December 2025; 16(2): 147-161
Jurnal Kesehatan Reproduksi
DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)
https://journaliakmitangsel2.iakmi.or.id/index.php/kespro

UNCOVERING PATHOGENIC MISSENSE VARIANTS IN ENDOMETRIOSIS USING A GENOME-WIDE ASSOCIATION STUDY

Ichtiarini Nurullita Santri[1], Wirawan Adikusuma[2], Petrina Theda Philothra[3], Nurul Fadhliati Maulida[4], Rockie Chong[5], Ilker Ates[6], Yohane Vincent Abero Phiri[7], Lalu Muhammad Irham[8,9*]

[1] Faculty of Public Health, Unversitas Ahmad Dahlan, Yogyakarta, Indonesia
[2] Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong, Indonesia
[3] Department of Rehabilitation Medicine, General Hospital Yogyakarta City, Yogyakarta, Indonesia
[4] Residents of Obstetrics and Gynaecology, Faculty of Medicine, Syiah Kuala University, Banda Aceh, Indonesia
[5] Department of Chemistry and Biochemistry, University of California, Los Angeles, USA
[6] Department of Pharmaceutical Toxicology, Faculty of Pharmacy, Ankara University, Ankara, Turkey
[7] School of Public Health, Department of Epidemiology and Biostatistics, University of Nevada, Las Vegas, USA
[8] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[9] Faculty of Pharmacy, Silpakorn University, Thailand

Email: lalu.irham@pharm.uad.ac.id
* Corresponding author

Article Info

Keywords:
Endometriosis;
GWAS;
Pathogenic variants;
Protein Prediction;
Genetic Pathway Enrichment;

Abstract

Background: Endometriosis is a complex gynecological disorder with a strong genetic component. Although genome-wide association studies (GWAS) have identified numerous risk loci, the functional interpretation of protein-altering missense variants remains limited. **Objective:** This study identified pathogenic missense variants linked to endometriosis risk using publicly available GWAS data and explored implications for genetic risk detection, particularly in underrepresented populations such as Indonesia. **Methods:** Endometriosis-associated missense single nucleotide polymorphisms (SNPs) were identified from GWAS data, and a total of eight missense SNPs were analyzed. Functional effects were evaluated in silico using PolyPhen-2 and SIFT. Allele frequency distributions were assessed across global populations, and pathway enrichment analysis was conducted using the Reactome database. **Results:** Several missense variants were significantly associated with increased endometriosis risk (e.g., rs75801644, OR = 3.88; rs144824657, OR = 3.52), while rs2341097 showed a potential protective effect. Functional prediction prioritized variants in genes such as KCNG2 and BSG as potentially damaging. Population analyses revealed marked allele frequency differences, and enriched pathways were related to potassium channel activity, metabolism, extracellular matrix organization, and signal transduction. **Conclusion:** This study identifies missense variants contributing to endometriosis susceptibility and provides insight into biological pathways. Further experimental validation and clinical studies are warranted.

Received 13/11/2025 Revised 18/11/2025 Accepted 03/03/2026

How to Cite:

Santri IN, Adikusuma W, Philothra PT, Maulida NF, Chong R, Ates I, Phiri YVA, Irham LM. UNCOVERING PATHOGENIC MISSENSE VARIANTS IN ENDOMETRIOSIS USING A GENOME-WIDE ASSOCIATION STUDY. J. Kesehat. Reproduksi [Internet]. 2025 Dec. 16(2):147-161. Available from: https://doi.org/10.58185/jkr.v16i2.407

INTRODUCTION

Endometriosis is a prevalent, estrogen-dependent gynecological disorder characterized by the presence of endometrial-like tissue outside the uterine cavity, causing chronic pelvic pain, infertility, and reduced quality of life.[1, 2] Although environmental and hormonal factors contribute to disease development, genetic predisposition plays a significant role in susceptibility and may account for approximately 50% of risk variability.[1] Understanding the genetic basis of endometriosis is important for the development of improved diagnostic tools and targeted therapies.[3]

Genome-wide association studies (GWAS) have transformed the study of complex disease genomics by enabling the identification of common genetic variants associated with a wide range of traits, including endometriosis. Recent GWAS meta-analyses involving diverse populations have identified numerous risk loci across nearly all autosomal chromosomes, with chromosomes 1, 6, and 11 harboring variants showing highly significant associations with endometriosis susceptibility.[4,5] These loci often include genes involved in hormone metabolism, inflammatory signaling, and cellular adhesion, which are all relevant to endometriosis pathophysiology.[6] However, a major challenge is that most associated variants are located in non-coding regions, such as introns, intergenic regions, and regulatory elements, making functional interpretation difficult.[7-9]

Non-coding variants are likely to exert their effects by modulating gene expression rather than altering protein structure, thereby influencing when, where, and how much a gene is expressed.[10] These regulatory regions act as control nodes that orchestrate complex transcriptional programs important for cellular functions, including the maintenance and invasiveness of ectopic endometrial tissue. Nevertheless, some exonic variants with protein-altering consequences, including missense or synonymous changes, have been identified and may have direct functional effects relevant to disease pathogenesis.[11] For example, somatic mutations in the KRAS oncogene identified in endometriotic lesions underscore the potential contribution of coding variants to disease progression.[10,12]

This study leverages publicly available GWAS datasets to specifically identify pathogenic variants that result in protein changes, as these are more likely to directly affect the molecular mechanisms underlying endometriosis.[13] By prioritizing protein-altering variants through

comprehensive bioinformatics analyses, this study aims to identify potential causal mutations that may disrupt endometrial cell function or immune modulation.[14] This strategy complements regulatory variant analysis by focusing on variants with a higher likelihood of direct functional impact.

Identification of such pathogenic variants is important for advancing precision medicine in endometriosis, including the development of molecularly targeted therapies and personalized risk prediction models.[15] Previous studies have reported somatic mutations, such as those in KRAS, in ectopic endometrial tissue, highlighting the relevance of protein-coding changes in endometriosis pathogenesis.[10,12] By elucidating inherited variants with similar protein-altering effects, this study seeks to address gaps in the current understanding of the genetic contributors to endometriosis.[15] Combining GWAS data with protein function prediction may provide a promising approach for dissecting the genetic complexity of endometriosis. This investigation may also have translational value for improving disease management and therapeutic development.

Despite the growing number of GWAS identifying loci associated with endometriosis, the functional characterization of protein-altering missense variants remains relatively underexplored. This knowledge gap limits the translation of genetic associations into mechanistic understanding and clinical application. Therefore, prioritizing pathogenic missense variants is an important step toward elucidating molecular mechanisms underlying endometriosis and advancing precision medicine approaches.

Accordingly, this study aims to identify and prioritize pathogenic missense variants associated with endometriosis using GWAS data. By integrating protein function prediction, population allele frequency analysis, and pathway enrichment analysis, this study seeks to identify biologically relevant variants that may contribute to endometriosis susceptibility. The findings are expected to provide a foundation for future genetic risk assessment and molecular-based early detection strategies, particularly in underrepresented populations such as Indonesia.

METHODS

Retrieval and Annotation of Endometriosis-associated SNPs

In this study, the primary goal was to identify pathogenic genetic variants implicated in endometriosis that alter protein sequences, with a specific focus on missense mutations that cause amino acid substitutions with potential functional consequences. Endometriosis-associated SNPs were retrieved from the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/) using the keyword "*Endometriosis*".[16] The database search, conducted on November 12, 2025, yielded a total of 729 reported variants and corresponding risk alleles associated with endometriosis.

To enhance biological interpretability, the initial dataset was systematically refined by restricting the analysis to exonic missense variants, defined as single nucleotide substitutions within protein-coding regions that result in amino acid changes. These variants were prioritized because missense mutations are more likely to directly affect protein structure, stability, or function than non-coding variants. Pathogenic missense variants were further prioritized because they are more likely to exert functional effects relevant to disease susceptibility and progression. This targeted filtering strategy enabled the identification of protein-altering variants with increased potential to contribute to the molecular mechanisms underlying endometriosis.[17]

Bioinformatic Prediction of Missense Variant Effects

Endometriosis-associated SNPs obtained from the GWAS Catalog were functionally annotated using SNPnexus (https://www.snp-nexus.org/), which integrates multiple predictive algorithms, including PolyPhen-2 [18] and SIFT,[19] to evaluate the potential effects of genetic variants on protein function.[20] PolyPhen-2 (Polymorphism Phenotyping v2) estimates the potential impact of amino acid substitutions on protein structure and function by combining sequence- and structure-based features and classifies variants as *benign, possibly damaging,* or *probably damaging*.[18, 20, 21] SIFT (Sorting Intolerant From Tolerant) evaluates the functional consequences of substitutions based on evolutionary conservation and amino acid properties and categorizes variants as *tolerated* or *deleterious*.[19,20] The combined use of these tools enables a more comprehensive assessment of the potential pathogenicity of missense variants and provides insight into their possible biological relevance to endometriosis.

Population Distribution of SNP Alleles Based on 1000 Genomes Data

Population allele frequencies of the selected SNPs were analyzed using the SNPnexus platform, incorporating data from the 1000 Genomes Project. This database classifies individuals into five major ancestral groups: African, American, European, East Asian, and South Asian populations, and provides comprehensive information on human genetic variation. Through this analysis, allele frequencies were estimated for each continental population, enabling a broader understanding of global genetic diversity.[22, 23]

Pathway Enrichment Analysis

Pathway enrichment analysis and SNP annotation were performed using SNPnexus (https://www.snp-nexus.org/). This platform functionally annotates submitted variants by integrating information from multiple biological databases. Each SNP was mapped to its corresponding genes and biological pathways based on the Reactome database.[20,24] Enrichment analysis was then performed to identify pathways significantly associated with genes containing the analyzed variants, with statistical significance determined using the hypergeometric test to calculate p-values.

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

Ethical Considerations

This study used secondary data obtained from publicly available databases, including the NHGRI-EBI GWAS Catalog, SNPnexus, the 1000 Genomes Project, and Reactome. All data were anonymized and did not involve direct interaction with human participants or access to individual-level identifiable information. Therefore, approval from an institutional review board or ethics committee was not required.

RESULTS

Genetic Associations of Missense Variants with Endometriosis Risk Revealed by GWAS

This analysis identified eight missense variants significantly associated with endometriosis risk, with odds ratios ranging from a potential protective effect to more than a threefold increase in risk, highlighting variants with potential functional relevance. This study aimed to identify genetic variants, particularly pathogenic missense mutations, associated with increased endometriosis risk using a GWAS approach. Missense variants are mutations in which a single nucleotide change results in a different amino acid in the encoded protein, which may alter protein function and influence disease risk. **Table 1** presents several single-nucleotide polymorphisms (SNPs) identified as missense variants associated with endometriosis risk. Each SNP is listed with its p-value and odds ratio (OR), indicating the strength and significance of its association with the disease. Variants such as rs75801644 and rs144824657 have very low p-values and high odds ratios, suggesting strong associations with increased endometriosis susceptibility. In contrast, some variants show more modest risk increases, while rs2341097 appears to have a protective effect, with an OR substantially below 1.

Table 1 highlights the genetic complexity underlying endometriosis by showing both high-risk and potentially protective missense variants. Odds ratios above 1 for most SNPs indicate increased odds of disease, with some variants associated with more than a threefold increase in risk. The p-values indicate that these associations are unlikely to be due to chance. These results provide evidence for the involvement of specific protein-altering variants in endometriosis and support further studies to investigate their biological effects and potential clinical relevance.

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

Table 1. Missense Variants of Endometriosis-Associated Risk SNPs

SNPs	*p*-value	Odds ratio
rs116250606	1×10^{-7}	3.24
rs3820445	6×10^{-7}	2.04
rs75801644	8×10^{-10}	3.88
rs144824657	1×10^{-6}	3.52
rs2276314	7×10^{-6}	1.31
rs75339189	3×10^{-6}	1.1
rs2341097	1×10^{-5}	0.05
rs2278868	5×10^{-7}	1.14

Polyphen-2 Analysis of Endometriosis-Associated Missense Variants and Predicted Effects on Protein Function

The section evaluates how different missense variants may affect the function of proteins produced by genes linked to endometriosis (**Table 2**). Each row in the table lists a variant identifier (SNP), its chromosome position, the specific nucleotide alteration, the related gene, the amino acid position impacted, the Polyphen-2 score, and the predicted effect on protein function. Polyphen-2 scores near zero suggest that the variant is likely harmless and probably does not significantly alter protein function, while higher scores imply a possible detrimental effect. Most of the variants, including those in CALHM3, C18orf21, and GPNMB genes, are predicted to be benign based on low scores, indicating these mutations are unlikely to cause major changes in the protein's structure or function.

However, there are a few variants with higher Polyphen-2 scores, such as rs75339189 in the KCNG2 gene and one of the changes involving rs144824657 in the BSG gene, categorized as "possibly damaging," indicating these mutations could potentially impair protein function and contribute more directly to endometriosis pathogenesis. This prediction is critical because damaging missense variants may alter protein stability, binding, or activity, thereby influencing disease mechanisms. Overall, **Table 2** provides valuable insights by combining genetic association data with functional predictions, helping prioritize variants for further experimental validation in understanding endometriosis at the molecular level.

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

Table 2. PolyPhen-2 Predictions for Endometriosis-Associated Missense Variants

Variation ID	Chromosome	Position	Variant	Gene	AA Position	Score	Prediction
rs116250606	chr10	103473496	C/T	*CALHM3*	251	0.006	Benign
rs2276314	chr18	35977503	A/T	*C18orf21*	44	0.085	Benign
rs2276314	chr18	35977503	A/G	*C18orf21*	44	0.001	Benign
rs75339189	chr18	79899553	C/A	*KCNG2*	380	0.673	Possibly Damaging
rs144824657	chr19	577782	G/A	*BSG*	26	0.001	Benign
rs144824657	chr19	577782	G/T	*BSG*	26	0.593	Possibly Damaging
rs2341097	chr19	45765644	C/T	*SIX5*	112	0.317	Benign
rs2341097	chr19	45765644	C/T	*SIX5*	693	0.169	Benign
rs75801644	chr7	23266522	G/A	*GPNMB*	354	0.003	Benign
rs75801644	chr7	23266522	G/A	*GPNMB*	342	0.015	Benign
rs75801644	chr7	23266522	G/A	*GPNMB*	370	0.013	Benign

SIFT Analysis of Endometriosis-Associated Missense Variants and Predicted Effects on Protein Function

The SIFT analysis presented in **Table 3** evaluates the functional impact of missense variants associated with endometriosis by predicting whether changes in amino acid sequences will affect protein function. Each entry in the table includes information such as the variation ID (SNP), chromosome location, nucleotide change, gene name, amino acid position, SIFT score, and the predicted effect on the protein. The SIFT score ranges from 0 to 1, where scores closer to 1 indicate that the variant is likely "tolerated," meaning it probably does not have a significant impact on protein function. Conversely, scores near 0 suggest a "deleterious" effect, implying that the variant may impair protein activity. This functional prediction is essential for understanding how genetic variations contribute to disease mechanisms at the molecular level.

Most variants listed in **Table 3** are predicted to be tolerated, suggesting these mutations are unlikely to significantly disrupt protein function. For example, missense variants in genes such as CALHM3, SKAP1, C18orf21, and GPNMB have high SIFT scores (close to or equal to 1), indicating that the amino acid substitutions they cause are probably benign. This tolerance implies that these proteins can maintain their normal structure and function despite the genetic changes, which means these variants may only play minor roles, if any, in the pathogenesis of endometriosis. Identifying tolerated variants helps filter out less likely candidates for causal involvement in the disease.

In contrast, several variants show SIFT scores approaching zero, classifying them as deleterious. Notably, variants in genes like BSG (rs144824657) and SIX5 (rs2341097) receive very low scores, indicating potentially harmful effects on protein function. These deleterious predictions highlight variants that may disrupt critical protein activities such as binding, signaling, or structural

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

stability, thereby contributing more directly to the development and progression of endometriosis. Pinpointing these potentially damaging variants is crucial for prioritizing variants for further experimental validation, functional studies, and potential therapeutic targeting. Overall, the SIFT analysis in **Table 3** complements other protein function prediction tools to deepen our understanding of the molecular basis of endometriosis.

Table 3. SIFT Predictions for Endometriosis-Associated Missense Variants

Variation ID	Chromosome	Position	Variant	Gene	AA Position	Score	Prediction
rs116250606	chr10	103473496	C/T	*CALHM3*	251	1	Tolerated
rs2278868	chr17	48184809	C/T	*SKAP1*	161	1	Tolerated
rs2276314	chr18	35977503	A/G	*C18orf21*	44	0.81	Tolerated
rs2276314	chr18	35977503	A/T	*C18orf21*	44	0.98	Tolerated
rs75339189	chr18	79899553	C/A	*KCNG2*	380	0.07	Tolerated
rs144824657	chr19	577782	G/A	*BSG*	26	0.89	Tolerated
rs144824657	chr19	577782	G/T	*BSG*	26	0	Deleterious
rs2341097	chr19	45765644	C/T	*SIX5*	693	0.01	Deleterious
rs2341097	chr19	45765644	C/T	*SIX5*	693	0.01	Deleterious
rs75801644	chr7	23266522	G/A	*GPNMB*	354	0.6	Tolerated
rs75801644	chr7	23266522	G/A	*GPNMB*	342	0.76	Tolerated
rs75801644	chr7	23266522	G/A	*GPNMB*	370	0.43	Tolerated

Allele Frequencies of Endometriosis-associated SNPs Across Global Populations

Table 4 provides valuable insight into the genetic variability of endometriosis risk variants among different ethnic groups around the world. This table lists the reference (REF) and alternative (ALT) alleles for several SNPs identified as associated with endometriosis, along with their allele frequencies in diverse populations, including African, American, East Asian, European, and South Asian populations. The frequency data indicate how common or rare these variants are in each population, highlighting important differences that can influence genetic susceptibility to the disease across ethnicities. For example, the variant rs2278868 shows a very high ALT allele frequency in East Asian (0.7827) and South Asian (0.7536) populations, whereas its frequency in African populations is lower (0.357). This disparity suggests a population-specific distribution that could affect the overall risk profile and prevalence of endometriosis in these groups.

Moreover, some variants like rs116250606 and rs75801644 demonstrate low frequencies across most populations or are completely absent in certain groups, implying that these variants might have a more restricted or localized impact on disease risk. Absence of data (noted as "None") for some SNPs in certain populations could reflect low variant occurrence, limited sampling, or lack of comprehensive genomic studies in those populations. Understanding these allele frequency patterns

is crucial for epidemiological studies and for the design of genetically informed diagnosis and treatment strategies. This genetic diversity underlines the importance of including multiple ethnic groups in genetic research to ensure the findings are broadly applicable and to identify population-specific genetic risk factors that may guide personalized medicine approaches for endometriosis**.**

Table 4. Allele Frequencies of Endometriosis-Associated SNPs Across Global Populations

Variation ID	REF Allele	ALT Allele	Allele Frequency				
			Africa	America	East Asia	Europe	South Asia
rs116250606	C	T	0.0923	0.0072	None	None	None
rs2278868	C	T	0.357	0.6484	0.7827	0.5885	0.7536
rs2276314	A	G	0.323	0.1859	0.248	0.2137	0.2352
rs75339189	C	A	None	None	None	None	None
rs75339189	C	T	0.0083	0.0706	0.0169	0.1093	0.0256
rs144824657	G	A	None	None	None	None	None
rs144824657	G	T	None	0.0072	None	0.0089	None
rs2341097	C	T	0.3298	0.2305	0.2937	0.3231	0.2822
rs75801644	G	A	0.0008	0.0202	None	0.0298	0.0051

The pathway enrichment analysis

The pathway enrichment analysis shown in **Table 5** highlights the biological pathways that are significantly associated with the endometriosis-related missense variants uncovered in this study. The table lists pathway IDs, their functional descriptions, parent pathway categories, p-values indicating statistical significance, genes involved, and corresponding variant IDs. Notably, several pathways associated with neuronal systems, metabolism, extracellular matrix organization, transport of small molecules, disease, and signal transduction are enriched with genetic variants, suggesting diverse molecular mechanisms underlying endometriosis.

A primary focus is on pathways related to potassium channels and voltage-gated potassium channels within the neuronal system, both involving the KCNG2 gene and its variant rs75339189. These channels play critical roles in regulating membrane potential and cellular signaling, processes that may be disrupted in endometriosis. Metabolic pathways, including the citric acid (TCA) cycle, respiratory electron transport, integration of energy metabolism, and pyruvate metabolism, are also enriched, driven largely by variants in the BSG gene (e.g., rs144824657). The involvement of metabolic pathways suggests that energy production and mitochondrial function may be altered in endometriosis pathophysiology.

Additionally, pathways involved in extracellular matrix organization such as degradation of the extracellular matrix and integrin cell surface interactions, again linked to BSG, point to changes in tissue remodeling and cell adhesion, which are central to endometriosis lesion formation. Signal transduction pathways involving the GPNMB gene and variant rs75801644 relate to PTK6 signaling

and HIF1A stabilization, processes important for cellular responses to stress, hypoxia, and proliferation. Together, these enriched pathways reveal how variants may affect various biological systems, offering potential targets for therapeutic intervention and advancing the molecular understanding of endometriosis.

Table 5. Pathway Enrichment Analysis of Endometriosis-Associated Missense Variants

Pathway ID	Description	Parent(s)	p-Value	Genes Involved	Variation IDs
R-HSA-1296071	Potassium Channels	Neuronal System	0.028407	*KCNG2*	rs75339189
R-HSA-1296072	Voltage gated Potassium channels	Neuronal System	0.011926	*KCNG2*	rs75339189
R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	Metabolism	0.044704	*BSG*	rs144824657
R-HSA-1474228	Degradation of the extracellular matrix	Extracellular matrix organization	0.038478	*BSG*	rs144824657
R-HSA-163685	Integration of energy metabolism	Metabolism	0.029772	*KCNG2*	rs75339189
R-HSA-210991	Basigin interactions	Hemostasis	0.006945	*BSG*	rs144824657
R-HSA-216083	Integrin cell surface interactions	Extracellular matrix organization	0.023482	*BSG*	rs144824657
R-HSA-381676	Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	Metabolism	0.011649	*KCNG2*	rs75339189
R-HSA-422356	Regulation of insulin secretion	Metabolism	0.021562	*KCNG2*	rs75339189
R-HSA-425366	Transport of bile salts and organic acids, metal ions and amine compounds	Transport of small molecules	0.023756	*BSG*	rs144824657
R-HSA-433692	Proton-coupled monocarboxylate transport	Transport of small molecules	0.00167	*BSG*	rs144824657
R-HSA-5619070	Defective SLC16A1 causes symptomatic deficiency in lactate transport (SDLT)	Disease	0.000557	*BSG*	rs144824657
R-HSA-5619102	SLC transporter disorders	Disease	0.027314	*BSG*	rs144824657
R-HSA-5619115	Disorders of transmembrane transporters	Disease	0.048211	*BSG*	rs144824657
R-HSA-70268	Pyruvate metabolism	Metabolism	0.008607	*BSG*	rs144824657
R-HSA-71406	Pyruvate metabolism and Citric Acid (TCA) cycle	Metabolism	0.015237	*BSG*	rs144824657
R-HSA-8848021	Signaling by PTK6	Signal Transduction	0.014961	*GPNMB*	rs75801644
R-HSA-8857538	PTK6 promotes HIF1A stabilization	Signal Transduction	0.00167	*GPNMB*	rs75801644

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

R-HSA-9006927	Signaling by Non-Receptor Tyrosine Kinases	Signal Transduction	0.014961	*GPNMB*	rs75801644

DISCUSSION

This study advances the understanding of the genetic basis of endometriosis by identifying missense variants strongly associated with disease risk using a GWAS-based approach. Compared with prior GWAS efforts, such as the meta-analysis by Rahmioglu et al. (2014),[25] which confirmed several loci associated with endometriosis across diverse populations, this study highlights high-impact missense variants (e.g., rs75801644 and rs144824657) with relatively large effect sizes (OR > 3). While earlier studies often reported variants with modest effect sizes (OR around 1.1-1.3), this analysis emphasizes variants that may exert stronger biological effects and therefore helps narrow the gap between association signals and functional interpretation. Unlike studies focused mainly on non-coding or regulatory variants, this research prioritizes protein-altering missense mutations, providing a clearer link to altered protein function and disease-related pathways.[26]

The integration of functional prediction tools such as PolyPhen-2 and SIFT is a strength of this study because it helps prioritize variants that are more likely to affect protein structure and function, an approach aligned with recent work advocating in silico prioritization followed by experimental validation of GWAS hits.[27, 28] Predicting deleterious or damaging effects adds biological plausibility beyond statistical association alone.[29] However, this study is limited by its reliance on in silico predictions without direct experimental validation, which is still needed to confirm the actual effects of these variants on protein function and disease phenotypes. In addition, although allele frequency analysis across global populations provides useful insight into genetic diversity, the absence of some variants in certain populations highlights the need for broader and more inclusive studies, particularly in underrepresented populations.[30]

A further strength of this study is the pathway enrichment analysis, which links prioritized variants to biological processes such as potassium channel function, metabolism, extracellular matrix organization, and signal transduction. This systems-level view is consistent with emerging evidence that endometriosis is a multifactorial disease involving altered metabolism and tissue microenvironment remodeling. By identifying pathways such as PTK6 signaling and hypoxia-related signaling, this study suggests potential directions for future therapeutic investigation.[31] However, pathway analysis depends on currently available annotation databases and may not capture novel or context-specific interactions relevant to disease pathology. Future studies integrating multi-omics data and functional assays will be needed to validate these pathway-level hypotheses and clarify gene-environment interactions.[32]

This study also addresses gaps in previous GWAS reports by focusing on missense variants rather than broader SNP associations, thereby providing more direct mechanistic hypotheses regarding how specific protein changes may contribute to endometriosis. However, the analysis does not stratify findings by endometriosis subtype or disease severity, a limitation that has been noted in prior work, including Rahmioglu et al., who reported stronger genetic effects in severe phenotypes.[25] Incorporating detailed phenotypic data in future studies may improve genotype-phenotype interpretation and support more accurate risk prediction. Moreover, despite the functional predictions presented here, the study does not include longitudinal or clinical correlation analyses to assess the effects of these variants on disease progression or treatment response.

From a public health perspective, these findings contribute to the growing evidence base for genetic risk stratification in endometriosis. In countries such as Indonesia, where delayed diagnosis remains common, identifying functionally relevant genetic variants may support the future development of early detection strategies and precision reproductive health approaches. However, this study remains limited by the absence of experimental validation, phenotype-stratified analyses, and sufficient representation of some ethnic groups in existing GWAS datasets. Future studies that integrate functional assays and diverse cohorts are needed to validate and extend these findings.

CONCLUSION

This study strengthens current knowledge of endometriosis genetics by identifying high-impact missense variants and characterizing their predicted functional effects and pathway associations. Compared with previous literature, this work narrows the focus from general association signals to potentially functional protein-altering variants, representing a step toward molecularly informed diagnostics and therapeutics. This study also demonstrates an integrative genomics approach that combines GWAS data, protein function prediction, population genetics, and pathway enrichment analysis to generate biologically meaningful hypotheses about endometriosis pathogenesis. Its main strengths lie in variant prioritization and biological contextualization, whereas its key limitations include the lack of experimental validation, limited clinical phenotyping, and incomplete population representation. Further studies using functional assays and ethnically diverse cohorts are required to validate these findings and improve their broader applicability. These results may serve as an initial reference for future genetic studies and early risk detection strategies for endometriosis, particularly in settings such as Indonesia where genomic-based reproductive health research is still limited.

DOI: 10.58185/jkr.v16i2.407
ISSN: 2354-8762 (Online)

RECOMMENDATION

Future studies should validate the identified variants through functional assays and population-specific analyses, particularly in underrepresented groups, to support translation into genetic risk assessment and personalized endometriosis management.

ACKNOWLEDGEMENT

The authors acknowledge the NHGRI-EBI GWAS Catalog for providing publicly accessible GWAS summary data. The authors also thank SNPnexus, the 1000 Genomes Project, and the Reactome database for enabling the functional annotation, population allele frequency analysis, and pathway enrichment analysis used in this study.

REFERENCES

1. Rathod S, Shanoo A and Acharya N. Endometriosis: A Comprehensive Exploration of Inflammatory Mechanisms and Fertility Implications. Cureus 2024; 16: e66128. 2024/09/04. Available from: https://doi.org/10.7759/cureus.66128
2. Miranda JA, Fabrini E, Coelho FMA, et al. Giant endometrioma in an asymptomatic patient. Radiol Case Rep 2024; 19: 1945-1948. 20240227. Available from: https://doi.org/10.1016/j.radcr.2024.01.064
3. Garmendia JV, De Sanctis CV, Hajdúch M, et al. Endometriosis: An Immunologist's Perspective. Int J Mol Sci 2025; 26 2025/06/13. Available from: https://doi.org/10.20944/preprints202504.1726.v1
4. Naik S, Sudan J, Urwat U, et al. Genome-wide SNP discovery and genotyping delineates potential QTLs underlying major yield-attributing traits in buckwheat. Plant Genome 2024; 17: e20427. 2024/01/19. Available from: https://doi.org/10.1002/tpg2.20427
5. Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet 2019; 20: 467-484. 2019/05/10. Available from: https://doi.org/10.1038/s41576-019-0127-1
6. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 2019; 570: 514-518. 2019/06/21. Available from: https://doi.org/10.1038/s41586-019-1310-4
7. Uimari O, Rahmioglu N, Nyholt DR, et al. Genome-wide genetic analyses highlight mitogen-activated protein kinase (MAPK) signaling in the pathogenesis of endometriosis. Hum Reprod 2017; 32: 780-793. 2017/03/24. Available from: https://doi.org/10.1093/humrep/dex024
8. Holdsworth-Carson SJ, Fung JN, Luong HT, et al. Endometrial vezatin and its association with endometriosis risk. Hum Reprod 2016; 31: 999-1013. 2016/03/24. Available from: https://doi.org/10.1093/humrep/dew047
9. Adewuyi EO, Mehta D, Sapkota Y, et al. Genetic analysis of endometriosis and depression identifies shared loci and implicates causal links with gastric mucosa abnormality. Hum Genet 2021; 140: 529-552. 2020/09/23. Available from: https://doi.org/10.1007/s00439-020-02223-6
10. Koppolu A, Maksym RB, Paskal W, et al. Epithelial Cells of Deep Infiltrating Endometriosis Harbor Mutations in Cancer Driver Genes. Cells 2021; 10 2021/04/04. Available from: https://doi.org/10.3390/cells10040749
11. Suda K, Nakaoka H, Yoshihara K, et al. Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell Rep 2018; 24: 1777-1789. 2018/08/16. Available from: https://doi.org/10.1016/j.celrep.2018.07.037
12. Schulte LA, Beck A, Marienfeld R, et al. Unveiling the intriguing relationship: oncogenic KRAS, morphological shifts, and mutational complexity in pancreatic mucinous cystic neoplasms. J Pathol 2025; 265: 401-407. 2025/02/05. Available from: https://doi.org/10.1002/path.6397

13. Sun Q, Crowley CA, Huang L, et al. From GWAS variant to function: A study of ~148,000 variants for blood cell traits. HGG Adv 2022; 3: 100063. 2022/01/21. Available from: https://doi.org/10.1101/2021.02.16.431409
14. Kaňovská I, Biová J and Škrabišová M. New perspectives of post-GWAS analyses: From markers to causal genes for more precise crop breeding. Curr Opin Plant Biol 2024; 82: 102658. 2024/11/17. Available from: https://doi.org/10.1016/j.pbi.2024.102658
15. Cousins FL, McKinnon BD, Mortlock S, et al. New concepts on the etiology of endometriosis. J Obstet Gynaecol Res 2023; 49: 1090-1105. 2023/02/07. Available from: https://doi.org/10.1111/jog.15549
16. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019; 47: D1005-d1012. 2018/11/18. Available from: https://doi.org/10.1093/nar/gky1120
17. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 2023; 381: eadg7492. 20230922. Available from: https://doi.org/10.1126/science.adg7492
18. Adzhubei I, Jordan DM and Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013; Chapter 7: Unit7 20. Available from: https://doi.org/10.1002/0471142905.hg0720s76
19. Vaser R, Adusumalli S, Leng SN, et al. SIFT missense predictions for genomes. Nat Protoc 2016; 11: 1-9. 20151203. Available from: https://doi.org/10.1038/nprot.2015.123
20. Oscanoa J, Sivapalan L, Gadaleta E, et al. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res 2020; 48: W185-W192. Available from: https://doi.org/10.1093/nar/gkaa420
21. Dayem Ullah AZ, Oscanoa J, Wang J, et al. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res 2018; 46: W109-W113. Available from: https://doi.org/10.1093/nar/gky399
22. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015; 526: 68-74. Available from: https://doi.org/10.1038/nature15393
23. Belsare S, Levy-Sakin M, Mostovoy Y, et al. Evaluating the quality of the 1000 genomes project data. BMC Genomics 2019; 20: 620. 20190816. Available from: https://doi.org/10.1186/s12864-019-5957-x
24. Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 2017; 18: 142. 20170302. Available from: https://doi.org/10.1186/s12859-017-1559-2
25. Rahmioglu N, Nyholt DR, Morris AP, et al. Genetic variants underlying risk of endometriosis: insights from meta-analysis of eight genome-wide association and replication datasets. Hum Reprod Update 2014; 20: 702-716. 2014/03/29. Available from: https://doi.org/10.1093/humupd/dmu015
26. Guare L, Humphrey LA, Rush M, et al. Enhancing Genetic Association Power in Endometriosis through Unsupervised Clustering of Clinical Subtypes Identified from Electronic Health Records. medRxiv 2024 2024/05/07. Available from: https://doi.org/10.1101/2024.04.22.24306092
27. Lei L, Xu X, Gong C, et al. Integrated analysis of genome-wide gene expression and DNA methylation profiles reveals candidate genes in ovary endometriosis. Front Endocrinol (Lausanne) 2023; 14: 1093683. 2023/04/11. Available from: https://doi.org/10.3389/fendo.2023.1093683
28. Flanagan SE, Patch A-M and Ellard S. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. Genetic Testing and Molecular Biomarkers 2010; 14: 533-537. Available from: https://doi.org/10.1089/gtmb.2010.0036
29. Sim NL, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012; 40: W452-457. 2012/06/13. Available from: https://doi.org/10.1093/nar/gks539

30. Suybeng V, Koeppel F, Harlé A, et al. Comparison of Pathogenicity Prediction Tools on Somatic Variants. The Journal of Molecular Diagnostics 2020; 22: 1383-1392. DOI: https://doi.org/10.1016/j.jmoldx.2020.08.007
31. Zhao F, Zhong H, You L, et al. PTK6 mediated immune signatures revealed by single cell transcriptomic and multi omics big data analysis in cervical cancer. Discov Oncol 2025; 16: 1566. 2025/08/16. Available from: https://doi.org/10.1007/s12672-025-03365-7
32. Datkhayeva Z, Iskakova A, Mireeva A, et al. The Multifactorial Pathogenesis of Endometriosis: A Narrative Review Integrating Hormonal, Immune, and Microbiome Aspects. Medicina (Kaunas) 2025; 61 2025/05/28. Available from: https://doi.org/10.3390/medicina61050811

Declarations

Author contribution	:	Conceptualized, conceived, and designed the study: Ichtiarini Nurullita Santri, Wirawan Adikusuma, and Lalu Muhammad Irham. Performed the analysis: Ichtiarini Nurullita Santri. Curated the data: Wirawan Adikusuma and Lalu Muhammad Irham. Wrote the original draft: Ichtiarini Nurullita Santri. Interpreted the data: Ichtiarini Nurullita Santri, Petrina Theda Philothra, Wirawan Adikusuma, Lalu Muhammad Irham, and Nurul Fadhliati Maulida. Revised the manuscript: Ichtiarini Nurullita Santri, Wirawan Adikusuma, Lalu Muhammad Irham, Petrina Theda Philothra, Nurul Fadhliati Maulida, Rockie Chong, Ilker Ates, and Yohane Vincent Abero Phiri. All authors have read and agreed to the published version of the manuscript.
Funding statement	:	This research did not receive any funding.
Conflict of interest	:	The authors declare that they have no competing interests.
Additional information	:	No additional information is available for this paper.