CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

In silico Analysis and Functional Annotation of Genes Associated With Ovarian Cancer Pathogenesis

by

Andleeb Munawar

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

I would like to dedicate this thesis to Allah Almighty, my parents and teachers.

CERTIFICATE OF APPROVAL

In silico Analysis and Functional Annotation of Genes Associated With Ovarian Cancer Pathogenesis

by

Andleeb Munawar

(MBS233015)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Shahid Hussain	KU, Murree
(b)	Internal Examiner	Dr. Sania Riaz	CUST, Islamabad
(c)	Supervisor	Dr. Shaukat Iqbal Malik	CUST, Islamabad

Dr. Shaukat Iqbal Malik
Thesis Supervisor
September, 2025

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. & Biosciences
September, 2025

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
September, 2025

Author's Declaration

I, **Andleeb Munawar** hereby state that my MS thesis titled "***In silico* Analysis and Functional Annotation of Genes Associated With Ovarian Cancer Pathogenesis**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Andleeb Munawar)

Registration No: MBS233015

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "***In silico* Analysis and Functional Annotation of Genes Associated With Ovarian Cancer Pathogenesis**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Andleeb Munawar)

Registration No: MBS233015

Acknowledgement

Alhamdulillah! all praise and thanks to Allah Almighty, the Most Gracious and the Most Merciful, who blessed me with strength, patience, and guidance to complete this thesis successfully. I am sincerely grateful to my respected supervisor, Dr. Shaukat Iqbal Malik, for his guidance, encouragement, and continuous support throughout my research work. His insightful suggestions and constructive feedback greatly enriched the quality of this work. I am deeply indebted to my beloved family, especially my father Munawar Hussain, my mother, and my brothers and sisters, whose unwavering love, prayers, and belief in me kept me motivated and steadfast. Their patience and encouragement have been my greatest source of strength during this journey. I would also like to extend my heartfelt thanks to my senior, Arifa Siddiqa, for her kind support, guidance, and help whenever I needed it. Finally, I am thankful to all those who directly or indirectly contributed to the completion of this thesis and made this journey more enjoyable and motivating.

(Andleeb Munawar)

Abstract

Ovarian cancer (OC) remains one of the most lethal gynecological malignancies worldwide, largely due to its late diagnosis, complex pathogenesis, and limited molecular understanding beyond a few well-studied genes such as *BRCA1, BRCA2,* and *TP53*. This study aimed to uncover novel candidate genes potentially associated with OC pathogenesis through an in silico approach, integrating multiple bioinformatics tools for gene identification, network construction, clustering, functional annotation, and interaction prediction. An initial list of approximately 2000 genes associated with malignant neoplasm of the ovary was retrieved using Coremine Medical.The genes were analyzed through STRING to construct a protein–protein interaction (PPI) network and clustered using k-means. Clusters containing both multiple and single genes. Some clusters comprised single genes, suggesting possible outliers or unique interactions were manually verified in KEGG pathways to distinguish known OC-related genes from uncharacterized candidates. Forty-nine mismatch genes, not mapped to KEGG OC pathways, were subjected to functional annotation clustering in DAVID, revealing enrichment in biological processes such as keratinization, ABC transporter activity, transcriptional regulation, and calcium ion binding. Further PPI analysis in STRING examined interactions between mismatch genes and key OC proteins (*BRCA1, BRCA2, TP53, PTEN, PARP1, AKT1, PIK3CA*). Notably, *KRT5* exhibited a high-confidence interaction with TP53, suggesting a potential novel link to OC pathogenesis. Several medium-confidence interactions, including those involving *ABCG2* and *KRT6B*, were also observed. These findings highlight previously uncharacterized genes that may contribute to OC development through structural, regulatory, or transport-related mechanisms. The study demonstrates the utility of in silico tools in prioritizing candidate genes for further investigation. Future work should validate these findings experimentally and explore the clinical relevance of these genes as potential biomarkers or therapeutic targets.

Keywords: Ovarian cancer, Bioinformatics, Pathogenesis, Gene Annotation, Coremine Medical, STRING Database, KEGG, DAVID, Protein-protein Interactions.

Contents

List of Figures

List of Tables

Abbreviations

DAVID	Database for Interactions, Visualization and Integrated Discovery
DEGs	Differentially Expressed Genes
EOC	Epithelial Ovarian Cancer
FIGO	International Federation of Gynaecology and Obstetrics
GEO	Gene Expression Omnibus
GPVs	Germline Pathogenic Variations
HGSOC	High-grade Serous Ovarian Cancer
KEGG	Kyoto Encyclopedia of Genes and Genome
LGSOC	Low-grade Serous Ovarian Cancer
OC	Ovarian Cancer
STRING	Search Tool for Retrieval of Interacting Genes
TCGA	The Cancer Genome Atlas Program

Chapter 1

Introduction

Cancer is a collective term for a number of distinct diseases rather than a single disease. The uncontrolled growth and development of cells in the body is termed as cancer. Cancer is a multifactorial disease where a multifarious genome changes are associated with interactive events of the individual's environment. Multifactorial malady cancer involves multifarious changes in the genome because of interaction with the individual's environment [1]. Disruptions in cell signaling pathways that cause rapid cell growth and metastasis are the pathogenesis of cancer. Cancer is still one of the leading causes of death worldwide, despite significant progress in lowering its incidence, and both its frequency and death rates are predicted to rise in the coming years. As many as 10 million deaths were due to cancer in 2023; therefore, there is an urgency to develop strategies to better understand its mechanisms, identify it at an early stage, and create more effective treatments [2].

According to the World Health Organization, cancer is the second most Lethal disease in the world, accounting for roughly 10 million deaths per year. Up to 2030, 22.2 million people are expected to receive a cancer diagnosis [3]. Cancer is one of the leading causes of mortality globally, and chemotherapy, whether administered alone or in combination with other treatments, is considered the gold standard [4].

Given pathogenesis of cancer or its cause includes combination of gene and epigenetic change (oncogenes, tumor suppressor genes, and DNA repair defects) and

alter cellular signaling pathway resulting in genomic instability, triggering unregulated cell growth, evasion of apoptosis, and enhancing in metastatic potential that contribute to development of the cancer.

The most frequently researched and diagnosed cancers encompass haematologic malignancies, lung, colorectal, prostate, cutaneous, breast, uterine, thyroid, lymphatic, and ovarian cancers, among others.

Numerous techniques are utilized in the treatment of cancer, like stem cell transplantation, chemotherapy, radiation therapy, hormone therapy, targeted therapy, surgery, and precision medicine [5].

The deadliest gynaecological cancer is ovarian cancer. Worldwide, ovarian cancer (OC) ranks eighth in cancer-related deaths among women and seventh in malignant tumors.

In 2012, the United States and Northern Europe had the highest ovarian cancer prevalence rates, while Japan had the lowest [6]. Ovarian cancer comprises a diverse array of malignancies characterized by cell or location of origin, pathological grade, risk factors, prognosis, and therapeutic options.

Early detection of ovarian cancer is difficult since most women are unaware that the disease is progressing in their body. About 1.3% of women, or 1 in 78, will get ovarian cancer at some point in their lives. Unfortunately, ovarian cancer has a very low survival rate, mostly because it is diagnosed at a late stage. It causes 5% of female cancer deaths even though it makes up 2.5% of all female cancers [7].

Still, it's the sixth most common cancer killer of women, behind pancreatic, colon, breast, and lung cancers. The incidence and mortality rates of ovarian cancer are known to rise with age. A personal or family history of the condition, along with any related genetic disorders, are risk factors for ovarian cancer.

Women whose families have a history of ovarian cancer have a 5% increased risk of developing the disease themselves, while women whose families have two or more cases have a 7% increased risk. Some genetic anomalies put women at a high risk of developing ovarian cancer (Figure 1.1) [8].

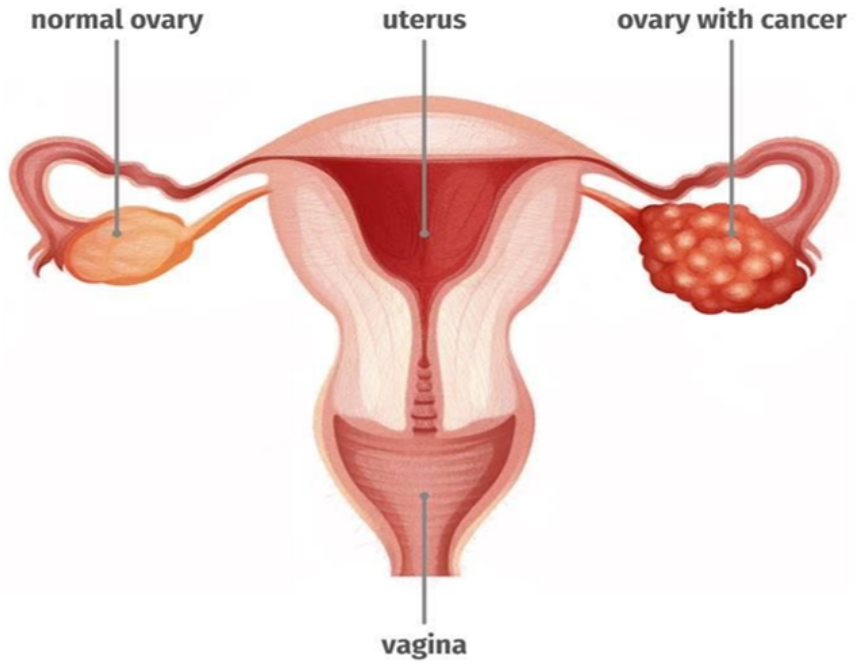


FIGURE 1.1: Atypical illustration of a normal ovary and ovary with cancer [9].

Women of all ages can be affected, although the most prevalent age group to be diagnosed is 55 to 64. One, two Epithelial ovarian malignancies, which mostly affect women after menopause, account for over 90% of tumors. Only about 45% of women who get ovarian cancer are able to survive five years after their diagnosis. The five-year survival rate for patients with advanced epithelial ovarian cancer was 17% to 28%, compared to 92% for those with stage I tumors [9]. Approximately 90% of all occurrences of ovarian cancer are epithelial ovarian cancer (EOC) making it the most frequent subtype of the illness. In most cases, stromal cells (5-8%), germ cells (3-5%), and epithelial cells (85-95 percent) are the cells that help bring about ovarian cancer (Figure 1.2). The age of the patient determines the specific kind of ovarian tumour that is present. Women above the age of 50 are more likely to develop epithelial cell tumors. Although androblastomas and other stromal cell tumors may be more prevalent in teenage girls, they can affect women of any age. Typically, germ cell tumor manifest in individuals between the ages of 15 and 19, as well as in those younger than one year [8]. A number of important genes are associated with ovarian cancer through genetics. These genes include *BRCA1*, *BRCA2*, *TP53*, *PTEN*, and *KRAS*, which play a role in cell cycle regulation, tumor suppression, and DNA repair [10].

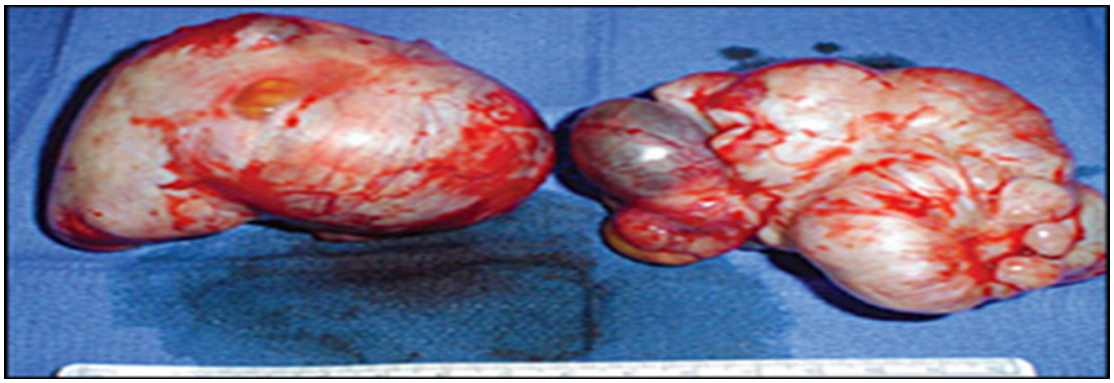


FIGURE 1.2: Epithelial ovarian tumors [8].

Ovarian cancer is known to be influenced by a combination of hereditary abnormalities (such as *BRCA1* and *BRCA2*) and environmental and lifestyle variables. Ovarian cancer risk factors include reproductive history, hormone exposure, food, obesity, talc exposure, and smoking [11].

Many risk factors associated with OC have been detailed. Some ways to classify these potential dangers are as follows: demographic, reproductive, gynecologic, hormonal, hereditary, and lifestyle-related. The topic has been thoroughly researched, although the association between some of the stated characteristics and the development of OC is debatable. While a small percentage of OC cases have a familial basis, the vast majority are sporadic. A big risk factor for OC is having a close relative with the disease, though, so that shouldn't be overlooked. There are additional factors that have been found to mitigate the risk of OC, such as breastfeeding. Reducing the risk may be as simple as nursing for three months. Women who used contraceptives and subsequently had children also showed an association of protective variables [12].

According to the National Cancer Registry of Pakistan (2015–2019), ovarian cancer accounted for 8,823 cases, comprising 6.09% of all female cancers diagnosed during this period [13]. These findings are further supported by studies done in the hospital. For example, a study carried out at a tertiary care hospital in Pakistan reported that ovarian cancer comprised 49.18% of all gynecological malignancies diagnosed with gynecologic cancer most common gynecologic cancer in that population [14].

Like most other countries, Pakistan has a high rate of epithelial ovarian cancer diagnoses. Nevertheless, absence of national screening programs, insufficient public awareness, and socio- cultural barriers hamper the diagnosis and clinical outcomes at delayed stages. Ovarian cancer is a serious issue in Pakistan, and these trends highlight the need for education and awareness initiatives, early detection programs, and better healthcare facilities to combat the disease.

Unlike the degree of its effectiveness, traditional cancer treatments like chemotherapy, radiation therapy, and surgery are unable to avoid the weightiness of side effects, applied from case to case, according to the type and stage of the cancer.

There is still a need for newer and safer safer and newer safer more effective means of targeting cancer cells without damaging the healthy tissues. For this reason, natural compounds have been evaluated as cancer prevention and therapy as they are known to have multitask biological activities [15].

Both incidence and mortality have been slightly reduced in the past 4 decades 1 perhaps because of increases of use of short and long term hormonal contraceptives and decreases in post menopausal hormone use [9].

In recent time, bioinformatics has improved the progress of cancer research by being able to perform large scale genomic analysis, characterize novel biomarkers and therapeutic targets [16].

The researchers make use of bioinformatics approaches in order to know the changes in the progression of ovarian cancer that are brought about by mutations, changes in gene expression, and alterations in signaling networks. Molecular research databases that are essential for the study of ovarian cancer [17].

To address this knowledge gap, this study employs a plethora of bioinformatics tools, including COREMINE Medical, STRING, KEGG and DAVID, to conduct a comprehensive analysis of genes associated with ovarian cancer [18].

Ovarian cancer's complicated pathogenesis and lack of early detection markers make it one of the most deadly gynecological cancers. Although several key genes,

such as *BRCA1*, *TP53*, and *PTEN*, are well-characterized in ovarian cancer pathways, many potentially relevant genes remain unrepresented in curated pathway databases like KEGG.

This gap limits our understanding of the full molecular network involved in disease development and progression. The absence of these genes from standard cancer pathways raises critical questions:

Are these "mismatched" genes functionally relevant to ovarian cancer? Do they participate in regulatory or interaction networks that influence tumor biology?.

This research aims to explore the role of such mismatch genes using integrated bioinformatics tools, including DAVID for functional enrichment, STRING for protein-protein interaction analysis, and for predictive interaction modeling.

By characterizing the biological functions and interaction potential of these underrepresented genes, this study seeks to uncover novel candidates that may contribute to ovarian cancer pathogenesis and broaden the scope of gene panels used in future diagnostic or therapeutic strategies.

1.1 Problem Statement

Ovarian cancer remains one of the deadliest gynecological malignancies due to its complex pathogenesis and limited early detection markers. Many genes potentially involved in its pathogenesis remain uncharacterized in curated databases. This study aims to identify, functionally annotate, and predict interactions of such genes using *in silico* tools to uncover their potential roles in ovarian cancer progression.

1.2 Aim and Objectives

This study aims at unraveling the hidden roles of uncharacterize genes in ovarian cancer pathogenesis.

1. To identify ovarian cancer-associated genes using biomedical literature mining through the Coremine Medical database.

2. To retrieve and analyze known and predicted protein-protein interactions of selected genes using the STRING database.

3. To manually verify gene involvement in ovarian cancer pathways by using KEGG Pathway Database.

4. To functionally annotate the genes not represented in KEGG pathways using the DAVID Functional Annotation Tool.

5. To predict potential novel protein-protein interactions for selected uncharacterized or mismatch genes using *in silico* tools.

6. To interpret the biological significance of these genes and their interactions, with the aim of uncovering novel contributors to ovarian cancer pathogenesis.

Chapter 2

Literature Review

When abnormal cells start in the ovary and start to multiply out of control, it's called ovarian cancer (OC). When these cells spread to other parts of the body, they can cause tumours to grow harmfully. Each of the three cell types found in the ovaries has the potential to develop into a different type of cancer. All types of ovarian malignancies, including endometrioid, mucinous, clear cell, and high and low-grade serous carcinomas, are epithelial in origin, accounting for over 90% of these cases. Stromal ovarian cancers account for 7% of all ovarian cancers, but germ cell tumours are the most seldom cause of this disease [19].

There are numerous subtypes of ovarian cancer, each with its own set of risk factors, cellular origins, molecular composition, clinical features, and treatment options. At least five subtypes have been identified. A dependable screening approach is lacking for ovarian cancer, which is a global concern and typically detected at an advanced stage [20]. When it comes to gynaecological cancers, epithelial ovarian cancer kills more people than any other type. Symptoms, including as abdominal distention and pain that lasts for months, usually show up in women who have gone through menopause. The majority of women have surgery and traditional chemotherapy with platinum-based agents after their cancer has progressed to stage III, as defined by the International Federation of Gynaecology and Obstetrics [FIGO]. The majority of patients with early-stage cancer may

find a cure with this treatment; nevertheless, the majority of women with advanced disease will have multiple recurrence episodes with shorter and shorter disease duration. Gap-free periods [21]. Developing ovarian cancer is more likely among older women, those with a personal or family history of the disease, those with certain genetic traits, and those with abnormalities in the BRCA gene. A comprehensive assessment of the incidence, origins, and trends of ovarian cancer is necessary for enhancing global response to the illness. Ovarian cancer incidence and mortality rates for 185 nations worldwide in 2020 were derived from the database of the International Agency for Research on Cancer (IARC, WHO, Lyon, France) [22]. Although there are often indicators and warning indications of OC, the first symptoms might be difficult to discern because of the overlap with other gynaecological, genitourinary, and gastrointestinal disorders [23].

This chapter provides a comprehensive review of the existing literature relevant to ovarian cancer pathogenesis, with a specific focus on the genetic components and *in silico* approaches used for gene analysis and functional annotation. It begins with the molecular and genetic background of ovarian cancer, followed by an exploration of relevant bioinformatics tools and databases, and concludes with gaps in the current knowledge that this thesis aims to address.

2.1 Types of Ovarian Cancer

There are numerous subtypes of ovarian cancer that may be classified based on the cell type that originated in the body. The three most common types are sex cord-stromal tumors, epithelial tumors, and germ cell tumors [24]. The vast majority of ovarian neoplasms are epithelial ovarian cancers (EOCs), which account for more than 90% of all cases [25].

2.1.1 Epithelial Tumors

These originate from the ovarian surface epithelium and encompass many subtypes:

1. The most common and malignant type is high-grade serous carcinoma (HGSC).

2. Low grade serous carcinoma.
3. Endometrioid carcinoma
4. Mucinous carcinoma
5. Clear cell carcinoma [26]
6. *TP53* and *BRCA1/2* mutations are usually associated with HGSC, which is frequently detected at an advanced stage [27].

2.1.2 Germ Cell Tumors

Derived from the cells that form eggs, these tumors are more common in younger women and include types like dysgerminomas, yolk sac tumors, and teratomas. They typically have a better prognosis with early detection and treatment [28].

2.1.3 Sex Cord-Stromal Tumors

These uncommon neoplasms arise from the ovarian connective tissue. Notable kinds include granulosa cell tumors and Sertoli-Leydig cell tumors, which have the potential to release hormones like oestrogen or androgens [28].

2.2 Epidemiology and Global Burden

Ovarian cancer (OC) is the eighth most common cancer among women worldwide and the leading cause of death from gynecologic malignancies. Ovarian cancer continues to have a generally unfavourable prognosis, especially in low-resource environments. It is essential to consistently assess the impact of ovarian cancer to pinpoint inequities. Past study provides a thorough evaluation of the worldwide impact of OC by utilising the GLOBOCAN 2020 estimations that are classified by country, global region, and HDI levels, as well as by projecting these levels into the year 2040. Predictions for new cases and deaths in 2040 were derived using HDI-categorized demographic estimates.

In 2020, about 314,000 women received an ovarian cancer diagnosis, and 207,000 of them died from this disease. Countries in Europe with a relatively high Human Development Index (HDI) had the highest incidence rates, whilst those in Africa with the lowest HDI had the lowest rates. This disparity was statistically significant. There was little difference in mortality rates among the four HDI tiers. In contrast to nations with a very high HDI, where new ovarian cancer cases are expected to climb by 19% and fatalities by 100% by 2040, low HDI nations are projected to see an increase of approximately 96% in new cases and 28% in fatalities [29]. A significant burden of ovarian cancer was borne by China (6.6 per 100,000) and India (4.9 per 100,000), accounting for 40.5% (125,629/310,991) of all ovarian cancer cases worldwide in 2020. Despite advancements in treatments like surgery and chemotherapy, the five-year survival rate remains below 50%, largely because 70% of cases are detected at advanced stages (III or IV).

Early detection and molecular profiling remain crucial for improving prognosis and patient outcomes (Figure 2.1) [30]. These statistics underscore the urgent need for improved strategies in prevention, early detection, and targeted therapy to reduce the global burden of this lethal disease. The burden of OC is expected to rise in the coming decades because of increasing life expectancy, population growth, and the adoption of risk factors like obesity, hormone replacement therapy, and low parity.

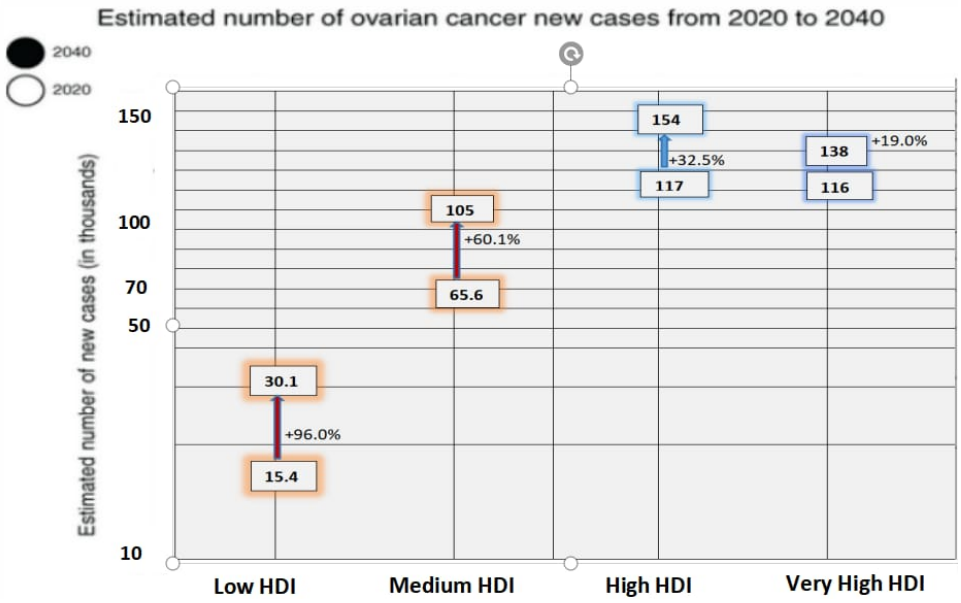


FIGURE 2.1: Ovarian Cancer new cases from 2020 to 2040 [29].

2.3 Symptoms of OC

Due to delayed detection and frequent chemotherapy resistance in patients, ovarian cancer is a hidden enemy, it is the leading cause of mortality among the different forms of gynaecological cancer. In the last ten years, substantial evidence has emerged indicating that women with ovarian cancer exhibit discernible symptoms prior to diagnosis.

Symptoms most commonly associated with ovarian cancer include gas, a larger waist circumference, pelvic pain, abdominal pain, early fullness, and difficulties with eating. The frequency of urinary symptoms is generally high, according to several studies (Figure 2.2).

Ovarian cancer should be considered as a possible diagnosis if these symptoms appear more than 12 times per month and are new. Ovarian cancer is not a "silent disease," and both women and medical professionals must acknowledge this [31].

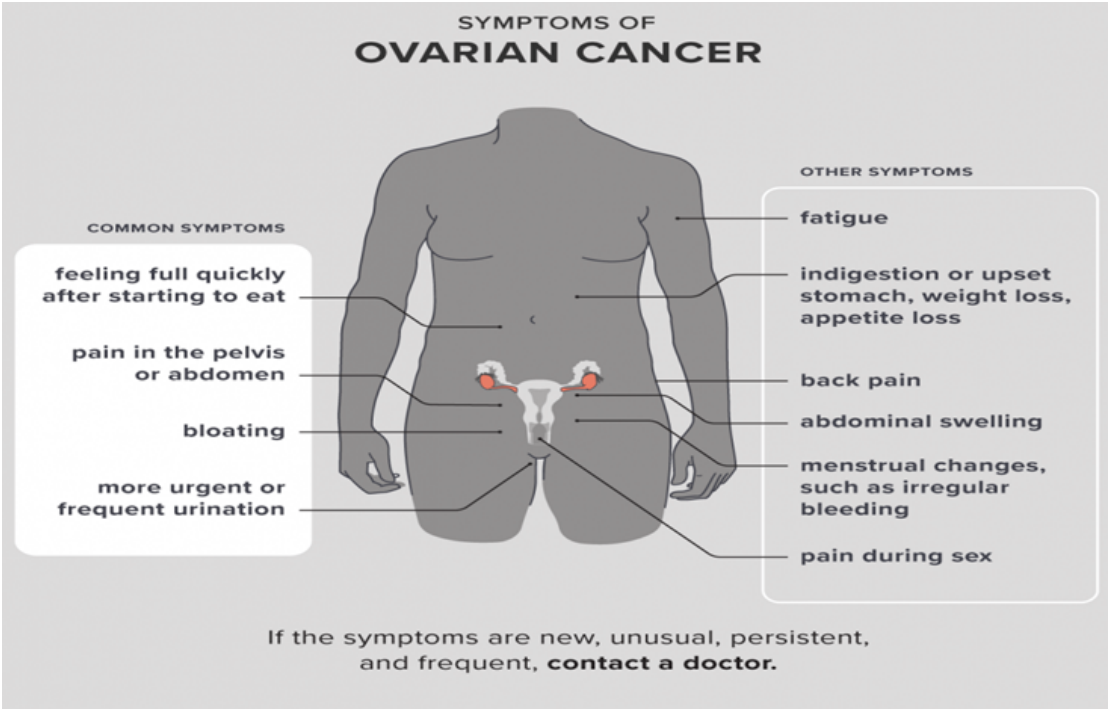


FIGURE 2.2: Symptoms of Ovarian Cancer [31].

2.4 Pathogenesis of Ovarian Cancer

Of all gynaecological diseases, ovarian cancer presents the greatest challenge in terms of both diagnosis and therapy. This gynaecological cancer is either found by chance, like during a diagnostic laparoscopy, or diagnosed at an advanced stage when symptoms begin to appear, because its early stages are completely symptomless [32].

Imbalances in the *BRCA2* or *BRCA1* gene, various genetic tendencies, unexplained start and advancement, molecular tumour diversity, and disease staging can all affect how well these indicators work [33].

It Is primarily categorized into two principal types: **Type I** tumors (low-grade serous, endometrioid, clear cell, and mucinous) and **Type II** tumors (predominantly high-grade serous carcinomas), the latter being the most aggressive and prevalent subtype [33].

The molecular pathogenesis of high grade serous ovarian cancer (HGSOC) is mostly linked to mutations in the *TP53* gene, found in more than 95% of instances [34]. Additional significant genetic abnormalities encompass mutations or dysfunctions in *BRCA1* and *BRCA2*, which hinder homologous recombination DNA repair and lead to genomic instability [35].

Oncogenic pathway dysregulation, such as that of PI3K/AKT/mTOR, RAS/MAPK, and p53 signaling, is a major factor in tumor growth, angiogenesis, and resistance to apoptosis [36]. Furthermore, epigenetic alterations, non-coding RNAs, and interactions within the tumor microenvironment are acknowledged as essential factors in ovarian tumor biology [27].

2.4.1 Role of Risk Factors in Initiating Pathogenesis

Multiple risk factors are involved in the development of ovarian cancer.

2.4.1.1 Genetic Predisposition in Ovarian Cancer

Over 20% of all OC cases are caused by germline pathogenic variations (GPVs) in Ovarian cancer predisposition genes (such as *BRCA1, BRCA2, BRIP1, RAD51C, RAD51D, Lynch syndrome genes, or BRIP1*). These GPVs are the most important risk factor for OC development. The majority of GPVs occur in *BRCA1 / BRCA2*. Patients with OC and their families can benefit from individualized plans to lower their cancer risk when a GPV is present [37]. Inherited breast and ovarian cancer (HBOC) syndrome is most often caused by mutations in the BRCA1 and BRCA2 HR genes, which are germline-altered cancer predisposition genes. Almost 20% of OC patients have GPVs in both genes [38].

After the ages of 35 and 45, respectively, the Ovarian Cancer risk for *BRCA1* and *BRCA2* carriers exceeds 58% and 29%. In comparison to *BRCA2, RAD51D, RAD51C*, and *BRIP1* GPV carriers, the median age of OC onset in *BRCA1* GPV carriers is 53 years, whereas in BRCA2, 57, 62, and 65 years, respectively, it is considerably higher. From very low percentages (*MSH6* and *PMS2*) to twenty percent (MLH1) and thirty-eight percent (*MSH2/EPCAM)*, the lifetime risk ranges widely (Table 2.1). A modest risk of late-onset epithelial ovarian cancer has been linked to GPVs in other genes that are involved in DNA repair, such as PALB2 and ATM [37].

Table 2.1: Established OC predisposition genes [37].

Genes	Associated OC Histology	Risk of OC	GPV Identified in early-Onset
BRCA1	Epithelial	39 - 58%	Yes
BRCA2	Epithelial	13 - 29%	Yes
BRIP1	Epithelial	5- 15%	Yes
DICER1	Sex-cord stromal	NA	NA
MLH1	Epithelial	4 - 20 %	No
MSH2	Epithelial	8 - 38%	Yes
RAD51C	Epithelial	10 - 15%	Yes
STK11	Non-epithelial	10%	Yes

Although candidate OC predisposition genes have not been linked to OC as of yet, they can be suggested based on their propensity to other cancer kinds or related

disorders [39]. Some of candidates genes are *ATR, APC, BPA1, BARD1, BLM, BMPR1A, BRAT, CNKSR1, CDKN2A, CHEK2, ERCC3, FANCA, FANCC, FANCL, FANCM, FH, MEN1, MRE11, NBN, PIK3C2G.*

2.4.1.2 Incessant Ovulation Theory

The Incessant Ovulation Theory is a well-established hypothesis explaining the origin of epithelial ovarian cancer, especially low-grade serous carcinoma. It postulates that repeated ovulatory cycles throughout a woman's reproductive life lead to cumulative trauma, inflammation, and cellular proliferation of the ovarian surface epithelium (OSE), which increases the likelihood of DNA damage and malignant transformation [40].

During ovulation, the ovarian follicle ruptures through the OSE to release the oocyte, causing local inflammatory responses, oxidative stress, and activation of cytokine-mediated pathways. The subsequent healing process necessitates epithelial cell proliferation, which introduces opportunities for DNA replication errors and somatic mutations in oncogenes and tumor suppressor genes. Epidemiological data support this theory. Ovarian cancer is substantially more likely to occur in women who have more lifetime ovulatory cycles (either as a result of early menarche, late menopause, nulliparity, or infertility). On the other hand, ovulation suppression factors like pregnancy, lactation, and oral contraceptive use are linked to a lower risk [41]. At the molecular level, incessant ovulation has been linked to mutations in key genes such as *TP53*, *BRCA1*, and *BRCA2*, which are frequently found altered in ovarian cancer cases. However, with the discovery that many high grade serous ovarian cancers (HGSOC) may originate in the fallopian tube epithelium (particularly the fimbrial end), the theory has been partially revised [42]. While it may not fully explain the origin of all subtypes of ovarian cancer, it remains highly relevant for ovary-derived epithelial tumors and contributes significantly to our understanding of disease etiology. The implications of this theory are substantial. Preventive strategies, such as the use of ovulation-suppressing hormonal therapies are partly based on this hypothesis. Furthermore, it underscores the importance of investigating ovulation-associated

gene expression, inflammation-related markers, and repair pathways in ovarian cancer research.

2.4.1.3 Hormonal Exposure in OC Development

Hormonal factors play a major role in the pathogenesis of ovarian cancer. Estrogen and other sex steroids influence the proliferation, differentiation, and apoptosis of ovarian surface epithelial cells, which are key in the development of epithelial ovarian cancer [43]. Prolonged exposure to endogenous estrogens as observed in women with early menarche, late menopause, nulliparity, or infertility has been associated with increased risk of ovarian cancer. Estrogen promotes epithelial proliferation, which may lead to accumulation of genetic mutations and malignant transformation [44].

Conversely, progesterone has shown protective effects against ovarian carcinogenesis. Studies indicate that pregnancy and oral contraceptive use, which increase progesterone levels and reduce the number of ovulatory cycles, are linked to a lower incidence of ovarian cancer [45]. The estrogen receptor (ER) and progesterone receptor (PR) status of ovarian tumors is also of clinical significance. ER-positive tumors often exhibit more aggressive phenotypes, while PR expression is associated with better prognosis and chemosensitivity.

2.4.1.4 Inflammation and Oxidative Stress in the Development of Ovarian Cancer

Chronic inflammation and oxidative stress are increasingly recognized as critical contributors to the initiation and progression of ovarian cancer (OC). Repeated ovulation is associated with localized inflammation, which leads to the production of pro-inflammatory cytokines, chemokines, and reactive oxygen species (ROS). These factors create a tumor-promoting microenvironment that can induce DNA damage, mutagenesis, and epithelial-mesenchymal transition (EMT) [46]. Inflammatory mediators such as interleukin-6 (IL-6), tumor necrosis factor-alpha (TNF-α), and prostaglandins have been found at elevated levels in ovarian cancer tissues

and ascites. These molecules stimulate angiogenesis, cell proliferation, and resistance to apoptosis, thereby facilitating tumor progression. Oxidative stress, resulting from an imbalance between ROS generation and antioxidant defense mechanisms, further exacerbates genomic instability. ROS can oxidize DNA bases, lipids, and proteins, promoting mutations in tumor suppressor genes like *TP53* and *BRCA1/2*, both commonly altered in OC [47]. The disease commnly progresses asymptomtically, and by the time of diagnosis, it often involves widespread peritoneal metastasis, due to transcoelomic spread.

2.5 Environmental Risk Factors

Evidence suggests that ovarian cancer is more likely in people who are exposed to certain environmental risk factors. Researchers have discovered that talcum powder, which may have been the first risk factor to be documented, raises the likelihood of ovarian tumours, particularly serous tumours. Talc and asbestos are physically similar, and histologic studies have shown that asbestos exposure causes serous adenocarcinomas that are identical to mesotheliomas [48].

Many members of the chemical family known as polycyclic aromatic hydrocarbons (PAHs) are suspected carcinogens and contribute to environmental pollution. In addition to other known carcinogens, PAHs are found in cigarette smoke. In the presence of PAHs, DNA adducts can be formed, leading to cellular mutations and, ultimately, tumor development. Western research has extensively documented environmental risk factors, particularly asbestos exposure, but Asian ovarian cancer research has not. In industrialized areas of China and India, our study finds a possible association between ovarian cancer risk and exposure to industrial contaminants. Climate change in itself stems from human activity and hence, as noted by the harmful exposures combined with poor regulation could make some populations more prone to cancer. Globally, the present findings support research showing a deniable human carcinogenic role for environmental pollutants such as inclusion of three persistent organic compounds (POPs) in groups1 and 2A i.e., Human baseline risk classification versus environment pollutants. However, more

focused research is advised to elucidate all possible risks in Asia, where industrialization is happening quickly and environmental protections are not being strictly monitored [49].

2.6 Molecular Mechanisms and Gene Signature in Ovarian Cancer

Molecular mechanisms refer to the genetic and biochemical alterations at the cellular level that derive ovarian cancer development, progression, and resistance to therapy. Determining biomarkers, therapeutic targets, and interpreting functional annotation results from *in silico* analysis all depend on an understanding of these mechanisms.

2.6.1 Genetic Alterations And Dysregulation

Since *TP53* gene mutations are present in more than 96% of HGSOC cases, they are one of the defining characteristics of ovarian cancer. By triggering cell cycle arrest and apoptosis in reaction to DNA damage, *TP53* suppresses tumors.Loss of *TP53* activity leads to genomic instability and uncontrolled cell proliferation. Additionally, mutations or epigenetic silencing of *BRCA1* and *BRCA2*, key components of the homologous recombination repair (HRR) pathway, result in defective DNA repair and increased mutation burden. Other oncogenes and tumor suppressor genes frequently implicated in ovarian cancer include *KRAS*, *PIK3CA*, *CTNNB1*, and *PTEN*. Mutations in *KRAS* and *PIK3CA* promote constitutive activation of the MAPK and PI3K/AKT pathways, respectively, both of which support cell survival and proliferation [50].

Table 2.2: Significantly Mutated Genes in HGS-OvCa [50]

Gene	Number of Mutations	Validated	Unvalidated
TP53	302	294	8
CSMD3	19	19	0

Table 2.2 continued from previous page

Gene	Number of Mutations	Validated	Unvalidated
NF1	13	13	0
BRCA1	11	10	1
FAT3	19	18	1
CDK12	9	9	0
BRCA2	10	10	0
GABRA6	6	6	0
RB1	6	6	0

Mutations that have been verified by an independent assay are known as validated mutations. A second independent WGA sample from the same tumor is used to validate the majority of them. Unvalidated mutations are very likely to be real mutations even though they haven't been independently verified. By hand curation, another 25 *TP53* mutations were found. The following genes are among the most commonly mutated or dysregulated in ovarian cancer, particularly high grade serous ovarian carcinoma (HGSOC).

2.6.1.1 *TP53* Mutations in Ovarian Cancer

TP53, located on chromosome 17p13.1, is a tumor suppressor gene encoding the p53 protein, which plays a crucial role in cell cycle regulation, apoptosis, DNA repair, and genomic stability [51]. Mutation of *TP53* is the most frequent genetic alteration in high-grade serous ovarian carcinoma (HGSOC), present in over 96% of cases, making it a hallmark of this ovarian cancer subtype. Mutations in *TP53* typically result in a loss of function or gain of oncogenic function, leading to impaired cellular responses to DNA damage and uncontrolled cell proliferation. These mutations are mostly missense and cluster in the DNA-binding domain of the gene, altering the protein's ability to activate transcription of downstream targets like *CDKN1A* (p21) and *BAX* [52].

Growing complexity within the p53 pathway, where *p53* can be involved in disparate and even contradictory responses, is a result of recent recognition of *p53*'s

role in controlling glycolysis, oxidative stress, and cell survival. These p53 functions draw attention to an intriguing paradox discussed previously: if certain *p53* activities that typically aid in tumor suppression are not appropriately controlled, they may "switch sides" and aid in the development of cancer (Figure 2.3). In certain situations, *p53*'s nontranscriptional activities, such as its inhibition of autophagy, may also aid in the development of tumors. Interestingly, cancer-associated mutant p53 proteins maintain this specific function of wild-type *p53* [51].

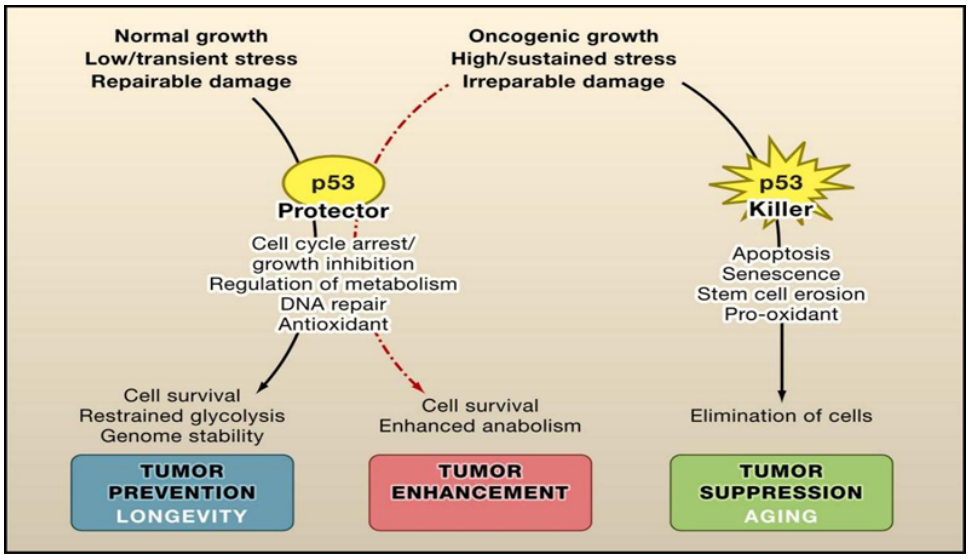


FIGURE 2.3: Dual Mechanisms of p53 Function in Tumor Suppression and Aging [51].

The Cancer Genome Atlas (TCGA) integrated genomic study on ovarian carcinoma confirmed that nearly all HGSOC tumors harbor somatic mutations in *TP53*, often as early driver events in tumorigenesis. *TP53* mutations are ubiquitous and often correlate with chromosomal instability, poor prognosis, and resistance to platinum-based chemotherapy. Immunohistochemistry (IHC) studies have used *p53* protein accumulation or absence as a surrogate marker for *TP53* mutation status, with strong nuclear staining indicating a mutant phenotype. Experimental models have shown that mutant p53 proteins can exhibit gain-of-function properties, such as promoting metastasis and chemotherapy resistance through interaction with oncogenic transcription factors like NF-Y and ETS2 [53].

Targeting *TP53*-related pathways remains an active area of research, including efforts to restore wild-type p53 activity using small molecules like PRIMA-1MET (APR-246) or to exploit synthetic lethality in *TP53*-deficient cells [54]. *TP53* mutation status is now being considered for molecular classification, prognostic stratification, and treatment response prediction in ovarian cancer patients. It also serves as a foundation for *in silico* pathway analysis, as many *TP53*-regulated genes and networks are involved in cell cycle arrest, DNA repair, and apoptotic signaling key features investigated in this thesis.

2.6.1.2 *BRCA1* Mutations

Previous research indicates that germline mutations in *BRCA1* and *BRCA2* are linked to 13% to 18% of OC. A woman with a BRCA2 mutation has a 20% lifetime risk of developing OC, while a woman with a *BRCA1* mutation has a 50% lifetime risk. Risk-reducing salpingo-oophorectomy has been demonstrated to significantly lower the risk of OC and all cause mortality. The tumor suppressor protein encoded by the *BRCA1* gene, which is found on chromosome 17q21.31, is essential for genomic stability, cell cycle checkpoint regulation, and homologous recombination (HR) DNA repair [55]. Whole-exome sequencing has revealed somatic *BRCA1* mutations and epigenetic silencing (e.g., promoter methylation) in sporadic HGSOC cases, emphasizing that BRCA1 dysfunction is not limited to inherited mutations.

Germline or somatic mutations in *BRCA1* significantly increase the risk of developing high-grade serous ovarian carcinoma (HGSOC) and are observed in up to 15-20% of ovarian cancer cases. Most pathogenic *BRCA1* mutations are frameshift, nonsense, or splice-site mutations, leading to truncated, non-functional proteins. A 2022 multi-center study demonstrated that *BRCA1* mutation carriers with ovarian cancer had significantly better responses to platinum-based chemotherapy and PARP inhibitors like olaparib, owing to homologous recombination deficiency (HRD) [56]. Common hotspot mutations include: 185delAG (c.68_69delAG): a founder mutation in Ashkenazi Jewish populations. 5382insC (c.5266dupC): prevalent in multiple ethnic groups and associated with high penetrance. c.4035

delA: frequently reported in European populations. Recent CRISPR-based functional assays have been used to validate Variants of Uncertain Significance (VUS) in *BRCA1*, improving the classification of risk-associated mutation Finlay [57]. *BRCA1* mutation status is now used for risk stratification, preventive interventions (e.g., risk-reducing salpingo-oophorectomy), and personalized therapy, particularly with PARP inhibitors in HR-deficient tumors.

2.6.1.3 *BRCA2* Gene Mutations in Ovarian Cancer

BRCA2, located on chromosome 13q13.1, is a key tumor suppressor gene involved in homologous recombination (HR), an error-free DNA double-strand break repair pathway. *BRCA2* interacts directly with *RAD51* to mediate strand invasion and repair of damaged DNA. Loss-of-function mutations in *BRCA2* impair genomic stability, resulting in the accumulation of mutations and a higher risk of tumorigenesis [55]. Germline *BRCA2* mutations are responsible for approximately 5–10% of ovarian cancers, predominantly high-grade serous ovarian carcinoma (HGSOC). Most pathogenic mutations are frameshift or nonsense variants that lead to truncated, non-functional proteins. Common mutations include:

1. 6174delT (c.5946delT) – a founder mutation in Ashkenazi Jewish populations.
2. c.9097dupA and c.9976A¿T (p.K3326) – frequently observed in Western populations

In addition to germline mutations, somatic *BRCA2* mutations and epigenetic silencing are found in sporadic cases, suggesting that BRCA2 inactivation can occur through multiple mechanisms. *BRCA2*-mutated ovarian tumors often display homologous recombination deficiency (HRD), making them particularly sensitive to DNA-damaging agents (e.g., platinum) and PARP inhibitors. Recent studies have shown that *BRCA2*-mutated ovarian cancers are associated with better prognosis and longer overall survival compared to *BRCA1*-mutated or non-BRCA-mutated tumors, possibly due to greater responsiveness to platinum-based

chemotherapy and PARP inhibitors. Moreover, *BRCA2* mutation status is a predictive biomarker for treatment response to PARP inhibitors such as olaparib, niraparib, and rucaparib. These therapies exploit the concept of synthetic lethality, selectively killing *BRCA2* deficient tumor cells by blocking their backup DNA repair pathways.

TABLE 2.3: Comparison of *BRCA1* and *BRCA2* Mutations in Ovarian Cancer [55, 58]

Feature	***BRCA1***	***BRCA2***
Gene Location	Chromosome 17q21.31	Chromosome 13q13.1
Protein Function	DNA repair, homologous recombination, transcriptional regulation	DNA repair, homologous recombination via RAD51 recruitment
Type of Mutations	Mostly frameshift, nonsense, or splice-site (e.g., 185delAG, 5382insC)	Mostly frameshift, nonsense (e.g., 6174delT, c.9097dupA)
Prevalence in OC	∼10–15% of epithelial OC	∼5–10% of epithelial OC
Lifetime OC Risk	39–46%	10–27%
Common Subtypes	High grade serous carcinoma (HGSOC)	High grade serous carcinoma (HGSOC)
Age of Onset	Typically earlier onset (before age 50)	Usually later onset compared to *BRCA1* carriers
Prognosis	Intermediate prognosis, variable depending on therapy	Often associated with better overall survival
Therapy Sensitivity	Sensitive to platinum agents and PARP inhibitors	Highly sensitive to platinum agents and PARP inhibitors
Synthetic Lethality Target	Effective (e.g., olaparib, niraparib, rucaparib)	Effective, often more responsive
VUS Burden	Higher number of Variants of Uncertain Significance (VUS)	Lower VUS rate compared to *BRCA1*

2.6.1.4 *PTEN* Gene Mutations in Ovarian Cancer

PTEN (phosphatase and tensin homolog) is a tumor suppressor gene located on chromosome 10q23.31, encoding a dual-specificity phosphatase that antagonizes the PI3K/AKT/mTOR signaling pathway. *PTEN* negatively regulates cell survival, proliferation, and metabolism by dephosphorylating PIP3 to PIP2, thereby

preventing AKT activation. Loss of *PTEN* function leads to uncontrolled AKT signaling, contributing to tumorigenesis, including in ovarian cancer [59]. *PTEN* is frequently mutated, deleted, or epigenetically silenced in ovarian cancer, particularly in the endometrioid and clear cell carcinoma subtypes. Common mutation types include Frameshift mutations, Nonsense mutations, Missense mutations affecting catalytic domains, Promoter hypermethylation leading to transcriptional silencing, Loss of heterozygosity (LOH).

A 2021 study using whole-exome sequencing found *PTEN* inactivating mutations and deletions in 35–40% of endometrioid ovarian carcinoma cases, often co-occurring with PIK3CA mutations, suggesting a cooperative effect in tumor development *In silico* pathway analysis have also confirmed *PTEN* loss as a key event in PI3K pathway activation, a feature observed in many ovarian tumor datasets (e.g., TCGA, cBioPortal), supporting its role in therapeutic targeting.

2.6.1.5 *KRAS* Mutations in Ovarian Cancer

On chromosome 12p12.1, the proto-oncogene *KRAS* (Kirsten rat sarcoma viral oncogene homolog) encodes a GTPase that is essential to the MAPK/ERK signaling pathway, which regulates cell survival, differentiation, and proliferation. *KRAS* mutations cause constitutive activation of downstream signaling cascades, which promotes oncogenesis and unchecked cell proliferation. *KRAS* mutations are uncommon in high-grade serous carcinoma (HGSC), but they are more common in mucinous ovarian carcinoma and low-grade serous ovarian carcinoma (LGSOC). These mutations usually affect exon 2 at codons 12 or 13, resulting in substitutions of a single amino acid that impair GTPase activity:

1. G12D (Gly12Asp)
2. G12V (Gly12Val)
3. G12C (Gly12Cys)
4. G13D (Gly13Asp)

These missense mutations lock *KRAS* in a GTP-bound (active) state, preventing its inactivation and continuously stimulating the RAF-MEK-ERK and PI3K-AKT signaling pathways, promoting cell proliferation and survival. A 2021 genomic profiling study revealed *KRAS* mutations in ∼35–40% of LGSOC and up to 50% of mucinous ovarian carcinomas, highlighting their diagnostic and therapeutic relevance [60]. CRISPR/Cas9-based editing in ovarian cancer models has shown that *KRAS* mutant cells exhibit increased vulnerability to MEK inhibitors (e.g., trametinib), supporting ongoing trials.

2.6.1.6 *PIK3CA* Gene and Its Role in Ovarian Cancer

The p110α catalytic subunit of class I phosphatidylinositol 3-kinase (PI3K), a crucial element of the PI3K/AKT/mTOR signaling pathway, is encoded by *PIK3CA*. Important cellular processes like growth, proliferation, metabolism, and survival are regulated by this pathway. *PIK3CA* mutations cause constitutive activation of PI3K signaling, which aids in the growth, metastasis, and resistance to chemotherapy of ovarian tumors [50]. *PIK3CA* is frequently mutated or amplified in specific ovarian cancer subtypes:

1. Most common in clear cell carcinoma (∼33%) and endometrioid carcinoma (∼20%)

2. Rare in high-grade serous ovarian carcinoma (HGSOC)

Hotspot missense mutations in the helical (exon 9) and kinase (exon 20) domains E545K (exon 9), E542K (exon 9), and H1047R (exon 20) are the most commonly found mutations. Increased phosphorylation of AKT, which promotes cell survival and proliferation, constitutive activation of PI3K independent of receptor tyrosine kinase (RTK) signals, resistance to apoptosis, and improved angiogenesis are the outcomes of these mutations. This oncogenic signaling contributes to the aggressive behavior and therapy resistance of *PIK3CA*-mutated ovarian tumors.

2.7 Dysregulted Signaling Pathways

Dysregulation of intracellular signaling pathways plays a central role in the initiation, progression, and therapeutic resistance of ovarian cancer. Multiple pathways that regulate cell proliferation, survival, differentiation, angiogenesis, and immune evasion become aberrantly activated due to genetic and epigenetic alterations. Among these, the PI3K/AKT/mTOR pathway is the most frequently disrupted, particularly due to *PIK3CA* activating mutations and PTEN loss, leading to uncontrolled cell growth and chemo-resistance [61][62]. The RAS/RAF/MEK/ERK (MAPK) pathway, commonly altered via *KRAS* mutations, is especially relevant in low-grade serous and mucinous ovarian cancers. Additionally, *TP53* mutations, present in over 95% of high-grade serous ovarian carcinomas (HGSOC), disrupt the p53 signaling pathway, undermining genomic stability and apoptosis [60]. Other dysregulated pathways include Wnt/β-catenin, Notch, TGF-β, NF-κB, and JAK/-STAT, all of which contribute to tumour cell plasticity, epithelial–mesenchymal transition (EMT), inflammation, and immune evasion [62]. Understanding the complex network of these dysregulated pathways is essential for identifying novel biomarkers and developing targeted therapeutic strategies in ovarian cancer.

TABLE 2.4: Dysregulated Signaling Pathways in Ovarian Cancer [60, 62].

Signaling Pathway	**Key Dysregulated Genes**	**Associated OC Subtypes**	**Functional Role in OC**
PI3K/AKT/mTOR	*PIK3CA, PTEN, AKT1, mTOR*	Clear cell, endometrioid	Cell survival, proliferation, chemoresistance
RAS / sRAF / MEK / ERK (MAPK)	*KRAS, BRAF, NRAS*	Low-grade serous, mucinous	Cell proliferation, differentiation
p53 pathway	*TP53*	High-grade serous	DNA damage response, apoptosis, genomic stability
Wnt/β-catenin	*CTNNB1, APC, AXIN1*	Endometrioid	EMT, migration, cell polarity
Notch pathway	*NOTCH1, NOTCH3, DLL4*	Serous, endometrioid	Angiogenesis, cell differentiation
TGF-β pathway	*SMAD2, SMAD4, TGFBR2*	Advanced OC stages	EMT, metastasis, immune modulation

Table 2.4 continued from previous page

Signaling Pathway	**Key Dysregulated Genes**	**Associated OC Subtypes**	**Functional Role in OC**
NF-κB pathway	*RELA, NFKB1, IKK complex*	Inflammatory microenvironment	Inflammation, immune evasion
JAK/STAT pathway	*JAK1, STAT3, IL6*	Serous	Cytokine signaling, immune response
VEGF/VEGFR	*VEGFA, FLT1 (VEGFR1), KDR (VEGFR2)*	High-grade serous	Angiogenesis, vascular remodeling
Hedgehog pathway	*SHH, PTCH1, SMO, GLI1*	Clear cell, some HGSOC	Stemness, proliferation

2.8 Epigenetic Modifications in Ovarian Cancer

The pathophysiology and progression of ovarian cancer are significantly influenced by epigenetic modifications, such as DNA methylation, histone modifications, and non-coding RNA regulation, which change gene expression without affecting the underlying DNA sequence. The transcriptional silencing caused by aberrant promoter hypermethylation of tumor suppressor genes, including *BRCA1*, *PTEN*, *RASSF1A*, and *MLH1*, compromises DNA repair, cell cycle regulation, and apoptosis. For example, in roughly 10–15% of cases of sporadic high-grade serous ovarian carcinoma (HGSOC), *BRCA1* promoter methylation is found. This mimics the effects of BRCA1 mutations and contributes to homologous recombination deficiency (HRD). Genes involved in tumor *suppression* are also repressed and chromatin condensation is encouraged by histone deacetylation mediated by enzymes such as *HDAC1* and *HDAC2*. Furthermore, microRNAs (miRNAs), especially those in the miR-200 family and miR-21, are often. These epigenetic changes not only contribute to early tumorigenesis and metastasis but also serve as potential diagnostic biomarkers and therapeutic targets, especially in combination with epigenetic drugs such as DNA methyltransferase inhibitors (e.g., decitabine) and HDAC inhibitors (e.g., vorinostat) under clinical investigation for ovarian cancer [63].

2.9 Clinical Challenges in Ovarian Cancer

Ovarian cancer remains one of the most lethal gynecologic malignancies, largely due to a multitude of unresolved clinical challenges that hinder early detection, effective treatment, and long-term disease control. One of the foremost issues is the late diagnosis of the disease; nearly 75% of ovarian cancer cases are detected at advanced stages (stage III or IV), primarily because early-stage disease is often asymptomatic or presents with vague symptoms that mimic benign conditions [64]. Furthermore, tumor heterogeneity, both intertumoral and intratumoral, complicates treatment planning, as different histological subtypes such as high-grade serous, clear cell, and endometrioid carcinomas exhibit variable genetic profiles and therapeutic responses [65]. A major therapeutic obstacle is the development of chemoresistance although many patients initially respond well to platinum-based chemotherapy, most eventually relapse with drug-resistant tumors. Resistance to novel agents like PARP inhibitors and taxanes further limits treatment options in recurrent disease. Another significant limitation is the lack of reliable biomarkers; while CA-125 is widely used, its sensitivity and specificity are insufficient for early detection, highlighting the need for more effective molecular markers. The limited efficacy and accessibility of targeted therapies, such as anti-angiogenic agents and DNA repair inhibitors, pose additional barriers, particularly in low-resource settings. Moreover, the complexity of metastasis, which in ovarian cancer often occurs via peritoneal dissemination rather than hematogenous spread, remains poorly understood and difficult to manage clinically [66]. Surgical challenges also persist, as optimal cytoreductive surgery linked to improved outcomes is technically demanding and often achievable only in specialized centers. Additionally, psychosocial challenges, including treatment-induced infertility, menopausal symptoms, and reduced quality of life, are frequently under-addressed. The high recurrence rate, with up to 80% of advanced-stage patients relapsing, and the limited availability of precision medicine tools, such as next-generation sequencing, further underscore the need for integrative and personalized management strategies. Addressing these multifaceted clinical hurdles is essential to improving outcomes in ovarian cancer patients.

2.10 Bioinformatics in Addressing Clinical Challenges in Ovarian Cancer

Bioinformatics has become an essential tool in overcoming many of the clinical challenges associated with ovarian cancer by enabling the integration, analysis, and interpretation of high-throughput biological data.

One of the major contributions is in the early detection and biomarker discovery, where bioinformatics pipelines analyze transcriptomic, proteomic, and epigenomic data to identify non-invasive diagnostic markers, such as circulating miRNAs, lncRNAs, and methylation patterns.

These computational approaches help refine early diagnostic models with improved sensitivity and specificity over conventional CA-125 testing. In terms of tumor heterogeneity and treatment resistance, multi-omics integration tools and machine learning models allow for the classification of ovarian cancer into molecular subtypes with distinct mutation profiles, drug sensitivities, and prognoses.

For example, The Cancer Genome Atlas (TCGA) and cBioPortal platforms have enabled researchers to identify actionable mutations (e.g., in *PIK3CA*, *PTEN*, *BRCA1/2*) and to predict drug response patterns *in silico*.

Additionally, bioinformatics tools support the analysis of single-cell RNA sequencing (scRNA-seq) data to dissect tumor microenvironment complexity, revealing key players in immune evasion, angiogenesis, and chemo-resistance. Furthermore, pathway analysis tools such as KEGG, Reactome, and STRING aid in understanding dysregulated signaling pathways and in identifying potential therapeutic targets.

Bioinformatics also facilitates virtual drug screening, network pharmacology, and synthetic lethality prediction, contributing to the development of precision medicine strategies for ovarian cancer [67]. These advancements collectively accelerate research, reduce laboratory costs, and improve the clinical decision-making process in ovarian cancer management (Table 2.5).

TABLE 2.5: Treatments for Ovarian Cancer and Bioinformatics Applications [68].

Treatment	**Description**	**Bioinformatics Role**
Surgery	Primary method for tumor removal.	Gene profiling helps predict tumor spread and surgical outcomes.
Chemotherapy	Platinum-based drugs (e.g., cisplatin, carboplatin).	BI identifies resistance genes and predicts therapy response.
PARP Inhibitors (Targeted)	Used in BRCA-mutated or HR-deficient tumors.	BI detects BRCA1/2 mutations and homologous recombination deficiency (HRD) scores.
Hormonal Therapy	Block estrogen in hormone sensitive tumors.	BI assesses hormone receptor gene expression.
Immunotherapy	Uses checkpoint inhibitors to enhance immune response.	BI evaluates tumor immune profiles and predicts responsiveness.
Anti-angiogenic Therapy	Inhibits tumor blood vessel formation	BI analyzes VEGF-related gene expression and pathways.
Combination Therapy	Uses two or more therapies together.	BI models synergistic effects and helps optimize treatment combinations.

2.11 Risk Factors of OC in Asian Populations

The risk of ovarian cancer is increasing in Asian populations due to a number of risk factors, including genetic predisposition, lifestyle and reproductive patterns, and environmental exposures. The prevalence of *BRCA1* and *BRCA2* mutations varies in some Asian countries, despite the fact that genetic data on these mutations increases exposure risks globally. For additional reasons, other population-specific genetic markers are also important, such as mutations linked to Lynch syndrome, *RAD51*, and oocyte-specific gamma-tubulin. The increased risk of ovarian cancer is directly linked to reproductive patterns like delayed childbirth, decreased fertility, and nulliparity, which have recently been spreading throughout Asian nations as a result of societal and economic changes. For instance, women in South Korea and Japan are more likely to put off having children in favor of concentrating on their education and careers [69]. The lifestyle patterns, such as

obesity, sedentary behavior, and high-fat diets, which are rapidly spreading across many Asian countries, are the known risk causes of ovarian cancer. The increased cancer risk due to dietary and behavioral patterns is observed in many countries willing to integrate into the Western lifestyle.

Finally, some environmental and occupational exposures such as asbestos and talc use, which is observed in some Asian countries but not sufficiently proven through evidence, may also contribute to increased risks. Obesity, sedentary lifestyles, and high-fat diets are the known risk factors for ovarian cancer, and they are rapidly spreading throughout many Asian countries.

Many nations that are eager to adopt the Western way of life have seen an increase in the risk of cancer as a result of dietary and behavioral changes. Last but not least, certain occupational and environmental exposures, like the use of talc and asbestos, which have been noted in some Asian nations but have not been adequately supported by data, may also raise risks.

2.12 Prognosis

Stage III or IV is seen in as many as 70% of women with epithelial ovarian cancer. While the stage at diagnosis is a powerful predictor of prognosis, the histologic grade is equally important, especially for recurrence prediction. Epithelial ovarian cancer has a 95% to 99% 10-year survival rate, is frequently detected at stage I, and in 15% of cases is found to have little malignant potential according to histology. Table 2.6 provides a summary of the poor prognostic variables [70].

TABLE 2.6: Poor Prognostic Factors in Ovarian Cancer [8]

Sr. No.	**Prognostic factor**
1	Age older than 65 years
2	Clear cell or mucinous tumors (histology)
3	Extensive disease (advanced stage)
4	Large volume of residual tumor
5	Reduced overall quality of life score
6	Low quality cell differentiation

2.13 Role of Bioinformatics in Ovarian Cancer Gene Analysis

Diagnostic and prognosis prediction tests have become much more accurate with the use of AI. Several malignancies, including prostate, stomach, and breast cancers, have shown encouraging outcomes [71]. Thanks to bioinformatics advancements, large-scale genomic analysis are now possible, which has greatly improved cancer research by allowing for the discovery of new biomarkers and treatment targets [72]. Gene expression, mutations, and signalling pathways linked to the course of ovarian cancer can be better understood with the use of bioinformatics techniques. Public databases like GEO and The Cancer Genome Atlas (TCGA) offer researchers with excellent datasets for molecular level studies of ovarian cancer [73]. Using sophisticated algorithms, Coremine Medical correlates user enquiries with gene profiles and returns a prioritised list of relevant genes. Gene profiling enhances the accuracy of gene identification in response to user queries by assigning a keyword profile to each gene based on keywords taken from biomedical literature [74].

As one of widely used bioinformatics databases, PPIs or the related bioinformatics **STRING** (Search Tool for the Retrieval of Interacting Genes/Proteins) allows protein interactions to share with the thousands of users. The importance of STRING for the understanding of functional associations between genes and proteins functionally important in tumorigenesis, progression or metastasis of ovarian cancer research has been underscored in [75]. Usually, to identify groups of genes having similar expression pattern or functional roles, clustering is one of the most common methods used in bioinformatics. By clustering we can also identify functionally related genes in ovarian cancer that may act together in cancer development or progression. Since these gene clusters are now verifiable and interpretable, they can now be verified and interpreted using tools such as **KEGG** (Kyoto Encyclopedia of Genes and Genomes) [76]. It is database for annotation, visualization, and integrated discovery (**DAVID**) which is a very powerful bioinformatics tools to perform the functional annotation of gene lists.

In ovarian cancer research, **DAVID** helps to assign biological meaning to large sets of differentially expressed genes (DEGs) by mapping them to gene ontology (GO) terms, biological pathways, and functional clusters.

After identifying gene clusters from expression datasets in ovarian cancer, one essential step in bioinformatics analysis is validating these clusters through protein–protein interaction (PPI) networks. PPIs help confirm that the clustered genes not only share similar expression patterns but also participate in shared biological functions or pathways by directly or indirectly interacting at the protein level.

PPI network construction is commonly performed using databases such as STRING, which compile interaction data from experimental studies, computational predictions, co-expression evidence, and curated biological knowledge [77]. By inputting the clustered gene list into these tools, researchers can visualize the interconnectedness of the proteins they encode.

2.14 Limitation of *In Silico* Analysis

Despite their usefulness, these tools have several limitations that must be considered. Literature mining platforms like Coremine are limited by the availability and quality of published data, often missing recently discovered genes or context-specific interactions.

STRING may include predicted interactions that lack experimental validation and does not account for tissue or stage-specific expression patterns critical in OC heterogeneity. Similarly, DAVID and KEGG are based on predefined, static databases that may not capture dynamic changes in gene function or regulation, particularly under treatment pressure or in recurrent tumors [60]. Tools like Protein Prompt offer computational predictions without biological validation, which can lead to false positives. Therefore, while these bioinformatics platforms significantly enhance our ability to study ovarian cancer, their outputs must be interpreted cautiously and, where possible, supported by experimental evidence or clinical data.

2.15 Comparative Studies and Recent Advances

Here are three recent comparative *in silico* studies (2015–2024) that have significantly advanced our understanding of ovarian cancer through bioinformatics-driven, gene-based analysis:

2.15.1 Identification of Four Hub Genes via WGCNA & PPI Networks

A 2020 study conducted a weighted gene co expression network analysis (WGCNA) on microarray data (e.g., GSE49997), followed by protein–protein interaction (PPI) network integration using STRING and DAVID for enrichment analysis. The authors identified four key hub genes *COL6A3*, *CRISPLD2*, *FBN1*, and *SERPINF1* that were strongly associated with poor prognosis in ovarian cancer patients [78].

These genes were validated in independent datasets, suggesting robust prognostic value derived entirely from *in silico* modeling.

2.15.2 Integrated Bioinformatics for Driver Gene & miRNA Discovery

In a 2023 study, researchers combined differential expression analysis, miRNA-mRNA network modeling, and GO/KEGG pathway enrichment to identify prognostic biomarkers in ovarian cancer. Using datasets like GSE119055, they pinpointed several up and down-regulated miRNAs (e.g., hsa-miR-182-5p, hsa-miR-501-3p) and linked these to hub genes such as *BCL2, CADM1, ELAVL2,* and *PRKACB*, which may influence tumor progression [79].

This demonstrates the power of integrated *in silico* pipelines to reveal novel regulatory networks.

2.15.3 Multi-Omics Driver Gene Discovery via cBioPortal & Molecular Docking

A 2024 study employed cBioPortal to analyze TCGA ovarian serous carcinoma datasets, identifying frequently altered genes. They then used molecular docking and molecular dynamics simulations to examine protein ligand interactions for potential drug repositioning. This combined genomic and structural *in silico* approach enabled pinpointing of therapeutic targets and assessing drug-binding dynamics all computationally [80].

TABLE 2.7: Comparision of BI methods and their findings [79, 80].

Study	**Methods Used**	**Key Findings**
WGCNA + PPI (2020)	Co-expression + STRING + DAVID	Hub genes: *COL6A3, CRISPLD2, FBN1, SERPINF1*
miRNA–mRNA network (2023)	GEO data + Enrichr + Network Analysis	miRNAs & target genes: *BCL2, CADM1, ELAVL2, PRKACB*
cBioPortal + Docking (2024)	TCGA + cBioPortal + Docking + MD Simulations	Identified driver genes and potential inhibitors

Each study highlights how *in silico* analysis from co-expression networks to structural modeling can yield valuable candidate biomarkers and therapeutic targets in a hypothesis-driven yet cost-effective manner. Integrating these approaches in methodology will strengthen the rigor and biomedical relevance of thesis.

2.16 Research Gaps and Future Directions in *In Silico* Analysis of OC

Over the past decade, *in silico* approaches have significantly advanced the understanding of ovarian cancer (OC) by enabling high-throughput analysis of genomic data and functional interpretation of disease-associated genes. Numerous studies have identified key genes such as *BRCA1/2, TP53, PIK3CA, KRAS,* and *PTEN*, using databases like TCGA, and analyzed them with tools such as

STRING, DAVID, KEGG, and Coremine Medical. These genes' roles in DNA repair, apoptosis, signal transduction, and cell adhesion all of which are crucial for the development and progression of OC have been identified through functional annotation. PI3K/AKT/mTOR, p53, Wnt/β-catenin, and MAPK/ERK are examples of dysregulated pathways that have been repeatedly linked to drug resistance, tumor growth, and metastasis. Furthermore, the identification of epigenetic regulators that contribute to gene silencing in OC, such as histone modifiers and DNA methylation markers, has been made possible in large part by *in silico* tools.

Despite significant advances in understanding the molecular mechanisms of ovarian cancer (OC), there remain substantial gaps in the comprehensive functional annotation and validation of genes potentially involved in its pathogenesis. Most studies have focused predominantly on a limited set of well-characterized genes such as *BRCA1, BRCA2, TP53,* and *PTEN* [81]-[82]. Consequently, many other genes that may play contributory or modulatory roles in tumor development remain uncharacterized. While *in silico* tools like STRING, DAVID, and KEGG pathways provide valuable platforms to predict gene functions and interactions, their reliance on existing curated data inherently biases results toward already-studied genes [83]. This leaves a gap in discovering truly novel genes or unannotated interactions. Moreover, the data in public databases is often incomplete or outdated, and functional annotations may lack specificity for tissue type or cancer subtype [84]. Another critical gap lies in the lack of integration of multi-omics data, most analyses, including many *in silico* studies, rely solely on genomics or transcriptomics without considering proteomics, epigenomics, or metabolomics, which could provide a more holistic view of OC pathogenesis [85]. Furthermore, computational findings are often not experimentally validated, and rare OC subtypes remain understudied. Finally, while *in silico* predictions are valuable, the lack of experimental validation of predicted interactions and pathways limits their clinical translatability [86]. Comparative studies from recent years have shown that combining gene expression profiling, miRNA-target prediction, co-expression networks, and protein–protein interaction mapping enhances biomarker discovery

and therapeutic target prediction. Future directions should prioritize the integration of single-cell data, machine learning models, and clinical metadata to create predictive tools for patient-specific diagnosis and treatment. Bioinformatics platforms are thus indispensable for bridging genotype–phenotype gaps and guiding translational research in ovarian cancer. Therefore, future studies should aim to combine robust computational predictions with experimental confirmation and incorporate integrative multi-omics approaches to fill these knowledge gaps.

Chapter 3

Material and Method

In this study, a systematic *in silico* approach was used to identify and functionally annotate genes implicated in ovarian cancer pathogenesis. Gene candidates were retrieved using Coremine Medical, based on co-occurrence with the term "ovarian cancer." Protein–protein interactions were constructed using the STRING database, and key hub genes were identified based on network topological parameters. Functional annotation was carried out using DAVID to obtain enriched GO terms and KEGG pathways. Pathway mapping was validated using KEGG Mapper. Additionally, STRING multiple protein feature was employed to predict protein functionality based on amino acid sequences. The integration of these tools allowed for the comprehensive annotation of gene functions and their mechanistic roles in ovarian tumorigenesis.

This study follows a translational research approach, utilizing insilico bioinformatics tools to bridge computational predictions with potential clinical applications. The integration of these complementary bioinformatics tools facilitated a comprehensive analysis of gene functions, interactions, and their potential mechanistic roles in ovarian tumorigenesis, laying the groundwork for identifying novel biomarkers and therapeutic targets. Visual representations, including flowchart, is generated to clearly illustrate the workflow and results (Figure 3.1).

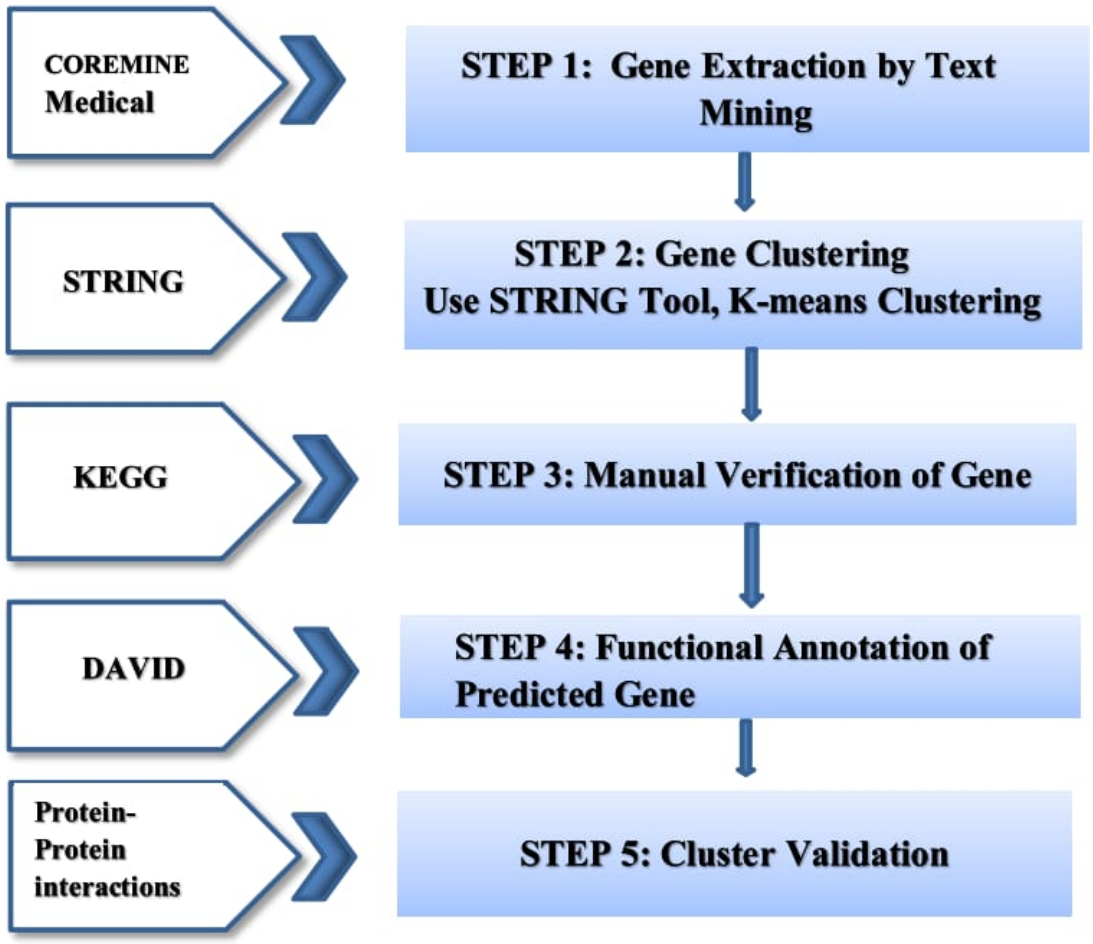


FIGURE 3.1: Five Bioinformatics tools unraveling ovarian cancer genes [74][76].

3.1 Retrieval of Genes Related to Ovarian Cancer by Using COREMINE MEDICAL

Using the firm's one-of-a-kind text mining algorithms and vast biomedical data integration and analysis pipeline, the Norwegian bioinformatics company PubGene developed a suite of tools known as COREMINETM. These products were built by the company. On top of the COREMINE Platform, Coremine MedicalTM is the first information community tailored to a particular domain. This is a free Internet service that allows users to search for, update, and share medical information online. It functions as both a search engine and a social network.

3.1.1 Functionality of Coremine Medical

Using sophisticated algorithms, Coremine Medical correlates user enquiries with gene profiles and returns a prioritised list of relevant genes. Gene identification via user queries becomes more accurate with the help of keyword pofiling, which involves assigning a set of biomedical literature-derived keywords to each gene. When it comes to text mining, having access to material on the proper topic is absolutely necessary.

The principal source, the PubMed database, provides access to this information in the form of a published corpus about ovarian cancer in the field of biomedical literature. The Gene/Protein database was searched for the main topic "Ovarian Cancer" for the purpose of retrieving genes related to this cancer. A collection of texts that were published during the last five years i.e. 2019 to 2024 was gathered.

3.1.2 Significance of Coremine Medical in Gene Identification

The use of Coremine Medical in this study played a pivotal role in the initial identification and prioritization of ovarian cancer-associated genes. As a biomedical text-mining platform, Coremine Medical leverages co-occurrence patterns from vast scientific literature to establish semantic associations between diseases and genes. By querying terms such as "*Ovarian Cancer and Genes*", the tool enables rapid extraction of biologically relevant gene candidates that are frequently reported in relation to the disease. This step ensures that the analysis is grounded in evidence-based literature, which is particularly important when constructing gene lists for downstream bioinformatics analysis. Moreover, Coremine's scoring system helps in filtering genes based on association strength, allowing the inclusion of highly relevant genes for further functional annotation and interaction mapping. Thus, this step not only improves the biological relevance and accuracy of the study but also enhances the efficiency of data curation by focusing on well-supported genetic factors implicated in ovarian cancer pathogenesis.

3.2 STRING Tool in Ovarian Cancer Analysis

To investigate the potential functional associations among the selected mismatch genes, the STRING database (version 12.0) was used to build a protein-protein interaction (PPI) network [87]. To guarantee species-specific interaction mapping, ovarian cancer gene symbols were uploaded into the STRING web interface and the organism was set to Homo sapiens. A cutoff of 0.7 for the high confidence interaction score was used to screen for biologically significant and robust associations. The resulting interaction network was visualized using STRING's interactive viewer, displaying both direct (physical) and indirect (functional) protein associations based on curated databases, experimental data, and computational predictions. For network refinement, clustering analysis was performed using the K-means clustering algorithm, as provided within STRING's clustering tools [88].

This grouped proteins into clusters based on interaction similarity and network topology. Each cluster was examined to identify potential biological themes or pathway involvement, with special attention given to clusters enriched in cancer-related functions such as signal transduction, DNA repair, drug resistance, and transcriptional regulation. The network's statistical significance was confirmed by STRING's internal enrichment analysis, which yielded a p-value of 1.0×10^{-16}, indicating that the observed interactions were highly unlikely to have occurred by random chance.

3.2.1 Selection and Documentation of STRING Clusters

Following the generation of the STRING interaction network and its clustering via K-means, each resulting cluster was manually examined based on gene content, functional annotations, and interaction density. Clusters showing high internal connectivity and involvement in biological processes relevant to ovarian cancer such as epithelial structure maintenance, drug resistance, transcriptional regulation, and DNA repair were prioritized for further study. The number of clusters

was incrementally adjusted starting from 3, 4, 5, 6, and so on, and then implemented. Each selected cluster was visualized using STRING's built-in layout options, and the resulting network images were exported as high-resolution screenshots for inclusion in the Results chapter. Gene members of each cluster were recorded and further analyzed using DAVID functional annotation clustering and KEGG pathway mapping to identify enriched molecular functions and biological pathways.

K-means clustering was performed on the STRING-derived interaction network of the 2000 genes identified from Coremine, with the number of clusters (k) empirically set to 24 to capture distinct functional modules. The clustering assigned all genes into one of the 24 clusters, as required by the algorithm. Notably, clusters beyond the 15th contained only a single gene each. These single-gene clusters were retained in the analysis for completeness and were manually verified in KEGG to assess potential functional relevance. However, in subsequent interpretation and discussion, only clusters containing two or more genes were considered biologically meaningful. This integrative approach helped confirm the functional relevance of the clusters and strengthened the association of these genes with potential ovarian cancer mechanisms. Clusters with unclear or minimal functional coherence were excluded from detailed discussion but archived for reference.

3.3 Clusters of Ovarian Cancer Associated Genes Verified by KEGG

The KEGG (Kyoto Encyclopedia of Genes and Genomes) database was used to manually verify the involvement of specific genes in biological pathways linked to ovarian cancer. The KEGG Medicus portal was used to conduct individual searches for the genes that were shortlisted from STRING clustering (www.genome.jp/kegg), and their presence was verified in the "Pathways in cancer" (hsa 05200) and "Ovarian cancer" (hsa05213) pathway maps. Each gene symbol was checked for its occurrence within the pathway diagrams by either entering the

gene name directly into the KEGG search bar or pasting the complete gene list into the "Search Pathway" tool. The KEGG pathway comprised 160 ovarian cancer pathways. In all 160 pathways, we verified clusters ranging from 3 to 24, as obtained from the STRING database. We identified matching genes in 15 pathways. Out to total 65 cluster genes, 16 genes were matched in pathways and these are *KRT14, KRT17, TPTEP2, DNAH12, DNAH2, DNAH15, DNAH8, ABCB11, ABCC2, SLC22A1, ATP7B, HOXA10, MAGT1, RPN2, HTR1E, GALT*.

Genes that were highlighted in KEGG diagrams were labeled as matched genes, indicating their known or curated involvement in cancer pathways. In contrast, genes that did not appear were marked as mismatched and considered for further analysis. These mismatched genes, while not listed in the KEGG ovarian cancer map, may still hold functional relevance and represent novel or less-characterized candidates potentially involved in disease mechanisms.

This manual verification step added an extra layer of validation and helped distinguish between genes with established cancer associations and those with unexplored roles, forming the basis for subsequent functional annotation clustering in DAVID and interaction prediction . This workflow ensured that both known and potentially novel ovarian cancer genes were systematically evaluated.

3.4 Functional Annotation of Ovarian Cancer Genes Using DAVID

To functionally characterize the genes that did not appear in KEGG ovarian cancer pathway maps (mismatched genes), the bioinformatics resource DAVID (Database for Annotation, Visualization and Integrated Discovery) (https://david.ncifcrf.gov/) was used to perform annotation clustering on a list of 49 genes. The analysis was performed with the following settings: species selected as Homo sapiens, and the gene identifier format set to official gene symbols. The gene list was uploaded into the "Functional Annotation Tool," and DAVID performed enrichment analysis using integrated databases such as Gene Ontology (GO), InterPro, and SMART

to identify shared biological functions, molecular domains, and pathway associations. The tool generated eight annotation clusters, each representing a group of functionally related genes. For each cluster, DAVID provided an enrichment score (indicating the strength of shared annotations) and associated p-values for statistical significance of term enrichment.

3.4.1 Presentation of DAVID Functional Annotation Results

The output from DAVID functional annotation clustering was systematically organized and presented in the Results chapter. A total of eight annotation clusters were generated from the 49 mismatched genes, each accompanied by an enrichment score, gene count, and statistical p-values for associated biological terms. For clarity and visual representation, screenshots of all eight clusters were included in the results section, displaying the enriched functional terms and their significance levels. In addition to the screenshots, a summary table was prepared showing each cluster's number, constituent genes, enrichment score, top enriched terms, and potential significance in ovarian cancer pathogenesis. Clusters with higher enrichment scores and cancer-relevant functions, such as those involving keratin family proteins, ABC transporters, and transcription factors, were emphasized in the discussion. This approach ensured that both visual evidence and detailed interpretation of functional patterns were available to support the exploration of novel gene involvement in ovarian cancer.

3.5 Cluster Validation Through Protein-Protein Interactions

Thus given expression datasets in ovarian cancer, it is necessary to validate gene cluster from bioinformatics analysis through protein–protein interaction (PPI) networks. The clustered genes are then confirmed in PPIs to prove that the clustered

genes do not only exhibit similar expression patterns but also participate in similar biological functions or pathways by directly or indirectly interacting at the protein level [89]. Commonly, PPI network construction is performed using databases like STRING that contains interaction data from experimental studies, computational predictions, coexpression evidence and from curated biological knowledge. Researchers are able to visualize the interconnectedness of the proteins encoded by the inputted clustered gene list through inputting the clustered gene list into these tools [90].

These PPI validated clusters are then further assessed through the use of topological analysis, as well as betweenness, degree centrality and clustering coefficient. This step backs up that found gene clusters are not random but are biological meaningful units in the molecular brain of ovarian cancer [91]. PPI analysis for cluster validation improves biological interpretation, robustness and relevance of gene clustering results in ovarian cancer studies.

These PPI validated clusters are then further assessed through the use of topological analysis, as well as betweenness, degree centrality and clustering coefficient. This step backs up that found gene clusters are not random but are biological meaningful units in the molecular brain of ovarian cancer [92]. PPI analysis for cluster validation improves biological interpretation, robustness and relevance of gene clustering results in ovarian cancer studies.

3.5.1 Gene Selection for Protein-Protein Interaction Prediction

Protein-protein interaction (PPI) analysis was carried out using the STRING database (version 12.0) to look into possible functional relationships between unidentified genes and known ovarian cancer-related proteins. A total of 49 mismatch genes from Cluster 3 to 24, identified through KEGG pathway comparison and DAVID annotation, were tested individually against seven well-established ovarian cancer proteins: B*RCA1, BRCA2, TP53, PTEN, PARP1, AKT1,* and

PIK3CA. The "Multiple Proteins" option in STRING was used, with Homo sapiens selected as the organism. Each mismatch gene was input alongside the known OC gene to assess potential interaction. STRING provided a combined interaction score based on evidence from experimental data, curated databases, text mining, and co-expression. A combined score threshold of ≥ 0.700 was considered indicative of a high-confidence functional interaction. All interaction scores and evidence types were recorded, and network visualizations were captured for interactions meeting the high-confidence criteria. Visual STRING network diagrams were captured as screenshots for high-scoring interactions, and results were tabulated for further interpretation and integration with functional annotation findings.These STRING-derived interactions suggest that several genes not currently mapped in KEGG ovarian cancer pathways may still functionally contribute to ovarian tumor biology and should be explored as novel targets.

Chapter 4

Results

4.1 Retrieval of Genes by Coremine Medical

The initial stage was extracting genes using the COREMINE tool. This was accomplish by accessing the search bar of the tool and conducting a search for the specific disease, Ovarian Cancer. After selecting the "extracted associations" option in the COREMINE interface, we proceeded to extract the associations spanning from the year 2019 to 2024. COREMINE generated a list of 2000 genes associated with ovarian cancer when the query "ovarian cancer (Genes/Proteins)" was used. A comprehensive Excel spreadsheet with the names of the corresponding genes, disorders, and their respective descriptions was acquired (Table 4.1).

Table 4.1: List of genes of Ovarian Cancer generated by Coremine Medical

Gene Symbol	**Gene Symbol**	**Gene Symbol**	**Gene Symbol**
BRCA1	*MUC16*	*NBR1*	*PALB2*
WFDC2	*COL11A2*	*PARP1*	*RAD51C*
MRPL36	*CHEK2*	*BLID*	*RAD51D*
TP53	*BARD1*	*BRIP1*	*RAD51*
ARID1A	*FANCD2*	*ERBB2*	*BRCA3*
MSLN	*PTEN*	*MSH2*	*FANCB*
MLH1	*TSC1*	*BRAP*	*FANCF*
MSH6	*FOLR1*	*PIK3CA*	*PMS2*
CDH1	*FANCG*	*FANCE*	*PAX8*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CDK12	*RAD50*	*FZR1*	*FOXL2*
C11orf30	*XRCC2*	*CCNE1*	*BCCIP*
WT1	*SLC13A5*	*MRE11A*	*FANCM*
FANCL	*ERCC1*	*MUC1*	*DIRAS3*
CDKN2A	*EPCAM*	*RAD51B*	*SLC13A2*
KRAS	*VIM*	*CT55*	*VEGFA*
KRT7	*STK11*	*FAM175A*	*CHEK1*
CRSA	*CDKN1A*	*ABCB1*	*BCL2*
C10orf2	*SHFM1*	*ESR1*	*RAD52*
DICER1	*BIRC5*	*CALB2*	*RB1*
NBN	*CD274*	*FANCI*	*EGFR*
H3F3AP6	*TCEAL1*	*HOXB13*	*CCND1*
XRCC3	*FANCA*	*AFP*	*ALDH1A1*
MMP2	*TP53BP1*	*RMI2*	*CD44*
PARP2	*PGR*	*CASP3*	*SMARCA4*
TBC1D9	*CTNNB1*	*LOC100508689*	*ZNF217*
FANCC	*MTOR*	*NLRP2*	*BCL2L1*
CRISP3	*PPP2R1A*	*MYC*	*BRAF*
POLQ	*TIPARP*	*CD24*	*MMP9*
STAT3	*ATR*	*PDCD1*	*AKT1*
DPH1	*KRT20*	*VTCN1*	*PROM1*
FSHR	*ERBB3*	*RAD54L*	*UIMC1*
NF1	*XIAP*	*WEE1*	*PLAU*
AKT2	*SLX4*	*RMI1*	*KLK10*
RBBP8	*ZEB1*	*TERT*	*CLDN4*
CDX2	*CTAG1B*	*HIF1A*	*RPL17*
FHIT	*MSH3*	*MDM2*	*EZH2*
BAP1	*DDR1*	*PAK3*	*SMAD4*
RASSF1	*CSF3*	*CLDN3*	*TOX3*
TCF12	*ANIB1*	*ERCC4*	*NME1*
CASP9	*NR4A1*	*CDK4*	*CYP19A1*
EIF3A	*CCNB1*	*BAX*	*MAGI1*
RNF2	*POU5F1*	*CDK2*	*CD99*
CDKN1B	*DMC1*	*CD99L2*	*ERCC2*
H2AFX	*SALL4*	*SLC31A1*	*PAX2*
KIT	*BACH1*	*PRKDC*	*PMS1*
ATRX	*HID1*	*PPM1D*	*CDK1*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CASP8	*FOXM1*	*MET*	*MCL1*
TTF1	*GSTP1*	*POLK*	*TCHP*
NBR2	*MCPH1*	*XRCC1*	*WWOX*
MDM4	*AMHR2*	*NEURL1*	*HMG20B*
NOTCH1	*EXO1*	*AURKA*	*TNFRSF10B*
PRSS58	*MLH3*	*GINS2*	*KMT2C*
PTK2	*NOTCH3*	*SWSAP1*	*RECQL*
ERBB4	*RAD51AP1*	*NKX2-1*	*COL11A1*
BUB1B	*AR*	*SNAI1*	*LRP1B*
ABHD8	*ZEB2*	*TNS1*	*TNFSF10*
ABCG2	*NCOA3*	*MIR145*	*FBXW7*
KDR	*HMGA2*	*HNF1B*	*CDK6*
KLK4	*RNF43*	*FGFR2*	*MRC1*
SNAI2	*PTCH1*	*AMH*	*KMT2D*
POLR1A	*NAPSA*	*XRCC4*	*BECN1*
XRCC5	*FAAP24*	*SKP2*	*NR5A1*
INHBE	*ABCC2*	*PECAM1*	*GPC3*
AXL	*HSF2BP*	*MAP3K1*	*BRE*
HELQ	*NCAM1*	*HOTAIR*	*LAG3*
POLE	*PCNA*	*SLFN11*	*PLK1*
ARL11	*ARID1B*	*SOX2*	*NRAS*
E2F5	*MDC1*	*BRCC3*	*MUS81*
LGR6	*GADD45A*	*TXK*	*BABAM1*
DES	*TOPBP1*	*SERPINB5*	*RP1-4G17.5*
E2F1	*CDC25A*	*OR2A1*	*OR2T35*
ZNF703	*MMP14*	*GATA3*	*APC*
PIK3R1	*FSCN1*	*CCNA2*	*CDKN3*
BRMS1L	*MKI67*	*GLI1*	*RHOC*
LIG4	*TYMS*	*ZNF93*	*CSF1*
PIP	*LIN28B*	*RAD54B*	*DIABLO*
MME	*PRSS1*	*IGF2BP3*	*BRMS1*
RAD18	*XRCC6*	*TUBB3*	*COL18A1*
GPER1	*METTL3*	*MIR21*	*ERCC3*
PRKCI	*INPP4B*	*HIC1*	*HRAS*
FAT3	*CTAG2*	*WTAP*	*CXCL12*
MECOM	*FOXA1*	*WNT5A*	*PGRMC1*
MAD2L2	*POLD1*	*FOXO3*	*KLLN*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
ESR2	*PVRL4*	*E2F3*	*SMARCA2*
EPHA2	*PTGS2*	*SNHG20*	*SLC34A2*
PAEP	*AKT3*	*ABCF2*	*BNC2*
THBS1	*SMARCAL1*	*LSP1*	*ARHGAP24*
HAVCR2	*SMARCB1*	*PSMC3IP*	*TGFB1*
HDAC9	*PARP4*	*IGHD2-15*	*MS*
FASLG	*ETV4*	*ZNF283*	*SMAD2*
HFM1	*RECQL4*	*ZNF350*	*PDPK1*
IMP3	*UBE2C*	*MIB1*	*CFLAR*
RNF168	*NTHL1*	*CXCR4*	*CDH13*
APP	*TWIST1*	*MEN1*	*MALAT1*
CTLA4	*BMPR1A*	*PRSS8*	*RET*
EBAG9	*ERCC6*	*WDR70*	*RNF8*
HOXA9	*RFWD3*	*CDC25C*	*CCND2*
BLM	*ABCC3*	*IL2*	*SYP*
IGF2	*SPATA2*	*BRD4*	*RASSF7*
NEAT1	*FGFR3*	*SOHLH2*	*CIB1*
TFF1	*FGFR1*	*BMI1*	*TGFA*
SPPL2A	*TOP1*	*TACSTD2*	*RIF1*
HOXA10	*APEX1*	*MIR100*	*SMAD3*
BUB1	*KLF6*	*PDGFRA*	*PLG*
TNFRSF10A	*GAS5*	*PDGFRB*	*FOXP1*
LIG3	*SLC7A11*	*GNAS*	*PDK1*
ADRM1	*OR51F1*	*OR51I1*	*PRSS3*
KHDRBS1	*E2F4*	*HBEGF*	*DEPDC1B*
GJA8	*MAP2K4*	*MAMSTR*	*CD53*
CDH17	*DCC*	*INS*	*CAT*
VWA5A	*PPP2R1B*	*STIL*	*CXCL11*
CTNNA1	*SFRP1*	*CASZ1*	*TNFRSF10D*
YAP1	*ETS1*	*LRIF1*	*IGF2BP2*
RPA1	*KMT2B*	*CYP1B1*	*PALLD*
EIF6	*MIR93*	*TG*	*HMCN1*
DNA2	*EGF*	*PRIMPOL*	*LAMA4*
RECQL5	*MAD2L1*	*KIF4A*	*PAXIP1*
HUS1	*ELAVL1*	*INSL6*	*USP11*
HSPA9	*CXCL1*	*LIN37*	*PTK2B*
EPHA1	*CSF1R*	*TP73*	*GRB7*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CDH3	*NUSAP1*	*HSPB2*	*USP1*
CD163	*DDB2*	*AMBP*	*TOP2A*
CALCR	*FAAP20*	*URI1*	*MIR4435-2*
KLK1	*MGMT*	*PPP2R2A*	*PARK2*
JUN	*FGF3*	*KIF11*	*TPX2*
MUL1	*TEK*	*TYMP*	*UCHL3*
PSEN1	*SETD2*	*ANXA4*	*ROS1*
ZNF645	*APAF1*	*RPA2*	*PAGR1*
ATAD2	*PRDX5*	*WNK2*	*DNMT1*
DAPK1	*MAPK1*	*HSPB1*	*FLT4*
CDKN2B-AS1	*ERCC5*	*THRA*	*CTCFL*
OLA1	*SPP1*	*PLA2G2D*	*ELF3*
MAP2K1	*MUC4*	*SERPINE1*	*VEGFC*
PCAT1	*ZRANB3*	*NUP62*	*SMARCD1*
IGF1R	*CDK7*	*KIF14*	*CREM*
MLANA	*CDKN2B*	*ELAC2*	*SPOP*
FAT4	*HSP90AA1*	*GALNT3*	*HSD17B1*
CTSD	*NR1D1*	*LATS2*	*CELSR2*
CSF2	*BBC3*	*FN1*	*IGF2BP1*
PTP4A1	*ABCA8*	*BIRC3*	*TGFBR2*
RINT1	*NOTCH2*	*ITGB4*	*DUSP1*
PDCD4	*CYP17A1*	*CDK13*	*CCNG1*
VHL	*FAS*	*AURKB*	*SLC2A1*
DCK	*BIRC2*	*EP300*	*IGFBP2*
TONSL	*BNC1*	*SQSTM1*	*CRP*
AXIN2	*TWIST2*	*SP1*	*CHD1*
HGF	*STMN1*	*MAGI2-AS3*	*SMYD3*
SRA1	*PIK3R2*	*NOTCH4*	*FGF13*
PIK3CB	*GSTT1*	*SCLC1*	*PSG1*
REV1	*IL1B*	*FOXP3*	*CXCL8*
EIF4EBP1	*LHCGR*	*CENPF*	*DAND5*
MAP3K3	*ANGPTL4*	*CHGA*	*CEP55*
CD34	*LDHA*	*RNMT*	*GPX4*
SOX4	*NRP1*	*MAGEA4*	*GSTM1*
CCNB2	*SMUG1*	*DDR2*	*TSG101*
ENO2	*ISG15*	*NEIL3*	*SPINK1*
CASP7	*CCNE2*	*UBE2N*	*KLK3*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
LRIG3	*ATRIP*	*CD276*	*USP9X*
KPNA2	*CDC20*	*POLN*	*PAK1*
DNMT3A	*ALKBH5*	*RPS6KB1*	*PSAT1*
LGALS3BP	*IGFBP3*	*TTC21A*	*MIR503HG*
EPHA5	*FAN1*	*KRT19*	*KITLG*
SRY	*NTRK1*	*HMGA1*	*ARPC1B*
PARP10	*GCG*	*CBX2*	*BDNF*
RICTOR	*MORF4L1*	*TRRAP*	*CCNA1*
SLU7	*SERPINB2*	*KIFC1*	*GATA6*
BMND7	*BMND8*	*TUBB2A*	*DROSHA*
SPARC	*FBXO5*	*MIR4435-2HG*	*PTTG1*
VAT1	*PKMYT1*	*FEN1*	*STAG3*
NBEAL1	*CFAP52*	*ZNF385B*	*UBE2T*
HDAC1	*PARP3*	*PRKAR1A*	*HLA-DRB1*
TTK	*HES1*	*LATS1*	*ZFPM2*
RBM10	*MIR107*	*GALT*	*USP15*
KDM6A	*USP44*	*MXI1*	*RAD21*
RBBP6	*WNT2B*	*ADAM17*	*PIK3R3*
TUBB	*TERC*	*SNHG14*	*HP*
RUNX3	*PTH*	*CHD1L*	*MAGED2*
LRIG1	*NSG1*	*AGR2*	*HPC1*
RBMS3	*PTPN13*	*FOXO1*	*TRIM28*
FIGF	*SMARCC1*	*WNT4*	*AREG*
SDHB	*SCD*	*ZNF365*	*UBE2D1*
SOX17	*CUL4A*	*FLT1*	*WTIP*
RAF1	*TGFBR1*	*MEG3*	*TCF4*
MCM7	*TAP1*	*STAT1*	*BRF2*
SNHG1	*CCDC170*	*EPHA3*	*ZMYND11*
ARNT	*ATG2B*	*HMGN4*	*QRICH2*
GGN	*PTBP1*	*ANAPC2*	*MIRLET7D*
MAP3K15	*TACC1*	*GGNBP2*	*NDC80*
ZC3H13	*TSC2*	*PIWIL1*	*KIF15*
TMEM158	*SCGB2A1*	*DECR1*	*FZD3*
CTTN	*MCM2*	*E2F6*	*APOE*
GTF2H1	*FASN*	*ANGPT2*	*HDAC2*
MYCL	*JAK2*	*E2F8*	*DSG2*
MND1	*MPP7*	*MAPK3*	*MUC6*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
MUTYH	*POLR1B*	*GATA4*	*MAP3K4*
PKN1	*DYNLL1*	*FGF2*	*CSTF1*
ZNF536	*ANXA2*	*RAC1*	*MIR26B*
RHOA	*TP63*	*MUC5AC*	*DUSP12*
PRSS55	*PDS5B*	*BIRC7*	*APOBEC3A*
FGFR4	*GDF15*	*KDM5A*	*BCHE*
MYBL2	*EME1*	*PEG3*	*ENG*
RBM15	*CCND3*	*SLC9A6*	*SPEN*
BMPR1B	*MAGEA5*	*RGCC*	*LINC-PINT*
CYP1A1	*FBXL7*	*NOS3*	*CREBBP*
MUC17	*CISH*	*RCBTB1*	*C1D*
ATG7	*TKTL1*	*FCGBP*	*CHD4*
MLF2	*TNFRSF1B*	*REV3L*	*APC2*
KAT2B	*EPHB4*	*BCOR*	*KIAA0101*
ABCC10	*MYCN*	*ARID2*	*ERCC8*
RELA	*DDB1*	*ANGPT1*	*TACC3*
ZNF395	*POU5F1B*	*UTP14A*	*DIO2*
PSMD14	*RAD9A*	*TNKS*	*XPC*
COL6A6	*HUWE1*	*NEK2*	*TELO2*
MBD2	*NCOA4*	*FRA3B*	*PMAIP1*
NF2	*SLC12A9*	*POLD3*	*STC2*
IDH2	*SCGB3A1*	*DNAH12*	*DNPH1*
TTC39B	*TRAF4*	*CD8A*	*MAP3K5*
RFC3	*ATM*	*TOP2B*	*SLC22A3*
FZD1	*MNX1*	*ELF1*	*AKR1C3*
RSPO1	*SLC25A23*	*SRD5A2*	*HSPA5*
CTBP1	*IFNG*	*KIFAP3*	*PTHLH*
LYPD1	*RTEL1*	*HN1*	*CXCL10*
METTL16	*CROT*	*CYP2C8*	*CAV1*
ABCA2	*NCOR1*	*CCL22*	*SCGB1D2*
ALCAM	*DDIT3*	*LAMA3*	*CTCF*
TEX14	*KLF5*	*RACGAP1*	*KAT6A*
B2M	*BRWD1*	*POF1B*	*EEPD1*
DNAH2	*LARP1*	*NCOA2*	*IGFBP1*
AXIN1	*AMPH*	*ECT2*	*ELF5*
DDX11	*KRT5*	*KIAA1524*	*ELOF1*
ATMIN	*MYO18B*	*DEGS1*	*CEACAM6*

Table 4.1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
NFKB1	*UBR5*	*CLU*	*ZNF385A*
OIP5	*PIK3CG*	*ALPPL2*	*PPP2R3A*
APOBEC3B	*EFCAB13*	*MIR369*	*EIF3E*
BAK1	*PRKAA1*	*MPO*	*ACVR1*
TNFAIP3	*LGR4*	*MIR146A*	*HMMR*
TDP1	*FOXN3*	*IDH1*	*RUNX1*
DCLRE1A	*TCF7*	*RAD17*	*IKBKB*
FHL2	*AGR3*	*RNASEH2B*	*DUSP5*
HDAC4	*MPV17*	*MAGEC1*	*FHL1*
TRIM31	*GATA5*	*GLI2*	*KRT18*
SERPINA1	*CXXC1*	*TXN*	*BDNF-AS*
THEG	*TSPAN8*	*FAT1*	*CDKN1C*
PRKAA2	*ID4*	*MAPK4*	*CLIP2*
LOXL3	*ALB*	*IL4*	*CA9*
ASPM	*PRR13*	*PRKAB1*	*RPA3*
HLTF	*DNAH8*	*PPIA*	*FZD10*
CADM1	*CNTLN*	*TTI1*	*MDN1*
COL23A1	*DACH2*	*C21orf33*	*PUF60*
MSH5	*DNMT3B*	*SET*	*CTNND1*
LAMB3	*ACVR1B*	*WNK1*	*HES6*
PGAM4	*FLG2*	*DPP9*	*POLB*
LINC00673	*RNF212*	*IMMP2L*	*COPS8*
PXN	*XPO1*	*SCN1A*	*EIF1AX*
PHGDH	*MAP2K6*	*CDKN2C*	*PRSS2*
NRG1	*NAV3*	*ZGLP1*	*LPL*
MIR205	*MIR17*	*TGFB2*	*CYP3A43*
GEN1	*VWF*	*RNASE1*	*SAT2*
GAS1	*KRT6C*	*HAS2*	*NELFE*
DCLRE1C	*WASF3*	*LRRC32*	*CRISP2*
KDM1A	*KLF4*	*DDX23*	*CENPO*
GIPC3	*WRN*	*COL3A1*	*TP53AIP1*
MCM9	*XPA*	*UHMK1*	*SUV39H1*
SETD1B	*KRT14*	*NCOA1*	*DYNC1I2*
DCAF15	*PPIC*	*APEX2*	*OGG1*
DNAH11	*CIZ1*	*ATG12*	*CD68*

Using Coremine Medical, a total of approximately 2000 genes were retrieved that showed association with ovarian cancer.

For readability and clarity, only the half genes are presented in the main body of this thesis (Table 4.1). The complete list of all retrieved genes is provided in Appendix A for reference.

4.2 Network Building using STRING

Using the STRING database (**v12.0**), a protein-protein interaction (PPI) network was created in order to investigate the functional connections between the identified genes in more detail.

The gene list was submitted to STRING, and interactions with a high confidence score ($\geq$0.7) were selected for analysis. Accessing the STRING database and entering a list of two thousand OVARIAN CANCER genes were both required steps in the course of the process. In this particular instance, the procedure of selecting organisms, namely Homo sapiens, was carried out.

Using STRING's built-in Markov Cluster (**MCL**) algorithm, the network was automatically clustered based on interaction strength and connectivity. The resulting network consisted of 1935 nodes and 80479 edges, and having PPI enrichment p-value <1.0e-16. The STRING-generated protein interaction network revealed statistically significant enrichment of interactions among the input genes (p-value $= 1.0 \times 10^{-16}$), strongly suggesting that the observed interactions are not due to random chance but reflect meaningful biological associations relevant to ovarian cancer.

Each cluster grouped proteins with similar functions, co-expression patterns, or shared pathway involvement. These clusters were further analyzed to identify biological themes such as DNA repair, signal transduction, transport, and transcription regulation, suggesting modular functional organization among the genes potentially involved in ovarian cancer pathogenesis (Figure 4.1).

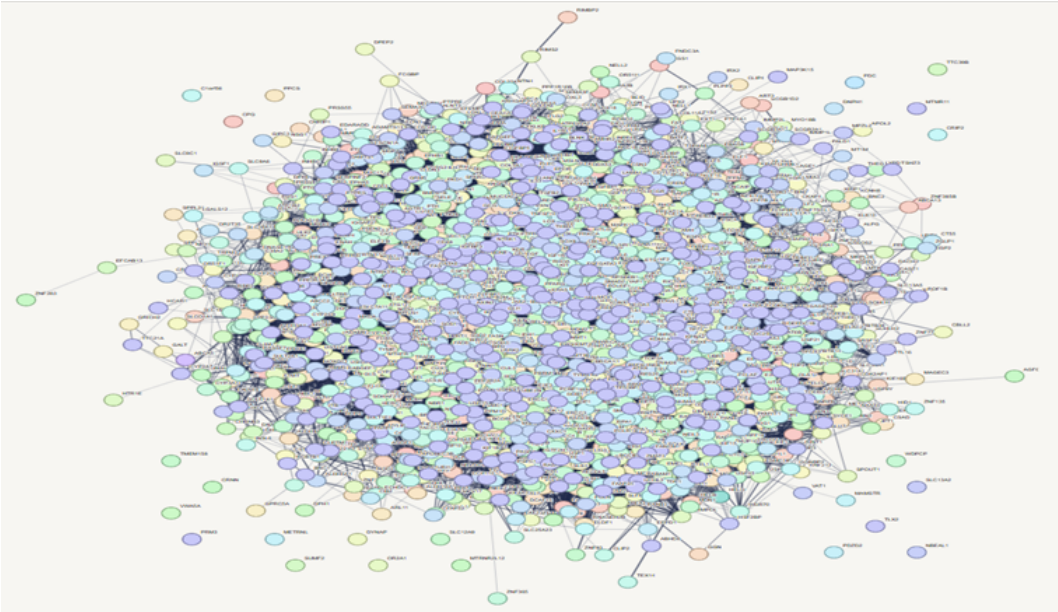


FIGURE 4.1: Network building of 2000 genes of Malignant Neoplasm of Ovary and Network Status

4.3 Cluster Generation

Subsequently, the process entailed the formation of clusters using STRING. To achieve this, selection of clusters was done and then proceeded to perform K-means clustering. Several different clustering techniques, including K-means, hierarchical clustering, and DBSCAN, are used to automatically find these groupings. The number of clusters was incrementally adjusted starting from 3, 4, 5, 6, and so on, and then implemented. Clusters ranging from 3 to 24 were verified based on the KEGG pathway (Figure 4.2). The clusters generated are shown (Table 4.2).

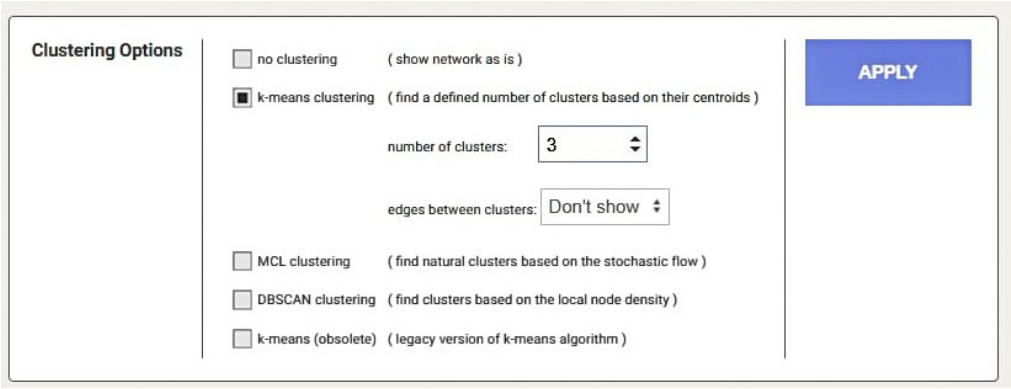


FIGURE 4.2: Executing K-means clustering.

TABLE 4.2: Cluster Structures and their PPI Enrichment p-value

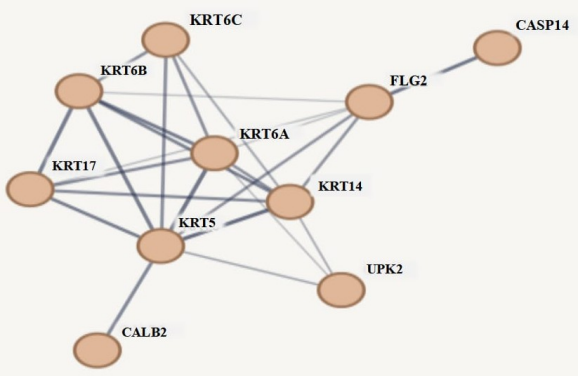
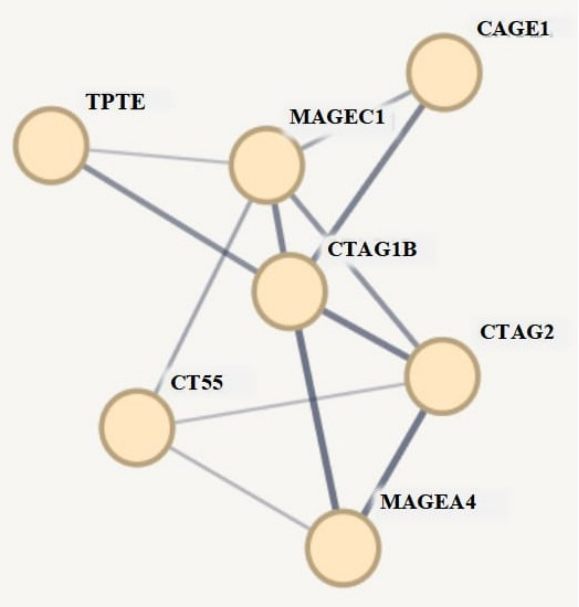
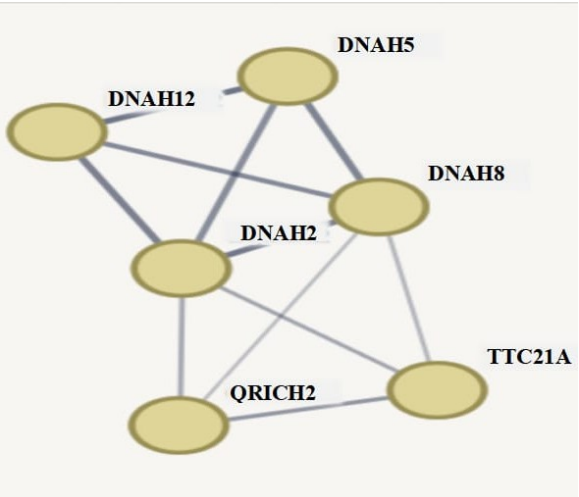
Sr. No.	Cluster No.	p-value	Cluster Structure
1	3	<1.0-16	
2	4	1.00E-16	
3	5	1.00E-16	

Table 4.2 continued from previous page

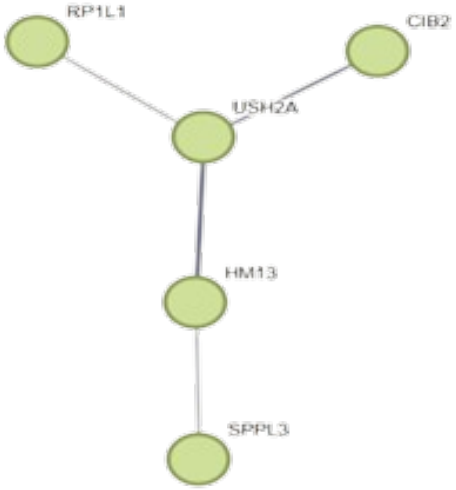
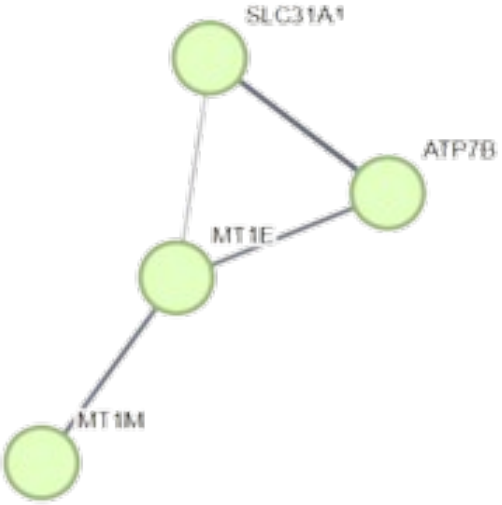
Sr. No.	Cluster No.	p-value	Cluster Structure
4	6	1.00E-16	
5	7	1.94E-08	
6	8	2.91E-10	

Table 4.2 continued from previous page

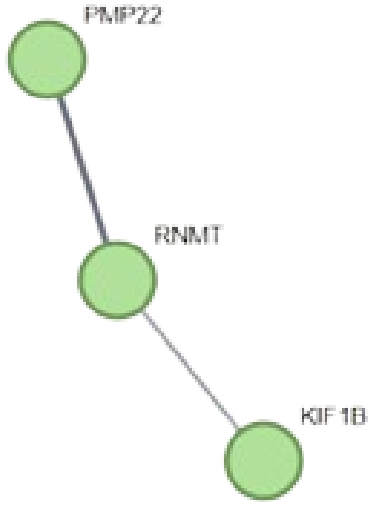
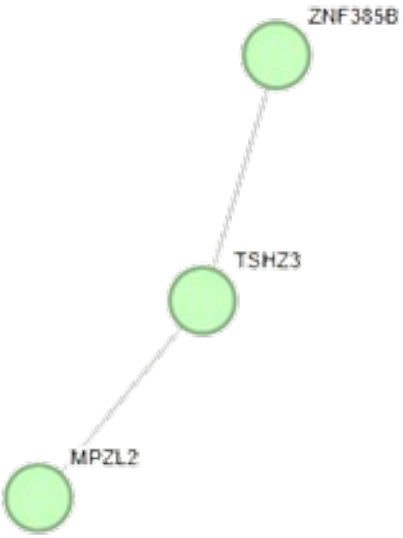
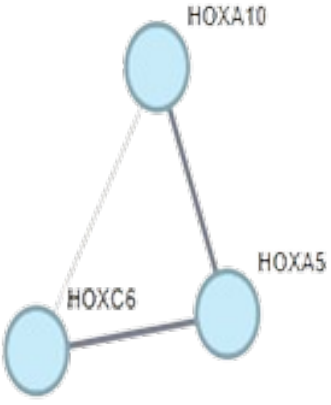
Sr. No.	Cluster No.	p-value	Cluster Structure
7	9	*7.80E-05*	
8	10	*6.84E-07*	
9	11	2.45E-08	

Table 4.2 continued from previous page

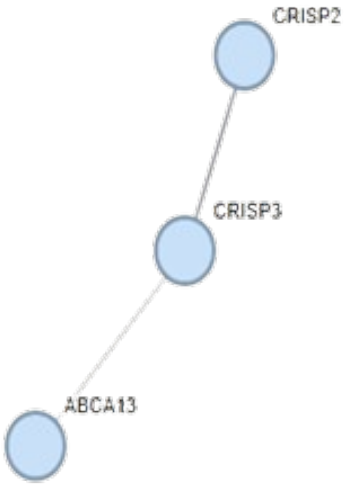
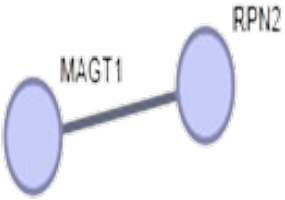
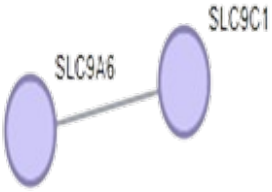
Sr. No.	Cluster No.	p-value	Cluster Structure
10	12	6.23E-06	
11	13	0.000383	
12	14	0.00519	
13	15	0.000725	
14	16	1	
15	17	1	

Table 4.2 continued from previous page

Sr. No.	Cluster No.	p-value	Cluster Structure
16	18	1	
17	19	1	
18	20	1	
19	21	1	
20	22	1	
21	23	1	
22	24	1	

K-means clustering of the STRING-derived interaction network for the 2000 genes

identified through Coremine produced a total of 24 clusters. Among these, the first 15 clusters contained multiple genes (ranging from 2 to several), indicating meaningful groupings based on functional and interaction similarities.

However, clusters 16 to 24 each contained only a single gene, reflecting weak or no detectable associations with other genes in the dataset.

While these singletons were included for manual verification in KEGG to assess their potential relevance, they were not interpreted as functional clusters due to the absence of gene–gene groupings.

4.4 Manual Verification by KEGG Database

KEGG pathway was utilized to verify clusters generated while working with STRING database. The KEGG pathway comprised 160 ovarian cancer pathways. In all 160 pathways, we verified clusters ranging from 3 to 24, as obtained from the STRING database.

We identified matching genes in 15 pathways. Out to total 65 cluster genes, 16 genes were matched in pathways and these are *KRT14, KRT17, TPTEP2, DNAH12, DNAH2, DNAH15, DNAH8, ABCB11, ABCC2, SLC22A1, ATP7B, HOXA10, MAGT1, RPN2, HTR1E, GALT* (Table 4.3). A subset of genes not currently mapped in the KEGG ovarian cancer pathway were identified.

These genes, although not established in canonical pathways, represent potential novel candidates for further investigation in ovarian cancer pathogenesis based on their expression and predicted interactions. These 49 genes i.e. *CALB2, CASP14, FLG2, KRT5, KRT6A, KRT6B, KRT6C, UPK2, CAGE1, CT55, CTAG18, CTAG2, MAGEA4, MAGEC1, QRICH2, TTC21A, ABCB4, ABCG2, SLCO4A1, CIB2, HM13, RP1L1, SPPL3, USH2A, MT1E, MT1M, SLC3A1, KIF1B, PMP22, RNMT, MPZL2, TS HZ3, ZNF3858, HOXA5, HOXC6, ABCA13, CRISP2, CRISP3, SCGB1D2, SCGB2A1, SLC9A6, SLC9C1, AGFC2, MAMSTR, CLIP4, PRSS55, LYPD1, ZNF365, ZNF283* were mismatched are novel candidate genes for ovarian cancer.

TABLE 4.3: Matching genes and their associated KEGG pathways in OC

Pathway ID	**Pathway Name**	**Cluster No.**	**Matching genes**
hsa01100	Metabolic pathway	Cluster no 14	*RPN2*
		Cluster no 21	*GALT*
hsa01524	Platinum drug resistance	Cluster no 6	*ABCC2*
		Cluster no 8	*ATP7B*
hsa04390	HIPPO signalling pathway	Cluster no 4	*TPTE*
hsa04915	Estrogen signalling pathway	Cluster no 3	*KRT14*
			KRT17
hsa05022	Neurodegeneration-multiple diseases	*Cluster no 5*	*DNAH12*
			DNAH2
			DNAH5
			,DNAH8
hsa01522	Endocrine resistance	Cluster no 6	*ABCB11*
hsa04917	Prolactin signalling pathway	Cluster no 21	*GALT*
hsa05202	Transcriptional misregulation in cancer	Cluster no 11	*HOXA10*
hsa05014	Amyotrophic lateral sclerosis	Cluster no 5	*DNAH2*
			DNAH5
			DNAH8
			DNAH12
hsa05231	Choline metabolism in cancer	Cluster no 6	*SLC22A1*
hsa04024	CAMP signalling pathway	Cluster no 17	*HTR1E*
hsa04141	Protein processing in endoplasmic reticulum	Cluster no 14	*MAGT1*
			RPN2
Hsa04726	Serotoneric synapse	Cluster no 17	*HTR1E*
Hsa05016	Huntington disease	Cluster no 5	*DNAH2*
			DNAH5
			DNAH8
			DNAH12
Hsa04082	Neuroactive ligan-receptor interaction	Cluster no 17	*HTR1E*

4.5 Functional Annotation of Novel Candidate Genes by DAVID Tool

DAVID functional annotation analysis of the 49 mismatch genes (those not mapped in known KEGG ovarian cancer pathways) yielded 8 significant annotation clusters, each representing groups of related biological functions, molecular features, or gene ontology terms. These clusters provide insights into the potential roles and common characteristics of the genes, despite their absence in classical ovarian cancer pathways. The enrichment of specific clusters suggests that some of these genes may share involvement in key biological processes such as transcription regulation, cell signaling, membrane transport, or protein binding. This clustering highlights the possibility that these genes could represent novel candidates or indirect contributors to ovarian cancer pathogenesis, warranting further investigation. The results are shown as follows:

4.5.1 Functional Annotation Clustering

4.5.1.1 Cluster 1 Annotation

Cluster 1 Annotation, with an enrichment score of 2.89, represents a functionally coherent group of genes involved in keratinization, intermediate filament formation, and structural organization of the cytoskeleton. This cluster includes 5 genes *CASP14, KRT5, KRT6C, KRT6B, KRT6A* contributing to major annotations such as "keratinization" (P-value = 2.2E-5), "*Keratin_2_head*" (P = 2.2E-5), and "*Keratin_II*" (P = 3.6E-5), indicating strong statistical enrichment for keratin-related pathways.

Several other annotations, including "keratin filament", "intermediate filament", and "*IF_rod_dom*", further emphasize the structural roles of these proteins in maintaining epithelial integrity. The presence of terms like "*Palmoplantar keratoderma*" under disease associations also links this gene cluster to disorders of epithelial differentiation.

Additionally, the presence of "cytosol" and "extracellular exosome" components suggests potential involvement in intracellular transport and communication.

Overall, the significant enrichment of this cluster suggests that these mismatch genes though not part of classical ovarian cancer pathways may play supportive or novel roles in tumor progression via cytoskeletal alterations and epithelial dynamics (Figure 4.3).

Annotation Cluster 1	Enrichment Score: 2.89		Count	P_Value	Benjamini
GOTERM_BP_DIRECT	keratinization	RT	5	2.2E-5	4.5E-3
INTERPRO	Keratin_2_head	RT	4	2.4E-5	2.2E-3
INTERPRO	Keratin_II	RT	4	3.0E-5	2.2E-3
UP_SEQ_FEATURE	SITE:Stutter	RT	4	5.6E-5	1.8E-2
GOTERM_CC_DIRECT	keratin filament	RT	5	5.7E-5	6.0E-3
GOTERM_MF_DIRECT	structural constituent of skin epidermis	RT	4	6.6E-5	4.6E-3
UP_SEQ_FEATURE	REGION:Coil 2	RT	4	3.4E-4	2.0E-2
UP_SEQ_FEATURE	REGION:Linker 12	RT	4	3.4E-4	2.0E-2
INTERPRO	IF_conserved	RT	4	3.6E-4	1.8E-2
UP_SEQ_FEATURE	REGION:Coil 1B	RT	4	4.3E-4	2.0E-2
UP_SEQ_FEATURE	REGION:Linker 1	RT	4	4.5E-4	2.0E-2
UP_SEQ_FEATURE	REGION:Coil 1A	RT	4	4.5E-4	2.0E-2
UP_SEQ_FEATURE	REGION:Head	RT	4	5.0E-4	2.0E-2
GOTERM_BP_DIRECT	intermediate filament organization	RT	4	5.2E-4	5.2E-2
UP_SEQ_FEATURE	REGION:Tail	RT	4	5.5E-4	2.0E-2
UP_SEQ_FEATURE	DOMAIN:IF rod	RT	4	5.7E-4	2.0E-2
UP_KW_CELLULAR_COMPONENT	Intermediate filament	RT	4	5.9E-4	1.1E-2
INTERPRO	IF_rod_dom	RT	4	6.2E-4	2.3E-2
SMART	Filament	RT	4	1.1E-3	3.1E-2
UP_KW_CELLULAR_COMPONENT	Keratin	RT	4	4.8E-3	4.4E-2
UP_KW_DISEASE	Palmoplantar keratoderma	RT	3	8.3E-3	1.5E-1
GOTERM_MF_DIRECT	structural constituent of cytoskeleton	RT	3	2.4E-2	3.1E-1
GOTERM_CC_DIRECT	intermediate filament	RT	3	3.4E-2	9.4E-1
UP_KW_DOMAIN	Coiled coil	RT	10	2.5E-1	1.0E0
GOTERM_CC_DIRECT	extracellular exosome	RT	7	3.4E-1	1.0E0
GOTERM_CC_DIRECT	cytosol	RT	9	9.4E-1	1.0E0
UP_KW_PTM	Acetylation	RT	5	9.5E-1	1.0E0

FIGURE 4.3: Functional Annotation clustering of OC (Cluster 1)

4.5.1.2 Cluster 2 Annotation

Cluster 2 Annotation, with an enrichment score of 1.59, includes the genes *RP1L1, CIB2,* and *USH2A*. This cluster is associated with functions such as photoreceptor cell maintenance (p = *5.0E-2*), the disease term Retinitis pigmentosa (p = *5.8E-2*), and the cellular component cell projection (p = *5.9E-2*). Although these genes are primarily known for their roles in retinal function and sensory signaling, their structural features and membrane localization suggest potential involvement in broader biological contexts. *USH2A* and *RP1L1* encode large extracellular matrix or scaffold proteins essential for cell integrity, while *CIB2* functions in calcium-binding and integrin-associated signaling. These roles may indirectly contribute to

cancer-related mechanisms such as cell adhesion, polarity, migration, or stromal interactions.Although not part of canonical ovarian cancer pathways, the functional clustering of these genes points to possible supportive roles in epithelial dynamics or tumor microenvironment interactions, warranting further investigation.

However, a limitation of this cluster is that its enrichment is based on tissue-specific terms (e.g., photoreceptors), which may reduce direct applicability unless further validated in ovarian contexts (Figure 4.4).

Annotation Cluster 2	Enrichment Score: 1.59	G		Count	P_Value	Benjamini
GOTERM_BP_DIRECT	photoreceptor cell maintenance	RT		3	5.0E-3	2.5E-1
UP_KW_DISEASE	Retinitis pigmentosa	RT		3	5.8E-2	3.5E-1
UP_KW_CELLULAR_COMPONENT	Cell projection	RT		7	5.9E-2	2.7E-1

FIGURE 4.4: Functional Annotation clustering of OC (Cluster 2)

4.5.1.3 Cluster 3 Annotation

Cluster 3 Annotation, with an enrichment score of 1.47, includes the genes *RP1L1, CIB2, QRICH2, SLC9C1,* and *CT55.* This cluster is enriched in terms related to cilium structure and cell projections, such as "*cilium*" (p = *8.4E-3*), "*cell projection*" (p = *5.9E-2*), and "*flagellum*" (p = *7.8E-1*).

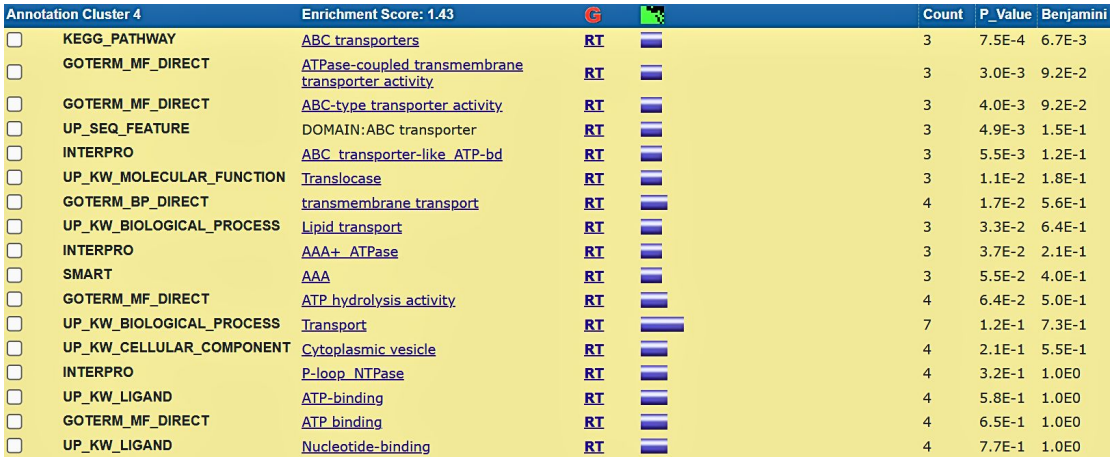
These annotations indicate a possible involvement of these genes in cellular movement, polarity, or sensory functions. While these roles are classically associated with ciliated or sensory tissues, emerging evidence suggests that cilia-related and projection-related genes can influence tumor progression, particularly through signaling pathways, microtubule dynamics, and intercellular communication. For example, *QRICH2* and *RP1L1* are known for roles in axoneme and microtubule stability, and *CIB2* may influence calcium-dependent pathways that affect cell adhesion. Although these genes are not mapped in canonical ovarian cancer pathways, their clustering around ciliary structures may reflect non-traditional mechanisms by which they contribute to tumor microenvironment modulation or metastatic potential (Figure 4.5).

Annotation Cluster 3	Enrichment Score: 1.47			Count	P_Value	Benjamini
UP_KW_CELLULAR_COMPONENT	Cilium	RT		5	8.4E-3	5.0E-2
UP_KW_CELLULAR_COMPONENT	Cell projection	RT		7	5.9E-2	2.7E-1
UP_KW_CELLULAR_COMPONENT	Flagellum	RT		3	7.8E-2	2.8E-1

FIGURE 4.5: Functional Annotation clustering of OC (Cluster 3)

4.5.1.4 Cluster 4 Annotation

Annotation Cluster 4, with an enrichment score of 1.43, includes the genes *ABCB4, ABCA13,* and *ABCG2*. This cluster is significantly enriched in annotations related to ATP-binding cassette (ABC) transporters, including "*ABC transporters*" ($p = 7.5E\text{-}4$), "*ATPase-coupled transmembrane transporter activity*" ($p = 3.0E\text{-}3$), and "*ABC-type transporter activity*" ($p = 4.0E\text{-}3$). These genes encode membrane-bound proteins that use ATP hydrolysis to transport various molecules across cellular membranes, including lipids, xenobiotics, and metabolic products. ABCG2, in particular, is well-known for its role in multidrug resistance in cancer, contributing to chemotherapy efflux and resistance mechanisms. The presence of domains related to "*ATP binding*," "*AAA+ ATPase*," and "*translocase*" suggests functional involvement in energy-dependent transport and signaling regulation. Though not part of canonical ovarian cancer pathways, the clustering of these genes highlights their potential roles in tumor drug resistance, cellular detoxification, and metabolic adaptation which are crucial features of cancer progression and treatment response (Figure 4.6).



Annotation Cluster 4	Enrichment Score: 1.43			Count	P_Value	Benjamini
KEGG_PATHWAY	ABC transporters	RT		3	7.5E-4	6.7E-3
GOTERM_MF_DIRECT	ATPase-coupled transmembrane transporter activity	RT		3	3.0E-3	9.2E-2
GOTERM_MF_DIRECT	ABC-type transporter activity	RT		3	4.0E-3	9.2E-2
UP_SEQ_FEATURE	DOMAIN:ABC transporter	RT		3	4.9E-3	1.5E-1
INTERPRO	ABC_transporter-like_ATP-bd	RT		3	5.5E-3	1.2E-1
UP_KW_MOLECULAR_FUNCTION	Translocase	RT		3	1.1E-2	1.8E-1
GOTERM_BP_DIRECT	transmembrane transport	RT		4	1.7E-2	5.6E-1
UP_KW_BIOLOGICAL_PROCESS	Lipid transport	RT		3	3.3E-2	6.4E-1
INTERPRO	AAA+_ATPase	RT		3	3.7E-2	2.1E-1
SMART	AAA	RT		3	5.5E-2	4.0E-1
GOTERM_MF_DIRECT	ATP hydrolysis activity	RT		4	6.4E-2	5.0E-1
UP_KW_BIOLOGICAL_PROCESS	Transport	RT		7	1.2E-1	7.3E-1
UP_KW_CELLULAR_COMPONENT	Cytoplasmic vesicle	RT		4	2.1E-1	5.5E-1
INTERPRO	P-loop_NTPase	RT		4	3.2E-1	1.0E0
UP_KW_LIGAND	ATP-binding	RT		4	5.8E-1	1.0E0
GOTERM_MF_DIRECT	ATP binding	RT		4	6.5E-1	1.0E0
UP_KW_LIGAND	Nucleotide-binding	RT		4	7.7E-1	1.0E0

FIGURE 4.6: Functional Annotation clustering of OC (Cluster 4)

4.5.1.5 Cluster 5 Annotation

Cluster 5 Annotation, with an enrichment score of 0.78, includes the genes *CALB2, FLG2, and CIB2*, all of which are associated with calcium-binding and metal-binding functionalities. The cluster is enriched in annotations such as "*EF-hand 1 calcium-binding site*" (p = *5.8E-2*), "*EF-hand domain 1*" (p = *6.1E-2*), and "*EF-hand domain 2*" (p = *6.2E-2*), which suggest that these genes share structural domains capable of binding calcium ions. Calcium signaling plays a central role in cell proliferation, migration, apoptosis, and metastasis, all of which are critical in cancer development. *CALB2* (Calretinin) is a well-characterized calcium-binding protein involved in cell cycle regulation and apoptosis, while *CIB2* also functions in calcium-dependent integrin signaling. Although *FLG2* is more commonly associated with epithelial barrier function, its inclusion here may reflect interactions with calcium-mediated signaling networks. While the p-values are modest and the overall enrichment is lower compared to other clusters, this grouping suggests that dysregulation of calcium signaling or buffering could contribute to ovarian cancer pathophysiology in ways not yet fully understood (Figure 4.7).

Annotation Cluster 5	Enrichment Score: 0.78	G		Count	P_Value	Benjamini
INTERPRO	EF_Hand_1_Ca_BS	RT		3	5.8E-2	3.1E-1
UP_SEQ_FEATURE	DOMAIN:EF-hand 1	RT		3	6.1E-2	9.4E-1
UP_SEQ_FEATURE	DOMAIN:EF-hand 2	RT		3	6.2E-2	9.4E-1
INTERPRO	EF_hand_dom	RT		3	9.0E-2	4.1E-1
INTERPRO	EF-hand-dom_pair	RT		3	1.2E-1	5.1E-1
GOTERM_MF_DIRECT	calcium ion binding	RT		3	4.7E-1	1.0E0
UP_KW_LIGAND	Calcium	RT		3	6.1E-1	1.0E0
UP_KW_LIGAND	Metal-binding	RT		8	8.4E-1	1.0E0

FIGURE 4.7: Functional Annotation clustering of OC (Cluster 5)

4.5.1.6 Cluster 6 Annotation

Cluster 6 Annotation, with an enrichment score of 0.61, comprises the genes *UPK2, ABCB4, SLC3A1, USH2A,* and *ABCG2*. This cluster is enriched in structural and localization features primarily associated with membrane-bound and transmembrane proteins, including terms like "*apical plasma membrane*" (p = *1.1E-2*), "*transport*" (p = *1.2E-1*), and "*carbohydrate: N-linked asparagine*" (p = *1.3E-1*). Additional annotations such as "*transmembrane helix*," "*glycoprotein*," "*extracellular membrane*," and "*disulfide bond*" indicate that these genes encode proteins

localized at or across the plasma membrane, with roles in cell signaling, transport, and intercellular communication. *ABCB4* and *ABCG2* are known ABC transporters, with ABCG2 also associated with drug efflux and chemoresistance in cancer. *SLC3A1* encodes a solute carrier protein involved in amino acid transport, and *UPK2* contributes to urothelial cell membrane structure, while USH2A is a large extracellular protein with diverse domain architecture. Though not traditionally associated with ovarian cancer, the enrichment of membrane-related features in this cluster suggests these genes may contribute to tumor cell-environment interactions, drug resistance, or nutrient transport, all of which are important aspects of cancer biology (Figure 4.8).

Annotation Cluster 6	Enrichment Score: 0.61	G		Count	P_Value	Benjamini
GOTERM_CC_DIRECT	apical plasma membrane	RT		5	1.1E-2	5.9E-1
UP_KW_BIOLOGICAL_PROCESS	Transport	RT		7	1.2E-1	7.3E-1
UP_SEQ_FEATURE	CARBOHYD:N-linked (GlcNAc...) asparagine	RT		14	1.3E-1	1.0E0
UP_KW_CELLULAR_COMPONENT	Cell membrane	RT		13	1.4E-1	4.2E-1
UP_KW_PTM	Glycoprotein	RT		14	1.5E-1	1.0E0
GOTERM_CC_DIRECT	plasma membrane	RT		16	1.8E-1	1.0E0
UP_SEQ_FEATURE	TRANSMEM:Helical	RT		15	2.5E-1	1.0E0
UP_SEQ_FEATURE	TOPO_DOM:Cytoplasmic	RT		11	3.2E-1	1.0E0
UP_KW_CELLULAR_COMPONENT	Membrane	RT		20	3.8E-1	8.6E-1
UP_SEQ_FEATURE	TOPO_DOM:Extracellular	RT		8	4.6E-1	1.0E0
GOTERM_CC_DIRECT	membrane	RT		13	4.8E-1	1.0E0
UP_KW_PTM	Disulfide bond	RT		9	5.7E-1	1.0E0
UP_KW_DOMAIN	Signal	RT		12	5.9E-1	1.0E0
UP_KW_DOMAIN	Transmembrane helix	RT		15	6.7E-1	1.0E0
UP_KW_DOMAIN	Transmembrane	RT		15	6.9E-1	1.0E0

FIGURE 4.8: Functional Annotation clustering of OC (Cluster 6)

4.5.1.7 Cluster 7 Annotation

Cluster 7 Annotation, with an enrichment score of 0.47, includes the genes *TSHZ3, SLC9C1, HOXA5, CT55,* and *HOXC6.* This cluster is enriched in terms related to transcriptional regulation and developmental processes, including "*developmental protein*" ($p = 6.5E\text{-}2$), "*homeobox*" ($p = 1.4E\text{-}1$), and "*DNA-binding transcription factor activity*". The presence of *HOXA5* and i, both members of the homeobox (HOX) gene family, suggests a strong link to embryonic patterning, differentiation, and tumorigenesis. HOX genes are often dysregulated in various cancers, including ovarian cancer, where they can influence cell fate, proliferation, and

invasion. *TSHZ3* is a zinc finger transcription factor also implicated in developmental regulation, while *CT55* is a cancer-testis antigen with roles still being explored in oncogenesis. Although the p-values are modest, the clustering of these genes suggests potential roles in epigenetic modulation, chromatin interaction, and transcriptional reprogramming within the tumor microenvironment, particularly in driving differentiation or metastatic potential in ovarian cancer (Figure 4.9).

Annotation Cluster 7	Enrichment Score: 0.47	G		Count	P_Value	Benjamini
UP_KW_MOLECULAR_FUNCTION	Developmental protein	RT		5	6.5E-2	2.2E-1
UP_KW_MOLECULAR_FUNCTION	Developmental protein	RT		5	6.5E-2	2.2E-1
INTERPRO	HD	RT		3	1.1E-1	4.7E-1
UP_KW_DOMAIN	Homeobox	RT		3	1.4E-1	1.0E0
SMART	HOX	RT		3	1.5E-1	7.2E-1
GOTERM_MF_DIRECT	DNA-binding transcription factor activity, RNA polymerase II-specific	RT		4	5.1E-1	1.0E0
GOTERM_BP_DIRECT	regulation of transcription by RNA polymerase II	RT		5	5.3E-1	1.0E0
UP_KW_BIOLOGICAL_PROCESS	Transcription regulation	RT		5	5.9E-1	1.0E0
UP_KW_BIOLOGICAL_PROCESS	Transcription	RT		5	6.2E-1	1.0E0
UP_KW_MOLECULAR_FUNCTION	DNA-binding	RT		4	6.4E-1	1.0E0
GOTERM_CC_DIRECT	chromatin	RT		3	7.0E-1	1.0E0
GOTERM_MF_DIRECT	RNA polymerase II cis-regulatory region sequence-specific DNA binding	RT		3	7.2E-1	1.0E0
GOTERM_MF_DIRECT	DNA binding	RT		3	8.4E-1	1.0E0
UP_KW_CELLULAR_COMPONENT	Nucleus	RT		8	9.8E-1	1.0E0

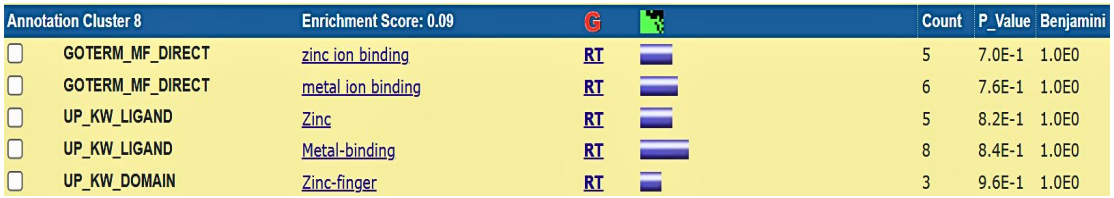
FIGURE 4.9: Functional Annotation clustering of OC (Cluster 7)

4.5.1.8 Cluster 8 Annotation

Cluster 8 Annotation, with an enrichment score of 0.09, includes the genes *TSHZ3, ZNF283, MT1M, ZNF365,* and *MT1E*. This cluster is weakly enriched in general metal ion binding functions, particularly "*zinc ion binding*" (p = *7.0E-1*), "*metal ion binding*" (p = *7.6E-1*), and "*zinc finger*" domains (p = *9.6E-1*). Genes such as *ZNF283* and *ZNF365* are zinc finger transcription factors, which are commonly involved in DNA binding, transcriptional regulation, and chromatin structure remodeling.

MT1M and *MT1E* belong to the metallothionein family, which plays roles in metal homeostasis, oxidative stress response, and potentially in tumor resistance to metal-based drugs. *TSHZ3*, also present in Cluster 7, is included again here due to overlapping zinc-related functional domains. While this cluster shows low statistical enrichment, the shared structural motifs among these genes suggest a

possible contribution to transcriptional regulation, redox balance, or metal-related cellular stress responses, all of which may be subtly involved in ovarian cancer pathogenesis or resistance mechanisms (Figure 4.10).



Annotation Cluster 8	Enrichment Score: 0.09			Count	P_Value	Benjamini
GOTERM_MF_DIRECT	zinc ion binding	RT		5	7.0E-1	1.0E0
GOTERM_MF_DIRECT	metal ion binding	RT		6	7.6E-1	1.0E0
UP_KW_LIGAND	Zinc	RT		5	8.2E-1	1.0E0
UP_KW_LIGAND	Metal-binding	RT		8	8.4E-1	1.0E0
UP_KW_DOMAIN	Zinc-finger	RT		3	9.6E-1	1.0E0

FIGURE 4.10: Functional Annotation clustering of OC (Cluster 8)

TABLE 4.4: DAVID Annotation Clusters, Genes and Their Significance in OC

Cluster No.	**Enrichment Score**	**Genes**	**Significance in Ovarian Cancer Pathogenesis**
1	2.89	*KRT5, KRT6A, KRT6B, KRT6C, CASP14*	Structural integrity, keratinization, possible involvement in epithelial-mesenchymal transition (EMT) and tumor invasion.
2	1.59	*RP1L1, CIB2, USH2A*	Membrane and photoreceptor-related roles; may influence cell polarity, microtubule organization, or stromal interactions.
3	1.47	*RP1L1, CIB2, QRICH2, SLC9C1, CT55*	Cilium and cell projection-related processes; potential roles in cellular communication and polarity regulation.
4	1.43	*ABCB4, ABCA13, ABCG2*	ABC transporters associated with drug efflux and chemoresistance mechanisms in tumor cells.
5	0.78	*CALB2, FLG2, CIB2*	Calcium ion binding proteins; may influence signal transduction and apoptosis.
6	0.61	*UPK2, ABCB4, SLC3A1, USH2A, ABCG2*	Membrane-localized proteins involved in transport, signaling, and potential drug resistance.
7	0.47	*TSHZ3, SLC9C1, HOXA5, CT55, HOXC6*	Transcription factors and developmental regulators; may impact differentiation and tumor progression.

Table 4.4 continued from previous page

Cluster No.	Enrichment Score	Genes	Significance in Ovarian Cancer Pathogenesis
8	0.09	*TSHZ3, ZNF283, MT1M, ZNF365, MT1E*	Zinc finger and metal-binding proteins; roles in transcription regulation and oxidative stress response.

Functional annotation clustering of 49 mismatch genes using the DAVID tool yielded 8 annotation clusters, highlighting a range of biological themes that may be relevant to ovarian cancer pathogenesis despite these genes not being mapped in KEGG ovarian cancer pathways. Cluster 1 revealed strong enrichment in keratinization and intermediate filament-related genes (*KRT5, KRT6A-C, CASP14*), indicating possible roles in epithelial structure and tumor progression. Clusters 2 and 3 included genes associated with photoreceptor function, cilium organization, and cellular projections, suggesting involvement in membrane signaling or cell polarity. Cluster 4 identified ABC transporter genes (*ABCB4, ABCA13, ABCG2*), with potential relevance to drug resistance. Clusters 5 and 6 involved calcium-binding proteins and membrane-associated transporters, respectively, implicating roles in signal transduction and cellular communication. Cluster 7 highlighted transcription factors and developmental regulators, including *HOXA5* and *HOXC6*, pointing to transcriptional control mechanisms in cancer. Finally, Cluster 8 grouped zinc finger and metallothionein genes, suggesting functions related to metal ion homeostasis and gene regulation. Together, these findings suggest that the mismatch genes, although not part of canonical ovarian cancer pathways, may have functional significance in tumor biology and represent potential targets for further study.

4.6 Cluster Validation through Protein-Protein Interactions

STRING analysis revealed that among the 49 mismatch genes tested, only *KRT5* demonstrated a high-confidence interaction (combined score = 0.737) with the

tumour suppressor protein *TP53*. This interaction was primarily supported by co-mentioning in literature and experimental evidence, suggesting a potential functional link between cytoskeletal remodeling and tumor suppression pathways in ovarian cancer.

In addition, three other genes *KRT6B, ABCG2, TPTE*, and *MT1E* showed medium-confidence interactions with key ovarian cancer proteins, with combined scores ranging from 0.5 to 0.69. The majority of mismatch genes did not exhibit significant interaction scores, indicating that they may either play indirect roles or represent novel elements yet to be functionally characterized. Only the high-confidence interaction was visualized and included in the results as a STRING network figure, while medium-confidence interactions were summarized in a tabulated format. The analysis revealed several notable interactions.

Overall, STRING-based analysis provided insight into the potential involvement of certain mismatch proteins in ovarian cancer pathogenesis through their direct or indirect association with established tumor suppressors and oncogenes.

4.6.1 KRT5-TP53 High-Confidence Interaction(s)

KRT5: Keratin, type II cytoskeletal 5, a structural protein, typically involved in the intermediate filament network of epithelial cells and TP53: The primary regulator of cell cycle arrest is the tumor suppressor p53, DNA repair, and apoptosis, frequently mutated in cancers including ovarian cancer. The interaction has a high combined confidence score of 0.737, supported primarily by experimental data (score: 0.165) and co-mentioning in PubMed abstracts (score: 0.695).

While no curated database or co-expression evidence was found. This suggests that while the interaction may not be fully experimentally validated, it is predicted to be functionally relevant, possibly through shared pathways like epithelial stress responses, apoptosis, or tumor suppression mechanisms.it also showed association with othher cancer protein like BRCA1, PIK3CA, BRCA2 and PTEN but have medium scores (Figure 4.11).

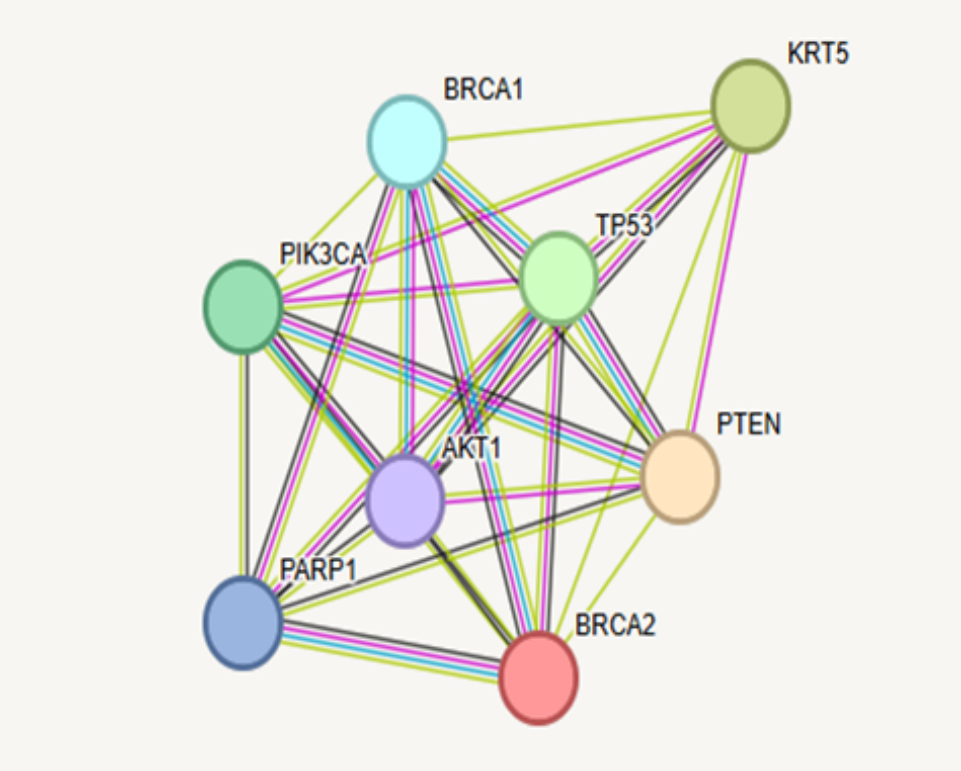


FIGURE 4.11: STRING interaction evidence between the KRT5 and TP53 proteins.

4.6.1.1 Medium-Confidence Interactions (Score 0.5–0.69)

Additional medium-confidence interactions (score 0.5–0.69) identified between mismatch genes and ovarian cancer proteins.Although these interactions lacked strong experimental or curated database support, their presence suggests possible indirect roles or emerging relevance in ovarian cancer pathways. These interactions are summarized in Table 4.5.

TABLE 4.5: Summarizes Medium-confidence protein-protein interactions (score 0.5-0.69) identified by STRING analysis between mismatch genes and key ovarian cancer proteins.

Mismatch Gene (Protein A)	**OC Gene (Protein B)**	**STRING Score**	**Confidence Level**
KRT6B	*PIK3CA*	0.617	Medium
ABCG2	*TP53*	0.615	Medium
ABCG2	*AKT1*	0.619	Medium
MT1E	*TP53*	0.679	Medium
KRT5	*BRCA1*	0.587	Medium
KRT5	*PTEN*	O.513	Medium

Table 4.5 continued from previous page

Mismatch Gene (Protein A)	**OC Gene (Protein B)**	**STRING Score**	**Confidence Level**
TPTE	*PTEN*	0.506	Medium
ABCG2	*PTEN*	0.514	Medium
KRT6B	*TP53*	0.519	Medium

Overall, the STRING-based interaction analysis highlighted a limited number of strong associations between mismatch genes and known ovarian cancer proteins. The high-confidence interaction observed between *KRT5* and *TP53* suggests a potential structural or regulatory link, while medium-confidence interactions point to additional genes that may be functionally relevant but require further validation. These findings contribute to understanding the extended molecular landscape potentially involved in ovarian cancer pathogenesis.

Chapter 5

Discussion

Ovarian cancer (OC) is a complex and multifactorial disease characterized by late diagnosis, high mortality, and diverse molecular mechanisms. The deadliest gynaecological cancer is ovarian cancer. Worldwide, ovarian cancer (OC) ranks eighth in cancer-related deaths among women and seventh in malignant tumours. Still, it's the sixth most common cancer killer of women, behind pancreatic, colon, breast, and lung cancers. The researchers make use of bioinformatics approaches in order to investigate the changes in the progression of ovarian cancer that are brought about by mutations, changes in gene expression, and alterations in signaling networks.

In the past few years, several researchers have applied bioinformatics strategies to uncover novel genes and pathways involved in ovarian cancer, supporting the significance of the present study. For example, Chao Dong, Xiaoqian Tian, and Fang He (2021) identified key molecular markers such as *CCNB1* and *CDK1* through integrated analysis of GEO datasets, highlighting their prognostic value in epithelial ovarian cancer. More recently, Feng Zhan, Yihong Guo, and Lihua He (2024) developed a programmed cell death–related gene signature to predict prognosis in serous ovarian cancer, demonstrating the power of functional annotation in revealing novel biomarkers. Similarly, Jie Zhang, Ming Li, and Wei Sun (2024) applied network-based methods to unravel potential driver genes as therapeutic targets, emphasizing the importance of hub gene analysis in ovarian cancer

research. While these studies have provided valuable insights, they largely concentrated on either differentially expressed genes or pathway-enriched markers. In contrast, the present work specifically addresses the gap of "mismatch genes" those retrieved from literature mining but absent from KEGG ovarian cancer pathways thereby extending current knowledge and offering a complementary perspective to existing bioinformatics-driven discoveries.

Although numerous genes such as *BRCA1*, *TP53*, and *PTEN* have been extensively studied in the context of OC pathogenesis, many potentially relevant genes remain uncharacterized in curated pathway databases like KEGG. The absence of these genes from standard cancer pathways raises critical questions: Are these "mismatched" genes functionally relevant to ovarian cancer?. Do they participate in regulatory or interaction networks that influence tumor biology?.

This study was designed to bridge this knowledge gap through an *in silico* approach, focusing on the functional annotation and interaction analysis of genes not currently mapped in ovarian cancer pathways. By integrating multiple bioinformatics tools including Coremine Medical, STRING, KEGG, DAVID, the research aimed to identify novel or overlooked genes that may play indirect or supporting roles in ovarian tumor development, progression, or treatment response. The following discussion interprets the major findings, compares them with existing literature, and highlights their potential biological relevance and implications.

Using sophisticated algorithms, Coremine Medical correlates user enquiries with gene profiles and returns a prioritised list of relevant genes. The Gene/Protein database was searched for the main topic "Ovarian Cancer" for the purpose of retrieving genes related to this cancer. A collection of texts that were published during the last five years i.e. 2019 to 2024 was gathered. The use of Coremine Medical in this study played a pivotal role in the initial identification and prioritization of ovarian cancer-associated genes. Thus, this step not only improves the biological relevance and accuracy of the study but also enhances the efficiency of data curation by focusing on well-supported genetic factors implicated in ovarian cancer pathogenesis.

To investigate the potential functional associations among the selected mismatch genes, a protein-protein interaction (PPI) network was constructed using the **STRING database**. Following the generation of the STRING interaction network and its clustering via K-means.

Clusters showing high internal connectivity and involvement in biological processes relevant to ovarian cancer such as epithelial structure maintenance, drug resistance, transcriptional regulation, and DNA repair were prioritized for further study. The number of clusters was incrementally adjusted starting from 3, 4, 5, 6, and so on, and then implemented. Each selected cluster was visualized using STRING's built-in layout options.

The presence of single-gene clusters observed beyond the 15th cluster likely reflects the limitations of k-means clustering when the number of specified clusters exceeds the true number of biologically meaningful groups. These singleton clusters represent genes with weak or no clear connections to others in the network, suggesting they may function independently or remain uncharacterized in the context of ovarian cancer pathways. While they were verified manually in KEGG for completeness, their lack of clustering indicates that they do not contribute to the collective functional modules identified in this study.

To validate the involvement of selected genes in ovarian cancer-related biological pathways, manual pathway verification of clusters was performed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. The KEGG pathway comprised 160 ovarian cancer pathways. In all 160 pathways, we verified clusters ranging from 3 to 24, as obtained from the STRING database. We identified matching genes in 15 pathways. Out to total 65 cluster genes, 16 genes were matched in pathways and these are *KRT14, KRT17, TPTEP2, DNAH12, DNAH2, DNAH15, DNAH8, ABCB11, ABCC2, SLC22A1, ATP7B, HOXA10, MAGT1, RPN2, HTR1E, GALT*. In contrast, 49 genes that did not appear were marked as mismatched or novel candidate genes and considered for further analysis. These mismatched genes, while not listed in the KEGG ovarian cancer map, may still hold functional relevance and represent novel or less-characterized candidates potentially involved in disease mechanisms.

DAVID functional annotation analysis of the 49 mismatch genes (those not mapped in known KEGG ovarian cancer pathways) yielded 8 significant annotation clusters. These clusters provide insights into the potential roles and common characteristics of the genes, despite their absence in classical ovarian cancer pathways. Clusters 1, 4 and 7 with higher enrichment scores and cancer-relevant functions, such as those involving keratin family proteins, ABC transporters, and transcription factors, were emphasized in the discussion. This clustering highlights the possibility that these genes could represent novel candidates or indirect contributors to ovarian cancer pathogenesis, warranting further investigation. In the final phase of the study, specific proteins from Clusters 1 and 4 were identified by examining their interactions with other proteins. The *KRT5* gene showed a high-confidence interaction (score = 0.737) with *TP53*, a key tumor suppressor protein in OC. This suggests that *KRT5*, typically involved in epithelial cell structure, may contribute to tumour suppression or stress responses, potentially through cytoskeletal regulation of cell cycle checkpoints. Other genes such as *ABCG2,* *KRT6B,* and *TPTE* exhibited medium-confidence interactions, indicating possible indirect roles in processes like drug resistance, transcriptional regulation, and cancer cell differentiation. The presence of functional clusters related to keratinization, ABC transporters, and transcription factors further supports the biological relevance of these genes in OC pathogenesis.

The findings of this study complement previous research that highlights the involvement of cytoskeletal proteins, transporters, and transcription factors in ovarian cancer biology. For instance, *KRT5*, although not commonly studied in ovarian cancer, has been reported in literature as overexpressed in various epithelial tumors, suggesting a role in structural remodeling and cellular stress adaptation which may align with its predicted interaction with *TP53* in this study. Similarly, *ABCG2* is a well-documented drug transporter often implicated in multidrug resistance; its moderate association with TP53 and *PIK3CA* in STRING supports its possible indirect role in treatment resistance mechanisms. *KRT6B*, a member of the keratin family, has been linked to epithelial differentiation and is often overexpressed in squamous cell carcinomas. Its predicted interaction with key OC proteins suggests a possible involvement in maintaining epithelial Although

TPTE is less studied, its homology to *PTEN* and predicted interactions in this study suggest it may be involved in phosphoinositide signaling and cell survival regulation. These findings provide functional clues about genes not currently included in KEGG OC pathways and point to their possible relevance in the broader molecular landscape of ovarian cancer.integrity or modulating cellular stress responses during tumor development.

While only a few genes showed strong STRING interactions, this may be due to limited experimental data or underrepresentation in cancer-specific pathway databases. The study's *in silico* nature also restricts experimental confirmation.Identifying unannotated but potentially relevant genes could expand the known molecular framework of ovarian cancer and guide future biomarker discovery or therapeutic target validation.

Chapter 6

Conclusion, Limitations and Future Directions

6.1 Conclusion

This study aimed to explore genes potentially involved in ovarian cancer pathogenesis through *in silico* analysis and functional annotation. The comprehensive investigation of genes associated with ovarian cancer through the application of a variety of bioinformatics techniques has resulted in the acquisition of significant knowledge regarding the molecular underpinnings of the disorder. Initially, the COREMINE tool was utilized in order to generate a substantial list of two thousand genes that are associated with Ovarian Cancer. The development of an interaction network consisting of 1935 nodes was made possible by additional investigation that made use of the STRING database. This network served as the basis for a more extensive functional study.

The identification of gene clusters that are related with ovarian cancer was accomplished through the use of K-means clustering. Based on the findings of this investigation, significant biological pathways and processes that may have a role in the development of this disease were identified. These clusters were confirmed by the use of the KEGG pathway analysis, which found a total of 160 pathways. Although there were no pathways directly related to OC. Out of 65 genes, 16

genes were matched and 49 genes were mismatched and considered as novel candidate genes. DAVID functional annotation analysis of the 49 mismatch genes (those not mapped in known KEGG ovarian cancer pathways) yielded 8 significant annotation clusters. Clusters 1, 4 and 7 with higher enrichment scores and cancer-relevant functions, such as those involving keratin family proteins, ABC transporters, and transcription factors, were emphasized. In the final phase of the study, specific proteins from Clusters 1 and 4 were identified by examining their interactions with other proteins. The *KRT5* showed a high-confidence interaction (score = 0.737) with *TP53*, a key tumor suppressor protein in OC. This suggests that *KRT5*, typically involved in epithelial cell structure, may contribute to tumour suppression or stress responses, potentially through cytoskeletal regulation of cell cycle checkpoints. Other genes such as *ABCG2*, *KRT6B*, and *TPTE* exhibited medium-confidence interactions, indicating possible indirect roles in processes like drug resistance, transcriptional regulation, and cancer cell differentiation. These findings expand the current understanding of ovarian cancer gene networks and propose several genes, not previously mapped in KEGG OC pathways, as potential contributors to disease biology. In conclusion, this research highlights the value of *in silico* approaches in uncovering potentially important, yet understudied, genes in ovarian cancer. Through functional clustering and interaction network analysis, several candidates were identified that may serve as future biomarkers or therapeutic targets pending experimental validation.

6.2 Limitations

This study has certain limitations inherent to its design and methodology. First, the use of k-means clustering with a predefined number of clusters (k=24) may not fully capture the true modular structure of the protein–protein interaction network, resulting in several single-gene clusters that likely represent outliers rather than biologically meaningful groups. Additionally, the analysis relied entirely on *in silico* tools and publicly available databases, which, while powerful, are limited by the completeness and accuracy of current annotations and experimental data.

As a result, some predicted interactions or functional associations may be overestimated or missed due to gaps in existing knowledge. Furthermore, no experimental validation was performed in this study to confirm the predicted interactions or biological roles of the identified genes. Future studies incorporating gene expression profiling, protein assays, and functional experiments are recommended to substantiate and expand upon these findings.

6.3 Future Directions

Building on the findings of this *in silico* study, future research should focus on experimental validation of the identified genes and interactions to confirm their roles in ovarian cancer pathogenesis.

1. Experimental validation of the predicted *KRT5–TP53* interaction, using techniques such as co-immunoprecipitation, western blotting, and immunohistochemistry, to confirm the physical and functional association observed *in silico*.

2. Assess the biological role of *KRT5* in ovarian cancer cell lines through gene knockdown or over-expression experiments, evaluating its impact on *TP53* activity, cell cycle regulation, apoptosis, and response to DNA damage.

3. Investigate the expression and clinical relevance of *KRT5* and *TP53* in patient-derived ovarian cancer tissues, including correlation with *TP53* mutation status and patient outcomes, to evaluate their potential as biomarkers or therapeutic targets.

4. Evaluate medium-confidence genes identified in this study, such as *ABCG2* and *KRT6B*, for their potential indirect roles in ovarian cancer pathogenesis, as they may represent secondary regulators or modulators of disease mechanisms.

5. Explore alternative clustering and network analysis methods, such as community detection or modularity-based algorithms, to better capture biologically meaningful gene modules and avoid artifacts such as single-gene clusters.

6. Integrate multi-omics datasets, including transcriptomics, proteomics, and epigenomics, with the predicted networks and clusters to gain a more comprehensive understanding of ovarian cancer pathogenesis and to identify robust biomarkers or targets.

7. Extend bioinformatics predictions with experimental and clinical validation, bridging genotype–phenotype gaps and strengthening the translational relevance of *in silico* findings in ovarian cancer research.

These directions emphasize the importance of combining computational predictions with functional and clinical studies to validate novel candidate genes and interactions implicated in ovarian cancer and to refine our understanding of its molecular underpinnings.

Bibliography

[1] M. S. Chandraprasad, A. Dey, and M. K. Swamy, "Introduction to cancer and treatment approaches," in Paclitaxel: Elsevier, 2022, pp. 1-27.

[2] S. R. Chinreddy et al., "Unraveling TRPV1's Role in Cancer: Expression, Modulation, and Therapeutic Opportunities with Capsaicin," Molecules, vol. 29, no. 19, p. 4729, 2024.

[3] V. Johariya, A. Joshi, N. Malviya, and S. Malviya, "Introduction to cancer," in Medicinal plants and cancer chemoprevention: CRC Press, 2023, pp. 1-28.

[4] H. M. Alkhalidi et al., "Revolutionizing Cancer Treatment: Biopolymer-Based Aerogels as Smart Platforms for Targeted Drug Delivery," Macromolecular Rapid Communications, vol. 45, no. 9, p. 2300687, 2024.

[5] B. Koul and B. Koul, "Types of cancer," Herbs for cancer treatment, pp. 53-150, 2019.

[6] Z. Momenimovahed, A. Tiznobaik, S. Taheri, and H. Salehiniya, "Ovarian cancer in the world: epidemiology and risk factors," International journal of women's health, pp. 287-299, 2019.

[7] L. A. Torre et al., "Ovarian cancer statistics, 2018," CA: a cancer journal for clinicians, vol. 68, no. 4, pp. 284-296, 2018.

[8] M. A. Roett and P. Evans, "Ovarian cancer: an overview," American family physician, vol. 80, no. 6, pp. 609-616, 2009.

[9] C. A. Doubeni, A. R. Doubeni, and A. E. Myers, "Diagnosis and management of ovarian cancer," American family physician, vol. 93, no. 11, pp. 937-944, 2016.

[10] L. Wang, Q. Zhang, H. H. Tong, X. Yao, H. Liu, and G. Li, "Computational methods for unlocking the secrets of potassium channels: Structure, mechanism, and drug design," Wiley Interdisciplinary Reviews: Computational Molecular Science, vol. 14, no. 1, p. e1704, 2024.

[11] F. Salehi, L. Dunfield, K. P. Phillips, D. Krewski, and B. C. Vanderhyden, "Risk factors for ovarian cancer: an overview with emphasis on hormonal factors," Journal of Toxicology and Environmental Health, Part B, vol. 11, no. 3-4, pp. 301-321, 2008.

[12] J. Li et al., "Identification of BRCA1: c. 5470_5477del as a founder mutation in Chinese ovarian cancer patients," Frontiers in Oncology, vol. 11, p. 655709, 2021.

[13] A. Ikram et al., "National Cancer Registry of Pakistan: First Comprehensive Report of Cancer Statistics 2015-2019," (in eng), J Coll Physicians Surg Pak, vol. 33, no. 6, pp. 625-632, Jun 2023.

[14] T. Wasim, J. Mushtaq, A. Z. Wasim, and G. E. Raana, "Gynecological malignancies at tertiary care hospital, Pakistan: A five-year review," (in eng), Pak J Med Sci, vol. 37, no. 3, pp. 621-627, May-Jun 2021.

[15] U. Sherefedin et al., "Physicochemical properties and drug likeness of hydroxycinnamic acids and their molecular docking with caffeine and amoxicillin: Potential anticancer drugs," Results in Chemistry, vol. 13, p. 101996, 2025.

[16] A. J. Clark and J. W. Lillard Jr, "A comprehensive review of bioinformatics tools for genomic biomarker discovery driving precision oncology," Genes, vol. 15, no. 8, p. 1036, 2024.

[17] J. Huang et al., "Worldwide Burden, Risk Factors, and Temporal Trends of Ovarian Cancer: A Global Study," (in eng), Cancers (Basel), vol. 14, no. 9, Apr 29 2022.

[18] F. Zalfa et al., "Genome-wide identification and validation of gene expression biomarkers in the diagnosis of ovarian serous cystadenocarcinoma," Cancers, vol. 14, no. 15, p. 3764, 2022.

[19] S. Akter et al., "Recent advances in ovarian cancer: therapeutic strategies, potential biomarkers, and technological improvements," Cells, vol. 11, no. 4, p. 650, 2022.

[20] U. A. Matulonis, A. K. Sood, L. Fallowfield, B. E. Howitt, J. Sehouli, and B. Y. Karlan, "Ovarian cancer (Primer)," Nature Reviews: Disease Primers, vol. 2, no. 1, 2016.

[21] U. A. Matulonis, A. K. Sood, L. Fallowfield, B. E. Howitt, J. Sehouli, and B. Y. Karlan, "Ovarian cancer," Nature reviews Disease primers, vol. 2, no. 1, pp. 1-22, 2016.

[22] J. Huang et al., "Worldwide burden, risk factors, and temporal trends of ovarian cancer: a global study," Cancers, vol. 14, no. 9, p. 2230, 2022.

[23] Y. Xu et al., "Mining TCGA database for gene expression in ovarian serous cystadenocarcinoma microenvironment," PeerJ, vol. 9, p. e11375, 2021.

[24] S. M. McGregor, "Pathologic classification of ovarian cancer," in Ovarian Cancer: Methods and Protocols: Springer, 2021, pp. 11-40.

[25] D. D. Bowtell et al., "Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer," Nature reviews Cancer, vol. 15, no. 11, pp. 668-679, 2015.

[26] I. A. Cree, V. A. White, B. I. Indave, and D. Lokuhetty, "Revising the WHO classification: female genital tract tumours," Histopathology, vol. 76, no. 1, pp. 151-156, 2020.

[27] A.-M. Patch et al., "Whole–genome characterization of chemoresistant ovarian cancer," Nature, vol. 521, no. 7553, pp. 489-494, 2015.

[28] D. M. Gershenson, "Update on malignant ovarian germ cell tumors," Cancer, vol. 71, no. S4, pp. 1581-1590, 1993.

[29] C. J. Cabasag et al., "Ovarian cancer today and tomorrow: A global assessment by world region and Human Development Index using GLOBOCAN 2020," International journal of cancer, vol. 151, no. 9, pp. 1535-1541, 2022.

[30] R. M. Pallar et al., "Lectin: a molecular tool in cancer diagnosis and therapy with special reference to reproductive cancers," Molecular Biotechnology, vol. 67, no. 2, pp. 456-468, 2025.

[31] B. Goff, "Symptoms associated with ovarian cancer," Clinical obstetrics and gynecology, vol. 55, no. 1, pp. 36-42, 2012.

[32] M. A. Shihab and N. E. Daim, "Ovarian Cancer," Eastern Journal of Agricultural and Biological Sciences, vol. 4, no. 1, pp. 32-42, 2024.

[33] M. Mahima et al., "An in-depth analysis of ovarian cancer: pathogenesis and clinical manifestation," Drug Research, vol. 72, no. 08, pp. 424-434, 2022.

[34] Y. Zhang, L. Cao, D. Nguyen, and H. Lu, "TP53 mutations in epithelial ovarian cancer," Translational cancer research, vol. 5, no. 6, p. 650, 2016.

[35] D. D. Bowtell, "The genesis and evolution of high-grade serous ovarian cancer," Nature Reviews Cancer, vol. 10, no. 11, pp. 803-808, 2010.

[36] D.-H. Jung et al., "A focused review on molecular signalling mechanisms of ginsenosides anti-Lung cancer and anti-inflammatory activities," Anti-Cancer Agents in Medicinal Chemistry-Anti-Cancer Agents), vol. 23, no. 1, pp. 3-14, 2023.

[37] K. Horackova, M. Janatova, P. Kleiblova, Z. Kleibl, and J. Soukupova, "Early-onset ovarian cancer¡ 30 years: What do we know about its genetic predisposition?," International journal of molecular sciences, vol. 24, no. 23, p. 17020, 2023.

[38] A. Toss et al., "Hereditary ovarian cancer: not only BRCA 1 and 2 genes," BioMed research international, vol. 2015, no. 1, p. 341723, 2015.

[39] K. Lhotova et al., "Multigene panel germline testing of 1333 Czech patients with ovarian cancer," Cancers, vol. 12, no. 4, p. 956, 2020.

[40] J. Casagrande, "'INCESSANT OVULATION' AND OVARIAN CANCER," The Lancet, vol. 314, no. 8135, pp. 170–173, Jul. 1979.

[41] V. Beral, R. Doll, C. Hermon, R. Peto, and G. Reeves, "Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23 257 women with ovarian cancer and 87 303 controls," The Lancet, vol. 371, no. 9609, pp. 303–314, Jan. 2008.

[42] R. J. Kurman and I.-M. Shih, "The Origin and Pathogenesis of Epithelial ovarian Cancer: A proposed unifying theory," The American Journal of Surgical Pathology, vol. 34, no. 3, pp. 433–443, Feb. 2010.

[43] Modugno, R. B. Ness, and J. E. Wheeler, "Reproductive risk factors for epithelial ovarian cancer according to histologic type and invasiveness," Annals of Epidemiology, vol. 11, no. 8, pp. 568–574, Nov. 2001.

[44] A. Lukanova and R. Kaaks, "Endogenous hormones and ovarian Cancer: Epidemiology and current hypotheses," Cancer Epidemiology Biomarkers & Prevention, vol. 14, no. 1, pp. 98–107, Jan. 2005.

[45] H. A. Risch, "Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone," JNCI Journal of the National Cancer Institute, vol. 90, no. 23, pp. 1774–1786, Dec. 1998.

[46] R. B. Ness and C. Cottreau, "Possible role of ovarian epithelial inflammation in ovarian cancer," JNCI Journal of the National Cancer Institute, vol. 91, no. 17, pp. 1459–1467, Sep. 1999.

[47] A. Agarwal, S. Gupta, and R. K. Sharma, "Role of oxidative stress in female reproduction," Reproductive Biology and Endocrinology, vol. 3, no. 1, Jul. 2005.

[48] C. Vo and M. E. Carney, "Ovarian Cancer Hormonal and Environmental Risk Effect," Obstetrics and Gynecology Clinics of North America, vol. 34, no. 4, pp. 687-700, 2007/12/01/ 2007.

[49] A. Ali et al., "The burden of cancer, government strategic policies, and challenges in Pakistan: A comprehensive review," Frontiers in nutrition, vol. 9, p. 940514, 2022.

[50] D. Bell et al., "Integrated genomic analyses of ovarian carcinoma," Nature, vol. 474, no. 7353, pp. 609–615, Jun. 2011.

[51] K. H. Vousden and C. Prives, "Blinded by the Light: The Growing Complexity of p53," Cell, vol. 137, no. 3, pp. 413–431, May 2009

[52] M. Olivier, M. Hollstein, and P. Hainaut, "TP53 mutations in human cancers: origins, consequences, and clinical use," Cold Spring Harbor Perspectives in Biology, vol. 2, no. 1, p. a001008, Nov. 2009.

[53] V. J. N. Bykov, S. E. Eriksson, J. Bianchi, and K. G. Wiman, "Targeting mutant p53 for efficient cancer therapy," Nature Reviews. Cancer, vol. 18, no. 2, pp. 89–102, Dec. 2017.

[54] B. M. Norquist et al., "Inherited mutations in women with ovarian carcinoma," JAMA Oncology, vol. 2, no. 4, p. 482, Dec. 2015.

[55] J. A. Ledermann and F. El-Khouly, "PARP inhibitors in ovarian cancer: Clinical evidence for informed treatment decisions," British Journal of Cancer, vol. 113, no. S1, pp. S10–S16, Dec. 2015.

[56] G. M. Findlay et al., "Accurate classification of BRCA1 variants with saturation genome editing," Nature, vol. 562, no. 7726, pp. 217–222, Sep. 2018.

[57] K. B. Kuchenbaecker et al., "Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers," JAMA, vol. 317, no. 23, p. 2402, Jun. 2017.

[58] D. Garcia-Carracedo et al., "Loss of PTEN Expression Is Associated with Poor Prognosis in Patients with Intraductal Papillary Mucinous Neoplasms of the Pancreas," Clinical Cancer Research, vol. 19, no. 24, pp. 6830–6841, Oct. 2013.

[59] S. Jones et al., "Low-grade serous carcinomas of the ovary contain very few point mutations," The Journal of Pathology, vol. 226, no. 3, pp. 413–420, Nov. 2011.

[60] K.-T. Kuo et al., "Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma," American Journal of Pathology, vol. 174, no. 5, pp. 1597–1601, Apr. 2009.

[61] K.-T. Kuo et al., "Frequent activating mutations of PIK3CA in ovarian clear cell carcinoma," American Journal of Pathology, vol. 174, no. 5, pp. 1597–1601, Apr. 2009.

[62] C. Balch, T. H. -m. Huang, R. Brown, and K. P. Nephew, "The epigenetics of ovarian cancer drug resistance and resensitization," American Journal of Obstetrics and Gynecology, vol. 191, no. 5, pp. 1552–1572, Nov. 2004.

[63] D. Jelovac and D. K. Armstrong, "Recent progress in the diagnosis and treatment of ovarian cancer," CA a Cancer Journal for Clinicians, vol. 61, no. 3, pp. 183–203, Apr. 2011.

[64] S. Vaughan et al., "Rethinking ovarian cancer: recommendations for improving outcomes," Nature Reviews. Cancer, vol. 11, no. 10, pp. 719–725, Sep. 2011.

[65] E. Lengyel, "Ovarian cancer development and metastasis," American Journal of Pathology, vol. 177, no. 3, pp. 1053–1064, Jul. 2010.

[66] A.-M. Patch et al., "Whole–genome characterization of chemoresistant ovarian cancer," Nature, vol.521, no. 7553, pp. 489–494, May 2015.

[67] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA a Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, Feb. 2021.

[68] T.-K. Jenssen, A. Lægreid, J. Komorowski, and E. Hovig, "A literature network of human genes for high-throughput analysis of gene expression," Nature Genetics, vol. 28, no. 1, pp. 21–28, May 2001.

[69] M. C. Brinton and E. Oh, "Babies, work, or both? Highly educated women's employment and fertility in East Asia," American Journal of Sociology, vol. 125, no. 1, pp. 105-140, 2019.

[70] A. Sevinc, M. Adli, M. E. Kalender, and C. Camci, "Benign causes of increased serum CA-125 concentration," The lancet oncology, vol. 8, no. 12, pp. 1054-1055, 2007.

[71] J. Zhou, W. Cao, L. Wang, Z. Pan, and Y. Fu, "Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer," Computers in biology and medicine, vol. 146, p. 105608, 2022.

[72] S. Saraf, A. De, and B. Tripathy, "Effective use of computational biology and artificial intelligence in the domain of medical oncology," in Computational Intelligence for Oncology and Neurological Disorders: CRC Press, 2024, pp. 228-252.

[73] C. G. A. R. Network, "Integrated genomic analyses of ovarian carcinoma," Nature, vol. 474, no. 7353, p. 609, 2011.

[74] D. Szklarczyk et al., "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," Nucleic acids research, vol. 51, no. D1, pp. D638-D646, 2023.

[75] A. Franceschini, "STRINGdb package vignette," Nucleic Acids Res, vol. 41, pp. D808-D815, 2013.

[76] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," Kanehisa laboratories, 2012.

[77] D. Szklarczyk et al., "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," Nucleic acids research, vol. 47, no. D1, pp. D607-D613, 2019.

[78] A. Beg, R. Parveen, H. Fouad, M. E. Yahia, and A. S. Hassanein, "Identification of Driver Genes and in Ovarian Cancer through an Integrated In-Silico Approach," Biology, vol. 12, no. 2, p. 192, Jan. 2023.

[79] A. Beg, R. Parveen, H. Fouad, M. E. Yahia, and A. S. Hassanein, "Unravelling driver genes as potential therapeutic targets in ovarian cancer via integrated bioinformatics approach," Journal of Ovarian Research, vol. 17, no. 1, Apr. 2024.

[80] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," Genome Biology, vol. 18, no. 1, May 2017, doi: 10.1186/s13059-017-1215-1.

[81] R. J. Kurman and I.-M. Shih, "The dualistic model of ovarian carcinogenesis," American Journal of Pathology, vol. 186, no. 4, pp. 733–747, Mar. 2016.

[82] S. Vaughan et al., "Rethinking ovarian cancer: recommendations for improving outcomes," Nature Reviews. Cancer, vol. 11, no. 10, pp. 719–725, Sep. 2011.

[83] D. Szklarczyk et al., "The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets," Nucleic Acids Research, vol. 49, no. D1, pp. D605–D612, Nov. 2020.

[84] C. J. Alves et al., "Bone injury and repair trigger central and peripheral NPY neuronal pathways," PLoS ONE, vol. 11, no. 11, p. e0165465, Nov. 2016.

[85] X. Xu et al., "MiRNA–MRNA associated with survival in endometrial cancer," Frontiers in Genetics, vol. 10, Aug. 2019.

[86] J. Zhu, T. Zhang, J. Jiang, M. Yang, N. Xia, and Y. Chen, "Genetic variation perspective reveals potential drug targets for subtypes of endometrial cancer," Scientific Reports, vol. 14, no. 1, Nov. 2024.

[87] F. Mehryary, K. Nastou, T. Ohta, L. J. Jensen, and S. Pyysalo, "STRING-ing together protein complexes: Corpus and methods for extracting physical protein interactions from the biomedical literature," Bioinformatics, Sep. 2024.

[88] S. Lloyd, "Least squares quantization in PCM," IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[89] T. Zhang, L. Zhang, and F. Li, "Integrative network analysis identifies potential targets and drugs for ovarian cancer," BMC Medical Genomics, vol. 13, pp. 1-11, 2020.

[90] P. Tabrizi-Nejhadi, H. MotieGhader, and M. Maleki, "Protein-Protein Interaction network analysis for Discovering Potent Candidate Drugs in Female Infertility," International Journal of Biophotonics and Biomedical Engineering (IJBBE), vol. 2, no. 1, pp. 1-14, 2022.

[91] Y. Zhao, J. Pi, L. Liu, W. Yan, S. Ma, and L. Hong, "Identification of the hub genes associated with the prognosis of ovarian cancer patients via integrated bioinformatics analysis and experimental validation," Cancer Management and Research, pp. 707-721, 2021.

[92] J. Chen, Y. Cai, R. Xu, J. Pan, J. Zhou, and J. Mei, "Identification of four hub genes as promising biomarkers to evaluate the prognosis of ovarian cancer in silico," Cancer cell international, vol. 20, pp. 1-11, 2020.

Appendix - A

List of Genes Retrieved From Coremine Medical

A complete list of approximately 2000 ovarian cancer related genes was retrieved using Coremine Medical. To maintain clarity in the main body of the thesis, only the first half of the genes was presented in the Results chapter (Table 4.1). The remaining genes are provided here in tabular form for reference and completeness.

Table 4.1: (Continued).

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CLPTM1	*XPO7*	*HSP90AB1*	*TOPORS*
ELAVL2	*POT1*	*EFEMP1*	*NR1H2*
CDK9	*IL17RB*	*ZNRF3*	*MGAT5*
RDM1	*RIMS2*	*CCK*	*NTN4*
IMMP1L	*SATB1*	*HNF1A*	*SYT1*
F8	*TEAD3*	*VPS13C*	*MIR192*
DOT1L	*SF3B1*	*ATP7B*	*PAX9*
NKX3-1	*COPS5*	*RRM2*	*DKK1*
CYP4F3	*MIR183*	*LPAL2*	*DDX53*
LY9	*HEPACAM2*	*ACVR2A*	*CA2*
SMARCC2	*PPP2R4*	*MIF*	*FOXC2*
TBX3	*ADRA2B*	*LAMP3*	*AGGF1*
ZNF750	*MSR1*	*PRF1*	*INHBC*
SLC45A3	*VAMAS6*	*FLNA*	*CENPI*
LIG1	*SMARCA5*	*CUL7*	*HERC1*
CDK16	*PREX2*	*MYH13*	*XAB2*
HLA-DRB4	*PGAM1*	*TRIP12*	*MAGEC3*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CTC1	*CSTB*	*MUC5B*	*PALD1*
ADIPOQ	*SLC3A2*	*OGDHL*	*CLIP4*
MIR23A	*RAD23B*	*PRC1*	*NR5A2*
JTB	*HSD17B12*	*SNRPE*	*C1QA*
IFI16	*CD63*	*SHBG*	*TP53I3*
NTF3	*HMGB1*	*CCNO*	*KDM5B*
AP1S2	*EAF2*	*HIST1H2AH*	*PCBP1*
TPTE	*CACUL1*	*SOX9*	*SDHA*
GRB2	*SPDL1*	*CBR3-AS1*	*MAP3K7*
SCNN1A	*RARB*	*TAF4*	*FGF1*
CKAP2	*MTHFR*	*WNT3A*	*DAXX*
NBL1	*HOXD11*	*CLSPN*	*NOS2*
ULK2	*CDC25B*	*FADD*	*MIR196A2*
SIK1	*CLDN7*	*APOB*	*GATA2*
CYBB	*DHX9*	*ELK1*	*DDX5*
PAF1	*CARD11*	*MRPL28*	*ULK1*
BRD7	*NQO2*	*ID2*	*CEP250*
RAD1	*BCAR1*	*GREM1*	*EIF4A3*
USP22	*KLK2*	*MTA3*	*KEAP1*
MTUS1	*RIMBP2*	*CD4*	*PTPRE*
PTPRM	*CNRIP1*	*CXCL2*	*IGF1*
BNIP3L	*SYCE1*	*LZTR1*	*KRT17*
MTAP	*NCOR2*	*HPCA*	*PDCD5*
UBASH3A	*SDHC*	*MBD4*	*NOS1*
AGTR1	*ITGAV*	*PTPN1*	*MMP1*
ING1	*CRKL*	*JAK1*	*ZBTB38*
CD70	*IGFBP5*	*SART1*	*UPK2*
THY1	*RNASEL*	*NUP205*	*CBX6*
CUX1	*DNASE1*	*DDX3X*	*DYNAP*
MS4A1	*RRM1*	*EWSR1*	*STAT5A*
CLGN	*MPZL2*	*ADAMTS1*	*CS*
MACF1	*TRPS1*	*SIK3*	*FOXJ2*
PLEC	*BCL3*	*MKL1*	*CDCA2*
PSMC2	*NAA50*	*NTRK2*	*USP12P1*
DCAF17	*UGT1A1*	*RBL2*	*TLX2*
MCM8	*KRT6B*	*ARTN*	*STAB1*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
CLTCL1	*PHLDA3*	*COL12A1*	*OBSCN*
CXCL16	*G6PD*	*PHB*	*AP3S1*
PPIP5K2	*WDR62*	*MTMR11*	*BCL2L14*
ART3	*EREG*	*SEMA3F*	*VTRNA2-1*
CRH	*USP7*	*JAG1*	*DHPS*
TET2	*CDK2AP1*	*HOXA5*	*SLC5A2*
CDH5	*GLS*	*SYK*	*GSTO1*
DRD2	*HDAC8*	*YY1*	*TBCE*
CPEB1	*HTT*	*RNF20*	*SUB1*
IFNB1	*MCM4*	*ATP4A*	*BCLAF1*
FMN1	*FUBP1*	*ABCA5*	*ABCA13*
HPLH1	*ITGB3*	*HDAC3*	*HSD17B4*
CDC27	*DUSP4*	*CD96*	*EEF1D*
HR	*MC1R*	*VDR*	*CENPE*
DACH1	*USP12*	*CYP27B1*	*NAALADL2*
TNFSF11	*CREB3L4*	*SMO*	*SRSF1*
CYP2C19	*UBC*	*SPPL3*	*DPEP2*
MKNK1	*LLGL2*	*BRD9*	*PRMT1*
CREB3	*SULT1A1*	*ACSL3*	*KANSL1*
CALM2	*MYD88*	*MYCBP*	*ROBO1*
CALM3	*HTRA1*	*ZBTB14*	*SOHLH1*
PDCD6	*YWHAE*	*ATF2*	*PLA2G1B*
CUL3	*MT1E*	*TIA1*	*BCL2L11*
PTPRD	*WHSC1*	*FLNB*	*DIAPH2*
BIN2	*SPAM1*	*CCL4*	*WIF1*
USP28	*RBX1*	*KDM5C*	*MTRNR2L2*
CYP24A1	*MEFV*	*LOX*	*PRKACB*
UBA52	*AMER1*	*GSK3B*	*PTCH2*
MIA3	*TERF2IP*	*SOX18*	*TCERG1*
TFE3	*RUNX2*	*MIR19A*	*WISP2*
MAML2	*SIRT1*	*IL17A*	*TNFRSF1A*
MRPL23	*UBE2E1*	*LAMC3*	*STK33*
IFI6	*MELK*	*MT1M*	*NANOS2*
BNIP3	*RBBP4*	*NUP214*	*ALDH16A1*
GAGE7	*MIR17HG*	*EIF3B*	*CCR2*
PBRM1	*NDUFS1*	*ZMAT3*	*IRX2*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
KAT6B	*ATP6V0A2*	*MMP3*	*IL2RG*
KDM6B	*CKAP4*	*CRIP1*	*SLIT3*
CYP2C9	*PYCARD*	*LRP1*	*HNRNPA1*
TBX2	*CDC42*	*MED1*	*SUMF2*
ALDH2	*APOL2*	*CALM1*	*EPHB2*
MITF	*TCF7L2*	*KIAA0319*	*HSD3B2*
INO80	*TFDP1*	*STAR*	*AP2B1*
RBMX	*DPP8*	*C3*	*LEP*
NFE2L2	*ZFP36L2*	*SETDB1*	*IL1A*
ITGB1	*GCLC*	*MED12*	*EPHX1*
RPS27A	*HTR1E*	*RP1L1*	*GEMIN2*
SOD1	*TRERF1*	*KMT2A*	*PPP1R16B*
ZNF135	*S1PR1*	*PGK1*	*ERCC6L*
BCL10	*DKC1*	*GNA11*	*APPL1*
FAT2	*ASXL1*	*HYAL1*	*STK3*
IRS2	*NFYA*	*SETD7*	*MMS19*
PPARGC1A	*H3F3B*	*HSPA1B*	*PSMB3*
MIR146B	*ACTB*	*BATF2*	*CASP10*
EGR1	*ATP1B2*	*FNDC3A*	*TCF3*
CXCR6	*PTPRS*	*CASP1*	*CDC45*
COL6A2	*MTUS2*	*MYH9*	*LMTK2*
TDP2	*FGF4*	*FLT3*	*CELF2*
SMG1	*JAK3*	*PRDX1*	*RAP1A*
PRMT5	*IL3*	*MMRN1*	*PERP*
NCL	*SVIL*	*EIF4ENIF1*	*PCSK9*
A2M	*BCL6*	*HPGDS*	*FCN3*
CCAR2	*CD1C*	*MSX1*	*TFAP2A*
CHTF18	*CDT1*	*RARA*	*PPP1CA*
GABRA6	*SNRPB*	*CDCA7*	*TOP3A*
SMC1A	*PPAP2B*	*LARP7*	*ALDH8A1*
ADAM9	*MAPK13*	*RBP2*	*NELFCD*
HSPA2	*SLC18A1*	*USH2A*	*SYNPO2*
HIF1AN	*RUNX1T1*	*IFNA17*	*CYP2E1*
CRIP2	*RAE1*	*PRKCA*	*PLAG1*
IRAK2	*SRSF10*	*COL6A1*	*PPP1R1A*
RRP1B	*MAD1L1*	*F5*	*SERPINA5*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
RABGEF1	*USP21*	*TRIB1*	*TCF7L1*
BCORL1	*TNNT2*	*GADD45B*	*HDAC5*
ELANE	*HBA1*	*PPCS*	*EPOR*
PHF6	*SDC1*	*SNCAIP*	*TNFRSF11B*
BCL2A1	*ABCD2*	*CYP3A*	*CFTR*
IGSF1	*ALDH4A1*	*CIB2*	*RASA1*
SPAG5	*WDR77*	*ACACB*	*USP45*
ENAH	*NAT2*	*CABIN1*	*UBE2S*
MYOD1	*MAP3K13*	*GPLD1*	*CLASP1*
CAGE1	*KCNQ1OT1*	*DSC3*	*MT-CYB*
RYR2	*AKAP9*	*CD1B*	*CSDE1*
WNT1	*MAGED1*	*FGF19*	*CNGB1*
HIST2H2AA3	*PROC*	*LEF1*	*ABL1*
RAPGEF4	*PF4V1*	*ELSPBP1*	*HBB*
SLC38A1	*PPAT*	*ARHGEF28*	*PRL*
PDZD2	*CBX3*	*ADD3*	*DNAJA3*
CCP110	*DAPK3*	*PRTN3*	*CDKN2D*
UGT1A	*USP9Y*	*EFNA2*	*FGF23*
NDUFS2	*PCAP*	*E2F7*	*NHP2*
IRX1	*FGG*	*TAS2R64P*	*MARCKS*
BAZ1A	*NEIL1*	*SULT1E1*	*ZNF77*
HSPG2	*SUZ12*	*ARHGAP1*	*OPHN1*
IQGAP2	*ZMIZ1*	*SELL*	*METTL14*
PIAS4	*ASPH*	*HOXC6*	*HOOK3*
WDPCP	*MCM10*	*KPNA4*	*SCN5A*
TERF2	*CBLC*	*FTO*	*DNAH5*
CORO1A	*SNCA*	*ZFP36L1*	*APOBEC3G*
ATAD5	*KLKB1*	*ADAMTS13*	*CUL1*
CYP2A	*EPHB1*	*STAG2*	*HCAR1*
CYP11A1	*HIST2H2AC*	*GRM5*	*LRRK2*
EFEMP2	*MYLIP*	*CEACAM1*	*PSCA*
HEXIM1	*ESCO1*	*PLK3*	*CSF3R*
SHH	*RBM39*	*GFI1*	*MICB*
METTL1	*PTGFR*	*SDHAF2*	*SELE*
TRPM3	*SEMA3C*	*RCHY1*	*NPLOC4*
ADH1B	*GTF2H2*	*UHRF1*	*CYP2D6*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
UBE2L6	*MB21D1*	*CYP3A5*	*CHD8*
XIRP1	*HOXA1*	*NELL1*	*GRIA1*
KCNQ1	*PSG5*	*ASH2L*	*ALDH6A1*
COL1A2	*BHLHE40*	*PSMD4*	*MFSD2A*
WISP1	*RGS6*	*CREB1*	*FLCN*
ARG1	*FTMT*	*PGF*	*ABCG1*
SELP	*FGL2*	*PTPN6*	*FOXP2*
COL4A2	*SULT2A1*	*ZNF331*	*CBL*
PON1	*ARL2BP*	*STMN2*	*MGP*
VLDLR	*TRAPPC10*	*MAGT1*	*STAT5B*
DGCR8	*YTHDC1*	*RNF115*	*HSD17B7*
EGLN2	*THM*	*ZBTB16*	*USF1*
PKHD1	*KRT6A*	*TINF2*	*CRNN*
HT	*CHL1*	*PSPC1*	*PPP6R3*
RRM2B	*RPN2*	*RPS20*	*LTBP4*
PIN1	*SLCO4A1*	*FOXC1*	*RBBP5*
PDE11A	*MAPK12*	*CPA1*	*SNX9*
CD86	*GTF2H4*	*ECHDC1*	*SKP1*
KIF16B	*NUP107*	*SLC38A2*	*ALDH5A1*
SPO11	*PRM3*	*SMAD6*	*PDIK1L*
CYCS	*GPRC5A*	*FYN*	*NELL2*
SFTPC	*PDC*	*LDLR*	*HPRT1*
IRS1	*PLCG1*	*PCK1*	*NPM1*
KLRC1	*SOD2*	*KAT5*	*SIAH1*
ALDH9A1	*METRNL*	*GAPDH*	*AIRE*
AIF1	*TSHZ3*	*RIC8A*	*HELB*
OTC	*CHAF1B*	*GDNF*	*PRPF8*
CD151	*KIF1B*	*TXNRD2*	*ADAR*
EIF1	*HM13*	*S100B*	*FLI1*
GFAP	*ERG*	*MAP2K5*	*INSL3*
CD160	*BUB3*	*PMP22*	*GFI1B*
ALDH3A1	*NISCH*	*GPX1*	*LGALS12*
AHR	*QKI*	*SDHD*	*NUMA1*
EDARADD	*GABPA*	*TNFRSF13B*	*STXBP2*
UBE2E2	*TLR4*	*PIGA*	*USP39*
CASP14	*KLF15*	*ROBO2*	*DLEU1*

continued on next page

Table 1 continued from previous page

Gene Symbol	Gene Symbol	Gene Symbol	Gene Symbol
NPRL2	*HOXD13*	*CLDN1*	*SLC9C1*
ANPEP	*PSMC1*	*BMP2*	*CST6*
ANTXR1	*NES*	*HNRNPC*	*SERPINF2*
OBFC1	*GSDMA*	*CEBPD*	*FOXO4*
RCOR1	*PPARG*	*SOS1*	*CD19*
RNF146	*ABCB4*	*PRKACA*	*CSN2*
HK1	*PRDX2*	*ELAVL3*	*DBF4*
GPX6	*CPQ*	*LGALS4*	*KLHL2*
H3F3A	*CHRNB3*	*ZBTB17*	*MANEA*
AGFG2	*CCL7*	*PTGER2*	*MYO10*
ROCK2	*EEF1A1*	*SLIT1*	*BLNK*
SLC39A1	*CYP2B6*	*PIAS1*	*SRCAP*
BRF1	*IL6*	*RPL23*	*CYP1A2*
HNRNPH1	*LLGL1*	*NEIL2*	*HFE*
EXT1	*U2AF1*	*POLH*	*HSPA1A*
USP25	*RND2*	*CDK8*	*PRPF19*
SPI1	*IL33*	*QRSL1*	*CHDM*
DIAPH1	*GFRA1*	*SHC1*	*PDCD6IP*
KLF2	*LY6E*	*CDK5RAP3*	*TUBA1A*
KLF1	*RGPD2*	*NFIL3*	*CD40*
ABCB11	*SLC22A1*	*NTRK3*	*KCNH8*
OTX1	*FGFR1OP*	*MT-ATP6*	*IL1RL1*
PSEN2	*FABP3*	*C1orf56*	*DYRK1A*
SCAI	*PAX3*	*PKD1*	*YBX3*
TOMM40	*ALOX15B*	*SOAT1*	*CRYGC*
LRIG2	*HIST2H2BE*	*POLR2A*	*FOS*
FKBP4	*NOX4*	*MAVS*	*ADH1C*
SYNGAP1	*POU2F2*	*GZMA*	*POTEF*
GTPBP3	*ARNTL*	*LTBR*	*UNG*
MAST1	*PTPRC*	*MYB*	*NRP2*
NPY4R	*PAX5*	*LOXL1*	*CCT3*