus, pituitary, etc.), ovaries, adipose tissue, muscle tissue, liver, etc.

Expression QTL

Five SNP from the 16 loci associated with endometriosis were also associated ($P < 5 \times 10^{-5}$, $P_{FDR} < 0.05$) with the transcription level of six genes (*CENPO*, *ADCY3*, *DNAJC27*, *KIAA0368*, *C11orf46* (*ARL14EP*), *MANBAL*) in peripheral blood (*cis*-eQTL) (Supplementary Table 17). Locus rs713586 *RBJ* is strongly linked ($r^2 \geq 0.8$) to six SNP affecting transcription of genes *CENPO* ($P \leq 1.5 \times 10^{-17}$, $P_{FDR} = 0$), *ADCY3* ($P \leq 1.5 \times 10^{-17}$, $P_{FDR} = 0$) and *DNAJC27* ($P \leq 1.83 \times 10^{-5}$, $P_{FDR} < 0.01$) in peripheral blood (*cis*-eQTL). Among the studied SNP, no statistically significant *trans*-eQTL effects were detected ($P_{FDR} \leq 0.05$).

According to data from the Genotype-Tissue Expression (GTEx) project, 10 loci from the 16 associated with endometriosis had *cis*-eQTL significance in various tissues and organs ($P < 8 \times 10^{-5}$, $P_{FDR} \leq 0.05$) (Supplementary Table 18). The following loci demonstrated possible effects on gene transcription in organs and tissues related to the development of endometriosis: rs1782507 and rs555621 (*ARL14EP*) in the basal ganglia of the brain, rs1782507 and rs555621 (*FSHB*) in the hypothalamus, rs7759938 (*LIN28B*), rs713586 (*ADCY3*), rs1782507 and rs555621 (*ARL14EP*) in the pituitary gland, rs1782507 (*ARL14EP*) in the ovaries, rs713586 (*CENPO*, *ADCY3*), rs11031010 (*ARL14EP*) in subcutaneous adipose tissue, rs713586 (*ADCY3*, *DNAJC27-AS1*), rs6438424 (*RP11-384F7.2*), rs1782507 (*ARL14EP*) in visceral adipose tissue, rs1514175 (*LRRIQ3*), rs7579411 (*STON1-GTF2A1L*), rs1782507, rs11031010 and rs555621 (*ARL14EP*) in the thyroid gland, rs6438424 (*RP11-384F7.2* and *LSAMP*), rs1782507 (*ARL14EP*) in the adrenals, rs713586 (*CENPO*, *ADCY3*, *DNAJC27-AS1*), rs555621 (*ARL14EP*) in peripheral blood. The highest *cis*-eQTL value was determined for the rs713586 *RBJ* locus associated with expression of six genes in more than 10 different organs and tissues, and for polymorphisms rs555621, rs11031010 and rs1782507 of the *FSHB* gene associated with expression of three genes in more than 25 different organs and tissues (Supplementary Table 18).

Twelve endometriosis-associated loci manifested strong linkage disequilibrium ($r^2 \geq 0.8$) with more than 230 SNP affecting transcription of genes ($P < 8.5 \times 10^{-5}$, $P_{FDR} \leq 0.05$) in various organs and tissues (Supplementary Table 19). The most notable effect was suggested for the rs555621, rs11031010 and rs1782507 loci of *FSHB*, which were in strong linkage disequilibrium with more than 120 SNP affecting the *ARL14EP*, *FSHB* and *RP4-710M3.1* genes in more than 25 different organs and tissues. Then, rs7759938 *LIN28B* was strongly linked to 27 SNP associated with the transcription of this gene in the pituitary gland. Polymorphism rs2090409 *TMEM38B* was linked to 25 SNP affecting the expression level of the *SLC44A1* gene in the large intestine; rs1514175 *TNNI3K* was linked to 23 loci associated with the transcription of the *LRRIQ3* gene in the thyroid gland; rs713586 *RBJ* was linked to 20 SNP affecting the expression of the *DNAJC27*, *CENPO*, *ADCY3*, *DNAJC27-AS1*, *EFR3B*, *POMC* and *NCOA1* genes in more than 10 different organs and tissues. Finally, a significant effect on gene expression of *RP11–384F7.2* in the visceral adipose tissue and adrenal glands, *LSAMP* in the thyroid gland, adrenal glands and testicles was suggested for eight SNP that are linked to rs6438424 *3q13.32*.

Overall, 13 of the 16 loci associated with endometriosis have important *cis*-eQTL significance; they affect the expression of 21 different genes. Of these loci, one SNP is independently associated with the mRNA expression level, one SNP is in linkage disequilibrium with other eQTL SNP, and 11 SNP are both individually associated with and strongly linked to other SNP affecting gene transcription levels.

Pathway analyses

The genes associated with endometriosis (in total 11) and those whose expression is affected by the endometriosis-associated SNP according to the eQTL analysis (in total 21, including four from the above list, Supplementary Table 14) were analysed *in silico* for their role in metabolic pathways. Of the 28 genes under consideration, information on *STON1-GTF2A1L*, *RP4-710M3.1*, *DNAJC27-AS1* and *RP11-384F7* was not available from the Gene Ontology

databases. Only two biological pathways that characterize the functional significance of the 24 genes under consideration were identified: G alpha signalling pathway (fold enrichment 31.09, P_{FDR} = 0.001) and responses to endogenous stimuli (fold enrichment 16.01, P_{FDR} = 0.027).

The above 28 genes were analysed for their gene–gene interactions using GeneMANIA (http://genemania.org). The data on three genes, *DNAJC27-AS1*, *RP11-384F7* and *RP4-710M3*, were not available in GeneMANIA. Therefore, the inferred network included 25 endometriosis-associated genes and 20 other genes that interact with them most significantly (FIGURE 1). The gene–gene interactions are realized through co-expression (70.42%), physical (23.96%) and genetic (4.96%) interactions, and common protein domains (0.67%) (Supplementary Table 20).

DISCUSSION

This study reports for the first time associations of 16 candidate loci for age at menarche with endometriosis. Among these, 14 SNP were also associated with either age at menarche (7 SNP) or height and/or BMI (10 SNP) in the studied sample. These results suggest that the proportion of pleiotropic loci contributing to the studied phenotypes (age at menarche, anthropometric characteristics, endometriosis) may be quite large. In turn, this may suggest a shared genetic architecture for these phenotypes. This assumption is supported by the estimates of genetic correlations (r_g) between various phenotypic characteristics and endometriosis obtained by *Sapkota et al. (2017a)* based on GWAS data. They reported significant genetic correlations ($P < 0.05$) between endometriosis and age at menarche, height, extreme height and obesity. *Day et al. (2015a)* determined significant genetic correlations between age at menarche and BMI ($r_g = -0.35$, $P = 1.4 \times 10^{-91}$) and height ($r_g = 0.13$, $P = 2.5 \times 10^{-10}$). Another study by *Day et al. (2015b)* conducted on the data of 250,037 women from the UK Biobank reported a strong association between age at menarche and endometriosis (linear trend models, $P < 7.48 \times 10^{-5}$). These correlations suggest the probable shared genetic architecture of the phenotypes, which may include pleiotropic loci for age at menarche among the others.

However, despite using a large number of instrumental variables, neither significant MR signal of age at menarche on endometriosis nor significant genetic correlation between them was detected. One of the reasons for this is that the available GWAS for endometriosis contained a relatively small number of cases and consequently was not powerful enough. Using only 14 instrumental variables from the present study, it was not possible to detect a negative effect of age at menarche on the risk of endometriosis, although this effect was not statistically significant. More powerful GWAS for endometriosis using a larger number of cases may yield significant results and shed more light on the genetic relationships of this disorder with age at menarche.

Among the analysed loci, rs6589964 of the *BSX* gene manifested the most significant associations: its allele A increased risk for endometriosis according to the additive and recessive models (OR 1.27–1.47). This locus is also associated with endometriosis within 7 of the 12 most significant models of gene–gene interactions. Association of this polymorphism with age at menarche and height has been reported previously (*Elks et al., 2010*; *Ponomarenko et al., 2019*). Importantly, *Elks et al. (2010)* determined allele A of this locus to be associated with early menarche and short stature, whereas the present study identified this allele as a risk factor for endometriosis. Together these results suggest a shared genetic basis for early menarche and the risk of endometriosis.

According to the HaploReg database (v4.1), allele A of rs6589964 *BSX* increases affinity to the transcription factors CDP_1, Foxa_known1 and Foxa_known2. Transcription factor Foxa binds to histone proteins and causes chromatin remodelling, which is one of the important epigenetic mechanisms of gene transcription regulation. According to GeneCards: The Human Gene Database (http://www.genecards.org/), BSX (brain-specific homeobox) is a DNA binding protein that is expressed in some regions of the central nervous system (including the hypothalamus and pituitary) and acts as a transcriptional activator. It affects the phenotypic effects of neuropeptide Y (NPY), agouti-related peptide (AgRP), ghrelin and leptin, which regulate eating behaviour, is important in controlling the body's energy balance and is necessary for normal postnatal growth (*Nogueiras et al., 2008*). One of the possible mechanisms of BSX-mediated regulation of the agouti-linked peptide expression is the interaction of BSX with the Foxa transcription factor (*Lee et al., 2013*).

According to data from this study, two polymorphisms of the *LHCGR* gene (rs7579411 and rs4374421) are involved in the largest number of the most significant models of gene–gene and gene–environment interactions associated with endometriosis. These polymorphisms manifested pleiotropic effects; they were also suggested to contribute to age at menarche, height (rs7579411) and BMI (rs4374421) in the cohort studied here (*Ponomarenko et al., 2019*). The association of these polymorphisms with age at menarche was previously demonstrated by *He et al. (2010)*.

The *LHCGR* gene encodes a receptor for LH and gonadotrophin (http://www.genecards.org/). LH produces multiple effects on female physiology, including the induction of follicle development and androgen synthesis in the ovaries, initiation of ovulation and menarche (*Abreu and Kaiser, 2016*; *Plant, 2015*). Polymorphisms located in the region of the *LHCGR* gene and the genes themselves may be related to reproductive traits. There are data about their association with polycystic ovary syndrome (PCOS) in Chinese patients (*Chen et al., 2011b*; *Shi et al., 2012*) and elevated expression in the subcutaneous fat of PCOS patients (*Jones et al., 2015*). *Davis et al. (2013)* reported different expression levels of *STON1-GTF2A1L* in uterine myoma samples from both black and white women.

Three polymorphisms (rs11031010, rs555621 and rs1782507) of the *FSHB* gene were involved in the largest number of the most significant models of gene–environment interactions associated with endometriosis (4 out of 10 models). They are closely linked and form a single haploblock (D' = 1; LOD > 2). Previously we reported the association of the rs555621 polymorphism with BMI (*Ponomarenko et al., 2019*). Other studies also reported the association of the rs555621, rs11031010 and rs1782507 loci with menarcheal age (*He et al., 2010*). Furthermore, rs11031010 was associated with PCOS and the level of LH in patients (*Tian et al., 2016*).

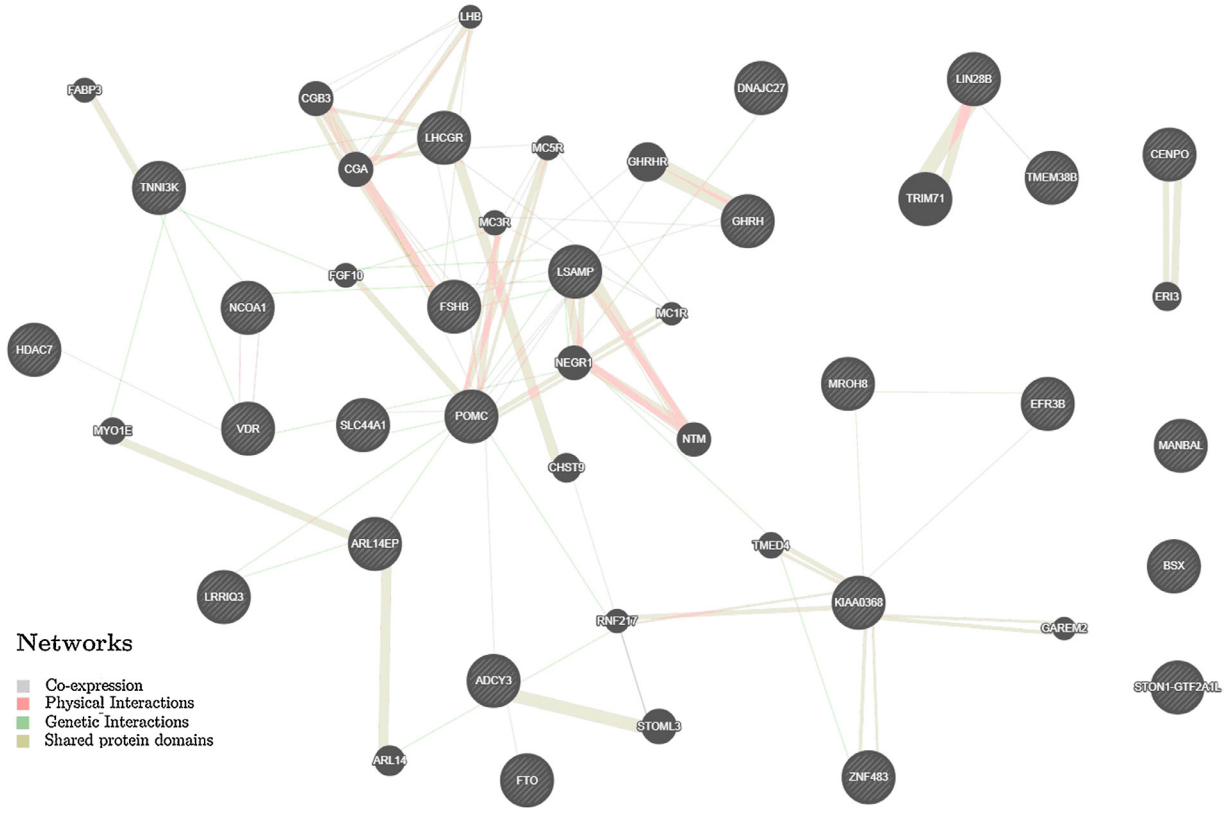

FIGURE 1 The interaction networks of the candidate genes for endometriosis inferred using GeneMANIA (http://genemania.org). The candidate genes determined in the present study are cross-shaded.

The above loci of the *FSHB* gene not only have an independent eQTL value but are also linked to more than 120 SNP, which affect expression of the *ARL14EP*, *FSHB* and *RP4-710M3.1* genes in nearly 25 tissues. According to the Ensembl database (http://www.ensembl.org/), the FSHB gene product is a beta-subunit of FSH. FSH causes proliferation of cells in the granulosa layer in follicles and the growth of follicles in the ovaries, induces the synthesis of aromatases that convert androgens into oestrogens (oestradiol) and stimulates the synthesis of receptors for LH on the granulosa cells of the follicle before ovulation.

The recent meta-analysis of 11 GWAS in a sample of 17,045 patients and 191,596 controls determined 19 SNP associated with endometriosis, including the region of chromosome 11 (11p14.1, rs74485684) in which the *FSHB* gene is located (*Sapkota et al., 2017a*). Importantly, rs11031010 *FSHB* is located just 1.9 kb from the rs74485684 polymorphism, previously determined by GWAS to be associated with endometriosis (*Sapkota et al., 2017a*).

The results of GWAS studies have suggested that several loci linked to rs11031010 might be related to various reproductive parameters, such as the content of FSH (rs11031005, $r^2 = 0.64$) and LH (rs11031002, $r^2 = 0.76$) in blood plasma (*Ruth et al., 2016*) as well as menopausal age (rs12294104, $r^2 = 0.61$) (*Stolk et al., 2012*). The key importance of the SNP located upstream of the *FSHB* transcription start site in the development and function of many reproductive processes has been suggested by *Gajbhiye et al. (2018)*.

Candidate genes for age at menarche are associated with endometriosis and some anthropometric phenotypes. These results suggest a pleiotropic effect of the genes and shared genetic architecture of the studied phenotypes. The determined associations are probably underlined by multiple biological effects of more than 310 SNP linked to the above polymorphisms and implicated in regulatory mechanisms in organs and tissues that are important for both menarche and development of endometriosis.

ACKNOWLEDGEMENTS

This is a self-funded work with no external sponsorship.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rbmo.2020.04.016.

REFERENCES

Abreu, A.P., Kaiser, U.B **Pubertal development and regulation. Lancet. Diabetes.** Endocrinol. 2016; 3: 254–264

Adachi, S., Tajima, A., Quan, J., Haino, K., Yoshihara, K., Masuzaki, H., Katabuchi, H., Ikuma, K., Suginami, H., Nishida, N., Kuwano, R., Okazaki, Y., Kawamura, Y., Sasaki, T., Tokunaga, K., Inoue, I., Tanaka, K. **Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population.** J. Hum. Genet. 2010; 12: 816–821

Albertsen, H.M., Chettier, R., Farrington, P., Ward, K **Genome-wide association study link novel loci to endometriosis.** PLoS One 2013; 3: e58257

American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil. Steril. 1997; 5: 817–821

Ashrafi, M., Sadatmahalleh, S.J., Akhoond, M.R., Talebi, M. **Evaluation of risk factors associated with endometriosis in infertile women.** Int. J. Fertil. Steril. 2016; 1: 11–21

Borghese, B., Tost, J., de Surville, M., Busato, F., Letourneur, F., Mondon, F., Vaiman, D., Chapron, C. **Identification of susceptibility genes for peritoneal, ovarian, and deep infiltrating endometriosis using a pooled sample-based genome-wide association study.** BioMed. Research International 2015; 2015461024

Buck Louis, G.M., Hediger, M.L., Peña, J.B. **Intrauterine exposures and risk of endometriosis.** Hum. Reprod. 2007; 12: 3232–3236

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R. **An atlas of genetic correlations across human diseases and traits.** Nat Genet. 2015; 47: 1236–1241. doi:10.1038/ng.3406

Burney, R.O., Giudice, L.C. **Pathogenesis and pathophysiology of endometriosis.** Fertil. Steril. 2012; 3: 511–519

Calle, M.L., Urrea, V., Vellalta, G., Malats, N., Steen, K.V. **Improving strategies for detecting genetic patterns of disease susceptibility in association studies.** Stat. Med. 2008; 30: 6532–6546

Calle, M.L., Urrea, V., Malats, N., Van Steen, K. **mbmdr: an R package for exploring gene-gene interactions associated with binary or quantitative traits.** Bioinformatics 2010; 17: 2198–2199

Candiani, G.B., Danesino, V., Gastaldi, A., Parazzini, F., Ferraroni, M. **Reproductive and menstrual factors and risk of peritoneal and ovarian endometriosis.** Fertil. Steril. 1991; 2: 230–234

Che, R., Jack, J.R., Motsinger-Reif, A.A., Brown, C.C.. **An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use.** BioData Min. 2014; 7: 9. doi:10.1186/1756-0381-7-9

Chen, G.B., Xu, Y., Xu, H.M., Li, M.D., Zhu, J., Lou, X.Y. **Practical and theoretical considerations in study design for detecting gene-gene interactions using MDR and GMDR approaches.** PLoS One 2011; 6: e16981

Chen, Z.J., Zhao, H., He, L., Shi, Y., Qin, Y., Shi, Y., Li, Z., You, L., Zhao, J., Liu, J., Liang, X., Zhao, X., Zhao, J., Sun, Y., Zhang, B., Jiang, H., Zhao, D., Bian, Y., Gao, X., Geng, L., Li, Y., Zhu, D., Sun, X., Xu, J.E., Hao, C., Ren, C.E., Zhang, Y., Chen, S., Zhang, W., Yang, A., Yan, J., Li, Y., Ma, J, Zhao, Y. **Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3.** Nat. Genet. 2011; 1: 55–59

David, P.H., Reichenbach, L., Savelieva, I., Vartapetova, N., Potemkina, R. **Women's reproductive health needs in Russia: what can we learn from an intervention to improve post–abortion care?** Health Policy Plan 2007; 2: 83–94

Davis, B.J., Risinger, J.I., Chandramouli, G.V.R., Bushel, P.R., Baird, D.D., Peddada, S.D. **Gene expression in uterine leiomyoma from tumors likely to be growing (from black women over 35) and tumors likely to be non-growing (from white women over 35).** PLoS One 2013; 6: e63909

Day, F.R., Bulik-Sullivan, B., Hinds, D.A., Finucane, H.K., Murabito, J.M., Tung, J.Y., Ong, K.K., Perry, J. **Shared genetic aetiology of puberty timing between sexes and with health-related outcomes.** Nat. Commun. 2015; 6: 8842

Day, F.R., Elks, C.E., Murray, A., Ong, K.K., Perry, J.R. **Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study.** Sci. Rep. 2015; 5: 11208

Day, F.R., Thompson, D.J., Helgason, H., Chasman, D.I., Finucane, H., Sulem, P., Ruth, K.S., Whalen, S., Sarkar, A.K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C.M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J.J., Karlsson, R., Kolcic, I., Loh, P.R. **Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk.** Nat. Genet. 2017; 6: 834–841

Douglas, N.I., Pavlova, T.U., Burtseva, T.E., Rad, Y.G., Petrova, P.G., Odland, J.Ø **Women's reproductive health in the Sakha Republic (Yakutia).** Int. J. Circumpolar Health 2014; 73: 10

Elks, C.E., Perry, J.R.B., Sulem, P., Chasman, D.I., Franceschini, N., He, C., Lunetta, K.L., Visser, J.A., Byrne, E.M., Cousminer, D.L., Gudbjartsson, D.F., Esko, T., Feenstra, B., Hottenga, J.J., Koller, D.L., Kutalik, Z., Lin, P., Mangino, M., Marongiu, M., McArdle, P.F., Smith, A.V., Stolk, L. **Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies.** Nat. Genet. 2010; 12: 1077–1085

Gajbhiye, R., Fung, J.N., Montgomery, G.W. **Complex genetics of female fertility.** NPJ. Genom. Med. 2018; 3: 29

Gao, X., Outley, J., Botteman, M., Spalding, J., Simon, J.A., Pashos, C.L. **Economic burden of endometriosis.** Fertil. Steril. 2006; 6: 1561–1572

He, C., Kraft, P., Buring, J.E., Chen, C., Hankison, S.E., Pare, G., Chanock, S., Ridker, P.M., Hunter, D.J. **A large-scale candidate-gene association study of age at menarche and age at natural menopause.** Hum. Genet. 2010; 5: 515–527

Hemani, G, Zheng, J, Elsworth, B, Wade, KH, Baird, D, Haberland, V, Laurin, C, Burgess, S, Bowden, J, Langdon, R, Tan, VY, Yarmolinsky, J, Shihab, HA, Timpson, NJ, Evans, DM, Relton, C, Martin, RM, Davey Smith, G, Gaunt, TR, Haycock, PC **The MR-Base Collaboration. The MR-Base platform supports systematic causal inference across the human phenome.** eLife 2018; 7: e34408. doi:10.7554/eLife.34408

Jones, M.R., Brower, M.A., Xu, N., Cui, J., Mengesha, E., Chen, Y.D., Taylor, K.D., Azziz, R., Goodarzi, M.O. **Systems genetics reveals the functional context of PCOS loci and identifies genetic and molecular mechanisms of disease heterogeneity.** PLoS Genet 2015; 8e1005455

Lee, B., Kim, S.G., Kim, J., Choi, K.Y., Lee, S., Lee, S.K., Lee, J.W. **Brain-specific homeobox factor as a target selector for glucocorticoid receptor in energy balance.** Mol. Cell. Biol. 2013; 14: 2650–2658

Lee, S.H., Harold, D., Nyholt, D.R. **ANZGene Consortium; International Endogene Consortium; Genetic and Environmental Risk for Alzheimer's disease Consortium, Goddard, M.E., Zondervan, K.T., Williams, J., Montgomery, G.W., Wray, N.R., Visscher, P.M. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis.** Hum. Mol. Genet. 2012; 4: 832–841

Lou, X.Y., Chen, G.B., Yan, L., Ma, J.Z., Zhu, J., Elston, R.C., Li, M.D. **A generalized combinatorial approach for detecting gene by gene and gene by environment interactions with application to nicotine dependence.** Am. J. Hum. Genet. 2007; 6: 1125–1137

Mahachie, J.M., Cattaert, T., Van Lishout, F., Gusareva, E.S., Van Steen, K. **Lower-order effects adjustment in quantitative traits model-based multifactor dimensionality reduction.** PLoS One 2012; 7: e29594

Nnoaham, K.E., Webster, P., Kumbang, J., Kennedy, S.H., Zondervan, K.T. **Is early age at menarche a risk factor for endometriosis? A systematic review and meta-analysis of case-control studies.** Fertil. Steril. 2012; 3: 702–712

Nnoaham, K.E., Hummelshoj, L., Webster, P., d'Hooghe, T., de Cicco Nardone, F., de Cicco Nardone, C., Jenkinson, C., Kennedy, S.H., Zondervan, K.T. **World Endometriosis Research Foundation Global Study of Women's Health consortium. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries.** Fertil. Steril. 2011; 2: 366–373

Nogueiras, R., López, M., Lage, R., Perez-Tilve, D., Pfluger, P., Mendieta-Zerón, H., Sakkou, M., Wiedmer, P., Benoit, S.C., Datta, R., Dong, J.Z., Culler, M., Sleeman, M., Vidal-Puig, A., Horvath, T., Treier, M., Diéguez, C., Tschöp, M.H. **Bsx, a novel hypothalamic factor linking feeding with locomotor activity, is regulated by energy availability.** Endocrinology 2008; 6: 3009–3015

Nyholt, D.R., Low, S.K., Anderson, C.A., Painter, J.N., Uno, S., Morris, A.P., MacGregor, S., Gordon, S.D., Henders, A.K., Martin, N.G., Attia, J., Holliday, E.G., McEvoy, M., Scott, R.J., Kennedy, S.H., Treloar, S.A., Missmer, S.A., Adachi, S., Tanaka, K., Nakamura, Y., Zondervan, K.T., Zembutsu, H., Montgomery, G.W. **Genome-wide association meta-analysis identifies new endometriosis risk loci.** Nat. Genet. 2012; 12

O'Connor, L.J., Price, A.L. **Distinguishing genetic correlation from causation across 52 diseases and complex traits.** Nat. Genet. 2018; 50: 1728–1734. doi:10.1038/s41588–018–0255–0

Pagliardini, L., Gentilini, D., Sanchez, A.M., Candiani, M., Vigano, P., Di Blasio, A.M. **Replication and meta-analysis of previous genome-wide association studies confirm vezatin as the locus with the strongest**

evidence for association with endometriosis. Hum. Reprod. 2015; 4: 987–993

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

Painter, J.N., Anderson, C.A., Nyholt, D.R., Macgregor, S., Lin, J., Lee, S.H., Lambert, A., Zhao, Z.Z., Roseman, F., Guo, Q., Gordon, S.D., Wallace, L., Henders, A.K., Visscher, P.M., Kraft, P., Martin, N.G., Morris, A.P., Treloar, S.A., Kennedy, S.H., Missmer, S.A., Montgomery, G.W., Zondervan, K.T. **Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis.** Nat. Genet. 2011; 1: 51–54

Peterson, C.M., Johnstone, E.B., Hammoud, A.O., Stanford, J.B., Varner, M.W., Kennedy, A., Chen, Z., Sun, L., Fujimoto, V.Y., Hediger, M.L., Buck Louis, G.M. **ENDO Study Working Group. Risk factors associated with endometriosis: importance of study population for characterizing disease in the ENDO Study.** Am. J. Obstet. Gynecol. 2013; 6: 451

Plant, T.M. **Neuroendocrine control of the onset of puberty.** Front. Neuroendocrinol. 2015; 38: 73–88

Ponomarenko, I.V. **Selection of polymorphic loci for association analysis in genetic-epidemiological studies.** Research Result Medicine and Pharmacy 2018; 2: 40–54

Ponomarenko, I.V., Reshetnikov, E.A., Altuchova, O.B., Polonikov, A.V., Sorokina, I.N., Yermachenko, A., Dvornyk, V.Ya., Golovchenko, O.V., Churnosov, M.I. **Association of genetic polymorphisms with age at menarche in Russian women.** Gene 2019; 686: 228–236

Rahmioglu, N., Macgregor, S., Drong, A.W., Hedman, Å.K., Harris, H.R., Randall, J.C., Prokopenko, I **International Endogene Consortium (IEC), The GIANT Consortium, Nyholt, D.R., Morris, A.P., Montgomery, G.W., Missmer, S.A., Lindgren, C.M., Zondervan, K.T. Genome-wide enrichment analysis between endometriosis and obesity-related traits reveals novel susceptibility loci.** Hum. Mol. Genet. 2015; 4: 1185–1199

Rahmioglu, N., Nyholt, D.R., Morris, A.P., Missmer, S.A., Montgomery, G.W., Zondervan, K.T. **Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets.** Hum, Reprod. Update 2014; 5: 702–716

Ruth, K.S., Campbell, P.J., Chew, S., Lim, E.M., Hadlow, N., Stuckey, B.G., Brown, S.J., Feenstra, B., Joseph, J., Surdulescu, G.L., Zheng, H.F., Richards, J.B, Murray, A., Spector, T.D., Wilson, S.G., Perry, J.R. **Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes.** Eur. J. Hum. Genet. 2016; 2: 284–290

Sapkota, Y., Fassbender, A., Bowdler, L., Fung, J.N., Peterse, D.O.D., Montgomery, G.W., Nyholt, D.R., D'Hooghe, T.M. **Independent Replication and Meta-Analysis for Endometriosis Risk Loci.** Twin. Res. Hum. Genet. 2015; 5: 518–525

Sapkota, Y., Steinthorsdottir, V., Morris, A.P., Fassbender, A., Rahmioglu, N., De Vivo, I., Buring, J.E., Zhang, F., Edwards, T.L., Jones, S., Peterse, D., Rexrode, K.M., Ridker, P.M., Schork, A.J., MacGregor, S., Martin, N.G., Becker, C.M., Adachi, S., Yoshihara, K., Enomoto, T., Takahashi, A., Kamatani, Y., Matsuda, K., Kubo, M., Thorleifsson, G., Geirsson, R.T., Thorsteinsdottir, U., Wallace, L.M. **iPSYCH-SSI-Broad Group, Yang, J., Velez Edwards, D.R., Nyegaard, M., Low, S.K., Zondervan, K.T., Missmer, S.A., D'Hooghe, T., Montgomery, G.W., Chasman, D.I., Stefansson, K., Tung, J.Y., Nyholt, D.R. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism.** Nature Commun 2017; 8: 15539

Sapkota, Y., Vivo, I.D., Steinthorsdottir, V., Fassbender, A., Bowdler, L., Buring, J.E., Edwards, T.L., Jones, S., O., D., Peterse, D., Rexrode, K.M., Ridker, P.M., Schork, A.J., Thorleifsson, G., Wallace, L.M. **iPSYCH-SSI-Broad Group, Kraft, P., Morris, A.P., Nyholt, D.R., Edwards, D.R.V., Nyegaard, M., D'Hooghe, T., Chasman, D.I., Stefansson, K., Missmer, S.A., Montgomery, G.W. Analysis of potential protein-modifying variants in 9000 endometriosis patients and 150000 controls of European ancestry.** Sci. Rep. 2017; 1: 11380

Sedgh, G., Henshaw, S.K., Singh, S., Bankole, A., Drescher, J. **Legal abortion worldwide: incidence and recent trends.** Int. Fam. Plan Perspect 2007; 3: 106–116

Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z., Zhang, B., Liang, X., Li, T., Chen, J., Shen, J., Zhao, J., You, L., Gao, X., Zhu, D., Zhao, X., Yan, Y., Qin, Y., Li, W., Yan, J., Wang, Q., Zhao, J., Geng, L., Ma, J., Zhao, Y., He, G., Zhang, A., Zou, S., Yang, A., Liu, J., Li, W., Li, B., Wan, C., Qin, Y., Shi, J., Yang, J., Jiang, H., Xu, J.E., Qi, X., Sun, Y., Zhang, Y., Hao, C., Ju, X., Zhao, D., Ren, C.E., Li, X., Zhang, W., Zhang, Y., Zhang, J., Wu, D., Zhang, C., He, L., Chen, Z.J. **Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome.** Nat. Genet. 2012; 9: 1020–1025

Simoens, S., Dunselman, G., Dirksen, C., Hummelshoj, L., Bokor, A., Brandes, I., Brodszky, V., Canis, M., Colombo, G.L., DeLeire, T., Falcone, T., Graham, B., Halis, G., Horne, A., Kanj, O., Kjer, J.J., Kristensen, J., Lebovic, D., Mueller, M., Vigano, P., Wullschleger, M., D'Hooghe, T. **The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres.** Hum. Reprod. 2012; 5: 1292–1299

Sobalska-Kwapis, M., Smolarz, B., Słomka, M., Szaflik, T., Kępka, E., Kulig, B., Siewierska-Górska, A., Polak, G., Romanowicz, H., Strapagiel, D., Szyłło, K. **New variants near RHOJ and C2, HLA-DRA region and susceptibility to endometriosis in the Polish population-The genome-wide association study.** Eur. J. Obstet. Gynecol. Reprod. Biol. 2017; 217: 106–112

Sorokina, I.N., Rudykh, N.A., Bezmenova, I.N., Polyakova, I.S. **Population genetic characteristics and genetic epidemiological research of candidate genes associations with multifactorial diseases.** Research Results in Biomedicine 2018; 4: 20–30

Stefansson, H., Geirsson, R.T., Steinthorsdottir, V., Jonsson, H., Manolescu, A., Kong, A., Ingadottir, G., Gulcher, J., Stefansson, K. **Genetic factors contribute to the risk of developing endometriosis.** Hum. Reprod. 2002; 3: 555–559

Stolk, L., Perry, J.R., Chasman, D.I., He, C., Mangino, M., Sulem, P., Barbalic, M., Broer, L., Byrne, E.M., Ernst, F., Esko, T., Franceschini, N., Gudbjartsson, D.F., Hottenga, J.J., Kraft, P., McArdle, P.F., Porcu, E., Shin, S.Y., Smith, A.V., van Wingerden, S., Zhai, G., Zhuang, W.V., Albrecht, E., Alizadeh, B.Z., Aspelund, T., Bandinelli, S., Lauc, L.B., Beckmann, J.S. **Meta-analyses identify 13 novel loci associated with age at menopause and highlights DNA repair and immune pathways.** Nat. Genet. 2012; 3: 260–268. doi:10.1038/ng.1051

Sundqvist, J., Xu, H., Vodolazkaia, A., Fassbender, A., Kyama, C., Bokor, A., Gemzell-Danielsson, K., D'Hooghe, T.M., Falconer, H. **Replication of endometriosis-associated single-nucleotide polymorphisms from genome-wide association studies in a Caucasian population.** Hum. Reprod. 2013; 3: 835–839

Tian, Y., Zhao, H., Chen, H., Peng, Y., Cui, L., Du, Y., Wang, Z., Xu, J., Chen, Z.J. **Variants in FSHB are associated with polycystic ovary syndrome and luteinizing hormone level in Han Chinese women.** J. Clin. Endocrinol. Metab. 2016; 5: 2178–2184

Treloar, S.A., O'Connor, D.T., O'Connor, V.M., Martin, N.G. **Genetic influences on endometriosis in an Australian twin sample.** Fertil. Steril. 1999; 4: 701–710

Treloar, S.A., Bell, T.A., Nagle, C.M., Purdie, D.M., Green, A.C. **Early menstrual characteristics associated with subsequent diagnosis of endometriosis.** Am. J. Obstet. Gynecol. 2010; 6: 534

Uimari, O., Rahmioglu, N., Nyholt, D.R., Vincent, K., Missmer, S.A., Becker, C., Morris, A.P., Montgomery, G.W., Zondervan, K.T. **Genome-wide genetic analyses highlight mitogen-activated protein kinase (MAPK) signaling in the pathogenesis of endometriosis.** Hum. Reprod. 2017; 4: 780–793

Uno, S., Zembutsu, H., Hirasawa, A., Takahashi, A., Kubo, M., Akahane, T., Aoki, D., Kamatani, N., Hirata, K., Nakamura, Y. **A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese.** Nat. Genet. 2010; 8: 707–710

Verbanck, M., Chen, C., Neale, B **Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases.** Nat Genet 2018; 50: 693–698. doi:10.1038/s41588–018–0099–7

Wang, W., Li, Y., Li, S., Wu, Z., Yuan, M., Wang, T., Wang, S. **Pooling-based genome-wide association study identifies risk loci in the pathogenesis of ovarian endometrioma in Chinese Han women.** Reprod. Sci. 2017; 3: 400–406

Ward, L.D., Kellis, M. **HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.** Nucleic Acids Res. 2012; 40: D930–D934

Zondervan, K.T., Rahmioglu, N., Morris, A.P., Nyholt, D.R., Montgomery, G.W., Becker, C.M., Missmer, S.A **Beyond endometriosis genome-wide association study: from genomics to phenomics to the patient.** Semin. Reprod. Med. 2016; 4: 242–254

Received 19 November 2019; received in revised form 25 February 2020; accepted 21 April 2020.