MOLECULAR PROFILING OF PREVENTION STRATEGIES USING RODENT MODELS OF COLORECTAL CANCER

A Dissertation

by

YING-SHIUAN CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Roderick H. Dashwood
Committee Members,	David W. Threadgill
	Arul Jayaraman
	Susanne U Mertens-Talcott
	John P. Hagan
Head of Program,	Warren E. Zimmer

May 2019

Major Subject: Medical Sciences

ABSTRACT

Colorectal cancer (CRC) is the third most lethal cancer worldwide, caused by both genetic and environmental exposures, with underlying mechanisms that dovetail genetic, epigenetic, metabolomic, and gut microbiome influences. Adenomatous polyposis coli (Apc) is a tumor suppressor and a negative regulator of Wnt/β-catenin signaling, found mutated in over 70% of CRC cases. The current dissertation used two rodent models, one genetic and one environmental, in the context of cancer prevention to study the etiology of CRC and provide preclinical mechanistic insights. The environmental model incorporated a cooked meat derived mutagen, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP), to mediate multi-organ carcinogenesis in the rat. The study focused on microRNA (miRNA) profiling in the PhIP model. miRNAs are stable negative regulators of gene targets, and play important roles in cancer and pluripotency. Rat tumor profiles and human pan-cancer datasets defined a miRNA signature of PhIP-induced multi-organ tumorigenesis. The miR-21^{high}/miR-126^{low}/miR-29c^{low}/miR-215^{low}/miR-145^{low} signature was associated with poor survival and reduced Klf4 levels, being predictive of possible environmental exposure and paradigm-shifting from 'genotoxic' to epigenetic regulation of dietary heterocyclic amines. A genetic angle was provided by the Apc-mutant polyposis in rat colon (Pirc) model, which harbors a large number of polyps in the colon, and recapitulates human familial adenomatous polyposis (FAP). The second study employed sequencing technology for mRNA, miRNA and 16S rRNA to assess the crosstalk between host, gut microbiota, and a cancer preventive diet in the Pirc model. Dietary spinach

reduced tumor outcomes significantly in the Pirc model, and reversed host genetic effects on microbiota. mRNA and miRNA analyses revealed the importance of an inflammatory response in Pirc tumorigenesis, and implicated specific miRNA-mRNA associations, such as miR-145/*Serpine1* and miR-34a/*Klf4*. The third study utilized untargeted metabolomics to investigate metabolic changes in the Pirc model along with the prevention effects of dietary spinach. The preliminary results indicated purine and lipid metabolism are important for tumorigenesis and prevention of colorectal cancer. Current technologies are providing new insights at the molecular level, incorporating 'big data' with genetic and phenotypic read-outs, to identify underlying leads associated with a designed dietary prevention strategy, which might help to mitigate the worldwide burden of CRC.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Dashwood, for the kind and invaluable mentorship throughout my Ph.D. training. I also would like to thank my committee members, Dr. Threadgill, Dr. Jayaraman, Dr. Talcott, and Dr. Hagan, for their constructive guidance, insights, time, and support throughout the course of this research.

Many thanks go to the current and former lab members, Wan-Mohaiza Dashwood, Gavin Johnson, Dr. Praveen Rajendran, Dr. Eunah Kim, Ashley Perkins, and Li Li, for providing fruitful discussion and experimental assistance. Special thanks go to Dr. Tien-Shun Yeh in National Yang-Ming University, Dr. Sheau-Yann Shieh in Academia Sinica, and Dr. Barbara Taylor in Oregon State University for mentoring my goal to receive PhD training in the US. Also, thanks to my friends in Taiwan, Oregon, and Texas Medical Center, for being my unconditional cheerleaders.

Thanks to the department program coordinator, Cindy Lewis, department graduate program faculties, Dr. David Reiner, Dr. Margie Moczygemba, Dr. Julian Hurdle, and Dr. Jiang Chang, and the IBT directors Drs. Cheryl Walker and Peter Davies for welcoming me and making my time at Texas A&M University a great experience.

Finally, I extend thanks and appreciation for my family's love and support. Thank you to my parents for their encouragement and blessing on my decision to pursue my PhD abroad. Also, thank you to my brother, Dr. BoGaun Chen, for being a role model, as a great scientist and an excellent communicator, guiding my personal and research growth.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Dashwood of the Center for Epigenetics & Disease Prevention (CEDP), Dr. Threadgill of the Department of Veterinary Pathobiology, Dr. Jayaraman of the Department of Chemical Engineering (ChemE), Dr. Talcott of the Department of Nutrition & Food Science (NFSC) at Texas A&M University, and external committee member Dr. Hagan of the Department of Neurosurgery, the University of Texas Health Science Center at Houston (UT-HSC).

Dr. John Hagan and Dr. Miriam Rose Menezes of the Department of Neurosurgery, UT-HSC were consulted on Chapter 2. Dr. Rong Wang, Dr. Christiane V. Löhr, Dr. David E. Williams, and Dr. Emily Ho of the Linus Pauling Institute, Oregon State University, and Dr. Susanne Talcott of NFS were consulted on Chapter 4, and Chapter 4 was published in 2017 in *Archive of Toxicology*. The colonoscopy training was offered by Dr. Furkan Ertem and Ms. Wan-Mohaiza Dashwood. The primary data analyzed for Chapter 5 was provided by Dr. Jia Li and Dr. Deqiang Sun of CEDP. The sequencing technology was guided by Dr. Yu Huang of CEDP. The microbiome profiling was consulted with Dr. Nadim Ajami, Baylor College of Medicine. Instrumental support and primary data analysis for Chapter 6 were provided by Dr. Rani Menon and Dr. Arul Jayaraman of ChemE and Dr. Cory Klemashevich of the Integrated Metabolomics Analysis Core in Texas A&M University. All animal work in Chapters 4, 5, and 6 could not be possible without Ms. Wan-Mohaiza Dashwood, and prior IACUC approval.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate assistant fellowship from Texas A&M University and a graduate student fellowship from Mr. Robert Hawk.

This research was supported in part by NIH Grants P01 CA090890, R01 CA122959, P30 ES00210, and ES023512, the John S. Dunn Foundation, and a Chancellor's Research Initiative.

NOMENCLATURE

ACF	Aberrant crypt foci
AFB1	Aflatoxin B1
AOM	Azoxymethane
$Apc^{Min/+}$	Multiple intestinal neoplasia in mouse
BPA	Bisphenol A
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CMS	Consensus molecular subtypes
COAD	Colon adenocarcinoma subset of TCGA database
CRC	Colorectal cancer
DDT	Dichloro-diphenyl-trichloroethane
DEGs	Differentially expressed genes
DEmiRs	Differentially expressed miRNAs
DMH	1,2-dimethylhydrazine
DSS	Dextran sodium sulfate
EGCG	Epigallocatechin-3-gallate
EMT	Epithelial-mesenchymal transition
FAP	Familial adenomatous polyposis
FMT	Fecal microbiota transplantation
GC	Gas chromatography

GI	Gastrointestinal
HCAs	Heterocyclic amines
HMDB	Human Metabolome Database
I3C	Indol-3-carbinol
ITC	Isothiocyanates
IBD	Inflammatory bowel disease
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LEfSe	LDA effect size
LncRNA	Long non-coding RNA
miRNA	micro-RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSI	Microsatellite instability
NGS	Next generation sequencing
OTA	Ochratoxin A
OTUs	Operational taxonomic units
PAHs	Polycyclic aromatic hydrocarbons
PANCAN	Pan-cancer subset of TCGA database
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PhIP	2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine

PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
Pirc	Polyposis in rat colon
PLSDA	Partial least squares discriminant analysis
RISC	RNA-induced silencing complex
RPKM	Reads per kilo base per million mapped reads
RT-PCR	Real-time quantitative reverse transcription PCR
SPI	AIN93 diet containing 10% freeze dried spinach by weight
STAMP	Statistical Analysis of Metagenomic Profiles
TCGA	The Cancer Genome Atlas

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Introduction and background of colorectal cancer (CRC)

The burden of CRC is a leading issue worldwide and has been associated with countries with a high human development index. The estimated incidence rates in males, carrying a greater risk than female, varies from <2 per 100,000 people in sub-Saharan Africa to over 60 per 100,000 people in central Europe.[1] Although incidence and mortality rates of CRC in the US are gradually declining, the overall prevalence is estimated to increase from 1,216,000 in 2010 to 1,517,000 in 2020. Medical expenditure for CRC is the second highest in cancer care, only behind female breast cancer. With a cost of $14.14 billion in 2010, the national cost of CRC care could reach $20 billion in 2020.[2] In the mission of fostering fundamental discoveries in disease etiology to improving clinical prognosis, the US National Institute of Health funded various priority areas, including CRC for $270 million in 2017.[3] Healthcare systems and the research community have made substantial efforts towards lowering the disease burden, but there are still great needs to be met for improved understanding of a complex malignancy, using advanced technologies and multidisciplinary collaborations. Notably, Welch and Robertson used incidence and mortality rates among the US population, 50 years of age or older, between 1975-2012, to demonstrate that the decline in CRC attributes to 1) improved therapies to reduce mortality, 2) early detection to avoid advanced disease, and 3) risk management, such as through diet/lifestyle changes and anticancer therapeutics, such as nonsteroidal

anti-inflammatory drugs (NSAIDs).[4] Therefore, addressing the needs in CRC has the potential for significant impact on the economic and health burden of the population.[5]

Multiple risk factors are associated with the incidence of CRC, beyond geographic variations. Some of these are generally considered to be non-modifiable, such as age, gender, and hereditary factors. Other risk factors involve environmental and lifestyle aspects that an individual can modify to reduce or increase the potential for CRC.[6] In general, men have 30% higher incidence rates of CRC than women, but lifetime risks are similar between genders (1 in 22 in men and 1 in 24 in women) due to longer life expectancy of women.[7] Diagnosis of CRC increases after the age of 40, and the incidence rate peaks between the ages of 60-79.[8] Notably, the incidence rate of those younger than 40 years old has been increasing in the past two decades. A substantial proportion of patients have a family history of CRC, and about 5% of those carry an inherited genetic abnormality. Familial CRC includes Lynch syndrome and Familial adenomatous polyposis (FAP). Lynch syndrome is at 50% risk of CRC between age 40-70 and also inherits higher risk for other cancers, both endometrial and gastrointestinal.[9] FAP has early onset of polyposis at the age of 10-12, and almost 100% at risk for CRC by age 40.[10,11] In addition, people who have other medical history, such as inflammatory bowel disease and type 2 diabetes, are more likely to develop CRC at a younger age.[12] It could be because the diseases share similar risk factors, or one disease subsequently became a risk factor for another. Environmental factors and lifestyle, including physical activities, diet, nutrition, smoking, and alcohol consumption, play important roles in CRC etiology.[13]

Thus, modifying and managing risk provides an avenue to prevent a major proportion of human malignancies due to CRC.

Symptoms of CRC vary among patients depending on size and location of the tumor. Symptoms include changing bowel habits, blood in the stool, persistent abdominal discomfort, fatigue, and weight loss, but many patients experience no symptoms in early stages. Screening tools have been established for early detection of CRC. According to the American Cancer Society Guideline for Colorectal Cancer Screening[14], either an annual stool-based test or a visual examination (endoscopy) every five years is recommended for people at risk, starting with screening at age 45. Currently, National Colorectal Cancer Roundtable (NCCRT) is seeking to reach an 80% colorectal screening rate for adults 50 and older by 2018, to lower the mortality burden of CRC in the US.[6] Diagnostic tests of CRC include blood tests related to complete blood count, liver enzymes, and tumor markers. Diagnostic colonoscopy and proctoscopy can provide biopsy samples for genetic testing, regarding mutations and genomic instability for further disease management. Additional imaging examinations can detect advance of metastasis. Tumor-node-metastasis (TNM) system defined by the American Joint Committee on Cancer is widely used for cancer staging in the clinic. The TNM system emphasizes invasiveness and spread of the tumor from stage 0 to stage IV, corresponding to the earliest intra-mucosal carcinoma (cancer-in-situ) to metastases in distant organs or peritoneum.[15]

The majority of CRC cases are ‘sporadic’, potentially tied to diet/lifestyle or environmental factors. Other cases arise from familial risk factors and inheritance, including hereditary nonpolyposis colon cancer (HNPCC/Lynch syndrome) and FAP (Fig.

I-1). The defective DNA mismatch repair (MMR) genes in HNPCC and the inactivation of *APC* gene in FAP are highly penetrant, inherited traits leading to lifetime risk of CRC between 50% to 100%.[16] The underlying dysregulations are also observed in sporadic cases, indicating the essential genetic drivers of CRC.[17,18] Depending on the genomic instability, CRC is categorized to microsatellite instability (MSI) and chromosomal instability (CIN). MSI is caused by the defective DNA mismatch repair machinery leading to hypermutations in nucleotide repeat sequences, occurring in ~15% of CRC cases. The MMR defects can be caused by germline mutation or somatic gene epigenetic silencing of MLH1, MSH2, and MSH6. On the other hand, CIN contributes to the other ~85% of cases involving point mutations and/or chromosomal rearrangements. These changes activate oncogenes, such as *KRAS*, *BRAF*, *PTEN*, and *PIK3CA*, or inactivate tumor suppressor genes, such as *APC*, *CTNNB1*, *TP53*, *SMAD4* and *TGFBR2*, via loss of heterozygosity or epigenetic silencing (an extension of the Knudson two-hit rule). According to the multi-stage carcinogenesis model by Fearon and Vogelstein in 1990[19], early *APC* mutation was associated with aberrant hyperplasia, *KRAS* mutation caused clonal expansion, and subsequent loss of suppressor genes (e.g. *TP53*) leads to later stage adenocarcinoma. The specific sequence and timing of these events has been open to debate, and adjusted on a case-by-case basis in individual patients presenting at the clinic or hospital.

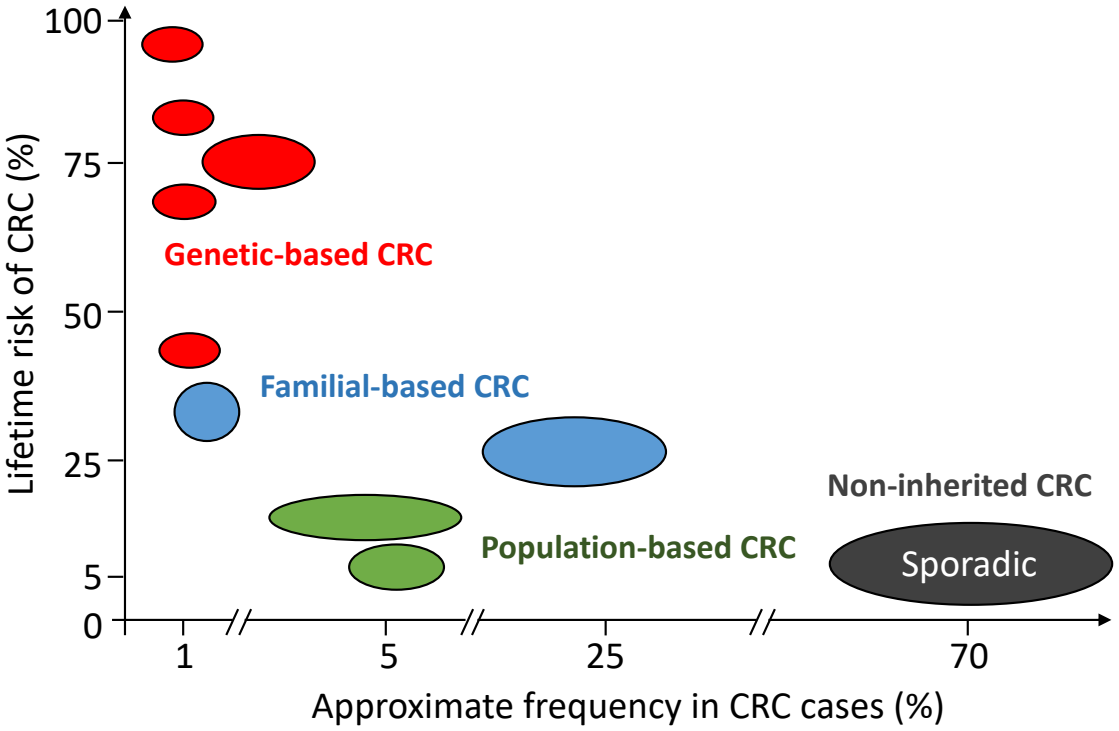


Figure I-1 Model of colon cancer risk susceptibility.
Colorectal cancer (CRC) cases are mainly divided into sporadic and familial/genetically inherited, which corresponds to high to low prevalence with relatively low to high lifetime risk. Figure is adapted from *Gastroenterology*. 2010; 138: 2044-58[16].

The heterogeneous nature of CRC has been recognized from anatomical sites, morphological phenotypes, and additional genetic and epigenetic changes beyond the relatively simple linear tumorigenesis steps originally conceived (Fig. I-2). Epigenetic regulators are emerging as key players in CRC, modulating gene expression, protein expression and biological functions.[20] The CpG island methylator phenotype (CIMP) subtype described by Feinberg and Vogelstein in 1983[21] showed overall depletion of DNA methylation and 5-methylcytosine in colorectal tissues, but increased CpG island

methylation specific promoter regions was found to silence critical tumor suppressors.[22] Cytosine hydroxymethylation (5hmC) was also globally reduced in colon tumors and resisted to cancerous hypermethylation in the promoter region.[23] microRNAs (miRNA) are small negative regulators of protein expression, stably detected in various tissues, including tumors and body fluids. Dysregulation of miR-21, miR-34a, and let-7 are associated with cell proliferation, Wnt pathway activation, and inflammation during CRC development[24]. Long-noncoding RNAs (LncRNAs), on the other hand, act as a signal, guide, sequester or scaffold for proteins. For example, CCAT1-L and HOTAIR are oncogenic LncRNAs in CRC.[25] CCAT1-L positively regulates *MYC* expression by incorporating CTCF binding[26], and HOTAIR participates in CRC progression through altered chromatin organization.[27]

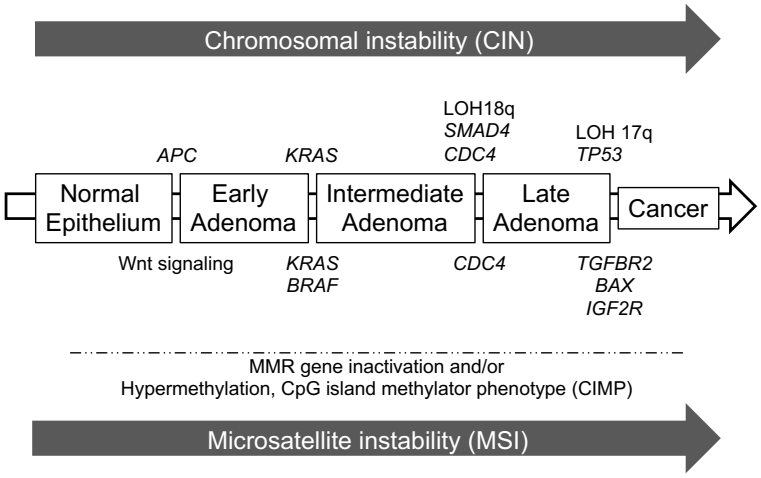


Figure I-2 Multistep sequences of colorectal adenoma-carcinoma development.
The APC mutation is believed the first step transforming normal colorectal epithelium to adenoma, and it takes years with accumulated genetic and epigenetic events to develop colorectal carcinoma. Figure is adapted from Oncol Lett. 2018; 16: 9-18[28].

Classification of CRC provides mechanistic categories for investigating causation, pathogenesis, therapeutics and prevention. Jass proposed more than a decade ago[29] five subgroups according to the molecular, clinical, and morphological features. Molecular features included MSI, CIMP, and genetic alterations in *APC*, *KRAS*, *BRAF*, and *TP53*. Clinical and morphological features, such as mucin secretion, differentiation status, invasive tumor budding, and lymphocytic infiltration, were correlated to the molecular features, implicating the underlying mechanisms. About the same time, the consensus coding sequences and the genomic landscapes of 11 human breast and 11 colorectal cancers were revealed.[30,31] The studies had more than 250,000 PCR primer pairs for protein coding exon amplification followed by Sanger sequencing. The results identified 299 and 131 validated mutations according to two cohorts. The somatic mutations are related to cellular adhesion and motility, signal transduction and transcriptional regulation, providing opportunities for further research on etiology, biomarker development, as well as personalized medicine. Oncogenic mutations and genomic rearrangements were identified by Next generation sequencing (NGS) around 2011[32], and The Cancer Genome Atlas (TCGA) Network performed a comprehensive characterization of somatic alternations in colorectal carcinoma in 2012.[33] The analyses included exome sequence, DNA copy number, promoter methylation, and messenger RNA and microRNA expression profiles, and a subset with whole genome sequencing. The report mainly focused on hypermutation of mismatch-repair genes, POLE silencing, high-frequent mutations including APC, TP53, SMAD4, PI3KCA, KRAS, novel mutations in ARID1A, SOX9, and FAM123B, copy number alternations and chromosomal translocations. The

proteomic analyses of 95 TCGA colorectal tumor specimens were performed by the Clinical Proteomic Tumour Analysis Consortium (CPTAC).[34] CPTAC identified 124,823 distinct peptides which assemble 7,526 protein groups corresponding to 7,211 genes by liquid chromatography–tandem mass spectrometry (LC-MS/MS)-based shotgun proteomic analyses. The analyses detected 796 single amino acid variants, including 152 unreported variants, and 544 accrued in 3 databases (TCGA, COSMIC and dbSNP), plus a strong copy number alteration and amplification of transcription factor protein HNF4A. The proteome proposed A-E subtypes in which type C was associated with epithelial-to-mesenchymal transition markers and leading to poor outcome. Although positive correlation of mRNA to protein was observed, poor prediction of mRNA abundances and variants to protein indicated the biological function of gene products is much more dynamics and further quantification of all biological stages are needed.[35,36] Soon after multiple efforts toward colorectal classification, Guinney et al. consolidated four consensus molecular subtypes (CMS1-4) from 17 gene expression microarray and TCGA RNA sequencing data sets.[37] CMS1(14%) is characterized by MSI and immune activation. CMS2(37%) is representing the classic colorectal carcinogenesis with upregulation of Wnt pathway and c-Myc targets. CMS3(13%) displayed metabolism adaptation and KRAS activating mutations. CMS4(23%) carries epithelial-mesenchymal transition (EMT), TGF-β signaling, stroma remodeling and invasive features. Mixed (13%) or non-consensus reflects the potentials of phenotype transition or intra-tumor heterogeneity. The consensus subtypes were also compared to copy number, DNA methylation, proteome, miRNA profile, and clinicopathological parameters and the study found the worse overall

survival in CMS4 patients which was consistent with the previous report by Calon et al., emphasizing the importance of stroma cells and microenvironment in CRC etiology[38].

Three primary treatments for CRC are surgery, chemotherapy, and radiation. Surgery is useful to remove minimally invasive lesions and also to remove affected colon and lymph node when cancer invasion occurs to the colon wall. Chemotherapy reduces the risk of cancer spread, and radiation therapy helps to shrink large tumors before surgery. Major chemotherapy regimens for CRC incorporate 5-fluorouracil and capecitabine for inhibition of thymidylate synthase[39], and oxaliplatin and irinotecan to prevent DNA replication and synthesis.[40,41] Additionally, several target therapies have been developed for treating advanced cases. For example, antiangiogenic agents are currently used for metastatic CRC. Bevacizumab (Avastin), Ziv-aflibercept (Zaltrap), and Ramucirumab (Cyramza) are antibodies to target VEGF ligands, and Regorafenib (Stivarga) is a receptor tyrosine kinase inhibitor.[42] Patients with wild-type RAS (both KRAS and NRAS) benefit from anti-EGFR (epidermal growth factor receptor) drugs, such as Cetuximab (Erbitux) and Panitumumab (Vectibix), to block the growth and proliferation of cancer cells. Immunotherapies, such as pembrolizumab (Keytruda) and nivolumab (Opdivo)[43–45], are indicated to re-activate the immune system to attack cancer cells for MSI-High or defected MMR patients.

Because it takes several years to develop, and many of the factors are modifiable, there are numerous clinical trials targeting CRC with therapeutic drugs and dietary components for general and high-risk populations.[46,47] According to *Clinicaltrials.gov*, many of the active trials use NSAIDs, such as cyclooxygenase (COX) inhibitors and

aspirin, to evaluate efficacy, while dietary agents like omega-3 fatty acids, folic acid, vitamins, antioxidants, and probiotic bacteria also are undergoing clinical evaluation.

Multiple rodent models have been established to study CRC in rats and mice, from genetic and carcinogen-induced perspectives. Genetic models emphasize *Apc* mutation, including $Apc^{Min/+}$ mice and polyposis in rat colon (Pirc) (Fig. I-3).[48] Chemical carcinogenesis models are characterized by early mutation events in DNA.[49] Methylazoxymethanol, 1,2-dimethylhydrazine (DMH) and azoxymethane (AOM) cause G to A transition. Heterocyclic amines (HCAs), such as 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) and 2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ), leads to G to T transversion or G deletion. Additionally, inflammatory agents, such as dextran sodium sulfate (DSS) or 2,4,6-trinitrobenzene sulfonic acid[50], strongly promote colorectal carcinogenesis in rodents in conjunction with genetic or carcinogen-induced drivers.[51,52] Two such models were selected for use in this dissertation, namely, the PhIP/high-fat diet and Pirc models. The PhIP/high-fat diet model was designated as 'environmental' and the Pirc model as 'genetic'. Compared to $Apc^{Min/+}$ mice, Pirc has a prevalent colon tumor yield, plus some polyps in the duodenum, and more closely mimics human FAP.[5348] The Pirc model also has a longer life span, suitable for intervention studies. An added feature of the dissertation project is the application of a murine colonoscopy procedure to monitor disease progression, tumor suppression, and resection of polyps (polypectomy) for sequential tracking of molecular biomarkers.[54,55]

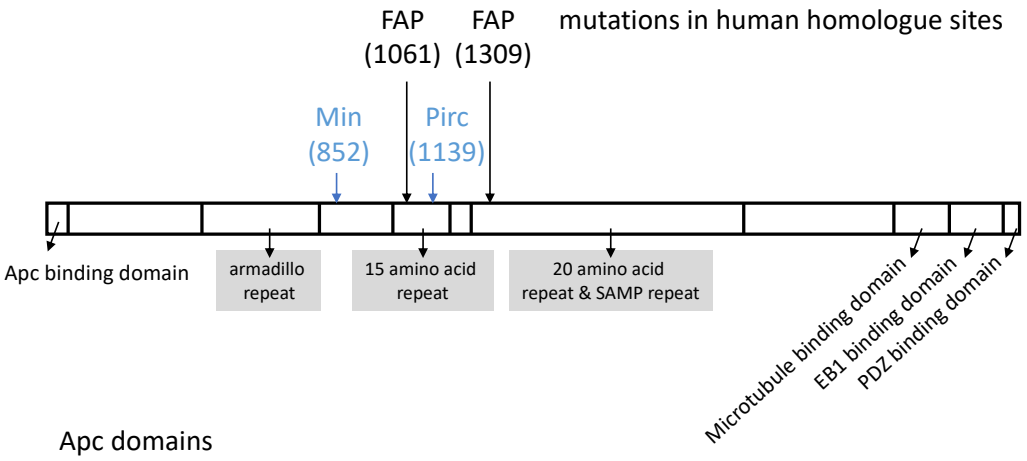


Figure I-3 APC protein and mutations that correlate with disease severity.
Protein domains of the protein are indicated in colored bars, and the lower color bar indicates the correlation of protein truncation sites to tumor multiplicity in human patients. Top arrows indicate orthologous mutation sites of rodent model of CRC. FAP, familial adenomatous polyposis; Min, multiple intestinal neoplasia; Pirc, polyposis in rat colon. Figure is adapted from Adv Exp Med Biol. 2009; 656: 85-106[56].

Scope

The main focus of the dissertation is to understand the molecule profiles of tumors in both environmental (PhIP/high-fat diet) and genetic (Pirc) models of CRC, and to clarify the key drivers of tumor suppression by dietary spinach. The dissertation research utilizes real-time quantitative reverse transcription PCR (RT-PCR), human TCGA database mining, RNA sequencing, small RNA sequencing, 16S rRNA sequencing, and untargeted metabolomics. The study in PhIP-treated rats concentrated on pre-selected miRNAs and archival tumor tissue samples, whereas the Pirc experiments were conducted by candidate, during the course of the research, incorporating unbiased profiling/omic technologies. Both studies used male rats, which are more susceptible to tumor formation than female, and group numbers that were based on published or in-house data and the expected tumor outcomes. Analysis of efficacy (prevention) considered multiple variables, such as genetic background, dietary intervention, duration of intervention, and tissue types, but housing and littermate factors were not included as variables. As detailed in the individual chapters that follow, molecular profiling identified potential new leads for future validation and mechanistic investigation.

CHAPTER II

MICRORNAS AND ENVIRONMENTAL EXPOSURE *IN VIVO*: CARCINOGENS AND PREVENTIVE AGENTS IN ANIMAL MODELS AND HUMAN SUBJECTS

Overview

MicroRNA (miRNAs) are stable short oligonucleotides found in various tissues and body fluids, and are functional negative regulators of multiple mRNA targets. Although cell-based assays have demonstrated the underlying mechanisms of (de)regulation in cancer etiology, including the impact of diet and environmental agents, few investigations have corroborated the implicated molecular phenotypes in preclinical models or in human subjects. Reports have accrued in the literature on miRNA changes from environmental chemical exposures, and on miRNAs as potential mediators of the anti-cancer effects of dietary interventions. We review the evidence *in vivo* on the use of miRNAs, or miRNAs 'signatures', as biomarkers of environmental carcinogenesis and prevention. The scope of work encompasses mycotoxins, endocrine disruptors, foodborne mutagens/carcinogens, and heavy metals among agents commonly associated with cancer and chronic disease, as well as preventive agents such as curcumin and polyphenols from dietary intakes, beverages, and whole foods. Many of the affected miRNAs are related to deregulated disease mechanisms or anti-inflammatory properties, implicating a role for miRNAs in prevention outcomes by dietary agents. However, cause-and-effect remain open to debate in the vast majority of such reports. Due to the variety of target tissues and different criteria for determining exposures, there is limited consistency among the

miRNAs prioritized and the dietary exposures in animal and human studies. Future work assessing potential tissue-specific miRNA 'signatures' altered during disease progression, and the dynamic changes in miRNAs with specific environmental/dietary exposures, should be conducted in a more concerted manner, assisted by omics and bioinformatics-based approaches. This should advance the field by prioritizing biomarkers for further evaluation in the clinical setting, as well as delineating new prevention and therapeutic strategies

Introduction

Environmental exposure can be viewed, broadly, as a mixture of detrimental and beneficial influences that have significant impacts on human health.[57] Detrimental exposures typically encompass airborne mutagens, pesticides, endocrine disruptors, food borne carcinogens, heavy metals, and other contaminants in food or water. These hazards have been implicated in metabolic disorders, reproductive system abnormalities, degenerative diseases, and cancers[58–60]. Conversely, dietary intake of fruits, vegetables, and other plant-based foods ('botanicals') offer the promise of protection against disease development through their constituent nutrients, vitamins, fiber, phytochemicals, and other bioactive components.[61–63] Here, we focus on dietary exposure as an important 'environmental' influence on human lifespan. In 1981, Doll and Peto[64] estimated that 35% of cancer incidence could be attributed to diet. Alternatively, in 2016, Blot and Taron proposed a dietary contribution of 20%[65], while Song and Giovannucci emphasized that the effect of early life malnutrition on later cancer risk may require a closer examination.[66]

Diet and nutrition also contribute to risks of other major chronic conditions, including obesity, diabetes, and cardiovascular diseases.[67] Therefore, numerous studies have attempted to investigate the impacts of our living environment through dietary exposure.

It is now well accepted that most common human diseases involve interactions, or 'crosstalk', between genes and environment.[68] Starting from four decades ago, studies in environmental health focused mainly on aspects of genotoxicity and mutagenicity, leading to a *genetic* perspective in risk assessment.[69,70] Emerging evidence in *epigenetic* research suggests that environmental influences exert consequences beyond gene mutations, chromosome aberrations, and DNA damage responses.[71,72] Epigenetic regulations are essential processes in physiological development, and also participate in pathological progression.[73] Importantly, the epigenetic mechanisms are potentially reversible, unlike the genetic changes affecting DNA sequence. Cancer and other pathologies involve deregulation of histone modifications, DNA methylation states, and non-coding RNAs with key roles in gene regulation and homeostasis. MicroRNAs (miRNAs), a class of small non-coding RNAs, are post-transcriptional regulators that were first identified through developmental studies using the nematode *C. elegans* in 1993[74]. Thousands of mammalian miRNAs have now been identified, where they are estimated to regulate 30% to 60% of coding genes.[75,76] Although many miRNAs are conserved in terms of their sequences and central mechanisms across phylogeny, their expression and regulation are highly context- and tissue-specific. During the past 15 years, the roles of mammalian miRNAs have been reported in development, proliferation, apoptosis, differentiation, pluripotency, oncology, and other pathophysiological conditions. Notably, miRNAs are

gaining attention as potential early biomarkers in epidemiological studies and clinical applications for the following reasons: (i) miRNA profiling can distinguish normal and pathological lesions; (ii) miRNAs can classify different cancer subtypes; (iii) miRNAs provide diagnostic and prognostic insights[77]; (iv) miRNAs are relatively stable molecules; and (v) miRNAs can be detected in fresh tissues, blood, body fluid,[78] and archived formalin-fixed paraffin-embedded (FFPE) tissues.[79]

In previous reviews on the topic, the primary focus was on miRNA biogenesis and results from cell-based assays, seeking to elucidate molecular mechanisms of miRNA's, sometimes under the influence of environmental[80,81] or nutritional/dietary agents.[82,83] Here, we highlight findings from preclinical models and from human subjects, in which miRNA changes were examined in the context of exposures to dietary carcinogens or preventive agents.

MicroRNA biogenesis and decay

MiRNAs are a class of small, non-coding RNAs that act as post-transcriptional negative regulators of coding gene expression. MiRNA was first described in *C. elegans* as lin-4 RNA, a negative regulator of lin-14 protein translation during temporal post-embryonic development.[74] Subsequently, thousands of miRNAs have been identified in metazoans and plants, and their sequences and the conserved functional mechanisms of gene regulation have emerged.[84] There are 439 mature miRNAs in *C. elegans*, 2,654 miRNAs in *H. sapiens*, 1,978 in *M. musculus*, and 764 in *R. norvegicus* (March 12, 2018,

www.mirbase.org). During the past two decades, we have learned the significance of miRNAs in cellular processes, embryonic development, and disease progression.

The canonical biogenesis of miRNAs (Fig. II-1) is initiated from the transcription of independent genes or introns of coding genes by RNA polymerase II (Pol II) in the nucleus, where the primary microRNA (pri-miRNA) transcript can be an independent unit or a part of its host gene.[85] The pri-miRNAs contain a characteristic hairpin structure that is cleaved by Microprocessor complex to yield the precursor miRNA (pre-miRNA),[86,87] The Microprocessor complex consists of double-stranded binding protein DGCR8, double-stranded RNase III enzyme DROSHA, and other cofactors.[88] Methylation of N6-adenosine (m^6A) of pri-miRNA, a post-transcriptional modification, was found to be recognized by DGCR8 and promotes DROSHA recruitment.[89] A ribonucleoprotein, HNRNPA2B1, was also found to bind to the methylated mark and facilitated pri-miRNA processing.[90] The cleaved pre-miRNA is then exported to the cytoplasm by exportin 5 (XPO5) protein[91–93] and further processed by RNase III enzyme Dicer[94] and transactivation-responsive RNA binding protein TARBP2.[95] One strand of the mature miRNA duplex, with 2 nucleotide 3'overhangs, gets degraded and one is loaded to Argonaute (AGO) protein to assemble the miRNA-induced silencing complex (miRISC).[96] In addition to the canonical pathway, Drosha-independent,[97] Dicer-independent,[98] and terminal uridylyl transferase (TUTase)-dependent pathways[99] also contribute to miRNA maturation. The asymmetric miRISC complex undergoes RNA cleavage or translational repression of its target RNA, depending on the pairing complementarity.[100] MiRNAs can target coding sequence, 3'UTR, and 5'UTR with

bulged or mismatched seed interactions and specific non-seed base pairing.[101–103] Interestingly, the m^{6}A marks on RNA were detected across coding and non-coding RNA regions by m^{6}A specific methylated RNA immunoprecipitation (MeRIP)-seq, and the marks were enriched near stop codons and 3'UTR of coding genes. The study also showed an inverse correlation between m^{6}A marks and miRNA-binding sites within 3' UTRs, suggesting an epigenetic regulation involving miRNA targeting events.[104] Deadenylation, decapping, and exonucleolytic decay of the target RNA are commonly required to achieve gene silencing.[105] The miRNA impact on target gene output was further established through stable isotope labeling with amino acids (SILAC) in cell culture experiments by Selbach *et al.*[106] and Baek *et al.*[107] Thousands of genes were suppressed by a single miRNA alteration. However, miRNA repression of gene expression is often complex. One mRNA transcript can be targeted by multiple miRNAs,[108] and miRNA can also up-regulate protein translation[109]. The field has evolved rapidly, and we are still a long way from fully elucidating the many, complex interrelationships in miRNA-mRNA networks.

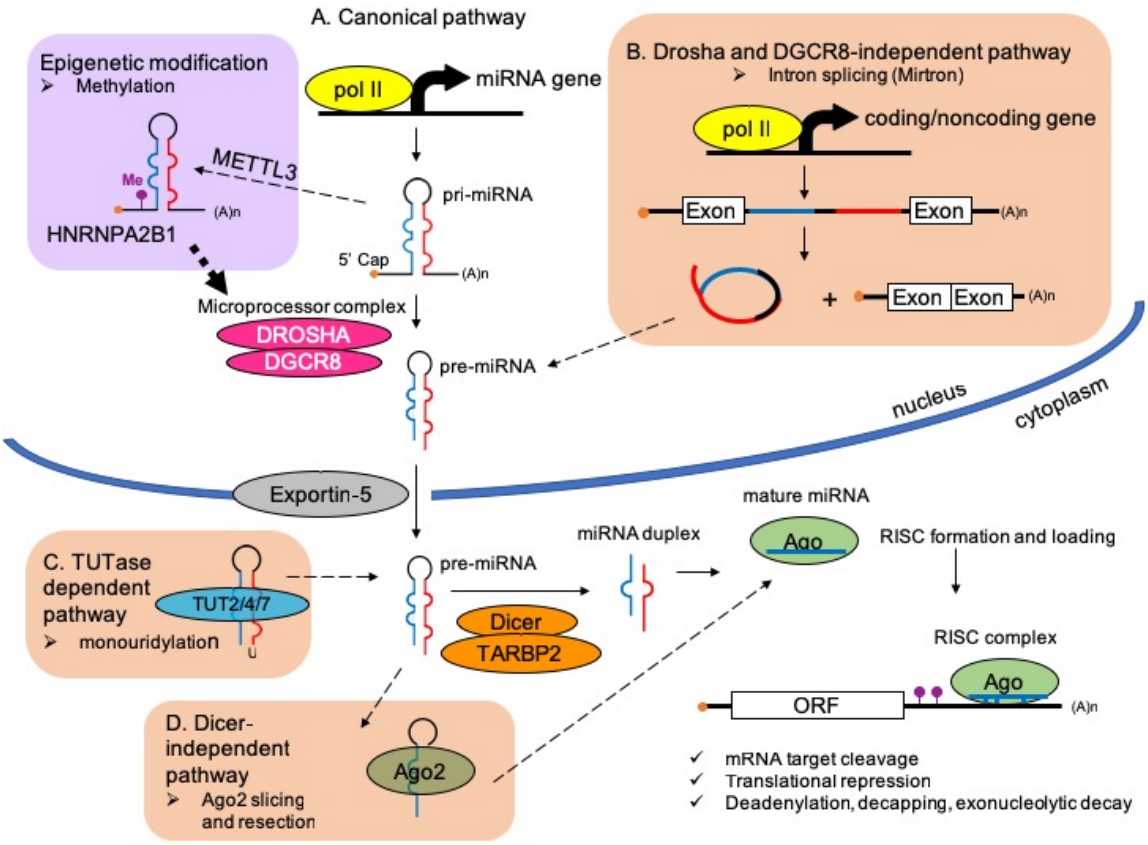


Figure II-1 miRNA biogenesis pathways and epigenetic modifications in mammals. A, Canonical pathway of miRNA biogenesis through DROSHA and DICER dependent pathway. Three examples of alternative miRNA biogenesis pathways are marked in orange and the post-transcriptional methylation of RNAs were marked in purple including facilitating pri-miRNA processing and epigenetic modifications on stop codon and 3'UTR of targeted coding RNA. B, Mirtron as an example of Drosha and DGCR8-independent pathway by generating pre-miRNA through slicing and lariat debranching from coding or non-coding RNA. C, For Group II miRNAs (let-7 family), TUTases incorporate mono-uridylation of pre-miRNA to generate 1-nt 3' overhang (TUT4 and TUT7 utilized UTP, while TENT2 could use UTP, GTP, or ATP) for further Dicer processing. D, Ago2 protein slices and resects pre-miRNA in Dicer-independent pathway. ORF, open reading frame.

Defective microRNA biogenesis in cancer

Cancer is the second-leading cause of death in the US.[110] The first cancer-related miRNAs were identified in the majority of B cell chronic lymphocytic leukemia patients that were located on a minimally deleted region of 13q14, carrying two non-coding RNAs. Loss of miR-15 and miR-16[111,112] were inversely correlated with the increased anti-apoptotic *BCL2* gene expressison[113]. Through a comprehensive analysis of 334 tumor samples across 11 tumor types, Lu *et al.* demonstrated significant changes of miRNAs in human and murine cancer, which in part reflected loss of 'suppressor' miRNAs and poor differentiation in malignant cells.[77] To test the causality of the global effects of miRNAs in cancer, Kumar *et al.* demonstrated that disruption of miRNA processing proteins, including DGCR8, DROSHA and DICER1, increased oncoprotein expression in a mouse lung adenocarcinoma cell line. Conditional knockout of *Dicer1* promoted *in vivo* tumor burden in a *Kras* mutated mouse lung cancer model.[114] Loss of *Dicer1* induced tumor development and reduced survival rates. Further investigation suggested that *Dicer1* is a haplosufficient tumor suppressor, linked to severe tumor development and worsened survival rates in heterozygous than homozygous conditioned mice.[115] Diverse gene mutations of miRNA biogenesis proteins and dysregulation of the biogenesis pathways in human cancer were identified.[116,117]

Although some human cancer-cell-based experiments and analyses *in silico* indicated that environmental toxins might affect miRNA biogenesis proteins, the current literature remains unclear on the subject[118,119]. Moreover, very few rodent preclinical studies have demonstrated a clear and unambiguous link between environmental exposure

and dysregulation of miRNA biogenesis pathways. One example is the altered expression of *Dicer1* and *Drosha,* suppressed significantly by Orchratoxin A (OTA) in a dose-dependent manner in the kidney of rats[120]. A second example would be chronic dichloro-diphenyl-trichloroethane (DDT) exposure in female rats leading to a reduction in *Dicer1* expression in uterine tissues[121].

miRNAs as biomarkers

Microarray and sequencing technologies coupled with computational analysis have identified recurrent miRNA targets across various cancer types. The miR-106 family modulates *TGFBP2,* as identified in a study of 540 solid tumors,[122] and statistical analysis of over 3000 tumors from The Cancer Genome Atlas (TCGA) database,[123] which indicated the critical role of TGF-β signaling in the progression of various types of cancer. MiRNAs are altered across all types of cancer, and the profiles can be characterized by their tissue-specific dysregulation.[77,124]

Multiple studies have explored the potential of miRNA profiles as *biomarkers for diagnosis and prognosis*. MiR-155 was found to be elevated in lung cancer tissues compared to adjacent normal tissue, and was correlated with poor survival, using multivariate analysis in an adenocarcinoma and squamous cell carcinoma cohort.[125] In a study investigating colorectal cancer using two cohorts, one in the US with 84 patients and another in Hong Kong with 113 patients, high miR-21 expression was determined to be associated with advanced stage, poor prognosis, and poor therapeutic outcome after adjuvant chemotherapy.[126]

MiRNAs also can be detected in serum and plasma, which offers noninvasive biomarker assessment. The miRNAs in serum and plasma from healthy subjects were first characterized and the profile was compared to serum miRNA from patients with non-small cell lung carcinoma (NSCLC), colorectal cancer, ovarian cancer, and type 2 diabetes. Based on the analysis, the induction of miR-25 and miR-223 was verified in serum of NSCLC[127]. Resnick *et al.* also identified miRs-21, 92 and 93 overexpression in serum samples of ovarian cancer patients.[128] Later, Weber *et al.* demonstrated the successful detection of miRNAs in 12 human body fluids, which gave rise to the new era of biomarker discovery of health and disease.[78] For example, plasma miR-122 was identified as a liver injury biomarker, which correlated with liver disease alanine aminotransferase (ALT) elevation but not with muscle injury ALT level[129]. One of the specific sources of miRNA in plasma is circulating plasma macrovesicles (PMVs). Badrnya *et al.* identified decreasing PMV numbers and miR-233 level but significantly upregulated miR-29b in 20 young, apparently healthy smokers. [130]

Dietary carcinogens and miRNA

Major human carcinogenic influences include cigarette smoking, environmental/ dietary factors, and inflammation/infection.[64,131] Some of these agents possess genotoxic activity, and others induce tumor promoting effects such as DNA damage response,[132] oxidative stress, and inflammation.[133,134] Recent studies have contributed an additional layer of knowledge regarding how these carcinogens affect miRNA profiles and their involvement in cancer and other diseases. Five categorized agents are addressed in the

discussion. These include naturally-occurring mutagen, endocrine disruption chemicals, foodborne carcinogens, heavy metal contaminates, and alcohol consumption. Table II-1 presents a summary of dietary carcinogens involved in miRNA modulation in animal models or human subjects.

Mycotoxins (aflatoxins and ochratoxins). Mycotoxins are toxic secondary metabolites produced by fungi. Hundreds of mycotoxins have been identified, but only a few regularly contaminate food. Aflatoxin and ochratoxin contaminations are often found in cereals, nuts, flours, vine fruit, coffee and cocoa.[135] miR-429 is one member of the miR-200 family, which can regulate transcription factor ZEB. miR-429 was identified in a cohort that included 138 pathology-diagnosed hepatocellular carcinoma (HCC) subjects. miR-429 upregulation was associated with larger tumor size, poor survival rate, and higher aflatoxin(AF)-DNA adducts in tumor tissues.[136] In a F344 rat dosing experiment, 200 μg/kg aflatoxin B1(AFB1) exposure perturbed the expression of 36 miRNAs in liver tissue, while increased miR-429 was identified among 10 validated miRNAs. The serum miRNA profile was rather different in the same study, but it nevertheless shared the upregulation of miR-34a and miR-181 by AFB1 in both serum and liver tissue[137]. Liu *et al.* suggested a joint effect of high miR-24 and high AFB1-DNA adducts in tumor samples on disease outcome of HCC.[138] Small RNA sequencing of rat liver tissue after a three-day exposure to 1.5mg/kg AFB1 revealed changes in the miRNA profile, seed sequence, and potential novel miRNA expression. miR-34a was validated in the study as being upregulated as an oncogenic miRNA after AFB1 exposure.[139] While AF primarily causes

liver toxicity, ochratoxin A (OTA) prominently causes nephrotoxicity. Heme oxygenase-1(HO-1) was reported participating in OTA-induced toxicity in mouse kidney. Wild-type and *HO-1* knockout mice were tested with OTA exposure[140]. NQO1 and NRF2 proteins were significantly decreased by OTA in *HO-1* knockout mice, while NQO1 was also reduced by OTA in wild-type mice. On the other hand, the dysregulation of increasing p53-dependent miR-34a and pro-fibrotic/anti-apoptotic miR-21, and decreasing anti-fibrotic/tumor suppressive miR-29b in *HO-1* knockout mouse can be further exacerbated by OTA, suggesting a role of OTA antioxidant pathways through miRNA modulation. Similarly, in a 26-week OTA exposure of F344 rats, *Nrf2* reduction and *Keap1* induction were observed in kidney, which indicated the involvement of oxidative stress and detoxification pathways.[120] In the latter study, miRNA biogenesis genes *Drosha* and *Dicer1* were downregulated, and further led to the suppression of most miRNAs. The study illustrated the deleterious effect of OTA on miRNA maturation, which may be a novel mechanism of OTA toxicity.

Endocrine disruption chemicals (BPA and DDT). Endocrine disruption chemicals (EDCs) originate from both natural and man-made sources. EDCs can interfere with diverse pathways in the endocrine system, reproductive system, neurological and immunological development of humans and wildlife. Common substances in our environment include solvent byproducts, pesticides, and plasticizers such as dioxin, dichlorodiphenyltrichloroethane and bisphenol A. Dichloro-diphenyl-trichloroethane (DDT) and bisphenol A (BPA) are found in ground water, vegetables grown under

contaminated soil, and meat coming from the top of the food chain. In the placental tissue of pregnant women, miR-146a was overexpressed and significantly correlated with placental BPA level.[141] In a Northern Swedish population, a blood leucocyte miRNA signature, including 193a-3p, miR-152, miR-31-5p, and miR-34a-5p, was positively associated with persistent organic pollutants (POPs) levels, which were calculated from measurements of polychlorinated biphenyls (PCBs), hexachlorobenzene (HCB) and DDT metabolites in serum. The study suggested that the miRNA signature crosstalks with oncogenes, such as *MYC*, *CCND1*, *BCL2*, and *VEGFA*, and further implied a connection between miRNAs and human cancer[142] through p53 and Wnt pathways.

In addition to the exposure of EDCs observed in the reproductive system and blood, research has also demonstrated the toxic effects in local organs. In a male Wistar rat oral gavage study, BPA induced miR-122 expression through activation of JNK, ERK1/2, MAPK, and Akt pathways. The induced miR-122 was further predicted to interfere lipid metabolism and liver homeostasis.[143] BPA also has been demonstrated to influence miR-338 through different upstream regulators, *Gpr30* and *Glp1r*, in mouse pancreatic islets following short-term and long-term exposures, respectively. By targeting the *Pdx1* gene, the dysregulated miR-338 interrupted the compensation of glucose-stimulated insulin secretion and ATP production.[144] On the other hand, in liver tissues, intraperitoneally (i.p.) administered DDT to female Wistar rats increased *Cyp2b1* gene and protein expression as well as CYP substrates after 72-hour exposure. Reductions of miR-21, 221, 222, and 429, were predicted to target *Cyp2b1* gene, which also were observed in DDT-treated liver

tissues. The results suggested the involvement of miRNAs in DDT-induced CYP enzymes in liver.[145]

Foodborne mutagens (PAH and HCA). Polycyclic aromatic hydrocarbons (PAHs) and heterocyclic amines (HCAs) are chemicals found in high-temperature cooked meat[146]. The metabolites of PAHs and HCAs can form DNA-adducts which contribute to their mutagenic activity and DNA damage responses.[147,148] Several epidemiologic studies have illustrated that meat consumption and cooking methods are associated with increased risks of cancers[149,150]. PAHs also may be airborne pollutants. By measuring urinary monohydroxy-PAHs and plasma benzo[a]pyrene (BaP)-albumin adduct in coke plant workers, Deng et el. demonstrated an inverse association between the increasing plasma albumin adducts and decreasing miR-24-3p, miR-27a-3p, miR-142-5p, and miR-28-5p levels. On the contrary, miR-150-5p was increased and positively associated with urinary monohydroxy-PAHs.[151] BaP is one of the most widely studied PAHs and can bind to aryl hydrocarbon receptors (AhR) to trigger downstream gene regulation[152,153]. A systematic evaluation of BaP-DNA adducts, gene expression, and miRNA profile were conducted in BaP-treated mice. BaP–DNA adduct was detected in the stomach, spleen, lung, liver, and colon, which indicated a systematic toxic effect. The gene profile indicated a tissue-specific response which demonstrated that only Ccng1, *Cyp1a1* and *Hspa8* were consistently upregulated in BaP-targeted organs. Notably, the miRNA profile was modestly changed in the study,[154] and appears to be consistent with another short-term

exposure study in male mice which indicated a lack of change in miRNA through BaP exposure in rodent models[155].

2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) is one of the most common HCAs. Rat oral gavage studies indicated a 13-miRNA signature in PhIP-induced rat colon tumors. Let-7 miRNA downregulation and the upregulation of its target genes, including c-Myc and Lin28, were identified.[156] A follow-up study which focused on the other target organs of PhIP-induced tumorigenesis in the rat identified a five miRNA 'PhIP signature' involving miR-21, miR-126, miR-29c, miR-215, and miR-145, associated with reduced *Klf4* levels in PhIP-induced tumors and correlated with a subgroup of human cancer patients.[157] Assigning this subgroup as potential HCA consumers was not feasible.

Heavy metals (As and Hg). Arsenic is a natural element but is highly toxic in its inorganic form. Contamination of groundwater by arsenic poses risks for irrigation, food preparation, and drinking water. Exposure to inorganic arsenic increases risk of skin cancer, bladder cancer, lung cancer, metabolic syndrome, cardiovascular disease and defective fetus development. When sodium arsenite was added in drinking water to male Sprague Dawley rats, 26 miRNAs were differently expressed in liver tissue[158]. The activity of hepatic glutamate-cysteine ligase (GCL) was increased by low doses but decreased by high doses of sodium arsenite exposure. Interestingly, the gene expression of *Gclc* and *Gclm*, the subunits of GCL, were affected in the same manner. miR-25 was identified as a conserved miRNA, which regulates *Gclc* and *Gclm* in rats and humans, and is responsible for arsenic-induced toxicity in the liver. In a cohort in Mexico which consisted

of 200 pregnant women, twelve miRNAs were upregulated in cord blood of newborns and were associated with in utero arsenic exposure by detecting maternal urine arsenic level. The miRNAs were predicted to regulate immune response in blood cells suggesting the miRNAs were prenatal exposure responders.[159] In a fetal growth focused birth cohort of 1,141 cases in Bangladesh, Rahman *et al.* identified the increasing miR-1290 in the placenta was significantly associated with preterm birth, lower birthweight, and high cord blood arsenic exposure in the 466 mother-infant pairs.[160]

Mercury is another natural element that possesses toxicity through vapor exposure, mercury-containing products, and food source contamination. For example, mercury-containing dental material, plants grown in contaminated soil, and accumulated toxins in the food chain, particularly in seafood, enter the body through unintentional or dietary exposure. Mercury can cause cardiomyopathy, bowel immune response, renal disease, and neurotoxicity.[161] Plasma miR-92a-3p and miR-486-5p are upregulated in mercury exposed workers, and the two miRNAs can suppress the expression of *Klf4* and *Cezanne* to further sustain NFκB pathway activation.[162] *In utero* exposure of environmental pollutants in the US population was linked to let-7 miRNA family downregulation, and was negatively correlated with the increasing mercury and lead exposure levels in placenta tissue.[163]

Alcohol. Alcohol consumption is associated with steatosis, hepatitis, fibrosis, and cirrhosis, which contribute to mortality in western countries[164]. miR-122 has been reported to be a highly abundant miRNA in mature hepatocytes and was inversely correlated with the prognosis of hepatocellular cancer (HCC). Liver miR-122 was suppressed in alcoholic

liver disease (ALD) patients and in an ALD mouse model.[165] Loss of miR-122 expression was mechanistically due to activation of the transcriptional repressor Grainyhead like transcription factor 2 (GRHL2) via an alternative splicing event induced by ethanol exposure. The suppression of miR-122 led to hepatic steatosis and inflammation through HIFα. In serum samples, however, miR-122 was elevated, implying miR-122 released from injured liver tissue from an investigation of recreational alcohol consumption.[166] A follow up-study from the same group used unbiased small RNA sequencing and analysis of serum miRNAs. By comparing small RNA sequencing profiles between 48-hours before drinking and 3-5 hours post-drinking in 16 healthy participants, the study found that a fifth of the total detected miRNAs, including miR-122, were upregulated, and only miR-185-5p was downregulated.[167] Additionally, a non-human primate study illustrated that chronic alcohol consumption disrupted the homeostasis of peripheral blood and colonic immunity. The results suggested the reduced expression of transcription factors, such as *Stat3* and *Arnt*, and growth factors, such as *Vegf*, *Hgf*, and *G-csf*, may be due to the dysregulation of miR-181-5p and miR-221-3p.[168]

Chemopreventive agents and miRNA

Dietary patterns, macronutrients, and phytochemicals have been reported to affect cancer risk.[169] Although the benefits of natural agents are not limited to cancer prevention, the majority of bioactive components in the food have been historically studied in the context of cancer. Phytochemicals are non-nutrient plant-derived chemicals which possess anti-oxidation, anti-inflammation, and anti-proliferation capabilities against neoplasia

stages.[170–172] The miRNA regulation *in vivo* by phytochemicals, including curcumin, resveratrol, polyphenols, and cruciferous vegetable-derived compounds, other macronutrients and whole food are discussed and listed in Table II-2.

Curcumin. Curcumin and other curcuminoid analogs, such as demethoxycurcumin and bisdemethoxycurcumin, are polyphenolic compounds found in turmeric spice (*Curcuma longa*). Although low bioavailability of curcumin has been reported,[173,174] curcumin has been demonstrated to contain anti-inflammatory, anti-oxidation, anti-microbial, and anti-cancer properties in numerous reports.[175,176] In 2012, Bao *et al.* indicated that diflourinated-curcumin, a curcumin analog, restored the expression of let-7, miR-26a, and miR-101, which are commonly lost in pancreatic cancer, and inhibited pancreatic tumor growth in cell-based assays and in a xenografted mouse model through intragastric treatment. The inhibition effects also involved downregulation of histone methyltransferase *Ezh2* and other factors, such as *Notch-1, Cd44, Epcam*, and *Nanog*.[177] Since inflammation is a critical mediator for cancer progression[178], a study combined curcumin and another anti-inflammatory agent, 3 acetyl-11-keto-β-boswellic acid (AKBA), to combat colorectal cancer tumor growth in a HCT116-cell-derived xenograft mouse model. The intraperitoneal treatments of the two agents worked together to suppress the tumor growth in mice. Induction of miR-34a and the reduction of targets such as *MYC*, *CDK6* and *CCNE1*, and the repression of miR-27a with the increasing *FBXW7* were observed in tumor growth inhibition[179]. A more recent study demonstrated the systemic effects of curcumin in the lipopolysaccharide (LPS)-induced sepsis mouse

model. Pretreatment with curcumin via oral gavage ameliorated LPS damage in the liver and kidney, with reduced serum biomarkers, including AST and BUN, decreased cytokine IL-6 and TNF-α level, and improved histology. Curcumin also inhibited the elevation of miR-155 through the PI3K/Akt pathway in LPS-induced liver and kidney inflammation.[180] miR-155 was one of the first identified miRNAs in B cell lymphoma and was derived from BIC RNA induced by avian leukosis virus.[181] miR-155 was also proven to perform a multifunctional role in cell-specific inflammatory responses. Together, the literature suggests that curcumin could modulate inflammatory conditions through miR-155. Moreover, curcumin also showed a protective effect in renal podocytes against fructose-induced metabolic syndrome, proteinuria, and insulin resistance in rats.[182] The upregulation of miR-206 and downregulation of protein tyrosine phosphatase 1B (*PTP1B*) by curcumin, with consequentially improved insulin signaling, suggested curcumin can be a potential dietary intervention to combat diabetes, metabolic syndrome, and fructose-associated podocyte injury involved in defective insulin signaling.

Resveratrol and stilbene. Resveratrol and its glucoside are bioactive stilbenoids derived from plants such as grape skin, grape seeds, mulberry[183], pistachio and peanut.[184] In a previous report on tumor inhibition by resveratrol in Transgenic Adenocarcinoma Mouse Prostate (TRAMP) mouse model through reducing IGF-1 and ERK1/2 signaling,[185] Seth *et al.* extended the mechanistic study from signaling to miRNA alterations.[186] Oral administration of resveratrol reduced tumor growth and metastasis of aggressive PC-3M-MM2 cells in immune compromised mouse models. Resveratrol reduced Akt signaling

resulting in reduced miR-21 and elevated *PDCD4* expression. The study also explored resveratrol protective effects on other prostate cancer cell lines, and the results indicated tumor suppression by resveratrol can be either miR-21 dependent or independent. Another prostate cancer study tested the effects of resveratrol and its analog, pterostilbene, on xenografted prostate cancer tumor growth. The results indicated that pterostilbene could restore the loss of *PTEN* expression in DU145 and 22Rv1 human prostate cancer cells through suppressing miR-17, miR-20a and miR-106b.[187]

Additionally, anti-inflammatory miRNAs, miR-101b and miR-455, were upregulated by resveratrol in intestinal tissues after dextran sodium sulfate-induced colitis-associated tumorigenesis in $Apc^{Min/+}$ mouse. Functionally, resveratrol mitigated intestinal epithelial proliferation, pro-inflammatory cytokines such as IL-6 and TNF-a, and modulated immune cell infiltration to lymph nodes.[188] Resveratrol also demonstrated an anti-inflammatory response through downregulation of RELA, a NF-κB subunit, resulting in reduced miR-221 transcriptionally. TRK-fused gene (*TFG*), a tumor suppressor, was identified as a miR-221 target. Peritoneal administered resveratrol suppressed xenografted melanoma via repression of RELA and miR-221 and the de-repression of *TFG.*[189]

A one-year supplementation of grape extract with enriched resveratrol was performed in type 2 diabetes patients. Downregulation of cytokine expression such as CCL3, IL-1b, and TNF-a, and modulation of miR-181b, miR-155, miR-34a and miR-30c2 were observed in peripheral blood mononuclear cells (PBMCs), which indicated an immune modulating role of resveratrol in human subjects[190].

Other polyphenols such as EGCG, quercetin, genistein, and anthocyanin. Epigallocatechin-3-gallate (EGCG) is a highly abundant tea catechin with anti-cancer potency. It can scavenge reactive oxygen species (ROS) to modulate gene expression and signaling.[191] The effects of EGCG on miRNA and mRNA expression were evaluated by microarrays in a tobacco carcinogen-induced mouse lung adenoma. The results indicated a modest upregulation of 21 miRNAs and gene networks associated with IGFBP5 and AKT/MAPK regulation.[192] miR-210 was one of the elevated tumor suppressor miRNA by EGCG in mouse lung, and was investigated in lung carcinogenesis via cell-based assays,[193] indicating the role of mi-210 was consistent in both *in vitro* and *in vivo* conditions.

Quercetin is a flavonoid found in several fruits and vegetables such as apples, plums, peppers, and onions. Quercetin has anti-inflammatory and proapoptotic properties. Epidemiological studies have used paraffin-embedded lung cancer tissues and quercetin-rich food frequency questionnaires to 264 lung cancer patients to assess the miRNA level and quercetin exposure. The study identified 33 miRNAs associated with quercetin exposure. The tumor suppressor let-7 family and pro-apoptotic miR-26 were significantly higher in the frequent quercetin consumption group.[194]

Pomegranate contains other types of polyphenols, such as ellagitannins and anthocyanins, which are precursors of anti-oxidative ellagic acids and phenolic acids. Pomegranate extracts showed anti-tumor activities through downregulation of PI3K/AKT and NF-κB pathways in lung cancer cells.[195,196] Banerjee *et al.* later demonstrated the anti-cancer effect on reducing aberrant crypt foci (ACF) by pomegranate polyphenol in

azoxymethane (AOM), a colon carcinogen, treated rat model. Drinking pomegranate juice increased miR-126 and reduced *Vcam1* and *Pik3ca* expression and further suppress PI3K/AKT and NF-κB pathway activation in an AOM-treated rat colon.[197]

Cruciferous vegetables. Cruciferous vegetables have four cross-shaped petals, and the common edible cruciferous vegetable are cabbage, broccoli, Brussels sprouts and mustard greens. They have been determined to be associated with reduced colorectal cancer risk.[198] Particularly, glucosinolate precursors and isothiocyanates (ITC), such as sulforaphane, and indole-3-carbinol (I3C) and its dimeric analog 3,3'-diindolylmethane (DIM), have been reported to possess substantial chemoprevention potential[199,200] through modulating liver metabolism enzymes, anti-oxidation machinery involving in nuclear factor E2-factor related factor (Nrf2) pathway, and histone deacetylase (HDAC) inhibition.[201] Izzotti *et al.* used oral administration of chemopreventive agents, including phenethyl isothiocyanate (PEITC) and I3C, and evaluated early carcinogenesis changes from environmental cigarette smoke (ECS) exposure. The study first identified 25 miRNAs modulated by chemopreventive agents in rat lung, while the combination of PEITC and I3C showed the most protective effect on normalizing miRNA changes, including let-7 family, miR-30 family, miR-34 family and miR-99b.[202] Another study used *i.p.* injected vinyl carbamate (VC) to induce lung tumors in A/J mice to assess the miRNA regulation by dietary administered I3C. Although the study did not reveal tumor prevention, I3C reduced 5 VC-induced miRNAs in lung tumors. The induction of miR-21 targets, *Pten*, *Pdcd4*, and *Reck* by VS were observed while miR-21 was downregulated by I3C treatment. The effect of

I3C was also recapitulated in A549 lung cancer cell lines.[203] More recently, Wang *et al.* demonstrated I3C suppressed HCC cell-based assays and xenografted tumors with intraperitoneal I3C administration. The result indicated that I3C inhibited miR-21 activated PTEN/Akt pathway both *in vitro* and *in vivo*.[204]

Macronutrients and whole foods. n-3 polyunsaturated fatty acids (PUFA), but not n-6 PUFA, and dietary fiber possess chemopreventive efficacy against colon cancer development. This work demonstrated in a series of studies using fish oil and dietary fiber miRNA changes in AOM-induced rat colon tumors.[205–207] Fish oil reduced the colon tumor incidence regardless of the combination with dietary fibers, cellulose or pectin, and selectively modulated let-7d, miR-15b, miR-107, miR-191 and miR-324- 5p in the rat colon. The second study focused on the upregulation of miR-26b and miR-203 after fish oil and pectin combination. An inverse correlation of the miRNAs and the reduced target genes *Ptk2b, Pde4b*, and *Tcf4*, supported the chemopreventive outcomes. The most recent report indicated that miR-26b and miR-203 regulation was specifically in the $Lgr5^{hi}$ stem cell population from mouse colon.

The whole food approach is also an emerging topic in cancer prevention research[208,209] due to its multi-targeting effects. Parasramka *et al.* demonstrated that dietary spinach inhibited PhIP-induced rat tumorigenesis across various organs, and normalized the miRNA alterations in PhIP-induced colon tumors. The protection included de-repression of several let-7 family members and the association of reduced pluripotency targets such as *Myc, Lin28,* and *Hmga2*.[156] Jeyabalan *et al.* demonstrated the

chemopreventive and therapeutic effects of dietary blueberries in female ACI rats with 17β-estradiol (E_2) induced breast cancer.[210] Dietary blueberries were supplied in the beginning of the study for evaluating preventive effects while therapeutic effect was supplied with blueberry after the appearance of the first palpable tumor. In the end of the 25-week study, 5% blueberry and had both preventive and therapeutic outcomes, with downregulation of *Cyp1a1* and *Esr1* gene expression. Among four estrogen specific miRNAs, significant down-regulation of miR-18a and miR-34c in the therapeutic mode of mammary tissues were observed.

Conclusion and future perspectives

MiRNAs are critical gene regulators in physiological and pathological processes, as well as being implicated in environmental exposures. This review summarized the impact of *in vivo* dietary exposures on miRNA biogenesis, and miRNA profiling with selected agents that exert detrimental or beneficial effects on human health. The literature provides evidence of miRNAs responding to external exposures, in locally affected organs or in body fluids. In addition, the influence of dietary components on miRNAs provides deeper insight into novel disease prevention mechanisms via nutritional intervention.

Our living environment is complex and dynamic, which exacerbates the difficulty of discerning the contributions of individual exposures. In addition, it remains unknown whether miRNAs are exposure-specific or disease-specific read-outs. Ideally, monitoring environmental exposure through biomarkers should enable the early detection of exposure for intervention before disease progression Developing miRNAs as biomarkers of

environmental exposure presents several challenges. First, few comprehensive and cross validated studies have been published on body fluids and disease sites in relation to dietary toxin exposure. Therefore, specificity and reproducibility are difficult to assess. Notably, few consistent results are found in the literature. For example, in a rodent aflatoxin exposure model, miR-429 was elevated by aflatoxin exposure and miR-429 was correlated with AF-adduct in human liver tumors. However, other cases appeared to be either non-specific to various exposures or study-specific findings. For instance, let-7, a tumor suppressor across various cancers[211], was upregulated by quercetin[194], curcumin[177] and spinach[156] in lung and colonic tumors indicating its non-specific roles in multiple cancer types in response to dietary exposure. Thus, detecting let-7 level may not be sufficient to reflect a specific disease or environmental exposure.

Study-specific results may be due to the complexity of content-dependent miRNA regulation. miRNAs in different sample types could have distinct meanings. For example, miR-122 was downregulated in the liver tissue of alcohol treated mice, but was elevated in human serum after recreational alcohol consumption. Therefore, one must be careful in interpretations across studies and cognizant of their reproducibility and widespread significance. Secondly, there are technical issues involved in assessing miRNAs. The biomarker source (e.g. tissue, tumor, or a particular biofluid) that serves as the miRNA starting-material varies depending upon particular research interests. Interestingly, Haider *et al.* analyzed 104 publicly available studies focusing on miRNA profiling in non-neoplastic diseases and identified significant methodology issues, including sample preparation, lysis of hematologic cells, and data normalization, in serum, plasma and

PBMC-based miRNA biomarker research[212]. The study also demonstrated a high inconsistency of results across the literature, in which non-disease-specific miRNA biomarkers were identified, such as in the case of organ injuries. For example, miR-122 was associated with liver injuries regardless of damage types, including viral infection and drug-induced toxicity. miR-370 was only identified in two independent studies of coronary artery disease and was not found in other included reports.

While the research investment regarding miRNAs as biomarkers is increasing, the studies to date suggest that a careful sample acquisition and interpretation are necessary to facilitate clinical biomarker development. Specificity and reproducibility issues in miRNAs as biomarkers have also been recognized in the cancer field.[213] Although circulating cell-free miRNAs were reported since 2008[127,214,215], a profound effect on miRNA profiles from platelets during plasma processing was addressed by Cheng *et al.*[216] For miRNA biomarker development across multi-institutional operation and validation, the necessities of protocol optimization and quality control (including data normalization) are substantial challenges to overcome.

Second, whether the miRNA changes truly reflect a specific environmental exposure remains unclear in the current literature, since the timing of environmental exposure and disease progression could be overlapped. Many studies in this review included either case-control human studies or xenograft mouse models for toxin exposure and testing health-promoting agents. There are apparent demands for longitudinal exposure studies with miRNA profiling and disease prevalence analysis in human subjects. On the other hand, rodent models with comparable human etiology are more

suitable to justify *in vivo* study for functional dissection[217,218]. Exposure studies in rodents also require consideration of the delivery method, the dose, and the duration to be relevant to human conditions. Testing prevention in various rodent backgrounds, mimicking wider human exposure scenarios, might provide more insights and allow knowledge go be translated from rodent models to medicine.[219,220] Collectively, the limitations in miRNAs as biomarkers could be tackled through in-depth study designs which are targeted towards a specific purpose, such as exposure-driven or disease mechanism-driven designs, with proper sample acquisition and standardized processing in a relevant longitudinal setting.

Third, dietary impacts on miRNAs involves layers of new knowledge. Recent reports have indicated that diet-derived miRNAs may affect human health through exosome uptake[221,222]. Gut microbiota also has been proven to play a role in bioactive compound releasing,[223,224] cross-talking with miRNA modulation,[225,226] and even interaction with environmental exposure.[227] Relative to toxin exposure measurement for risk management, dietary or phytochemical intervention for health promotion presents other difficulties in the clinic, such as long-term compliance and the single molecule mindset.[208] Moreover, the need for relevant end point measurements benefiting patients' health must be further developed and evaluated.[228] In summary, achieving dietary intervention of miRNAs or the combination with other aspects would require further interdisciplinary efforts.

Although dietary intervention and phytochemicals are proposed to be effective in combating cancer through miRNA modulation[83,229], careful evaluation is necessary. The the bioavailability of active compounds is one of the barriers to transfer the cell-based

successes to human cases. Various methods can improve the delivery,[230–232] such as derived analogues, conjugation, nanoparticle, and liposome. Also, it is challenging to determine whether a single phytochemical or a mixture, either of enriched components or whole food, would be a superior strategy. Phytochemicals possess multiple functions in addition to modulating miRNAs, including preventing genotoxicity, regulating immune response, and reducing proliferation. Ricciardiello *et al.*,[208] Hussain *et al.*,[209] and Potter[233] have all proposed a potential for whole food intervention for cancer prevention. However, without a direct comparison, their decisive benefits from one or the other remain open to debate.

Understanding the modulation of miRNA profiles through environmental exposure in circulation or in at-risk organs provides future directions for monitoring and disease prevention. Utilizing dietary components for disease intervention remains optimistic, due to modulation of miRNAs, which further regulates multiple targets and exerts other anti-cancer properties. The current review summarizes evidence of miRNA alterations from detrimental and beneficial environmental exposures linked to human health. Future investigations with longitudinal cohorts and with standardized processes and procedures will offer enhanced solutions for disease prevention strategies centered on miRNAs as targets for precision medicine.

Table II-1 miRNA modulation by dietary carcinogens.

Abbreviation:

AF, aflatoxin; BaP, Benzo[a]pyrene; BPA, bisphenol A; DRC-ICP-MS, dynamic reaction cell inductively coupled plasma mass spectrometer; ELISA, enzyme-linked immunosorbent assay; GC-MS, Gas chromatography–mass spectrometry; HCC, hepatocellular carcinoma; HG-AAS, hydride generation-atomic absorption spectrometry; Hg, mercury; OTA, ochratoxin A; PAHs, polycyclic aromatic hydrocarbons; Pb, lead; PBMC, peripheral blood mononuclear cell; PhIP, 2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine; POPs, persistent organic pollutants (POPs) level was calculated 6 polychlorinated biphenyls (PCBs), 1 hexachlorobenzene (HCB) and dichlorodiphenyltrichloroethane metabolite in serum; RT-PCR, real time PCR

Dietary toxins	Tissue source	Toxin exposure Or measurement	miRNA profiling methods	Identified miRNA (validated target)	Ref.
Mycotoxins					
AF	• 138 HCC patients • Liver tumor tissue	AFB1 DNA adduct, ELISA	RT-PCR	miR-429⇑(*Zeb*)	Huang *et al.*[136]
AF	• Male rats • Liver and serum	• 200 μg/kg/d AFB1 • 4 weeks • oral gavage	• Small RNA sequencing • RT-PCR	Liver: miR-434⇑, miR-411⇑, miR-221⇑, miR-127⇑, miR-205⇑, miR-429⇑, miR-34a⇑, miR-181c⇑, moir-200b⇑, miR-541⇑ Serum: miR-221⇑, miR-34a⇑ and miR-181c⇑	Livingstone *et al.*[137]
AF	• 207 HCC cases • Liver tumor tissue	• AFB1-DNA adducts • ELISA	• TaqMan RT-qPCR	miR-24⇑	Liu *et al.*[138]
AF	• Male rats • Liver	• 1.5mg/kg/d AFB1 • 3-days • oral gavage	• Small RNA sequencing • RT-PCR	miR-34a⇑	Yang *et al.*[139]
OTA	• HO-1+/+ andHO-1− / −mice • Kidney	• 2.5 mg/kg OTA • every other day for 20 days • intraperitoneal injection	• RT-PCR	miR-34a⇑ and miR-21⇑, miR-29b⇓	Loboda *et al.*[140]
OTA	• Male rats • Kidney	• 0, 70 or 210 μg/kg • 5 days per week for 26 weeks • oral gavage	Small RNA sequencing and RT-PCR	DROSHA⇓, DICER⇓ And Pan-miRNA⇓	Dai *et al.*[120]

Table II-1 Continued

Dietary toxins	Tissue source	Toxin exposure Or measurement	miRNA profiling methods	Identified miRNA (validated target)	Ref.
Endocrine disruption chemicals					
BPA	• 80 pregnant women • Placenta	• placental BPA • GC-MS	• Small RNA sequencing • RT-PCR	miR-146a⇑	De Felive *et al.*[141]
DDT	• 229 lymphoma case and 327 controls • Blood leucocyte	• serum POPs • GC-MS	• miRNA microarray • genome microarray	193a-3p⇑, miR-152⇑, miR-31-5p⇑, and miR-34a-5p⇑	Krauskopf *et al.*[142]
BPA	• Male rats • Liver	• 0.5, 5, 50 mg/kg/d • 30 days • oral gavage	• RT-PCR	miR-122⇑	Vahdati Hassani *et al.*[143]
BPA	• Male mice • Pancreatic islets	• 50, 500 m g/kg/d • 8-weeks • oral gavage	• RT-PCR	miR-338⇓(*Pdx1*)	Wei *et al.*[144]
DDT	• Female rats • Liver and ovaries	• 50 mg/kg • once • intraperitoneal injection	• RT-PCR	Liver: miRs-21⇓, 221⇓, 222⇓, and 429⇓ Ovaries: miR-21⇑	Chanyshev *et al.*[145]
DDT	• Female rats • Uterine and ovaries	• 10,50,75 mg/kg/d • 12 weeks • intraperitoneal injection	• RT-PCR	Uterus: miR-221⇑, and Dicer⇓ Ovaries: miR-222⇓, miR-205⇓, miR-126a⇓, and Dicer⇓	Kalinina *et al.*[121]
Foodborne mutagens					
PAHs and BaP	• 135 exposure cases and 256 controls • Plasma	• plasma albumin adducts by ELISA • urinary monohydroxy-PAHs by clinical chemistry analysis	• Small RNA sequencing • RT-PCR	Urine: miR-24-3p⇓, miR-27a-3p⇓, miR-142-5p⇓, miR-28-5p⇓ and miR-150-5p⇑ Plasma: miR-24-3p⇓, miR-28-5p⇓ and miR-150-5p⇑	Deng *et al.*[151]
BaP	• Mice • Stomach, spleen, lung, liver, and colon	• 125 mg/kg/d • 5 days • oral gavage	• miRNA microarray	miR-290⇑, miR-298⇑, miR-346⇑, and miR-351⇑ tissue-specific responses, modest changes	Zuo *et al.*[154]
BaP	• Male mice • Liver	• 150 mg/kg/d • 3 days • oral gavage	• miRNA microarray • RT-PCR	Lack of change in miRNA profiling	Yauk *et al.*[155]
PhIP	• Male rats • Colon	• 40 mg/kg/d PhIP and 4-weeks HF diet cycling • 2-weeks for 3 times • oral gavage	• miRNA microarray • RT-PCR	miR-126⇑, miR-145⇑, miR-21⇑, let-7 family⇓, miR-29c⇓, miR-215⇓	Parasramka *et al.*[156]
PhIP	• Male rats • Small intestine, liver, skin, lung, zymbal's gland	• 40 mg/kg/d PhIP and 4-weeks HF diet cycling • 2-weeks for 3 times • oral gavage	• RT-PCR	miR-21⇑, miR-126⇓, miR-29⇓, miR-215⇓, miR-145⇓	Chen *et al.*[157]

Table II-1 Continued

Dietary toxins	Tissue source	Toxin exposure Or measurement	miRNA profiling methods	Identified miRNA (validated target)	Ref.
Metals					
Sodium arsenite	• Male rats • Liver	• 0.1, 1, 10, or 100 mg/L, 60 days drinking water	• Small RNA sequencing • RT-PCR	miR-151⇑, miR-183⇑, miR-25⇑ (Gclc and Gclm)	Ren *et al.*[158]
Total arsenic in maternal urine	• 200 pregnant women • Maternal urine and cord blood	• HG-AAS with cryo-trapping	• miRNA microarray • RT-PCR	Cord blood: let-7a⇑, miR-107⇑, miR-126⇑, miR-16⇑, miR-17*⇑, miR-195⇑, miR-20a⇑, miR-20b⇑, miR-26b⇑, miR-454⇑, miR-96⇑, and miR-98⇑	Rager *et al.*[159]
Total inorganic arsenic	• 466 mother-infant pairs, • Placenta and cord blood	• DRC-ICP-MS	• TaqMan RT-qPCR Open Array	placenta: miR-1290⇑	Rahman *et al.*[160]
Urine mercury	• 60 cases and 30 controls • Plasma	NA	• RT-PCR	Plasma: miR-92a-3p⇑ and miR-486-5p⇑	Ding *et al.*[162]
Hg and Pb	• 110 newborns • Placenta	• inductively coupled plasma mass spectrometry (ICP-MS)	• nCounter Analysis System	Let-7 family⇓, miR-151-5p⇓, miR-190b⇓, miR-193b⇓, miR-423-5p⇓, miR-520d-3p⇓, and miR-96⇓	Li *et al.*[163]
Alcohol					
Alcohol	• 18 healthy subjects • Blood	• medical lab	• RT-PCR	miR-122⇑	McCrae *et al.*[166]
Alcohol	• 16 modest alcohol drinkers • Blood	• medical lab	• Small RNA sequencing	Serum: miR-122⇑, miR-185-5p⇓	Ten Berg *et al.*[167]
Alcohol	• Rhesus macaques • Blood and colon	• 4% (w/v) ethanol, 12 months • self-administration	• RT-PCR	PBMC: miR-181-5p⇑ and miR-221-3p⇑ Colon: miR-155⇑	Asquith *et al.*[168]

Table II-2 miRNA modulation by chemopreventive agents.

Abbreviation:

AKBA, 3 acetyl-11-keto-β -boswellic acid; AOM, Azoxymethane; DSS, Dextran sulfate sodium; ECS, environmental cigarette smoke; EGCG, Epigallocatechin-3-gallate; I3C, indol-3-carbinol; LPS, lipopolysaccharide; NNK, tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; PBMC, peripheral blood mononuclear cell; EITC, phenethyl isothiocyanate; PhIP, 2-amino-1-methyl-6-phenylimidazo [4,5-b] pyridine; RT-PCR, real time PCR; SCID, Severe combined immunodeficiency; T2DM, type 2 diabetes mellitus; VC, vinyl carbamate

Chemopreventive agents	Tissue source	Agent exposure	miRNA Profiling methods	Identified miRNA (validated target)	Ref.
Curcumin					
Diflourinated-curcumin	• Female SCID mice • Pancreatic cancer xenografted tumor	• 2.5, 5 mg/kg/d • 3 weeks • oral gavage	• RT-PCR	let-7⇑, miR-26a⇑, and miR-101⇑ (EZH2 and EpCAM)	Bao *et al.*[177]
Curcumin and AKBA	• Male nude mice • Colorectal cancer xenografted tumor	• 25mg/kg/d curcumin and/or 75mg/kg/d AKBA • 3 weeks • intraperitoneal injection	• miRNA microarray • RT-PCR	miR-34a⇑ (c-Myc, CDK6 and CyclinE1) miR-27a⇓ (FBXW7)	Toden *et al.*[179]
Curcumin	• LPS-induced mice sepsis model • Liver and kidney	• 20mg/kg curcumin • 3-day pretreatment before LPS injection • oral gavage	• RT-PCR	miR-155⇓	Tili *et al.*[181]
Curcumin	• Male rats • Kidney	• 15, 30 and 60 mg/kg • 6 weeks • oral gavage	• RT-PCR	miR-206⇑ (PTP1B)	Ding *et al.*[182]
Resveratrol					
Resveratrol	• Male SCID mice • Prostate cancer xenografted	• 20 mg/kg • every other day for 5 weeks • oral gavage	• miRNA microarray • RT-PCR	miR-21⇓ (PDCD4)	Sheth *et al.*[186]
Pterostilbene	• Male mice • Prostate cancer xenografted tumor	• 50 mg/kg/d • 5 days a week for 39 days • intraperitoneal injection	• RT-PCR	miR-17⇓, miR-20a⇓ and miR-106b⇓ (PTEN)	Dhar *et al.*[187]
Resveratrol	• DSS-Apc Min+/- mouse • Intestine	• 100 mg/kg/d • 5 days a week for 5 weeks • oral gavage	• miRNA microarray • RT-PCR	miR-101b⇑ and miR-455⇑	Altamemi *et al.*[188]
Resveratrol	• SCID mice • Melanoma xenografted	• 30 mg/kg/d • 7 days • intraperitoneal injection	• RT-PCR	miR-221⇓ (TFG)	Wu *et al.*[189]
Grape extract with resveratrol	• 35 T2DM and hypertensive medicated male patients • PBMC	• 8 mg/d/first 6 month and 16mg/d/second 6 month • dietary supplement	miRNA microarray and RT-PCR	miR-21⇑, miR-181b⇑, miR-663⇑, miR-30c2⇑, miR-34a⇓, miR-155⇓	Tomé-Carneiro *et al.*[190]

Table II-2 Continued

Chemopreventive agents	Tissue source	Agent exposure	miRNA Profiling methods	Identified miRNA (validated target)	Ref.
Other polyphenols					
EGCG	• NNK treated mice • Lung	• 0.4% EGCG in diet • 1 week • dietary consumption	• miRNA microarray	modest miRNAs dysregulation, miR-210⇑	Zhou *et al.*[192]
Quercetin	• 264 lung cancer cases • Lung FFPE samples	• quercetin-rich food frequency questionnaires	• miRNA microarray • RT-PCR	let-7 family⇑, miR-17 family⇓, miR-26⇑, and miR-146⇑	Lam *et al.*[194]
Pomegranate juice	• AOM-treated male rats • Colon	• pomegranate juice (2504.74 mg gallic acid equivalents/l) • 10 weeks • drinking juice	• RT-PCR	miR-126⇑ (VCAM-1)	Banerjee *et al.*[197]
Cruciferous vegetable					
PEITC and I3C	• ECS-treated male rats • Lung	• 500 mg/kg diet (PEITC) and/or 2,500 mg/kg diet (I3C) • 28 days • oral gavage	• miRNA microarray	let-7 family⇑, miR-30 family⇑, miR-34 family⇓ and miR-99b⇑	Izzotti *et al.*[202]
I3C	• VC-treated mice • Lung	• I3C supplemented diet • 15 weeks • dietary consumption	• miRNA microarray • RT-PCR	miR-21⇓ (PTEN, PDCD4, and RECK), miR-31⇓, miR-130a⇓, miR-146b⇓, and miR-377⇓	Melkamu *et al.*[203]
I3C	• Female nude mice • Hepatocellular carcinoma xenografted tumor	• I3C at 25 mg/kg, • twice per week for 4 weeks • intraperitoneal injection	• RT-PCR	miR-21⇓ (PTEN), miR-221⇓ and miR-222⇓	Wang *et al.*[204]
Macronutrient and whole food					
Fish oil	• AOM-treated male rats • Colon	• 11.5 g fish oil/100g • 10 and 34 weeks • dietary consumption	• miRNA microarray	let-7d⇑, miR-15b⇑, miR-107⇑, miR-191⇑ and miR-324- 5p⇑	Davidson *et al.*[205]
Fish oil and pectin	• AOM-treated male rats • Colon	• 11.5 g fish oil/100g • 6 mg pectin/100g • 10 and 34 weeks • dietary consumption	• miRNA microarray	miR-26b⇑ (PDE4B) and miR-203⇑ (TCF4)	Shah *et al.*[206]
Spinach	• PhIP-treated male rats • Colon	• 10% wt/wt freeze-dried baby spinach • 34 weeks • dietary consumption	• miRNA microarray • RT-PCR	let-7⇑, miR-29c⇑, miR-215⇑ and miR-145⇑	Parasramka *et al.*[156]
Blueberries	• 17β-estradiol-treated female rats • Mammary tissue	• 2.5 or 5 % wt/wt freeze-dried whole blueberry in diet • 25 weeks • dietary consumption	• RT-PCR	miR-18a⇓ and miR-34c⇓	Jeyabalan *et al.*[210]

CHAPTER III

MULTI-OMIC APPROACHES FOR TARGET IDENTIFICATION AND NOVEL INTERVENTION FOR COLORECTAL CANCER

Overview

The burden of colorectal cancer (CRC) demands better understanding of the disease etiology and improved intervention strategies. Environmental exposures, genetics, and diet/life style factors all contribute to CRC risk. Diet/lifestyle style factors influence several areas, including the gut microbiome. There is increasing appreciation of the complex interactions that exist between diet, microbiota, bile acids, and host epithelial cells in the gut. Recent advances in sequencing and mass spectrometry-based technologies have rapidly expanded the availability of large biological datasets and an appreciation for various 'omics' approaches, including genomics, epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics. Multi-omic approaches hold the promise of uncovering exciting and important new scientific directions, from dynamic multivariate analyses of large biological datasets, sometimes called 'Big Data', integrated with established molecular pathways and networks. This has been characterized as the era of hypothesis-generating research, rather than hypothesis-testing, and the 'one gene/one drug' philosophy applied historically to cancer treatment. We review multi-omic advances associated with CRC etiology, focusing on diet/host/microbiome interactions in the gut. We highlight challenges in the field linked to emerging interventions for CRC. Lastly, we discuss the 'multi-gene/multi-target' concept as a paradigm for CRC precision medicine.

Introduction

Colorectal cancer (CRC) is the third leading lethal cancer in the US[234]. Over many years, the accumulation of deregulated genetic and epigenetic factors provides a driver for the progressive transformation from normal colonic epithelium to invasive cancer[20,235]. Although ~30% of CRC cases are familial and genetically related, the majority (~70%) are described as sporadic. The latter cases have been linked to environmental exposures, including nutritional status and dietary intake patters[61,131,236–239], along with physical activity[240–242]. There growing appreciation for the importance of environmental exposures interfacing with microorganisms living in the human gut[243–247]. The human-gut microbiome field has been aided by advances in various key technologies[248]. Gut dysbiosis is implicated in CRC, colitis-associated cancer, inflammatory bowel disease (IBD), and other intestinal disorders/syndromes[249–252 253,254].

The molecular landscapes of CRC have evolved enormously in the past decade with improved understanding of the basic mechanisms and clinical applications, yet this remains a complex disease requiring new advances on multiple fronts. The refined "Hallmarks of Cancer" by Hanahan and Weinberg in 2011[255] added two categories, 'Energy re-programming' and 'Evasion from immunity', and advanced newer concepts related to genomic instability and tumor-promoting inflammation. The recognition of signaling networks/circuits and tumor microenvironment further highlight the intrinsic, diverse interplay of factors in the neoplastic setting. Advances in CRC and other human malignancies support the important hallmark of Epigenetics[71,256,257], including noncoding RNAs, such as microRNAs[258] and LncRNAs[259]. All of these hallmarks interface with the

emerging fields of host-microbiota interactions and infectious pathogenesis, emphasizing the complex nature of CRC tied to extrinsic factors[260].

Omic technologies hold the promise of identifying critical regulatory molecules in biological samples[261]. These approaches have prompted a rapid growth of 'Big Data' investigations in (epi)genomics, transcriptomics, proteomics, and metabolomics for various diseases[262–264]. Here, we discuss the advances made in understanding CRC, linked to single 'omic' determinations. We propose that integrated multi-omic approaches are needed to advance the field towards improved intervention and therapeutic strategies for CRC, while acknowledging that this is by no means a trivial pursuit.

Advancements in technologies

DNA sequencing. The fields of genomics and transcriptomics strongly support inherited aspects of biological information, regulatory machinery, and functional activities within the cell. Microarrays were used frequently to detect transcripts by hybridization[265,266] before the advancement of DNA sequencing technologies. The sequencing field evolved over three generations during the past 40 years[267]. First-generation sequencing refers to Maxam-Gilbert sequencing (chemical sequencing)[268] and Sanger chain-termination sequencing[269] in gel-based systems. The primer guided DNA polymerase reaction of Sanger sequencing circumvented several complex methodological steps and extensive radiolabeling. Improvements, including the polymerase chain reaction (PCR), recombinant DNA technologies, and capillary array electrophoresis, led to automated DNA sequencing. By combining sheared DNA segments, cloning into vectors, Sanger

sequencing, and assembly of overlapping sequence reads results, *de novo* 'shotgun' sequencing generated the first human genome 'draft' copy, accomplished by the Human Genome Project[270,271].

The next (second) generation sequencing (NGS) also has been referred to as 'massively parallel' sequencing. Breakthroughs on increasing speed and throughput capacity crucially reduced sequencing cost. The principle involved converting the entire genome into small fragments and generating DNA reads from annealed adaptors via real time image acquisition. Common methods include pyrosequencing (Roche 454), sequence by ligation (ABI SOLiD), or sequencing by synthesis (illumine Solexa)[272,273]. NGS generates 50-500 base reads, which are much shorter than traditional Sanger sequencing, ranging from 700-1200 bases. However, in principle, mapped reads from high throughput sequencing can provide coverage across the whole genome.

Third-generation sequencing involves single molecule sequencing[274] by nanopore technologies[275] to reduce the bias and error from DNA amplification, and to achieve long reads within a short time[276]. Notably, NGS technology has grown exponentially since 2010, providing a cost-efficient solution for diverse molecular profiling applications[277]. Applications range from genome (whole genome, exome, or targeted regions), transcriptome (mRNA, small and long non-coding RNAs), and epigenome (chromatin topology, histone and chromatin modification, transcriptome machinery, and DNA modification)[278]. The sequencing technologies also enable determination of microorganisms in particular environments[279–281], such as soil, water, or human body sites.

The human microbiome is an emerging field based on sequencing technology, used to investigate the colonized bacteria influencing human health and disease[282].

Analytical mass spectrometry. Mass spectrometry (MS) evolved to understand the mass of particles, and has been adopted to study a wide range of molecules in the life sciences[283]. In contrast to DNA sequencing as an alignment-based data assembly, MS-based proteomics and metabolomics approaches are often exploratory, due to their complex nature and insufficient characterization of biological molecules for comparative and database searching. As pioneers of MS-based metabolite detection, Horning developed gas chromatography (GC)-MS to examine steroids, drugs, and drug metabolites in human urine samples[284], whereas Pauling *et al.* used GC-MS to profile about 250 substances in breath and urine vapor[285]. The successful measurements of human metabolites in the 1970s to 1980s led to further development of several analytical techniques to achieve high-resolution/sensitivity, to advance metabolic profiling of human diseases. Many MS and nuclear magnetic resonance (NMR)[286] configurations seek to achieve low sample requirement and high sensitivity, and have been used for both qualitative and quantitative studies to discover changes in proteomics[287,288] and metabolomics[289,290]. In general, the procedure is initiated from ionization of an individual sample, or series of samples, followed by separation techniques such as liquid (LC) or gas chromatography (GC), capillary electrophoresis (CE), and ion mobility spectrometry (IMS)[291]. The chemical species then can be sorted according to their mass to charge ratio by various mass selection methods, which include time-of-flight (TOF), quadrupole mass filter, ion traps, and

Fourier transform ion cyclotron resonance. A scanning instrument produces the mass spectrum from the obtained signals for further interpretation, and the final detector records the charge or current from the passing ion, or ion hits on a surface with an electron multiplier. Coupling MS instruments can increase the sensitivity of detection through product-ion scan, precursor-ion scan, neutral-loss scan, and selected reaction monitoring[292]. For proteomics, samples are usually applied with prior gel-based separation and enzyme digestion before MS analysis. Analytically, proteins are considered gene products consisting of a combination of defined amino acids. Although there is basic knowledge accrued of protein composition, interpretation of proteomics remains complicated, due to functional aspects such as conformation, interaction, activity, subcellular location, and post-translational modifications[293,294]. Unlike proteins, metabolites are not linked directly to genomic information, and are highly dynamic, relying on enzymatic reactions for formation, flux and degradation. Metabolomics can be briefly divided into targeted and untargeted approaches, emphasizing specific metabolites with closely related biochemical properties, such as lipidomics, or global profiling of samples[295]. Modern MS approaches combine imaging technologies and direct ionization methods, leading to spatial distribution of targeted metabolites in tissues[296,297].

From genetics to omics

Classification of CRC typically involves genetic and epigenetic aspects, including genomic instability and DNA methylation[29,298]. Mutations of DNA mismatch repair machinery results in microsatellite instability (MSI) phenotype with hypermutations.

Chromosomal instability (CIN) refers to chromosomal mutation and rearrangement leading to activation of oncogenes and inactivation of tumor suppressors. Epigenetic gene silencing by DNA hypermethylation is described as CpG island methylator phenotype (CIMP)[29,298–301]. Coupled to sequencing technology advancement, several milestones have been made in the cancer arena that led to the evolution of omic concepts.

In 2006, Sjöblom et al. used more than 250,000 PCR primers to assess exomes via Sanger sequencing[30,31]. The study validated hundreds of somatic mutations associated with cellular adhesion/motility, signal transduction, and transcriptional regulation in breast and CRC patients. Beginning in 2008, the cost for whole genome sequencing had declined due to advances in NGS, and again in 2012[302]. In 2012, The Cancer Genome Atlas (TCGA) Network reported a subset of whole genome sequencing results, confirming known genetic and epigenetic abnormalities, and identifying novel chromosomal alternations, gene mutations, and gene silencing after comprehensive characterization of genome, epigenome and transcriptome read-outs in CRC[33]. After multiple efforts toward unified colorectal classification, Guinney et al. consolidated four consensus molecular subtypes (CMS1-4)m from 17 gene expression microarray and TCGA RNA sequencing datasets[37]. In 2013, the first epidemiologic study identified decreased microbial community diversity and Clostridia, and increased *Fusobacterium* and *Porphyromonas* by 16S rRNA sequencing in human CRC cases compared to health subjects[303].

In parallel, MS-based measurements contributed towards enhanced understanding of the proteome and metabolome. In 2009, Chan et al. applied NMR and GC/MS to paried normal and tumor samples from CRC patients. The results indicated an increase of

hypoxia, glycolysis, lipid metabolism, and immune response in tumors, and the potential of using *metabolites as biomakers* for disease mangement and prognosis[304]. Around the same time, Hirayama et al. utilized CE-TOF to profile metabolome of colon and gastric cancer, and uncovered the dysregulation of glycolysis, glutaminolysis, aerobic repiration, and nucleotide metabolism[305]. Two years after the original TCGA study of CRC, the Clinical Proteomic Tumour Analysis Consortium performed LC-MS/MS based proteomic analysis in CRC specimens. The restuls identified more than a hundred novel amino acid variants, and indicated poor protein prediction from mRNA[34]. In 2016, a comprehensive metabolome was generated for CRC, adjacent normal tissue, and stool using untargeted GC-MS/MS and ultra-performance (UP)-LC-MS/MS[306]. The results indicated that short chain fatty acid metabolism was the most critical dysregulated metabolic pathway, and stool metabolome differed from patient tissue samples. Thus, molecular profiling of CRC has progressed tredmendously over the last decade.

Omics provides opportunities to identify new leads based on robust and high throughput information, as well as expanding knowledge beyond a central dogma to seek new and important gene-environment interactions at a global level[307] (Figure III-1).

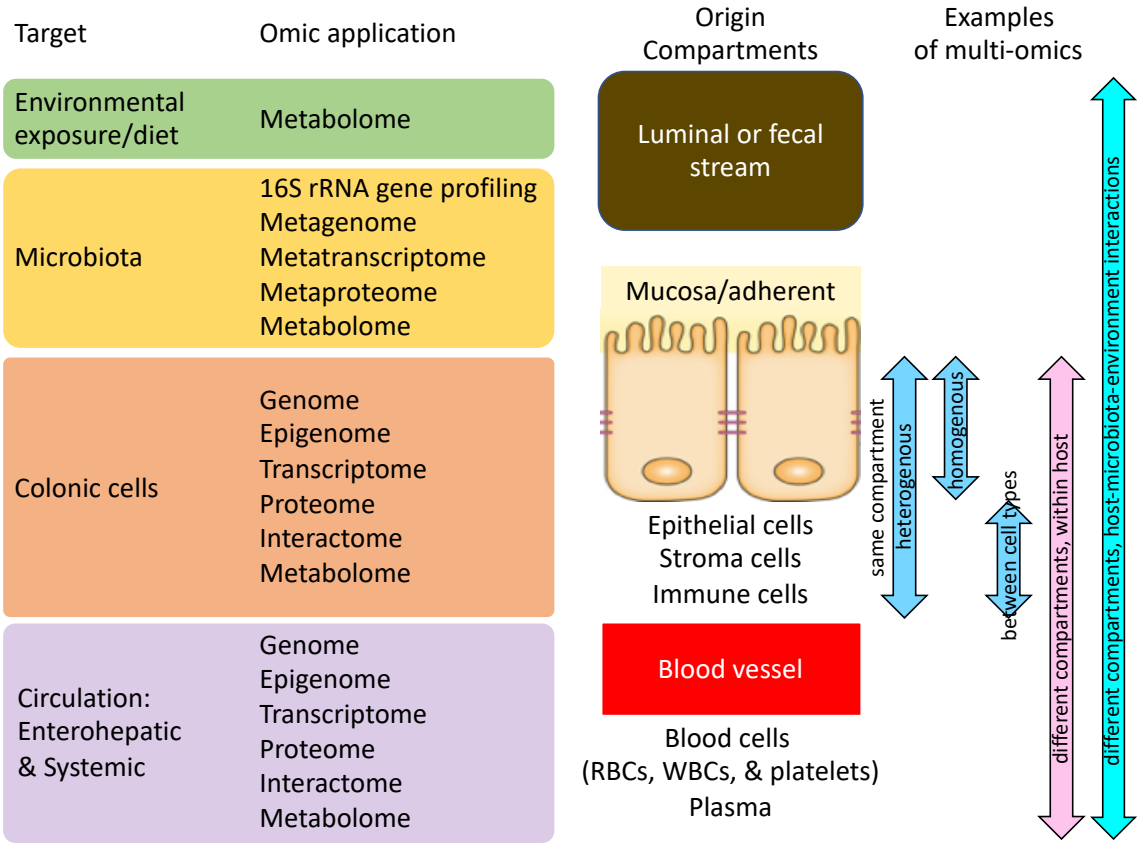


Figure III-1 Diet/host/microbe interactions coupled to -omics
The overview figure presents targets of detection, omic applications, and sample origins of colonic compartments. Multi-omic approaches can encompass multiple molecular levels, multiple targets, or multiple compartments.

Methods and tools for multivariate and network analysis

Several key databases and resources are available to facilitate biomedical research[308,309] (Table III-1). Genome related resources include Ensembl, NCBI (The National Center for Biotechnology Information), ENCODE (Encyclopedia of DNA Elements), UCSC Genome Browser, and KEGG (Kyoto Encyclopedia of Genes and

Genomes). Omic database repositories include GEO (Gene expression Ominibus) and TCGA. Pathway analysis tools include DAVID (The Database for Annotation, Visualization and Integrated Discovery), GO (Gene Ontology Consortium), MSigDB (Molecular signature database), and Reactome. For microbiome and metabolome research there are HMP (NIH Human Microbiome Project), HMDB (The Human Metabolome Database), PubChem, FooDB (The Food Database), METLIN, ChemSpider, WEIZMASS[310], and KEGG Compound, plus PeptideAtlas and PRIDE (PRoteomics ID Entifications) Archive for proteome work. For global analysis and data visualization, resources are available through Bioconductor, Github or Galaxy, and others can be accessed by web-based or software-based platforms such as Metaboanalyst, XCMS, IPA, MetaCore, IMPALA (Integrated Molecular Pathway Level Analysis), and Cytoscape.

With increasing high throughput data from omics, bioinformatic tools and systems biology are in great demand for data processing, interpretation, and integration. Ge et al. pointed out in 2003 the need of systems biology for moving from 'molecule biology' to 'modular biology', that combines co-expression, localization, and phenotypic patterns of protein functional network[311]. Joyce and Palsson summed omics integration techniques into identifying scaffold, scaffold decomposition, and cellular modeling and analysis, involving multiple algorithms and databases in model organisms[312]. Buescher and Driggers reiterated the challenges in omics integration. They included technical and statistical caveat of each -omics, biological relationships across multiple -omics, and the dynamics in time or conditions[313]. Overcoming those difficulties would expand our knowledge by either co-analyzing or merged analysis of the integrated multi-omics.

Learning from bacteria or model organisms such as *C. elegans* and yeast may provide insights to resolve relationships of multi-omic layers[311–313]. Yugi et al. later proposed five trans-omics technologies and three network concepts to pursue trans-omic network reconstruction, which could be helpful for multifactorial diseases[314]. Data heterogeneity, pitfalls of the individual analytical tools, and lack of systematic bioinformatics, will need further improvements to better dissect biology via multi-omic integration. Towards this end, integrating multi-omic experimental data remains challenging.

Multi-omic integration for CRC

Using matched colonic tissues to study epigenetic alterations leading to gene expression dysregulation revealed molecular subtypes and underlying regulatory networks in CRC. Tang et al. identified 72 miRNA-mRNA pairs in CRC tissues compared to matched adjacent normal[315]. Bioinformatic analysis uncovered miRNA-mRNA regulation of Wnt signaling, and confirmed the association of Klf4 downregulation and miR-29a upregulation. Vishnubalaji[316] et al. used microarrays for detecting mRNA and miRNA to confirm dysregulation of the Wnt pathway, and identified epigenetic regulators, miR-26a and let-7, associated with upregulation of *EZH2*, an important enzymatic chromatin component. Inhibition of *EZH2* by pharmacological chemicals and siRNA, or ectopically overexpressed pre-miRNA regulators, led to changes in viability of colorectal cell line, suggesting anti-cancer roles of miRNA-mRNA regulations on chromatin.

Public resources also provide valuable and robust analysis pipelines. Although data heterogeneity and subtype heterogeneity were observed, Cantini et al. employed

stepwise microRNA master regulator analysis to identify miRNA drivers in TCGA data. The analysis found that downregulation of miR-194, miR-200b, miR-203, and miR-429 was associated with poor prognosis in stem/serrated/mesenchymal subtype, along with dysregulation of TNFα signaling through NF-κB, inflammatory response, and extracellular matrix remodeling pathway in CRC[317].

Tumor heterogeneity and tumor microenvironment are associated with disease progression and the resistance to therapies[318,319]. Laser-capture microdissection, flow cytometry-assisted cell sorting techniques, and microfluidic systems are increasingly used for subpopulation analysis[320–322]. Human colorectal intratumor heterogeneity was elucidated by single-cell transcriptomics, enabling the separation of epithelial, endothelial, fibroblast, T cell, Mast cell , B cell and other myeloid cells with their cell type specific gene dysregulation[323]. Compared to bulk transcriptome, the study identified EMT related genes that were upregulated specifically in cancer-associated fibroblast (CAFs) subtype, and provided better stratification of prognosis, indicating the critical role of CAFs in CRC. To further increase omic resolution and investigate the nature of mutational diversification, Roerink et al. utilized clonal organoids derived from patient normal and neoplastic colorectal epithelial cells[324]. The study involved whole genome sequencing, RNA-seq, methylation chip array, and identified high somatic mutation rates, and correlated DNA methylation and gene expression to drugs on CRC primary organoids.

Metabolic reprogramming is an evolving but poorly understood hallmark of cancer[255]. Satoh et al. assessed both transcriptome and metabolome data of human CRC across different stages, and found early global metabolic reprogramming in adenoma

stages, which was independent from mutation status but was associated with c-Myc expression[325]. The investigation discovered the roles of MYC-dependent metabolic gene expression, and metabolites associated with purine and pyrimidine metabolism, with further validation by rodent and cell-based assays.

IBD are considered possible precursors of CRC, and inflammation has been implicated in CRC progression and survival[326–330]. Haberman et al.[331] characterized transcriptome and microbiome in pediatric Crohn's disease patients and identified increased dual oxidase (*DUOX2*) correlated with an expansion of *Proteobacteria* and decreased *APOA1* in association with *Firmicutes*, which led to severe mucosal immune response by Th1 polarization[331]. Morgan et al. sought to identify host-microbe interactions by transcriptome and microbiome in pouchitis from IBD and familial adenomatous polyposis (FAP) patients[332]. With co-variants in biopsy location, inflammation score, and antibiotic use, the interaction between host and microbiome was unexpectedly modest. Enriched pathways in Complement cascade and IL-12 pathway were inversely correlated to *Sutterella, Akkermansia, Bifidobacterial*, and *Rosebura* abundance suggesting host transcriptome was not the major regulatory component.

When using 100 inbred mice strains to assess host genetic and microbe factors in response to high-fat, high-sucrose diet, Parks et al. revealed a stain-specific single nucleotide polymorphisms (SNPs) effect on genetically determined obesity, regardless of diet, and demonstrated a robust microbiome shift by diet with host strain-specific plasticity[333]. To investigate the influence of high-protein diet on gut health, Mu et al. employed microbial and host colonic gene expression analyses in rats, and suggested an

increased risk of colonic disease by dietary intervention[334]. The high-protein diet shifted microbiome leading to dysbiosis, reduced butyrate levels, and downregulation of genes related to oxidative phosphorylation, *O*-linked glycosylation of mucins, and innate immune response.

It remains open to explore structure and microbial gene expression of microbiota in the gut[335,336] and how diet can play a part shaping the gut microbiota[337,338]. David et al. showed a replicable short-term microbiome alteration by animal- and plant-based diets in human subjects[245]. The animal-based diet increased bile acids, the expressions of bile salt hydrolases and sulphite reductases, and the abundance of *Alistipes, Bilophila* and *Bacteroides* linking to IBD-prone microbiota. It was reported that the luminal contents and mucus layer had distinguishable microbial community[339]. Mono-colonized bacteria also had different transcriptional profiles from the content and mucus, and the defined microbiota showed differences in functional metabolism and proliferative adaptation, suggesting a specified spatial and functional role of certain bacteria in the host gut.

To address the diet effect on human microbiome, Desai et al. colonized a gnotobiotic mouse model with human microbiome and supplemented fiber-rich, Fiber-free and prebiotic diets. With 16S-seq profiling of the bacteria community and microbial RNA-seq, the results indicated fiber-depleted microbiome promoted mucus-degrading bacteria growth, and induced carbohydrate active enzymes to degrade mucus polysaccharide. A lack of complex fiber led to reduced barrier function resulting in susceptibility to pathogen infection and severe colitis[340]. To understand the impact of diet and obesity on CRC risk, a diet-induced obesity mouse model with transcriptome and

methylome analysis was conducted. Li et al.[341] identified diet-induced obesity caused reprogramming of DNA methylation leading to metabolic gene dysregulation in fatty acid metabolism at young age and cancer-prone gene signature.

Gut microbiota can co-metabolize with host mammalian cells[342–344]. Weir et al. utilized a case-control study to employ microbiome and metabolome profiling of stool samples[345]. There was no significant difference in bacterial community but reduced butyrate-producing species and increased mucin-degrading species were observed. Higher levels of acetate, unsaturated fatty acids, amino acids and a conjugated-bile acid were detected in stools from CRC patients. The correlation analysis also identified strong correlations between certain microbes and metabolites such as glutamine, cholesterol derivatives, serine, linoleic acid, and oleic acid.

Other studies engaged mutli-omics approaches to dissect the interactions in the complex gut ecosystem. Aidy et al. used a conventionalization of germfree mouse to elucidate the host-microbe dynamics in a time series multi-omic analyses[346]. The results indicated a rapid microbiota adaptation after day 4, reduced succinate and lactate, increased acetate, propionate and butyrate throughout the course, induction of genes related to glycolysis, amino acid and nucleotide pathways in late time points, and increased urine creatinine and formate in early time points. The multi-omic correlation models also suggested fructose and mannose metabolism and folate biosynthesis connecting the interaction among microbes-host transcript-tissue metabolites.

To determine the impacts of host genetics and diet on gut microbiome and metabolome, stool samples from 30 strains of mouse were investigated by 16S sequencing

and GC-MS with integrated SNPs information[347]. The study demonstrated early environmental exposure and host genetic variants as contributing to microbiome structure, related to human disease and dietary impacts. The study also suggested that non-dietary metabolites were more closely regulated by microbial metabolism. A mono-colonization of E. coli commensal strain was indicated to affect lipid metabolism through bile acids and inflammation via a genome-scale metabolic model involving 696 differentially altered mRNAs, 400 reactions and 704 metabolites[348].

Out microbiome also participates in epigenetic regulation of host gene expression[349,350]. Richards et al. treated primary human colonocytes with suspended human microbes and utilized RNA-seq, Assay for transposase-accessible chromatin sequencing (ATAC-seq), and 16S rRNA-seq to address the interaction between mammalian cells and microbes[351]. Microbiome taxa was associated with mammalian genes related to hemidesmosome and membrane targeting, and acute exposure to microbes altered the mammalian chromatin accessibility leading to changes in gene expression. Qin et al. [352] showed the microbiota changes prior to host phenotype in a sex-dependent manner by high fat diet. The diet intervention led to alterations in transcriptome and the occupancy of genome enhancer regions. Manipulating microbiome by fecal transplantation recapitulated the dietary impacts, suggesting gut microbiome is a key mediator of dietary reprogramming of host epigenomes leading to phenotypic consequences.

The multi-omic studies are summarized in Table III-2.

Emerging interventions for CRC

Preventive interventions for CRC can be generalized into primary prevention (reducing or eliminating risk factors), secondary (screening, early diagnosis, and treatment), and tertiary (disease management and improving survival)[353–355]. Lifestyle modification and medication for primary prevention can have benefits[46,356]. Nonsteroidal anti-inflammatory drugs (NSAIDs), such as aspirin, reduce the recurrence of human colorectal adenomas[357,358] through modulating prostaglandin synthesis, Wnt pathway inhibition, and host immune responses[359]. However, long-term adverse event management and patient stratification are necessary. A multi-target stool deoxyribonucleic Acid (MT-sDNA, Cologuard™) screening test for CRC was developed by the Mayo Clinic and Exact Sciences (Madison, WI)[360] and was adopted by several national preventive and cancer communities[361]. The test detects quantitative KRAS mutations, aberrant NDRG4 and BMP3 methylation, and hemoglobin with a weighted algorithm calculation. Although the benefits of MT-sDNA must be evaluated in the future, this work establishes that improved molecule profiling will likely add value to clinical biomarker discovery.

With increasing assessments of host-diet-microbiota interactions, several intervention methods targeting gut microbial community were proposed. Probiotics and prebiotics are implicated for CRC prevention through modulating immune responses[362,363]. While probiotic means the live microorganism providing health benefits on the host, prebiotic is an ingredient allowing changes in gut microflora composition[364]. Through inducing the production of short chain fatty acids by food grade dairy *Propionibacteria*, Cousin et al. demonstrated a synergistic pro-apoptotic cytotoxicity with

co-treatment of TNF-related apoptosis-inducing ligand in human colon cell line[365]. A probiotic strain *Lactobacillus casei BL23* was used against inflammation induced CRC in mouse model, and a Th17 response was identified as the mechenism[366]. Additionally, a mixture of oligofructose-maltodextrin-enriched bacteria strains were treated to an inflammation-induced rat model. Improved barrier function and reduced inflammation via TLR2 signaling were suggested as contributing to decreased tumor outcomes[367]. Beyond probiotics and prebiotics, fecal microbiota transplantation (FMT) has shown efficacy on *C. difficile* infection[368–370] and may have beneficial potential towards IBD[371,372] as well as CRC[373,374]. On the other hand, engineered non-pathogenic bacteria also has been proposed for bacteria-directed enzyme prodrug therapy[375–377]. Several bacteria showed selective tumor localization and proliferation in tumors after oral administration, supporting the delivery potential[378–380]. Hsieh et al. utilized engineered *E. coli* to produce β-glucuronidase to enhance the anticancer activity of a camptothecin analog by facilitating conversion to the therapeutic SN-38, a topoisomerase poison[381]. Murphy et al. demonstrated TNFα expressing *E. coli* produce therapeutic molecules inside tumors, and reduce tumor growth[382]. Similarly, Zheng et al. reported a cancer immunotherapy using nonvirulent and flagellin secreting *Salmonella* to trigger immune response via TLR4 signaling and M1 macrophage phenotype activation[383]. More recently, an engineered *E. coli* carrying myrosinase activity mixed with cruciferous vegetable showed anti-cancer activity by tumor targeting through proteoglycan binding and anticancer activity via sulforaphane production[384].

Conclusion and perspectives

The colon is a major organ in contact with dietary exposures, and harbors up to 10^{12} cfu/ml bacteria, leading to a complicated gut ecosystem[385]. Given that CRC is a complex disease involving gene-environment interactions[68,386], comprehensive mechanistic insights are needed to develop improved prevention and treatment strategies. After the consensus molecular subtype described by transcriptome in 2015[37], several studies contributed to expand the molecular understanding in immune subtype and heterogeneity. With molecular profiling we are moving towards precision medicine. However, shifting the 'one gene, one drug' paradigm to 'multi-genes, multi-drugs' may offer novel perspectives and new strategies against multiple aberrant targets[387]. Dietary intervention could offer 'multi-gene, multi-target' potential[208,209,233]. Additionally, the recent literature reveals the capabilities of gut microbiota-mediated dietary benefits, drug activity and metabolism, towards host physiological modulation[388–390]. Together, diet-microbe intervention could provide a novel angle for CRC intervention.

From the literature reviewed, we noted that the majority of studies failed to integrate adequately multi-omic datasets, indicating a need for improved statistical methods, modeling, integration, and visualization[313]. Most omics involves constant *de novo* construction of the reference information, comparative analysis to separate signal and noise, and alignment/match to a database. Therefore, technologies and comprehensive, representative databases require constant advancements for better data interpretation[307]. As more high-throughput data from different technologies and tissue

sources are generated, improvements from sampling to measurements remain challenging and need to be further standardized.

Multi-omic approaches in CRC can provide novel mechanistic leads and indicate the substantial role of diet-microbe on host physiology and pathology[391–393]. This review highlights the increasing use of multi-omic analyses in CRC research, with an urgent need for enhanced integration, as well as development of novel interventions based on advanced technologies and multi-omic approaches. Understanding the interplay among diet-microbe-host with systematic interpretation will provide important future directions for disease monitoring and intervention.

Table III-1 Commonly used database and resources for genome, transcriptome, proteome, metabolome and microbiome.

Resources	URL
Bioconductor	www.bioconductor.org
ChemSpider	www.chemspider.com
Cytoscape	cytoscape.org
DAVID	david.ncifcrf.gov
ENCODE	www.encodeproject.org
Ensembl	useast.ensembl.org/index.html
Galaxy	huttenhower.sph.harvard.edu/galaxy/
GenomeBrowser	genome.ucsc.edu/index.html
GEO	www.ncbi.nlm.nih.gov/geo/
GitHub	github.com
GO	www.geneontology.org
HMDB	www.hmdb.ca
HMP	hmpdacc.org/ihmp/
IMPALA	impala.molgen.mpg.de
IPA	www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/
KEGG	www.genome.jp/kegg/
Metaboanalyst	www.metaboanalyst.ca
MetaCore	clarivate.com/products/metacore/
METLIN	metlin.scripps.edu
MSigDB	software.broadinstitute.org/gsea/msigdb/index.jsp
NCBI	www.ncbi.nlm.nih.gov
PeptideAtlas	www.peptideatlas.org
PRIDEArchive	www.ebi.ac.uk/pride/archive/
PubChem	pubchem.ncbi.nlm.nih.gov
Reactome	reactome.org
TCGA	portal.gdc.cancer.gov
XCMS	xcmsonline.scripps.edu

Table III-2 Relevant multi-omic studies related to colorectal cancer.

Abbreviations:

CRC, human colorectal cancer; T/N, matched tumor and adjacent normal tissues from patients; UC, ulcerative colitis, FAP, familial adenomatous polyposis; WGS, whole genome sequencing; WGBS, whole-genome bisulfite sequencing; CE-TOF/MS, capillary electrophoresis time of flight mass spectrometer; MaAsLin, multivariate analysis with linear modeling; LDA, linear discriminant analysis; EMMA, efficient mixed model association; REML, restricted maximum likelihood; GF, germ-free; Conv, conventional; NIST, National Institute for Technology Standards; GEM, Genome-scale Metabolic model.

Sampling type & number	Profiling tech	Resources	Integration/Statistical approach	Ref
8 CRC T/N	mRNA microarray miRNA microarray	TargetScan miRnada	Pearson correlation	Tang et al.[1]
13 CRC T/N	mRNA microarray miRNA microarray	TargetScan	Pearson correlation	Vishnubalaji et al.[2]
TCGA 450 primary CRC tumors	RNA-seq Small RNA-seq	miRTarBase doRiNA-PicTar microRNA.org PITA TargetScan	Kolmogorov–Smirnov test ARACNe Algorithm Stepwise linear regression	Cantini et al.[3]
11 CRC T/N 7 cell lines	Single cell RNA-seq	BioGPS Cancer Genome Browser	Eight clustering methods including All-HC, HiLoadG-HC, BackSPIN, RaceID2, Seurat, VarG-HC, VarG-PCAproj-HC and VarG-tSNEproj-HC Wilcoxon rank-sum test	Li et al.[4]
C57BL/6J mice Colonic epithelial	WGBS RNA-seq	Reactome	Rao Scott Likelihood Ratio Test	Li et al.[5]
3 CRC T/V	WGS, RNA-seq, Methylation array		Subtree prune and regraft distance Shearwater Algorithm	Roerink et al.[6]
275 T/N $Apc^{+/\Delta 716}$ mice HCT116 cell line	RNA-seq CE-TOF/MS	MasterHands DAVID KEGG	XLStat	Satoch et al.[7]
316 pediatric cases and 43 controls Ilieal biopsies	RNA-seq 16S rRNA-seq	ToppGene TppCluster IPA	Pearson correlation Galaxy/multivariate analysis	Haberman et al.[8]
265 UC and FAP patients Pelvic pouch biopsies	RNA microarray 16S rRNA-seq		Bread-crumbs software package Spearman correlation MaAsLin LDA	Morgan et al.[9]

Table III-2 Continued

Sampling type & number	Profiling tech	Resources	Integration/Statistical approach	Ref
100 mice strains Cecal samples Epididymal adipose	RNA microarray 16S rRNA-seq		EMMA REML	Parks et al.[10]
Rats Luminal contents Proximal colon	RNA microarray 16S rRNA-seq		Spearman correlation	Mu et al.[11]
18 human subjects	16S rRNA-seq Microbial RNA-seq			David et al.[12]
Gnotobiotic mouse Luminal content Mucus layer	16S rRNA-seq Microbial RNA-seq	STRING-DB	Tree Of Life iTOL	Li et al.[13]
Gnotobiotic mouse Cecal and fecal samples	16S rRNA-seq Microbial RNA-seq			Desai et al.[14]
10 CRC cases 11 control Stool samples	16S rRNA-seq GC-MS	XCMC AMDIS NIST Golm	Pearson's correlation XLStat	Weir et al.[15]
GF and Conv C57BL/6J mice Luminal contents Colon tissue Urine samples	RNA microarray 16S rRNA array NMR	IPA KEGG GO STEM	Matlab O2-PLS regression	Aidy et al.[16]
30 strains mice Stool samples Blood	16SrRNA-seq GC-MS	NIST	MIMOSA	Snijders et al.[17]
GF and mono-inoculated mice Ileum tissues	RNA microarray UPLC-MS	LIPID MAPS Reactome XCalibur	GEM	Chakrabarti et al.[18]
Primary human colonocyte	RNA-seq ATAC-seq 16S rRNA-seq			Richards et al.[19]
C57BL/6J mice Stool samples Colonic tissues	RNA-seq 16SrRNA-seq ChIP-seq	MSigDB/GSEA IPA		Qin et al.[20]

CHAPTER IV

A MIRNA SIGNATURE FOR AN ENVIRONMENTAL HETEROCYCLIC AMINE DEFINED BY A MULTI-ORGAN CARCINOGENESITY BIOASSAY IN THE RAT[1]

Overview

Heterocyclic amines (HCAs) produced during high-temperature cooking have been studied extensively in terms of their genotoxic/genetic effects, but recent work has implicated epigenetic mechanisms involving non-coding RNAs. Colon tumors induced in the rat by 2-amino-1- methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP) have altered microRNA (miRNA) signatures linked to dysregulated pluripotency factors, such as c-Myc and Krüppel-like factor 4 (KLF4). We tested the hypothesis that dysregulated miRNAs from PhIP-induced colon tumors would provide a "PhIP signature" for use in other target organs obtained from a 1-year carcinogenicity bioassay in the rat. Downstream targets that were corroborated in the rat were then investigated in human cancer datasets. The results confirmed that multiple let-7 family members were downregulated in PhIP-induced skin, colon, lung, small intestine, and Zymbal's gland tumors, and were associated with c-myc and Hmga2 upregulation. PhIP signature miRNAs with the profile miR-21^{high}/ miR-126^{low}/miR-29c^{low}/miR-215^{low}/miR-145^{low} were linked to reduced Klf4 levels in rat tumors, and in human pan-cancer and colorectal cancer. It remains to be determined

[1] Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature Archives of Toxicology (A miRNA signature for an environmental heterocyclic amine defined by a multi-organ carcinogenicity bioassay in the rat, Ying-Shiuan Chen, Rong Wang, Wan-Mohaiza Dashwood et al.), Springer Nature Archives of Toxicology (2017)

whether this PhIP signature has predictive value, given that more than 20 different genotoxic HCAs are present in the human diet, plus other agents that likely induce or repress many of the same miRNAs. Future studies should define more precisely the miRNA signatures of other HCAs, and their possible value for human risk assessment.

Introduction

Epidemiological evidence has linked the consumption of cooked and processed meat to increased risk for cardiovascular diseases[394], and for malignancies of the digestive system[395], breast[396], and prostate[397]. Heterocyclic amines (HCAs) found in high-temperature cooked meat can undergo metabolic activation to form bulky DNA adducts that are precursors to mutagenesis and carcinogenesis in various organs of the body [398–400]. For this reason, HCAs have been investigated historically in terms of their genetic aspects linked to genotoxicity and tumor initiation in target tissues[401,402].

One of the most abundant HCAs found in cooked protein-rich foods is 2-amino-1-methyl-6- phenylimidazo[4,5-b]pyridine(PhIP)[403–405]. Metabolic intermediates of PhIP have been detected in various tissues, including serum, urine, and hair samples[406–409], indicating wide systemic distribution in the human body. The US National Toxicology Program (NTP) classified PhIP as "reasonably anticipated to be a human carcinogen" (NTP 12th Report on Carcinogens)[410], and the International Agency for Research on Cancer categorized PhIP as a Group 2B carcinogen (IARC 1993)[411].

In preclinical models, PhIP serves as a multi-organ carcinogen[412–414], and can affect newborns through trans-placenta and neonatal exposure[415]. In PhIP-induced colon

tumors, genetic changes in Ctnnb1 constitutively activate β-catenin and β-catenin/T-cell factor (Tcf) signaling, dysregulating the β-catenin/c-Myc/E2F1/Bcl-2 axis and downstream apoptosis pathways[416–418]. In rat colon and prostate carcinogenesis models, PhIP exposure has been linked to chronic inflammation via changes in molecular endpoints such as NADPH oxidase, nuclear factor kappa-B (NFκB)[419], cyclooxygenases[420], and glutathione S-transferases[421]. PhIP-induced mammary carcinogenesis models identified changes in the expression of transcription factors, cyclins, annexins, matrix metalloproteinases, and extracellular signal regulated (ERK) kinases[422].

In addition to the genetic aspect mentioned above, research with PhIP and other HCAs has moved into the epigenetic realm. Non-coding RNAs have gained increasing attention for their roles in epigenetic regulation, and as signatures of environmental exposure [423–427]. For example, in human-derived prostaspheres, bisphenol A silenced small nucleolar RNAs with C/D motif (SNORDs) through histone modifications[428]. Industrial exposure to metal-rich particulate matter altered microRNAs (miRNAs) linked to oxidative stress and inflammatory responses in blood[429]. It is well established that miRNAs are conserved, negative regulatory small RNAs comprising 21–23 nucleotides, which act via mRNA cleavage or translational inhibition to influence phenotypic outcomes such as cell proliferation, differentiation, autophagy, and apoptosis[81,430].

Dysregulation of miRNAs has been linked to various pathologies, including cancer, neurodegeneration, obesity, and chronic inflammation[431]. Consequently, altered miRNA signatures are regarded as possible biomarkers for disease diagnosis and

prognosis, due to their relative stability and accessibility in various tissues and biofluids[127]. For instance, miR-21 regulates phosphatase and tensin homolog (PTEN) and programmed cell death 4 (PDCD4) in tumors, leading to enhanced cell proliferation and stemness maintenance. In blood, miR-21 is viewed as a diagnostic and prognostic biomarker, whereas miR-141, a regulator of epithelial-to-mesenchymal transition, is a proposed circulating biomarker of late stage colorectal cancer[432]. Circulating and tissue miRNAs also respond to environmental exposures, such as air pollution[433] and toxic chemicals[129].

In cultured human breast cancer cells, PhIP treatment was reported to alter miRNAs regulating the response to xeno-estrogen exposure[434]. In PhIP-induced rat colon tumors, dysregulated miRNAs were linked to the altered expression of cMyc and other pluripotency factors, such as Sox2, Nanog, Oct-3/4, and Krüppel-like factor 4 (KLF4)[435]. We were intrigued by the possibility that miRNAs altered in PhIP-induced rat colon tumors might serve as biomarkers of PhIP exposure in other target organs for tumorigenesis, and by extension, might represent a 'HCA miRNA signature' in the general population. The present investigation reports, for the first time, on miRNAs and their validated downstream targets that are dysregulated in PhIP-induced tumors of the colon, small intestine, skin, lung, spleen, and Zymbal's gland, and correlates with the corresponding human datasets.

Materials and methods

Animal treatment. PhIP originally was defined as a prostate and colon carcinogen in male rats, and a mammary carcinogen in female rats, after continuous exposure in the diet

for a year[413,436]. Subsequently, a modified protocol was devised that reduced the overall carcinogen exposure markedly, starting with short-term administration of PhIP followed by a high-fat diet[414,417,437]. For the current investigation, PhIP was given by daily oral gavage (40 mg/kg body weight) for 2 weeks, alternating with a high fat diet for 4 weeks, and generated tumor incidence outcomes as follows: Zymbal's gland 2%, liver 8%, spleen 8%, lung 10%, small intestine 27%, skin 38%, and colon 58%[156]. The 1-year carcinogenicity bioassay involved male F344 rats, purchased at 3–4 weeks of age from the National Cancer Institute. At termination, each rat (n = 40) was euthanized by CO2 inhalation, following a protocol that was approved by the Institutional Animal Care and Use Committee. After a thorough necropsy examination, tumor samples and other tissues were collected and flash-frozen before storage at −80 °C. Controls (n = 12 rat) that received vehicle by oral gavage for 2 weeks, alternating with a high-fat diet for 4 weeks, had no detectable tumors in any of the organs examined at the end of the study[156].

RNA extraction and profiling. The key methodologies were described in previous studies that defined the major miRNAs altered in PhIP-induced rat colon tumors[156]. In brief, vehicle controls, PhIP-induced tumors, and tumor-matched normal tissues from colon, small intestine, skin, Zymbal's gland, spleen, and lung were homogenized on ice. Total RNA was extracted in Trizol reagent (Life Technologies), whereas, miRNA was extracted using the miRNeasy kit (Qiagen, Valencia, CA, USA). RNA quantity and purity were verified by Nanodrop ND-1000, from the absorbance at 260 and 280 nm (260/280 ratio >1.9), as reported[55,438,439]. RNA (1 μg) from each rat tissue was reverse transcribed

using the Superscript III RT kit (Life Technologies) in 10- µl reaction buffer, and diluted 10-fold before adding to a 10-µl reaction containing SYBR Green I Master mix (Roche Applied Science) and gene-specific primers. For miRNA analyses, 1-µg RNA from each rat tissue was reverse transcribed using the miScript II RT Kit (Qiagen) and diluted fivefold before adding to a 20-µl miScript Primer assay (Qiagen) containing primers specific for selected mature rat miRNAs. Real-time qPCR data were acquired on a Light-Cycler 480 II (Roche Applied Science). Relative expression levels of miRNAs and mRNAs were calculated using U6B and Gapdh as internal references for normalization, respectively.

Principal component analysis (PCA). PCA used the online tool MetaboAnalyst 3.0[440], and clustering was performed using Pearson correlation and average linkage.

Biological network analyses. MicroRNAs and their putative target mRNAs were examined via MetaCore pathway analysis (GeneGo Inc., St Joseph, MI, USA). For pathway enrichment analysis, p values were calculated using the formula for hypergeometric distribution, reflecting the probability for a pathway to arise by chance. Pathway maps were prioritized based on statistical significance, for tumors from animals given PhIP versus normal-looking tissue for each target organ. Interaction networks of miRNAs of interest were generated using Metacore (Thomson Reuters), as reported[156]. Functional annotation and subsequent analyses were conducted using DAVID[441] and TargetScan[442].

Surveying the cancer genome atlas (TCGA). Pan-cancer (PANCAN) miRNA (n = 11,010) and gene-specific RNA-seq data (n = 9755) were downloaded on 8 April 2016 from the UCSC Cancer Genomics Browser[443]. To be consistent with the rat preclinical model, only primary tumor and solid normal tissue data were analyzed, whereas metastatic, advanced metastatic, and recurrent cancers were not included. The miRNA database contained 364 primary colon adenocarcinomas and 8 solid normal colon tissues, plus 8402 primary tumors and 635 normal tissues from other origins. The gene expression database for colon and rectal cancers contained 380 primary tumors and 50 solid normal tissues, plus 8079 primary tumors and 626 solid normal tissues from other origins. A survival plot was generated by Graphpad6.07 (GraphPad Software Inc.) according to the relative expression level of miRNAs in PANCAN datasets for solid primary tumors. Proportion views of miRNA expression were captured from the Cancer Browser, and statistical analyses were performed using Student's t test with Benjamin– Hochberg correction, designating significantly upregulated (red) or downregulated (green) targets.

Statistical analyses. Unless indicated otherwise, results were presented as mean ± SEM with n = 3 per group. A paired two-tail Student's t test was performed for tumor and tumor-matched normal tissue. Unpaired two-tail Student's t test was performed for TCGA data, comparing primary tumor and normal tissue. An asterisk in the figure designates significance at *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, except when the exact p value is shown.

Results

Defining "PhIP signature" miRNAs in multiple target organs. The miRNAs most highly dysregulated in PhIP-induced colon tumors[156] were examined in other target tissues of the rat, namely the small intestine, skin, spleen, lung, and Zymbal's gland. Hierarchical clustering of miRNAs grouped the colon and small intestine as separate from the other tissues (Fig. IV-1A, upper blue panels). The miRNA pattern was similar for skin and Zymbal's gland (Fig. IV-1A, bold red panels at center), and clustered spleen and lung separately, except for replicate Lung-PhIP-2 T (Fig. IV-1A, lower panels).

Principal component analysis (PCA) was conducted on triplicate samples of each normal tissue, obtained from control rats that had received vehicle but no PhIP (Fig. IV-1B). Colon and small intestine were clustered in one area, separate from spleen, and from skin, Zymbal's gland and lung samples that were grouped together. When PCA was repeated for PhIP treated animals (Fig. IV-1C), analyzing both tumor and adjacent normal-looking tissues, colon and small intestine continued to segregate from the other tissues examined. The results indicated that a panel comprising of miR-21, miR-126, miR-29c, miR-135, miR-215, plus eight members of the let-7 family could distinguish among the different target organs of PhIP-induced tumorigenesis in the rat.

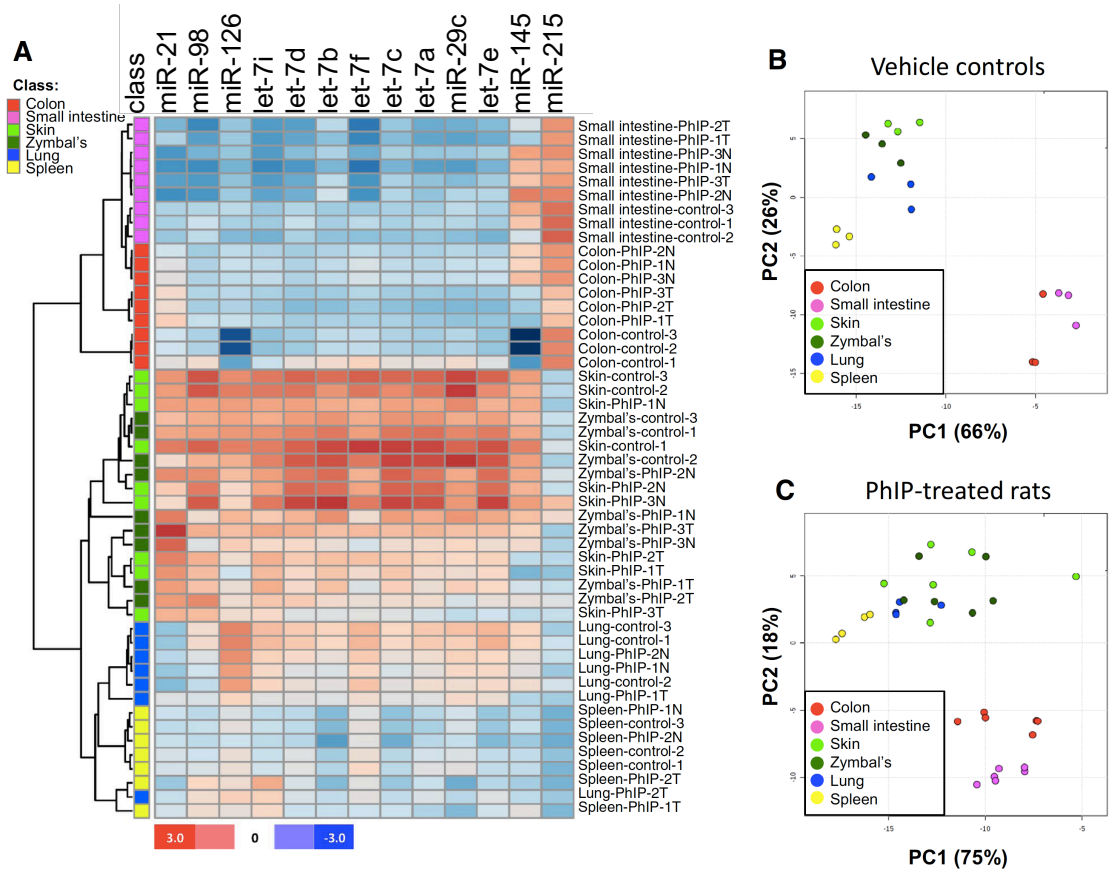


Figure IV-1 Target tissue miRNA profiling in the rat.
A, Heatmap showing the relative expression level of miRNAs, with Pearson average clustering of "PhIP signature miRNAs" for the different subgroups indicated in the figure. For each tissue, n = 3 individual replicates, except for spleen and lung tumors, where n = 2. T tumor; N normal-looking tissue adjacent to tumor. B, Principal component analysis (PCA) of miRNAs in normal tissues from control rats given vehicle but no PhIP. C, The corresponding PCA data for tumor and adjacent normal-looking tissues from PhIP-treated rats.

The let-7 family is dysregulated in multiple target organs of PhIP tumorigenesis. The let-7 family is highly conserved across species and function[444], and loss of let-7 expression during cancer development has been linked to changes in differentiation[445] and oncogenic

transformation[446]. Using quantitative real-time RT-PCR, PhIP signature miRNAs that were examined in multiple target organs of the rat including let-7 family members let-7a, let-7b, let-7c, let-7d, let-7e, let-7f, let-7i, and miR-98. These miRNAs varied in expression among the various target tissues (Fig. IV-2A), with five members generally downregulated and two others marginally upregulated in PhIP-induced tumors (Fig. IV-2B).

Targets of let-7 were validated in tissue samples from PhIP-treated animals. When all six target organs were assessed, collectively, both c-myc and Hmga2 mRNA levels, normalized to Gapdh, were increased significantly in PhIP-induced tumors compared with matched normal tissue (Fig. IV-2C, D, respectively). Notably, when peripheral blood mononuclear cells (PBMCs) were collected as negative controls from the same study, representing a site that did not develop tumors, no marked differences were detected between PhIP- and vehicle treated rats with respect to the relative expression levels of c-myc, Hmga2, and the PhIP signature miRNAs (data not shown).

In human colorectal cancers, and in "all other" cancers from the TCGA pan-cancer database, MYC and HMGA2 levels also were increased significantly compared with tissue-matched normal controls (Fig. IV-2E, F). The results are consistent with the tumor suppressor functions of let-7 family members, linked to upregulation of oncogenic targets in both rat and human primary tumors.

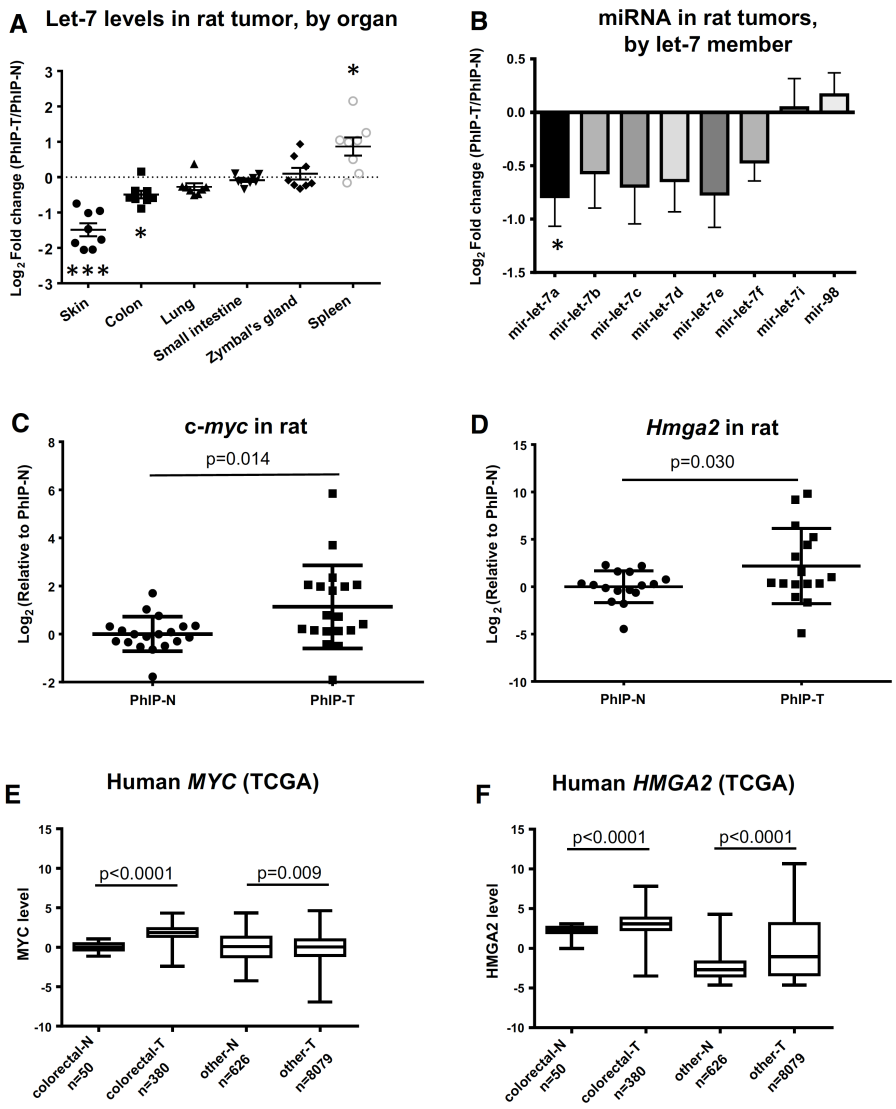


Figure IV-2 Let-7 family is altered in multiple target organs of PhIP-induced tumorigenesis.

A, Expression of let-7 members according to tissue of origin. Each point represents an individual let-7 member in PhIP-induced tumor (T), normalized to the corresponding expression in adjacent normal-looking tissue (N). B, Let-7 family members were downregulated in PhIP-induced tumors (all sites combined), with the exception of lowabundance members let-7i and miR-98. C, D, RT-PCR data for let-7 targets c-myc and Hmga2, normalized to Gapdh, in PhIP-induced tumors and adjacent normal tissue in the rat. E, F, The corresponding human MYC and HMGA2 expression data from The Cancer Genome Atlas (TCGA), for Pan-Cancer primary tumors and normal tissue samples.

PhIP signature miRNAs predict worse prognosis. A second group of highly dysregulated PhIP signature miRNAs included miR-21, a well-known oncomiR[447], and reported tumor suppressors miR-145, miR-126, miR-29c, and miR-215[156,448]. When miRNAs were compared in PhIP-induced tumors (n = 16) versus adjacent normal-looking tissue (n = 16), there was a trend towards upregulation of miR-21 and downregulation of miR-126, miR-29c, miR-145, and miR-215 (Fig. IV-3A). Statistically significant differences were noted for miR-21 ($p < 0.01$), miR-29c ($p < 0.001$), and miR-145 ($p < 0.01$).

This miRNA signature was used to profile the human pan-cancer database in TCGA, and predicted a significantly worse overall survival in cancer patients (Fig. IV-3B, red line). Metacore analysis for the corresponding miRNAs revealed a network of interactions centered on KLF4 (Fig. IV-3C, red square). Subsequent investigation confirmed that the predicted target was significantly lower in the majority of PhIP-induced rat tumors (Fig. IV-3D), and in human primary colorectal and "other" cancers (Fig. IV-3E). Thus, tumors harboring a PhIP signature miRNA profile predicted worse overall prognosis in cancer patients.

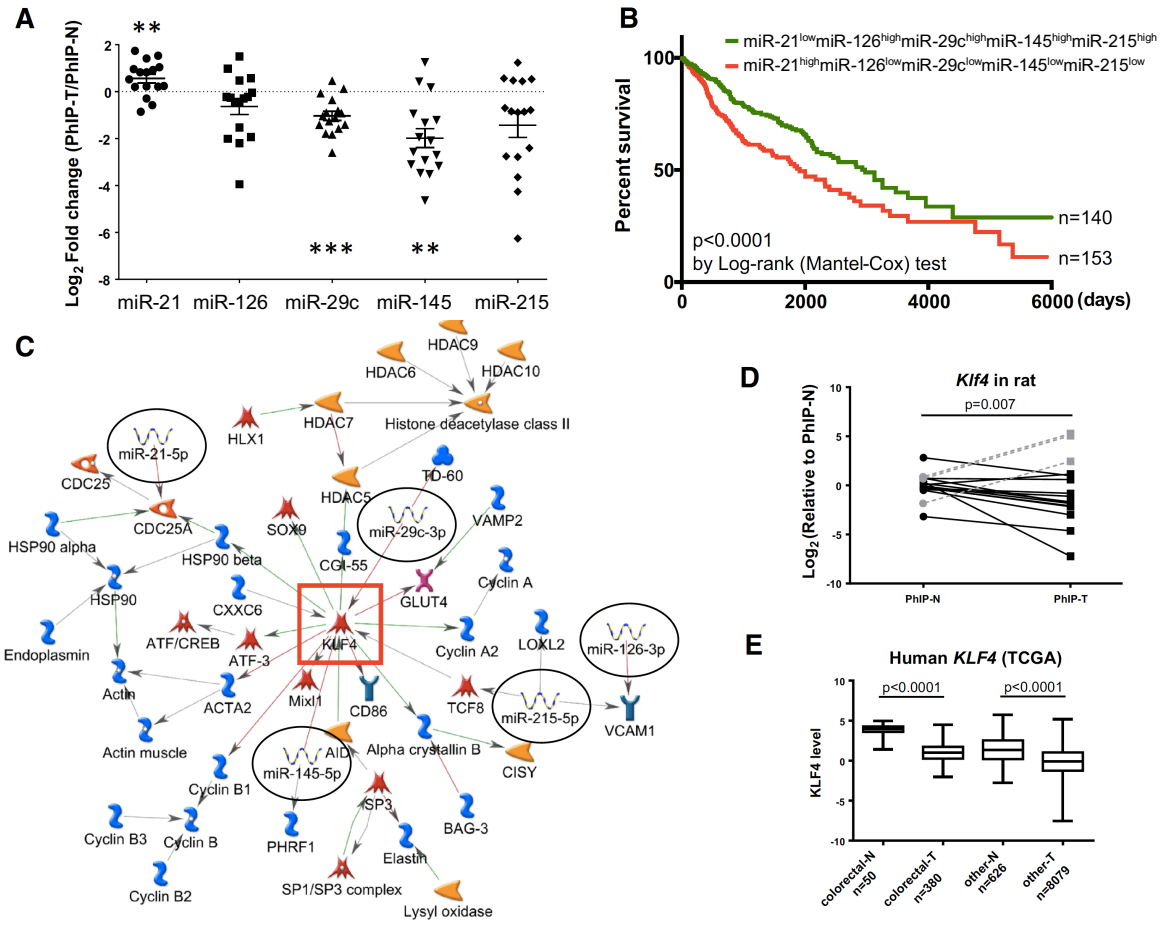


Figure IV-3 PhIP signature miRNAs predict poor prognosis.
A, Expression profile of five major PhIP signature miRNAs in tumors compared with the corresponding adjacent normal tissue; n = 16, **p < 0.01, ***p < 0.001. B, Survival curves of patients listed with primary tumors in the Pan-Cancer TCGA database, grouped according to the relative expression of PhIP signature miRNAs indicated in the figure. C, Metacore pathway analysis of five PhIP signature miRNAs defined a central node involving KLF4 (red square). D, In RT-PCR analyses, Klf4 was downregulated in PhIP-induced tumors compared with adjacent normal tissue in the rat, except for skin tumor samples, where the reverse trend was observed (gray dotted lines). E, TCGA data showing downregulation of KLF4 in human colorectal cancer and pan-cancer primary tumors (T), compared with normal tissues (N).

Overlapping PhIP signature miRNAs in human primary tumors. When TCGA data were examined in more detail, among the 13 most highly dysregulated miRNAs in PhIP-induced rat tumors, 9 were similarly changed in human colon adenocarcinomas (Fig. IV-4A), and 10 were correspondingly altered in human pan-cancer datasets (Fig. IV-4B). Functional annotation identified key pathways for these miRNAs, including MAPK signaling, endocytosis, apoptosis, Wnt signaling, and colorectal cancer (Fig. IV-4C).

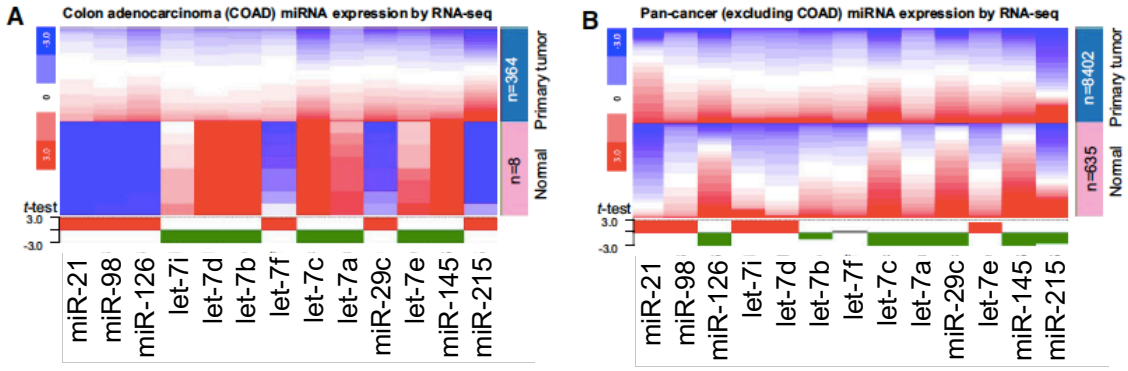


Figure IV-4 PhIP miRNA signature from rats recapitulated in human primary tumors.

A, TCGA data were mined for PhIP signature miRNA levels in human colon primary adenocarcinomas (COAD) compared with normal tissue. The heatmap is color-coded to indicate the relative gene expression level within each dataset. The expression profile of each miRNA that recapitulated the corresponding outcome in rat is shown in bold. B, Corresponding data for pan-cancer, excluding the colon adenocarcinoma (COAD) data already analyzed in A. C, Functional annotation of gene targets of PhIP signature miRNAs. miR-21, miR-145, and let-7 family members were entered into TargetScan (http://www.targetscan.org), which identified 180 predicted target genes. These predicted targets were uploaded into DAVID, and the functional annotation results for the top five KEGG pathway terms were listed, as shown.

Figure IV-4 Continued

C

Pathway	p Value	Genes
MAPK signaling	0.002	*FGF18, NTF3, RASGRP1, TGFBR1, DUSP16, TP53, FASLG, PPP3CA, DUSP8, NGF*
Endocytosis	0.0475	*ADRB2, DAB2, TFRC, TGFBR1, RAB22A, ARF6*
Colorectal cancer	0.062	*MSH2, TGFBR1, TP53, FZD3*
Apoptosis	0.067	*TP53, FASLG, PPP3CA, NGF*
Wnt signaling	0.079	*CCND2, TP53, FZD3, SKP1, PPP3CA*

Discussion

Environmental factors can impact epigenetic mechanisms that regulate the crosstalk between DNA methylation, histone modifications, and non-coding RNAs[72,156,449–452]. Historically, HCAs have been investigated in terms of their genotoxic/genetic influences and various prevention strategies[401,453–456], but few studies to date have explored the possible epigenetic aspects linked to human intake of PhIP and structurally related compounds in the diet.

Starting with the colon tumors, we performed an unbiased screen and validated the most highly dysregulated miRNAs and their downstream targets in the rat[156]. The current investigation builds upon those initial observations, testing the hypothesis that PhIP signature miRNAs in colon tumors might be applicable to other target organs of the rat.

Indeed, a relatively small and focused 13-panel miRNA profile clustered colon and small intestine separately from skin, Zymbal's gland, spleen, and lung. Similar groupings were reported following genome-wide miRNA profiling of 55 different organs and tissues in the rat[457].

Dysregulated miRNAs in tumors trigger upregulation or downregulation of mRNA targets, depending on the circumstances. Overexpression of miR-21, for example, causes loss of tumor suppressors PDCD4 and PTEN to promote cell survival[458], epithelial-to-mesenchymal transition[459], and chemoresistance[460]. In addition to targeting PDCD4, PTEN, and FASL to inhibit apoptosis [447], miR-21 can coordinate with miR-145 in regulating colon cancer stemness and chemoresistance through CD44, SOX-2, and β-catenin [461]. Interestingly, both miR-21 and miR-145 were among the PhIP signature miRNAs dysregulated in multiple target organs of the rat.

Mir-21 is one of a cadre of miRNAs altered by environmental exposures such as cigarette smoking, air pollution, diesel exhaust, nanoparticles, particulate matter, toxic metals, and diverse chemicals[81]. Environmental pollutants such as bisphenol A, polychlorinated biphenyls, arsenic, mercury, lead, and cadmium also downregulate members of the let-7 family of suppressor miRNAs [163], which can impact IL6-STAT3 signaling and epigenetic transformation[446].

Multiple let-7 family members were among the PhIP signature miRNAs, with let-7a, let-7b, let-7c, let-7d, led-7e, and let-7f being downregulated in the different target organs of PhIP tumorigenesis. In accordance with the tumor suppressor functions ascribed to let-7 members[462], we confirmed upregulation of their oncogenic targets, c-myc and

Hmga2, in the corresponding rat tumors. Interestingly, human breast cancer cells treated with bisphenol A or dichlorodiphenyltrichloroethane (DDT) also had attenuated let-7 levels[463], highlighting a potential broader impact of environmental influences on these suppressor miRNAs.

Four additional suppressor miRNAs were downregulated at multiple tumor sites in PhIP treated rats, namely miR-126, miR-29c, miR-145, and miR-215. In human stage II and III colon cancers, miR-215 is decreased and is associated with poor prognosis[464], but miR-215 also plays a role in chemoresistance[465]. As discussed above, miR-215 can coordinate with miR-21 to regulate CD44, SOX2, and β-catenin in colon cancer cells[461]. Mir-126 inhibits pancreatic[466] and breast cancer progression and metastasis[467] by targeting ADAM9 and PI3K, whereas downregulation of miR-126 is associated with poor prognosis in non-small cell lung cancer, due to dysregulated EGFL7[468]. The miR-29 family regulates DNA methylation and metastasis in lung cancer and nasopharyngeal carcinoma by targeting DNA methyltransferases[469,470].

We postulated that the miRNA profile prioritized in PhIP-induced rat tumors might serve as a "fingerprint" of environmental HCA exposure in the human population. Insufficient miRNA data were available in TCGA to interrogate the entire panel of 13 PhIP signatures miRNAs, collectively. However, a focused panel comprising of miR-21^{high}/ miR-126^{low}/ miR-29c^{low}/miR-215^{low}/miR-145^{low} predicted significantly worse overall patient survival (Fig. VI-3b). Interestingly, this subgroup consisted of 153/9723 available cases, giving an estimated frequency of 1.57% of pan-cancers in the human population. However, we are cautious not to over-interpret these findings as being directly

associated with PhIP exposure. There are more than 20 different genotoxic HCAs present in the diet[398,405]. Also, polycyclic aromatic hydrocarbons, N-nitroso compounds, lipid peroxides, and reactive oxygen species are thought to contribute to meat-associated DNA damage and cancer etiology[471,472], and might induce or repress the same miRNAs as PhIP.

Nonetheless, we were interested in the fact that PhIP signature miRNAs defined KLF4 as a central player (Fig. VI-3c), a transcription factor with known links to cell proliferation, autophagy, transformation, metastasis, and pluripotency[473]. Expression of KLF4 was reduced significantly in human primary colorectal and human pan-cancer, and Klf4 also was decreased in the majority of PhIP-induced rat tumors. In PhIP-induced skin tumors, however, Klf4 was overexpressed rather than downregulated. It is known that KLF4 can serve as a bivalent transcription factor, with activating or repressive functions on gene expression according to the circumstances[474]. In the skin, nuclear KLF4 was linked to squamous epithelial dysplasia[475], but KLF4 deficiency also correlated with increased cell proliferation and enhanced tumor outcomes in a classical mouse skin carcinogenesis model[476]. Cross-talk with other transcription factors, such as the glucocorticoid receptor[477], likely contributes to the divergent actions of KLF4 in the skin.

Finally, when PhIP signature miRNAs were compared, as a group, with TCGA data for colon adenocarcinoma and pan-cancer, most of the miRNAs in human subjects were upregulated or downregulated in the same direction as in the rat tumors. Bioinformatics analyses implicated gene ontology pathways linked to endocytosis, apoptosis, MAPK signaling, Wnt signaling, and colorectal cancer. We did observe a relatively modest change in c-myc and Hmga2 expression in PhIP-induced tumors,

compared with adjacent normal tissue, whereas the expression range of these proteins was larger in human tumors from the same organs. The wide range of expression in humans probably reflects different types of environmental exposure. Thus, it is difficult to make firm conclusions about the relative abundance of these proteins and attribute them specifically to PhIP, or other HCAs, in humans. Nonetheless, the current investigation has laid the groundwork for follow-up studies on other environmental HCAs, their miRNA signatures, and the downstream targets implicated in cancer development.

CHAPTER V

GUT MICROBIOME AND HOST TRANSCRIPTOME CHANGES IN A PRECLINICAL MODEL OF COLORECTAL CANCER GENETIC VERSUS DIETARY DRIVERS OF DISEASE OUTCOMES

Overview

There is great interest in delineating the role of gut microbes, the host, and their responses to specific dietary exposures that regulate physiological and pathophysiological development. We employed a multi-omics approach to examine these interrelationships in an *Apc*-mutant model of colorectal cancer following tumor suppression by dietary spinach. For the intervention groups, increasing diversity coincided with reversal of microbiome taxonomic composition in both wild type rats and in the polyposis in rat colon (Pirc) model of human familial adenomatous polyposis (FAP). Metagenomic prediction implicated pathways in cancer and the TCA cycle, as well as butanoate and α-linoleic acid metabolism. Transcriptomic analyses of mRNA and miRNA profiles prioritized, among other targets, inflammatory responses and TNFα signaling via NF-κB in normal-looking Pirc colon, which were further dysregulated in Pirc colon polyps, along with E2F targets, G2M checkpoint, and oncogenic signaling networks. Tumor suppression was associated with altered TNFα signaling, hypoxia, epithelial mesenchymal transition, and KRAS signaling, along with decreased *Serpine1* and increased miR-143/145 levels. Microbiome diversity was negatively associated with colon tumor outcome, and several Ruminococcaceae genera were correlated with anti-tumor gene signatures involving

Serpine1, *Plcd1*, *Tcf7l1* and *Slc30a10*. We conclude that dietary spinach reshapes the gut microbiota and impacts host transcriptome changes, contributing to tumor suppression outcomes in the colon.

Introduction

The mammalian gut microbiome is strongly implicated in host physiological and pathophysiological development[478–480]. Studies in germ-free models of colorectal cancer (CRC) revealed decreased bowel inflammation and carcinogenesis[481,482]. Fecal microbiota transplantation has been used successfully to treat recurrent *Clostridium difficile* infection[483], and provides benefit to patients with inflammatory bowel diseases, functional gastrointestinal (GI) disorders, and obesity[484]. There is increasing interest in defining specific dietary interventions that might alter the gut microbiota in a beneficial manner for disease prevention and treatment. Epidemiological studies show that low consumption of green vegetables and fiber is associated with increased CRC risk, whereas intake of dark leafy vegetables is linked to decreased risk[485]. Little is known, however, about how these dietary intake patterns influence the crosstalk between gut microbiota, host transcriptomic profiles, and disease outcomes. Spinach is a dark leafy vegetable with a high chlorophyll content and appreciable levels of bioactive phytochemicals such as carotenoids, betaine, flavonoids, and polyunsaturated fatty acids[486–488]. While evaluating anticancer strategies in the *Apc*-mutant polyposis in rat colon (Pirc) model[55,489,490], we observed that dietary spinach was associated with significant suppression of adenomatous polyps in the Gastrointestinal (GI) tract. The Pirc model mimics disease progression in human familial

adenomatous polyposis (FAP) patients, involving spontaneous tumor development both in the colon and in the small intestine[53,491]. This genetic model circumvents the need for carcinogen treatment, which in prior studies revealed anticancer effects of dietary spinach in the colon[156]. We report here, for the first time, a multi-omics approach that examined gut microbiome and host transcriptomic changes in the Pirc model, and the implications for genetic *versus* dietary drivers of colon tumorigenesis.

Materials and methods

Preclinical studies. Experiments were approved by the Institutional Animal Care and Use Committee. After weaning, Pirc (F344/NTac-*Apc*^{am1137}, Taconic Farms, Inc. USA) and wild-type F344 male rats were assigned randomly to basal AIN93 control diet (Ctrl) or AIN93 diet containing 10% freeze dried spinach by weight (SPI). Diets were replenished every 2 to 3 days, and animal body weights were monitored weekly. Monthly endoscopy was employed for temporal tracking of polyp development in the colon, as described in prior studies[55]. Each polyp was assigned a unique 'address' in the colon, based on a reported methodology.[54] Prior to termination, 3-4 rats in each group were injected with bromodeoxyuridine (BrdU, 100 mg/kg body weight), and animals were euthanized 1 h later by CO_2 inhalation. A thorough necropsy examination was performed, and tissue samples were taken for histopathology and molecular analysis, as reported[55].

Histology, immunostaining and immunoblotting. Tissue sections (5 μm) were stained by hematoxylin & eosin (H&E). Immunostaining with antibodies for β-catenin (BD

Biosciences # 610153), cleaved caspase-3 (Cell signaling #9661), and BrdU (BD Biosciences # 347580) was conducted at the Research Histology, Pathology and Imaging Core, The University of Texas MD Anderson Cancer Center. A BrdU labeling index was determined, as described[417], and cleaved caspase-3 was quantified as percent positive-stained crypts in a given field. Two to three biological replicates were employed for WT-Ctrl, WT-SPI, Pirc-Ctrl, and Pirc-SPI groups. For each tissue section, at least 15 independent fields were quantified in the proximal, middle, and distal regions of the colon. Other proteins were immunoblotted as reported[492], using Cell Signaling primary antibodies for PARP (#9542), β-catenin (#9581), p53 (#9282), and cyclin D1 (#2926), with β-actin as loading control (Sigma, A1978).

Microbiome analysis. Frozen cecum content from rats in each group was submitted for bacterial genomic DNA extraction at the Alkek Center for Metagenomics and Microbiome Research (CMMR), Baylor College of Medicine. The 16S rDNA V4 region was amplified and barcoded through PCR and sequenced using the MiSeq platform (Illumina) with the 2x250 bp paired-end protocol. Operational Taxonomic Units (OTUs) at a similarity cutoff value of 97% were generated by the UPARSE algorithm and mapped to SILVA database. OTU tables and Agile Toolkit for Incisive Microbial Analyses (ATIMA) were provided by CMMR for primary data visualization. ATIMA microbiome data were subjected to the Kruskal-Wallis test, with the adjusted P-values indicated in the figures. LDA Effect Size (LEfSe) used rarefied OTU data from CMMR, and results were generated with LDA>2 and $p<0.05$ (Kruskal-Wallis test)[493]. Phylogenetic Investigation of Communities by

Reconstruction of Unobserved States (PICRUSt) used raw BIOM files to map Greengenes for metagenome prediction[494], coupled to Statistical Analysis of Metagenomic Profiles (STAMP)[495]. Results were replotted using Prism or Excel. To associate microbiome and disease outcome, we calculated the corresponding diversity and OTUs with matched tumor multiplicity and tumor volume datasets. At the family and genus level, tumor outcomes were plotted by linear regression analysis, coupled to Pearson correlation coefficient. To integrate microbiome OTUs with host gene predisposition and anticancer outcomes, we calculated the Pearson correlation by group for enriched microbiome OTUs from LEfSe (genus level) and the reads per kilo base per million mapped reads (RPKM) of the host transcriptome.

RNA analyses. Flash-frozen rat tissues (n=3 replicates per group) were subjected to total RNA extraction using a Qiagen miRNeasy mini kit, according to the manufacturer's instructions. After DNase digestion and purity/integrity confirmation, as reported[55], RNA samples were run on a 1% denaturing agarose gel prior to library preparation for RNA and small RNA sequencing. For RNA library preparation, NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) and NEBNext® Ultra™ Directional RNA Library Prep Kit (NEB, E7420) were performed with 3μg RNA input as the starting material. For small RNA library preparation, NEBNext® Small RNA Library Prep Set for Illumina® (NEB, E7300) were performed with 1μg RNA input as the starting material. The RNA went through 3' ligation, primer hybridization, 5' ligation, 1st strand synthesis, PCR amplification and size selection (MinElute Gel Extraction Kit, Qiagen). During

library construction, each sample was tagged with a specific index during at the PCR amplification step. The library concentration was measured by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and the size of the library was accessed by 2% agarose gel with 2ng library. The samples were diluted to 1pmol for further library pooling. Before loading the pooled library into the NextSeq 550/500 system (Illumina), 1nM library was denatured and diluted according to NextSeq System Denature and Dilute Libraries Guide (Illumina). Stranded RNA sequencing was conducted by paired-end sequencing with 75 bp read length and small RNA sequencing was conducted by single-end sequencing with 50 bp read length. Fastq files were generated on BaseSpace for further bioinformatics processing and analysis.

Bioinformatics and differentially expressed genes (DEGs). RNA-seq data were mapped to rn6 genome assembly using bowtie2.2.5[496] with default parameters. The count table for each gene were calculated by using HTseq-count[497]. The DEGs were called using DESeq2 with an adjusted p value less than 0.05. miRNA-seq raw fastq data were first trimmed the adaptor sequence using cutadaptor[498]. Reads with exactly the same sequence were collapsed and counted. Mature and hairpin microRNA sequence data for rat were downloaded from miRbase (http://www.mirbase.org/ftp.shtml) on June 23rd,2016. Collapsed reads were mapped to rat mature microRNA sequences using blastn. The counts table was input to DESeq2 to call differential expressed miRNAs (DEmiRs) with adjust p value <0.05. Principal Component Analysis (PCA) was performed using DESeq2 for both RNA-seq and miRNA-seq. DEGs were further analyzed by GSEA Java Desktop

application v3.0 Beta 2. The cutoff for significance was using FDR<25%. 6 miRNAs were prioritized for further target RNA analysis. Target prediction was conducted by Targetscan 7.2 algorithm[442] for rat specific analysis, and only the miRNA-mRNA pairs with correlation value less than -0.5 were retained in the Cytoscape view[499].

Quantitative RT-PCR. The basic methodology for RNA extraction and purification was as reported, using three replicates per group[201,490,500–502]. After SuperScript III (Thermo Fisher Scientific) or miScript RT II (Qiagen) kits, quantitative PCR (qPCR) reactions were performed by LightCycler® FastStart DNA Master SYBR Green (Roche Applied Science) on a LightCycler96 instrument. Primers listed in Table V-1 for mRNA qPCR were designed by NCBI-BLAST, whereas primers for miRNA qPCR were purchased from Qiagen miScript Primer Assays. Internal controls were *Gapdh* and *U6B* small nuclear RNA, for mRNA and miRNA analyses, respectively. Selected RNAs and miRNAs were based on prior validation and ranking in GSEA data. Original gene lists for tight junction and anti-microbial function were obtained from KEGG and GO, respectively, and further sorted by the sequencing data to generate genes of interest for RT-PCR. We focused on six miRNAs consistently altered during tumor formation in Pirc, and filtered miRNA-mRNA pairs by conserved UTR target site in human and rat, with a linear correlation <-0.7 in sequencing data. We verified miRNAs and mRNAs by qPCR (Pearson's test, with $r < -0.5$).

Statistical analyses. Statistical analysis of two-group comparisons was performed using an unpaired two-tailed *t*-test. Correlation analysis was performed by linear and Pearson's correlation, for miRNA-RNA target and microbiome-host correlations (tumor and diet groups). For matched tumor outcome-microbiome correlation, seven biological replicates of Pirc/Ctrl or Pirc/SPI were considered, whereas for spinach responsive genes-microbiome correlations, average levels of the three groups were considered.

Results

Dietary spinach exerts anti-tumor activity in a genetic model of FAP. Pirc and wild type (WT) male rats were fed basal AIN93M diet (Ctrl), or basal diet containing 10% freeze-dried spinach (SPI), starting at 4 weeks of age (Fig. V-1A). No significant treatment-related effects were observed with respects to food consumption and body weight throughout the study (Fig. V-1B). Applying an established endoscopy methodology in the Pirc model,[54,55] SPI was found to suppress colon polyps as early as 12-weeks into the experiment (Fig. V-1C), *i.e.*, after 8-weeks of SPI treatment. During this period and at later times, endoscopy revealed significant inhibition of small colon polyps, and of large colon polyps after week 20 (Fig. V-1D). When the experiment was terminated at 30 weeks, tumor multiplicity was decreased significantly both in the colon and in the small intestine, and tumor volume also was reduced significantly in the colon by SPI treatment (Fig. V-1E). No marked changes were observed in the histology of normal-looking colon or colon polyps in the treatment groups (data not shown), and the modest increase in cleaved-caspase 3 for SPI-treated rats was not statistically significant

(data not shown). In Pirc and WT controls, BrdU labeling suggested reduced cell proliferation rates by SPI, although this was not consistent for all regions of the colonic crypt (data not shown). Immunohistochemical analyses confirmed the anticipated overexpression of β-catenin in colon polyps and adjacent normal-looking colon and, notably, this was unaffected by SPI treatment (Fig. V-1F). We concluded that although the Wnt/β-catenin signaling pathway is deregulated in the *Apc*-mutant Pirc model, other pathways or mechanisms likely accounted for the tumor suppression observed by dietary SPI.

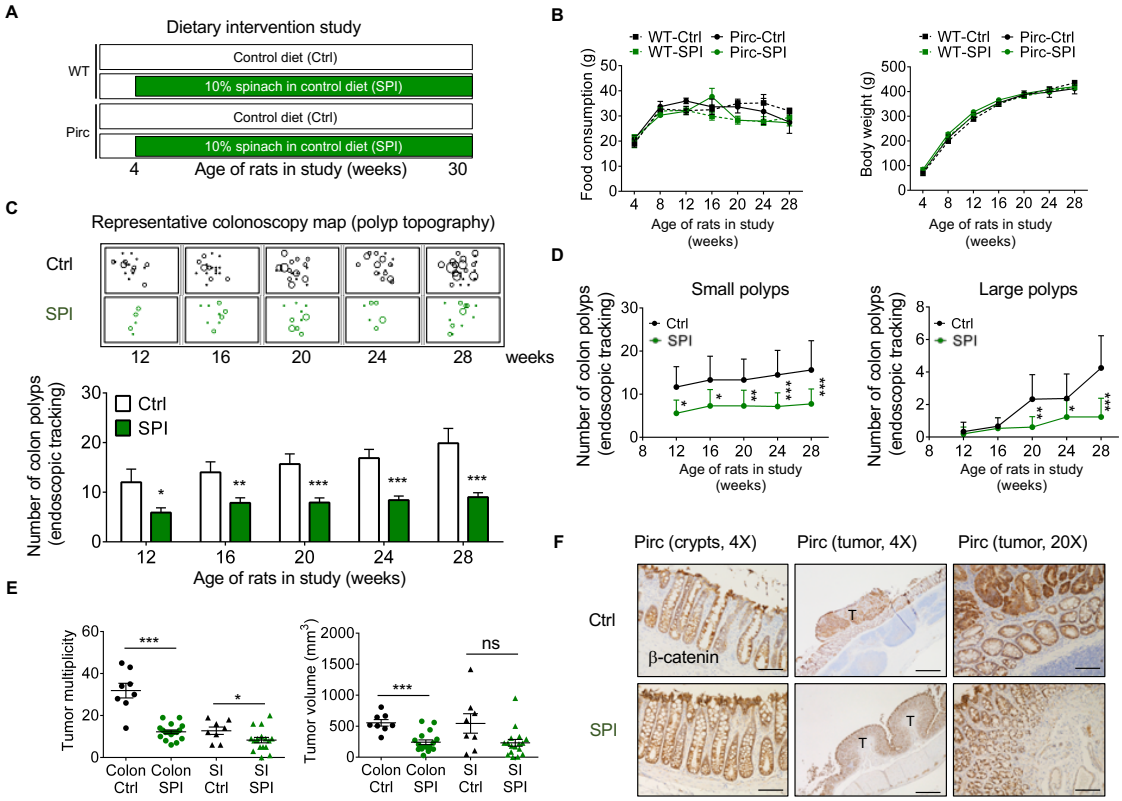


Figure V-1 Dietary spinach suppresses polyp formation in a preclinical model of FAP.

A, Study design in Pirc and WT male rats; n=10 for WT-Ctrl and WT-SPI, n=8 for Pirc-Ctrl, and n=17 for Pirc-SPI. B, Food consumption and animal body weight gain were plotted for monthly changes among the 4 groups. C, Monthly endoscopic tracking of colon polyps. Topography images reflect progressive changes observed in polyp size and location, whereas the bar graph represents average multiplicity via endoscopy. D, Number of small (size grade 1-3) and large polyps (size grade 4-5) by monthly endoscopy, based on a reported grading system[54]. E, Final tumor outcomes for colon and small intestine (SI) polyps at the 30-wk necropsy. F, Immunohistochemistry staining of β-catenin in Pirc males given Ctrl or SPI diets. Scale bar indicates 100 μm in 20X images, or 500 μm in 4X images. Numerical data are presented as mean ±SEM; *p<0.05, **p<0.01, ***p<0.001, ns statistically non-significant.

The gut microbiota is profoundly altered by dietary spinach. We performed 16S rRNA sequencing of the cecum/gut microbial community in Pirc and WT rats. The observed Operational Taxonomic Units (OTUs) and Shannon index revealed that α-diversity was not altered by host genotype, but was significantly increased by SPI in Pirc and WT rats (Fig. V-2A). There was no clustering separation between Pirc and WT rats for weighted UniFrac principal coordinates analysis (PCoA) (Fig. V-2B), but a significant separation was observed in unweighted UniFrac PCoA (Fig. V-2C). The gut microbiome in both host genotypes clustered separately in weighted UniFrac PCoA between Ctrl and SPI groups (Fig. V-2B), with a marked shift in unweighted UniFrac PCoA (Fig. V-2C). The results were consistent with previous studies that used a diverse range of inbred and outbred mouse strains[338], and concluded that diet plays a dominant role over genetic background in shaping interindividual variations in host-associated microbial communities[245]. Pirc had a higher abundance of Bacteroidetes and Proteobacteria than WT rats, while Firmicutes and other phyla were lower (Fig. V-2D), which is consistent with prior findings from the proximal colonic mucosa microbiome in $Apc^{Min/+}$ mice[503]. Interestingly, these abundances were reversed by SPI intake in Pirc, such that Pirc and WT rats responded similarly to SPI intake, independent of host genotypes (Fig. V-2E). On Ctrl diet, Pirc had a reduced abundance of Lachnospiraceae and increased abundances of Ruminococcaceae, Bacteroidaceae, Desulfovibrionaceae and Erysipelptrichaceae compared to WT, and these patterns were reversed by SPI (Fig. V-2F). As with the phylum- and family- level observations (Fig. V-2E-F), at the genus level, abundances were also reversed by SPI in Pirc, and WT rats also responded in a similar manner (Fig. V-2G). Our results suggested

that SPI re-shapes the microbiome composition, reversing the effects of host genetic predisposition in the Pirc model.

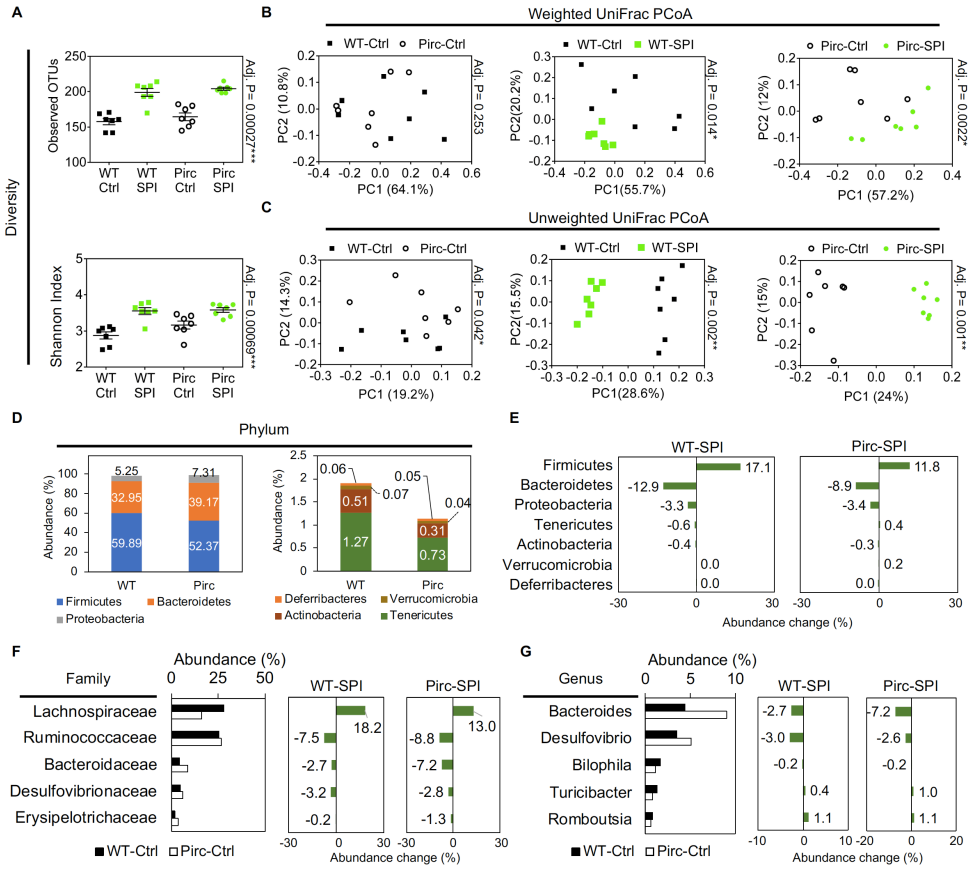


Figure V-2 Microbial composition in relation to host genetics and dietary intervention.

A, α-diversity measured by observed OTUs and Shannon index; n=7 replicates for 16S sequencing in each group. Weighted UniFrac (B) and unweighted UniFrac (C) Principal coordinate analysis (PCoA) plots displaying β-diversity of host genetics and dietary intervention. D, Taxonomic composition in Pirc and WT rats at the phylum level. E, Relative abundance changes by dietary spinach in Pirc and WT rats at the phylum level. F, Left, top five most abundant family members. Right, abundance changes by SPI in Pirc and WT rats. G, Left, top five most abundant genera. Right, abundance changes by SPI in Pirc and WT rats. Adjusted P values were calculated by Kruskal-Wallis test.

We used linear discriminant analysis (LDA) effect size (LEfSe) to analyze the OTU microbiome data. Twenty-three, sixty-nine, and seventy-two differentially abundant genera were identified, respectively, between Pirc and WT rats on Ctrl diet, WT-Ctrl *vs.* WT-SPI, and Pirc-Ctrl *vs.* Pirc-SPI. From the corresponding cladograms (Figs. V-3A-C), host genotype and dietary SPI intake both influenced Ruminococcaceae and Lachnospiraceae family members. In response to SPI, LEfSe analyses revealed that Pirc and WT rats shared ~50% common changes at the genus level (Fig. V-3D). We next applied Phylogenetic Investigation of Communities by Reconstruction of Unobserved State (PICRUSt) and categorized 328 terms following Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Two terms, ether lipid metabolism and alpha-linolenic acid metabolism were altered significantly in Pirc *vs.* WT rats. Notably, PICRUSt analysis revealed a marked effect of SPI intake. In response to SPI, 85 terms (57 decreased and 28 increased) and 112 terms (71 decreased and 41 increased) were changed significantly in Pirc and WT rats, respectively, and seventy-five terms overlapped between the two genotypes, *i.e.*, 54 decreased and 21 increased (Fig. V-3E). We identified increased membrane transporters, cell motility, signal transduction, transcription, carbohydrate metabolism, and kinases among the top 10 terms prioritized by KEGG analysis in Pirc and WT rats given SPI (Fig. V-3F). The corresponding decreased terms were related to translation, replication/repair, and energy/nucleotide metabolism. Porphyrin and chlorophyll metabolism, identified after PICRUSt/KEGG, likely reflects the chloroplast DNA detected during 16S sequencing, excluded in the microbiome analysis. Notable pathway changes included an increase in alpha-Linolenic acid and

butanoate (butyrate) metabolism, and a decrease in TCA cycle and Pathways in cancer (Fig. V-3G).

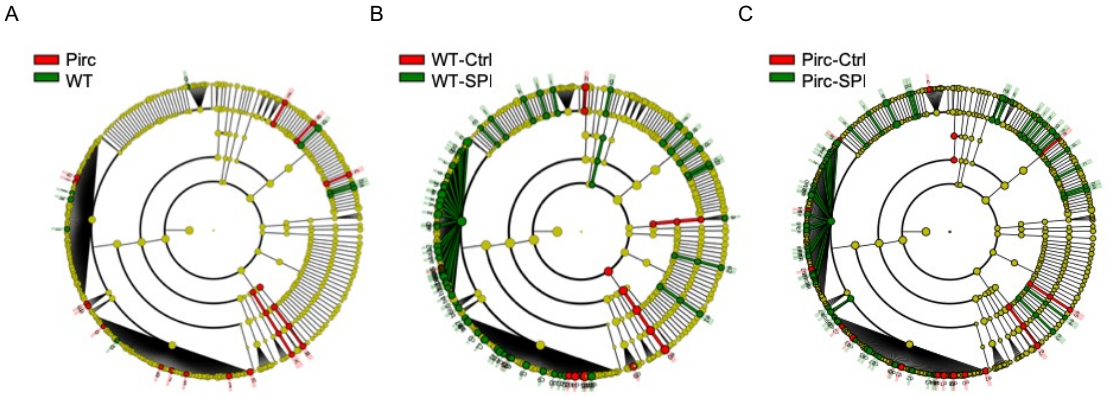


Figure V-3 Dietary spinach shapes the microbiome and its predicted metagenome. Cladograms and LDA scores showed enriched clades between two-group comparisons from LEfSe: WT vs Pirc (A), WT responding to SPI (B), and Pirc responding to SPI (C). Cladograms illustrate all significant changes in taxonomy, whereas bar graphs below show representative enriched bacteria taxonomy. D, Venn diagrams of enriched clades in Pirc and WT rats responding to SPI. E, Venn diagrams of metagenome prediction from PICRUSt in Pirc and WT rats responding to SPI. F. Top ten functional predictions of metagenome in Pirc and WT rats given SPI. G, Alpha-linolenic acid (p=2.19e-5) and butanoate metabolism (p=1.44e-9) were significantly increased, and TCA cycle (p=9.86e-4) and Pathways in Cancer (p=2.85e-3) were decreased by SPI Pirc and WT rats. Statistical analysis was performed in STAMP with ANOVA.

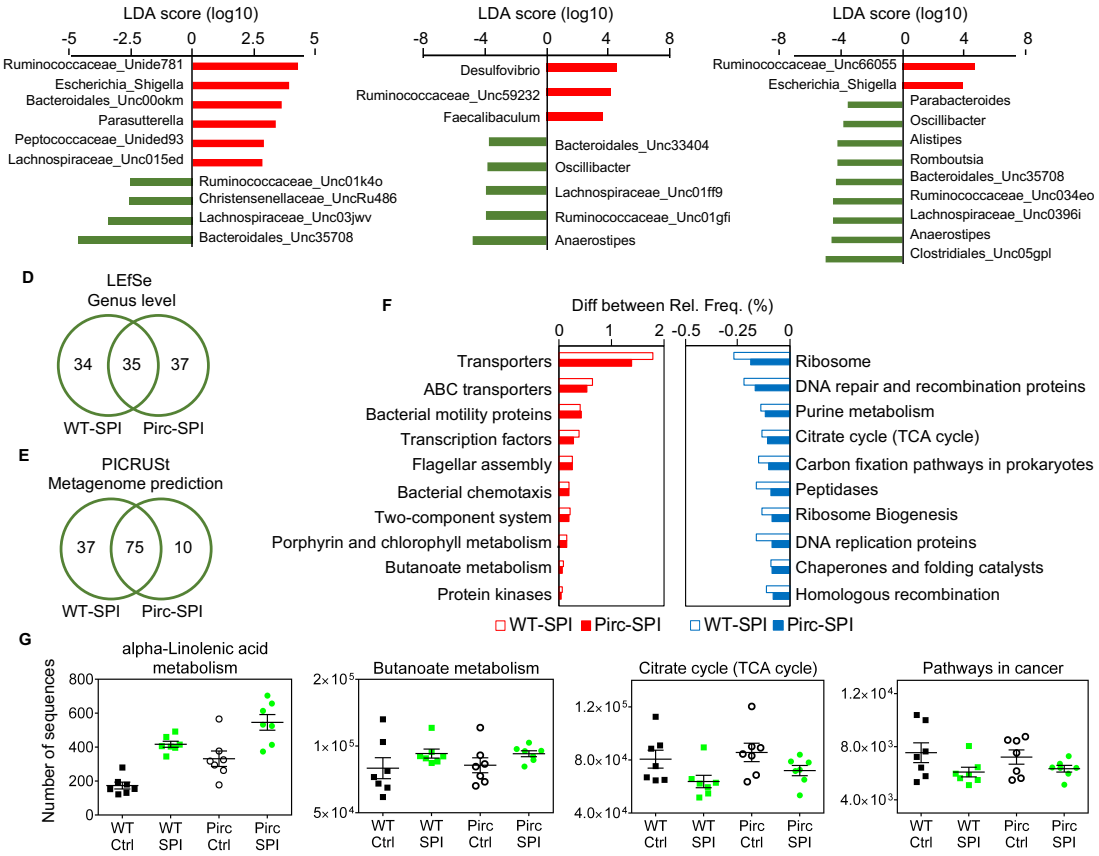


Figure V-3 Continued.

Spinach impacts key genes associated with tumorigenesis in the Pirc model. We utilized RNA-seq to map 17378 transcripts in colonic tissues from Pirc and WT rats. Principal component analysis (PCA) segregated tumor tissues from normal tissues (Fig. V-4A). Host genetics (Pirc *vs.* WT rats) and SPI consumption had less effect on overall transcriptome levels than tumor development in the colon. There were 261 differentially expressed genes (DEGs) identified between Pirc and WT normal-looking tissues, half of which (138 genes) overlapped with the 2180 DEGs associated with tumorigenesis (Fig. V-4B). Heatmaps of DEGs showed a distinct tumor feature compared to the expression

profiles from Pirc and WT normal-looking tissues (Fig. V-4C). Gene Set Enrichment Analysis (GSEA) overlapping with HALLMARK identified nine significantly altered pathways in Pirc adjacent normal-looking colon (PCN) compared to WT normal colonic tissue (WCN) (Fig. V-4D, upper panel), indicating differences in host genetics for Pirc *vs.* WT, and five of these pathways were further altered during tumorigenesis (Fig. V-4D, PCN → PCT, lower panel, underlined). The five pathways included Inflammatory Response, Allograft Rejection, TNFα signaling through NF-κB, Adipogenesis, and Myogenesis. As expected in the *Apc*-mutant Pirc model, Wnt/β-catenin signaling was upregulated significantly in Pirc tumors (Fig. V-4D, PCN → PCT, lower panel, red font). Other pathways of note were related to cell cycle changes, immune response, oxidative stress, and metabolism. Immunoblotting corroborated changes at the protein level for PARP cleavage, β-catenin overexpression, decreased p53, and increased cyclin D1 in Pirc tumors compared to adjacent matched normal and WT normal colonic tissue (data not shown). After further validation by real-time PCR (RT-PCR), genes upregulated significantly in Pirc colon tumors *vs*. Pirc adjacent normal-looking colon included *Cxcl6*, *Serpine1*, and *Il-1b*, whereas *Hspb8*, *Tpm2* and *Fhl1* were downregulated significantly (Fig. V-4E). Compared to adjacent normal colon, upregulation of *Defa6* and *Bcl3* and downregulation of *App*, *Myh11*, and *Myl9* in Pirc tumors indicated changes in tight junctions and anti-microbial activity in Pirc tumorigenesis (Fig. V-4F). We also mapped 559 miRNAs with small RNA-seq and revealed the separation of tumor profiles from normal-looking profiles (Fig. V-4G). Similar to RNA profiles, the heatmap indicated that 'normal' Pirc colonic mucosa was similar to WT normal colon regardless of diets, while

marked changes occurred during tumorigenesis (Fig. V-4H). Among the 115 differentially expressed miRNAs (DEmiRs) associated with tumorigenesis in Pirc (Fig. V-4H), we used TargetScan and both RNA and small RNA sequencing datasets to identify miRNA-RNA pairs of the most altered miRNAs during Pirc tumorigenesis (Fig. V-4I), and validated significant changes in miR-215, miR-143, and miR-145 in Pirc normal colon compared to colon tumors (Fig. V-4J). Other candidates, such as miR-146b, miR-34a, and miR-21 did not quite reach statistical significance.

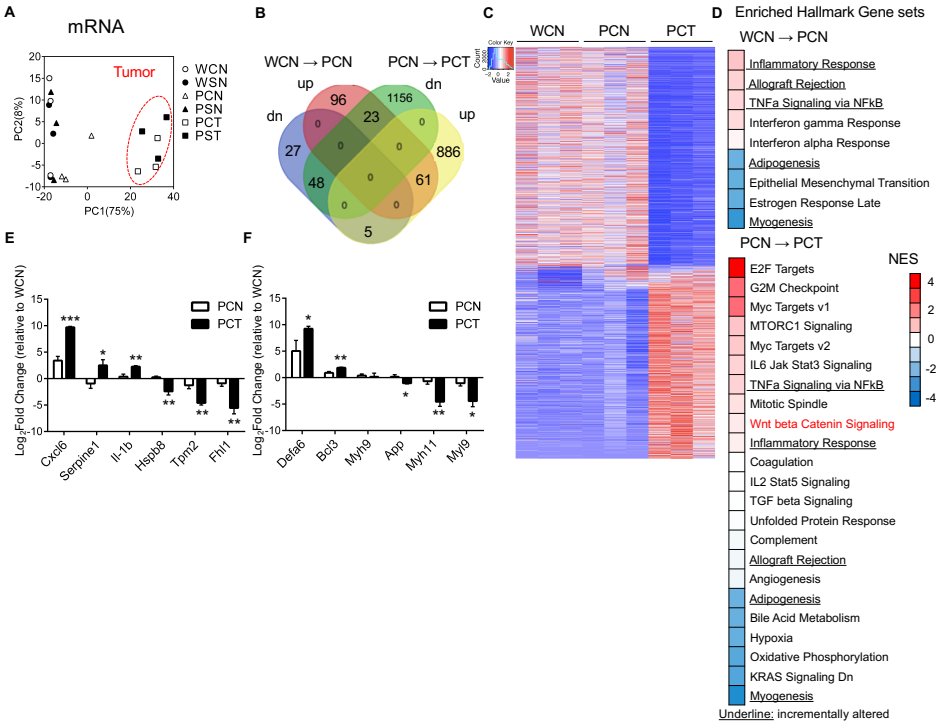


Figure V-4 Transcriptomics of colonic tissues from Pirc and WT rats.
Transcriptome analyses were performed at the 30-wk timepoint (Fig V-1A). WCN, WT-control diet-normal colon mucosa; WSN, WT-SPI diet-normal colon mucosa; PCN, Pirc-control diet-normal looking colonic mucosa; PCT, Pirc-control diet-colon tumor; PSN, Pirc-SPI diet-normal looking colonic mucosa; PST, Pirc-SPI diet-colon tumor. Each group contain 3 biological replicates, and Pirc normal-looking colon was matched to Pirc tumor. A, PCA highlighted the distribution of RNA-seq. B, Venn diagrams of DEGs. C, Heatmap of overall DEGs of RNA-seq data from WCN, PCN and PCT. D, Enriched GSEA HALLMARK analysis of DEGs between WCN and PCN (upper panel), and between PCN and PCT (lower panel). Underlined terms indicate incrementally altered pathways for WCN → PCN to and PCN → PCT group comparisons. Normalized enrichment score (NES) is indicated by color scale; -4 (blue) to +4 (red). Wnt pathway (in red font) was enriched, as expected in the Apc-mutant Pirc model. E, Validation by qPCR of genes in the five incrementally altered pathways (underlined in d). F, Validation by qPCR of genes related to tight junctions and antimicrobial function. G, PCA highlighted the distribution of small RNA-seq. H, Heatmap of overall DEmiRs. I, Cytoscape view of conserved miRNA-RNA targets and the negative associations in the rat RNA-seq and small-RNA seq data. J, Validation of selected miRNAs in Pirc colon tumors. *p<0.05, **p<0.005, ***p<0.001. Two-tailed t-test was performed for qPCR results in two-group comparisons. Numerical data are presented as mean±SEM, n=3.

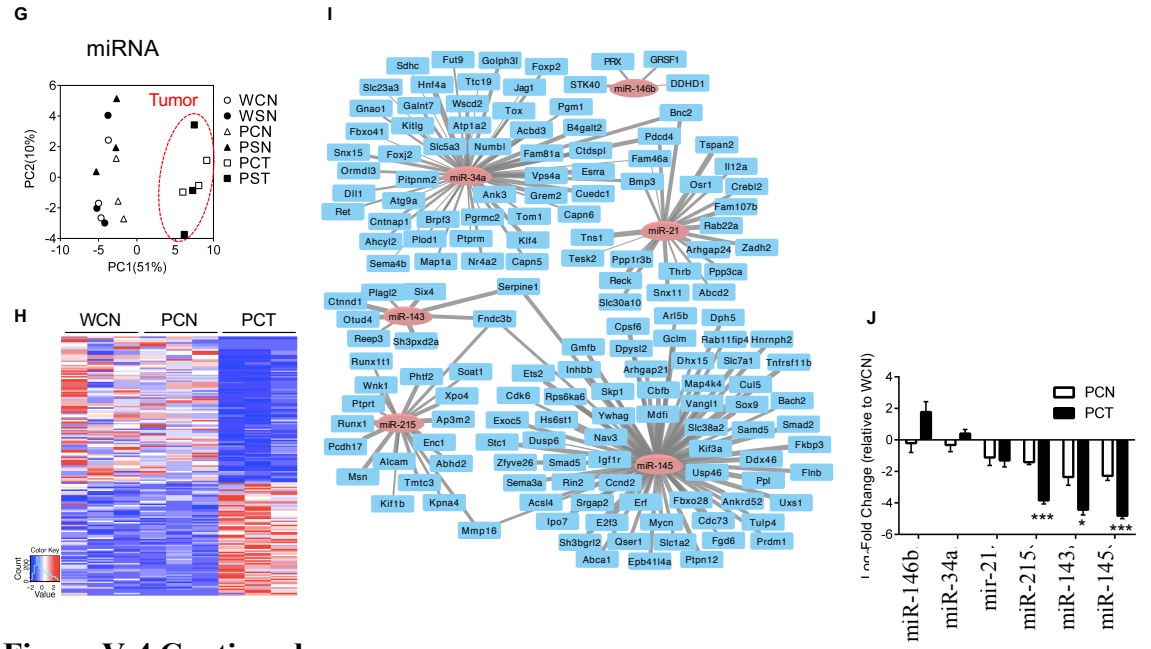


Figure V-4 Continued.

The prior analyses did not consider spinach effects on predicted targets, and the corresponding data are summarized in Fig. V-5. Compared to the control diet group, dietary spinach altered 4 genes in common among the 101 DEGs in WT rats and 80 DEGs in Pirc normal colon (Fig. V-5A). GSEA indicated significant downregulation of cell cycle-related pathways and upregulation of immune-related pathways in Pirc and WT rats fed SPI, with five pathways in common among the genotypes (Fig. V-5B). The latter pathways included TNFα Signaling through NFκB, Hypoxia, Epithelial Mesenchymal Transition, Apoptosis, and KRAS Signaling Up. After SPI consumption, 2945 DEGs were identified in Pirc colon tumors compared to adjacent normal colon, and 1754 of the DEGs also were detected in tumors *versus* adjacent normal colon from rats given control diet

(Fig. V-5C). Among the pathways most strongly implicated were IFN-α and IFN-γ for tumors from SPI-fed rats compared to rats given control diet.

While evaluating the preventive actions of SPI, we considered two scenarios: (i) genes that were up- or downregulated in Pirc colon tumors, relative to adjacent normal colon, and were reversed by SPI in Pirc normal colon, and (ii) genes that were 'normalized' in colon tumors from Pirc rats given SPI compared with colon tumors from Pirc rats given control diet. The first scenario involves putative gene targets altered by SPI in normal colon, representing 'primary prevention' of early lesions (e.g., aberrant crypt foci and microadenomas), before they advance to frank adenoma or carcinoma stages. Gene candidates related to scenarios (i) and (ii), above, were validated by RT-PCR (Figs. V-5D and V-5E, respectively). Tumor-associated genes that were reversed in Pirc normal colon by SPI consumption included *Serpine1*, *Itga6* and *Duoxa2*, whereas *Tcf7l1*, *Plcd1* and *Slc30a10* were upregulated (Fig. V-5D). Comparing tumor to tumor, *Ccl21* and *Klf7* were partially normalized by SPI ingestion relative to basal diet controls (Fig. V-5E). On the other hand, 66 DEmiRs were found in tumors from SPI-fed rats, with 41 DEmiRs also detected in tumors from animals on Ctrl diet (Fig. V-5F).

After investigating the miRNA-RNA pairs conserved in humans and with negative associations in the rat (Fig. V-4J), loss of miR-145 was associated with increased *Serpine1*, whereas gain of miR-34a was associated with reduced *Klf4* (Fig. V-5G) in Pirc tumors. The negative correlation for miR-145/*Serpine1* was further enhanced by SPI consumption, whereas the trend for miR-34a/*Klf4* was reversed by SPI in Pirc colon tumors (Fig. V-5G).

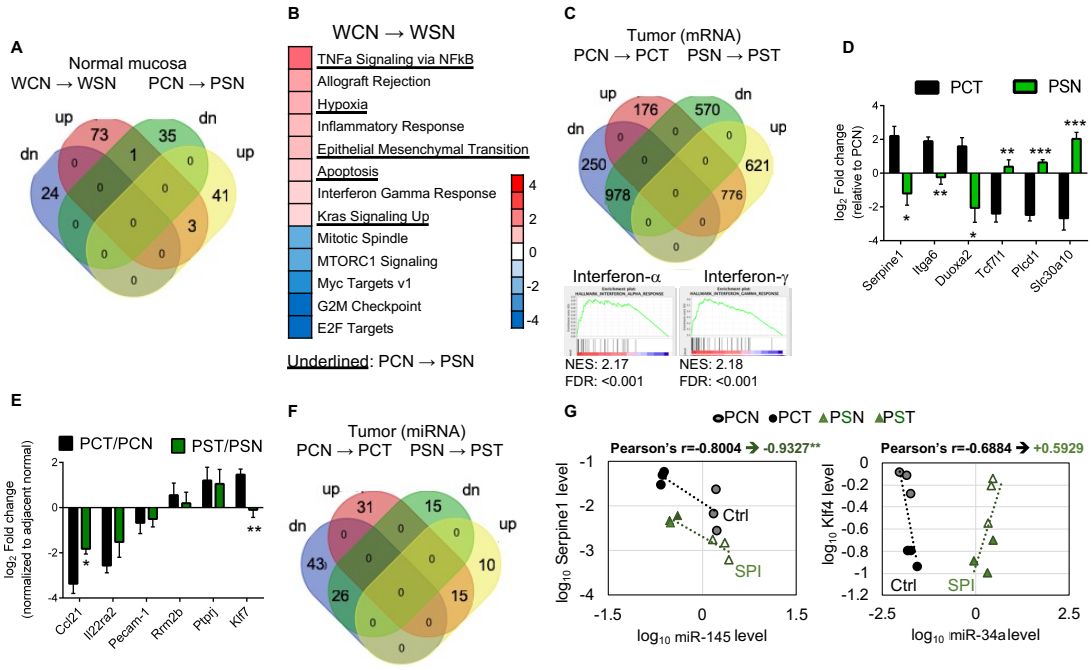


Figure V-5 Transcriptome changes due to SPI treatment in Pirc and WT rats.
A, Venn diagram of DEGs for RNAs altered by SPI in normal mucosa of Pirc and WT rats. B, Statistically significant pathways from GSEA HALLMARK analysis, highlighting SPI effects on WT normal colon. Underlined terms indicate affected pathways in Pirc normal-looking colon. C, Venn diagram of tumor DEGs in Pirc tumors under Ctrl or SPI diet consumption, and GSEA enrichment in Pirc tumors arising despite SPI treatment (i.e., putative 'SPI-resistant' lesions). D, Validation by qPCR of genes reversed in PCT vs. PSN groups. E, Partially normalized targets associated with oncogenic effects in tumors. F, Venn diagram of tumor DEmiRs in Pirc tumors from rats given Ctrl or SPI diet. G, Correlation of miRNA-mRNA target relationships, validated by RT-PCR in Pirc tissues from Ctrl and SPI groups. See Table V-2 for further miRNA-mRNA prediction data. *p<0.05, **p<0.005, ***p<0.001. Numerical data are presented as the mean±SEM, n=3. For Pearson's correlation, degree of freedom(N-2)=4 and the critical value=0.917 (p=0.01)

Association between microbiome and host responses. We integrated the data on tumor outcomes (Fig. V-1) with α-diversity calculated by observed OTUs (Fig. V-2), and observed a significant negative association for tumor multiplicity but not tumor volume (Fig. V-6A). Tumor multiplicity was inversely correlated with three unclassified Bacteroidales families (Fig. V-6B). At the genus level, one unclassified Lachnospiraceae and one unclassified Ruminococcaceae genus were negatively correlated to tumor multiplicity, while one unclassified Ruminococcaceae genus was positively correlated, based on 16S sequencing (Fig. V-6C). Metagenome predictions in relation to tumor multiplicity outcomes found significant inverse correlations for butanoate metabolism and calcium signaling, and positive associations for peptidases and pathways in cancer (Fig. V-6D). In addition to the altered patterns in the bacteria populations correlated to tumor outcomes (Figs. V-6A-6D), we also compared microbiome to host gene expression changes (Fig. V-6E). A significant positive correlation was observed for (i) Lachnospiraceae (Unc0396i) and *Slc30a10*, a manganese efflux transporter; (ii) Bacteroidales (Unc00krl) and *Plcd1*, a tumor suppressor; (iii) Ruminococcaceae (Unc01k4o) and *Plcd1*; (iv) Ruminococcaceae (Unide781) and *Serpine1*, a serine protease inhibitor and (v) Ruminococcaceae (Unc00vst) and *Serpine1*. Negative correlations were detected for (i) Ruminococcaceae (Unide781) and *Tcf7l1*, a transcription factor; and (ii) Ruminococcaceae (Unc00vst) and *Tcf7l1*. We concluded that the microbiome contributes to host tumor outcome and host transcription modulation in response to dietary spinach.

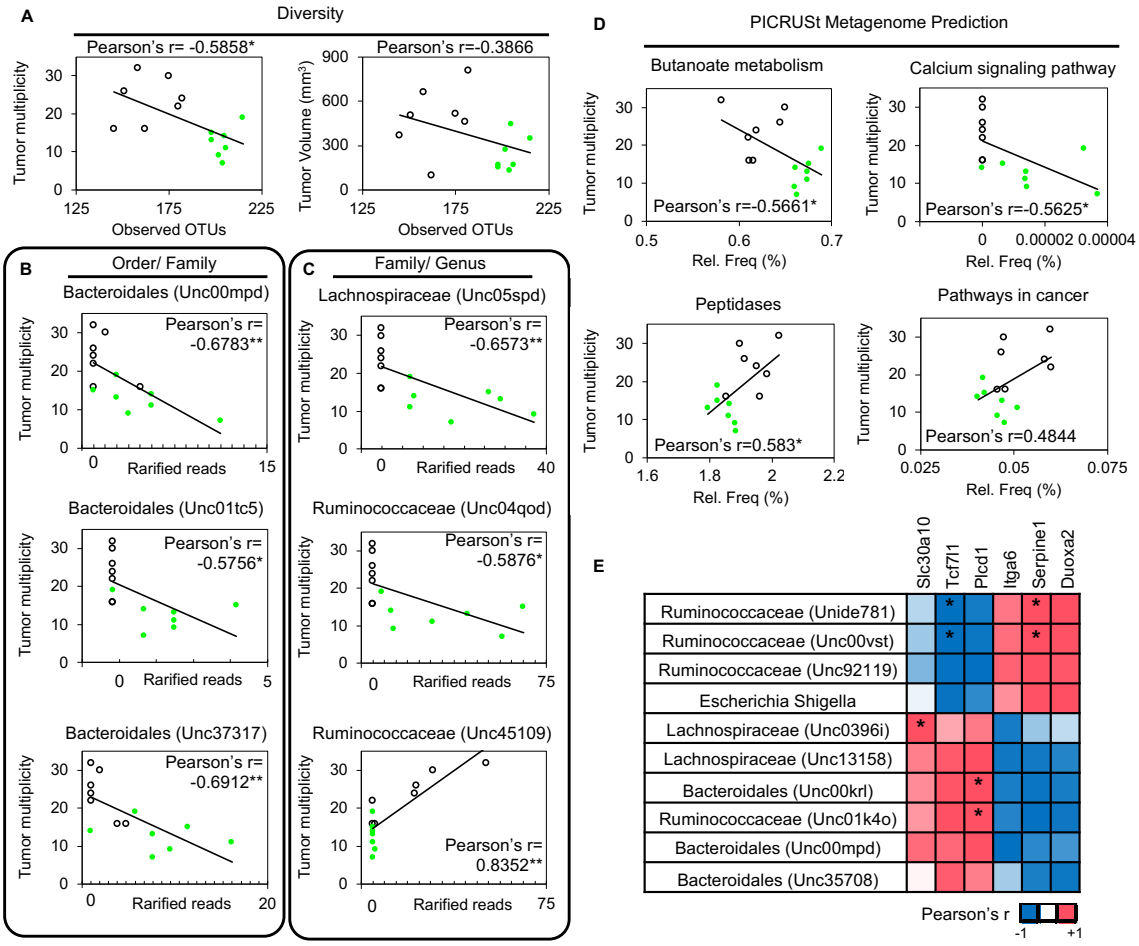


Figure V-6 Integrating gut microbiome, gene expression, and tumor outcomes in response to SPI intervention.

A, Associations between α-diversity of gut microbiome and host disease outcomes for tumor multiplicity and tumor volume. Associations also were probed between tumor multiplicity and the relative abundance of bacterial OTUs at the family level (B) and genus level (C). d, Correlations of tumor multiplicity and metagenome prediction. E, Associations between relative abundance of bacterial OTUs and preventive genes identified in Pirc normal-looking colon. Pearson's correlation coefficient was indicated on the right (red, positive correlation, blue, negative correlation; asterisk indicates statistical significance). For Pearson's correlation in A-D, degree of freedom(N-2) =12 and the critical value=0.533 (p=0.05), and 0.661 (p=0.01). For Pearson's correlation in E, degree of freedom(N-2) =1 and the critical value=0.997 (p=0.05).

Discussion

There continues to be much interest in modulating the human microbiome as an approach to prevention and treatment of cancer[504,505]. We report here on the interrelationships between gut microbial composition, host genetics, dietary exposure, and disease pathogenesis in a preclinical model of FAP. The *Apc*-mutant genetic background of the Pirc rat predisposed to a lower α-diversity in the cecum microbiome, and this was circumvented by feeding dietary spinach under conditions in which significant suppression of adenomatous polyps occurred in the GI tract. This work extends prior observations on the decreased microbial diversity and loss of Firmicutes, Clostridia, and Lachnospiraceae in the $Apc^{Min/+}$ mouse[503] and in human colorectal cancers[303]. In a short-term clinical study, Firmicutes was decreased by consumption of an animal-based diet, whereas Bacteroidetes was decreased by a plant-based diet[245]. Ruminococcaceae phylotypes were increased by resistant starch in obese men, whereas Lachnospiraceae phylotypes were increased by non-starch polysaccharides[506]. These trends were recapitulated in the current investigation, with decreasing Bacteroidetes and increasing Firmicutes and Lachnospiraceae after intervention with dietary spinach. Spinach consumption also reduced Desulfovibrio sulfur-reducing bacteria, implicating inflammatory and barrier functions[507,508].

In contrast to the marked effect on the gut microbiome, prolonged spinach consumption had a relatively mild effect on host transcriptomics, based on mRNA and miRNA profiling. We prioritized genes associated with adenomatous polyp suppression by spinach, which also were correlated with microbiome abundance. *Serpine1* encodes

plasminogen activator inhibitor type 1 (PAI-1), a protein that plays an essential role in extracellular matrix proteolysis and matrix metalloproteinase activity[509], and PAI-1 elevation was observed in sporadic CRC[510] and FAP tumors[511]. Mucosal gene expression profiling of *SERPINE1*, along with other inflammatory regulators such as *CXCL1*, *STAT3* and *IL*s, were negatively correlated with decreased abundance of Firmicutes and Bacteroidetes subsets in human CRC[512]. *DUOXA2* is a maturation factor for the epithelial antimicrobial dual oxidase *DUOX2,* which is among several NADPH oxidase/dual-oxidase family members dysregulated in Crohn's disease and CRC, acting via NFκB[331,502]. *DUOX2* expression was negatively correlated with Bacteroides, Lachnospiraceae, and Blautia, but positively correlated with Pasteurellaceae, Enterobacteriaceae, and Gammaproteobacteria[331].

We also report here on the negative associations between miRNAs and their mRNA targets in the Pirc genetic model, extending prior work in a carcinogen-induced rat model that also exhibited significant colon tumor suppression by dietary spinach[156]. The latter investigation and the current study both prioritized multiple miRNAs, including miR-215 downregulation in colon tumors. Mir-145 was among the most highly downregulated miRNAs in Pirc colon tumors, consistent with its proposed tumor suppressor role in human CRC. This "suppressor miR" was unexpectedly upregulated in a prior report, possibly related the mechanisms associated with the carcinogen treatment[156]. Previous studies prioritized *SERPINE1* as a target of miR-143/145 in bladder cancer[513], as well as miR-34a in liver cancer regulating *KLF4*[514]. *KLF4* can be regulated by different miRNAs depending on circumstances, including mi-206[435] and miR-34a, and

in the Pirc model the miR-34a/*Klf4* axis was disrupted by spinach treatment. *KLF4* is an important zinc-finger transcription factor involved in cell cycle regulation, somatic cell reprogramming, and tumorigenesis. Reduction of *KLF4* is well documented in rat[157,435] and human colon tumors[515,516], providing a potential mechanistic target for spinach, in the context of CRC prevention or treatment[517]. In a mouse colitis model[225], fecal miRNAs also were implicated in dysbiosis and altered disease outcomes, revealing a complex crosstalk between host cells, intestinal content, and gut microbiota. Collectively, the present investigation and published reports provide support for the interplay of microbiome and host factors in the protection afforded by anticancer agents against CRC. This warrants further investigation using both carcinogen-induced and genetic models, where spinach and other dietary interventions have proven highly effective. Interestingly, polyethylene glycol, an osmotic laxative that reduces GI transit time and increases bowel frequency, reduced colon tumor multiplicity significantly in the Pirc model[518]. Although the mechanism was unclear, colon transit time modulating microbiome and bacterial metabolism[519] could be a possible explanation.

Limitations of the current investigation will be overcome in future studies that advance 16S sequencing towards meta-genomic or meta-transcriptomic analyses, and extend the statistical associations to cause-and-effect determination in the context of host-microbiota-diet interactions. Chen *et al.* suggested that lumen- and mucosa-associated microbiota might play different roles in CRC progression through indirect co-metabolism and direct interaction, respectively[520]. Gut microbes can affect, in a site specific manner, host transcriptome immunity, cell proliferation, and metabolism[521]. To advance the field,

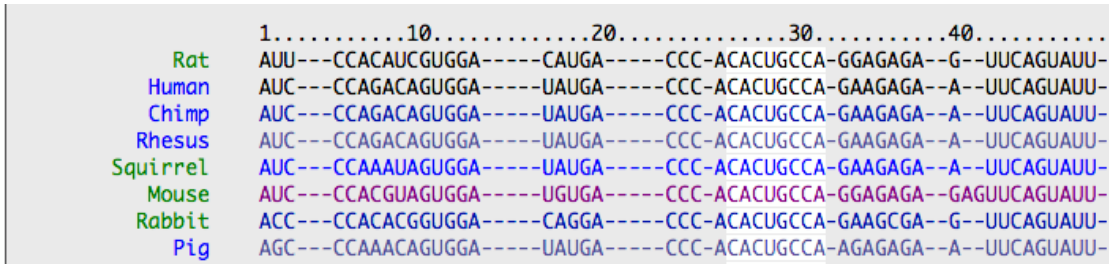
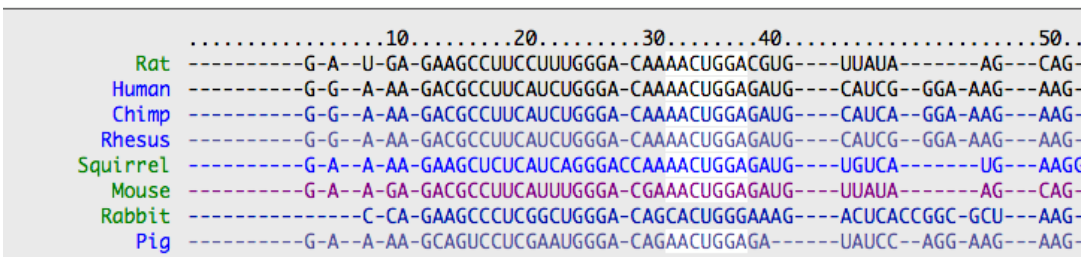
the complex interactions that affect normal homeostasis and pathology in the gut will require integration of multi-omics data, beyond what has been reported here,[522,523] including optimized sample acquisition, data generation and interpretation, and network integration[262,263].

In conclusion, we report here on the efficacy of dietary spinach in a murine model of FAP, linked to marked changes in the gut microbiome, and associated changes in host transcriptomics. Taken together, the data reaffirm the importance of the complex interactions between host genetics, immune response, and microbiota during malignancy. Future studies should seek to enhance the integration of multi-omics datasets, for both natural dietary interventions and standard of care therapeutic agents, with a view to prioritizing potential new targets for prevention and treatment of CRC.

Table V-1 Primers for rat mRNA qPCR.

Rat gene	Forward primer	Reverse Primer
Fhl1	TCTCAAAGACACTCAGGTCCC	CTTATGGGCTTGCGGCATTC
Tpm2	CAGGCTCTCAAGTCGCTGAT	GCACTAGCCAAAGTCTCTTCCA
Hspb8	TGGAACCCTAAGGTCTGGCA	CTCTTCGTGCTTCCCTGACAC
Il1b	CTGTGACTCGTGGGATGATG	GGGATTTTGTCGTTGCTTGT
Serpine1	AGTCTTTCCGACCAAGAGCAG	GTTGGATTGTGCCGAACCAC
Cxcl6	GGTGGAAGTCATAGCTAAGTTGA	GCAAGTGCATTCCGCTTTGT
Myl9	GAGCGCTACGTCCAATGTCT	GGGCCTCGTTCATCATACCC
Myh11	ATCACGGGGGAGCTGGAAAA	AATGAACTTGCCAAAGCGGG
App	CAACCGTGGCATCCTTTTGG	CGTCGACAGGCTCAACTTCA
Myh9	AGGCTCAAGAACCTGATCTGC	TGACTTATGGCCAGGACCTGAA
Bcl3	CCGCAGGGTCATTGATATCCTAA	AAGCTTGGAGCATCCTTCGG
Defa6	GATTTTGAGATAGGAAGGCCAGT	CAGAACGCAGTGGCATATTCTT
Itga6	GGCACTCAGGTTCGAGTGA	GACGGCTGAGTGTGGATCTC
Duoxa2	CACTCAGAGGAACACCAATGC	GGTGGTACAATCCGCAAGGA
Tcf7l1	TCCAGCACACTTGTCCAACA	GGAGAAGTGGTCGTTGCTGTA
Plcd1	CGTGTCCGGATCATCTCTGG	CACCGTGGGTTGAAACCATT
Slc30a10	GTAGCAGGTGATTCCCTGAACA	TGCCAGTTACACGGGTCTTC
Ccl21	CACCCCGGAAGCAATCTCAG	GTCTGTTCAGTCCCCTTGCAG
Il22Ra2	CACTGGAGAAGGAACAAAAAGCC	CTCCCGGTTTCAGACCAGTT
Pecam1	GGTAATAGCCCCGGTGGATG	TTCTTCGTGGAAGGGTCTGC
Rrm2B	CAGCAGAGGAGGTTGACTTGT	CTGGAACCTGCACCTCTTGA
Ptprj	CCAAGCAGGCTCAGGACTAC	AGTTGATGAGCAGGTCGGTG
Klf4	TACCCTCCTTTCCTGCCAGA	TTTGCCACAGCCTGCATAGT
Gapdh	ATGGGAGTTGCTGTTGAAGTC	CCGAGGGCCCACTAAAGG

Table V-2 Prediction of rat miRNA and mRNA pairs.

mRNA	miRNA	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length	P_{CT}
Position 32-38 of Serpine1 3' UTR	rno-miR-145-5p	7mer-m8	-0.27	96	-0.27	4.129	0.59
							
Position 25-32 of KLF4 3' UTR	rno-miR-34a-5p	8mer	-0.48	97	-0.48	2.411	0.33
							

CHAPTER VI

UNTARGETED COLONIC METABOLOME IN APC MUTATED RATS WITH DIETARY SPINACH

Overview

The adenomatous polyposis coli (*APC*) gene is a tumor suppressor gene which is mutated inactively in the majority of colorectal cancer patients. Polyposis in rat colon (Pirc) model is a heterozygous *APC* mutation rodent model, and develops predominantly colon polyps, which are similar to the symptom observed in human familial adenomatous polyposis (FAP). Freeze-dried spinach has demonstrated chemoprevention effects in dietary intervention trials with Pirc and carcinogen-induced colorectal cancer rat models. The aim of the study is to use untargeted metabolomics to characterize metabolic changes of tumorigenesis, as well as impacts from dietary spinach, in the Pirc model. Multivariate analysis indicated metabolic differences between genetic backgrounds, tissue types, and dietary spinach exposures. We found decreased lipids and fatty acids, and increased nucleic acids and vitamins and cofactors in Pirc tumors compared to adjacent normal and wild-type normal colonic tissue. Preliminary mummichog analysis revealed linoleate, arachidonic acid and purine metabolism were most affected metabolic changes in Pirc tumors. Dietary intervention had mild influences on metabolomics. Although the responses to spinach were different in wild-type and Pirc rats, fatty acid related pathways were commonly affected. We sought to explore the metabolome of scraping samples regarding the interests in gut microbiome interaction with host and environment, and the

preliminary data suggestion purine, lipid, and bile acid metabolism were altered. Additionally, purine metabolism was also enriched in tumors growing under spinach intervention indicating a central role in Pirc tumorigenesis. After long-term exposure to dietary spinach, amino acids and vitamins and cofactors were downregulated, and lipids and fatty acids were upregulated. We conclude that lipid and fatty acid and purine metabolism were important in Pirc tumorigenesis, and dietary spinach could reverse the tumor metabolic signatures to achieve chemoprevention of colorectal cancer.

Introduction

Colorectal cancer is one of the most commonly diagnosed cancers with the third leading incidence rate and the forth mortality rate in the world[524]. In 2015, Colorectal cancer accounted for more than 0.14 million new cases and 52,396 deaths in the United States[234]. To reduce the prevalence of colorectal cancer, several screening methods for early diagnosis and treatment options for primary and recurrent diseases have been developed. There are more than 1,200 ongoing clinical trials for therapy and 139 trials for prevention in primary and advanced colorectal cancers across the world (Clinicaltrials.gov, database accessed in September, 2018) indicating the vigorous effort to against colorectal cancer. The causes of colorectal cancer can be categorized by sporadic (>65%), familial (<30%), and hereditary (<5%) factors. Genetic and epigenetic alternations are important features of colorectal carcinogenesis, including chromosomal instability (CIN), microsatellite instability (MSI), CpG island methylation (CIMP) and global DNA hypomethylation. *KRAS*, *BRAF* and *PIK3CA* are common mutated genes

related to oncogenic activation. Tumor suppressors, such as *APC*, *TP53* and *TGFBR2*, are found inactivated mutated in colorectal cancer patients[525]. Adenomatous Polyposis Coli (*APC*) gene is a key tumor suppressor gene, and *APC* mutation is believed to be one of the early events during the multi-step sporadic colorectal carcinogenesis[19]. APC protein is a negative regulator of Wnt signaling by modulating protein interaction, subcellular localization and turnover of β-catenin, and plays important role in cell adhesion and motility. Germline gene predisposition in *APC* causes Familial adenomatous polyposis (FAP) while more than 80% of sporadic colorectal cancer tumors carry somatic APC mutations[526] indicating the essential role of APC in colorectal cancer. Currently, there are two well-established rodent models utilizing mutant *APC* for facilitating colorectal cancer research. They are $Apc^{Min/+}$ (multiple intestinal neoplasia) mouse model and Pirc (polyposis in rat colon) rat model. Pirc model has several advantages over $Apc^{Min/+}$ model to be closer mimicking human conditions. Pirc rats have closer genome structure to human than mouse and spontaneously develop predominant colonic tumor[53,527]. Also, Pirc rats have longer lifespan that longitudinal intervention study can be conducted during different stages of the multi-step colorectal carcinogenesis.

Mass-spectrometry based metabolomics has enabled the small molecular profiling in various specimens to advance research in cancer metabolism and novel biomarker identification[528,529]. The metabolites represent the functional products of gene expression, protein enzymatic processes, environmental exposure and microbial co-metabolism. Thus, Metabolome is a sum of the critical biological processes connecting genome to phenome. It has been proposed that understanding the system levels of metabolites provides

additional insights for mechanistic studies for human diseases[530,531]. A study including almost 3000 participants from German KORA F4 and British TwinsUK cohorts revealed 37 genetic loci associated with serum metabolites using ultrahigh-performance(UHP)-liquid-phase chromatography (LC) and gas-chromatography (GC), coupled with tandem mass spectrometry[532]. Many of the metabolites were found correlated with chronic diseases, novel biological functions, and pharmacotoxicity suggesting new leads for researches. In colorectal cancer patients, metabolomic profiles were found to distinguish normal and cancerous tissue[304] and to predict the prognosis after treatments[533,534]. Short chain fatty acid, amino acids, nucleotide and energy metabolism were implicated in colorectal cancer tumors comparing to adjacent mucosa[306]. MYC-activated global metabolic reprogramming was identified by capillary electrophoresis time-of-flight mass spectrometry (CE-MS) in adenoma stage patients, and knockdown MYC lead to reset of metabolic gene expression related to transporter, purine and pyrimidine metabolisms[325]. In $Apc^{Min/+}$ mouse model, choline phospholipid metabolism was altered in non-tumor gastrointestinal tissues[535], and increasing TCA cycle metabolites was reported in $APC^{min/+}$ polyp tissues[536].

Accordingly to the epidemiology studies, western lifestyle and environmental exposures were significantly associated with colorectal carcinogenesis, changing lifestyle such as diet habits, avoiding known risk factors such as carcinogenic chemicals, and managing precancerous conditions also provide the prevention approaches lowering the risks of getting colorectal cancer[46]. Several metabolomic based study focusing on dietary prevention of colorectal cancer in $Apc^{Min/+}$ model. Nutmeg was reported to have

antimicrobial activity by reducing uremic toxin level and further decreased IL-6 level and normalized the dysregulated lipid metabolism in $Apc^{Min/+}$ colon tumor[537]. In serum samples, American ginseng attenuated the impaired amino acids, carbohydrates and lipid metabolism to extend the lifespan of the $Apc^{Min/+}$ mouse[538]. Freeze-dried blackberries have been demonstrated in a clinical phase 1 pilot trial on demethylation of *SFRP2* and *WIF1* and downregulation of β-catenin and E-cadherin protein[539] that a follow-up study using $Apc^{Min/+}$ mouse identified the prevention associated metabolites in colon, liver and fecal samples were involved in amino acids, glutathione, lipid and nucleotide metabolism[540]. Collectively, lipid metabolism played important role in tumorigenesis in $Apc^{Min/+}$ mouse and could be a potential chemoprevention target.

We have not seen report on metabolomic study using rat model, and our previous studies indicated dietary spinach can reduce 50% tumor outcome in carcinogen-induced[156] and genetic driven colorectal tumorigenesis(CHAPTER V). In this study, we hypothesized that the untargeted analysis of metabolites in Pirc colonic tissues would provide novel metabolic fingerprints reflecting the APC mutation driven colorectal tumorigenesis and prevention effects from dietary spinach which shown drastic modification of the cecum microbiome in rats. Here we applied the UHPLC- Quadrupole-Orbitrap™ Mass Spectrometer to profile the metabolome from rat colonic samples, and reveal the contributions of multiple factors, including genotype, phenotype and diet intervention, in Pirc tumorigenesis and chemoprevention of dietary spinach through fatty acid and purine metabolism.

Materials and methods

Rats and sample collections. In brief, Pirc (F344/NTac-Apcam1137, Taconic Farms, Inc. USA) and wild-type F344 male rats were assigned randomly to basal AIN93 control diet or AIN93 diet containing 10% freeze dried spinach by weight (SPI) for a 30-week colorectal cancer prevention study as previously described (CHAPTER V). The 3-day exposure arms were added to both wild type and Pirc rats to obtain the acute responses in for dynamic metabolomic and transcriptome evaluation. All samples were collected under normoxia during rat necropsy. The samples were flash-frozen and stored at -80°C till sample preparation for further analysis. Experiments were approved by the Institutional Animal Care and Use Committee.

Sample preparation for aqueous phase extraction. The tissue samples, including normal snip (N, mucosa biopsy), mucosa scraping (S), and polyps (T, tumors), were weighed and homogenized in 0.5 ml of cold methanol and 0.2 ml chloroform in pre-cooled garnet bead tubes using Precellys®24 beadbeater. The samples were centrifuged at 3000 rpm for 10 minutes at 4 ° C and 0.7 ml of cold water was added to the supernatant. The aqueous phase was collected by centrifugation at 3000 rpm for 1 minute, and the extraction procedure was repeated again. The pooled aqueous (upper) phase was passed through a sterile nylon cell strainer and. and lyophilized (Labconco). The lyophilized sample was re-constituted in 50ul of methanol/water (1:1, v/v). The samples were stored at −80°C until the analysis on LC-MS/MS.

Untargeted metabolomic data acquisition. The untargeted liquid chromatography high resolution accurate mass spectrometry (LC-HRAM) profiling was conducted on a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific) coupled to a Dionex UltiMate 3000 HPLC system (Thermo Scientific). Synergi Fusion-RP C-18 column was used with a gradient of methanol/acetonitrile solvent system, and the mass scanning was range from 50 to 750. The data was collected in positive mode. MS1 and MS1-dependent MS2 spectra were collected at a *m/z* resolution of 70,000 and 17,500, respectively. Autosampler was maintained at 4 °C. Blank (methanol and water at 1:1 v/v) samples were injected between each sample to prevent any sample carryover. The raw data acquisition was performed by Chromeleon™ (Thermo Scientific).

Metabolomic data analysis. Data was imported to Progenesis QI software (Waters) for alignment, peak picking, compound identification and data processing. There were 17243 features detected, and many were identified using Human Metabolome Database (HMDB) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Raw abundance data were normalized to each sample weights. All features were taking consideration for Partial least squares Discriminant Analysis (PLSDA) analysis. Features were further filtered by 1) at least appeared in 3 groups, 2) at least 3 biological replicates, and 3) significant by ANOVA test resulting in 5946 differential features. The significant features were subjected to clustering and correlation by MetaboAnalyst 4.0[541]. p-values (two-tailed T-test) and t-scores (standardized test statistic) were generated for multiple two-group comparisons among the 15 groups for accessing metabolic networks and prediction of

functional metabolites by mummichog version 2 in R[542]. Mummichog processed multiple two-group comparisons with various significant features list. The feature p-value cutoff and feature numbers were included in the figure legend. Empirical compounds were collected from multiple mummichog predictions to form a consensus predicted compound identification of 883 features. The primary prediction of the 883 compound names were mapped to KEGG Compound collections of Compounds with biological roles, Lipids, and Phytochemical compound with total 5575 small molecules. The three categories were furfure breakdown to eight sub-categories for the compound characterization evaluation. Pathway analyses by mummichog were ranked according to the p-value and using p=0.05 cutoff.

RNA sequencing and bioinformatics. RNA sequencing data was acquired previously with additional normal and tumor samples under 3-day spinach exposure (N=2 for each group). RNA analyses and bioinformatics were previously described (CHAPTER V).

Visualization of metabolites and gene interaction. The metabolite feature information was extracted according to the empirical compound predictions corresponding to the enriched pathway. The gene expression profile was extracted from the previous study in transcriptome (CHAPTER V) according to the KEGG defined genes, associated to the rat metabolic pathway. The metabolites and gene associated with the enriched pathways were displayed in heatmaps. For the assumption that the scraping samples contain a mixture of high content of microbe-specific metabolites, rat-specific metabolites, and co-metabolites,

we used KEGG Search&Color Pathway tool to map the predicted metabolites from scraping samples without defining the species-specific information.

Results

Untargeted metabolomic profiles can distinguish genotypes, tissue types and dietary exposure. The study included wild-type (WT) and polyposis in rat colon (Pirc) rat in three dietary conditions: AIN control diet, AIN control diet containing 10% freeze dried spinach for 3-day intervention (short-term, 3d), and AIN control diet containing 10% freeze dried spinach for 26 weeks (long-term, SPI). The sampling tissue types were normal snip(N), mucosa scraping(S), and polyps(T) (Fig. VI-1A). LC-MS/MS was indicated a robust metabolomic platform for identification and quantification of small molecules[543]. We applied untargeted liquid chromatography-quadrupole-orbitrap mass spectrometry (LC-MS/MS) to determine the metabolites in the total 84 rat colonic samples, and the instrumental detections resulted in over 17,000 metabolite features. Partial least squares discriminant analysis (PLSDA) of WT Normal, WT scraping, Pirc Normal, Pirc scraping, and Pirc Tumor revealed intra-group variations and mild clustering effects of short- and long-term exposure of dietary spinach (Fig. VI-1B). For each diet comparisons, long-term spinach groups were more different from AIN control groups than 3-day spinach groups. For the five groups in control diet, tumors were the most different group clustered away from other four groups. We further calculated the 5946 significant features for overall statistical analyses. The one-way hierarchical clustering revealed distinct small molecular signatures of each group (Fig. VI-1C). The sample-to-sample correlation analysis

indicated tumor profiles were the most different tissue type among the three tissue types from both WT and Pirc rats indicating the distinct metabolic profiles of tumorigenesis (Fig. VI-1D). Interestingly, the Normal snip profiles and scraping profiles from WT and Pirc were more similar in the same tissue types than source genotypes suggesting the tissue collecting methods harvested different metabolites in the gut surface, potentially carrying different ratio of either rat-derived, microbiota-derived, or co-derived metabolites.

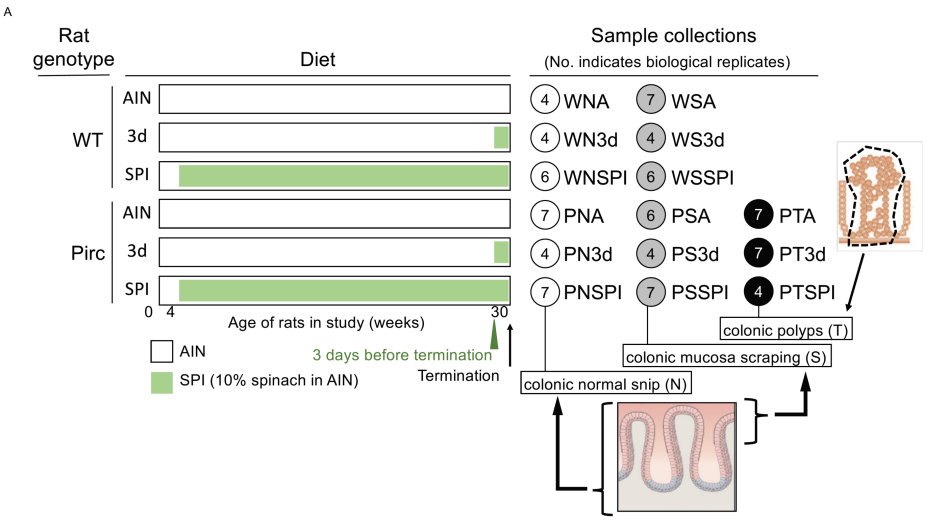


Figure VI-1 Study design and an overview of the untargeted metabolomic analyses of all sample collections.
A, Study design included two rat genotypes: WT and Pirc, three diet conditions: AIN (control diet), 3d (3-day spinach diet), and SPI (26-week spinach diet), and three sample types: colonic normal snip(N), mucosa scraping(S), and polyps(T) leading to total 15 groups. Numbers in the circle indicated the biological replicates of LC-MS/MS data included in the study for each group. B, Partial least squares Discriminant Analysis of 17243 LC-MS/MS features. The comparisons included the diet effects on the five major tissue types and a comparison of the five tissue types on AIN diet. C, Heat map of hierarchically clustered 5946 significant features among the 15 groups of rat samples. Data represented the average of each group and were auto-scaled by features. D, Correlation analyses across all individual samples in five major tissue types: WT Normal, WT Scraping, Pirc Normal, Pirc Scraping, and Pirc Tumor.

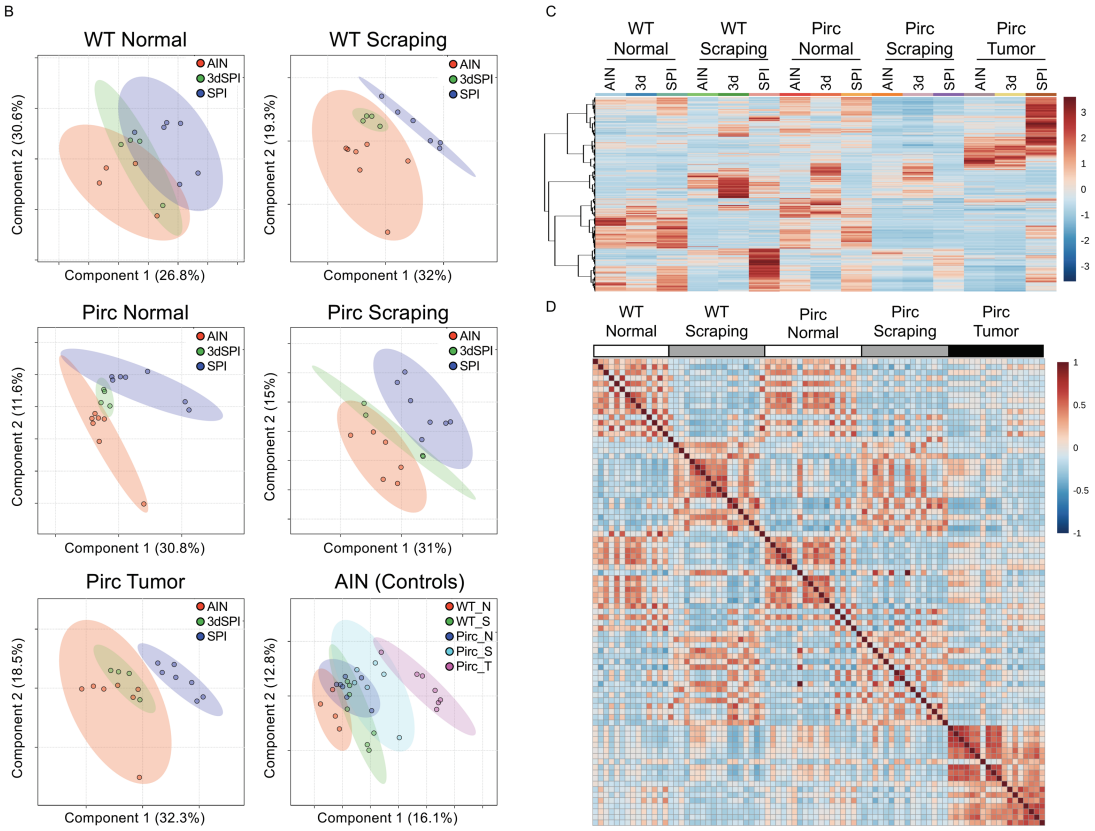


Figure VI-1 Continued

Lipid, nucleic acid and vitamin metabolism were altered during Pirc colon tumorigenesis. To evaluate the metabolic molecules and pathways of the untargeted metabolomics during Pirc tumorigenesis, we primarily used mummichog to reduce the burden of compound identification and confirmation in a high throughput dataset. The consensus empirical compounds were mapped to KEGG Compound database leading to eight categories of small molecule. The categories included lipid and fatty acids, amino acid derivatives, phytochemicals, carbohydrates, nucleic acids, vitamins and cofactors, hormones and steroids, and organic acids. The results showed the average proportion of

each category in WT Normal, Pirc Normal and Pirc Tumor samples, and the lipid and fatty acids, amino acid derivatives, and phytochemicals were the most abundant molecules in the consensus predictions across all samples (Figs, VI-2A-2C). There was no significant difference between Pirc Normal and WT Normal metabolic profiles. However, Lipid and fatty acid level was decreased while nucleic acids and vitamins and cofactors were increased during Pirc tumorigenesis (Fig. VI-2D). Two group analyses by mummichog[542] provided step-wise metabolic changes for *Apc* mutant driven tumorigenesis in Pirc model. Comparing to WT Normal, significant features in Pirc Normal were enriched for pathways involved in fatty acid metabolism (i.e. arachidonic acid, omega-3 fatty acid and Leukotriene metabolism), nucleotide metabolism (i.e. pentose phosphate pathway), and vitamin metabolism (i.e. retinol, biotin, and CoA metabolism) (Fig. VI-2E). Comparing to Pirc Normal, significant features in Pirc Tumors were enriched for pathways related to fatty acid metabolism (i.e. Linoleate, Arachidonic acid, omega-3 metabolism), purine metabolism, carnitine shuttle and glutathione metabolism (Fig. VI-2F). Our findings in lipid metabolism dysregulation in step-wise Pirc tumorigenesis recapitulated the human[306] and APC $^{min/+}$ studies[540]. In order to understand the directional changes of the enriched pathways during Pirc tumorigenesis, we retrieved the metabolites for further investigation. The majority of metabolites associated with linoleate metabolism, arachidonic acid metabolism purine metabolism, and metabolites from eicosapentaenoic acid were decreased in Pirc tumors while metabolites related to omega-3 fatty acid metabolism, beta-alanine metabolism, carnitine shuttle, and glutathione metabolism were increased (Fig. VI-2G). With the previous RNA sequencing data, we also retrieved the genes related to

the enriched pathways in Pirc tumors (Fig. VI-2H). Notably, gene expression profile of the Pirc tumor metabolic pathways was different from WT and Pirc Normal, and the changes seemed to have a balance of the up and down dysregulation, unlike the same directional changes of metabolites of a certain pathway.

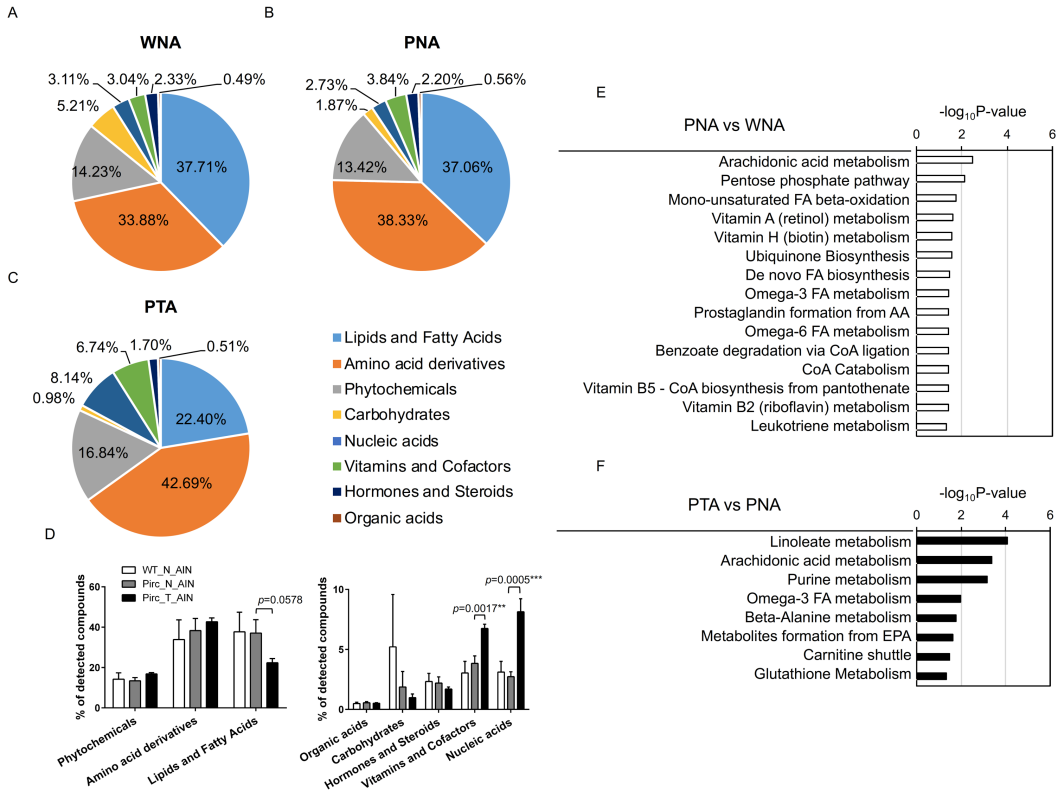


Figure VI-2 Predicted metabolites and pathways altered by Apc mutation in Pirc and tumorigenesis in Pirc colon.

A-C, Features were annotated according to KEGG Compounds and further divided to eight biological compound categories. D, Statistical increases were observed in nucleic acids and vitamins and cofactors in tumor samples. E, Enriched metabolic pathways by significant metabolites features between Pirc and WT normal snip samples (654 features, mummichog feature cutoff at 0.05). F, Enriched metabolic pathways by significant metabolites features between Pirc tumor and adjacent normal snip samples (807 features, mummichog feature cutoff at 0.005). G, Heat map of the associated LC-MS/MS features contributing to the enriched pathways in Figure V-2F. H, Heat map of the gene expression profile according to the genes associated with the enriched pathways in Figure V-2F. Individual column represented individual samples. FA, fatty acid; AA, arachidonic acid; EPA, eicosapentaenoic acid.

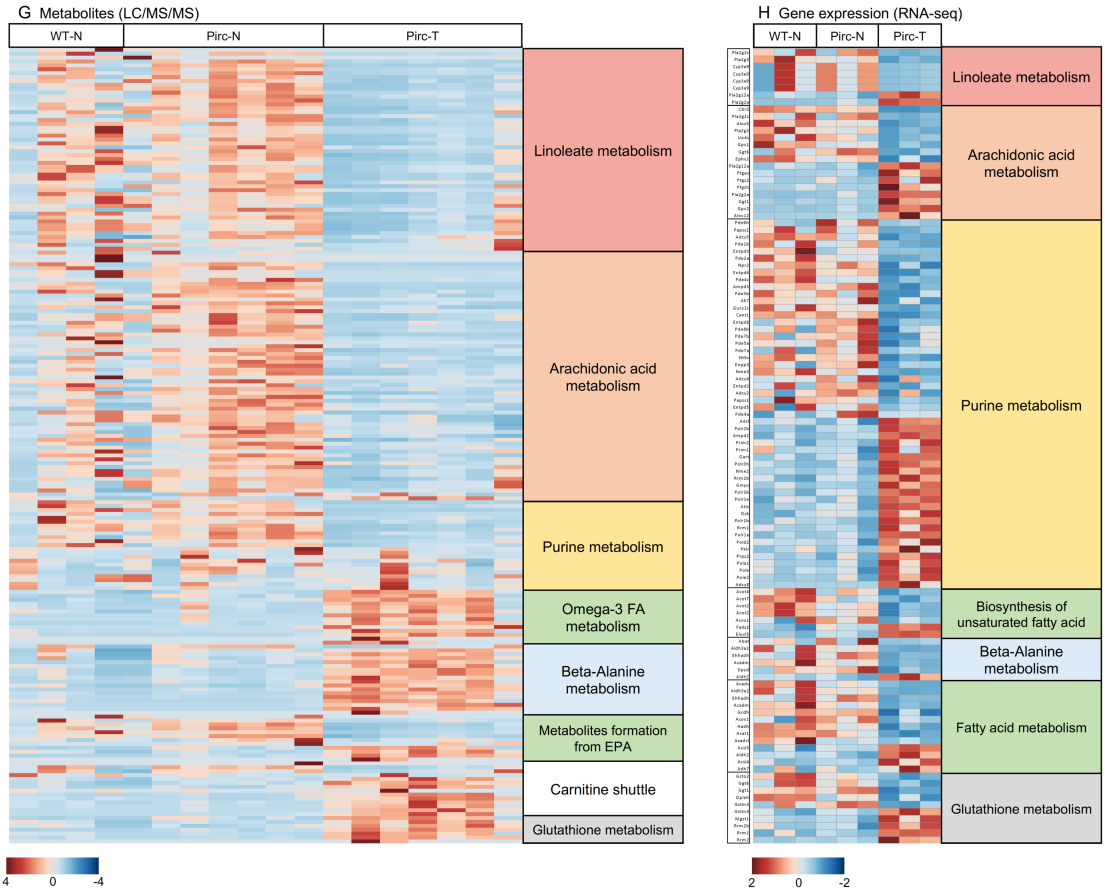


Figure VI-2 Continued.

Dietary spinach exposure altered fatty acid metabolism in Pirc and WT rats. To investigate the dietary spinach effects on Pirc and WT rats, we compared the short- and long-term spinach effects on WT (Fig. VI-3A-C) and Pirc (Fig. VI-3D-E) respectively. Short-term spinach had a greater influence than long-term exposure on metabolomic changes in both Pirc and WT rats (Fig. VI-3A and 3D). However, more detected metabolites were downregulated in WT rats while more detected metabolites were

upregulated in Pirc rats. In the categorical view, there was no statistical differences in both Pirc and WT Normal samples when subjected to dietary spinach exposure (Fig. VI-3B and 3E). Only hormone and steroids and nucleic acids had similar decreasing trends in Pirc and WT rats by dietary spinach. Other metabolites were affected differently in Pirc and WT rats. In WT rat colon, mummichog pathway enrichments indicated that omega-3 fatty acids were the most changed metabolic pathway by short-term spinach along with other alanine and vitamin related metabolisms, and linoleate metabolism was the most altered pathway by long term spinach (Fig. VI-3C). For Pirc, linoleate metabolism was more affected by short-term spinach, and C21-steroid biosynthesis and metabolism was the most altered pathway by long term spinach along with other fatty acids related metabolisms (Fig. VI-3F). Taken together, lipid and fatty acid metabolism seemed critically involving in the metabolism modulation by dietary spinach.

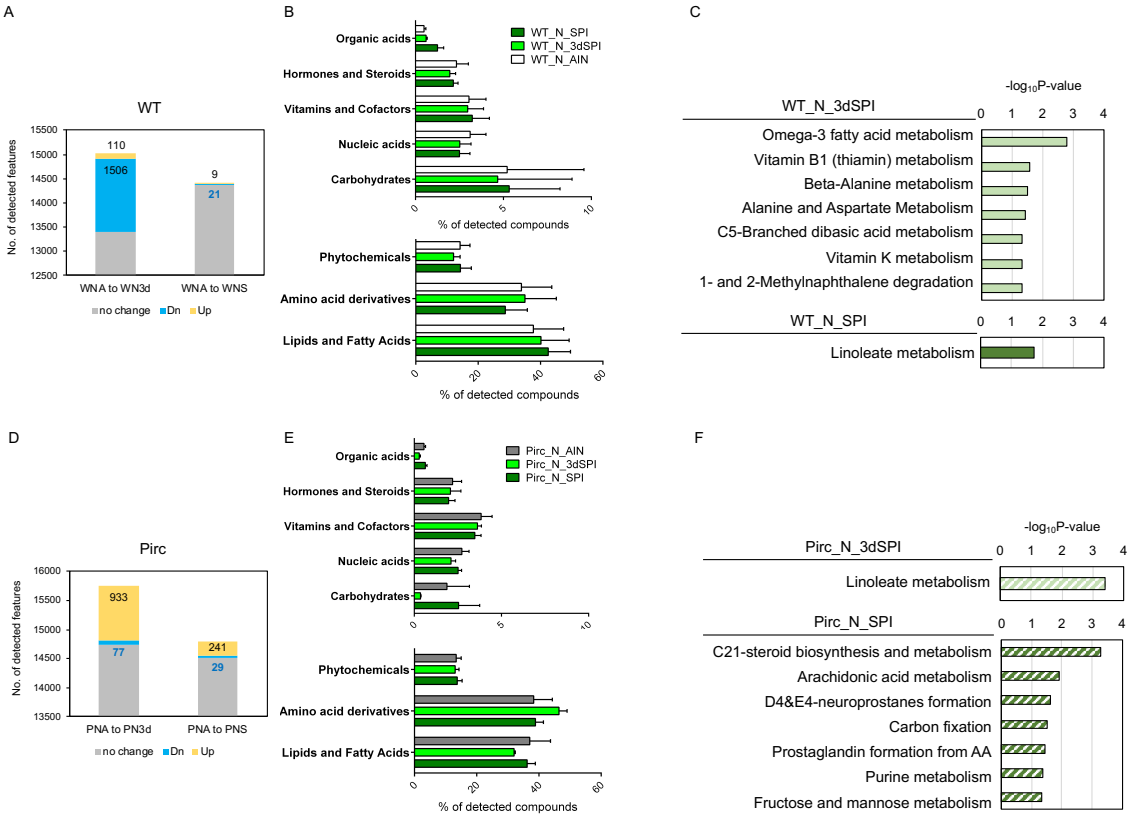


Figure VI-3 Spinach effects on WT and Pirc normal snip tissues.

A, The significant ($p<0.05$) up and downregulated features of all detected features in short- and long-term dietary spinach exposure. B, Features in WT normal snip tissues were categorized according to the eight biological compound categories. C, Enriched metabolic pathways by significant metabolites features between short- or long-term dietary spinach exposure compared to AIN control diet. (3dSPI: 461 features, mummichog feature cutoff at 0.01; SPI: 1036 features, mummichog feature cutoff at 0.2) D, The significant ($p<0.05$) up and downregulated features of all detected features in short- and long-term dietary spinach exposure. E, Features in Pirc normal snip tissues were categorized according to the eight biological compound categories. F, Enriched metabolic pathways by significant metabolites features between short- or long-term dietary spinach exposure compared to AIN control diet. (3dSPI: 1010 features, mummichog feature cutoff at 0.05; SPI: 894 features, mummichog feature cutoff at 0.1) AA, arachidonic acid.

Pirc and WT Scraping samples exhibit distinct metabolic changes by dietary spinach. Brown et al. observed low conserved metabolites overlapped between colonic mucosa and stool samples indicating microbial or other mechanism contributing to the complexity of colorectal cancer biology[306]. It also has been shown the host-microbiota interactions are niche specific[339,521]. Therefore, we included mucosa scraping samples in the study. In the Scraping samples of Pirc and WT rats, dietary spinach affected the metabolites very differently and few metabolites were changed by dietary intervention in the current LC-MS/MS analysis. Among the detected features, 669 features were upregulated and 56 were downregulated in WT Scraping samples under short-term spinach, and 818 features were upregulated and 63 were downregulated in the long-term spinach exposure. However, very few were affected in Pirc Scraping samples. 8 upregulation and 22 downregulation in short-term group while 54 upregulation and 11 upregulation in long-term group were observed. Using KEGG reference metabolic pathway, we were able to reveal the effects of dietary spinach in Pirc and WT Scraping samples. The significant metabolites did not cluster to a specific metabolic pathway but rather affecting several functional biosynthesis and degradation pathways, including bile acid and steroid hormone biosynthesis, phosphatidyl-ethanolamine biosynthesis, sphingosine degradation, lysin and valine biosynthesis, purine degradation, urea cycle and biosynthesis of secondary metabolites (Fig. VI-4).

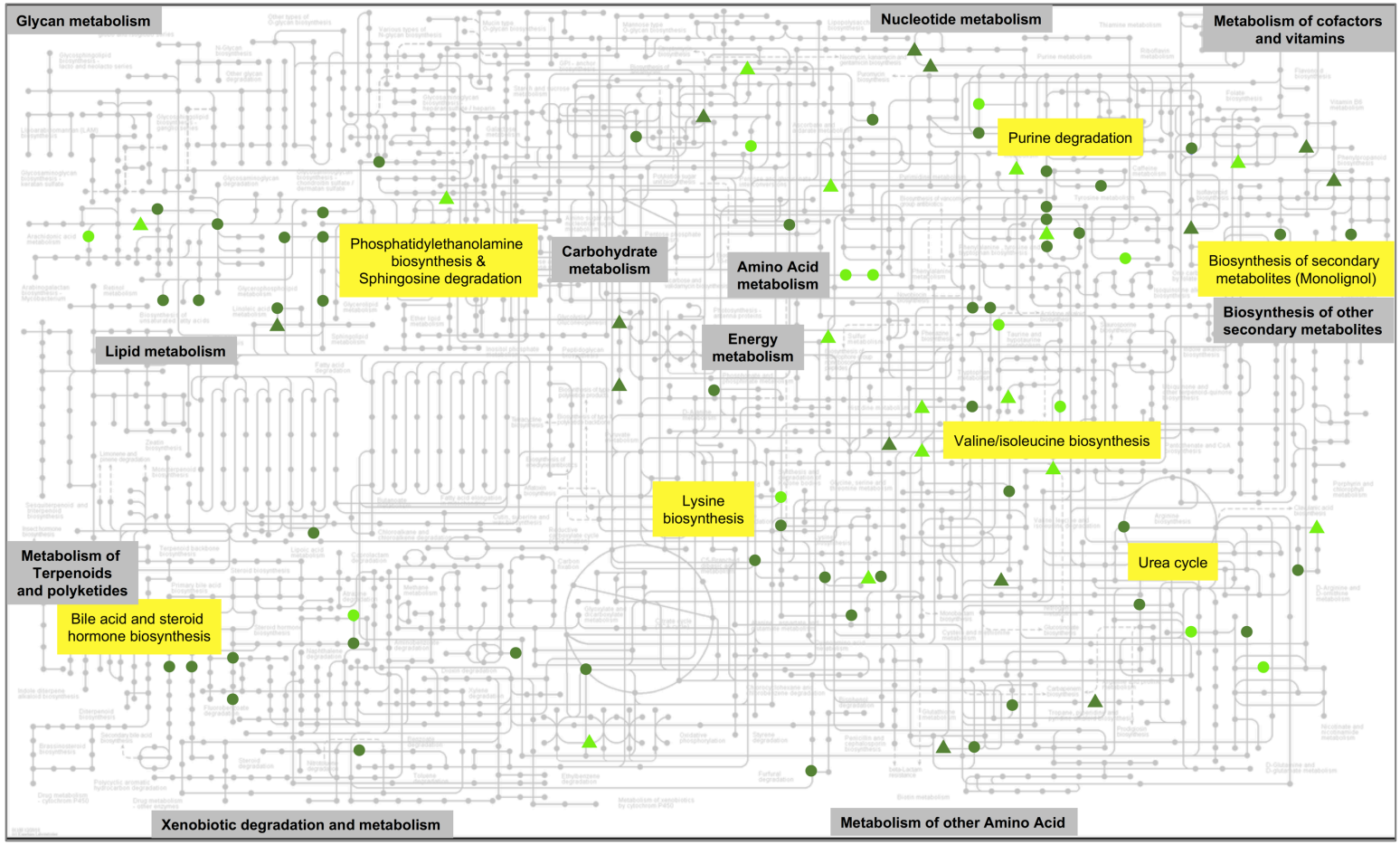


Figure VI-4 Spinach effects on WT and Pirc normal scraping tissues.
Predicted metabolites were mapped to KEGG Reference pathway-metabolic pathway. Light green, metabolites affected by short-term spinach; green, metabolites affected by long-term spinach; circle, WT scraping sample; triangle, Pirc scraping sample. Black boxes indicated the primary category title from KEGG, and the yellow boxes indicated the modules associated with the compounds in the sub-area of the Reference pathway map.

Purine metabolism was essential for tumorigenesis in Pirc model while fatty acid metabolism may be the long-term benefit of dietary spinach. In the previous reports, dietary spinach showed about 50% efficacy to reduce tumor outcomes in the male Pirc model (CHAPTER V). Here we thought to assess the effects of spinach on tumor tissue and to explore the metabolic signature of the tumors resisted to dietary spinach. Comparing to the adjacent normal snip tissue, the tumors growing under long-term spinach exposure exhibited enrichments of four metabolic pathways (Fig. VI-5A). Among the four, purine metabolism was the one persisted altered in tumors both in AIN control (Fig. VI-2F) and SPI diet. Other lipid regulated pathways, such as linoleate, arachidonic acid, and omega-3 fatty acid metabolisms, were not enriched in the tumors under dietary spinach intervention. From a tumor to tumor comparison, nucleic acids were downregulated by short-term spinach exposure while lipid and fatty acids were increased, amino acid derivatives and vitamins and cofactors were decreased by long-term spinach exposure (Fig. VI-5B). In the mummichog pathway analysis, steroid hormones, fatty acid metabolites, and TCA cycle were most affected by short-term spinach exposure in Pirc tumor tissues (Fig. VI-5C) while long-term spinach affected omega-3 fatty acids and butanoate metabolisms (Fig VI-5D). After retrieved the individual metabolites, steroid hormone related metabolites were decreased in Pirc tumor by short-term spinach, and omega-3 fatty acids and butanoate related metabolites were increased by long-term spinach (Figs. VI-5E, light green and green colored, respectively). When looking into the tumor metabolites under spinach diet, pyrimidine and purine metabolites were further increased (Fig. VI-5E, yellow colored) suggesting a tumor signature. There were no

obvious trends on gene expression associated with the spinach effects (Fig. VI-5F, light green and green colored), but there were differences in the tumor-normal comparisons (Fig. VI-5F, yellow) suggesting the potential resistant tumor signature, especially in purine metabolism.

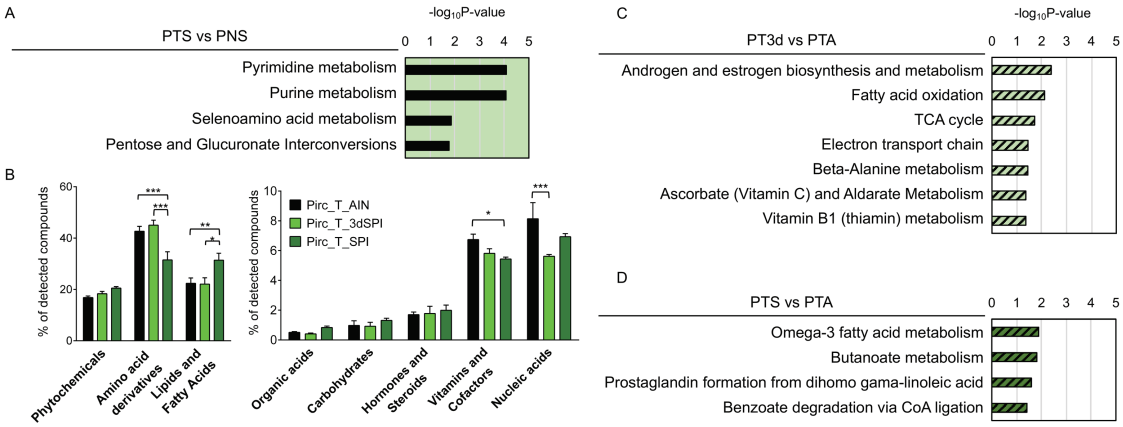


Figure VI-5 Spinach effects on Pirc tumor tissues.
A, Enriched metabolic pathways by significant metabolites features between Pirc tumor and adjacent normal snip samples under long term spinach exposure. (318 features, mummichog feature cutoff at 0.001) B, Features in Pirc tumor tissues were categorized according to the eight biological compound categories. C, Enriched metabolic pathways by significant metabolites features between short-term dietary spinach exposure compared to AIN control diet. (3dSPI: 363 features, mummichog feature cutoff at 0.05) D, Enriched metabolic pathways by significant metabolites features between long-term dietary spinach exposure compared to AIN control diet. (SPI: 822 features, mummichog feature cutoff at 0.001) E, Heat map of the associated LC-MS/MS features contributing to the enriched pathways in Figs. V-5A, 5C and 5D. Individual column represented individual samples. F, Heat map of the gene expression profile according to the genes associated with the enriched pathways in Figs. 5 V-A, 5C and 5D. D-E, light green: short-term spinach effect on tumors, green: long-term spinach effect on tumors, yellow: tumor effect to adjacent normal snip under long term spinach.

Figure VI-5 Continued

Discussion

Colorectal carcinogenesis is a multi-step process and usually takes years to develop without any symptoms. Therefore, early diagnosis of molecular events and effective intervention could provide significant benefits to reduce the burden of tumor initiation and formation. According to Satoch *et al.*, significant metabolic changes occur at the adenoma stage, and not necessarily directly associated with genomic mutation status[325]. Glycolysis, gluconeogenesis, pentose phosphate pathway (PPP), tricarboxylic

acid (TCA) cycle, and fatty acid metabolism were significantly altered in colorectal clinical samples. Similarly, Tian *et al*. identified significantly increased glycolysis and decreased lipids in colorectal cancer (CRC) samples across stages I to IV, and most differences were found between low-grade tumors and adjacent normal tissues[544]. Dysregulation of glycolysis implicated the Warburg effect, via a high rate of glycolysis followed by lactate accumulation, and enhanced cell division with biomass incorporation[545,546]. The PPP interconnects with glycolysis, fatty acid metabolism, and mitochondrial oxidative phosphorylation, and was indicated as a potential target for cancer therapy[547,548]. Lipid and fatty acid metabolism contributes to energy, signaling, oxidative stress, and inflammation in cancer biology[549,550]. From the current dissertation project, arachidonic acid metabolism and PPP were the most highly altered metabolic pathways in Pirc normal looking colonic tissues as compared to normal colon from control WT rats, indicating that the mutant Apc status can cause deregulated metabolic phenotypes before morphological changes that result in tumor formation. Extending these observations from normal looking colon to Pirc tumors, several fatty acid pathways were implicated, including linoleate, arachidonic acid, and omega-3 metabolism. Additionally, beta-alanine metabolism was enriched. Alanine is implicated in glycolysis and glutamate oxidation[551], and previous a report in prostate cancer identified significant increases of lactate and alanine in tumor tissue by metabolic imaging[552].

Notably, while spinach reduced polyp burden significantly in the Pirc model, the anticancer outcomes reset (‘normalized’) certain aspects of the tumor metabolomic profiles. While Pirc colon polyps exhibited reduced metabolites related to lipid and fatty

acids, and increased metabolites related to amino acid derivatives, nucleic acids and vitamin and cofactors (Fig. V-2B), dietary spinach reversed these metabolites changes (Fig. V-5B), *i.e.,* lipid and fatty acids were increased and amino acid derivatives, nucleic acids and vitamin and cofactors were decreased by the dietary intervention. Surprisingly, tumor grown under long-term spinach treatment showed significant increases of pyrimidine and purine metabolism and their metabolites (Fig. V-5A. and Fig. V-5F). In the context of tumor grown under the basal control diet (Fig. V-2F), purine metabolism was a central metabolic feature, while pyrimidine metabolism featured as a resistance pathway in Pirc. Purine and pyrimidine metabolism were both implicated in *MYC*-regulated global metabolic reprogramming in human CRC[325] and were reversed by knockdown of *MYC* expression. Purine metabolism also has been implicated in gut inflammation and gut microbiota interactions[553–556]. Among inflammatory bowel diseases, Crohn's and ulcerative colitis had disease-specific purine gene dysregulation, and correlated with inflammation scores in biopsy samples[557]. Colonization of certain microbes, such as *S. cerevisiae*, promoted purine metabolism and inflammation in mice[558].

In Pirc rats given dietary spinach, several metabolic pathways were implicated in colonic mucosa scrapings of normal looking tissues, which likely capture microbiota-derived metabolites in close proximity to the host, as opposed to gut contents. However, without further careful separation of rat cells and mucus layer, conclusions cannot be made regarding microbiome *versus* host aspects of the scraping samples. To dissect the interactions of the host-microbe-environment, various techniques and additional models are required. For example, fluorescent labeling under anaerobic conditions, for bacteria

with metabolic oligosaccharide engineering (MOE) and bio-orthogonal click-chemistry (BCC), proved useful to study bacteria-derived polysaccharides and glycoproteins in bacterial-host communication[559,560]. Indole-3- acetate and tryptamine were identified as key microbiota-dependent gut metabolites, which reduced liver inflammation through macrophage chemotaxis and hepatocyte fatty acid metabolism, by comparative study using germ-free animals[344]. Isotopic glucose labeling also determined the metabolic incorporation of ^{13}C in metabolic products, such as lactate, formate, acetate and butyrate, during human intestinal fermentation modeling *in vitro*[561]. In addition to purine metabolism, glycerophospholipid and sphingosine degradation, bile acid and steroid hormone biogenesis were implicated in Pirc scrapings after dietary spinach intake (Fig. V-5). Both bile acids and cellular lipids, such as glycerophospholipid, steroids, and sphingolipids, play important roles in gut mucosal interactions and cellular membrane function[562]. Intestinal bacteria can transform primary bile acids to unconjugated hydrophobic secondary bile acids, contributing to microbiota maintenance[563]. There are known inter-species differences for human and murine bile acid synthesis and metabolism, and diet-microbiota interactions were reported critical for cross talk with host FXR signaling[564]. The dynamics of secondary bile acids and FXR signaling determined glucose and lipid regulation, and bile acid synthesis, between intestinal and hepatic feedback regulation[565,566]. Phosphatidylethanolamine participates in membrane constitution, membrane function, mitochondrial integrity, and protein folding[567], and the metabolite ethanolamine provided nutrient advantages for pathogenic *Salmonella* growing in inflamed gut[568]. On the other hand, sphingolipid biogenesis can happen in bacteria, virus,

yeast, fungi, and mammalian cells, and dysregulation of sphingolipid homeostasis has been linked to neurodegeneration, cancer, and inflammatory diseases[569]. Upregulation of mammalian sphingosine kinase 1 (*Sphk1*) regulated NF-κB-induced IL-6 and STAT3 signaling in colitis-associated cancer[570], and Sphk1 also was found overexpressed in CRC, contributing to cetuximab resistance[571].

Molecular signatures of CRC in The Cancer Genome Atlas (TCGA) define MSI/CIMP and CIN categories[33]. However, mass-spectrometry-based proteomic characterization identified five subtypes of cohort with various clinical outcomes in TCGA[34]. Importantly, the study also indicated poor prediction from mRNA information to protein abundance. This infers the need to profile different metabolic aspects to better understand complex biological events in CRC development. In 2015, consensus molecular subtypes (CMS) aggregated several transcriptomic analyses into four classifiers, namely, CMS1: MSI-immune; CMS2: Canonical; CMS3: Metabolic; and CMS4: Mesenchymal[37,387]. The Apc-mutant Pirc model is related to CMS2 signature in WNT activation, MYC target activation, and epithelial markers with poor immunogenic property. CMS2 also contributes to 37% of human CRC, which have better prognosis compared to other CMS categories. However, it is unclear how environmental exposures[337,572] and microbiota interactions affect these CRC subgroups[573–575]. Purcell et al. recently examined colon tumor microbiome and CMS categories,[576] noting that the abundance of *Selenomonas sp.*, *Porphyromonas gingivalis*, and other bacteria were correlated with CMS1, CMS2, and CMS3. This provided further support for microbiota-derived metabolites modulating host health and disease[344,577].

The current dissertation project extended previous work by the Dashwood lab which focused on host gene expression changes in FAP patient samples and Pirc colon polyps (ref IJC) by examining microbiome changes in the rat during tumorigenesis and chemoprevention by spinach. Integrating host transcriptome, microbiome and metabolomics may offer novel insights to mechanistic processes. For instance, Gardinassi et al. identified innate cell immunity and platelet activation by integrated metabolomics and transcriptomes in parasite infection[578]. The study also found different blood metabolites, such as tryptophan, kynurenine, and glutamate, and gene dysregulation in Type 1 interferon, neutrophil and coagulation in partial tolerance patients (semi-immune). Another study utilized blood transcriptome, plasma metabolomics, plasma cytokines and cell populations to understand the vaccination response in humans[579]. With detailed clinical measurements, Li *et al.* identified that IFN-γ+ T cells are the most significant predictor of IgG response, and the transcription factor SREBP1 was critical to determine intracellular sterol metabolism and extracellular humoral responses.

The current dissertation project revealed several limitations and challenges in the study design. 1) High intragroup variability in the Pirc model leads to less statistical power among the sample size. Compared to the tumor efficacy outcome study, metabolomics acquires many more variables, with high degree of diversity due to high sensitivity of the instrument detection. To achieve better statistical analysis, larger sample size and additional internal controls, such as isotope labeling or internal control of extraction, may improve 'Big Data' approaches towards precision medicine. 2) Compound identification is still a major bottleneck of metabolomics. The number of samples and replicates

collected from preclinical studies in Pirc, with and without dietary spinach intervention, makes this one of the most detailed and comprehensive studies of its kind. Indeed, untargeted LC-MS/MS metabolomics detected more than 17,000 features in these experiments. However, annotation remains challenging to match the MS/MS spectra to the reference database with high confidence. Moreover, the microbial metabolites and co-metabolites with host metabolism are still incomplete. In the current dissertation project, mummichog was used to primarily analyze the data focusing on rat metabolism, and generated 883 consensus predictions, mapping to KEGG Compound collections. Current work is focused on annotation and identification of putative compounds agreed by two or more databases. Positive mode LC-MS/MS was applied to the aqueous phase of MeOH/chloroform extracts, which could miss negative mode metabolites, or volatile and low polarity metabolites frequently monitored by GC-MS detection, or certain metabolites requiring specific organic extraction (*e.g.,* chlorophyll-derived metabolites[580,581] and polyphenol phase-II metabolites[582])

To our knowledge, this is the first investigation of global, non-targeted metabolomics in the Pirc model. The study identified critical metabolic pathway networks associated with Apc mutation in normal looking rat colon, and further dysregulated during tumorigenesis events. In addition, dietary spinach intervention led to diverse responses in rat colon according to genotype, as well as in colonic surface niches (within mucosal scrapings), highlighting the complexity of host-microbe-environment interactions in the gut. Findings from the study reiterate the variability of individual animals, and the importance of lipid metabolism and nucleic acids metabolism in CRC, as well as dietary

intervention. To confirm the individual metabolites and to understand their interplay with the enzymatic proteins will provide powerful insights for further cancer prevention and treatment through precision medicine approaches. The large datasets generated from this project can be mined in the future for important and valuable new leads – thus, the dissertation project might represent the 'tip of the iceberg' rather than a complete and final treatise of the subject (see next chapter).

CHAPTER VII

SUMMARY

Discussion and future directions

A previous study of carcinogen-induced tumors in the rat and prevention by dietary spinach implicated dysregulation of the Let-7/c-Myc/Lin28 pathway[156]. In the beginning of this dissertation research, I sought to identify component(s) or fraction(s) of spinach extract that had anti-cancer properties in human CRC lines. Chlorophyll dissolved in DMSO, water extracts, and isopropanol extracts of spinach were shown to decrease LIN28A and LIN28B protein levels along with β-catenin level in Caco2 cells, but the let-7 family was not altered significantly after 48 hours of treatment. Dr. William Bisson at Oregon State University performed *in silico* modeling of chlorophylls binding to murine Lin28A Arg50 cold shock domain, which implicated inhibitory effects on pre-let-7 miRNA interaction. Dr. Arthur Laganowsky and Dr. Wen Liu assisted in purifying human and mouse full-length Lin28A and the cold shock domain for further biochemical testing. A crystallography trial in 96x6 conditions produced no crystal formation when combined with specific chlorophyll derivatives. The LIN28A proteins complexed with chlorophyll derivatives in classic spectrophotometry titration experiments, indicating an affinity of 200 μM (equilibrium dissociation constant). Taken together, these experiments did not provide a clear answer to the focused mechanistic question at the outset; that chlorophylls in spinach might interact directly with components of the Lin/Let pathway, upregulating bioactive suppressor miRNAs (let7 members) to suppress tumorigenesis in the rat.

However, these studies were discontinued and omitted from the dissertation for the simple reason that *in silico*, *in vitro*, and cell-based assays were misleading, and did not predict the situation in the Pirc model with respect to the let-7/ Lin28 axis. Unlike the carcinogen-induced model, the Pirc model showed relatively mild downregulation of let-7 family after spinach treatment, and colon polyps had almost undetectable Lin28 proteins, despite the excellent preventive efficacy of spinach in both genetic and carcinogen models.

Subsequent detailed molecular profiling of Pirc the model sought to identify the 'true' targets of spinach for further mechanistic study. Interestingly, Lin28 levels were associated with invasive tumor growth in human CRC, but the authors were unable to detect Lin28 proteins in $Apc^{Min/+}$ and DSS/AOM-induced gastrointestinal tumors[583]. The study indicated independent roles of Wnt pathway and Lin28/let-7 axis in colorectal tumor development. We did not observe β-catenin downregulation in the colon polyps from Pirc rats fed with spinach, despite evidence to the contrary from cell-based assays (data not presented in the dissertation). This suggested the following: 1) spinach solvent extracts functioned differently from the digesta of spinach diet in rodent digestive system, along with the presence of microbiome; 2) rat colonic cells with heterozygous *Apc* mutation behaved differently from immortalized human CRC cell lines growing a in petri dish; 3) re-focus the goals and overall scope of the dissertation to findings *in vivo,* from the Pirc model.

The first molecular candidate considered was Krüppel-like factor 4 (KLF4), one of four pluripotency factors and an important transcription factor regulating cell proliferation and differentiation[584]. *Klf4* was found epigenetically regulated by

hypermethylation[585,586] and miRNA regulation[315,587]. Together with OCT4, and SOX2, KLF4 was identified as a direct target of miR-145 and can determine self-renewal and differentiation in human embryonic stem cells[587]. KLF4 is a downstream mediator of Wnt signaling in human gastrointestinal cancer, and has crosstalk with oncogenic pathways and epithelial to mesenchymal transition[588]. KLF4 was reduced in colorectal cancer patient samples as well as colorectal cell lines, and Zhao et al. found evidence for loss of heterozygosity, hypermethylation of 5'-untranslated region, and point mutations of *Klf4* leading to p21 induction, indicating a tumor suppressive role[515,589,590]. *Klf4* was downregulated in all PhIP-induced tumors in the rat (except skin tumors), and divergent functions of *Klf4* in skin was discussed in Chapter IV. Klf4 also was implicated as a miR-34a target in Pirc colon tumors, and reversed by dietary spinach. However, unambiguous confirmation of miRNA-mRNA pairing would require additional crosslinking and immunoprecipitation (CLIP) methods to verify the targeting events[591]. Although Klf4 has not been widely proposed as a therapeutic target, an inhibitor strategy is in development involving arginine protein methyltransferase-mediated KLF4 protein turnover[592]. Additional work on the response of Klf4 to environmental exposures and epigenetic regulation might corroborate Klf4 as a valuable target for cancer intervention[473].

Dietary spinach is among the most potent and effective anti-cancer regimens tested thus far in the Pirc model, and reaffirmed the potential of dietary intervention for disease prevention[83,593,594]. It was important to persist with experiments that might provide further clues; thus, the dissertation project switched from hypothesis-testing to hypothesis-generating, and to multi-omics approaches. Notably, the majority of changes in

microbiome and metagenome prediction affected by dietary spinach were shared between Pirc and WT rats, implying that diet indeed has profound effects on the microbiome[245,595]. However, there are still differences between the responses in host transcriptome and a subset of microbiome, indicating that host genetic influences on microbiome cannot be ignored. Technically, whether filtering out low abundant OUT read can provide a more critical insights would be worth tested in the current dataset and how different locations (cecum versus colon and content versus mucus) would affect the microbiome profiling can be tested in the future.

Spinach effects were marked and compelling in terms of reversing Pirc microbiome deregulation (dysbiosis) and metabolomics, despite the incomplete nature of the latter. The results raise fundamental questions as to how exactly microbiota and metabolites contribute to spinach cancer prevention. Is one or the other the driving force? What is cause and what is effect? These questions should be addressed in future studies. Recently, fecal material transplantation (FMT) emerged as a strategy for disease management and therapeutics[369,596]. Carefully assessing host genetics, microbiome intervention, and dietary patterns in long-term FMT studies might assure the desirable outcome and improve future intervention designs. To validate whether microbiome is required for prevention by dietary spinach, one can perform experiments in germ-free or gnotobiotic animals to evaluate tumor outcomes. Alternatively, these questions might be answered by FMT from rats consuming dietary spinach to Pirc rats with control diet, or vice versa, via FMT from healthy WT rats to Pirc rats. The results could provide value

insights into the roles of microbiota and metabolites for designed dietary interventions against human CRC.

This thesis project also challenged the definition of "prevention" among the many groups evaluated. Dietary spinach reduced the tumor outcomes ~50% in the Pirc model, but there were still polyps detected in all the rats. Chapter V described two scenarios for prevention by spinach: 1) against polyp formation at the outset, such that normal looking colon in Pirc rats remained 'normal' and with an opposite molecular profile of polyps, and 2) polyps that still formed in the present of spinach treatment, under the strong Wnt pathway driving force, had 'partially-normalized' molecular signatures, leading to delayed onset and/or smaller lesion size. In fact, the *Apc*-mutant Pirc model is a genetically driven counterpart of human FAP, which makes all interventions technically secondary or tertiary prevention [353,597,598], and the latter scenario is closely aligned to cancer 'therapy'. Therefore, ambiguity exists regarding prevention *versus* therapeutic outcomes for spinach.

There are limitations of the carcinogen and genetic rat models used in this dissertation. For the cooked meat heterocyclic amine project, unbiased analysis of tumors from different organs was not undertaken, and it is highly likely that there are tissue specific deregulation of miRNAs, and other epigenetic players. The original PhIP study[156] did not have enough tissue samples for further analysis, so assessing the prevention effects of dietary spinach became more dependent on the Pirc model. In the latter case, sample size was calculated according to published tumor outcomes, but there was high polyp-to-polyp variability of some molecular read-outs, inferring the need for a larger sample size and more detailed measurements to achieve better statistical confidence. Additionally, due

to the tissue requirements of each technology, samples were subdivided into two main groups in order to have matched data for the most sought-after correlated measurements. Thus, one group was taken for RNA and miRNA, and another group for microbiome and metabolomics. Using group average data limited the analysis power and lost the information from individuality. There were relatively fewer features changed by dietary spinach exposure, compared to tumor-normal comparison, in the current LC/MS-MS work using aqueous phase extracts. Additional insights might come from organic phase extracts and more time points, especially for early time points, for example, during weaning, and right when the dietary intervention started, etc.) and sample numbers, to obtain a comprehensive picture of the dynamic changes of metabolites by dietary spinach exposure leading to tumor inhibition.

Sample collection is another aspect that can be improved in the future, and there are several factors worth considering. 1) Tumor *heterogeneity* plays an important role in most human malignancies, and immune cells or stem cells are minor but important sub-populations[599–601]. Transcriptome analysis after cell sorting or utilizing single cell sequencing technology[602,603] can provide population structure and cell-specific regulation information. 2) There is a proposed driver-passenger model of microbial agents in human CRC[520,575,604,605], and the *region of colon* also contributes to noticeable differences in microbiome profiles[606,607]. 3) Gut environment is an ecological complex system with host cells, microbiota, and digestive content or stool. *Separating host cells* from extracellular content (surface mucus layer or lumen content/stool) may be helpful to concentrate on either host-cellular metabolism or host-microbe co-metabolism in extracellular contents.

Metabolomic imaging is another powerful technology to detect biodistribution and metabolism in tissue sections[608–610]. In this dissertation work, there were nine groups for transcriptomics, four groups for cecum microbiomics, and 15 groups for metabolomics in the Pirc model, which lead to a high complexity of analysis and interpretation. Acquiring new insights from additional data mining is recommended, for follow-up hypothesis driven studies with targeted groups *in vivo*.

Moreover, since there are emerging interests in monitoring miRNAs for environmental exposure and disease[78,81,142,611], detection methods might involve cell-free circulating miRNAs or cell-based (PBMC or platelet) miRNA in blood samples of PhIP and Pirc models[214,216,612,613]. A longitudinal study on miRNA changes by environmental exposure or disease progression could provide valuable insights for disease management. There also is the concept of detecting circulating cancer cells by digital PCR[614–617], but it may not be suitable due to the fact that both rat models mimic early CRC stages before invasion and metastasis. However, if additional rodent models of metastasis are tested with spinach intervention, circulating cancer cells might be of interest[618].

Among the various omic data acquired, untargeted metabolomics is the most difficult to analyze and extract information. This is mainly because identification and validation of compounds is a critical bottleneck[291,543], despite the fact that LC/MS-MS is a robust approach. Unlike sequencing technology and sequence information, metabolomics has higher resolution/dynamics to detect subtle differences, but lacks a well-established and finite 'reference map'[619,620]. This thesis project focused, to the maximum extent possible, on identifying high confidence compounds supported by

statistical analysis and associated enrichment metabolic pathway evaluations. For instance, a recurring theme was the identification of linolenic acid, stearidonic acid, leukotriene A4, and glutamate in fatty acid related pathways. This suggests the need to pursue lipid metabolism and purine metabolism, and to spike in standards to provide sufficient evidence for final identification confirmation. On the other hand, hit validation also can be performed on nuclear magnetic resonance spectroscopy (NMR)[621,622] as a gold standard. With data in the form of 'features', 'putative identification' and 'confirmed identification', the study can be reported according to the Metabolomics Standards Initiative[619,623], and provide leads for future mechanistic studies in cell-based assays of spinach inhibition.

Despite the limitations of the preclinical models, and insufficient matched molecular profiling across all biological individuals, there are opportunities to integrate multivariate data to provide insights from multi-omic measurements. Bioinformatics is in demand for Big Data analysis, and there are several potential methods utilizing dimension reduction techniques and correlation structure to integrate multi-omic data[624,625]. Buescher and Driggers emphasized the importance of statistical behavior, relationship between layers, and timing differences of the layers[313] in integration of omics, or in most cases, "co-analysis", and suggested the condition-dependent unbiased co-clustering on the same timescale to offer valuable insight from the literature. Beyond correlation analysis, more sophisticated methods exists, including identifying the scaffold, decomposing the scaffold, cellular modeling, and trans-omics[312,314]. However, the current data include omics from different entities, namely host, microbiota and host-microbiota co-derived

information, so unbiased co-clustering with reduction techniques may be a practical method to explore for further understand the overall measurements.

Summary and conclusion

The dissertation project focused on utilizing various molecular profiling techniques to understand fundamental aspects cancer etiology in preclinical models of colorectal carcinogenesis. Both genetic and carcinogen-induced rat models have pros and cons, and intrinsic value for the kind of work pursued here. The carcinogen-induced model reflects aspects of environmental exposure, which results in multiple genetic mutations and epigenetic dysregulation, capable of driving carcinogenesis in a 'moderate' fashion, akin to sporadic cancer in humans. The genetic model exemplifies the single penetrant gene-driver concept, via inactivation of murine *Apc,* leading to extensive polyposis in the intestine and colon, mimicking human FAP.

An initial focus was on whether miRNA profiles of colon tumors, the primary target organ of foodborne carcinogen PhIP, had predictive value for other PhIP-induced tumors (small intestine, skin, liver), and to validate miRNA and mRNA targets implicated in a subset of human cancers (Chapter IV). Downregulation of the let-7 family and upregulation of their targets, *c-myc* and *Hmga2*, were consistent in rat tumors and human primary tumors. A five-miRNA 'PhIP signature' was identified in the carcinogen-induced rat tumorigenesis model, and revealed the signature correlated to poor prognosis of a subgroup human cancer patients. The 'PhIP signature' also was linked to downregulation of the pluripotency gene *Klf4*, and pathway predictions associated with MAPK, endocytosis, colorectal cancer, apoptosis, and Wnt pathway, implicating a role for

epigenetic regulation. This extends the traditional 'genotoxin' perspective of PhIP and other environmental heterocyclic amines.

In order to broaden this work towards genetic predisposition syndromes, the Pirc model was chosen to recapitulate aspects of human FAP, including a colon tumor burden and lifespan suitable for intervention and disease monitoring. Sequencing technologies and mass spectrometry facilitated comprehensive analysis of biological datasets, and enabled extensive molecular profiling in Pirc tumors and normal tissues, including RNA, miRNA, metabolomics, and microbiomics (Chapters V-VI).

The dissertation project prioritized inflammatory pathways, lipid metabolism, and purine metabolism in Pirc colon polyps, which recapitulated human CRC features. Microbiomics in Pirc rats indicated a reduced relative abundance of *Firmicutes* and increased *Bacteroidetes,* without a significant change in α-diversity, and small RNA sequencing indicated significant downregulation of suppressor miRNAs. Dietary spinach started after weaning, followed by 6.5-month intervention, lead to ~50% reduction of tumor outcome in Pirc rats. Spinach is among the most effective prevention strategies tested thus far in the Pirc model, and given that the Apc-mutant background is considered a 'strong' oncogenic driver, has potentially important clinical implications for FAP patients. It is unlikely that a person with a family history of FAP will put off standard-of-care, indefinitely, but consumption of dietary spinach might help to 'buy time' before total colectomy and anticancer drug treatments.

Transcriptomics suggested immune related pathways, extracellular matrix remodeling, transcription factors, and miRNA targeting events may be affected by spinach

consumption. Spinach was shown to reshape the metabolome against changes in the Pirc microbiome changes, and the metabolome also conferred reversing effects in lipid and steroid metabolism. This investigation provided the fundamental molecular profiling of the Pirc model for further intervention design, and revealed potential targets of genes, metabolic pathways, and microbiota for further validation.

This thesis project identified step-wise, temporal molecular signatures associated with colorectal carcinogenesis, and provided support for PhIP and Pirc preclinical models, that were complementary, which each offering useful insights. The complex interactions within host biological system and the interactions between host genetics, microbiota and environmental exposure require further integration to clarify the multi-dimensional networks, leading to discovery of novel strategies for CRC prevention, management and treatment.

REFERENCES

1. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66,** 683–691 (2017).

2. Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J. & Brown, M. L. Projections of the cost of cancer care in the United States: 2010-2020. *J. Natl. Cancer Inst.* **103,** 117–128 (2011).

3. NIH Categorical Spending -NIH Research Portfolio Online Reporting Tools (RePORT). at <https://report.nih.gov/categorical_spending.aspx#tab1>

4. Welch, H. G. & Robertson, D. J. Colorectal Cancer on the Decline--Why Screening Can't Explain It All. *N. Engl. J. Med.* **374,** 1605–1607 (2016).

5. Cassidy, S. & Syed, B. A. Colorectal cancer drugs market. *Nat. Rev. Drug Discov.* **16,** 525–526 (2017).

6. Haggar, F. A. & Boushey, R. P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin. Colon Rectal Surg.* **22,** 191–197 (2009).

7. American Cancer Society. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019 . *American Cancer Society, Atlanta* (2017).

8. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA Cancer J Clin* **67,** 7–30 (2017).

9. Lynch, H. T. *et al.* Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* **76,** 1–18 (2009).

10. Galiatsatos, P. & Foulkes, W. D. Familial adenomatous polyposis. *Am. J. Gastroenterol.* **101,** 385–398 (2006).

11. Powell, S. M. *et al.* Molecular diagnosis of familial adenomatous polyposis. *N. Engl. J. Med.* **329,** 1982–1987 (1993).

12. Huxley, R. R. *et al.* The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int. J. Cancer* **125,** 171–180 (2009).

13. Anand, P. *et al.* Cancer is a preventable disease that requires major lifestyle changes. *Pharm. Res.* **25,** 2097–2116 (2008).

14. The American Cancer Society medical and editorial content team. Colorectal

Cancer. at <https://www.cancer.org/cancer/colon-rectal-cancer.html>

15. American Joint Committee on Cancer. *Colon and Rectum Cancer Staging*. (American Cancer Society, 2010). at <https://cancerstaging.org/references-tools/quickreferences/documents/colonmedium.pdf>

16. Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138,** 2044–2058 (2010).

17. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6,** 479–507 (2011).

18. Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N. Engl. J. Med.* **361,** 2449–2460 (2009).

19. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767 (1990).

20. Okugawa, Y., Grady, W. M. & Goel, A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology* **149,** 1204–1225.e12 (2015).

21. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301,** 89–92 (1983).

22. Coppedè, F. Epigenetic biomarkers of colorectal cancer: Focus on DNA methylation. *Cancer Lett.* **342,** 238–247 (2014).

23. Uribe-Lewis, S. *et al.* 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome Biol.* **16,** 69 (2015).

24. Strubberg, A. M. & Madison, B. B. MicroRNAs in the etiology of colorectal cancer: pathways and clinical implications. *Dis. Model. Mech.* **10,** 197–214 (2017).

25. Yang, Y., Junjie, P., Sanjun, C. & Ma, Y. Long non-coding RNAs in Colorectal Cancer: Progression and Future Directions. *J. Cancer* **8,** 3212–3225 (2017).

26. Xiang, J.-F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24,** 513–531 (2014).

27. Kogo, R. *et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* **71,** 6320–6326 (2011).

28. Nguyen, H. T. & Duong, H.-Q. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy. *Oncol. Lett.* **16,** 9–18 (2018).

29. Jass, J. R. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* **50,** 113–130 (2007).

30. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314,** 268–274 (2006).

31. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).

32. Bass, A. J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nat. Genet.* **43,** 964–968 (2011).

33. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).

34. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513,** 382–387 (2014).

35. Guo, Y. *et al.* How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim. Biophys. Sin. (Shanghai)* **40,** 426–436 (2008).

36. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).

37. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21,** 1350–1356 (2015).

38. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47,** 320–329 (2015).

39. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3,** 330–338 (2003).

40. Raymond, E., Faivre, S., Woynarowski, J. M. & Chaney, S. G. Oxaliplatin: mechanism of action and antineoplastic activity. *Semin. Oncol.* **25,** 4–12 (1998).

41. Xu, Y. & Villalona-Calero, M. A. Irinotecan: mechanisms of tumor resistance and novel strategies for modulating its activity. *Ann. Oncol.* **13,** 1841–1851 (2002).

42. Clarke, J. M., Hurwitz, H. I. & Rangwala, F. Understanding the mechanisms of action of antiangiogenic agents in metastatic colorectal cancer: a clinician's perspective. *Cancer Treat. Rev.* **40,** 1065–1072 (2014).

43. Ivashko, I. N. & Kolesar, J. M. Pembrolizumab and nivolumab: PD-1 inhibitors for advanced melanoma. *Am. J. Health Syst. Pharm.* **73,** 193–201 (2016).

44. Marginean, E. C. & Melosky, B. Is There a Role for Programmed Death Ligand-1 Testing and Immunotherapy in Colorectal Cancer With Microsatellite Instability? Part II-The Challenge of Programmed Death Ligand-1 Testing and Its Role in Microsatellite Instability-High Colorectal Cancer. *Arch. Pathol. Lab. Med.* **142,** 26–34 (2018).

45. Marginean, E. C. & Melosky, B. Is There a Role for Programmed Death Ligand-1 Testing and Immunotherapy in Colorectal Cancer With Microsatellite Instability? Part I-Colorectal Cancer: Microsatellite Instability, Testing, and Clinical Implications. *Arch. Pathol. Lab. Med.* **142,** 17–25 (2018).

46. Chan, A. T. & Giovannucci, E. L. Primary prevention of colorectal cancer. *Gastroenterology* **138,** 2029–2043.e10 (2010).

47. Gravitz, L. Prevention: Tending the gut. *Nature* **521,** S6-8 (2015).

48. Zeineldin, M. & Neufeld, K. L. Understanding phenotypic variation in rodent models with germline Apc mutations. *Cancer Res.* **73,** 2389–2399 (2013).

49. Rosenberg, D. W., Giardina, C. & Tanaka, T. Mouse models for the study of colon carcinogenesis. *Carcinogenesis* **30,** 183–196 (2009).

50. Kiesler, P., Fuss, I. J. & Strober, W. Experimental models of inflammatory bowel diseases. *Cell. Mol. Gastroenterol. Hepatol.* **1,** 154–170 (2015).

51. Tanaka, T. *et al.* Dextran sodium sulfate strongly promotes colorectal carcinogenesis in Apc(Min/+) mice: inflammatory stimuli by dextran sodium sulfate results in development of multiple colonic neoplasms. *Int. J. Cancer* **118,** 25–34 (2006).

52. Tanaka, T. *et al.* A novel inflammation-related mouse colon carcinogenesis model induced by azoxymethane and dextran sodium sulfate. *Cancer Sci.* **94,** 965–973 (2003).

53. Amos-Landgraf, J. M. *et al.* A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. *Proc. Natl. Acad. Sci. USA* **104,** 4036–4041 (2007).

54. Ertem, F. *et al.* Development of a murine colonoscopic polypectomy model (with videos). *Gastrointest. Endosc.* **83,** 1272–1276 (2016).

55. Ertem, F. U. *et al.* Oncogenic targets Mmp7, S100a9, Nppb and Aldh1a3 from transcriptome profiling of FAP and Pirc adenomas are downregulated in response to tumor suppression by Clotam. *Int. J. Cancer* **140,** 460–468 (2017).

56. Kwong, L. N. & Dove, W. F. APC and its modifiers in colon cancer. *Adv. Exp. Med. Biol.* **656,** 85–106 (2009).

57. Briggs, D. Environmental pollution and the global burden of disease. *Br Med Bull* **68,** 1–24 (2003).

58. Diamanti-Kandarakis, E. *et al.* Endocrine-disrupting chemicals: an Endocrine Society scientific statement. *Endocr. Rev.* **30,** 293–342 (2009).

59. Hodjat, M. *et al.* Environmental toxicants, incidence of degenerative diseases, and therapies from the epigenetic point of view. *Arch. Toxicol.* **91,** 2577–2597 (2017).

60. Belpomme, D. *et al.* The multitude and diversity of environmental carcinogens. *Environ. Res.* **105,** 414–429 (2007).

61. Surh, Y.-J. Cancer chemoprevention with dietary phytochemicals. *Nat. Rev. Cancer* **3,** 768–780 (2003).

62. Su, Z.-Y. *et al.* A perspective on dietary phytochemicals and cancer chemoprevention: oxidative stress, nrf2, and epigenomics. *Top Curr Chem* **329,** 133–162 (2013).

63. Cragg, G. M., Grothaus, P. G. & Newman, D. J. Impact of natural products on developing new anti-cancer agents. *Chem. Rev.* **109,** 3012–3043 (2009).

64. Doll, R. & Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl. Cancer Inst.* **66,** 1191–1308 (1981).

65. Blot, W. J. & Tarone, R. E. Doll and Peto's quantitative estimates of cancer risks: holding generally true for 35 years. *J. Natl. Cancer Inst.* **107,** (2015).

66. Song, M. & Giovannucci, E. L. RE: doll and peto's quantitative estimates of cancer risks: holding generally true for 35 years. *J. Natl. Cancer Inst.* **107,** (2015).

67. WHO | Diet, nutrition and the prevention of chronic diseases Report of the joint WHO/FAO expert consultation. at <http://www.who.int/dietphysicalactivity/publications/trs916/summary/en/>

68. Hunter, D. J. Gene-environment interactions in human diseases. *Nat. Rev. Genet.* **6,** 287–298 (2005).

69. Weisburger, J. H. & Williams, G. M. The distinct health risk analyses required for genotoxic carcinogens and promoting agents. *Environ. Health Perspect.* **50,** 233–245 (1983).

70. Ames, B. N. Identifying environmental chemicals causing mutations and cancer.

Science **204,** 587–593 (1979).

71. Baccarelli, A. & Bollati, V. Epigenetics and environmental chemicals. *Curr. Opin. Pediatr.* **21,** 243–251 (2009).

72. Hou, L., Zhang, X., Wang, D. & Baccarelli, A. Environmental chemical exposures and human epigenetics. *Int. J. Epidemiol.* **41,** 79–105 (2012).

73. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128,** 635–638 (2007).

74. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75,** 843–854 (1993).

75. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).

76. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).

77. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834–838 (2005).

78. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin. Chem.* **56,** 1733–1741 (2010).

79. Liu, A. & Xu, X. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. *Methods Mol. Biol.* **724,** 259–267 (2011).

80. Izzotti, A. & Pulliero, A. The effects of environmental chemical carcinogens on the microRNA machinery. *Int J Hyg Environ Health* **217,** 601–627 (2014).

81. Vrijens, K., Bollati, V. & Nawrot, T. S. MicroRNAs as potential signatures of environmental exposure or effect: a systematic review. *Environ. Health Perspect.* **123,** 399–411 (2015).

82. Ross, S. A. & Davis, C. D. The emerging role of microRNAs and nutrition in modulating health and disease. *Annu. Rev. Nutr.* **34,** 305–336 (2014).

83. Banikazemi, Z. *et al.* Diet and cancer prevention: Dietary compounds, dietary MicroRNAs, and dietary exosomes. *J. Cell Biochem.* **119,** 185–196 (2018).

84. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39,** D152-7 (2011).

85. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23,** 4051–4060 (2004).

86. Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432,** 231–235 (2004).

87. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432,** 235–240 (2004).

88. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18,** 3016–3027 (2004).

89. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519,** 482–485 (2015).

90. Alarcón, C. R. *et al.* HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162,** 1299–1308 (2015).

91. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17,** 3011–3016 (2003).

92. Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303,** 95–98 (2004).

93. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10,** 185–191 (2004).

94. Park, J.-E. *et al.* Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475,** 201–205 (2011).

95. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436,** 740–744 (2005).

96. Tang, G. siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* **30,** 106–114 (2005).

97. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–86 (2007).

98. Cheloufi, S., Dos Santos, C. O., Chong, M. M. W. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465,**

584–589 (2010).

99. Heo, I. *et al.* Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151,** 521–532 (2012).

100. Wilczynska, A. & Bushell, M. The complexity of miRNA-mediated repression. *Cell Death Differ.* **22,** 22–33 (2015).

101. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27,** 91–105 (2007).

102. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153,** 654–665 (2013).

103. Lytle, J. R., Yario, T. A. & Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc. Natl. Acad. Sci. USA* **104,** 9667–9672 (2007).

104. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149,** 1635–1646 (2012).

105. Iwakawa, H.-O. & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* **25,** 651–665 (2015).

106. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).

107. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).

108. Wu, S. *et al.* Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* **29,** 2302–2308 (2010).

109. Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318,** 1931–1934 (2007).

110. Murphy, S. L., Xu, J., Kochanek, K. D., Curtin, S. C. & Arias, E. Deaths: final data for 2015. *Natl. Vital Stat. Rep.* **66,** 1–75 (2017).

111. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA* **99,** 15524–15529 (2002).

112. Calin, G. A. *et al.* MiR-15a and miR-16-1 cluster functions in human leukemia. *Proc. Natl. Acad. Sci. USA* **105,** 5166–5171 (2008).

113. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. USA* **102,** 13944–13949 (2005).

114. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* **39,** 673–677 (2007).

115. Kumar, M. S. *et al.* Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* **23,** 2700–2704 (2009).

116. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* **15,** 321–333 (2015).

117. Hata, A. & Kashima, R. Dysregulation of microRNA biogenesis machinery in cancer. *Crit Rev Biochem Mol Biol* **51,** 121–134 (2016).

118. Zhu, L. *et al.* miR-34a screened by miRNA profiling negatively regulates Wnt/β-catenin signaling pathway in Aflatoxin B1 induced hepatotoxicity. *Sci. Rep.* **5,** 16732 (2015).

119. Ligorio, M., Izzotti, A., Pulliero, A. & Arrigo, P. Mutagens interfere with microRNA maturation by inhibiting DICER. An in silico biology analysis. *Mutat. Res.* **717,** 116–128 (2011).

120. Dai, Q. *et al.* MicroRNA profiling of rats with ochratoxin A nephrotoxicity. *BMC Genomics* **15,** 333 (2014).

121. Kalinina, T. S., Kononchuk, V. V. & Gulyaeva, L. F. Expression of Hormonal Carcinogenesis Genes and Related Regulatory microRNAs in Uterus and Ovaries of DDT-Treated Female Rats. *Biochemistry. (Mosc)* **82,** 1118–1128 (2017).

122. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA* **103,** 2257–2261 (2006).

123. Jacobsen, A. *et al.* Analysis of microRNA-target interactions across diverse cancer types. *Nat. Struct. Mol. Biol.* **20,** 1325–1332 (2013).

124. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6,** 857–866 (2006).

125. Yanaihara, N. *et al.* Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9,** 189–198 (2006).

126. Schetter, A. J. *et al.* MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* **299,** 425–436 (2008).

127. Chen, X. *et al.* Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18,** 997–1006 (2008).

128. Resnick, K. E. *et al.* The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. *Gynecol. Oncol.* **112,** 55–59 (2009).

129. Zhang, Y. *et al.* Plasma microRNA-122 as a biomarker for viral-, alcohol-, and chemical-related hepatic diseases. *Clin. Chem.* **56,** 1830–1838 (2010).

130. Badrnya, S., Baumgartner, R. & Assinger, A. Smoking alters circulating plasma microvesicle pattern and microRNA signatures. *Thromb. Haemost.* **112,** 128–136 (2014).

131. Sugimura, T. Nutrition and dietary carcinogens. *Carcinogenesis* **21,** 387–395 (2000).

132. Basu, A. K. DNA damage, mutagenesis and cancer. *Int. J. Mol. Sci.* **19,** (2018).

133. Ferguson, L. R. Chronic inflammation and mutagenesis. *Mutat. Res.* **690,** 3–11 (2010).

134. El-Kenawi, A. & Ruffell, B. Inflammation, ROS, and mutagenesis. *Cancer Cell* **32,** 727–729 (2017).

135. Alshannaq, A. & Yu, J.-H. Occurrence, toxicity, and analysis of major mycotoxins in food. *Int. J. Environ. Res. Public Health* **14,** (2017).

136. Huang, X.-Y. *et al.* MicroRNA-429 Modulates Hepatocellular Carcinoma Prognosis and Tumorigenesis. *Gastroenterol Res Pract* **2013,** 804128 (2013).

137. Livingstone, M. C., Johnson, N. M., Roebuck, B. D., Kensler, T. W. & Groopman, J. D. Profound changes in miRNA expression during cancer initiation by aflatoxin B1 and their abrogation by the chemopreventive triterpenoid CDDO-Im. *Mol. Carcinog.* **56,** 2382–2390 (2017).

138. Liu, Y.-X. *et al.* MicroRNA-24 modulates aflatoxin B1-related hepatocellular carcinoma prognosis and tumorigenesis. *Biomed Res. Int.* **2014,** 482926 (2014).

139. Yang, W. *et al.* Genome-wide miRNA-profiling of aflatoxin B1-induced hepatic injury using deep sequencing. *Toxicol. Lett.* **226,** 140–149 (2014).

140. Loboda, A. *et al.* Effect of heme oxygenase-1 on ochratoxin A-induced nephrotoxicity in mice. *Int. J. Biochem. Cell Biol.* **84,** 46–57 (2017).

141. De Felice, B. *et al.* Genome-wide microRNA expression profiling in placentas from pregnant women exposed to BPA. *BMC Med. Genomics* **8,** 56 (2015).

142. Krauskopf, J. *et al.* MicroRNA profile for health risk assessment: Environmental exposure to persistent organic pollutants strongly affects the human blood microRNA machinery. *Sci. Rep.* **7,** 9262 (2017).

143. Vahdati Hassani, F., Mehri, S., Abnous, K., Birner-Gruenberger, R. & Hosseinzadeh, H. Protective effect of crocin on BPA-induced liver toxicity in rats through inhibition of oxidative stress and downregulation of MAPK and MAPKAP signaling pathway and miRNA-122 expression. *Food Chem. Toxicol.* **107,** 395–405 (2017).

144. Wei, J., Ding, D., Wang, T., Liu, Q. & Lin, Y. MiR-338 controls BPA-triggered pancreatic islet insulin secretory dysfunction from compensation to decompensation by targeting Pdx-1. *FASEB J.* **31,** 5184–5195 (2017).

145. Chanyshev, M. D., Kosorotikov, N. I., Titov, S. E., Kolesnikov, N. N. & Gulyaeva, L. F. Expression of microRNAs, CYP1A1 and CYP2B1 in the livers and ovaries of female rats treated with DDT and PAHs. *Life Sci.* **103,** 95–100 (2014).

146. Viegas, O., Novo, P., Pinto, E., Pinho, O. & Ferreira, I. M. P. L. V. O. Effect of charcoal types and grilling conditions on formation of heterocyclic aromatic amines (HAs) and polycyclic aromatic hydrocarbons (PAHs) in grilled muscle foods. *Food Chem. Toxicol.* **50,** 2128–2134 (2012).

147. Schut, H. A. & Snyderwine, E. G. DNA adducts of heterocyclic amine food mutagens: implications for mutagenesis and carcinogenesis. *Carcinogenesis* **20,** 353–368 (1999).

148. Abdel-Shafy, H. I. & Mansour, M. S. M. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum* **25,** 107–123 (2016).

149. Sinha, R. *et al.* Meat, meat cooking methods and preservation, and risk for colorectal adenoma. *Cancer Res.* **65,** 8034–8041 (2005).

150. Butler, L. M. *et al.* Heterocyclic amines, meat intake, and association with colon cancer in a population-based study. *Am. J. Epidemiol.* **157,** 434–445 (2003).

151. Deng, Q. *et al.* Plasma microRNA expression and micronuclei frequency in workers exposed to polycyclic aromatic hydrocarbons. *Environ. Health Perspect.* **122,** 719–725 (2014).

152. Shimizu, Y. *et al.* Benzo[a]pyrene carcinogenicity is lost in mice lacking the aryl

hydrocarbon receptor. *Proc. Natl. Acad. Sci. USA* **97,** 779–782 (2000).

153. Hockley, S. L. *et al.* AHR- and DNA-damage-mediated gene expression responses induced by benzo(a)pyrene in human cell lines. *Chem. Res. Toxicol.* **20,** 1797–1810 (2007).

154. Zuo, J., Brewer, D. S., Arlt, V. M., Cooper, C. S. & Phillips, D. H. Benzo pyrene-induced DNA adducts and gene expression profiles in target and non-target organs for carcinogenesis in mice. *BMC Genomics* **15,** 880 (2014).

155. Yauk, C. L., Jackson, K., Malowany, M. & Williams, A. Lack of change in microRNA expression in adult mouse liver following treatment with benzo(a)pyrene despite robust mRNA transcriptional response. *Mutat. Res.* **722,** 131–139 (2011).

156. Parasramka, M. A. *et al.* MicroRNA profiling of carcinogen-induced rat colon tumors and the influence of dietary spinach. *Mol. Nutr. Food Res.* **56,** 1259–1269 (2012).

157. Chen, Y.-S. *et al.* A miRNA signature for an environmental heterocyclic amine defined by a multi-organ carcinogenicity bioassay in the rat. *Arch. Toxicol.* **91,** 3415–3425 (2017).

158. Ren, X. *et al.* Arsenic responsive microRNAs in vivo and their potential involvement in arsenic-induced oxidative stress. *Toxicol. Appl. Pharmacol.* **283,** 198–209 (2015).

159. Rager, J. E. *et al.* Prenatal arsenic exposure and the epigenome: altered microRNAs associated with innate and adaptive immune signaling in newborn cord blood. *Environ Mol Mutagen* **55,** 196–208 (2014).

160. Rahman, M. L. *et al.* Regulation of birthweight by placenta-derived miRNAs: evidence from an arsenic-exposed birth cohort in Bangladesh. *Epigenetics* 1–18 (2018). doi:10.1080/15592294.2018.1481704

161. Rice, K. M., Walker, E. M., Wu, M., Gillette, C. & Blough, E. R. Environmental mercury and its toxic effects. *J. Prev. Med. Public Health* **47,** 74–83 (2014).

162. Ding, E. *et al.* MiR-92a and miR-486 are potential diagnostic biomarkers for mercury poisoning and jointly sustain NF-κB activity in mercury toxicity. *Sci. Rep.* **7,** 15980 (2017).

163. Li, Q. *et al.* Exploring the associations between microRNA expression profiles and environmental pollutants in human placenta from the National Children's Study (NCS). *Epigenetics* **10,** 793–802 (2015).

164. Bruha, R., Dvorak, K. & Petrtyl, J. Alcoholic liver disease. *World J Hepatol* **4,** 81–90 (2012).

165. Satishchandran, A. *et al.* MicroRNA 122, Regulated by GRLH2, Protects Livers of Mice and Patients From Ethanol-Induced Liver Disease. *Gastroenterology* **154,** 238–252.e7 (2018).

166. McCrae, J. C., Sharkey, N., Webb, D. J., Vliegenthart, A. D. B. & Dear, J. W. Ethanol consumption produces a small increase in circulating miR-122 in healthy individuals. *Clin. Toxicol.* **54,** 53–55 (2016).

167. Ten Berg, P. W. *et al.* Attending a social event and consuming alcohol is associated with changes in serum microRNA: a before and after study in healthy adults. *Biomarkers* 1–6 (2018). doi:10.1080/1354750X.2018.1499128

168. Asquith, M. *et al.* Chronic ethanol consumption modulates growth factor release, mucosal cytokine production, and microRNA expression in nonhuman primates. *Alcohol. Clin. Exp. Res.* **38,** 980–993 (2014).

169. Key, T. J. *et al.* Diet, nutrition and the prevention of cancer. *Public Health Nutr.* **7,** 187–200 (2004).

170. Mehta, R. G., Murillo, G., Naithani, R. & Peng, X. Cancer chemoprevention by natural products: how far have we come? *Pharm. Res.* **27,** 950–961 (2010).

171. Pan, M.-H., Lai, C.-S., Dushenkov, S. & Ho, C.-T. Modulation of inflammatory genes by natural dietary bioactive compounds. *J. Agric. Food Chem.* **57,** 4467–4477 (2009).

172. Sun, S.-Y., Hail, N. & Lotan, R. Apoptosis as a novel target for cancer chemoprevention. *J. Natl. Cancer Inst.* **96,** 662–672 (2004).

173. Anand, P., Kunnumakkara, A. B., Newman, R. A. & Aggarwal, B. B. Bioavailability of curcumin: problems and promises. *Mol. Pharm.* **4,** 807–818 (2007).

174. Prasad, S., Tyagi, A. K. & Aggarwal, B. B. Recent developments in delivery, bioavailability, absorption and metabolism of curcumin: the golden pigment from golden spice. *Cancer Res. Treat.* **46,** 2–18 (2014).

175. Gupta, S. C., Patchva, S., Koh, W. & Aggarwal, B. B. Discovery of curcumin, a component of golden spice, and its miraculous biological activities. *Clin. Exp. Pharmacol. Physiol.* **39,** 283–299 (2012).

176. Amalraj, A., Pius, A., Gopi, S. & Gopi, S. Biological activities of curcuminoids,

other biomolecules from turmeric and their derivatives - A review. *J Tradit Complement Med* **7,** 205–233 (2017).

177. Bao, B. *et al.* Curcumin analogue CDF inhibits pancreatic tumor growth by switching on suppressor microRNAs and attenuating EZH2 expression. *Cancer Res.* **72,** 335–345 (2012).

178. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420,** 860–867 (2002).

179. Toden, S. *et al.* Novel Evidence for Curcumin and Boswellic Acid-Induced Chemoprevention through Regulation of miR-34a and miR-27a in Colorectal Cancer. *Cancer Prev Res (Phila Pa)* **8,** 431–443 (2015).

180. Ma, F. *et al.* Anti-inflammatory effects of curcumin are associated with down regulating microRNA-155 in LPS-treated macrophages and mice. *Pharm Biol* **55,** 1263–1273 (2017).

181. Tili, E., Croce, C. M. & Michaille, J.-J. miR-155: on the crosstalk between inflammation and cancer. *Int Rev Immunol* **28,** 264–284 (2009).

182. Ding, X.-Q. *et al.* Curcumin protects against fructose-induced podocyte insulin signaling impairment through upregulation of miR-206. *Mol. Nutr. Food Res.* **59,** 2355–2370 (2015).

183. Shrikanta, A., Kumar, A. & Govindaswamy, V. Resveratrol content and antioxidant properties of underutilized fruits. *J. Food Sci. Technol.* **52,** 383–390 (2015).

184. Tokuşoglu, O., Unal, M. K. & Yemiş, F. Determination of the phytoalexin resveratrol (3,5,4'-trihydroxystilbene) in peanuts and pistachios by high-performance liquid chromatographic diode array (HPLC-DAD) and gas chromatography-mass spectrometry (GC-MS). *J. Agric. Food Chem.* **53,** 5003–5009 (2005).

185. Harper, C. E. *et al.* Resveratrol suppresses prostate cancer progression in transgenic mice. *Carcinogenesis* **28,** 1946–1953 (2007).

186. Sheth, S. *et al.* Resveratrol reduces prostate cancer growth and metastasis by inhibiting the Akt/MicroRNA-21 pathway. *PLoS One* **7,** e51655 (2012).

187. Dhar, S., Kumar, A., Rimando, A. M., Zhang, X. & Levenson, A. S. Resveratrol and pterostilbene epigenetically restore PTEN expression by targeting oncomiRs of the miR-17 family in prostate cancer. *Oncotarget* **6,** 27214–27226 (2015).

188. Altamemi, I. *et al.* Role of microRNAs in resveratrol-mediated mitigation of colitis-associated tumorigenesis in Apc(Min/+) mice. *J. Pharmacol. Exp. Ther.* **350,** 99–109 (2014).

189. Wu, F. & Cui, L. Resveratrol suppresses melanoma by inhibiting NF-κB/miR-221 and inducing TFG expression. *Arch Dermatol Res* **309,** 823–831 (2017).

190. Tomé-Carneiro, J. *et al.* One-year supplementation with a grape extract containing resveratrol modulates inflammatory-related microRNAs and cytokines expression in peripheral blood mononuclear cells of type 2 diabetes and hypertensive patients with coronary artery disease. *Pharmacol. Res.* **72,** 69–82 (2013).

191. Singh, B. N., Shankar, S. & Srivastava, R. K. Green tea catechin, epigallocatechin-3-gallate (EGCG): mechanisms, perspectives and clinical applications. *Biochem. Pharmacol.* **82,** 1807–1821 (2011).

192. Zhou, H. *et al.* Gene regulation mediated by microRNAs in response to green tea polyphenol EGCG in mouse lung cancer. *BMC Genomics* **15 Suppl 11,** S3 (2014).

193. Wang, H., Bian, S. & Yang, C. S. Green tea polyphenol EGCG suppresses lung cancer cell growth through upregulating miR-210 expression caused by stabilizing HIF-1α. *Carcinogenesis* **32,** 1881–1889 (2011).

194. Lam, T. K. *et al.* Influence of quercetin-rich food intake on microRNA expression in lung cancer tissues. *Cancer Epidemiol. Biomarkers Prev.* **21,** 2176–2184 (2012).

195. Khan, N. *et al.* Pomegranate fruit extract inhibits prosurvival pathways in human A549 lung carcinoma cells and tumor growth in athymic nude mice. *Carcinogenesis* **28,** 163–173 (2007).

196. Khan, N., Afaq, F., Kweon, M.-H., Kim, K. & Mukhtar, H. Oral consumption of pomegranate fruit extract inhibits growth and progression of primary lung tumors in mice. *Cancer Res.* **67,** 3475–3482 (2007).

197. Banerjee, N., Kim, H., Talcott, S. & Mertens-Talcott, S. Pomegranate polyphenolics suppressed azoxymethane-induced colorectal aberrant crypt foci and inflammation: possible role of miR-126/VCAM-1 and miR-126/PI3K/AKT/mTOR. *Carcinogenesis* **34,** 2814–2822 (2013).

198. Wu, Q. J. *et al.* Cruciferous vegetables intake and the risk of colorectal cancer: a meta-analysis of observational studies. *Ann. Oncol.* **24,** 1079–1087 (2013).

199. Clarke, J. D., Dashwood, R. H. & Ho, E. Multi-targeted prevention of cancer by sulforaphane. *Cancer Lett.* **269,** 291–304 (2008).

200. Gupta, P., Kim, B., Kim, S.-H. & Srivastava, S. K. Molecular targets of isothiocyanates in cancer: recent advances. *Mol. Nutr. Food Res.* **58,** 1685–1707 (2014).

201. Rajendran, P. *et al.* Nrf2 status affects tumor growth, HDAC3 gene promoter associations, and the response to sulforaphane in the colon. *Clin. Epigenetics* **7,** 102 (2015).

202. Izzotti, A. *et al.* Chemoprevention of cigarette smoke-induced alterations of MicroRNA expression in rat lungs. *Cancer Prev Res (Phila Pa)* **3,** 62–72 (2010).

203. Melkamu, T., Zhang, X., Tan, J., Zeng, Y. & Kassie, F. Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol. *Carcinogenesis* **31,** 252–258 (2010).

204. Wang, X. *et al.* Indole-3-carbinol inhibits tumorigenicity of hepatocellular carcinoma cells via suppression of microRNA-21 and upregulation of phosphatase and tensin homolog. *Biochim. Biophys. Acta* **1853,** 244–253 (2015).

205. Davidson, L. A. *et al.* n-3 Polyunsaturated fatty acids modulate carcinogen-directed non-coding microRNA signatures in rat colon. *Carcinogenesis* **30,** 2077–2084 (2009).

206. Shah, M. S. *et al.* Integrated microRNA and mRNA expression profiling in a rat colon carcinogenesis model: effect of a chemo-protective diet. *Physiol. Genomics* **43,** 640–654 (2011).

207. Shah, M. S. *et al.* Comparative effects of diet and carcinogen on microRNA expression in the stem cell niche of the mouse colonic crypt. *Biochim. Biophys. Acta* **1862,** 121–134 (2016).

208. Ricciardiello, L., Bazzoli, F. & Fogliano, V. Phytochemicals and colorectal cancer prevention--myth or reality? *Nat. Rev. Gastroenterol. Hepatol.* **8,** 592–596 (2011).

209. Hussain, S. S., Kumar, A. P. & Ghosh, R. Food-based natural products for cancer management: Is the whole greater than the sum of the parts? *Semin. Cancer Biol.* **40–41,** 233–246 (2016).

210. Jeyabalan, J. *et al.* Chemopreventive and therapeutic activity of dietary blueberry against estrogen-mediated breast cancer. *J. Agric. Food Chem.* **62,** 3963–3971 (2014).

211. Balzeau, J., Menezes, M. R., Cao, S. & Hagan, J. P. The LIN28/let-7 Pathway in Cancer. *Front. Genet.* **8,** 31 (2017).

212. Haider, B. A. *et al.* A critical evaluation of microRNA biomarkers in non-neoplastic disease. *PLoS One* **9,** e89565 (2014).

213. Witwer, K. W. Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin. Chem.* **61,** 56–63 (2015).

214. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. USA* **105,** 10513–10518 (2008).

215. Rabinowits, G., Gerçel-Taylor, C., Day, J. M., Taylor, D. D. & Kloecker, G. H. Exosomal microRNA: a diagnostic marker for lung cancer. *Clin. Lung Cancer* **10,** 42–46 (2009).

216. Cheng, H. H. *et al.* Plasma processing conditions substantially influence circulating microRNA biomarker levels. *PLoS One* **8,** e64795 (2013).

217. Corpet, D. E. & Pierre, F. How good are rodent models of carcinogenesis in predicting efficacy in humans? A systematic review and meta-analysis of colon chemoprevention in rats, mice and men. *Eur. J. Cancer* **41,** 1911–1922 (2005).

218. Le Magnen, C., Dutta, A. & Abate-Shen, C. Optimizing mouse models for precision cancer prevention. *Nat. Rev. Cancer* **16,** 187–196 (2016).

219. Threadgill, D. W. & Churchill, G. A. Ten years of the Collaborative Cross. *Genetics* **190,** 291–294 (2012).

220. Bogue, M. A., Churchill, G. A. & Chesler, E. J. Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database. *Mamm. Genome* **26,** 511–520 (2015).

221. Zhang, L. *et al.* Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res.* **22,** 107–126 (2012).

222. Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* **13,** 423–433 (2011).

223. Laparra, J. M. & Sanz, Y. Interactions of gut microbiota with functional food components and nutraceuticals. *Pharmacol. Res.* **61,** 219–225 (2010).

224. Marín, L., Miguélez, E. M., Villar, C. J. & Lombó, F. Bioavailability of dietary polyphenols and gut microbiota metabolism: antimicrobial properties. *Biomed Res. Int.* **2015,** 905215 (2015).

225. Liu, S. *et al.* The host shapes the gut microbiota via fecal microrna. *Cell Host Microbe* **19,** 32–43 (2016).

226. Dalmasso, G. *et al.* Microbiota modulate host gene expression via microRNAs. *PLoS One* **6,** e19293 (2011).

227. DeLuca, J. A. *et al.* Bisphenol-A alters microbiota metabolites derived from aromatic amino acids and worsens disease activity during colitis. *Exp. Biol. Med.* **243,** 864–875 (2018).

228. Ricciardiello, L., Ahnen, D. J. & Lynch, P. M. Chemoprevention of hereditary colon cancers: time for new strategies. *Nat. Rev. Gastroenterol. Hepatol.* **13,** 352–361 (2016).

229. Ross, S. A. & Davis, C. D. MicroRNA, nutrition, and cancer prevention. *Adv. Nutr.* **2,** 472–485 (2011).

230. Aqil, F., Munagala, R., Jeyabalan, J. & Vadhanam, M. V. Bioavailability of phytochemicals and its enhancement by drug delivery systems. *Cancer Lett.* **334,** 133–141 (2013).

231. Padhye, S. *et al.* Perspectives on chemopreventive and therapeutic potential of curcumin analogs in medicinal chemistry. *Mini Rev. Med. Chem.* **10,** 372–387 (2010).

232. Li, Z., Jiang, H., Xu, C. & Gu, L. A review: Using nanoparticles to enhance absorption and bioavailability of phenolic phytochemicals. *Food Hydrocoll.* **43,** 153–164 (2015).

233. Potter, J. D. The failure of cancer chemoprevention. *Carcinogenesis* **35,** 974–982 (2014).

234. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999-2015): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. (2018). at <https://gis.cdc.gov/Cancer/USCS/DataViz.html>

235. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319,** 525–532 (1988).

236. Magalhães, B., Peleteiro, B. & Lunet, N. Dietary patterns and colorectal cancer: systematic review and meta-analysis. *Eur J Cancer Prev* **21,** 15–23 (2012).

237. Vargas, A. J. & Thompson, P. A. Diet and nutrient factors in colorectal cancer risk.

Nutr Clin Pract **27,** 613–623 (2012).

238. Fung, T. *et al.* Major dietary patterns and the risk of colorectal cancer in women. *Arch. Intern. Med.* **163,** 309–314 (2003).

239. Sachse, C. A pharmacogenetic study to investigate the role of dietary carcinogens in the etiology of colorectal cancer. *Carcinogenesis* **23,** 1839–1850 (2002).

240. Thune, I. & Furberg, A. S. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. *Med. Sci. Sports Exerc.* **33,** S530–50; discussion S609 (2001).

241. Samad, A. K. A., Taylor, R. S., Marshall, T. & Chapman, M. A. S. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis* **7,** 204–213 (2005).

242. Whittemore, A. S. *et al.* Diet, physical activity, and colorectal cancer among chinese in north america and china. *JNCI Journal of the National Cancer Institute* **82,** 915–926 (1990).

243. Allen, J. M. *et al.* Exercise alters gut microbiota composition and function in lean and obese humans. *Med. Sci. Sports Exerc.* **50,** 747–757 (2018).

244. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15,** R89 (2014).

245. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2014).

246. Singh, R. K. *et al.* Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15,** 73 (2017).

247. Tindall, A. M., Petersen, K. S. & Kris-Etherton, P. M. Dietary Patterns Affect the Gut Microbiome-The Link to Risk of Cardiometabolic Diseases. *J. Nutr.* **148,** 1402–1407 (2018).

248. Gevers, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biol.* **10,** e1001377 (2012).

249. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Health Dis.* **26,** 26191 (2015).

250. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474,** 327–336 (2011).

251. Sobhani, I. *et al.* Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One* **6,** e16393 (2011).

252. Mayer, E. A., Knight, R., Mazmanian, S. K., Cryan, J. F. & Tillisch, K. Gut microbes and the brain: paradigm shift in neuroscience. *J. Neurosci.* **34,** 15490–15496 (2014).

253. Grivennikov, S. I. Inflammation and colorectal cancer: colitis-associated neoplasia. *Semin Immunopathol* **35,** 229–244 (2013).

254. Rubin, D. C., Shaker, A. & Levin, M. S. Chronic intestinal inflammation: inflammatory bowel disease and colitis-associated colon cancer. *Front. Immunol.* **3,** 107 (2012).

255. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

256. Momparler, R. L. Cancer epigenetics. *Oncogene* **22,** 6479–6483 (2003).

257. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357,** (2017).

258. Larrea, E. *et al.* New Concepts in Cancer Biomarkers: Circulating miRNAs in Liquid Biopsies. *Int. J. Mol. Sci.* **17,** (2016).

259. Rao, A. K. D. M., Rajkumar, T. & Mani, S. Perspectives of long non-coding RNAs in cancer. *Mol. Biol. Rep.* **44,** 203–218 (2017).

260. Fulbright, L. E., Ellermann, M. & Arthur, J. C. The microbiome and the hallmarks of cancer. *PLoS Pathog.* **13,** e1006480 (2017).

261. Horgan, R. P. & Kenny, L. C. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist* **13,** 189–195 (2011).

262. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18,** 83 (2017).

263. Franzosa, E. A. *et al.* Sequencing and beyond: integrating molecular "omics" for microbial community profiling. *Nat. Rev. Microbiol.* **13,** 360–372 (2015).

264. Ferguson, J. F. *et al.* Nutrigenomics, the Microbiome, and Gene-Environment Interactions: New Directions in Cardiovascular Disease Research, Prevention, and Treatment: A Scientific Statement From the American Heart Association. *Circ Cardiovasc Genet* **9,** 291–313 (2016).

265. Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. M. Expression profiling using cDNA microarrays. *Nat. Genet.* **21,** 10–14 (1999).

266. Hoheisel, J. D. Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* **7,** 200–210 (2006).

267. Zhang, J., Chiodini, R., Badr, A. & Zhang, G. The impact of next-generation sequencing on genomics. *J Genet Genomics* **38,** 95–109 (2011).

268. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74,** 560–564 (1977).

269. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467 (1977).

270. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

271. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).

272. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107,** 1–8 (2016).

273. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145 (2008).

274. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **100,** 3960–3964 (2003).

275. Haque, F., Li, J., Wu, H.-C., Liang, X.-J. & Guo, P. Solid-State and Biological Nanopore for Real-Time Sensing of Single Chemical and Sequencing of DNA. *Nano Today* **8,** 56–74 (2013).

276. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).

277. Kamps, R. *et al.* Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* **18,** (2017).

278. Chaitankar, V. *et al.* Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog Retin Eye Res* **55,** 1–31 (2016).

279. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).

280. Tringe, S. G. & Rubin, E. M. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* **6,** 805–814 (2005).

281. Belizário, J. E. & Napolitano, M. Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. *Front. Microbiol.* **6,** 1050 (2015).

282. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13,** 260–270 (2012).

283. Griffiths, J. A brief history of mass spectrometry. *Anal. Chem.* **80,** 5678–5683 (2008).

284. Horning, E. C. & Horning, M. G. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin. Chem.* **17,** 802–809 (1971).

285. Pauling, L., Robinson, A. B., Teranishi, R. & Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. USA* **68,** 2374–2376 (1971).

286. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* **27,** 228–237 (2008).

287. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).

288. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell Proteomics* **1,** 845–867 (2002).

289. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **137,** 293–300 (2012).

290. Becker, S., Kortz, L., Helmschrodt, C., Thiery, J. & Ceglarek, U. LC-MS-based metabolomics in the clinical laboratory. *J. Chromatogr. B, Analyt. Technol. Biomed. Life Sci.* **883–884,** 68–75 (2012).

291. Dunn, W. B. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.* **5,** 011001 (2008).

292. Griffiths, W. J. & Wang, Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem. Soc. Rev.* **38,** 1882–1896 (2009).

293. Altelaar, A. F. M., Munoz, J. & Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48

(2013).

294. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537,** 347–355 (2016).

295. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13,** 263–269 (2012).

296. Miura, D., Fujimura, Y. & Wariishi, H. In situ metabolomic mass spectrometry imaging: recent advances and difficulties. *J. Proteomics* **75,** 5052–5060 (2012).

297. Amstalden van Hove, E. R., Smith, D. F. & Heeren, R. M. A. A concise review of mass spectrometry imaging. *J. Chromatogr. A* **1217,** 3946–3954 (2010).

298. Lao, V. V. & Grady, W. M. Epigenetics and colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **8,** 686–700 (2011).

299. Ogino, S. & Goel, A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn* **10,** 13–27 (2008).

300. Worthley, D. L. & Leggett, B. A. Colorectal cancer: molecular features and clinical opportunities. *Clin Biochem Rev* **31,** 31–38 (2010).

301. Nazemalhosseini Mojarad, E., Kuppen, P. J., Aghdaei, H. A. & Zali, M. R. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol. Hepatol. Bed Bench* **6,** 120–128 (2013).

302. Wetterstrand, K. A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) . at <http://www.genome.gov/sequencingcostsdata>

303. Ahn, J. *et al.* Human gut microbiome and risk for colorectal cancer. *J. Natl. Cancer Inst.* **105,** 1907–1911 (2013).

304. Chan, E. C. Y. *et al.* Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J. Proteome Res.* **8,** 352–361 (2009).

305. Hirayama, A. *et al.* Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res.* **69,** 4918–4925 (2009).

306. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab* **4,** 11 (2016).

307. Manzoni, C. *et al.* Genome, transcriptome and proteome: the rise of omics data and

their integration in biomedical sciences. *Brief. Bioinformatics* **19,** 286–302 (2018).

308. Misra, B. B., Fahrmann, J. F. & Grapov, D. Review of emerging metabolomic tools and resources: 2015-2016. *Electrophoresis* **38,** 2257–2274 (2017).

309. Putignani, L. & Dallapiccola, B. Foodomics as part of the host-microbiota-exposome interplay. *J. Proteomics* **147,** 3–20 (2016).

310. Shahaf, N. *et al.* The WEIZMASS spectral library for high-confidence metabolite identification. *Nat. Commun.* **7,** 12423 (2016).

311. Ge, H., Walhout, A. J. M. & Vidal, M. Integrating "omic" information: a bridge between genomics and systems biology. *Trends Genet.* **19,** 551–560 (2003).

312. Joyce, A. R. & Palsson, B. Ø. The model organism as a system: integrating "omics" data sets. *Nat. Rev. Mol. Cell Biol.* **7,** 198–210 (2006).

313. Buescher, J. M. & Driggers, E. M. Integration of omics: more than the sum of its parts. *Cancer Metab* **4,** 4 (2016).

314. Yugi, K., Kubota, H., Hatano, A. & Kuroda, S. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple "Omic" Layers. *Trends Biotechnol.* **34,** 276–290 (2016).

315. Tang, W. *et al.* MicroRNA-29a promotes colorectal cancer metastasis by regulating matrix metalloproteinase 2 and E-cadherin via KLF4. *Br. J. Cancer* **110,** 450–458 (2014).

316. Vishnubalaji, R. *et al.* Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. *Cell Death Dis.* **6,** e1614 (2015).

317. Cantini, L. *et al.* MicroRNA-mRNA interactions underlying colorectal cancer molecular subtypes. *Nat. Commun.* **6,** 8878 (2015).

318. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19,** 1423–1437 (2013).

319. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501,** 328–337 (2013).

320. Espina, V., Heiby, M., Pierobon, M. & Liotta, L. A. Laser capture microdissection technology. *Expert Rev Mol Diagn* **7,** 647–657 (2007).

321. Navin, N. E. Cancer genomics: one cell at a time. *Genome Biol.* **15,** 452 (2014).

322. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33,** 155–160 (2015).

323. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49,** 708–718 (2017).

324. Roerink, S. F. *et al.* Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556,** 457–462 (2018).

325. Satoh, K. *et al.* Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC. *Proc. Natl. Acad. Sci. USA* **114,** E7697–E7706 (2017).

326. Itzkowitz, S. H. & Yio, X. Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* **287,** G7-17 (2004).

327. Munkholm, P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. *Aliment. Pharmacol. Ther.* **18 Suppl 2,** 1–5 (2003).

328. McMillan, D. C., Canna, K. & McArdle, C. S. Systemic inflammatory response predicts survival following curative resection of colorectal cancer. *Br. J. Surg.* **90,** 215–219 (2003).

329. Beaugerie, L. & Itzkowitz, S. H. Cancers complicating inflammatory bowel disease. *N. Engl. J. Med.* **372,** 1441–1452 (2015).

330. Lasry, A., Zinger, A. & Ben-Neriah, Y. Inflammatory networks underlying colorectal cancer. *Nat. Immunol.* **17,** 230–240 (2016).

331. Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* **124,** 3617–3633 (2014).

332. Morgan, X. C. *et al.* Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol.* **16,** 67 (2015).

333. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* **17,** 141–152 (2013).

334. Mu, C., Yang, Y., Luo, Z., Guan, L. & Zhu, W. The Colonic Microbiome and Epithelial Transcriptome Are Altered in Rats Fed a High-Protein Diet Compared with a Normal-Protein Diet. *J. Nutr.* **146,** 474–483 (2016).

335. Mann, E., Wetzels, S. U., Wagner, M., Zebeli, Q. & Schmitz-Esser, S. Metatranscriptome Sequencing Reveals Insights into the Gene Expression and Functional Potential of Rumen Wall Bacteria. *Front. Microbiol.* **9,** 43 (2018).

336. Abu-Ali, G. S. *et al.* Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* **3,** 356–366 (2018).

337. Bultman, S. J. Interplay between diet, gut microbiota, epigenetic events, and colorectal cancer. *Mol. Nutr. Food Res.* **61,** (2017).

338. Carmody, R. N. *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17,** 72–84 (2015).

339. Li, H. *et al.* The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6,** 8292 (2015).

340. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167,** 1339–1353.e21 (2016).

341. Li, R. *et al.* Transcriptome and DNA Methylome Analysis in a Mouse Model of Diet-Induced Obesity Predicts Increased Risk of Colorectal Cancer. *Cell Rep.* **22,** 624–637 (2018).

342. Zheng, X. *et al.* The footprints of gut microbial-mammalian co-metabolism. *J. Proteome Res.* **10,** 5512–5522 (2011).

343. Krishnan, S., Alden, N. & Lee, K. Pathways and functions of gut microbiota metabolism impacting host physiology. *Curr. Opin. Biotechnol.* **36,** 137–145 (2015).

344. Krishnan, S. *et al.* Gut Microbiota-Derived Tryptophan Metabolites Modulate Inflammatory Response in Hepatocytes and Macrophages. *Cell Rep.* **23,** 1099–1111 (2018).

345. Weir, T. L. *et al.* Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One* **8,** e70803 (2013).

346. El Aidy, S. *et al.* Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. *ISME J.* **7,** 743–755 (2013).

347. Snijders, A. M. *et al.* Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nat. Microbiol.* **2,** 16221 (2016).

348. Chakrabarti, A. *et al.* Transcriptomics-driven lipidomics (TDL) identifies the

microbiome-regulated targets of ileal lipid metabolism. *npj Syst. Biol. Appl.* **3,** 33 (2017).

349. Cortese, R., Lu, L., Yu, Y., Ruden, D. & Claud, E. C. Epigenome-Microbiome crosstalk: A potential new paradigm influencing neonatal susceptibility to disease. *Epigenetics* **11,** 205–215 (2016).

350. Qin, Y. & Wade, P. A. Crosstalk between the microbiome and epigenome: messages from bugs. *J. Biochem.* **163,** 105–112 (2018).

351. Richards, A. *et al.* Variation in gut microbiota composition impacts host gene expression. *BioRxiv* (2017). doi:10.1101/210294

352. Qin, Y. *et al.* An obesity-associated gut microbiome reprograms the intestinal epigenome and leads to altered colonic gene expression. *Genome Biol.* **19,** 7 (2018).

353. Tárraga López, P. J., Albero, J. S. & Rodríguez-Montes, J. A. Primary and secondary prevention of colorectal cancer. *Clin. Med. Insights Gastroenterol.* **7,** 33–46 (2014).

354. Demark-Wahnefried, W., Rock, C. L., Patrick, K. & Byers, T. Lifestyle interventions to reduce cancer risk and improve outcomes. *Am. Fam. Physician* **77,** 1573–1578 (2008).

355. Brenner, H. & Chen, C. The colorectal cancer epidemic: challenges and opportunities for primary, secondary and tertiary prevention. *Br. J. Cancer* **119,** 785–792 (2018).

356. Crosara Teixeira, M., Braghiroli, M. I., Sabbaga, J. & Hoff, P. M. Primary prevention of colorectal cancer: myth or reality? *World J. Gastroenterol.* **20,** 15060–15069 (2014).

357. Sandler, R. S. *et al.* A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. *N. Engl. J. Med.* **348,** 883–890 (2003).

358. Bibbins-Domingo, K. & U.S. Preventive Services Task Force. Aspirin use for the primary prevention of cardiovascular disease and colorectal cancer: U.S. preventive services task force recommendation statement. *Ann. Intern. Med.* **164,** 836–845 (2016).

359. Drew, D. A., Cao, Y. & Chan, A. T. Aspirin and colorectal cancer: the promise of precision chemoprevention. *Nat. Rev. Cancer* **16,** 173–186 (2016).

360. Imperiale, T. F. *et al.* Multitarget stool DNA testing for colorectal-cancer

screening. *N. Engl. J. Med.* **370,** 1287–1297 (2014).

361. Colon cancer screening made easy | Cologuard®. at <https://www.cologuardtest.com/>

362. Hardy, H., Harris, J., Lyon, E., Beal, J. & Foey, A. D. Probiotics, prebiotics and immunomodulation of gut mucosal defences: homeostasis and immunopathology. *Nutrients* **5,** 1869–1912 (2013).

363. Hirayama, K. & Rafter, J. The role of probiotic bacteria in cancer prevention. *Microbes Infect.* **2,** 681–686 (2000).

364. Ambalam, P., Raman, M., Purama, R. K. & Doble, M. Probiotics, prebiotics and colorectal cancer prevention. *Best Pract Res Clin Gastroenterol* **30,** 119–131 (2016).

365. Cousin, F. J. *et al.* The probiotic Propionibacterium freudenreichii as a new adjuvant for TRAIL-based therapy in colorectal cancer. *Oncotarget* **7,** 7161–7178 (2016).

366. Lenoir, M. *et al.* Lactobacillus casei BL23 regulates Treg and Th17 T-cell populations and reduces DMH-associated colorectal cancer. *J. Gastroenterol.* **51,** 862–873 (2016).

367. Kuugbee, E. D. *et al.* Structural change in microbiota by a probiotic cocktail enhances the gut barrier and reduces cancer via TLR2 signaling in a rat model of colon cancer. *Dig. Dis. Sci.* **61,** 2908–2920 (2016).

368. Lee, C. H. *et al.* Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. *JAMA* **315,** 142–149 (2016).

369. Kassam, Z., Lee, C. H., Yuan, Y. & Hunt, R. H. Fecal microbiota transplantation for Clostridium difficile infection: systematic review and meta-analysis. *Am. J. Gastroenterol.* **108,** 500–508 (2013).

370. Bakken, J. S. *et al.* Treating Clostridium difficile infection with fecal microbiota transplantation. *Clin. Gastroenterol. Hepatol.* **9,** 1044–1049 (2011).

371. Kostic, A. D., Xavier, R. J. & Gevers, D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* **146,** 1489–1499 (2014).

372. Damman, C. J., Miller, S. I., Surawicz, C. M. & Zisman, T. L. The microbiome and inflammatory bowel disease: is there a therapeutic role for fecal microbiota

transplantation? *Am. J. Gastroenterol.* **107,** 1452–1459 (2012).

373. Wong, S. H. *et al.* Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology* **153,** 1621–1633.e6 (2017).

374. Ianiro, G., Bibbò, S., Gasbarrini, A. & Cammarota, G. Therapeutic modulation of gut microbiota: current clinical applications and future perspectives. *Curr Drug Targets* **15,** 762–770 (2014).

375. Lehouritis, P., Springer, C. & Tangney, M. Bacterial-directed enzyme prodrug therapy. *J. Control. Release* **170,** 120–131 (2013).

376. Forbes, N. S. Engineering the perfect (bacterial) cancer therapy. *Nat. Rev. Cancer* **10,** 785–794 (2010).

377. Cheng, C. M. *et al.* Tumor-targeting prodrug-activating bacteria for cancer therapy. *Cancer Gene Ther* **15,** 393–401 (2008).

378. Yazawa, K., Fujimori, M., Amano, J., Kano, Y. & Taniguchi, S. Bifidobacterium longum as a delivery system for cancer gene therapy: selective localization and growth in hypoxic tumors. *Cancer Gene Ther* **7,** 269–274 (2000).

379. Pawelek, J. M., Low, K. B. & Bermudes, D. Tumor-targeted Salmonella as a novel anticancer vector. *Cancer Res.* **57,** 4537–4544 (1997).

380. Stritzker, J. *et al.* Tumor-specific colonization, tissue distribution, and gene induction by probiotic Escherichia coli Nissle 1917 in live mice. *Int. J. Med. Microbiol.* **297,** 151–162 (2007).

381. Hsieh, Y.-T. *et al.* Impediments to enhancement of CPT-11 anticancer activity by E. coli directed beta-glucuronidase therapy. *PLoS One* **10,** e0118028 (2015).

382. Murphy, C., Rettedal, E., Lehouritis, P., Devoy, C. & Tangney, M. Intratumoural production of TNFα by bacteria mediates cancer therapy. *PLoS One* **12,** e0180034 (2017).

383. Zheng, J. H. *et al.* Two-step enhanced cancer immunotherapy with engineered Salmonella typhimurium secreting heterologous flagellin. *Sci. Transl. Med.* **9,** (2017).

384. Ho, C. L. *et al.* Engineered commensal microbes for diet-mediated colorectal-cancer chemoprevention. *Nat. Biomed. Eng.* **2,** 27–37 (2018).

385. Evaldson, G., Heimdahl, A., Kager, L. & Nord, C. E. The normal human anaerobic

microflora. *Scand J Infect Dis Suppl* **35,** 9–15 (1982).

386. Hutter, C. M. *et al.* Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res.* **72,** 2036–2044 (2012).

387. Dienstmann, R. *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **17,** 79–92 (2017).

388. Sommer, F. & Bäckhed, F. The gut microbiota--masters of host development and physiology. *Nat. Rev. Microbiol.* **11,** 227–238 (2013).

389. Takasuna, K. *et al.* Involvement of beta-glucuronidase in intestinal microflora in the intestinal toxicity of the antitumor camptothecin derivative irinotecan hydrochloride (CPT-11) in rats. *Cancer Res.* **56,** 3752–3757 (1996).

390. Zou, J. *et al.* Fiber-Mediated Nourishment of Gut Microbiota Protects against Diet-Induced Obesity by Restoring IL-22-Mediated Colonic Health. *Cell Host Microbe* **23,** 41–53.e4 (2018).

391. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* **57,** 1–24 (2018).

392. Tilg, H., Adolph, T. E., Gerner, R. R. & Moschen, A. R. The intestinal microbiota in colorectal cancer. *Cancer Cell* **33,** 954–964 (2018).

393. Ogino, S., Nowak, J. A., Hamada, T., Milner, D. A. & Nishihara, R. Insights into Pathogenic Interactions Among Environment, Host, and Tumor at the Crossroads of Molecular Pathology and Epidemiology. *Annu. Rev. Pathol.* (2018). doi:10.1146/annurev-pathmechdis-012418-012818

394. Micha, R., Wallace, S. K. & Mozaffarian, D. Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation* **121,** 2271–2283 (2010).

395. Le, N. T. *et al.* A Prospective Analysis of Meat Mutagens and Colorectal Cancer in the Nurses' Health Study and Health Professionals Follow-up Study. *Environ. Health Perspect.* **124,** 1529–1536 (2016).

396. Fu, Z. *et al.* Well-done meat intake and meat-derived mutagen exposures in relation to breast cancer risk: the Nashville Breast Health Study. *Breast Cancer Res. Treat.* **129,** 919–928 (2011).

397. Cross, A. J. *et al.* A prospective study of meat and meat mutagens and prostate cancer risk. *Cancer Res.* **65,** 11779–11784 (2005).

398. Adamson, R. H., Thorgeirsson, U. P. & Sugimura, T. Extrapolation of heterocyclic amine carcinogenesis data from rodents and nonhuman primates to humans. *Arch. Toxicol. Suppl.* **18,** 303–318 (1996).

399. Turesky, R. J., Gross, G. A., Stillwell, W. G., Skipper, P. L. & Tannenbaum, S. R. Species differences in metabolism of heterocyclic aromatic amines, human exposure, and biomonitoring. *Environ. Health Perspect.* **102 Suppl 6,** 47–51 (1994).

400. Alexander, J. Formation and processing of reactive metabolites of the heterocyclic amines. *Arch. Toxicol. Suppl.* **18,** 275–285 (1996).

401. Dashwood, R. H., Suzui, M., Nakagama, H., Sugimura, T. & Nagao, M. High frequency of beta-catenin (ctnnb1) mutations in the colon tumors induced by two heterocyclic amines in the F344 rat. *Cancer Res.* **58,** 1127–1129 (1998).

402. David, R., Ebbels, T. & Gooderham, N. Synergistic and Antagonistic Mutation Responses of Human MCL-5 Cells to Mixtures of Benzo[a]pyrene and 2-Amino-1-Methyl-6-Phenylimidazo[4,5-b]pyridine: Dose-Related Variation in the Joint Effects of Common Dietary Carcinogens. *Environ. Health Perspect.* **124,** 88–96 (2016).

403. Nicken, P., Schröder, B., von Keutz, A., Breves, G. & Steinberg, P. The colon carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) is actively secreted in the distal colon of the rat: an integrated view on the role of PhIP transport and metabolism in PhIP-induced colon carcinogenesis. *Arch. Toxicol.* **87,** 895–904 (2013).

404. Augustsson, K., Skog, K., Jägerstad, M. & Steineck, G. Assessment of the human exposure to heterocyclic amines. *Carcinogenesis* **18,** 1931–1935 (1997).

405. Nagao, M. *et al.* Human exposure to carcinogenic heterocyclic amines and their mutational fingerprints in experimental animals. *Environ. Health Perspect.* **104 Suppl 3,** 497–501 (1996).

406. Busquets, R. *et al.* Biomonitoring of dietary heterocyclic amines and metabolites in urine by liquid phase microextraction: 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), a possible biomarker of exposure to dietary PhIP. *Chem. Res. Toxicol.* **26,** 233–240 (2013).

407. Le Marchand, L. *et al.* Dose validation of PhIP hair level as a biomarker of heterocyclic aromatic amines exposure: a feeding study. *Carcinogenesis* **37,** 685–691 (2016).

408. Peng, L. & Turesky, R. J. Mass spectrometric characterization of 2-amino-1-

methyl-6-phenylimidazo[4,5-b]pyridine N-oxidized metabolites bound at Cys34 of human serum albumin. *Chem. Res. Toxicol.* **24,** 2004–2017 (2011).

409. Turteltaub, K. W., Vogel, J. S., Frantz, C. E. & Shen, N. Fate and distribution of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine in mice at a human dietary equivalent dose. *Cancer Res.* **52,** 4682–4687 (1992).

410. National Toxicology Program. NTP 12th report on carcinogens. *Rep. Carcinog.* **12,** iii-499 (2011).

411. International Agency for Research on Cancer (IARC). PhIP (2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine)(Group 2B) (IARC Summary & Evaluation, Volume 56, 1993). at <http://www.inchem.org/documents/iarc/vol56/08-phip.html>

412. Ghoshal, A., Preisegger, K. H., Takayama, S., Thorgeirsson, S. S. & Snyderwine, E. G. Induction of mammary tumors in female Sprague-Dawley rats by the food-derived carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine and effect of dietary fat. *Carcinogenesis* **15,** 2429–2433 (1994).

413. Shirai, T. *et al.* The prostate: a target for carcinogenicity of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) derived from cooked foods. *Cancer Res.* **57,** 195–198 (1997).

414. Ubagai, T. *et al.* Efficient induction of rat large intestinal tumors with a new spectrum of mutations by intermittent administration of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine in combination with a high fat diet. *Carcinogenesis* **23,** 197–200 (2002).

415. Hasegawa, R. *et al.* Increased risk of mammary carcinoma development following transplacental and trans-breast milk exposure to a food-derived carcinogen, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), in Sprague-Dawley rats. *Cancer Res.* **55,** 4333–4338 (1995).

416. Li, Q., Dashwood, W. M., Zhong, X., Nakagama, H. & Dashwood, R. H. Bcl-2 overexpression in PhIP-induced colon tumors: cloning of the rat Bcl-2 promoter and characterization of a pathway involving beta-catenin, c-Myc and E2F1. *Oncogene* **26,** 6194–6202 (2007).

417. Wang, R. *et al.* beta-catenin is strongly elevated in rat colonic epithelium following short-term intermittent treatment with 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and a high-fat diet. *Cancer Sci.* **99,** 1754–1759 (2008).

418. Kwon, I. K. *et al.* PKG inhibits TCF signaling in colon cancer cells by blocking

beta-catenin expression and activating FOXO4. *Oncogene* **29,** 3423–3434 (2010).

419. Wang, H. *et al.* Genetic variation in the bioactivation pathway for polycyclic hydrocarbons and heterocyclic amines in relation to risk of colorectal neoplasia. *Carcinogenesis* **32,** 203–209 (2011).

420. Wiese, F. W., Thompson, P. A. & Kadlubar, F. F. Carcinogen substrate specificity of human COX-1 and COX-2. *Carcinogenesis* **22,** 5–10 (2001).

421. Nelson, C. P. *et al.* Protection against 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine cytotoxicity and DNA adduct formation in human prostate by glutathione S-transferase P1. *Cancer Res.* **61,** 103–109 (2001).

422. Choudhary, S., Sood, S., Donnell, R. L. & Wang, H.-C. R. Intervention of human breast cell carcinogenesis chronically induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. *Carcinogenesis* **33,** 876–885 (2012).

423. Bolleyn, J. *et al.* MicroRNAs as key regulators of xenobiotic biotransformation and drug response. *Arch. Toxicol.* **89,** 1523–1541 (2015).

424. Wei, J., Li, F., Yang, J., Liu, X. & Cho, W. C. MicroRNAs as regulators of airborne pollution-induced lung inflammation and carcinogenesis. *Arch. Toxicol.* **89,** 677–685 (2015).

425. Yu, H. W. & Cho, W. C. The role of microRNAs in toxicology. *Arch. Toxicol.* **89,** 319–325 (2015).

426. Melis, J. P. M. *et al.* In vivo murine hepatic microRNA and mRNA expression signatures predicting the (non-)genotoxic carcinogenic potential of chemicals. *Arch. Toxicol.* **88,** 1023–1034 (2014).

427. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12,** 861–874 (2011).

428. Ho, S.-M. *et al.* Exposure of human prostaspheres to bisphenol A epigenetically regulates SNORD family noncoding rnas via histone modification. *Endocrinology* **156,** 3984–3995 (2015).

429. Bollati, V. *et al.* Exposure to metal-rich particulate matter modifies the expression of candidate microRNAs in peripheral blood leukocytes. *Environ. Health Perspect.* **118,** 763–768 (2010).

430. Parasramka, M. A., Ho, E., Williams, D. E. & Dashwood, R. H. MicroRNAs, diet, and cancer: new mechanistic insights on the epigenetic actions of phytochemicals. *Mol. Carcinog.* **51,** 213–230 (2012).

431. Dimmeler, S. & Nicotera, P. MicroRNAs in age-related diseases. *EMBO Mol. Med.* **5,** 180–190 (2013).

432. Yin, J. *et al.* Differential expression of serum miR-126, miR-141 and miR-21 as novel biomarkers for early detection of liver metastasis in colorectal cancer. *Chin. J. Cancer Res.* **26,** 95–103 (2014).

433. Hou, L. *et al.* Particulate air pollution exposure and expression of viral and human micrornas in blood: the beijing truck driver air pollution study. *Environ. Health Perspect.* **124,** 344–350 (2016).

434. Papaioannou, M. D., Koufaris, C. & Gooderham, N. J. The cooked meat-derived mammary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) elicits estrogenic-like microRNA responses in breast cancer cells. *Toxicol. Lett.* **229,** 9–16 (2014).

435. Parasramka, M. A. *et al.* A role for low-abundance miRNAs in colon cancer: the miR-206/Krüppel-like factor 4 (KLF4) axis. *Clin. Epigenetics* **4,** 16 (2012).

436. Ito, N. *et al.* Carcinogenicity of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) in the rat. *Mutat. Res.* **376,** 107–114 (1997).

437. Wang, R. *et al.* Protective versus promotional effects of white tea and caffeine on PhIP-induced tumorigenesis and beta-catenin expression in the rat. *Carcinogenesis* **29,** 834–839 (2008).

438. Simonich, M. T. *et al.* Natural chlorophyll inhibits aflatoxin B1-induced multi-organ carcinogenesis in the rat. *Carcinogenesis* **28,** 1294–1302 (2007).

439. Jubert, C. *et al.* Effects of chlorophyll and chlorophyllin on low-dose aflatoxin B(1) pharmacokinetics in human volunteers. *Cancer Prev Res (Phila Pa)* **2,** 1015–1022 (2009).

440. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.* **43,** W251-7 (2015).

441. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

442. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4,** (2015).

443. Cline, M. S. *et al.* Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. *Sci. Rep.* **3,** 2652 (2013).

444. Barh, D., Malhotra, R., Ravi, B. & Sindhurani, P. MicroRNA let-7: an emerging next-generation cancer therapeutic. *Curr. Oncol.* **17,** 70–80 (2010).

445. Boyerinas, B., Park, S.-M., Hau, A., Murmann, A. E. & Peter, M. E. The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* **17,** F19-36 (2010).

446. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* **139,** 693–706 (2009).

447. Buscaglia, L. E. B. & Li, Y. Apoptosis and the target genes of microRNA-21. *Chin. J. Cancer* **30,** 371–380 (2011).

448. Sachdeva, M. & Mo, Y.-Y. miR-145-mediated suppression of cell growth, invasion and metastasis. *Am J Transl Res* **2,** 170–180 (2010).

449. Rajendran, P., Williams, D. E., Ho, E. & Dashwood, R. H. Metabolism as a key to histone deacetylase inhibition. *Crit Rev Biochem Mol Biol* **46,** 181–199 (2011).

450. Rajendran, P., Ho, E., Williams, D. E. & Dashwood, R. H. Dietary phytochemicals, HDAC inhibition, and DNA damage/repair defects in cancer cells. *Clin. Epigenetics* **3,** 4 (2011).

451. Tellez-Plaza, M. *et al.* Association of global DNA methylation and global DNA hydroxymethylation with metals and other exposures in human blood DNA samples. *Environ. Health Perspect.* **122,** 946–954 (2014).

452. Johnson, G. S. *et al.* A functional pseudogene, NMRAL2P, is regulated by Nrf2 and serves as a coactivator of NQO1 in sulforaphane-treated colon cancer cells. *Mol. Nutr. Food Res.* **61,** (2017).

453. Hayashi, S., Møller, M. E. & Thorgeirsson, S. S. Genotoxicity of heterocyclic amines in the Salmonella/hepatocyte system. *Jpn. J. Cancer Res.* **76,** 835–845 (1985).

454. Hernaez, J. F., Xu, M. & Dashwood, R. H. Antimutagenic activity of tea towards 2-hydroxyamino-3-methylimidazo[4,5-f]quinoline: effect of tea concentration and brew time on electrophile scavenging. *Mutat. Res.* **402,** 299–306 (1998).

455. Xu, M. & Dashwood, R. H. Chemoprevention studies of heterocyclic amine-induced colon carcinogenesis. *Cancer Lett.* **143,** 179–183 (1999).

456. Kim, S., Guo, J., O'Sullivan, M. G., Gallaher, D. D. & Turesky, R. J. Comparative DNA adduct formation and induction of colonic aberrant crypt foci in mice exposed to 2-amino-9H-pyrido[2,3-b]indole, 2-amino-3,4-dimethylimidazo[4,5-

f]quinoline, and azoxymethane. *Environ Mol Mutagen* **57,** 125–136 (2016).

457. Minami, K. *et al.* miRNA expression atlas in male rat. *Sci. Data* **1,** 140005 (2014).

458. Li, X. *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor PDCD4 and promotes cell transformation, proliferation, and metastasis in renal cell carcinoma. *Cell Physiol. Biochem.* **33,** 1631–1642 (2014).

459. Ferraro, A. *et al.* Epigenetic regulation of miR-21 in colorectal cancer: ITGB4 as a novel miR-21 target and a three-gene network (miR-21-ITGB4-PDCD4) as predictor of metastatic tumor potential. *Epigenetics* **9,** 129–141 (2014).

460. Chao, T.-F., Xiong, H.-H., Liu, W., Chen, Y. & Zhang, J.-X. MiR-21 mediates the radiation resistance of glioblastoma cells by regulating PDCD4 and hMSH2. *J. Huazhong Univ. Sci. Technolog. Med. Sci.* **33,** 525–529 (2013).

461. Yu, Y. *et al.* miR-21 and miR-145 cooperation in regulation of colon cancer stem cells. *Mol. Cancer* **14,** 98 (2015).

462. Wang, X. *et al.* Regulation of let-7 and its target oncogenes (Review). *Oncol. Lett.* **3,** 955–960 (2012).

463. Tilghman, S. L. *et al.* Endocrine disruptor regulation of microRNA expression in breast carcinoma cells. *PLoS One* **7,** e32754 (2012).

464. Karaayvaz, M. *et al.* Prognostic significance of miR-215 in colon cancer. *Clin Colorectal Cancer* **10,** 340–347 (2011).

465. Song, B. *et al.* Molecular mechanism of chemoresistance by miR-215 in osteosarcoma and colon cancer cells. *Mol. Cancer* **9,** 96 (2010).

466. Hamada, S. *et al.* MiR-126 acts as a tumor suppressor in pancreatic cancer cells via the regulation of ADAM9. *Mol. Cancer Res.* **10,** 3–10 (2012).

467. Zhu, N. *et al.* Endothelial-specific intron-derived miR-126 is down-regulated in human breast cancer and targets both VEGFA and PIK3R2. *Mol. Cell. Biochem.* **351,** 157–164 (2011).

468. Liu, B., Peng, X.-C., Zheng, X.-L., Wang, J. & Qin, Y.-W. MiR-126 restoration down-regulate VEGF and inhibit the growth of lung cancer cell lines in vitro and in vivo. *Lung Cancer* **66,** 169–175 (2009).

469. Fabbri, M. *et al.* MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl. Acad. Sci. USA* **104,** 15805–15810 (2007).

470. Sengupta, S. *et al.* MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins. *Proc. Natl. Acad. Sci. USA* **105,** 5874–5878 (2008).

471. Knize, M. G., Salmon, C. P., Pais, P. & Felton, J. S. Food heating and the formation of heterocyclic aromatic amine and polycyclic aromatic hydrocarbon mutagens/carcinogens. *Adv. Exp. Med. Biol.* **459,** 179–193 (1999).

472. Hamidi, E. N., Hajeb, P., Selamat, J. & Abdull Razis, A. F. Polycyclic Aromatic Hydrocarbons (PAHs) and their Bioaccessibility in Meat: a Tool for Assessing Human Cancer Risk. *Asian Pac. J. Cancer Prev.* **17,** 15–23 (2016).

473. Farrugia, M. K., Vanderbilt, D. B., Salkeni, M. A. & Ruppert, J. M. Kruppel-like Pluripotency Factors as Modulators of Cancer Cell Therapeutic Responses. *Cancer Res.* **76,** 1677–1682 (2016).

474. Rowland, B. D., Bernards, R. & Peeper, D. S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nat. Cell Biol.* **7,** 1074–1082 (2005).

475. Foster, K. W. *et al.* Induction of KLF4 in basal keratinocytes blocks the proliferation-differentiation switch and initiates squamous epithelial dysplasia. *Oncogene* **24,** 1491–1500 (2005).

476. Li, J. *et al.* Deficiency of the Kruppel-like factor KLF4 correlates with increased cell proliferation and enhanced skin tumorigenesis. *Carcinogenesis* **33,** 1239–1246 (2012).

477. Sevilla, L. M. *et al.* Glucocorticoid receptor and Klf4 co-regulate anti-inflammatory genes in keratinocytes. *Mol. Cell. Endocrinol.* **412,** 281–289 (2015).

478. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9,** 313–323 (2009).

479. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148,** 1258–1270 (2012).

480. Ghaisas, S., Maher, J. & Kanthasamy, A. Gut microbiome in health and disease: Linking the microbiome-gut-brain axis and environmental factors in the pathogenesis of systemic and neurodegenerative diseases. *Pharmacol. Ther.* **158,** 52–62 (2016).

481. Tlaskalová-Hogenová, H. *et al.* The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of

human diseases. *Cell Mol Immunol* **8,** 110–120 (2011).

482. Vannucci, L. *et al.* Colorectal carcinogenesis in germ-free and conventionally reared rats: different intestinal environments affect the systemic immunity. *Int. J. Oncol.* **32,** 609–617 (2008).

483. Kelly, C. R. *et al.* Effect of Fecal Microbiota Transplantation on Recurrence in Multiply Recurrent Clostridium difficile Infection: A Randomized Trial. *Ann. Intern. Med.* **165,** 609–616 (2016).

484. Gupta, S., Allen-Vercoe, E. & Petrof, E. O. Fecal microbiota transplantation: in perspective. *Therap. Adv. Gastroenterol.* **9,** 229–239 (2016).

485. Nomura, A. M. Y. *et al.* Association of vegetable, fruit, and grain intakes with colorectal cancer: the Multiethnic Cohort Study. *Am. J. Clin. Nutr.* **88,** 730–737 (2008).

486. Erfani, F., Hassandokht, M. R., Jabbari, A. & Barzegar, M. Effect of cultivar on chemical composition of some Iranian spinach. *Pak. J. Biol. Sci.* **10,** 602–606 (2007).

487. Roberts, J. L. & Moreau, R. Functional properties of spinach (Spinacia oleracea L.) phytochemicals and bioactives. *Food Funct* **7,** 3337–3353 (2016).

488. Singh, A. *et al.* Detection of flavonoids from Spinacia oleracea leaves using HPLC-ESI-QTOF-MS/MS and UPLC-QqQLIT-MS/MS techniques. *Nat Prod Res* 1–4 (2018). doi:10.1080/14786419.2018.1489395

489. Okonkwo, A. *et al.* Heterocyclic analogs of sulforaphane trigger DNA damage and impede DNA repair in colon cancer cells: interplay of hats and hdacs. *Mol. Nutr. Food Res.* **62,** e1800228 (2018).

490. Rajendran, P., Johnson, G. S. & Dashwood, R. H. CCAR2 acetylation establishes a BET/BRD9 acetyl switch in response to combined deacetylase and bromodomain inhibition. *Cancer Res.* (2018).

491. Irving, A. A. *et al.* The utility of Apc-mutant rats in modeling human colon cancer. *Dis. Model. Mech.* **7,** 1215–1225 (2014).

492. Myzak, M. C., Karplus, P. A., Chung, F.-L. & Dashwood, R. H. A novel mechanism of chemoprotection by sulforaphane: inhibition of histone deacetylase. *Cancer Res.* **64,** 5767–5774 (2004).

493. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12,** R60 (2011).

494. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31,** 814–821 (2013).

495. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30,** 3123–3124 (2014).

496. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

497. Anders, S., Pyl, P. T. & Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *BioRxiv* (2014). doi:10.1101/002824

498. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17,** 10 (2011).

499. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).

500. Rajendran, P. *et al.* Histone deacetylase turnover and recovery in sulforaphane-treated colon cancer cells: competing actions of 14-3-3 and Pin1 in HDAC3/SMRT corepressor complex dissociation/reassembly. *Mol. Cancer* **10,** 68 (2011).

501. Wang, R. *et al.* Epigenetic inactivation of endothelin-2 and endothelin-3 in colon cancer. *Int. J. Cancer* **132,** 1004–1012 (2013).

502. Wang, R. *et al.* NADPH oxidase overexpression in human colon cancers and rat colon tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). *Int. J. Cancer* **128,** 2581–2590 (2011).

503. Son, J. S. *et al.* Altered Interactions between the Gut Microbiome and Colonic Mucosa Precede Polyposis in APCMin/+ Mice. *PLoS One* **10,** e0127985 (2015).

504. Tsilimigras, M. C. B., Fodor, A. & Jobin, C. Carcinogenesis and therapeutics: the microbiota perspective. *Nat. Microbiol.* **2,** 17008 (2017).

505. Ericsson, A. C. *et al.* Differential susceptibility to colorectal cancer due to naturally occurring gut microbiota. *Oncotarget* **6,** 33689–33704 (2015).

506. Salonen, A. *et al.* Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J.* **8,** 2218–2230 (2014).

507. Rey, F. E. *et al.* Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc. Natl. Acad. Sci. USA* **110,** 13582–13587 (2013).

508. Jakobsson, H. E. *et al.* The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep.* **16,** 164–177 (2015).

509. Ghosh, A. K. & Vaughan, D. E. PAI-1 in tissue fibrosis. *J. Cell Physiol.* **227,** 493–507 (2012).

510. Sakakibara, T. *et al.* Plasminogen activator inhibitor-1 as a potential marker for the malignancy of colorectal cancer. *Br. J. Cancer* **93,** 799–803 (2005).

511. Ganesh, S. *et al.* Prognostic relevance of plasminogen activators and their inhibitors in colorectal cancer. *Cancer Res.* **54,** 4065–4071 (1994).

512. Flemer, B. *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* **66,** 633–643 (2017).

513. Villadsen, S. B. *et al.* The miR-143/-145 cluster regulates plasminogen activator inhibitor-1 in bladder cancer. *Br. J. Cancer* **106,** 366–374 (2012).

514. Chen, Q. *et al.* MiR-34a regulates apoptosis in liver cells by targeting the KLF4 gene. *Cell Mol Biol Lett* **19,** 52–64 (2014).

515. Xu, J. *et al.* Dynamic down-regulation of Krüppel-like factor 4 in colorectal adenoma-carcinoma sequence. *J. Cancer Res. Clin. Oncol.* **134,** 891–898 (2008).

516. Patel, N. V., Ghaleb, A. M., Nandan, M. O. & Yang, V. W. Expression of the tumor suppressor Krüppel-like factor 4 as a prognostic predictor for colon cancer. *Cancer Epidemiol. Biomarkers Prev.* **19,** 2631–2638 (2010).

517. Chakraborty, C., Sharma, A. R., Sharma, G., Doss, C. G. P. & Lee, S.-S. Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine. *Mol. Ther. Nucleic Acids* **8,** 132–143 (2017).

518. Wali, R. K., DeLaCruz, M. & Kunte, D. Polyethylene Glycol (PEG) Suppresses Adenomas and Aberrant Crypt Foci (ACF) in the Polyposis in Rat Colon Model (Pirc): Implications for Colorectal Cancer Prevention. *Gastroenterology* (2012).

519. Roager, H. M. *et al.* Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat. Microbiol.* **1,** 16093 (2016).

520. Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One* **7,** e39743 (2012).

521. Sommer, F., Nookaew, I., Sommer, N., Fogelstrand, P. & Bäckhed, F. Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome*

Biol. **16,** 62 (2015).

522. Wooden, B., Goossens, N., Hoshida, Y. & Friedman, S. L. Using Big Data to Discover Diagnostics and Therapeutics for Gastrointestinal and Liver Diseases. *Gastroenterology* **152,** 53–67.e3 (2017).

523. Weersma, R. K., Xavier, R. J., Vermeire, S. & Barrett, J. C. Multiomics analyses to deliver the most effective treatment to every patient with inflammatory bowel disease. *Gastroenterology* (2018). doi:10.1053/j.gastro.2018.07.039

524. Bray, F., Jemal, A., Grey, N., Ferlay, J. & Forman, D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *Lancet Oncol.* **13,** 790–801 (2012).

525. Pritchard, C. C. & Grady, W. M. Colorectal cancer molecular biology moves into clinical practice. *Gut* **60,** 116–129 (2011).

526. Fearnhead, N. S., Britton, M. P. & Bodmer, W. F. The ABC of APC. *Hum. Mol. Genet.* **10,** 721–733 (2001).

527. Zhao, S. *et al.* Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. *Genome Res.* **14,** 1851–1860 (2004).

528. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26,** 51–78 (2007).

529. Crutchfield, C. A., Thomas, S. N., Sokoll, L. J. & Chan, D. W. Advances in mass spectrometry-based clinical biomarker discovery. *Clin. Proteomics* **13,** 1 (2016).

530. Armitage, E. G. & Barbas, C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87,** 1–11 (2014).

531. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17,** 451–459 (2016).

532. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477,** 54–60 (2011).

533. Qiu, Y. *et al.* A distinct metabolic signature of human colorectal cancer with prognostic potential. *Clin. Cancer Res.* **20,** 2136–2146 (2014).

534. Jiménez, B. *et al.* 1H HR-MAS NMR spectroscopy of tumor-induced local metabolic "field-effects" enables colorectal cancer staging and prognostication. *J. Proteome Res.* **12,** 959–968 (2013).

535. Backshall, A. *et al.* Detection of metabolic alterations in non-tumor gastrointestinal

tissue of the Apc(Min/+) mouse by (1)H MAS NMR spectroscopy. *J. Proteome Res.* **8,** 1423–1430 (2009).

536. Yoshie, T. *et al.* Regulation of the metabolite profile by an APC gene mutation in colorectal cancer. *Cancer Sci.* **103,** 1010–1021 (2012).

537. Li, F. *et al.* Modulation of colon cancer by nutmeg. *J. Proteome Res.* **14,** 1937–1946 (2015).

538. Xie, G. *et al.* Metabonomic profiling reveals cancer chemopreventive effects of american ginseng on colon carcinogenesis in apc(min/+) mice. *J. Proteome Res.* **14,** 3336–3347 (2015).

539. Wang, L.-S. *et al.* Modulation of genetic and epigenetic biomarkers of colorectal cancer in humans by black raspberries: a phase I pilot study. *Clin. Cancer Res.* **17,** 598–610 (2011).

540. Pan, P. *et al.* Black raspberries suppress colonic adenoma development in ApcMin/+ mice: relation to metabolite profiles. *Carcinogenesis* **36,** 1245–1253 (2015).

541. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46,** W486–W494 (2018).

542. Li, S. *et al.* Predicting network activity from high throughput metabolomics. *PLoS Comput. Biol.* **9,** e1003123 (2013).

543. Zhou, B., Xiao, J. F., Tuli, L. & Ressom, H. W. LC-MS-based metabolomics. *Mol. Biosyst.* **8,** 470–481 (2012).

544. Tian, Y. *et al.* Tissue metabonomic phenotyping for diagnosis and prognosis of human colorectal cancer. *Sci. Rep.* **6,** 20790 (2016).

545. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324,** 1029–1033 (2009).

546. Liberti, M. V. & Locasale, J. W. The warburg effect: how does it benefit cancer cells? *Trends Biochem. Sci.* **41,** 211–218 (2016).

547. Patra, K. C. & Hay, N. The pentose phosphate pathway and cancer. *Trends Biochem. Sci.* **39,** 347–354 (2014).

548. Cho, E. S., Cha, Y. H., Kim, H. S., Kim, N. H. & Yook, J. I. The pentose phosphate pathway as a potential target for cancer therapy. *Biomol. Ther. (Seoul)*

26, 29–38 (2018).

549. Santos, C. R. & Schulze, A. Lipid metabolism in cancer. *FEBS J.* **279,** 2610–2623 (2012).

550. Kuhajda, F. P. Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology. *Nutrition* **16,** 202–208 (2000).

551. Smith, B. *et al.* Addiction to Coupling of the Warburg Effect with Glutamine Catabolism in Cancer Cells. *Cell Rep.* **17,** 821–836 (2016).

552. Tessem, M.-B. *et al.* Evaluation of lactate and alanine as metabolic biomarkers of prostate cancer using 1H HR-MAS spectroscopy of biopsy tissues. *Magn. Reson. Med.* **60,** 510–516 (2008).

553. Ni, J., Wu, G. D., Albenberg, L. & Tomov, V. T. Gut microbiota and IBD: causation or correlation? *Nat. Rev. Gastroenterol. Hepatol.* **14,** 573–584 (2017).

554. Eltzschig, H. K., Sitkovsky, M. V. & Robson, S. C. Purinergic signaling during inflammation. *N. Engl. J. Med.* **367,** 2322–2333 (2012).

555. Burnstock, G. Purinergic signalling in the gastrointestinal tract and related organs in health and disease. *Purinergic Signal* **10,** 3–50 (2014).

556. Crittenden, S. *et al.* Purine metabolism controls innate lymphoid cell function and protects against intestinal injury. *Immunol. Cell Biol.* **96,** 1049–1059 (2018).

557. Rybaczyk, L. *et al.* New bioinformatics approach to analyze gene expressions and signaling pathways reveals unique purine gene dysregulation profiles that distinguish between CD and UC. *Inflamm. Bowel Dis.* **15,** 971–984 (2009).

558. Chiaro, T. R. *et al.* A member of the gut mycobiota modulates host purine metabolism exacerbating colitis in mice. *Sci. Transl. Med.* **9,** (2017).

559. Boyce, M. & Bertozzi, C. R. Bringing chemistry to life. *Nat. Methods* **8,** 638–642 (2011).

560. Geva-Zatorsky, N. *et al.* In vivo imaging and tracking of host-microbiota interactions via metabolic labeling of gut anaerobic bacteria. *Nat. Med.* **21,** 1091–1100 (2015).

561. de Graaf, A. A. *et al.* Profiling human gut bacterial metabolism and its kinetics using [U-13C]glucose and NMR. *NMR Biomed.* **23,** 2–12 (2010).

562. Basson, A., Trotter, A., Rodriguez-Palacios, A. & Cominelli, F. Mucosal Interactions between Genetics, Diet, and Microbiome in Inflammatory Bowel

Disease. *Front. Immunol.* **7,** 290 (2016).

563. Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30,** 332–338 (2014).

564. Wahlström, A., Sayin, S. I., Marschall, H.-U. & Bäckhed, F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab.* **24,** 41–50 (2016).

565. Fang, S. *et al.* Intestinal FXR agonism promotes adipose tissue browning and reduces obesity and insulin resistance. *Nat. Med.* **21,** 159–165 (2015).

566. Parséus, A. *et al.* Microbiota-induced obesity requires farnesoid X receptor. *Gut* **66,** 429–437 (2017).

567. Calzada, E., Onguka, O. & Claypool, S. M. Phosphatidylethanolamine metabolism in health and disease. *Int. Rev. Cell Mol. Biol.* **321,** 29–88 (2016).

568. Thiennimitr, P. *et al.* Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. USA* **108,** 17480–17485 (2011).

569. Harrison, P. J., Dunn, T. M. & Campopiano, D. J. Sphingolipid biosynthesis in man and microbes. *Nat Prod Rep* **35,** 921–954 (2018).

570. Liang, J. *et al.* Sphingosine-1-phosphate links persistent STAT3 activation, chronic intestinal inflammation, and development of colitis-associated cancer. *Cancer Cell* **23,** 107–120 (2013).

571. Rosa, R. *et al.* Sphingosine kinase 1 overexpression contributes to cetuximab resistance in human colorectal cancer models. *Clin. Cancer Res.* **19,** 138–147 (2013).

572. O'Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13,** 691–706 (2016).

573. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* **6,** 6528 (2015).

574. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10,** 766 (2014).

575. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat. Rev. Microbiol.* **10,** 575–582 (2012).

576. Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S. & Frizelle, F. A. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci. Rep.* **7,** 11590 (2017).

577. Lee, W.-J. & Hase, K. Gut microbiota-generated metabolites in animal health and disease. *Nat. Chem. Biol.* **10,** 416–424 (2014).

578. Gardinassi, L. G. *et al.* Integrative metabolomics and transcriptomics signatures of clinical tolerance to Plasmodium vivax reveal activation of innate cell immunity and T cell signaling. *Redox Biol* **17,** 158–170 (2018).

579. Li, S. *et al.* Metabolic phenotypes of response to vaccination in humans. *Cell* **169,** 862–877.e17 (2017).

580. Baxter, J. H. & Steinberg, D. Absorption of phytol from dietary chlorophyll in the rat. *J. Lipid Res.* **8,** 615–620 (1967).

581. Ferruzzi, M. G., Failla, M. L. & Schwartz, S. J. Assessment of degradation and intestinal cell uptake of carotenoids and chlorophyll derivatives from spinach puree using an in vitro digestion and Caco-2 human cell model. *J. Agric. Food Chem.* **49,** 2082–2089 (2001).

582. Passon, M. *et al.* Polyphenol Phase-II Metabolites are Detectable in Human Plasma after Ingestion of 13 C Labeled Spinach-a Pilot Intervention Trial in Young Healthy Adults. *Mol. Nutr. Food Res.* **62,** e1701003 (2018).

583. Tu, H.-C. *et al.* LIN28 cooperates with WNT signaling to drive invasive intestinal and colorectal adenocarcinoma in mice and humans. *Genes Dev.* **29,** 1074–1086 (2015).

584. Ghaleb, A. M. & Yang, V. W. Krüppel-like factor 4 (KLF4): What we currently know. *Gene* **611,** 27–37 (2017).

585. Cho, Y. G. *et al.* Genetic and epigenetic analysis of the KLF4 gene in gastric cancer. *APMIS* **115,** 802–808 (2007).

586. Yang, W.-T. & Zheng, P.-S. Promoter hypermethylation of KLF4 inactivates its tumor suppressor function in cervical carcinogenesis. *PLoS One* **9,** e88827 (2014).

587. Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. & Kosik, K. S. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137,** 647–658 (2009).

588. Cui, J., Shi, M., Quan, M. & Xie, K. Regulation of EMT by KLF4 in gastrointestinal cancer. *Curr. Cancer Drug Targets* **13,** 986–995 (2013).

589. Zhao, W. *et al.* Identification of Krüppel-like factor 4 as a potential tumor suppressor gene in colorectal cancer. *Oncogene* **23,** 395–402 (2004).

590. Hu, J., Li, H., Wu, C., Zhao, X. & Liu, C. The Prognostic Value of Decreased KLF4 in Digestive System Cancers: A Meta-Analysis from 17 Studies. *Dis. Markers* **2017,** 3064246 (2017).

591. Grosswendt, S. *et al.* Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell* **54,** 1042–1054 (2014).

592. Hu, D. *et al.* Interplay between arginine methylation and ubiquitylation regulates KLF4-mediated genome stability and carcinogenesis. *Nat. Commun.* **6,** 8419 (2015).

593. Shoaie, S. & Nielsen, J. Elucidating the interactions between the human gut microbiota and its host through metabolic modeling. *Front. Genet.* **5,** 86 (2014).

594. Lukens, J. R. *et al.* Dietary modulation of the microbiome affects autoinflammatory disease. *Nature* **516,** 246–249 (2014).

595. Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555,** 210–215 (2018).

596. Smits, L. P., Bouter, K. E. C., de Vos, W. M., Borody, T. J. & Nieuwdorp, M. Therapeutic potential of fecal microbiota transplantation. *Gastroenterology* **145,** 946–953 (2013).

597. Segura, P. P. *et al.* SEOM guide to primary and secondary prevention of cancer: 2014. *Clin Transl Oncol* **16,** 1072–1078 (2014).

598. Primary, secondary and tertiary prevention | Institute for Work & Health. at <https://www.iwh.on.ca/what-researchers-mean-by/primary-secondary-and-tertiary-prevention>

599. Shackleton, M., Quintana, E., Fearon, E. R. & Morrison, S. J. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* **138,** 822–829 (2009).

600. Fidler, I. J. Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res.* **38,** 2651–2660 (1978).

601. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313,** 1960–1964 (2006).

602. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based

technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14,** 618–630 (2013).

603. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).

604. Geng, J. *et al.* Co-occurrence of driver and passenger bacteria in human colorectal cancer. *Gut Pathog* **6,** 26 (2014).

605. Geng, J., Fan, H., Tang, X., Zhai, H. & Zhang, Z. Diversified pattern of the human colorectal cancer microbiome. *Gut Pathog* **5,** 2 (2013).

606. Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc. Natl. Acad. Sci. USA* **111,** 18321–18326 (2014).

607. Lepage, P. *et al.* Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm. Bowel Dis.* **11,** 473–480 (2005).

608. Swales, J. G., Hamm, G., Clench, M. R. & Goodwin, R. J. A. Mass spectrometry imaging and its application in pharmaceutical research and development: A concise review. *Int. J. Mass Spectrom.* (2018). doi:10.1016/j.ijms.2018.02.007

609. Wu, C.-L. *et al.* Metabolomic imaging for human prostate cancer detection. *Sci. Transl. Med.* **2,** 16ra8 (2010).

610. Stoeckli, M., Chaurand, P., Hallahan, D. E. & Caprioli, R. M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat. Med.* **7,** 493–496 (2001).

611. Cortez, M. A. *et al.* MicroRNAs in body fluids--the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* **8,** 467–477 (2011).

612. Ward, J. A. *et al.* Circulating Cell and Plasma microRNA Profiles Differ between Non-ST-Segment and ST-Segment-Elevation Myocardial Infarction. *Fam Med Med Sci Res* **2,** 108 (2013).

613. Moldovan, L. *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell Mol. Med.* **18,** 371–390 (2014).

614. Postel, M., Roosen, A., Laurent-Puig, P., Taly, V. & Wang-Renault, S.-F. Droplet-based digital PCR and next generation sequencing for monitoring circulating tumor DNA: a cancer diagnostic perspective. *Expert Rev Mol Diagn* **18,** 7–17 (2018).

615. Hudecova, I. Digital PCR analysis of circulating nucleic acids. *Clin. Biochem.* **48,** 948–956 (2015).

616. Chen, W. W. *et al.* BEAMing and Droplet Digital PCR Analysis of Mutant IDH1 mRNA in Glioma Patient Serum and Cerebrospinal Fluid Extracellular Vesicles. *Mol. Ther. Nucleic Acids* **2,** e109 (2013).

617. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83,** 8604–8610 (2011).

618. Kitz, J., Lowes, L. E., Goodale, D. & Allan, A. L. Circulating tumor cell analysis in preclinical mouse models of metastasis. *Diagnostics (Basel)* **8,** (2018).

619. Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **27,** 1897–1905 (2016).

620. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88,** 524–545 (2016).

621. Wu, B. *et al.* High-Throughput Screening by Nuclear Magnetic Resonance (HTS by NMR) for the Identification of PPIs Antagonists. *Curr. Top. Med. Chem.* **15,** 2032–2042 (2015).

622. Kostidis, S., Addie, R. D., Morreau, H., Mayboroda, O. A. & Giera, M. Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Anal. Chim. Acta* **980,** 1–24 (2017).

623. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3,** 211–221 (2007).

624. Meng, C. *et al.* Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief. Bioinformatics* **17,** 628–641 (2016).

625. Stone, E. A. & Ayroles, J. F. Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet.* **5,** e1000479 (2009).