RESEARCH ARTICLE

Identification of Novel Intronic Variants Implicating *L3MBTL4* and *AOAH* in Indonesian Ovarian Endometriosis

Trijayani[1,*], Surya Dwira[1,2], Raden Muharam[3], Rafika Indah Paramita[1,2]

[1]Master's Program in Biomedical Science, Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No.6, Jakarta 10430, Indonesia
[2]Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No.6, Jakarta 10430, Indonesia
[3]Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, Faculty of Medicine, Universitas Indonesia, Jl. Salemba Raya No.6, Jakarta 10430, Indonesia

*Corresponding author. Email: trijayani@ui.ac.id

Received date: Jun 25, 2025; Revised date: Nov 28, 2025; Accepted date: Dec 5, 2025

Abstract

BACKGROUND: Genetic factors are contributing substantially to endometriosis risk, however, there is limited data on genetic variant associations specific to ovarian endometriosis, especially in Indonesian populations. Understanding the genetic variants associated with this condition is essential for improving diagnosis and identifying potential therapeutic targets. Therefore, this study was conducted to analyze the association of genetic variants with ovarian endometriosis in Indonesian women using the single nucleotide polymorphisms (SNP)-array method, followed by an *in-silico* functional enrichment and gene expression analysis to explore their biological context.

METHODS: This case-control study utilized the Infinium Asian Screening Array microarray platform to examine 46 samples, consisting of 22 ovarian endometriosis and 24 controls. The analysis included quality control, genetic association testing, and enrichment analysis. Genotypes were analyzed under dominant and recessive inheritance models, and functional insights were explored using gene expression databases.

RESULTS: Ten intronic SNPs were significantly associated with endometrioma (p<0.05). Eight variants conferred increased risk (OR>1): rs77360595 (*IFNLR1*), rs2325558 (*KLF12*), rs1654499 (*NLRP2*), rs4809494 (*LOC105376996*), rs168482 (*TMPRSS11A*), rs17026725 (*STPG2-AS1*), rs59330070 (*GFOD1*), and rs58909364 (*AOAH*). Two variants were protective (OR<1): rs180732 (*L3MBTL4*) and rs7356507 (*CRMP1*). Notably, *AOAH* variant showed the highest odds ratio among risk variants, while *L3MBTL4* variant showed the strongest statistical significance among protective variants in the dominant model. *In-silico* expression analysis confirmed that key implicated genes (*KLF12*, *L3MBTL4*, *AOAH*, and *GFOD1*) are expressed in relevant female reproductive tissues, generally at low-to-moderate levels.

CONCLUSION: Ten novel genetic variants associated with ovarian endometriosis in Indonesian were identified. In particular, variants in *L3MBTL4* and *AOAH* represent promising candidates that may play roles in disease pathophysiology of endometriosis, suggesting the importance of population-specific genetic studies and these specific loci as potential biomarkers or therapeutic targets.

KEYWORDS: ovarian endometriosis, SNP, microarray, genetic association, plink, bioinformatics

Indones Biomed J. 2025; 17(6): 561-74

Introduction

Endometriosis is a gynecological condition marked by the growth of endometrial-like tissue outside the uterus. Globally, it affects 5–10% of women of reproductive age. In Indonesia, hospital-based studies report an incidence ranging from 13.6%–69.5% among infertile women and 6%–10% among women of reproductive age.(1) Diagnosis of endometriosis generally relies on direct visualization of

lesions through laparoscopy and histological confirmation via biopsy.(2-4) Endometriosis is typically grouped into three forms, one of which is ovarian endometrioma. Ovarian endometrioma is the most frequently diagnosed form and is characterized by ovarian cysts filled with dark brown fluid, which can vary in size.(5,6)

The pathogenesis of endometriosis is multifactorial, involving hormonal dysregulation, immune dysfunction, epigenetic modifications, and genetic predisposition. Advances in molecular technologies have enabled deeper exploration of the genetic basis of endometriosis. Polymerase chain reaction (PCR) technologies allow targeted amplification of specific genetic regions, while DNA microarrays and sequencing enable efficient, high-throughput detection of single nucleotide polymorphisms (SNPs) across the genome, supporting case-control comparisons to identify disease-associated variants. These approaches have also been widely applied in studying other complex diseases, revealing genetic variant associations with conditions such as heart disease, age-related disorders, and various cancers.(7-9)

Genetic studies on endometriosis date back to the 1980s, with heritability estimates ranging from 47% to 52%. Since 2010, genome-wide association studies (GWAS) have identified over 600 SNPs significantly associated with endometriosis, implicating genes involved in hormone metabolism, immune response, and inflammatory pathways. Among the most robust and frequently reported of these are variants near the *WNT4* gene, which are hypothesized to contribute to endometriosis by disrupting developmental pathways of the Müllerian duct. Other key loci include variants near *GREB1*, an estrogen-responsive gene, and *IL1A*, which implicates pro-inflammatory pathways in the disease's etiology. These findings confirm that the genetic basis of endometriosis involves complex interplay between hormonal, developmental, and immune/inflammatory pathways.(10-12)

As the vast majority of genetic studies have been conducted in populations of European and East Asian ancestry; however, genetically distinct populations, such as in Indonesia, remain significantly underrepresented in endometriosis research. To help address this gap, the present study employed the Infinium Asian Screening Array platform which specifically designed for Asian populations, as it combines highly optimized multi-ethnic genome-wide content with curated clinical research variants and a discovery panel of markers derived from East and Southeast Asian populations. The use of this targeted platform provides a greater likelihood of identifying population-specific variants relevant to Indonesia population. This study is therefore among the first to leverage this technology to specifically investigate genetic variants associated with ovarian endometriosis in an Indonesian cohort, aiming to identify variants that have not been previously examined in this population.(13)

Identifying associated genetic variants is the first step; understanding their functional consequence is also critical. Large-scale GWAS have consistently shown that while some disease-associated SNPs are in protein-coding (exonic) regions, a vast majority are located in non-coding regions of the genome, such as introns or intergenic areas. While variants in coding regions can directly alter protein structure, these more common non-coding variants are hypothesized to exert their effects by regulating gene expression, for example by altering transcription factor binding sites. Bridging the gap between a statistical association and its biological function, however, can be challenging. To provide initial context and help prioritize variants for future validation, in-silico tools have become invaluable. Publicly available gene expression databases, for instance, can be leveraged to explore whether genes implicated by nearby SNPs are expressed in disease-relevant tissues. Furthermore, functional enrichment analysis can help reveal shared molecular pathways, offering clues to the biological plausibility of the genetic findings.(14)

Associations from endometriosis studies are often reported for the disease as a whole, without distinguishing between subtypes, despite known phenotypic heterogeneity, yet ovarian endometrioma is a distinct phenotype that may involve unique genetic mechanisms. For example, the unique cystic microenvironment within the ovary may alter local tissue biology and promote genomic instability, contributing to disease progression and possibly increasing the risk of malignant transformation. Despite these unique characteristics, specific genetic variants and mechanisms driving endometrioma development remain poorly understood.(15,16) In this study, an analysis of genetic variants in endometriosis with endometrioma phenotypes will be carried out using a DNA microarray approach, followed by an *in-silico* functional enrichment and gene expression analysis to explore their biological context.

Methods

Study Population

A case-control study was conducted on 22 cases of ovarian endometriosis (endometrioma) and 24 controls.

For this genetic association study, samples were collected from peripheral blood to analyze germline DNA. The case group consisted of female patients with ovarian endometriosis who had previously undergone laparoscopic or laparotomy surgery and/ received a confirmed diagnosis of ovarian endometriosis (endometrioma) based on surgical visualization and/or histopathological evidence. The control group consisted of healthy women with no history of ovarian endometriosis (endometrioma) or symptoms implicated in endometrioma, such as dysmenorrhea, dysuria, dyspareunia, or fertility problems. Exclusion criteria for both groups included the presence of other complex diseases, such as known malignancies, autoimmune disorders, or cardiovascular disease. The collection of blood samples of both case and control participants was carried out for 6 months from July 2023 to December 2023 at Dr. Cipto Mangunkusumo Hospital, Jakarta, Indonesia by obtaining informed consent from all research subjects. This study has received approval to pass ethical review from the Health Research Ethics Commission of FKUI-RSCM (No. KET-930/UN2.F1/ETIK/PPM.00.02/2023).

DNA Extraction and Quantification

The collected blood samples were then subjected to DNA extraction, DNA concentration quantification, and DNA quality control checks. DNA concentration was calculated based on the fluorescent intensity value of the specific fluorescent dye that binds to the DNA double chain. DNA quality control checks are calculated based on the ratio of absorbance to wavelengths of 260 nm and 280 nm as well as at wavelengths of 260 nm and 230 nm, where good DNA purity is in the range A260/A280 1.8 - 2.0 and A260/A230 2.0 - 2.2.

DNA Microarray

DNA samples were first normalized to a concentration of 50 ng/µL before undergoing microarray processing using the Infinium HTS Assay and Infinium Asian Screening Array (ASA) - 24 v1.0 BeadChip (Illumina, San Diego, CA, USA). This high-density array interrogates a total of 659,184 markers, providing broad genome-wide coverage as well as variants with known disease associations. Marker selection was based on curated data from authoritative institutions and databases, including ClinVar (>7,000 disease and trait associations), the NHGRI-EBI GWAS Catalog (~12,000 variants), and the ExAC database for diverse exonic content. Additionally, the array covers pharmacogenomic variants associated with absorption, distribution, metabolism, and excretion (ADME) phenotypes based on PharmGKB and Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. The microarray protocol was completed over three days. During the first day, DNA underwent amplification followed by roughly 20 hours of incubation. On the second day, the procedure included enzymatic fragmentation, precipitation with alcohol, resuspension of DNA, and loading onto the BeadChip for hybridization, followed by an additional 16-hour incubation. On the third day, enzymatic extension and fluorescent staining were performed, and fluorescence signal intensities were read using the Illumina iScan system.

Quality Control

The quality control test of SNP-Array (microarray) work was carried out by calculating the quality value based on several Genome Studio software cluster algorithm parameters with predetermined limitation values. Some of the recommended parameters are call rate/call frequency, GenTrain Score, Cluster Separation Score, and AA/AB/BB R Mean. An important step that should be part of any research related to genetic association studies is proper quality control tests, because without extensive controls, the study results cannot provide good and reliable results because by nature, raw genotype data are imperfect. Errors in data analyzed without passing quality control tests can arise due to poor sample quality, poor signal intensity, genotyping probes that do not work optimally, contaminated samples and so on. In this study, 7 stages of quality control protocol was implemented, adapted from established guidelines for genetic association analysis using PLINK.(17)

Bioinformatic Analysis

Genetic association analysis was conducted using PLINK through a logistic regression approach to evaluate the association between SNPs and endometriosis risk in case and control groups. The additive model (1 degree of freedom) was applied, assessing the linear effect of risk alleles (minor *vs.* major) on disease status.(17) Genotype distribution was further analyzed using MedCalc under dominant and recessive models to calculate odds ratios. (18) To validate the expression of candidate genes in relevant human tissues, RNA and protein expression data were retrieved from The Human Protein Atlas (HPA). (19) Furthermore, to investigate potential functional interactions among the implicated genes, a Protein-Protein Interaction (PPI) network analysis was constructed using the STRING database, with a confidence threshold of 0.400. (20) Enrichment and functional annotation analysis were performed for variants with p-values ranging from 1×10^{-4}

to 5×10^{-8} using FUMA.(21) SNPs were mapped to genes using the SNP2GENE tool, followed by tissue expression analysis via MAGMA. Genes linked to significant GWAS loci were analyzed using GENE2FUNC for functional enrichment, excluding major histocompatibility complex (MHC) regions. Additional enrichment analysis was carried out using GARFIELD, applying the significance thresholds of $p<1\times10^{-5}$ and $p<1\times10^{-8}$ to assess regulatory and functional variant enrichment. Pathway and disease associations were explored using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.(22) Finally, variant pathogenicity was interpreted following ACMG guidelines using the Franklin by Genoox tool for clinical classification.(23)

Results

Subjects' Characteristics

A total of 22 cases and 24 controls were included in this study. A significant difference in age was observed between the two groups. The average age of the case group in the study was 35±7.17 years with an age range of 27–47 years, while the control group was 46±7.37 years with an age range of 33–8 years. Due to this age discrepancy, age was included as a covariate in all subsequent genetic association analyses to adjust the results.

DNA Quantity and Quality

Collected blood samples underwent DNA extraction, followed by quantification and quality control. DNA purity was assessed using absorbance ratios at 260/280 nm and 260/230 nm, with optimal values ranging from 1.8–2.0 and 2.0–2.2, respectively. The target DNA concentration was >50 ng/µL, meeting the minimum requirement for Illumina-based genotyping, as specified in the manufacturer's guidelines.(23) All extracted samples met these quality and concentration standards.

Quality Control Test of Microarray

Microarray quality control was performed by evaluating key statistical parameters in accordance with Illumina guidelines, including call rate (call frequency), GenTrain Score, Cluster Separation Score, and AA/AB/BB R Mean. (24) In this study, all quality metrics met recommended thresholds, indicating that the microarray analysis produced high-quality data suitable for further genetic association analysis. Detailed QC results were presented in Table 1.

PLINK Quality Control Test

The results of the quality control test in this study have successfully filtered 396,032 probes from a total of 659,184 probes on the microarray chip to 263,152 probes for further analysis. Besides eliminating low-quality array probes, the PLINK quality control steps also filtered SNPs with a Minor Allele Frequency (MAF) of less than 0.05 (5%). This procedure is standard practice and removes non-polymorphic (monomorphic) or very rare variants that are uninformative for a statistical association study. Given that the Infinium Asian Screening Array is a multi-ethnic platform designed for many Asian populations, it is an expected outcome that a large portion of its SNPs will be uninformative in our specific Indonesian cohort. In addition, the quality test has also filtered out samples that were judged to have a deviation value>3 due to extreme heterozygosity rates. Details of the quality control test results could be seen in Table 2.

Genetic Variant Association Analysis

Genetic variant association analysis identified 15,905 SNPs significantly associated with ovarian endometriosis ($p<0.05$) out of 263,152 SNPs that passed quality control. The odds ratio (OR) for each SNP was used to assess its association with disease risk: OR=1 indicates no association, OR>1 suggests increased risk, and OR<1 indicates a potential protective effect. The association results were visualized using a Manhattan plot, which displays the $-\log_{10}$ p-values of SNPs across chromosomal positions (Figure 1). The most statistically significant variant was rs17026725, located on chromosome 4, with a p-value of 1.24×10^{-5}. In terms of effect size, rs58909364 showed the highest OR of 25.31 ($p=4.34\times10^{-5}$), indicating a strong association with increased risk of endometrioma. Conversely, rs7356507 had an OR of 0.04 ($p=3.58\times10^{-5}$), suggesting a potential

Table 1. Microarray quality control test results.

Quality Score	Quality Control Parameters					
	Call Rate	GenTrain	Cluster Separation	AA R Mean	AB R Mean	BB R Mean
Mean	0.9935	0.8299	0.8780	10.700	12.600	10.200
Std Dev	0.0010	0.0828	0.1777	0.4027	0.474	0.311

Table 2. GWAS quality control test results based on the number of array probes.

Quality Control Stages	Number of Probes After Quality Control	Number of Samples After Quality Control
Eliminates low-quality array probes	511,365	46
Missingness of SNPs and individuals	489,198	46
Gender check	489,198	46
Minor allele frequency (MAF)	263,152	46
Hardy–Weinberg equilibrium (HWE)	263,152	46
Heterozygosity	263,152	44
Relatedness	263,152	44

protective effect. Detailed genetic variants associated results were summarized in Table 3.

Further association analysis was conducted on the top 10 SNPs by comparing genotype distributions between case and control groups (Table 4). All 10 variants showed statistically significant associations (p<0.05) under at least one genetic model; dominant, recessive, or both. Among the top 10 SNPs, eight were identified as risk variants: rs77360595 (*IFNLR1*), rs2325558 (*KLF12*), rs1654499 (*NLRP2*), rs4809494 (*LOC105376996*), rs168482 (*TMPRSS11A*), rs17026725 (*STPG2-AS1*), rs59330070 (*GFOD1*), and rs58909364 (*AOAH*). Two variants were found to be protective: rs180732 (*L3MBTL4*) and rs7356507 (*CRMP1*). Notably, rs180732 had the highest statistical significance (OR=0.05, p=0.0001) in the dominant model, suggesting that individuals carrying the A allele may have reduced risk for ovarian endometriosis compared to those with the G allele. Additionally, rs2325558 and rs4809494 showed significance in both dominant and recessive models, indicating that two risk alleles (A for rs2325558 and G for rs4809494) are required to increase disease risk under the recessive model.

Enrichment Analysis

Annotation of the top 10 SNPs using the dbSNP database revealed that all variants were intronic and classified as benign according to the American College of Medical Genetics and Genomics (ACMG) guidelines. Using the Franklin by Genoox tool, all 10 variants were classified as benign based on two key criteria: "Benign Stand-Alone (BA1)" criterion and "Benign Supporting (BP7)" criterion. To explore the biological plausibility of the associated variants, we used The Human Protein Atlas to examine the expression of the implicated genes. The results showed that there were four genes that were expressed in female tissues such as the endometrium and/or ovaries including the genes *KLF12*, *L3MBTL4*, *AOAH*, and *GFOD1*. As shown in Figure 2, these genes were also expressed in a wide variety of human tissues. However, this analysis confirmed that all four genes are also expressed in female reproductive tissues relevant to endometriosis. This supports their potential role in the disease's pathophysiology. Notably, Notably, *KLF12* showed medium expression in endometrial stromal cells; *L3MBTL4* showed low expression in endometrial glandular cells; and *AOAH* showed medium expression in endometrial glandular cells. *GFOD1* showed high expression in endometrial glandular cells and medium expression in endometrial stromal cells.

Protein–protein interaction analysis (Figure 3) revealed that *KLF12* interacts with nine proteins, mainly transcription factors and enzymes. *AOAH* interacts with *DLG2*, *L3MBTL4* with *NINL*, and *GFOD1* with 17 proteins,

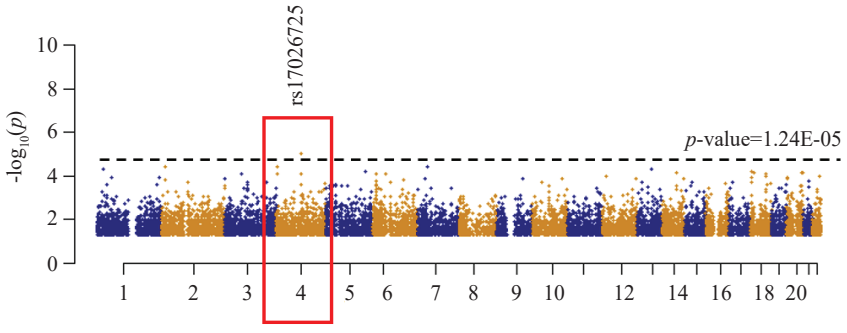

Figure 1. Manhattan plot of SNPs across chromosomal positions.

Table 3. Association results of genetic variants associated with ovarian endometriosis.

rsID	Chr	Position (hg38)	Allele 1	Allele 2	Allele Freq 1	OR	*p*- value
rs77360595	1	24184085	A	C	0.063	11.09	5.34E-05
rs2325558	13	73764078	G	A	0.292	6.40	5.14E-05
rs1654499	19	54993598	C	T	0.021	20.14	2.38E-04
rs180732	18	6357139	A	G	0.500	0.11	6.04E-05
rs4809494	20	63129732	G	A	0.167	6.77	6.53E-05
rs168482	4	67943249	C	T	0.021	15.67	1.21E-03
rs17026725	**4**	**97412767**	**C**	**T**	**0.083**	**11.00**	**1.24E-05***
rs7356507	4	5859179	G	A	0.396	0.04	3.58E-05
rs59330070	6	13479701	A	C	0.021	17.83	5.41E-04
rs58909364	7	36683419	C	T	0.021	25.31	4.34E-05

Allele 1 is the minor allele, while Allele 2 is the major allele. The frequency of Allele 1 was calculated based on the control group. Results were considered significant if the *p*-value<0.05. *The variant with the highest significance value. Abbreviations: Chr, Chromosome; hg38, Human Genome Assembly GRCh38; Allele Freq 1, Frequency of Allele 1; OR, Odds Ratio.

including enzymes and transporters. These findings suggest potential regulatory roles of these genes in ovarian endometriosis pathophysiology.

To visualize the biological context of the associated variants, a gene expression heatmap was generated display consensus expression levels (Figure 4). Our analysis focused on the expression that implicated by our significant SNPs across various female reproductive tissues, such as the ovary, cervix, uterus, and vagina. It is a key limitation that a specific dataset for endometrium was not available at the time of analysis. However, given that the endometrium is the inner mucosal lining of the uterus, the uterus dataset serves as the most relevant available biological proxy for this analysis. Furthermore, expression in the ovary is also highly relevant, as this is the primary anatomical site of the ovarian endometrioma phenotype studied. It is noteworthy that while the overall heatmap shows some implicated genes with high (red) expression in these reproductive tissues, the key genes that were the focus of our study generally showed low to moderate expression levels.

A gene-set enrichment analysis was performed to tests for overrepresentation in numerous gene sets, including many pathways already known to be correlated with the pathophysiology of endometriosis, such as the Wingless-related int-1 (WNT) signaling pathway, mitogen-activated protein kinase (MAPK) signaling pathway, Cell Adhesion Molecules (CAMs), and extracellular matrix (ECM)-receptor interaction. While our analysis identified these and other relevant pathways for testing (48 in total), after applying corrections for multiple comparisons, our specific list of implicated genes (including *KLF12*, *L3MBTL4*, *AOAH*, and *GFOD1*) did not show a statistically significant enrichment (adj-*p*<0.05) in any of these pathways (Figure 5). This important negative finding suggests that the genetic variants identified in our study may influence ovarian endometriosis risk through novel mechanisms that are distinct from these well-established pathways.

Following the identification of these novel variants, a comparative analysis was performed to determine whether established endometriosis risk loci identified in previous global studies were also present in the current dataset. This involved cross-referencing the variant-associated genes with previously reported endometriosis-associated genes in the NHGRI-EBI GWAS catalog. A total of 13 genes known to be linked to endometriosis were filtered, including *RND3*, *KDR*, *RNF144B*, *ESR1*, *SYNE1*, *NFE2L3*, *PDE1C*, *CSMD1*, *IL33*, *CDKN2B-AS1*, *SORCS1*, *VEZT*, and *KSR2* (Table 5). Among them, rs3020314 (*ESR1*) and rs2813503 (*SYNE1*) had the highest significance values (1.19E-03 and 8.29E-04) and OR>1, indicating a possible increased risk of ovarian endometriosis in individuals carrying these variants.

Discussion

This study provides the first genetic association analysis of ovarian endometriosis specifically within an Indonesian cohort. Our primary finding is the identification of ten novel SNPs significantly associated with disease susceptibility (*p*<0.05). Most notably, we identified a potent risk variant in the AOAH gene, which exhibited the highest effect size in our cohort (OR=25.31), and a highly significant protective variant in the *L3MBTL4* gene (*p*=0.0001). These findings, supported by *in-silico* expression analysis confirming

Table 4. Genetic association results based on top 10 SNP genotype distribution.

rsID	Gene	Genotype	Control (n)	Case (n)	*p*-value	Odd Ratio (95% CI)
rs77360595	*IFNLR1*	CC	21	6	**0.0004***	**16.33 (3.494 to 76.350)**
		AA+AC	3	14		
		AC+CC	24	17	0.1393	9.80 (0.475 to 202.038)
		AA	0	3		
rs2325558	*KLF12*	GG	12	1	**0.0077***	**19.00 (2.182 to 165.456)**
		AA+AG	12	19		
		AG+GG	22	10	**0.0055***	**11.00 (2.025 to 59.749)**
		AA	2	10		
rs1654499	*NLRP2*	TT	23	11	**0.0085***	**18.82 (2.112 to 167.707)**
		TC+CC	1	9		
		TT+TC	24	17	0.1393	9.80 (0.475 to 202.038)
		CC	0	3		
rs180732	*L3MBTL4*	GG	4	16	**0.0001***	**0.05 (0.011 to 0.232)**
		AA+AG	20	4		
		AG+GG	20	20	0.1491	0.11 (0.006 to 2.199)
		AA	4	0		
rs4809494	*LOC105376996*	AA	17	3	**0.0007***	**13.76 (3.039 to 62.321)**
		AG+GG	7	17		
		AA+AG	23	14	**0.0433***	**9.86 (1.072 to 90.654)**
		GG	1	6		
rs168482	*TMPRSS11A*	TT	23	10	**0.0049***	**23.00 (2.585 to 204.612)**
		TC+CC	1	10		
		TT+TC	24	20	0.9298	1.20 (0.023 to 62.923)
		CC	0	0		
rs17026725	*STPG2-AS1*	TT	20	5	**0.0003***	**15.00 (3.430 to 65.594)**
		TC+CC	4	15		
		TT+TC	24	15	0.059	17.39 (0.8972to 336.939)
		CC	0	5		
rs7356507	*CRMP1*	AA	10	19	**0.0030***	**0.038 (0.004 to 0.329)**
		AG+GG	14	1		
		AA+AG	19	20	0.1051	0.087 (0.005 to 1.670)
		GG	5	0		
rs59330070	*GFOD1*	CC	23	11	**0.0085***	**18.82 (2.112 to 167.707)**
		AA+AC	1	9		
		AC+CC	24	18	0.2314	6.622 (0.300 to 146.378)
		AA	0	2		
rs58909364	*AOAH*	TT	23	9	**0.0028***	**28.11 (3.154 to 250.526)**
		TC+CC	1	11		
		TT+TC	24	17	0.1393	9.80 (0.475 to 202.038)
		CC	0	3		

their presence in relevant reproductive tissues, suggest that the genetic architecture of ovarian endometriosis in the Indonesian population involves unique loci distinct from those typically reported in European or East Asian populations.

This current study employed a two-step analysis to first identify and then characterize genetic variants. The initial allelic association test (Table 3) successfully identified 10 novel SNPs significantly associated with ovarian endometrioma. To better understand the mode of inheritance for these 10 variants, we then performed a secondary genotype analysis using dominant and recessive models (Table 4). This second analysis provided deeper insights into how these variants might influence disease risk. The two-step approach allowed us to not only discover novel loci but also to build a more complete picture of their potential inheritance patterns. In total, 10 top SNPs were prioritized for further analysis: rs58909364 (*AOAH*), rs77360595 (*IFNLR1*), rs2325558 (*KLF12*), rs1654499 (*NLRP2*), rs4809494 (*LOC105376996*), rs168482 (*TMPRSS11A*), rs17026725 (*STPG2-AS1*), rs59330070 (*GFOD1*), rs180732 (*L3MBTL4*), and rs7356507 (*CRMP1*).

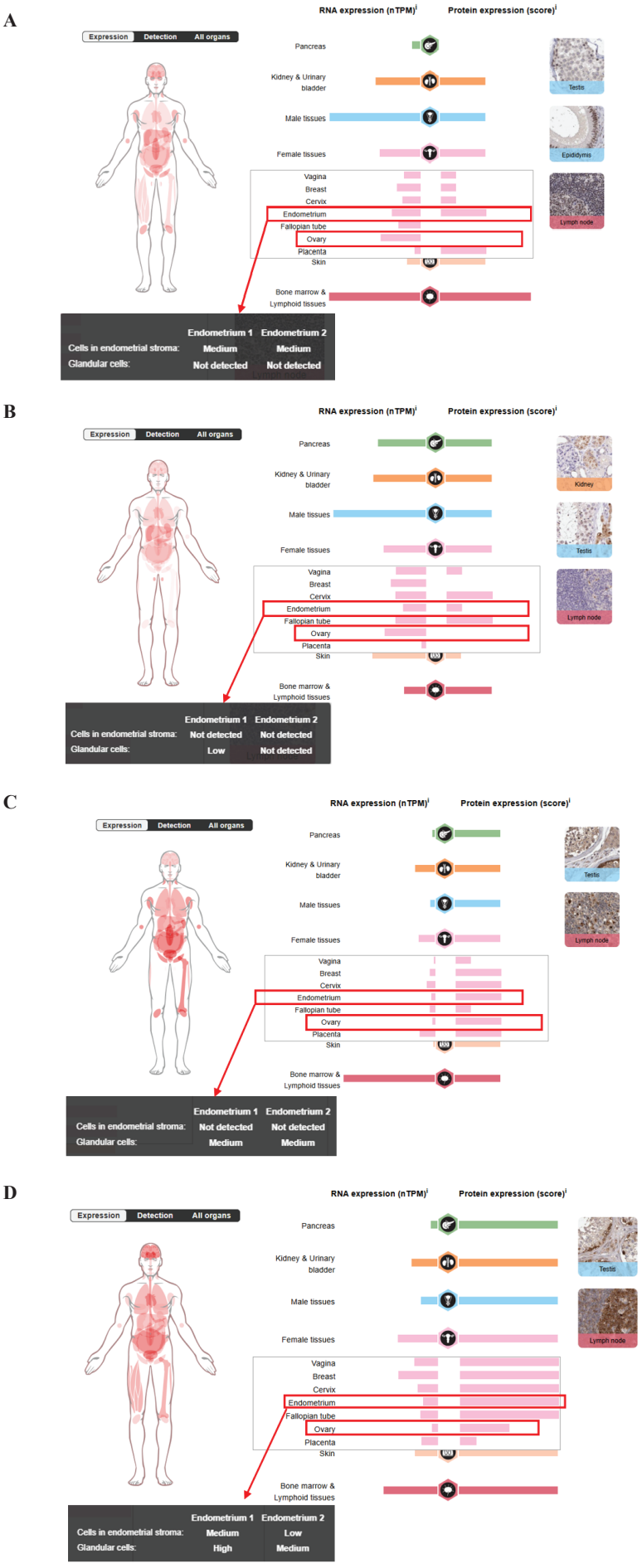

Figure 2. Gene expression analysis results for (A) *KLF12*, (B) *L3MBTL4*, (C) *AOAH*, and (D) *GFOD1* in female tissues, particularly the endometrium and ovaries. Data and visualizations were derived from The Human Protein Atlas. The bar charts display consensus RNA expression levels (nTPM), with pink bars indicating expression in female tissues and red boxes highlighting the endometrium and ovary. The inset tables detail specific expression levels (Low/Medium/High) within endometrial stromal and glandular cells.

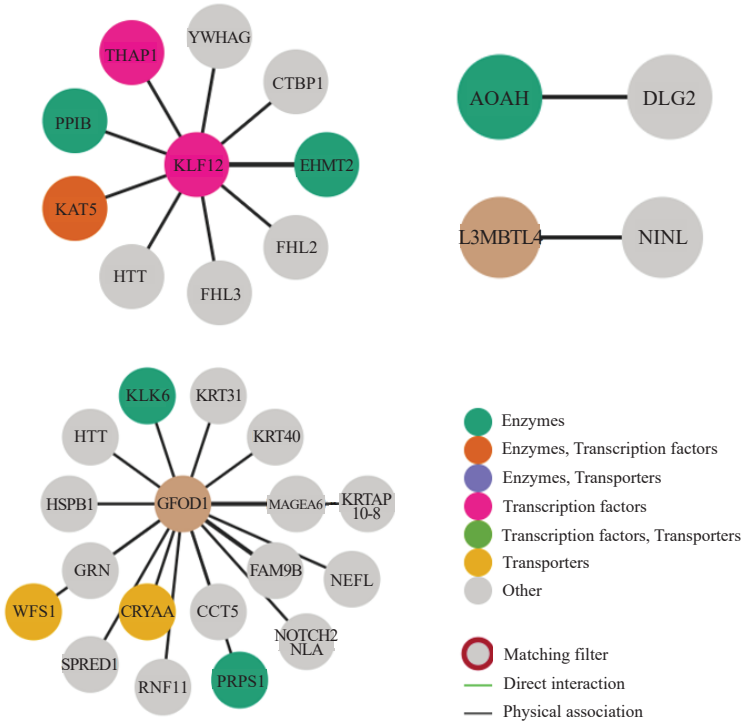

Figure 3. Results of protein–protein interaction analysis for the *KLF12*, *L3MBTL4*, *AOAH*, and *GFOD1* genes.

All variants are located in intronic regions and likely function as cis-regulatory elements (*e.g.*, enhancers or silencers) that modulate the quantitative expression of the gene. This mechanistic hypothesis aligns with our *in-silico* expression analysis (Figure 2), which confirms that these genes are actively transcribed in endometrial and ovarian tissues. Consequently, the pathogenicity of these variants is likely driven by dysregulated gene dosage, either an overexpression or underexpression of the normal protein, rather than the production of a structurally defective protein. (25,26) All top variants were further examined using gene expression analysis via the Human Protein Atlas. Of the 10 prioritized genes, *KLF12, L3MBTL4, AOAH*, and *GFOD1* were confirmed to be expressed in female reproductive tissues, particularly the endometrium and ovaries. These genes were also analyzed using PPI networks to understand their interaction with other functional proteins, strengthening their biological plausibility in endometriosis pathogenesis.

KLF12 gene belongs to the Krüppel-like factor (KLF) family of transcriptional regulators and essential for proliferation, differentiation, inflammation, and apoptosis. Consistent with our PPI network analysis, which linked *KLF12* to cellular proliferation pathways, this protein acts as a transcriptional repressor involved in cell cycle regulation. In the context of endometrial physiology, *KLF12* appears to function as a critical negative regulator of decidualization. Recent studies indicate that its overexpression can repress the secretion of key biomarkers, such as Prolactin (PRL) and insulin-like growth factor-binding protein 1 (IGFBP-1), by binding to GC/CACCC-rich promoter sequences. Therefore, we hypothesize that the risk variants identified in our study may lead to dysregulated *KLF12* activity. This dysfunction could contribute to the 'progesterone resistance' and defective decidualization often observed in endometriosis, potentially impairing endometrial receptivity and contributing to clinical issues such as recurrent implantation failure.(27,28)

The rs58909364 in the AOAH gene had the highest OR (25.31; p=4.34×10^{-5}), indicating a strong risk association. *AOAH* encodes acyloxyacyl hydrolase, an enzyme that deactivates lipopolysaccharides (LPS) from gram-negative bacteria, preventing excessive immune activation. *AOAH* is expressed in immune-related tissues such as the spleen and bone marrow, and notably in endometrial glandular cells and ovarian follicular cells. This expression pattern supports a potential role in modulating local immune responses in reproductive tissues. PPI analysis revealed that *AOAH* interacts with *DLG2*, a protein that regulates N-methyl-D-aspartate (NMDA) receptor function and has been linked to GnRH signaling in the hypothalamus, implicating hormonal pathways in the gene's broader function. Given emerging evidence linking dysbiosis and bacterial LPS with endometriosis, *AOAH* may contribute to disease risk by modulating immune responses in the reproductive tract.(29-31)

In addition to the *AOAH* gene, the rs59330070 variant in *GFOD1* also shows high levels of gene expression in

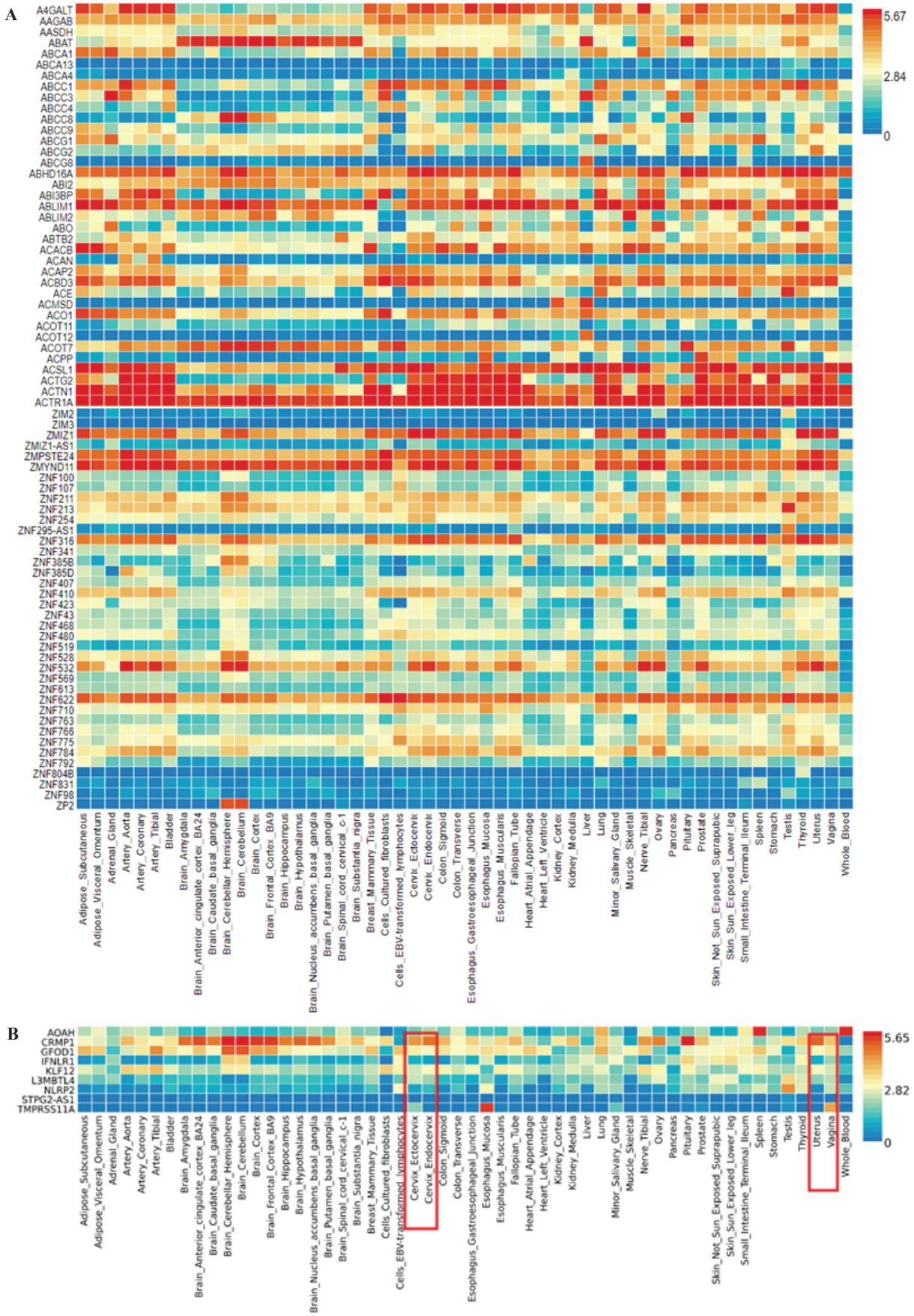

Figure 4. Gene expression heatmap of implicated genes. The color scale indicates normalized expression levels: high (red), medium (white), and low (blue). A: Overview of all key implicated genes. B: Detailed view of the four genes discussed in the text: *KLF12*, *L3MBTL4*, *AOAH*, and *GFOD1*.

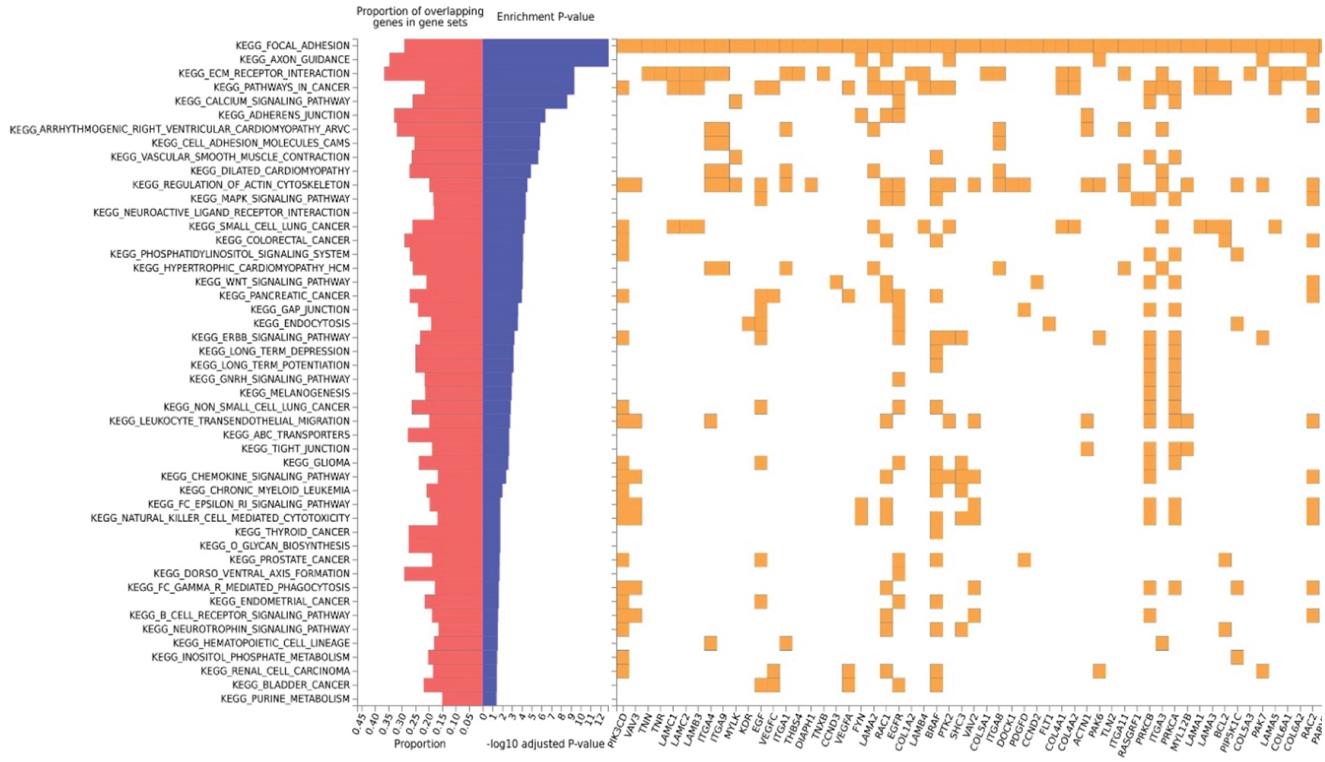

Figure 5. FUMA gene-set enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. This chart shows the results of the functional enrichment analysis for KEGG pathways. The blue bars indicate the statistical significance (FDR-adjusted *p*-value, -$\log_{10}$). The red bars represent the proportion of input genes overlapping with the pathway. The orange squares identify the specific input genes that are part of the corresponding enriched pathway.

glandular cells and medium levels in stromal cells in the endometrium, as well as medium levels of expression in ovarian follicular cells. *GFOD1* is a gene that encodes an enzyme glutamine: fructose-6-phosphate amidotransferase (GFAT), which plays a role in the biosynthetic pathway of glucosaminoglycans (GAGs) and is a large group of polysaccharide molecules found in the ECM as well as various tissues in the body. GAGs are known to have various structural and regulatory functions in various biological processes. The role of the *GFOD1* gene in endometriosis

Tabel 5. Association results of 13 genes and common genetic variants linked to endometriosis based on GWAS data.

rsID	Location (hg38)	Allele 1	Allele 2	Allele Freq 1	*p*-value	OR	Gene	Variant Type
rs1528429	150480328	A	G	0.2292	4.95E-02	2.49	*RND3*	Intron
rs4576072*	55120071	C	T	0.1458	4.96E-02	0.15	*KDR*	Intron
rs9396865*	18426344	C	T	0.0208	2.58E-02	8.29	*RNF144B*	Intron
rs3020314*	**151949537**	**T**	**C**	**0.1042**	**1.19E-03**	**5.73**	***ESR1***	**Intron**
rs2813503*	**152174676**	**A**	**G**	**0.1667**	**8.29E-04**	**5.00**	***SYNE1***	**Intron**
rs3753098	26158420	C	T	0.1458	4.58E-02	2.82	*NFE2L3*	Intron
rs13241693	32221481	T	C	0.1875	4.92E-02	2.60	*PDE1C*	Intron
rs1217542	4497229	A	C	0.2083	4.97E-02	2.53	*CSMD1*	Intron
rs1891385	6219845	C	A	0.2708	4.33E-02	0.30	*IL33*	Intron
rs7341786	22112242	A	C	0.4583	4.31E-02	0.39	*CDKN2B-AS1*	Intron
rs4918280	107031000	T	C	0.3333	4.80E-02	0.35	*SORCS1*	Intron
rs2367305	95272322	A	G	0.0625	1.48E-03	7.22	*VEZT*	Intron
rs74432428	117629480	A	G	0.08333	8.69E-03	4.71	*KSR2*	Intron

Allele 1 represents the minor allele, while Allele 2 is the major allele. Allele 1 frequency is calculated based on the control group. Results are considered significant if $p<0.05$. *Indicates the variant with the highest significance.

disease has not been widely studied, but it is likely that this gene plays a role in cellular functions such as cell proliferation, growth, migration, and cell differentiation. Based on the results of interactions between proteins, the *GFOD1* gene has interactions with as many as 17 other proteins, one of which is the Kallikrein-related peptidase 6 (KLK6) protein. KLK6 protein is a serine protease protein that is known to be involved in several types of malignancies. This protein is a pro-enzyme in various tissues in adults and is activated by cleavage by other proteases which are then secreted into biological fluids. In its mature form, it plays a role in degrading the extracellular matrix and basement membrane of tissues, so overexpression of this protein is often associated with malignancies, such as ovarian cancer. Although there is no definitive evidence regarding how *GFOD1* and *KLK6* interact, based on the functions and roles of the genes above, it is possible that both variants also play a role in the pathophysiological mechanism of endometriosis.(32,33)

A significant finding of this study was the identification of rs180732 in the *L3MBTL4* gene as a novel protective variant (OR=0.05). This variant showed the highest statistical significance in our dominant model analysis. To our knowledge, *L3MBTL4* has not been previously implicated in endometriosis risk. The gene itself functions as a Polycomb group (PcG) protein, acting as a transcriptional repressor by binding to methylated histones and promoting chromatin compaction. While *L3MBTL4* is a novel gene in the context of endometriosis, intriguing correlations exist. For example, previous studies have found that *L3MBTL4* expression was downregulated in control groups but upregulated in patients with Polycystic Ovary Syndrome (PCOS), a hormonal disorder that shares some risk factors and potential disease pathways with endometriosis. This suggests a potential role for *L3MBTL4* in female reproductive health and pathology. (34) Furthermore, a plausible mechanistic correlation exists through the gene's known interaction with the MAPK signaling pathway. Previous studies on other diseases have demonstrated that *L3MBTL4* can modulate the activation of the MAPK pathway, which is a well-established pathway involved in the inflammation, proliferation, and cell survival of endometriotic lesions. It is important to note that the MAPK signaling pathway did not appear as a statistically significant finding in our own KEGG gene-set enrichment analysis (Figure 6). This is not necessarily contradictory. The current KEGG analysis tested for enrichment across our entire list of associated genes and may not have the statistical power to detect a single, novel gene-pathway interaction. The hypothesized link between *L3MBTL4* and MAPK signaling is a specific, protein-level mechanism that may not be captured by a gene-set enrichment test.(35)

A critical outcome of our enrichment analysis was the comparison between our identified variants and those previously reported in large-scale GWAS. Notably, our analysis revealed a distinct genetic profile; the genes identified in this Indonesian cohort do not overlap with the well-established endometriosis loci found in the GWAS Catalog, such as *WNT4, GREB1*, or *ID4*. Furthermore, while our *in-silico* analysis flagged pathways like *WNT* and *MAPK* for testing, our specific gene list did not show statistically significant enrichment in these classic endometriosis pathways.

This divergence is likely attributable to two main factors. First, the vast majority of existing GWAS data is derived from European and East Asian populations. The genetic architecture of the Indonesian population involves unique linkage disequilibrium (LD) blocks and allele frequencies, which can mask global risk variants or reveal novel, population-specific ones. Second, our study specifically targeted the ovarian endometrioma phenotype, whereas many cataloged GWAS studies aggregate all endometriosis subtypes. Therefore, the lack of overlap in our enrichment analysis highlights the unique genetic etiology of ovarian endometriosis in Indonesian women, suggesting that disease susceptibility in this population may be driven by novel mechanisms rather than the standard pathways observed in other populations.(36)

Therefore, the lack of overlap in our enrichment analysis highlights the unique genetic etiology of ovarian endometriosis in Indonesian women, suggesting that disease susceptibility in this population may be driven by novel mechanisms rather than the standard pathways observed in other populations. However, this distinction does not imply a complete divergence from established disease biology. When we expanded our scope to examine the broader list of 2,497 genes identified in our enrichment analysis, we recovered significant associations with several well-known genes frequently reported in global GWAS studies. These include *ESR1* (rs3020314), *SYNE1* (rs2367305), *IL33*, *VEZT*, and *KSR2*. Among these, the strongest significance was observed for rs3020314 (*ESR1*) and rs2367305 (*SYNE1*), confirming that our Indonesian cohort still shares fundamental genetic risk factors with other populations, even if the primary drivers differ.

Both *ESR1* and *SYNE1* were previously identified in a large-scale meta-analysis on genetic and inflammatory comorbidities in endometriosis. *SYNE1* encodes nesprin-1, a nuclear envelope protein involved in nuclear structure,

function, and signaling. It is expressed in epithelial cells and has been associated with endometriosis symptoms such as dysmenorrhea and dyspareunia. Alongside *ESR1*, *SYNE1* also plays a role in hormone signaling.(37)

The rs3020314 variant is located in the intronic region of the *ESR1* gene, which encodes a receptor activated by 17β-estradiol (E2), a key hormone in reproductive function and endometriosis lesion development. E2 is produced locally in endometriotic tissue, promoting chronic activation of estrogen receptors and inflammation. There are two isoforms: ERα (*ESR1*) and ERβ (*ESR2*). In healthy endometrium, ERα expression dominates, while in ectopic lesions, ERβ is overexpressed. This abnormal ERβ/ERα ratio suppresses ERα and is linked to progesterone resistance, chronic inflammation, and pelvic pain. Notably, *ESR1* mRNA and ERα protein expression are reduced in ovarian endometrioma lesions compared to eutopic endometrial tissue, while ERβ expression is elevated, supporting its role in endometriosis pathogenesis.(38-40)

Despite our novel findings, this study has several important limitations that must be acknowledged. This study was limited by the content of the SNP array. While the Infinium Asian Screening Array is optimized for Asian populations, it does not provide the same resolution as whole-genome sequencing (WGS). It is possible that the intronic variants we identified are not the true causal variants themselves, but are instead in high LD with a different, un-genotyped causal variant. The gene expression and pathway analyses from platforms like FUMA provide biological plausibility but do not constitute functional proof. Future *in-vitro* functional studies are required to determine if these intronic variants do, in fact, alter the expression of their implicated genes.

Conclusion

This study identified ten novel genetic variants associated with ovarian endometriosis in an Indonesian cohort. Among the significant findings, the intronic variant *AOAH* rs58909364 conferred the highest risk magnitude, while the variant in *L3MBTL4* rs180732 exhibited the strongest protective association. Functional annotation and *in-silico* analysis indicate that these loci are involved in immune-inflammatory regulation and cell cycle control, providing a biological basis for their involvement in the genetic susceptibility to ovarian endometrioma. Furthermore, expression analysis confirmed that these key candidate genes are actively transcribed in the endometrium and ovary.

Acknowledgments

This research was supported by the *Publikasi Terindeks Internasional* (PUTI) Pascasarjana Q3 Grant from Universitas Indonesia. We gratefully acknowledge the funding assistance provided under the PUTI Grant scheme for supporting this study and its publication.

Authors Contribution

T was responsible for data and sample collection, methodology design, and data analysis and interpretation. T, SD, and RIP contributed to drafting, reviewing, and editing the manuscript. RM provided supervision throughout the study. All authors contributed to the critical revision and final approval of the manuscript.

Conflict of Interest

The authors declare no conflicts of interest or competing interests related to the content of this manuscript.

References

1. Mukti P. Faktor Risiko Kejadian Endometriosis. Unnes J Public Health. 2014; 3(3): 1–10. doi: 10.15294/ujph.v3i3.3542.
2. Kloeve-Mogensen K, Rohde PD, Twisttmann S, Nygaard M, Koldby KM, Steffensen R, *et al.* Polygenic risk score prediction for endometriosis. Front Reprod Health. 2021; 3: 793226. doi: 10.3389/frph.2021.793226.
3. Lalami I, Abo C, Borghese B, Chapron C, Vaiman D. Genomics of endometriosis: From genome wide association studies to exome sequencing. Int J Mol Sci. 2021; 22(14): 7297. doi: 10.3390/ijms22147297.
4. Sari V, Jenie RI, Widad S, Dewanto A. Mmp-9 and timp-1 promote extracellular matrix remodeling in the formation of ovarian endometrioma: In vitro study on chicken chorioallantoic membrane. Indones Biomed J. 2023; 15(1): 69–76.
5. Febriyeni F, Hestiantoro A, Tulandi T, Muharam M, Asmarinah, A, Sandra F. Mrna expression and dna methylation of cxcl16 in menstrual blood and endometrium tissue of subjects with endometriosis and pelvic pain. Indones Biomed J. 2024; 16(2): 180–8.
6. Dun EC, Taylor RN, Wieser F. Advances in the genetics of endometriosis. Genome Med. 2010; 2(10): 75. doi: 10.1186/gm196.
7. Gresham D, Dunham MJ, Botstein D. Comparing whole genomes using DNA microarrays. Nat Rev Genet. 2008; 9(4): 291302.
8. Tella S, Chukkayapalli SG, Akka J, Uppala S. Association between maternal fut2 204a>g (Rs492602) genetic polymorphism and congenital heart disease in the indian population: A study in maternal-fetal dyads. Mol Cell Biomed Sci. 2023; 7(2): 65–9.

9. Ramadhani N, Soeroso NN, Tarigan SP, Eyanoer PC, Hidayat H. Correlation between genetic polymorphism of cyp2a13 genotype and lung cancer in female passive smokers. Mol Cell Biomed Sci. 2022; 6(2): 89–95.
10. GWAS Catalog [Internet]. The NHGRI-EBI Catalog of human genome-wide association studies: "endometriosis" [cited 2024 Jan 4]. Available from: https://www.ebi.ac.uk/gwas/search?query=endometriosis
11. Giacomini E, Minetto S, Li Piani L, Pagliardini L, Somigliana E, Viganò P. Genetics and inflammation in endometriosis: improving knowledge for development of new pharmacological strategies. Int J Mol Sci. 2021; 22(16): 9033. doi: 10.3390/ijms22169033.
12. Zondervan KT, Rahmioglu N, Morris AP, Nyholt DR, Montgomery GW, Becker CM, *et al.* Beyond endometriosis GWAS: From genomics to phenomics to the patient. Semin Reprod Med. 2016; 34(4): 242–54.
13. Illimina [Internet]. Infinium Asian Screening Array-24 Kit [cited 2025 Nov 3]. Available from: https://www.illumina.com/products/by-type/microarray-kits/infinium-asian-screening.html.
14. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017; 8(1)1: 1826. doi: 10.1038/s41467-017-01261-5.
15. Borghese B, Tost J, de Surville M, Busato F, Letourneur F, Mondon F, *et al.* Identification of susceptibility genes for peritoneal, ovarian, and deep infiltrating endometriosis using a pooled sample-based genome-wide association study. BioMed Res Int. 2015; 2015: 461024. doi: 10.1155/2015/461024.
16. Sanchez AM, Viganò P, Somigliana E, Panina-Bordignon P, Vercellini P, Candiani M. The distinguishing cellular and molecular features of the endometriotic ovarian cyst: From pathophysiology to the potential endometrioma-mediated damage to the ovary. Hum Reprod Update. 2014; 20(2): 217–30.
17. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018; 27(2): e1608. doi: 10.1002/mpr.1608.
18. MedCalc [Internet] Odds Ratio - Free MedCalc Online Statistical Calculator [cited 2025 Nov 28]. Available from: https://www.medcalc.org/en/calc/odds_ratio.php.
19. The Human Protein Atlas [Internet]. The open access resource for human proteins [cited 2025 Nov 28]. Available from: https://www.proteinatlas.org/.
20. STRING: Functional Protein Association Networks [Internet]. Protein-Protein Interaction Networks [cited 2025 Nov 28]. Available from: https://string-db.org/.
21. FUMA GWAS [Internet]. FUMA GWAS: Functional Mapping and Annotation of Genome-Wide Association Studies [cited 2024 Jan 5]. Available from: https://fuma.ctglab.nl/.
22. Franklin by Qiagen [Internet]. The Future of Genomic Medicine [cited 2025 Nov 3]. Available from: https://franklin.genoox.com/clinical-db/home.
23. Illumina. Infinium HTS Assay Reference Guide. San Diego: Illumina; 2019.
24. Illumina. 2014. Technical Note: Genotyping. Infinium Genotyping Data Analysis. San Diego: Illumina; 2019.
25. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017; 136(9): 1093–111.
26. Jo BS, Choi SS. Introns: The functional benefits of introns in genomes. Genomics Inform. 2015; 13(4): 112–8.
27. Shen X, Hu Y, Jiang Y, Liu H, Zhu L, Jin X, *et al.* Krüppel-like factor 12 negatively regulates human endometrial stromal cell decidualization. Biochem Biophys Res Commun. 2013; 433(1): 11–7.
28. Huang C, Sun H, Wang Z, Liu Y, Cheng X, Liu J, *et al*. Increased Krüppel-like factor 12 impairs embryo attachment via downregulation of leukemia inhibitory factor in women with recurrent implantation failure. Cell Death Discov. 2018; 4: 23. doi: 10.1038/s41420-018-0088-8.
29. ScienceDirect Topics [Internet]. Ovarian Follicle [cited 2024 Jan 19]. Available from: https://www.sciencedirect.com/topics/medicine-and-dentistry/ovarian-follicle.
30. Zizolfi B, Foreste V, Gallo A, Martone S, Giampaolino P, Di Spiezio Sardo A. Endometriosis and dysbiosis: State of art. Front Endocrinol. 2023; 14: 1140774. doi: 10.3389/fendo.2023.1140774.
31. Jee YH, Won S, Lui JC, Jennings M, Whalen P, Yue S, *et al.* DLG2 variants in patients with pubertal disorders. Genet Med Off J Am Coll Med Genet. 2020; 22(8): 1329–37.
32. NIH NCBI [Internet]. GFOD1 Gfo/Idh/MocA-like oxidoreductase domain containing 1 [Homo sapiens (human)] [cited 2024 Jan 19]. Available from: https://www.ncbi.nlm.nih.gov/gene/54438.
33. Tamir A, Jag U, Sarojini S, Schindewolf C, Tanaka T, Gharbaran R, *et al.* Kallikrein family proteases KLK6 and KLK7 are potential early detection and diagnostic biomarkers for serous and papillary serous ovarian cancer subtypes. J Ovarian Res. 2014; 7: 109. doi: 10.1186/s13048-014-0109-z.
34. Liu X, Hu C, Bao M, Li J, Liu X, Tan X, *et al.* Genome wide association study identifies L3MBTL4 as a novel susceptibility gene for hypertension. Sci Rep. 2016; 6: 30811. doi: 10.1038/srep30811.
35. Nahdi S, Arafah M, Harrath AH. Integrated bioinformatics and experimental analysis revealed crosstalk between IL-6, autophagy, ubiquitination, and key miRNAs in female infertility: Insights from ovarian endometriosis and polycystic ovary syndrome. Cells. 2025; 14(21): 1693. doi: 10.3390/cells14211693.
36. Koyama S, Liu X, Koike Y, Hikino K, Koido M, Li W, *et al*. Population-specific putative causal variants shape quantitative traits. Nat Genet. 2024; 56(10): 2027–35.
37. Rahmioglu N, Mortlock S, Ghiasi M, Møller PL, Stefansdottir L, Galarneau G, *et al.* The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. Nat Genet. 2023; 55(3): 423–36.
38. Chantalat E, Valera MC, Vaysse C, Noirrit E, Rusidze M, Weyl A, e*t al.* Estrogen receptors and endometriosis. Int J Mol Sci. 2020; 21(8): 2815. doi: 10.3390/ijms21082815.
39. Wang W, Li Y, Maitituoheti M, Yang R, Wu Z, Wang T, *et al.* Association of an oestrogen receptor gene polymorphism in Chinese Han women with endometriosis and endometriosis-related infertility. Reprod Biomed Online. 2013; 26(1): 93–8.
40. Yuane E, Dewanto A, Widad S. High expression of pr-a and low expression of pr-b is correlated with inflammation in endometrioma cases. Indones Biomed J. 2023; 15(1): 85–93.