
**Arab American University**

**Faculty of Graduate Studies**

# Functional assessment of a novel mutation (*SMAD4*, p. Cys115Arg) in a Palestinian patient with Colorectal cancer

By

**Rawan Nael Mohammed Obeid**

Supervisor

**Dr. Nouar Mohammad Abdel Muti Qutob**

Co- Supervisor

**Dr. Zaidoun Mahmoud Hasan Salah**

**This thesis was submitted in partial fulfillment of the requirements for the Master`s degree in**

**Molecular Genetics and Genetic Toxicology**

**February / 2024**

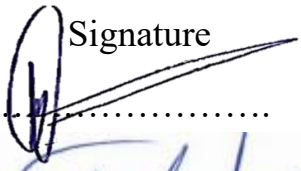
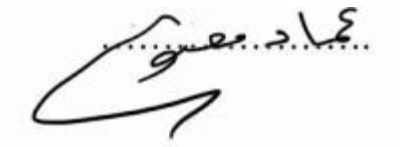
# Thesis approval

## Functional assessment of a novel mutation (*SMAD4*, p. Cys115Arg) in aPalestinian patient with Colorectal cancer

By

**Rawan Nael Obeid**

This thesis was defended successfully on 29/2/2024 and approved by:

| Committee members | Signature |
|---|---|
| 1. Dr. Nouar Qutob: Supervisor |  |
| 2. Dr. Zaidoun Salah: Co-supervisor |  |
| 3. Dr. Feras Al Batta: Internal Examiner |  |
| 4. Dr. Imad Matouk: External Examiner |  |

# Declaration

Hereby, I declare that this thesis has been composed by my own research work as a master's student in the molecular genetics and genetic toxicology program at the Arab American University of Palestine and has not been previously submitted for any other degree or professional qualification, and every effort has been made to appropriately indicate, mention, and acknowledge other contributions.

Name: Rawan Nael Mohammed Obeid

ID No: 201912495

Signature: 

Date: 26/9/2024

# Dedication

To my beloved father, whose unwavering belief in my potential pushed me toward the pursuit of a master's degree. Your memory and enduring inspiration will forever guide me in my life. I love you, dad.

To my remarkable sister, Sarah, a beacon of strength and resilience, whose two-time triumph over cancer kindled the fire within me to dedicate myself to cancer research. Your courage and determination will continue to light the way for countless others.

To my cherished mother, Nuha, my sister Lama, and brother Hamza, your love and constant support have been the bedrock upon which my journey was built. Your sacrifices and encouragement have been immeasurable.

To my loving husband, confidant, and best friend, your unwavering patience and steadfast support have made this challenging path feel more like a walk in the park. Your belief in me has been my greatest motivation.

To my precious son, Mustafa, who brought new challenges and joys to my journey. Your presence has added meaning and depth to my accomplishments.

To my best friends, Dana and Adleen, whose companionship and unwavering support have turned this arduous journey into an enjoyable and seamless adventure. Your presence has been a constant source of encouragement and laughter.

This thesis is dedicated to all of you, the pillars of my life, who have made this academic pursuit not only possible but immensely rewarding. Your love, faith, and support have been the driving force behind my success, and I am eternally grateful for the roles you have played in this journey.

## Acknowledgment

I would like to express my heartfelt gratitude to several individuals who have been instrumental in the completion of this thesis.

First and foremost, I extend my sincere appreciation to my dedicated supervisors, Dr. Nouar Al-Qutob, and Dr. Zaidoun Salah, for their unwavering support, patience, and invaluable guidance throughout this academic journey. Your expertise and mentorship have been pivotal to the success of this work.

I am also deeply thankful to Husam Sallam and Rua' Thawabteh, the exemplary members of the medical laboratory staff at AAUP. Their professionalism, tireless efforts, and constant assistance were indispensable in carrying out the research. Your commitment to excellence has significantly contributed to the quality of this study.

I would like to acknowledge my boss, Dr. Mahmoud Ruzayqat, for his understanding and support, which made it possible for me to pursue my education at the university. Your flexibility and encouragement have been a cornerstone in this endeavor.

Finally, I would like to extend my gratitude to the patient and their family who participated in this study. Their willingness to be a part of this research is deeply appreciated, as their contributions are pivotal in advancing medical knowledge and improving patient care.

This thesis would not have been possible without the collective efforts, support, and cooperation of these outstanding individuals. I am truly grateful for the roles each of you played in this academic endeavor.

# Abstract

Colorectal carcinoma (CRC) is a globally common cancer that leads to mortality and morbidity. Genetic testing is becoming an essential tool for colorectal cancer (CRC) diagnosis and treatment pathway. In a variety of colorectal tumors, *SMAD4* is somatically inactivated. Moreover, it was shown that expression loss of *SMAD4* is correlated with both; metastatic development, a poor response to chemotherapy, and decreased immune infiltrate, reinforcing its use as a prognostic marker for CRC patients.

In this study, Next-Generation Sequencing (NGS) was utilized to identify genetic variants that contribute to the tumorigenesis of colorectal cancer. Our analysis identified the following variants: (*NOTCH1*, p.Asp1533Gly), (*EGFR*, p.Leu619Pro), (*MAP2K1*, p.Pro265Ser), (*ALK*, p.Phe174Leu), (*ERBB2*, p.Asp1252His), (*FBXW7*, p.Gly477Arg), (*FGFR2*, p.Met538Ile), (*GNA11*, p.Leu170Phe), (*MET*, p.Asp153Asn), (*NOTCH1*, p.Ala1610Thr), (*PIK3CA*, p.Asn107Thr), (*SMAD4*, p.Cys115Arg). In silico analysis of the previously mentioned variants presumed (*SMAD4*, p.Cys115Arg) as a highly likely pathogenic variant. To further investigate the function of the identified variant, luciferase assay was used to assess the effect of the mutation on *SMAD4* gene transcriptional activity.

Our results revealed that the mutation decreased *SMAD4* transcriptional activity by four-fold compared to the wild-type *SMAD4*.

This research provides a better understanding of the cancer genetic landscape in Palestine by characterizing the functional effects of (*SMAD4*, p. Cys115Arg). Moreover, it might contribute to developing personalized and targeted therapies and treatments for those diagnosed with CRC.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| °C | Degree Celsius |
| µL | Microliter |
| AKT1 | AKT Serine/Threonine Kinase 1 |
| ALK | Anaplastic lymphoma kinase |
| APC | Adenomatous polyposis coli |
| BMP | Bone Morphogenetic Protein |
| BRAF | v-Raf murine sarcoma viral oncogene homolog B |
| BRCA1 | Breast Cancer gene 1 |
| BRCA2 | Breast Cancer gene 2 |
| CRC | Colorectal Cancer |
| CTNNB1 | Catenin Beta 1 |
| DNA | Deoxyribonucleic Acid |
| DDR2 | Discoidin Domain Receptor Tyrosine Kinase 2) |
| DPC4 | Deleted in pancreatic cancer 4 |
| EDTA | Ethylenediaminetetraacetic acid |
| EGFR | Epidermal growth factor receptor |
| EMT | Epithelial-Mesenchymal Transition |
| ERBB2 | Erb-b2 receptor tyrosine kinase 2 |
| ERBB3 | Erb-B2 Receptor Tyrosine Kinase 3 |
| ERBB4 | Erb-B2 Receptor Tyrosine Kinase 4 |
| ERK | Extracellular Signal-Regulated kinase |
| ESR1 | Estrogen Receptor 1 |

| | |
|---|---|
| FAP | Familial adenomatous polyposis |
| FBXW7 | F-Box and WD Repeat Domain Containing 7 |
| FFPE | Formalin-Fixed Paraffin-Embedded |
| FGFR1 | Fibroblast Growth Factor Receptor 1 |
| FGFR2 | Fibroblast growth factor receptor 2 |
| FGFR3 | Fibroblast growth factor receptor 3 |
| FLT3 | Fms-like tyrosine kinase 3 |
| GNA11 | G Protein Subunit Alpha 11 |
| GNAQ | G Protein Subunit Alpha Q |
| HNPCC | Hereditary Non Polyposis Colorectal Cancer |
| HRAS | Harvey rat sarcoma viral oncogene homolog |
| Jagged-*1* | Jagged Canonical Notch Ligand 1 |
| KIT | Proto-oncogene receptor tyrosine kinase |
| KRAS | Ki-ras2 Kirsten rat sarcoma viral oncogene homolog |
| MAP2K1 | Mitogen-Activated Protein Kinase 1 |
| MAP2K2 | Mitogen-Activated Protein Kinase 2 |
| MET | MET proto-oncogene, receptor tyrosine kinase |
| miRNAs | Micro Ribonucleic acids |
| mL | Milliliter |
| *MLH1* | MutL Protein Homolog 1 |
| MMR | Mismatch Repair |
| MSH2 | MutS Homolog 2 |
| ng | Nano Gram |

| | |
|---|---|
| NGS | Next Generation Sequencing |
| NOTCH1 | Notch homolog 1, translocation-associated |
| NRAS | Neuroblastoma RAS viral oncogene homolog |
| NSCLC | Non-Small Cell Lung Cancer |
| PAK3 | p21-Activated Kinase 3 |
| PCR | Polymerase Chain Reaction |
| PDGFRA | Platelet-Derived Growth Factor Receptor Alpha |
| PIK3CA | Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha |
| PMS2 | Postmeiotic Segregation Increased 2 |
| PTEN | Phosphatase and Tensin Homolog deleted on Chromosome 10 |
| RAF | Rapidly Accelerated Fibrosarcoma |
| RAS | Rat Sarcoma |
| rpm | Round Per Minute |
| SMADs | Suppressor of Mothers against Decapentaplegic |
| SMAD4 | Mothers against decapentaplegic homolog 4 |
| STK1 | Serological Thymidine Kinase 1 |
| TGF-β | Transforming Growth Factor Beta |
| TGFβR1 | Transforming Growth Factor Beta Receptor 1 |
| TGFβR2 | Transforming Growth Factor Beta Receptor 2 |
| TP53 | Tumor Protein P53 |
| WHO | World Health Organization |
| Wnt | Wingless-related integration site |

# Chapter One: Introduction

Cancer is a complex and multifactorial disease, of which, colorectal cancer is considered as one of the most prevalent of all cancer types worldwide. Understanding the underlying causes of cancer, including colorectal cancer, is crucial to develop more effective diagnostic and therapeutic approaches, and preventive measures, leading to early detection and improved patient outcomes. Employing cutting-edge methodologies like next-generation sequencing and advanced molecular techniques in colon cancer research shows great potential in deepening our insights in the prevention and diagnosis of CRC.

## 1.1. Cancer Development and Progression

Cancer is a group of diseases characterized by uncontrolled cell growth that leads to the formation of tumors in various parts of the body. It is a complex process that involves multiple genetic and epigenetic alterations that drive the transformation of normal cells into cancerous cells. Pre-cancerous lesions typically progress to malignant tumors through a multi-stage mechanism that results in the transformation of normal cells into tumor cells. Such changes result from a combination of the person's genetic factors interacting with a variable of external agents such as physical carcinogens (UV radiation), chemical carcinogens (Asbestos), and biological carcinogens, such as infections with specific viruses, bacteria, or parasites (Parsa, 2012). The diversity of neoplastic diseases is remarkable; however, several cancer hallmarks contribute to the development and progression of human tumors (Negrini et al., 2010). Cancer cells can acquire mutations or epigenetic changes in their DNA that alter normal cellular processes and lead to the hallmark characteristics of cancer. The hallmarks of cancer encompass six

fundamental biological capabilities that emerge over the course of the complex development of human tumors. These hallmarks serve as a fundamental framework for understanding the intricacies of cancer. They encompass the ability to sustain proliferative signaling, evade growth-inhibiting mechanisms, resist programmed cell death, attain replicative immortality, induce the formation of new blood vessels (angiogenesis), and activate processes related to invasion and metastasis (Carmeliet & Jain, 2011,Weinberg, 2000). These hallmarks are underpinned by genome instability, which leads to genetic diversity that facilitates their acquisition, and inflammation, which supports multiple hallmark functions. In recent years, two additional emerging hallmarks have gained recognition—reprogramming of energy metabolism and evading immune destruction (Hanahan & Weinberg, 2011a). Finally, cancer cells can invade and metastasize to other tissues and organs, which is the main cause of cancer mortality (Isaiah J. Fidler, 2003). This process requires the activation of specific genes and signaling pathways that enable cancer cells to break a way from the primary tumor, survive in the bloodstream, and colonize distant sites (Valastyan & Weinberg, 2011).

### 1.2 Global Cancer Statistics

According to the (WHO, 2022), cancer is the second leading cause of death globally, with approximately 10 million fatalities registered in the year 2020. Cancer remains a significant global health challenge, with colorectal cancer emerging as the third most commonly diagnosed cancer and the second leading cause of cancer death worldwide; by 2030, it is anticipated that there will be 1.1 million deaths and more than 2.2 million new cases of colorectal cancer (CRC) (Arnold et al., 2017; Sung et al., 2021). In terms of new cancer cases in 2020, the three most common were: 2.26 million cases of breast cancer;

2.21 million cases of lung cancer; and 1.93 million cases of colon and rectum cancer. Moreover, the three most frequent causes of cancer death were: lung cancer (1.80 million deaths); colon and rectum (935,000 deaths); and liver cancer (830,000 deaths) (WHO, 2022). Colorectal cancer stands as the second leading cause of cancer-related fatalities on a global scale. These statistics revealed substantial disparities across regions, with the highest incidence rates recorded in Europe and Australia, while Eastern Europe witnessed the highest mortality rates.

In a study conducted in 2016, it was found that among cancer types in the West Bank, Palestine, lung cancer was the most widespread cause of mortality in males, accounting for 22.8% of deaths. For females, breast cancer was the leading cause, with a mortality rate of 21.5%, followed by colon cancer at 11.4%. Prostate cancer was the least common cause of death among males, with a mortality rate of 9.5% (Abu-Rmeileh et al., 2016).

Projections into the future indicate a significant surge in the burden of colorectal cancer, with anticipated figures reaching 3.2 million new cases annually, marking a 63% increase, along with 1.6 million deaths per year, representing a 73% escalation by the year 2040 (WHO, 2023).

### 1.3 The Genetics of Cancer

Genetic factors play a pivotal role in cancer development, with numerous genes and mutations contributing to uncontrolled cell growth and division, ultimately resulting in the formation of tumors (Stratton et al., 2009). Genes can be classified as oncogenes, tumor suppressor genes or DNA repair genes. Oncogenes are a mutated form of a gene that is highly expressed which can promote cell proliferation when activated by genetic changes (Vogelstein et al., 2013). Contrarily, tumor suppressor genes regulate cell growth

and inhibit tumor formation, mutations in these genes can disable their normal functions, leading to uncontrolled cell growth (Hanahan & Weinberg, 2011b). DNA repair genes encode DNA-repair enzymes and is mostly associated with cellular response to DNA damage.

In cancer genetics, the concepts of passenger and driver mutations are essential for understanding the molecular landscape of the disease. Driver mutations are genetic changes that provide cancer cells a selective growth advantage that promotes the development, progression, and metastasis of tumors. These mutations occur in genes critical for regulating cell growth, DNA repair, or signaling pathways (Vogelstein et al., 2013). It stands in contrast to passenger mutations, which do not inherently dictate the onset of cancer. The process of somatic evolution in cancer is marked by genomic instability and elevated mutation rates, leading to the accumulation of a multitude of mutations and chromosomal changes. The majority of these changes are classified as passenger mutations. While the field of cancer biology has placed greater emphasis on driver mutations, emerging evidence suggests that passenger mutations hold greater significance. They exert influence in critical domains such as epigenetics, mitochondrial DNA, immunogenicity, and the response to chemotherapy (Aparisi F et al., 2019). The distinction between passenger and driver mutations in various cancer types has been clarified by several research. Extensive research has identified several well-known cancer-causing genes. For instance, the *TP53* gene, commonly referred to as the "guardian of the genome," is frequently mutated in various cancer types, including breast, lung, and colorectal cancer (Muller & Vousden, 2014). Moreover, research of colorectal cancer tumor genomes revealed frequent driver mutations in *APC, TP53,* and *KRAS*, *PIK3CA* and *FBXW7* genes, while the bulk of mutations were found to be passengers. (Muzny et

al., 2012). Driver mutations in cancer have significant implications for targeted therapy; the identification of driver mutations in genes such as *KRAS* and *EGFR* in non-small cell lung cancer (NSCLC) has paved the way for the development of targeted therapies. *KRAS* inhibitors, such as Sotorasib (AMG510), as well as *EGFR* inhibitors, such as erlotinib and gefitinib have shown clinical efficacy against NSCLC (Jg et al., 2009; Mok et al., 2009; Santarpia et al., 2023). The identification of passenger and driver mutations is crucial for identifying targetable alterations and developing personalized cancer therapies.

To understand the genetic basis of cancer, it is essential to distinguish between somatic and germline alterations. Somatic mutations take place in non-reproductive cells and are acquired during a person's lifespan as a result of factors like exposure to the environment or abnormalities in DNA replication. Most cancer cases are caused by these mutations, which are not passed down through generations (Stratton et al., 2009). In 518 genes across 210 different cancers, 82 percent show somatic mutations of which most were also identified as passenger mutations, while 18 percent show driver mutations (Greenman et al., 2007). In contrast, germline mutations are heritable changes found in reproductive cells that can be passed down from parents, these mutations have been discovered in between 5% and 10% of all malignancies (von Stedingk et al., 2021). Germline mutations are linked to a genetic cancer predisposition syndrome and greatly raise the risk of developing specific cancers, such as the hereditary breast and ovarian cancer syndrome caused by the *BRCA1* and *BRCA2* mutations. (Miki et al., 1994; Richard Wooster, Graham Bignell, Jonathan Lancaster, Sally Swift, Sheila Seal, Jonathan Mangion, Nadine Collins, Simon Gregory, Curtis Gumbs, Gos Micklem, Rita Barfoot, Rifat Hamoudi, Sandeep Patel, Catherine Rices, Patrick Biggs, Yasmin Hashim, Amanda Sm, 1995).

Modern diagnosis and prognosis of cancer patients include identifying individuals (and possible relatives) at risk of hereditary cancer syndromes and delivering evidence-based care for those with a known mutation (Van Cott, 2020).

### 1.4. Colorectal Cancer Epidemiology and Development

Colorectal cancer refers to malignant tumors that originate in the colon or rectum. It is the second most deadly cancer (Hossain et al., 2022), third most common cancer globally and 4th most widely diagnosed cancer worldwide, accounting for a significant burden of morbidity and mortality (Miller et al., 2020; Rawla et al., 2019a). Colorectal cancer is a complex disease that is influenced by both environmental and genetic factors. High-fat diets, sedentary lifestyles, obesity, smoking, and binge drinking are all lifestyle choices that have been linked to an increased risk of colorectal cancer (Johnson et al., 2013; S. Y. Park et al., 2019).

Up to 95% of instances of colorectal cancer are adenocarcinoma; other types of CRC including: carcinoid tumors, gastrointestinal stromal tumors, lymphomas, and sarcomas are less common (Mattiuzzi et al., 2019). Colorectal adenocarcinoma mostly develops from the epithelial cells of the large intestine. This kind of cancer arises when certain epithelial cells acquire a set of genetic or epigenetic modifications that provide them a selective advantage. These hyper-proliferative cells lead to the development of benign adenoma with abnormally elevated replication and survival, which can then turn into carcinoma and metastasize over time (Ewing et al., 2014a; Rawla et al., 2019b). It is crucial to understand the molecular basis of CRC since it can establish variables that relate to growth initiation and progression maintenance. Moreover, the molecular understanding of colorectal cancer has led to the establishment of predictive markers that

are used to select patients who are susceptible to specific forms of therapy, the development of molecular diagnostic tests to identify early non-invasive cancers, and data used for genetic testing of family forms (Munteanu & Mastalier, 2014).

### 1.4.1 Genetics and Epigenetics of CRC

Genetic alterations play a crucial role in colorectal cancer pathogenesis. The adenoma-carcinoma sequence, proposed by Fearon and Vogelstein, describes the stepwise accumulation of genetic alterations during tumor development (E. F. Fearon & Vogelstein, 1990). The pathogenesis of CRC involves a variety of genetic alterations that contribute to tumor initiation, progression, and metastasis. While most cases of CRC are sporadic and result from somatic mutations, a subset of cases can be attributed to germline mutations in specific genes. Germline mutations are hereditary and can be passed down through generations, increasing an individual's susceptibility to developing CRC. Understanding the genetic basis of CRC is crucial for identifying high-risk individuals, implementing appropriate screening strategies, and developing targeted therapies (Jasperson et al., 2010).

A key driver in CRC development is the mutation of the adenomatous polyposis coli (*APC*) gene, which plays an integral role in the Wnt signaling pathway. The germline *APC* mutations cause Familial adenomatous polyposis (FAP), which is a hereditary disorder characterized by the growth of multiple polyps in the colon and rectum (Leoz & Moreira, 2015). These polyps have a high potential to progress into cancer if left untreated. Mutations in *APC*, as well as other genes such as *KRAS*, *TP53*, and *PIK3CA*, have been extensively implicated in the initiation and progression of CRC (E. R. Fearon, 2011; Fodde, 2002; T. Tanaka, 2009; Vogelstein et al., 2013). True inherited mutations linked to colorectal cancer, like those involving mismatch repair genes (*MLH1, MSH2,*

*PMS2*) and *APC* genes, are relatively rare and account for approximately 5% of CRC cases. Nevertheless, the examination of these hereditary mutations, along with sporadic mutations in genes like *APC* and DNA mismatch repair genes, has yielded invaluable insights into the gradual genetic evolution from precancerous localized polyps to full-fledged metastatic cancer (Simon, 2016). Another commonly mutated gene in hereditary CRC is the DNA mismatch repair (MMR), resulting in Lynch syndrome (also known as hereditary nonpolyposis colorectal cancer, HNPCC) (Kempers et al., 2011). Lynch syndrome accounts for approximately 2-4% of all CRC cases and is associated with a higher risk of developing other cancers, including endometrial, ovarian, and gastric cancers. (Bhattacharya & McHugh, 2023) Several studies have elucidated the role of specific MMR gene mutations in Lynch syndrome-associated colorectal cancer. For instance, mutations in the *MLH1* gene have been linked to a high risk of developing colorectal cancer. It was found that *MLH1* mutations interfere with the MMR mechanism, causing an increase in DNA mismatches and a subsequent tumorigenic process (Peltomäki et al., 1993, Wagner et al., 2003). Similarly, mutations in the *MSH2* gene have also been associated with Lynch syndrome-related colorectal cancer (Shao et al., 2020).

A multitude of genes are associated with the signaling pathways that are frequently proven to be dysregulated in colorectal cancer (CRC) due to mutations or changes in the function of their products. Genes including *EGFR, RAS, RAF*, *Notch-1*, *Jagged-1*, *PIK3CA, PTEN, TGFBR1, TGFBR2, SMADs*, *AXIN*, and *CTNNB1* have been identified as being responsible for enhanced proliferation, invasion, progression, or suppression of apoptosis in CRC cells in several research (Jeong et al., 2018; Li et al., 2014; Liao et al., 2018; W. K. K. Wu et al., 2012). Tumor-Suppressor genes and oncogenes, that are usually

associated with CRC, include but are not limited to: *APC, MLH1, MSH2, MSH6, PMS2, TP53, TGFBR2, KRAS, SMAD4, PTEN, BRAF* (Munteanu & Mastalier, 2014).

Sporadic colorectal cancers frequently arise from preneoplastic lesions through the activation of oncogenes (*KRAS* and *BRAF*) as well as the inactivation of tumor suppressor genes (*APC, p16, p53,* and *DCC*) and mismatch repair genes, such as *MLH1* and *MSH2* and, to a lower extent, *PMS2* and *hMSH6* (H. Tanaka et al., 2006). Moreover, studies have demonstrated that *PTEN* mutations and loss of function promote cell proliferation, block apoptosis, and have a role in cetuximab and panitumumab resistance, all of which contribute to the progression of colorectal cancer (Salvatore et al., 2019).

*KRAS* encodes a GTPase protein involved in cell signaling. Activating mutations in *KRAS* have been extensively studied and are associated with poor prognosis and resistance to targeted therapies (Santarpia et al., 2023). Research have shown that *KRAS* mutations lead to the constitutive activation of downstream signaling pathways, promoting cell survival and proliferation in colorectal cancer (Hsu et al., 2016, László et al., 2021).

In colorectal cancer, *BRAF*, a downstream *KRAS* effector, is commonly mutated, especially in tumors with microsatellite instability. A worse prognosis is indicated by *BRAF* mutations, which are connected to a specific molecular subtype of colorectal cancer. The MAPK/ERK signaling system is dysregulated by *BRAF* mutations, which contribute to tumor growth and invasiveness.(Samowitz et al., 2005; Ogino et al., 2012).

TGF-signaling pathway, which controls cell proliferation, differentiation, and death, also includes *TGFBR2* and *SMAD4*. The onset and spread of colorectal cancer have been linked to mutations in the genes *SMAD4* and *TGFBR2*. The inactivation of *TGFBR2* and *SMAD4* results in the dysregulation of downstream signaling events, which promotes cell proliferation and invasion in colorectal cancer (Biankin et al., 2002; Zhao et al., 2018)

The epigenome of colon cancer has been studied, and it has been discovered that nearly all CRCs have abnormally methylated genes, with the average CRC methylome having hundreds to thousands of aberrantly methylated genes (Lao & Grady, 2011; Yoshinaga Okugawa, William M. Grady, 2017). By altering chromatin structure and gene expression patterns, epigenetic changes play a significant role in the onset and spread of colorectal cancer. The pathogenesis of some malignancies, including CRC, is heavily influenced by epigenetics, which are heritable changes in gene expression without alterations in DNA sequence (Goel & Boland, 2012). Histone modifications, such as changes in histone acetylation, which are mediated by enzymes like histone deacetylases (HDACs), have been observed in colorectal cancer, studies have shown that aberrant histone deacetylation promotes colorectal tumorigenesis by silencing genes (Qin et al., 2019; Ropero et al., 2006). Moreover, chromosome instability has also been demonstrated to result from worldwide hypomethylation in Colorectal Cancer (Suter et al., 2010). Additionally, non-coding RNAs, such as microRNAs (miRNAs), have become essential players in the control of colorectal cancer's epigenetics. MiRNA dysregulation has been connected to several mechanisms that contribute to the growth and spread of tumors. Specific miRNAs such as (miR-29a, miR-101, miR-125b, miR-146a and miR-155) are often dysregulated in colorectal cancer (Ždralević et al., 2023). miRNAs can act as either oncogenes or tumor suppressors, influencing important pathways and processes such cell proliferation, apoptosis, and metastasis. (Schetter et al., 2012; Wang, 2020)

### 1.5. *SMAD4*: Navigating the Landscape of Multiple Signaling Pathways in Cancer.

*SMAD4* is one of the most frequently altered genes in colorectal malignancies, with inactivating mutations present in around 5.0–24.2% of CRCs (Fang et al., 2021). *SMAD4*,

also known as *DPC4* (deleted in pancreatic cancer 4), is a tumor suppressor gene that plays a critical role in various cellular processes. It serves as a key component of the transforming growth factor-beta (TGF-β) signaling pathway, which regulates numerous cellular functions, including pluripotency and plasticity (Chen et al., 2019), metastasis (Ding et al., 2011), cell growth, differentiation, apoptosis, immune response (Mccarthy & Chetty, 2018), cell proliferation (Tong, 2019), as well as, autophagy (Newman et al., 2017) and invasion (Voorneveld et al., 2014).

The development and progression of many malignancies, including pancreatic, colorectal, and gastric cancers, have been linked to *SMAD4* loss or inactivation (Alazzouzi et al., 2005).

SMADs are a family of proteins that directly transmit extracellular signals to the nucleus. Mammalian cells have eight distinct SMADs, which can be categorized into three groups: (i) the receptor-regulated SMAD (R-SMAD), which consists of the SMADs 1, 2, 3, and 5 and SMADs 8 and 9. (ii) The common-mediator SMAD (Co-SMAD) includes only SMAD4, which is the central mediator of both TGF- and BMP signaling pathways; (iii) the inhibitory SMAD (I-SMAD) includes SMAD6 and SMAD7, which primarily function to inhibit receptor-mediated R-SMAD phosphorylation, preventing the complex formation with Co-SMAD (Zhao et al., 2018).

While SMAD2/3 primarily regulate signaling from TGF- subfamily members, SMAD4 also functions as a transcription factor that mediates the TGF-β signaling cascade (Shi & Massague, 2003). Upon activation of TGF-β receptors, the receptors phosphorylate and activate SMAD proteins, including SMAD4. Activated SMAD4 then forms a complex with other phosphorylated SMADs and translocate into the nucleus, where it interacts with transcription factors and regulates the expression of target genes (Hata & Chen,

2016). This process modulates cellular responses to TGF-β, exerting both tumor-suppressive and pro-oncogenic effects depending on the context (Tang et al., 2003).

SMAD4 participates in different signaling pathways downstream of TGF-β, canonical and non-canonical pathways. Using the canonical SMAD pathway; the activated TGF-β receptor phosphorylates SMAD2 and SMAD3, which then form complexes with SMAD4. This complex translocates into the nucleus and regulates gene expression (Ahmed et al., 2017). Another pathway involves the interaction of SMAD4 with other transcription factors and co-regulators. The binding of the ligand to type I and type II serine-threonine kinase receptors initiates the bone morphogenetic protein 2/4 (BMP2/4) signaling pathway. The recruited and phosphorylated receptor Smads (Smad1, Smad5, or Smad8) are then associated with SMAD4 by the activated receptor complex. The complex subsequently moves to the nucleus where it takes part in modulating transcription (Hussein et al., 2003). Moreover, the site of SMAD3 and SMAD4 specific binding was found to be in the goosecoid promoter to GC-rich areas. The structural basis for the binding site illustrated a consensus sequence (5CG) GGCGC; this sequence presents a binding site for BMP and TGF-β activated Smads as well as SMAD4. The 5CG motifs are illustrated as clusters in genome wide regions that are Smad-bound.(Martin-Malpartida et al., 2017)

SMAD4 also plays a role in non-canonical TGF-β signaling, where it interacts with various proteins to regulate cellular functions independently of the SMAD pathway (Itoh & ten Dijke, 2007). In addition, SMAD4 plays a crucial role in preventing the suppression of Natural Killer (NK) cells by TGF-β. It achieves this by acting as an unexpected suppressor of non-canonical TGF-β signaling through TGFβR1 in conventional NK cells (Cortez et al., 2017).

### 1.5.1 The Impact of *SMAD4* Loss on Cancer Development and Progression

*SMAD4*/*DPC4* is a critical gene involved in the development and progression of various cancers. It acts as a tumor suppressor by regulating key cellular processes. Loss or inactivation of *SMAD4* has been implicated in the pathogenesis of pancreatic cancer (Hahn et al., 1996), colorectal cancer (Alazzouzi et al., 2005), gastric cancer (D. M. Wu et al., 2010), and other malignancies.

In pancreatic duct adenocarcinoma, deletion of *SMAD4* was first identified and then detected in different types of cancers, including colorectal cancer, cholangiocarcinoma, gastric cancer, prostate cancer, and many others (Zhao et al., 2018). In pancreatic cancer, *SMAD4* alterations are prevalent and associated with poor prognosis. Studies have shown that *SMAD4* loss or mutation occurs during the early stages of pancreatic cancer progression, promoting tumor growth and invasion (Wilentz et al., 2000). Additionally, *SMAD4* inactivation is involved in the resistance to therapeutic interventions in pancreatic cancer (Bardeesy et al., 2006).

Colorectal cancer is another malignancy where *SMAD4* plays a critical role. *SMAD4* mutations and loss of expression are observed in a subset of colorectal tumors (Ma et al., 2014). Such alterations contribute to tumor progression and metastasis through dysregulated TGF-β signaling, leading to uncontrolled cell growth, epithelial-mesenchymal transition (EMT), and angiogenesis (Goumans et al., 2009; Hao et al., 2019).

In gastric cancer, *SMAD4* alterations are associated with an aggressive phenotype and poor patient prognosis. Studies have demonstrated that *SMAD4* loss contributes to increased tumor invasiveness, lymph node metastasis, and reduced patient survival.

Moreover, SMAD4 deficiency in gastric cancer disrupts TGF-β signaling, resulting in the dysregulation of genes involved in cell cycle control and metastasis (Wan et al., 2021; X. Xu et al., 2000).

Furthermore, *SMAD4* alterations have been implicated in other cancer types. In lung adenocarcinoma, *SMAD4* loss or inactivation promotes aggressive tumorigenesis and metastasis by activating *PAK3* via miRNA regulation (Tan et al., 2021). In head and neck squamous cell carcinoma (HNSCC), knockdown of *SMAD*4 is associated with enhanced migration and invasion of HNSCC cells (Lin et al., 2019). In hepatocellular carcinoma, *SMAD4* inactivation is involved in tumor progression and poor patient survival (Hernanda et al., 2014).

### 1.5.2 *SMAD4* Loss Contributes to the Development and Progression of Colorectal Cancer

*SMAD4* is somatically inactivated in most colorectal cancers. Frameshift mutations and point mutations with functional implications have been found in addition to deletions, providing strong evidence that *SMAD4* undergoes specific genetic alterations in carcinogenesis (Salovaara et al., 2002). Several data show that loss of *SMAD4* is associated worse clinical results, resistance to chemotherapy, and reduced immune infiltration, justifying its utilization in patients with colorectal cancer as a prognostic marker (Wasserman et al., 2019).

Loss of *SMAD4* expression in colorectal cancer is often linked to aggressive tumor behavior and increased metastatic potential. Reduced *SMAD4* expression has been associated with lymph node metastasis and distant metastasis in colorectal cancer patients (T. Tanaka et al., 2008; Zhang et al., 2010). Additionally, SMAD4 deficiency is correlated

with the activation of the epithelial-mesenchymal transition (EMT) process, which promotes tumor invasion and metastasis (Bardeesy et al., 2006; Link et al., 2022)

Furthermore, *SMAD4* alterations interact with other molecular pathways involved in colorectal cancer progression. The interaction between SMAD4 and Wnt signaling pathway components, such as β-catenin, contributes to the dysregulation of cell growth and differentiation processes (Romero et al., 2008). *SMAD4* also interacts with *p53*, a critical tumor suppressor gene, and alterations in both genes have been observed in colorectal cancer (J. W. Park et al., 2022). Moreover, according to a study, *SMAD4*, *SMAD2*, and *SMAD3* mutations were found in 8.6%, 3.4%, and 4.3% of sporadic colorectal cancer cases, respectively (Fleming et al., 2013).

In the TGF-β pathway, SMAD4 is an integral mediator. TGF-β performs two roles in cancer; in earlier stages, it suppresses tumor development and in more advanced stages, it supports tumor progression. Moreover, in 2.1-20.0% of colorectal cancers, sporadic *SMAD4* mutations are present (Sarshekeh et al., 2017). According to the Cancer Genome Atlas data, *SMAD4* mutations accompanied with *KRAS, NRAS, and BRAF* mutations in metastatic patients are at higher risk of disease progression. (Sarshekeh et al., 2017).

*SMAD4* alterations also play a role in the resistance to chemotherapy, recurrence, and loss of immune infiltrate in colorectal cancer (Wasserman et al., 2019). Studies have shown that loss of *SMAD4* expression is associated with resistance to 5-fluorouracil (5-FU), a commonly used chemotherapy drug in colorectal cancer treatment (Zhang et al., 2014). The dysregulation of SMAD4-mediated signaling pathways contributes to the resistance mechanisms (Xu et al., 2021), highlighting the importance of *SMAD4* in therapeutic response.

### 1.6. Thesis Statement

Colorectal cancer is identified as the second most deadly cancer, and the third most common cancer globally. In Palestine, colorectal cancer ranks first in males and second most common in females with 440 new cases in 2022 and an incidence rate of 15.3 per 100,000 population. These statistics signify the importance of investigating the underlying genetic mutations in colorectal cancer among Palestinians. This research utilizes NGS technique and mutation functional analysis to identify the significance and the pathogenicity of previously unknown mutations found in a Palestinian patient. The findings will help better understand and build the genetic landscape of colorectal cancer in Palestine.

### 1.7. Study Objectives

The study employs scientifically advanced techniques to identify novel mutations in a Palestinian Colorectal cancer patient. In-silico tools determined the most prominent pathogenic novel mutation (*SMAD4*, p. Cys115Arg) followed by functional analysis to confirm pathogenicity. The gene *SMAD4*, a key component of the TGF-β signaling pathway, has been linked to the pathophysiology of colorectal cancer. However, the specific mutation profile of *SMAD4* in the Palestinian population remains poorly understood. By identifying a novel mutation in *SMAD4*, this study seeks to (i) contribute to the growing body of knowledge on the genetic landscape of colorectal cancer, (ii) providing insights into the molecular mechanisms underlying the disease and (iii) potentially paving the way for targeted therapies tailored to the Palestinian population. The study's findings may have important implications for the diagnosis, prognosis, and treatment of colorectal cancer patients in Palestine and beyond.

### 1.8. Study Significance

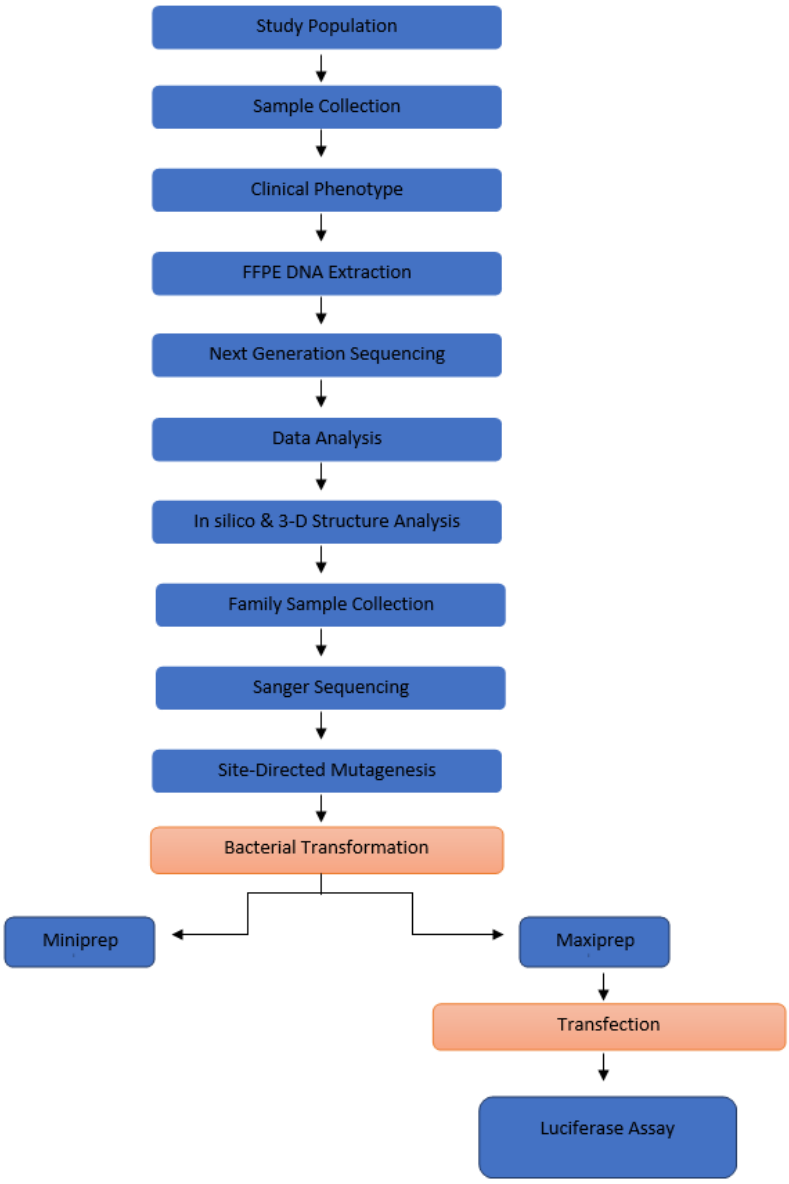
The research's significance underlies in the understanding of the underlying genetic makeup of colorectal cancer in Palestinian patients. Identifying novel mutations in CRC patients offers invaluable insights into the precise mechanisms involved in tumor start and progression. Moreover, by characterizing the specific mutation in *SMAD4*, researchers can gain insights into the functional consequences of the mutation and its impact on tumor development and progression. Finally, uncovering novel mutations in CRC within the Palestinian population can contribute to a better understanding of the global diversity of colorectal cancer and enhance our knowledge of the disease at a local and global scale.

# Chapter Two: Materials and Methods

A colorectal cancer patient was referred to the National Center for Cancer Diagnostics and Human Genetics (NCCDGH) for genetic testing. The colonic tissue embedded in paraffin was used to detect genetic variations with GeneRead QIAact Actionable Insights Tumor Panel; sequenced using the GeneReader platform (QIAGEN, Hilden, Germany). The patient's clinical information was obtained to facilitate the variant analysis process for clinical diagnosis using CLC workbench and QCI-I. When no known pathogenic variants were detected, in-silico analysis and protein 3-D structure tools were implemented to predict the pathogenicity of the variants of unknown significance (VUS). In-silico analysis identified (*SMAD4,* p. Cys115Arg) as a pathogenic variant. Next, functional assays were performed to determine the mutation's effect on *SMAD4* transcriptional activity.

## 2.1 Study Design

This study was designed as shown by the schematic below:




Our study was designed to target a Palestinian colorectal cancer patient with no known significant variants in order to identify a novel mutation that could be the underlying cause of the development or progression of patient's disease. Subsequently, functional analysis was done to verify the pathogenicity of the variant and its role in the loss of the gene's transcriptional activity.

### 2.2 Study Subject and Sample Collection

The NCCDGH received a colonic formalin-Fixed Paraffin-Embedded (FFPE) tissue resected from the large intestine from a colorectal cancer patient. The patient is a 39-Year male that had complains about chronic abdominal pain and constipation, he was referred to an oncologist where he was diagnosed in 2019 with malignant neoplasm of sigmoid colon (colorectal adenocarcinoma). The patient underwent chemotherapy using the FOLFOX regimen (5-Fluorouracil/5-FU, Oxaliplatin, and Leucovorin Calcium) followed by targeted therapy with Herceptin.

A field visit was made to the patient's family for blood samples collection. EDTA Whole-Blood samples (around 2.5 mL) were collected from the patient and the mother following standard health protocols.

### 2.3 Ethical Considerations

Patient and families' members recruited in this study signed consent forms prior to participation in the study voluntarily agreeing to participate in research. Participants were assigned ID numbers to ensure confidentiality. Approval was obtained by the Helsinki Committee for Ethical Approval, Palestinian Health Research Council, Gaza, Palestine. [Ethical approval number PHRC/HC/1216/22].

### 2.4 Next-Generation Sequencing

GeneRead QIAact AIT DNA UMI Kit includes 2 primer mix tubes with roughly 500 primers each. The Kit uses 40–160 ng of DNA to enrich target areas in a wide range of genes, including: *AKT1, ALK1, BRAF, CTNNB1, DDR2, EGFR, ERBB2, ERBB3, ERBB4, ESR1, FBXW7, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAQ, HRAS, KIT, KRAS* and *MAP2K1*.


Figure 1: Overview of the complete NGS workflow. Summary of NGS workflow steps using the GeneRead QIAact Actionable Insights Tumor Panel.

#### 2.4.1 Nucleic Acid Extraction from FFPE tissue

A microtome was used to cut 10 µm thick sections from FFPE sample (excess paraffin was trimmed using a scalpel). The sections were then placed in a 1.5 mL microcentrifuge (Eppendorf, Hamburg, Germany) and 160 µl Deparaffinization Solution was added. The mixture was vortexed vigorously, centrifuged briefly, and incubated for 3 minutes at 56°C. 55 µl RNase-free water, 25 µl Buffer FTB and 20 µl proteinase K were then added respectively. The mixture was then vortexed, centrifuged briefly, then incubated for 1 h at 56°C followed by another 1h incubation at 90°C. The clear phase was transferred into a new 1.5 mL microcentrifuge tube, 115 µl RNase-free water and 35 µl UNG were added to the sample. The mixture was vortexed and incubated for one hour at 50°C.

An hour later, 250 µl of both lysis Buffer AL and (96–100%) ethanol were added to the sample and mixed thoroughly by vortex. 700 µl of the lysate was transferred to the QIAamp MinElute column (in a 2 ml collection tube) and centrifuged at 20,000 x g; 14,000 rpm for 1 minute, flow-through was discarded. 500 µl wash Buffer AW1, 500 µl

Buffer AW2 and 250 µl ethanol (96–100%) were added respectively to the spin column. Each of them was centrifuged at maximum speed (20,000 x g;14,000 rpm) for 1 minute, and the flow-through was discarded each time. The spin column was then placed into a new 2 ml collection tube and centrifuged at maximum speed for 1 min to remove any residual liquid. Finally, the QIAamp MinElute column was placed in a clean 1.5 ml microcentrifuge tube and 40 µl elution Buffer ATE was added to the center of the membrane. To collect the nucleic acid in the microcentrifuge tube, the tube was centrifuged at maximum speed (20,000 x g:14,000 rpm) for 1 minute. The DNA yielded was used as a template for NGS.

### 2.4.2 Target Enrichment and Library preparation

Fragmentation:

Using a single, carefully controlled multi-enzyme procedure, genomic DNA samples were first fragmented, end-repaired, and A-tailed. 16 µL of diluted DNA (2.5 ng) was used as a template for fragmentation. A reaction mix was prepared using 2.5 µL of Fragmentation buffer (10X), 0.75 µL FERA solution, and 0.75 µL Nuclease-Free water for each sample. The mix was added to the sample followed by 5 µL of Fragmentation enzyme. The final reaction was mixed by pipetting and was placed on the thermocycler with the following conditions: 1 minute at 4°C, 24 minutes at 32°C, 30 minutes at 72°C, and cool down/hold at 4°C.

Adapter Ligation:

The prepared 25 µL DNA fragments are then ligated at their 5’ ends to a GeneReader-specific adapter containing a UMI and a 9 base-pair (bp) sample-specific bar code. The ligation master mix was prepared as follows: 10 µL of 5X Ligation buffer, 5 µL of DNA

ligase, 2.8 µL of QIAact Adapter and 7.2 µL Ligation Solution (PEG6000, 50%) for each sample. 25 µL of each fragmentation product was transferred into a 0.2 ml PCR tube, followed by 22.2 µL of ligation mix to each sample. The final solution was mixed by pipetting up and down then the mixture was placed on the thermocycler for 15 minutes at 20°C.

Cleanup of Adapter-ligated DNA with QIAact Beads:

Next, 50 µl of ligation reaction from "Protocol: Adapter Ligation" was transferred to a 1.5 ml LoBind tube (Eppendorf®). 50 µl of nuclease-free water was added to bring the volume to 100 µl. 1.0x volume QIAact Beads to 100 µl diluted DNA solution, mixed well by pipetting and incubated for 5 minutes at room temperature. The tubes were then placed on the magnetic rack, (DynaMag™-2 Magnet (Thermo Fisher Scientific, cat. no. 12321D)), the clear supernatant was discarded, the beads were washed twice with 2.0x volume (200 µl) 80% Ethanol, and the product was eluted with 52 µl. 50 µl of the elute was transferred to a clean 1.5 ml tube where 1.0x volume QIAact Beads (50 µl) was added to the product and a second clean-up was performed. DNA was finally eluted from beads in 22 µl nuclease-free water and 9.4 µl of the volume was used as a template for each target enrichment reaction.

Target Enrichment:

Ligated DNA molecules are subjected to a restricted number of target enrichment PCR cycles. One gene-specific primer targets a region, while the other is a universal forward primer that works in conjunction with an adapter sequence. Through this reaction, it is assured that the targeted targets and UMIs are adequately enriched to be included in the final library.

Two reactions were prepared in separate 1.5 ml LoBind tubes: forward target enrichment PCR, and reverse target enrichment PCR. Both reactions were prepared as follows: 4 µl of 5X TEPCR buffer, 3.75 µl Nuclease-Free water, 0.8 µl GeneReader TE-PCR primer, 0.8 µl HotStarTaq DNA polymerase, 9.4 µl of the cleaned-up product, and 1.75 µl of either GeneRead QIAact AIT DNA UMI Kit Forward Primers or GeneRead QIAact AIT DNA UMI Kit Reverse Primers. The thermal cycler (Veriti™ 96-Well Fast Thermal Cycler, Applied Biosystems™) was programmed as follows: 13 minutes at 95°C, 2 minutes at 98°C, (15 seconds at 98°C, 10 minutes at 68°C – 8 cycles), 5 minutes at 72°C, 5 minutes at 4°C then hold at 4°C.

The target enrichment product is then cleaned up with QIAact beads and washed twice with 80% ethanol, the DNA is eluted with 16 µl nuclease-free water and 13.4 µl is used as a template for Universal PCR Amplification protocol.

Universal PCR Amplification

The targets are then amplified with Universal PCR protocol using GeneReader-specific sequences to build the library. 13.4 µl of the target enrichment cleaned-up product is used as a template along with 6.6 µl of the reaction mix; the reaction mix was prepared as follows: 4 µl of UPCR buffer, 5X, 0.8 µl of both GeneReader Universal PCR Primer A and B, 1 µl of HotStarTaq DNA Polymerase. The reaction was set up on the following conditions: 13 minutes at 95°C , 2 minutes at 98°C, (15 seconds at 98°C, 2 minutes – 60°C, 21 cycles), 5 minutes at 72°C, then 5 minutes at 4°C, and hold at 4°C.

When the reaction is complete, we proceed to the cleanup of Universal PCR with QIAact beads protocol which is identical to the cleanup of target enrichment. However, the library was eluted with 30 µl of nuclease-free water and 28 µl of those were transferred to clean 0.2 mL tubes.

### 2.4.3 Library Quantification

The library was quantified using Qubit™ 4 Fluorometer (Thermo Fisher Scientific™ cat. no. Q33238) with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific cat. no. Q32851). Calibration of the Qubit® Fluorometer requires the standards to be inserted into the instrument in the right order, the sample was then measured as per the kit's instructions. The Qubit reading for the sample was 20.8 ng/µl.

### 2.4.4 Library Analysis using QIAxcel

To check the fragment size and concentration, QIAxcel Advanced in combination with the QIAxcel DNA High Resolution Kit (QIAQEN, Germany, cat. no. 929002) were used. The typical peak height for DNA extracted from FFPE tissue is 350 bp on average. The bulk of the library fragments range in size from 200 to 800 bp. Libraries can be quantified using the amount of DNA under the peak.

### 2.4.5 Clonal Amplification and sequencing

The GeneRead Clonal Amp Q Kit is used to create sequencing templates utilizing clonal amplification after the library preparation completion using GeneRead DNA Library Q Kit. Finally, for the preparation of DNA sequencing using the QIAGEN GeneReader® instrument, QIAGEN GeneRead® UMI Advanced Sequencing Q Kit Handbook was used for the setup instructions.

**2.5 In-Silico Analysis**

A list of all the mutations that were found by NGS were studied thoroughly to choose the most impactful and deleterious mutation. In silico Analysis was done to the following mutations in various genes: (*NOTCH1*, p.Asp1533Gly), (*EGFR*, p.Leu619Pro), (*MAP2K1*, p.Pro265Ser), (*ALK*, p.Phe174Leu), (*ERBB2*, p.Asp1252His), (*FBXW7*, p.Gly477Arg), (*FGFR2*, p.Met538Ile), (*GNA11*, p.Leu170Phe), (*MET*, p.Asp153Asn), (*NOTCH1*, p.Ala1610Thr), (*PIK3CA*, p.Asn107Thr), (*SMAD4*, p.Cys115Arg) using the following tools: SIFT(https://sift.bii.a-star.edu.sg/), PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), Align GVGD (http://agvgd.hci.utah.edu/agvgd_input.php) and Provean (https://www.jcvi.org/research/provean#downloads). Moreover, the mutations reported were viewed across various tumor types using cBioPortal (https://www.cbioportal.org/) and COSMIC (https://cancer.sanger.ac.uk/cosmic). ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) was also used to verify which conditions were accompanied by the list of mutations we have. BLAST alignment tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to illustrate the conservation of our mutation site. Finally, RCSB Protein Data Bank (RCSB PDB) (https://www.rcsb.org/) was used to visualize and analyze the 3D-Protein structure of (*SMAD4*, p. Cys115Arg).

**2.6 Sample Collection**

Field visits were done to collect EDTA Whole-Blood samples from the patient and the mother. This was done for the purpose of identifying the mutation as somatic or germline before proceeding to functional analysis.

### 2.6.1 DNA Extraction

DNA was extracted from both the patient and his mother by using Wizard® Genomic DNA Purification Kit (Catalog No. PR-A1120). First, blood samples were centrifuged at 4000 rpm for 10 minutes. 300 µL from the buffy coat was then taken to an Eppendorf tube where 900 µL of cell lysis solution was added to it. The tube was inverted 5-6 times to mix the solution and then incubated for 10 minutes at room temperature. After the incubation period was over, the solution was centrifuged at (20,000 x g;14,000 rpm) for 1 minute; the supernatant was discarded. The pellet was resuspended in the remaining supernatant by vortex, followed by adding 300 µL Nuclei lysis solution to the resuspended pellet and mixed by pipetting. Then, 100 µL of Protein Precipitation solution was added to the lysate and vortexed for 10-20 seconds which was centrifuged at (20,000 x g;14,000 rpm) at room temperature for 3 minutes. The supernatant is transferred to a new Eppendorf tube, after which 300µL of cold isopropanol was added to it. The tube was inverted to mix the solution and then placed in the centrifuge for 1 minute at (20,000 x g:14,000 rpm). The supernatant was discarded and 300µL of cold 70% Ethanol was added to the precipitate. The tube was inverted and then placed in the centrifuge for 1 minute at (20,000 x g:14,000 rpm). Finally, the supernatant was discarded, the tube was placed in a heat block for 10 minutes at 37°C, and the precipitate was dissolved in 50µL of Nuclease free water. The samples were measured on the NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™) for quality check.

### 2.7 Variant validation by Sanger Sequencing

To validate the findings of NGS mutation (*SMAD4*: c.343T>C, p. Cys115Arg) and detect the presence of the mutation of interest in the patient and family members, Sanger sequencing was used.

#### 2.7.1 Primer Design

Primers to detect the target mutation were designed using Primer3 tool (https://primer3.ut.ee/). First, the target FASTA sequence was obtained using the reference sequence on NCBI (https://www.ncbi.nlm.nih.gov/), the sequence was then inserted in the Primer3 design tool. The best pair of primers were chosen based on the following parameters: product size, GC% content, melting temperature, GC clamp, and others. The chosen primer pair sequence is as follows: Forward 5'ctgagttggtaggattgtgaggatt 3', reverse 5' taaagtcgcgggctatcttcc 3'. Finally, to test the primers' specificity, UCSC in-silico PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) was used.

#### 2.7.2 Polymerase Chain Reaction

PCR was carried out to amplify the *SMAD4* target region in the samples. PCR reactions were prepared using 8 µL GoTaq® Green Master Mix, 2X (Promega, USA, Cat no. M7122), 0.5 µL of 10µM upstream primer, 0.5 µL of 10µM downstream primer, 2 µL of DNA and 9 µL dH2O. The ready reactions were then placed on the thermal cycler (Biometra TONE thermal cycler, Analytik Jena, Germany) with the following conditions: 5 minutes at 95 °C, [30 seconds at 95 °C, 30 seconds at 57 °C, 30 seconds at 72 °C] for 35 cycles, 7 minutes at 72 °C then cool down at 4 °C.

### 2.7.3 Detection of PCR products

The PCR products were run on gel electrophoresis to illustrate the targeted bands and their size. A 1% agarose gel was made by adding 100mLTAE Buffer (Tris-acetate-EDTA) to 1 gram of Agarose (SeaKem® LE Agarose). The mixture was boiled until the solution became clear and homogenized. Finally, 30 µL of Ethidium Bromide was added to the solution and swirled to mix, the gel was poured in a gel cast with combs and left to cool down and solidify. 4 µL of the sample was mixed with 1 µL of 5X loading dye (Quick load, Biological Industries) and loaded into the wells of the gel, along with a 1000 bp ladder. The voltage was set to 400 Volts and 100 Amp and was run for 30 minutes.

### 2.7.4 PCR Products Purification and Sequencing

EPPiC Fast Reagent (A&A Biotechnology) was used to hydrolyze excess primers and nucleotides in order to be able to use the product for DNA Sequencing. 1 µL of EPPiC Fast was mixed with 5 µL of each of the post-PCR reactions, briefly centrifuged and placed on the thermocycler for 10 minutes at 37 °C to activate the enzyme, then for 1 minute at 80 °C for enzyme deactivation.

Next,1 µL of the purified PCR product was used as a template for BigDye™ Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems™, Catalog no. 4337450) protocol, which contains a set of dye terminators labeled with novel, high-sensitivity dyes that enable us to illustrate the bases of our target. 1 µL of 10 pmol forward or reverse primer was added to 1µL of the purified PCR product, with 2 µL BigDye™ Terminator v1.1Ready Reaction Mix, 2 µL BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer and 4 µL of Deionized water (RNase/Dnase-free) for a total volume of 10 µL per each

reaction. The reactions were placed on the thermocycler with the following conditions: 1 minute at 96 °C, [10 seconds at 96 °C, 5 seconds at 55 °C, and 4 minutes at 60 °C] for 25 cycles then hold at 4 °C.

The sequencing reactions were purified with BigDye Xterminator™ (Applied Biosystems™, Catalog no., 4376486). Following the kit's manual, 10 µL of BigDye Xterminator™

bead solution and 45 µL of SAM solution were added to 10 µL of the product. The samples were then vortexed for 30 minutes before they were sequenced on Applied Biosystems 3500 Genetic Analyzer (8-capillary) (Catalog no. 4405673).

### 2.8 Functional Assay: Luciferase Assay

Luciferase assay was used to determine if the mutation (*SMAD4*: c.343T>C, p. Cys115Arg) has an effect on the transcriptional activity of the *SMAD4* gene.

#### 2.8.1 Plasmid Miniprep

Plasmids with the target gene, SMAD4: pBabe-puro-Smad4-Flag plasmid and pCMV5 DPC4-HA plasmid (map N/A), as well as two reporters for the luciferase assay: SBE4-Luc plasmid and TRS6E1b-luc (-732/-721) plasmid (map N/A) were obtained from Addgene(https://www.addgene.org/).

1. pBabe-puro-Smad4-Flag plasmid

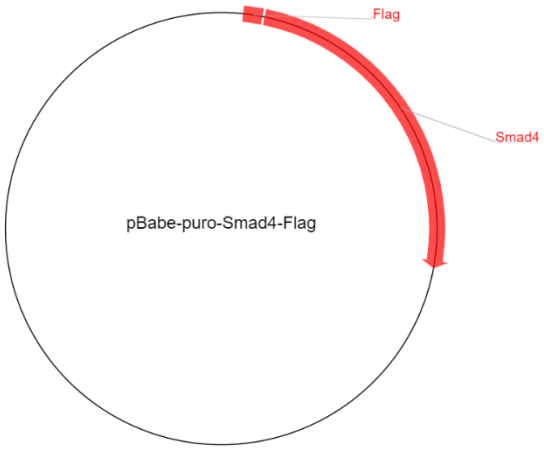



Figure 2: plasmid map for pBabe-puro-Smad4-Flag.
Plasmid number on https://www.addgene.org/: 37041

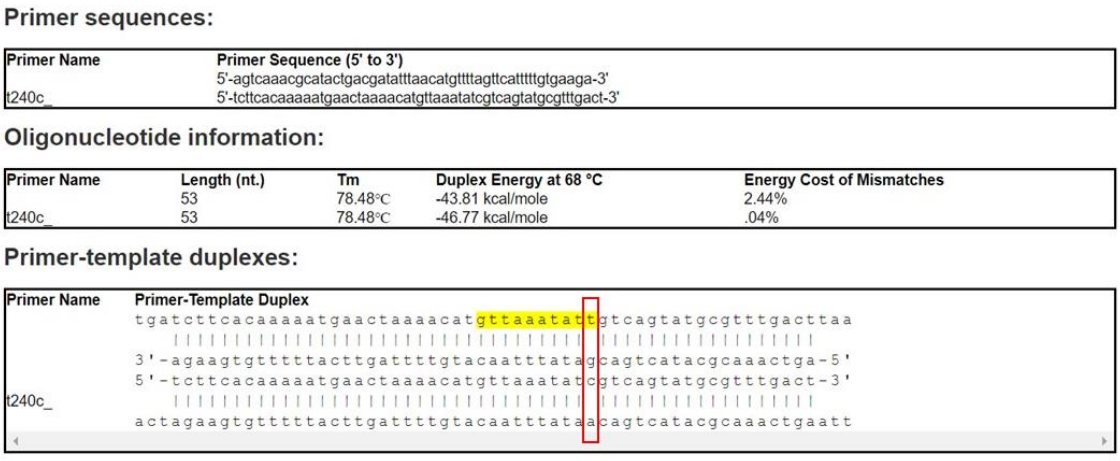
2. SBE4-LUC Plasmid


Figure 3: Plasmid map of SBE4-Luc. plasmid number on https://www.addgene.org/:16495

The aforementioned plasmids were sent in bacteria as agar stabs. Minipreps were prepared using Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit (Catalog no., 740615.250). Each of the transformed bacteria was scraped and placed in 5mLbroth. 100mLof the broth was prepared as follows: 1 gram of Tryptone, 0.5 grams of Yeast, and 1 gram of NACL. The broth containing the transformed bacteria was grown in a shaker incubator for 7-8 hours at 37 °C. The 15-ml conical tubes containing the bacteria were then placed in the centrifuge at 1500 x g for 10 minutes. The supernatant was discarded and 150 µL of A1 resuspension buffer was used to resuspend the pellet. The suspension was transferred to an Eppendorf tube. 250 µL of A2 buffer was added to the suspension, the mixture was inverted 5-6 times and incubated for 2 minutes. Next, 350 µL of A3 buffer was added; the lysate was inverted until it became colorless. The Eppendorf tube containing the lysate was placed in the centrifuge at 12,000 x g for 3 minutes. 700 µL of the supernatant was then transferred into the columns provided in the Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit, the columns were placed in collection tubes and then centrifuged for 30 seconds at 2000 x g. After the supernatant has passed through the column to the collection tube, the collection tube was discarded, and the column was inserted into a new collection tube. Then, 450 µL of AQ buffer was added to the column and centrifuged for 1 minute at 12,000 x g. The column was then placed in a new collection tube and centrifuged at the same conditions again to dry out the column before elution. Finally, 50 µL of AE buffer was added to the spin column and left to incubate at room temperature for 3 minutes. The spin column was then centrifuged for 1 minute at 12,000 g to elute the plasmids. Each of the plasmids were then measured on the NanoDrop™ 2000/2000c Spectrophotometer (Thermo Scientific™) using AE as a blank.

### 2.8.2 Site-directed Mutagenesis

Site-directed mutagenesis was done using QuikChange Lightning Site-Directed Mutagenesis Kit (Catalog no., 210518) to insert the mutation (*SMAD4*: c.343T>C, p. Cys115Arg) into the following plasmids: pBabe-puro-Smad4-Flag #37041 and pCMV5 DPC4-HA #14038. The primers that were used to introduce the mutation were designed using the Quickchange primer design website: www.agilent.com/genomics/qcpd. The best pair of primers were chosen according to appropriate parameters: Forward 5'-agtcaaacgcatactgacgatatttaacatgttttagttcatttttgtgaaga-3',Reverse5'-tcttcacaaaaatgaactaaaacatgttaaatatcgtcagtatgcgtttgact-3' (Figure 4). The Site-directed mutagenesis reaction was prepared as follows: 0.4 µL (125 ng) of each forward and reverse primer, 5 µL of 10X reaction buffer, 0.6 µL (100 ng) of DNA template, 1 µL dNTP mix, 42.6 µL of double distilled water, and 1 µL of QuikChange Lightning Enzyme. The reactions were placed on the thermocycler with the following conditions: 2 minutes at 95 °C, [20 seconds at 95 °C, 10 seconds at 60 °C, 7 minutes, and 30 seconds (30 sec/kb of plasmid) at 68 °C] 18 cycles, 5 minutes at 68 °C, hold at 4 °C.



**Primer sequences:**

| Primer Name | Primer Sequence (5' to 3') |
|---|---|
| | 5'-agtcaaacgcatactgacgatatttaacatgttttagttcatttttgtgaaga-3' |
| t240c_ | 5'-tcttcacaaaaatgaactaaaacatgttaaatatcgtcagtatgcgtttgact-3' |

**Oligonucleotide information:**

| Primer Name | Length (nt.) | Tm | Duplex Energy at 68 °C | Energy Cost of Mismatches |
|---|---|---|---|---|
| | 53 | 78.48°C | -43.81 kcal/mole | 2.44% |
| t240c_ | 53 | 78.48°C | -46.77 kcal/mole | .04% |

**Primer-template duplexes:**

| Primer Name | Primer-Template Duplex |
|---|---|
| | tgatcttcacaaaaatgaactaaaacatgttaaatattgtcagtatgcgtttgacttaa |
| | 3'-agaagtgttttacttgattttgtacaatttatagcagtcatacgcaaactga-5' |
| | 5'-tcttcacaaaaatgaactaaaacatgttaaatatcgtcagtatgcgtttgact-3' |
| t240c_ | actagaagtgttttacttgattttgtacaatttataacagtcatacgcaaactgaatt |

Figure 4: Mutagenesis primers designed with www.agilent.com/genomics/qcpd. Shown in the figure above are the primer sequences used for inducing the *SMAD4* mutatnt (*SMAD4*: c.343T>C). The primer-template duplex illustrates the primer position in reference to *SMAD4* exon 3.

#### 2.8.2.1 Dpn I digestion of the amplified products

The product was treated with Dpn I. The Dpn I endonuclease contains a target sequence: 5′-Gm6ATC-3′. This target is specific for hemi-methylated and methylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. 1 µL of DPN I (10 U/ µL) was added to each amplification reaction. Each reaction was then gently and thoroughly mixed by pipetting the solution up and down several times followed by a spin down for 1 minute. Each reaction was incubated at 37 °C for 1 hour to digest the parental (non-mutated) supercoiled dsDNA. The nicked vector DNA containing the desired mutations is then transformed into XL10-Gold ultracompetent cells.

#### 2.8.2.2 Bacterial Transformation

XL10-Gold ultracompetent cells were thawed on ice for 30 minutes. 45 µL of the bacterial stock was transferred into 4 Eppendorf tubes for each reaction. The 4 reactions were as follows: 2 tubes for pBabe-puro-Smad4-Flag Wild Type and Mutant (previously done by site-directed mutagenesis), and 2 tubes for pCMV5 DPC4-HA Wild type and Mutant plasmids. 2 µL of each Dpn I-treated plasmids were added to each tube and swirled for 2 minutes then left on ice 30-45 minutes. The cells were then heat shocked at 42 °C for 30 seconds and directly transferred to ice for 2 minutes. The reaction was then placed at room temperature for 5 minutes and later seeded on freshly prepared agar plates for 24 hours. 500 mL of 2% agar was prepared as follows: 5 grams of tryptone, 2.5 grams of yeast extract, 5 grams of NACL and 10 grams of Bactoagar. After cooling down the agar to 50-60 °C, 50 µL of 100mg/ml of ampicillin was added for selection.

### 2.8.3 Mutation detection

To select transformed cells, 5 µL of 100mg/ml ampicillin was added to each 5 mL of Luria-broth that was previously prepared. The broth was then inoculated with the 24-hour colonies of transformed DH5α cells by scraping one colony per plate and inserting it into the 15 mL conical tube with 5 mL LB and 5 µL ampicillin. The conical tubes were placed in the shaker-incubator at 37 °C for 8 hours. The plasmids were then collected from each tube by miniprep using Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit protocol mentioned in previous steps. To verify plasmid mutagenesis, plasmid concentration was first measured using Nanodrop. The concentration of each plasmid, (i. pBabe-puro-Smad4-Flag Wild Type ii. pBabe-puro-Smad4-Flag Mutant iii. pCMV5 DPC4-HA Wild type ix. pCMV5 DPC4-HA Mutant), was diluted to 10 ng/µL for sequencing.

The samples were Sanger sequenced using a pair of primers for mutation detection designed using Primer 3 software (URL: http://primer3.ut.ee/). Forward primer: 5' cggaaaggatttcctcatgt 3', Reverse Primer: 5' gtccctcaaagtcatgcaca 3'. The mutagenesis primers were used to amplify each of the 10 ng/µL plasmids using the following reaction set up: 0.5 µL Forward primer, 0.5 µL Reverse Primer, 1 µL of 10ng/µL plasmid template, 8 µL of GoTaq® Green Master Mix and 10 µL of Nuclease Free water. The reactions were placed on the thermocycler with the following conditions: 5 minutes at 95 °C, [20 seconds at 95°C, 20 seconds at 60°C, 20 seconds at 72°C] x25 cycles, 5 minutes at 72 °C, and Hold at 4°C. The PCR products were then cleaned by EPPiC Fast Reagent followed by dying the bases of our sequence using BigDye™ Terminator v1.1 Cycle Sequencing Kit, finally, purifying the product from unincorporated BigDye® terminators and salts by

using BigDye XTerminator™ Purification Kit. The final product was then placed into the sanger sequencer for capillary electrophoresis.

### 2.8.4 Plasmid Maxiprep

PureLink™ HiPure Plasmid Maxiprep Kit (Catalog no., K210007) was used to isolate transfection-grade plasmid DNA from XL10-Gold® Ultracompetent Cells. Transformed E. coli was first grown overnight on a shaker incubator in 5 mL of LB medium with 5 µL of 100 mg/ml ampicillin. Then, 5-10 µL (according to turbidity) of the liquid culture was inserted into 50 ml- conical tubes with 45 mL broth and 45 µL of 100mg/ml ampicillin. The tubes were placed in the shaker incubator at 37 °C overnight. Maxiprep was done as per manual instructions: First, RNase A was added to the resuspension Buffer (R3) and the bottle was stored at 4 °C. Then, 30 mL of Equilibration buffer (EQ1) was added to the HiPure Maxi Columns; the solution was allowed to drain through the column by gravity flow. The cells were harvested and sedimented by centrifuging the overnight LB culture at 4000 x g for 10 minutes and the supernatant was discarded. Next, the cell pellet was resuspended until homogenous using 10 mL of Resuspension buffer (R3). The cells were lysed using 10 mL of Lysis Buffer (L7), the mixture was mixed gently by inverting the capped tube 5 times. Then, 10 mL of Precipitation Buffer (N3) was added, the lysate was mixed immediately by inverting until the mixture until homogenous. The lysate was centrifuged at 12,000 x g for 10 minutes at room temperature. Next, the supernatant was loaded onto the equilibrated column with a pipette and the solution was allowed to drain by gravity flow. The column was washed using 60 mL Wash Buffer (W8); the flow-through was discarded. Then, a sterile 50-mL centrifuge tube was placed under the column and 15 mL of Elution Buffer (E4) was added to the column. The solution was

drained by gravity flow. 10.5 mL of Isopropanol was later added to the eluate tube, mixed well, and centrifuged at 12,000 x g for 30 minutes at 4 °C; the supernatant was discarded, and the DNA pellet was washed in 5 mL 70% Ethanol. The tube was then centrifuged at 12,000 x g for 5 minutes at 4 °C and the supernatant was removed. Finally, the pellet was air-dried for 10 minutes, then the purified plasmid DNA was resuspended in 200-500 μL (according to pellet size) of TE Buffer (TE). The concentration and purity of the purified plasmids were then measured on NanoDrop™ 2000/2000c Spectrophotometers. The sequence of the plasmids was also verified by SangerSeq in the same manner mentioned above. Because of the presence of a mammalian selective marker in pCMV5 DPC4-HA plasmid and the four copies of the Smad binding site in the SBE4-LUC Luciferase, they were preferred to be utilized in following cell transfection experiment.

### 2.8.5 Cell Culture

HEK293T cell line is used in this assay for a combination of reasons including high transfection efficiency, stability, adaptability to growth conditions, ease of culture, comes from human origin, and versatility.

#### 2.8.5.1 Media Preparation

The cell culture media RPMI 1640 Medium with L-Glutamine (500 mL) (Catalog no., 11875093) was prepared by adding 10% Fetal Bovine Serum (50 mL), 1% Streptomycin (5 mL) and 1% L-glutamine (5 mL). This media was used to culture HEK 293T cells, which are daughter cell lines derived from HEK293 original cell line that is transfected with a plasmid vector carrying SV40 origin of replication.

#### 2.8.5.2 Cell Thawing

HEK293T Cells were removed from liquid nitrogen and directly placed in water bath at 37 °C. 5 mL of media was placed in a conical tube; the cells were thawed by mixing with previously prepared RPMI media. The media was transferred to a conical tube. Next, the cells were centrifuged at 1600 rpm for 10 minutes. The supernatant was then aspirated using the suction device to get rid of the 10% DMSO. Cells were resuspended with 2 mL fresh media and transferred to a culture plate with 6 mL media. The plate was placed in a CO2 incubator with the following conditions: 37 ℃ and 5% CO2.

#### 2.8.5.3 Cell Passaging

To split HEK293T cells, the media was aspirated from the plate. This was followed by adding 1 mL of Trypsin, the plate was then placed in the 5% CO2 incubator at 37 ℃ for 2 minutes. The plate was then removed and tapped to remove any attached cells. Next, the appropriate amount of media was added according to the needed concentration of cells. The volume prepared was then split into two plates, and media was added to the plates to have a final total volume of 8 mL in each plate. The plates were finally placed in 37 ℃, 5% CO2 incubator.

#### 2.8.5.4 Cell Cryopreservation

Freezing media was prepared by adding 10% DMSO and 20% Fetal Bovine Serum to RPMI 1640 growth media. Media was first aspirated from the plate, followed by trypsinization of the HEK293T cells. 8 mL of fresh media was added to the trypsinized cells, and then transferred into a 15-mL conical tube to be centrifuged for 10 minutes at 1,000 rpm at room temperature. Next, the supernatant was aspirated, and the freezing

media was used to resuspend the cell pellet. Finally, the cryotubes were preserved in -80 ℃, freezers.

#### 2.8.5.5 Mycoplasma Detection

EZ-PCR Mycoplasma Detection Kit (Biological Industries, Catalog no., 20-700-20) was used to detect the presence of mycoplasma in the passaged cells before transfection. The test sample was first prepared by transferring 1.0 mL of cell culture supernatant into a 1.5 mL sterile centrifuge tube. The cellular debris was then pelleted by centrifuging the sample at 250 g briefly. Next, the supernatant was centrifuged at 20,000 g for 10 minutes to sediment mycoplasma (if present). The supernatant was later decanted, and the pellet was re-suspended with 50 µL of the Buffer Solution supplied with the kit. The test sample and negative control preparation were done per protocol as follows: 5 µL Internal control primer mix, 1 µL internal control DNA template, 10 µL Reaction mix, 29 µL Nuclease free water, and 5 µL of test sample or Negative control (Buffer solution). The positive control was prepared differently as per the manual's instructions: 5 µL internal control primer mix, 1 µL internal control DNA template, 10 µL Reaction mix, 33 µL Nuclease free water, and 1 µL positive control DNA.

The reactions prepared were placed on the thermal cycler with the following conditions: 30 seconds at 94 °C, [30 seconds at 94°C, 2 minutes at 60°C, 1 minute at 72°C] x 35 cycles, 30 seconds at 94 °C, 2 minutes at 60°C, 5 minutes at 72 °C, hold at 4°C.

The products were then analyzed by gel electrophoresis; 2% agarose was prepared by adding 2 grams of agarose to 100 mL of TAE buffer. 20 µL of each of the PCR products (Test sample, negative control, and positive control) mixed with loading buffer, was loaded into the gel for electrophoresis.

#### 2.8.5.6 Cell Seeding

The incubated HEK293T cells prepared previously were cultured to be 70-80% confluent. The media was then removed by using the suction device. 1 mL of Trypsin was added to the cells, the plate was incubated and gently tapped to remove all attached cells. 4 mL of warm media was added to the cells, subsequently, 1.2 mL of the cells were transferred to a falcon tube and 24 mL of warm media was added. 1 mL of the final concentration was transferred to two 12-well plates. The plates were incubated for 48 hours to reach the targeted cell confluence before transfection.

#### 2.8.5.7 Cell Transfection

HEK 293 T cells were transfected with 0.4 µg of Empty vector plasmid (plasmid backbone), pCMV5 DPC4-HA *SMAD4* Wild type or pCMV5 DPC4-HA *SMAD4* Mutant plasmid. The plasmids were co-transfected with 200 ng of the Luciferase reporter, SBE4 – LUC and 20 ng of Renilla luciferase plasmid. 3 µl of TransIT-LT1 Reagent (MIRUS) per 1 µg of DNA volumes needed were all mixed and incubated for 5 minutes before distributing to the wells containing the plasmid and reporter. This experiment was first calibrated using different plasmid concentrations and 2 luciferase reporters shown in Figure 5 (A) and (B), Figure 5 (C) shows the actual experiment ran to obtain the results.

A.

| | WELL 1 | WELL 2 | WELL 3 | WELL 4 | WELL 5 | WELL 6 | WELL 7 | WELL 8 | WELL 9 |
|---|---|---|---|---|---|---|---|---|---|
| Empty Vector (plasmid backbone) | 0.4 µg | *** | *** | *** | *** | *** | *** | *** | *** |
| SMAD4 WT | *** | 0.1 µg | 0.2 µg | 0.4 µg | 1.0 µg | *** | *** | *** | *** |
| SMAD4 MUT | *** | *** | *** | *** | *** | 0.1 µg | 0.2 µg | 0.4 µg | 1.0 µg |
| SBE4-Luciferase Reporter | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng |
| Serum-Free Media | 60 µL | 30 µL | 40 µL | 60 µL | 120 µL | 30 µL | 40 µL | 60 µL | 120 µL |
| MIRUS Transfection Reagent | 1.2 µL | 0.6 µL | 0.8 µL | 1.2 µL | 2.4 µL | 0.6 µL | 0.8 µL | 1.2 µL | 2.4 µL |

B.

| | WELL 1 | WELL 2 | WELL 3 | WELL 4 | WELL 5 | WELL 6 | WELL 7 | WELL 8 | WELL 9 |
|---|---|---|---|---|---|---|---|---|---|
| Empty Vector (plasmid backbone) | 0.4 µg | *** | *** | *** | *** | *** | *** | *** | *** |
| SMAD4 WT | *** | 0.1 µg | 0.2 µg | 0.4 µg | 1.0 µg | *** | *** | *** | *** |
| SMAD4 MUT | *** | *** | *** | *** | *** | 0.1 µg | 0.2 µg | 0.4 µg | 1.0 µg |
| TRS6E-Luciferase Reporter | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng |
| Serum-Free Media | 60 µL | 30 µL | 40 µL | 60 µL | 120 µL | 30 µL | 40 µL | 60 µL | 120 µL |
| MIRUS Transfection Reagent | 1.2 µL | 0.6 µL | 0.8 µL | 1.2 µL | 2.4 µL | 0.6 µL | 0.8 µL | 1.2 µL | 2.4 µL |

C.

| | WELL 1 | WELL 2 | WELL 3 | WELL 4 | WELL 5 | WELL 6 |
|---|---|---|---|---|---|---|
| Empty Vector (plasmid backbone) | 0.4 µg | 0.4 µg | *** | *** | *** | *** |
| SMAD4 WT | *** | *** | 0.4 µg | 0.4 µg | *** | *** |
| SMAD4 MUT | *** | *** | *** | *** | 0.4 µg | 0.4 µg |
| SBE4-Luciferase Reporter | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng | 200 ng |
| Serum-Free Media | 60 µL | 60 µL | 60 µL | 60 µL | 60 µL | 60 µL |
| MIRUS Transfection Reagent | 1.2 µL | 1.2 µL | 1.2 µL | 1.2 µL | 1.2 µL | 1.2 µL |

Figure 5: Luciferase assay run calibration and plate arrangement. (A) and (B) are calibration plates using different luciferase reporters SBE4-LUC (A) and TRS6E-LUC (B) and different concentrations of *SMAD4* WT and MUT concentrations. After calibration, SBE4-LUC reporter gave the best results using 0.4 ug of WT and MUT *SMAD4* plasmids; the assay was then duplicated for validation as shown in (C)

### 2.8.6 Measuring SMAD4 transcriptional activity

#### 2.8.6.1 Preparing Lysates

Dual-Luciferase® Reporter Assay (Promega, Catalog no. E1910) was used in this experiment. Growth media was removed from cultured cells. 1X Lysis reagents was prepare by adding 4 volumes of water to 1 volume of 5X lysis reagent. Next, 250 µL of the PLB (Passive Lysis Buffer) prepared was dispensed in each well of the 19 wells prepared. The plate was gently rocked for 15 minutes at room temperature. The lysates were then transferred into Eppendorf tubes, which were centrifuged at 12,000 rpm for 5

minutes to remove cell debris. The supernatant was transferred to new Eppendorf tubes and stored at –20 °C until ready to measure.

#### 2.8.6.2 Measuring Lysates' Activity

The Luciferase Assay Reagent (LAR) was prepared by adding Luciferase Assay Buffer (10ml) to the vial of lyophilized Luciferase Assay Substrate; the LARII was then distributed into aliquots and stored at – 70 °C. To the required amount of Stop & Glo® Buffer (3.3 mL), 50X Stop & Glo® Substrate (66 µL) was added to a final 1X concentration.

The two injectors of the luminometer were set to dispense 100µl of LAR II and Stop & Glo® Reagent, respectively to 20 µl of the lysates prepared previously. For measurements, a 1- to 2-second delay was used and a 5- to 10-second read time.

# Chapter Three: Results

This study was conducted on a colorectal cancer patient in Palestine. The application of advanced molecular techniques, including next-generation sequencing (NGS), Sanger sequencing was used to investigate the genetic profile of the patient and luciferase assay was used to evaluate the functional consequences of identified mutations. Using NGS, a novel mutation was discovered in the gene *SMAD4*, a key component of the TGF-β signaling pathway known for its involvement in colorectal cancer. Sanger sequencing confirmed that the identified mutation, is somatic. The luciferase assay measured the transcriptional activity of *SMAD4* and revealed that the identified mutation in *SMAD4* resulted in aberrant transcriptional activity and disrupted the normal signaling cascade. These functional alterations suggested the pathogenic nature of the mutation, indicating its potential contribution to the development and progression of colorectal cancer in the patient.

## 3.1 Next-Generation Sequencing and the Identification of Candidate Mutations

The patient's sample was sequenced using NGS to identify the underlying mutations of colorectal cancer. In order to obtain the sample's data for analysis, Fasta Q files were generated from the sequencer. the CLC Genomic Workbench toolkit by QIAgen was used to convert the Fasta Q to VCS files and was compared against Human reference Genome 19 (HG19), also known as the genome reference consortium human build 37 (GRCh37). Then, Qci software was used for variant filtering and interpretation based on clinical phenotype. However, the known pathogenic variants were negative for this panel. The raw data was then re-analyzed to show all variants with unknown significance (VUS)

before using in-silico tools to predict their pathogenicity. The table below illustrates all VUS detected:

Table 1: AIT panel results of Variant of unknown significance. This table lists all VUS generated by the NGS

| *Mutation* | *c. variant* |
|---|---|
| ***EGFR*, p. Arg521Lys** | c.1562G>A |
| ***NOTCH1*, p. Asp1533Gly** | c.4598A>G |
| ***PIK3CA*, p. Ile391Met** | c.1173A>G |
| ***EGFR*, p. Leu619Pro** | c.1856_1857delTGinsCA |
| ***GNAQ*, p. Gln197Pro** | c.590A>C |
| ***KIT*, p. Met541Leu** | c.1621A>C |
| ***FGFR2*, p. Phe276Leu** | c.828T>G |
| ***MAP2K1*, p. Pro265Ser** | c.793C>T |
| ***ALK*, p. Arg214His** | c.641G>A |
| ***ALK*, p. Phe174Leu** | c.522C>A |
| ***ALK*, p. Thr680Ile** | c.2039C>T |
| ***ERBB2*, p. Asp1252His** | c.3754G>C |
| ***ERBB2*, p. Gly292Ser** | c.874G>A |
| ***ERBB2*, p. Thr875Pro** | c.2623A>C |
| ***ERBB4*, p. Arg81Gln** | c.242G>A |
| ***ERBB4*, p. Arg95Cys** | c.283C>T |
| ***FBXW7*, p. Gly477Arg** | c.1429G>C |
| ***FGFR2*, p. Met538Ile** | c.1614G>T |
| ***FGFR3*, p. Gly692Cys** | c.2073_2074delGGinsCT |
| ***FGFR3*, p. Ser679Phe** | c.2036C>T |
| ***GNA11*, p. Leu170Phe** | c.510G>C |
| ***KRAS*, p. Gly77Ser** | c.229G>A |
| ***MET*, p. Asp153Asn** | c.457G>A |
| ***MET*, p. Asp372Glu** | c.1116C>A |
| ***NOTCH1*, p. Ala1610Thr** | c.4828G>A |
| ***NOTCH1*, p. Val1578del** | c.4732_4734delGTG |
| ***PIK3CA*, p. Asn107Thr** | c.320A>C |
| ***PIK3CA*, p. Leu455Val** | c.1363T>G |
| ***SMAD4*, p. Cys115Arg** | c.343T>C |
| ***SMAD4*, p. His92Gln** | c.276T>A |
| ***SMAD4*, p. Leu536Arg** | c.1607T>G |
| ***ALK*, p. Arg209Cys** | c.625C>T |

### 3.2 In silico Analysis

Before performing functional investigations, in silico prediction software enables evaluating the impact of amino acid alterations on the structure or functionality of a protein. The following In Silico Tools were used to determine the effects of 12 of the previously mentioned variants in Table 1: SIFT(https://sift.bii.a-star.edu.sg/), PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), Align GVGD (http://agvgd.hci.utah.edu/agvgd_input.php) and Provean https://www.jcvi.org/research/provean#downloads).

Table 2: In Silico analysis results for specified VUS with various analysis tools. Results show that SMAD4, p. Cys115Arg is more likely to have damaging effects on protein function.

| Mutations | Sift | Polyphen | Provean | Align-GVGD | Conservation | A.A change |
|---|---|---|---|---|---|---|
| *NOTCH1,* p.Asp1533Gly | N/A | 0.895, damaging | (-)6.218, deleterious | 93.77, Class C65, highly likely to affect protein function | conserved | acidic to nonpolar |
| *EGFR,* p.Leu619Pro | 0.98, tolerated | 0.001, benign | 1.605, neutral | 97.78, Class C65, highly likely to affect protein function | conserved | nonpolar to nonpolar |
| *MAP2K1,* p.Pro265Ser | 0.08, tolerated | 0.242, benign | (-)3.499, deleterious | 73.35, Class C65, highly likely to affect protein function | conserved | nonpolar to polar |
| *ALK,* p.Phe174Leu | 0.00, likely to affect protein function | 0.115, benign | (-)0.919, neutral | 21.82, Class C15, less likely to affect protein function | conserved | nonpolar to nonpolar |
| *ERBB2,* p.Asp1252His | 0.10, tolerated | 0.999, damaging | (-)1.056, neutral | 81.24 Class C65, highly likely to affect protein function | conserved | acidic to basic |

| Mutations | Sift | Polyphen | Provean | Align-GVGD | Conservation | A.A change |
|---|---|---|---|---|---|---|
| *FBXW7,* p.Gly477Arg | 0.00, likely to affect protein function | 1.00, damaging | (-)7.943, deleterious | 125.13, Class C55, highly likely to affect protein function | conserved | nonpolar to basic |
| *FGFR2,* p.Met538Ile | 0.00, likely to affect protein function | 1.00, damaging | (-)3.432, deleterious | 10.12, Class 0, less likely to affect protein function | conserved | nonpolar to nonpolar |
| *GNA11,* p.Leu170Phe | 0.00, likely to affect protein function | 0.012, benign | (-)0.881, neutral | 21.82, Class C15, less likely to affect protein function | not conserved | nonpolar to nonpolar |
| *MET,* p.Asp153Asn | 0.03, likely to affect protein function | 0.961, damaging | (-)0.901, neutral | 23.01, Class C15, less likely to affect protein function | conserved | acidic to polar |
| *NOTCH1,* p.Ala1610Thr | N/A | 0.073, bengin | (-)1.171, neutral | 58.02, Class C55, highly likely to affect protein function | conserved | nonpolar to polar |
| *PIK3CA,* p.Asn107Thr | 0.08, tolerated | 0.991, damaging | (-)2.570, deleterious | 64.77, Class C55, highly likely to affect protein function | conserved | polar to polar |
| ***SMAD4,* p.Cys115Arg** | **0.00, likely to affect protein function** | **1.00, damaging** | **(-) 10.850, deleterious** | **179.53 Class C65, highly likely to affect protein function** | **conserved** | **polar to basic** |

Based on the preliminary in silico results, (*SMAD4*, p. Cys115Arg) is most likely a deleterious mutation that is highly likely to affect protein function. Therefore, further in-silico analysis was done by visualization of the mutation and its effect on the protein 3D structure using RCSB Protein Data Bank (RCSB PDB) (https://www.rcsb.org/) (Figure 6)

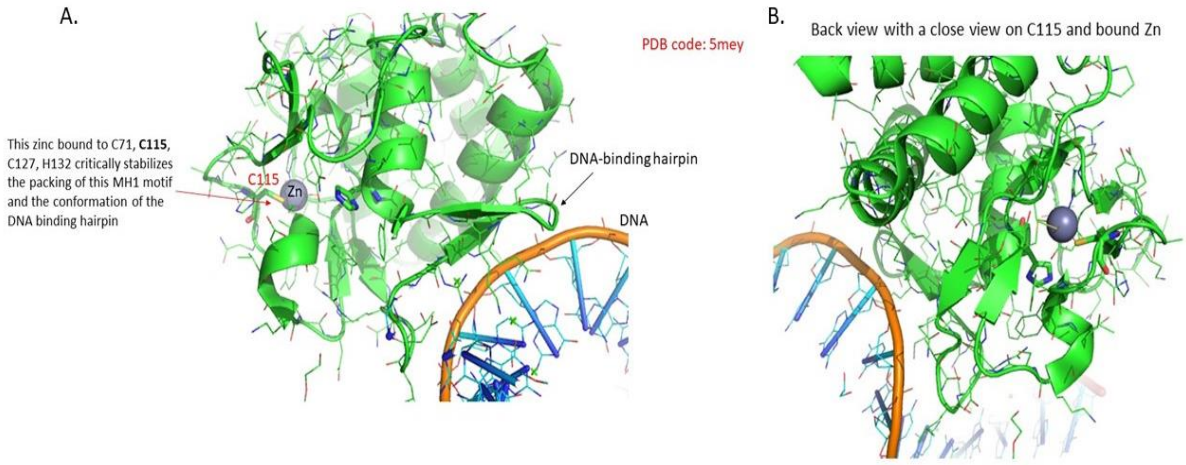



Figure 6: Predicted 3D structure of SMAD4 protein (PDB:5mey). Visual comparison (front (A) and back (B)) to show the placement of the C115 residue and its importance in stabilizing the MH1 motif by forming the zinc finger.

SMAD4 amino acid Cysteine position 115 is one of 4 residues that bind Zinc (C71, C115, C127, H132) and form a zinc-finger that stabilizes the conformation of the DNA binding arm (Figure 6). Therefore, (*SMAD4*, p. Cys115Arg) change might hinder the binding of zinc ion (Zn2+) and thus will not be able to bind DNA. Consequently, SMAD4 protein is predicted to be functionally dead. Since SMAD4, p. (18-142 MH1) is required for DNA binding, the mutation may severely inhibit DNA binding.

### 3.3 Sanger Sequencing of peripheral blood DNA

To study whether the (*SMAD4*, p. Cys115Arg) mutation identified in our proband is a germ line or somatic mutation, extracted DNA from peripheral blood of the colorectal cancer patient and the patient's mother were sequenced and the results were analyzed by Finch TV (Version 1.5.0). The electrographs for both patient and mother (Figure 7) confirmed that (*SMAD4*: c.343T>C, p. Cys115Arg) is a somatic mutation.

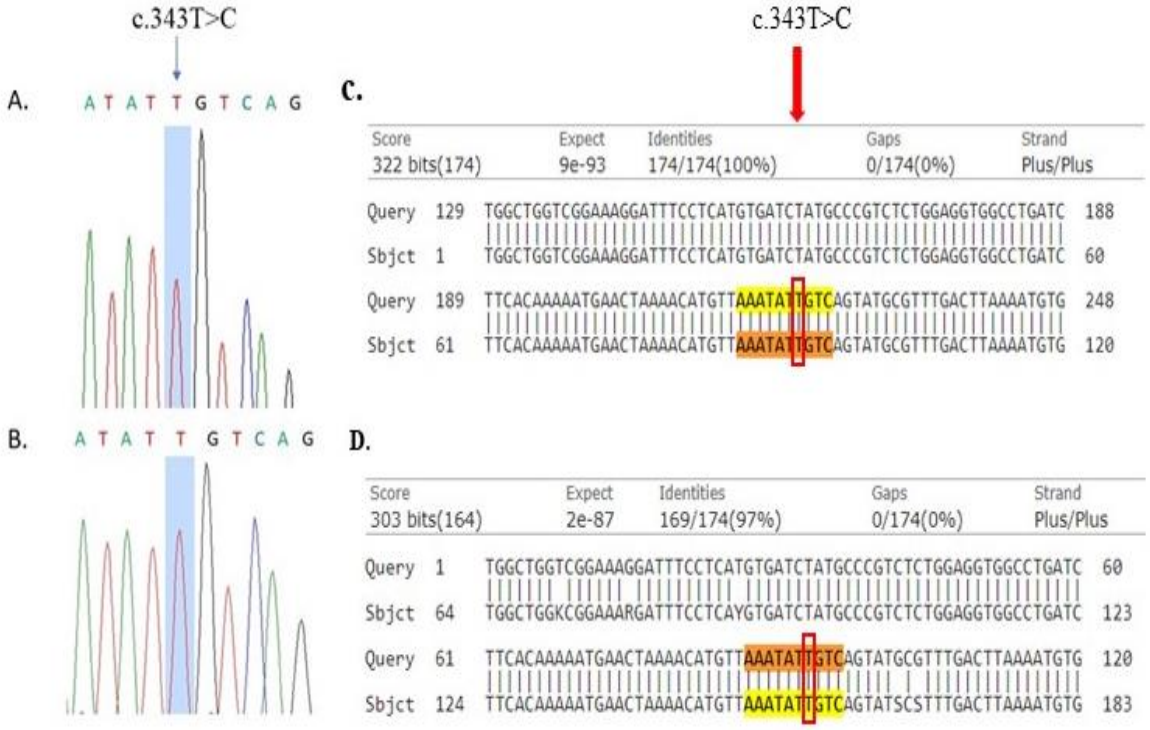



Figure 7: Segregation and alignment of SMAD4 mutation in both the patient (A) and his mother (B). The electrographs for both the patient and the mother show negative results for the mutation (*SMAD4*: c.343 T>C, p. Cys115Arg). The blue highlighted dNTP T shown in both electrographs confirms that the mutation is not familial and is somatic. Alignment of the FASTA sequence of the patient (C) (subject) and the mother (D) (subject) against the normal FASTA sequence of *SMAD4* exon3 obtained from NCBI (query) Both alignments confirm that both the patient and the mother do not harbor the mutation (*SMAD4*: c.343 T>C, p. Cys115Arg)

### 3.4 *SMAD4*, p. Cys115Arg mutation and its effect on SMAD4 transcriptional activity

To assess the effect of *SMAD4* mutation on the transcriptional activity of the *SMAD4* gene, luciferase assay was used. The results showed that the tested mutation (*SMAD4*, p. Cys115Arg) decreased *SMAD4* transcriptional activity by four-fold when compared to

the wild-type *SMAD4*. These results were normalized and compared to an empty vector plasmid. Furthermore, the overexpression of the *SMAD4* wild type was found to increase SMAD4 transcriptional activity. Overall, these findings suggest that the identified *SMAD4* mutation has a negative impact on the transcriptional activity of the *SMAD4* gene. (Figure 8).

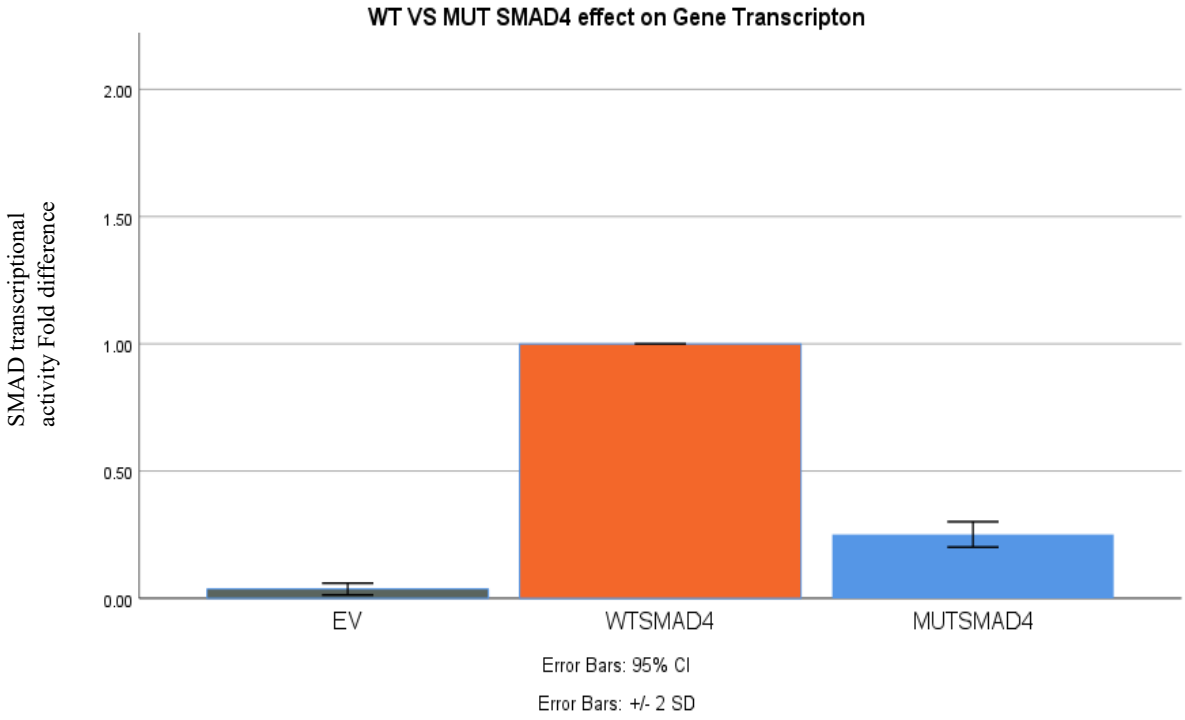



Figure 8: WT vs MUT *SMAD4* effect on transcriptional activity of SMAD4. The empty vector (plasmid backbone) gives a near-zero reading (0.037). The over-expressed *SMAD4* wild type (in orange) shows increased transcriptional activity of the *SMAD4* gene in comparison to the MUT (*SMAD4*, p. Cys115Arg) (in blue). The 4-fold difference (p-value <0.001) between wild-type and mutant indicates that the *SMAD4* mutation studied affects SMAD4 transcriptional activity

# Chapter Four: Discussion

Our study investigated clinical manifestations and genetic makeup of colorectal cancer in a Palestinian CRC. Bioinformatics analysis of WES data revealed variants of unknown significance that have deleterious effects on the protein structure and function of various genes. Upon further investigation, (*SMAD4*, p. Cys115Arg) was selected as the candidate mutation for severe pathogenic effects on CRC; the mutation was found to downregulate the transcriptional activity of the *SMAD4* gene, indicating its potential role in the disruption of normal gene function.

In this study, advanced bioinformatics tools such as cBioPortal and COSMIC [(https://www.cbioportal.org/) - (https://cancer.sanger.ac.uk/cosmic)] were first utilized to analyze mutations across various tumor types. Both platforms for mutation distribution across cancer types in Figure 9 and Figure 10 illustrate that *SMAD4* mutations are prevalent mainly in pancreatic and colorectal cancer, which further encouraged us to proceed in the functional analysis of the (*SMAD4*, p. Cys115Arg) mutation. Additionally, we employed ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) to identify the specific conditions associated with the mutation identified in our analysis. It was found that *SMAD4* mutations are predominantly causing Juvenile polyposis syndrome (Table 3). Juvenile polyposis (JP) is an autosomal-dominant condition that is linked to colorectal cancer and is characterized by the growth of gastrointestinal polyps. For patients with JP, the overall lifetime risk for colorectal cancer was 38.7%. (Brosens et al., 2007)

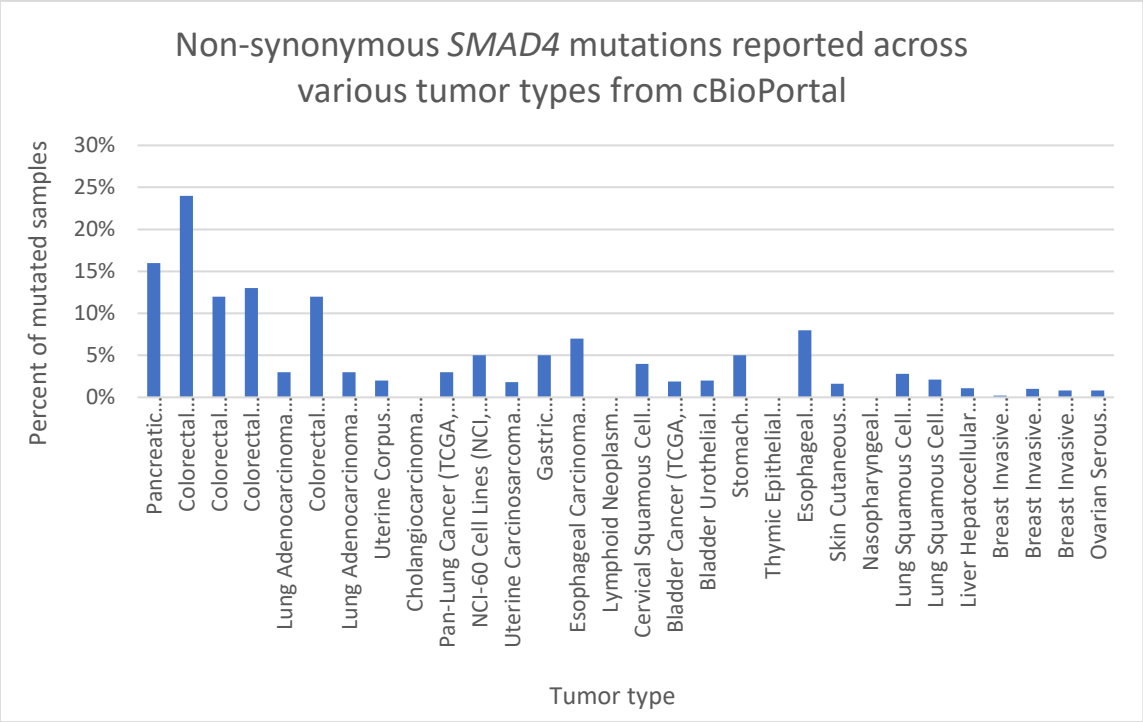



Figure 9: Non-synonymous *SMAD4* mutations reported across various tumor types from cBioPortal. The chart illustrates that *SMAD4* mutations are mostly present in patients with colorectal cancer

.

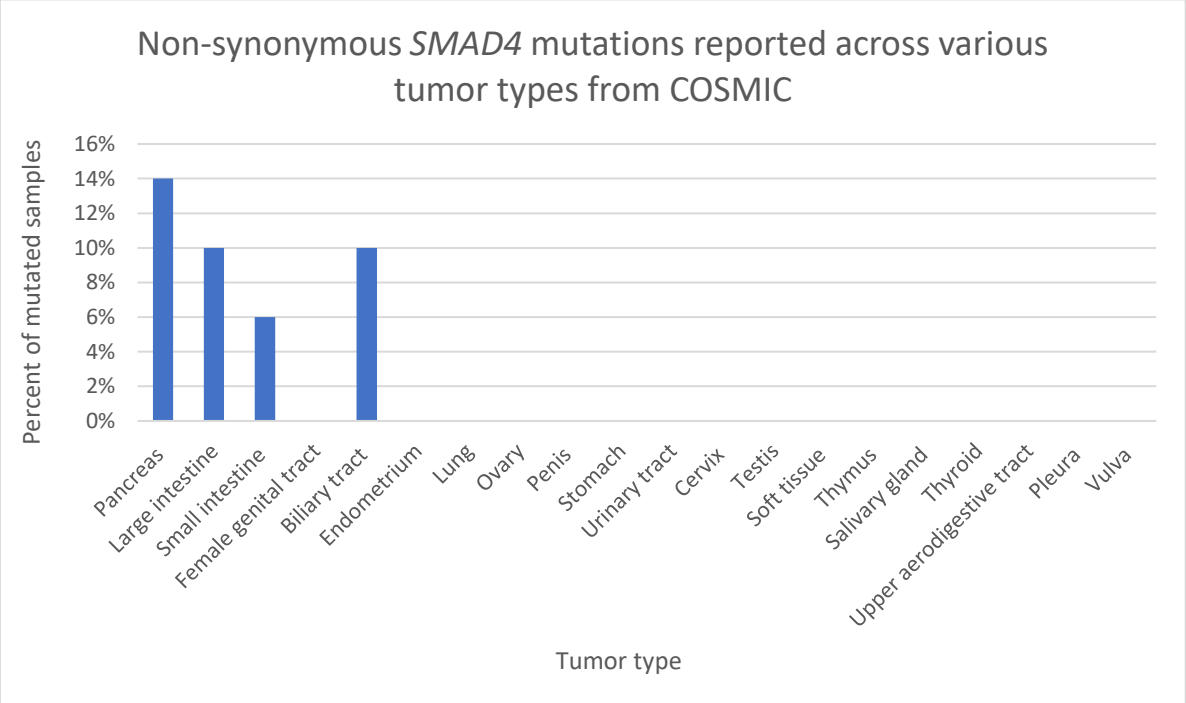



Figure 10: Non-synonymous *SMAD4* mutations reported across various tumor types from COSMIC. *SMAD4* mutations are distributed in several tumor types, mainly in pancreas and the large intestine

Table 3: ClinVar results for conditions related to *SMAD4* mutations. *SMAD4* mutations are predominant in both, Juvenile polyposis syndrome and Hereditary cancer-predisposing syndrome.

| **Conditions** | **Associated *SMAD4* mutations** |
|---|---|
| *Juvenile polyposis syndrome* | M24fs, D52fs, T72fs, Q83*, W101*, H111*, S125fs, V128fs, Y133*, R135*, S144fs, L146*, L148fs, S154*, F166fs, Q169*, S178*, Q180*, Q183*, Y195*, P203fs, S232fs, G243fs, Y276*, W302*, F310fs, Q311*, I314fs, C324fs, Y328fs, P342fs, P346fs, V348fs, G352R, G352E, Y353S, S357fs, P356fs, R361S, R361C, R361L, R361H, C363del, C363R, Q366*, S368fs, |
| *Hereditary cancer-predisposing syndrome* | D52fs, H92fs, W99*, W101*, Y133*, S138*, P139fs, S144fs, S154*, Q169*, S178*, S232fs, P296fs, W302fs, W302*, Q311*, I314fs, P342fs, S343*, G352R, G352E, R361C, R361L, R361H, C363del, Q366*, S368fs |

To understand the structural implications of the mutation, a 3D protein model was employed. The model revealed that the residue Cys115 was one of four residues involved in binding zinc (Cys71, Cys115, Cys127, and His132) to form a zinc-finger motif that stabilizes the DNA binding arm conformation. It was found that this alteration renders the protein functionally inactive and severely inhibits its ability to bind DNA, as SMAD4's MH1 domain (amino acids 18-142) is crucial for DNA binding. The MAD-homology 1 (MH1) domain is a highly conserved domain of SMAD4 and plays a crucial role in DNA binding and transcriptional regulation (Ruiz et al., 2021). This domain is responsible for the interaction with specific DNA sequences, enabling SMAD4 to modulate gene expression. Several studies have demonstrated the significance of the

MH1 domain in DNA binding and transcriptional activity (Fleming et al., 2013; Shi et al., 1998). Therefore, a crucial step in deciding which set of genes will be activated in response to a TGF- stimulation is SMAD4 recruitment to DNA. The MH1 domain is utilized by R-Smads and Co-Smads for DNA-binding (Kim et al., 1997; Shi et al., 1998). DNA binds to the MH1 region only in specific sequences. SMAD4's capacity to bind DNA was significantly diminished by mutations in the region from L43 to R135. It was discovered that the entire MH1 domain is extremely vulnerable to structural alterations and that the mutations in the L43–R135 area would certainly target DNA-binding.(Jones & Kern, 2000).

Mutations in the MH1 domain of SMAD4 can have profound effects on DNA binding and gene transcription. The impact of MH1 domain mutations on SMAD4 function was explored. The study showed that specific mutations within the MH1 domain resulted in impaired DNA binding affinity, compromising the ability of SMAD4 to interact with its target DNA sequences (Kuang & Chen, 2004) . Similarly, another study investigated the functional consequences of MH1 domain mutations in SMAD4 associated with colorectal and pancreatic cancer. They observed that these mutations disrupted the DNA binding ability of SMAD4, leading to the truncation of gene transcription and dysregulation of downstream target genes (J. Xu & Attisano, 2000). These findings emphasize the critical role of the MH1 domain in SMAD4-mediated DNA binding and transcriptional control. Primary cancer lesions in colorectal tumors contain a range of mutations, including nonsense, frameshift, missense, insertion, and deletion; and most commonly, missense mutations, which typically have adverse outcomes (Figure 11) (Fleming et al., 2013).

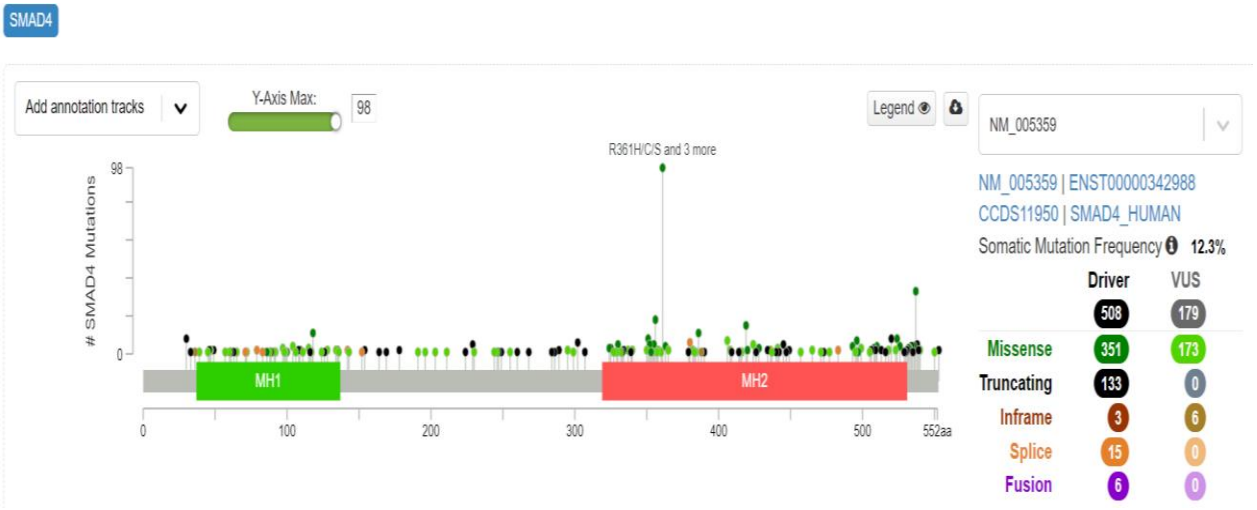



Figure 11: Mutations distributed across SMAD4 domains. Using all cBioportal's colorectal adenocarcinoma databases, *SMAD4* gene mutations are shown across the gene's domains. The figure illustrates that the bulk of the mutations are missense mutations and concentrated on the MH1 and MH2 domains.

Furthermore, studies have provided insights into the structural basis of the MH1 domain's contribution to DNA binding and transcriptional regulation. A structural analysis revealed the presence of a DNA-binding loop within the MH1 domain that directly interacts with DNA. This loop, along with other critical residues in the MH1 domain, forms a highly specific interface for DNA recognition (Massagué et al., 2005). Moreover, the study by (Chacko et al., 2001) demonstrated that mutations in the MH1 domain DNA-binding loop can impair the ability of SMAD4 to bind DNA effectively.

Functional assays play a crucial role in validating in silico findings and providing experimental evidence to support computational predictions. In the field of molecular biology, one commonly used functional assay is the luciferase assay, which allows for the evaluation of the transcriptional activity of a gene. This assay measures the expression level of a luciferase reporter gene driven by a promoter region of interest, reflecting the transcriptional activity of the target gene. To investigate the impact of a novel mutation (*SMAD4*, p. Cys115Arg), we performed mutagenesis of the plasmid carrying the *SMAD4*

gene and transfected both the wild-type and mutant versions of the plasmid into HEK293T cells. By employing the luciferase assay, we were able to detect and compare the transcriptional activity in the transfected cells. The results of the luciferase assay aligned with the in-silico findings and 3D structural analysis. The mutant *SMAD4* carrying the novel mutation exhibited significantly reduced transcriptional activity compared to the wild-type *SMAD4*. Specifically, the mutant variant demonstrated approximately 4-fold lower activity than the wild-type variant, as supported by a significant p-value of <0.001These findings further confirmed the deleterious effects of the identified mutation on the transcriptional activity of the *SMAD4* gene.

Missense mutations in *SMAD4* have been shown to have significant effects on the transcriptional activity of the gene. Several studies have explored the functional consequences of these missense mutations and their impact on SMAD4-mediated gene expression. For instance, (Blackford et al., 2009) investigated the effects of missense mutations in *SMAD4* associated with juvenile polyposis syndrome. They found that these mutations disrupted the binding affinity of SMAD4 to its DNA target sequences, resulting in reduced transcriptional activity. This study highlights the detrimental impact of missense mutations on SMAD4-mediated gene transcription. The main function of SMAD4 in the TGF-/Smad signaling pathway is to prevent cellular proliferation. Reduced TGF-/Smad signaling pathway activity can result from *SMAD4* mutations (Xiao et al., 2003).

Furthermore, studies by (Hahn et al., 1996; Itoh et al., 2000) examined the effects of missense mutations in *SMAD4* in pancreatic cancer and colorectal. They observed that these mutations impaired the interaction of SMAD4 with its co-activator proteins, leading to a decrease in the transcriptional activity of SMAD4 target genes. Similarly, (Lanauze

et al., 2021; Zhao et al., 2018) investigated the functional consequences of missense mutations in *SMAD4* associated with colorectal cancer. They reported that these mutations disrupted the binding of SMAD4 to DNA, resulting in reduced gene transcriptional activity and dysregulation of downstream target genes.

Moreover, research by (Lin et al., 2019) examined the effects of missense mutations in *SMAD4* in the context of head and neck squamous cell carcinoma. They demonstrated that these mutations in *SMAD4* impair its ability to interact with transcriptional co-factors and result in reduced transcriptional activity. Additionally, a study by (Schutte, 1999) investigated the impact of missense mutations in *SMAD4* gene on SMAD4 protein's interaction with other SMAD proteins. They found that these mutations disrupted the formation of SMAD4-SMAD2/3 complexes, leading to impaired transcriptional activity.

Colon cancer is a prevalent malignancy worldwide, and its treatment often involves a combination of chemotherapy and surgery. One important consideration in determining the effectiveness of chemotherapy regimens is the presence of specific genomic alterations in patients, such as *SMAD4* mutations. The presence of *SMAD4* mutations in colon cancer patients has been associated with poorer prognosis and resistance to certain chemotherapy agents. One of the most widely used medications for the treatment of metastatic colon cancer is 5-fluorouracil (5-FU), which is the primary component of practically all chemotherapy regimens. It is frequently used as FOLFOX (5-fluorouracil, leucovorin [folinic acid], oxaliplatin) as the conventional first- and second-line chemotherapeutic treatments available to combat metastatic colon cancer and improve patient survival (Douillard et al., 2000; Neugut et al., 2019).

In this study, our patient with a *SMAD4* mutation undergoing FOLFOX treatment with 5-fluorouracil (5-FU) may face challenges related to drug resistance. Studies have indicated

a potential association between *SMAD4* loss and resistance to 5-FU-based chemotherapy regimens in various cancer types. For instance, research by (Alhopuro et al., 2005; B. Zhang et al., 2014; Binhao Zhang et al., 2016) demonstrated that *SMAD4* loss induce resistance to 5-FU treatment in colorectal cancer patients through activation of Akt pathway. This suggests that the presence of a *SMAD4* mutation in the patient raises concerns about potential resistance to 5-FU therapy. Therefore, personalized treatment approaches considering the patient's genetic profile and alternative treatment options may be necessary to overcome potential drug resistance and improve treatment outcomes.

In summary, missense mutations in *SMAD4* have been shown to significantly affect and reduce the transcriptional activity of the gene. These mutations disrupt the binding of SMAD4 to DNA, impair its interaction with co-activators and SMAD proteins, and alter its protein conformation. Understanding the functional consequences of missense mutations in *SMAD4* provides valuable insights into the molecular mechanisms underlying their impact on *SMAD4* transcriptional activity and may have implications for the development of targeted therapies.

# Chapter Five: Conclusion

In conclusion, the identification and characterization of novel mutations in the *SMAD4* gene through next-generation sequencing have proven to be crucial in understanding the genetic landscape of colorectal cancer, both globally and within specific populations like the Palestinian community. The significance of the *SMAD4* gene in colorectal cancer has been well-established, as it plays a critical role in regulating key signaling pathways involved in cell growth, differentiation, and apoptosis. By elucidating the functional consequences of *SMAD4* mutations, we have gained insights into the mechanisms by which these alterations contribute to the development and progression of colorectal cancer.

The advent of next-generation sequencing technologies has revolutionized our ability to screen and identify genetic mutations on a large scale. This technology makes it possible to find unique mutations that were previously hard to find using conventional sequencing techniques. In the context of colorectal cancer and *SMAD4*, next-generation sequencing has provided a valuable tool for uncovering new variants and their functional implications. This knowledge enhances our understanding of colorectal cancer biology.

Our knowledge of the pathophysiology of colorectal cancer may be greatly expanded by future research concentrating on the functional effects of new mutations in *SMAD4*. Examining how these mutations affect *SMAD4's* transcriptional activity, DNA binding, and interactions with other signaling molecules will shed light on the fundamental processes that underlie the onset of disease. Additionally, studying the prevalence and clinical significance of specific *SMAD4* mutations in different populations, including the

Palestinian community, can contribute to the development of targeted screening strategies and personalized treatment approaches.

The discovery of a novel mutation, such as the (*SMAD4*, p. Cys115Arg) mutation identified in our study, carries significant importance. It not only adds to the growing body of knowledge regarding *SMAD4* mutations in colorectal cancer but also underscores the importance of considering rare and unique variants in diverse populations. Making informed decisions about treatment and prognosis can be aided by knowledge of the functional effects and clinical implications of these alterations. Overall, the identification and characterization of novel *SMAD4* mutations hold promise for advancing our knowledge of colorectal cancer biology and improving patient outcomes in the future.

# References

Abu-Rmeileh, N. M. E., Gianicolo, E. A. L., Bruni, A., Mitwali, S., Portaluri, M., Bitar, J., Hamad, M., Giacaman, R., & Vigotti, M. A. (2016). Cancer mortality in the West Bank, Occupied Palestinian Territory Chronic Disease epidemiology. *BMC Public Health*, *16*(1), 1–10. https://doi.org/10.1186/s12889-016-2715-8

Ahmed, S., Bradshaw, A., Gera, S., Dewan, M. Z., & Xu, R. (2017). *The TGF-β / Smad4 Signaling Pathway in Pancreatic Carcinogenesis and Its Clinical Significance*. https://doi.org/10.3390/jcm6010005

Alazzouzi, H., Alhopuro, P., & Salovaara, R. (2005). *SMAD4 as a Prognostic Marker in Colorectal Cancer SMAD4 As a Prognostic Marker in Colorectal Cancer*. 2606–2611. https://doi.org/10.1158/1078-0432.CCR-04-1458

Aparisi F, A.-L. H. C.-F. S. T. S. H.-P. A. J.-L. E. B. A. I. V. and C. C., H, A.-L., S, C.-F., S, T., A, H.-P., E, J.-L., A, B., V, I., & C, C. (2019). Passenger mutations in cancer evolution. *Cancer Reports and Reviews*, *3*(3). https://doi.org/10.15761/crr.1000188

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, *66*(4), 683–691. https://doi.org/10.1136/gutjnl-2015-310912

Bardeesy, N., Cheng, K. H., Berger, J. H., Chu, G. C., Pahler, J., Olson, P., Hezel, A. F., Horner, J., Lauwers, G. Y., Hanahan, D., & DePinho, R. A. (2006). Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes and Development*, *20*(22), 3130–3146. https://doi.org/10.1101/gad.1478706

Biankin, A. V., Morey, A. L., Lee, C. S., Kench, J. G., Biankin, S. A., Hook, H. C., Head, D. R., Hugh, T. B., Sutherland, R. L., & Henshall, S. M. (2002). DPC4/Smad4 expression and outcome in pancreatic ductal adenocarcinoma. *Journal of Clinical Oncology*, *20*(23), 4531–4542. https://doi.org/10.1200/JCO.2002.12.063

Blackford, A., Serrano, O. K., Wolfgang, C. L., Parmigiani, G., Zhang, X., Parsons, D. W., Lin, J. C., Leary, R. J., Eshleman, J. R., Goggins, M., Jaffee, E. M., Iacobuzio-donahue, C. A., Maitra, A., Cameron, J. L., Olino, K., Schulick, R., Winter, J., Herman, J. M., Laheru, D., … Hruban, R. H. (2009). SMAD4 Gene Mutations Are Associated with Poor Prognosis in Pancreatic Cancer. *Clinical Cancer Research*, *15*(17), 4674–4680. https://doi.org/10.1158/1078-0432.CCR-09-0227

Brosens, L. A. A., Van Hattem, A., Hylind, L. M., Iacobuzio-Donahue, C., Romans, K. E., Axilbund, J., Cruz-Correa, M., Tersmette, A. C., Offerhaus, G. J. A., & Giardiello, F. M. (2007). Risk of colorectal cancer in juvenile polyposis. *Gut*, *56*(7), 965–967. https://doi.org/10.1136/gut.2006.116913

Chacko, B. M., Qin, B., Correia, J. J., Lam, S. S., Caestecker, M. P. De, & Lin, K. (2001). The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. *Nature Structural Biology*, *8*(3), 4–9.

Chen, L., Toke, N. H., Luo, S., Vasoya, R. P., Fullem, R. L., Parthasarathy, A., Perekatt, A. O., & Verzi, M. P. (2019). A reinforcing HNF4–SMAD4 feed-forward module stabilizes enterocyte identity. In *Nature Genetics* (Vol. 51, Issue 5, pp. 777–785). Springer US. https://doi.org/10.1038/s41588-019-0384-0

Cortez, V. S., Ulland, T. K., Cervantes-Barragan, L., Bando, J. K., Robinette, M. L., Wang, Q., White, A. J., Gilfillan, S., Cella, M., & Colonna, M. (2017). SMAD4 impedes the conversion of NK cells into ILC1-like cells by curtailing non-canonical TGF-β signaling. *Nature Immunology*, *18*(9), 995–1003. https://doi.org/10.1038/ni.3809

Ding, Z., Wu, C., Chu, G. C., Xiao, Y., Ho, D., Zhang, J., Perry, S. R., Labrot, E. S., Wu, X., Lis, R., Hoshida, Y., Hiller, D., Hu, B., Jiang, S., Zheng, H., Stegh, A. H., Scott, K. L., Signoretti, S., Bardeesy, N., & Wang, Y. A. (2011). SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, *470*(7333), 269–273. https://doi.org/10.1038/nature09677

Ewing, I., Hurley, J. J., Josephides, E., & Millar, A. (2014a). CURRICULUM BASED CLINICAL REVIEWS The molecular genetics of colorectal cancer. *Frontline Gastroenterology*, *5*(1), 26–30. https://doi.org/10.1136/flgastro-2013-100329

Ewing, I., Hurley, J. J., Josephides, E., & Millar, A. (2014b). CURRICULUM BASED CLINICAL REVIEWS The molecular genetics of colorectal cancer. *Frontline Gastroenterology*, *5*(1), 26–30. https://doi.org/10.1136/flgastro-2013-100329

Fearon, E. F., & Vogelstein, B. (1990). *A Genetic Model for Colorectal Tumorigenesis*. *61*, 759–767.

Fearon, E. R. (2011). Molecular genetics of colorectal cancer. *Annual Review of Pathology: Mechanisms of Disease*, *6*, 479–507. https://doi.org/10.1146/annurev-pathol-011110-130235

Fleming, N. I., Jorissen, R. N., Mouradov, D., Christie, M., Sakthianandeswaren, A., Palmieri, M., Day, F., Li, S., Tsui, C., Lipton, L., Desai, J., Jones, I. T., McLaughlin, S., Ward, R. L., Hawkins, N. J., Ruszkiewicz, A. R., Moore, J., Zhu, H. J., Mariadason, J. M., … Sieber, O. M. (2013). SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Research*, *73*(2), 725–735. https://doi.org/10.1158/0008-5472.CAN-12-2706

Fodde, R. (2002). *The APC gene in Colorectal Cancer*. *38*, 867–871.

Goel, A., & Boland, C. R. (2012). Epigenetics of Colorectal Cancer. *REVIEWS IN BASIC AND CLINICAL GASTROENTEROLOGY AND HEPATOLOGY*, *143*(6), 1442-1460.e1. https://doi.org/10.1053/j.gastro.2012.09.032

Goumans, M. J., Liu, Z., & Ten Dijke, P. (2009). TGF-β signaling in vascular biology and dysfunction. *Cell Research*, *19*(1), 116–127. https://doi.org/10.1038/cr.2008.326

Hahn, S. A., Schutte, M., Hoque, A. T. M. S., Moskaluk, C. A., Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., & Kern, S. E. (1996). *DPC4, A Candidate Tumor Suppressor Gene at Human Chromosome 18q21 .I. 3*, 11–14.

Hanahan, D., & Weinberg, R. A. (2011a). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hanahan, D., & Weinberg, R. A. (2011b). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hao, Y., Baker, D., & Dijke, P. Ten. (2019). TGF-β-mediated epithelial-mesenchymal transition and cancer metastasis. *International Journal of Molecular Sciences*, *20*(11). https://doi.org/10.3390/ijms20112767

Hata, A., & Chen, Y. (2016). *TGF- b Signaling from Receptors to Smads*.

Hernanda, P. Y., Chen, K., Das, A. M., Sideras, K., Wang, W., Li, J., Cao, W., Bots, S. J. A., Kodach, L. L., & Man, R. A. De. (2014). *SMAD4 exerts a tumor-promoting role in hepatocellular carcinoma*. *July*, 1–14. https://doi.org/10.1038/onc.2014.425

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

Hsu, H. C., Thiam, T. K., Lu, Y. J., Yeh, C. Y., Tsai, W. S., You, J. F., Hung, H. Y., Tsai, C. N., Hsu, A., Chen, H. C., Chen, S. J., & Yang, T. S. (2016). Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. *Oncotarget*, *7*(16), 22257–22270. https://doi.org/10.18632/oncotarget.8076

Hussein, S. M., Duff, E. K., & Sirard, C. (2003). Smad4 and β-catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of Msx2. *Journal of Biological Chemistry*, *278*(49), 48805–48814. https://doi.org/10.1074/jbc.M305472200

Isaiah J. Fidler. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Reviews Cancer*, *3*(1), 453–458. https://doi.org/10.1111/j.1937-5956.1995.tb00040.x

Itoh, S., Moustakas, A., Dijke, P., Heldin, C., & More, A. (2000). *Functional consequences of tumorigenic missense mutations in the amino-terminal domain of Smad4*. 4396–4404.

Itoh, S., & ten Dijke, P. (2007). Negative regulation of TGF-β receptor/Smad signal transduction. *Current Opinion in Cell Biology*, *19*(2), 176–184. https://doi.org/10.1016/j.ceb.2007.02.015

Jasperson, K. W., Tuohy, T. M., Neklason, D. W., & Burt, R. W. (2010). Hereditary and Familial Colon Cancer. *Gastroenterology*, *138*(6), 2044–2058. https://doi.org/10.1053/j.gastro.2010.01.054

Jeong, W., Ro, E. J., & Choi, K. (2018). Interaction between Wnt / β -catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of β -catenin and RAS by targeting the Wnt / β -catenin pathway. *Npj Precision Oncology*, *January*, 1–10. https://doi.org/10.1038/s41698-018-0049-y

Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B., & Berry, D. A. (2013). *Meta-analyses of colorectal cancer risk factors*. 1207–1222. https://doi.org/10.1007/s10552-013-0201-5

Jones, J. B., & Kern, S. E. (2000). Functional Mapping of the MH1 DNA-binding domain of DPC4/SMAD4. *Nucleic Acids Research*, *28*(12), 2363–2368.

Kempers, M. J. E., Kuiper, R. P., Ockeloen, C. W., Chappuis, P. O., Hutter, P., Rahner, N., Schackert, H. K., Steinke, V., Holinski-feder, E., Morak, M., Kloor, M., Büttner, R., Verwiel, E. T. P., Krieken, J. H. Van, Nagtegaal, I. D., Goossens, M., Bunyan, D. J., Syngal, S., Stoff, E. M., … Oláh, E. (2011). *Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome : a cohort study*. 49–55. https://doi.org/10.1016/S1470-2045(10)70265-5

Kim, J., Johnson, K., & Chen, H. J. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature*, *388*(July), 1132–1136.

Kuang, C., & Chen, Y. (2004). Tumor-derived C-terminal mutations of Smad4 with decreased DNA binding activity and enhanced intramolecular interaction. *Oncogene*, *23*(June 2003), 1021–1029. https://doi.org/10.1038/sj.onc.1207219

Lanauze, C. B., Sehgal, P., Hayer, K., Torres-Diz, M., Pippin, J. A., Grant, S. F. A., & Thomas-Tikhonenko, A. (2021). Colorectal cancer-associated smad4 R361 hotspot mutations boost Wnt/β-catenin signaling through enhanced smad4-LEF1 binding. *Molecular Cancer Research*, *19*(5), 823–833. https://doi.org/10.1158/1541-7786.MCR-20-0721

Lao, V. V., & Grady, W. M. (2011). Epigenetics and colorectal cancer. *Nature Reviews Gastroenterology and Hepatology*, *8*(12), 686–700. https://doi.org/10.1038/nrgastro.2011.173

Leoz, M. L., & Moreira, L. (2015). *The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management*. 95–107.

Li, D., Masiero, M., Banham, A. H., & Harris, A. L. (2014). *The Notch ligand Jagged1 as a target for anti-tumor therapy*. *4*(September), 1–14. https://doi.org/10.3389/fonc.2014.00254

Liao, W., Li, G., You, Y., & Wan, H. (2018). *Antitumor activity of Notch - 1 inhibition in human colorectal carcinoma cells*. 1063–1071. https://doi.org/10.3892/or.2017.6176

Lin, L., Chang, K., Cheng, H., Liu, C. J., & Hospital, M. M. (2019). *Somatic SMAD4 mutations in head and neck carcinoma are associated with tumor progression*. https://doi.org/10.3389/fonc.2019.01379

Link, M., Racu, M., Lebrun, L., Schiavo, A. A., Campenhout, C. Van, Clercq, S. De, Absil, L., Perez, E. M., Maris, C., Decaestecker, C., Salmon, I., & Haene, N. D. (2022). *The Role of SMAD4 Inactivation in Epithelial – Mesenchymal Plasticity of Pancreatic Ductal Adenocarcinoma : The*.

Ma, Y., Yan, F., Li, L., Liu, L., & Sun, J. (2014). Deletion and down-regulation of SMAD4 gene in colorectal cancers in a Chinese population. *Chinese Journal of Cancer Research*, *26*(5), 525–531. https://doi.org/10.3978/j.issn.1000-9604.2014.09.02

Martin-Malpartida, P., Batet, M., Kaczmarska, Z., Freier, R., Gomes, T., Aragón, E., Zou, Y., Wang, Q., Xi, Q., Ruiz, L., Vea, A., Márquez, J. A., Massagué, J., & Macias, M. J. (2017). Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. *Nature Communications*, *8*(1), 1–15. https://doi.org/10.1038/s41467-017-02054-6

Massagué, J., Seoane, J., & Wotton, D. (2005). Smad transcription factors. *Genes and Development*, *19*, 2783–2810. https://doi.org/10.1101/gad.1350705.embryo

Mccarthy, A. J., & Chetty, R. (2018). *Smad4 / DPC4*. 1–4. https://doi.org/10.1136/jclinpath-2018-205095

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., … Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, *266*(5182), 66–71. https://doi.org/10.1126/science.7545954

Miller, K. D., Sauer, A. G., & Fedewa, S. A. (2020). *Colorectal Cancer Statistics , 2020*. *0*(0), 1–20. https://doi.org/10.3322/caac.21601

Muller, P. A. J., & Vousden, K. H. (2014). Mutant p53 in cancer: New functions and therapeutic opportunities. *Cancer Cell*, *25*(3), 304–317. https://doi.org/10.1016/j.ccr.2014.01.021

Munteanu, I., & Mastalier, B. (2014). Genetics of colorectal cancer. *Journal of Medicine and Life*, *7*(4), 507–511. https://doi.org/10.1016/b978-0-12-091075-5.50016-0

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y. Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., … Thomson., E. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330–337. https://doi.org/10.1038/nature11252

Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, *11*(3), 220–228. https://doi.org/10.1038/nrm2858

Newman, A. C., Kemp, A. J., Drabsch, Y., Behrends, C., & Wilkinson, S. (2017). Autophagy acts through TRAF3 and RELB to regulate gene expression via antagonism of SMAD proteins. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00859-z

Park, J. W., Jeong, J. M., & Kim, H. K. (2022). Smad4 and p53 synergize in suppressing autochthonous intestinal cancer. *Cancer Medicine*, *November 2021*, 1925–1936. https://doi.org/10.1002/cam4.4533

Park, S. Y., Wilkens, L. R., Setiawan, V. W., Monroe, K. R., Haiman, C. A., & Le Marchand, L. (2019). Alcohol Intake and Colorectal Cancer Risk in the Multiethnic Cohort Study. *American Journal of Epidemiology*, *188*(1), 67–76. https://doi.org/10.1093/aje/kwy208

Parsa, N. (2012). Environmental factors inducing human cancers. *Iranian Journal of Public Health*, *41*(11), 1–9.

Peltomäki, P., Aaltonen, L. A., Sistonen, P., Pylkkänen, L., Mecklin, J. P., Järvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., Hamilton, S. R., De La Chapelle, A., & Vogelstein, B. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science*, *260*(5109), 810–812. https://doi.org/10.1126/science.8484120

Qin, J., Wen, B., Liang, Y., Yu, W., & Li, H. (2019). *Histone Modifications and their Role in Colorectal Cancer ( Review )*.

Rawla, P., Sunkara, T., & Barsouk, A. (2019a). Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przeglad Gastroenterologiczny*, *14*(2), 89–103. https://doi.org/10.5114/pg.2018.81072

Rawla, P., Sunkara, T., & Barsouk, A. (2019b). Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przeglad Gastroenterologiczny*, *14*(2), 89–103. https://doi.org/10.5114/pg.2018.81072

Richard Wooster, Graham Bignell, Jonathan Lancaster, Sally Swift, Sheila Seal, Jonathan Mangion, Nadine Collins, Simon Gregory, Curtis Gumbs, Gos Micklem, Rita Barfoot, Rifat Hamoudi, Sandeep Patel, Catherine Rices, Patrick Biggs, Yasmin Hashim, Amanda Sm, A. A. & M. R. S. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature*, *378*(7), 789–792.

Romero, D., Iglesias, M., Vary, C. P. H., & Quintanilla, M. (2008). *Functional blockade of Smad4 leads to a decrease in b -catenin levels and signaling activity in human pancreatic carcinoma cells*. *29*(5), 1070–1076. https://doi.org/10.1093/carcin/bgn054

Ropero, S., Fraga, M. F., Ballestar, E., Hamelin, R., Yamamoto, H., Boix-chornet, M., Caballero, R., Alaminos, M., Setien, F., Paz, M. F., Herranz, M., Palacios, J., Arango, D., Orntoft, T. F., Aaltonen, L. A., Jr, S., & Esteller, M. (2006). *A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition*. *38*(5), 566–569. https://doi.org/10.1038/ng1773

Ruiz, L., Kaczmarska, Z., Gomes, T., Aragon, E., Torner, C., Freier, R., Baginski, B., Martin-malpartida, P., Martin, N. De, Marquez, J. A., Cordeiro, T. N., Pluta, R., & Macias, M. J. (2021). Unveiling the dimer / monomer propensities of Smad MH1-DNA complexes. *Computational and Structural Biotechnology Journal*, *19*, 632–646. https://doi.org/10.1016/j.csbj.2020.12.044

Salovaara, R., Roth, S., Loukola, A., Launonen, V., Sistonen, P., Avizienyte, E., Kristo, P., Järvinen, H., Souchelnytskyi, S., Sarlomo-Rikala, M., & Aaltonen, L. A. (2002). Frequent loss of SMAD4/DPC4 protein in colorectal cancers. *Gut*, *51*(1), 56–59. https://doi.org/10.1136/gut.51.1.56

Samowitz, W. S., Sweeney, C., Herrick, J., Albertsen, H., Levin, T. R., Murtaugh, M. A., Wolff, R. K., & Slattery, M. L. (2005). Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. *Cancer Research*, *65*(14), 6063–6070. https://doi.org/10.1158/0008-5472.CAN-05-0404

Sarshekeh, A. M., Advani, S., Overman, M. J., Manyam, G., Kee, B. K., Fogelman, D. R., Dasari, A., Raghav, K., Vilar, E., Manuel, S., Shureiqi, I., Wolff, R. A., Patel, K. P., Luthra, R., Shaw, K., Eng, C., Maru, D. M., Routbort, M. J., Meric-Bernstam, F., & Kopetz, S. (2017). Association of SMAD4 mutation with patient demographics, tumor characteristics, and clinical outcomes in colorectal cancer. *PLoS ONE*, *12*(3), 1–14. https://doi.org/10.1371/journal.pone.0173345

Schetter, A. J., Sohn, J. J., Zanetti, K. A., Bowman, E. D., Chan, T. L., Kwong, D. L. W., Au, G. K. H., & Harris, C. C. (2012). *MicroRNA Expression Profiles Associated With Prognosis and Therapeutic Outcome in Colon Adenocarcinoma*. *299*(4), 425–436.

Schutte, M. (1999). DPC4/SMAD4 gene alterations in human cancer, and their functional implications. *Annals of Oncology*, *10*, 56–59.

Shi, Y., & Massague, J. (2003). *Mechanisms of TGF-beta Signaling from Cell Membrane to the Nucleus*. *113*, 685–700.

Shi, Y., Wang, Y., Jayaraman, L., Yang, H., & Massague, J. (1998). Crystal Structure of a Smad MH1 Domain Bound to DNA : Insights on DNA Binding in TGF- ␤ Signaling. *Cell*, *94*(2), 585–594.

Simon, K. (2016). Colorectal cancer development and advances in screening. In *Clinical Interventions in Aging* (Vol. 11, pp. 967–976). Dove Medical Press Ltd. https://doi.org/10.2147/CIA.S109285

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719–724. https://doi.org/10.1038/nature07943

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Suter, C. M., Martin, D. I., & Ward, R. L. (2010). *Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue*. *2004*, 95–101. https://doi.org/10.1007/s00384-003-0539-3

Tan, X., Tong, L., Li, L., Xu, J., Xie, S., Ji, L., Fu, J., & Liu, Q. (2021). Loss of Smad4 promotes aggressive lung cancer metastasis by de-repression of PAK3 via miRNA regulation. *Nature Communications*, 1–11. https://doi.org/10.1038/s41467-021-24898-9

Tanaka, H., Deng, G., Matsuzaki, K., Kakar, S., Kim, G. E., Miura, S., Sleisenger, M. H., & Kim, Y. S. (2006). BRAF mutation, CpG island methylator phenotype and microsatellite instability occur more frequently and concordantly in mucinous than non-mucinous colorectal cancer. *International Journal of Cancer*, *118*(11), 2765–2771. https://doi.org/10.1002/ijc.21701

Tanaka, T. (2009). *Colorectal carcinogenesis : Review of human and experimental animal studies*. https://doi.org/10.4103/1477-3163.49014

Tanaka, T., Watanabe, T., & Kazama, Y. (2008). Loss of Smad4 Protein Expression and 18qLOH as Molecular Markers Indicating Lymph Node Metastasis in Colorectal Cancer — A Study Matched for Tumor Depth and Pathology. *Journal of Surgical Oncology*, *September 2007*, 69–73. https://doi.org/10.1002/jso

Tang, B., Vu, M., Booker, T., Santner, S. J., Miller, F. R., Anver, M. R., & Wakefield, L. M. (2003). *TGF- β switches from tumor suppressor to prometastatic factor in a model of breast cancer progression*. *112*(7). https://doi.org/10.1172/JCI18899.TGF-

Tong, B. D. (2019). *The Smad2 / 3 / 4 complex binds miR - 139 promoter to modulate TGF β - induced proliferation and activation of human Tenon ' s capsule fibroblasts through the Wnt pathway*. *November 2018*, 1–11. https://doi.org/10.1002/jcp.28011

Valastyan, S., & Weinberg, R. A. (2011). Tumor metastasis: Molecular insights and evolving paradigms. *Cell*, *147*(2), 275–292. https://doi.org/10.1016/j.cell.2011.09.024

Van Cott, C. (2020). Cancer Genetics. *Surgical Clinics of North America*, *100*(3), 483–498. https://doi.org/10.1016/j.suc.2020.02.012

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, *340*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Voorneveld, P. W., Kodach, L. L., Jacobs, R. J., Liv, N., Zonnevylle, C., Hoogenboom, J. P., Biemond, I., Verspaget, H. W., Hommes, D. W., Rooij, K. De, Noesel, C. J. Van, Morreau, H., Van, T., Offerhaus, G. J., Brink, G. R. Van Den, Peppelenbosch, M. P., & Hardwick, J. C. H. (2014). Loss of SMAD4 Alters BMP Signaling to Promote Colorectal Cancer Cell Metastasis via Activation of Rho and ROCK. *Gastroenterology*. https://doi.org/10.1053/j.gastro.2014.03.052

Wagner, A., Barrows, A., Wijnen, J. T., Van Der Klift, H., Franken, P. F., Verkuijlen, P., Nakagawa, H., Geugien, M., Jaghmohan-Changur, S., Breukel, C., Meijers-

Heijboer, H., Morreau, H., Van Puijenbroek, M., Burn, J., Coronel, S., Kinarski, Y., Okimoto, R., Watson, P., Lynch, J. F., … Fodde, R. (2003). Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: High mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the MSH2 gene. *American Journal of Human Genetics*, *72*(5), 1088–1100. https://doi.org/10.1086/373963

Wan, R., Feng, J., & Tang, L. (2021). Consequences of mutations and abnormal expression of smad4 in tumors and t cells. *OncoTargets and Therapy*, *14*, 2531–2540. https://doi.org/10.2147/OTT.S297855

Wang, H. (2020). MicroRNAs and Apoptosis in Colorectal Cancer. *International Journal of Molecular Sciences*.

Wasserman, I., Lee, L. H., Ogino, S., Marco, M. R., Wu, C., Chen, X., Datta, J., Sadot, E., Szeglin, B., Guillem, J. G., Paty, P. B., Weiser, M. R., Nash, G. M., Saltz, L., Barlas, A., Manova-Todorova, K., Uppada, S. P. B., Elghouayel, A. E., Ntiamoah, P., … Joshua Smith, J. (2019). Smad4 loss in colorectal cancer patients correlates with recurrence, loss of immune infiltrate, and chemoresistance. *Clinical Cancer Research*, *25*(6), 1948–1956. https://doi.org/10.1158/1078-0432.CCR-18-1726

Weinberg, D. H. and R. A. (2000). The Hallmarks of Cancer. *La Medicina Del Lavoro*, *100*(4), 57–70.

WHO. (2022, February 3). *Cancer*. World Health Organization. (2022, February 3). Cancer. World Health Organization. Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Cancer . World Health Organization. (2022, February 3). Cancer. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/cancer

WHO. (2023). *Colorectal Cancer* . World Health Organization. (2023, July 11). Colorectal Cancer. World Health Organization. Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Colorectal-Cancer. https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer#:~:text=Colon%20cancer%20is%20the%20second,and%20mortality%20rates%20were%20observed.

Wilentz, R. E., Su, G. H., Le Dai, J., Sparks, A. B., Argani, P., Sohn, T. A., Yeo, C. J., Kern, S. E., & Hruban, R. H. (2000). Immunohistochemical labeling for Dpc4 mirrors genetic status in pancreatic adenocarcinomas: A new marker of DPC4 inactivation. *American Journal of Pathology*, *156*(1), 37–43. https://doi.org/10.1016/S0002-9440(10)64703-7

Wu, D. M., Zhu, H. X., Zhao, Q. H., Zhang, Z. Z., Wang, S. Z., Wang, M. L., Gong, W. Da, Tan, M., & Zhang, Z. D. (2010). Genetic variations in the SMAD4 gene and gastric cancer susceptibility. *World Journal of Gastroenterology*, *16*(44), 5635–5641. https://doi.org/10.3748/wjg.v16.i44.5635

Wu, W. K. K., Wang, X. J., Cheng, A. S. L., Luo, M. X. M., Ng, S. S. M., To, K. F., Chan, F. K. L., Cho, C. H., Sung, J. J. Y., & Yu, J. (2012). *Dysregulation and crosstalk of cellular signaling pathways in colon*. 1–27. https://doi.org/10.1016/j.critrevonc.2012.11.009

Xiao, Z., Latek, R., & Lodish, H. F. (2003). An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity. *Oncogene*, *22*, 1057–1069. https://doi.org/10.1038/sj.onc.1206212

Xu, J., & Attisano, L. (2000). Mutations in the tumor suppressors Smad2 and Smad4 inactivate transforming growth factor ß signaling by targeting Smads to the ubiquitin – proteasome pathway. *Proceedings of the National Academy of Sciences*, *97*(9), 2–7.

Xu, X., Brodie, S. G., Yang, X., Im, Y. H., Parks, W. T., Chen, L., Zhou, Y. X., Weinstein, M., Kim, S. J., & Deng, C. X. (2000). Haploid loss of the tumor suppressor Smad4/Dpc4 initiates gastric polyposis and cancer in mice. *Oncogene*, *19*(15), 1868–1874. https://doi.org/10.1038/sj.onc.1203504

Zhang, B., Halder, S. K., Kashikar, N. D., Cho, Y. J., Datta, A., Gorden, D. L., & Datta, P. K. (2010). Antimetastatic Role of Smad4 Signaling in Colorectal Cancer. *GASTROENTEROLOGY*, *138*(3), 969-980.e3. https://doi.org/10.1053/j.gastro.2009.11.004

Zhang, B., Zhang, B., Chen, X., Bae, S., Singh, K., Washington, M. K., & Datta, P. K. (2014). Loss of Smad4 in colorectal cancer induces resistance to 5-fluorouracil through activating Akt pathway. *British Journal of Cancer*, *110*(4), 946–957. https://doi.org/10.1038/bjc.2013.789

Zhao, M., Mishra, L., & Deng, C. X. (2018). The role of TGF-β/SMAD4 signaling in cancer. *International Journal of Biological Sciences*, *14*(2), 111–123. https://doi.org/10.7150/ijbs.23230

# Appendices

## Appendix A: Consent Forms

الجامعة العربية الأمريكية
ARAB AMERICAN UNIVERSITY

### موافقة للمشاركة في دراسة بحثية للمخصصات البيولوجية و الطفرات الجينيه التي لها دور في حدوث السرطان

الهدف من هذا المشـروع هو تجميع عدد كافي من العينات النسـيجية وعينات الدم من المرضـى المصـابين بالسـرطان في فلسـطين. قد تكون هناك حاجه للحصـول على عينه دم من بعض أفراد العائلة بقرار يحدده البحث وبموافقه الفرد لتحقيق الأهداف التاليه:

1- دراسـة *المخصـصـات البيولوجية* و الطفرات الجينيه التي لها دور في حدوث السـرطان بين المرضى في فلسطينن.
2- تحليل المعلومات الجينيه لغرض البحث عن طفرات غير معروفه ممكن ان تسـبب الاورام يليها تجارب لمعرفه اليه عمل الطفرات.

**المطلوب منك**

1- سـيقوم الجراح بإزالة خزعه لديك كجزء من خط العلاج و الفحص حسـب المقرر من قبللك و من قبل الطبيب.
2- بالإضافة إلى هذه العينة، نحن نأمل استخدام جزء من هذه العينة بالضافه الى عينة دم لأهداف بحثية لدراسه السرطان .
3- نـأمل موافقتك ان نراجع ملفك الطبي في حين بدء البحث و لاحقا بهدف تجميع المعلومات الضروريه للبحث مع العلم انا سنحاقظ على السريه **التامه**.

سوف يتم حفظ عينه النسيج في بنك النسيج الخاص في الجامعه و الوزاره. نحن نطمح استخدامها لاجراء تجارب قد تسـاعدنا على تحديد دور الطفرات الجينيه في التسـبب بالسـرطان. إذا اخترت المشاركة، ستبقي في الدراسة فقط لمدة سحب عينتك الدم و اخذ العينيه النسيجيه فقط.

في حال ظهور حاجه إلى استخدام عينه مخزنه مستقبلا فان ذلك يعتمد على رغبت المريض في استخدامها وللمريض الحق في الرفض أو الموافقة منذ الان و لن يتم استخدام العينة لأي هدف غير البحث

**المخاطر المترتبة على على المشاركة بهذه الدراسة**

المخاطرالمترتبة على سحب عينة دم من ذراعك تتضمن انزعاج لحظي فإن المخاطر الزائدة المترتبة على مشاركتك في البحث تكون ضئيلة.

إن عينة النسيج التي ستقوم بوهبنا إياها للبحث، ستكون من النسيج الذي سيتم إزالته خلال عمليتك الجراحية، وبالتالي لن يكون هناك عدم ارتياح إضافي ولا أي تغيير على العملية الجراحية المعتادة. هناك خطر من فقدان السرية (انظري قسم الخصوصية والسرية).

**الخصوصيه و السريه**

ستحفظ المعلومات البحثية والطبية الخاصة بهذه الدراسة تحديداً في ملف خاص بالبحث منفصل عن الملفات الطبية الخاصة بالمشفى. للمحافظة على الخصوصية، فإن اسمك أو أية معلومات قد تقوم بالتعريف بك، ستزال من عينة الدم ومن كل المعلومات عن وضعك الطبي وسيرتك المرضية التي ترافق عينتك. كل المعلومات الخاصة بالتعريف بك ستكون عن طريق إعطاء رمز خاص. هذا الرمز سيكون معروفا فقط لدى الباثين واضاء الطاقم الطبي المشارك في البحث.

**فوائد للمشاركة في هذه الدراسة؟**

1- الابحاث و بنك المعلومات الجينيه سوف يساعد في اكتشاف طفرات جينيه و عوامل بيولوجيه مسببه للمرض. دراسه هذه العوامل قد يساعد في ايجاد علاج افضل للمرضى في المستقبل.

**التكاليف والتعويضات**

لن يكون هناك أية تكاليف إضافية عليك إثر مشاركتك بهذه الدراسة. كل التكاليف المتعلقة باخذ عينة الدم وعمل التحليلات الوراثيه اللازمه سيتم تغطيتها من الموازنه المتوفره للبحث من الجامعه. لن نقوم بالدفع لك لمشاركتك بهذه الدراسة.

**الحق في رفض المشاركة بهذه الدراسة أو الانسحاب منها**

المشاركة بهذه الدراسة هي مشاركة **طوعية بحتة**. يمكنك رفض المشاركة في هذه الدراسة أو في حال وافقت، يمكنك الانسحاب منها إن رغبت في أي وقت. قرارك بعدم المشاركة أو الانسحاب من المشاركة بهذه الدراسة لن يترتب عليه أية عقوبة أو خسارة لأية امتيازات تتمتع بها، ولن يؤثر ذلك على مقدرتك في الحصول على الرعاية الصحية.

**بمن أتصل في حال كانت لدي أسئلة أو واجهتني مشاكل؟**

للسؤال عن الدراسة أو عن إصابة ناتجة عن المشاركة بالدراسة، أو إذا كانت لديك مشاكل، مخاوف، أسئلة أو اقتراحات حول البحث، الرجاء الاتصال ب _____

**نص الموافقة**

تم شرح لي الهدف من هذه الدراسة، الخطوات التي سيتم اتباعها، المخاطر والفوائد المترتبة على المشاركة بها. لقد تم السماح لي بسؤال أسئلة، وتمت الإجابة عن تساؤلاتي لحد يرضيني. لقد تم إخباري بمن أتصل إذا كانت لدي تساؤلات، أو لمناقشة مشاكل، أو مخاوف، أو اقتراحات متعلقة بالبحث، أو للحصول على معلومات أو إعطاء أية إضافات حول البحث. لقد قمت بقراءة وثيقة الموافقة هذه وأوافق على المشاركة بهذه الدراسة، مع العمل أنه بإمكاني الانسحاب أنّى شئت.

☐ أوافق على أن يتم أخذ عينة دم مني لفحص المادة الوراثية.

الاسم الثلاثي للمشارك/ة_____

توقيع المشارك/ة_____ التاريخ _____

الاسم للباحثين:

السيد محمود رزيقات _____

الدكتوره نوار القطب _____

الدكتور زيدون صلاح _____

التاريخ _____

## الملخص

يتزايد انتشار سرطان القولون والمستقيم على مستوى العالم، خاصة في الدول الناشئة. يعتبر سرطان القولون والمستقيم ثالث أكثر سرطان انتشارا وثاني أكثر سرطان مسبب للوفيات عالميا. حيث في دولة فلسطين يأتي بالمرتبة الثانية من السرطانات الأكثر شيوعا بعد سرطان الثدي. بسبب هذا الانتشار الشاسع، أصبح الكشف عن مسببات سرطان القولون أساسيا. حيث أصبحت الفحوصات الجينية من أهم وأحدث الطرق للكشف عن السرطان وللمساعدة في تحديد مسار العلاج.

في هذه الدراسة، تم استخدام تقنية تسلسل الجيل التالي (NGS) لتحديد المتغيرات الجينية التي تساهم في تكوين الأورام السرطانية في سرطان القولون والمستقيم. حدد تحليلنا المتغيرات التالية: (NOTCH1, p.Asp1533Gly), (EGFR, p.Leu619Pro), (MAP2K1, p.Pro265Ser), (ALK, p.Phe174Leu), (ERBB2, p.Asp1252His), (FBXW7, p.Gly477Arg), (FGFR2, p.Met538Ile), (GNA11, p.Leu170Phe), (MET, p.Asp153Asn), (NOTCH1, p.Ala1610Thr), (PIK3CA, p.Asn107Thr), (SMAD4, p.Cys115Arg).

كشفت دراستنا أنه يتم تعطيل جين SMAD4 في الخلايا الجسدية في أورام القولون والمستقيم، علاوة على ذلك، فقد تبين أن فقدان التعبير عن SMAD4 يرتبط بكل من تطور السرطان وانتشاره، والاستجابة الضعيفة للعلاج الكيميائي، وانخفاض التسلل المناعي، مما يعزز استخدامه كعلامة إنذار لمرضى سرطان القولون والمستقيم.

كشفت الدراسة أيضا ان طفرة (SMAD4, p. Cys115Arg) هي المسبب المحتمل في تطور مرض سرطان القولون وانتشاره. وللتحقيق في وظيفة المتغير المحدد، تم استخدام تقنية Luciferase assay لتقييم تأثير الطفرة على النسخ الجيني ل SMAD4 و قد كشفت أن الطفرة قللت من نشاط النسخ لجين SMAD4 بمقدار أربعة أضعاف مقارنة بالمورث شائع النمط من جين SMAD4.

يوفر هذا البحث فهمًا أفضل للأسباب الجزيئية للسرطان في فلسطين من خلال توصيف التأثيرات الوظيفية ل (SMAD4, p. Cys115Arg). علاوة على ذلك، قد يساهم في تطوير علاجات شخصية ومستهدفة لأولئك الذين تم تشخيص إصابتهم بسرطان القولون والمستقيم.