J Genet Resour 2023;9(1): 59-68 *Homepage: http://sc.journals.umz.ac.ir/*

RESEARCH ARTICLE *DOI: 10.22080/jgr.2023.24736.1340*

J|G|R

COL12A1-rs970547, Mutant Allele Advantage by Being Protective Against Recurrent Spontaneous Abortion

Abasalt Hosseinzadeh Colagar[1*], Mohammadmahdi Arjmand[1], Mohammadkazem Heydari[1], and Gholam Ali Jorsaraei[2]

[1] Department of Molecular and Cell Biology, Faculty of Science, University of Mazandaran, Babolsar, Mazandaran, Iran

[2] Fateme Zahra Infertility and Health Reproductive Research Center, Babol University of Medical Sciences, Babol, Mazandaran, Iran

ARTICLE INFO

Article history:
Received 19 August 2022
Accepted 09 October 2022
Available online 29 October 2022

Keywords:
CD44
COL12A1
MMPs
Physical interaction network
RSA

Corresponding authors:
✉ A Hosseinzadeh Colagar
acolagar@yahoo.com;
ahcolagar@umz.ac.ir

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Recurrent spontaneous abortion (RSA) is one of the causes of infertility following fetus creation. This can lead to the reduction of the fertility rate in various populations. Collagens' structure, their ratios to each other, and their connections to various receptors are some of the key players in successful fetus implantation. Among them, COL12A1 has a special role because of its very high expression in the uterus. Also, the CD44 protein as a cell adhesion molecule which is a receptor for some of the collagens plays a significant role in RSA. The aim of the study is to investigate the association of *CD44-rs13347* and *CoL12A1-rs970547* with RSA in a case-control base, and the impacts of COL12A1-rs970547 on protein structure. To genotype single nucleotide polymorphisms (SNPs) of the genes, the PCR-RFLP method was performed on the 124 RSA and 124 control samples. The results were analyzed by the binary-logistic regression method (p-value≤0.05). The SNPs' effect on the proteins' structure was analyzed by PSIPRED, HOPE, LOMET, and chimera-USF. Proteins signaling pathway and physical interaction between COL12A1 and CD44 were investigated by KEGG-pathway and GeneMANIA, respectively. Results showed that TT (P= 0.032) genotype of *CD44-rs13347* increased the risk of RSA while the CT (P= 0.027) genotype of CD44-rs13347, TT (P= 0.044) genotype, and T (P= 0.019) mutant allele of *COL12A1-rs970547* decreased the risk of RSA. Moreover, 3D structures investigation indicated that *COL12A1-rs970547* may affect the structure of COL12A1 and its interaction with Integrins. The analysis of the signaling pathway and proteins' physical interaction network also revealed the interaction of COL12A1 and CD44 with MMP2 and MMP9. On this base, we recommen that T allele of *COL12A1-rs970547* has a protective feature against RSA, especially in homozygous form by improving their interaction with Integrins and probably MMPs, too. On the other hand, the *CD44-rs13347* probably has an indirect influence on the attachment of the fetus to the extracellular matrix by affecting the MMPs and finally leading to a greater risk of RSA.

Please cite this paper as: Hosseinzadeh Colagar A, Arjmand M, Heydari M, Jorsaraei GhA. 2023. COL12A1-rs970547, mutant allele advantage by being protective against recurrent spontaneous abortion. *J Genet Resour* 9(1): 59-68. doi: *10.22080/jgr.2023.24736.1340*

Introduction

Recurrent spontaneous abortions (RSAs), which are a common concern and complication of pregnancy, refer to the loss of fetuses in consecutive pregnancies before the 20th week for of least three times (Larsen *et al*., 2013). About 1% to 2% of women experience RSA, and 50% of them are caused by reasons such as uterine abnormalities, chromosomal abnormalities, immune system problems, metabolism disorders, infections, lifestyle, and

genetic factors (Andersen *et al*., 2012; Hyde and Schust, 2015; Kaur *et al*., 2016). Genetic factors involved in RSA include gene variation and the changes in structure and expression of some genes such as *GSTT1*, *MTHFR*, *NOS3*, *VEGFA* (Gu *et al*., 2021), *CDH I* (Yang *et al*., 2017), and *CD44* (Jacques *et al*., 1993). Previous studies have shown that the abnormal expression of collagen in the fetus can be associated with RSA. Expression of Collagen type IV, which is an important ligand for Cluster of differentiation 44 (CD44) and type V, was decreased in RSA (Fu *et al.*, 2014). MMP-2 and MMP-9 have an important role in endometrial tissue remodeling during pregnancy. They are thought to play a crucial role in pregnancy as they can degrade components of the ECM thereby facilitating trophoblast migration and angiogenesis (Lin *et al*., 2020). These two MMPs are associated with various types of collagens including COL12A1. They also have a connection with CD44 (Vastrik *et al*., 2007).

The COL12A1 protein is a type of fibril-associated collagen with an interrupted triple helix that can play an important role in cell binding. This protein may also be involved in cell proliferation, growth, and survival. The *Col12A1* gene is on chromosome 6 at position q13-q14.1 with a length of 121728bp (Gerecke *et al.*, 1997) which is expressed in the extracellular matrix (Kalmari *et al*., 2022; Li *et al*., 2022).

Three isoforms are expressed under the process of alternative splicing of *Col12A1*, among which isoform 2 was detected in the placenta (Wessel *et al.*, 1997). Due to the important role of cellular adhesion genes in the attachment of the fetus to the extracellular matrix (uterus), defects or changes in the related genes such as *CD44* and *Col12A1* can be important.

CD44 protein is a cell surface glycoprotein receptor that is located on chromosome 11 (11p13) with a length of about 93533 bp. This gene has 11 isoforms, which are generated from alternative splicing of 10 exons. The CD44 expression, involved in cell-cell and cell-matrix adhesion, began in the placenta after 16 weeks of gestation (Marzioni *et al*., 2001). Its expression plays a significant role in RSA as a cell-adhesion factor (Hu *et al*., 2018). Important extracellular matrix ligands for this protein receptor are hyaluronic acid (HA), osteopontin, matrix metalloproteinases (MMPs), and collagens (Goodfellow *et al*., 1982; Stamenkovic *et al*., 1989; Jackson *et al*., 1992; Konig *et al*., 1998; Goodison *et al*., 1999).

A previous study showed that *COL12A1* has two missense SNPs rs970547 and rs240736 with minor allele frequency (MAF) ≥ 0.1. Among them, *COL12A1-rs970547* was more important and effective based on *in silico* study (Kalmari *et al*., 2022). Among SNPs of CD44, the rs13347 is an important SNP because of its role in CD44 gene expression. This SNP is located in 3'UTR of this gene. Due to the regulatory role of the 3'UTR region in the function of genes, *CD44*-rs13347 was selected for the study, too. The aim of this study is to investigate the association of Col12A1- rs970547 and CD44- rs13347 with RSA in the Mazandaran province (Iran) population.

Materials and Methods

Blood sample collection

In the case-control study, blood samples were collected from 124 cases that had 18 to 42 years old with at least three times of RSA before the 20th week of pregnancy. The blood of 124 samples that had 18 to 42 years old with at least 2 live deliveries without a history of pathology in the reproductive period was also selected as the control group. The minimum number of samples for this study is calculated based on the prevalence of the disease (Daniel *et al.*, 1999), which is measured by the standard biostatistics' equation: $n = ([Z^2P(1-P)] \div d^2)$.

In this equation, P is abortion's prevalence (0.7%). The n, Z, and d symbols are minimum number, confidence statistic (95%), and degree of confidence (0.05 and 0.01), respectively. The minimum number of case samples for our study is 75.81 (~76). All of the blood samples were collected from Qaemshahr and Babol hospitals in Mazandaran province (Iran) from the year 2018 to 2022. This study was approved by the Ethics Committees of the University of Mazandaran (Offline code: #IR.UMZ.REC.1397.052, which was later registered online: IR.UMZ.REC.1400.052). All of the study's subjects signed an informed consent before entering the study.

RFLP and genotyping

Genomic DNA was extracted from white blood cells (WBC) by the salting-out method from all samples (MWer *et al*., 1998). Designed primers with Oligo 7 software (Table 1) were synthesized by SinaColon (SinaColon Co., Iran), and then PCR amplification was done by a thermal cycler (Eppendorf Co, Germany). It was performed by PCR MasterMix 2X (Cat NO: YT1551, Yekta Tajhiz Azma Co, Iran) in the final volume of 25 µl, which consisted of 10 pM of each primer and 10-100 pM DNA template.
For *Col12A1*- rs970547, 2.5 µl of PCR products digested with 1 µl of Alu I restriction enzyme (#FD0014, ThermoFisherScientific C0, UK) and 10X FastDigest buffer at 37°C for 15min. Although rs13347 and rs970547 are introduced as triallelic (C>A, G, T) variations in NCBI, the A and G frequencies are 0.000 and 0.000, both globally and in the Asian population. This shows that these two were probably very rare mutations. Therefore, these two SNPs are considered biallelic (C>T) polymorphisms. Digested fragments with *Alu I* were electrophoresed in 1% agarose gel, and then stained by 1µg/ml ethidium bromide stock solution (Agha Heydar Ali Naghash *et al*., 2022). The *Alu I* digestion of PCR products produced 419 bp, 266 bp, and 153 bp fragments. The *Alu I* restriction enzyme digested fragments of the PCR products to "153, 266" and 419bp, which are T and C alleles, respectively (Fig. 1A1). The PCR products of *CD44*-rs13347 flanking were digested with 1 µl of the *Hin1 II* restriction enzyme (#FD1834, ThermoFisherScientific C0, UK) as described above manual. Digested fragments of *Hin1 II* are 206, 359, and 565 bp, which are T and C alleles, respectively (Fig. 1B1). For final confirmation of genotypes, three samples from each genotype of case and control were selected randomly and sequenced (GenFanAvaran Co., Iran).

Table 1. PCR primer sequences of *Col12A1-rs970547* and *CD44-rs13347*.

Gene name	SNP ID	Primer name	Primer sequences (5'→3')	aTm (ºC)	PCR Prod.
Col12A1	rs970547	F C12A1	F: AGAATCCAGAACAGGTCCAC	59	417 bp
		R C12A1	R: CCAAGAGAAGTAAGGGGAGG		
CD44	rs13347	F CD44	F: CTGTTGTAGTCCCTCACTTGG	59	565 bp
		R CD44	R: TTCCTCTCTCCTACTCCTCTG		

aTm= annealing Tm; PCR Prod= PCR products

Statistical analysis between genotyping case and control was done by logistic regression analysis (SPSS Statistics ver.16.0). This study was calculated by odds ratio (OR) with 95% confidence interval (95% CI) and a p-value of less than 0.05, which was considered statistically significant.

Physical interaction and structural analysis

For this study, single nucleotide polymorphism (SNP) selection confirmed by NCBI databank (https://www.ncbi.nlm.nih.gov/), and SNP effect on the 2D structure of protein analyzed by PSIPRED (http://bioinf.cs.ucl.ac.uk/PSIPRED/) were utilized. HOPE (https://www3.cmbi.umcn.nl/hope/) and LOMETS (https://zhanggroup.org/LOMETS/) were also used to analyze and predict the 3D structure of the protein. Since LOMETS gives only one best-predicted structure for each sequence, there was no need for an omnibus test or other tests that are needed for the selection or prediction of the best-predicted structure. Because HOPE and LOMETS are limited in the number of amino acids, the 3D structure was studied from amino acids 2241 to 3118). In addition, 3D structure design and assessment were done by Chimera USF software. Moreover, to investigate proteins pathway and physical interaction prediction, KEGG pathway database (https://www.genome.jp/kegg/pathway.html) and GeneMANIA (https://genemania.org/) were used.

Result

SNPs genotyping by RFLP

The RFLP profile on 1% agarose gel showed that the *COL12A1-rs970547* profile has three fragments: 153, 266, and 417bp were CT; 153 and 264bp were TT; and fragments without a digest (a 417bp) were CC genotypes (Fig. 1A2). Additionally, it demonstrated that *CD44-rs13347*

profile has also three fragments: 206, 359, and 565bp were CT; 206 and 359bp were TT; and fragments without a digest (a 565bp) were CC genotypes (Fig. 1B2).

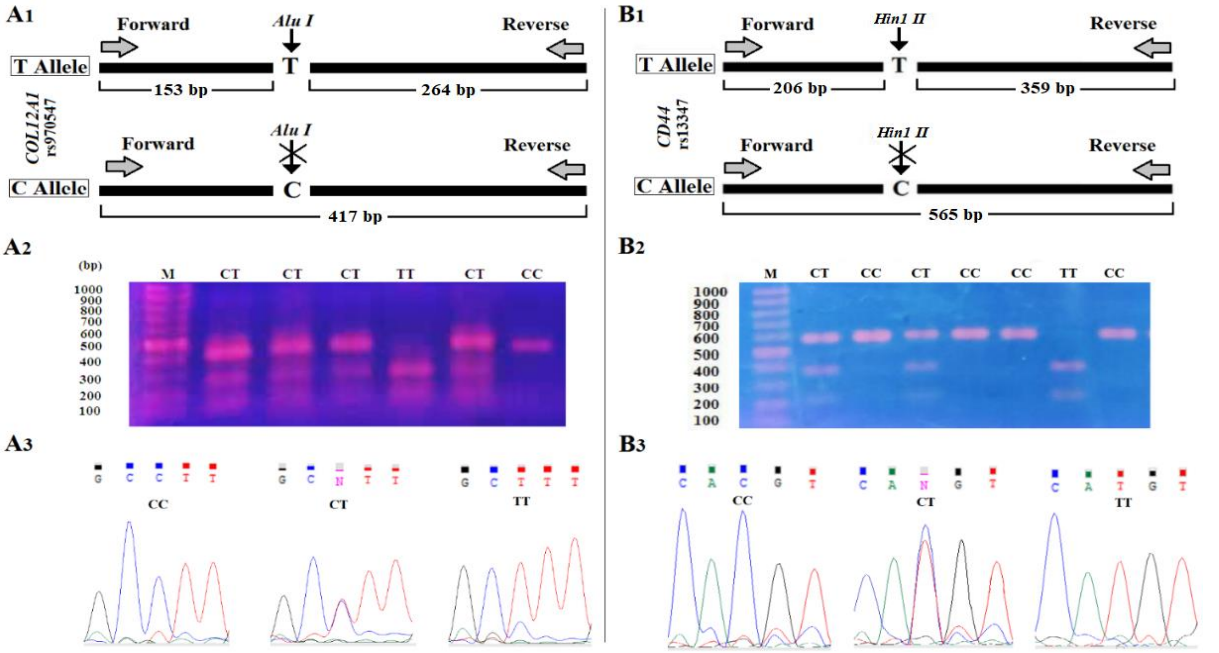

Fig. 1. RFLP agarose gel profiles, and DNA sequencing of *COL12A1-rs970547* and *CD44-rs13347* SNPs: A1 and B1) RFLP maps of SNPs by digestion of PCR products by *Alu I* and *Hin1 II* restriction enzymes; A2 and B2); RFLP profiles of digested PCR products by *Alu I* and *Hin1 II* in the 1% agarose gel electrophorese of *COL12A1-rs970547* and *CD44-rs13347*, respectively. CC, CT, and TT lanes indicate genotypes, M= 100-3000bp DNA marker (SinaColon Co, Iran); A3 and B3) chromatograms of PCR direct-sequencing from *COL12A1-rs970547* and *CD44-rs13347*genotypes, respectively.

All of these genotypes were confirmed by PCR direct-sequencing (Fig. 1 A3 and B3). The results of our study showed that the TT homozygous genotype in patients is 16.13%, while the frequency of the same genotype in the control group is 36.29%, which is more than twice (p-value equal to 0.044) in patients. In addition, the T allele in the group with RSA is significantly lower than the control group (p-value equals to 0.019). These results show that the T allele, especially in the homozygous, is a protective form against RSA. Moreover, TT and CT genotypes of the *CD44-rs13347* are associated with RSA (*P*= 0.032 and 0.027, respectively). Other genotypes and alleles of SNPs were not associated with this disease (Table 2).

Table 2. Genotype and allele frequencies of *COL12A1*-rs970547 and *CD44*-rs13347.

Gene name	Genotype	Control (n=124)	Case (n=124)	*p*-value: OR (95% CI) *
Col12A1	**rs970547 C>T**			
	CC	17(13.71%)	18(14.52%)	Ref
	CT	62(50%)	86 (69.35%)	0.474: 1.310(0.626-2.743)
	TT	45(36.29%)	20 (16.13%)	**0.044**: 0.420(0.180-0.979)
	CT+TT	107(86.29%)	106(42.74%)	0.855: 0.936(0.458-1.913)
	C- allele	96(38.71%)	122(49.20%)	Ref
	T- allele	152(61.29%)	126(50.80%)	**0.019**: 0.652(0.457-0.932)
CD44	**rs13347 C>T**			
	CC	65(52.42%)	74(59.68%)	Ref
	CT	55(44.35%)	34(27.42%)	**0.027:** 0.543(0.316-0.934)
	TT	4(3.23%)	16(12.90%)	**0.032:** 3.514(1.118-11.043)
	CT+TT	59(96.77%)	50(87.10%)	0.250: 0.744(0.450-1.231)
	C-allele	185(74.6%)	182(73.39%)	Ref
	T-allele	63(25.4%)	66(26.61%)	0.759: 1.065(0.713-1.591)

*OR: Odds ratio; CI: Confidence interval.

Physical interaction and structural analysis

The study of two- and three-dimensional (2D, 3D) structures was performed by PSIPRED, LOMET, HOPE web-tools, and chimera USF software. Analysis of the 2D structure by PSIPRED showed that the replacement of the serine instead of glycine residue (G3058S) in *COL12A1-rs970547* could change the 2D structure. The changes occurred in 15 motifs, most of which are related to changes in the coil-to-strand structure (Fig. 2).

The HOPE showed G3058S caused an enlargement in the size of mutant residue compared to the wild type. In addition, it might lead to bumps and changes in torsion angles that are unusual. On the other hand, glycine is flexible to make correct torsion angles but its substitution with another amino acid results in an incorrect conformation and disturbs the local structure.

The LOMET and chimera USF software were used to design and analyze the 3D structure since HOPE did not design the 3D structure. The result of LOMET showed that *COL12A1-rs970547* changed the number and size of domains as the number of domains increased in the mutant residue of COL12A1 compared to the wild type (Fig. 3A1 and B1). FU-score, the quality value of the discontinuous domain, altered in the mutant variant in 500 to 878 residues (Fig. 3A2 and B2). The effect results of 15 changed 2D motifs in designed 3D structure by chimera showed which G3058S led to A2907 and V2908, in the non-helical domain, replacing near mutated residue S3058 (Fig. 3A3 and B3).

The results of the KEGG pathway database showed that by binding to different cell surface proteins, collagens can be involved in different pathways such as integrins, Syndecan, GPVI, and CD44 (Fig. 4A). Investigation of physical interaction between CD44 and COL12A1 proteins by GeneMANIA free web-tool showed these two proteins were not interacting together directly, but interacted with 20 other proteins directly and indirectly (Fig. 4B). In addition, CD44 and COL12A1 directly interacted with 16 (SLC3A2, SLC9A1, VCAN, SELE, HYAL2, IGFBP3, ABCB5, ABCC5, SPP1, MMP2, MMP7, MMP9, CD9, COL14A1, PDPN, and ITGF4) and five (ID1, MMP2, MMP7, MMP9, and COL11A1) proteins, respectively (Fig. 4C and D). Moreover, the results indicated that CD44 and COL12A1 together physically interacted with MMP2, 7, and 9 (Fig. 4C and D).

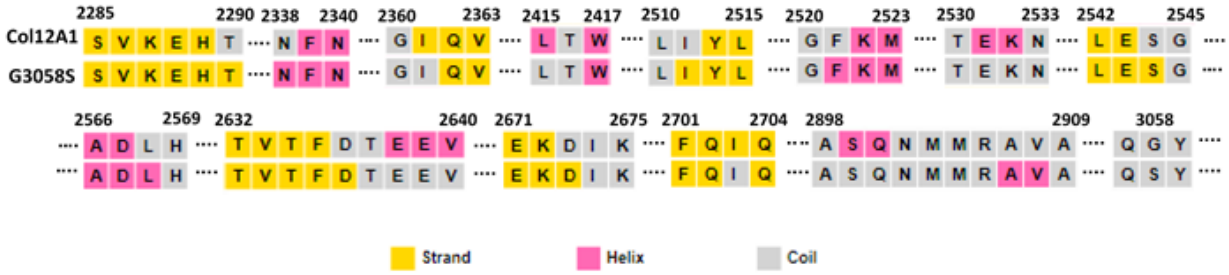

Fig. 2. Effect of *COL12A1-rs970547* on the 2D structure of proteins by PSIPRED web-tool: Glycine as a wild type is at position 3058. conversion of the glycine residue to serine in rs970547 altered the second structure of the COL12A1 protein in 15 motifs, including 5 conversions of the coil to strand, 4 conversions of the coil to helix, 4 conversions of the helix-to-coil, and 2 conversions of the strand to coil structure.

Discussion

Based on this study, the T allele of COL12A1-rs970547, especially in the homozygous, is a protective form against RSA. In CD44-rs13347, even though the CT genotype in the control group is more than the RSA group, the TT genotype showed a significant association with RSA.

According to previous studies, the COL12A1 expression is remarkably high in the uterus, which makes it an important gene factor in RSA (Alegina *et al*., 2016). Additionally, Aberkane et al. (2018) asserted that the correlation of COL12A1 has a significantly low expression in fetus without connection to trophoblast compared to the fetus with weak connection.

Therefore, the alterations leading to lower expression of COL12A1 could be related to higher RSA occurrence.

Previous *in silico* analysis showed that the *COL12A1*-rs970547 as a missense SNP is deleterious and damaging (Kalmari *et al*., 2022) while the global MAF of the T allele is 0.8, which is the opposite of being deleterious. This exhibits that at some point T allele gains an advantage over the ancestral allele. Probably, this advantage is related to its protective association against RSA compared to the C allele. Previous researchers predicted the association of rs1107946-COL1A1 with RSA. On the other hand, collagen XII is related to collagen I and plays a prominent role in the function of this collagen (Alegina *et al*., 2016) by which rs970547-COL12A1 can play its role against RSA.

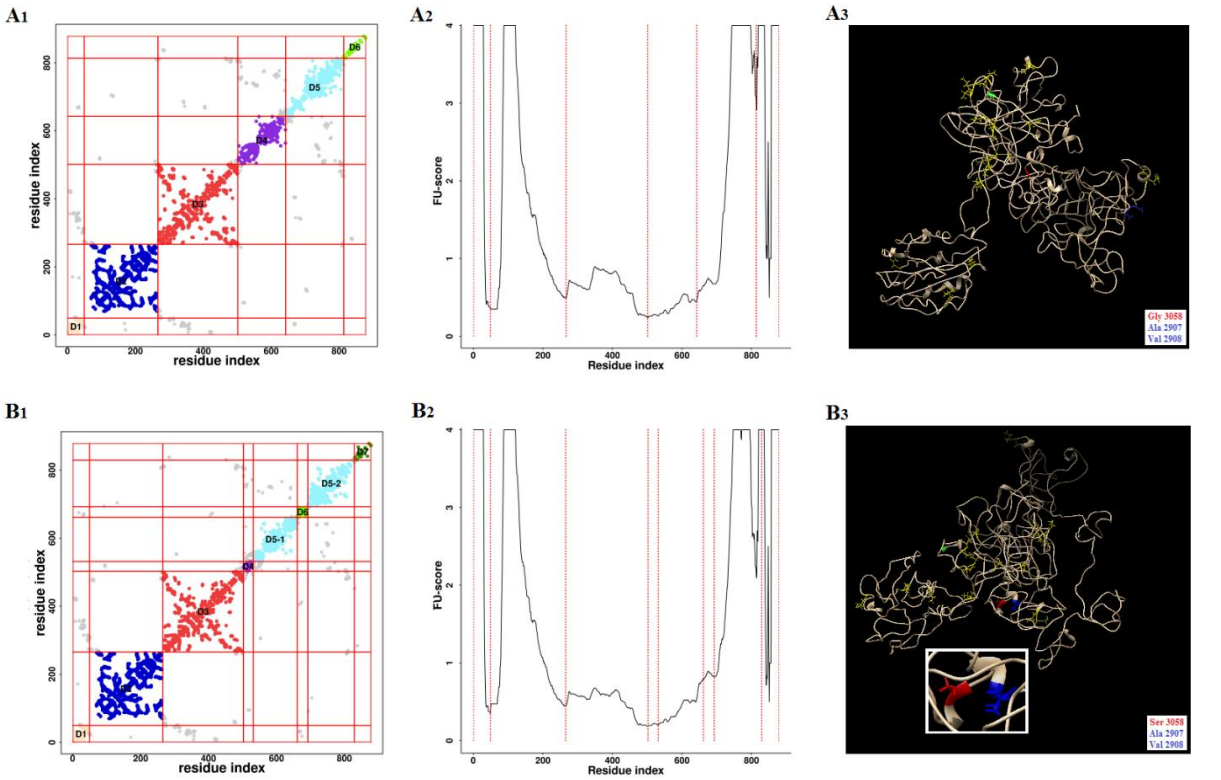

Fig. 3. Effect of *COL12A1-rs970547* on the 3D structure of a protein by LOMET: A1 and B1) The analysis of glycine 3058, as wild type amino acid, in LOMET online server showed this variant has 6 domains (A1) whereas this amino acid substitution with serine created 8 domains (B1), D= Domain; A2 and B2) FU-score analysis of LOMET server displayed substitution of serine with glycine in 3058 positions causing FU-score to change for wild type (A2) than mutate variant in 500- 878 residue from D4 to D7 (B2). A3 and B3) Analysis of LOMET and HOPE results by chimera software indicated mutated amino acid inflexible, enlargement in size, and changes in torsion angles (A3) than wild type amino acid (B3) in 3D structure. It showed residues of A2907 and V2908 (blue color) are farther away from G3058 (red color) in the wild-type variant (A3) than S3058 (red color) as the mutant variant (B3), the yellow color is 2D changed motifs.

Our finding in 2D and 3D protein structure indicated that *COL12A1*-rs970547 converts glycine to serine and changes protein structure since glycine is more flexible and has a smaller size than serine. This study displayed that due to the alteration of Glycine 3058 by Serine, its placement in the 3D structure has changed and come closer to the A2907-V2908 motif which is one of the 15 altered motifs in the 2D structure. A2907-V2908 motif located in the non-helical domain that is essential for triplex assembly in COL12A1 as fibrillar collagen. This region, which according to the results of this study goes under changes in the 3D structure, has the important sequence of RGD (amino acids 2895 to 2897). Based on previous studies, RGD is a connective sequence by which collagens are binding to the surface of the cells by connecting

to Integrins (Barczyk *et al.*, 2010). Therefore, one of the reasons for the protective effect of rs970547C>T can be caused by changes in COL12A1 and Integrin connection. Increasing the adhesion of Integrins to the extracellular matrix improves the successful implantation of the fetus (Smart and Riley, 2008). Therefore, the alteration of rs970547-COL12A1 may reduce RSA by increasing its binding to integrins.

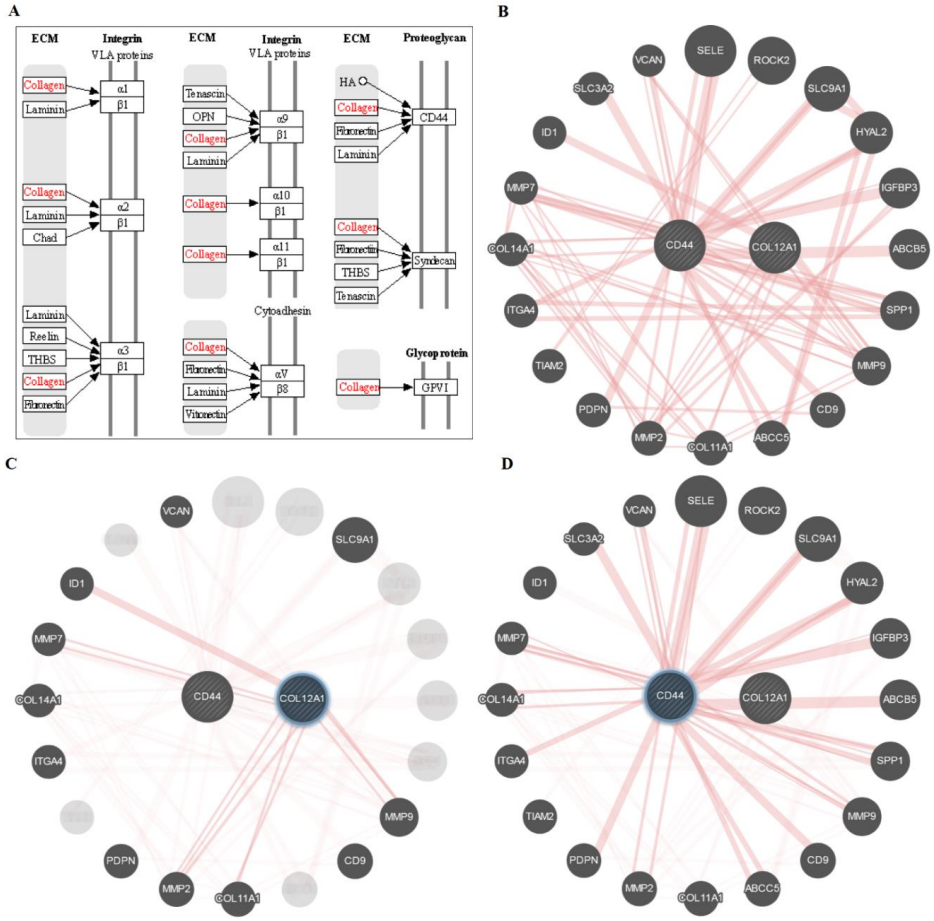

Fig. 4. *Col12A1* and *CD44* gene pathway and proteins interaction of KEEG pathway and GeneMANIA web-tools. A) Result of Collagens gene pathway of KEEG pathway; B) Physical proteins network interaction of GeneMANIA showed *Col12A1* and *CD44* gene correlated with the 20 genes to directly and indirectly: *CD9*, *CD44*, *COL11A1*, *COL12A1*, *COL14A1*, *MMP2*, *MMP7*, *MMP9*, *ABCB5*, *ABCC5*, *VCAN*, *SELE*, *ROCK2*, *HYAL2*, *IGFBP3*, *SPP1*, *PDPN*, *TIAM2*, *ITGA4*, *ID1*, *SLC3A2*, and *SLC9A1*; C)The physical proteins network interaction of *Col12A1* and *CD44* showed that *CD44* is directly related to four genes: *COL12A1*, *VCAN*, *ITGA4*, and *HYAL2*; D) Also, the physical proteins network interaction showed that the *Col12A1* gene is related to *SLC9A1*, *CD44*, *CD9*, *COL11A1*, *PDPN*, *COL14A1*, *VCAN*, *MMP2*, and *MMP9*, directly. *CD9*= Cluster of differentiation 9, *CD44*= Cluster of differentiation 44, *COL11A1*= Collagen type XI alpha 1 chain, *COL12A1*= Collagen type xii alpha 1 chain, *COL14A1*= Collagen type XIV alpha 1 chain, *MMP2*= Matrix metallopeptidase 2, *MMP7*= Matrix metallopeptidase 7, *MMP9*= Matrix metallopeptidase 9, *ABCB5*= ATP binding cassette subfamily B member 5, *ABCC5*= ATP binding cassette subfamily C member 5, *VCAN*= Versican, *SELE*= Selectin E, *ROCK2*= Rho-associated coiled-coil containing protein kinase 2, *SLC9A1*= Solute carrier family 9 member A1, *HYAL2*= Hyaluronidase 2, *IGFBP3*= Insulin-like growth factor binding protein 3, *SPP1*= Secreted phosphoprotein 1, *PDPN*= Podoplanin, *TIAM2*= TIAM Rac1, associated GEF 2, *ITGA4*= Integrin subunit alpha 4, *ID1*= Inhibitor Of DNA binding 1, *SLC3A2*= Solute carrier family 3 member 2.

According to the type of association of rs970547C>T-COL12A1 with RSA and its role in the 3D structure of the protein, all proteins that are directly or indirectly related to COL12A1 can be affected by it. In general, one of the receptors of the collagen family is CD44. The *CD44* gene, known as stem cells marker (SCs), has various roles such as cell-cell interactions and cell adhesion, which is expressed in the placenta (Choi *et al*., 2006). The CD44 is a cell surface glycoprotein that could attach to many extracellular matrix components such as collagens. Previous studies showed that CD44 is related to RSA (Zhu *et al*., 2013) since the localization of CD44 in the endometrial epithelium and stroma raises the possibility of trophoblastic implantation (Aplin, 1991; Glasser *et al*., 1991). In addition, some studies demonstrated that changes in some CD44 variants' expression might lead to RSA by inhibiting the proliferation and migration of trophoblastic cells (Hu *et al*., 2018). The *CD44-rs13347* can modulate CD44 expression by attaching to transcription factors, long non-coding RNAs and miRNAs (Jeyapalan *et al*., 2011). It can weaken the binding of miRNAs, changing the CD44 expression, and possible alteration of CD44 function (Lou *et al*., 2014). Some researchers reported that the T allele, as a mutant allele, in *CD44-rs13347* could associate with various diseases (Qi *et al*., 2016). Our results showed that the TT genotype is related to RSA in our population. However, the heterozygote genotype is significantly higher in the control group; it seems that substitution of the T allele with the C allele in the homozygote form of *CD44-rs13347* will cause RSA by disrupting the cellular adhesion of the uterus to the fetus.

Some important proteins, such as MMP2 and MMP9, interact with both proteins. On the other hand, researchers reported in previous studies that balancing the expression and function between these two MMPs and their inhibitors causes a successful pregnancy. In the early stages of invasion, increasing the level of MMP2 is effective for successful embryo implantation, although its high level in the later stages can negatively affect the balance of vessels (Yan *et al*., 2021).

The results of the KEGG pathway confirmed that COL12A1 is related to CD44 but the nature of its binding was unknown. Based on GeneMANIA results, the physical interaction between CD44 and COL12A1 does not exist. In addition, the results indicated that both CD44 and COL12A1 physically interacted with MMP2, MMP7, and MMP9. As mentioned before, researchers reported the association of MMP2 and MMP9 proteins with RSA (Nissi *et al*., 2013; Radulovic *et al*., 2010; Jiang and Qi, 2015). Therefore, it seems that COL12A1 does not cause RSA by attaching directly to CD44 but each of COL12A1 and CD44 individually causes RSA by physical interaction with some MMPs. Thus, these interactions could cause a change in cell binding attachment in the fetus and as a result with RSA.

Conclusion

Our findings showed that the TT genotype of *CD44-rs13347*, as a cell adhesion receptor, associates with women's RSA, while the T mutant allele and TT genotype of *COL12A1-rs970547* decrease the risk of RSA, which makes it a protective allele. This effect can be due to its impact on the structure and function of COL12A1. In addition, the proteins' physical interaction network showed that COL12A1 and CD44 interact with MMP2 and MMP9, which are related to successful or unsuccessful pregnancy. On this base, we recommend that *CD44-rs13347* and *COL12A1-rs970547* play their role in fetus implantation and successful pregnancy by their influence on MMPs activity.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors appreciate all the colleagues who collaborated in this study.

Reference

Aberkane A, Essahib W, Spits C, De Paepe C, Sermon K, Adriaenssens T, ..., Van de Velde H. (2018). Expression of adhesion and extracellular matrix genes in human blastocysts upon attachment in a 2D co-culture system. *Mol Hum Reprod* 24(7): 375- 387.

Agha Heydar Ali Naghash HR, Hoshmand M, Salehzadeh A, Ansari Nejad N. 2022. The rs4415084 and rs10941679 Might be Risk Conferring Factors for Breast Cancer among Iranian Women. *J Genet Resour* 8(1): 1-6.

Alegina EV, Tetruashvili NK, Agadzhanova AA, Trofimov DY, Donnikov AE. (2016). Role of collagen gene polymorphisms in the structure of early gestation loss. *Bull Exp Biol Med* 160(3): 360-363.

Andersen AMN, Andersen PK, Olsen J, Grønbæk M, Strandberg-Larsen K. 2012. Moderate alcohol intake during pregnancy and risk of fetal death. *Int J Epidemiol* 41(2): 405-413.

Aplin JD. 1991. Implantation, trophoblast differentiation and haemochorial placentation: mechanistic evidence in vivo and in vitro. *J Cell Sci* 99: 681-692.

Barczyk M, Carracedo S, Gullberg D. (2010). Integrins. *Cell Tissue Res* 339(1): 269-280.

Choi CH, Roh CR, Kim TJ, Choi YL, Lee JW, Kim BG, Lee JH, Bae DS. 2006. Expression of CD44 adhesion molecules on human placentae. *Eur J Obstet Gynecol* 128(1-2): 243-247.

Daniel WW. 1999. Biostatistics: a foundation for analysis in the health sciences. 7th ed. John Wiley & Sons, New York, USA.

Fu Q, Tao Y, Piao H, Du MR, Li DJ. 2014. Trophoblasts and decidual stromal cells regulate decidual NK cell functions via interaction between collagen and LAIR-1. *Am J Reprod Immunol* 71: 368-78

Gerecke DR, Olson PF, Koch M, Knoll JH, Taylor R, Hudson DL, …, Burgeson RE. 1997. Complete primary structure of two splice variants of collagen XII, and assignment of α1 (XII) collagen (COL12A1), α1 (IX) collagen (COL9A1), and α1 (XIX) collagen (COL19A1) to human chromosome 6q12–q13. *Genomics* 41(2): 236-242.

Glasser SR, Mulholland J, Mani SK, Julian J. 1991. Blastocyst-endometrial relationships: reciprocal interactions between uterine epithelial and stromal cells and blastocysts. *Placenta* 5: 229-229.

Goodfellow PN, Banting G, Wiles MV, Tunnacliffe A, Parkar M, Solomon E, ..., Fabre JW. 1982. The gene, MIC4, which controls expression of the antigen defined by monoclonal antibody F10. 44.2, is on human chromosome 11. *Eur J Immunol* 12(8): 659-663.

Goodison S, Urquidi V, Tarin D. 1999. CD44 cell adhesion molecules. *Mol Pathol* 52(4): 189-196.

Gu H, Li L, Du M, Xu H, Gao M, Liu X, …, Zhong X. 2021. Key gene and functional pathways identified in unexplained recurrent spontaneous abortion using targeted RNA sequencing and clinical analysis. *Front Immunol* 12: 717832. doi: 10.3389/fimmu.2021.717832.

Hu M, Cheng YX, Yang X, Yu J, Huang J, Hong L. 2018. Dysregulation of CD44v6 may lead to recurrent spontaneous abortion by inhibiting the proliferation and migration of trophoblast cells. *Int J Clin Exp Pathol* 11(4): 2072–2079.

Hyde KJ, Schust DJ. 2015. Genetic considerations in recurrent pregnancy loss. *Cold Spring Harb Perspect Med* 5(3): a023119. doi: 10.1101/cshperspect.a023119

Jackson DG, Buckley J, Bell JI. 1992. Multiple variants of the human lymphocyte homing receptor CD44 generated by insertions at a single site in the extracellular domain. *J Biol Chem* 267(7): 4732-4739.

Jeyapalan Z, Deng Z, Shatseva T, Fang L, He C, Yang BB. 2011. Expression of CD44 3′-untranslated region regulates endogenous microRNA functions in tumorigenesis and angiogenesis. *Nucleic Acids Res* 39(8): 3026-3041.

Jiang G, Qi Y. 2015. Detection of MMP-9 and TIMP-3 mRNA expression in the villi of patients undergoing early spontaneous abortion: A report of 30 cases. *Exp Ther Med* 9(5): 1939-1943.

Kalmari A, Heydari M, Hosseinzadeh Colagar A, Arash V. 2022. In silico analysis of collagens missense SNPs and human abnormalities. *Biochem Genet* 60(5): 1630-1656.

Kaur R, Gupta K. 2016. Endocrine dysfunction and recurrent spontaneous abortion: an overview. *Int J Appl Basic Med* 6(2): 79-83.

Konig H, Ponta H, Herrlich P. 1988. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *Eur Mol Biol Organiz J* 17(10): 2904-2913.

Larsen EC, Christiansen OB, Kolte AM, Macklon N. 2013. New insights into mechanisms behind miscarriage. *BMC Med* 11(1): 1-10.

Li J, Li Z, Xu Y, Huang C, Shan B. 2022. METTL3 facilitates tumor progression by COL12A1/MAPK signaling pathway in esophageal squamous cell carcinoma. *J Cancer* 13(6): 1972-1984.

Lin C, He H, Cui N, Ren Z, Zhu M, Khalil RA. 2020. Decreased uterine vascularization and

uterine arterial expansive remodeling with reduced matrix metalloproteinase-2 and -9 in hypertensive pregnancy. *Am J Physiol Heart Circ Physiol* 318: H165-H180.

Lou F, Ma HN, Xu L, Chen M, Zhu YB. 2014. Two polymorphisms of CD44 3'UTR weaken the binding of miRNAs and associate with naso-pharyngeal carcinoma in a Chinese population. *Eur Rev Med Pharmacol Sci* 18(17): 2444-2452.

Marzioni D, Crescimanno C, Zaccheo D, Coppari R, Underhill CB, Castellucci M. 2001. Hyaluronate and CD44 expression patterns in the human placenta throughout pregnancy. *Eur J Histochem* 45(2): 131-40.

MWer S, Dykes D, Polesky H. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3): 1215.

Nissi R, Talvensaari-Mattila A, Kotila V, Niinimäki M, Järvelä I, Turpeenniemi-Hujanen T. 2013. Circulating matrix metalloproteinase MMP-9 and MMP-2/TIMP-2 complex are associated with spontaneous early pregnancy failure. *Reprod Biol Endocrinol* 11(1): 1-6.

Qi Q, Wang J, Chen A, Huang B, Li G, Li X, Wang J. 2016. Associations of five polymorphisms in the CD44 gene with cancer susceptibility in Asians. *Sci Rep* 6:39485. doi: 10.1038/srep39485

Radulovic NV, Ekerhovd E, Abrahamsson G, Norström A. 2010. Cervical tissue changes in women with miscarriage: a morphological and biochemical investigation. *Acta Obstet Gynecol Scand* 89(1): 54-64.

Smart N, Riley PR. (2008). The stem cell movement. *Circ Res* 102(10): 1155-1168.

St Jacques S, Dadi HK, Letarte M. 1993. CD44 in human placenta: localization and binding to hyaluronic acid. *Placenta* 14(1): 25-39.

Stamenkovic I, Amiot M, Pesando JM, Seed B. 1989. A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell* 56(6): 1057-1062.

Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, de Bono B, …, Stein L. 2007. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol* 8(3): R39. doi: 10.1186/gb-2007-8-3-r39.

Wessel H, Anderson S, Fite D, Halvas E, Hempel J, SundarRaj N. 1997. Type XII collagen contributes to diversities in human corneal and limbal extracellular matrices. *Investig Ophthalmol Vis Sci* 38(11): 2408-2422.

Yan Y, Fang L, Li Y, Yu Y, Li Y, Cheng JC, Sun YP. (2021). Association of MMP2 and MMP9 gene polymorphisms with the recurrent spontaneous abortion: A meta-analysis. *Gene* 767: 145173.

Yang Y, Chen X, Saravelos SH, Liu Y, Huang J, Zhang J, Li TC. 2017. HOXA-10 and E-cadherin expression in the endometrium of women with recurrent implantation failure and recurrent miscarriage. *Fertil Steril* 107(1): 136-143.

Zhu R, Wang SC, Sun C, Tao Y, Piao HL, Wang XQ, …, Li DJ. 2013. Hyaluronan-CD44 interaction promotes growth of decidual stromal cells in human first-trimester pregnancy. *PLoS One* 8(9): e74812. doi:10.1371/journal.pone.0074812.