CONSTRUCTION OF GENE EXPRESSION MODEL ON RELATIONSHIP BETWEEN *CANDIDA ALBICANS* AND COLORECTAL CANCER

ZHANG HAOLING

UNIVERSITI SAINS MALAYSIA

2024

CONSTRUCTION OF GENE EXPRESSION MODEL ON RELATIONSHIP BETWEEN CANDIDA ALBICANS AND COLORECTAL CANCER

by

ZHANG HAOLING

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

July 2024

ACKNOWLEDGEMENT

As I gaze upon the completed manuscript, an overwhelming sense of gratitude washes over me, reminiscent of the cherished moments spent in pursuit of knowledge during my graduate studies. Throughout this journey, I cannot help but reflect upon the invaluable guidance provided by my esteemed supervisor, Associate Professor Dr Doblin Sandai. Under the meticulous mentorship of Assoc Prof. Dr Doblin Sandai, every facet of my research endeavour, from the inception of my graduation thesis to its defence was meticulously nurtured and supported. Assoc Prof. Dr Doblin Sandai's unwavering commitment to academic excellence and rigorous scientific inquiry served as a beacon, illuminating my path to intellectual growth and professional development. His profound wisdom, coupled with relentless encouragement and astute guidance, not only propelled me towards significant strides in my academic and scientific pursuits but also instilled within me enduring principles of integrity and fortitude, which I shall carry forward throughout my lifetime. I extend my heartfelt gratitude to Dr Mohammad Syamsul Reza Harun for his unwavering assistance, steadfast support, nurturing care, and unwavering encouragement throughout this transformative journey. I am deeply grateful to Professors Song Zhijing, Zhang Zhongwen, and Wang Wei for their unwavering support and guidance throughout my academic journey, spanning from my undergraduate studies to graduate school. Their enthusiasm and dedication have played a pivotal role in shaping my academic growth and success. Furthermore, I extend my sincere

gratitude to Dr. Rosline Sandai for her nurturing care and steadfast support, both academically and personally. Her mentorship has left an indelible mark on my life, influencing not only my academic pursuits but also my personal growth. Dr. Rosline Sandai holds a special place in my heart as my most esteemed and admired teacher. Heartfelt appreciation is also extended to my parents, brother and sisters whose unwavering financial and emotional support have been my bedrock throughout this journey. Their encouragement and belief in my abilities have been a constant source of motivation. This odyssey transcends the realms of academia, imparting invaluable life lessons and timeless principles that shall continue to enrich my journey, both personally and professionally.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate
ANOVA	Analysis of Variance
AKT	Protein Kinase B
AUC	Area Under the Curve
BME	Basement membrane extracellular matrix
βmRNA	Beta Messenger Ribonucleic Acid
C. albicans	*Candida albicans*
CRA	Colorectal Adenoma
CRC	Colorectal Cancer
CCK8	Cell Counting Kit-8
CD4+	Cluster of Differentiation 4 positive
CD8+	Cluster of Differentiation 8 positive
CO_2	Carbon Dioxide
cDNA	Complementary DNA
Cox	Cox proportional hazards model
CI	Confidence Interval
CXCL8	C-X-C Motif Chemokine Ligand 8
CXCL12	C-X-C Motif Chemokine Ligand 12
CXCL13	C-X-C Motif Chemokine Ligand 13
CLDN23	Claudin-23
DNA	Deoxyribonucleic acid
DCs	Dendritic cells
DFS	Disease-Free Survival
eATP	Extracellular ATP
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EHD1	EH domain-containing protein 1
EHD2	EH domain-containing protein 2

EHD4	EH domain-containing protein 4
expmRNAn	Experimental messenger RNA normalized
EHD	Eps15 Homology Domain
ES2	Endometrial stromal sarcoma cell line 2
EMT	Epithelial-mesenchymal transition
FC	Full change
FDFT1	Farnesyl-diphosphate farnesyltransferase 1
FBS	Fetal bovine serum
FABP4	Fatty acid-binding protein 4
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GEO	Gene Expression Omnibus
GADD45B	Growth Arrest and DNA Damage-inducible Beta
GSEA	Gene Set Enrichment Analysis
GPCRs	G protein-coupled receptors
HC	healthy controls
HR	Hazard Ratio
HIF1α	Hypoxia-Inducible Factor 1 Alpha
ICAM-1	Intercellular Adhesion Molecule 1
IECs	In intestinal epithelial cells
ISCU	Iron-Sulfur Cluster Assembly Enzyme
IFN-γ	Interferon-gamma
IL-4	Interleukin-4
IL13RA2	Interleukin-13 receptor alpha-2
KM	Kaplan-Meier
KEGG	Kyoto Encyclopedia of Genes and Genomes
LIME1	Lck-interacting transmembrane adaptor 1
LASSO	Least Absolute Shrinkage and Selection Operator
LM	liver Metastasis
mTOR	Mechanistic Target of Rapamycin
MyD88	Myeloid differentiation primary response 88
MALDI-ToF	Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry
MBD3	Methyl-CpG Binding Domain Protein 3
MELK	Maternal Embryonic Leucine Zipper Kinase
MMP10	Matrix Metalloproteinase 10

NPs	Nanoparticles
OS	Overall Survival
OD	Optical Density
OSCC	Oral Squamous Cell Carcinoma
PI3K-Akt	Phosphatidylinositol 3-Kinase-Akt
PCR	Polymerase Chain Reaction
PP	postoperative CRC patients
PBS	Phosphate-Buffered Saline
PCDHGC3	Protocadherin Gamma Subfamily C, 3
PFS	Progression-Free Survival
POP1	Processing of Precursor 1
qRT-PCR	Quantitative Reverse Transcription Polymerase Chain Reaction
ROC	Receiver Operating Characteristic
RPX	Risk Prediction Index
SCFAs	Short-chain fatty acids
SMARCA5	Actin-Dependent Regulator of Chromatin, Subfamily A, Member 5
Th17	T Helper 17
TME	Tumor Microenvironment
TGF-β	Transforming Growth Factor Beta
TIMP1	Tissue Inhibitor of Metalloproteinases 1
TNM	Tumor Node Metastasis
TIMP-1	Tissue Inhibitor of Metalloproteinase-1
VEGF	Vascular Endothelial Growth Factor
VPS53	Vacuolar Protein Sorting 53 Homolog
YM	Yield Monitoring

PEMBINAAN MODEL EKSPRESI GEN TENTANG HUBUNGAN ANTARA *CANDIDA ALBICANS* DAN KANSER USUS

ABSTRAK

Kajian dasar ini adalah melalui penerokaan paradoks samaada *Candida albicans* (*C. albicans*) menggalakkan atau menghalang pertumbuhan Kanser Kolorektal (CRC) dengan memfokuskan hujah yang berkaitan dengan metabolit. Analisis pembezaan gen yang dimulai dengan pembinaan satu model gen prognostik yang berkaitan dengan perbezaan gen yang disahkan secara luaran dan dalaman oleh maklumat pangkalan data biologi. Kesan metabolit pada daya maju sel CRC telah dinilai menggunakan kaedah CCK-8 pada selang masa (12, 24, 48 dan 72 jam) dan kepekatan (OD_{600} lebih kurang 0.2, 0.3, 0.4), dengan menyaringkan masa optima dan kepekatan pendorong perosak oleh metabolit. qRT-PCR digunakan untuk mengesan tahap ekspresi mRNA gen prognostik *EHD4, LIME1, GADD45B, TIMP1*, and *FDFT1*(Farnesyl-diphosphate farnesyltransferase 1) dalam setiap kumpulan sel CRC. Keupayaan pencerobohan dan penghijrahan sel telah dinilai melalui eksperimen Transwell. Akhir sekali, kandungan Extracellular ATP (eATP) ditentukan dengan menggunakan kit pengesanan ATP. Sebanyak 213 gen yang nyatakan secara berbeza telah dikenalpasti dalam kajian ini. Model prognostik telah dibina mengandungi 5 penanda mRNA khusus iaitu *EHD4, LIME1, GADD45B, TIMP1*, dan *FDFT1*. Metabolit *C. albicans* didapati telah mengurangkan aktiviti sel CRC. Keputusan

qPT-PCR menunjukkan gen *LIME* dan *EHD4* telah menurun dalam sel CRC berbanding dengan sel epitelium kolon biasa, manakala tahap ekpresi *FDFT1* meningkat secara ketara. Tahap ekspresi gen *TIMP1* meningkat secara ketara dalam sel HT29, manakala menurun secara ketara dalam sel HCT116. Tambahan pula, analisis pasca intervensi menunjukkan penurunan ketara dalam tahap ekspresi gen dalam sel HT29, manakala ekspresi *TIMP1, EHD4*, dan *GADD45B* meningkat dalam sel HCT116, dan *LIME* dan sel-sel CRC menunjukkan penurunan ekspresi secara sepadan. Dalam sel epitelium kolon normal NCM460, tahap ekspresi gen *GADD45B, TIMP1*, dan *FDFT1* meningkat secara ketara, manakala tahap ekspresi *LIME* dan *EHD4* menurun secara ketara. Selepas metabolit intervensi, keupayaan pencerobohan dan penghijrahan sel NCM460, HT29 dan sel HCT116 menurun. Selain itu, pengukuran kuantitatif tahap eATP selepas intervensi menunjukkan peningkatan ketara ($P<0.01$). Metabolit *C. albicans* memainkan peranan sebagai pelindung dalam kejadian dan perkembangan CRC, mempamerkan interaksi dinamik dengan energetik selular.

CONSTRUCTION OF GENE EXPRESSION MODEL ON RELATIONSHIP BETWEEN *CANDIDA ALBICANS* AND COLORECTAL CANCER

ABSTRACT

This study primarily explores the paradox of whether *Candida albicans* (*C. albicans*) promotes or inhibits the development of Colorectal cancer (CRC), focusing on its metabolites mixture for relevant arguments. Differential gene analysis was initially employed to construct a model of differential gene-related prognostic genes (mRNAs) and was internally and externally validated within biological information databases. The impact of metabolites mixture on CRC cell viability was assessed utilizing the CCK-8 assay at different time intervals (12, 24, 48, and 72 hours) and concentrations (OD_{600} approximately 0.2, 0.3, 0.4). qRT-PCR was employed to assess the mRNA expression levels of prognostic genes *EHD4, LIME1, GADD45B, TIMP1*, and *FDFT1* in each group of CRC cells. Cell invasion and migration capabilities were evaluated through Transwell experiments. Finally, the eATP content was determined using an ATP detection kit. This study identified a total of 213 differentially expressed genes. A prognostic model containing 5 specific mRNA markers, namely *EHD4, LIME1, GADD45B, TIMP1*, and *FDFT1*, was constructed. *C. albicans* metabolites mixture reduced CRC cell activity. qRT-PCR results showed that compared to normal colonic epithelial cells, *LIME* and *EHD4* were downregulated in CRC cells, while *FDFT1* expression was significantly upregulated.

Notably, the *TIMP1* gene was significantly upregulated in HT29 cells, while it was significantly downregulated in HCT116 cells. Furthermore, post-intervention analysis showed a significant decrease in gene expression levels in HT29 cells, while the expression of *TIMP1, EHD4*, and *GADD45B* increased in HCT116 cells, with *LIME* and other CRC cells showing a corresponding decrease in expression. In NCM460 normal colonic epithelial cells, the expression levels of *GADD45B, TIMP1*, and *FDFT1* genes were significantly upregulated, while the expression levels of *LIME* and *EHD4* showed a significant downward trend. After metabolite intervention, the invasion and migration capabilities of NCM460 cells, HT29 cells, and HCT116 cells decreased. Additionally, quantitative measurement of eATP levels after intervention showed a significant increase *($P<0.01$)*. The metabolites mixture of *C. albicans* play a protective role in the onset and progression of CRC, exhibiting dynamic interactions with cellular energetics.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Currently, colorectal cancer (CRC) is the third most common cancer worldwide and ranks as the second leading cause of cancer-related deaths. As of 2020, the projected global incidence of CRC encompasses 1,993,590 newly diagnosed cases, with 935,173 resultant deaths, constituting 10.0% and 9.4% of the overall incidence and mortality rates, respectively (J. Weng et al., 2022). The delineation of cancer patterns across different regions and time periods serves as a valuable compass for understanding risk factors, prevalence dynamics, and the formulation of comprehensive cancer control strategies. Our research delved into the CRC burden across 185 countries worldwide, spanning from 2020 into the envisioned landscape of 2040. In 2020 alone, an estimated 1.9 million novel instances of CRC emerged, culminating in 930,000 fatalities. Remarkably, the highest incidence rates were observed in Australia/New Zealand and Europe, peaking at 4.06×10^4 cases per 10^5 men, while the lowest rates were recorded in Africa and South Asia, registering at 4.4×10^3 cases per 10^5 women. Analogous trends were observed in mortality rates, with Eastern Europe demonstrating the highest rate at 2.02×10^4 deaths per 10^5 men, while South Asia exhibited the lowest at 2.5×10^3 deaths per 10^5 women. Forecasts suggest that the burden of CRC is poised to escalate to 3.2×10^6 new cases and 1.6×10^6 deaths by 2040, primarily concentrated in nations with high or very high human development

indices (Morgan et al., 2023). Noteworthy is the status of CRC as the second most prevalent cancer in Malaysia, often diagnosed belatedly (Schliemann et al., 2022).

Prognostic models possess the capability to anticipate the likelihood of forthcoming events for an individual patient or a population, enabling the stratification of patients based on these prognostic risks. An exemplary model demonstrates the adeptness to judiciously and dependably categorize patients into distinct prognostic risk strata (Altman, Vergouwe, Royston, & Moons, 2009). Furthermore, within the medical domain, prognostic models serve as instrumental tools for scrutinizing patient outcomes in correlation with both patient-specific and disease-related characteristics (Altman & Royston, 2000).

Adenosine triphosphate (ATP) is a high-energy phosphorylated compound that plays a crucial role in storing and releasing energy within cells, ensuring the energy supply required for various cellular activities by interconverting with ADP (Adenosine Diphosphate). Due to its ease of regeneration within cells, ATP can continuously provide energy. The ATP cycle refers to the continuous utilization of energy through ATP hydrolysis and synthesis, cycling between energy-releasing and energy-absorbing reactions. As ATP is widely utilized as an energy carrier within cells, it is often referred to as the "currency of the cell." As a pivotal biochemical constituent within the tumor microenvironment (TME), ATP has significant effects on tumor progression, with its role in promoting or inhibiting tumors depending on

its concentration and the expression of specific extracellular nucleotide enzymes and receptors on immune and cancer cells (Vultaggio-Poma, Sarti, & Di Virgilio, 2020).

Remarkably, in rats with CRC subjected to chemotherapy drug treatment along with *Lactobacillus plantarum* and *C. albicans*, a notable reduction in cancer cell volume was noted, accompanied by the nucleus displaying heightened darkness, indicative of apoptosis. Notably, serum concentrations of IFN-γ, IL-4, and TGF-β were markedly diminished compared to the control cohort. Noteworthy benefits in CRC management were observed with the administration of *Lactobacillus plantarum* and *C. albicans*, sourced from the gastrointestinal microbiota of both elderly individuals and healthy subjects (Shams, Larypoor, & Salimian, 2021). Recent discoveries suggest that *C. albicans* could potentially exert a favorable influence on CRC; however, the exact mechanism underlying this phenomenon remains elusive. Nevertheless, as of present, there is a dearth of studies delving into the impact of *C. albicans* metabolites mixture on CRC. Hence, the principal aim of this investigation is to scrutinize the potential prognostic implications of *C. albicans* metabolites mixture in CRC, along with an exploration of the interplay between *C. albicans* metabolites mixture and the mRNA associated with CRC prognosis-related genes. At the same time, the effects of *C. albicans* metabolites mixture on eATP content in CRC were investigated to evaluate cell energy homeostasis. The outcomes posit that the influence of *C. albicans* metabolites mixture on correlated mRNA could emerge as a novel focal point for both the diagnosis and therapeutic interventions in CRC.

This discovery is poised to offer invaluable insights into the realm of precise treatment methodologies and the meticulous prognostic evaluation of CRC.

1.2 Study objectives

In this investigation, a prognostic model for CRC was formulated employing *C. albicans* mRNA. Through meticulous *in vitro* experimentation, the repercussions of *C. albicans* metabolites mixture on the invasion and migration of CRC cells were elucidated. Does this influence contribute to the protective or detrimental aspects of CRC development? Effects of *C. albicans* metabolites mixture on eATP Content in CRC?

Objectives:

i. To develop a prognostic model for CRC by leveraging the intricate characteristics of *C. albicans* mRNA.

ii. To assess how *C. albicans* metabolites mixture influence the invasion and migration of CRC cells.

iii. To evaluate the effects of *C. albicans* metabolites mixture on eATP content in CRC cells.

CHAPTER 2

LITERATURE REVIEW

2.1 Colorectal cancer

Presently, CRC stands as the third most prevalent cancer globally, ranking second in cancer-related fatalities. In 2020, there were 1,931,590 new cases of CRC reported, resulting in 935,173 deaths globally, accounting for 10.0% and 9.4% of the total global cancer incidence and mortality, respectively (J. Weng et al., 2022). Early-onset CRC comprises roughly 10% of all new diagnoses within this cancer category, with an observable rise in CRC-related mortality among younger patients over the last decade (Sinicrope, 2022). In Malaysia, CRC holds the position of the second most common cancer, typically being identified at an advanced stage (Schliemann et al., 2022). According to predictions, the number of new cases of cancer worldwide would reach 22 million per year by 2030, a 75% increase since 2008. Of notable concern, this upward trajectory is notably apparent in Malaysia, where over 60% of CRC cases are identified at advanced stages, specifically stages III and IV (Su & Donnelly, 2022). According to statistical data, Malaysia exhibits an overall CRC mortality rate of 9.8 per 100,000 individuals, with a higher mortality rate among males than females. A study conducted across multiple hospitals revealed a 5-year survival rate of 48.7% for cases diagnosed in 2008-2009. Simultaneously, an investigation encompassing all state cancer registries unveiled a 5-year OS (overall survival) rate of 40.8% for cases diagnosed between 2007 and 2011 (Amir et al., 2022).

Colorectal cancer typically progresses at a slow rate and often remains asymptomatic until it reaches a substantial size, several centimeters in diameter, potentially obstructing stool passage and causing cramping, pain, or bleeding. This bleeding can present as visible blood in the stool or, less commonly, as black "tar" stool. The progression of most colorectal tumors is a multistep process involving a series of histological, morphological, and genetic alterations that accumulate over time. Globally, the incidence rate is notably high in European countries, with Hungary exceeding 45 per 100,000 and Portugal surpassing 103 per 100,000. Conversely, African countries, such as Guinea, exhibit relatively low rates, with considerable variation observed across Asia, the Americas, and Oceania. For instance, Japan's rate is relatively high at 38.5 per 100,000, whereas Bangladesh's rate is merely 3.8 per 100,000. By 2020, China alone accounted for nearly 29% (555,477 cases) of new colorectal cancer cases worldwide. The gender distribution of colorectal cancer incidence is largely consistent globally. In China, the increase in colorectal cancer incidence among men is the most significant, followed by Costa Rica and Ecuador. Similarly, the largest rise in female incidence was noted in Ecuador, followed by China and Grenada. In contrast, 11 countries, including Austria and Germany, have observed a decline in the incidence of colorectal cancer among men. For women, a decrease in CRC incidence has been reported in 13 countries, including Austria, Iceland, and Germany (Lu et al., 2021). There is substantial evidence indicating that factors such as overweight and obesity, physical inactivity, smoking, alcohol consumption, and improper dietary patterns (low in fiber, fruits,

vegetables, calcium, and dairy products, and high in red and processed meats) elevate the risk of colorectal cancer. Additionally, intestinal microbiota, age, sex, race, and socioeconomic status are known to influence colorectal cancer risk(Sawicki et al., 2021). For every 8 kg/m^2 increase in body mass index, the risk of colorectal cancer rises by 10%. A large-scale meta-analysis involving 8,091 colorectal cancer cases revealed that individuals with a family history of colorectal cancer have almost twice the average risk compared to those without such a family history (Hossain et al., 2022).

Endoscopic procedures including endoscopic mucosal excision and endoscopic submucosal dissection have been used in the early stages of CRC. However, lymph node dissection has become a crucial part of surgical therapies for advanced cases due to the increased risk of lymph node metastases. Laparoscopic and robotic surgery have replaced the conventional paradigm of open surgery. Although the safety and feasibility of laparoscopic surgery in the treatment of CRC have been confirmed by prospective studies, more research and validation are required to determine the appropriate therapeutic modalities for rectal and cross-sectional colon cancer. Moreover, achieving a comprehensive cure for CRC mandates a multidisciplinary approach (Shinji et al., 2022). In recent developments, nanoparticles (NPs) have emerged as auspicious candidates for enhancing diagnostics and therapeutics by augmenting drug targeting, solubility, and bioavailability. Notably, NPs exhibit the potential to mitigate drug toxicity through enhanced solubility and can be

meticulously engineered for malignancy-specific applications, thereby minimizing undesired side effects (Mauri et al., 2022).

As our understanding of CRC advances, it becomes increasingly apparent that the intestinal microbiota, comprised of microorganisms residing in close proximity to the rectum and colon, exerts a pivotal influence on the onset and advancement of CRC. Research indicates that metabolites mixture produced by the intestinal microbiota can influence the formation of colorectal tumors. Specific types of bacteria may impact the evolution of CRC. Dietary factors have the capacity to modulate the composition and metabolic function of the gut microbiota, thereby exerting a significant impact on CRC (Han, Zhuang, Wu, Wu, & Yang, 2020). A wealth of studies demonstrates the presence of characteristic microorganisms in the intestines of CRC patients, including Clostridium, Streptococcus, and Clostridium species. Intestinal microorganisms participate in the onset and development of CRC through various pathways, such as inducing inflammation and synthesizing genotoxins. Recent studies suggest that the occurrence of CRC is not merely a function of individual microorganisms, but rather a consequence of disrupted intestinal microbial equilibrium. Short-chain fatty acids, secondary bile acids, and various microbial metabolites mixture constitute essential constituents of the intestinal microbiota, playing pivotal roles in the etiology and progression of CRC (Han, Wu, et al., 2020).

Furthermore, the intestinal mucosa intimately interacts with diverse microbial communities comprising bacteria, bacteriophages, viruses, archaea, and fungi, collectively forming the intricate intestinal microbiota. CRC is linked to the aberrant makeup of these microbial communities, as well as alterations in the amounts of microbial metabolites mixture (such as butyrate and polyamines) obtained from diet and ambient chemical concentrations. Certain bacteria, such as pks+ *Escherichia coli or Fusobacterium nucleatum*, participate in the occurrence of CRC through different pathological mechanisms, such as inducing genetic mutations in epithelial cells and modulating the TME. Pattern recognition receptors and GPCRs(G protein-coupled receptors) are present in intestinal mucosal epithelial cells and immune cells, suggesting that their activation may be regulated by gut microbiota and metabolites mixture (Hanus et al., 2021).

2.2 *C. albicans*

Fungal pathogens like *C. albicans* exhibit widespread distribution and have the propensity to compromise integumentary and mucosal barriers, thereby posing a risk of systemic infections. *C. albicans* stands as the predominant pathogen responsible for mucosal and systemic infections, comprising nearly 70% of fungal infections worldwide (Talapko et al., 2021). *C. albicans* disruption can result in the proliferation of fungi and the development of mucosal infections, like vulvovaginal candidiasis or oral thrush. Patients who are immunocompromised are more vulnerable to potentially fatal systemic infections (d'Enfert et al., 2021). The

importance of interactions between fungi, hosts, and microbial communities in driving *C. albicans* from commensalism to pathogenicity has been widely recognized. *C. albicans* exhibits remarkable adaptability, transitioning from a commensal microorganism to a pathogenic entity through the modulation of various virulence factors. Specifically, *C. albicans* pathogenic processes rely on its capacity to change shape and create biofilms (Pereira, dos Santos Fontenelle, De Brito, & De Morais, 2021). An often overlooked yet significant risk factor contributing to the transition of *C. albicans* from a commensal organism to a pathogenic one is the utilization of broad-spectrum antibiotics. Studies have shown that feeding mice with β-lactam antibiotics can induce a peptidoglycan storm, altering the gut from an inhibitory commensal state for *C. albicans* to one that encourages invasive growth and leads to systemic dissemination (Tan, Xu, Qiao, & Wang, 2021).

C. albicans exists in three biological stages: yeast, pseudohyphae, and hyphae. Hyphae are a critical stage in the illness process, capable of inflicting tissue damage by infiltrating mucosal epithelial cells, resulting in bloodstream infections (H. Chen, Zhou, Ren, & Cheng, 2020). Yeast cells exhibit a rounded to oval morphology, stemming from budding and nuclear division. Conversely, hyphae consist of tubular cells that remain firmly interconnected post-cytoplasmic division, lacking constriction at the division site. Pseudohyphae possess characteristics resembling both yeast and hyphae, appearing as elongated branching chains of yeast cells with constriction at septa. Yeast and hyphal forms perform important complimentary

functions in infection. In the early stages of infection, for instance, the hyphal form is necessary for tissue penetration and phagocytosis evasion, whereas the yeast form is essential for endothelial cell attachment and spread into the circulation (Chow, Pang, & Wang, 2021).

2.3 *C. albicans* and Colorectal cancer

C. albicans, an opportunistic pathogenic fungus, easily infects hosts with reduced immune function, including cancer patients. Increasing data suggests that *C. albicans* infection enhances the host's susceptibility to malignancies such are oral, gastric, and CRCs. Cancer and anticancer therapies may also influence the colonisation of *C. albicans*. *C. albicans* may produce carcinogens, cause chronic inflammation, particularly Th17 cell-mediated immunological responses, and promote cancer formation by altering the mucosal epithelium (D. Yu & Liu, 2022).

C. albicans has been shown to induce metabolic reprogramming in monocytes and macrophages, a phenomenon contingent upon the fungal morphotype. Both yeast and hyphae can stimulate the expression of genes associated with glycolysis and glutamine metabolism, although hyphal stimulation generally elicits a comparatively lower response. These findings demonstrate that metabolites mixture from *C. albicans* after inactivation may influence the occurrence and development of CRC (Pellon, Sadeghi Nasab, & Moyes, 2020). Numerous metagenomic studies suggest a potential association between microbial composition and CRC. Fifty-two individuals

recently diagnosed with adenoma/CRC and Fifty-two age-matched controls were enlisted and sampled for the study. Rectal swabs were cultured on various types of media and adequately incubated under both aerobic and anaerobic conditions. Using MALDI-ToF MS, all colony morphologies were subcultured and identified. While our investigation did not reveal any significant bacterial species linked to CRC, we were surprised to see a high percentage of the yeast *C. albicans* in cases. In our cohort, potential confounders were unrelated to CRC (personal history of CRC in a first-degree relative, appendectomy, cholecystectomy, increasing BMI, and percentage of males) and the presence of *C. albicans* (pre-existing diabetes and PPI medication) (Starý et al., 2020). Recently, in Dectin-3 knockout mice (Dectin-3-/-mice), the loss of the Dectin-3 gene significantly increased the development of CRC, and the fecal fungal load in Dectin-3-/-mice was significantly higher than that in wild-type mice. Interestingly, *in vivo* and *in vitro* experiments also demonstrated that the absence of the Dectin-3 gene impaired macrophage clearance of *C. albicans* and increased fungal load (D. Yu & Liu, 2022). The abundance of *C. albicans* notably increased in the intestines of CRC patients. Dectin-1, a C-type lectin receptor expressed across various cell types including DCs(dendritic cells), macrophages, and monocytes, recognizes β-1,3-glucan present in the cell walls of most fungi. However, the regulatory mechanism governing dectin-1 expression and its function in intestinal epithelial cells (IECs) remains elusive. Furthermore, the potential impact of *C. albicans* on IECs remains speculative. By activating the Wnt signaling pathway, *C. albicans* stimulates the proliferation of IECs, a pivotal event in CRC development.

Mice infected with *C. albicans* demonstrated activation of the Wnt pathway (Y. Wang et al., 2021). Using 16S rRNA and internal transcribed spacer amplicon sequencing, a study examined the combination of intestinal bacteria and fungi in Chinese healthy controls (HC), colorectal adenoma patients, and postoperative CRC patients (PP). Our analysis showed significant differences in β diversity among the four groups based on intestinal bacteria and fungi data. Fifty-one bacterial and eight fungal species were identified in the HC, CRA (Colorectal Adenoma), CRC, and PP groups. qPCR analysis of a large cohort of HC, CRC, CRA, and PP patients showed that the numbers of *Fusobacterium nucleatum, Bacteroides fragilis*, *C. albicans*, and *Saccharomyces cerevisiae* in the feces of CRC patients were significantly higher than those in HC patients. However, CRA and PP patients had much lower abundance levels than CRC patients. These findings support the view that *C. albicans* indeed contributes to human cancers (X. Li, Feng, & Wang, 2023).

Over the past decade, research has demonstrated that dysbiosis of intestinal bacteria, fungi, viruses, and archaea accompanies the development of colorectal tumors, and these microbial changes may be causal. The dynamic yet fundamentally stable intestinal microbiota is generally considered beneficial to health. Disruption of this ecosystem, known as dysbiosis, may therefore threaten the host's health. Beyond colorectal cancer, elevated levels of oral pathogens, including F. nucleatum, S. moorei, and Lachnoclostridium spp., have been detected in the intestinal microbiota of patients with colorectal adenomas, indicating that intestinal dysbiosis is an early

event in colorectal tumor development. Bacteria frequently depleted in colorectal cancer include several probiotics deemed beneficial, such as Streptococcus thermophilus, Streptococcus salivarius, Lactobacillus gallinarum, Clostridium butyricum, and Carnobacterium divergens. These probiotics can form co-exclusive networks with pathogens, suggesting competitive or antagonistic interactions between pathogens and probiotics (Wong & Yu, 2023). An imbalance in the intestinal microbiota can significantly impact the intestinal microenvironment. Normally, the intestinal microbiota maintains a symbiotic state with the host, but when this balance is disrupted, the overgrowth of certain harmful bacteria can trigger inflammation, abnormal immune responses, and subsequently induce cancer development. Alterations in the balance of intestinal bacteria can lead to changes in levels of intestinal microbial metabolites such as short-chain fatty acids (SCFAs), polyphenols, vitamins, tryptophan catabolites, and polyamines, which may be linked to cancer invasion and metastasis(Sánchez-Alcoholado et al., 2020). The intestinal microbiota includes Escherichia coli, Enterococcus, Bacteroidetes, and Clostridium, which can promote CRC by enhancing 1,2-dimethylhydrazine-induced aberrant crypt foci. Colorectal cancer patients exhibit an imbalanced taxonomic structure, characterized by the upregulation and downregulation of cancer-promoting bacteria (such as Bacteroidetes, Escherichia, Fusobacterium, and Porphyromonas) and potential protective bacteria (such as Roseburia), respectively (Qu et al., 2023).

It is worth noting that in induced CRC rats treated with chemotherapy drugs,

Lactobacillus plantarum, and *C. albicans* therapy, cancer cell shrinkage and significantly darkened cell nuclei, indicative of apoptosis, were observed. Compared to the control group, serum levels of *IFN-γ, IL-4,* and *TGF-β* were significantly reduced. *Lactobacillus plantarum* and *C. albicans* sourced from the gastrointestinal tract of both elderly and healthy individuals have shown significant efficacy in the treatment of CRC (Shams et al., 2021).

These findings present two contrasting viewpoints: (1) *C. albicans* indeed contributes to the occurrence of CRC; (2) *C. albicans* can indeed effectively ameliorate CRC. While our understanding of the association between the microbiota and CRC has increased in recent years, there are significant gaps in current research that must be addressed before this information can be implemented to assist patients.

2.4 Prognosis model of Colorectal cancer

A prognostic model refers to a study conducted to predict the probabilities of future outcomes such as disease recurrence, death, disability, etc., in subjects who have already been diagnosed with a certain disease. Such models are typically employed in prospective studies, such as cohort studies. They are based on the current health status of patients to forecast potential outcomes after treatment. A prognostic model, also known as a signature, serves as a predictive tool that evaluates potential treatment outcomes based on a patient's present health condition. Genetic prognostic models utilize gene expression data from patients to anticipate

disease advancement and survival rates, providing more refined and personalized insights to guide therapeutic strategies.

Jingyu Chen developed a nomogram that facilitates the prediction and identification of high-risk tumor deposit-positive patients, serving as a tool to alert surgeons and pathologists to meticulously examine surgical fields and resection specimens, thereby detecting occult tumor deposit lesions. A prognostic chart was established to aid clinicians in identifying individuals with adverse prognosis in CRC. For those in high or moderate risk groupings, extra chemotherapy and close follow-up for tumor deposit-positive patients may be more effective than radiation (J. Chen et al., 2022). Z Wu created a 4-FRL signature and identified two risk categories using the risk scores resulting from this signature. In colon cancer patients, the feature-based risk score demonstrated a stronger predictive ability for survival compared to traditional clinicopathological features. Additionally, significant differences in immune cells such as CD4+ and CD8+ T cells and macrophages were observed between the two groups (Wu et al., 2022). J Nie built a prognostic model consisting of five ferroptosis-related genes (*AKR1C1, ALOX12, FDFT1, ATP5MC3,* and *CARS1*). The area under the curve (AUC) for 1-year, 2-year, and 3-year survival rates was 0.668, 0.678, and 0.686, respectively. Significant differences in survival rates were identified between high and low-risk patients, and principal component analysis (PCA) and t-distributed stochastic neighbor embedding analysis could clearly discriminate high and low-risk patients (Nie et al., 2021). Exploring the role

of ferroptosis in CRC prognosis and uncovering relevant targets associated with ferroptosis for additional experimental and clinical studies aims to improve the clinical management of colon cancer.

2.5 Extracellular ATP

ATP and its final breakdown product, adenosine, serve as potent extracellular signaling molecules, triggering many pathological activities in the kidney by activating P2 and P1 purinergic receptors. In fact, changes in extracellular nucleotide and adenosine signaling dictate the course of inflammation and healing processes. Extracellular purines have the ability to regulate immune responses, balance inflammatory processes, and suppress the immune system (Dwyer, Kishore, & Robson, 2020). Extracellular adenosine 5'-triphosphate ATP can promote the invasion of breast cancer cells. *In vitro* invasion and migration tests show that removing *SOX9* can reduce ATP-driven invasive capabilities. Mass spectrometry and co-immunoprecipitation studies suggest a connection between *SOX9* and *JAK1*. Subsequently, the IL-6-JAK1-STAT3 signaling pathway has been shown to promote *SOX9* expression and ATP-mediated invasion. Notably, the ATP-IL-6-SOX9 signaling cascade has been demonstrated to stimulate chemoresistance in breast cancer cells. Chromatin immunoprecipitation assays have identified some potential *SOX9* target genes, among which carcinoembryonic antigen-related cell adhesion molecule 5/6 has been confirmed to mediate ATP-induced invasion, while ATP-binding cassette subfamily B member 1 and ATP-binding cassette subfamily G

member 2 mediate. Additionally, in naked mice, *SOX9* knockdown and apyrase (an ATP hydrolase) therapy of MDA-MB-231 cells result in decreased tumor growth and increased medication sensitivity (Yang et al., 2020). eATP and adenosine work as neuromodulators in the brain, regulating a variety of neuronal processes. Neuronal activity and brain traumas include ischemia and traumatic injuries upregulate these neuromodulators, which work through purinergic receptor activation. Moreover, eATP/adenosine signaling is critical in the pathophysiology of neurological diseases. Indeed, every cell type in the brain contributes to the rise of ATP/adenosine, and numerous mechanisms for this increase have been proposed (Shigetomi, Sakai, & Koizumi, 2024). An essential element of the TME, ATP controls the growth of tumors, the proliferation of cells, and the immune system's reaction to them (De Marchi, Orioli, Pegoraro, Adinolfi, & Di Virgilio, 2020). Moreover, ATP, a crucial energy molecule essential for cellular functions, can be generated within exosomes, and it has been elucidated that one of the cargoes carried by exosomes is mitochondria. These findings strengthen the prospect of exosomes communicating through mitochondrial transfer and recipient cell energetics. Accumulation of lactate and ATP in the TME may facilitate exosome entrance into cancer cells, hence boosting metastasis and helping to direct cancer cells to tumor locations (Thakur et al., 2022).

2.6 The role of different cell lines in colorectal cancer

NCM460 cells are employed to investigate the physiological functions of normal intestinal epithelial cells and their responses to carcinogens. HT29 cells have been

validated as tumor glandular cells that exhibit the morphological and functional characteristics of intestinal epithelial cells. These cells are typically utilized to study invasive, non-metastatic, relatively early or advanced stages of colorectal cancer (Ngamkham et al., 2020; Watson et al., 2020). HCT116 cells are used to examine metastatic and advanced colorectal cancer (Dikeocha, Al-Kabsi, Chiu, & Alshawsh, 2022; Sabolova, Kristian, & Kozurkova, 2020). In Malaysia, prostate cancer and colorectal cancer are the second and third most common cancers among men, while liver cancer and colorectal cancer are the second and third leading causes of cancer-related deaths. For women, lung cancer and colorectal cancer rank second and third in terms of new cases and mortality, respectively. The 5-year survival rate for colorectal cancer in Malaysia stands at 18.4% (Fernández-Tomé, Xu, Han, Hernández-Ledesma, & Xiao, 2020). Colorectal cancer is the most prevalent cancer among Malaysian men (approximately 15 per 100,000 individuals) and the second most common cancer among Malaysian women (11.1 per 100,000 individuals). Individuals who are obese or maintain a high-fat and unhealthy diet are at an elevated risk of developing colorectal cancer.

CHAPTER 3

METHODOLOGY

3.1 Experimental material

3.1.1 Database and analysis software

The database and analysis software were used in this study in Table 3.1.

Table 3.1. Database and analysis software

Database and analysis software	Website
GEO database	https://www.ncbi.nlm.nih.gov/geo/
TCGA-CRC database	https://portal.gdc.cancer.gov/
R language software	https://www.r-project.org
Online server in primer sequence	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Microbial informatics analysis tools	https://www.bioinformatics.com.cn/

3.1.2 Experimental cells and main reagents

The experimental cells and main reagents were used in this study in Table 3.2.

Table 3.2. Experimental cells and main reagents

Name	Production company	Number
NCM460 (Human Normal Colon Epithelial Cells) HT29 (Human Colon Cancer cells) HCT 116 cells (Human colon Cancer cells)	Advanced Medical & Dental Institute (AMDI), Universiti Sains Malaysia (USM)	
RPMI 1640 medium	Beijing Solaibao Technology Co., LTD	AJ30725668
MCCOY 5A medium		20230630
C. albicans	North Na Chuanglian Biotechnology Co., LTD	BNCC263676
Fetal bovine serum (FBS)	Xcel Biotechnology Co., LTD	12B308
Phosphate-Buffered Saline (PBS)	cytiva Corporation	AH30695633
Trypsin-edta Digestive Solution (free of phenol red)	Beijing Solaibao Technology Co., LTD	2312022
Penicillin-streptomycin - amphotericin B mixture (100×)		P1400
Crystal Violet Dye (1%)		G1062
4% tissue cell fixative		P1110
ATP content test kit		2311003
Cell Counting Kit-8 (CCK8)	Shanghai Yisheng Biotechnology Co., LTD	HB181114
Complementary DNA (cDNA) reverse transcription kit	Shenzhen Shangwei	1334606081

qRT-PCR kit	Biotechnology Co., LTD	1174624162
Base adhesive	CORNING	1363003

3.1.3 Main instruments and equipment

The main instruments and equipment were used in this study in Table 3.3.

Table 3.3. Main instruments and equipment

Instrument and equipment	Production company	Number
Cell incubator	Shenzhen Reward Life Technology Co., LTD	D180
Thermostatic water bath	Shenzhen Dingxin Yi experimental equipment Co., LTD	DXY-2
Enzyme-labeled instrument	Nanjing De Tie experimental Equipment Co., LTD	HBS-1096A
Fluorescent quantitative PCR instrument	ROCGENE	201901003
PCR amplification instrument 0	Hangzhou Langji Scientific Instrument Co., LTD	A30
Inverted fluorescence microscope	Sunny Optical Technology Co., LTD	ICX4I
Low temperature high speed centrifuge	Sichuan Shuke Instrument Co., LTD	TD-420

3.2 Experimental method

3.2.1 Source of dataset

The flow chart of this study is shown in Figure 3.1. Using the Gene Expression Omnibus (GEO) dataset (https://www.ncbi.nlm.nih.gov/geo/), this study obtained the gene expression

profiles altered by *C. albicans* from GSE42606. The retrieval was conducted on August 16, 2023, encompassing 25 samples infected with *C. albicans* for 4 hours and 34 samples infected for 24 hours. Simultaneously, mRNA expression profiles and clinical data were sourced from The Cancer Genome Atlas CRC Dataset (TCGA-CRC; https://portal.gdc.cancer.gov/). This retrieval, conducted on August 16, 2023, involved 44 samples of both normal and tumor samples derived from a pool of 571. The dataset was randomly divided into a training set (70%) and an internal validation set (30%) in order to build and validate the model successfully.

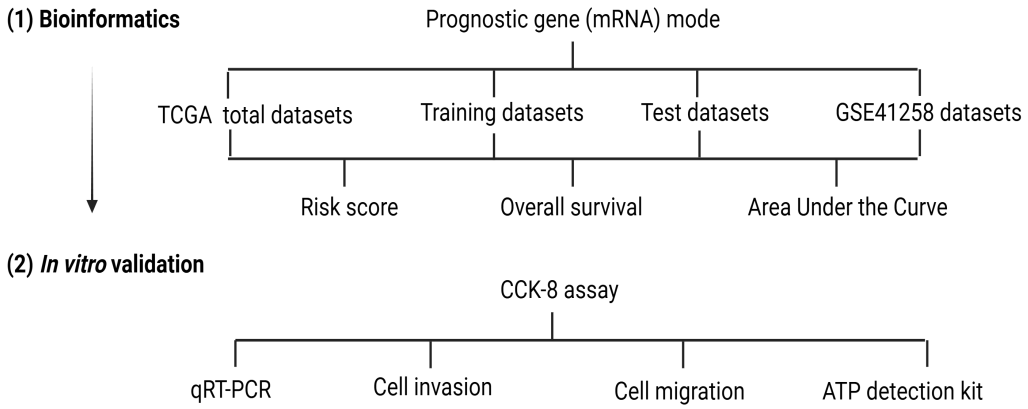

Figure 3.1 Flow Chart. This study mainly includes two parts: bioinformatics and external validation.

Furthermore, for external validation, GSE41258 (retrieved on August 16, 2023) was acquired from the GEO database, comprising 390 samples of CRC mRNA profiles along with corresponding clinical information. Throughout the analysis, meticulous attention was given to excluding samples with missing clinical data, as

well as those with a survival duration of less than 10 days, ensuring the study's reliability and precision.

3.2.2 Differential gene analysis

The limma package was used for the differential analysis of mRNA expression matrices between the 4-hour *C. albicans* samples and the 24-hour *C. albicans* infection samples (Tong, 2021). The criteria for identifying significant mRNAs were set as follows: |log2 (Full Change)| > 1 and a false discovery rate (FDR) < 0.05 (Y. Li, Ge, Peng, Li, & Li, 2022).

3.2.3 Construction and validation of mRNAs related prognostic models

In this study, R software version 4.1.0 was utilized for comprehensive data analysis. Initially, the glmnet (version 2.0.18) and survival (version 2.44.1.1) R packages were used to perform univariate Cox Proportional Hazards Model (Cox) regression, multifactor Cox regression analysis, and Less Absolute Shrinkage and Selection Operator (LASSO) regression. The evaluation of the relationship between mRNA expression levels and overall survival was made possible via univariate Cox regression, where P-values less than 0.05 were regarded as statistically significant (Jin et al., 2021).

The mRNAs that satisfied the criteria were subjected to LASSO regression analyses to refine their features. Subsequently, using multivariate Cox regression analysis, the prognostic effect and hazard ratio (HR) of the prediction model were