Gene 757 (2020) 144933

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Candidate genes for age at menarche are associated with endometrial hyperplasia

Irina Ponomarenko[a], Evgeny Reshetnikov[a,*], Alexey Polonikov[b], Inna Sorokina[a], Anna Yermachenko[c,d], Volodymyr Dvornyk[e], Mikhail Churnosov[a]

[a] Department of Medical Biological Disciplines, Belgorod State University, 308015 Belgorod, Russia
[b] Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, 305041 Kursk, Russia
[c] Department of Social Epidemiology, Pierre Louis Institute of Epidemiology and Public Health, 75571 Paris, France
[d] Sorbonne Universités, 75320 Paris, France
[e] Department of Life Sciences, College of Science and General Studies, Alfaisal University, 11533 Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:
Age at menarche
Association
Endometrial hyperplasia
Gene-gene interaction
Gene-environment interactions
Single nucleotide polymorphism

ABSTRACT

Objectives: To study associations candidate genes for age at menarche with a risk of endometrial hyperplasia (EH).

Methods: 52 candidate loci for age at menarche were analyzed for possible association with EH in a sample of 520 patients and 981 controls. Association of the polymorphisms was analyzed using the method of logistic regression. The gene-gene and gene-environment interactions were analyzed using MB-MDR. 21 polymorphisms, which were associated with EH, and 397 polymorphisms linked to them ($r^2 \geq 0.8$) were analyzed *in silico* for their functional significance.

Results: 21 out of the 52 studied polymorphisms had association with EH. Locus rs11031010 *FSHB* was individually associated with the disease according to the dominant (OR = 0.62, p_{perm} = 0.001) and additive (OR = 0.67, p_{perm} = 0.002) models. Haplotype GAA of loci rs555621-rs11031010-rs1782507 *FSHB* were associated with the EH (OR = 0.66, p_{perm} = 0.007). Seventeen loci were associated with EH within 12 most significant models of intergenic interactions ($p_{perm} \leq 0.001$). Locus rs4374421 of the *LHCGR* gene appeared in the largest number of models (four models).

Nine loci involved in 14 most significant models of interactions between SNP, induced abortions, and chronic endometritis were associated with EH. The polymorphisms of genes *FTO* (rs12324955) and *FSHB* (rs11031010) appeared in the largest number of the models (9 and 6, respectively).

Among the 21 loci associated with EH, 16 manifested association also with either age at menarche (7 SNPs) or height and/or BMI (13 SNPs). The above 21 SNPs and 397 SNPs linked to them have non-synonymous, regulatory and eQTL significance for 25 genes, which play roles in the pathways related to development of the female reproductive organs and hormone-mediated signaling (FDR ≤ 0.05).

Conclusions: Candidate genes for age at menarche are associated with endometrial hyperplasia.

1. Introduction

Endometrial hyperplasia (EH) is caused by the proliferation and subsequent thickening of the endometrium. This process results in an increase of the glands/stroma ratio (Montgomery et al., 2004). Estimates range from 133 to 208 cases per 100,000 woman-years for overall EH incidence, 16.8 cases per 100,000 woman-years for atypical EH and 121 cases per 100,000 woman-years for non-atypical EH (Reed et al., 2009; Lacey et al., 2012). In developed countries, about 200 000 new cases of EH are reported annually (Ozdegirmenci et al., 2011). EH

Abbreviations: AAM, age at menarche; BMI, body mass index; EH, endometrial hyperplasia; FSHB, follicle stimulating hormone subunit beta; FTO, fat mass and obesity-associated; GMDR, generalized multifactor dimensionality reduction; LHCGR, luteinizing hormone/choriogonadotropin receptor; MB-MDR, model based multifactor dimensionality reduction; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; MDR, multifactor dimensionality reduction; LD, linkage disequilibrium; LOD, log-odds; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; FDR, false discovery rate; CVC, Cross-Validation Consistency; S_e, sensitivity; S_p, specificity

* Corresponding author at: Belgorod State Universite, Department of Medical Biological Disciplines, 85, Pobedy St., Belgorod 308015, Russia.
E-mail address: reshetnikov@bsu.edu.ru (E. Reshetnikov).

https://doi.org/10.1016/j.gene.2020.144933
Received 10 February 2020; Received in revised form 12 April 2020; Accepted 1 July 2020
Available online 05 July 2020
0378-1119/

is a significant segment in the structure of gynecological morbidity in women of reproductive age and is among the most common causes of gynecological hospitalization (10% to 18%) (Goncharenko et al., 2013).

EH has increased the interest of researchers primarily due to its risk of transforming into malignant tumors (atypical hyperplasia/EIN) and frequent menstrual irregularities (Goncharenko et al., 2013; Emons et al., 2015; Chandra et al., 2016). EH is a pathologically heterogeneous and varies from histologically subtle and spontaneously reversible proliferative lesions to incipient carcinoma (Montgomery et al., 2004; Mazur, 2005). A typical diagnosis is based on the endometrial biopsy or curettage when a woman suffers from abnormal uterine bleeding (Montgomery et al., 2004; Chen et al., 2013).

One of the primary factors of EH pathophysiology is the continuous overstimulation of the endometrium by estrogen (Epplein et al., 2008). Other risk factors of EH include obesity, polycystic ovary syndrome, chronic anovulation, tamoxifen therapy, estrogen-only hormone therapy, postmenopausal status, early menarche or late menopause (Chandra et al., 2016; Epplein et al., 2008; Armstrong et al., 2012; Sanderson et al., 2017). Age at menarche (AAM) has been suggested as a risk factor for EH (Chandra et al., 2016; Sanderson et al., 2017). Women with an early age at menarche have, on average, longer period of menstrual cycling in their life and thus a greater lifelong exposure to estrogens, which may promote EH (Chandra et al., 2016). Importantly, chronic exposure to estrogens in the absence of progesterone is thought to be particularly important for the development of EH (Sanderson et al., 2017).

Although family history was suggested as an important risk factor for EH (Sanderson et al., 2017), the role of hereditary factors in the development of EH have been poorly studied; no family or twin studies have been conducted so far, whereas candidate gene association studies are scarce (Pachomov et al., 2014; Aban et al., 2006; Dieudonné et al., 2014; Demakova, 2018; Ivanova et al., 2015). On the other hand, > 350 loci have been reported by genome-wide association study (GWAS) for association with age at menarche (Day et al., 2017). There is data showing that candidate genes for age at menarche (e.g., *FTO, LIN28B, MAP2K5, TNNI3K, GPRC5B, FANCL,* and the others) may contribute to various anthropometric traits (e.g., body mass index (BMI), height, and the others) (Ong et al., 2009; Elks et al., 2010; Fernandez-Rhodes et al., 2013; Perry et al., 2014), while some of these traits, for example, an increased BMI, may be a risk factor for EH (Chandra et al., 2016; Sanderson et al., 2017). This suggests that candidate genes for AAM may also be associated with EH. However, no studies on association of candidate genes for age at menarche with EH have been conducted so far. This work is aimed to fill this gap.

The present study analyzed 52 candidate loci for menarcheal age as to their possible association with EH. Recently, we studied the association of these loci with some phenotypic traits (including age at menarche, height, and BMI) in the same sample of Caucasian females (Ponomarenko et al., 2019). Using the same sample for the analysis of different traits eliminates between-sample heterogeneity and makes comparison and interpretation of the results more meaningful and robust.

2. Subjects and methods

2.1. Study subjects

The study protocol was approved by the Regional Ethics Committee of Belgorod State University. All participants signed informed consent documents before entering the study. The participants were recruited through the Perinatal Center of the Belgorod Regional Clinical Hospital of St. Joasaph during 2008–2013. All patients with EH underwent endometrial biopsy or curettage followed by morphological verification of the diagnosis. Patients with hyperplasia without atypia (according to the 4th edition of *Classification of Tumours of the Female Reproductive Organs*, WHO 2014) (Emons et al., 2015; Kurman et al., 2014) were included in the case sample. The control group included women without clinical (an asymptomatic women without abnormal uterine bleeding, etc.) and ultrasound signs of benign proliferative diseases of the reproductive organs. The participants of the control sample were recruited during regular medical examinations at the above Perinatal Center. A total of 1560 women were recruited: 540 patients with EH and 1020 controls.

The participants were enrolled under the following exclusion criteria: self-declared non-Russian descent, a birthplace outside of Central Russia, chronic severe disorders of the vital organs (heart, respiratory or renal failure), cancers of a small pelvis and breast, severe autoimmune disorders.

The following data was collected for each participant: physical characteristics (height, weight, BMI), age at menarche (determined as described elsewhere (Ponomarenko et al., 2019), characteristics of the menstrual cycle (length, duration of menses), other data about reproductive health (number of pregnancies and childbirths, age at first birth, time since last birth, spontaneous and induced abortions, infertility), marital status, family history of EH, use of oral contraceptives and age at first oral contraceptive use, smoking and alcohol use. The participants were also examined for the presence of gynecological disorders (endometrial hyperplasia, endometriosis, adenomyosis, etc.).

2.2. Blood sample collection and DNA handling

The phlebotomy was performed by a certified nurse. DNA was extracted from a buffy coat according to the protocol used in our previous gene association studies (Ponomarenko et al., 2019).

2.3. SNP selection

The criteria and the online tools used to select the single nucleotide polymorphisms (SNPs) for the study were described elsewhere (Ponomarenko et al., 2019). Briefly, the criteria included: 1) Associations with AAM or traits, which have common biological pathways with menarche (e.g, some anthropometric characteristics, obesity, etc.), 2) Effect on gene expression - eQTL (eSNP), 3) Regulatory potential (regSNP), 4) Tag value (tagSNP), and 5) Minor allele frequency (MAF) > 5%.

In total 52 loci were selected for the present study (Supplementary Tables 1-4). All selected SNPs have a significant regulatory potential (Supplementary Table 2), 43 SNPs are eSNPs (Supplementary Table 3), 29 SNPs are tagSNPs.

Fourteen loci among the 52 selected showed association with AAM according to the results of GWAS and 28 loci – according to the results of gene association studies (Supplementary Table 4). Seventeen SNPs were associated with anthropometric traits (Supplementary Table 4). Ten more polymorphisms were not directly associated with AAM, but manifested association or are tagged with the menarche-related traits (e.g., polycystic ovary syndrome, vitamin D metabolism, physical characteristics, etc., Supplementary Table 5).

Several selected loci were previously reported as being associated with AAM (14 SNPs), BMI (15 SNPs), and height (16 SNPs) in the studied sample (Ponomarenko et al., 2019).

2.4. SNP genotyping and data quality control

Genotyping of the DNA samples was conducted using the Sequenom MassARRAY® iPLEX platform (Centre of Genomic Sciences, University of Hong Kong). The protocols of DNA sample preparation and data quality control are described elsewhere (Ponomarenko et al., 2019). All DNA samples met the following quality control criteria: call rate > 95%, the success rate of duplicate check > 99.5%, and the success rate of the blank check > 90%. Finally, the proportion of the determined genotypes for the 52 SNPs was 98.86%. The samples with < 95% of determined genotypes were excluded from the analysis (n = 59). Thus,

Table 1
Characteristics of participants from the case and control groups.

Parameters	Cases (n = 520) $\bar{X}$ ± SD/%(n)	Controls (n = 981) $\bar{X}$ ± SD/%(n)	p
Age, years	41.78 ± 10.04	40.73 ± 8.60	> 0.05
Height, m	1.66 ± 0.06	1.65 ± 0.06	> 0.05
Weight, kg	73.67 ± 14.66	72.49 ± 13.37	> 0.05
BMI, kg/m2	26.94 ± 5.56	26.66 ± 4.61	> 0.05
Proportion of the participants by relative BMI, % (n):			
underweight (< 18.50)	2.69 (14)	1.12 (11)	> 0.05
normal weight (18.50–24.99)	39.23 (204)	42.41 (416)	
overweight (25.00–29.99)	31.92 (166)	30.49 (299)	
obese (> 30.00)	26.15 (136)	25.99 (255)	
Family history of endometrial hyperplasia (mother had endometrial hyperplasia)	32.88 (171)	17.53 (172)	< **0.001**
Married	85.76 (446)	85.93 (843)	> 0.05
Smoking (yes)	15.96 (83)	17.33 (170)	> 0.05
Drinking alcohol (≥7 drinks per week)	3.27 (17)	3.06 (30)	> 0.05
Oral contraceptive use	9.88 (51)	10.09 (99)	> 0.05
Age at first oral contraceptive use (mean, years)	23.26 ± 2.32	23.64 ± 2.36	> 0.05
Age at menarche and menstrual cycle			
Age at menarche, years	13.34 ± 1.28	13.27 ± 1.25	> 0.05
Proportion of the participants by relative age at menarche, % (n)			
early (< 12 years)	5.23 (27)	6.42 (63)	> 0.05
average (12–14 years)	83.53 (431)	79.51 (780)	
late (> 14 years)	11.24 (58)	14.07 (138)	
Duration of bleeding menstrual (mean, days)	5.13 ± 1.39	4.94 ± 0.94	> 0.05
Menstrual cycle length (mean, days)	27.94 ± 2.15	28.15 ± 2.24	> 0.05
Reproductive characteristic			
Age at first birth (mean, years)	21.12 ± 2.37	21.71 ± 3.49	> 0.05
No of gravidity (mean)	2.84 ± 2.45	2.45 ± 1.55	> 0.05
No of births (mean)	1.23 ± 0.88	1.51 ± 0.67	< **0.001**
No of spontaneous abortions (mean)	0.22 ± 0.53	0.24 ± 0.51	> 0.05
No of induced abortions (mean)	1.35 ± 1.55	0.67 ± 0.99	< **0.001**
No of induced abortions:			
0	37.88 (197)	58.92 (578)	< **0.001**
1	25.38 (132)	23.75 (233)	
2	18.85 (98)	10.40 (102)	
3	8.65 (45)	5.40 (53)	
≥4	9.23 (48)	1.53 (15)	
History of infertility	11.92 (62)	5.20 (51)	< **0.001**
Gynecological pathologies			
Cervical disorders	26.54 (138)	25.08 (246)	> 0.05
History of sexually transmitted disease	26.35 (137)	26.91 (264)	> 0.05
Chronic endometritis	14.04 (73)	5.71 (56)	< **0.001**
Chronic inflammation of adnexa	34.23 (178)	31.91 (313)	> 0.05
Uterine leiomyoma	51.54 (268)	–	–
Endometriosis	35.19 (183)	–	–
Adenomyosis	20.58 (107)	–	–

the final study sample included 1501 participants: 520 women with EH and 981 control subjects.

2.5. Statistical analysis

Association of clinical anamnestic risk factors with EH was assessed using the logistic regression analysis as implemented in the *epicalc* package in the R software environment (version 3.4.0 (2017-04-21)).

The allele frequencies were checked for correspondence to the Hardy-Weinberg equilibrium (HWE) using the chi-square test. The logistic regression method was used to analyze associations of the SNPs with EH assuming additive, recessive and dominant genetic models. To account for possible confounding effects, the following set of covariates was applied: family history of EH, history of infertility, the presence of induced abortions in the anamnesis, and chronic endometritis as qualitative variables (yes / no), whereas number of births and induced abortions in the anamnesis as quantitative variables (value of the trait) (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 given sample size (520 patients with EH and 981 controls) was sufficient to detect differences in allelic frequencies between affected subjects and controls at OR = 1.25–1.62 for the additive model, OR = 1.38–1.66 for the dominant model and OR = 1.44–6.18 for the recessive model (at 80% power, α = 0.05 for 2-sided test).

The haplotype blocks were identified using HaploView v.4.2 (Barrett et al., 2005). The association analyses and adaptive permutation test were performed using the PLINK v. 2.050 software (Purcell et al., 2007) (http://zzz.bwh.harvard.edu/plink/).

The interactions between genes were analyzed assuming the two-, three-, and four-locus models and using the MB-MDR method (Model Based Multifactor Dimensionality Reduction) (Calle et al., 2010; Ponomarenko, 2019) and the namesake software (v. 2.6). The permutation test was demonstrated to be efficient for analysis of large massifs of GWAS data without reduction of the power (Che 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 SNPs) were adopted for models of gene-gene interactions: $p < 3.8*10^{-5}$ (< 0.05/1326) for two-locus models, $p < 2.3*10^{-6}$ (< 0.05/22100) for three-locus models, and $p < 1.8*10^{-7}$ (< 0.05/270725) for four-locus models. The significance level was set at $p_{perm} \leq 0.001$.

The interactions of the genes with induced abortions and chronic endometritis were analyzed for their possible effect on EH. Induced abortions and chronic endometritis were included in the analysis because (i) they were determined as risk factors for EH in the study sample (Table 1), (ii) induced abortions are a common birth control method of Russian women as compared to other countries (Sedgh et al., 2007), and (iii) induced abortions are a common cause of post-abortion endometritis and, subsequently, chronic endometritis in Russian women (Douglas et al., 2014). The analysis was performed using MB-MDR with adjustment for covariates (family history of EH, history of infertility) and multiple comparisons (1000 permutations) as described above. The permutation test was applied to the selected best models of gene-environment interactions with the significance level of $p < 1*10^{-12}$. The significance level was set at $p_{perm} < 0.001$.

The most significant models of gene-gene and gene-environment interactions associated with EH were cross-validated using GMDR (Generalized Multifactor Dimensionality Reduction) (Lou et al., 2007; Chen et al., 2011) (http://www.ssg.uab.edu/gmdr) and the respective software (Beta 0.9) (http://sourceforge.net/projects/gmdr). The main parameters of the validation were adjusted for covariates and multiple comparisons using the permutation test (1,000 permutations with 10-fold cross-validation, which provided statistical significance at $p_{perm} < 0.001$ for a validated model).

The identified interactions and proportion of their contribution to the total variance of the trait were visualized using the Multifactor Dimensionality Reduction (MDR) method (http://www.multifactordimensionalityreduction.org/), and the MDR v. 3.0.2 software (http://sourceforge.net/projects/mdr).

2.5.1. Functional SNPs

The HaploReg v. 4.1 (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) and the data of the European population from the 1000 Genomes Project Phase were utilized to determine SNPs linked to the EH-associated polymorphisms. Then all these loci were analyzed *in silico* for their functional significance (non-synonymous SNPs, regulatory potential, and eQTLs).

Table 2
Associations of the 52 SNPs with endometrial hyperplasia.

Chr	SNP	n	Additive model				Dominant model				Recessive model			
			OR	95%CI		P	OR	95%CI		P	OR	95%CI		P
				L95	U95			L95	U95			L95	U95	
1	rs1514175	1496	0.98	0.83	1.15	0.789	0.95	0.76	1.19	0.655	1.02	0.74	1.39	0.926
1	rs466639	1495	0.87	0.68	1.11	0.255	0.85	0.65	1.12	0.242	0.87	0.34	2.17	0.758
1	rs7538038	1495	1.09	0.9	1.31	0.395	1.09	0.87	1.37	0.439	1.16	0.69	1.96	0.578
2	rs713586	1493	0.98	0.84	1.15	0.82	0.95	0.75	1.21	0.683	1.01	0.76	1.34	0.936
2	rs2164808	1495	1.01	0.86	1.19	0.88	0.94	0.73	1.21	0.654	1.1	0.85	1.44	0.467
2	rs7589318	1490	0.97	0.82	1.15	0.739	0.94	0.75	1.18	0.588	1.04	0.71	1.53	0.844
2	rs4374421	1431	0.96	0.81	1.14	0.637	0.86	0.68	1.08	0.181	1.26	0.87	1.84	0.222
2	rs7579411	1485	0.91	0.78	1.07	0.268	0.8	0.63	1.02	0.066	1.03	0.78	1.36	0.846
2	rs6729809	1449	0.92	0.78	1.1	0.364	0.87	0.7	1.1	0.244	1	0.68	1.46	0.98
2	rs4953616	1489	0.96	0.81	1.16	0.695	0.92	0.74	1.15	0.459	1.12	0.72	1.75	0.605
2	rs6732220	1493	0.94	0.79	1.14	0.543	0.92	0.74	1.15	0.477	0.98	0.62	1.57	0.942
2	rs4953655	1493	0.91	0.76	1.09	0.311	0.89	0.71	1.12	0.316	0.88	0.55	1.42	0.61
2	rs887912	1438	0.82	0.67	0.99	0.036	0.8	0.64	1	0.054	0.7	0.42	1.17	0.176
2	rs12617311	1496	0.93	0.79	1.1	0.413	0.87	0.7	1.09	0.222	1.04	0.73	1.47	0.848
3	rs6438424	1486	1.1	0.94	1.29	0.22	1.2	0.93	1.55	0.153	1.08	0.83	1.4	0.574
4	rs2013573	1493	0.87	0.7	1.08	0.202	0.82	0.65	1.04	0.101	1.26	0.63	2.52	0.516
4	rs13111134	1494	0.9	0.74	1.09	0.282	0.84	0.67	1.05	0.12	1.21	0.71	2.05	0.489
4	rs222003	1497	0.76	0.54	1.06	0.108	0.77	0.54	1.09	0.134	0	0	inf	0.999
4	rs222020	1497	1.21	0.95	1.54	0.132	1.26	0.97	1.63	0.081	0.7	0.21	2.31	0.554
4	rs3756261	1479	0.9	0.66	1.24	0.537	0.91	0.66	1.26	0.555	0.67	0.05	8.49	0.759
5	rs757647	1483	0.9	0.74	1.09	0.269	0.9	0.72	1.13	0.377	0.75	0.43	1.31	0.316
6	rs7766109	1498	0.91	0.77	1.07	0.24	0.87	0.68	1.1	0.245	0.9	0.68	1.2	0.475
6	rs4946651	1497	0.9	0.77	1.06	0.202	0.82	0.65	1.03	0.085	0.97	0.73	1.3	0.842
6	rs7759938	1496	0.83	0.69	0.99	0.036	0.75	0.6	0.94	0.011	0.95	0.63	1.44	0.822
6	rs314280	1452	0.91	0.77	1.08	0.272	0.83	0.66	1.05	0.124	1	0.73	1.37	0.985
6	rs314276	1454	0.87	0.73	1.04	0.118	0.77	0.62	0.97	0.027	1.05	0.72	1.53	0.796
6	rs3020394	1497	0.9	0.76	1.06	0.204	0.89	0.71	1.11	0.284	0.82	0.55	1.21	0.307
6	rs1884051	1497	0.91	0.76	1.07	0.255	0.91	0.73	1.13	0.381	0.81	0.54	1.2	0.29
6	rs7753051	1494	0.91	0.76	1.09	0.308	0.86	0.69	1.07	0.185	1.01	0.67	1.53	0.945
7	rs1079866	1496	1.01	0.82	1.24	0.913	1.01	0.79	1.28	0.958	1.07	0.56	2.06	0.841
8	rs2288696	1500	1.01	0.83	1.23	0.908	1.03	0.82	1.3	0.783	0.89	0.49	1.63	0.717
9	rs2090409	1428	1.05	0.89	1.24	0.576	1.21	0.96	1.53	0.104	0.81	0.57	1.14	0.222
9	rs10980926	1492	1.02	0.86	1.21	0.829	0.97	0.78	1.22	0.821	1.18	0.81	1.71	0.387
9	rs10441737	1395	0.97	0.81	1.15	0.712	0.91	0.73	1.15	0.447	1.09	0.75	1.58	0.652
11	rs10769908	1477	0.95	0.81	1.12	0.53	0.89	0.7	1.14	0.355	0.99	0.76	1.3	0.954
11	rs555621	1493	0.91	0.77	1.07	0.241	0.87	0.69	1.1	0.252	0.89	0.65	1.23	0.487
11	rs11031010	1475	**0.67**	**0.52**	**0.86**	**0.002**	**0.62**	**0.46**	**0.82**	**< 0.001**	0.81	0.33	1.96	0.635
11	rs1782507	1493	1.14	0.97	1.35	0.117	1.13	0.9	1.42	0.28	1.31	0.94	1.83	0.114
11	rs6589964	1496	1.05	0.9	1.23	0.507	1.07	0.84	1.37	0.586	1.08	0.83	1.4	0.59
12	rs1544410	1492	1.04	0.88	1.22	0.666	0.98	0.78	1.23	0.869	1.19	0.87	1.62	0.281
14	rs999460	1497	0.89	0.76	1.06	0.193	0.83	0.67	1.04	0.109	0.96	0.67	1.37	0.825
14	rs4986938	1495	1.09	0.92	1.28	0.325	1	0.8	1.25	0.991	1.43	1.02	2	0.039
15	rs2241423	1485	0.79	0.64	0.98	0.033	0.79	0.62	1.01	0.06	0.57	0.28	1.15	0.116
16	rs12444979	1491	0.9	0.72	1.12	0.338	0.92	0.72	1.19	0.524	0.58	0.26	1.3	0.186
16	rs9939609	1496	0.94	0.8	1.1	0.449	0.87	0.69	1.1	0.253	1	0.75	1.34	0.986
16	rs12324955	1497	1.02	0.86	1.21	0.858	1.1	0.88	1.37	0.42	0.81	0.54	1.21	0.307
18	rs1398217	1487	1.1	0.94	1.3	0.231	1.16	0.91	1.48	0.225	1.11	0.83	1.48	0.499
19	rs2252673	1489	1	0.82	1.21	0.969	1	0.79	1.26	0.995	0.97	0.55	1.73	0.921
20	rs1073768	1496	0.95	0.81	1.11	0.487	0.86	0.67	1.1	0.218	1.02	0.78	1.32	0.894
22	rs4633	1496	0.98	0.84	1.15	0.843	1.05	0.82	1.35	0.706	0.91	0.7	1.18	0.475
23	rs5930973	1473	1.19	0.85	1.66	0.304	NA	NA	NA	NA	NA	NA	NA	NA
23	rs3092921	1497	1.01	0.75	1.36	0.93	NA	NA	NA	NA	NA	NA	NA	NA

All results were obtained after adjustment for covariates.
Statistically signifcant values after adjustment by the adaptive permutation test.
OR, odds ratio, 95%CI, 95% confidence interval.

2.5.2. Non-synonymous SNPs

The SIFT online tool (http://sift.jcvi.org/) and PolyPhen-2 online tool (http://genetics.bwh.harvard.edu/pph2/index.shtml) were used to analyze non-synonymous SNPs and their functional predictions.

2.5.3. Regulatory effects

The regulatory potential of the candidate loci for EH was analyzed using SNP Function Prediction (FuncPred) (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html), rSNPBase (http://rsnp.psych.ac.cn/index.do), RegulomeDB (Version 1.1) (http://regulome.stanford.edu/), and HaploReg (v4.1).

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 log-odds (LOD) scores of the alternative (alt) and reference (ref) alleles (Ward and Kellis, 2016): ΔLOD = LOD (alt) − LOD (ref). A negative value of this indicator suggests an increase in the affinity of this motif by the reference allele; on the contrary, a positive value demonstrates a connection between the alternative allele and an increase in the affinity of the analyzed DNA motif.

The possible regulatory effects of polymorphisms, which are in strong linkage disequilibrium (LD) ($r^2 \geq 0.8$) with those associated

with endometriosis, were analyzed using HaploReg (v4.1).

2.5.4. Expression QTLs

The data from the Blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) was used to estimate the effect of the candidate SNPs for endometriosis on gene expression level (*cis*- and *trans*-eQTL) in peripheral blood; the estimates for other organs and tissues were obtained 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). Likewise, eQTL values of SNPs in the strong LD ($r^2 \geq 0.8$) with the endometriosis-associated polymorphisms were estimated.

2.5.5. Pathway analyses

The functional significance of the EH-associated genes in metabolic pathways were analyzed *in silico* using the tools available on the Gene Ontology Portal (PANTHER Overrepresentation Test от 13.04.2017; PANTHER version 12.0 от 10.07.2017, http://geneontology.org) and the false discovery rate (FDR) test to correct the results for multiple comparisons. GeneMANIA (http://genemania.org, version 3.5.0, accessed on 13 March 2017) and the automatic weighting were used to construct the gene interaction networks.

3. Results

3.1. Study participants characteristics

The groups of patients with EH (n = 520) and controls (n = 981) did not differ by age and BMI (p > 0.05) (Table 1). The patients with EH had higher rates of family history of EH (OR = 2.30 95% CI 1.79–2.97p < 0.001), history of infertility (OR = 2.47 95% CI 1.65–3.70p < 0.001), presence (OR = 2.35 95% CI 1.88–2.94p < 0.001) and the number of induced abortions in the anamnesis (OR = 1.55 95% CI 1.41–1.70p < 0.001), chronic endometritis (OR = 2.70 95% CI 1.84–3.95p < 0.001), and lower parity (OR = 0.84 95% CI 0.76–0.92p < 0.001) (Table 1). These risk factors for EH were used as covariates in the association analyses.

3.2. SNP and haplotype association analysis

The data about the analyzed loci are presented in Supplementary Tables 5 and 6. All loci had MAF > 5% and were in correspondence with the HWE ($p_{bonf} < 0.001$).

Allele A of the rs11031010 *FSHB* locus was associated with EH according to the dominant (OR = 0.62, 95% CI 0.46–0.82, p < 0.001, p_{perm} = 0.001, power − 94.13%) and additive (OR = 0.67, 95% CI 0.52–0.86, p = 0.002, p_{perm} = 0.002, power − 89.93%) models (Table 2). Haplotype GAA of loci rs555621-rs11031010-rs1782507 (*FSHB*) was associated with EH (OR = 0.66, p = 0.001, p_{perm} = 0.007) (Supplementary Table 7).

3.3. SNP × SNP interactions

In total 12 most significant 2, 3 and 4-locus models of gene-gene interactions associated with EH were determined ($p_{perm} \leq 0.001$, Cross-Validation Consistency (CVC) = 10/10, testing balanced accuracy 52.67–57.66%, sensitivity (S_e) and specificity (S_p) of the best models were 89.42% and 72.27%, respectively) (Table 3 and Supplementary Table 8). These models included 17 SNPs. Locus rs4374421 of the *LHCGR* gene appeared in the largest number of models (four models). Loci rs1782507 *FSHB,* rs6589964 *BSX,* rs7766109 *F13A1* contributed to the most significant gene-gene interaction models at all the levels considered. The following genotype combinations were determined to have the most significant association with EH: rs13111134 GA *UGT2B4* × rs7759938 TC *LIN28B* (*beta* = -0.65, p = 0.0001), rs10441737 TT *ZNF483* × rs4374421 TT (*beta* = 0.49, p = 0.0002), rs4986938 GA *ESR2* × rs10441737 TC *ZNF483* × rs4374421 TT *LHCGR* (*beta* = -0.64, p = 0.0003) (Supplementary Table 9).

The graph of SNP-SNP interactions of 17 polymorphisms (Fig. 1) suggests these interactions are concerted; the highest contribution to the entropy is made by interactions rs1782507 *FSHB* × rs7766109 *F13A1* (0.55%), rs11031010 *FSHB* × rs6589964 *BSX* (0.48%), rs555621 *FSHB* × rs2090409 *TMEM38B* (0.47%) and polymorphism rs11031010 *FSHB* (0.46%).

3.4. Gene-environment interactions

In total nine loci were determined to contribute to gene-environment interactions with induced abortions and chronic endometritis, which were significantly associated with EH. The analysis yielded 14 best two-, three-, and four-order models($p_{perm} < 0.001$, CVC = 10/10, Testing balanced accuracy 60.68–63.87%, sensitivity (S_e) and specificity (S_p) of the best models − 67.69% and 73.09% respectively) (Table 4 and Supplementary Table 10). Two loci, rs12324955 *FTO* and rs11031010 *FSHB*, contributed to the largest number of the best models (nine and six, respectively) (Table 4). Loci rs4953655 *FSHR,* rs6732220 *FSHR,* rs12324955 *FTO*, and rs999460 *NKX2-1* were significantly associated with EH through the interaction with induced abortions and chronic endometritis. The following combinations of genotypes with chronic endometritis and induced abortions were associated with EH: rs11031010 CC *FSHB* × abortion (*beta* = 0.78, p = $4.53*10^{-12}$), rs11031010 CC *FSHB* × chronic endometritis × abortion (*beta* = 0.71, p = $1.02*10^{-9}$) (Supplementary Table 11).

The graph of the interactions between the studied loci, induced abortions, and chronic endometritis suggests that the abortions account for the largest proportion (2.91%) of the trait entropy (Fig. 2). In general, the contribution of the most significant gene-environment interactions to EH is 0.40–1.15%, on average.

3.5. Snps associated with endometrial hyperplasia are also associated with age at menarche, height and body mass index in adults

The analyses determined that 21 loci are associated with EH, either individually or through gene-gene and gene-environment interactions. Previously, we also analyzed whether these SNPs were associated with the other phenotypic characteristics, namely menarcheal age, height and BMI in the studied sample (Ponomarenko et al., 2019). Out of the 21 SNPs, 16 SNPs (76.19%) also manifested association with age at menarche, height and / or BMI (Supplementary Table 12). Locus rs4633 *COMT* manifested association with all above phenotypes, and 3 loci (rs4374421 *LHCGR,* rs6589964 *BSX* and rs2090409 *TMEM38B*) were also associated with at least 2 of the 3 phenotypes analyzed (i.e., age of menarche and / or height and / or BMI).

3.6. Functional SNP

3.6.1. Non-synonymous SNPs

None of the 21 polymorphisms associated with EH was missense. However, two of these were in linkage disequilibrium with non-synonymous SNPs. Polymorphism rs4633 was linked to rs4680 (r^2 = 0.99), which is an amino acid substitution Val18Met in the COMT protein. This replacement has SIFT Score = 0.02, which corresponds to the predictive value “deleterious”. Another locus, rs7538038 (1q32.1), was linked to rs4889 (r^2 = 0.98), a missense variant (Pro81Arg) in the KISS1 protein. This amino acid change has a predictive value of “possibly damaging” (PolyPhen-2 score = 0.524, sensitivity 0.88, specificity 0.90).

3.6.2. Regulatory effects

The data on regulatory effects of the EH-associated SNPs are given in Supplementary Table 13. The most significant regulatory potential

Table 3
The most significant models of SNP-SNP interactions associated with endometrial hyperplasia.

N	Models of SNP × SNP interactions	NH	*beta*H	WH	NL	*beta*L	WL	P_{perm}
Two-locus models ($p < 3*10^{-5}$)								
1	rs2013573 *UGT2B4* × rs7759938 *LIN28B*	1	0.361	6.02	2	−0.748	19.47	< 0.001
2	rs13111134 *UGT2B4* × rs7759938 *LIN28B*	1	0.303	4.80	2	−0.695	19.33	< 0.001
3	rs10441737 *ZNF483* × rs4374421 *LHCGR*	2	0.544	19.07	1	−0.474	9.27	< 0.001
4	rs1782507 *FSHB* × rs7766109 *F13A1*	2	0.573	18.98	2	−0.408	8.12	< 0.001
5	rs6589964 *BSX* × rs11031010 *FSHB*	1	0.244	4.82	2	−0.710	17.37	0.001
Three-locus models ($p < 5*10^{-8}$)								
1	rs6589964 *BSX* × rs1782507 *FSHB* × rs7766109 *F13A1*	7	0.725	32.49	5	−0.744	28.00	< 0.001
2	rs2090409 *TMEM38B* × rs6438424 *3q13.32* × rs555621 *FSHB*	6	0.721	31.97	2	−1.332	14.93	< 0.001
3	rs314276 *LIN28B* × rs10441737 *ZNF483* × rs4374421 *LHCGR*	3	0.565	18.69	4	−0.675	30.84	< 0.001
4	rs4986938 *ESR2* × rs10441737 *ZNF483* × rs4374421 *LHCGR*	5	0.683	29.73	3	−0.732	30.69	< 0.001
5	rs4986938 *ESR2* × rs10980926 *ZNF483* × rs4374421 *LHCGR*	4	0.644	20.29	3	−0.704	29.98	< 0.001
Four-locus models ($p < 5*10^{-13}$)								
1	rs2090409 *TMEM38B* × rs6589964 *BSX* × rs1782507 *FSHB* × rs7766109 *F13A1*	11	0.990	51.19	6	−0.930	23.36	< 0.001
2	rs2090409 *TMEM38B* × rs314276 *LIN28B* × rs4633 *COMT* × rs7538038 *KISS1*	12	0.958	64.12	4	−0.728	15.60	< 0.001

The results were obtained using the MB-MDR method with adjustment for covariates.
NH, number of significant high-risk genotypic combinations associated with endometrial hyperplasia.
beta H, coefficient of the logistic regression for significant high-risk genotypic combinations associated with endometrial hyperplasia.
WH, the Wald test value for significant high-risk genotypic combinations associated with endometrial hyperplasia.
NL, number of significant low-risk genotypic combinations associated with the endometrial hyperplasia.
beta L, coefficient of the logistic regression for significant low-risk genotypic combinations associated with the endometrial hyperplasia.
WL, the Wald test value for significant low-risk genotypic combinations associated with endometrial hyperplasia.
$P_{perm,}$ P value for the permutation test (1.000 permutations).

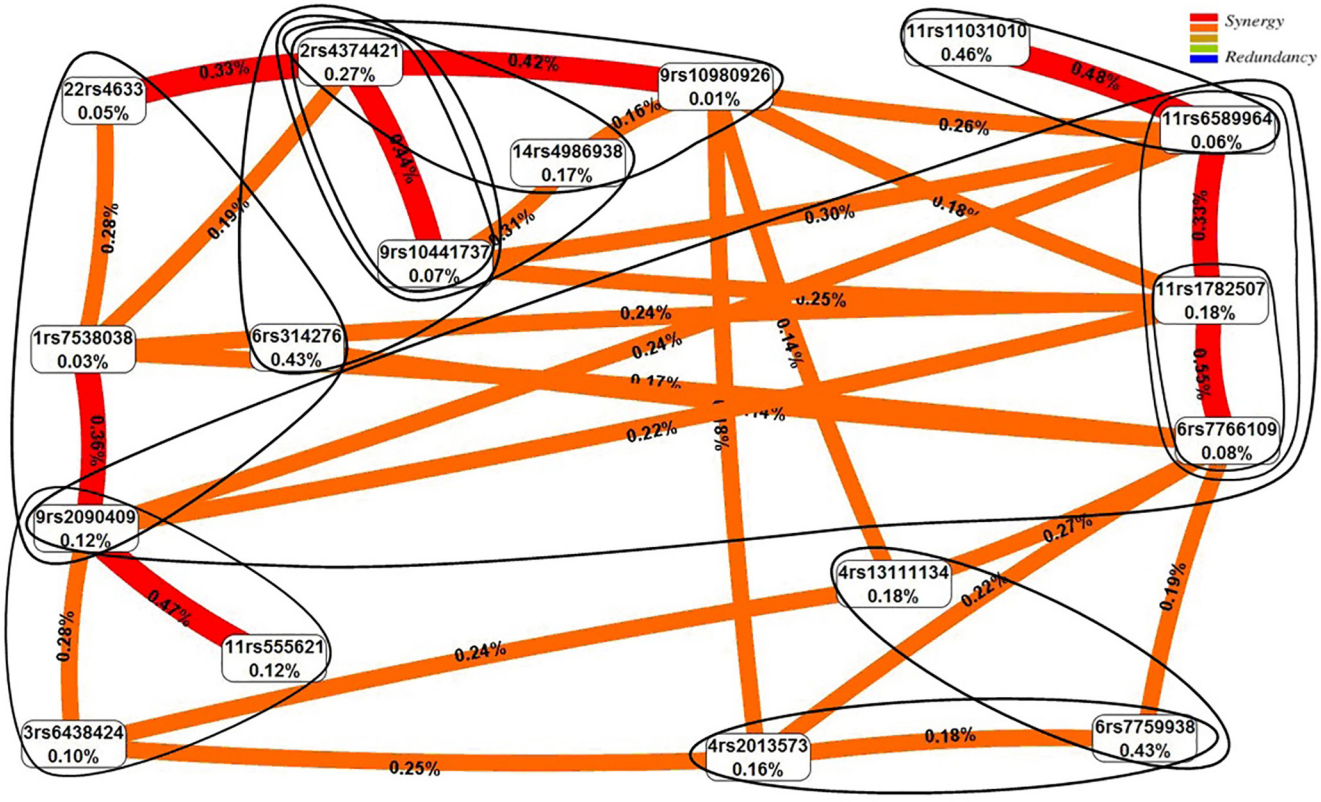

Fig. 1. The entropy graph of SNP-SNP interactions with endometrial hyperplasia based on the multifactor dimensionality reduction (MDR) analysis. Positive values of entropy indicate synergistic interactions. The polymorphisms are denoted by the chromosome number and rs SNP ID. The red and orange colors denote strong and moderate synergism, respectively. The figures outline the SNP × SNP interactions within the 2-, 3-, and 4-locus models obtained by the MB-MDR method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was determined for rs4633 *COMT* and rs4986938 *ESR2*. According to the HaploReg (v4.1), allele A of rs11031010 *FSHB* increases affinity to transcription factors (TFs) Foxp3 (ΔLOD = 3.2) and Foxa_known2 (ΔLOD = 1.7), and decreases affinity to TFs TCF4_disc2 (ΔLOD = -12.0) and Foxd1_1 (ΔLOD = -1.0). Also, allele A rs1782507 *FSHB* decreases affinity to TFs HDAC2_disc5 (ΔLOD = -9.7), HNF4_disc1 (ΔLOD = -9.5), Lmo2-complex_1 (ΔLOD = -11.9), TCF12_disc1 (ΔLOD = -12.0), ZEB1_known3 (ΔLOD = -11.9).

The 397 SNPs linked to the EH-associated loci were analyzed for their regulatory potential (Supplementary Table 14). Out of these, four SNPs (two non-synonymous and two synonymous) were located in exons of the studied genes, four polymorphisms were found in 3′UTR,

Table 4
The most significant models of gene-environment interactions associated with endometrial hyperplasia.

N	Models of gene-environment interactions	NH	*beta*H	WH	NL	*beta*L	WL	P_{perm}
Two-order interaction models ($p < 1*10^{-12}$)								
1	rs11031010 *FSHB* × abortions	1	0.839	55.48	2	−0.732	42.85	< 0.001
2	rs12324955 *FTO* × abortions	2	0.818	53.69	2	−0.785	48.33	< 0.001
Three-order interaction models ($p < 1*10^{-14}$)								
1	rs11031010 *FSHB* × abortions × chronic endometritis	4	0.985	76.18	2	−0.893	61.80	< 0.001
2	rs12324955 *FTO* × abortions × chronic endometritis	5	0.950	71.56	2	−0.949	67.70	< 0.001
3	rs12324955 *FTO* × rs555621 *FSHB* × abortions	5	0.901	63.67	3	−0.786	40.16	< 0.001
4	rs4633 *COMT* × rs4374421 *LHCGR* × abortions	5	0.899	63.07	4	−0.688	31.49	< 0.001
5	rs11031010 *FSHB* × rs12324955 *FTO* × abortions	2	0.905	62.19	3	−0.785	37.09	< 0.001
Four-order interaction models ($p < 1*10^{-17}$)								
1	rs6589964 *BSX* × rs11031010 *FSHB* × rs6732220 *FSHR* × abortions	7	1.090	84.24	5	−0.650	25.42	< 0.001
2	rs6589964 *BSX* × rs11031010 *FSHB* × rs4953655 *FSHR* × abortions	7	1.065	80.81	4	−0.761	22.10	< 0.001
3	rs11031010 *FSHB* × rs12324955 *FTO* × rs999460 *NKX2-1* × abortions	6	1.026	79.79	5	−0.855	43.33	< 0.001
4	rs12324955 *FTO* × rs4374421 *LHCGR* × rs4953655 *FSHR* × abortions	9	1.047	78.09	4	−1.001	40.29	< 0.001
5	rs12324955 *FTO* × rs4953655 *FSHR* × abortions × chronic endometritis	6	0.979	62.76	7	−1.008	77.14	< 0.001
6	rs12324955 *FTO* × rs4633 *COMT* × abortions × chronic endometritis	6	0.877	51.07	7	−1.033	76.96	< 0.001
7	rs12324955 *FTO* × rs555621 *FSHB* × abortions × chronic endometritis	7	0.992	75.62	5	−1.002	66.40	< 0.001

The results were obtained using the MB-MDR method with adjustment for covariates.
NH, number of significant high-risk genotypic, aborts chronic and endometritis combinations associated with endometrial hyperplasia.
beta H, coefficient of the logistic regression for significant high-risk genotypic, aborts and chronic endometritis combinations associated with endometrial hyperplasia.
WH, the Wald test value for significant high-risk genotypic, aborts and chronic endometritis combinations associated with endometrial hyperplasia.
NL, number of significant low-risk genotypic, aborts and chronic endometritis combinations associated with the endometrial hyperplasia.
beta L, coefficient of the logistic regression for significant low-risk genotypic, aborts and chronic endometritis combinations associated with the endometrial hyperplasia.
WL, the Wald test value for significant low-risk genotypic, aborts and chronic endometritis combinations associated with endometrial hyperplasia.
$P_{perm,}$ P value for the permutation test (1.000 permutations).

158 were located in introns and 252 in intergenic regions. There were 16 SNPs located in the evolutionarily conserved regions (Supplementary Table 14).

Among these, > 350 had regulatory significance. The most pronounced regulatory potential was determined for loci linked to polymorphisms rs11031010, rs555621, and rs1782507 of the *FSHB* gene (4 SNPs), rs6732220 and rs4953655 of the *FSHR* gene (5 SNPs), rs4633 of the *COMT* gene (3 SNPs), rs314276 and rs7759938 of the *LIN28B* gene (2 SNPs) (Supplementary Table 14).

For example, rs1222218 is linked to three loci (rs555621, rs11031010, and rs1782507) of the *FSHB* gene (r^2 = 0.28, r^2 = 0.21, r^2 = 0.82, respectively) associated with EH. This SNP is located in the region of histones marking promoters in 24 tissues, the DNAase- 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 them (r^2 = 0.26, r^2 = 0.20, r^2 = 0.80, respectively) is located in the region of 32 regulatory DNA motifs.

More importantly, the loci that are linked to the EH-associated SNPs, showed their regulatory effects in organs and tissues related to pathogenesis of EH, i.e., ovaries, muscle tissue, adipose tissue, various brain regions (hypothalamus, pituitary, etc.), liver, etc. (Sanderson et al., 2017).

3.6.3. Expression QTLs

Five SNPs from the 21 loci associated with EH were also associated ($p < 5*10^{-5}$, $p_{FDR} < 0.05$) with the transcription level of four genes (*F13A1*, *COMT*, *KIAA0368, C11orf46* (*ARL14EP*)) in peripheral blood (*cis*-eQTL) (Supplementary Table 14Three of these SNPs are also linked to the other *cis*-eQTL loci affecting mRNA transcription level in blood (*KIAA0368, COMT)* (Supplementary Table 16). No *trans*-eQTL SNPs were determined (FDR > 0.05).

According to the data of the Genotype-Tissue Expression (GTEx) project, 13 SNPs of the 21 associated with EH had *cis*-eQTL significance in various tissues and organs ($p < 8*10^{-5}$, $p_{FDR} \leq 0.05$) (Supplementary Table 17). The following loci demonstrated possible effects on gene transcription in organs and tissues pathogenetically significant for the development of endometriosis: rs1782507 and rs555621 (*ARL14EP*) - in the basal ganglia of the brain, rs1782507 and rs555621 (*FSHB*) - hypothalamus, rs7759938 and rs314276 (*LIN28B*), rs1782507 and rs555621 (*ARL14EP*) - the pituitary gland, rs1782507 (*ARL14EP*) - the ovaries, rs11031010 (*ARL14EP*) - subcutaneous adipose tissue, rs6438424 (*RP11-384F7.2*), rs1782507 (*ARL14EP*) - visceral adipose tissue, rs1782507, rs11031010 and rs555621 (*ARL14EP*) - thyroid gland, rs6438424 (*RP11-384F7.2* and *LSAMP*), rs1782507 (*ARL14EP*) - adrenals, rs555621 (*ARL14EP*) - peripheral blood. The highest *cis*-eQTL value was determined for the polymorphisms rs555621, rs11031010 and rs1782507 of the *FSHB* gene associated with expression of three genes in > 25 different organs and tissues (Supplementary Table 17).

Allele A (alt) of rs11031010 *FSHB* (protective effect for EH, OR = 0.62–0.66) is associated with the high expression of the *ARL14EP* gene in 10 different organs and tissues (β = 0.17–0.76, $p = 4.2*10^{-5}$–$8.7*10^{-15}$, $p_{FDR} \leq 0.05$). Allele G (alt) of rs555621 and allele A (alt) of rs1782507 *FSHB* (protective effect for EH within the haplotype, OR = 0.66) are associated with the high expression of the genes (*ARL14EP, RP4-710 M3.1, FSHB* and *ARL14EP, FSHB,* respectively) in more than 25 different organs and tissues (Supplementary Table 17).

Sixteen EH-associated loci manifested strong linkage to more than 280 polymorphisms affecting gene expression ($p < 8.5*10^{-5}$, $p_{FDR} \leq 0.05$) in various organs and tissues (Supplementary Table 18). Loci rs1782507, rs11031010 and rs555621 of *FSHB* appeared to have the most pronounced effect as they were linked to > 120 SNPs affecting expression of the *FSHB*, *ARL14EP*, and *RP4-710 M3.1* genes in > 25 organs and tissues.

Overall, 17 of the 21 EH-associated loci have *cis*-eQTL value: they affect the expression of 19 genes. Among these, one SNP was associated with the mRNA transcription level independently, one SNP was in linkage disequilibrium ($r^2 \geq 0.8$) with other eQTL SNPs, and 15 SNPs were both individually associated with and linked to other loci affecting gene expression levels.

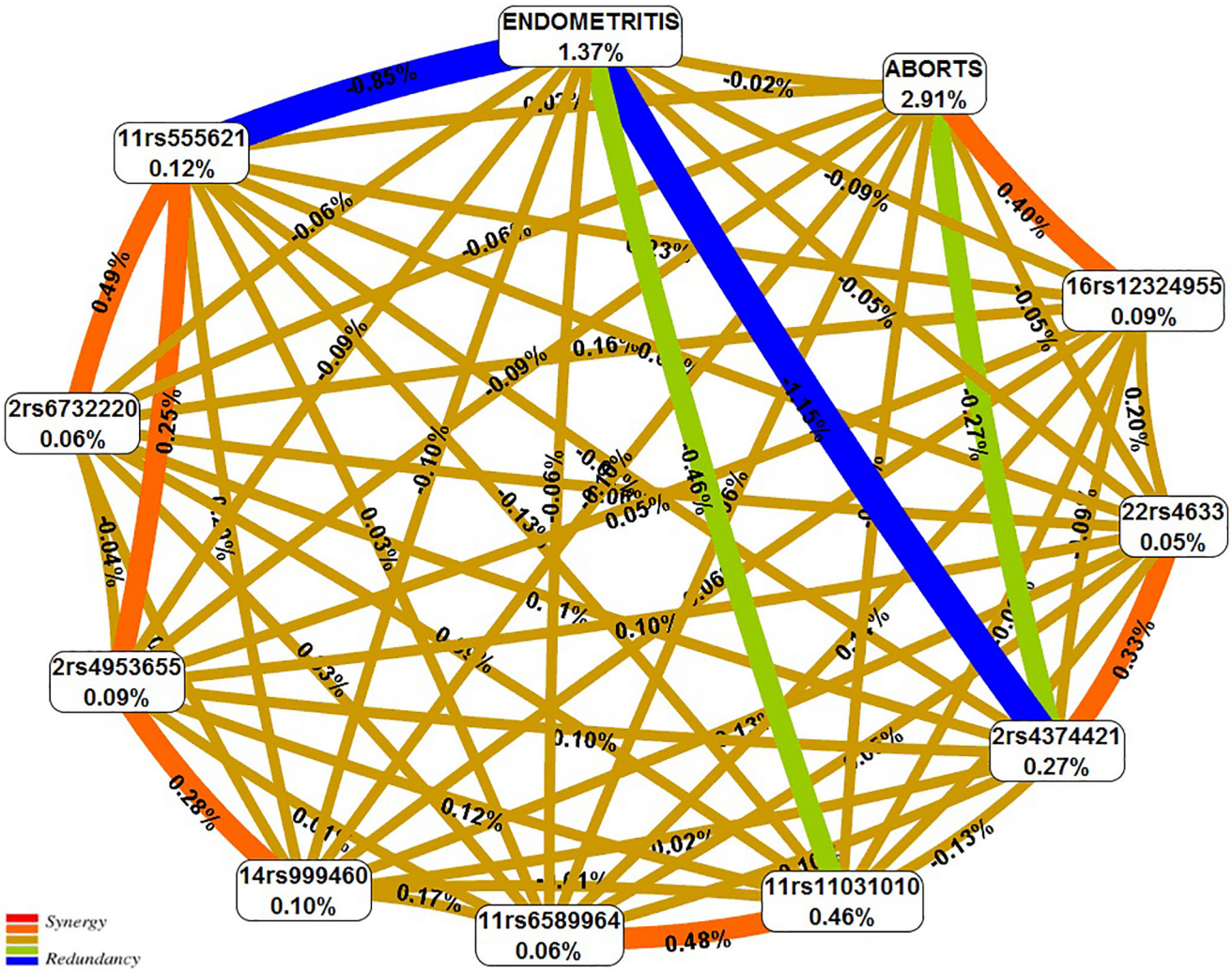

Fig. 2. The entropy graph of SNP × abortions × chronic endometritis interactions with endometrial hyperplasia based on the multifactor dimensionality reduction (MDR) analysis. Positive values of entropy indicate synergistic interactions while the negative values indicate redundancy. The polymorphisms are denoted by the chromosome number and rs SNP ID. The red and orange colors denote strong and moderate synergism, respectively, brown color denotes the independent effect, green and blue colors denote moderate and strong antagonism. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.7. Pathway analyses

The roles in biological processes or molecular functions of the EH-associated genes (Supplementary Table 13) and those whose expression is affected by the EH-associated 21 loci according to the eQTL analysis (Supplementary Tables 15–17) were analyzed *in silico* using annotations from the Gene Ontology databases. We found evidence of enrichment for pathways involved in the follicle-stimulating hormone signaling pathway (FDR = 0.001), hormone ligand-binding receptors (FDR = 0.001), ovulation cycle process (FDR = 0.005), and other pathways related to development of the female reproductive organs and hormone-mediated signaling pathways (FDR ≤ 0.05) (Supplementary Table 19).

The intergenic interactions of the above genes were analyzed using GeneMANIA (http://genemania.org). The data on three genes, *RP11-384F7, UGT2A3P7* and *RP4-710 M3* were absent in the database. Therefore, the resulting network included 22 EH-associated genes and 20 other genes significantly interacting with them (Fig. 3). The intergenic interactions are executed through physical interactions (30.89%), co-expression (25.34%), pathway (21.98%), predicted interactions (11.92%), co-localization (7.18%) and shared protein domains (2.69%) (Supplementary Table 20).

4. Discussion

For the first time, this study reports multiple associations of 21 candidate loci for age at menarche with EH. Locus rs11031010 of the *FSHB* gene demonstrated the most significant associations: its allele A decreased the risk for EH according to the dominant and additive models (OR = 0.62–0.67). This locus is also associated with EH within the most significant two-locus model intergenic interactions and six models of polymorphisms interactions with induced abortions and chronic endometritis. According to the GTExportal database, rs11031010 have the *cis*-eQTL significance and may affect expression of the *ARL14EP* gene in 10 different organs and tissues (allele A of rs11031010 increases expression of *ARL14EP*). Additionally, according to the HaploReg (v4.1), allele A of rs11031010 *FSHB* increases affinity to transcription factors Foxp3, Foxa_known2 and decreases affinity to TFs TCF4_disc2, Foxd1_1. Association of this SNP with menarcheal age was reported previously (He et al., 2010), and rs11031010 was suggested to be a risk factor of polycystic ovary syndrome and affect a level of luteinizing hormone in patients (Tian et al., 2016).

In addition to rs11031010, two other popymorphisms of *FSHB* are associated with EH: haplotype GAA of loci rs555621-rs11031010-rs1782507 is a protective factor against EH (OR = 0.66). Also rs555621

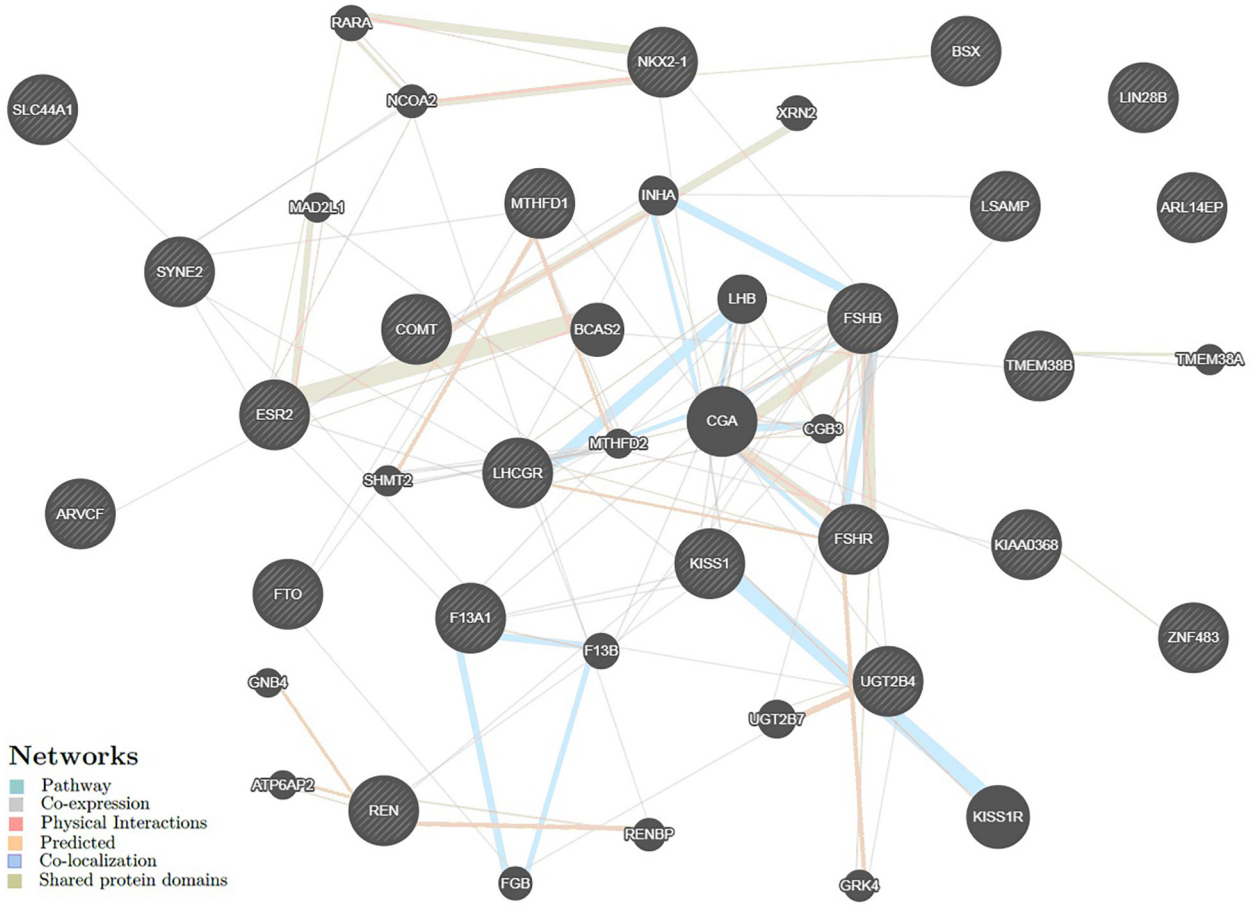

Fig. 3. The interaction networks of the candidate genes for endometrial hyperplasia inferred using GeneMANIA (http://genemania.org). The candidate genes determined in the present study are cross-shaded.

FSHB is associated with EH within the most significant three-locus model of intergenic interactions and two models of polymorphisms interactions with induced abortions and chronic endometritis. Locus rs1782507 contributed to the most significant gene-gene interaction models at all levels considered. According to the GTExportal database, allele G of rs555621 and allele A of rs1782507 *FSHB* are associated with the high expression of the *ARL14EP, RP4-710 M3.1, FSHB* genes in > 25 different organs and tissues. Also, according to the HaploReg (v4.1), allele A rs1782507 *FSHB* decreases affinity to TFs HDAC2_disc5, HNF4_disc1, Lmo2-complex_1, TCF12_disc1, ZEB1_known3. Recently we determined the association of rs555621 with the BMI of adults (Ponomarenko et al., 2019). The above two loci were also associated with age at menarche (He et al., 2010).

Importantly, alleles G rs555621 *FSHB*, A rs11031010 *FSHB*, A rs1782507 *FSHB* were previously determined as being associated with later menarche (He et al., 2010). On the other hand, the present study suggested this allele is a protective factor for EH, meaning age at menarche and a risk for EH may have a shared genetic basis.

The above loci of the *FHSB* gene not only have independent regulatory effects and eQTL value, but also linked to more than 120 SNPs, which affect expression of the *ARL14EP, FSHB, RP4-710 M3.1* genes and regulatory effects in more than 25 different organs and tissues. More importantly, the loci that are linked to the EH-associated SNPs, showed their regulatory effects and *cis*-eQTL value in organs and tissues related to pathogenesis of EH, i.e., ovaries, muscle tissue, adipose tissue, various brain regions (hypothalamus, pituitary, etc.), etc. (Sanderson et al., 2017).

According to the Ensembl database (http://www.ensembl.org/), the *FSHB* gene product is a beta-subunit of the follicle-stimulating hormone. The follicle-stimulating hormone causes proliferation of cells of the granulosa layer in follicles and the growth of follicles in the ovaries, induces the synthesis of aromatases that convert androgens into estrogens (estradiol), stimulates the synthesis of receptors for the luteinizing hormone on the granulose cells of the follicle before ovulation, etc. Gene *ARL14EP* (*ADP ribosylation factor like GTPase 14 effector protein*) encodes a protein with GTPase and ADP-ribosylating activity, which interacts with the proteins of the cell actin network (beta-actin (ACTB), myosin 1E (MYO1E)) and controls the export MHC class II molecules (http://www.genecards.org/). Information about the *RP4-710 M3.1* gene (pseudogen) in the Ensembl (http://www.ensembl.org/) and GeneCards (http://www.genecards.org/) databases is not available.

The recent meta-analysis of 11 GWAS determined the region on chromosome 11 (11p14.1, rs74485684) harboring the *FSHB* gene as that associated with endometriosis (Sapkota, et al., 2017). Importantly, rs11031010 *FSHB* analyzed in the present study, is located just 1.9 kb from the above locus. On the other hand, the first GWAS of UL in European populations (Rafnar et al., 2018) replicated three of the 19 previously reported endometriosis variants (Sapkota, et al., 2017).

There is evidence that some loci linked to rs11031010 may contribute to various reproductive characters. For example, rs11031002 was suggested to influence the concentration of the luteinizing and rs11031005 - follicle-stimulating hormone in blood plasma (Ruth et al., 2016); rs11031006 was implicated in polycystic ovary syndrome and luteinizing hormone levels (Hayes et al., 2015); rs12294104 was associated with age at menopause (Stolk et al., 2012). The SNPs upstream of the *FSHB* transcription start were implicated into a key role in many reproductive processes (Gajbhiye et al., 2018).

We did not find a direct association of age at menarche with EH, but determined that 21 candidate loci for age at menarche might be associated with EH. Among these, several were also associated with age of menarche, height and/or BMI of adults in the study sample. This may suggest that menarcheal age *per se* is not a primary risk factor for EH in the studied sample (the population of Russia) and other factors may be of greater importance. Specifically, we found that the history of induced abortions significantly increased the risk of EH in the study sample (OR = 2.35). The risk of EH directly correlates with the number of the abortions in the anamnesis (OR = 1.55): it increases from OR = 2.00 95%CI 1.46–2.73 ($p < 0.001$) in a case of two abortions in the history up to OR = 6.54 95%CI 3.51–12.35 ($p < 0.001$) with four and more abortions. Induced abortions have a high prevalence among women in Russia (and the countries of the former USSR) as a birth control method (Sedgh et al., 2007; Douglas et al., 2014; David et al., 2007). The frequency of induced abortions in Russia is the highest among all countries of Eastern Europe, and Russia (along with Cuba) is a leader by this indicator in the world; the average number of induced abortions per woman during her reproductive life is 1.3 and 1.7, respectively, while in countries with low levels of induced abortions (Belgium, Germany, Switzerland, etc.), this figure is only 0.2 (Sedgh et al., 2007). David et al. (David et al., 2007) reported even higher rates of the average number of induced abortions per woman in Russia, 2.2–2.4, with the average number of pregnancies per woman being 3.3–3.5. Induced abortions in the first trimester of pregnancy may result in hyperprolactinemia, polycystic ovary syndrome, anovulatory menstruation, disorders of the luteal phase of the menstrual cycle, post-abortion endometritis, and immunological disorders (Douglas et al., 2014).

5. Conclusions

Candidate genes for age at menarche are associated with endometrial hyperplasia. Locus rs11031010 *FSHB* is associated with EH independently; 17 loci are associated within the 12 most significant intergenic interaction models, and 9 loci may contribute to the risk of the disorder through the 14 most significant gene-environment interactions. These associations probably result from multiple effects of > 397 polymorphisms linked to the above SNPs and implicated in various metabolic processes in organs and tissues that are related to pathogenesis of endometrial hyperplasia.

Funding

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

CRediT authorship contribution statement

Irina Ponomarenko: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. **Evgeny Reshetnikov:** Methodology, Formal analysis, Supervision, Writing - review & editing. **Alexey Polonikov:** Methodology, Formal analysis, Writing - review & editing. **Inna Sorokina:** Methodology, Data curation, Formal analysis, Writing - review & editing. **Anna Yermachenko:** Methodology, Formal analysis, Writing - review & editing. **Volodymyr Dvornyk:** Data curation, Methodology, Formal analysis, Writing - review & editing, Supervision, Methodology. **Mikhail Churnosov:** Project administration, Funding acquisition, Formal analysis, Software, Visualization, Supervision, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2020.144933.

References

Aban, M., Arslan, M., Tok, E., Tekes, S., Budak, T., Altintas, A., 2006. CYP17 genetic polymorphism in patients with endometrial hyperplasia and cancer. Int. J. Gynecol. Cancer 16, 448–451.

Armstrong, A.J., Hurd, W.W., Elguero, S., Barker, N.M., Zanotti, K.M., 2012. Diagnosis and management of endometrial hyperplasia. J. Minim Invasive Gynecol. 19, 562–571. https://doi.org/10.1016/j.jmig.2012.05.009.

Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.

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

Chandra, V., Kim, J.J., Benbrook, D.M., Dwivedi, A., Rai, R., 2016. Therapeutic options for management of endometrial hyperplasia. J Gynecol Oncol. 27, e8. https://doi.org/10.3802/jgo.2016.27.e8.

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

Chen, Y.L., Wang, K.L., Chen, M.Y., Yu, M.H., Wu, C.H., Ke, Y.M., Chen, Y.J., Chang, Y.Y., Hsu, K.F., Yen, M.S., 2013. Risk factor analysis of coexisting endometrial carcinoma in patients with endometrial hyperplasia: a retrospective observational study of Taiwanese Gynecologic Oncology Group. J. Gynecol. Oncol. 24, 14–20.

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

David, P.H., Reichenbach, L., Savelieva, I., Vartapetova, N., Potemkina, R., 2007. Women's reproductive health needs in Russia: what can we learn from an intervention to improvepost–abortion care? Health Policy Plan 22, 83–94.

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., et al., 2017. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nat. Genet. 49, 834–841.

Demakova, N.A., 2018. Molecular and genetic characteristics of patients with hyperplasia and endometric polyps. Res. Result. Med. Pharmacy 2, 26–39. https://doi.org/10.18413/2313-8955-2018-4-2-0-4. (In Russian).

Dieudonné, A.S., Lambrechts, D., Smeets, D., Belmans, A., Wildiers, H., Paridaens, R., Hyonil, C., Timmerman, D., Christiaens, M.R., Vergote, I., et al., 2014. The rs1800716 variant in CYP2D6 is associated with an increased double endometrial thickness in postmenopausal women on tamoxifen. Ann. Oncol. 25, 90–95. https://doi.org/10.1093/annonc/mdt399.

Douglas, N.I., Pavlova, T.U., Burtseva, T.E., Rad, Y.G., Petrova, P.G., Odland, J.Ø., 2014. Women's reproductive health in the Sakha Republic (Yakutia). Int. J. Circumpolar Health 73. https://doi.org/10.3402/ijch.v73.25872.

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., et al., 2010. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat. Genet. 12, 1077–1085.

Emons, G., Beckmann, M., Schmidt, D., Mallmann, P., 2015. New WHO classification of endometrial hyperplasias. Geburtshilfe Frauenheilkd. 75, 135–136.

Epplein, M., Reed, S.D., Voigt, L.F., Newton, K.M., Holt, V.L., Weiss, N.S., 2008. Risk of complex and atypical endometrial hyperplasia in relation to anthropometric measures and reproductive history. Am. J. Epidemiol. 168, 563–570. https://doi.org/10.1093/aje/kwn168.

Fernandez-Rhodes, L., Demerath, E.W., Cousminer, D.L., Tao, R., Dreyfus, J.G., Esko, T., Smith, A.V., Gudnason, V., Harris, T.B., Launer, L., et al., 2013. Association of adiposity genetic variants with menarche timing in 92,105 women of European descent. Am. J. Epidemiol. 3, 451–460.

Gajbhiye, R., Fung, J.N., Montgomery, G.W., 2018. Complex genetics of female fertility. NPJ. Genom. Med. 3, 29. https://doi.org/10.1038/s41525-018-0068-1.

Goncharenko, V.M., Beniuk, V.A., Kalenska, O.V., Demchenko, O.M., Spivak, M.Y., Bubnov, R.V., 2013. Predictive diagnosis of endometrial hyperplasia and personalized therapeutic strategy in women of fertile age. EPMA J. 4, 24.

Hayes, M.G., Urbanek, M., Ehrmann, D.A., Armstrong, L.L., Lee, J.Y., Sisk, R., Karaderi, T., Barber, T.M., McCarthy, M.I., Franks, S., et al., 2015. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations [published correction appears in Nat Commun. 2016;7: 10762]. Nat. Commun. 6, 7502.

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

Ivanova, T.I., Krikunova, L.I., Ryabchenko, N.I., Mkrtchyan, L.S., Khorokhorina, V.A., Salnikova, L.E., 2015. Association of the apolipoprotein E 2 allele with concurrent occurrence of endometrial hyperplasia and endometrial carcinoma. Oxid. Med. Cell Longev 2015, 593658. https://doi.org/10.1155/2015/593658.

Kurman, R., Carcangiu, M., Herrington, C., Young, R., 2014. World Health Organisation classification of tumors of female reproductive organs, 4th edn. International Agency for Research on Cancer (IARC) Press, Lyon, France.

Lacey, J.V., Chia, V.M., Rush, B.B., Carreon, D.J., Richesson, D.A., Ioffe, O.B., Ronnett,

B.M., Chatterjee, N., Langholz, B., Sherman, M.E., et al., 2012. Incidence rates of endometrial hyperplasia, endometrial cancer and hysterectomy from 1980 to 2003 within a large prepaid health plan. Int. J. Cancer 131, 1921–1929. https://doi.org/10.1002/ijc.27457.

Lou, X.Y., Chen, G.B., Yan, L., Ma, J.Z., Zhu, J., Elston, R.C., Li, M.D., 2007. A generalized combinatorial approach for detecting gene by gene and gene by environment interactions with application to nicotine dependence. Am. J. Hum. Genet. 80, 1125–1137.

Mazur, M.T., 2005. Endometrial hyperplasia/adenocarcinoma. A conventional approach. Ann. Diagn. Pathol. 9, 174–181.

Montgomery, B.E., Daum, G.S., Dunton, C.J., 2004 May. Endometrial hyperplasia: a review. Obstet. Gynecol. Surv. 59, 368–378.

Ong, K.K., Elks, C.E., Li, S., Zhao, J.H., Luan, J., Andersen, B., Bingham, S.A., Brage, S., Smith, G.D., Ekelund, U., et al., 2009. Genetic variation in LIN28B is associated with the timing of puberty. Nat. Genet. 6, 729–733.

Ozdegirmenci, O., Kayikcioglu, F., Bozkurt, U., Akgul, M.A., Haberal, A., 2011. Comparison of the efficacy of three progestins in the treatment of simple endometrial hyperplasia without atypia. Gynecol. Obstet. Invest. 72, 10–14.

Pachomov, S.P., Altuchova, O.B., Demakova, N.A., Krivoshei, I.V., Kolesnikov, Y.V., Sobyanin, F.I., 2014. Study of cytokines polymorphous loci connections with rise of endometrium proliferative diseases. Res. J. Pharm. Biol. Chem. Sci. 6, 1473–1476.

Perry, J.R., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D., Esko, T., Thorleifsson, G., et al., 2014. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. Nature 7520, 92–97.

Ponomarenko, I.V., 2019. Using the method of Multifactor Dimensionality Reduction (MDR) and its modifications for analysis of gene-gene and gene-environment interactions in genetic-epidemiological studies (review). Res. Results Biomed. 1, 4–21. https://doi.org/10.18413/2313-8955-2019-5-1-0-1. (In Russian).

Ponomarenko, I.V., Reshetnikov, E.A., Altuchova, O.B., Polonikov, A.V., Sorokina, I.N., Yermachenko, A., 2019. Dvornyk VYa, Golovchenko OV, Churnosov MI. Association of genetic polymorphisms with age at menarche in Russian women. Gene 686, 228–236. https://doi.org/10.1016/j.gene.2018.11.042.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al., 2007. PLINK: a tool set for wholegenome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Rafnar, T., Gunnarsson, B., Stefansson, O.A., Sulem, P., Ingason, A., Frigge, M.L., Stefansdottir, L., Sigurdsson, J.K., Tragante, V., Steinthorsdottir, V., et al., 2018. Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. Nat. Commun. 9, 3636. https://doi.org/10.1038/s41467-018-05428-6.

Reed, S.D., Newton, K.M., Clinton, W.L., Epplein, M., Garcia, R., Allison, K., Voigt, L.F., Weiss, N.S., 2009. Incidence of endometrial hyperplasia. Am. J. Obstet. Gynecol. 200 (678), e1–e6. https://doi.org/10.1016/j.ajog.2009.02.032.

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., et al., 2016. Genome–wide association study with 1000 genomes imputation identifies signals for nine sex hormone–related phenotypes. Eur. J. Hum. Genet. 24, 284–290.

Sanderson, P.A., Critchley, H.O., Williams, A.R., Arends, M.J., Saunders, P.T., 2017. New concepts for an old problem: the diagnosis of endometrial hyperplasia. Hum Reprod Update. 23, 232–254. https://doi.org/10.1093/humupd/dmw042.

Sapkota, Y., Steinthorsdottir, V., Morris, A.P., Fassbender, A., Rahmioglu, N., De Vivo, I., Buring, J.E., Zhang, F., Edwards, T.L., Jones, S., et al., 2017. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. Nat Commun 8, 15539. https://doi.org/10.1038/ncomms15539.

Sedgh, G., Henshaw, S.K., Singh, S., Bankole, A., Drescher, J., 2007. Legal abortion worldwide: incidence and recent trends. Perspect. Sex Reprod. Health 39, 216–225.

Stolk, L., Perry, J.R., Chasman, D.I., He, C., Mangino, M., Sulem, P., Sulem, P., Barbalic, M., Broer, L., Byrne, E.M., et al., 2012. Meta-analyses identify 13 novel loci associated with age at menopause and highlights DNA repair and immune pathways. Nat. Genet. 44, 260–268.

Tian, Y., Zhao, H., Chen, H., Peng, Y., Cui, L., Du, Y., Wang, Z., Xu, J., Chen, Z.J., 2016. Variants in FSHB are associated with polycystic ovary syndrome and luteinizing hormone level in Han Chinese Women. J. Clin. Endocrinol. Metab. 101, 2178–2184.

Ward, L.D., Kellis, M., 2016. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 44 (D1), D877–D881.