Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* genes underlying risk of Endometriosis in Pakistani Patients

Emaan Imtiaz

00000364678

MS Healthcare Biotechnology

Supervisor:

Prof. Dr. Attya Bhatti

Atta-ur-Rahman School of Applied Biosciences (ASAB),

National University of Sciences and Technology (NUST),

H-12, Islamabad, Pakistan

2021-2023

Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* genes underlying risk of Endometriosis in Pakistani Patients

By

Emaan Imtiaz

(Fall 2021 – MS HCB – 00000364678)

Supervisor:

Prof. Dr. Attya Bhatti

A thesis submitted in partial fulfillment of the requirement of the degree of

Master of Sciences Healthcare Biotechnology

Atta-ur-Rahman School of Applied Biosciences (ASAB),

National University of Sciences and Technology (NUST),

H-12, Islamabad, Pakistan

FORM TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by:

Student Name: **Emaan Imtiaz** Reg No: **00000364678**

Titled: **Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* underlying risk of Endometriosis in Pakistani Patients** be accepted in partial fulfillment of the requirements for the award of MS degree in Healthcare Biotechnology with (A grade).

Examination Committee Members

1. Name: Prof. Dr. Muhammad Ansar Signature: _____

2. Name: Prof. Dr. Sobia Manzoor Signature: _____

3. Name: Dr. Maria Shabbir Signature: _____

Supervisor's Name: Prof. Dr. Attya Bhatti Signature: _____

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Date: 25/08/2023

Co-supervisor's name: Prof. Dr. Peter John Signature: _____

Date: 25/8/2023

Date: 07/09/2023

Head of Department

DR. HUSSAIN M. WAHEDI
Associate Professor
Deptt. of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

COUNTERSIGNED

Date: 07/09/2023

Dean/Principal

Principal
Atta ur Rahman School of Applied Biosciences (ASAB)
NUST Islamabad

Thesis Acceptance Certificate

Certified that final copy of MS thesis entitled **"Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* genes underlying risk of Endometriosis in Pakistani Patients"** submitted by Emaan Imtiaz (Registration # 00000364678), of Atta-ur-Rahman School of Applied Biosciences (ASAB), has been verified by undersigned, found complete in all respects as per NUST regulations, and is accepted as fulfillment for the award of MS degree.

Supervisor: _____

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Prof. Dr. Attya Bhatti

Atta ur Rahman School of Applied Biosciences,

National University of Sciences and Technology

Head of Department: _____

DR. HUSSAIN M. WAHEDI
Associate Professor
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

For

Prof. Dr. Sobia Manzoor

Atta ur Rahman School of Applied Biosciences,

National University of Sciences and Technology

Principal: _____

Principal
Atta ur Rahman School of Applied Biosciences (ASAB)
NUST Islamabad

Prof. Dr. Muhammad Asghar

Atta ur Rahman School of Applied Biosciences,

National University of Sciences and Technology

Dated: 07/09/2023

Certificate for Plagiarism

It is certified that MS thesis titled "**Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* Underlying Risk of Endometriosis in Pakistani Patients**" by Emaan Imtiaz has been examined:

I undertake the follows:

i. Thesis has significant new work/ knowledge as compared to already published or are under consideration to be published work. No sentence, equation, diagram, table, paragraph, or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.

ii. The work presented is original and the own work of the author (i.e., there is no plagiarism). No ideas, process, results, or words of others have been presented as own work.

iii. There is no fabrication of data or results, which have been compiled/ analyzed.

iv. There is no falsification by manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

v. The thesis has been checked using TURNITIN (copy of original report attached) and found within limits as per HEC plagiarism Policy and instructions.

Signature of Student: _____

Emaan Imtiaz

Signature of Supervisor: _____

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Prof. Dr. Attya Bhatti

Declaration

I certify that this research work titled "**Molecular Investigation of Candidate Gene Polymorphisms in Patients with Endometriosis Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* genes underlying risk of Endometriosis in Pakistani Patients** " is my work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged/referred to.

Signature of the Student

Emaan Imtiaz

00000364678

Dedicated to my parents.

Acknowledgements

It is through the blessings of Allah Almighty that I am what I am today, and no amount of gratefulness can be enough for all that He has given me. My family, my friends, and the resources to get to where I stand today are all because of Him, and no one is more praiseworthy than Him.

I thank my parents for all they have given me. Their belief in me as well as their constant and unwavering support is what has allowed me to face every hurdle I've ever encountered. I am beyond lucky to have been raised by two of the most talented people I have ever known, who have taught me more than any school did.

I thank my brother for giving me free food whenever I was down all through MS, for never letting me worry about getting stuck anywhere because he's always been a call away. I thank my sisters who weren't with me in person but never let me feel alone even in the hardest days.

I am grateful to my supervisor Prof. Dr. Attya Bhatti for being an incredible mentor and teacher. The relationship she developed with her students made it so easy to ask for help without hesitation, something I had not experienced before meeting her. Her constant encouragement and gentle support are what led me to the end of this journey. I am grateful to Prof. Dr. Peter John for teaching us what is really important: honesty with ourselves and others. He has truly given me a new insight into life with his incredible adherence to good morals, and has become a role model for all of us.

For my best friend and partner Ameema Khan, words of thanks are hardly enough. She has been with me through good days and bad, through college, BS, and now MS. From shopping sprees, and riding in the back of an open Suzuki to collect blood samples, to escaping together for lunch out somewhere. Every hurdle and obstacle was never too hard because she was there with me through it all. She was the one who got me into MS, and she's the one who got me through it.

I thank my emotional support Bisma Khalil, who has been there for me even if I have not been there for her. Who has the magical superpower to steal away all your

frustrations and worries with just a good ear. We have come so far together, and I hope we go even further.

I am forever grateful to my seniors Maria Sharif and Kashaf Rasool, two incredible ladies I had the delight of getting to know. The love and care they have given me and all their other juniors is awe-inspiring, and they are the biggest part of what I will miss about IGL.

To all my lab fellows: Faria, Tayyaba, Saliha, Rabia, Huda, and Iraj, I am so grateful to you all for making this research year such an enjoyable experience. Our experiments might not have always gone right, but I knew that you guys were always by my side.

Lastly, to the inanimate room named IGL, I am grateful that it gave me a family I never imagined I would have. There were days of frustrations, of feeling wronged, or lost, or useless, but in that lab there was always laughter, the occasional singing and dancing, and a shoulder to lean on. Thank you for everything.

Emaan Imtiaz

Contents

List of Figures

List of Tables

List of Acronyms

Alanine	Ala, A
Amplification refractory mutation system polymerase chain reaction	ARMS-PCR
Asparagine	Asn, N
Atta ur Rahman School of Applied Biosciences	ASAB
Base pairs	bp
Body mass index	BMI
Chronic kidney disease	CKD
Deep infiltrating endometriosis	DIE
Deoxyribonucleic acid	DNA
Differences of sex development	DSD
Estrogen receptor 1	ESR1, ERα
Estrogen receptor 2	ESR2, ERβ
Ethylenediaminetetraacetic acid	EDTA
Follicle stimulating hormone	FSH
Follicle stimulating hormone receptor	FSHR
Genome-wide association studies	GWAS
Grams	g
Hydrochloric acid	HCl
Luteinizing hormone	LH
Magnesium chloride	$MgCl_2$
Mayer-Rokitansky-Kuster-Hauser	MRKH
Messenger ribonucleic acid	mRNA
Microliter	µl
Milliliter	ml
Millimolar	mM
Molar	M
Mullerian duct	MD
Nanometer	nm

National Center for Biotechnology Information	NCBI
National University of Sciences and Technology	NUST
Ovarian endometriomas	OMAs
Ovarian hyperstimulation syndrome	OHSS
Polycystic ovarian syndrome	PCOS
Polymerase chain reaction	PCR
Post-translational modifications	PTMs
Potential of hydrogen	pH
Premature ovarian failure	POF
Revolutions per minute	RPM
Root-mean-square deviation	RMSD
Serine	Ser, S
Sex determining region Y	SRY
Sex reversion, kidneys, adrenal and lung dysgenesis	SERKAL
Single nucleotide polymorphisms	SNPs
Small ubiquitin-like modifiers	SUMOs
Sodium dodecyl sulfate	SDS
Sodium hydroxide	NaOH
SUMO-interacting motifs	SIMs
Superficial peritoneal endometriosis	SPE
Support vector machine	SVM
Threonine	Thr, T
Tris-acetic acid-EDTA	TAE
Tris-EDTA	TE
Ultraviolet-visible	UV-Vis
Untranslated region	UTR
Uterosacral ligaments	USL
Volts	V
Volume per volume	V/V
Weight per volume	W/V
Wingless type MMTV integration site	WNT

Abstract

Endometriosis is a complex and painful disorder in which endometrial tissue of the uterus grows in ectopic locations in the form of endometriotic lesions. It affects upto 10% females of reproductive age around the world, affecting their quality of life and causing infertility in many cases. No curative treatment exists, and management involves suppressing the symptoms with painkillers, contraceptives, or surgery. Bypassing invasive diagnostic procedures might be possible by the discovery of genetic biomarkers that are associated with incidence of endometriosis. In the present study, three single nucleotide polymorphisms (SNPs) (rs6166 of *FSHR*, rs7521902 near *WNT4,* and rs4986938 of *ESR2*) were selected to investigate their association with endometriosis in Pakistani patients. The study involved thirty-eight subjects and amplification refractory mutation system – polymerase chain reaction (ARMS-PCR) was used; however, no significant association was found. In-silico analysis of *FSHR* and *ESR2* was also performed by utilizing information from online data sets and webtools. Missense SNPs of *FSHR* and 3' and 5' UTR variants of *ESR2* were analyzed in-silico to filter other potentially pathogenic polymorphisms: rs386833515, rs121909661, rs371482817, and rs1256120. Further in-vitro experimentation is required to validate the obtained results in the future. Candidate SNPs associated with endometriosis may be proposed as biomarkers and molecular drug targets to improve diagnosis and treatment in the future.

Chapter 1 – Introduction

Endometriosis

Endometriosis is a painful, estrogen-dependent disease that is more commonly diagnosed in females of reproductive age. It is defined by the presence and growth of ectopic, functional endometrial tissue that is associated with severe and chronic pain (Greene et al., 2016). Endometrium is the tissue lining the inside of a uterus and is the mucosal tissue which thickens in anticipation of pregnancy. However if no fertilized egg is implanted within a specific period of time, the thickened lining is shed in the form of menstrual periods (Goldstein, 2010). When this endometrial tissue is found in aberrant locations outside the uterus, it has a significant detrimental effect on the day-to-day life, personal and sexual relationships, productivity, and mental health of affected women (Chapron et al., 2019). The presence of endometriotic lesions also increases the risk of ovarian cancer and may act as precursors to endometrioid ovarian carcinomas (Saavalainen et al., 2018).

1.1. Prevalence

Around 2-10% of all reproductive age women are affected by endometriosis (Parasar et al., 2017), however, a delay in diagnosis often happens and significantly reduces the quality of life. The average age at which endometriosis gets diagnosed is 28 years (Falcone & Flyckt, 2018). It is a misconception that pre-menarchal or post-menopausal women cannot have the disease; endometriosis has also been reported in neonates in a handful of cases (Dekker et al., 2021). Endometriosis has also been called a disease of adolescence, or by ancient physicians as ‘the disease of young girls’ and ‘the disease of virgins’ (“Endometriosis of Young Girls and Teenagers,” n.d.). It affects 70-73% of adolescent girls reporting pelvic pain, and is often unresponsive to medical treatment (Dun et al., 2015). In addition, 20-50% of women presenting infertility or subfertility, and 30-80% of women with chronic pelvic pain get diagnosed with endometriosis (Kho et al., 2018). The age-specific prevalence and yearly incidence of endometriosis has been shown in Figure 1.1. (Eisenberg et al., 2018)

1.2. Classification of Endometriosis

There is no one way to classify endometriosis; the disease can be classified based

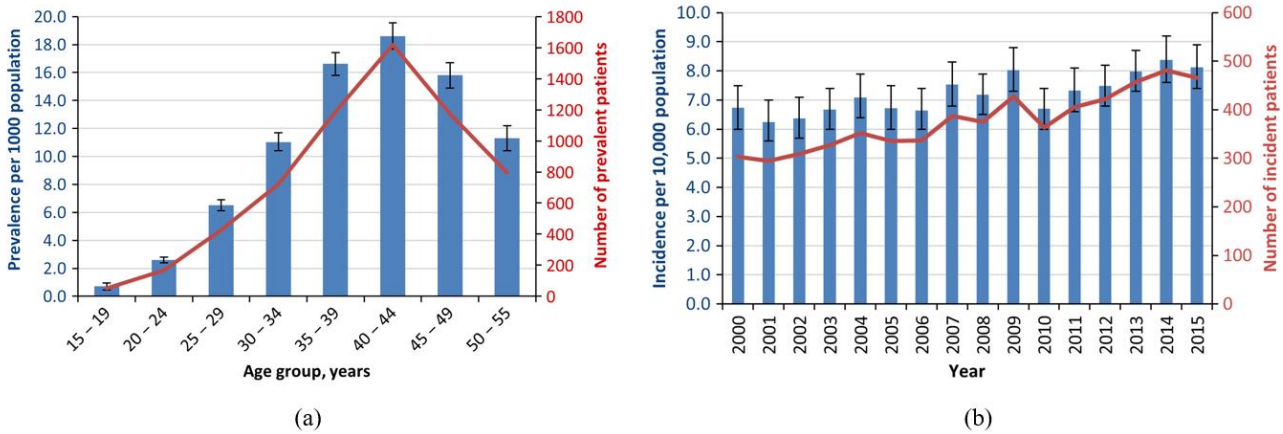


Figure 1.1. (a) Age specific prevalence of endometriosis and (b) yearly incidence of new cases of diagnosed endometriosis.

(*Copied from Eisenberg et al., 2018*)

on its progression, location, morphology, as well as number and size of lesions (Parasar et al., 2017). Based on whether the patient is experiencing minimal, mild, moderate, or severe endometriosis, it can be labelled as stage I, II, III, or IV respectively (American Society for Reproductive Medicine, 1997).

The common sites for endometriosis are the 'pelvic' locations: ovaries, peritoneum, fallopian tubes, and uterosacral ligaments (USL). Endometriosis might also be found rarely in 'extra-pelvic' sites such as the gastrointestinal tract, soft tissues, urinary tract, or chest (Lee et al., 2015).

Three distinct phenotypes of endometriosis have also been recognized based on the type and localization of endometrial lesions. The first, superficial peritoneal endometriosis (SPE), is associated with three types of superficial peritoneal lesions (red, white, or black) and is the least severe form. Second, ovarian endometriomas (OMA) also called 'chocolate cysts', are ovarian cysts filled with dark, chocolate-colored endometrial fluid. These cysts resist medical treatment and can potentially destroy all the ovarian tissue. Lastly, the most severe phenotype is the deep infiltrating endometriosis (DIE), associated with endometriotic lesions that penetrate greater than 5 mm into the peritoneum or organs surrounding the uterus (Imperiale et al., 2023)

1.3. Causes and Risk Factors

1.3.1. Reproductive and Lifestyle Aspects

Several factors have been constantly associated with the incidence of endometriosis in women, although the mechanisms behind them, or even the nature of the relationship, are not well understood. Short menstrual cycles, heavy bleeding, and early age of menarche have been linked to an increased risk of endometriosis, while the use of oral contraceptives, parity, and a high BMI (body mass index) interestingly has been associated with a decreased risk (Parasar et al., 2017). Linking other lifestyle factors with the disease has been even more confusing. Smoking is generally known to be harmful to health, however it has shown to reduce the risk of endometriosis in some women, probably due to estrogen suppression (Baron et al., 1990). A similar reason is behind the increased risk of disease in women who consume large amounts of caffeine and alcohol, although once again, the exact relationship cannot be defined. Increased physical activity and the inclusion of omega-3 fatty acids in diet has also some poorly understood role in decreasing the risk of endometriosis (Missmer et al., 2010).

1.3.2. Retrograde Menstruation

To date, the most widely accepted model for the occurrence of endometriosis is Sampson's retrograde menstruation theory (Sampson, 1927), according to which the backwards flow of menstrual blood through the fallopian tubes and its deposition into the peritoneal cavity leads to the adhesion of endometrial cells and formation of lesions (Persoons et al., 2020). This is also backed by the observation that the factors that increase menstrual flow (shorter cycles and longer periods with heavy bleeding) are also risk factors for endometriosis (Chapron et al., 2019). However, retrograde menstruation is a phenomenon observed in 90% of menstruating women and is generally regarded as non-pathological. This implies that in women with the disease, the migrating cells have acquired some additional qualities for adhesion, proliferation, and immune evasion that leads to the formation of pathogenic lesions (Persoons et al., 2020).

1.3.3. Hormonal Imbalances

Estrogens are a group of hormones responsible for the normal sexual and reproductive development in women. These hormones function by binding to their

receptors which are of two types: estrogen receptor α and estrogen receptor β, that are both persistently expressed by the ectopic endometrium. High estradiol levels are also found in the menstrual blood of endometriosis patients even though levels in peripheral blood are normal. This is mostly the result of the action of aromatase (cytochrome P450), that converts androgens to estrogens, and thus stimulates the growth of endometriotic lesions (Barbosa et al., 2011).

Progesterone is a hormone involved in maintaining menstrual cycles and the early stages of pregnancy. It can antagonize the estrogen-induced proliferation of endometriomal cells; the isoform A of the progesterone receptor can inactivate estrogen receptors. However, endometriosis is often associated with progesterone-resistance and loss of progesterone signaling (Patel et al., 2017). Luteinizing hormone (LH) has an important role in the maturation of oocytes, ovulation, production of androgens, and the release of estradiol and progesterone. High concentrations of LH have been observed in the peritoneal fluid of patients as well as a higher expression of the luteinizing hormone receptor on the surface of endometriotic lesions (Barbosa et al., 2011).

A correlation between incidence of endometriosis and thyroid dysfunction has also been reported. This might be caused in part by the progesterone resistance that is a hallmark of endometriosis (Peyneau et al., 2019).

1.3.4. Immunopathophysiology

As mentioned previously, endometrial cells that migrate out of the uterus have altered qualities that allow them to implant in ectopic lesions. Immune dysfunction has been theorized to be one of the mechanisms that facilitates implantation and endometrial lesion formation. The immunoinflammatory characteristics of ectopic endometrial cells are different from the eutopic endometrium. Transcriptional analysis of these cells shows increased expression of genes involved in apoptosis, adhesion, immune cell and cytokine recruitment, and other immunological pathways. Moreover, inflammatory cells associated with the micro-environment of a tumor have been seen in that of endometriotic lesions. Higher populations of neutrophils, natural killer cells, and dendritic cells have been majorly implicated in the pathophysiology of endometriosis (Symons et al., 2018).

1.3.5. Genetic Factors

The genetic basis for endometriosis has now been proven, although the identity of the genetic mutations has not. It is a complex and polygenic disorder, meaning that it is not caused by one single culprit gene, but by a complex combination of different gene mutations that confer susceptibility to the patient, in addition to environmental factors (Rahmioglu et al., 2015). Familial aggregation of the disease has been illustrated by many studies, showing that first-degree relatives are at a 3-15% higher risk of being affected (Saha et al., 2015). The best way to understand heritability however is through twin studies. A Sweden-based study showed that an additive genetic component of about 47% contributes to the manifestation of endometriosis (Saha et al., 2015)

While the biological mechanisms causing the disease are not exactly understood, many genetic variants that may be the reason behind them have been identified. This has been done through both hypothesis-based and hypothesis-free methods. Hypothesis-based approach, or the candidate-gene based approach, relies on prior understanding of biological mechanisms that may be related to the disease phenotype, and so the genes involved in these biological mechanisms are surveyed for correlation to disease manifestation. Hypothesis-free approach (including linkage studies and genome-wide association studies) on the other hand targets the whole genome instead of preselecting a specific area that might be involved in the disease (Rahmioglu et al., 2015). Candidate gene studies have approached many different gene groups with mixed negative and positive results of association; some of these are galactose-1-phosphate uridyl transferase gene, drug metabolism genes, steroid-related genes, DNA mismatch repair genes, and tumor suppressor genes (Simpson et al., 2003). To date, five genome-wide association studies (GWAS) have been conducted and have discovered several single nucleotide polymorphisms (SNPs) that have a robust association with endometriosis. These include missense, intronic, and intergenic mutations, and the genes containing them, or present near them are *WNT4*, *ETAA1*, *GREB1*, *FN1*, *ID4*, *NFE2L3*, *miRNA_148a*, *HOXA10*, *CDKN2B-AS1*, and *VEZT*. However, these genetic variants contribute very little to the heritability of the disease and its manifestation, and cannot be used as suitable diagnostic biomarkers (Rahmioglu et al., 2015).

1.4. Symptoms and their Mechanisms

The major symptoms presented by women suspected of having endometriosis are chronic pain and subfertility. However, a less quantifiable but no less detrimental effect is on the quality of life of the patients. Women experience loss of productivity, depression, decreased participation in social activities and hence diminished relationships, as well as associated high healthcare costs (Agarwal et al., 2019).

Pain symptoms that women often experience are dysmenorrhea (painful menses), dysuria (painful urination), dyspareunia (painful sexual intercourse), and dyschezia (painful passing of stool)(Parasar et al., 2017). No symptoms may be present as well, and as such there is a poor correlation between severity of pain and progression of disease. However, deep infiltrating endometriosis is often associated with more severe symptoms, and even in OMAs the deeper endometriomas cause stronger pain regardless of the size of the cyst. The increase in pro-inflammatory cytokines and growth factors linked with chronic inflammation and their long-term exposure can lead to peripheral and central sensitization. This can cause a persisting hyperalgesic state that harms quality of life to a worrying extent (Falcone & Flyckt, 2018).

The reason behind subfertility and infertility is not well understood, although multiple theories are present. The increase of inflammatory cells in the endometriomas and peritoneal cells may affect oocyte production and ovulation. These inflammatory cells may also affect fallopian tube mobility and function, impairing gamete transport and fertilized egg implantation. In addition, alteration in the environment of the endometrium due to genetic factors and other unknown reasons may cause implantation failure, leading to infertility (Macer & Taylor, 2012).

1.5. Diagnosis

There is an average delay of 7-8 years between the first sign of symptoms, and diagnosis of disease (Falcone & Flyckt, 2018). This has multiple reasons; a poor understanding of the causes of endometriosis, heterogeneity in its phenotypes, possibility of asymptomatic disease, and most importantly: the multifarious nature of its major symptom. Pelvic pain is hard to correlate with any single disorder, and finding the root cause behind it is a challenge faced by healthcare professionals (Chapron et al., 2019).

Initial diagnosis usually involves a physical examination in which a clinician palpates for tenderness and/or pelvic masses. Appropriate diagnostic tests are carried out to rule out other causes of pelvic pain such as gastrointestinal and urologic disorders, pregnancy, or other gynecological problems. A pelvic, transvaginal, or transabdominal ultrasound may be carried out to facilitate the diagnosis of ovarian cysts and endometriomas (Parasar et al., 2017).

The gold standard yet for the diagnosis of endometriosis is the invasive visualization of lesions through a laparoscope, coupled with a histological analysis for confirmation. Still, the correlation between the extent of symptoms experienced by the patient and the findings of the laparoscopy is poor (Agarwal et al., 2019). To date, there has been no reliable biomarker identified that could aid in the early and non-invasive diagnosis of endometriosis, and prevent its progression (Kiesel & Sourouni, 2019).

1.6. Management of Endometriosis

Currently, there is no cure for endometriosis. Existing treatment options are designed to suppress the symptoms and must be continued indefinitely to avoid recurrence. Available medical therapies include non-hormonal treatments (painkillers and non-steroidal anti-inflammatory drugs) or hormonal contraceptives (such as progestins and ethyl estradiol) to decrease inflammation and relieve pain. However, hormonal treatments are not recommended for patients who are trying for pregnancy. Assistive reproductive technologies are an option for women with endometriosis-associated infertility. (Bedaiwy et al., 2017).

Conservative surgery is another option for endometriosis patients, especially those who are trying to conceive naturally. It involves the complete or partial removal of endometriotic lesions through a laparoscopy. Definitive surgery on the other hand, includes a hysterectomy, and a possible oophorectomy, in addition to complete removal of lesions. Unfortunately, ongoing hormonal suppression is necessary even after surgery since pain

symptoms likely recur, and at a 50% risk after five years (Falcone & Flyckt, 2018). Figure 1.2. summarizes the basic information regarding endometriosis.

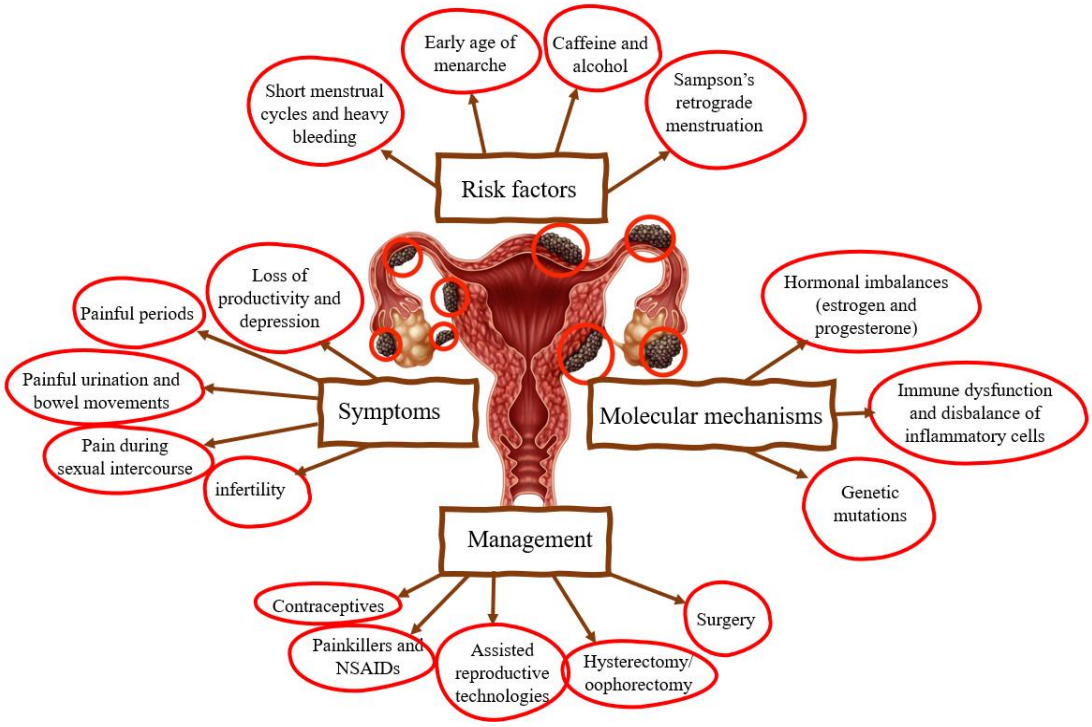


Figure 1.2. Risk factors, molecular mechanisms, symptoms, and treatments of endometriosis.

1.7. Selected SNPs for Association Analysis

There is an urgent need for the identification of genetic biomarkers that can aid in the diagnosis and treatment of endometriosis. These can be discovered by performing association studies to find out which gene mutations, specifically single nucleotide polymorphisms, are persistently present in patients suffering from endometriosis, while absent in people who are not. Such studies are being done all over the world and new potential biomarkers and drug targets are discovered every year.

In this study, three previously reported SNPs were selected to confirm their association with endometriosis in samples taken from Pakistani women. Two SNPs, rs6166 (Liaqat et al., 2021) and rs4986938 (Liaqat et al., 2013), were reported in studies involving Pakistani patients. The third polymorphism, rs7521902 (Matalliotakis et al., 2017), has been reported by GWAS and international studies but has not yet been studied in Pakistan.

rs6166 is a missense mutation (a genetic mutation changing an amino acid in the translated protein) present in the gene FSHR, that codes for a follicle stimulating hormone receptor. rs4986938 is present in the 3' UTR region of the gene ESR2, which codes for the estrogen receptor β and is persistently expressed in ectopic endometrial cells. Lastly, rs7521902, is an intronic variant present near the gene WNT4, which has a critical role in female sexual development.

1.8. Objective of the Study

Due to several reasons, the understanding of endometriosis in Pakistan is very low, even when the number of cases is increasing day by day. Women do not get checked until their symptoms are very severe, or they are facing difficulty in getting pregnant. As mentioned previously, the diagnosis itself is painful and invasive, and there is still no curative treatment available. The discovery of genetic biomarkers that can serve as tools for non-invasive diagnosis, or even as potential drug targets in the future, would improve the quality of life of countless women suffering from endometriosis.

The objectives of this study are:

1. To identify potential disease associated SNPs in selected genes (*FSHR* and *ESR2*) through in-silico analysis
2. To investigate the association between genetic variants in selected genes (*FSHR, WNT4* and *ESR2*) and patients with endometriosis.

Chapter 2 – Literature Review

Endometriosis is a complex, estrogen-dependent disorder, and little is known about its direct causes. The disease is becoming increasingly prevalent in females of reproductive age around the world. It is difficult to diagnose, and no curative treatment yet exists, and existing medical interventions only aim to suppress the symptoms. Severe chronic pain that heavily affects the quality of life of patients is one of the major symptoms, along with other possible complications and infertility. A cost-analysis study conducted by the US army showed a loss of 2.6 million dollars over a period of 6 years because of the time lost to surgical hospitalization (Eskenazi & Warner, 1997). Over the last few decades, many studies have been conducted to understand the causes of endometriosis. While little success has been had, several risk factors have been discovered that confer susceptibility to the disease. These include environmental factors as well as genetic mutations and polymorphisms in the genomes of subjects. Two GWAS performed recently in Japanese and European populations have discovered eight genomic regions with strong evidence of association with risk of endometriosis (Fung et al., 2015). Multiple studies have been conducted to confirm the results of GWAS in different populations, with varying levels of success. Other genes have also been studied and have been associated with disease occurrence and progression risk. The extent of work done on the genes selected for this study has been described below.

2.1. Follicle Stimulating Hormone Receptor

The mature follicle stimulating hormone receptor (FSHR) is a 678 amino acid long G protein coupled receptor that binds to and mediates the action of the follicle stimulating hormone (FSH). This hormone is secreted by the pituitary gland and plays a vital role in the reproduction of mammals (Coss, 2020). FSH is a 35.5 kDa protein consisting of two dimers, α and β, both of which are necessary for biological activity. In females, this hormone is necessary for follicular growth while in males it stimulates the proliferation of Sertoli cells. These cells interact with testosterone and thus regulate spermatogenesis. As such, men and women suffering from subfertility are treated with FSH routinely. In males the treatment has been seen to increase sperm count and improve morphology (Santi et al.,

2020). Assisted reproduction in females can involve treatment with FSH to stimulate collection of multiple eggs by the ovaries.

The *FSHR* gene coding for the receptor is located on chromosome 2 and is 54 kb long. It contains 9 introns and 10 exons, in addition to a promoter region. The gene has been seen expressed in extragonadal tissues as well as on the surface of tumors. Due to alternative splicing, four different isoforms of the FSHR protein exist, with isoform 1 being the full-length canonical protein. Many mutations and polymorphisms in the FSHR gene have been reported that can increase or decrease the function of the receptor, and hence its responsiveness to the hormone, causing disorders like ovarian hyperstimulation syndrome (OHSS) and premature ovarian failure (POF) respectively (Bhartiya & Patel, 2021). Approximately 1800 single nucleotide polymorphisms have been documented that affect the responsiveness of the receptor to its hormone, as well as the risk of developing infertility or OHSS among other reproductive disorders. Understandably, these mutations have a worse effect in women, leading to infertility, while men might become subfertile. Much research has been done to correlate FSHR polymorphisms to reproductive disorders and to uncover the molecular mechanisms beneath them, but without notable success. With so many variables coming into play, including ethnicity, maternal age, and combinations of genotypes, no clear results have yet been obtained.

2.1.1. Single nucleotide polymorphism of FSHR (rs6166) and associated disorders

Of the 1800 reported SNPs for the *FSHR* gene, only 8 are exonic and non-synonymous, causing an amino acid change in the resultant protein. Seven of these are in exon 10, which codes for the intracellular domain, the transmembrane domain, and the C-terminus of the extracellular domain. Two of these seven SNPs have been widely studied and implicated in several reproductive disorders; rs6165 (Ala307Thr) and rs6166 (Ser680Asn). These have been well characterized and are common in terms of ethnic distribution and allele frequencies (Bhartiya & Patel, 2021). The codons for these two positions seem to be in linkage disequilibrium, since they occur together in a way that is not entirely random. This has generated specific allelic variants that are found within different populations at specific frequencies; of these, Ala307/Ser680 (AS genotype) and

Thr307/Asn680 (TN genotype) are the most common. Some of the disorders associated with these two polymorphisms, along with others, are discussed below.

a. Infertility

Globally, 8-12% of couples face infertility issues that are unequally contributed by the male and female partner (Vander Borght & Wyns, 2018). Different FSHR associated medications are available, such as follicle stimulating drugs, gonadotropins, and medicines for ovarian stimulation that females are given to treat their infertility. Mutations in the *FSHR* gene, and the resultant changes in the receptor structure and function, may affect the capacity for reproduction. A study conducted in Iraq tested the response of infertile women to ovarian stimulating drugs. When genotyped, poor responders were homozygous for the rs6166 mutant allele at a high frequency. The polymorphism can thus be considered as one of the genetic factors behind infertility in Iraqi women (Hussein et al., 2022). Another study that recruited 106 infertile women was conducted in India to relate the co-occurrence of rs6166 Ser680Asn with infertility, however no significant results were obtained (Rai et al., 2019). There is still a clinically relevant relationship between the presence of this polymorphism and low fecundity that can be addressed for patients experiencing unexplained infertility.

b. Ovarian Hyperstimulation Syndrome (OHSS)

Studies have shown that people with Asn at position 680, either in homozygous (NN) or heterozygous (SN, NS) form, often respond very well to exogenous FSH or gonadotropins, and potentially have a higher risk of developing OHSS (Lussiana et al., 2008). The subjects with SS genotype (Ser680 homozygous) are more likely to be poor responders to ovarian stimulating drugs (Pabalan et al., 2014). These results are not consistent among all ethnicities, since a study done on Caucasian patients found a surprising number of OHSS patients that had the SS genotype (Daelemans et al., 2004). However, the same study did also find that patients suffering with a more severe case of OHSS had the Asn680 allele. It could be thus concluded that the NN or NS genotype can be used to predict the severity of the disease rather than its onset.

c. Polycystic Ovarian Syndrome (PCOS)

PCOS is a leading cause of female infertility that affects up to 12% of females worldwide, and is characterized by multiple ovarian cysts, androgen excess, and ovulatory dysfunction. Causative links between *FSHR* polymorphisms and PCOS have been studied extensively with inconsistent results. Chinese studies conducted on both rs6165 and rs6166 collectively failed to show any significant relationship between the disease and the genotypes (Fu et al., 2013; X. Wu et al., 2014). An Egyptian study showed different results by showing that the Ala307Thr polymorphism was associated with PCOS and may act as a causative factor (Abdel-Aziz et al., 2015). Another study conducted in India showed association between the risk of developing PCOS and NN genotype, along with elevated body mass index (BMI) and levels of LH (Sujatha et al., 2016). Other *FSHR* polymorphisms: rs2268361, rs2349415, rs1922476, and rs11692782 have also been studied and linked with PCOS in multiple European, Asian, and American studies (Laven, 2019).

d. Endometriosis

Studies have also been done on relating rs6166 with endometriosis. One conducted in Pakistan showed that NN genotype for rs6166 and TT genotype for rs6165 showed a high incidence in patients suffering from endometriosis compared to controls. The co-incidence of the two was 45.5% in endometriosis patients and 11% in control subjects (Liaqat et al., 2021). Another study showed different results; it found no significant correlation between the polymorphisms and the disease itself. However, when the endometriosis patients were further divided into groups, analysis showed that the AS genotype (Ala307Ser680) was associated with fertile women diagnosed with endometriosis. The AN haplotype (Ala307Asn680) was rarely found in the endometriosis group, regardless of the fertility status (André et al., 2018). A Chinese study concluded that subjects homozygous for the TN genotype (Thr/Thr307Asn/Asn680) had a higher risk of developing endometriosis compared to those that contained at least one of the converse alleles (H.-S. Wang et al., 2012)

2.2. Wingless-type MMTV Integration Site 4

Among the members of the wingless type MMTV integration site family (WNT) that are involved in embryonic development, WNT4 is a signaling protein involved in organogenesis forming the adrenals, the urinary tract and kidneys, the mammary glands, and the female reproductive tract. It is a 351 amino acid long protein encoded by the *WNT4* gene present on chromosome 1 that consists of 7 exons. The gene codes for two isoforms of the protein, isoform 1 being the full-length canonical protein. WNT4 protein shows 98% amino acid identity to that found in mice and rats. It is also highly conserved and found in both mammals and other vertebrates. Members of the WNT family act in a paracrine signaling pathway to induce a number of developmental changes. The proteins bind to specific cell surface receptors of the frizzled family; a family of atypical G protein-coupled receptors. Their activation in turn activates dishevelled phosphoprotein which recruits the cytosolic scaffold Axin. Axin binds to the cytoplasmic tail of lipoprotein receptor-related protein 5 and 6 and is then degraded. As a result, β-catenin (a protein responsible for triggering gene transcription) degradation is decreased, T-cell factor transcription factor is activated, and WNT-responsive genes are induced (Biason-Lauber & Konrad, 2008). A non-canonical pathway that does not involve β-catenin has also been reported. WNT4 can also activate Jun N-terminal kinase, also known as the stress-activated protein kinase, which in turn activates the transcription factor Pax2. Pax2 activation is necessary for the development of human embryonic kidney cells, and the Mullerian structures in females (Pellegrino et al., 2010).

The expression of the WNT4 protein is consistent in females, but dramatically decreases in males after the SRY (sex-determining region Y) transcription factor starts being expressed (Pellegrino et al., 2010). *WNT4* is expressed in fetal granulosa cells in humans, and in oocytes and granulosa cells after birth. This upregulates a gene responsible for antagonizing the nuclear-receptor steroid factor 1, and as a result steroidogenic enzymes are inhibited. It also works with *Rspo1* to decrease the expression of *Sox9*, a downstream effector of the male-specific SRY (Biason-Lauber, 2012).

2.2.1. Müllerian Duct and SERKAL Syndrome

Müllerian duct (MD) is an embryonic structure running down the side of the urogenital ridge, that develops into the fallopian tubes, the uterus, and the upper part of the vagina at about 8 weeks of gestation in females. While present in early male embryo as well, selective ablation later happens due to the action of Anti-Müllerian Hormone secreted by the Sertoli cells. in *WNT4* knock-out mice this structure fails to develop, indicating the major role of *WNT4* in its regulation. Along with *WNT4*, *WNT7* and *WNT9* have also shown some involvement in the morphogenesis of MD (Biason-Lauber & Konrad, 2008). If *WNT4* is overexpressed, or a duplication of the gene-containing region on chromosome 1 is present, severe hypospadias along with remains of both MD and Wolffian ducts have been reported. This is a characteristic of male-to-female sex reversal.

SERKAL syndrome (SEx Reversion, Kidneys, Adrenal and Lung dysgenesis) is the opposite of this situation, characterized by the inactivation of both copies of *WNT4* on homologous chromosomes. This recessive disorder has been described in 46XX fetuses and involves masculinization of fetal sex organs, adrenal hypoplasia, ambiguous genitals, underdeveloped kidneys, and several lung and heart defects. A missense mutation causing loss of function of *WNT4* by disrupting mRNA structure has been identified as the culprit (Mandel et al., 2008).

2.2.2. Mayer-Rokitansky-Kuster-Hauser (MRKH) Syndrome

In females with primary amenorrhea and normal secondary external genitalia, MRKH syndrome is suspected. Complete uterus and vaginal aplasia or hypoplasia coupled with malformed tubes and kidney problems are characteristic of the disorder. A French study identified an L12P (leucine to proline) mutation in exon 1 of the *WNT4* gene after analyzing the DNA samples of 28 adolescent girls suffering from MRKH syndrome. Functional analysis of the mutation showed increased expression of enzymes involved in androgen biosynthesis, which explains the hyperandrogenism symptoms experienced by the patients (Philibert et al., 2008). Another study on a patient with absent uterus and vagina along with other MRKH syndrome symptoms revealed a missense mutation in exon 5 of the gene. E226G (glutamate to glycine) caused the formation of a mutant WNT4 protein that was not palmitoylated at its secondary strucuture and thus remained trapped inside the

cells. When the mutant ovarian cells were cultured, no WNT4 was secreted and β-catenin was also not detected (Biason-Lauber et al., 2004).

2.2.3. Kidney Diseases

Correct expression of *WNT4* is also crucial in proper nephrogenesis (kidney development). Mice displaying loss of function of this gene showed renal agenesis and renal failure (Pellegrino et al., 2010). Expression of the gene was found to be altered in chronic kidney disease (CKD) experimental models. This was also tested in 98 human patients diagnosed with different glomerular diseases and at different stages of chronic kidney disease. It was found that the level of mRNA was higher in patients with membranous nephropathy, compared to those with lupus nephritis or IgA nephropathy. Patients at stage 3 of CKD also showed higher levels of *WNT4* mRNA compared to CKD stage 2 patients. This study also showed a correlation between *WNT4* expression and serum levels of albumin, cholesterol, sodium and potassium, creatinine, and urea in patients at different stages of CKD. It can be thus suggested that *WNT4* expression plays a role in disease progression and development (Kiewisz et al., 2019). Its expression was also found to be upregulated in acute kidney injury and the protein was detected earlier than serum creatinine and epidermal growth factor in the urine of patients with tubular injury, suggesting its possible use as a noninvasive biomarker (S.-L. Zhao et al., 2016). Some studies suggest that kidney lesions in glomerular diseases might be caused by increased WNT4 expression and its inhibition would help in kidney injury treatment (Q. Zhang et al., 2021).

2.2.4. Bone Diseases

Maintenance of bone mineral density is done by osteoclasts and osteoblasts that are responsible for bone resorption and bone formation respectively. Lower levels of WNT4 have been linked to a higher risk of fracture while higher levels prevent skeletal aging and osteoporosis by inhibiting bone resorption and promoting bone formation (Hendrickx et al., 2017). A study involving patients with primary osteoporosis showed a downregulation of *WNT4, WNT3a,* and *WNT5a* in their bone tissues. According to another study, mesenchymal stem cells that have been genetically engineered to express WNT4 can

potentially be used to enhance osteogenesis and repair craniofacial defects (Chang et al., 2007).

2.2.5. Single Nucleotide Polymorphism of *WNT4* (rs7521902) and Endometriosis

Recently conducted genome wide association studies conducted to understand endometriosis identified several SNPs, including those present in or near *WNT4* gene. The meta-analysis of two of these studies produced rs7521902; a C>A polymorphism located in the regulatory region at 1p36.12 and involved in post-natal uterine development. Different studies have been done to confirm this finding in different populations. Independent studies conducted in Italian (Pagliardini et al., 2013), Brazilian (Mafra et al., 2015), Chinese (Z. Wu et al., 2015), and Sardinian (Angioni et al., 2020) populations failed to find any significant correlation between occurrence of the polymorphism and increased risk or progression of endometriosis. A Greek study that recruited 166 women diagnosed with endometriosis and 150 controls reported association between the AC genotype and severe stages of the disease only (Matalliotakis et al., 2017).

2.3. Estrogen Receptor 2

ESR2, also known as ERβ or estrogen receptor 2, is a member of the nuclear receptor transcription factor family that binds to 17-beta-estradiol and other ligands. When the *ESR2* gene is translated, it forms a protein that binds specific ligands at its C-terminal and subsequently interacts with DNA through its N-terminal to activate transcription. The gene is located on chromosome 14q23.1 and consists of 9 exons spanning approximately 62 kb. Two other members of the estrogen receptor family are ERα and ERγ. ERα and ERβ have been widely investigated while ERγ has only been discovered recently. While the two former receptors have similar DNA and ligand binding sites, and interact with the same DNA response elements, their activation function domains and hence the protein co-regulatory complexes they recruit are very different, which might explain the differences in their biological function. Relative expression of the two receptors is also different in different cells. Estrogen effects can be mediated either by binding of the receptors to estrogen response elements in the target genes, or by the post-translational modifications (followed by the assembly of a protein complex) of the ERs localized at the plasma

membranes instead of in the nucleus (Gregorio et al., 2021). Overview of the estrogen pathway is shown in Figure 2.2. (Dimauro et al., 2021)

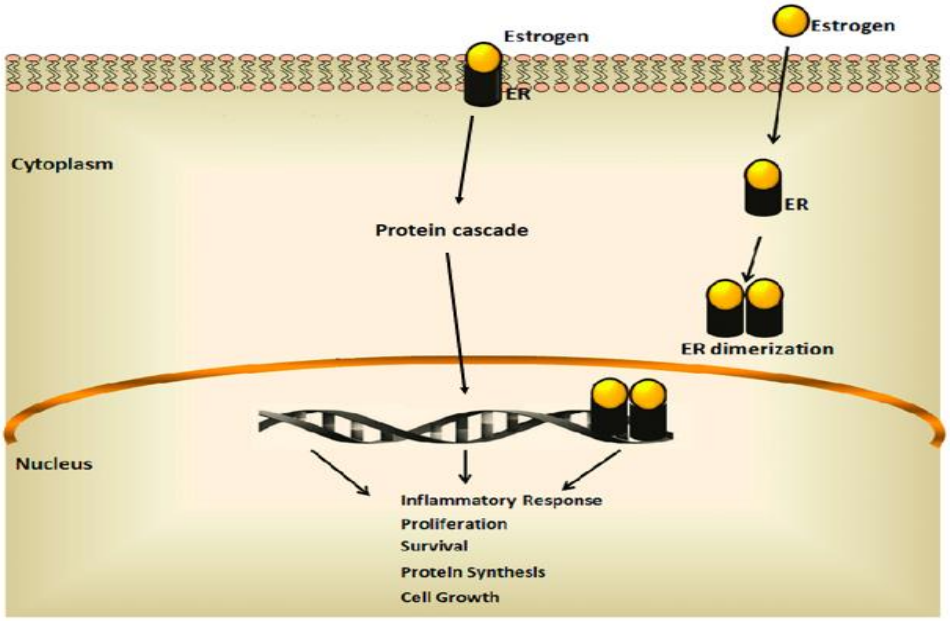


Figure 2.1. Overview of the estrogen signaling pathway.

Estrogen binds to and activates intracellular estrogen receptor which dimerizes, enters the nucleus, and binds to estrogen response elements to activate genes responsible for various cellular functions. Alternatively, estrogen binds to plasma membrane ER and activates a protein cascade which regulates gene expression of estrogen target genes (*Adapted from Dimauro et al., 2021*)

Estrogens (majorly estrone, estradiol, estriol, and estetrol) being steroid hormones, can passively enter through the plasma membrane of any cell, therefore it is the presence of the ER that is necessary for the action of the estrogen. While predominantly expressed in females, they also play important reproductive roles in males, such as maturation of sperm cells. Development of secondary sexual characteristics of females and thickening of the endometrium during the menstrual cycle is also dependent on estrogen (Fuentes & Silveyra, 2019).

2.3.1. Reproductive Disorders

Sex determination depends on the activation and inactivation, or presence and absence of several sex-specific genes. Changes in this delicate balance can lead to

differences of sex development (DSD) where individuals with atypical sexual characteristics are born. Homozygosity mapping and whole exome sequencing revealed that a homozygous deletion mutation in the *ESR2* gene, causing significantly increased transcriptional activation, was the reason behind syndromic DSD in a 46,XY individual (Baetens et al., 2018). While mutations in *ESR1* have been commonly linked with reproductive disorders, delayed puberty, and osteoporosis in females, a study conducted in 2018 showed that *ESR1* alone is not sufficient for correct sexual development in women. Functional transactivation studies and whole exome sequencing showed that a 16-year-old girl with primary amenorrhea, osteoporosis, lack of breast development (signs similar to *ESR1* mutation consequences) had a loss-of-function mutation in a conserved residue of *ESR2*, which disrupted estradiol signaling and caused complete ovarian failure (Lang-Muritano et al., 2018).

2.3.2. Hypertension and Cardiovascular Disease

Estrogen deficiency plays an important role in the path to cardiovascular disease, and hypertension is one of mediators involved. To study the association between a specific CA repeat polymorphism in the *ESR2* flanking region and hypertension, a study was conducted on postmenopausal Japanese women. Women that had at least one allele for 26 CA repeats had significantly high systolic blood pressure than those who did not, although the molecular mechanism is not understood (Ogawa et al., 2000). Another study tested the relationship between SNP rs4986938 and risk of hypertension in post-menopausal Chinese Han women, and found that the wild type allele was significantly associated with decreased risk of high blood pressure (Xin et al., 2013). A study also showed that in women with hypertension, two *ESR2* SNPs in linkage disequilibrium with one another were associated with left ventricular hypertrophy and left ventricular wall thickness. No association was found in men (Peter et al., 2005).

2.3.3. Single Nucleotide Polymorphism of *ESR2* (rs4986938) and Endometriosis

Endometrial stromal cells display abnormal expression of a wide range of nuclear receptors. NR5A1 and NR2F2 are two nuclear receptors that compete to maintain the regulation of steroid-synthesizing genes in these cells. Dominance of NR5A1 leads to the formation of excessive estrogen. Moreover, the ratio ESR1/ESR2 in the endometriotic

stromal cells is abnormally low due to high expression of *ESR2* (Yilmaz & Bulun, 2019). Multiple studies have shown that in various endometrial lesions, the mRNA levels of *ESR2* are significantly higher than that of *ESR1* and *PGR*, corresponding to the low levels of ESR1 and progesterone receptors that have been observed (Smolarz et al., 2020). Different mutations in the *ESR2* gene, such as SNP rs17179740 (Smolarz & Romanowicz, 2021), SNP rs4986938 (Liaqat et al., 2013), and a CA repeat polymorphism (Lamp et al., 2011) have been associated with a higher risk of endometriosis.

Chapter 3 - Methodology

3.1. In-silico Analysis of *FSHR* and *ESR2*

3.1.1. Retrieval of Missense SNPs of FSHR

The total reported non-synonymous SNPs of *FSHR* gene were analyzed for identification of other possibly deleterious polymorphisms that may play a role in the causation of endometriosis. For this purpose, missense SNPs of this gene were retrieved from the NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp/).

a. Filtering deleterious SNPs

The obtained SNPs were filtered through a series of online tools that predicted whether they were deleterious or not. The list of these tools, along with their operating principles, is given in Table 3.1. Amino acid FASTA sequence of the canonical FSHR protein was obtained from UniProt (https://uniprot.org/) to be added as input sequence where required.

Table 3.1. Online tools used to filter deleterious missense SNPs of the FSHR gene.

	Tool	**Principle**	**Weblink**
1.	SIFT (Sorting Intolerant From Tolerant)	It is a sequence-homology based tool that predicts the effect of non-synonymous genetic polymorphisms on protein function by assuming important amino acids to be evolutionarily conserved (Kumar et al., 2009).	https://sift.bii.a-star.edu.sg/.
2.	PolyPhen-2 (Polymorphism Phenotyping v2)	This tool considers comparative evolutionary and structural information to predict the effect of non-synonymous mutations on the stability and function of the protein (Adzhubei et al., 2013).	http://genetics.bwh.harvard.edu/pph2/
3.	PANTHER (Protein	PANTHER uses phylogenetic trees as a set of reference protein sequences that have	http://www.pantherdb.org/

	Analysis Through Evolutionary Relationships)	been obtained from sequencing genomes of organisms. This information is used to make inferences on the gain or loss or protein function due to amino acid substitutions (Thomas et al., 2022).	
4.	PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms)	This tool is an SVM (Support Vector Machine) -based classifier that classifies missense substitutions into disease and neutral polymorphisms (Capriotti et al., 2006).	https://snps.biofold.org/phd-snp/phd-snp.html
5.	SNPs&GO	SNPs&GO is a GO-integrated predicting software that operates on an SVM-based classifier principle. It takes the input protein sequence and functional information to predict probability of deleterious mutation (Capriotti et al., 2013).	https://snps.biofold.org/snps-and-go/snps-and-go.html

b. Predicting Structural and Functional Alterations

Only the SNPs that were predicted to be deleterious by all the filtering tools were then entered into the MutPred2 online tool (http://mutpred.mutdb.org/#qform). Along with giving a general prediction about the pathogenicity of a polymorphism, the tool also provided a list of potential molecular alterations that an SNP may cause in the resultant protein (Pejaver et al., 2020).

To understand the effect of each polymorphism on the stability of the protein, I-Mutant 2.0 (https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi) was used. This tool is also an SVM-based webserver that uses the protein sequence or structure to accurately predict whether a single point mutation stabilizes or destabilizes the protein (Capriotti et al., 2005).

The evolutionary conservation of the amino acids being substituted by filtered missense mutations was estimated using the online ConSurf webserver (https://consurf.tau.ac.il/consurf_overview.php). Highly conserved amino acids are often

those that have important roles in maintaining the structure and function of a protein, therefore knowing the degree of conservation of specific amino acids can help in understanding their significance in the protein (Ashkenazy et al., 2016).

c. Protein Modelling of Mutant and Wild-type Protein

For protein modelling, trRosetta (transform-restrained Rosetta) online tool was used (https://yanglab.nankai.edu.cn/trRosetta/). The wild protein was modelled first using the canonical protein's amino acid FASTA sequence. Separate models for each filtered SNP were generated subsequently. trRosetta is a fast and accurate modelling tool that uses both de-novo and homology modelling (Du et al., 2021), and thus provided far more reliable results than other online webservers. Each protein model result was obtained within a few hours of query submission, and multiple entries were made at the same time.

The quality and confidence levels of the designed models were tested using two different protein validation webservers; ERRAT (https://saves.mbi.ucla.edu/), and MolProbity (http://molprobity.manchester.ac.uk/).

d. Protein Superimposition

Root-mean-square deviation (RMSD) values are often used as a good indicator of structural variability between two similar proteins. The value is zero for two identical proteins and increases as the level of variability increases. This value can be used to predict the structural change caused by a point mutation in the translated protein. If the change of a single amino acid results in a protein that is notably different in structure compared to the wild-type protein, it can be assumed that the function of that protein would be significantly altered as well.

To find out the RMSD value of each mutant/wild protein pair, TM-align software was used (https://zhanggroup.org/TM-align/). The algorithm aligns the two uploaded protein structures residue-by-residue, providing an optimal superimposition as well as the RMSD and TM scores (Y. Zhang & Skolnick, 2005). The TM-score value works in reverse to the RMSD; a higher value shows greater similarity between the two proteins, with the score of 1 indicating a perfect match. Mutants with the highest RMSD values were selected as final outputs. Superimposed proteins were also visualized using the PyMOL 2.5 software (Yuan et al., 2017).

e. Protein-protein Interactions

The protein-protein interactions, including physical and functional associations, of the selected protein with others were predicted by the STRING (Search Tool for the Retrieval of Interacting Genes/proteins) software (https://string-db.org/). The data contained in this server is obtained from literature, reported experiments, and computer predictions, and can be used to understand the complex protein interactions (Szklarczyk et al., 2023).

f. Post-translational Modifications

Post-translational modifications (PTMs) are covalent modifications, performed by enzymes, that occur on specific amino acid residues of synthesized proteins and can significantly alter their properties. Different tools were used to check the effect of selected polymorphisms on the PTMs of the protein.

i. Phosphorylation

Phosphorylation is one of the most important post-translational modifications that can activate, deactivate, or otherwise alter the function of a protein. Protein kinases covalently attach a phosphate group to the side chains of serine, threonine, and tyrosine residues in the protein. Any change in the phosphorylation of a protein can have significant effects on the protein. GPS 6.0, available at http://gps.biocuckoo.cn/online.php, was used to check phosphorylated sites in the wild and mutant proteins (Chen et al., 2023).

ii. Glycosylation

Protein glycosylation is a crucial process that is important for cell adhesion, recognition, protein stability, solubility, transport, and immunogenicity among many others. In this process, a carbohydrate is attached to specific functional groups of amino acids. N-glycosylation involves the addition of glycans to the nitrogen of asparagine or arginine side chains. The N-glycosylation of wild and mutant protein were predicted by NetNGlyc 1.0 (https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/). O-glycosylation is the attachment of sugar group to oxygen atom of serine or threonine residues of the protein, and was testing using NetOGlyc 4.0 (https://services.healthtech.dtu.dk/services/NetOGlyc-4.0/)

iii. Methylation

Methyltransferases catalyze the transfer of methyl groups to the N-groups of lysine and arginine side chains, or the C- and N-termini of different proteins. The methylated sites of the wild and mutant protein were predicted using the GPS-MSP 1.0 webserver (http://msp.biocuckoo.org/) (Deng et al., 2017).

iv. Ubiquitination

Ubiquitination is an ATP-dependent process performed by a cascade of enzymes and has several biological functions. GPS-Uber webserver available at http://gpsuber.biocuckoo.cn/ was used for prediction (C. Wang et al., 2022).

v. SUMOylation

Small ubiquitin-like modifiers (SUMOs) can covalently modify lysine residues on different proteins, and thus regulate transcription, DNA repair, cell cycle, and other aspects of protein functioning. These sites, along with SUMO-interacting motifs (SIMs) were identified by using GPS-SUMO 2.0 (https://sumo.biocuckoo.cn/) (Q. Zhao et al., 2014)

3.1.2. Non-coding Transcript Variants of *ESR2*

The non-coding transcript variants of *ESR2* were retrieved from NCBI dbSNP. Only the 3' and 5' UTR variants were selected and analyzed for potentially pathogenic SNPs. Selected SNPs were first entered into the RegulomeDB webserver (https://regulomedb.org/regulome-search/). RegulomeDB is a tool used to select and prioritize functionally significant single nucleotide variants present in the non-coding regions of the genome. The tool ranks SNPs from 1-7, variants with a score of 1 have the most supporting data available and are thus most likely to be functional. A probability score of 0 to 1 is also given, with 1 being the highest probability of an SNP being a regulatory variant (Dong et al., 2023). From the returned result, only the SNPs with a score of 1a-1f or 2a were selected.

All 3' and 5' UTR variants were then entered into SNPinfo, a webserver that identifies SNPs with alternative alleles that have differential effects (https://snpinfo.niehs.nih.gov/). We selected the SNPinfo FuncPred tool to filter SNPs that may be functionally significant . These functional effects include changing transcriptional activity by binding to transcription factor binding sites (TFBS), affecting stop codons, and causing premature termination of protein, disrupting splice sites, exonic splicing silences

(ESS), and exonic splicing enhancers (ESE), and affecting the translation of proteins by changing the sites on which miRNAs bind. A regulatory potential score and conservation score ranging from 0 to 1 is also given (Xu & Taylor, 2009). SNPs that were found to be functionally significant by both RegulomeDB and SNPinfo were finalized.

3.2. In-vitro Analysis of FSHR (rs6166), WNT4 (rs7521902), and ESR2 (rs4986938)

This study was performed to investigate the association of three candidate gene polymorphisms (rs6166 of *FSHR*, rs4986938 of *ESR2*, and rs7521902 of *WNT4*) with endometriosis in patients of Pakistani origin. Ethical approval for this study was obtained from the Institutional Review Board (IRB), ASAB.

3.2.1. Sample Collection

A total of 38 samples were collected from collaborative hospitals. The sampling was completed over a period of 8 months.

The study was a case-control study, and the collected samples were categorized into the following groups:

1. Cases: Subjects diagnosed with endometriosis (19 samples)
2. Controls: Healthy subjects without endometriosis (19 samples)

3.2.2. Selection of Subjects

The subjects were chosen based on the inclusion and exclusion criteria given in Table 3.2.

Table 3.2. Inclusion and exclusion criteria for designed association study

Inclusion criteria		**Exclusion criteria**
Cases	**Controls**	
Diagnosed with Endometriosis	Not diagnosed with endometriosis	Males
Ages 15-45	Ages 18-45	Other reproductive disorders besides endometriosis
		Crossed menopause
		Below age of menarche

3.2.3. Sample Type

Whole blood was collected from the study subjects after informing them about the nature and purpose of the study, as well as taking verbal and/or written consent. The blood samples were collected in 5ml EDTA (Ethylenediaminetetraacetic acid) tubes by a trained phlebotomist. The tubes were labelled with the names and ages of the patients, and the tube ID numbers were entered into a list for future reference. Blood collected in the hospitals was stored safely in an ice box and brought back to the ASAB Immunogenetics Lab where it was immediately stored at 4°C.

3.2.4. DNA Extraction

Genomic DNA was extracted from the whole blood samples using the phenol-chloroform method. The composition of the different solutions used in the process is given in Table 3.3.

Table 3.3. Solutions required for DNA extraction using Phenol-Chloroform method.

Solution	Method for making	Function
Solution A (blood cell lysis buffer)	For 1000 ml solution • 109.44 g Sucrose (320 mM) • 12.114 g Tris (10 mM) • 0.476 g $MgCl_2$ (5mM) • 10 ml Triton X-100 (1% V/V) • The above chemicals were combined in 600 ml double-distilled water and the pH was set to 7.5 using HCl/NaOH. Volume was made up to 1000 ml with double-distilled water.	This solution breaks the cell membranes of the blood cells allowing cell contents to be released into solution.
Solution B	For 1000 ml solution • 12.114 g Tris (10 mM) • 23.37 g NaCl (400mM) • 0.58 g EDTA (2mM)	This solution separates the protein content and precipitates the DNA present in the mixture.

	• The above chemicals were combined in 600 ml double-distilled water and the pH was set to 7.5 using HCl/NaOH. Volume was made up to 1000 ml with double-distilled water.	
20% Sodium Dodecyl Sulfate (SDS)	For 100 ml solution • 20 g SDS was weighed out in a graduated bottle and volume made up to 100 ml with distilled water	SDS contributes to cell lysis, promoting high yield and purity of the extracted DNA.
Proteinase K		It degrades contaminating proteins and any nucleases that may attack the DNA.
Solution C	Phenol	Phenol helps in the separation of DNA into a separate, aqueous layer.
Solution D	For 50 ml solution • 48 ml Chloroform (24 parts) • 2 ml iso-amyl alcohol (1 part)	This solution helps in the coagulation of protein so that the aqueous layer of pure DNA can be easily collected.
Isopropanol		Isopropanol efficiently precipitates DNA from large volumes of solution.
Sodium Acetate	For 100 ml solution: • 24.609 g sodium acetate was weighed out in a graduated bottle and volume made up to 100 ml with distilled water (3M)	Sodium acetate is also used to precipitate the DNA.
Ethanol	100%	Ethanol decreases the solubility of DNA allowing it to precipitate. Pure ethanol is

		used as it evaporates completely.
PCR water/TE buffer		Used to resuspend the DNA pellet and for storage.

Phenol-chloroform extraction protocol

Day 1

- Using a reliable micropipette and autoclaved tips, 750 µl of whole blood was transferred to a 2 ml autoclaved microcentrifuge tube. 750 µl of Solution A was also transferred to the same tube and the tube was inverted 4-6 times to ensure proper mixing of the two solutions.
- The tube was then kept at room temperature for 10 minutes, after which it was centrifuged at 13000 rpm for 10 minutes.
- After centrifugation, the supernatant was discarded, leaving behind a small pellet at the bottom of the tube. This pellet was then completely dissolved in 400 µl of Solution B by the application of mechanical force. The tube was again centrifuged at 13000 rpm for 10 minutes.
- The supernatant was again discarded and 400 µl of Solution B, 12 µl of 20% SDS, and 5 µl of proteinase K was added to the pellet. The pellet was again dissolved by application of mechanical force.
- The tube was incubated overnight at 37°C.

Day 2

- 250 µl each of Solution C and Solution D was mixed in a separate tube and then added carefully without mixing to the tube taken out from the incubator after overnight incubation. It was then centrifuged at 13000 rpm for 10 minutes.
- After the tube was centrifuged, the upper aqueous layer containing DNA was carefully pipetted out into a new tube. Care was taken as to not disturb the sticky protein layer directly in contact with the aqueous layer, as mixing affects the purity of the extracted DNA. In case of too much mixing, 500 µl of Solution D was added to the removed aqueous layer, centrifuged, and the aqueous layer was removed again.

- The aqueous layer was then submerged in 55 µl of 3M sodium acetate and 500 µl iso-propanol. The tube was inverted several times to promote the precipitation of DNA. At this stage, DNA was observed as a white thread suspended in the solution.
- The tube was again centrifuged at 13000 rpm for 10 minutes, and the supernatant was discarded. The DNA pellet was then resuspended in 200 µl of chilled pure ethanol.
- After the tube was again centrifuged at 13000 rpm for 10 minutes, the tube was gently tipped over to discard the excess ethanol and the remaining was left to evaporate through air-drying.
- Once the tube was completely dry and no ethanol was left behind, the DNA was submerged in 500-100 µl of TE buffer. This was then stored at -20°C in the freezer.

3.2.5. Quantificaiton of DNA

a. Nanodrop Spectrophotometer

The concentration and purity of the extracted DNA was tested using the ThermoScientific Nanodrop 2000 UV-Vis Spectrophotometer at ASAB. The following steps were followed:

- The Nanodrop software was started and the 'Nucleotides' application was selected. In the next option 'DNA' was selected. The arm was lifted over the pedestal and 1 µl of TE buffer (the same which was used to resuspend the DNA pellet) was used as a blank.
- Once it was confirmed that the device gave a quantification in negative values when blanked, the arm was lifted, and the drop of TE buffer was wiped off.
- 1 µl of the DNA sample was now applied and the arm was gently closed.
- Pressing 'Measure' then gave the result for DNA concentration in ng/µl. The device also gave the ratio of absorbances at 260 nm and 280 nm. This ratio is used to assess the purity of the DNA. A ratio of around 1.8 is accepted as relatively pure for DNA, and a lower value might indicate the presence of proteins, phenol, or other contaminants that absorb strongly at 280 nm.
- The values for the DNA concentration and 260/280 ratio were noted for each sample.

b. Agarose Gel Electrophoresis

Gel electrophoresis is used for separating out DNA strands according to their size, as well as for the qualitative and semi-quantitative detection of DNA. The extracted DNA was run on a 2% (w/v) gel according to the protocol below. Table 3.4. describes the composition of the solutions used in this technique.

Table 3.4. Composition of solutions required for Agarose gel electrophoresis.

Solution	**Method for making**
0.5 M EDTA stock solution	For 500 ml • 93.05 g of EDTA disodium salt was weighed out and dissolved in 400 ml of deionized water. • pH was adjusted to 8.0 using NaOH pellets to dissolve the salt. • Volume was topped up to 500 ml and the solution was autoclaved.
50 x TAE buffer stock solution	For 1000 ml • 242 g of Tris-base was dissolved in around 600 ml deionized water. • 57.1 ml of glacial acetic acid and 100 ml of 0.5 M EDTA stock solution was added. • Volume was made up to 1000 ml.
1 x TAE working solution	For 1000 ml • 20 ml of the 50 x TAE stock buffer was combined with 980 ml of distilled water.

Protocol

- 2 g agarose powder was weighed out using a digital weighing balance and added to a glass flask. 100 ml of 1 x TAE solution was then added using a measuring

cylinder. To allow the agarose to dissolve, the flask was heated in the microwave for approximately 2 minutes.

- Once the solution had stopped steaming, 4 µl of ethidium bromide solution (8 µg/ml) was added and mixed.
- The solution was then poured into a gel casting tray and a comb was immediately inserted to allow wells to form.
- The comb was removed when the gel had cooled and solidified. The gel was then taken out of the casting tray and placed in a gel tank connected to a power supply. It was positioned so that the wells were on the side of the negative terminal.
- 1 x TAE buffer was poured into the tank, allowing it to cover the gel. 8 µl of a 1000 bp DNA ladder were loaded into the first well. 2 µl of loading dye were combined with 6 µl of the DNA sample and then loaded into the next well.
- The gel was electrophoresed at 90 V for 30 minutes, and then viewed in a Wealtec UV transilluminator as well as a Cleaver Scientific omniDOC gel documentation system.

3.2.6. Primer Designing

To carry out ARMS-PCR (Amplification-Refractory Mutation System-Polymerase Chain Reaction), three primers each were manually designed for the three candidate gene polymorphisms; rs6166 in the FSHR gene on chromosome 2, rs7521902 an intron variant near WNT4 gene on chromosome 1, and rs4986938 located in the 3' UTR near ESR2 gene on chromosome 14. Among the three primers designed, one was the common primer that would bind to DNA template, both in the presence and absence of target mutation. The 'wild' primer would bind only if the wild-type allele was present in the DNA target and the 'mutant' primer would bind only with the mutant allele of the SNP.

FASTA sequences of the three SNPs were obtained from dbSNP in NCBI (National Center for Biotechnology Information). An online software, OligoCalc, was used to predict the properties of oligonucleotide primers, including self-complementarity, hair-pin formation, and annealing temperatures. Primer BLAST software was used to check the specificity of the designed primers to make sure that they did not bind on other DNA regions in the genome. Finally, UCSC in-silico PCR browser was used to predict the

efficiency of the primers when running PCR in the wet-lab, as well as a secondary test for primer specificity and product size. Table 3 shows the details of the designed primers.

Primers were obtained in lyophilized form, and the reconstituted stock was used to make dilutions in separate tubes by adding nuclease-free water. The optimum annealing temperature for each primer set was determined using gradient PCR. The primer pair for rs6166 (FSHR) showed best amplification at 59°C, that of rs7521902 (WNT4) at 58°C, and that of rs4986938 (ESR2) at 57°C.

Table 3.5. Primers designed for ARMS-PCR.

Gene and SNP-ID	**Nucleotide change**	**Primer sequence – 5'- 3'**	**Primer length – bp**	**Melting temperature – °C**	**PCR amplicon size – bp**
FSHR rs6166	C>T	F-wild: GGACAAGTATGTAAGTGGAACCAC	24	55.7	371
		F-mutant: GGACAAGTATGTAAGTGGAACCAT	24	54	
		R: CTCTAGTGACACCAGGATCGC	21	56.3	
WNT4 rs7521902	C>A	F-wild: GGAATTGCGAGGTGAAACC	19	51.5	328
		F-mutant: GGAATTGCGAGGTGAAACA	19	49.3	
		R: TTTCTCAGGATAACGATTTCACC	23	51.9	
ESR2 rs4986938	C>T	F-wild: CTGGAGTTCACGCTTCAGCC	20	56.5	245
		F-mutant: CTGGAGTTCACGCTTCAGCT	20	54.4	
		R: TAACAAGGGCATGGAACATCTGC	23	55.5	

Note: F-wild – forward primer with wild-type allele of SNP at 3'; F-mutant – forward primer with mutant allele of SNP at 3'; R – common reverse primer.

3.2.7. Amplification Refractory Mutation System – Polymerase Chain Reaction

To carry out the Amplification Refractory Mutation System – Polymerase Chain Reaction (ARMS-PCR), each DNA sample was amplified in two separate PCR tubes. One tube contained the 'wild' forward primer, the other tube was given the 'mutant' forward primer. The common reverse primer was added to both the tubes. Amplification only happened if the target allele was present in the DNA sample. If amplification occurred in both the tubes, it indicated a heterozygous DNA sample that contained both the wild and the mutant allele.

The reaction mixture that was made in each tube is given in table 4. Preparation of the reaction mixture was always done in a laminar flow biosafety cabinet, and the tubes were constantly kept on ice during the process. Finally, the tubes were spun for a few seconds in a short-spin equipment to allow proper mixing of all the components and then loaded in a Thermocycler 2720. The thermocycling conditions for the three separate primer pairs are given in Figures 3.1., 3.2., and 3.3.

Table 3.6. Components and volumes used in the reaction mixture for PCR.

Order	**Reagents**	**Amounts**
1	PCR water	10 µl
2	Forward primer	1 µl
3	Reverse primer	1 µl
4	DNA template	2 µl
5	Master mix	6 µl
		Total: 20 µl

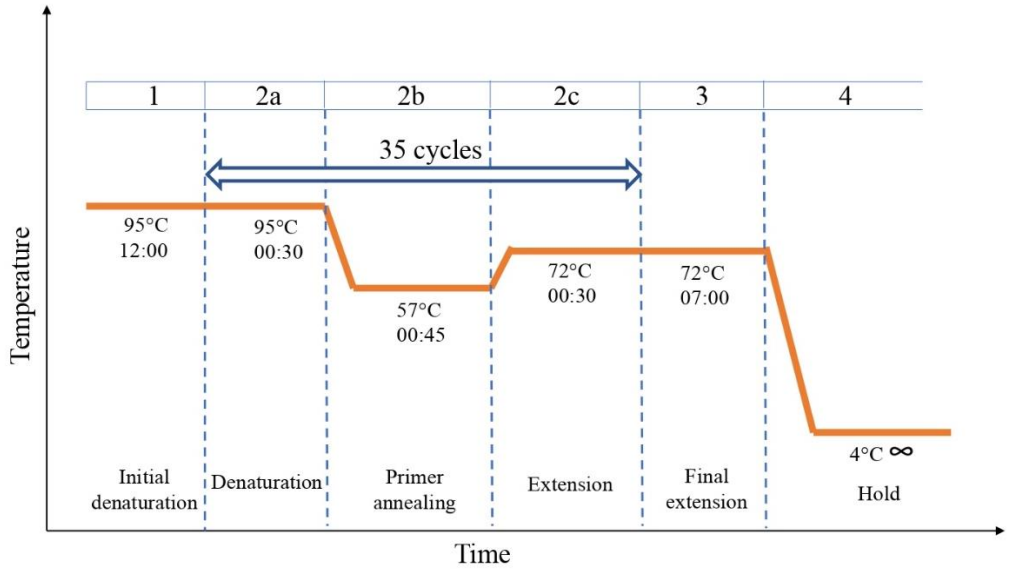


Figure 3.1. PCR profile for ESR2 rs4986938

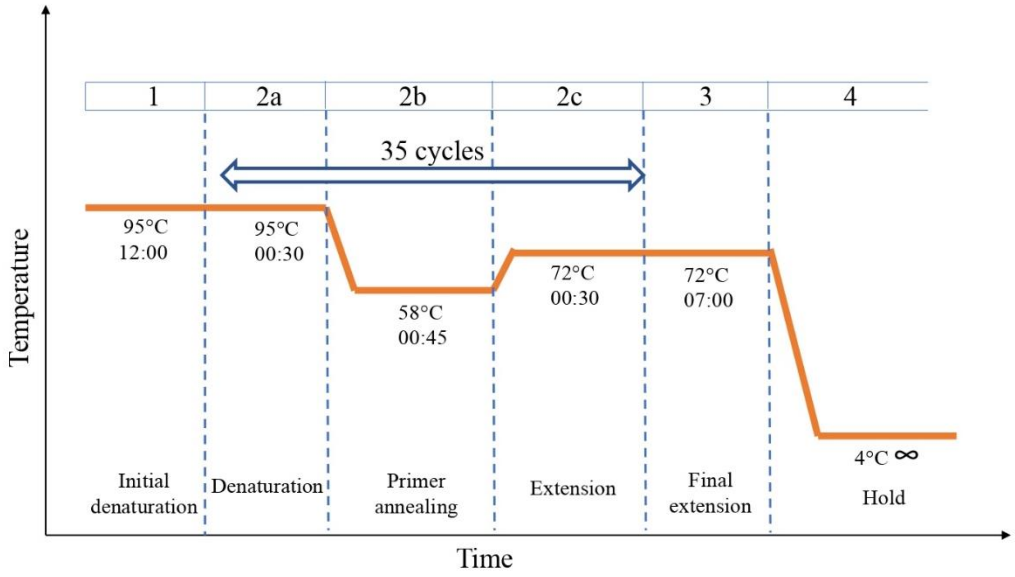


Figure 3.2. PCR profile for WNT4 rs7521902

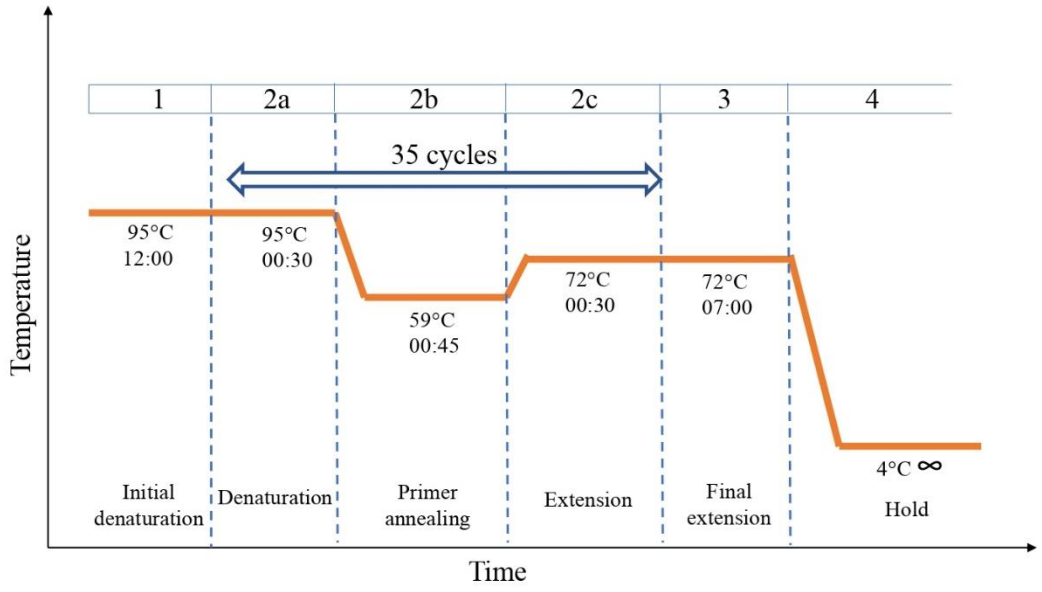


Figure 3.3. PCR profile for FSHR rs6166

3.2.8. Agarose Gel Electrophoresis

Gel electrophoresis was carried out again to visualize the results of the amplification. While the PCR was running, a 2% agarose gel was made and set as described before. Once the PCR was completed, the samples were immediately loaded into the wells of the gel (6 µl sample + 2 µl loading dye) alongside a 50 bp DNA ladder. The gel was run at 80 V for 45 minutes and visualized using the gel documentation system.

3.2.9. Statistical Analysis

Statistical analysis was performed using GraphPad Prism 10.0 software and 3 x 2 contingency table. Genotypic and allelic frequencies were calculated from the obtained in-vitro analysis data. The Chi-square test and Fisher's exact test were applied to calculate the chi-square value and P-value for each case-control study, and level of association was deduced from these results.

Chapter 4 - Results

4.1. In-silico Analysis of *FSHR* and *ESR2*

4.1.1. Missense SNPs of FSHR

a. SNP Retrieval

Out of the 86,184 SNPs that have been reported for the *FSHR* gene in the NCBI database, 743 non-synonymous/missense SNPs were retrieved. Other functional classes present are inframe deletions, insertions, and indels as well as intronic SNPs, non-coding transcript variants, initiator codon variants, and synonymous SNPs.

b. Filtering Deleterious SNPs

The 743 retrieved missense SNPs were entered into SIFT tool, which predicted 111 SNPs to be deleterious. Out of these, only 35 were SNPs present in the canonical isoform 1 of the FSHR protein and were selected. Information obtained from SIFT has been summarized in Figure 4.1.

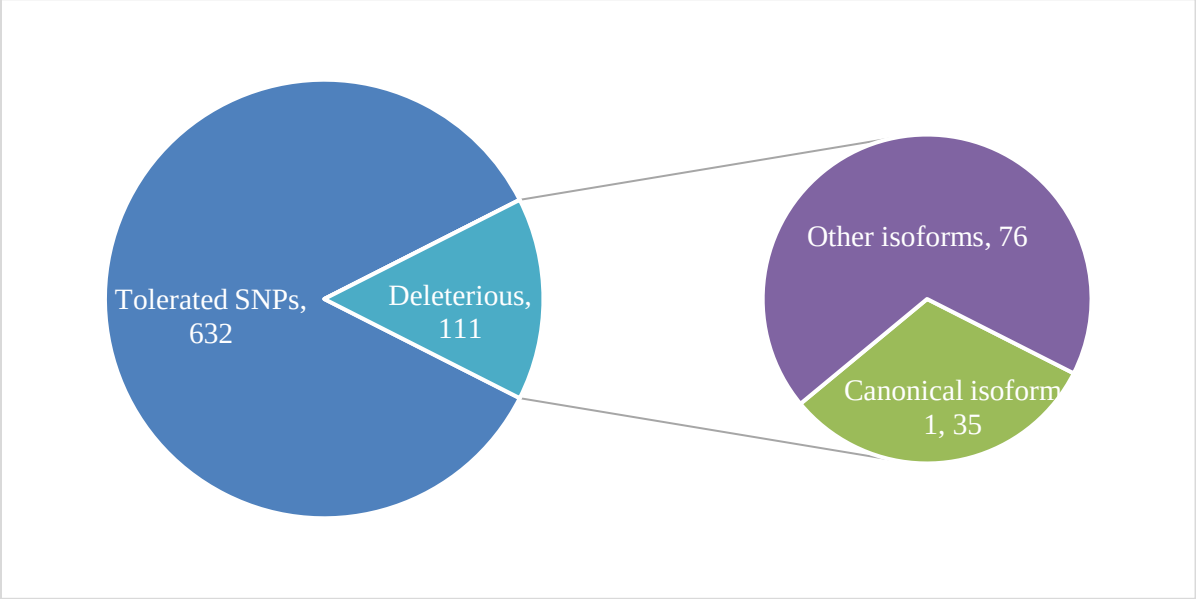


Figure 4.1. Summary of the result obtained when 743 missense SNPs of *FSHR* were entered into the SIFT tool.

The 36 potentially deleterious SNPs of isoform 1 were further entered into four different filtering tools; PolyPhen-2, PANTHER, PhD-SNP, and SNPs&GO. The results returned by each of these have been listed in Tables 4.1. and 4.2.

Table 4.1. Effects of 35 missense SNPs of *FSHR* predicted by PolyPhen-2 and PANTHER webservers.

SNP	Amino acid change	PolyPhen predicted effect	PolyPhen score	PANTHER predicted effect	PANTHER score
rs6167	S524R	probably damaging	0.993	probably benign	0.24
rs28928870	T449I	probably damaging	1	probably damaging	0.74
rs28928871	D567N	probably damaging	1	probably damaging	0.85
rs121909658	A189V	probably damaging	0.999	probably damaging	0.74
rs121909659	I160T	possibly damaging	0.894	probably benign	0.19
rs121909660	R573C	probably damaging	1	probably damaging	0.85
rs121909661	A419T	probably damaging	1	probably damaging	0.85
rs121909662	P519T	probably damaging	1	probably damaging	0.89
rs121909663	T449A	probably damaging	1	probably damaging	0.74
rs121909664	I545T	probably damaging	1	probably damaging	0.57
rs200144377	T555I	probably damaging	1	probably damaging	0.57
rs371482817	G446S	possibly damaging	0.936	probably damaging	0.89
rs386833510	P348R	probably damaging	0.97	probably damaging	0.89
rs386833511	A575V	possibly damaging	0.936	probably damaging	0.85
rs386833512	P525H	probably damaging	0.995	probably damaging	0.89
rs386833512	P587H	probably damaging	1	probably damaging	0.89
rs386833513	L601V	probably damaging	1	probably damaging	0.85
rs386833513	L337V	possibly damaging	0.882	possibly damaging	0.5
rs386833515	D224V	probably damaging	1	probably damaging	0.85
rs140415279	I132T	probably damaging	0.992	possibly damaging	0.5
rs141135052	N118K	possibly damaging	0.762	probably damaging	0.57
rs142755430	L72Q	probably damaging	1	possibly damaging	0.5
rs145775250	A88V	possibly damaging	0.534	probably damaging	0.57
rs145815060	L535H	probably damaging	1	probably damaging	0.74
rs200191550	L611P	probably damaging	1	probably damaging	0.85
rs200328782	R283W	benign	0.008	possibly damaging	0.5
rs201078200	M532T	benign	0.121	probably benign	0.27
rs201115329	C644G	benign	0.118	probably benign	0.13

rs267599403	S527L	probably damaging	1	probably damaging	0.57
rs368367400	C548Y	probably damaging	1	probably damaging	0.85
rs369583512	P45L	probably damaging	1	probably damaging	0.74
rs374191266	R557W	probably damaging	1	probably damaging	0.74
rs374788399	L252F	benign	0.001	probably benign	0.27
rs375329429	D137H	possibly damaging	0.609	probably benign	0.19
rs375475417	R573H	probably damaging	1	probably damaging	0.85

Note: For each tool, the score describes the probability of a mutation being deleterious, with damaging mutations scoring higher and benign mutations scoring closer to 0.

Table 4.2. Effects of 35 missense SNPs of *FSHR* predicted by PhD-SNP and SNPs&GO webservers.

SNP	**Amino acid change**	**PhD-SNP effect**	**Reliability index**	**SNPs&GO effect**	**Reliability index**
rs6167	S524R	Neutral	3	neutral	6
rs28928870	T449I	Disease	4	disease	3
rs28928871	D567N	Disease	7	neutral	5
rs121909658	A189V	Disease	6	disease	4
rs121909659	I160T	Disease	7	disease	1
rs121909660	R573C	Disease	7	disease	1
rs121909661	A419T	Disease	5	disease	4
rs121909662	P519T	Disease	8	disease	5
rs121909663	T449A	Neutral	1	neutral	5
rs121909664	I545T	Disease	2	disease	1
rs200144377	T555I	Disease	4	neutral	4
rs371482817	G446S	Disease	7	disease	4
rs386833510	P348R	Disease	6	disease	6
rs386833511	A575V	Disease	7	disease	0
rs386833512	P525H	Disease	1	neutral	6
rs386833512	P587H	Disease	6	disease	3
rs386833513	L601V	Disease	2	neutral	8

rs386833513	L337V	Neutral	8	neutral	8
rs386833515	D224V	Disease	9	disease	7
rs140415279	I132T	Disease	5	disease	1
rs141135052	N118K	Disease	5	neutral	6
rs142755430	L72Q	Disease	5	disease	3
rs145775250	A88V	Neutral	2	neutral	6
rs145815060	L535H	Disease	3	neutral	1
rs200191550	L611P	Disease	6	neutral	2
rs200328782	R283W	Disease	8	disease	0
rs201078200	M532T	Disease	1	neutral	8
rs201115329	C644G	Disease	5	disease	0
rs267599403	S527L	Disease	1	neutral	6
rs368367400	C548Y	Disease	9	disease	8
rs369583512	P45L	Disease	4	disease	3
rs374191266	R557W	Disease	6	neutral	0
rs374788399	L252F	Neutral	1	disease	1
rs375329429	D137H	Disease	6	neutral	7
rs375475417	R573H	Disease	4	disease	1

Note: For each software, the reliability index is a score measuring how reliable the predicted result is, with a higher score being more reliable and 0 being least reliable.

SNPs predicted to be deleterious by all four tools, and having a reliability score greater than or equal to 4 were selected. This narrowed the 35 SNPs to 7, which are listed in Table 4.3.

Table 4.3. Seven shortlisted missense SNPs of *FSHR* after filtering.

SNP	**Amino acid change**
rs121909662	P519T
rs386833510	P348R
rs386833515	D224V
rs368367400	C548Y

rs121909658	A189V
rs121909661	A419T
rs371482817	G446S

c. Effect of Mutation on Protein Structure and Stability

The seven filtered SNPs were entered into the Mutpred2 webserver, which returned a result describing the potential effect each mutation could have on the structure and function of the protein. A general score (g) gives the probability of a mutation being pathogenic if the value is greater than or equal to 0.5. *Pr* is the probability of an alteration in structure/function happening because of the given substitution and can be evaluated for each individual alteration. The value *P* is the probability of a benign substitution (from the MutPred2 training set) having *Pr* greater than or equal to that of the given substitution; a lower *P* value corresponds to a higher chance of a specific alteration to be pathogenic. Two predicted alterations having the same *Pr* value does not mean they will have the same *P* value, since the *P* value is specific for each property. The result obtained has been displayed in Table 4.4.

Table 4.4. Results obtained from the Mutpred2 webserver for the seven filtered missense SNPs of *FSHR*.

SNP	Amino acid change	Score (g)	Predicted Effect
rs121909662	P519T	0.713	Altered Transmembrane protein (Pr = 0.26 | P = 1.1e-03); Gain of Disulfide linkage at C517 (Pr = 0.11 | P = 0.04)
rs386833510	P348R	0.768	Altered Transmembrane protein (Pr = 0.28 | P = 6.5e-04); Loss of Disulfide linkage at C346 (Pr = 0.25 | P = 5.9e-03)
rs386833515	D224V	0.738	Altered Transmembrane protein (Pr = 0.15 | P = 0.01)
rs368367400	C548Y	0.863	Loss of Helix (Pr = 0.28 | P = 0.03); Altered Ordered interface (Pr = 0.25 | P = 0.02)

rs121909658	A189V	0.68	Altered Transmembrane protein (Pr = 0.30 | P = 1.5e-04); Loss of N-linked glycosylation at N191 (Pr = 0.05 | P = 0.02)
rs121909661	A419T	0.635	Altered Transmembrane protein (Pr = 0.26 | P = 9.7e-04); Altered Ordered interface (Pr = 0.24 | P = 0.04); Altered Metal binding (Pr = 0.18 | P = 0.03)
rs371482817	G446S	0.841	Gain of Catalytic site at C442 (Pr = 0.29 | P = 1.6e-03); Gain of Relative solvent accessibility (Pr = 0.26 | P = 0.03); Altered Transmembrane protein (Pr = 0.19 | P = 6.8e-03); Gain of Disulfide linkage at C442 (Pr = 0.19 | P = 0.02)

Note: g – General probability of substitution being pathogenic; *Pr* – Probability of occurrence of the predicted change in property; *P* – Probability of the predicted change being benign.

Effect on the stability of the protein by the occurrence of each point mutation was predicted by the I-mutant2.0 webserver. Based on the sign of change in the change of Gibbs free energy (DDG, delta delta G, or ΔΔG), positive DDG referring to increased stability and negative DDG referring to decreased stability, the effect is predicted. A reliability index is also given, with higher scores corresponding to more reliable results. The results have been listed in Table 4.5.

Table 4.5. Effect on the stability of FSHR protein due to each missense SNP predicted by the I-Mutant2.0 webserver.

SNP	**Amino acid change**	**Predicted effect on stability**	**Reliability index**	**DDG**
rs121909662	P519T	Decrease	8	-1.69
rs386833510	P348R	Decrease	7	-0.88
rs386833515	D224V	Decrease	4	-1.28
rs368367400	C548Y	Decrease	3	0.17
rs121909658	A189V	Decrease	4	-0.73

rs121909661	A419T	Decrease	3	-0.22
rs371482817	G446S	Decrease	7	-1.31

Note: Higher score for reliability index refers to a more reliable result; DDG – delta delta G, change in the change of Gibbs free energy of a protein when point mutation is present.

The evolutionary conservation analysis of each residue in the FSHR protein was performed by the ConSurf webserver. Mutations in highly conserved residues have a higher chance of being pathogenic as compared to residues that have lower scores. The conservation score of each of the seven target residues being mutated has been listed in Table 4.6., along with its status prediction. All seven residues were found to be highly conserved which shows that they may have very important roles in determining the structure and function of the protein.

Table 4.6. Conservation scores of the seven target residues being substituted due to filtered missense SNPs of *FSHR*.

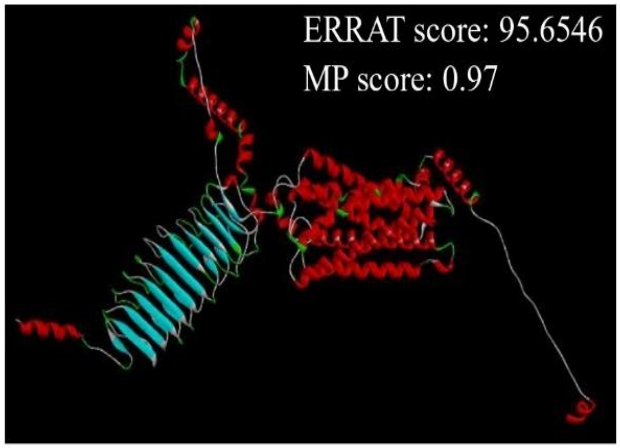
SNP	**AA**	**Color Score**	**Prediction**
rs121909662	P519T	9	structural, highly conserved, and buried
rs386833510	P348R	9	functional, highly conserved, and exposed
rs386833515	D224V	9	functional, highly conserved, and exposed
rs368367400	C548Y	9	structural, highly conserved, and buried
rs121909658	A189V	9	structural, highly conserved, and buried
Rs121909661	A419T	9	structural, highly conserved, and buried
rs371482817	G446S	9	structural, highly conserved, and buried

Note: The numbers on the conservation scale range from 1 to 9, with 9 being the most highly conserved. A score of '?' is lower than 1 and refers to highly variable residues.

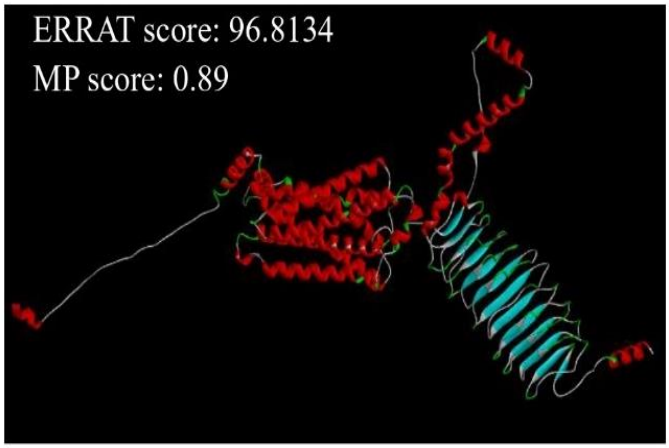
d. Protein Modelling and Model Validation

trRosetta was used for building the seven protein models. Each model of the highest confidence level was selected and uploaded to ERRAT and MolProbity for structural validation. ERRAT score is on a scale of 0 to 100, with higher scores being more favorable.

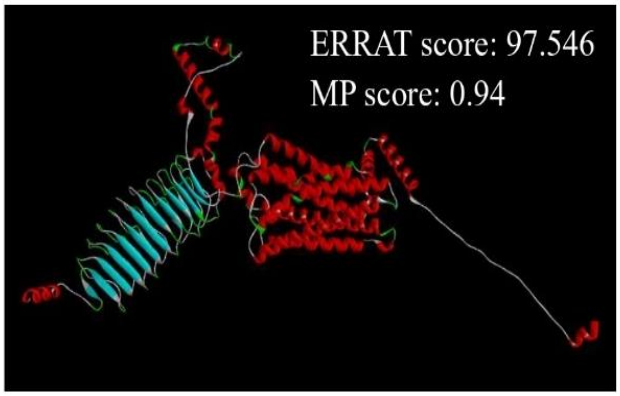
MolProbity returns a detailed report on the quality of the protein, giving quantitative information about poor and favored rotamers, Ramachandran outliers, bad bonds and angles, and other variables. A single value, the MolProbity score, is also given to summarize and condense the result into one value. A lower MolProbity score is given for better quality protein models. All seven of the trRosetta modelled proteins returned good scores when validated using the two webservers. The models, and their validation scores, are given in Figure 4.2.



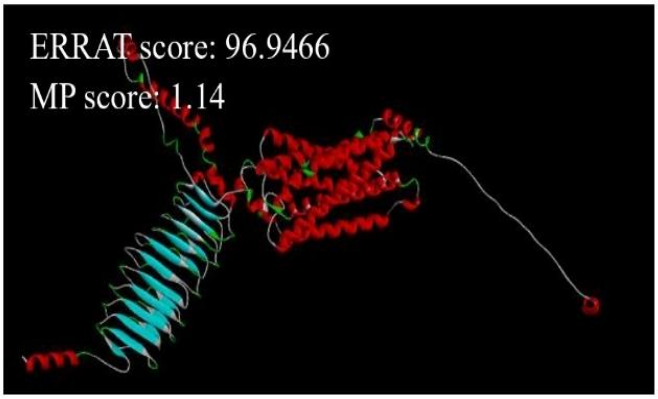
a) Wild FSHR protein



b) Mutant P519T



c) Mutant P348R



d) Mutant D224V

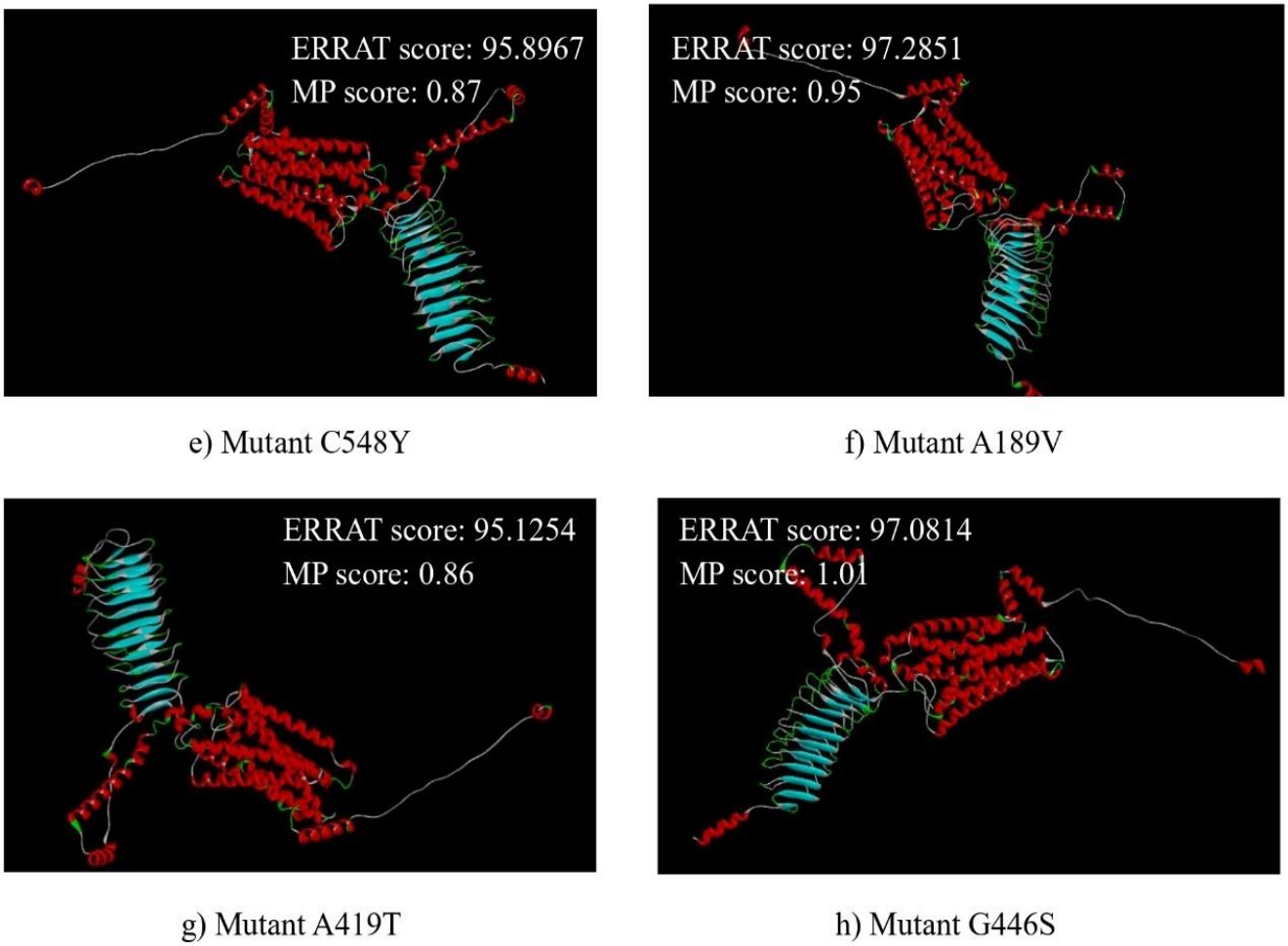
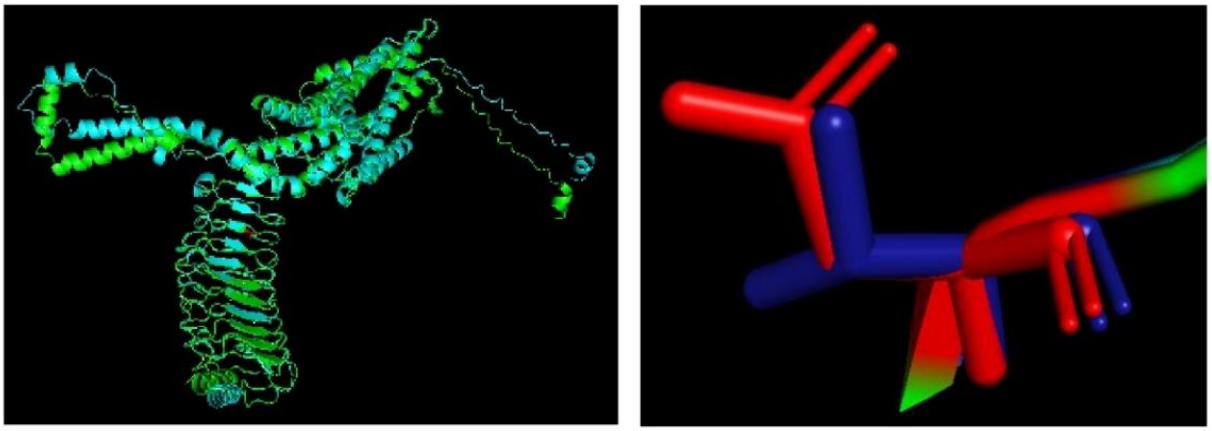


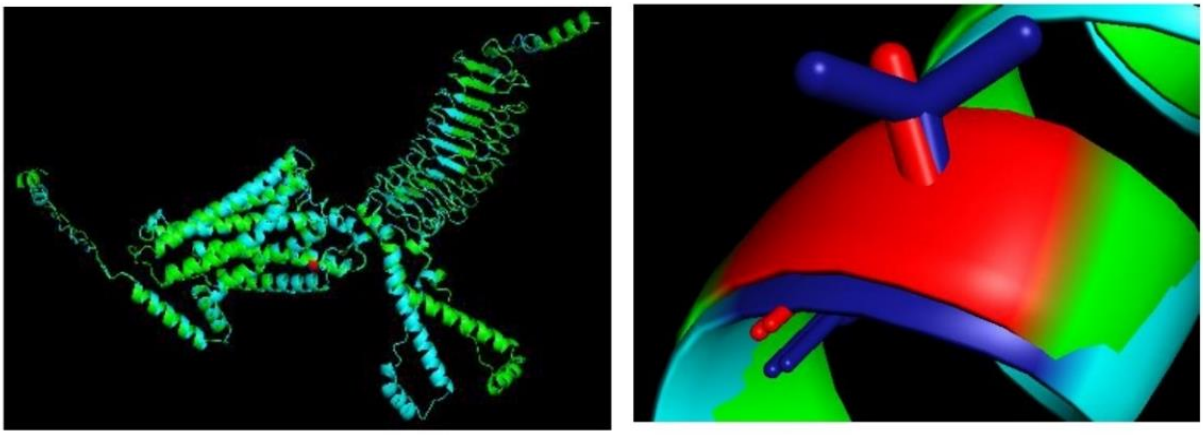
Figure 4.2. FSHR protein models predicted by trRosetta and their validation scores. MP score – MolProbity score.

e. Protein Superimposition

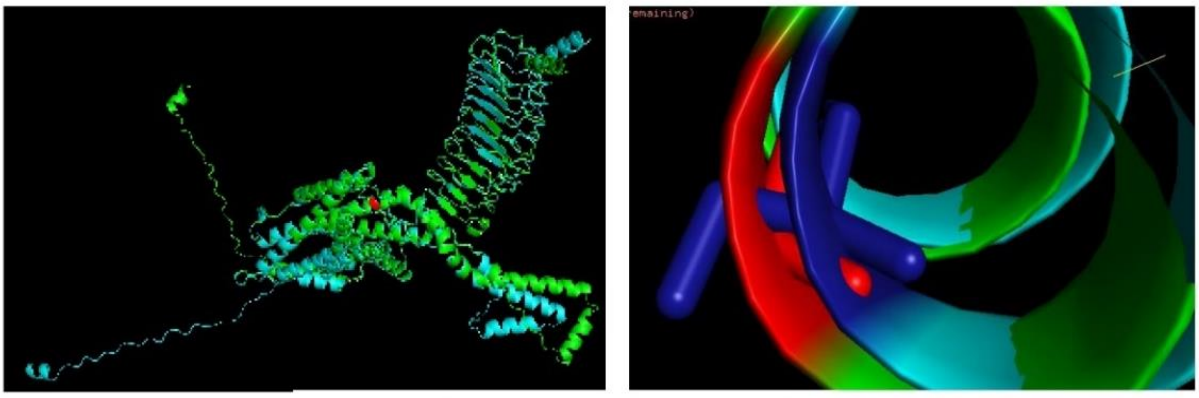
Each mutant protein was superimposed onto the wild protein using the TM-align software and the RMSD values were noted. Table 4.7. gives the RMSD values for each of the seven SNPs. Three superimpositions that gave the highest RMSD values were selected and visualized in PyMOL 2.5 and are shown in Figure 4.3.



a) D224V (Aspartic acid to Valine)



b) A419T (Alanine to Threonine)



c) G446S (Glycine to Serine)

Figure 4.3. The three short-listed SNPs of *FSHR* superimposed with the wild protein and visualized in PyMOL 2.5.

Table 4.7. RMSD (Root mean square deviation) values returned when each of the seven mutant protein models was superimposed onto the wild FSHR protein.

Protein-protein superimposition	RMSD score
Wild + P519T	1.87
Wild + P348R	1.89
Wild + D224V	2.33
Wild + C548Y	1.87
Wild + A189V	1.66
Wild + A419T	2.07
Wild + G446S	2.05

Note: The three superimpositions with highest RMSD scores have been highlighted.

f. Post-translational Modifications

The effect of the shortlisted SNPs on PTMs of the FSHR protein were tested using different webservers, and the results have been described below.

i. Phosphorylation

Phosphorylated sites of the wild and mutant proteins were checked using GPS 6.0. The wild FSHR protein was predicted to be phosphorylated at 27 sites by 4 different kinases. Mutant D224V resulted in a loss of phosphorylation at 2 sites: serine residue at position 232, and tyrosine residue at position 236. The other two SNPs did not affect phosphorylated sites of the protein.

ii. Glycosylation

NetNGlyc 1.0 predicted five sites in the FSHR wild protein to be N-glycosylated, and these were not affected by any of the target SNPs. O-glycosylation was predicted using NetOGlyc 4.0 and predicted 12 sites to be O-glycosylated in the wild protein. D224V resulted in loss of glycosylation at residue 38, 235, and 237, and the gain of O-glycosylation at positions 232, 312, and 661. A419T resulted in loss of glycosylation at positions 235 and 38 and gain at position 662. Finally, G446S was predicted to show a loss of glycosylation at position 38 and gain at position 232.

For each SNP, a zoomed-out image showing the whole proteins is given first, followed by an image zoomed into the amino acid change. Color key: Green – wild protein; Cyan – mutant protein; Red – wild residue; Dark blue – mutant residue

iii. Methylation

No sites were predicted to be methylated in the FSHR protein when checked by GPS-MSP 1.0.

iv. Ubiquitination

No sites were predicted to be ubiquitinated by the GPS-Uber webserver. None of the target SNPs had any effect on this status of the protein.

v. SUMOylation

Three short regions (residues 341-345, 173-177, and 221-225) in the FSHR wild protein were predicted to be SIMs by GPS-SUMO 2.0. One residue at position 74 was predicted to be SUMOylated. However, none of these PTMs were affected by the three shortlisted SNPs.

g. Protein-protein Interactions

The interactions between FSHR and other proteins inside the cell were evaluated using the STRING software. Co-expression of *FSHR* with *AMH*, *CYP19A1,* and *GNB1* were predicted, and experimentally determined interactions with *APPL1, CGA, CYP19A1, FOXO1,* and *FSHB* were also confirmed. The interactions of FSHR with other proteins have been summarized in Figure 4.4, obtained from the STRING webserver.

h. Single Nucleotide Polymorphism of *FSHR* (rs6166)

The result obtained for rs6166 has been summarized in Table 4.8. rs6166 was predicted to be a neutral polymorphism by all filtering tools, and as a variable (not conserved) residue by ConSurf. It was thus not continued for protein modelling and superimposition.

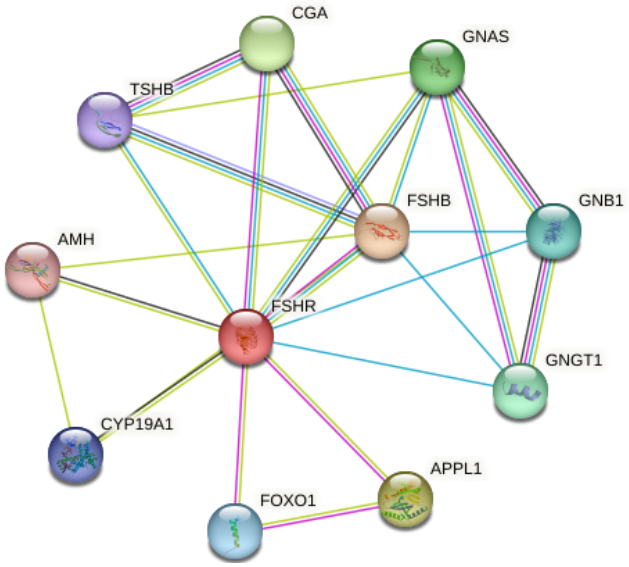


Figure 4.4. FSHR protein-protein interactions predicted by STRING webserver.

Table 4.8. In-silico analysis of literature-selected SNP rs6166.

Tool	**Score**	**Prediction**
SIFT	0.707	Tolerated
PolyPhen2	0.00	Benign
PANTHER	0.27	Probably benign
PhD-SNP	7 (RI)	Neutral
SNPs&GO	9 (RI)	Neutral
MutPred2	0.102	Not pathogenic
I-Mutant 2.0	6 (RI)	Decreased stability
ConSurf	Color score 2	Variable residue

Note: RI – Reliability index

4.1.2. Non-coding Transcript Variants of *ESR2*

Of the 45,132 SNPs of *ESR2,* 1,752 are non-coding transcript variants which were downloaded and further filtered for 3' and 5' variants. These 902 variants were entered into the RegulomeDB server and analyzed for functionally significant polymorphisms. The server ranked all the SNPs based on reported evidence of functional significance and

probability of being regulatory variants. Only the variants with a rank of 1a-1f or 2a (high probability of functional significance) were selected and have been listed in Table 4.9.

Table 4.9. 3' and 5' UTR SNPs of ESR2 filtered using RegulomeDB.

SNP ID	Rank	Probability
rs4986938	1b	1
rs553473793	2a	0.89591
rs540586598	1f	0.55436
rs919050282	2a	0.3145
rs1049589193, rs1174614695	2a	0.829
rs1253998613	2a	0.74451
rs1386817766	2a	0.96817
rs1309024818	2a	1
rs1256120	1f	0.70823

Note: A lower rank is given to an SNP with more supporting evidence of functional significance. Probability score is higher for variants that have a higher chance of being present in regulatory regions.

The 902 variants were also entered into the SNPinfo webserver for prediction of functionally significant SNPs. The result obtained has been shown in Table 4.9. SNPs that were predicted to be important by both tools are rs1256120 and rs4986938.

Table 4.10. 3' and 5' UTR SNPs of ESR2 with alternative alleles that have different functional consequences, as predicted by SNPinfo.

SNP ID	Functional prediction	Conservation score	Regulatory potential score
rs11622991	-	0	NA
rs1255998	miRNA	0	0
rs1256054	Splicing (ESE or ESS), Splicing (abolish domain)	0.998	0.352901
rs1256120	TFBS	0	0

rs2094914	-	0	NA
rs28440970	miRNA	0.333	0
rs45624541	-	NA	NA
rs4986938	miRNA	0.001	0
rs58262369	miRNA	NA	NA
rs8018687	miRNA	0.001	0
rs928554	miRNA	0.002	0.051291

Note: miRNA – micro RNA; ESE – exonic splicing enhancer; ESS – exonic splicing silencer; TFBS – transcription factor binding site.

4.2. In-vitro Analysis of FSHR (rs6166), WNT4 (rs7521902), and ESR2 (rs4986938)

This in-vitro study was performed to determine the level of association between three previously reported SNPs (rs6166, rs7521902, and rs4986938) and the diagnosis of endometriosis in females of reproductive age. 38 samples were collected from Benazir Bhutto Hospital, P.I.M.S. Hospital, and ASAB diagnostic lab, of which 19 were the diseased cases and 19 were controls. Phenol-chloroform extraction 2-day protocol was used to extract whole genomic DNA from each sample and was stored. Allele-specific primers for each of the three SNPs were designed and ordered, and ARMS-PCR was performed. Finally, gel electrophoresis was run after amplification to visualize and document the results. Statistical analyses were performed on each of the results obtained to reach a conclusion whether the polymorphisms are significantly associated with the occurrence of endometriosis or not.

4.2.1. Single Nucleotide Polymorphism of *FSHR* – rs6166

rs6166 is a missense SNP in exon 10 of the *FSHR* gene on chromosome 2, involving a C/T nucleotide change that belies a change from serine to asparagine in the translated protein. After ARMS-PCR was performed, cases and controls homozygous for CC or TT, or heterozygous CT were noted down. Figure 4.5. displays an electropherogram with seven case samples and seven control samples run to test for nucleotide change in rs6166. The size of the amplicon was 372 bp which was attested using a 50 bp DNA ladder in the first lane of the gel. Positive and negative controls for each primer set were also applied. Allelic

frequencies (frequency of C and T in the cases versus controls) and genotype frequencies (frequencies of CC, CT, and TT in the cases versus controls) were calculated and have been shown in Table 4.11. Figure 4.6. shows the genotypic distribution in cases and controls. The numbers of samples with each genotype were entered in GraphPad Prism 10.0, Chi-square test and Fisher's exact test were applied and the χ^2 and P-values were noted (Table 4.8.). A P-value of 0.5052 being greater than 0.05 points towards a result that is non-significant and χ^2 value of 1.366 shows that the distribution of the alternative alleles is fairly equal and does not stray much from the Hardy-Weinberg predicted expected frequencies.

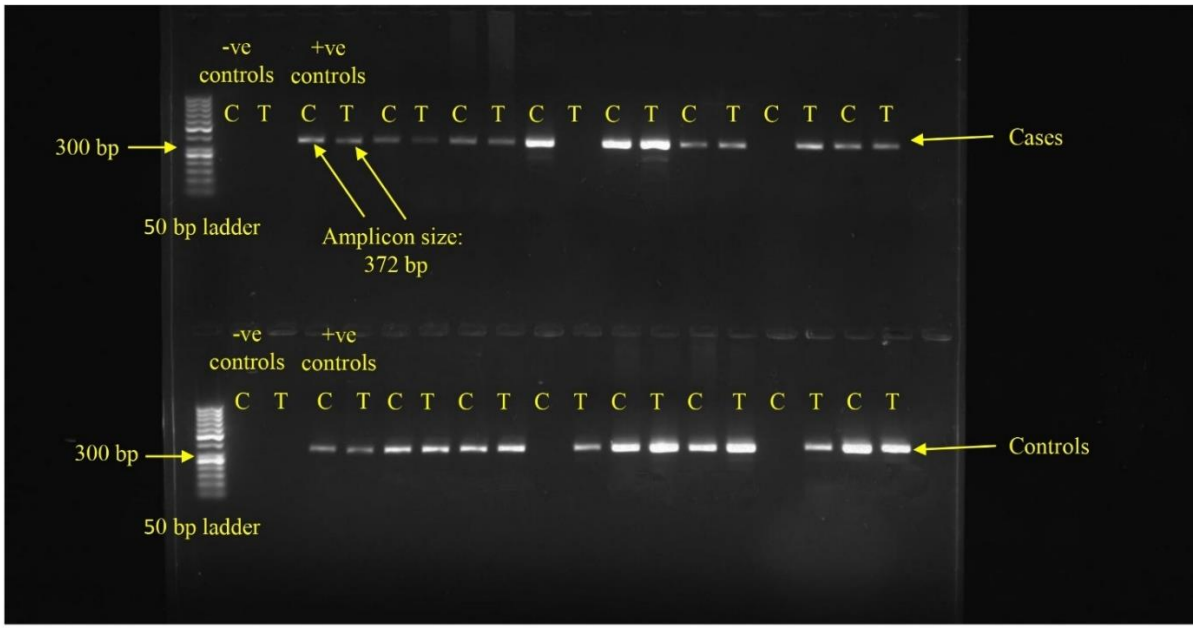


Figure 4.5. Gel image displaying ARMS-PCR results of seven diseased samples and seven healthy controls tested for C/T alleles of rs6166.

Table 4.11. Allelic and genotypic frequencies in subjects analyzed for rs6166.

	Allele frequency		**Genotype frequency**			χ^2	**P value**	**Degrees of freedom**
	C	**T**	**CC**	**CT**	**TT**			
Cases (n=19)	0.52	0.47	0.10	0.89	0.05	1.366	0.5052	2

Controls (n=19)	0.44	0.55	0.05	0.79	0.16			

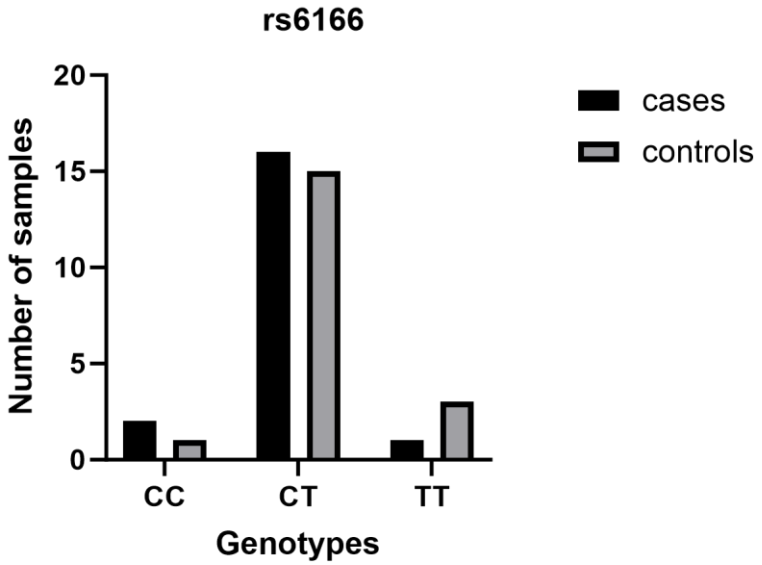


Figure 4.6. Genotypic distribution in 19 cases (endometriosis patients) and 19 controls for rs6166.

4.2.2. Single Nucleotide Polymorphism of *WNT4* – rs7521902

rs7521902 is an intron variant present upstream of the protein-coding *WNT4* gene in a regulatory region on chromosome 1. It involves a C/A nucleotide change that affects the activity of the enhancer region it is present in. After ARMS-PCR was performed, cases and controls homozygous for CC or AA, or heterozygous CA were noted down. Figure 4.7. displays an electropherogram with seven case samples and seven control samples run to test for nucleotide change in rs7521902. The size of the amplicon was 328 bp which was attested using a 50 bp DNA ladder in the first lane of the gel. Positive and negative controls for each primer set were also applied. Allelic frequencies (frequency of C and A in the cases versus controls) and genotype frequencies (frequencies of CC, CA, and AA in the cases versus controls) were calculated and have been shown in Table 4.12. Figure 4.8. shows the genotypic distribution in cases and controls. The numbers of samples with each genotype were entered in GraphPad Prism 10.0, Chi-square test and Fisher's exact test were applied and the χ^2 and P-values were noted (Table 4.9.). A P-value of 0.2185 being greater than

0.05 points towards a result that is non-significant and χ^2 value of 3.042 shows that the distribution of the alternative alleles is fairly equal and does not stray much from the Hardy-Weinberg predicted expected frequencies.

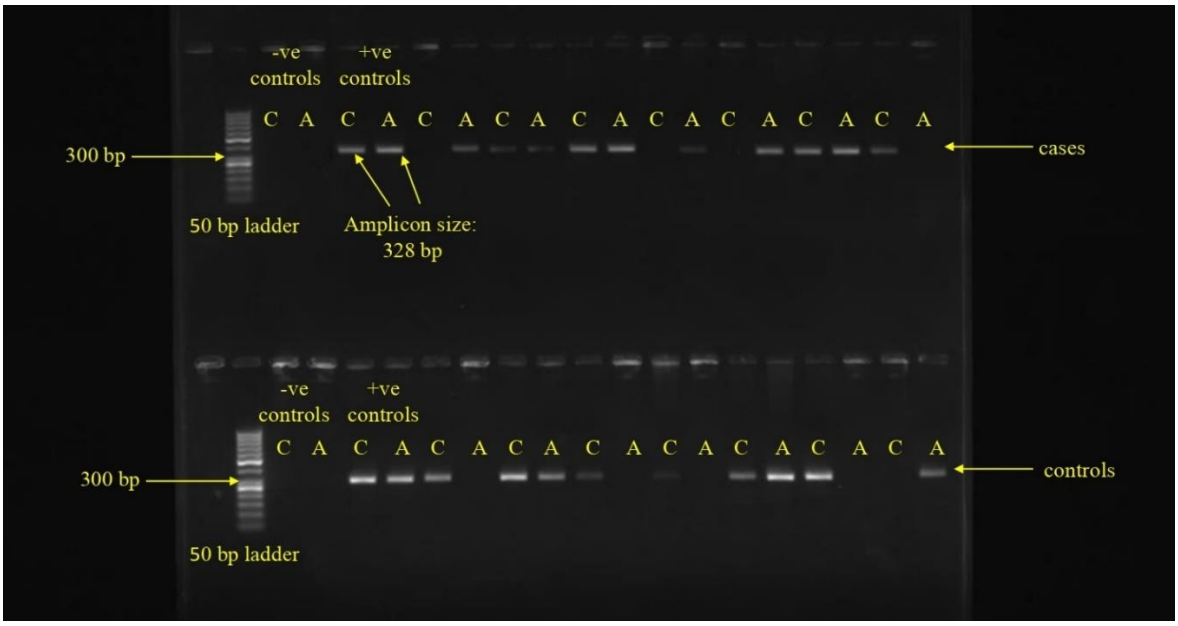


Figure 4.7. Gel image displaying ARMS-PCR results of seven diseased samples and seven healthy controls tested for C/A alleles of rs7521902.

Table 4.12. Allelic and genotypic frequencies in subjects analyzed for rs7521902.

	Allele frequency		**Genotype frequency**			χ^2	**P value**	**Degrees of freedom**
	C	**A**	**CC**	**CA**	**AA**			
Cases (n=19)	0.53	0.47	0.31	0.42	0.26	3.042	0.2185	2
Controls (n=19)	0.74	0.26	0.57	0.31	0.10			

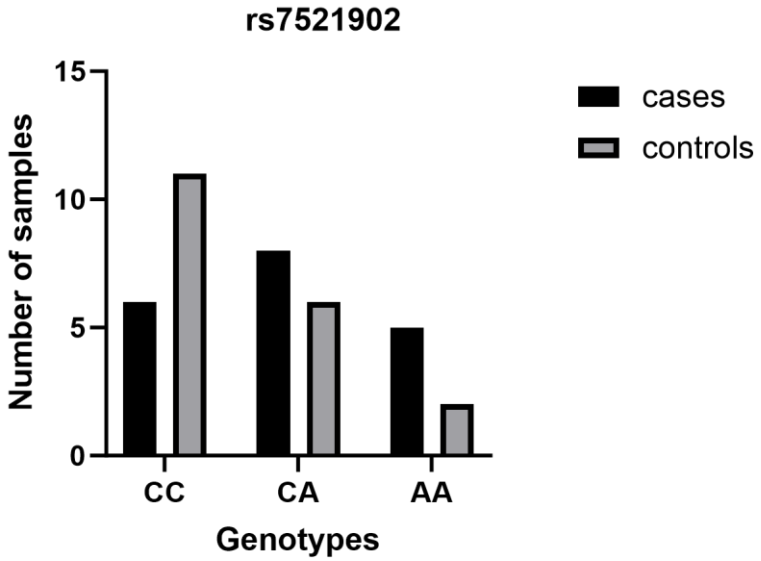


Figure 4.8. Genotypic distribution in 19 cases (endometriosis patients) and 19 controls for rs7521902.

4.2.3. Single Nucleotide Polymorphism of *ESR2* – rs4986938

rs4986938 is present in the 3' untranslated region of *ESR2* gene on chromosome 14 and involves a C/A nucleotide change. After ARMS-PCR was performed, cases and controls homozygous for CC or TT, or heterozygous CT were noted down. Figure 4.9. displays an electropherogram with seven case samples and seven control samples run to test for nucleotide change in rs4986938. The size of the amplicon was 245 bp which was attested using a 50 bp DNA ladder in the first lane of the gel. Positive and negative controls for each primer set were also applied. Allelic frequencies (frequency of C and T in the cases versus controls) and genotype frequencies (frequency of CC, CT, and TT in the cases versus controls) were calculated and have been shown in Table 4.13. Figure 4.10. shows the genotypic distribution in cases and controls. The numbers of samples with each genotype were entered in GraphPad Prism 10.0, Chi-square test and Fisher's exact test were applied and the χ^2 and P-values were noted (Table 4.10.). A P-value of 0.8338 being greater than 0.05 points towards a result that is non-significant and χ^2 value of 0.3636 shows that the distribution of the alternative alleles is fairly equal and does not stray much from the Hardy-Weinberg predicted expected frequencies.

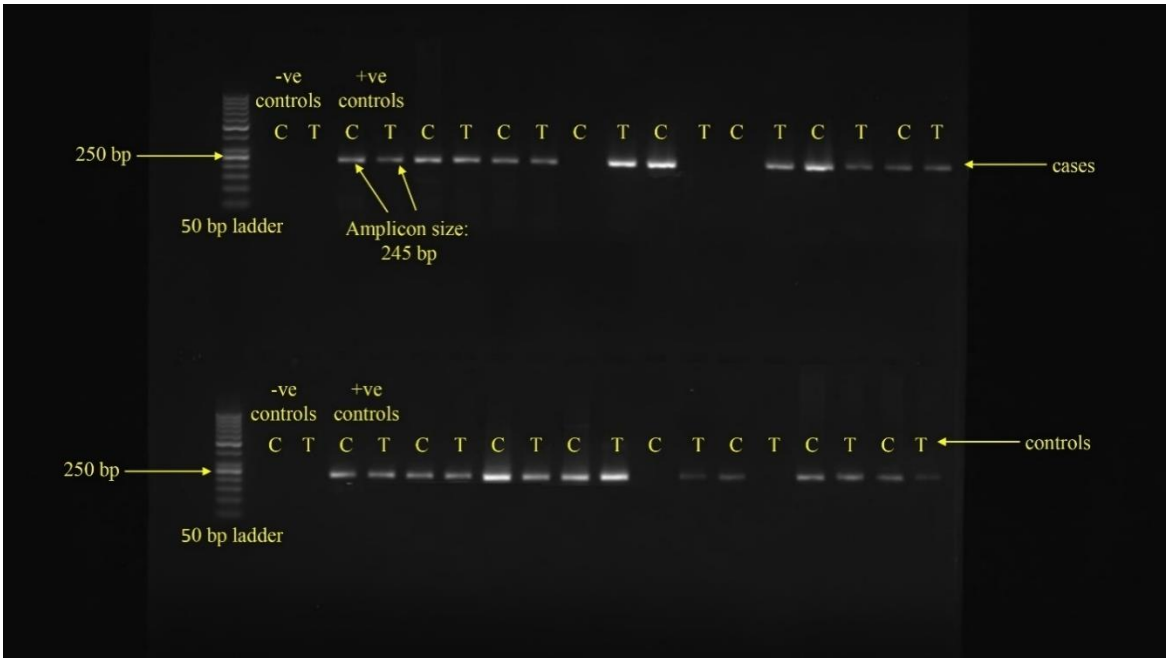


Figure 4.9. Gel image displaying ARMS-PCR results of seven diseased samples and seven healthy controls tested for C/T alleles of rs4986938.

Table 4.13. Allelic and genotypic frequencies in subjects analyzed for rs4986938.

	Allele frequency		**Genotype frequency**			**χ^2**	**P value**	**Degrees of freedom**
	C	**T**	**CC**	**CT**	**TT**			
Cases (n=19)	0.47	0.52	0.05	0.84	0.10	0.3636	0.8338	2
Controls (n=19)	0.5	0.5	0.05	0.89	0.05			

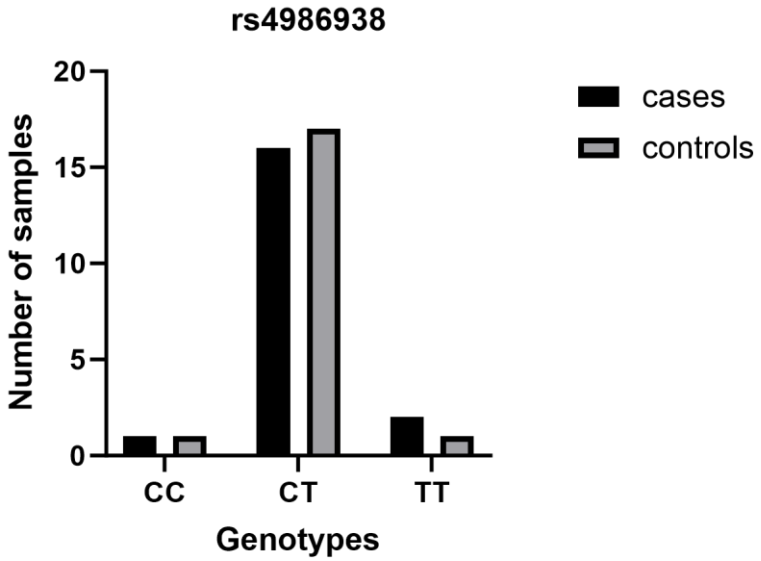


Figure 4.10. Genotypic distribution in 19 cases (endometriosis patients) and 19 controls for rs4986938.

Chapter 5 - Discussion

Endometriosis is the abnormal growth of endometrial tissue (tissue lining the inside of the uterus) in ectopic locations, causing severe and chronic pain (Chapron et al., 2019) It affects 2-10% females of reproductive age globally and causes several complications that severely lowers the quality of life. (Parasar et al., 2017). No clear causes of endometriosis have been determined, although different risk factors have been identified that include reproductive and lifestyle aspects as well as the most widely accepted theory: Sampson's retrograde menstruation model (Sampson, 1927). Recent advances have aimed to discover the genetic causes of the disease with little success because of its complex nature. Multiple mutations and polymorphisms that have been identified act as susceptibility factors rather than direct causes, and the results of association are not always identical in reproduced studies. From the existing literature on such association studies, three single nucleotide polymorphisms were selected that have been reported to be linked with endometriosis: rs6166, a missense SNP present in exon 10 of *FSHR* (Liaqat et al., 2021); rs7521902 , an intronic variant located upstream of *WNT4* in a CTCK-binding site (Matalliotakis et al., 2017); and rs4986938 located in the 3'UTR of *ESR2* (Liaqat et al., 2013). These three genes have important roles in the normal sexual development of females and have been implicated in endometriosis as well as other reproductive disorders.

In this study, the selected SNPs were first analyzed in-silico to understand their predicted effects on the respective proteins. All reported missense SNPs of *FSHR,* and 3' and 5' UTR variants of *ESR2* were also analyzed to identify new pathogenic polymorphisms that could be risk factors in causing endometriosis. In-vitro association study was next performed for the three SNPs to confirm previously reported results using samples taken from hospitals in Rawalpindi, Pakistan.

From the 86,184 SNPs reported for *FSHR* on NCBI, 743 missense SNPs were retrieved to perform an in-silico analysis. These SNPs were entered into SIFT tool which filtered out 111 deleterious SNPs. These were analyzed and only the SNPs present in the canonical isoform 1 were selected for further analysis. The selected 36 SNPs were filtered through a series of webservers: PolyPhen-2, PANTHER, PhD-SNP, and SNPs&GO. SNPs

that were predicted to be pathogenic by every one of these tools with a reliability score of ≥4 were picked, resulting in 7 SNPs. The effect of these shortlisted SNPs on the protein stability, structure, and function was then analyzed. The Mutpred2 webserver predicted different alterations in the protein that would be caused by each mutation. I-mutant2.0 was used to predict change in the stability of the protein, with each of the seven SNPs decreasing the stability. Evolutionary conservation analysis was performed using ConSurf, which shows the conservation status of each residue in a wild protein. If a mutation occurs in a residue that is highly conserved, it has a higher chance of being pathogenic because conserved residues usually have very important roles in determining the structure and function of a protein. All seven of the SNPs that were shortlisted up to this point were highly conserved, with two being functional and exposed, and five being structural and buried. Protein models with each of the seven SNPs were built using trRosetta and their quality validated through ERRAT, and MolProbity. TM-Align was used to superimpose mutant proteins onto the wild protein and the RMSD values were noted in each case. The three mutant models with the highest RMSD were finalized as potentially the most pathogenic point mutations. These are rs386833515 (D224V), rs121909661 (A419T), and rs371482817 (G446S). rs386833515, involving a change from aspartic acid (D) to valine (V), is especially of concern since D is a hydrophilic and acidic amino acid while V is highly hydrophobic and is often found buried deep within the protein. The D residue present on position 224 is functional and exposed as predicted by ConSurf, and so a replacement by a hydrophobic amino acid would probably affect the structure and function of the protein. Post-translational modification analysis of the D224V substitution also shows that it affects the phosphorylation (loss at position 232 and 236) and O-glycosylation (loss at position 38, 235, and 237, and gain at position 232, 312, and 661) of the mutant protein. The alanine (A) to threonine (T) change in rs121909661 involves a hydrophobic to hydrophilic amino acid substitution that could have structural and functional consequences, although both amino acids are neutral (neither acidic nor basic) and can be present inside or outside the protein. Glycine (G) is a small, neutral, and amphiphilic amino acid and so is serine (S), however the difference in molar mass and size of the two amino acids, as well as the flexibility of glycine could be a possible factor in the pathogenicity of

G446S substitution. A wet lab analysis would be needed to confirm these results in-vitro. The selected SNP rs6166, when analyzed through the in-silico tools, was predicted to be neutral. Our in-vitro association study is in agreement with these results, however further experimentation is required for validation.

The selected SNP rs4986938 and other 3' and 5' UTR variants of *ESR2* were entered into the RegulomeDB webserver for analysis. The tool is used to identify variants of functional significance located in non-coding regions of the genome. Each SNP entered was returned with a rank from 1-7 based on the probability of it being functionally significant (1a being most significant). A probability score from 0-1 was also given to describe the chance of the SNP being present in a regulatory region. 9 variants of the highest ranks were selected. 3' and 5' UTR variants were also entered into SNPinfo which predicted the SNPs that would have functional effects because of replacement of nucleotide. 11 SNPs were predicted with different predicted consequences. Only the SNPs that were predicted to have significance by both tools were finalized. These are rs1256120 (present in a transcription factor binding site) and rs4986938 (present in miRNA sequence region).

The in-vitro analysis involved designing allele-specific primers for the three selected SNPs, and then performing ARMS-PCR for amplification. The samples for study were taken from hospitals in Rawalpindi, Pakistan and the ASAB diagnostic lab in NUST, Islamabad, Pakistan. Statistical tools were applied to the obtained results to make a conclusion about the level of association between selected SNPs and incidence of endometriosis. Results showed that there is no significant association between any of the polymorphisms and endometriosis patients of Pakistani origin. Negative association for rs7521902 is consistent with results previously reported in Italian (Pagliardini et al., 2013), Brazilian (Mafra et al., 2015), Chinese (Z. Wu et al., 2015), and Sardinian populations (Angioni et al., 2020), and inconsistent with the result reported by GWAS conducted in Japanese and European populations, and an association analysis performed in a Greek population (Matalliotakis et al., 2017). rs6166 and rs4986938 were reported to be associated with endometriosis in a study conducted previously in Pakistani patients (Liaqat et al., 2013, 2021) but were not found to be associated in this study. Small sample size and

limited information on disease progression and diagnosis in patients are possible limitations of the current study, and different results might be obtained if the study is repeated with a larger sample size.

Conclusion

The association of estrogen-dependent disease endometriosis with Pakistani patients was investigated in this study. The three selected SNPs (rs6166 of *FSHR*, rs7521902 near *WNT4*, and rs4986938 of *ESR2*) showed no significant association in the selected subjects. Frequency of all three mutant alleles did not deviate significantly from the frequency of wild type alleles in selected samples. However, this study needs to be replicated on a larger sample size for affirmation. In-silico analysis of all reported missense SNPs of *FSHR* gene identified three point-mutations that could be potentially pathogenic: rs386833515, rs121909661, and rs371482817. Analysis of reported 3' and 5' UTR variants of *ESR2* identified rs1256120 and rs4986938 as functionally significant with high regulatory potential. Wet-lab experimentation is required to test the significance of these SNPs in causing disease.

The current study should be replicated in a larger sample size to confirm the relationship between the SNPs analyzed in this study and endometriosis. Using DNA sequencing side-by-side with ARMS-PCR is a good way of validating the results of the association analysis. Expression analysis of protein-coding genes in cases and controls would elucidate the effect of disease on protein expression. Wet-lab experimentation to validate the relationship between shortlisted SNPs of *FSHR* and *ESR2,* and endometriosis need to be conducted, and these mutations could be used as molecular targets for in-silico drug design. Identification of SNPs that can act as biomarkers would help in the early and non-invasive diagnosis of endometriosis. Further understanding of the molecular mechanisms of selected genes/proteins and others identified by GWAS will aid in management and treatment of the disease.

References

Abdel-Aziz, A., El-Sokkary, A., El-Refaeey, A., El-Sokkary, M., Osman, H., & El-Saeed, R. (2015). Association between Follicle Stimulating Hormone Receptor (FSHR) Polymorphism and Polycystic Ovary Syndrome among Egyptian Women. *International Journal of Biochemistry Research & Review*, *5*(3), Article 3. https://doi.org/10.9734/IJBCRR/2015/13896

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, *Chapter 7*, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Agarwal, S. K., Chapron, C., Giudice, L. C., Laufer, M. R., Leyland, N., Missmer, S. A., Singh, S. S., & Taylor, H. S. (2019). Clinical diagnosis of endometriosis: A call to action. *American Journal of Obstetrics and Gynecology*, *220*(4), 354.e1-354.e12. https://doi.org/10.1016/j.ajog.2018.12.039

American Society for Reproductive Medicine. (1997). Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertility and Sterility*, *67*(5), 817–821. https://doi.org/10.1016/S0015-0282(97)81391-X

André, G. M., Martins Trevisan, C., Pedruzzi, I. N., Fernandes, R. F. M., Oliveira, R., Christofolini, D. M., Bianco, B., & Barbosa, C. P. (2018). The Impact of FSHR Gene Polymorphisms Ala307Thr and Asn680Ser in the Endometriosis Development. *DNA and Cell Biology*, *37*(6), 584–591. https://doi.org/10.1089/dna.2017.4093

Angioni, S., D'Alterio, M. N., Coiana, A., Anni, F., Gessa, S., & Deiana, D. (2020). Genetic Characterization of Endometriosis Patients: Review of the Literature and a Prospective Cohort Study on a Mediterranean Population. *International Journal of Molecular Sciences*, *21*(5), Article 5. https://doi.org/10.3390/ijms21051765

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(Web Server issue), W344–W350. https://doi.org/10.1093/nar/gkw408

Baetens, D., Güran, T., Mendonca, B. B., Gomes, N. L., De Cauwer, L., Peelman, F., Verdin, H., Vuylsteke, M., Van der Linden, M., Atay, Z., Bereket, A., de Krijger, R. R., Preter, K. de, Domenice, S., Turan, S., Stoop, H., Looijenga, L. H., De Bosscher, K., Cools, M., & De Baere, E. (2018). Biallelic and monoallelic ESR2 variants associated with 46,XY disorders of sex development. *Genetics in Medicine*, *20*(7), Article 7. https://doi.org/10.1038/gim.2017.163

Barbosa, C., Souza, A., Bianco, B., & Christofolini, D. (2011). The effect of hormones on endometriosis development. *Minerva Ginecologica*, *63*, 375–386.

Baron, J. A., La Vecchia, C., & Levi, F. (1990). The antiestrogenic effect of cigarette smoking in women. *American Journal of Obstetrics and Gynecology*, *162*(2), 502–514. https://doi.org/10.1016/0002-9378(90)90420-C

Bedaiwy, M. A., Alfaraj, S., Yong, P., & Casper, R. (2017). New developments in the medical treatment of endometriosis. *Fertility and Sterility*, *107*(3), 555–565. https://doi.org/10.1016/j.fertnstert.2016.12.025

Bhartiya, D., & Patel, H. (2021). An overview of FSH-FSHR biology and explaining the existing conundrums. *Journal of Ovarian Research*, *14*(1), 144. https://doi.org/10.1186/s13048-021-00880-3

Biason-Lauber, A. (2012). WNT4, RSPO1, and FOXL2 in Sex Development. *Seminars in Reproductive Medicine*, *30*(05), 387–395. https://doi.org/10.1055/s-0032-1324722

Biason-Lauber, A., & Konrad, D. (2008). WNT4 and Sex Development. *Sexual Development*, *2*(4–5), 210–218. https://doi.org/10.1159/000152037

Biason-Lauber, A., Konrad, D., Navratil, F., & Schoenle, E. J. (2004). A WNT4 Mutation Associated with Müllerian-Duct Regression and Virilization in a 46,XX Woman. *New England Journal of Medicine*, *351*(8), 792–798. https://doi.org/10.1056/NEJMoa040533

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics (Oxford, England)*, *22*(22), 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, *14*(Suppl 3), S6. https://doi.org/10.1186/1471-2164-14-S3-S6

Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, *33*(Web Server issue), W306–W310. https://doi.org/10.1093/nar/gki375

Chang, J., Sonoyama, W., Wang, Z., Jin, Q., Zhang, C., Krebsbach, P. H., Giannobile, W., Shi, S., & Wang, C.-Y. (2007). Noncanonical Wnt-4 Signaling Enhances Bone Regeneration of Mesenchymal Stem Cells in Craniofacial Defects through Activation of p38 MAPK *. *Journal of Biological Chemistry*, *282*(42), 30938–30948. https://doi.org/10.1074/jbc.M702391200

Chapron, C., Marcellin, L., Borghese, B., & Santulli, P. (2019). Rethinking mechanisms, diagnosis and management of endometriosis. *Nature Reviews Endocrinology*, *15*(11), 666–682. https://doi.org/10.1038/s41574-019-0245-z

Chen, M., Zhang, W., Gou, Y., Xu, D., Wei, Y., Liu, D., Han, C., Huang, X., Li, C., Ning, W., Peng, D., & Xue, Y. (2023). GPS 6.0: An updated server for prediction of

kinase-specific phosphorylation sites in proteins. *Nucleic Acids Research*, *51*(W1), W243–W250. https://doi.org/10.1093/nar/gkad383

Coss, D. (2020). Commentary on the Recent FSH Collection: Known Knowns and Known Unknowns. *Endocrinology*, *161*(1), bqz035. https://doi.org/10.1210/endocr/bqz035

Daelemans, C., Smits, G., de Maertelaer, V., Costagliola, S., Englert, Y., Vassart, G., & Delbaere, A. (2004). Prediction of Severity of Symptoms in Iatrogenic Ovarian Hyperstimulation Syndrome by Follicle-Stimulating Hormone Receptor Ser680Asn Polymorphism. *The Journal of Clinical Endocrinology & Metabolism*, *89*(12), 6310–6315. https://doi.org/10.1210/jc.2004-1044

Dekker, J., Hooijer, I., Ket, J. C. F., Vejnović, A., Benagiano, G., Brosens, I., & Mijatovic, V. (2021). Neonatal Uterine Bleedings: An Ignored Sign but a Possible Cause of Early-Onset Endometriosis − A Systematic Review. *Biomedicine Hub*, *6*(1), 6–16. https://doi.org/10.1159/000512663

Deng, W., Wang, Y., Ma, L., Zhang, Y., Ullah, S., & Xue, Y. (2017). Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Briefings in Bioinformatics*, *18*(4), 647–658. https://doi.org/10.1093/bib/bbw041

Dimauro, I., Grazioli, E., Antinozzi, C., Duranti, G., Arminio, A., Mancini, A., Greco, E., Caporossi, D., Parisi, A., & Luigi, L. (2021). Estrogen-Receptor-Positive Breast Cancer in Postmenopausal Women: The Role of Body Composition and Physical Exercise. *International Journal of Environmental Research and Public Health*, *18*, 9834. https://doi.org/10.3390/ijerph18189834

Dong, S., Zhao, N., Spragins, E., Kagda, M. S., Li, M., Assis, P., Jolanki, O., Luo, Y., Cherry, J. M., Boyle, A. P., & Hitz, B. C. (2023). Annotating and prioritizing human non-coding variants with RegulomeDB v.2. *Nature Genetics*, *55*(5), Article 5. https://doi.org/10.1038/s41588-023-01365-3

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D., & Yang, J. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols*, *16*(12), Article 12. https://doi.org/10.1038/s41596-021-00628-9

Dun, E. C., Kho, K. A., Morozov, V. V., Kearney, S., Zurawin, J. L., & Nezhat, C. H. (2015). Endometriosis in Adolescents. *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, *19*(2), e2015.00019. https://doi.org/10.4293/JSLS.2015.00019

Eisenberg, V., Weil, C., Chodick, G., & Shalev, V. (2018). Epidemiology of endometriosis: A large population-based database study from a healthcare provider with 2 million members. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(1), 55–62. https://doi.org/10.1111/1471-0528.14711

Endometriosis of young girls and teenagers. (n.d.). *Endometriosis Specialist*. Retrieved June 18, 2023, from https://nezhat.org/endometriosis-of-young-girls-and-teenagers/

Eskenazi, B., & Warner, M. L. (1997). EPIDEMIOLOGY OF ENDOMETRIOSIS. *Obstetrics and Gynecology Clinics of North America*, *24*(2), 235–258. https://doi.org/10.1016/S0889-8545(05)70302-8

Falcone, T., & Flyckt, R. (2018). Clinical Management of Endometriosis. *Obstetrics & Gynecology*, *131*(3), 557–571. https://doi.org/10.1097/AOG.0000000000002469

Fu, L., Zhang, Z., Zhang, A., Xu, J., Huang, X., Zheng, Q., Cao, Y., Wang, L., & Du, J. (2013). Association study between FSHR Ala307Thr and Ser680Asn variants and polycystic ovary syndrome (PCOS) in Northern Chinese Han women. *Journal of Assisted Reproduction and Genetics*, *30*(5), 717–721. https://doi.org/10.1007/s10815-013-9979-z

Fuentes, N., & Silveyra, P. (2019). Chapter Three—Estrogen receptor signaling mechanisms. In R. Donev (Ed.), *Advances in Protein Chemistry and Structural Biology* (Vol. 116, pp. 135–170). Academic Press. https://doi.org/10.1016/bs.apcsb.2019.01.001

Fung, J. N., Rogers, P. A. W., & Montgomery, G. W. (2015). Identifying the Biological Basis of GWAS Hits for Endometriosis1. *Biology of Reproduction*, *92*(4), 87, 1–12. https://doi.org/10.1095/biolreprod.114.126458

Goldstein, S. R. (2010). Modern Evaluation of the Endometrium. *Obstetrics & Gynecology*, *116*(1), 168. https://doi.org/10.1097/AOG.0b013e3181dfd557

Greene, A. D., Lang, S. A., Kendziorski, J. A., Sroga-Rios, J. M., Herzog, T. J., & Burns, K. A. (2016). Endometriosis: Where are We and Where are We Going? *Reproduction (Cambridge, England)*, *152*(3), R63–R78. https://doi.org/10.1530/REP-16-0052

Gregorio, K. C. R., Laurindo, C. P., & Machado, U. F. (2021). Estrogen and Glycemic Homeostasis: The Fundamental Role of Nuclear Estrogen Receptors ESR1/ESR2 in Glucose Transporter GLUT4 Regulation. *Cells*, *10*(1), Article 1. https://doi.org/10.3390/cells10010099

Hendrickx, G., Boudin, E., Steenackers, E., Nielsen, T. L., Andersen, M., Brixen, K., & Van Hul, W. (2017). Genetic Screening of WNT4 and WNT5B in Two Populations with Deviating Bone Mineral Densities. *Calcified Tissue International*, *100*(3), 244–249. https://doi.org/10.1007/s00223-016-0213-8

Hussein, A., Oda, M. H., & Abdulwahab, H. H. (2022). The Impact of Follicle Stimulating Hormone Receptor (Rs6166) Gene Polymorphism on Infertility and Response to FSH Treatment in Iraqi Infertile Females. *HIV Nursing*, *22*(2), 2413–2418.

Imperiale, L., Nisolle, M., Noël, J.-C., & Fastrez, M. (2023). Three Types of Endometriosis: Pathogenesis, Diagnosis and Treatment. State of the Art. *Journal of Clinical Medicine*, *12*(3), Article 3. https://doi.org/10.3390/jcm12030994

Kho, R. M., Andres, M. P., Borrelli, G. M., Neto, J. S., Zanluchi, A., & Abrão, M. S. (2018). Surgical treatment of different types of endometriosis: Comparison of major society guidelines and preferred clinical algorithms. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *51*, 102–110. https://doi.org/10.1016/j.bpobgyn.2018.01.020

Kiesel, L., & Sourouni, M. (2019). Diagnosis of endometriosis in the 21st century. *Climacteric*, *22*(3), 296–302. https://doi.org/10.1080/13697137.2019.1578743

Kiewisz, J., Skowronska, A., Winiarska, A., Pawlowska, A., Kiezun, J., Rozicka, A., Perkowska-Ptasinska, A., Kmiec, Z., & Stompor, T. (2019). WNT4 Expression in Primary and Secondary Kidney Diseases: Dependence on Staging. *Kidney and Blood Pressure Research*, *44*(2), 200–210. https://doi.org/10.1159/000498989

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), Article 7. https://doi.org/10.1038/nprot.2009.86

Lamp, M., Peters, M., Reinmaa, E., Haller-Kikkatalo, K., Kaart, T., Kadastik, Ü., Karro, H., Metspalu, A., & Salumets, A. (2011). Polymorphisms in *ESR1* , *ESR2* and *HSD17B1* genes are associated with fertility status in endometriosis. *Gynecological Endocrinology*, *27*(6), 425–433. https://doi.org/10.3109/09513590.2010.495434

Lang-Muritano, M., Sproll, P., Wyss, S., Kolly, A., Hürlimann, R., Konrad, D., & Biason-Lauber, A. (2018). Early-Onset Complete Ovarian Failure and Lack of Puberty in a Woman With Mutated Estrogen Receptor β (ESR2). *The Journal of Clinical Endocrinology & Metabolism*, *103*(10), 3748–3756. https://doi.org/10.1210/jc.2018-00769

Laven, J. S. E. (2019). Follicle Stimulating Hormone Receptor (FSHR) Polymorphisms and Polycystic Ovary Syndrome (PCOS). *Frontiers in Endocrinology*, *10*. https://www.frontiersin.org/articles/10.3389/fendo.2019.00023

Lee, H. J., Park, Y. M., Jee, B. C., Kim, Y. B., & Suh, C. S. (2015). Various anatomic locations of surgically proven endometriosis: A single-center experience. *Obstetrics & Gynecology Science*, *58*(1), 53–58. https://doi.org/10.5468/ogs.2015.58.1.53

Liaqat, I., Arifa, N. A., Asif, S., & Lone, K. P. (2021). Association of FSHR gene polymorphisms with endometriosis in women visiting tertiary-care hospitals of Lahore, Pakistan. *JPMA. The Journal of the Pakistan Medical Association*, *71*(4), 1118–1122. https://doi.org/10.47391/JPMA.836

Liaqat, I., Jahan, N., Lone, K. P., Pakstis, A., & Taylor, H. S. (2013). Genetic Polymorphisms Associated with Endometriosis in Pakistani Women. *Journal of*

Endometriosis and Pelvic Pain Disorders, *5*(4), 134–143. https://doi.org/10.5301/je.5000165

Lussiana, C., Guani, B., Mari, C., Restagno, G., Massobrio, M., & Revelli, A. (2008). Mutations and Polymorphisms of the FSH Receptor (FSHR) Gene: Clinical Implications in Female Fecundity and Molecular Biology of FSHR Protein and Gene. *Obstetrical & Gynecological Survey*, *63*(12), 785–795. https://doi.org/10.1097/OGX.0b013e31818957eb

Macer, M. L., & Taylor, H. S. (2012). Endometriosis and Infertility: A review of the pathogenesis and treatment of endometriosis-associated infertility. *Obstetrics and Gynecology Clinics of North America*, *39*(4), 535. https://doi.org/10.1016/j.ogc.2012.10.002

Mafra, F., Catto, M., Bianco, B., Barbosa, C. P., & Christofolini, D. (2015). Association of WNT4 polymorphisms with endometriosis in infertile patients. *Journal of Assisted Reproduction and Genetics*, *32*(9), 1359–1364. https://doi.org/10.1007/s10815-015-0523-1

Mandel, H., Shemer, R., Borochowitz, Z. U., Okopnik, M., Knopf, C., Indelman, M., Drugan, A., Tiosano, D., Gershoni-Baruch, R., Choder, M., & Sprecher, E. (2008). SERKAL Syndrome: An Autosomal-Recessive Disorder Caused by a Loss-of-Function Mutation in WNT4. *The American Journal of Human Genetics*, *82*(1), 39–47. https://doi.org/10.1016/j.ajhg.2007.08.005

Matalliotakis, M., Zervou, M. I., Matalliotaki, C., Rahmioglu, N., Koumantakis, G., Kalogiannidis, I., Prapas, I., Zondervan, K., Spandidos, D. A., Matalliotakis, I., & Goulielmos, G. N. (2017). The role of gene polymorphisms in endometriosis. *Molecular Medicine Reports*, *16*(5), 5881–5886. https://doi.org/10.3892/mmr.2017.7398

Missmer, S. A., Chavarro, J. E., Malspeis, S., Bertone-Johnson, E. R., Hornstein, M. D., Spiegelman, D., Barbieri, R. L., Willett, W. C., & Hankinson, S. E. (2010). A prospective study of dietary fat consumption and endometriosis risk. *Human Reproduction*, *25*(6), 1528–1535. https://doi.org/10.1093/humrep/deq044

Ogawa, S., Emi, M., Shiraki, M., Hosoi, T., Ouchi, Y., & Inoue, S. (2000). Association of estrogen receptor β (ESR2) gene polymorphism with blood pressure. *Journal of Human Genetics*, *45*(6), 327–330. https://doi.org/10.1007/s100380070002

Pabalan, N., Trevisan, C. M., Peluso, C., Jarjanazi, H., Christofolini, D. M., Barbosa, C. P., & Bianco, B. (2014). Evaluating influence of the genotypes in the follicle-stimulating hormone receptor (FSHR) Ser680Asn (rs6166) polymorphism on poor and hyper-responders to ovarian stimulation: A meta-analysis. *Journal of Ovarian Research*, *7*(1), 285. https://doi.org/10.1186/s13048-014-0122-2

Pagliardini, L., Gentilini, D., Vigano', P., Panina-Bordignon, P., Busacca, M., Candiani, M., & Di Blasio, A. M. (2013). An Italian association study and meta-analysis with

previous GWAS confirm *WNT4* , *CDKN2BAS* and *FN1* as the first identified susceptibility loci for endometriosis. *Journal of Medical Genetics*, *50*(1), 43–46. https://doi.org/10.1136/jmedgenet-2012-101257

Parasar, P., Ozcan, P., & Terry, K. L. (2017). Endometriosis: Epidemiology, Diagnosis and Clinical Management. *Current Obstetrics and Gynecology Reports*, *6*(1), 34–41. https://doi.org/10.1007/s13669-017-0187-1

Patel, B. G., Rudnicki, M., Yu, J., Shu, Y., & Taylor, R. N. (2017). Progesterone resistance in endometriosis: Origins, consequences and interventions. *Acta Obstetricia et Gynecologica Scandinavica*, *96*(6), 623–632. https://doi.org/10.1111/aogs.13156

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H.-J., Mort, M., Cooper, D. N., Sebat, J., Iakoucheva, L. M., Mooney, S. D., & Radivojac, P. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-19669-x

Pellegrino, M., Maiorino, R., & Schonauer, S. (2010). WNT4 Signaling in Female Gonadal Development. *Endocrine, Metabolic & Immune Disorders - Drug Targets*, *10*(2), 168–174. https://doi.org/10.2174/187153010791213074

Persoons, E., De Clercq, K., Van den Eynde, C., Pinto, S. J. P. coutinho, Luyten, K., Van Bree, R., Tomassetti, C., Voets, T., & Vriens, J. (2020). Mimicking Sampson's Retrograde Menstrual Theory in Rats: A New Rat Model for Ongoing Endometriosis-Associated Pain. *International Journal of Molecular Sciences*, *21*(7), 2326. https://doi.org/10.3390/ijms21072326

Peter, I., Shearman, A. M., Vasan, R. S., Zucker, D. R., Schmid, C. H., Demissie, S., Cupples, L. A., Kuvin, J. T., Karas, R. H., Mendelsohn, M. E., Housman, D. E., & Benjamin, E. J. (2005). Association of Estrogen Receptor β Gene Polymorphisms With Left Ventricular Mass and Wall Thickness in Women*. *American Journal of Hypertension*, *18*(11), 1388–1395. https://doi.org/10.1016/j.amjhyper.2005.05.023

Peyneau, M., Kavian, N., Chouzenoux, S., Nicco, C., Jeljeli, M., Toullec, L., Reboul-Marty, J., Chenevier-Gobeaux, C., Reis, F. M., Santulli, P., Doridot, L., Chapron, C., & Batteux, F. (2019). Role of thyroid dysimmunity and thyroid hormones in endometriosis. *Proceedings of the National Academy of Sciences*, *116*(24), 11894–11899. https://doi.org/10.1073/pnas.1820469116

Philibert, P., Biason-Lauber, A., Rouzier, R., Pienkowski, C., Paris, F., Konrad, D., Schoenle, E., & Sultan, C. (2008). Identification and Functional Analysis of a New WNT4 Gene Mutation among 28 Adolescent Girls with Primary Amenorrhea and Müllerian Duct Abnormalities: A French Collaborative Study. *The Journal of Clinical Endocrinology & Metabolism*, *93*(3), 895–900. https://doi.org/10.1210/jc.2007-2023

Rahmioglu, N., Montgomery, G. W., & Zondervan, K. T. (2015). Genetics of Endometriosis. *Women's Health*, *11*(5), 577–586. https://doi.org/10.2217/whe.15.41

Rai, S., Ashish, Kumari, P., Singh, A., & Singh, R. (2019). Correlation of follicle-stimulating hormone receptor gene Asn 680 Ser (rs6166) polymorphism with female infertility. *Journal of Family Medicine and Primary Care*, *8*(10), 3356–3361. https://doi.org/10.4103/jfmpc.jfmpc_685_19

Saavalainen, L., Lassus, H., But, A., Tiitinen, A., Härkki, P., Gissler, M., Pukkala, E., & Heikinheimo, O. (2018). Risk of Gynecologic Cancer According to the Type of Endometriosis. *Obstetrics & Gynecology*, *131*(6), 1095–1102. https://doi.org/10.1097/AOG.0000000000002624

Saha, R., Pettersson, H. J., Svedberg, P., Olovsson, M., Bergqvist, A., Marions, L., Tornvall, P., & Kuja-Halkola, R. (2015). Heritability of endometriosis. *Fertility and Sterility*, *104*(4), 947–952. https://doi.org/10.1016/j.fertnstert.2015.06.035

Sampson, J. A. (1927). Metastatic or Embolic Endometriosis, due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation. *The American Journal of Pathology*, *3*(2), 93-110.43.

Santi, D., De Vincentis, S., Alfano, P., Balercia, G., Calogero, A. E., Cargnelutti, F., Coccia, M. E., Condorelli, R. A., Dal Lago, A., de Angelis, C., Gallo, M., Iannantuoni, N., Lombardo, F., Marino, A., Mazzella, M., Pallotti, F., Paoli, D., Pivonello, R., Rago, R., … Simoni, M. (2020). Use of follicle-stimulating hormone for the male partner of idiopathic infertile couples in Italy: Results from a multicentre, observational, clinical practice survey. *Andrology*, *8*(3), 637–644. https://doi.org/10.1111/andr.12748

Simpson, J. L., Bischoff, F. Z., Kamat, A., Buster, J. E., & Carson, S. A. (2003). Genetics of endometriosis. *Obstetrics and Gynecology Clinics of North America*, *30*(1), 21–40. https://doi.org/10.1016/S0889-8545(02)00051-7

Smolarz, B., & Romanowicz, H. (2021). Association between single nucleotide polymorphism of the CYP19A1 and ESR2 genes and endometriosis. *Archives of Gynecology and Obstetrics*, *304*(2), 439–445. https://doi.org/10.1007/s00404-021-06051-5

Smolarz, B., Szyłło, K., & Romanowicz, H. (2020). The Genetic Background of Endometriosis: Can ESR2 and CYP19A1 Genes Be a Potential Risk Factor for Its Development? *International Journal of Molecular Sciences*, *21*(21), Article 21. https://doi.org/10.3390/ijms21218235

Sujatha, T., Jayashankar, E., Addepally, U., Kodati, V., & Hasan, Q. (2016). Association of follicle-stimulating hormone receptor gene ser680 asn (rs6166) polymorphism with polycystic ovarian syndrome. *International Journal of Reproduction,*

Contraception, Obstetrics and Gynecology, 3126–3132. https://doi.org/10.18203/2320-1770.ijrcog20162999

Symons, L. K., Miller, J. E., Kay, V. R., Marks, R. M., Liblik, K., Koti, M., & Tayade, C. (2018). The Immunopathophysiology of Endometriosis. *Trends in Molecular Medicine*, *24*(9), 748–762. https://doi.org/10.1016/j.molmed.2018.07.004

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science*, *31*(1), 8–22. https://doi.org/10.1002/pro.4218

Vander Borght, M., & Wyns, C. (2018). Fertility and infertility: Definition and epidemiology. *Clinical Biochemistry*, *62*, 2–10. https://doi.org/10.1016/j.clinbiochem.2018.03.012

Wang, C., Tan, X., Tang, D., Gou, Y., Han, C., Ning, W., Lin, S., Zhang, W., Chen, M., Peng, D., & Xue, Y. (2022). GPS-Uber: A hybrid-learning framework for prediction of general and E3-specific lysine ubiquitination sites. *Briefings in Bioinformatics*, *23*(2), bbab574. https://doi.org/10.1093/bib/bbab574

Wang, H.-S., Wu, H.-M., Cheng, B.-H., Yen, C.-F., Chang, P.-Y., Chao, A., Lee, Y.-S., Huang, H.-D., & Wang, T.-H. (2012). Functional Analyses of Endometriosis-Related Polymorphisms in the Estrogen Synthesis and Metabolism-Related Genes. *PLOS ONE*, *7*(11), e47374. https://doi.org/10.1371/journal.pone.0047374

Wu, X., Xu, S., Liu, J., Bi, X., Wu, Y., & Liu, J. (2014). Association between FSHR polymorphisms and polycystic ovary syndrome among Chinese women in north China. *Journal of Assisted Reproduction and Genetics*, *31*(3), 371–377. https://doi.org/10.1007/s10815-013-0166-z

Wu, Z., Yuan, M., Li, Y., Fu, F., Ma, W., Li, H., Wang, W., & Wang, S. (2015). Analysis of WNT4 polymorphism in Chinese Han women with endometriosis. *Reproductive BioMedicine Online*, *30*(4), 415–420. https://doi.org/10.1016/j.rbmo.2014.12.010

Xin, J., Zhang, J., Yang, Y., Gao, Y., & Xiong, L. (2013). [Association of estrogen receptor β gene polymorphisms with hypertension in postmenopausal women]. *Zhonghua yi xue za zhi*, *93*(21), 1657–1659.

Xu, Z., & Taylor, J. A. (2009). SNPinfo: Integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Research*, *37*(suppl_2), W600–W605. https://doi.org/10.1093/nar/gkp290

Yilmaz, B. D., & Bulun, S. E. (2019). Endometriosis and nuclear receptors. *Human Reproduction Update*, *25*(4), 473–485. https://doi.org/10.1093/humupd/dmz005

Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *WIREs Computational Molecular Science*, *7*(2), e1298. https://doi.org/10.1002/wcms.1298

Zhang, Q., Pan, Y., Ji, J., Xu, Y., Zhang, Q., & Qin, L. (2021). Roles and action mechanisms of WNT4 in cell differentiation and human diseases: A review. *Cell Death Discovery*, *7*(1), Article 1. https://doi.org/10.1038/s41420-021-00668-w

Zhang, Y., & Skolnick, J. (2005). TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, *33*(7), 2302–2309. https://doi.org/10.1093/nar/gki524

Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., Liu, Z., Zhao, Y., Xue, Y., & Ren, J. (2014). GPS-SUMO: A tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Research*, *42*(W1), W325–W330. https://doi.org/10.1093/nar/gku383

Zhao, S.-L., Wei, S.-Y., Wang, Y.-X., Diao, T.-T., Li, J.-S., He, Y.-X., Yu, J., Jiang, X.-Y., Cao, Y., Mao, X.-Y., Wei, Q.-J., Wang, Y., & Li, B. (2016). Wnt4 is a novel biomarker for the early detection of kidney tubular injury after ischemia/reperfusion injury. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep32610

turnitin

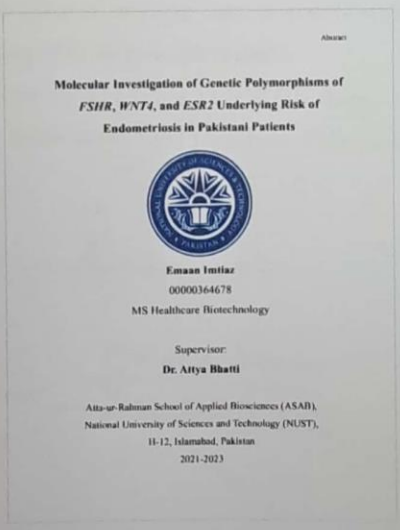
Digital Receipt

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Emaan Imtiaz
Assignment title:	ABCD
Submission title:	Emaan Thesis-01
File name:	Emaan_Thesis-01.docx
File size:	2.98M
Page count:	62
Word count:	15,608
Character count:	86,267
Submission date:	23-Aug-2023 12:30AM (UTC-0700)
Submission ID:	2149848605



Abstract

Molecular Investigation of Genetic Polymorphisms of *FSHR*, *WNT4*, and *ESR2* Underlying Risk of Endometriosis in Pakistani Patients

Emaan Imtiaz
00000364678
MS Healthcare Biotechnology

Supervisor:
Dr. Attya Bhatti

Atta-ur-Rahman School of Applied Biosciences (ASAB),
National University of Sciences and Technology (NUST),
H-12, Islamabad, Pakistan
2021-2023

Emaan Thesis-01

Prof. Dr. Attya Bhatti
Dept of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

ORIGINALITY REPORT

13%	8%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Higher Education Commission Pakistan Student Paper	2%
2	Submitted to The University of Manchester Student Paper	1%
3	prr.hec.gov.pk Internet Source	<1%
4	www.nature.com Internet Source	<1%
5	Submitted to Universiteit van Amsterdam Student Paper	<1%
6	Submitted to Imperial College of Science, Technology and Medicine Student Paper	<1%
7	www.mdpi.com Internet Source	<1%
8	Submitted to Universiti Sains Malaysia Student Paper	<1%