DOI: 10.24996/ijs.2026.67.2.13

ISSN: 0067-2904

Polymorphism and Haplotype of *Mucin four* gene in Iraqi women with polycystic ovarian syndrome

Abeer A. Faisal, Asmaa M. Salih Almohaidi*
Biology Department, College of Science for Women, the University of Baghdad, Ira

Received: 20/10/2023 Accepted: 13/2/2025 Published: xx

Abstract

The protein-coding gene *MUC4* (*Mucin 4*) is positioned on chromosome 3's long arm at position q29. The current study aims to investigate *MUC4* polymorphisms and haplotypes concerning polycystic ovarian syndrome (PCOS). The research involved 120 women, comprising 60 women with PCOS, and 60 healthy fertile individuals serving as a control group. This study focused on sequencing of certain segments (775 bp) from exon 11 in the *MUC4* gene to identify polymorphic SNPs and haplotypes. Four SNPs in the *MUC4* gene were examined. The SNP rs781407488 GG homozygous genotype showed a significant protective association with PCOS with a 0.46 odds ratio. In contrast, the GA heterozygous and AA homozygous genotypes exhibited nonsignificant risky effects with 2.02 and 3.05 odds ratios associated with PCOS. These findings suggest that the G allele acted as a protective allele, while the A allele may be associated with the disease. The other three SNPs, rs1172519849, rs1453997681, and rs1719865600 genotypes recorded no significant differences in genotype and allele frequencies. Additionally, a high linkage disequilibrium (LD) was observed among the four SNPs of the *MUC4* gene. The genotypes GA and AA of SNP rs781407488 increased frequency act as risky factors that may associated with increasing susceptibility for PCOS in Iraqi females, while the GG genotype acts as a protective factor. The current SNPs rs1172519849, rs781407488, rs1453997681, and rs1719865600 of the *MUC4* gene would inherit as one block through generation that will increase the susceptibility for PCOS and may associated with the development of PCOS.

Keywords: *MUC4* gene, Single nucleotide polymorphisms, haplotype, PCOS.

تعدد الطرزالجينية والنمط الوراثي الفردي للمورث الميوسين –4 في النساء العراقيات المصابات بمتلازمة تكيس المبايض

عبير عبد الحميد فيصل , اسماء محمد صالح المهيدي*
قسم علوم الحياة, كلية العلوم للبنات, جامعة بغداد, بغداد, العراق

الخلاصة

يقع المورث المشفر لبروتين ميوسين–4 (*MUC4*) على الذراع الطويل من الصبغي الثالث (3p29) . تهدف الدراسة الجديدة الحالية الى التحقق في تعدد الطرز الجينية مع النمط الوراثي الفردي للمورث *MUC4* المتعلقة بمتلازمة تكيس المبايض (PCOS) . شمل البحث 120 , 60 امرأة مصابة بمتلازمة تكيس

*Email: asmaams_bio@csw.uobaghdad.edu.iq

المبايض و60 فردا يتمتعون بصحة جيدة ويمثلون مجموعة سيطرة . ركزت هذه الدراسة بتحليل تسلسل اجزاء محددة (775pb) من الاكسون الحادي عشر للمورث *MUC4* لتحديد تعدد الطرز الجينية والنمط الوراثي الفردي . اظهرت الدراسة الحالية اربعة نيوكليدات مفردة (SNPs) للمورث *MUC4* .اظهرالطراز الجيني مماثل الزيجة GG للنيوكليوتيدة المفردة rs781407488 ارتباطا وقائيا مهما مع متلازمة تكيس المبايض بقيمة خطورة نسبية 0.46 ، بالمقابل اظهرت الطراز الجيني متغاير الزيجة GA والطراز الجيني مماثل الزيجة AA تاثيرات خطرة غير مهمة بقيمة خطورة نسبية 2.02 و 3.05 ارتباطا متلازمة تكيس المبايض ، تشير النتائج الى ان الاليل G يعمل كأليل وقائي بينما يمكن ان يكون للاليل A بقدرته على الارتباط بالمرض .لم تسجل النيوكليوتيدات المفردة الثلاثة الاخرى rs1172519849 ، rs1453997681 و rs1719865600 في مورث *MUC4* اي فروق معنوية للطرز الجينية وتردد الاليلات . بالاضافة الى ذلك، ظهر ارتباطا عاليا جدا (LD) بين النيوكليوتيدات المفردة الاربعة للمورث *MUC4* مع بعضها البعض . الطرز الجينية GA و AA للنيوكليوتيدة المفردة rs781407488 تزيد من احتمالية كونها عامل خطورة مرتبطة بزيادة قابلية الاصابة بمتلازمة تكيس المبايض في حين الطراز الوراثي GG يعتبر عامل وقائي . النيوكليوتيدات المفردة rs1172519849، rs781407488،rs1453997681 و rs1719865600 لمورث *MUC4* تورث كقطعة واحدة عبر الاجيال والتي سوف تزيد من قابلية الاصابة بمتلازمة تكيس المبايض وقد ترتبط بتطور متلازمة تكيس المبايض .

Introduction

Mucin is a component of mucus that serves to lubricate endothelial tissue. It is a large glycoprotein with high molecular weight, playing a crucial role in lubricating the epithelial surfaces of the respiratory, gastrointestinal, and reproductive tracts [1, 2]. The cervix and endometrium, two reproductive organs with important reproductive system roles, produce mucins. These molecules mediate inflammatory and immunological responses and regulate transcriptional and posttranscriptional processes [3-5].

Altered mucin composition or excessive viscosity of cervical mucus can create a barrier that prevents sperm from reaching the uterus [6]. Additionally, overexpression of mucin may hinder embryo attachment by obstructing access to adhesion molecules on the endometrial surface [7]. The thickness of the endometrium should be within normal range, any change in the thickness of the endometrium will affect the success of the embryo implantation process. [8]. Therefore, the presence of atypical mucus output may be a contributing factor to female infertility.

The mucin genes belong to a family of 24 members (*MUC1* to *MUC24*) and can be divided into secreted and transmembrane mucins [9, 10]. One of them is *MUC4* which is the primary mucin found in the endometrial epithelium of several species, alongside *MUC1* and *MUC16* [11, 12].

The *MUC4* gene is located on the long arm of chromosome 3 at band 9 (3q29) and comprises 25 exons/introns over 100 kb. It encodes a 930-kDa transmembrane mucin predicted to protrude over 2 μm above the cell surface [13-16].

MUC4 forms a protective barrier on the epithelial surfaces of reproductive organs, such as the cervix, uterus, and fallopian tubes in females. This barrier serves to shield mechanical injury, infections, and toxins [17]. Additionally, *MUC4* aids in lubricating mucosal membranes, which is particularly crucial during ovulation, and creating a suitable environment for aiding sperm transit past the cervix and into the uterus [18]. Furthermore, *MUC4* may play a role in the preparation of the endometrium for embryo implantation by altering the characteristics of the epithelial lining, *MUC4* might affect the uterus's receptivity for embryo attachment [19].

MUC4 polymorphism has a crucial role in the development of diseases, especially in cancer progression by multiple mechanisms such as enhancing cancer cell proliferation, inhibiting apoptosis, facilitating metastasis, and modulating tumor microenvironments. Research has

demonstrated that *MUC4* is associated with the onset and regulation of colorectal cancer (CRC) and has the potential as a cancer susceptibility indicator [20-22].
Polycystic ovarian syndrome (PCOS) is the most common endocrine condition among women of reproductive age [23]. PCOS encompasses four clinical conditions, referred to as "phenotypes," grouped due to shared clinical characteristics [24, 25].
Since PCOS is one of the types of infertility, multiple genetic factors are significant contributors to infertility in Iraqi women [26], and have a role in causing infertility among individuals in the Iraqi population [27, 28].
Previous studies have confirmed that polymorphisms in the *MUC4* gene are associated with recurrent pregnancy loss. This relationship is attributed to the role of mucin in facilitating blastocyst adhesion by creating a favorable environment for embryo implantation within the uterus endometrium [29, 30]. Even after much research on physiology and genetics, the pathophysiology of PCOS is still not fully understood. The role of *MUC1* gene polymorphism and expression with infertility has been studied in Iraqi infertile women [31], but the *MUC4* polymorphism and its association with PCOS in the Iraqi population is still under-studied. The current work aims to identify new factors that could serve as prognostic and diagnostic indicators for PCOS and enhance treatment strategies. Therefore, the present study explores the relationship between *MUC4* gene polymorphism and haplotype with PCOS in Iraqi females.

Materials and Methods:

Study Design:

The current study involved 120 women, including 60 women diagnosed with PCOS according to Rotterdam criteria, aged between 18 to 45 years, and sixty healthy fertile women of comparable age who had at least one previous childbirth. Samples were collected from AL-Elwiya Educational Hospital. Ethical approval was obtained from the Ministry of Health / Baghdad Health Department (permission number 117181, issued on June 8, 2023), and from the College of Science for Women/ University of Baghdad. The sample collection period extended from August 2023 until December 2023. Blood samples were collected on days 2-5 of the menstrual cycle. All females included in this study were within the reproductive age range and free from endocrine or other systemic diseases, while those with endocrine diseases and other diseases were excluded.

Collection of Blood Samples:

Blood samples were collected from each participant for both groups, with 2 ml of blood placed directly into an EDTA-containing tube for the genotyping study.

Genomic DNA Extraction and Genotyping:

The DNA was extracted from the study group subjects' whole blood using the protocol in the EasyPure® Blood Genomic DNA Kit (TransGen, biotech. EE121-01/China). The genotyping of the *MUC4* was carried out using polymerase chain reaction (PCR) and sequencing. The PCR cycles for the *MUC4* (775bp segment) reaction began at 94°C for 5 minutes, then 35 cycles of denaturation at 94°C for 30 seconds, annealing at 58°C for 30 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 5 minutes. PCR reactions were found in a 25- µL mixture that had 6 µL of genomic DNA, 4.5 µL of D.W., 12.5µL of master mix, and 1µL of each primer. The primers were designed using Primer 3plus, V4, and double-checked by the University Code of Student Conduct (UCSC) programs, and with their reference sequences in the National Center for Biotechnology Information (NCBI) database. They were synthesized and lyophilized by Alpha DNA Ltd. (Canada)The primers sequenced (775 bp) of DNA, the target region size of this primer in exon 11 of the *MUC4* gene. This gene is on chromosome 3 q29 and is shown in Table 1:

Table 1: Specific primer sequences for DNA sequencing.

Primer	Sequence (5′→3′ direction)	Templet length	Reference
	***MUC4* Human Sequencing Primer**		
Forward.	ATGGCAGTTTCCACAGTTCC	775 bp	This Study Design
Reverse	AGGTACAAAGCCCCTCCACT		

DNA sequencing:

The purified PCR products of the *MUC4* gene were sent to Macrogen Company in Korea for DNA sequencing. Additionally, the nucleotide sequences were compared with data from the gene bank of the National Center for Biotechnology Information (NCBI) website databases using the BLAST search tool.

Statistical Analysis:

The statistical significance of the P values ($P \leq 0.05$) was determined using Fisher's exact test, while the Odds Ratio was calculated through a particular $\chi 2$ analysis using the WINPEPI computer application (version 11.63), the odds ratio was used to estimate the likelihood of specific genotypes or alleles being associated with PCOS. The Hardy-Weinberg equilibrium was calculated to determine the expected genotype and allele frequencies in populations, and the Pearson chi-squared test was used to compare genotype distribution among groups conducted using the OEGE - Online Encyclopedia for Genetic Epidemiology studies [32].

Results and Discussion:

MUC4 Genotyping:

The current study is a pioneering investigation of these SNPs in the *MUC4* gene within the Iraqi population. It analyzes the total number of patients and fertile control and the genotype frequencies according to the number of PCOS patients (See Table 2). For the SNP rs1172519849 polymorphism, no significant differences were observed in the distribution of genotypes and the frequency of alleles between the patients and fertile control groups. The odds ratio for the GC hetero genotypes was 1.31, while it was 0.82 for the GG genotype and 0.33 for the CC genotype. These findings indicate that females with the GC genotype may have a higher susceptibility to PCOS, whereas the GG and CC genotypes may serve as a protective factor against this susceptibility. Additionally, the odds ratio for the C allele was 1.09, suggesting a possible association with the disease, the females who carry allele C may have more susceptibility to PCOS. Conversely, the odds ratio for the G allele was 0.91, showing a decrease in susceptibility and tend to be a protective effect with fertility.

For the rs1453997681 polymorphism, no significant differences were found in the distribution of genotypes and the frequency of alleles between the patients and fertile control groups. The odds ratio for the GG genotype was 0.76, while it was 1.22 for the GA genotype and 3.05 for the AA genotype. These findings indicate that females with the GA and AA genotypes may tend to progress PCOS, whereas the GG genotype acts as a protective factor and females who carry it tend to be fertile. Additionally, the odds ratio for the A allele was 1.26, indicating a potential association with the disease. In contrast, the odds ratio for the G allele was 0.79, showing a decrease in susceptibility to PCOS and tend to be fertility.

Table 2: Comparison of the genotype distribution and allele frequencies of *MUC4* gene polymorphism between the patient group and control group.

Genotypes	Patients [n 60, (%)]	Control [n 60, (%)]	Odds Ratio (OR)	Confidence Interval (95% C.I.)	P value
		rs1172519849			
GG	31 (51.67%)	34 (56.67%)	0.82	(0.39 to 1.67)	0.58
GC	29 (48.33%)	25 (41.67%)	1.31	(0.63 to 2.69)	0.46
CC	0 (0.0%)	1 (1.67%)	0.33	(0.01 to 8.21)	0.49
Allele G	91 (75.83%)	93 (77.5%)	0.91	(0.50 to 1.66)	0.76
Allele C	29 (24.17%)	27 (22.5%)	1.09	(0.60 to 1.99)	0.76
		rs1453997681			
GG	30 (50%)	34 (56.67%)	0.76	(0.43 to 1.44)	0.46
GA	29 (48.33%)	26 (43.33%)	1.22	(0.59 to 2.51)	0.58
AA	1 (1.67%)	0 (0.0%)	3.05	(0.12 to 76.4)	0.49
Allele G	89 (74.17%)	94 (78.17%)	0.79	(0.44 to 1.44)	0.45
Allele A	31 (25.83%)	26 (21.67%)	1.26	(0.69 to 2.28)	0.45
		rs781407488			
GG	30 (50%)	41 (68.33%)	0.46	(0.22 to 0.69)	**0.04***
GA	29 (48.335)	19 (31.67%)	2.02	(0.96 to 4.24)	0.05
AA	1 (1.67%)	0 (0.0%)	3.05	(0.12to 76.39)	0.49
Allele G	89 (74.17%)	101 (84.17%	0.54	(0.28 to 1.02)	0.06
Allele A	31 (25.83%)	19 (15.83%)	1.85	(0.97 to 3.50)	0.06
		rs1719865600			
CC	31 (51.67%)	41 (68.33%)	0.49	(0.24 to 1.04)	0.06
CA	29 (48.33%)	19 (31.67%)	2.02	(0.96 to 4.24)	0.06
AA	0 (0.0%)	0 (0.0%)	0	……	0
Allele C	91 (75.83%)	101 (84.17%)	0.59	(0.31 to 1.12)	0.11
Allele A	29 (24.17%)	19 (15.83%)	1.69	(0.89 to 3.22)	0.11

Non-significant P> 0.05, significant differences P ≤ 0.05*, OR: Odd Ratio (relative risk) OR=1: No association, the event is equally likely in both groups. OR>1: The event is more likely in patients than in control. OR<1: The event is less likely in patients than in control, CI: confidence interval. (CI: Is a range of values that is used to estimate an unknown population parameter with a specified level of confidence, it provides a measure of uncertainty or precision for a statical estimate).

For the rs781407488 polymorphism, significant differences were observed in the GG, while no significant differences were noted for the GA and AA genotype distribution and allele frequency between patients and fertile control groups. The odds ratios for the GA and AA genotypes were greater than 1 (2.02 and 3.05, respectively), indicating that these genotypes act as risk factors. In contrast, the GG genotype showed a significant difference (P ≤ 0.05) between the two groups, with an odds ratio of 0.46, suggesting that it has a protective effect against PCOS. The odds ratio for the G allele was 0.54 while the odds ratio for the A allele was 1.85, indicating that the A allele could be susceptible to association with the disease, but the G allele decreases this susceptibility by acting as a protective agent against PCOS. So when a female inherits both alleles, their protective effect will be more profound in protecting against PCOS and the female tend to be fertile.

Regarding the rs1719865600 polymorphism, no significant differences were found in the distribution of genotypes and the frequency of alleles between the patients and fertile control groups. The odds ratio for the CC genotype was 0.49, but for the CA genotype, it was 2.02. The findings indicate that females with the CA genotype may be more vulnerable to PCOS, whereas the CC genotype acts as a protective factor, reducing this susceptibility. Additionally, the odds ratio for the A allele was 1.69, suggesting a possible association with the disease. Conversely, the odds ratio for the C allele was 0.59, showing decreased susceptibility to infertility. Therefore, females with this allele tend to be fertile.

Table 3 presents the expected frequencies of *MUC4* genotypes compared to the observed frequencies, analyzed using Hardy-Weinberg Equilibrium. The allele frequencies agreed with HWE which means the studied group is consistent with HWE, indicating that the studied population is not affected by other external factors. The frequency of genotypes and alleles of SNP rs1172519849 polymorphism goes with the Hardy–Weinberg Equilibrium among the fertile control group. Whereas among patients' group genotypes, the frequencies of genotypes and alleles were inconsistent with the Hardy-Weinberg equilibrium. So, this locus may undergo genetic variation in the Iraqi population, and random changes in allele frequencies due to chance events, particularly in small populations can lead to loss of alleles or fixation of certain alleles over time [33]. From the same table, the total observed results indicate that the GG genotype may be considered a common genotype in the Iraqi female population with a total of 65 individuals from both patient and fertile control groups exhibiting this genotype. In contrast, the other genotype, the GC genotype, was recorded in 54 females, while the CC genotype was in one female.

Table 3: The comparison between observed and expected genotype according to Hardy Weinberg Equilibrium for several SNPs of the *MUC4* gene.

Group	**Genotypes frequencies**				**χ^2**
rs1172519849		GG	GC	CC	
Control	Observed	(34)	(25)	(1)	2.28 C
	Expected	36.03	20.92	3.03	
Patients	Observed	(31)	(29)	(0)	6.09 NC
	Expected	34.50	21.99	3.50	
Total observed		65	54	1	
rs1453997681		GG	GA	AA	
Control	Observed	(34)	(26)	(0)	4.59 NC
	Expected	36.81	20.37	2.81	
Patients	Observed	(30)	(29)	(1)	4.09 NC
	Expected	33.00	22.99	4.00	
Total observed		64	55	1	
rs781407488		GG	GA	AA	
Control	Observed	(41)	(19)	(0)	2.12 C
	Expected	42.50	15.99	1.50	
Patients	Observed	(30)	(29)	(1)	4.09 NC
	Expected	33.00	22.99	4.00	
Total observed		71	48	1	
rs1719865600		CC	CA	AA	
Control	Observed	(41)	(19)	(0)	2.12 C
	Expected	42.50	15.99	1.50	
Patients	Observed	(31)	(29)	(0)	6.09 NC
	Expected	34.50	21.99	3.50	
Total observed		72	48	0	

C: Distribution consistent with Hardy Weinberg's law at the level of significance: $X^2 < 3.84$, NC: distribution does not consistent with Hardy Weinberg's law at the level of significance: $X^2 > 3.84$.

The distribution of SNP rs1453997681 polymorphisms does not confirm to Hardy-Weinberg Equilibrium, as the observed genotype frequencies significantly differ from those expected values. This study faced several limitations, including sample size and population specificity, as the control sample was derived from related people. Moreover, the Iraqi population has a tendency for intermarriage among related families, which may suggest that this locus is going through evolutionary selection in the Iraqi population because of this kind of relationship between relative individuals. PCOS, distinct from HWE, might also impact the patient group.

The total observed results in Table 3 suggest that the GG genotype may be considered a common genotype in Iraqi females because the total observed record (patients with fertile control groups) was 64 females.

The frequency of genotypes and alleles of SNP rs781407488 polymorphisms goes with the Hardy–Weinberg Equilibrium among the fertile control group. Whereas among patients' group genotypes, the frequencies of genotypes and alleles were inconsistent with the Hardy–Weinberg Equilibrium. That means this disease causes deviation of expected genotypes from HWE. Based on the total observed results, the GG genotype may be considered a common genotype in Iraqi females because the total number of women who carry the GG genotype was 71.

Finally, the frequency of genotypes and alleles of SNP rs1719865600 polymorphisms conform to the Hardy–Weinberg Equilibrium among the fertile control group. In contrast, the genotype frequencies in the patient group do not align with Hardy–Weinberg Equilibrium. Besides the impact of the disease on genotype frequency, other factors such as genetic drift, natural selection, or migration in the population, make genotype distributions deviate from HWE [33].

The total observed CC genotype can be regarded as a common genotype among Iraqi females, as the combined number of patients and fertile control groups exhibiting this genotype was 72, while the other genotypes were lower. All SNP genotypes according to Pearson chi-square not consist of HWE which means diseases cause departure from HWE which means that these studied SNPs may make females tend to be more susceptible to PCOS.

MUC4, a membrane-bound mucin, plays a significant role in cellular signalling, immune response, and epithelial protection [17]. Several published researchers have found that variations in the *MUC4* gene are linked to an increased risk of developing lung cancer and colorectal Cancer, as well as the development of endometriosis and infertility related to endometriosis [34-38]. Therefore genetic variation of the *MUC4* gene is associated with different diseases.

Genetic variation as polymorphism in the current study is associated with PCOS since *MUC4* and *MUC1* are the two major mucins in the endometrial epithelium [39, 40]. Genetic diversity significantly influences disease susceptibility via several mechanisms, including regulatory disruptions and transcriptional changes. With altered gene expression, hence increases the likelihood of disease development. [41, 42] Therefore, any variation in the *MUC4* and *MUC1* genes could change their function and may cause infertility.

Earlier Iraqi research has demonstrated that *MUC1* gene polymorphisms are linked to infertility, supporting these current findings [26,31]. The current study shows that genetic polymorphism is associated with PCOS; this result goes with a few studies that have explored

the potential connection between *MUC4* gene polymorphism and reproductive health issues [12, 34]. *MUC4* is expressed in reproductive tissues, such as the endometrium and ovaries, and may influence ovarian function and the endometrial environment [35]. Therefore, any genetic variation in the *MUC4* gene sequence could change their function and cause infertility.

This study identifies several SNP polymorphisms and SNPs in exon 11 from *MUC4*. In contrast, Chang and his team found that women carrying the T allele of SNP rs882605 were less likely to develop infertility due to endometriosis. Additionally, SNPs rs2688513 and rs2246901 did not demonstrate any association with the reproductive ability of patients [34]. Another study looked at Korean women and found genetic links between *MUC4* rs882605 C>A and *MUC4* rs1104760 A>G polymorphisms and repeated pregnancy loss (RPL) [29, 43].

Linkage Disequilibrium:

Linkage Disequilibrium (LD) measures the degree to which alleles at different loci are inherited together more often than expected by chance. SNP alleles typically arise from rare mutations that occur on a specific inherited genetic background and are passed down from one generation to the next. As a result, SNPs often exhibited associations with each other, appearing in convergent positions on a single chromosome. When alleles at two loci are linked and inherited together throughout a population, they are said to be in linkage disequilibrium (LD)[44]. The association between the four SNP polymorphisms of the *MUC4* gene has been analyzed using linkage disequilibrium (LD). Figure 1 shows that two *MUC4* variants, rs1172519849, and rs1453997681, have a high linkage disequilibrium (LD) with a 0.97. Other tested genetic variants also have high LD rs1172519849 and rs781407488 with a D' of 0.88. rs1172519849 and rs1719865600 with a D' of 0.88, rs1453997681 and rs781407488 with a D' of 0.88, rs1453997681 and rs1719865600 with a D' of 0.88, and finally rs781407488 and rs1719865600 with a D' of 0.86 as moderate linkage comparing with others.

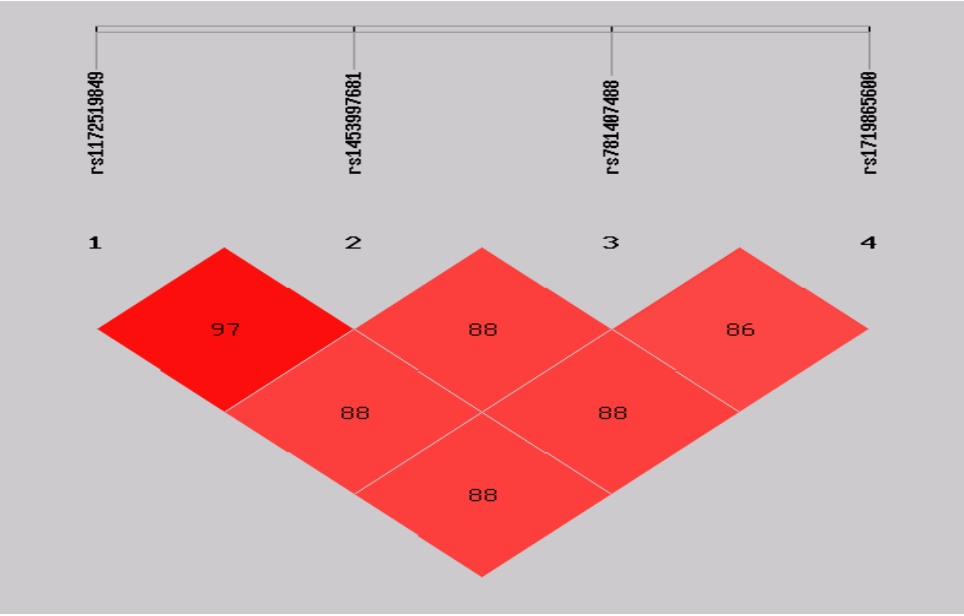

Figure 1: Linkage disequilibrium estimated between SNPs in mucin four genes. The blocks indicate haplotype blocks and the text above the horizontal numbers is the SNP names. The values in the red boxes are pair-wise SNP correlations (D′). The red-to-white gradient reflects higher to lower LD values.

The current study investigated the association between four SNPs of the *MUC4* gene (rs1172519849, rs1453997681, rs781407488, and rs1719865600) and PCOS susceptibility in

Iraqi women as a novel as a first local study focusing on this relation. Even *MUC4* polymorphisms of SNPs rs1172519849, rs1453997681, and rs1719865600 show significant differences according to HWE comparing patients to control groups. The SNP rs781407488 shows significant differences in the types of GG genotypes found in patients and control groups. The GG genotype is associated with increased PCOS prevalence. Furthermore, the polymorphisms of SNPs rs1172519849, rs1453997681, and rs1719865600 in the current study align with the findings of the previous study regarding the MUC1 gene [26,31].

This suggests that these polymorphisms of SNPs may associate effects on endometrium function and causes of diseases such as endometriosis [45]. As separated SNPs, when studied as one bloke these current findings suggest that these four SNPs constitute a single haplotype block. This means these four SNPs in the *MUC4* gene may be associated with PCOS when inherited as one block and make a new generation of women more susceptible to PCOS.

This result agrees with a previous study by Kim and his team who emphasized that the recurrent pregnancy loss in Korean women showed linkage disequilibrium for many SNPs within the *MUC4* gene, and their SNP haplotype may enhance susceptibility to recurrent pregnancy loss [29]. Pregnancy loss in previous studies refers to problems within the endometrium and these medical issues lead to infertility. This is similar to PCOS which may be one of the reasons that led to pregnancy loss. A different study about haplotype analysis of vitamin D receptors (VDR) confirmed that a strong linkage disequilibrium was recorded between close positions in the *VDR* gene [46]. The present result contradicts a recent study that emphasized no LD between the rs4889, rs12998, and rs35431622 SNPs in the *KISS1* gene and the development of PCOS [47].

Haplotype:

This Iraqi study focuses on *MUC4* polymorphism and haplotype to explore the role of inheritance of several alleles as one block. A haplotype can be defined as the combination of alleles of different polymorphisms that occur on the same chromosome, they are inherited together because they are very close and not by crossing or recombining between these loci. Table 4 shows the presence of haplotypes. CAGC (OR = 0.20, 95% Cl: 0.02~1.69, P = 0.10) and GGGC (OR = 0.75, 95% Cl: 0.39~1.43, P = 0.38) in the *MUC4* gene were significantly associated with a decrease in risk of PCOS. Simultaneously, the haplotype CAAA (OR = 1.76, 95% Cl: 0.88~3.52, P = 0.11) appears to function as a risk factor, as it increases susceptibility to PCOS. This means that the risk of PCOS is increased for those women carrying CAAA haplotype, and decreased for those carrying CAGC and GGGC haplotypes.

Table 4: analysis of haplotypes in the *MUC4* gene.

Haplotypes	Patients Frequency	Control frequency	Fisher's p	OR	95%CI
C A A A	24.93 (0.21)	15.96 (0.13)	0.11	1.76	[0.88~3.52]
C A G C	1.05 (0.01)	5.06 (0.04)	0.10	0.20	[0.02~1.69]
G G G C	84.88 (0.71)	91.90 (0.77)	0.38	0.75	[0.39~1.43]

Frequency<0.03 in both patients and control groups has been dropped, OR: Odd Ratio, CI: confidence interval.

The analysis of haplotypes in the *MUC4* gene is an important step. Because each SNP may have an influence on noticeably altering the phenotype, or an individual SNP in DNA may not directly impact the structure or function of proteins. Still, they have the potential to modify the levels of gene expression. The individual influence of a solitary regulatory SNP may be insignificant by itself. Still, when added to other SNPs as a collection block that controls gene function or expression, the overall effect can change the phenotype by changing gene structure

and expression [37]. This result aligned with a previous study on another member of the mucin gene family which showed a significant difference in the AAGC genotype in the *MUC1* gene between the groups of study and is considered a risk factor for the development of endometriosis [48].

However, considering all SNPs as a unit, a strong correlation with the PCOS is observed according to the odds ratio of 1.76 of the CAAA haplotype. Many previous Iraqi studies emphasized the association between SNPs and the genes responsible for developing many diseases in the Iraqi population like the association between the SNPs rs34402524 and rs2454206 in the *TET2* gene and Myeloid Leukemia.[49], and the association between the SNPs rs28416813, rs4803219, rs11881222, and rs8103142 in the *Interleukin-28β* gene and susceptibility to Hepatitis C Virus (HCV) infection [50].

This research suggests that the SNPs identified in the *MUC4* gene may be linked to PCOS and could be strongly linked as a single block, potentially increasing the likelihood of infertility among Iraqi women, particularly those with the CAAA haplotype. Further note besides haplotype the present findings recorded that all studied SNPs in patient groups significantly deviate from HWE, at the same time there are significant differences among genotype distribution which may associated with increasing females' susceptibility to PCOS. Further comprehensive investigations are required to explore the specific biochemical pathways controlled by *MUC4* throughout the progression of PCOS. The present finding submitted a suggestion that study the other mucin genes with their relationship with infertility and PCOS.

Conclusion

This novel study identified a significant association between GG genotype variations of SNP rs781407488 in the *MUC4* gene and PCOS, highlighting it as a potential cause of infertility in a sample of the Iraqi population. The GG genotype acts as a protective factor for PCOS which means that the females who carry the GG genotype tend to be fertile. The four SNPs in the *MUC4* gene that were studied also had a very high level of LD with each other. This suggests that these four SNPs in the *MUC4* gene may be linked to PCOS when passed down as a single unit (one block), making future female generations more likely to develop PCOs. The haplotype CAAA tends to act as a risk factor because it may be associated with PCOS, while CAGC and GGGC associated with fertile. The current genetic study can emphasize that the genotype polymorphism and haplotype could be very prognostic and diagnostic factors for PCOS and also could help to understand the possibility of transmitted susceptibility for disease through generations. Inherited genetic factors related to PCOS could determine the medical issue treatment, guiding whether specific medical interventions are necessary for affected females.

Acknowledgements:

The authors express their gratitude to all those who voluntarily participated in this study. In a declaration, the authors have no financial support.

Conflicts of Interest:

There are no conflicts of interest.

References

[1] B. A. Symmes, A. L. Stefanski, C. M. Magin, and C. M. Evans, "Role of mucins in lung homeostasis: regulated expression and biosynthesis in health and disease," *Biochemical Society Transactions*, vol. 46, no. 3, pp. 707-719, 2018.

[2] Y. Kang, H. Park, B.-H. Choe, and B. Kang, "The role and function of mucins and its relationship to inflammatory bowel disease," *Frontiers in Medicine*, vol. 9, p. 848344, 2022.

[3] M. Perrais, P. Pigny, M.-C. Copin, J.-P. Aubert, and I. Van Seuningen, "Induction of MUC2 and MUC5AC mucins by factors of the epidermal growth factor (EGF) family is mediated by EGF receptor/Ras/Raf/extracellular signal-regulated kinase cascade and Sp1," *Journal of Biological Chemistry*, vol. 277, no. 35, pp. 32258-32267, 2002.

[4] E. Gensch, M. Gallup, A. Sucher, D. Li, A. Gebremichael, H. Lemjabbar, A. Mengistab, V. Dasari, J. Hotchkiss, and J. Harkema, "Tobacco smoke control of mucin production in lung cells requires oxygen radicals AP-1 and JNK," *Journal of Biological Chemistry*, vol. 279, no. 37, pp. 39085-39093, 2004.

[5] M. Zhou, T. Tian, and C. Wu, "Mechanism underlying the regulation of mucin secretion in the uterus during pregnancy," *International Journal of Molecular Sciences*, vol. 24, no. 21, p. 15896, 2023.

[6] U. Schimpf, E. Caldas-Silveira, L. Katchan, C. Vigier-Carriere, I. Lantier, G. Nachmann, S. Gidlöf, A. F. Jonasson, L. Björndahl, and S. Trombotto, "Topical reinforcement of the cervical mucus barrier to sperm," *Science Translational Medicine*, vol. 14, no. 673, p. eabm2417, 2022.

[7] B. A. Symmes, A. L. Stefanski, C. M. Magin, and C. M. Evans, "Role of mucins in lung homeostasis: regulated expression and biosynthesis in health and disease," *Biochemical Society Transactions*, vol. 46, no. 3, pp. 707-719, 2018.

[8] Z. Liao, C. Liu, L. Cai, L. Shen, C. Sui, H. Zhang, and K. Qian, "The effect of endometrial thickness on pregnancy, maternal, and perinatal outcomes of women in fresh cycles after IVF/ICSI: a systematic review and meta-analysis," *Frontiers in endocrinology*, vol. 12, p. 814648, 2022.

[9] S. Kaur, S. Kumar, N. Momi, A. R. Sasson, and S. K. Batra, "Mucins in pancreatic cancer and its microenvironment," *Nature Reviews Gastroenterology & hepatology*, vol. 10, no. 10, pp. 607-620, 2013.

[10] R. Bhatia, S. K. Gautam, A. Cannon, C. Thompson, B. R. Hall, A. Aithal, K. Banerjee, M. Jain, J. C. Solheim, and S. Kumar, "Cancer-associated mucins: role in immune modulation and metastasis," *Cancer and Metastasis Reviews*, vol. 38, pp. 223-236, 2019.

[11] S. Yonezawa, M. Goto, N. Yamada, M. Higashi, and M. Nomoto, "Expression profiles of MUC1, MUC2, and MUC4 mucins in human neoplasms and their relationship with biological behavior," *Proteomics*, vol. 8, no. 16, pp. 3329-3341, 2008.

[12] I. Koscinski, S. Viville, N. Porchet, A. Bernigaud, F. Escande, A. Defossez, and M.-P. Buisine, "MUC4 gene polymorphism and expression in women with implantation failure," *Human Reproduction*, vol. 21, no. 9, pp. 2238-2245, 2006.

[13] J. K. Sheehan, D. J. Thornton, M. Somerville, and I. Carlstedt, "The structure and heterogeneity of respiratory mucus glycoproteins," *American Review of Respiratory Disease*, 2012.

[14] S. NOLLET, N. MONIAUX, J. MAURY, D. PETITPREZ, P. DEGAND, A. LAINE, N. PORCHET, and J.-P. AUBERT, "Human mucin gene MUC4: organization of its 5′-region and polymorphism of its central tandem repeat array," *Biochemical Journal,* vol. 332, no. 3, pp. 739-748, 1998.

[15] K. L. Carraway, G. Theodoropoulos, G. A. Kozloski, and C. A. Carothers Carraway, "Muc4/MUC4 functions and regulation in cancer," *Future oncology*, vol. 5, no. 10, pp. 1631-1640, 2009.

[16] F. Escande, L. Lemaitre, N. Moniaux, S. K. Batra, J. P. Aubert, and M. P. Buisine, "Genomic organization of MUC4 mucin gene: Towards the characterization of splice variants," *European Journal of Biochemistry*, vol. 269, no. 15, pp. 3637-3644, 2002.

[17] C. L. Hattrup and S. J. Gendler, "Structure and function of the cell surface (tethered) mucins," *Annu. Rev. Physiol.*, vol. 70, no. 1, pp. 431-457, 2008.

[18] Y. L. Lee, A. C. H. Chen, and W. S. B. Yeung, "Interaction of sperm and embryo with the female reproductive tract," in Human Reproductive and Prenatal Genetics: *Elsevier*, 2023, pp. 211-250.

[19] N. Dharmaraj, P. Chapela, M. Morgado, S. Hawkins, B. Lessey, S. Young, and D. Carson, "Expression of the transmembrane mucins, MUC1, MUC4, and MUC16, in normal endometrium and endometriosis," *Human Reproduction*, vol. 29, no. 8, pp. 1730-1738, 2014.

[20] M. J. Kwon, J. Y. Lee, E. J. Kim, E. J. Ko, C. S. Ryu, H. J. Cho, H. H. Jun, J. W. Kim, and N. K. Kim, "Genetic variants of MUC4 are associated with susceptibility to and mortality of colorectal cancer and exhibit synergistic effects with LDL-C levels," *PLoS One*, vol. 18, no. 6, p. e0287768, 2023.

[21] H. T. M. Nguyen, M. Gunathilake, J. Lee, J. H. Oh, H. J. Chang, D. K. Sohn, A. Shin, and J. Kim, "A higher dietary alpha-linolenic acid intake is associated with lower colorectal cancer risk based on MUC4 rs2246901 variant among Korean adults.: alpha-linolenic acid intake and the colorectal cancer risk," *Nutrition Research*, 2024.

[22] M. J. Kwon, J. Y. Lee, E. J. Kim, E. J. Ko, C. S. Ryu, H. J. Cho, H. H. Jun, J. W. Kim, and N. K. Kim, "Genetic variants of MUC4 are associated with prevalence and mortality of colorectal cancer and exhibit synergistic effects with LDL-C levels," 2022.

[23] N. F. Goodman, R. H. Cobin, W. Futterweit, J. S. Glueck, R. S. Legro, and E. Carmina, "American Association of Clinical Endocrinologists, American College of Endocrinology, and androgen excess and PCOS society disease state clinical review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome-part 1," *Endocrine Practice*, vol. 21, no. 11, pp. 1291-1300, 2015.

[24] M. I. Cedars, "Is it time to revisit Rotterdam?," *Fertility and Sterility*, vol. 117, no. 4, pp. 696-697, 2022.

[25] R. Azziz, "How polycystic ovary syndrome came into its own," *F&S Science*, vol. 2, no. 1, pp. 2-10, 2021.

[26] I. H. D. A.-J. A. M. S. Almohaidi, Ibrahim Jasim Hammadi Al-Janabi, "The rs1611770 SNP polymorphism and Expression of mucin one In Infertile Females in Baghdad," *Baghdad Science Journal*, 2024.

[27] R. A. Majeed, A. F. Shihab, and A. H. AL-Assei, "The Association of T45G Polymorphism in the Adiponectin Gene with Some Hormonal Parameters in Iraqi Women with Polycystic Ovary Syndrome," *Medico-legal Update*, vol. 20, no. 2, p. 775, 2020.

[28] M. Sharief, "Correlation of Estrogen Receptor Alpha Serum Level with Gene Polymorphism and Its Effect on Women with Unexplained Infertility, Basra, Iraq," *Archives of Razi Institute*, vol. 78, no. 2, p. 775, 2023.

[29] J.-H. Kim, H.-S. Park, J.-Y. Lee, E.-J. Ko, Y.-R. Kim, H.-Y. Cho, W.-S. Lee, E.-H. Ahn, and N.-K. Kim, "Association study between mucin 4 (MUC4) polymorphisms and idiopathic recurrent pregnancy loss in a Korean population," *Genes*, vol. 13, no. 6, p. 937, 2022.

[30] K. Neykova, V. Tosto, I. Giardina, V. Tsibizova, and G. Vakrilov, "Endometrial receptivity and pregnancy outcome," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 35, no. 13, pp. 2591-2605, 2022.

[31] S. F. Mirza, R. H. Saeed, A. M. S. Almohaidi, and I. J. H. Al-Janabi, "Evaluation mucin 1 polymorphism and expression with infertility in Iraqi females," *RES MILITARIS*, vol. 12, no. 2, pp. 6916-6927, 2022.

[32] S. Rodriguez, T. R. Gaunt, and I. N. Day, "Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies," *American Journal of Epidemiology*, vol. 169, no. 4, pp. 505-514, 2009.

[33] P. W. Hedrick, Genetics of populations. *Jones & Bartlett Publishers*, 2009.

[34] C. Y.-Y. Chang, H.-W. Chang, C.-M. Chen, C.-Y. Lin, C.-P. Chen, C.-H. Lai, W.-Y. Lin, H.-P. Liu, J. J.-C. Sheu, and F.-J. Tsai, "MUC4 gene polymorphisms associate with endometriosis development and endometriosis-related infertility," *BMC Medicine*, vol. 9, pp. 1-10, 2011.

[35] P. J. Chapela, R. R. Broaddus, S. M. Hawkins, B. A. Lessey, and D. D. Carson, "Cytokine stimulation of MUC4 expression in human female reproductive tissue carcinoma cell lines and endometrial cancer," *Journal of cellular biochemistry*, vol. 116, no. 11, pp. 2649-2657, 2015.

[36] Z. Zhang, J. Wang, J. He, Z. Zheng, X. Zeng, C. Zhang, J. Ye, Y. Zhang, N. Zhong, and W. Lu, "Genetic variants in MUC4 gene are associated with lung cancer risk in a Chinese population," *PloS one*, vol. 8, no. 10, p. e77723, 2013.

[37] J. Liu, R. Xing, J. Shao, and S. Jiao, "Relationship between MUC4 variants and metastatic recurrence in colorectal cancer," *International Journal of General Medicine*, pp. 5077-5087, 2023.

[38] E. M. Egashira, A. B. Trovo-Marqui, S. C. Tanaka, and M. T. Cintra, "Investigation of biomarkers in Endometriosis-associated infertility: Systematic Review," *Anais da Academia Brasileira de Ciências*, vol. 94, no. Suppl 3, p. e20211572, 2022.

[39] E. Lacunza, J. Bara, A. Segal-Eiras, and M. V. Croce, "Expression of conserved mucin domains by epithelial tissues in various mammalian species," *Research in veterinary science*, vol. 86, no. 1, pp. 68-77, 2009.

[40] J.-P. Audie, D. Tetaert, P. Pigny, M.-P. Buisine, A. Janin, J.-P. Aubert, N. Porchet, and A. Boersma, "Mucin gene expression in the human endocervix," *Human Reproduction*, vol. 10, no. 1, pp. 98-102, 1995.

[41] E. Ignatieva and E. Matrosova, "Disease-associated genetic variants in the regulatory regions of human genes: mechanisms of action on transcription and genomic resources for dissecting these mechanisms," *Vavilov Journal of Genetics and Breeding*, vol. 25, no. 1, p. 18, 2021.

[42] E. Antontseva, A. Degtyareva, E. Korbolina, I. Damarov, and T. Merkulova, "Human-genome single nucleotide polymorphisms affecting transcription factor binding and their role in pathogenesis," *Vavilov Journal of Genetics and Breeding*, vol. 27, no. 6, p. 662, 2023.

[43] J.-H. Kim, T.-H. Kim, Y.-S. Kim, W.-C. Jang, A. Ryu, J.-Y. Hwang, and H.-H. Lee, "Mucin gene polymorphisms are associated with endometriosis in Korean women," *Archives of Gynecology and Obstetrics*, vol. 301, pp. 801-807, 2020.

[44] Hettiarachchi, G. and Komar, A. A. (2022). GWAS to identify SNPs associated with common diseases and individual risk: Genome-Wide Association Studies (GWAS) to identify SNPs associated with common diseases and individual risk. In Single Nucleotide Polymorphisms: Human Variation and a Coming Revolution in Biology and Medicine, (pp. 51-76). Cham: *Springer International Publishing.*

[45] M. Qiao, H. Zhang, Y. Xue, and L. Yang, "Relationship between MUC17 Gene Polymorphisms and Endometriosis in Central Plains Chinese Women," *Clinical and Experimental Obstetrics & Gynecology*, vol. 49, no. 10, p. 225, 2022.

[46] J. Djurovic, G. Stamenkovic, J. Todorovic, N. Aleksic, and O. Stojkovic, "Polymorphisms and haplotypes in VDR gene are associated with female idiopathic infertility," *Human Fertility*, vol. 23, no. 2, pp. 101-110, 2020.

[47] M. Farsimadan, F. Moammadzadeh Ghosi, S. Takamoli, and H. Vaziri, "Association analysis of KISS1 polymorphisms and haplotypes with polycystic ovary syndrome," *British Journal of Biomedical Science*, vol. 78, no. 4, pp. 201-205, 2021.

[48] U. R. Budihastuti, D. Dasuki, A. H. Sadewa, and T. Utoro, "Endometrial receptivity defects MUC-1 and COX-2 polymorphisms in endometriosis," *Journal of Medicine and Life*, vol. 16, no. 10, p. 1503, 2023.

[49] N. H. Ismail, I. A. Abdulhassan, and A. H. M. Al-Faisal, "The Association of TET2 Gene Polymorphisms (rs34402524 and rs2454206) and their Haplotypes with Response to Treatment in Chronic Myeloid Leukemia Patients," *Iraqi Journal of Science*, 2024.

[50] S. A. Hussein and R. H. Al-azzawi, "Single Nucleotide Polymorphism of Interleukin-28β Subunit Genes Predict Host Susceptibility to Hepatitis C virus (HCV) Infection among Iraqi Patients," *Iraqi Journal of Science*, pp. 2385-2396, 2024.