Hypermutation in Microsatellite Stable Cancers

By

Aqsa Arshad

Reg. # 02282113009

National Centre for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2023

Hypermutation in Microsatellite Stable Cancers

By

Aqsa Arshad

A thesis submitted in the partial fulfillment of the

requirements for the degree of

Master of Philosophy in

Bioinformatics

Supervisor

Dr. Adnan Ahmad Ansari

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-i-Azam, University

Islamabad, Pakistan

2023


بسم الله الرحمن الرحيم

DEDICATION

I humbly dedicate this work to the Creator of the universe, Allah Almighty, for granting me good health, a sound mind, protection, inner strength, and the skills necessary to undertake this journey. I also dedicate this work to Hazrat Muhammad ﷺ, the role model for humanity, for whom this world was created. My deepest gratitude goes to my beloved parents, who have been my unwavering source of inspiration and strength. They have provided me with continuous moral, spiritual, emotional, and financial support, making this achievement possible. I want to express my heartfelt appreciation to my siblings, friends, teachers, and mentors, all of whom have generously shared their wisdom, guidance, and encouragement, helping me reach this significant milestone.

A special thanks to my supervisor, **Dr. Adnan Ahmad Ansari**, for their invaluable guidance, unwavering support, and encouragement throughout this challenging journey. Lastly, I extend my gratitude to Quaid-i-Azam University, Islamabad, for providing me with the opportunity to pursue my academic aspirations.

DECLARATION

I hereby solemnly declare that the work **"Hypermutation in Microsatellite Stable Cancers"** presented in the following thesis is my own research efforts, except where otherwise acknowledged, and that the thesis is my own composition.

The thesis has neither been published previously nor does it contain any material from the published resources that can be considered a violation of international copyright law. No part of the thesis has been previously presented for any other degree.

I also declare that I am aware of the term copyright and plagiarism I will be responsible for the consequences of violation of these rules (if any) found in this thesis. The thesis has been checked for plagiarism by Turnitin software.

Date: _____ _____

Aqsa Arshad

CERTIFICATE

This thesis is submitted by **Aqsa Arshad** from National Center for Bioinformatics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan, and is accepted in its present form as satisfying the thesis requirements for the Degree of Master of Philosophy in Bioinformatics.

Internal Examiner: _____

Dr. Adnan Ahmad Ansari

Assistant Professor & Supervisor

Quaid-i-Azam University, Islamabad

External Examiner: _____

Dr.

Designation

Corresponding Intuitional Address

Chairperson: _____

Dr. Syed Sikandar Azam

Professor

Quaid-i-Azam University, Islamabad

Date: _____

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious and the Most Merciful, I begin by expressing my gratitude for the strengths and blessings bestowed upon me by Allah, which have empowered me to successfully conclude this thesis. I deeply value the dedicated educators at the National Center for Bioinformatics who have been instrumental in my academic journey. Furthermore, my heartfelt acknowledgments to my parents, who not only provided moral guidance but also nurtured my growth with their unwavering love and support. Their wisdom and care have played a significant role in shaping who I am today.

I want to convey my deepest appreciation to my supervisor, **Dr. Adnan Ahmad Ansari**, whose unwavering support, guidance, and belief in my abilities have been instrumental in the successful completion of this research. His encouragement has instilled in me the confidence to work both independently and as a part of a team. I extend my gratitude to Dean Faculty of Biological Sciences and Chairman **Prof Dr. Syed Sikander Azam** for their role in shaping NCB into a resource-rich platform for learning. I would also like to express my appreciation to Quaid-i-Azam University for fostering an exceptional research atmosphere, which I value greatly. I am thankful to the faculty members, including **Dr. Amir Ali Abbasi, Dr. Sajid Rashid, and Dr. Rana Rehan Khalid,** for their contributions to this environment.

I would also like to express my heartfelt appreciation to my siblings for their unwavering moral support throughout my journey. My friends, *Aqsa, Farah, Ifrah, Irba, Laiba, Reem, Aliza, Muqaddas and Ayesha* have been a constant source of help, encouragement, advice, and affection, for which I am truly thankful.

This achievement wouldn't have been possible without the support of my family, friends, and educators. Their contributions have been invaluable. Lastly, I extend my sincere thanks to my senior, **Layla Sehar** and **Aqeel Irfan**, for their guidance and support throughout this research.

Aqsa Arshad

Table of Contents

List of Figures:

List of Tables:

List of Abbreviations

MSS	Microsatellite Stable
MSI	Microsatellite Instable
MSI-H	Microsatellite Instable-High
MSI-L	Microsatellite Instable-Low
UCEC	Uterine Corpus Endometrium Carcinoma
COAD	Colon Adenocarcinoma
READ	Rectal Adenocarcinoma
PAAD	Pancreatic Adenocarcinoma
CRC	Colorectal Carcinoma
ESCA	Esophageal Carcinoma
UCS	Uterine Carcinosarcoma
STAD	Stomach Adenocarcinoma
TCGA	The Cancer Genome Atlas
EdgeR	Empirical Analysis of Digital Gene Expression Data in R
GDC	Genomic Data Commons
DAVID	Database for Annotation, Visualization, and Integrated Discovery tool

Abstract

Cancer stands as one of the deadliest diseases worldwide, claiming a significant number of lives. It's a complex condition influenced by various factors, including physical, chemical, and biological carcinogens. Cancer can be broadly categorized into genetic or somatic forms, with somatic cancer involving DNA changes or damage as a primary cause. Hypermutation, characterized by an excessive number of mutations in cancer cells, further divides into microsatellite instability (MSI-High and MSI-Low) and microsatellite stability (MSS). In recent years, immunotherapy has emerged as a promising and relatively low-side-effect approach to cancer treatment. Recent findings reported promising results of immunotherapy in MSI. This study focuses on MSS samples with higher mutation burden has potential for immunotherapy. To achieve this, we examined hypermutated cancers, including UCEC, COAD, READ, PAAD, STAD, UCS and ESCA, utilizing TCGA data. This analysis includes transcriptomic and differential gene expression analysis using cBioPortal and the edgeR package in R for Next-Generation Sequencing (NGS) data. Among the key mutations contributing to MSS, the POLE gene mutation emerged as a major player, alongside other reported mutations like APC R1114* and ZNHIT6 R258*. This study incorporated a total of 42 hypermutated samples. From this analysis, we uncovered common upregulations in genes like DKK4, FGF20, and HMX2 corelated to POLE mutations, as well as TSIX, GNG4, GSTM1, ALDH1A1, and CPS1 corelated to APC mutations. Additionally, HOXC5 and LMO3 were upregulated corelated to ZNHIT6 mutations. Conversely, downregulation was observed in AKR1B10 and AKR1C2 corelated to POLE mutations, HOMER2 and FGF19 corelated to APC mutations, and IHH, UGT2B7, and OGDHL due to ZNHIT6 mutations. Furthermore, a differential expression analysis conducted in UCEC, STAD, and CRC identified MUC5AC and CXCL5 as genes involved in inflammation and apoptosis. Pathway analysis revealed their potential connection to the IL-17 signaling pathway and immune response. Therefore, this study suggests that the upregulation and downregulation of common genes in CRC and UCEC, along with the interaction of MUC5AC and CXCL5 with the IL-17 pathway, could serve as promising biomarkers for immunotherapy.

CHAPTER 01

INTRODUCTION

1.1 Cancer

Cancer, an intricate and formidable illness with far-reaching impact on millions of people worldwide. It is defined as uncontrolled division of body cells. According to the National Cancer Institute it is a disease in which some body cells grow uncontrollably and could spread to other body parts. (What Is Cancer? ,2021) Cell, the function unit of life, makes up whole human body. The human body is composed of trillions of cells. These cells have remarkable ability to undergo cell division process, in which cells grow and divide. The human body transmits special signals to genes residing inside cell nucleus for cell division initiation. These genes serve as regulators for cell division to produce new cells, as needed by body for growth, tissue repair or renewal. The newly produce cells are crucial for proper functioning of human body and multiple ongoing processes. At the same time these cells could be damaged by oxygen deprivation, genetic mutations, aging and physical, chemical, infectious agents etc. These newly synthesized cells replace damage or aged cells. Damaged cells are detected and controlled by immune system and have two potential outcomes, they either undergo necrosis, accidental cell death, or apoptosis, planned cell death. This process represents natural phenomena of life. Any disruption in this well-organized process could lead to uncontrolled division of damaged or abnormal cells resulting in cancer.

WHO (World Health Organization) defines cancer as rapid creation of abnormal cells beyond their boundary and forms a mass called **tumor**. Tumor could be **benign**, non-cancerous that grows slowly and does not spread to other parts of the body, but when it invades adjoining body parts and spread to other organs, through metastasis, such tumor is **malignant tumor**.

1.1.1 Causes/ Risk Factors of Cancer

Cancer is the leading cause of the death worldwide. According to the latest WHO statistics, nearly 10 million deaths due to cancer were reported worldwide in 2020. Whereas, only in Pakistan 117149 deaths and 178388 new cancers were reported. Cancer arises through a multistage process of transformation of normal cells into tumor cells. An individual's

genetic factors interactions with three external factors, mentioned below could lead to cancer.

- **Biological Carcinogens:** Such biological agents that have potential to cause cancer include any
    1. **Oncogenic viruses** that can integrate their DNA into host DNA and cause cancer, like *Hepatitis* virus. (Wild & Montesano, 2009)
    2. **Bacterial infections** are also associated with high risk of certain types of cancer. For example, *Helicobacter pylori (H. pylori)* is reported to be associated with gastric cancer. (Mohabati Mobarez et al., 2020)
    3. **Parasitic infections**, similar to bacterial infections are also associated with certain cancer development. For example, *Schistosoma haematobium* link to bladder cancer is reported in Middle east. (Mohammed et al., 2023) And *Opisthorchis viverrini* and *Clonorchis sinensis* association with cholangiocarcinoma is reported in Southeast Asia. (Herrick, 2023)
- **Physical Carcinogens:** As indicated by the name, they cause cancer through physical exposure or processes by directly damaging DNA or disrupting cellular processes.
    1. **Ionizing radiation** are high energy waves that are capable of removing tightly bound electrons of an atom and form ions. X-rays, gamma rays and radioactive material radiations are its examples. Prolong exposure to these radiations can cause certain solid tumors and leukemia. (Abalo et al., 2020; Mazzei-Abba et al., 2021)
    2. **Ultraviolent radiations** are electromagnetic radiation emitted by sun, prolong exposure to these can cause DNA damage in skin leading to different types of skin cancers, such as basal cell carcinoma, squamous cell carcinoma, melanoma etc. (Chan et al., 2019; Teng et al., 2021; Autier & Doré, 2020)
- **Chemical Carcinogens:** These substances can induce genetic alterations after interactions with DNA. These substances could be found in various environment, occupational or personal life style.
    1. **Tobacco smoke** contain numerous chemicals carcinogens such as polycyclic aromatic hydrocarbons (PAHs), nitrosamines, and aromatic amines. Smoking is the

major risk for lung, mouth, throat, esophagus, and bladder cancers. (Parascandola & Xiao, 2019; Lewis, 2020)

2. **Alcohol** consumption could also lead to cancers like mouth, throat, esophagus, liver, breast, and colon. (Rumgay et al., 2021)
3. **Asbestos** is the naturally occurring minerals found in the construction materials. Long term inhalation of asbestos could lead to lung cancer or other respiratory diseases.
4. **Aflatoxins** are toxins that are involved in food contamination produced by certain molds. These toxins are linked with high liver cancer risk. (Dabuo B., et al., 2022)

It is not necessary that all individuals exposed to these carcinogens may develop cancer. Complex interactions, intensity, exposure, individual susceptibility, environmental factors and lifestyle contribute equally along with these carcinogens.

1.2 Cancer Genetics

Cancer is a complex disease, caused by combination of multiple factors. Clear cause for all cancers is not known. Whereas, DNA sequence mutations are the fundamental cause of the cancer. Changes in DNA sequence of an individual usually do not affect the genetic mutations but some mutations can lead to genetic conditions. Among the multiple types of genetic mutations two main types are germline and somatic mutations.

Germline mutations, as suggested by their name, are heritable genetic alterations. These mutations originate in the very first cell, which subsequently divides and gives rise to a new individual. There exists a high likelihood that this mutation will be present in every cell of the body. These mutations are often referred to as hereditary mutations. In cases where germline mutations are associated with cancer, the risk of cancer development is notably heightened, particularly in the early stages of life. This is due to the fact that the defective gene carrying the hereditary mutation provides a predisposition for cancer initiation. When a gene alteration that elevates the risk of cancer is observed within a family, it is referred to as a family cancer syndrome. If multiple family members, especially from the same side of the family, have been afflicted with either rare or common forms of cancer, coupled with exposure to certain risk factors, it raises concerns for a potential

family cancer syndrome. Notable examples of cancers resulting from germline mutations include Lynch syndrome (also known as hereditary non-polyposis colorectal cancer), which is caused by damage to any of the MMR genes including MLH1, MSH2, MSH6, PMS2, and EPCAM, as well as Hereditary Breast and Ovarian Cancer (HBOC) syndrome, arising from mutations in the BRCA1 or BRCA2 genes.

Somatic or sporadic mutations are genetic changes that are not passed on to future generations. These mutations can occur at any stage of life. They originate in a single cell and can subsequently propagate to other cells. Factors such as exposure to ultraviolet (UV) radiation, chemical carcinogens, errors during DNA replication, and malfunctioning DNA repair mechanisms can contribute to the occurrence of these mutations. Skin cancer and lung cancer are examples of cancers resulting from somatic mutations, where mutations in specific genes relevant to these cancers play a significant role.

Not all mutations in genes, nor all types of mutations, result in cancer. Among these, somatic mutations are a more prevalent cause of cancer compared to germline mutations. This observation has been substantiated by various studies. For instance, a study conducted in 2021 by Zhaopei Li and colleagues examined the profiles of somatic and germline mutations in cancer. The findings revealed that the risk associated with germline mutations is around 9%, while somatic mutation risk stands at 73.1% within genes (Li et al., 2021).

Certain genes are susceptible to mutations that can lead to the development of cancer, and specific types of mutations have been implicated in this process. Understanding these genes and mutation types is vital in comprehending cancer's genetic basis and its various manifestations.

1.2.1 Genes Mutations linked to Cancer

Cancer is a genetic disease. It involves alterations or mutations in DNA of genes responsible for cell growth and division. For any healthy cell to turn cancerous, it requires more than one DNA change. Mutation in particular genes could lead to cancer including tumor suppressor genes, proto-onco genes and DNA- repair genes. (Smithgall et al., 2022)

- **Tumor suppresser genes** plays an important role in cell growth and differentiation and also inhibit cancer development. They work as a part of complex network of cellular processed to maintain genetic stability and control cell proliferation. Tumor suppresser genes are activated when cells are exposed to cancer causing factors, they identify damaged DNA and eliminate it before they have chance to divide by triggering apoptosis. P53, BRCA1, BRCA2, TP53 are most common reported tumor suppresser genes. (Zacharias, 2020)

- **Proto-onco genes** are the normal cellular genes involved in regulating normal cell growth, division and apoptosis but once mutated they become onco genes, cancer causing genes. Oncogenes lead to uncontrolled cell division, and are smart enough to escape cell death, leading to cancer. (Oncogene, 2023; Zacharias, 2020)

- **DNA repair genes** are of same importance as tumor suppresser genes. These genes identify and repair damages in DNA that is crucial for accurate DNA replication. They also have role in disease prevention as they do not let the mutation to accumulate. Hence, decreasing risk of cancer development. MMR (Mismatch Repair) a DNA repair pathway, recognizes error in DNA during synthesis of new cells and if match is not perfect it repairs it. In MMR pathway, MLH1, MSH2, MSH3, MSH6 genes have considerable importance in cancer development. (Lahtz & Pfeifer, 2011; Genes and Cancer, 2012)

1.2.2 Not all Mutations lead to Cancer

Every mutation has varying effect on gene functionality and protein production. Following are some mutation types that occur and contribute to cancer development.

- **Missense Mutation:** In protein sequence, s single nucleotide substitution takes place, whereas, the impact of the mutation depends on the location, and amino acid change. Most common missense reported mutation is of BRCA1, that hinder tumor suppresser gene function and escalates breast cancer risk. (Aljarf et al., 2022)

- **Nonsense Mutation:** In DNA sequence, nonsense mutation gives sudden rise to stop codon and as result protein is truncated. The translated protein is either short or non-

functional. P53 is crucial tumor suppresser genes, nonsense mutation in this gene results in non-functional P53 protein. (Wu et al., 2022)

- **Frameshift Mutation:** It involves insertion or deletion of one or more nucleotide and disturbs the normal reading frame of DNA sequence. Afterwards the synthesized protein has altered or no function. APC gene tumor suppresser functionality can be inactivated because of frameshift mutation in gene and develop CRC (Colorectal Cancer). (Aghabozorgi et al., 2019)

- **Duplication Mutation:** DNA segment is replicated, leading to increased number of gene copies. Duplication mutation of HER2 leads to its amplification, resulting in its overexpression and promoting aggressive tumor growth in breast cancer. (Yi et al., 2020)

- **Indel Mutation:** "Insertion and Deletion Mutation". In insertion mutation number of DNA bases is altered due to insertion of new DNA and deletion on the other hand is the removal of the DNA. The scale of these mutations could be small or massive and corresponding to that protein function could be disrupted. TP53 gene is involved in cell cycle regulation, gene repair and apoptosis. Indel mutations in TP53 are reported that result in loss of ability to control cell growth and division that leads to cancer. (Gurney et al., 2023)

1.2.3 Hotspot Mutation

Hotspot mutations are specific genetic alterations that occur at regions within the DNA known as "hotspots." These hotspots are positions in the DNA sequence where mutations happen at a notably high frequency. These mutations can be prompted by various factors, including genetic modifications and exposure to substances that can cause cancer. Hotspot mutations are influenced by the sequence and structure of DNA and can be advantageous in evolutionary processes. A study in 2017 proposed that a position could be considered a hotspot if it experiences at least two mutations of the same type. For example, the HTT gene was emphasized in a 2021 study, though its analysis dates back to 1983 when expandable CAG tracts were identified. These tracts were found to lead to nonfunctional

protein production and were designated as hotspot regions across the genome. Hotspots can significantly elevate the rate of DNA mutation, occasionally up to around 100 times more than the average rate.

The most prevalent genetic variations in the human genome are single nucleotide variants (SNVs), which account for roughly 80% of differences among individuals. SNVs can be categorized as transversions or transitions. However, mutations occurring in hotspot regions can encompass various types, including substitutions or insertions/deletions (indels). Numerous documented hotspot mutations are situated in CpG islands, with C to T transitions commonly occurring in germline variants. These CpG-rich sites, mostly located upstream of genes, represent about 1.5% of the human genome. Conversely, somatic hotspot mutations are driven by processes like activation-induced cytidine deaminase (AID) or apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC)-mediated deamination. When these mutations impact regulatory genes, they can lead to the generation of altered proteins or disrupted regulatory elements. Microsatellites, are also prone to mutations, making them hotspots in both germline and somatic events. Genes containing microsatellites can serve as hotspots in various diseases, including cancer. A comprehensive investigation into APOBEC's role in tumor evolution suggested that specific recurrent hotspot APOBEC mutations outside hairpin structures of tumor suppressor genes might contribute to tumor evolution, whereas other mutations might lack significant functional impact. Some of these hotspot mutations propel cancer development, exemplified by POLE P286R, POLE V411L, and PIK3CA R88Q. In conclusion, hotspot mutations play pivotal roles in shaping genetic diversity, fostering disease development, and influencing evolution. (Rogozin & Pavlov, 2003; Baeissa et al., 2017; Nesta et al., 2021)

These are just examples of few mutations types whose occurrence can contribute to cancer.

1.3 Types

In accordance with where it started, the primary site, there are four main forms of cancer, described below;

- **Carcinomas** originate in the epithelial cell lining, present at the surface and organs of the body. Epithelial cells are found in skin, lungs, breast, colon, and many other organs, upon cancer development they usually form solid tumors. Carcinomas are the most common type of cancer worldwide. Carcinomas include different skin cancers, breast cancer, CRC, lung, prostate cancer, etc.

- **Sarcomas** are comparatively less common than carcinomas. They arise from the connective tissues that support the body, such as muscle, fat, bone, nerves, tendons, joints, blood vessels, lymph vessels or cartilage. Sarcomas include bone cancers. (Tanaka & Nakamura, 2021)

- **Leukemias** is blood or bone marrow cancer, characterized by the excess WBCs (White Blood Cells) production in body. Acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML) are the four main types of leukemias. ALL originates in early lymphocytes and is more common in children than adults. CLL originates in mature lymphocytes, it progresses slowly and can be treated by "wait and watch" approach. It is common in adults. AML originates in myeloid cells. It is common in old adults and requires immediate treatment. CML is identified by the presence of a genetic order called Philadelphia chromosome, it affects myeloid cells. Initially it progresses slowly, but later on it can transform into blast crises (aggressive phase). (Burke & Startzell, 2008)

- **Lymphomas**, as indicated by name, is cancer of lymphatic system, which helps fight infections as it is the part of the immune system. These cancers develop in the lymph nodes, spleen, thymus, bone marrow, and other lymphatic tissues. Lymphomas are divided into Hodgkin lymphoma and non-Hodgkin lymphoma. Hodgkin lymphoma (HL) or Hodgkin disease is a rare lymphoma, in which large abnormal cells called *Reed-Sternberg cells* are found in lymph nodes and it can spread to nearby organs. (Küppers et al., 2012) On the other hand, non-Hodgkin lymphoma comprise of lymphomas that originate from WBCs (lymphocytes). Its outspread tendency is higher than HL. (Shankland et al., 2012; Nayak & Deschler, 2003)

Each of the cancer type is further divided into subtypes as mentioned above.

1.3.1 Subtypes

Cancer types are further subdivided based on cell type, growth pattern and genetic characteristics of cancer. Better understanding of originating points or subtypes of cancer would give better insight to its diagnosis, treatment, prevention etc.

Below discussed are some subtypes of the cancer used in my research.

- **Uterine Corpus Endometrial Carcinoma (UCEC):** It is the fourth most common reported cancer in females in USA. According to the 2013 study on integrated genomic characterization of UCEC by Douglas A. Levine, estimated 49,500 new cases and 8,200 deaths were reported in 2013. Not only in USA it is common gynecologic malignancy worldwide. Over 500,00 die due to UCEC yearly. (Pan et al., 2019) Epidemiological data have revealed increase in UCEC incidence globally in last two decades. Approximately 1.9% incidence increase is reported every year. (Hu et al., 2021) UCEC is the uterus lining (endometrium) cancer with low survival rate and poor prognosis. (Levine, 2013; Zhao & Li, 2023) It is common in post-menopausal women, due to high estrogen levels. Whereas, clinically UCEC is heterozygous and its heterozygosity is linked to tumor growth, clinical traits, and prognostic outcomes. (Hou et al., 2020) Fortunately, it can be detected at early stage with >90%, 5-year overall survival (OS) rate. But unfortunately, advance or recurrent UCEC have poor prognosis with 5-year OS rate >30%. Risk factors reported by different studies for generation and development include obesity, drinking, smoking and high blood pressure. Some studies also reported molecular changes or genetic alterations as a contributor to development and poor prognosis of UCEC. UCEC can be further divided into Type I and Type II UCEC. *Type I* is estrogenic-dependent UCEC with good prognosis and *Type II* is non-estrogen dependent UCEC with poor prognosis. (Zhao & Li, 2023) For early detected cases traditional treatment methods that include chemotherapy, radiotherapy, surgery and hormone therapy could be effective but for metastasized cases surgery and radiotherapy could not achieve satisfactory results (Zhao et al., 2021). In TCGA data,

the majority of hypermutated MSS samples, comprising 52.71%, were identified as UCEC samples.

- **Colorectal Cancer (CRC):** CRC comprises of colon and rectum cancer and is the second most deadly cancer after breast cancer. Yearly incidence and mortality of CRC is expected to rise. It is expected to be doubled by 2035. In 2020, 9.4% out all cancer-related deaths were due to CRC. CRC develops when epithelial cells undergo genetic or epigenetic changes that cause hyper-proliferation and form growths in the inner lining of colon and rectum, these growths are called polyps (adenoma), that may or may not develop into cancer over the period of time. CRC can progress through microsatellite instability (MSI), chromosomal instability (CIN), and serrated neoplasia pathways. Both modifiable and nonmodifiable risk factors influence CRC development. Modifiable risk factors include lifestyle and individual habits. On the other hand, non-modifiable risk factors include personal history, family history, inflammatory bowel disease history, adenomatous polyp' history. Some studies reported that the family history of CRC plays a significant role in increasing the risk. CRC deaths are more prevalent in high-income countries compared to developing countries. Although access to early detection and treatment is better in developed countries but recurrence rate is high and drug resistance is a hurdle in treatment. For early detected CRC at 0-II stage surgery is standard treatment option, for stage III surgery and adjuvant chemotherapy and for stage IV surgery, chemotherapy, and targeted therapy are used but unfortunately there is no known established cure (Hossain et al., 2022). Within TCGA data, 30.95% of hypermutated MSS samples were categorized as COAD.

- **Stomach Adenocarcinoma (STAD):** It is ranked as fifth most common whereas, third leading cause of cancer related mortalities worldwide. It is heterozygous cancer with highest incidence rate in Asia, specifically in China but 5-year OS rate is also higher that European countries i.e., 30-36%. OS rate could improve if early detected and treated. Development of STAD initiates when chronic inflammation of stomach lining takes place, often caused by *Helicobacter pylori (H. pylori)* or other risk factors like, alcohol consumption, smoking, obesity, family history. Further, these chronic

inflammations to formation of precancerous lesions, which are not cancerous but over the time accumulation of genetic alterations in precancerous lesions cells can transform it into cancerous cells. Due to this genetic instability, cells proliferate uncontrollably and metastasis. Around 90-95% of all stomach cancers are adenocarcinomas. Early detection based on its location it is further subdivided into *cardia*, located near gastro-esophageal junction and *non-cardia*, located away from gastro-esophageal junction. Treatment methods are similar as that of other cancers, surgeries for early detected and different therapies for later stages (Ren et al., 2020; Zhou et al., 2020). In the TCGA dataset, 7.14% of the samples were hypermutated MSS samples in STAD.

- **Pancreatic Adenocarcinoma (PAAD):** It is reported as the second most common gastrointestinal malignancy in USA with poor prognosis. In 2018, approximately 53,000 people were diagnosed with PAAD. Its incidence rate is also low comparatively but still it is the fourth leading cause of death. Development of PAAD takes place through genetic alterations in critical genes like KRAS, TP53, and CDKN2A, the cell undergoes abnormal growth leading to Pancreatic Intraepithelial Neoplasia (PanIN), which are early precancerous lesions that develop in the ducts of the pancreas with malignancy potential. Afterwards, some PanIN may transform into invasive carcinoma. Other than obesity, smoking, alcohol consumption, diabetes type 2 is suspected to be a risk factor in PAAD development, although exact mechanism is not clear yet. Like other cancers early detection and molecular mechanism comprehension are vital for advancing treatments. Only 10-15% of affected individuals are candidates for surgery due to advancing of disease without detection. Other than neoadjuvant, to optimize surgical recursion and adjuvant therapies, preventing recurrence, chemotherapy is used for PAAD treatment (Vareedayah AA, et al., 2018). Hypermutated MSS samples from PAAD accounted for 2.38% of the TCGA dataset.

- **Uterine Carcinosarcoma (UCS):** It is a rare and most aggressive subtype of cancer, also known as malignant mixed Müllerian tumor. In 2020, particularly in the USAtotal of 1.8 million new cases in the USA were reported out of which 417,367 were uterine malignancies. And UCS comprises of less than 5% of uterine malignancies. It is considered as subtype of endometrial cancer but unlike endometrial malignancy it

contains epithelial (carcinomatous) and mesenchymal (sarcomatous) components in tumor. Due to this unique behavior UCS tumors are known as biphasic tumors, as upon visualization under microscope one component may appear of mesenchymal and other of epithelial cells. Studies reveal that mutations in key regulatory genes such as PTEN and TP53 genes contribute to UCS tumor development. Because of the extremely aggressive behavior of the UCS, distant metastasis patients have a high recurrence and less survival rate. The incidence rate is 30-39%. Risk factors that may contribute to UCS development are not only genetic and environmental but also hormonal. Hormonal imbalance, especially long-term exposure of estrogen without counterbalance of progestogen can contribute to UCS particularly in postmenopausal women. Other risk factors are same as of other cancers. Surgery, including hysterectomy bilateral salpingo-oophorectomy and pelvic lymph node dissection is used as the primary treatment for UCS (Moukarzel et al., 2021; Pezzicoli et al., 2021). Similar to PAAD, UCS hypermutated MSS samples also constituted 2.38% of the total TCGA dataset.

1.4 Hypermutation

Hypermutation is a characteristic of tumors with **deficient mismatch repair (dMMR)** and is often associated with high mutation burden, observed in certain cancer types. The term "hypermutated tumors" refers to tumors that exhibit a significantly elevated number of mutations compared to normal cells (Molina et al., 2022). In the context of molecular subtyping, hypermutation can be further classified into two categories: **microsatellite instability (MSI)** and **microsatellite stability (MSS)** (Huang et al., 2015). Mismatch repair (MMR) plays a crucial role in maintaining genomic stability and is a prime cause of cancer, leading to microsatellite instability (MSI) (Huang et al., 2015).

In the early 1990s, while working on hereditary non-polyposis colorectal cancer (HNPCC), the nature of the HNPCC mapped genes was not known. However, based on observed clonal shifts in microsatellite allele sizes, occur approximately once per 100,000 base pairs in genomic DNA across the entire genome, researchers identified this as potential mechanisms that could lead to tumor development (Aaltonen et al., 1993).

Later, in the late 1990s, an international workshop on Microsatellite Instability and RER phenotypes in Cancer Detection and Familial Predisposition was held. During this workshop, it was announced that the form of genomic instability associated with defective DNA mismatch repair in tumors would be referred to as Microsatellite Instability (MSI). The workshop also proposed the use of five markers to classify tumors based on the frequency of instability. (Rodriguez-Bigas et al., 1997).

In addition to dysfunction in the mismatch repair system, other endogenous and exogenous mutagenic processes, such as ultraviolet (UV) radiations, tobacco smoke, MSI, excessive APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) activity, as well as mutations in the POLD1 and POLE genes, contribute to hypermutation. The occurrence of hypermutation can vary depending on the specific cancer type and sequencing techniques employed (Huang et al., 2015).

1.4.1 Microsatellite Instable (MSI)

Somatic or germline indel mutations in microsatellites, caused by malfunctioning DNA mismatch repair (MMR) or *MLH1* hypermethylation, are responsible for microsatellite instability (MSI) (Yuza et al., 2017; Ballhausen et al., 2019; Huang et al., 2015). Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), is characterized by germline mutations in MMR genes, particularly *MLH1, MSH2, MSH6*, and *PMS2* (Nature, 2012; Barbari and Shcherbakova, 2017).

Following the first reported case of MSI in 1996, an international conference on MSI was held, leading to the proposal of the Bethesda panel as a reference panel for future research (Boland et al., 1998). The Bethesda panel consists of five microsatellite loci: two mononucleotide markers (BAT25 and BAT26) and three dinucleotide markers (D5s346, D2s123, and D17s250) (Huang et al., 2015). Instability in any of these loci is classified as **microsatellite instability-low (MSI-L)**, while instability in two or more loci is classified as **microsatellite instability-high (MSI-H)** (Huang et al., 2015).

There are seven known MMR genes associated with MSI: MLH1, MSH2, MSH6, PMS2, EPCAM, APC, and MUTYH (Bonneville et al., 2020). Lynch syndrome, characterized by

deleterious germline mutations in MMR genes, primarily involves MLH1, MSH2, MSH6, and PMS2. Deletion of the EPCAM gene is associated with upstream MSH2 in MSI-H (Kim et al., 2014). In contrast, APC gene mutations are less frequently found in MSI-H tumors but are often associated with tumor budding, which is caused by beta-catenin overexpression and MSH1 hypermethylation (Bonneville et al., 2020). The MUTYH gene encodes the MYH glycosylase enzyme involved in the Base Excision Repair (BER) system (Bonneville et al., 2020).

MLH1 hypermethylation is considered the predominant cause of MSI, based on available literature (Nojadeh et al., 2018). The percentage of MSI varies among different cancer types, with approximately 15% in colorectal cancer, up to 30% in endometrial cancer, 10-22% in gastric cancer, and 10% in intrahepatic cholangiocarcinoma (Silva et al., 2016; Ballhausen et al., 2019; Kim et al., 2013).

1.4.2 Microsatellite Stable (MSS)

On the other hand, microsatellite stability (MSS) is typically associated with somatic or germline mutations in the DNA polymerase epsilon (POLE) gene (Nature, 2012; Barbari and Shcherbakova, 2017). POLE belongs to the DNA polymerase B family, which includes polymerase epsilon (Pol ε) and polymerase delta (Pol δ). These polymerases play a crucial role in DNA replication by accurately incorporating complementary nucleotides and proofreading the newly synthesized strands (Korona et al., 2010). They also participate in nucleotide excision repair and double-strand break repair processes. The exonuclease domain of POLE/POLD1 acts as a proofreader, identifying and removing incorrect bases during replication. Mutations in the exonuclease domain of POLE/POLD1 (POLE/POLD1-EDM) result in a loss of proofreading function, leading to the accumulation of mutations in cells (Loeb & Monnat, 2008; Ma et al., 2022). In Bethesda panel, tumors that showed no significant instability were classified as Microsatellite Stable (MSS), although using a larger number of markers can help differentiate between MSI-L and MSS

Studies have identified specific somatic and germline mutations in POLE. For instance, the somatic POLE P286R mutation was reported as a driver mutation for early-onset colorectal

cancer (CRC), causing the rapid accumulation of additional somatic mutations and accelerating cancer development (Ahn et al., 2016). High-frequency somatic mutations such as P286R and V411L, as well as low-frequency germline mutations like P286H/S, F367S, S459F, and V424L, have been observed in POLE (Nebot- Bral et al., 2017). Other pathogenic factors involved in cancer include somatic point mutations in the polymerase domain of PolE (R567C, K593C, S595P, E611K, and L621F) and frameshift deletion mutations in the middle region of the PolE gene (V1446fs *3) (Meng et al., 2020; Min et al., 2020). These mutations contribute to the development of cancer. The POLE P286R mutation, validated in a yeast model, exhibits a mutator phenotype similar to complete mismatch repair deficiency, further supporting its role in driving hypermutation (Kane & Shcherbakova, 2014).

1.5 Treatment

Cancer has become a major global health issue, causing a significant number of fatalities worldwide. Consequently, there is a growing emphasis on the importance of cancer prevention and treatment as crucial components of public health. In 2020, Cancer Research UK reported the identification of more than 200 different types of cancer, while the World Health Organization (WHO) revealed that approximately 10 million people lost their lives to cancer in the same year, accounting for one in every six deaths (Cancer, 2023). Over the past decade, a notable trend in cancer research has been the recognition of hypermutation as an emerging type of cancer. This phenomenon has been investigated through whole exome sequencing (WES), which has provided valuable insights into colorectal, endometrial, and lung cancers. The analysis revealed high mutation rates, exceeding 12 mutations per million bases (Mb), over 18 mutations per Mb, and at least 178 nonsynonymous mutations per tumor, classifying them as hypermutated (Yuza et al., 2017). The Cancer Genome Atlas (TCGA) has also identified cancers with mutation rates surpassing 12 mutations per one million bases as hypermutated (Huang et al., 2015).

While traditional cancer treatment methods such as surgery, radiotherapy, and chemotherapy have been developed and proven effective, they are not without side effects. Recent advancements in our understanding of the tumor microenvironment and tumor

immune mechanisms have led to the emergence of tumor immunotherapy as a novel and promising approach to cancer treatment (Disis, 2014).

1.5.1 Immunotherapy

Immunotherapy controls the power of the immune system to generate tumor-specific immune responses that can control tumor growth and eliminate cancer cells (Zhang et al., 2021). T cells are specialized to identify and attack foreign invaders. In cancer cells, neoantigens are the unique proteins found on cancer cell surfaces as result of mutations. They are novel proteins that distinguish cancer cells from normal. In immunotherapy these neoantigens are recognized as foreign bodies by our immune system and eliminated. However, cancer cells evade detection by immune system as neoantigens are not recognized. To overcome this evading one significant breakthrough in immunotherapy is the discovery of T-cell immune checkpoint inhibitors (ICIs) or inhibitory receptors, such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) (Esfahani et al., 2020). These inhibitors, when used in combination, block the PD-1/PD-L1 pathway and enhance antitumor immunity by activating CD8-positive T cells in cancer tissues, thus promoting their infiltration into tumors (Kudo, 2019). Tumor immunotherapy, particularly through the use of ICIs, has demonstrated significant improvements in the long-term survival of patients with advanced solid tumors, including lung cancer, melanoma, colorectal cancer, and esophageal cancer (Disis, 2014). PD-1 antibody treatment has shown particularly promising results. It has been associated with sustained remission even after discontinuing treatment and higher overall survival rates compared to traditional chemotherapy and targeted therapy (Disis, 2014; Zhang et al., 2021). However, it is important to note that the efficacy of ICIs remains limited to a subset of patients who positively respond to the treatment (Esfahani et al., 2020).

Immunotherapy works notably better in the MSI-H subtype of hypermutation compared to MSS due to distinct immune response differences. This is mainly because MSI-H patients have more neoantigens – special markers resulting from high mutation rates. These neoantigens are like flags that the immune system can recognize, making MSI-H tumors more vulnerable to immune attack. Additionally, the increased mutation rate in MSI-H tumors boosts the chances of presenting these neoantigens to immune cells. Immune

checkpoint inhibitor (ICI) drugs also excel in MSI-H cases because there are more antigens for them to target effectively. In simple terms, the combination of more neoantigens and effective ICIs makes immunotherapy a potent strategy for MSI-H tumors. (Nebot-Bral et al., 2019)

To optimize immunotherapy outcomes, there is a growing need to identify reliable molecular markers that can predict patient response to ICI treatment. This would enable the identification of individuals who would benefit the most from immunotherapy while reducing overtreatment in those who are less likely to respond (Esfahani et al., 2020). Additionally, in the realm of tumor immunotherapy, a noteworthy approach involves HLA (human leukocyte antigen)-dependent therapies. HLA molecules play a pivotal role in immune responses by presenting antigens to T cells. Thus, utilizing HLA-dependent strategies can hold potential for augmenting immune responses against tumors.

1.5.1.1 Human leukocyte antigen (HLA)

Human leukocyte antigen (HLA) molecules, also known as major histocompatibility complex (MHC) molecules, are vital for immune response as they present antigens to T cells (Mungall et al., 2019). HLA includes diverse types like HLA class I (e.g., HLA-A, HLA-B, HLA-C) presenting intracellular antigens to CD8+ T cells, and HLA class II (e.g., HLA-DR, HLA-DP, HLA-DQ) presenting extracellular antigens to CD4+ T cells (Petersdorf et al., 2020).

HLA typing, identifying specific alleles, is integral in predicting immunotherapy response. It's linked to T cell activity, immune checkpoint blockade effectiveness, and clinical outcomes. For example, certain HLA class I alleles were associated with better responses to immune checkpoint inhibitors in colorectal cancer (Le et al., 2015). Similar associations have been observed in melanoma and other cancers, emphasizing the significance of tailoring immunotherapy based on HLA genotypes (Chowell et al., 2018; Bell and Ellenson, 2019).

1.6 Cancer Research Tools and Databases

Cancer research is necessary for decoding the complexities of diagnosis, refining treatments and enhancing outcomes of disease itself. By understanding the genetic and molecular underpinnings of diverse cancers, new avenues for early detection and the development of personalized therapies are being paved, resulting in an improved survival rate. Tools and databases play a vital role in accomplishing these tasks by offering organized data in a user-friendly format for easy exploration. Following are some tools and databases that will be further used in this research:

1.6.1 The Cancer Genome Atlas Program (TCGA): It is a comprehensive and collaborative initiative by National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) to uncover genomic alteration in cancer. A large amount of genomic, epigenomic, transcriptomic and genomic data is not only generated by TCGA, but it is also freely available to scientific community around the world. TCGA data proves highly beneficial in comprehending cancer genomics and discerning multiple biomarkers for cancer therapies. TCGA store data for 33 cancer types that is accessible through Genomic Data Common (GDC) Data Portal. GDC data portal is a user-friendly interface to access TCGA collected data for analyzation and interpretation. GDC serves as central platform for storing, starting and distributing TCGA and its collaborative projects data.

1.6.2 cBioPortal: Another freely accessible online cancer genomics database, this platform offers a user-friendly interface for accessing and analyzing data. Moreover, it facilitates the interactive analysis of complex bulk data by enabling the creation of graphs and plots based on user-defined parameters. The database contains comprehensive genomic information, encompassing details such as genetic mutations, gene expression, protein abundance, and clinical outcomes. It not only provides open datasets for free access but also permits users to upload their own datasets.

1.6.3 Database for Annotation, Visualization and Integrated Discovery (DAVID):

It is a bioinformatics tool designed for the interpretation and analysis of large datasets. It serves as a pathway analysis tool, offering several functionalities such as gene ontology analysis, pathway analysis, functional annotation clustering, protein-protein interaction analysis, and disease association analysis. DAVID facilitates a better understanding of genes, proteins, and their interactions within high-throughput datasets.

1.6.4 Empirical Analysis of Digital Gene Expression Data in R (EdgeR):

It is a R computational language package extensively used for differential gene expression analysis in high-throughput data. It basically performs statistical analysis through various functions to identify differentially expressed genes. Some of the common statistical test conducted by edgeR are Negative Binomial Generalized Linear Model (GLM), distributes the normal read counts and incorporates experimental conditions, sample groups and potential sources of variation to differentiate biological changes from fluctuations. Afterwards, likelihood ratio test (LRT) is carried out to calculate p-value of each associated gene to check the significance of observed changes in gene expression. It also adjusts p-value to make sure identified genes are not by chance. It is a robust, differential gene expression analysis framework.

1.7 Problem Statement

Cancer's multifaceted nature has a global impact, leading to a range of treatment approaches. In the last two decades immunotherapy has emerged as the most effective and minimal side effect cancer treatment. Immunotherapy activates our body natural defense system to identify the unique markers on cancer cells i.e., neoantigens and mark them as foreign bodies. This signals specialized T cells to eliminate them. It has proved effective in hypermutation subclass microsatellite instability-high, but its effectiveness is constrained in microsatellite stable subclass due to availability of low number of biomarkers. To overcome this problem, there is need to identify new biomarkers for better prognosis of hypermutated microsatellite stable cancer.

1.8 Objectives

Following are the objectives of this research:

- To identify novel biomarkers for hypermutated microsatellite stable cancers.
- To enhance prognosis of hypermutated microsatellite stable cancers through identified biomarkers.
- To propose optimizing immunotherapy effectiveness on hypermutated microsatellite stable cancers.

1.9 Scope

Cancer is a widespread prevalent disease, whose quest for effective treatment remains paramount. Immunotherapy has arisen with notable effectiveness in hypermutated microsatellite instable high patients, while when applied on hypermutated microsatellite stable patients it reveals critical obstacles due to lack of available biomarkers. To overcome this obstacle, we aim to identify new biomarkers that could lead to better outcomes and predictions of hypermutated microsatellite stable patients. It will make personalized treatment effective for more people and improve overall cancer care.

1.10 Limitations

Some limitations of this project are that cancer is a heterogeneous disease, resulting in diverse response to treatment. Like other treatments, some patients could be nonresponsive to immunotherapy. While reported success rate of immunotherapy is 20-50%. Whereas, notably the response rate of immunotherapy is higher in developed countries than in under developing countries, due to resources limitations. Developing countries face financial crises that hinder their population from testing and treatment. On the contrary, developed countries possesses better resources and provide better accessibility to treatments. These variations lead to regional limitations in treatment applicability.

CHAPTER 02

MATERIALS AND METHODS

Hypermutated MSS cancer study for the identification and characterization of novel biomarkers, that hold the potential to transform the panorama of immunotherapy, addressing its current limitations and effectiveness within hypermutated MSS.

For above mentioned purpose, TCGA datasets of 7 different cancers were used. Its transcriptomics analysis was performed on cBioPortal and gene differential analysis was performed using EdgeR.

2.1 Data Retrieval

TCGA has been a game-changer in our understanding of cancer. It has transformed treatment approaches and paved the way for advancements in technology and research. This extensive database represents over 12 years of hard work and includes data from more than 11,000 patients. It's one of the most used freely available cancer databases, contributing significantly to our understanding of cancer. TCGA data can be easily accessed through GDC by customizing search parameters to fit specific research needs.

While focusing on the identification of new biomarkers like POLE P286R in hypermutated Microsatellite Stable (MSS) cancers, in this study seven cancers based on their high hypermutation rates were selected. These cancers included Uterine Corpus Endometrial Carcinoma (UCEC), Colorectal Adenocarcinoma (COAD), Rectal Adenocarcinoma (READ), Stomach Adenocarcinoma (STAD), Uterine Carcinosarcomas (UCS), Pancreatic Adenocarcinoma (PAAD), and Esophageal cancer (ESCA).

In the Genomic Data Commons (GDC), the 'Repository' tab was utilized to store, manage, and analyze data. Further exploration involved two additional tabs named 'Files' and 'Cases,' from which specific parameters were configured. In the 'Data Category' section, 'Simple Nucleotide Variation (SNV)' was selected, as SNVs are crucial for studying biomarker identification. The next parameter set was 'Masked Somatic Mutation' from the 'Data Type' dropdown and for mRNA files ‘Transcriptomic Profiling’ was used . This data type is instrumental in identifying hidden somatic mutations. For data format, 'MAF' (Masked Annotation Format) was chosen, which contains detailed annotation data for each mutation. The data selected for this research was all openly accessible.

Moving on to the 'Cases' tab, 'TCGA' was chosen from the 'Program' selection because all the data used in this research originates from TCGA. From the 'Project' dropdown, all seven

selected cancer studies were chosen, namely TCGA UCEC, TCGA COAD, TCGA READ, TCGA STAD, TCGA UCS, TCGA PAAD, and TCGA ESCA. The remaining parameters were left at their default settings.

The number of files downloaded from the GDC v36.0 repository for the selected seven cancer subtypes totaled 1,993.

Table 1: Number of files of each cancer subtype downloaded from GDC of TCGA data.

Cancer Subtype	Total number of files downloaded for each subtype	Total number of samples	Total number of mutations reported	Total number of mRNA files
TCGA-UCEC	519	518	626,945	589
TCGA-COAD	457	455	252,615	524
TCGA-READ	161	161	57,699	177
TCGA-STAD	434	432	183,109	448
TCGA-PAAD	180	173	24,849	183
TCGA-UCS	57	57	8,774	57
TCGA-ESCA	185	185	29,760	198

The table above represents the total number of files downloaded for each cancer and within those files total number of samples studied and mutations reported in all samples of each subtype along with their mRNA files count.

2.2 Data Filtration

After downloading the data, the files for each cancer subtype were extracted and organized into separate folders. To further filter these files into categories like hypermutated, non-hypermutated, MSI (Microsatellite Instability), MSI-High (MSI-H), MSI-Low (MSI-L), and MSS (Microsatellite Stable), Perl scripts were employed.

Initially, to merge all the files for each cancer subtype, the following Linux command was used:

*awk 'FNR > 8' *.maf > merge.txt*

This command read all "*.maf*" extension files within the folder the command is executed and merge them in a text file named as merge.txt and it skips headers of all files

To generate a master table that consolidates all the cancer data and separate master tables for hypermutated and non-hypermutated data, the following Perl script was executed.

Perl Script to Generate a Master Table for All, Hypermutated, and Non-Hypermutated Cancer Data Separately.

```
#!/usr/bin/perl
use Cwd;

@flow = grep { -d } glob "*";
$arrSize = @flow;
for ($p = 0; $p < $arrSize; $p++)
{

	@files=<$flow[$p]\\*.maf>;
	print $flow[$p],"\n";
	foreach $file(@files)
	{
		open(x,$file);
		%hash={};
		%tested={};
		%Mut_Count={};
		$hyper=0;
		$nonhyper=0;
		$Total=0;
		$line = <x>;
		$line = <x>;
		$line = <x>;
		$line = <x>;
		$line = <x>;
		$line = <x>;
		while($line = <x>)
		{
			@temp=split("\t",$line);

	$key=$temp[4]."_".$temp[5]."_".$temp[6]."_".$temp[10]."_".$temp[11]."_".$temp
[12]."_".$temp[15];
			push @{$hash{$key}}, "0";
			push @{$Mut_Count{$temp[15]}}, "0";
		}

		@nm=split(/\./,$file);
		open(x,$file);
		open(Out,">Results\\".$nm[1]."_MasterTable.txt");
		open(Out1,">Results\\".$nm[1]."_Hyper.txt");
		open(Out2,">Results\\".$nm[1]."_Nonhyper.txt");
		$line = <x>;
		$line = <x>;
		$line = <x>;
		$line = <x>;
		$line = <x>;
```

```
@temp=split(" ",$line);
		@temp2=split(",",$temp[1]);
		print Out "Gene\tMutation\tType\tdbSNP\tStatus\tPolyphen\tSift";
		print Out1 "Gene\tMutation\tType\tdbSNP\tStatus\tPolyphen\tSift";
		print Out2 "Gene\tMutation\tType\tdbSNP\tStatus\tPolyphen\tSift";

for($i=0;$i<scalar(@temp2);$i++)
		{
			print Out "\t",$temp2[$i];
			$Total++;
			if(scalar(@{$Mut_Count{$temp2[$i]}})>499)
			{
				print Out1 "\t",$temp2[$i];
				$hyper++;
			}
			else
			{
				print Out2 "\t",$temp2[$i];
				$nonhyper++;
			}
		}
		print Out "\tHyper\tPer\tNonhyper\tPer\tTotal\tPercentage\n";
		print Out1 "\tHyper\tPer\tNonhyper\tPer\tTotal\tPercentage\n";
		print Out2 "\tHyper\tPer\tNonhyper\tPer\tTotal\tPercentage\n";
		$line = <x>;
		while($line = <x>)

		{
			$c=0;
			$hy=0;
			$nonhy=0;
			@temp=split("\t",$line);

	$key=$temp[4]."_".$temp[5]."_".$temp[6]."_".$temp[10]."_".$temp[11]."_".$tem
p[12];
			if(!exists $tested{$key})
			{
				push @{$tested{$key}}, "0";
				print                                                    Out
$temp[0],"\t",$key,"\t",$temp[8],"\t",$temp[13],"\t",$temp[25],"\t",$temp[72],"\t",
$temp[73];
				print                                                  Out1
$temp[0],"\t",$key,"\t",$temp[8],"\t",$temp[13],"\t",$temp[25],"\t",$temp[72],"\t",
$temp[73];
				print                                                  Out2
$temp[0],"\t",$key,"\t",$temp[8],"\t",$temp[13],"\t",$temp[25],"\t",$temp[72],"\t",
$temp[73];
				for($i=0;$i<scalar(@temp2);$i++)
				{

	$key=$temp[4]."_".$temp[5]."_".$temp[6]."_".$temp[10]."_".$temp[11]."_".$tem
p[12]."_".$temp2[$i];
```

```
if(exists $hash{$key})
                                    {
                                           print Out "\t1";
                                           $c++;
                                    }
                                    else
                                    {
                                           print Out "\t0";
                                    }

if(scalar(@{$Mut_Count{$temp2[$i]}})>499)
                                    {
                                           if(exists $hash{$key})
                                           {
                                                  print Out1 "\t1";
                                                  $hy++;
                                           }
                                           else
                                           {
                                                  print Out1 "\t0";
                                           }
                                    }
                                    else
                                    {
                                           if(exists $hash{$key})
                                           {
                                                  print Out2 "\t1";
                                                  $nonhy++;
                                           }
                                           else
                                           {
                                                  print Out2 "\t0";
                                           }
                                    }
                             }
                             if($hyper!=0 && $nonhyper!=0 && $Total!=0)
                             {
                                    print Out
"\t",$hy,"\t",($hy/$hyper)*100,"\t",$nonhy,"\t",($nonhy/$nonhyper)*100,"\t"
,$c,"\t",($c/$Total)*100,"\n";
                                    print Out1
"\t",$hy,"\t",($hy/$hyper)*100,"\t",$nonhy,"\t",($nonhy/$nonhyper)*100,"\t"
,$c,"\t",($c/$Total)*100,"\n";
                                    print Out2
"\t",$hy,"\t",($hy/$hyper)*100,"\t",$nonhy,"\t",($nonhy/$nonhyper)*100,"\t"
,$c,"\t",($c/$Total)*100,"\n";
                             }
```

```
else
                                {
                                        print Out "\t",$hy,"\t-\t",$nonhy,"\t-
\t",$c,"\t-\n";
                                        print Out1 "\t",$hy,"\t-\t",$nonhy,"\t-
\t",$c,"\t-\n";
                                        print Out2 "\t",$hy,"\t-\t",$nonhy,"\t-
\t",$c,"\t-\n";
                                }
                        }
                }

                open(Out3,">Results\\".$nm[1]."_Summary.txt");
                print Out3
"Total\t$Total\nHypermutated\t$hyper\nNonhypermutated\t$nonhyper\n";
        }
}
```

Master tables serve as better representatives for data integration, simplifying analysis, and facilitating comparative analysis. The generated master tables in cancer genomics provide detailed representation of each mutation. In the script provided above, the data is categorized into hypermutated and non-hypermutated based on the following criteria: mutations above 500 were characterized as hypermutated, while those below are considered non-hypermutated. It's important to note that MSS (Microsatellite Stable) samples are present in both hypermutated and non-hypermutated data. However, this research specifically focuses on hypermutated MSS samples. To achieve this, a similar Perl code was executed on the previously generated hypermutated master table to further create MSS, MSI-High (MSI-H), and MSI-Low (MSI-L) master tables. MSI-H samples were segregated based on a mutation rate higher than 500, while others were categorized as MSS, as indicated by a given reference file.

Master tables generated for each subtype were sorted in descending order to identify genes with the highest number of mutations. At the end of each master table, their mutation ratios were compared with other subtypes. In the hypermutated MSS master table, the sorted genes with the greatest number of mutations were positioned at the top. However, in comparison to MSS, their MSI-L and MSI-H percentages were higher, indicating a potential presence of microsatellite instability. To address this issue, a filter was applied to the MSS master table to shortlist specific genes.

- **PER_MSS >= 16** (to prioritize MSS samples)
- **PER_MSIL < 10** (to minimize the presence of MSI-L samples)
- **PER_MSIH < 1** (to ensure a very low percentage of MSI-H samples)

Genes were shortlisted based on previous studies and literature reviews. To identify hotspot mutations in genes of interest, we created mutation maps using the GDC data visualization tab. Mutation mappers are algorithms designed to detect genetic alterations within the genome. These genetic changes, once identified, may be associated with various diseases, making them crucial for in-depth study.

2.3 cBioPortal Grouping

cBioPortal is a valuable, freely accessible platform for conducting transcriptomic analysis. This analysis involves the examination of gene expression patterns, specifically focusing on mRNA levels within cancerous cells. To investigate gene upregulation and downregulation across seven selected cancer subtypes, data from cBioPortal datasets was utilized. The next step involved performing a comparative analysis, which required the grouping of data. The groups were defined as follows:

- **Group 1** consisted of MSS samples with a specific hotspot mutation in a particular gene.
- **Group 2** comprised MSS samples without that mutation.
- **Group 3** included all MSS non-hypermutated samples.

Transcriptomic analysis was conducted to compare Group-1 vs. Group-2 and Group-1 vs. Group-3. This grouping and analysis focused on two primary cancer subtypes, UCEC and CRC. Upregulated and downregulated genes were distinguished based on their **log2ratio** values. Log2ratio is a mathematical representation of the fold change in gene expression levels between two conditions or groups. Positive values signify upregulation, while negative values indicate downregulation. In this study, the criteria for categorizing genes as upregulated or downregulated were as follows: genes with values greater than **+2** were considered upregulated, and those with values less than **-2** were considered downregulated.

Subsequently, a Perl script was employed to identify genes that were consistently upregulated and downregulated across both UCEC and CRC datasets. Pathway analysis was then conducted on the selected genes.

2.4 Empirical analysis of digital gene expression data in R (EdgeR)

“Empirical analysis of digital gene expression data in R" is an R package used for analyzing digital gene expression data. This type of data offers a digital readout of gene expression from High Throughput Sequencing experiments. It works by counting individual RNA reads, where each count represents the occurrence of a particular RNA sequence (transcript) in a sample. This digital approach significantly enhances sensitivity, allowing for the identification of low-abundance transcripts. Researchers often use the EdgeR package, a powerful tool for RNA-seq data analysis, to identify differentially expressed genes under various experimental conditions or groups.

EdgeR utilizes two input files. The first file is a contrast file, which is a structured document defining the basis for performing contrasts between groups. In this research, a contrast file was created using TCGA IDs of MSS and MSI-H samples, with the first column containing the TCGA IDs and the second column indicating the category of each sample (MSS or MSI-H).

The second file consists of STAR count files downloaded from the GDC data portal. STAR (Spliced Transcripts Alignment to a Reference) is a widely used alignment tool for mapping RNA sequence reads to a reference genome. The STAR count files contain count data representing the number of successfully mapped RNA sequences reads to the reference genome.

For each hypermutated MSS sample in UCEC, CRC, and STAD, STAR count files were individually downloaded from the GDC. All downloaded files were then merged into a single file, where the first column contained Ensembl IDs, and the remaining columns represented sample names and contained the count values obtained from the STAR count files.

This EdgeR analysis was conducted on only three subtypes due to the requirement for a minimum of three hypermutated MSS samples in each subtype. Specifically, in UCEC, there were 25 MSS and 29 MSI-H samples; in CRC, there were 10 MSS and 11 MSI-H samples; and in STAD, there were 3 MSS and 5 MSI-H samples that were grouped for the execution of the EdgeR script.

Differential Gene Expression Analysis Script using EdgeR in R

```
library(edgeR)
       library(limma)
       library(Glimma)
       library(org.Hs.eg.db)
       library(gplots)
       library(RColorBrewer)
       library(NMF)
       library(EnsDb.Hsapiens.v86)
seqdata <- read.csv("F:/MS_FYP/REVIEW-
DATA/DATA/Mine/Merge.maf/edgeR_Input_Data_CRC_UCEC.csv", stringsAsFactors = FALSE)
        sampleinfo <- read.csv("F:/MS_FYP/REVIEW-
DATA/DATA/Mine/Merge.maf/edgeR_Contrast_input_file.csv", stringsAsFactors = TRUE)
       ensembl_ids <- seqdata[, 1]
               countdata <- seqdata[,-(1)]
               head(countdata)
               rownames(countdata) <- ensembl_ids
               head(countdata)
              table(colnames(countdata)==sampleinfo$Case_Submitter_Id)
               y <- DGEList(counts = countdata)
               y
               names(y)
               y$samples
               group <- paste(sampleinfo$MSSvsMSIH)
               group
               group <- factor(group)
               group
               y$samples$group <- group
               y$samples
columns(EnsDb.Hsapiens.v86)
ann <-
select(EnsDb.Hsapiens.v86,keys=rownames(y$counts),columns=c("GENEID","SYMBOL","GENE
NAME"))
       head(ann)
       table(ann$GENEID==rownames(y$counts))
       y$genes <- ann
       myCPM <- cpm(countdata)
       head(myCPM)
       thresh <- myCPM > 0.5
       head(thresh)
       table(rowSums(thresh))
```

```
               keep <- rowSums(thresh) >= 2
               summary(keep)
               ensembl_ids <- seqdata[, 1]

               countdata <- seqdata[,-(1)]
               head(countdata)
               rownames(countdata) <- ensembl_ids
               head(countdata)
              table(colnames(countdata)==sampleinfo$Case_Submitter_Id)
               y <- DGEList(counts = countdata)
               y
               names(y)
               y$samples
               group <- paste(sampleinfo$MSSvsMSIH)
               group
               group <- factor(group)
               group
               y$samples$group <- group
               y$samples
columns(EnsDb.Hsapiens.v86)
ann <-
select(EnsDb.Hsapiens.v86,keys=rownames(y$counts),columns=c("GENEID","SYMBOL","GEN
ENAME"))
       head(ann)
       table(ann$GENEID==rownames(y$counts))
       y$genes <- ann
       myCPM <- cpm(countdata)
       head(myCPM)
       thresh <- myCPM > 0.5
       head(thresh)
       table(rowSums(thresh))
       keep <- rowSums(thresh) >= 2
       summary(keep)
       y <- y[keep, keep.lib.sizes=FALSE]
       y$samples$lib.size <- colSums(y$counts)
      y$samples$lib.size
       par(mfrow=c(1,2)) #not required for current analysis
       levels(sampleinfo$MSSvMSIH)
       y <- calcNormFactors(y)
       y$samples
       y1 <- estimateCommonDisp(y, verbose=T)
       names(y1)
       y1 <- estimateTagwiseDisp(y1)
       names(y1)
       design.mat <- model.matrix(~ 0 + y$samples$group)
       colnames(design.mat) <- levels(y$samples$group)
       y2 <- estimateGLMCommonDisp(y,design.mat)
       y2 <- estimateGLMTrendedDisp(y2,design.mat, method="power")
       y2 <- estimateGLMTagwiseDisp(y2,design.mat)
      et12 <- exactTest(y1, pair=c(1,2))
       topTags(et12, n=10)
      de1 <- decideTestsDGE(et12, adjust.method="BH", p.value=0.05)
       summary(de1)
```

```
et12_results <- et12$table
ensembl_ids <- rownames(et12_results)
ann_results <- ann[match(ensembl_ids, ann$GENEID), ]
output_data <- cbind(Ensembl_ID = ensembl_ids, ann_results, et12_results)
```

Initially, various libraries were loaded, providing essential tools for the analysis of differential gene expression. Following data loading and preparation, Ensembl annotation was utilized to retrieve information from EnsDb.Hsapiens. v86, which included gene identifiers (Ensembl ID), gene names, and symbols. Subsequently, quality filtering was applied by counting reads per million (CPM), with a CPM threshold of 0.5, resulting in the removal of values below this threshold.

Following quality filtering, data normalization and dispersion estimation were performed. Normalization ensures that the data can be effectively compared across different groups, while dispersion estimation quantifies how much gene activity varies across different conditions or groups.

Genes obtained as output of EdgeR are further taken for pathway analysis.

2.5 Pathway Analysis

After conducting transcriptomic and differential gene expression analyses, the next crucial step is pathway analysis. This analysis is performed on the shortlisted genes from the transcriptomic analysis and the output genes from the differential expression analysis. Pathway analysis plays a vital role in helping us comprehend the functionality of genes and their involvement in various diseases. By gaining insights into a gene's functionality and its participation in different biological processes, we piece together the larger puzzle surrounding the identified genes and their significance.

Among the various tools available for pathway analysis, one frequently used tool is DAVID, which stands for Database for Annotation, Visualization, and Integrated Discovery. DAVID offers a comprehensive set of functional annotation tools, including resources like Reactome and KEGG. In this research, we focused on studying Reactome pathways using DAVID. What distinguishes the Reactome database from others is its unique approach of concentrating solely on a single species, *Homo sapiens*, and its commitment to gathering information across all biological aspects using a unified

approach. In DAVID, we simply passed the list of genes obtained from *Homo sapiens* species for analysis, while all the remaining parameters were set to their default values.

CHAPTER 03

RESULTS

Identification of novel biomarkers in hypermutated MSS tumors has the potential to enhance the effectiveness of targeted therapies and improve patient selection for immunotherapy. These biomarkers play a crucial role in refining patient responses to immunotherapy, increasing treatment efficacy, overcoming resistance, and minimizing side effects. In essence, these biomarkers are indispensable for extending the lifespan of cancer patients.

3.1 TCGA Results

For the analysis of TCGA data, a total of 1,993 files from various cancer types, including UCEC, CRC, STAD, PAAD, UCS, and ESCA, were downloaded from the GDC data portal. Within these 1,993 files, there were a total of 1,979 samples. These samples were further categorized into two groups: hypermutated and non-hypermutated, resulting in 356 samples classified as hypermutated and 1,623 samples classified as non-hypermutated. Among the hypermutated samples, additional classifications were made into MSS, MSI-H, and MSI-L subtypes.

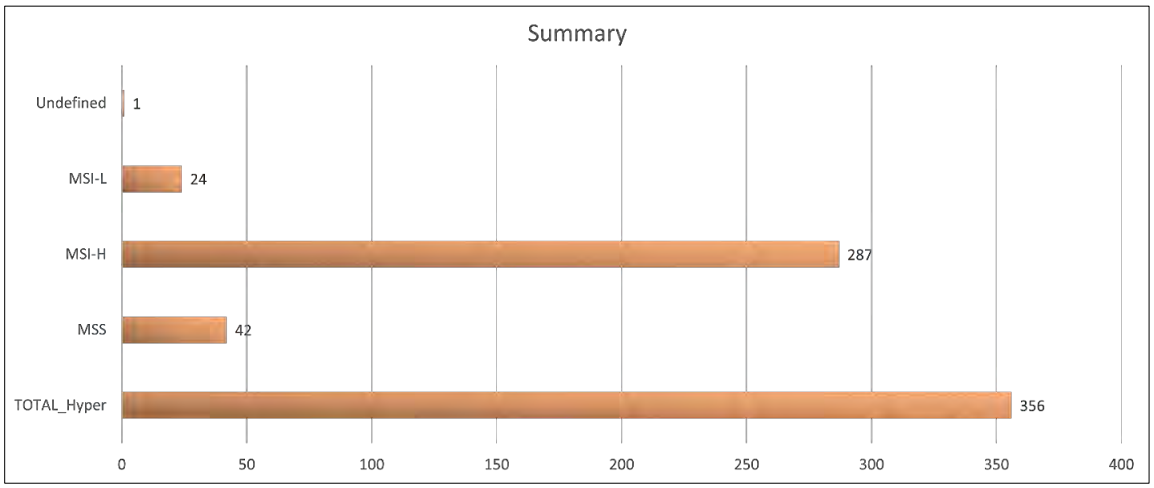

Figure 1: Hypermutated samples distribution graph.

This graph shows that there is total 356 hypermutated samples, along with 287 are MSI-H, 24 are MSI-L, 42 are MSS whereas, 1 sample is unidentified.

3.2 Microsatellite Stable (MSS)

This research particularly focuses on hypermutated MSS samples, we gained a total of 42 MSS hypermutated samples among the 7 most hypermutated cancers.

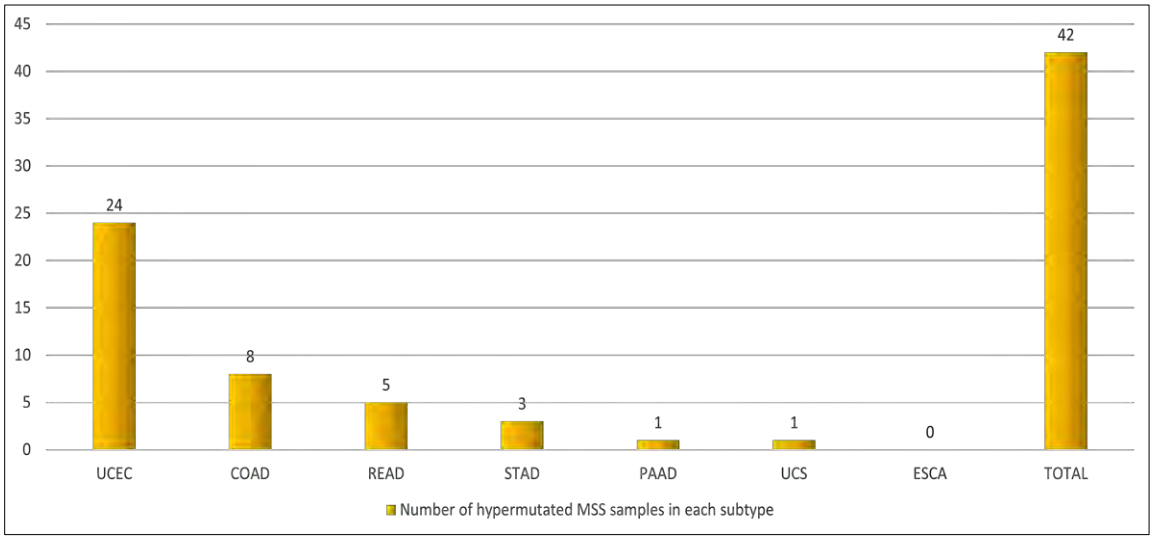

Figure 2: Hypermutated MSS samples distribution graph

Figure 2 graphs shows hypermutated MSS samples distribution, that 24 samples in UCEC, 8 in COAD, 5 in READ, 3 in STAD, 1 in UCS and 1 in PAAD makes total of 42 hypermutated MSS samples.

Sorted hypermutated MSS master table representing high number of MSI-H and MSI-L.

Table 2: Top hypermutated MSS genes.

Gene	Mutation	MSS	Per_MSS	MSI-L	Per_MSIL	MSI-H	Per_MSIH
POLE	chr12_132676598 _G_C	19	45.2381	5	20.83333	1	0.346021
ARID1A	chr1_26779863_C_T	16	38.09524	6	25	3	1.038062
PIK3CA	chr3_179199088 _G_A	14	33.33333	7	29.16667	18	6.228374
PTEN	chr10_87933148 _G_A	14	33.33333	12	50	5	1.730104
APC	chr5_112838934 _C_T	13	30.95238	0	0	1	0.346021
ZNHIT6	chr1_85706306 _G_A	11	26.19048	1	4.166667	1	0.346021
RFTN2	chr2_197615926 _G_A	9	21.42857	2	8.333333	0	0
C2CD5	chr12_22524531 _C_T	9	21.42857	0	0	1	0.346021
GABRA4	chr4_46928511 _C_T	9	21.42857	1	4.166667	0	0
NF1	chr17_31350209 _C_T	9	21.42857	1	4.166667	1	0.346021
PTEN	chr10_87864488 _G_T	9	21.42857	3	12.5	0	0
LRRC7	chr1_70038328 _C_T	8	19.04762	2	8.333333	0	0
TP53	chr17_7674894 _G_A	8	19.04762	1	4.166667	3	1.038062
LYST	chr1_235806698_C_C_T	8	19.04762	1	4.166667	0	0
POLE	chr12_132673703_C_A	8	19.04762	8	33.33333	1	0.346021
PTEN	chr10_87960987_G_T	8	19.04762	1	4.166667	2	0.692042
ASCC3	chr6_100848542_C_T	8	19.04762	1	4.166667	0	0
EFR3A	chr8_131970537_C_T	8	19.04762	1	4.166667	0	0
BMT2	chr7_112821884_G_A	8	19.04762	3	12.5	1	0.346021
LTV1	chr6_143863098_C_T	8	19.04762	3	12.5	2	0.692042
VPS50	chr7_93353688_C_T	7	16.66667	1	4.166667	1	0.346021
PUS7	chr7_105462682_G_A	7	16.66667	3	12.5	0	0
COL14A1	chr8_120216450_G_A	7	16.66667	1	4.166667	0	0
RFC5	chr12_118029791_C_T	7	16.66667	0	0	0	0
XRCC5	chr2_216190251_G_A	7	16.66667	2	8.333333	0	0
SKP2	chr5_36168343_G_A	7	16.66667	1	4.166667	0	0
TRIM23	chr5_65609421_C_T	7	16.66667	1	4.166667	0	0
DBX1	chr11_20159234_G_A	7	16.66667	5	20.83333	0	0
CCDC73	chr11_32635824 _C_A	7	16.66667	0	0	0	0
MS4A8	chr11_60700868_C_T	7	16.66667	4	16.66667	0	0
MYT1	chr20_64207872_G_A	7	16.66667	3	12.5	1	0.346021
SYNJ1	chr21_32673466_G_A	7	16.66667	1	4.166667	1	0.346021
SEZ6L	chr22_26292931 _C_T	7	16.66667	2	8.333333	0	0
PIK3R1	chr5_68293123_C_T	7	16.66667	9	37.5	4	1.384083
RALA	chr7_39706150_C_T	7	16.66667	1	4.166667	0	0
ZNF267	chr16_31915952_G_A	7	16.66667	1	4.166667	1	0.346021

Table 2 illustrates the leading genes in the MSS master table, ranked by the count in the MSS column. In the first column, there are gene names, and in the second column, details of the mutations, including chromosome number, start and end positions, as well as the reference and alternate alleles that have undergone mutation. The MSS column indicates the number of occurrences of each mutation within MSS samples, while per_MSS represents the percentage of these mutations relative to the total count of MSS samples. This same pattern extends to the MSI-L and MSI-H columns, revealing how frequently the same mutations were observed in MSI-L and MSI-H samples and their respective percentages concerning the total count for each sample type. After applying the filter, the master table was sorted into the one below.

This table unmistakably highlights that the percentages of MSI-H and MSI-L samples are significantly elevated in certain genes, potentially exerting an influence on our results. Therefore, after applying the filter, following table was obtained.

Table 3: Top Hypermutated MSS genes after applying filter.

Gene	Mutation	MSS	Per_MSS	MSI-L	Per_MSIL	MSI-H	Per_MSIH
APC	chr5_112838934 _C_T	13	30.95238	0	0	1	0.346021
ZNHIT6	chr1_85706306_G_A	11	26.19048	1	4.166667	1	0.346021
RFTN2	chr2_197615926_G_A	9	21.42857	2	8.333333	0	0
C2CD5	chr12_22524531_C_T	9	21.42857	0	0	1	0.346021
GABRA4	chr4_46928511_C_T	9	21.42857	1	4.166667	0	0
NF1	chr17_31350209_C_T	9	21.42857	1	4.166667	1	0.346021
LRRC7	chr1_70038328_C_T	8	19.04762	2	8.333333	0	0
LYST	chr1_235806698_C_T	8	19.04762	1	4.166667	0	0
PTEN	chr10_87960987_G_T	8	19.04762	1	4.166667	2	0.692042
ASCC3	chr6_100848542_C_T	8	19.04762	1	4.166667	0	0
EFR3A	chr8_131970537_C_T	8	19.04762	1	4.166667	0	0
VPS50	chr7_93353688_C_T	7	16.66667	1	4.166667	1	0.346021
COL14A1	chr8_120216450_G_A	7	16.66667	1	4.166667	0	0
RFC5	chr12_118029791_C_T	7	16.66667	0	0	0	0
XRCC5	chr2_216190251_G_A	7	16.66667	2	8.333333	0	0
SKP2	chr5_36168343_G_A	7	16.66667	1	4.166667	0	0
TRIM23	chr5_65609421_C_T	7	16.66667	1	4.166667	0	0
CCDC73	chr11_32635824_C_A	7	16.66667	0	0	0	0
SYNJ1	chr21_32673466_G_A	7	16.66667	1	4.166667	1	0.346021
SEZ6L	chr22_26292931_C_T	7	16.66667	2	8.333333	0	0
RALA	chr7_39706150_C_T	7	16.66667	1	4.166667	0	0
ZNF267	chr16_31915952_G_A	7	16.66667	1	4.166667	1	0.346021
MAEL	chr1_167005351_C_T	7	16.66667	1	4.166667	0	0
ABCB5	chr7_20643288_G_A	7	16.66667	2	8.333333	0	0
SELP	chr1_169609552_G_A	7	16.66667	0	0	1	0.346021
PCMTD1	chr8_51833615_C_T	7	16.66667	2	8.333333	0	0
MBOAT2	chr2_8958590_C_T	7	16.66667	2	8.333333	0	0
ZNF678	chr1_227655776_G_T	7	16.66667	1	4.166667	0	0
RASA1	chr5_87353188_G_T	7	16.66667	0	0	1	0.346021
SNX13	chr7_17796878_G_A	7	16.66667	0	0	1	0.346021

Table 3 below provides an overview of the all genes from the MSS master table, after applying filter on Per_MSS column >16, Per_MSI-H <1 and Per_MSI-L<10. In the first column, there are gene names, while the second column details the mutations, including chromosome number, start and end positions, along with the reference and alternate alleles affected. The MSS column quantifies how often each mutation appears within MSS

samples, and per_MSS expresses these frequencies as percentages relative to the total MSS sample count. This pattern extends to the MSI-L and MSI-H columns, showing the recurrence of the same mutations in MSI-L and MSI-H samples, along with their respective percentages relative to the total count for each sample type. There is clear difference in top genes here.

3.3 Mutation Mappers

Mutations mappers of the multiple genes were made for the identification of hotpot mutations. The number of samples are out of total of 42 MSS hypermutated samples;

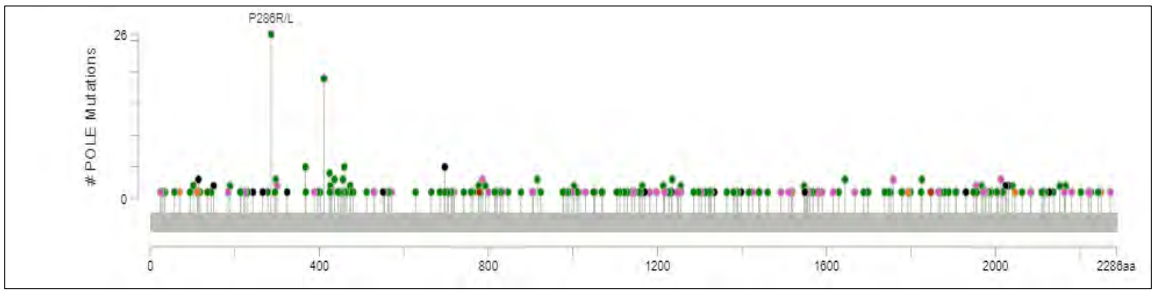

Figure 3: POLE Mutation Mapper showing P286R hotspot mutation in 2 samples

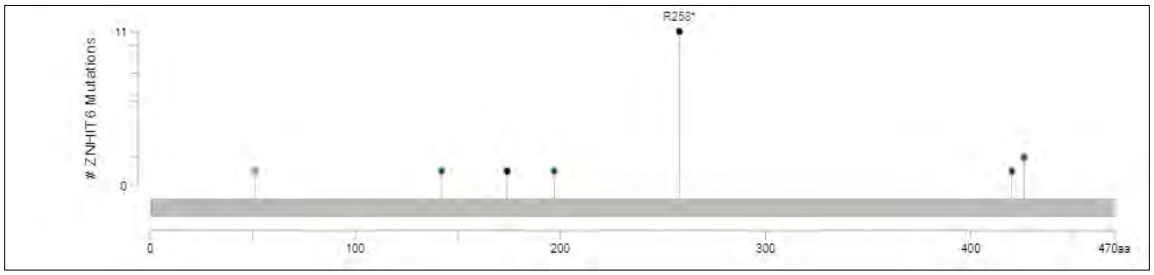

Figure 4: ZNHIT6 Mutation Mapper showing R258* hotpot mutation in 11 samples

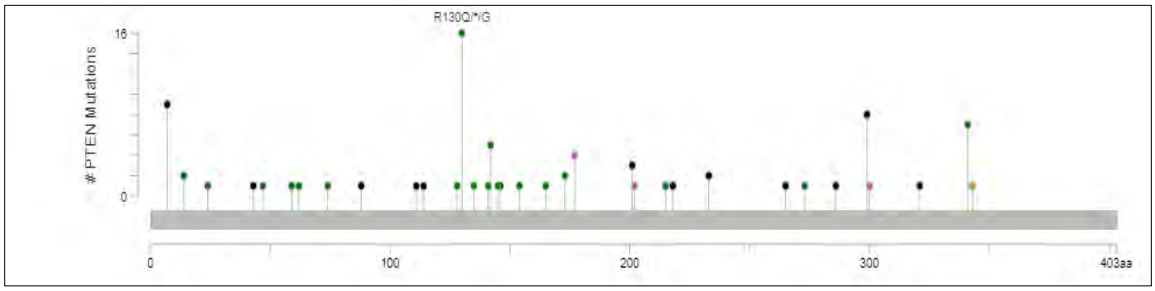

Figure 5: PTEN Mutation Mapper showing R130Q hotspot mutation in 16 samples

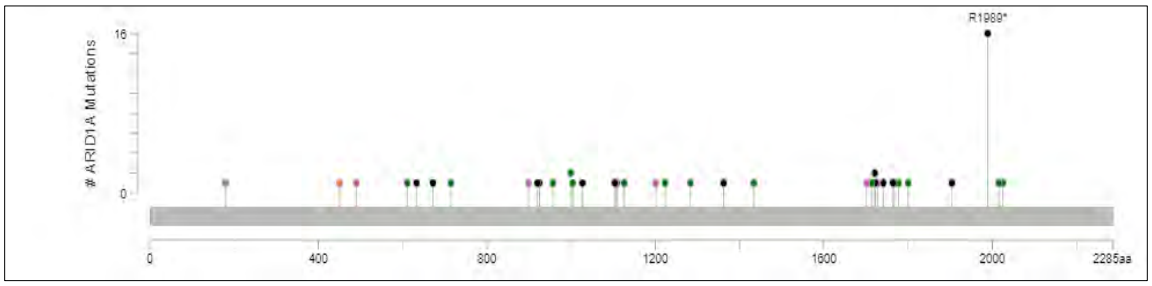

Figure 6: ARID1A Mutation Mapper showing R1114* hotspot mutation in 16 samples

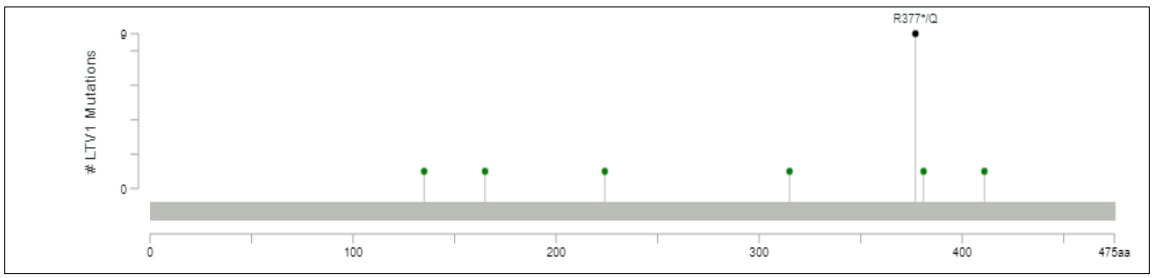

Figure 7: LTV1 Mutation Mapper showing R377* hotspot mutation in 9 samples

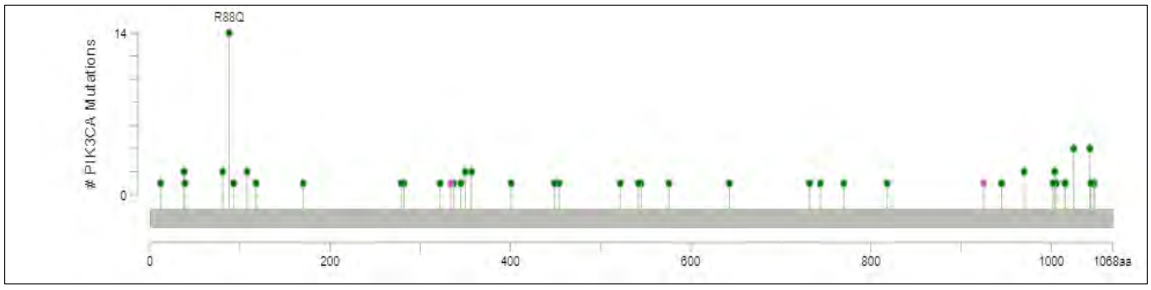

Figure 8: PIK3CA Mutation Mapper showing R88Q hotspot mutation in 14 samples

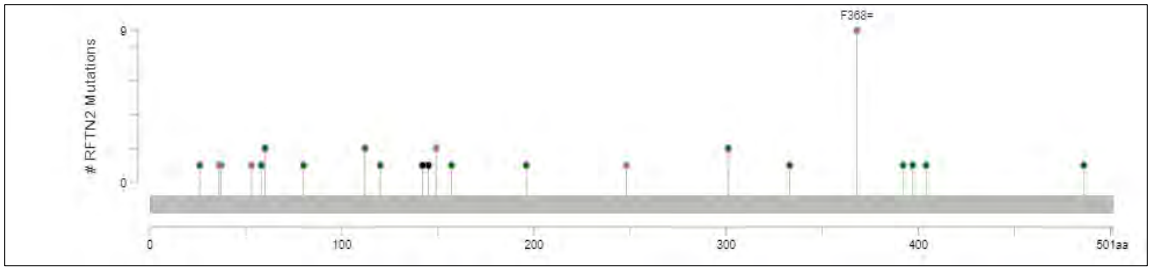

Figure 10: RFTN2 Mutation Mapper showing F368= hotspot mutation in 9 samples

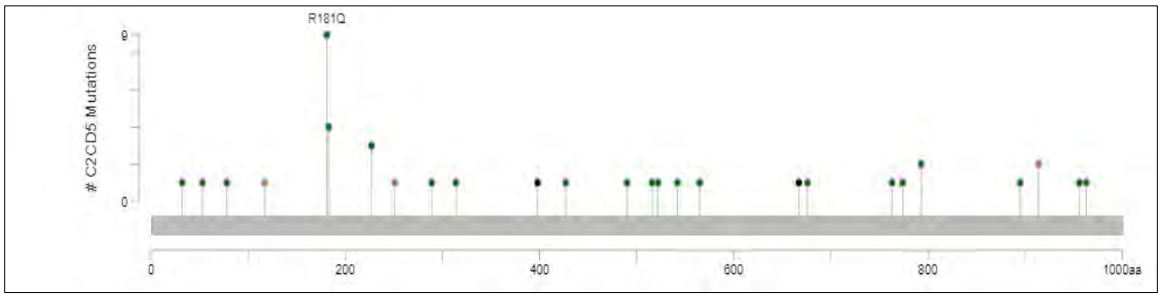

Figure 9: C2CD5 Mutation Mapper showing R181Q hotspot mutation in 9 samples

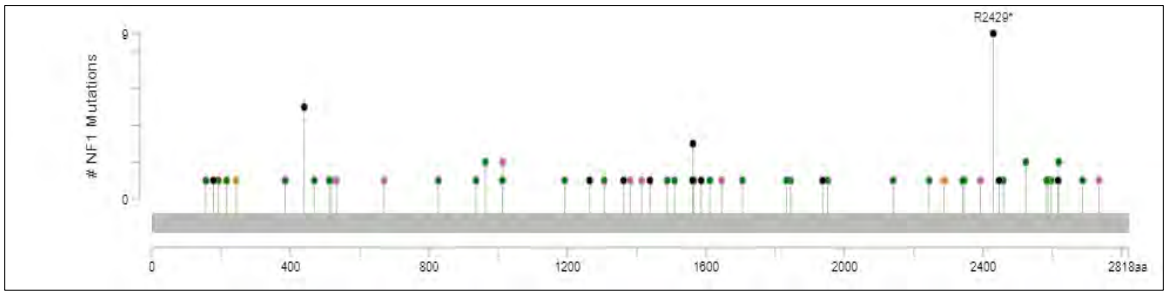

Figure 11: NF1 Mutation Mapper showing R2429* hotspot mutation in 9 samples

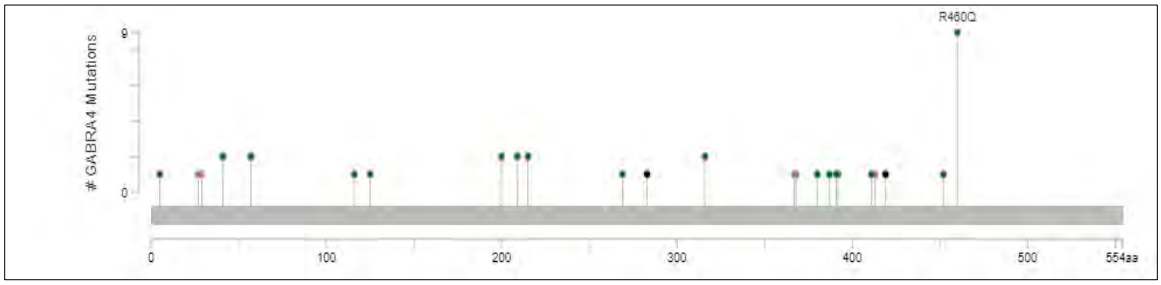

Figure 12: GABRA4 Mutation Mapper showing R460Q hotspot mutation in 9 samples

Table 4: Top Genes after filtration along with hotspot mutations.

Gene	Mutation	MSS	Per_MSS	MSI-L	Per_MSIL	MSI-H	Per_MSIH
POLE	P286R	19	45.2380	5	20.8333	1	0.346020
APC	R1114*	13	30.9523	0	0	1	0.346020
ZNHIT6	R258*	11	26.1904	1	4.16666	1	0.346020
C2CD5	R181Q	9	21.4285	0	0	1	0.346020
GABRA4	R460Q	9	21.4285	1	4.16666	0	0
NF1	R2450*	9	21.4285	1	4.16666	1	0.346020
RFTN2	F368=	9	21.4285	2	8.33333	0	0
ASCC3	R136Q	8	19.0476	1	4.16666	0	0
EFR3A	F351=	8	19.0476	1	4.16666	0	0
LRRC7	S835L	8	19.0476	2	8.33333	0	0
LYST	R813Q	8	19.0476	1	4.16666	0	0

Table 4 represent the name of the genes in the first column, there hotspot mutations in the second, number of MSS samples in which hotspot mutation is found. Whereas column 4 represent percentage of MSS in total MSS samples and same goes for MSI-L and MSI-H columns.

Even though the percentage of POLE mutations in the MSI-L category is relatively high, it remains an important focus for our study, particularly within hypermutated MSS samples. This is why it has been included as a significant target in our research.

Based on master table, the short-listed genes were APC, POLE, ZNHIT6, C2CD5, RFTN2 and NF1.

3.4 DNA Polymerase Epsilon, Catalytic Subunit (POLE)

Microsatellite stability (MSS) is often associated with mutations in the POLE gene. These mutations lead to a heightened mutation rate, characterized by over 100 mutations in the POLE and POLD1 genes per 10^6 bases [Nature, 2012; Barbari and Shcherbakova, 2017]. According to the Bethesda guidelines, CRC can be classified as either MSI-L or MSS based on instability, and modern hypermutation kits categorize samples lacking length mutations as MSS.

A study conducted by Barbari and Shcherbakova in 2017 identified specific mutations in the POLE gene, including high-frequency somatic mutations like P286R and V411L, along with low-frequency germline mutations such as P286H/S, F367S, S459F, and V424L

[Nebot-Bral et al., 2017]. Recent research has also pinpointed somatic point mutations in the polymerase domain of PolE (R567C, K593C, S595P, E611K, and L621F) and deletion frameshifts in the middle region of the PolE gene (V1446fs *3) as pathogenic factors contributing to cancer [Meng et al., 2020; Min et al., 2020].

MSS in CRC primarily arises from a high frequency of POLE mutations and a low frequency of POLD1 mutations [Barbari and Shcherbakova, 2017]. While the NCI panel has identified seven genes in which mutations result in forms of hypermutation, there are still cases of hypermutation that do not exhibit mutations in any of these panel genes. Investigating the cause of hypermutation in these cases remains an ongoing area of research. Key enzymes responsible for DNA replication and repair in vivo are DNA polymerases (Pols). In eukaryotes, three cooperating Pols are required for DNA replication: Pols α, δ, and ε. Human Pol ε is a heterotetramer composed of p261 (encoded by the POLE gene), p59 (POLE2), p17 (POLE3), and p12 (POLE4) subunits. The p261 subunit contains N-terminal domains with conserved polymerase and 3′→5′ exonuclease subdomains, as well as a large C-terminal domain required for interaction with the three smaller subunits. Human Pol δ is another heterotetramer, comprising p125 (POLD1), p66 (POLD3), p50 (POLD4), and p12 (POLD4) subunits. The p125 subunit shares similarities with the N-terminal half of Pol ε p261, containing polymerase and exonuclease domains. POLE is responsible for leading strand synthesis, while POLD1 handles lagging strand synthesis in DNA replication. Under specific conditions, POLD1 can also synthesize the leading strand alongside the lagging strand [Nature, 2012; Barbari and Shcherbakova, 2017].

The specific mutation P286R in the POLE gene has been identified as a significant biomarker for cases with MSS hypermutation [Nebot-Bral et al., 2017]. To investigate the impact of this mutation, a comparative analysis was conducted using the cBioPortal platform, focusing on samples from UCEC and CRC datasets. Among the samples, 19 carried the POLE P286R mutation, with the majority occurring in UCEC (15), while 3 were from CRC and 1 from pancreatic adenocarcinoma (PAAD). Conversely, 23 samples did not possess the POLE P286R mutation, including 9 from UCEC, 11 from CRC, and 3 from stomach adenocarcinoma (STAD).

Comparing the two groups across both cancer subtypes, common genes that were upregulated and downregulated were identified. Furthermore, to gain deeper insights into whether the specific mutation, POLE P286R, was responsible for observed gene expression changes, additional comparisons were conducted.

3.4.1 cBioPortal Results

In cBioPortal results the total number of upregulated genes in UCEC cancer in POLE vs no POLE hypermutated MSS samples (Group_1) were 44 with NOTUM as most upregulated gene. The total number of downregulated genes in same group were 70 with SCGB1A1 as most downregulated gene. Whereas in other group of POLE mutated vs non hyper samples (Group_2) in UCEC, 33 genes were upregulated in which CXCL9 was most upregulated and 60 genes were downregulated in which AOC1 was most downregulated gene.

On the other hand, in Group_1 in CRC total genes were upregulated 278 and SOX2 was most unregulated gene and 253 genes were downregulated with SI as most downregulated gene. In Group_2 analysis 415 genes were upregulated with SOX2 as most upregulated and 388 were down regulated in which CHP2 was most downregulated gene.

3.4.2 Common Upregulated Genes in UCEC and CRC

Following are the tables representing common upregulated genes in CRC and UCEC of POLE group:

Table 5: Common Upregulated Genes in UCEC, influenced by POLE Gene Mutations

UCEC_POLE Mut VS Non-POLE Mut			POLE_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
DKK4	3.14	Hyp_MSS_WITH_POLE	**DKK4**	1.35	Hyp_MSS_WITH_POLE
FGF20	2.13	Hyp_MSS_WITH_POLE	**FGF20**	1.30	Hyp_MSS__WITH_POLE
HMX2	2.07	Hyp_MSS_WITH_POLE	**HMX2**	2.13	Hyp_MSS_WITH_POLE

Table 6: Common Upregulated Genes in CRC, influenced by POLE Gene Mutations

CRC_POLE Mut VS Non-POLE Mut			POLE_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
DKK4	4.10	Hyp_MSS_WITH_POLE	**DKK4**	3.51	Hyp_MSS_WITH_POLE
FGF20	2.10	Hyp_MSS_WITH_POLE	**FGF20**	2.68	Hyp_MSS_WITH_POLE
HMX2	5.44	Hyp_MSS_WITH_POLE	**HMX2**	5.23	Hyp_MSS_WITH_POLE

Both these tables represent 3 common upregulated genes due to POLE mutation in CRC and UCEC. First group is comparison between hypermutated MSS POLE mutated and nonmutated samples and second group comparison is between POLE mutated samples with non-hyper samples. In each group first column represents gene name, second column represents its Log2Ratio value whereas third column represents the group in which gene expression is higher. HMX2 is highly upregulated in POLE mutated samples group.

3.4.2.1 Dickkopf WNT Signaling Pathway Inhibitor 4 (DKK4)

DKK4, a member of the Dickkopf protein family, that inhibits Wnt Signaling pathway by blocking the interaction between Wnt ligands and the Frizzled receptor. It may act as tumor suppressor gene and embryonic development, tissue homeostasis, and plays a part in disease pathways such as cancer and angiogenesis. In colorectal cancer (CRC), DKK4 exhibits complex expression patterns. It's often upregulated due to hypermethylation but can be silenced in some CRC cells, while remaining active in normal colon tissues. Studies yield mixed results, with some indicating increased DKK4 expression in CRC compared to normal tissue and others showing the opposite. DKK4 correlates positively with β-catenin and fibroblast growth factor-20 in CRC, and its knockdown leads to increased β-catenin expression. It acts as a T-cell factor (TCF)-dependent signaling inhibitor, impacting CRC cell growth by arresting the G0/G1 cell cycle phase and enhancing metastasis, aggression, and angiogenic potential in CRC. DKK4 may also play a role beyond suppressing the Wnt/canonical signaling pathway, and its inhibition by 1α, 25-dihydroxyvitamin D3 suggests potential regulatory mechanisms. (Cai et al., 2018; Lou et al., 2021)

3.4.2.2 Fibroblast Growth Factor 20 (FGF20)

FGF20, a member of the fibroblast growth factor family, plays a vital role in embryonic development, influencing the formation of tissues and organs such as the midbrain, inner ear, and limbs. FGF20 may also contribute to angiogenesis, the formation of new blood vessels that support tumor growth and facilitate cancer cell invasion and metastasis. The Mitogen-Activated Protein Kinase (MAPK) pathways regulate essential cellular functions like proliferation and apoptosis. When activated, FGF20 binds to its receptor known as FGFR1, initiating a cascade that activates the RAS protein, ultimately leading to the activation of the MAPK signaling pathway, which, in turn, can influence gene regulation. Therefore, any dysregulation in the MAPK signaling pathway can result in the upregulation of FGF20 and potentially contribute to the development of various carcinomas (Su et al., 2020). In a comprehensive pancancer study, FGF20 was identified as one of the most upregulated members of the FGF family (Li et al., 2020). Furthermore, FGF20 has been found to be upregulated not only in colorectal cancer (CRC) but also in different cancer types such as lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), and lung squamous cell carcinoma (LUSC) (Li et al., 2020). In a study examining the expression of long noncoding RNA and coding RNA during the initiation of secondary hair follicles in sheep skin, FGF20 exhibited significant differential expression within the Wnt, Shh, Notch, and BMP signaling pathways. Notably, it was found to be upregulated along with Wnt2 in these pathways (Yue et al., 2016).

3.4.3 Common Downregulated Genes in UCEC and CRC

Following are the tables representing common downregulated genes in CRC and UCEC of POLE group:

Table 7: Common downregulated Genes in UCEC, influenced by POLE Gene Mutations

UCEC_POLE Mut VS Non-POLE Mut			POLE_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
AKR1B10	-3.40	Hyp_MSS_NO_POLE	**AKR1B10**	-1.87	MSS_NonHyper
AKR1C2	-2.18	Hyp_MSS_NO_POLE	**AKR1C2**	-0.54	MSS_NonHyper

Table 8: Common downregulated Genes in CRC, influenced by POLE Gene Mutations

CRC_POLE Mut VS Non-POLE Mut			POLE_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
AKR1B10	-2.72	Hyp_MSS_NO_POLE	**AKR1B10**	-2.80	MSS_NonHyper
AKR1C2	-2.30	Hyp_MSS_NO_POLE	**AKR1C2**	-2.63	MSS_NonHyper

These tables, showcases common downregulated genes resulting from POLE mutations. It also distinguishes between two groups: hypermutated MSS POLE mutated samples and non-mutated samples, as well as POLE mutated samples versus non-hypermutated samples. It is noteworthy that AKR1B10 demonstrates the most prominent downregulation in the group of non-POLE mutated and non-hypermutated samples.

3.4.3.1 Aldo-Keto Reductase Family 1 Member B10 (AKR1B10)

AKR1B10, also known as Aldo-keto reductase family 1 member B10, is expressed in various tissues, including the lung, liver, and gastrointestinal tract. It has been linked to tumor invasion, growth, and drug resistance in different types of cancer, such as lung, breast, hepatic, and pancreatic carcinomas, where it is often overexpressed. The MAPK signaling pathway, of which ERK is a part, is involved in these processes. ERK is a kinase activated within the MAPK pathway and plays a role in influencing gene expression by phosphorylating transcription factors and other regulatory proteins (Cong et al., 2019)

All-trans-retinoic acid (atRA), an oxidized form of vitamin A (retinol), is known to have a significant role in various biological processes. Each step of this pathway has been reported as dysregulated in colorectal cancer (CRC), leading to a clear downregulation of AKR1B10. Consequently, a decrease in atRA levels can disrupt normal cell activities and contribute to the development of CRC or other cancers (Kropotova et al., 2014). In gastric cancer, lower levels of AKR1B10 expression have been reported (Zhou et al., 2023). In more than 60% of early-stage specimens, AKR1B10 expression was reported to be downregulated. The significant downregulation of AKR1B10 has been suggested as a potential biomarker for diagnosing CRC (Kropotova et al., 2010). Conversely,

upregulation of AKR1B10 is reported in breast cancer and hepatocellular carcinoma (Cong et al., 2019; Dai et al., 2021).

3.4.3.2 Aldo-Keto Reductase Family 1, Member C2 (AKR1C2)

The AKR1C2 gene, also known as Aldo-keto reductase family 1 member C2, encodes the AKR1C2 enzyme. Decreased AKR1C2 activity contributes to disease progression, while its dysregulation has been associated with various hormone-dependent cancers. Liu et al. proposed that downregulated AKR1C2 could serve as a potential predictive prognostic biomarker for immunotherapy in gastric cancer patients. Previous studies have reported that higher expression of AKR1C2 is related to a better prognosis in thyroid carcinoma and acute myeloid leukemia (AML), while it is associated with poor prognosis in esophageal squamous cell carcinoma (ESCC) (Liu et al., 2023; Jin et al., 2019). Downregulation of AKR1C2 is also observed in breast cancer studies, where this gene is proposed as a target for hormone-based therapies (Wenners et al., 2015). Li et al., in their study reported decreased AKR1C2 expression levels in CRC, based on their high CpG methylation levels. The exact function of AKR1C2 in tumorigenesis and prognosis has not been identified yet (Zhao et al., 2019)

3.5 Adenomatous Polyposis Coli (APC)

The adenomatous polyposis coli (APC) gene is a tumor suppressor gene that plays a critical role in the development of colorectal cancer (CRC). Approximately 80% of CRC cases are initiated by somatic mutations in the APC gene, highlighting the significance of dysregulated intracellular WNT signaling in CRC development (Grant et al., 2021). Interestingly, APC mutations are less frequent in early-onset CRC cases, suggesting that alternative pathways may drive tumorigenesis in these cases (Lieu et al., 2019; Willauer et al., 2010). In a recent study conducted by Grant et al. in 2021, the authors investigated the distinguishing characteristics of APC-mutated (APC mut+) and APC-nonmutated (APC mut-) MSS hypermutated CRCs. The findings revealed that early-onset CRC cases are more likely to be APC mut- and are often diagnosed at a later stage, beyond stage one (Grant et al., 2021).

In our own analysis of the data, we identified different APC mutations, with a total of 130 mutations observed. Among these mutations, the R1114* mutation was the most frequent. Out of the 35 samples analyzed, 31.4% (11) were diagnosed in the early-onset group (less than 50 years), while the remaining 68.5% (24) were diagnosed in the late-onset group (greater than 50 years).

Table 9: APC samples separated on basis of early and late onset

Late-Onset APC Mutated Samples		Early Onset APC Mutated Samples	
Sample-ID	**Diagnosis Age**	**Sample-ID**	**Diagnosis Age**
TCGA-A5-A0GP-01A	83	TCGA-AA-3510-01A	37
TCGA-A5-A2K5-01A	56	TCGA-AA-3977-01A	43
TCGA-AG-3892-01A	50	TCGA-AA-3977-01A	43
TCGA-AG-3892-01A	50	TCGA-AA-3984-01A	36
TCGA-AG-A002-01A	51	TCGA-AA-3984-01A	36
TCGA-AJ-A5DW-01A	57	TCGA-AZ-4315-01A	35
TCGA-AP-A0LM-01A	61	TCGA-BR-8680-01A	33
TCGA-AX-A05Z-01A	57	TCGA-CA-6717-01A	45
TCGA-B5-A0JY-01A	57	TCGA-CA-6718-01A	34
TCGA-B5-A11R-01A	83	TCGA-EI-6917-01A	46
TCGA-BK-A6W3-01A	65	TCGA-ND-A4WC-01A	33
TCGA-BS-A0UF-01A	76		
TCGA-D1-A16X-01A	64		
TCGA-E6-A1M0-01A	52		
TCGA-EO-A22X-01A	70		
TCGA-EO-A3AV-01A	56		
TCGA-EO-A3B0-01A	58		
TCGA-EY-A1G8-01A	56		
TCGA-EY-A1GD-01A	61		
TCGA-EY-A1GI-01A	65		
TCGA-F5-6814-01A	51		
TCGA-FI-A2D5-01A	55		
TCGA-IB-7651-01A	51		
TCGA-QF-A5YS-01A	54		

Furthermore, the study by Grant et al. also suggested that BRAF and RNF43 mutations are associated with the progression of MSS APC mut- CRCs. The mutation and expression

data indicated that CRCs with APC mut- tend to exhibit dysregulation of genes involved in responding to extracellular WNT signaling.

For upregulating and downregulating genes same analysis was performed using cBioPortal as for POLE gene. Out of total APC R1114* hotspot mutated 13 hypermutated MSS samples, 3 were UCEC and 10 were CRC. While on the other hand, 29 were total hypermutated MSS samples without APC R1114* mutation, from which 21 were UCEC, 3 were CRC, 3 STAD and for UCS and PAAD one each.

3.5.1 cBioPortal Results

In the results obtained from cBioPortal, when comparing hypermutated MSS samples with and without APC mutations (Group_1) in UCEC cancer, we identified 269 upregulated genes, with MMP10 being the most prominently upregulated. Additionally, we found 88 downregulated genes, with SCGB1A1 being the most significantly downregulated gene. In the same context, when examining APC mutated samples versus non-hypermutated samples (Group_2) in UCEC, we observed 145 upregulated genes, with MMP10 being the most highly upregulated. Furthermore, we identified 175 downregulated genes, with HOXB13 as the most prominently downregulated gene.

Shifting to CRC, within Group_1, there were 175 upregulated genes, with DEFA6 ranking as the most upregulated gene, while 148 genes were downregulated, with UGT1A8 as the most significantly downregulated gene. In Group_2 analysis for CRC, we noted 86 upregulated genes, with CALB1 standing out as the most upregulated, and 83 genes were downregulated, with XPNPEP2 as the most notably downregulated gene.

3.5.2 Common upregulated Genes in UCEC and CRC

Following are the tables representing common upregulated genes in CRC and UCEC of APC group:

Table 10: Common upregulated Genes in UCECC, influenced by APC Gene Mutations

UCEC_APC Mut VS Non-APC Mut			APC_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
TSIX	3.14	MSS_HYP_WITH_APC	**TSIX**	2.69	MSS_HYP_WITH_APC
GNG4	3.23	MSS_HYP_WITH_APC	**GNG4**	2.51	MSS_HYP_WITH_APC
GSTM1	3.18	MSS_HYP_WITH_APC	**GSTM1**	1.31	MSS_HYP_WITH_APC
ALDH1A1	2.73	MSS_HYP_WITH_APC	**ALDH1A1**	2.66	MSS_HYP_WITH_APC
CHIT1	2.54	MSS_HYP_WITH_APC	**CHIT1**	3.21	MSS_HYP_WITH_APC
CPS1	2.39	MSS_HYP_WITH_APC	**CPS1**	1.13	MSS_HYP_WITH_APC
NELL2	2.24	MSS_HYP_WITH_APC	**NELL2**	1.57	MSS_HYP_WITH_APC
RNF212	2.22	MSS_HYP_WITH_APC	**RNF212**	1.11	MSS_HYP_WITH_APC
COMP	2.15	MSS_HYP_WITH_APC	**COMP**	1.30	MSS_HYP_WITH_APC
GABRE	2.09	MSS_HYP_WITH_APC	**GABRE**	0.05	MSS_NonHyper
EPYC	2.09	MSS_HYP_WITH_APC	**EPYC**	2.23	MSS_HYP_WITH_APC

Table 11: Common upregulated Genes in CRC, influenced by APC Gene Mutations

CRC_APC Mut VS NonAPC Mut			APC_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
TSIX	3.05	MSS_HYP_WITH_APC	**TSIX**	0.22	MSS_HYP_WITH_APC
GNG4	2.15	MSS_HYP_WITH_APC	**GNG4**	2.13	MSS_NonHyp
GSTM1	2.62	MSS_HYP_WITH_APC	**GSTM1**	1.40	MSS_HYP_WITH_APC
ALDH1A1	3.17	MSS_HYP_WITH_APC	**ALDH1A1**	1.45	MSS_HYP_WITH_APC
CHIT1	2.51	MSS_HYP_WITH_APC	**CHIT1**	1.02	MSS_HYP_WITH_APC
CPS1	3.37	MSS_HYP_WITH_APC	**CPS1**	2.50	MSS_HYP_WITH_APC
NELL2	2.05	MSS_HYP_WITH_APC	**NELL2**	0.89	MSS_HYP_WITH_APC
RNF212	2.73	MSS_HYP_WITH_APC	**RNF212**	1.81	MSS_HYP_WITH_APC
COMP	2.02	MSS_HYP_WITH_APC	**COMP**	0.23	MSS_NonHyp
GABRE	2.84	MSS_HYP_WITH_APC	**GABRE**	0.91	MSS_NonHyper
EPYC	2.41	MSS_HYP_WITH_APC	**EPYC**	1.43	MSS_HYP_WITH_APC

These tables display the common genes that experience upregulation due to APC mutations. They further differentiate between two distinct groups: hypermutated MSS APC mutated samples and their non-mutated counterparts, as well as APC mutated samples versus those without hypermutation. Of particular note is the significant upregulation of GNG4 in UCEC and CPS1 in CRC observed in the APC mutated sample group.

3.5.2.1 X-Inactive Specific Transcript (TSIX)

TSIX is a long non-coding RNA gene known for its involvement in X chromosome inactivation. In a 2020 study by Salama et al., limited evidence suggested TSIX's role as a tumor suppressor. Moreover, TSIX has been identified as a potential biomarker for immunotherapy in breast cancer due to its upregulation. Interestingly, this study found that high PD-L1 expression in lymph nodes was significantly associated with TSIX downregulation (Salama et al., 2020).

Studies have also revealed TSIX's role in regulating mRNA stability associated with collagen. In normal skin cells, TSIX contributes to maintaining stable collagen levels, while in systemic sclerosis (SSc)-affected skin cells, it appears to be involved in the dysregulation of collagen production. This dysregulation is characteristic of SSc, a condition marked by excessive scarring and collagen production. Consequently, TSIX has been reported to be upregulated in scleroderma dermal fibroblasts (Wang et al., 2016).

TSIX is known to have a strong correlation with XIST, another long non-coding RNA. TSIX acts as an antisense or complementary sequence to XIST, and their expression levels are inversely related. When TSIX is upregulated, XIST is downregulated, and vice versa. Increased TSIX expression blocks X-chromosome inactivation, while increased XIST expression leads to X chromosome activation (Katopodis et al., 2020).

In the context of cancer, TSIX has been reported to be upregulated in lung adenocarcinoma (LUAD) and breast cancer. It is considered an important target in immunotherapy, including in the treatment of gastric cancer (Katopodis et al., 2020; Samir et al., 2021; Shen et al., 2021; Sun et al., 2021). However, TSIX's exact role in cancer is not yet fully understood.

3.5.2.2 G-Protein Subunit Gamma 4 (GNG4)

Guanine nucleotide-binding protein gamma subunit-4 (GNG4) is a member of the guanine nucleotide-binding protein complex. The role of GNG4 in tumors is not yet fully understood, but several studies have shown high expressions of GNG4, which may enhance the proliferation ability of cancer cells and affect their epithelial-mesenchymal transformation status, leading to changes in invasion and migration abilities. These factors can result in phenotypic changes that contribute to a poor prognosis (Zhou et al., 2021). Upregulation of GNG4 has been reported to be associated with poor prognosis in gallbladder cancer and colorectal cancer, with a negative correlation with overall survival rates. Silencing of GNG4 leads to increased apoptosis, potentially inhibiting cell proliferation. Reduction in GNG4 expression can also cause colorectal cancer (CRC) cells to become arrested in the S phase of the cell cycle. Studies by Yang et al. and Liang et al. have suggested that the upregulation of GNG4 in CRC may influence tumor prognosis through the PI3K-signaling pathway (Liang et al., 2021). However, clear evidence of the involvement of any specific pathway in GNG4's role in tumors is not yet known.

3.5.2.3 Glutathione S-Transferase Mu 1 (GSTM1)

GSTM1 refers to the Glutathione S-Transferase Mu 1 gene, which is responsible for producing the Glutathione S-Transferase Mu 1 enzyme. This gene is highly polymorphic, and one of the common variations is the GSTM1 null genotype, which results in the absence of GSTM1 enzyme production. This absence is associated with an increased risk of various cancers, including lung, colorectal, breast, bladder, cervical dysplasia, and head and neck cancers (Chen et al., 2012; Guo et al., 2012; Ryk et al., 2021). Downregulation of GSTM1 is frequently observed in different cancer types.

In the treatment of cervical cancer, a drug called doxorubicin is used, and it has been reported that this drug can upregulate GSTM1. This upregulation may be associated with drug resistance in some individuals. However, GSTM1 null individuals have shown a better response in breast and ovarian cancer (Drozd et al., 2016). Additionally, Ay et al. reported in their study that GSTM1 plays a role in cancer prevention, such as in colorectal cancer, through its involvement in the neutralization process. They also noted upregulation of GSTM1 in certain contexts.

When studying the association between the GSTM1 null genotype and specific populations, various studies have concluded that Turkish, Asian, Malaysian, Caucasian, Saudi Arabian, and Brazilian populations with homozygous GSTM1 deletions are at a higher risk of breast cancer, colorectal cancer, and other cancer types (Ay et al., 2021). However, the exact pathways and mechanisms underlying these associations are not yet clearly understood.

3.5.2.4 Aldehyde Dehydrogenase 1 Family Member A1 (ALDH1A1)

ALDH1A1 is involved in the synthesis of retinoic acid and plays a crucial role in cellular metabolism. It has diverse functions in tumorigenesis, contributing to stem cell cancer, tumor progression, chemotherapy resistance, and prognosis. Overexpression of ALDH1A1 is associated with poor clinical outcomes in breast, lung, liver, and ovarian cancer (Yong et al., 2018; Li et al., 2018; Moreb et al., 2017; Visus et al., 2011).

A study by Condello et al. reported that β-catenin, a protein involved in cell adhesion and gene transcription through the β-catenin pathway, regulates ALDH1A1. In ovarian cancer, up-regulation of ALDH1A1 has been reported, and abnormal activation of β-catenin is associated with tumor growth and prognosis. Therefore, targeting ALDH1A1 has been proposed as a potential treatment option in ovarian cancer, with the aim of inhibiting its activity (Condello et al., 2014). It is also considered a potential target for breast cancer treatment (Wang et al., 2018).

Upregulation of ALDH1A1 has been linked to chemotherapy resistance in various cancers, including breast, cervical, lung, and colorectal cancers (Poturnajova et al., 2021). While the clear mechanisms of ALDH1A1 are well-documented in colorectal cancer (CRC) and uterine corpus endometrial carcinoma (UCEC), different pathways have been reported in other cancers, such as the retinoic acid pathway in cancer stem cells (Yue et al., 2022).

Moreover, overexpression of ALDH1A1 has been found to play an important role in obesity, diabetes, and other diseases (Poturnajova et al., 2021).

3.5.2.5 Carbamoyl-Phosphate Synthase 1 (CPS1)

The CPS1 (carbamoyl phosphate synthase 1) gene encodes an enzyme called carbamoyl phosphate synthetase 1, which plays a crucial role in the metabolic pathway known as the

urea cycle. The urea cycle is responsible for removing toxic ammonia from the body. In several studies involving hepatic, colorectal, and pancreatic cancers, altered expression of CPS1 has been observed. Dysregulation of CPS1 has been linked to cell proliferation (Tsai et al., 2017; Biancur et al., 2017; Watanabe et al., 2016).

In particular, upregulation of CPS1 has been reported in various cancers, including stomach, melanoma, sarcoma, B cell lymphoma, lung cancer, glioma, and glioblastoma (GBM). In non-small cell lung carcinoma (NSCLC), KRAS mutation is responsible for the upregulation of CPS1. Conversely, silencing CPS1 has been shown to induce cell death and inhibit tumor growth. Furthermore, CPS1 upregulation beyond NSCLC has been associated with the loss of the P53 gene, as P53 is involved in CPS1 regulation.

Studies have demonstrated that the suppression of neurotensin (NTS), interleukin-6 (IL-6), or IL-6 signal transducer (IL6ST) can lead to the suppression of CPS1. Consequently, suppressing CPS1 expression is considered a potential biomarker and therapeutic target in the treatment of NSCLC (Hajaj et al., 2021).

3.5.3 Common Downregulated genes in UCEC and CRC

Following are the tables representing common downregulated genes in CRC and UCEC of APC group:

Table 12: Common downregulated Genes in UCEC, influenced by APC Gene Mutations

UCEC_APC Mut VS NonAPC Mut			APC_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
HOMER2	-2.8	MSS_Hyp_NO_APC	**HOMER2**	-1.78	MSS_NonHyper
FGF19	-2.04	MSS_Hyp_NO_APC	**FGF19**	-1.70	MSS_NonHyper

Table 13: Common downregulated Genes in CRC, influenced by APC Gene Mutations

CRC_APC Mut VS NonAPC Mut			APC_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
HOMER2	-4.83	MSS_Hyp_NO_APC	**HOMER2**	1.34	MSS_HYP_WITH_APC
FGF19	-2.46	MSS_Hyp_NO_APC	**FGF19**	-0.52	MSS_NonHyper

These tables reveal the shared downregulated genes resulting from APC mutations. They also discern between two specific groups: one composed of hypermutated MSS APC mutated samples and non-mutated samples, and the other comparing APC mutated samples with those lacking hypermutation. Notably, FGF19 exhibits substantial downregulation within the group of non-APC mutated samples.

3.5.3.1 Homer Scaffold Protein 2 (HOMER2)

HOMER2 (Homer Protein Homolog 2) is a gene that plays a crucial role in synaptic signaling and neuronal plasticity. Its function in cancer appears to be complex and context-dependent. In hepatocellular carcinoma (HCC), for example, HOMER2 is found to be downregulated, suggesting a potential tumor suppressive role (Luo et al., 2021). However, contradictory findings have been reported in endometrioid endometrial adenocarcinoma patients, where higher HOMER2 expression was associated with a better response (Mhawech et al.). Likewise, in Rectal Cancer (RC), increased HOMER2 expression was correlated with reduced overall survival (Sun et al., Luo et al., 2021). These studies suggest that HOMER2's function may differ in different types of carcinomas.

The dysregulation of the Homer protein family, including HOMER2, has been validated through various studies in multiple carcinomas. Therefore, the dysregulation of HOMER2 itself could potentially serve as a prognostic biomarker.

Aside from its role in cancer, HOMER2 is also involved in the regulatory processes of muscle cells. Muscle cells continuously undergo protein degradation of damaged proteins, and HOMER2 has been reported to have a significant effect on this degradation process. Decreased levels of HOMER2 enhance this degradation process, which can impact muscle health (Bortoloso et al., 2013; Kamal & Lawler, 2023).

3.5.3.2 Fibroblast Growth Factor 19 (FGF19)

FGF19 collaboratively interacts with various signaling pathways, including the epidermal growth factor receptor, Wnt/β-catenin, endoplasmic reticulum-related signaling pathway, STAT3/IL6, RAS, and extracellular signal-regulated protein kinase, which are involved in the development of primary liver cancer (Chen et al., 2021).

Inhibiting FGF19 holds promise as a treatment strategy for malignancies associated with FGF19, including colon and liver cancers (Desnoyers et al., 2007). In breast cancer, FGF19 upregulates the AKT signaling pathway and downstream PI3K/AKT signaling. Therefore, targeting FGF19 could be a potential treatment for basal-like breast cancer (Tiong et al., 2016).

In lung squamous cell carcinoma (LSCC), downregulation of FGF19 has been shown to significantly inhibit tumor cell growth and enhance apoptosis (Zhang et al., 2017). In ovarian cancer, FGF19 activates the p38 MAPK pathway, leading to increased chemoresistance. This suggests that FGF19 could be a promising therapeutic target for ovarian cancer (Zhu et al., 2023).

3.6 Zinc Finger HIT-Type Containing 6 (ZNHIT6)

ZNHIT6 is a protein coding gene that plays a role in breast cancer antigen immunogenicity and is involved in coding for the Nop56 protein, which is part of the box C/D small nucleolar RNA (snoRNA) complex (Souza Melo et al., 2015). It belongs to the zinc finger HIT protein family and is crucial for the proper biogenesis of small nucleolar ribonucleoproteins (snoRNPs) involved in ribosomal RNA modification and processing. The ZNHIT6 protein interacts with the R2TP pathway for snoRNP biogenesis (Verheggen et al., 2015; Cloutier et al., 2017).

Next, we performed pathway analysis for all these genes, including ZNHIT6, using the DAVID tool and validated the results through literature review. Interestingly, annotations were found for the other genes, indicating their involvement in various pathways. However, no pathway association was reported for ZNHIT6 in conjunction with any of the upregulated or downregulated genes within the same pathway.

These findings suggest that ZNHIT6 may have distinct functions or interactions that are not yet fully understood or reported in relation to the identified set of upregulated and downregulated genes. Further investigation is needed to elucidate the specific role of ZNHIT6 in the context of these pathways and its potential implications in MSS hypermutated samples.

3.6.1 cBioPortal Results

In the cBioPortal results, when comparing UCEC cancer samples with hypermutated MSS that have ZNHIT6 mutations to those without (Group_1), we identified a total of 89 upregulated genes, with LINC02381 being the most prominently upregulated. Additionally, there were 127 downregulated genes, and PIGR was the most significantly downregulated gene. In the same vein, when analyzing UCEC samples, specifically comparing those with ZNHIT6 mutations to non-hypermutated samples (Group_2), we found 234 upregulated genes, with TUBB2B being the most highly upregulated. Moreover, there were 100 downregulated genes, and CXCL9 was the most markedly downregulated gene.

Shifting our focus to CRC, within Group_1, we observed 256 upregulated genes, with GABRB2 ranking as the most upregulated gene. In contrast, 94 genes were downregulated, and ANKRD30B was the most significantly downregulated gene. Conversely, in the Group_2 analysis for CRC, we noted 167 upregulated genes, with CALB1 as the most prominently upregulated, and 166 genes were downregulated, with SLC39A5 being the most notably downregulated gene.

3.6.2 Common Upregulated Genes in UCEC and CRC

Following are the tables representing common upregulated genes in CRC and UCEC of ZNHIT6 group:

Table 14: Common upregulated Genes in UCEC, influenced by ZNHIT6 Gene Mutations

UCEC_ZNHIT6 Mut VS NonZNHIT6 Mut			ZNHIT6_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
DPT	2.81	MSS_HYP_WITH_ZNHIT6	**DPT**	3.11	MSS_HYP_WITH_ZNHIT6
FDCSP	2.97	MSS_HYP_WITH_ZNHIT6	**FDCSP**	2.66	MSS_HYP_WITH_ZNHIT6
HOXC4	2.03	MSS_HYP_WITH_ZNHIT6	**HOXC4**	0.58	MSS_HYP_WITH_ZNHIT6
HOXC5	3.23	MSS_HYP_WITH_ZNHIT6	**HOXC5**	1.47	MSS_HYP_WITH_ZNHIT6
LMO3	3.43	MSS_HYP_WITH_ZNHIT6	**LMO3**	1.15	MSS_HYP_WITH_ZNHIT6
LRRN4	2.37	MSS_HYP_WITH_ZNHIT6	**LRRN4**	1.45	MSS_HYP_WITH_ZNHIT6
PCOLCE2	2.29	MSS_HYP_WITH_ZNHIT6	**PCOLCE2**	0.16	MSS_HYP_WITH_ZNHIT6
RELN	2.19	MSS_HYP_WITH_ZNHIT6	**RELN**	1.38	MSS_HYP_WITH_ZNHIT6
RTN4RL1	2.26	MSS_HYP_WITH_ZNHIT6	**RTN4RL1**	1.29	MSS_HYP_WITH_ZNHIT6
SLC16A7	2.5	MSS_HYP_WITH_ZNHIT6	**SLC16A7**	1.88	MSS_HYP_WITH_ZNHIT6
SLITRK5	2.63	MSS_HYP_WITH_ZNHIT6	**SLITRK5**	2.17	MSS_HYP_WITH_ZNHIT6

Table 15: Common upregulated Genes in CRC, influenced by ZNHIT6 Gene Mutations

CRC_ZNHIT6 Mut VS NonZNHIT6 Mut			ZNHIT6_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
DPT	2.03	MSS_HYP_WITH_ZNHIT6	**DPT**	-0.17	MSS_NonHyp
FDCSP	2.17	MSS_HYP_WITH_ZNHIT6	**FDCSP**	1.45	MSS_HYP_WITH_ZNHIT6
HOXC4	2.55	MSS_HYP_WITH_ZNHIT6	**HOXC4**	1.54	MSS_HYP_WITH_ZNHIT6
HOXC5	2.01	MSS_HYP_WITH_ZNHIT6	**HOXC5**	1.96	MSS_HYP_WITH_ZNHIT6
LMO3	3.29	MSS_HYP_WITH_ZNHIT6	**LMO3**	-0.15	MSS_NonHyp
LRRN4	3.57	MSS_HYP_WITH_ZNHIT6	**LRRN4**	1.78	MSS_HYP_WITH_ZNHIT6
PCOLCE2	2.66	MSS_HYP_WITH_ZNHIT6	**PCOLCE2**	1.93	MSS_HYP_WITH_ZNHIT6
RELN	2.8	MSS_HYP_WITH_ZNHIT6	**RELN**	1.24	MSS_HYP_WITH_ZNHIT6
RTN4RL1	2.59	MSS_HYP_WITH_ZNHIT6	**RTN4RL1**	1.76	MSS_HYP_WITH_ZNHIT6
SLC16A7	2.87	MSS_HYP_WITH_ZNHIT6	**SLC16A7**	2.22	MSS_HYP_WITH_ZNHIT6
SLITRK5	2.24	MSS_HYP_WITH_ZNHIT6	**SLITRK5**	0.75	MSS_HYP_WITH_ZNHIT6

These tables highlight common upregulated genes attributed to ZNHIT6 mutations, while also distinguishing between two distinct groups: hypermutated MSS ZNHIT6 mutated samples versus non-mutated samples, and ZNHIT6 mutated samples versus non-hypermutated ones. Notably, within the group of non-ZNHIT6 mutated and non-hypermutated samples, LMO3 in UCEC and LRRN4 in CRC stands out as the genes with the most pronounced upregulation.

3.6.2.1 Homeobox C5 (HOXC5)

HOXC5 is a member of the highly conserved homeobox family of transcription factors, and its primary role is in embryogenesis and development. Generally, HOX proteins are known to be involved in the regulation of various noncoding RNAs, contributing to chemoresistance. This family has been reported to play a role in cancer resistance, specifically in lung cancer, breast cancer, hepatocellular carcinoma (HCC), and myeloid

leukemia. Interestingly, overexpression of HOXC5 has been found to suppress cancer cell growth.

In squamous cell carcinoma cells, HOXC5 overexpression has been observed compared to normal tissues. Further confirmation of HOXC5 upregulation was obtained by analyzing TCGA data, which revealed a similar pattern not only in oral squamous cell carcinoma (OSCC) but also in head and neck squamous cell carcinoma (HNSCC) cells. Detailed analysis results indicated that HOXC5 upregulation is correlated with oral carcinogenesis and significantly contributes to OSCC development. Therefore, HOXC5 may serve as a robust biomarker for OSCC (Moon et al., 2012).

Additionally, HOXC5 upregulation has been reported in other cancer types, including cervical cancer, breast cancer, prostate cancer, and bladder cancer (Bhatlekar et al., 2014)

3.6.2.2 LIM Domain Only 3 (LMO3)

LIM domain only 3 (LMO3) interacts with P53 and regulates its function. It also interacts with transcription factors to regulate targeted genes during embryonic development, thereby playing a role in nervous system development. In hepatocellular carcinoma (HCC), upregulation of LMO3 has been reported. LMO3 promotes HCC cell invasion by suppressing the Hippo signaling pathway. In gastric cancer, it induces cell invasion through the Akt-mTOR/GSK3β signaling pathway. LMO3's interaction with other genes, such as HEN2, also promotes cell growth in neuroblastoma. As a result, LMO3 holds promise as a potential therapeutic target for cancer treatment in the future (Cheng et al., 2018).

LMO3 is involved in cell proliferation in papillary thyroid carcinoma through the regulation of LIMK1-mediated cofilin and the β-catenin pathway. In almost all of these cancers, upregulation of LMO3 has been reported. Consequently, knocking down LMO3 has been shown to suppress cancer cell proliferation. In glioma cells, LMO3 knockdown promotes apoptosis (Qiu et al., 2018; Zhang et al., 2020).

3.6.3 Common Downregulated Genes in UCEC and CRC

Following are the tables representing common downregulated genes in CRC and UCEC of ZNHIT6 group:

Table 16: Common downregulated Genes in UCEC, influenced by ZNHIT6 Gene Mutations

UCEC_ZNHIT6 Mut VS NonZNHIT6 Mut			ZNHIT6_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
IHH	-2.74	Hyp_MSS_NO_ZNHIT6	**IHH**	-1.91	MSS_NonHyper
UGT2B7	-2.51	Hyp_MSS_NO_ZNHIT6	**UGT2B7**	-1.48	MSS_NonHyper
OGDHL	-2.22	Hyp_MSS_NO_ZNHIT6	**OGDHL**	-2.21	MSS_NonHyper

Table 17: Common downregulated Genes in CRC, influenced by ZNHIT6 Gene Mutations

CRC_ZNHIT6 Mut VS NonZNHIT6 Mut			ZNHIT6_Mut VS MSS Non_Hyper		
GENES	**Log2 Ratio**	**High Expression in**	**GENES**	**Log2 Ratio**	**High Expression in**
IHH	-2.21	Hyp_MSS_no_ZNHIT6	**IHH**	-3.11	MSS_NonHyper
UGT2B7	-2.6	Hyp_MSS_no_ZNHIT6	**UGT2B7**	-1.31	MSS_NonHyper
OGDHL	-2.58	Hyp_MSS_no_ZNHIT6	**OGDHL**	-0.17	MSS_NonHyper

These tables present a collection of common genes that undergo downregulation as a consequence of ZNHIT6 mutations. They further differentiate between two specific groups: hypermutated MSS ZNHIT6 mutated samples compared to non-mutated samples, and ZNHIT6 mutated samples contrasted with those lacking hypermutation. Worth noting is that OGDHL in UCEC and IHH in CRC displays significant downregulation, particularly within the subset of non-ZNHIT6 mutated samples.

3.6.3.1 Indian Hedgehog (IHH)

Indian Hedgehog (IHH) is involved in embryonic development, tissue maintenance, bone growth, and differentiation. It also plays a crucial role in regulating cellular processes such as cell differentiation and proliferation. Dysregulation of IHH has been linked to the development of cancer, including its involvement in the self-renewal of cancer stem cells.

In endometrial cancer, the downregulation of IHH inhibits the IHH pathway, subsequently leading to the activation of the Wnt pathway. This activation is accompanied by the suppression of its negative regulator, Wif1, which is responsible for tumor development (Lian et al., 2006). Similarly, IHH is downregulated in colon cancer, to the extent that its diminished expression serves as an early indicator of tumor development. Intriguingly, it has been observed that all colon cells exhibiting downregulation of IHH also harbor APC mutations, which are pivotal components of the Wnt signaling pathway. Therefore, it has been suggested that the upregulation of IHH, leading to the blockage of the Wnt pathway, could represent a potential target for cancer treatment (Fu et al., 2006).

3.6.3.2 UDP Glucuronosyltransferase Family 2 Member B7 (UGT2B7)

UDP-glucuronosyltransferase 2B7 (UGT2B7) is an enzyme involved in the glucuronidation process, which plays a crucial role in eliminating endogenous and exogenous compounds from the body. It helps regulate the levels of various substances, including steroid hormones, bile acids, and retinoids, all of which have implications in various diseases.

In endometrial cancer patients, downregulation of UGT2B7 has been reported to disrupt estrogen homeostasis and estrogen metabolism (Zhao et al., 2020). Similarly, in patients with hepatocellular carcinoma (HCC), UGT2B7 downregulation has been documented. Additionally, there have been reports of altered enzymatic activities of UGT2B7 in ovarian cancer and colorectal cancer (CRC) (Shen et al., 2019). An experimental study conducted on renal cell carcinoma (RCC) shed light on the downregulation of UGT2B7 in mRNA expression. This downregulation was associated with changes in its enzymatic activity and responses in RCC patients. These findings suggest that UGT2B7 may serve as a potential biomarker for endometrial cancer and renal cell carcinoma (Matsumoto et al., 2022).

3.7 C2 Calcium-Dependent Domain Containing 5 (C2CD5)

C2CD5 is essential for insulin-stimulated glucose transport. Strong clinical and epidemiological evidence indicates a significant association between hyperinsulinemia, obesity-related factors, and the development of various cancers, including breast, endometrial, colon, liver, esophageal, kidney, and pancreatic cancers. Insulin resistance resulting from excessive fat accumulation leads to elevated insulin levels, which promote tumor growth by enhancing cell proliferation and inhibiting cell growth regulation mechanisms. Moreover, elevated levels of Insulin-like growth factors (IGFs), chronic low-grade inflammation, and adipose tissue secretion further exacerbate hyperinsulinemia, contributing to cancer progression. (Kasprzak, 2021; Hua et al., 2020; Poloz & Stambolic, 2015) A modeling analysis of insulin signaling in insulin resistance states and cancer in 2016 suggests that disrupted insulin signaling, and insulin resistance play significant roles in tumor growth and progression. Elevated insulin levels in insulin resistance states activate pathways that promote cell growth, increase cell proliferation, and hinder cell death mechanisms, creating an environment conducive to cancer cell survival. (Bertuzzi et al., 2016)

In 2019, Molinaro et al. investigated insulin-driven PI3K-AKT signaling in hepatocytes and its regulation by PI3K isoforms (PI3Kα and PI3Kβ) and RAS. The study revealed that both PI3Kα and PI3Kβ redundantly transmit insulin signals, while RAS activation amplifies PI3K and AKT activation, promoting insulin signaling. These findings provide insights into the understanding and potential targeting of insulin signaling pathways in hepatocyte-related conditions such as diabetes and liver cancer. (Molinaro et al., 2019) Metabolic health is crucial for breast cancer development, and insulin signaling, and resistance may contribute to this association. Studies have identified insulin as a potential factor in breast cancer pathogenesis, highlighting the importance of considering metabolic health in cancer research. (Yee et al., 2020)

3.8 Raftlin Family Member 2 (RFTN2)

RFTN2 is a protein that plays a role in immune responses. When stimulated by bacterial lipopolysaccharide, it facilitates the internalization of TLR4, a receptor involved in

recognizing pathogens, in dendritic cells. This internalization process involves clathrin and leads to the activation of TICAM1-mediated signaling, ultimately resulting in the production of IFNB1, an important immune response molecule. RFTN2 is also associated with the regulation of B-cell antigen receptor (BCR)-mediated signaling. (Chaudhari et al., 2020) The BCR is expressed on the surface of B cells and is responsible for binding antigens. The BCR signaling pathway has been implicated in positive leukemia, and efforts have been made to target the BCR-ABL signaling pathway in therapy-resistant Philadelphia chromosome-positive leukemia. These studies highlight the importance of understanding and modulating these pathways in the context of immune responses and anti-tumor effects. (Hare T et al., 2011)

3.9 Neurofibromatosis Type 1 (NF1)

Mutations in the NF1 gene result in the loss of neurofibromin, a negative regulator of the RAS signaling pathway (Pearson et al., 2020). In breast cancer, NF1 mutations have been identified as drivers of increased tumor growth and resistance to endocrine therapy (Sokol et al., 2019).

These mutations lead to dysregulated signaling pathways involved in cell cycle regulation, DNA repair, and cell survival, contributing to enhanced tumor growth and therapy resistance (Pearson et al., 2020). The dysregulated RAS signaling pathway has emerged as a crucial driver in NF1 mutated tumors, and targeted therapies aimed at this pathway, such as RAS pathway inhibitors, are being developed to improve treatment outcomes (Tao et al., 2020). Understanding the molecular mechanisms and signaling pathways associated with NF1 mutations offers potential strategies for personalized and more effective treatments in breast cancer patients with NF1 mutations.

In pancreatic cancer, oncogenic KRAS activation has been found to engage the RSK1/NF1 pathway, leading to the inhibition of wild-type RAS signaling and promoting tumor growth (Cheng et al., 2021). This intricate interplay between oncogenic KRAS and the RSK1/NF1 pathway provides insights into the complex regulation of RAS signaling in pancreatic cancer and may hold promise for identifying therapeutic targets (Cheng et al., 2021). Furthermore, the loss of NF1 function has been implicated in acquired resistance to

endocrine therapy in lobular breast cancer (Sokol et al., 2019). These findings highlight the clinical significance of NF1 mutations and suggest that targeting NF1 or its associated signaling pathways could potentially overcome endocrine therapy resistance and improve treatment outcomes in lobular breast cancer patients (Sokol et al., 2019). Continued research in these areas may pave the way for tailored therapeutic approaches and personalized medicine strategies in NF1-mutated tumors.

In a study by Zheng et al. (2020), NF1 frameshift and nonsense mutations were found to be associated with a higher risk of breast cancer recurrence and death in ER+ patients. Among these mutations, R2450* was the most frequently observed non-silent NF1 mutation, leading to the absence of a small portion of the NF1 protein at the C-terminus (Ellison-Zelski & Alarid, 2010).

Other studies have also identified the R2450* mutation in patients, demonstrating reduced protein expression and non-functional NF1 protein (Whittaker et al., 2013; Gibney & Smalley, 2013). Additionally, a genomic analysis of endometrioid endometrial carcinoma (EEC) in Chinese patients revealed NF1 mutations in nine EEC patients, including the recurrent mutational site p.R2450 (Wang et al., 2019). These findings highlight the significance of NF1 mutations, particularly the R2450 mutation, in breast cancer and EEC, providing insights into their potential impact on disease progression and treatment response.

3.10 Empirical analysis of digital gene expression data in R (EdgeR) Results

Result of EgdeR script execution showed that common downregulated genes found in MSS group of UCEC, CRC and STAD are CXCL5 and MUC5AC.

3.10.1 C-X-C Motif Chemokine Ligand 5 (CXCL5)

CXCL5 belongs to chemokines class that functions with G-coupled proteins. It is involved in angiogenesis. CXCL5 has the capacity to attract immune cells like T/B lymphocytes and eosinophils to specific locations during immune responses. Moreover, it plays a role in facilitating the adhesion and restructuring of connective tissues. The expression of CXCL5 is reported higher in CRC, GC and HCC. It is reported to be induced by interleukin-17A

(IL-17A) or interleukin-1β (IL1β). It promotes programmed cell death protein 1 (PD-L1) in tumor cells. In breast cancer, CXCL5 secretion promotes tumor development. CXCL5 recruits and regulates the functions of neutrophils. Blockage of transforming growth factor-β (TGF-β) signaling, promotes tumor growth in melanomas and breast cancers. CXCL5 is secreted in RCC, NSCLC, GC, BC, prostrate, cervical, lung cancers, hepatoblastoma, and osteosarcoma is associated with proliferation. Blockage of CXCL5 through PI3K inhibitor in AKT and GSK-3β signaling pathway inhibits invasive behavior of tumor cells (Zhang et al., 2020).

3.10.2 Mucin 5AC (MUC5AC)

Mucin 5AC (MUC5AC), a member of the glycoprotein family, is expressed in epithelial cells in various locations, including the upper and lower respiratory tracts, stomach, and endocervix. Its interaction with integrin β4 has been implicated in facilitating lung cancer. Deregulation of p53 and β-catenin enhances colorectal cancer (CRC), and repression of apoptosis has been reported in pancreatic cancer. However, low or absent expression of MUC5AC has been observed in renal cancer, breast cancer, prostate cancer, sarcomas, lymphomas, endocrine tumors, and various skin tumors. Interestingly, many of the cancers associated with MUC5AC are found in organs where it is normally expressed. The regulation of cancer by MUC5AC involves MAPK and Akt/PI3K pathways. Its presence in multiple cancers suggests its potential utility as a biomarker for treatment (Rico et al., 2021).

3.10.3 Pathway Analysis

The pathway analysis of edgeR common genes found in UCEC, CRC and STAD on DAVID revealed only signal pathway common in both the genes I.e., MUC5AC and CXCL5 named IL-17 signaling pathway was found.

3.10.3.1 Interleukin-17 Signaling Pathway (IL-17)

I Interleukin-17 (IL-17) is a protein in our body that's involved in immune responses and inflammation. It's made by certain immune cells called T-helper 17 cells. Among its family members, IL-17A is the one we know most about. IL-17 can cause inflammation in tumors, which can affect cancer in different ways.

One important role of IL-17 in cancer is promoting the spread of cancer cells to other parts of the body. It does this by changing how cancer cells behave, making them better at moving around. It can also help cancer cells grow more and resist dying, which makes cancer worse. IL-17 can activate signals that support cancer cell growth, and it can make the body create substances that help tumors grow blood vessels. These blood vessels bring tumors the nutrients and oxygen they need to get bigger (Li et al., 2020).

In the IL-17 process, two other molecules, MUC5AC and CXCL5, are involved. When IL-17 connects with its receptor on target cells, it sets off a chain reaction. This leads to the creation of various molecules, including MUC5AC. So, when we see more MUC5AC, it can be because of IL-17 causing inflammation.

CXCL5, on the other hand, is like a messenger in the IL-17 pathway. When IL-17 binds to its receptor, it tells cells to produce CXCL5. Then, CXCL5 acts like a signal, calling immune cells like T and B lymphocytes and eosinophils to come to the place where there's inflammation. This is the body's way of fighting infections and repairing damaged tissue, and CXCL5 helps coordinate this process.

The IL-17 pathway has emerged as a potential target for cancer therapy. Blocking IL-17 or its receptors may be a strategy to mitigate the pro-tumorigenic effects of this pathway, studies on this topic are in progress. Particularly in immunotherapy, its blockage improves the efficacy of immunotherapy (Zhang et al., 2020; Shuai et al., 2020)

CHAPTER 04

DISCUSSION

The major findings of this research include the identification of multiple biomarkers in hypermutated microsatellite-stable cancers, the upregulation or downregulation of which can be potentially used as therapeutic agents for treatment. There is reported evidence of the involvement of these biomarkers in different pathways, which gives us a better understanding of how these biomarkers work and how they could be used in immunotherapy

Hypermutation in cancer refers to an elevated number of mutations compared to normal cells. It is divided into two main categories based on molecular subtyping: microsatellite instability (MSI) and microsatellite stability (MSS). MSI occurs when mutations in microsatellite regions result from a deficient mismatch repair mechanism. Hypermutated microsatellite stability is associated with germline or somatic mutations in the POLE gene. This gene plays a critical role in the DNA mismatch repair mechanism, which corrects errors in DNA sequences that occur during replication. Any malfunction in this repair mechanism can lead to an accumulation of mutations, increasing the risk of cancer (Nature, 2012; Barbari and Shcherbakova, 2017). Among the reported mutations, somatic mutations are the most common. Notably, P286R and V411L are among the highly reported somatic mutations in the POLE gene and are recognized as driver mutations in various types of cancer. These in-depth studies aim to discover better treatment options for cancer patients.

Immunotherapy, the latest and most advanced treatment method for cancer, is side-effect-free. It involves preparing an individual's immune system to combat cancer through oral or intravenous injections. Immunotherapy is significantly more effective in MSI patients than in MSS patients. One reason for this is the higher number of mutations in MSI, resulting in more available biomarkers to target. In contrast, MSS tumors are stable and have fewer mutations, leading to a lack of suitable biomarkers for immunotherapy targeting

In this study, we focused on seven cancers: UCEC, COAD, READ, STAD, PAAD, UCS, and ESCA. These cancers were selected based on previous reports highlighting their status as highly hypermutated cancers. We obtained their TCGA data from the GDC data portal. After separating the hypermutated MSS samples from the others, we found that POLE had the highest number of mutations, reaffirming its importance in MSS cancers. We shortlisted some genes from a total of 42 hypermutated MSS samples based on a literature

review. To identify new biomarkers specifically for hypermutated MSS cancers, we conducted a detailed analysis of POLE, APC, and ZNHIT6. APC is a tumor suppressor gene involved in the Wnt signaling pathway, often associated with early-onset CRC. ZNHIT6 is a zinc finger HIT family gene that plays a crucial role in the biogenesis of small nucleolar ribonucleoproteins. POLE, on the other hand, is the major gene responsible for MSS status.

In the cBioPortal transcriptomic analysis of these genes, we focused on UCEC and CRC cancers, as they had the highest number of hypermutated MSS samples in our data. Among POLE-mutated samples, a total of three genes—DKK4, FGF20, and HMX2—were found to be consistently upregulated. While their involvement in different cancers, including CRC, has been reported, significant literature suggests their association with UCEC cancer. Similarly, among APC-mutated samples, TSIX, GNG4, GSTM1, ALDH1A1, and CPS1 were consistently upregulated. While their involvement in cancer has been reported, there is no significant evidence linking them to specific pathways that could confirm their role in cancer. For ZNHIT6-mutated samples, HOXC5 and LMO3 were commonly upregulated. Although there is insufficient evidence of their involvement in CRC and UCEC cancers in the literature, our cBioPortal analysis results suggest that similar mechanisms could be at play in these cancers. Additionally, the up and downregulation observed in our results may be influenced by hotspot mutations in the reference genes used or vice versa, including POLE P286R, APC R1114*, and ZNHIT6 R2588*. While clear evidence is lacking, our results suggest a potential role for these genes in CRC and UCEC. Because of time constraints and the absence of conclusive evidence regarding the significant roles of other genes and their hotspot mutations in cancer, additional mRNA analysis for other genes was not carried out.

Likewise, the presence of a POLE P286R mutation corelated the downregulation of AKR1B10 and AKR1C2, while an R1114* APC mutation corelated with downregulation of HOMER2 and FGF19. Furthermore, an R258* ZNHIT6 mutation was associated with the downregulation of IHH, UGT2B7, and OGDHL. Consequently, the modulation of the expression levels of these genes due to specific mutations found in cancer holds promise as a potential avenue for therapeutic interventions.

For additional assurance and the identification of potential biomarkers in various cancers through differential gene expression analysis, EdgeR, a commonly used tool in Next-Generation Sequencing (NGS) studies, was employed. As a result, CXCL5 and MUC5AC were identified as common downregulated genes in hypermutated MSS samples of Uterine Corpus Endometrial Carcinoma (UCEC), Colorectal Cancer (CRC), and Stomach Adenocarcinoma (STAD). CXCL5 plays a vital role in immune responses by attracting T/B lymphocytes and eosinophils to specific locations, and it has been associated with programmed cell death of tumor cells. Additionally, its higher expression in CRC is induced by the IL-17A pathway. On the other hand, MUC5AC, typically expressed in both upper and lower respiratory tracts, has been implicated in the MAPK pathway. Dysregulation of this pathway is known to contribute to cancer development. Specifically in CRC, MUC5AC is involved in the deregulation of P5, a tumor suppressor gene.

Subsequently in pathway analysis, both MUC5AC and CXCL5 were found to share a common pathway, namely, the interleukin-17 (IL-17) signaling pathway. This pathway possesses a pro-inflammatory core that can either bolster or inhibit tumor growth, contingent on various contributing factors. In the context of inflammation induced by IL-17, MUC5AC has a propensity to overexpress. Consequently, it may have a role to play in cancer development. In contrast, CXCL5 is more directly linked to IL-17 and its role in promoting apoptosis. Elevated levels of CXCL5 have also been reported in the presence of IL-17. Hence, while MUC5AC's connection to the IL-17 pathway is indirect, CXCL5 could be considered a more direct participant in IL-17-mediated processes, including its potential involvement in cancer.

Conclusion

Immunotherapy is currently one of the most effective treatment methods for cancer. However, a major hurdle in its application to hypermutated microsatellite stable (MSS) cancers is the lack of biomarkers. Identifying more biomarkers is crucial to overcome this challenge.

Based on transcriptomic analysis, we observed the upregulation of certain genes as potential biomarkers. Upregulation of DKK4, FGF20, and HMX2 associated with POLE P286R mutation, upregulation of TSIX, GNG4, GSTM1, CPS1, and ALDH1A1 interconnected with the R1114* mutation of APC, and upregulation of HOXC5 and LMO3 interconnected to R258* ZNHIT6 mutation were noted. Additionally, downregulation of AKR1B10 and AKR1C2 corelated with POLE P286R mutation, downregulation of HOMER2 and FGF19 corelated with R1114* APC mutation, and downregulation of IHH, UGT2B7, and OGDHL corelated with R258* ZNHIT6 mutation were observed. These gene expression patterns may serve as therapeutic biomarkers for immunotherapy.

Furthermore, in transcriptomic analysis, we identified CXCL5 and MUC5AC, and their potential roles in the IL-7 signaling pathway as potential biomarkers for immunotherapy. However, it's worth noting that these genes are not directly affected in this pathway. Therefore, an in-depth study is required to understand the exact mechanism.

CHAPTER 05

REFERENCES

References

Abalo, K. D., Rage, E., Leuraud, K., Richardson, D. B., Le Pointe, H. D., Laurier, D., & Bernier, M. O. (2020, September 10). Early life ionizing radiation exposure and cancer risks: systematic review and meta-analysis. Pediatric Radiology, 51(1), 45–56. https://doi.org/10.1007/s00247-020-04803-0

Aghabozorgi, A. S., Bahreyni, A., Soleimani, A., Bahrami, A., Khazaei, M., Ferns, G. A., Avan, A., & Hassanian, S. M. (2019, February). Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. Biochimie, 157, 64–71. https://doi.org/10.1016/j.biochi.2018.11.003

Aljarf, R., Shen, M., Pires, D. E. V., & Ascher, D. B. (2022, June 21). Understanding and predicting the functional consequences of missense mutations in BRCA1 and BRCA2. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-13508-3

Autier, P., & Doré, J. F. (2020, April). Ultraviolet radiation and cutaneous melanoma: a historical perspective. Melanoma Research, 30(2), 113–125. https://doi.org/10.1097/cmr.0000000000000609

Baeissa, H., Benstead-Hume, G., Richardson, C. J., & Pearl, F. M. (2017, February 19). Identification and analysis of mutational hotspots in oncogenes and tumour suppressors. Oncotarget, 8(13), 21290–21304. https://doi.org/10.18632/oncotarget.15514

Bhatlekar, S., Fields, J. Z., & Boman, B. M. (2014, July 5). HOX genes and their role in the development of human cancers. Journal of Molecular Medicine, 92(8), 811–823. https://doi.org/10.1007/s00109-014-1181-y

Bortoloso, E., Megighian, A., Furlan, S., Gorza, L., & Volpe, P. (2013, January 1). Homer 2 antagonizes protein degradation in slow-twitch skeletal muscles. American Journal of Physiology-Cell Physiology, 304(1), C68–C77. https://doi.org/10.1152/ajpcell.00108.2012

Burke, V. P., & Startzell, J. M. (2008, November). The Leukemias. Oral and Maxillofacial Surgery Clinics of North America, 20(4), 597–608. https://doi.org/10.1016/j.coms.2008.06.011

Cai, X., Yao, Z., Li, L., & Huang, J. (2018). Role of DKK4 in Tumorigenesis and Tumor Progression. International Journal of Biological Sciences, 14(6), 616–621. https://doi.org/10.7150/ijbs.24329

Chan, J. W., Yeh, I., El-Sayed, I. H., Algazi, A. P., Glastonbury, C. M., Ha, P. K., Yom, S. S., & Zante, A. (2019, January 11). Ultraviolet light-related DNA damage mutation signature distinguishes cutaneous from mucosal or other origin for head and neck squamous cell carcinoma of unknown primary site. Head & Neck, 41(6), E82–E85. https://doi.org/10.1002/hed.25613

Cheng, Y. Y., Rath, E. M., Linton, A., Yuen, M. L., Takahashi, K., & Lee, K. (2020, January). The Current Understanding of Asbestos-Induced Epigenetic Changes Associated With Lung Cancer . Lung Cancer: Targets and Therapy, Volume 11, 1–11. https://doi.org/10.2147/lctt.s186843

Cheng, Y., Hou, T., Ping, J., Chen, T., & Yin, B. (2018, September 15). LMO3 promotes hepatocellular carcinoma invasion, metastasis and anoikis inhibition by directly interacting with LATS1 and suppressing Hippo signaling. Journal of Experimental & Clinical Cancer Research, 37(1). https://doi.org/10.1186/s13046-018-0903-3

Condello, S., Morgan, C. A., Nagdas, S., Cao, L., Turek, J., Hurley, T. D., & Matei, D. (2014, June 23). β-Catenin-regulated ALDH1A1 is a target in ovarian cancer spheroids. Oncogene, 34(18), 2297–2308. https://doi.org/10.1038/onc.2014.178

Cong, Z., Diao, Y., Xu, Y., Li, X., Jiang, Z., Shao, C., Ji, S., Shen, Y., De, W., & Qiang, Y. (2019, January 28). Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98. Cell Death & Disease, 10(2). https://doi.org/10.1038/s41419-019-1361-3

Dabuo, B., Wesome Avogo, E., Owusu Koomson, G., Akantibila, M., & Ayendo Gbati, D. (2022). Aflatoxins: Toxicity, Occurrences and Chronic Exposure. IntechOpen. doi: 10.5772/intechopen.105723

Dai, T., Ye, L., Yu, H., Li, K., Li, J., Liu, R., Lu, X., Deng, M., Li, R., Liu, W., Yang, Y., & Wang, G. (2021, August). Regulation Network and Prognostic Significance of Aldo-Keto Reductase (AKR) Superfamily Genes in Hepatocellular Carcinoma. Journal of Hepatocellular Carcinoma, Volume 8, 997–1021. https://doi.org/10.2147/jhc.s323743

Desnoyers, L. R., Pai, R., Ferrando, R. E., Hötzel, K., Le, T., Ross, J., Carano, R., D'Souza, A., Qing, J., Mohtashemi, I., Ashkenazi, A., & French, D. M. (2007, June 25). Targeting FGF19 inhibits tumor growth in colon cancer xenograft and FGF19 transgenic hepatocellular carcinoma models. Oncogene, 27(1), 85–97. https://doi.org/10.1038/sj.onc.1210623

Drozd, E., Krzysztoń-Russjan, J., Marczewska, J., Drozd, J., Bubko, I., Bielak, M., Lubelska, K., Wiktorska, K., Chilmonczyk, Z., Anuszewska, E., & Gruber-Bzura, B. (2016, October). Up-regulation of glutathione-related genes, enzyme activities and transport proteins in human cervical cancer cells treated with doxorubicin. Biomedicine & Pharmacotherapy, 83, 397–406. https://doi.org/10.1016/j.biopha.2016.06.051

Fu, X., Shi, L., Zhang, W., Zhang, X., Peng, Y., Chen, X., Tang, C., Li, X., & Zhou, X. (2014). Expression of Indian hedgehog is negatively correlated with APC gene mutation in colorectal tumors. International journal of clinical and experimental medicine, 7(8), 2150–2155

Genes and Cancer. (2012, March 26). Cancer.Net. https://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genes-and-cancer

Gurney, M., Mangaonkar, A. A., Lasho, T., Finke, C., Al-Kali, A., Gangat, N., Shah, M. V., Alkhateeb, H. B., Tefferi, A., Sallman, D., Xie, Z., Viswanatha, D., Reichard, K., Al Ali, N., Komrokji, R., Padron, E., & Patnaik, M. M. (2023, July 8). Somatic TP53 single nucleotide variants, indels and copy number alterations in chronic myelomonocytic leukemia (CMML). Leukemia. https://doi.org/10.1038/s41375-023-01964-3

Hajaj, E., Sciacovelli, M., Frezza, C., & Erez, A. (2021, September). The context-specific roles of urea cycle enzymes in tumorigenesis. Molecular Cell, 81(18), 3749–3759. https://doi.org/10.1016/j.molcel.2021.08.005

Herrick, J. A. (2023, May). Diagnosis of Indolent Clonorchis sinensis and Opisthorchis viverrini Infections as Risk Factors for Cholangiocarcinoma: An Unmet Medical Need. Federal Practitioner, 40(Suppl 1). https://doi.org/10.12788/fp.0376

Hossain, M. S., Karuniawati, H., Jairoun, A. A., Urbi, Z., Ooi, D. J., John, A., Lim, Y. C., Kibria, K. M. K., Mohiuddin, A. M., Ming, L. C., Goh, K. W., & Hadi, M. A. (2022, March 29). Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. Cancers, 14(7), 1732. https://doi.org/10.3390/cancers14071732

Hou, Y., Li, T., Gan, W., Lv, S., Zeng, Z., Yan, Z., Wang, W., & Yang, M. (2020, March). Prognostic significance of mutant-allele tumor heterogeneity in uterine corpus endometrial carcinoma. Annals of Translational Medicine, 8(6), 339–339. https://doi.org/10.21037/atm.2020.02.136

Hu, Y., Zheng, M., Zhang, D., Gou, R., Liu, O., Wang, S., & Lin, B. (2021, September 26). Identification of the prognostic value of a 2-gene signature of the WNT gene family in UCEC using bioinformatics and real-world data. Cancer Cell International, 21(1). https://doi.org/10.1186/s12935-021-02215-0

Kropotova, E. S., Tychko, R. A., Zinov'eva, O. L., Zyryanova, A. F., Khankin, S. L., Cherkes, V. L., Aliev, V. A., Beresten, S. F., Oparina, N. Y., & Mashkova, T. D. (2010, April). Downregulation of AKR1B10 expression in colorectal cancer. Molecular Biology, 44(2), 216–222. https://doi.org/10.1134/s0026893310020056

Kropotova, E. S., Zinovieva, O. L., Zyryanova, A. F., Dybovaya, V. I., Prasolov, V. S., Beresten, S. F., Oparina, N. Y., & Mashkova, T. D. (2014, March 6). Altered Expression of Multiple Genes Involved in Retinoic Acid Biosynthesis in Human Colorectal Cancer. Pathology & Oncology Research, 20(3), 707–717. https://doi.org/10.1007/s12253-014-9751-4

Küppers, R., Engert, A., & Hansmann, M. L. (2012, October 1). Hodgkin lymphoma. Journal of Clinical Investigation, 122(10), 3439–3447. https://doi.org/10.1172/jci61245

Lahtz, C., & Pfeifer, G. P. (2011, January 30). Epigenetic changes of DNA repair genes in cancer. Journal of Molecular Cell Biology, 3(1), 51–58. https://doi.org/10.1093/jmcb/mjq053

Levine, D., The Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. Nature 497, 67–73 (2013). https://doi.org/10.1038/nature12113

Lewis, M. (2020, November 2). Mouth Cancer – Risk Factors and Potentially Malignant Disorders. Dental Update, 47(10), 793–799. https://doi.org/10.12968/denu.2020.47.10.793

Li, L. L., Dai, B., Sun, Y. H., & Zhang, T. T. (2020, June). The activation of IL-17 signaling pathway promotes pyroptosis in pneumonia-induced sepsis. Annals of Translational Medicine, 8(11), 674–674. https://doi.org/10.21037/atm-19-1739

Li, Y., Wu, L., Tao, W., Wu, D., Ma, F., & Li, N. (2020, June 5). Expression Atlas of FGF and FGFR Genes in Pancancer Uncovered Predictive Biomarkers for Clinical Trials of Selective FGFR Inhibitors. BioMed Research International, 2020, 1–8. https://doi.org/10.1155/2020/5658904

Li, Y., Wu, L., Tao, W., Wu, D., Ma, F., & Li, N. (2020, June 5). Expression Atlas of FGF and FGFR Genes in Pancancer Uncovered Predictive Biomarkers for Clinical Trials of Selective FGFR Inhibitors. BioMed Research International, 2020, 1–8. https://doi.org/10.1155/2020/5658904

Li, Z., Wang, H., Zhang, Z., Meng, X., Liu, D., & Tang, Y. (2021, July). Germline and somatic mutation profile in Cancer patients revealed by a medium-sized pan-Cancer panel. Genomics, 113(4), 1930–1939. https://doi.org/10.1016/j.ygeno.2021.04.029

Lian, Z., De Luca, P., & Di Cristofano, A. (2006). Gene expression analysis reveals a signature of estrogen receptor activation upon loss ofPten in a mouse model of endometrial cancer. Journal of Cellular Physiology, 208(2), 255–266. https://doi.org/10.1002/jcp.20681

Liang, L., Huang, J., Yao, M., Li, L., Jin, X. J., & Cai, X. Y. (2021, October 15). GNG4 Promotes Tumor Progression in Colorectal Cancer. Journal of Oncology, 2021, 1–9. https://doi.org/10.1155/2021/9931984

Liu, W., Zhang, F., Yang, K., & Yan, Y. (2023, January 26). Comprehensive analysis regarding the prognostic significance of downregulated ferroptosis-related gene AKR1C2 in gastric cancer and its underlying roles in immune response. PLOS ONE, 18(1), e0280989. https://doi.org/10.1371/journal.pone.0280989

Lou, X., Meng, Y., & Hou, Y. (2021, February 14). A literature review on function and regulation mechanism of DKK4. Journal of Cellular and Molecular Medicine, 25(6), 2786–2794. https://doi.org/10.1111/jcmm.16372

Luo, P., Liang, C., Jing, W., Zhu, M., Zhou, H., Chai, H., Worley, P. F., & Tu, J. (2021). Homer2 and Homer3 Act as Novel Biomarkers in Diagnosis of hepatitis B virus-induced Hepatocellular Carcinoma. Journal of Cancer, 12(12), 3439–3447. https://doi.org/10.7150/jca.52118

Matsumoto, J., Nishimoto, A., Watari, S., Ueki, H., Shiromizu, S., Iwata, N., Takeda, T., Ushio, S., Kajizono, M., Fujiyoshi, M., Koyama, T., Araki, M., Wada, K., Zamami, Y., Nasu, Y., & Ariyoshi, N. (2022, December 26). Significance of UGT1A6, UGT1A9, and UGT2B7 genetic variants and their mRNA expression in the clinical outcome of renal cell carcinoma. Molecular and Cellular Biochemistry, 478(8), 1779–1790. https://doi.org/10.1007/s11010-022-04637-4

Mazzei-Abba, A., Folly, C. L., Kreis, C., Ammann, R. A., Adam, C., Brack, E., Egger, M., Kuehni, C. E., & Spycher, B. D. (2021, November). External background ionizing radiation and childhood cancer: Update of a nationwide cohort analysis. Journal of Environmental Radioactivity, 238–239, 106734. https://doi.org/10.1016/j.jenvrad.2021.106734

Mohabati Mobarez, A., Soleimani, N., Esmaeili, S. A., & Farhangi, B. (2020, October). Nanoparticle-based immunotherapy of breast cancer using recombinant Helicobacter pylori proteins. European Journal of Pharmaceutics and Biopharmaceutics, 155, 69–76. https://doi.org/10.1016/j.ejpb.2020.08.013

Mohammed, S. A., Hetta, H. F., Zahran, A. M., Tolba, M. E. M., Attia, R. A. H., Behnsawy, H. M., Algammal, A. M., Batiha, G. E. S., Mohammed, A. Q., & Ahmad, A. A. (2023, April 17). T cell subsets, regulatory T, regulatory B cells and proinflammatory cytokine profile in Schistosoma haematobium associated bladder cancer: First report from Upper Egypt. PLOS Neglected Tropical Diseases, 17(4), e0011258. https://doi.org/10.1371/journal.pntd.0011258

Moon, S. M., Ahn, M. Y., Kwon, S. M., Kim, S. A., Ahn, S. G., & Yoon, J. H. (2012, March 5). Homeobox C5 expression is associated with the progression of 4-nitroquinoline 1-oxide-induced rat tongue carcinogenesis. Journal of Oral Pathology & Medicine, 41(6), 470–476. https://doi.org/10.1111/j.1600-0714.2012.01133.x

Moukarzel, L. A., Ferrando, L., Da Cruz Paula, A., Brown, D. N., Geyer, F. C., Pareja, F., Piscuoglio, S., Papanastasiou, A. D., Fusco, N., Marchiò, C., Abu-Rustum, N. R., Murali, R., Brogi, E., Wen, H. Y., Norton, L., Soslow, R. A., Vincent-Salomon, A., Reis-Filho, J. S., & Weigelt, B. (2021, February 19). The genetic landscape of metaplastic breast cancers and uterine carcinosarcomas. Molecular Oncology, 15(4), 1024–1039. https://doi.org/10.1002/1878-0261.12813

Nayak, L. M., & Deschler, D. G. (2003, August). Lymphomas. Otolaryngologic Clinics of North America, 36(4), 625–646. https://doi.org/10.1016/s0030-6665(03)00033-1

Nebot-Bral, L., Coutzac, C., Kannouche, P. L., & Chaput, N. (2019, February). Why is immunotherapy effective (or not) in patients with MSI/MMRD tumors? Bulletin Du Cancer, 106(2), 105–113. https://doi.org/10.1016/j.bulcan.2018.08.007

Nesta, A. V., Tafur, D., & Beck, C. R. (2021, August). Hotspots of Human Mutation. Trends in Genetics, 37(8), 717–729. https://doi.org/10.1016/j.tig.2020.10.003

Oncogene. (2023, August 1). Genome.gov. https://www.genome.gov/genetics-glossary/Oncogene

Pan, Y., Jia, L. P., Liu, Y., Han, Y., & Deng, Q. (2019, August 31). Alteration of tumor associated neutrophils by PIK3CA expression in endometrial carcinoma from TCGA data. Journal of Ovarian Research, 12(1). https://doi.org/10.1186/s13048-019-0557-6

Parascandola, M., & Xiao, L. (2019, May). Tobacco and the lung cancer epidemic in China. Translational Lung Cancer Research, 8(S1), S21–S30. https://doi.org/10.21037/tlcr.2019.03.12

Pezzicoli, G., Moscaritolo, F., Silvestris, E., Silvestris, F., Cormio, G., Porta, C., & D'Oronzo, S. (2021, July). Uterine carcinosarcoma: An overview. Critical Reviews in Oncology/Hematology, 163, 103369. https://doi.org/10.1016/j.critrevonc.2021.103369

Poturnajova, M., Kozovska, Z., & Matuskova, M. (2021, November). Aldehyde dehydrogenase 1A1 and 1A3 isoforms – mechanism of activation and regulation in cancer. Cellular Signalling, 87, 110120. https://doi.org/10.1016/j.cellsig.2021.110120

Qiu, Y. S., Jiang, N. N., Zhou, Y., Yu, K. Y., Gong, H. Y., & Liao, G. J. (2018, February 8). LMO3 promotes gastric cancer cell invasion and proliferation through Akt-mTOR and Akt-GSK3β signaling. International Journal of Molecular Medicine. https://doi.org/10.3892/ijmm.2018.3476

Ren, N., Liang, B., & Li, Y. (2020, October). Identification of prognosis-related genes in the tumor microenvironment of stomach adenocarcinoma by TCGA and GEO datasets. Bioscience Reports, 40(10). https://doi.org/10.1042/bsr20200980

Rico, S. D., Mahnken, M., Büscheck, F., Dum, D., Luebke, A. M., Kluth, M., Hube-Magg, C., Hinsch, A., Höflmayer, D., Möller-Koop, C., Fraune, C., Möller, K., Menz, A., Bernreuther, C., Jacobsen, F., Lebok, P., Clauditz, T. S., Sauter, G., Uhlig, R., . . . Marx, A. H. (2021, January). MUC5AC Expression in Various Tumor Types and Nonneoplastic Tissue: A Tissue Microarray Study on 10 399 Tissue Samples. Technology in Cancer Research & Treatment, 20, 153303382110433. https://doi.org/10.1177/15330338211043328

Rogozin, I. B., & Pavlov, Y. I. (2003, September). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. Mutation Research/Reviews in Mutation Research, 544(1), 65–85. https://doi.org/10.1016/s1383-5742(03)00032-2

Rumgay, H., Shield, K., Charvat, H., Ferrari, P., Sornpaisarn, B., Obot, I., Islami, F., Lemmens, V., Rehm, J., & Soerjomataram, I. (2021, July 1). Abstract 43: Global Burden of Cancer in 2020 Attributable to Alcohol Consumption: A Population-Based Study. Cancer Epidemiology, Biomarkers & Prevention, 30(7_Supplement), 43–43. https://doi.org/10.1158/1538-7755.asgcr21-43

Salama, E. A., Adbeltawab, R. E., & El Tayebi, H. M. (2020, January 10). XIST and TSIX: Novel Cancer Immune Biomarkers in PD-L1-Overexpressing Breast Cancer Patients. Frontiers in Oncology, 9. https://doi.org/10.3389/fonc.2019.01459

Shankland, K. R., Armitage, J. O., & Hancock, B. W. (2012, September). Non-Hodgkin lymphoma. The Lancet, 380(9844), 848–857. https://doi.org/10.1016/s0140-6736(12)60605-9

Shen, C., Yang, C., Xia, B., & You, M. (2021, March). Long non-coding RNAs: Emerging regulators for chemo/immunotherapy resistance in cancer stem cells. Cancer Letters, 500, 244–252. https://doi.org/10.1016/j.canlet.2020.11.010

Shen, M. L., Xiao, A., Yin, S. J., Wang, P., Lin, X. Q., Yu, C. B., & He, G. H. (2019, July). Associations between UGT2B7 polymorphisms and cancer susceptibility: A meta-analysis. Gene, 706, 115–123. https://doi.org/10.1016/j.gene.2019.05.025

Smithgall, M. C., Remotti, H., Hsiao, S. J., Mansukhani, M., Liu-Jarin, X., & Fernandes, H. (2022, January). Investigation of discrepant mismatch repair immunohistochemistry and microsatellite instability polymerase chain reaction test results for gynecologic cancers using next-generation sequencing. Human Pathology, 119, 41–50. https://doi.org/10.1016/j.humpath.2021.10.004

Su, Y., Yang, L. M., & Ornitz, D. M. (2020, September 9). FGF20-FGFR1 signaling through MAPK and PI3K controls sensory progenitor differentiation in the organ of Corti. Developmental Dynamics, 250(2), 134–144. https://doi.org/10.1002/dvdy.231

Sun, Z., He, Z., Liu, R., & Zhang, Z. (2021, July 1). Cation Lipid-Assisted PEG6-PLGA Polymer Nanoparticles Encapsulated Knocking Down Long ncRNAs Reverse Non-Coding RNA of Xist Through the Support Vector Machine Model to Regulate the Molecular Mechanisms of Gastric Cancer Cell Apoptosis. Journal of Biomedical Nanotechnology, 17(7), 1305–1319. https://doi.org/10.1166/jbn.2021.3107

Tanaka, M., & Nakamura, T. (2021, July 15). Modeling fusion gene-associated sarcoma: Advantages for understanding sarcoma biology and pathology. Pathology International, 71(10), 643–654. https://doi.org/10.1111/pin.13142

Teng, Y., Yu, Y., Li, S., Huang, Y., Xu, D., Tao, X., & Fan, Y. (2021, July 22). Ultraviolet Radiation and Basal Cell Carcinoma: An Environmental Perspective. Frontiers in Public Health, 9. https://doi.org/10.3389/fpubh.2021.666528

Tiong, K. H., Tan, B. S., Choo, H. L., Chung, F. F. L., Hii, L. W., Tan, S. H., Khor, N. T. W., Wong, S. F., See, S. J., Tan, Y. F., Rosli, R., Cheong, S. K., & Leong, C. O. (2016, May 12). Fibroblast growth factor receptor 4 (FGFR4) and fibroblast growth factor 19 (FGF19) autocrine enhance breast cancer cells survival. Oncotarget, 7(36), 57633–57650. https://doi.org/10.18632/oncotarget.9328

Vareedayah AA, Alkaade S, Taylor JR. Pancreatic Adenocarcinoma. Mo Med. 2018 May-Jun;115(3):230-235. Erratum in: Mo Med. 2018 Nov-Dec;115(6):517. PMID: 30228728; PMCID: PMC6140147

Wang, Z., Jinnin, M., Nakamura, K., Harada, M., Kudo, H., Nakayama, W., Inoue, K., Nakashima, T., Honda, N., Fukushima, S., & Ihn, H. (2016, January 12). Long non-coding RNA TSIX is upregulated in scleroderma dermal fibroblasts and controls

collagen mRNA stabilization. Experimental Dermatology, 25(2), 131–136. https://doi.org/10.1111/exd.12900\

Wenners, A., Hartmann, F., Jochens, A., Roemer, A. M., Alkatout, I., Klapper, W., van Mackelenbergh, M., Mundhenke, C., Jonat, W., & Bauer, M. (2015, November 14). Stromal markers AKR1C1 and AKR1C2 are prognostic factors in primary human breast cancer. International Journal of Clinical Oncology, 21(3), 548–556. https://doi.org/10.1007/s10147-015-0924-2

What Is Cancer? (2021, October 11). National Cancer Institute. https://www.cancer.gov/about-cancer/understanding/what-is-cancer

Wild, C. P., & Montesano, R. (2009, December). A model of interaction: Aflatoxins and hepatitis viruses in liver cancer aetiology and prevention. Cancer Letters, 286(1), 22–28. https://doi.org/10.1016/j.canlet.2009.02.053

Wu, M. H., Lu, R. Y., Yu, S. J., Tsai, Y. Z., Lin, Y. C., Bai, Z. Y., Liao, R. Y., Hsu, Y. C., Chen, C. C., & Cai, B. H. (2022, November 16). PTC124 Rescues Nonsense Mutation of Two Tumor Suppressor Genes NOTCH1 and FAT1 to Repress HNSCC Cell Proliferation. Biomedicines, 10(11), 2948. https://doi.org/10.3390/biomedicines10112948

Yi, Z., Rong, G., Guan, Y., Li, J., Chang, L., Li, H., Liu, B., Wang, W., Guan, X., Ouyang, Q., Li, L., Zhai, J., Li, C., Li, L., Xia, X., Yang, L., Qian, H., Yi, X., Xu, B., & Ma, F. (2020, October 30). Molecular landscape and efficacy of HER2-targeted therapy in patients with HER2-mutated metastatic breast cancer. Npj Breast Cancer, 6(1). https://doi.org/10.1038/s41523-020-00201-9

Yue, H., Hu, Z., Hu, R., Guo, Z., Zheng, Y., Wang, Y., & Zhou, Y. (2022, June 22). ALDH1A1 in Cancers: Bidirectional Function, Drug Resistance, and Regulatory Mechanism. Frontiers in Oncology, 12. https://doi.org/10.3389/fonc.2022.918778

Zacharias, F. (2020). MicroRNAs Determining Carcinogenesis by Regulating Oncogenes and Tumor Suppressor Genes During Cell Cycle. MicroRNA, 9(2), 82–92. https://doi.org/10.2174/22115374mtawnotixw

Zhang, W., Wang, H., Sun, M., Deng, X., Wu, X., Ma, Y., Li, M., Shuoa, S. M., You, Q., & Miao, L. (2020, March). CXCL5/CXCR2 axis in tumor microenvironment as potential diagnostic biomarker and therapeutic target. Cancer Communications, 40(2–3), 69–80. https://doi.org/10.1002/cac2.12010

Zhang, X., Kong, M., Zhang, Z., Xu, S., Yan, F., Wei, L., & Zhou, J. (2017, September 14). FGF19 genetic amplification as a potential therapeutic target in lung squamous cell carcinomas. Thoracic Cancer, 8(6), 655–665. https://doi.org/10.1111/1759-7714.12504

Zhang, Y., An, J., & Pei, Y. (2020, July 1). LncRNA SNHG6 promotes LMO3 expression by sponging miR-543 in glioma. Molecular and Cellular Biochemistry, 472(1–2), 9–17. https://doi.org/10.1007/s11010-020-03772-0

Zhao, F., Wang, X., Wang, Y., Zhang, J., Lai, R., Zhang, B., & Zhou, X. (2020, June 26). The function of uterine UDP-glucuronosyltransferase 1A8 (UGT1A8) and UDP-glucuronosyltransferase 2B7 (UGT2B7) is involved in endometrial cancer based on estrogen metabolism regulation. Hormones, 19(3), 403–412. https://doi.org/10.1007/s42000-020-00213-x

Zhao, L., Fu, X., Han, X., Yu, Y., Ye, Y., & Gao, J. (2021, January 28). Tumor mutation burden in connection with immune-related survival in uterine corpus endometrial carcinoma. Cancer Cell International, 21(1). https://doi.org/10.1186/s12935-021-01774-6

Zhao, M., & Li, W. (2023, January 16). Metabolism-associated molecular classification of uterine corpus endometrial carcinoma. Frontiers in Genetics, 14. https://doi.org/10.3389/fgene.2023.955466

Zhao, S., Wagn, S., Zhao, Z., & Li, W. (2019, September 4). AKR1C1-3, notably AKR1C3, are distinct biomarkers for liver cancer diagnosis and prognosis: Database mining in malignancies. Oncology Letters. https://doi.org/10.3892/ol.2019.10802

Zhou, L., Huang, W., Yu, H. F., Feng, Y. J., & Teng, X. (2020, June 23). Exploring TCGA database for identification of potential prognostic genes in stomach adenocarcinoma. Cancer Cell International, 20(1). https://doi.org/10.1186/s12935-020-01351-3

Zhou, Y., Lin, Y., Li, W., Liu, Q., Gong, H., Li, Y., & Luo, D. (2023, February 22). Expression of AKRs superfamily and prognostic in human gastric cancer. Medicine, 102(8), e33041. https://doi.org/10.1097/md.0000000000033041

Zhu, W., Huang, M., Thakur, A., Yan, Y., & Wu, X. (2023, February 2). FGF19 promotes cell autophagy and cisplatin chemoresistance by activating MAPK signaling in ovarian cancer. PeerJ, 11, e14827. https://doi.org/10.7717/peerj.14827