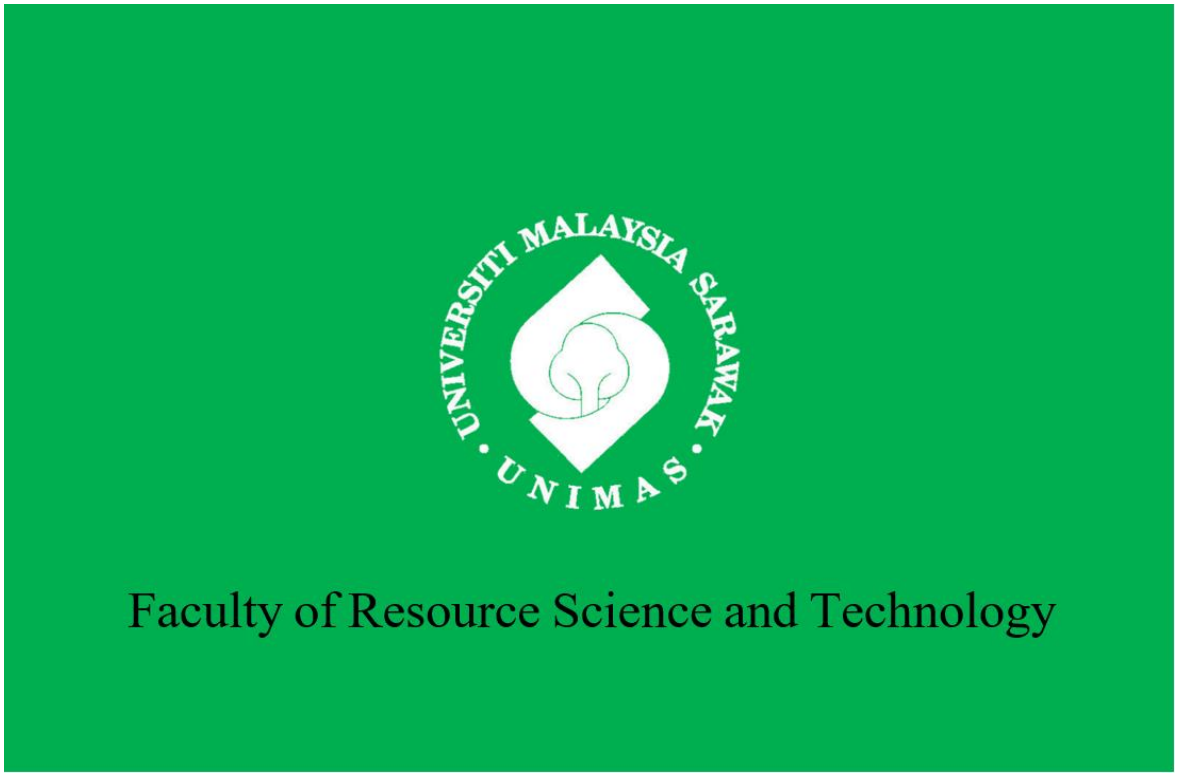

Bibliographic analysis of p53-gene family and gene sequence analysis of p73 in colorectal carcinoma.

Brennan Tang Yet Shen (69248)

Bachelor of Science with Honours
(Resource Biotechonology)
2022

Bibliographic analysis of p53-gene family and gene sequence analysis of p73 in colorectal carcinoma.

Brennan Tang Yet Shen

A thesis submitted in partial fulfilment of the Requirement of The Degree Bachelor of Science with Honours

(Resource Biotechnology)

SUPERVISOR: PROFESSOR DR. EDMUND SIM UI HANG

Programme of Resource Biotechnology
Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2022

DECLARATION

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)	
Final Year Project Report	√
Masters	
PhD	

DECLARATION OF ORIGINAL WORK

This declaration is made on the 15th day of June 2022.

Student's Declaration:

I Brennan Tang Yet Shen (69248), Faculty of Resource Science and Technology (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, Bibliographic analysis of p53-gene family and gene sequence analysis of p73 in colorectal carcinoma is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

15th June 2022 — Date submitted

Brennan Tang Yet Shen (69248) — Name of the student (Matric No.)

Supervisor's Declaration:

PROF. DR. EDMUND SIM UI HANG

I, ---------- (SUPERVISOR'S NAME), hereby certify that the work entitled, ----AS STATED IN THE STUDENT DECLARATION SECTION----- (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full ---------- fulfillment for the conferment of ---------- ---BSc Hons (Resource Biotechnology)---- -- (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

PROF DR EDMUND SIM UI HANG

Received for examination by: _____ Date: 14 JULY 2022

(Name of the supervisor)

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*** RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*****

☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitise the content to for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student himself / herself neither third party on this Project/Thesis once it becomes sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student's signature  (15th June 2022)

Supervisor's signature: Edmund MH Sim (14 July 2022)

Current Address:
88-10-02, Puncak Erskine, Jalan Fettes, Tanjong Tokong, 10470, Georgetown, Penang

Notes: ***** If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument was duly prepared by The Centre for Academic Information Services]

ACKNOWLEDGEMENT

Firstly, I would like to take this opportunity to thank my supervisor, Professor Dr. Edmund Sim Ui Hang, who had been constantly putting in all his time and effort supervising me throughout this project. Throughout the research, he had constantly been providing me advice and recommendations apart from sharing his knowledge and experiences on the steps and procedures related to this research. Without his guidance, I think I would not have come so far to completing this final year project.

Not only that, but I would also like to thank all the postgraduates inside the Immunology Molecular Genetics Laboratory, who are Ms. Adrianne Vong Dien Yu, Ms. Daphne Olivia Jawai Anak Sadai, and Ms. Alyaa Syafiqah binti Shamsuri who had been providing me their utmost care and attention as well as experiences on the lab works despite their pack schedules.

Not to forget, I would like to thank my parents, Mr. Tang Cheow Khuan and Mrs. Teh Gaik Kheng, who were always there for me whenever I am down and in need of mental support. They were there giving me their full encouragement and motivation for me to not give up although I had encountered numerous obstacles throughout this project. Finally, I am also forever grateful to my friends, Ms. Aliza Chung Yen Fang, Ms. Bonnie Tang Bao Chee, Ms. Emmanuella Nina Christina, Ms. Ling Zhao Wen, Ms. Michelle Yew Zhi Yan, and Mr. Thanes Gunasekaran, who were always there sharing their experiences and findings throughout this project.

Bibliographic analysis of p53-gene family and gene sequence analysis of p73 in colorectal carcinoma.

Brennan Tang Yet Shen

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Colorectal carcinoma (CRC) is one of the leading causes of cancer deaths yearly. Lately, there has been an influx of information regarding the p53-gene family in colorectal carcinoma in the form of publications. Therefore, there is a need to access and quantitatively analyze these publications in detail to obtain the parameters and trend of this topic of research. This information is said to be the key to various cancer treatments and act as a platform to frame future policies regarding cancer. Previous studies have also shown that mutations in the p73 gene have limited adverse effects leading to colorectal cancer. Bibliographic analysis of the p53-gene family was done using RStudio on the Scopus database to analyse this information by observing the parameters and trend of this topic of research throughout the years to predict its development. This study focused on accessing the parameters and trends of publications pertaining to the p53-gene family as well as to investigate the correlation between mutations in the p73 gene with colorectal cancer through comparative gene sequence analysis. Publications from the year 2010 to 2021 were used for the bibliographic analysis. The number of publications on the p53-gene family in colorectal cancer were found to be steadily increasing with China having the highest number of international collaborations at 534. In comparative gene sequence analysis, HT29 and HCT116 cell lines were used where the p73 gene were sequenced and compared with the wild type. There was limited correlation found between mutations of the p73 gene with the development of colorectal carcinoma.

Key words: Colorectal cancer, bibliographic analysis, p53, gene sequence analysis, p73.

ABSTRAK

Kanser kolorektal (CRC) merupakan salah satu punca utama yang menyebabkan kematian kanser setiap tahun. Sejak kebelakangan ini, kajian-kajian dalam bentuk penerbitan telah dijalankan terhadap keluarga gen p53 dalam kanser kolorektal dan telah memberi banyak informasi tentang gen tersebut. Oleh itu, terdapat keperluan untuk mengakses dan menganalisis penerbitan-penerbitan ini secara kuantitatif dengan lebih terperinci untuk mendapatkan parameter dan trend mengenai topik kajian. Dapatan daripada kajian tersebut menjadi kunci utama kepada pelbagai rawatan kanser dan bertindak sebagai sebuah platform untuk merangka polisi mengenai kanser. Juga, dapatan kajian sebelum ini menunjukkan bahawa mutasi yang berlaku pada gen p73 mempunyai kesan buruk terhad yang boleh menyebabkan berlakunya kanser kolorektal. Analisis bibliografi terhadap keluarga gen p53 telah dilaksanakan di pangkalan data Scopus dengan menggunakan RStudio untuk menganalisa maklumat dengan memerhatikan pelbagai parameter dan trend berkenaan dengan topik yang dikaji sepanjang kajian dilakukan untuk meramal perkembangannya. Kajian ini mengfokuskan pengaksesan parameter yang pelbagai dan trend penerbitan berkaitan dengan keluarga gen p53 dan juga, untuk mengkaji hubungan di antara mutasi yang berlaku di gen p73 dan kanser kolorektal melalui analisis perbandingan jujukan gen. Penerbitan dari tahun 2010 ke 2021 telah digunakan bagi tujuan analisis bibliografi. Terdapat peningkatan pada penerbitan tentang keluarga gen p53 dan kanser kolorektal, di mana China mempunyai jumlah kolaborasi tertinggi di peringkat antarabangsa iaitu 534. Manakala bagi analisis perbandingan jujukan gen, garisan sel HT29 dan HCT116 telah digunakan dimana gen p73 telah dijujukan dan dibandingkan dengan gen p73 jenis liar. Didapati bahawa hubungan di antara mutasi gen p73 dan pertumbuhan karsinoma kolorektal adalah terhad.

***Kata kunci**: Kanser kolorektal, analisis bibliografi, p53, analisis perbandingan jujukan gen, p73.*

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CRC	Colorectal cancer
LOH	Loss of heterozygosity
RT-PCR	Reverse transcription polymerase chain reaction
AGE	Agarose gel electrophoresis
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
IDE	Integrated Development Environment
BLAST	Basic Local Alignment Search Tool
PCR	Polymerase Chain Reaction
mRNA	Messenger RNA
SSP	Sessile serrated polyps
TSA	Traditional serrated adenomas
AJCC	American Joint Committee on Cancer
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
TAE	Tris-acetate EDTA
EtBr	Ethidium bromide
cDNA	Complementary DNA
dNTP	Deoxynucleoside triphosphate
NCBI	National Centre for Biotechnology
PBS	Phosphate-buffered saline
EDTA	Ethylenediaminetetraacetic acid

CHAPTER 1

INTRODUCTION

1.1 Background

Cancer is a category of disease that can manifest anywhere in the human body. Without doubt, cancer is known to be a major public health concern and is found to be the key cause of mortality worldwide, accounting for about 10 million deaths yearly. Among all the types of cancers, colorectal cancer (CRC) is one of the main causes of cancer death which contributes about 19% of all cancer deaths in 2020 (World Health Organization, 2021a). The rate of cancer occurrence is predicted to reach 28.4 million cases by 2040, that is a 47% rise from the year 2020 (Sung et al., 2021).

It is found that cancers can be regulated by multiple genes, such as by the p53-gene family. Due to this, various publications of elementary research, evaluation of curative effects, clinical trials, and many more on this gene have been done. However, this has caused an influx of information (Foy et al., 2018). Many data repositories, institution repositories, and databases have been developed with the introduction of new tools and techniques to perform citation analysis of a specific field (Haq and Al Fouzan, 2018). With this, system analysis of literature resource can be done to determine the development of this field of study, which could ultimately take part in framing future policies regarding CRC.

Up till now, there are so far limited bibliographic analysis done which focuses on the p53-gene family of cancers. It is not known whether bibliographic analysis of relevant literature is a valid approach to establish a comprehensive list of parameters regarding the p53-gene family, as well as to predict the development of this topic of study. These publications can be said to be the backbone for information resource pertaining to cancer and could be the key to various cancer treatments. Therefore, there is a need to access in detail and quantitatively

analyse these publications using mathematical and statistical methods to obtain the parameters and trend of this topic of research throughout the years. In this research, data collection of articles related to the p53-gene family on the Scopus database was firstly obtained, followed by the validation of the articles using the PRISMA Checklist and lastly, the analysis of the data.

Not only that, but accumulating evidence have also proved that neither complete loss nor mutations of the p73 gene will lead to CRC (Pfeifer et al., 2005). Research conducted by Bénard et al. (2003), where 1426 various human tumours were reviewed have shown that only 0.6% of the incidence has missense mutation and only about 20% had loss of heterozygosity (LOH) of the p73 gene. Thus, gene sequence analysis of the p73 gene in different CRC cell lines will prove whether there is a linkage between mutations of the p73 gene with the onset of colorectal carcinoma. Human CRC cell lines, HT29 and HCT116 were used in this research. p73 and GAPDH were isolated from the cell lines through the usage of reverse transcription PCR (RT-PCR). Then, agarose gel electrophoresis (AGE) was conducted for RNA quantification and visualization. The products were then outsourced for DNA sequencing. Lastly, data analysis and bioinformatics study were conducted.

In this research, bibliographic analysis of the p53-gene family was done where the parameters, i.e., annual scientific production, document types, the most relevant sources, the most productive author, the most productive country, and institution as well as the trend of publications was obtained and accessed. Not only that, the gene sequence analysis of the p73 gene was investigated as well and then compared to the wild type to find out whether there is a linkage between variations in the gene sequence with CRC. Through this research, the development of research on the p53-gene family in CRC was known by considering the parameters and trends of publications. Lastly, the correlation between the p73 gene and CRC was also investigated in depth.

1.2 Objectives

The main purpose of this research is:

1) To identify the research trend of the p53-gene family in colorectal carcinoma.
2) To perform comparative sequence analysis of the p73 gene in colorectal carcinoma using bioinformatics tools.

CHAPTER 2

LITERATURE REVIEW

2.1 Cancer

2.1.1 Background of cancer

Cancer can be defined as a disease that results from the uncontrolled growth and proliferation of cells in the body. Under normal circumstances, cells would grow and multiply through cell division as required for the healthy growth and development of the body while the old or damaged cells would often be degraded through a process called apoptosis. However, this process could in times break down and results in abnormal or damaged cells to grow and multiply uncontrollably which could lead to the formation of tumour (Cairns, 1981). Having a tumour does not indicate that the individual has cancer as tumours can be further divided into two, which are malignant (cancerous) or benign (non-cancerous). In cancer, the main concern would be on the malignant tumour as it can spread and invade surrounding tissues and could also potentially metastasize (travel to distant parts of the body to form new tumours) (Jones and Baylin, 2007).

In terms of cell growth, cancerous cells have several processes which allows it to grow uncontrollably. Among them is by having the ability to proliferate in the absence of signals, ignoring apoptosis, promote growth of blood vessels in tumours, invading surrounding cells and many more (Hanahan and Weinberg, 2011). This gives cancerous cells an advantage over normal cells which makes them detrimental and deadly. Moreover, cancer could initiate from anywhere in the body, such as from the lungs, colon, breast, liver, throat and many more. But the focus of this study would be on colorectal cancer which contributes one of the highest rates of death amongst all types of cancers.

2.1.2 Colorectal cancer

Colorectal cancer can be defined as a group of cancer that is started in the colon or rectum where the cells proliferate out of control. In short, it is oftenly referred to as colon cancer. At times, abnormal growth in the inner lining of the colon or rectum, also known as polyps, may develop in the colon or rectum. With time, the polyps may develop into cancer (Centers for Disease Control and Prevention, 2022). The probability of a polyp developing into cancer may vary depending on its type. There are basically three main types of polyps: hyperplastic polyps and inflammatory polyps, adenomatous polyps (adenomas), as well as traditional serrated adenomas (TSA) and sessile serrated polyps (SSP).

If cancer develops in a polyp, it can manifest into the wall of the colon or rectum with time. Although the walls consist of many layers, but colorectal cancer will often start at the innermost layer (the mucosa) and grows outwards through the other layers. It is worth to note that colorectal cancer metastasises by growing into blood vessels or lymph vessels, making it able to travel to adjacent lymph nodes or to other parts of the body (American Cancer Society, 2022).

In terms of staging (extent of spread), colorectal cancer is staged based on the extent it proliferates into the wall of the colon or rectum and whether it has metastasised. This will enable prediction on the seriousness of the cancer and the appropriate ways to treat it. In addition to that, survival statistics could also be obtained through cancer staging. The initial stage of colorectal cancer is regarded as stage 0 (a very early cancer), which later leads to stage I (1) to IV (4). Undeniably, the lower number signifies lower extent of spread of the cancer to other parts of the body. Although cancer in every individual is considered unique, but cancers with similar stages are usually found to have the same appearance and are often given the same treatment.

The staging system widely used for colorectal cancer is the American Joint Committee on Cancer (AJCC) TNM system, which are based on 3 pieces of key information: the degree of the tumour, the spread to adjacent lymph nodes, and metastasis to distant sites. Therefore, these 3 pieces of key information would always be the focus of researchers and medical officers when diagnosing the prognosis of the cancer.

2.1.3 HT29 cell line

HT-29 is a cell line that was isolated from a primary tumour belonging to a 44-year-old, White, female patient with colorectal adenocarcinoma in the year 1964. Since then, many cell lines of human colorectal cancer has been isolated as well. However, this cell line was still preferred due to its ability to express characteristics of mature intestinal cells, like mucus producing cells and enterocytes. As a result, this cell line is also receiving special interest in bioavailability and food digestion studies. To add on, it is also a cell line with epithelial morphology and has been a suitable transfection host and has numerous applications in cancer and toxicology research currently (National Library of Medicine, 2022). Due to this, this cell line has been chosen to be used as one of the samples to investigate the gene sequence analysis of the p73 gene in colorectal carcinoma which is one of the main focuses of this research. Figure 2.1 below shows the image of the HT29 cell line used for this research.

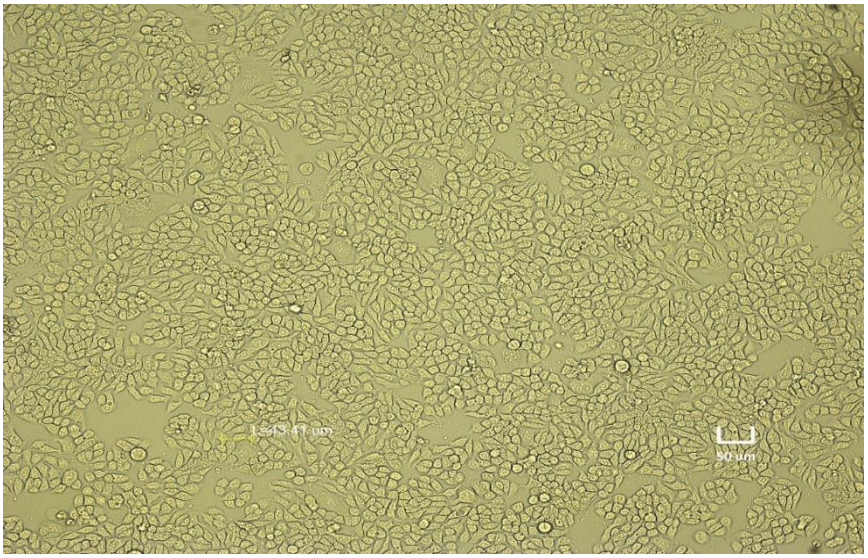

Figure 2.1: HT29 cell line.

To add on, these cells proliferate as a nonpolarized, undifferentiated multilayer under standard culture conditions (Martínez-Maqueda et al., 2015). However, they can also form a monolayer with tight junctions between cells and a distinguishing apical brush border in differentiated phenotypes.

2.1.4 HCT116 cell line

HCT116 is a cell line that was isolated from the colon of an adult male having colon cancer. In fact, it was found that there is a mutation in codon 13 of the ras proto-oncogene in this cell line and this makes it possible to be used as a positive control for PCR assays of mutations in this specific codon. Moreover, this cell line is also adherent with epithelial morphology and can metastasize in xenograft models. In terms of growth, HCT116 cells can grow *in vitro* at an optimum rate with a doubling time of approximately 18 hours. Therefore, this cell line is suitable for various *in vitro* and *in vivo* experimentation (Rajput et al., 2008).

In addition to that, this cell line is usually used in various studies involving colorectal carcinoma development and corresponding inhibitors. In this research, the HCT116 cell line

was also used as a sample. Figure 2.2 below shows the image of the HCT116 cell line used in this research.

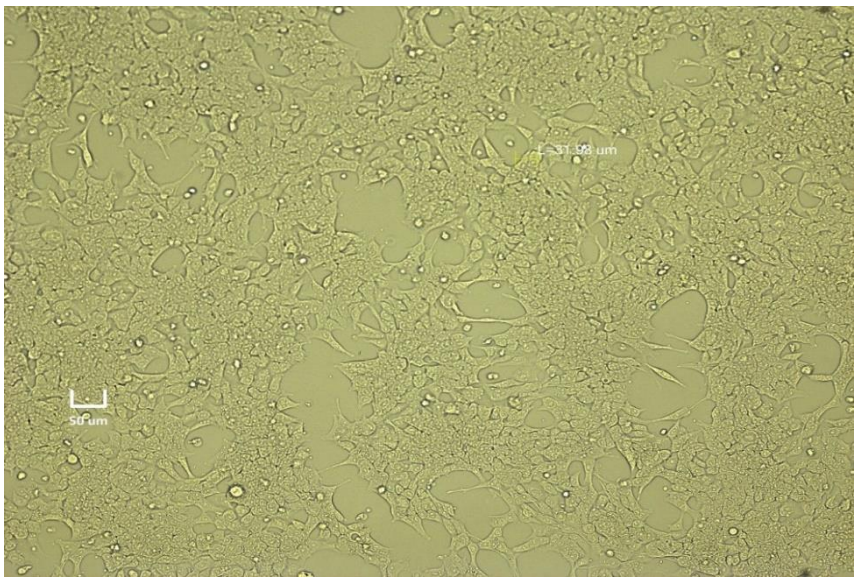

Figure 2.2: HCT116 cell line.

2.2 p53-gene family

The p53-gene family comprises of the TP53, TP63, and TP73 genes. Undoubtedly, each gene can synthesize protein isoforms through various mechanisms including extensive alternative mRNA splicing. Various studies have been made signifying the importance of these isoforms in the regulation of biological processes in normal cells. Clearly, abnormal expression of these genes will ultimately contribute to tumorigenesis (Wei et al., 2012).

The p53 gene, also known as TP53 because of its molecular mass, which breaks off in the 53 kilodalton fraction, is a tumour suppressor gene that encodes a protein that regulates the cell cycle. The gene is located on chromosome 17 (17p13.1). Not only that, but it also plays a tremendous role in multicellular organisms to suppress cancer by preventing genome mutation and conserving stability. This is made possible through the control of the cell cycle and apoptosis. A defect in the p53 gene could possibly allow abnormal cells to proliferate and

would at last, lead to cancer. Pflaum et al. (2014) mentioned that approximately 50% of human tumours contains p53 mutants, which is unusually high.

Moreover, it has been found that the p53 protein levels is lower in normal cells, and an increase of it is undeniably due to DNA damage and various stress signals. The p53 comprise of three main functions: DNA repair, growth arrest, and apoptosis. In growth arrest, the development of the cell cycle is terminated, hindering the replication of damaged DNA. However, the p53 present during growth arrest may also initiate the transcription of proteins associated in DNA repair. The last resort will be apoptosis where the cell is programmed for destruction to prevent the proliferation of cells containing the mutated DNA (Kaelin, 1999).

The p63 gene, also widely known as TP63 is highly involved in providing instructions for synthesizing a protein known as tumour protein p63. This protein functions as a transcription factor, which functions to control activities of genes by turning genes on and off at different times. To add on, the action of this protein is also important in regulating cell activities, including cell growth and proliferation, cell differentiation, maintenance, cell adhesion, as well as apoptosis. Not only that, the p63 protein also plays various roles in early development, such as on the normal development of ectodermal structures like the skin, teeth, hair, and nails (Westfall and Pietenpol, 2004).

The p73 gene, also known as TP73 is a gene that code for a member of the p53 family of transcription factors which are mainly involved in cellular responses to development and stress. The gene is mapped to a region on chromosome 1p36 that is often found to be deleted in tumours. Nonetheless, this region of the chromosome is also said to contain many tumour suppressor genes. It is also found that the p73 gene originates from the same family as the p53 and p63 gene as they are found to have high sequence similarity. Due to that, p73 gene could also transactivate p53-responsive genes causing cell cycle arrest and apoptosis. This is because

the family members can interconnect with one another in various ways, commonly on direct or indirect protein interactions. This results in regulation of the similar target gene promoter or either regulation of each other's promoters.

In addition to that, p73 levels are also found to be expressed at minute levels in normal tissues and instead found to be differentially expressed in several tumours. It also expresses approximately 35 mRNA variants through the existence of its alternate translation initiation sites, alternate promoters, and multiple splice variations. In layman's term, this can account for about 29 different p73 isoforms, but up till now, the nature and validity for most variants have not been determined yet.

2.3 GAPDH gene

The glyceraldehyde-3-phosphate dehydrogenase gene, known too as the GAPDH gene, is a highly conserved gene which encodes the glyceraldehyde-3-phosphate dehydrogenase protein family. In humans, there is only one functional gene for GAPDH. In fact, the encoded protein has been flagged as a moonlighting protein due to its potential to perform various functions. In terms of function, the product of this gene is vital in carbohydrate metabolism (the reverse oxidative phosphorylation of glyceraldehyde-3-phosphate) where it functions to catalyse an important energy-yielding step. Moreover, the encoded protein has also been found to exhibit uracil DNA glycosylase activity in the nucleus (Barber et al., 2005).

In terms of scientific usage, the GADPH gene is one of the most widely used housekeeping genes for comparison in gene expression studies. This is vital as for accurate quantitative RNA expression in reverse transcription polymerase chain reaction (RT-PCR) technique. Reference genes should always have constant expression levels between cells of different tissues and under varying experimental conditions (Zainuddin et al., 2010). In this

research, the GADPH gene was used as an internal control for the gene sequence analysis study of the p73 gene in colorectal carcinoma.

2.4 Focus of research

2.4.1 Bibliographic analysis: determine global research trends

Bibliographic analysis can be regarded as a content analysis of the publications in a certain field of study. It considers the author's research areas, the content, as well as the citation network in the paper. Moreover, various kind of information needs to be considered when doing models of bibliographic data. The articles often go along with metadata such as publication data, authors, categories, and time. Cited papers are often included as well. Through the modelling of the author's topic preferences, the topic information of the document is somehow associated with the author (Lim and Buntine, 2016). To be clear, it has helped many researchers in exploring and analysing large volumes of scientific data.

In bibliographic analysis, there are usually three steps involved: collection of related work, filtering of relevant work, as well as analysis and review of the related work. The first step usually includes a keyword-based search for articles and manuscripts from scientific databases. Next, big data which does not apply to the domain of interest is often filtered out. Then, the data of interest will be reviewed and analysed using various tools to ease the process of the analysis (Kamilaris et al., 2017). There are actually numerous software tools out there that supports bibliographic analysis, but most of this software are unable to assist scholars and authors in a thorough recommended workflow (Aria and Cuccurullo, 2017). Among most relevant software tools out there in the market, which are used widely by researchers are BibExcel, VOSviewer, CiteSpace, CitNetExplorer, PoP (Publish of Perish), RStudio and many more. But the software of choice must be the one that suites the analysis and study of the researcher to obtain optimal results (LiuPost, 2021).

Regardless of all those software tools out there, RStudio was found to be the best tool that suits the scope of this research. RStudio is a tool mainly focusing on analysing data relating to scientific research, data science and technical communication (Horton and Kleinman, 2015). With that, RStudio would be the tool of choice for this bibliographic analysis as it best suits scientific research which is the focus of this study.

In this research, bibliographic analysis of the p53-gene family in colorectal cancer will be accessed in detail. Clearly, this analysis would benefit all the previous studies which have been conducted regarding the p53-gene family as various parameters and the trend of publication will be known. Indeed, this will indirectly show the development of this topic.

2.4.1.1 RStudio: primary tool in bibliographic analysis

RStudio is an Integrated Development Environment (IDE) for R (a programming language for statistical computation and graphics). There are two formats of RStudio; RStudio Desktop (desktop application) and RStudio Server (runs on remote server) (Allaire, 2012). Both formats are available for free on the internet and could be downloaded from the RStudio website (https://www.rstudio.com). But there are also fee-based editions of this software which are mostly used for commercial purposes. However, for this research, we would only be using the free version of the software.

In the R environment, many packages have so far been made available on the official repository addressing the field of bibliometrics. Nonetheless, each package has its own specific analysis functions, and there is so far no package that can address the entire workflow (Verzani, 2011). Among such packages known are Shiny, Tidyverse, flexdashboard, Sparklyr, RMarkdown and many more (RStudio, 2021). Thus, it is important to know the packages that are best suited when dealing with bibliometrics to ensure that the documents and data are accurately analysed. This is because the bibliometric R-package is responsible in providing the