Cancer Genomics and Proteomics

Eden Dennis

Cancer Genomics and Proteomics

Edited by **Eden Dennis**

F A
FOSTER
A C A D E M I C S

New Jersey

Cancer Genomics and Proteomics
Edited by Eden Dennis

Cancer Genomics and Proteomics
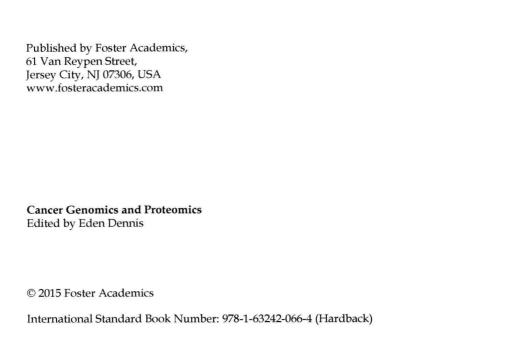
Preface

The book focuses on the field of cancer genomics and proteomics. It has been compiled with elaborative reviews of the most general cancers, from bench to bedside applications by an internationally renowned team of veterans and it will certainly contribute to the medical and scientific community by illustrating updated discoveries of oncogenomics and their potential applications in cancer translational research. This comprehensive book aims to serve as a useful resource for clinicians, students, oncologists, scientists and health professionals engaged in cancer research.

The information contained in this book is the result of intensive hard work done by researchers in this field. All due efforts have been made to make this book serve as a complete guiding source for students and researchers. The topics in this book have been comprehensively explained to help readers understand the growing trends in the field.

I would like to thank the entire group of writers who made sincere efforts in this book and my family who supported me in my efforts of working on this book. I take this opportunity to thank all those who have been a guiding force throughout my life.

Editor

Contents

Genomic Expression Profiling in Cancer

Genomic Expression Profiles: From Molecular Signatures to Clinical Oncology Translation

Norfilza M. Mokhtar, Nor Azian Murad, Then Sue Mian and Rahman Jamal

Additional information is available at the end of the chapter

1. Introduction

Study related to diseases such as cancer has changed tremendously for a decade. For many years, the study was restricted largely to a single gene or a few genes in cancer cells. The studies have uncovered the roles of individual genes in the uncontrolled behavior of cancer cells. Studying the functional roles of genes in cancer cells has deepened our understanding not only the cancer cells as well as normal cells. Since 2003 onwards, the trend of publications was focusing on the analysis of thousands of genes with related molecular pathways. Steps taken from this analysis is then translated to clinical practice for the biological markers for an early detection, monitoring, prognosis of the disease and response to therapy.

The completion of the Human Genome Project in 2003 enabled a new era in biological sciences, in particular molecular medicine. The availability of the database of full sequences of approximately 3 billion base pairs and approximately 30,000 genes in human DNA will lead to a better understanding of physiological and pathophysiological changes in human body. Genome-wide expression technology allows the simultenous analysis of thousands of genes in a single experiment. The availability of the technology alters the way biological experiments can be designed. This has resulted of so called 'discovery biology'. The large amount of data produced by microarray resulted to new and unexpected features of cellular functions.

Since it was first introduced, microarrays are widely used for basic research, the development of prognostic tests, target discovery or toxicology researchs. The new form of cancer screening utilizes the molecular data generated from microarray studies. We will discuss the application of gene profiling data in the clinical screening of cancer. It is hopefully will give a broad picture the pipeline required to discover biomarkers of cancer.

The chapter is subdivided into a series of sections; each will discuss the scientific evidence on the molecular and cellular studies in selected cancers. We will try to critically assess the evidence upon which the theory on the cancer was built. The conversion of normal cells into cancer cells is a complex process and multistep processes. Scientists for many years tried to uncover the causes of cancer and emphasize certain oncogenes, or tumor suppressor genes or other groups of genes. Further information on how these findings were translated to the clinical settings will be provided. To date, with the massive gene expression profile data available to the researchers, there are still major hurdles in validating and reproducing the results. We will discuss the major drawbacks associated with the use of molecular signatures as the biomarkers or response to treatment.

2. Molecular signatures in colorectal carcinoma

Colorectal cancer (CRC) is a type of cancers that develops in the colon or the rectum of the human digestive system or gastrointestinal tract (1).Colorectal cancer is the third leading cause of death in both men and women in the US with 141,210 new cases and 49,380 death expected in 2011 (2). CRC progresses slowly over a period of time usually between 10 to 15 years (3, 4). The tumor begins with noncancerous polyps where the tissues that form the lining of the colon or rectum differentiate into cancerous tissues (5). Approximately, 96% of colorectal cancers are adenocarcinomas, which arise from the glandular tissue (6). It can grow along the lining of the epithelium into the wall of the colon and rectum and invade the digestive system (7). In addition, the cancerous cells can also penetrate into the circulating systems, the blood and lymphatic systems which known as metastasis (7). Typically, the cancerous cells will first spread into the nearby lymph nodes and subsequently penetrate into other organs such as liver, lungs and ovary through blood vessels (8, 9). Colorectal cancer can be classified as tumors/nodes/metastasis (TMN) staging and Dukes classification (12). The TMN assigns the number based on three categories, T, M and N, which are the degree of invasion of the intestinal wall, lymph node involvement and the degree of metastasis, respectively (10). The higher number of TNM system indicates the advanced stage of colorectal cancer (10).

Unhealthy lifestyles such as alcohol consumption, high intake of red meat, obesity, smoking and lack of physical activities are among the risk factors for CRC (1, 11). Age and gender also play significant role in the development of CRC as the risk is higher in male and elderly(7). People with inflammatory bowel disease such as ulcerative colitis and Crohn's disease are also at high risk of getting CRC (12). Among the patients with Crohn's disease, approximately, 2%, 8% and 18% of the patients will develop CRC after 10, 20 and 30 years, respectively (12). About 20% of patients with ulcerative colitis develop CRC within the first 10 years (13). Mutations in genes such as *KRAS, APC,* and *MMR* are the well-documented genetic factor that contributes to colorectal cancer (3, 14, 15). Individual with family history of CRC in two or more first degree relatives have 2 or 3-fold greater risk of getting CRC and this has accounted for 20% of all cases (7). Examples of CRC involving genetic mutations are hereditary nonpolyposis colorectal cancer (HNPCC or Lynch Syndrome), Gardner syndrome and Familial adenomatous polyposis (16).

Diagnosis of CRC is based on tumor biopsy performed during the sigmoidoscopy or colonoscopy (7). CT scan of chest, abdomen and pelvis could be performed to determine the metastasis state and in certain cases, PET or MRI may be used to assist in the diagnosis (7).Molecular testing for patients with a strong family history can be performed to identify mutation, thus initiate early diagnosis and screening in family members. In addition, molecular characterization of mutations involved in CRC may help doctors to plan a better treatment strategy for the patients. Managing our lifestyles can help us to reduce our risk of getting CRC, for example by improving lifestyle through regular exercise, increasing the consumption of whole grains, fruits and vegetables and reducing the red meat intake (17). The treatments for CRC include surgery, chemotherapy and radiotherapy.

2.1. Molecular biology of colorectal cancer

Colorectal cancer is a multistep process that includes accumulation of several genetic and epigenetic alterations (18, 19). It is well characterized that the adenoma to carcinoma sequence is due to accumulation of the genomic alteration, which is induced by genomic instability (4, 20). Genomic instability is an event, which will increase tendency of the genome to acquire mutations when several important processes in maintaining and replicating the genome are malfunction. It is a hallmark of many human cancers (20). There are three well-reported genomic instability pathways that could lead to colorectal cancer, which will be discussed in details below.

a. **Chromosomal instability (CIN)**

Chromosomal instability lead to increase rate of losing or gaining chromosomes during cell division and accounts for 15% to 20% of sporadic CRC as well as Lynch Syndrome (Hereditary Non-Polyposis Colorectal Cancer) (21).There are three mechanisms involved in this process that includes structural chromosome instability, the chromosome breakage-fusion-bridge (BFB) cycles and numerical instability (22). Structural chromosome instability is caused by high incidences of DNA double-strand breaks, which may lead to abnormalities in chromosomal segregation during mitosis. Chromosomal damage may result in mitotically unstable chromosome, which may promote an event known as breakage-fusion-bridge (BFB) (22). An abnormal number of centrosome may be caused by abnormal mitotic polarity as well as unequal segregation of chromosomes during the anaphase stage (23). CIN promotes cancer progression by increasing clonal diversity (21). In the clinical perspective, large meta-analysis has shown that CIN is a marker of poor prognosis in colorectal cancer (20).

b. **Microsatelite instability (MIN)**

Microsatellites are repetitive sequences of DNA, which is highly varied between individuals (24). The most common microsatellites in human is a dinucleotide repeat of CA (25). MIN is a condition, which is manifested by damaged DNA due to defective in the DNA repair mechanism. CRC with the presence of MIN have a better prognosis compared to CRC with CIN (26). MIN involves the inactivation of the DNA Mismatch

Repair (MMR) genes via aberrant methylation or somatic mutation (26). HNPCC or Lynch Syndrome is an example of CRC, which is caused by MIN with 15% occurrence (27). MIN could cause CRC in 2 mechanisms; 1) mutations in the MMR genes where error in the microsatellite repeat replication is unfixed. This leads to the inactivation of tumor suppressor genes (TSG), a group of genes which is crucial in maintaining cell cycle progression and apoptosis induction (20). Inactivation of these genes may lead to tumorigenesis through uncontrolled cell division 2) epigenetic changes that silence the MMR genes (20).

c.   **CpG Island Methylation and CpG Island Methylator Phenotype (CIMP)**

Hypermethylation of the promoter region of a gene that contains CpG Island (CGI) and global DNA hypomethylation are associated with epigenetic instability in colorectal cancer (20). CGIs are short sequences rich in the CpG dinucleotides and are observed in the 5' region of almost half of all human genes (28). *In-vitro* study of BRAF in CRC cell lines showed no correlation between BRAF and CIMP (29).

2.2. Genome Wide Association Study (GWAS) in colorectal cancer

The completion of Human Genome Project in 2003 and the International HapMap Project in 2005 have opened up a new era in genetic and phenotype correlation study (30). The completion of these two projects has made the Genome wide association study (GWAS) possible. GWAS is considered as the most powerful tool to study the association between phenotypes and genotypes and also to identify common, low-penetrance susceptibility loci in a particular disease. In addition, GWAS can also be employed to investigate gene-environment interactions and the pooled analyses may also lead to the identification of novel modifying genes. Several GWAS studies have been performed in colorectal cancer and several loci were identified to be associated with CRC such as 8q24 (128.1-128.7 Mb, rs6983267) (31, 32). The *C-MYC (MYC)* oncogene is located approximately 300 kb from this region and is often over-expressed in CRC (33). Validation studies have confirmed that rs6983267 loci as the most promising variant in CRC, which has increased the chance of getting CRC by approximately 1.2 fold (33, 34). Recent publication has suggested that this variant is involved in enhancing the Wnt signaling and MYC regulation, which are known pathways in carcinogenesis (35). However, further functional analyses are still needed in order to determine the function of this variant. In the Japanese population, this variant leads to an increase risk of CRC with an allelic OR=1.22. Even after the adjustment for confounders, the OR remains significant (OR = 1.25). In the ARCTIC report, a locus at 9p24 was identified to be associated with CRC and was confirmed in the Colorectal Cancer Family Registry. Several numbers of loci that include 18q21:*SMAD7*; 15q13.3:*CRAC1*; 8q23.3: *E1F3H*; 14q22.2:*BMP4*; 16q22.1: *CDH1* and 19q13.1:*RHPN2* were also found to be associated with CRC. These genes have been shown to be involved in CRC progression. Studies conducted in Korean and Japanese patients with CRC have identified a novel susceptible locus in *SLC22A3*, which was significantly associated with distal colon cancer (36). The variant, rs7758229, was located on 6q26-q27 with OR=1.28. Three variants, rs7758229,

rs6983267 and rs4939827, in *SMAD7* together with alcohol consumption may increase the risk of CRC by approximately two-fold. Several variants including rs6983267, rs6695584, rs11986063, rs3087967, rs2059254 and rs72268855 showed evidence of association with CRC in Singaporean Chinese (31). sSNP rs3087967 at 11q23.1 was associated with increased risk of CRC in men (OR=1.34) compared to women (OR=1.07). The rs 10318 at locus 15q13 (GREM1) was also associated with CRC with OD =1.19 (37).

Almost half of the susceptibility loci in CRC are located nearby the transforming growth factor beta gene (*TGF-β1*), which is important in the carcinogenesis (38). An elevated level of *TGF-β1* was linked to tumor progression and recurrence in CRC. Germline mutations in components of *TGF-β1* signaling pathway such as SMAD4 is responsible for the high-penetrance juvenile polyposis syndrome. Other genes are *SMAD4, RHPN2, BMP4, BMP2* and *GREM1*.

2.3. Gene expression profiling in colorectal cancer

Gene expression profiling was performed to compare between colorectal adenomas and CRCs and the result showed that the level of six cancer-related gene sets were increased in CRCs compared to adenomas (FDR<0.05). These include genes that involved in chromosomal instability, proliferation, differentiation, angiogenesis, stroma activation and invasion. Changes in the activity of the chromosomal instability were the most significant gene set (FDR=0.004) (39). The key genes that are associated with colorectal adenoma to carcinoma progression are AURKA, TPX2 (Chromosomal instability), PLK1 (Proliferation), ADRM1 (Differentiation), SSCA1 (Stroma activation), SPARC and PDGFRB (Invasion). The expression levels of these genes were significantly higher in CRC compared to adenoma (p<1e-5). Overexpression of AURKA induces centrosome amplification, aneupploidy and cellular transformation *in vitro* (40). AURKA interacts with TPX2 and plays a role in centrosome maturation and spindle formation (41). The polo-like kinase 1 (PLK1) is important in spindle formation and cell cycle progression during the G2 and M phase (42).

Wu and colleagues showed that the extracellular matrix and metabolic pathways were activated and the genes related to cell homeostatsis were downregulated. In this study, they compared cancer transcriptome using massive parallel paired-end cDNA sequencing in 3 different tissues, CRC tissue (stage III), adjacent non-tumor tissue and normal tissue from a 57 years old female patient. They detected 1660, 1528 and 941 significant differential genes (DEGs) between the CRC and adjacent tissue, the CRC and normal tissue; and the adjacent and normal tissue respectively. 15-prostaglandin dehydrogenase (*15-PGDH*) was downregulated in cancer compared to normal tisssue, which is common oncogenic event in approximately 80% of CRC cases. The transition between adenoma and carcinoma processes involved inactivation of *TGFBR2*, thus progressive inactivation of this gene from cancer-adjacent and normal tissue was expected. In addition, *APC, MYH, CD133, IDH1* and *MINT2* were also dysregulated in CRC. They also identified many genes involved in extracellular matrix (ECM) receptor interactions were highly dysregulated in cancer. The findings showed that all collagen type proteins were overexpressed up to 1000-fold in cancer tissue.

In addition, members of MMP family, which degraded the ECM structures, were also induced significantly in tumor. These include MMP1, MMP3, MMP14 and MMP7. Other cell-cell adhesion-related molecules for examples laminins (LAMA4, LAMA5, LAMB1, LAMB2 and LAMC2) and integrins (ITGA5, ITGB5, ITGA11 and ITGBL1) were elevated in cancer tissues. It was suggested that "angiogenesis switch" was activated in tumor tissues since vascular endothelial growth factor (VEGF) was found to be upregulated. In conclusion, up-regulation of the ECM pathway and the angiogenic growth factors may lead to remodelling of the ECM pathways as well as expansion of the new vessel networks, which subsequently resulted in CRC progression. Since their results in concordance with previous studies that showed the ECM pathway was subjected to intensive epigenetic modification, therefore this ECM may be a good candidate as prognostic biomarkers in CRC (43).

3. Molecular signatures in ovarian cancer

Ovarian cancer is among the top ten leading cancers among women the United States. In this country alone, there are approximately 22,280 new cases and 15,500 estimated death in 2012 (44). At our local population, approximately 1627 women were diagnosed in 2003 to 2005 and the figure showed increasing trend in 2007(45).In Japan and Sweeden, the incidence of ovarian cancer per 100,000 women is 3.1 cases and 21 cases respectively (Green et al., 2012). Due to vague or absence of early signs and symptoms, patients suffer from this cancer seek late treatment (46). Therefore, the cancer is normally diagnosed late when the disease is not longer confined to the ovary. Based on different morphological characterisitcs of the cancer, it is divided into epithelial and nonepithelial types. The epithelial type is further subdivided into serous, mucinous, endometrioid and clear cells. On the other hand, the nonepithelial is granulosa cells, mixed germ cells tumour, immature teratoma, dysgerminoma and teratoma. The risk factor for this cancer is unclear, however the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study has recently documented that women who smoke more than 10 cigarettes a day had doubled the risk to develop mucinous ovarian cancer (47). This has suggested that the effect of smoking differs based on different histological subtypes of ovarian cancer(47). On the other hand, a study has shown that long period of breastfeeding seems to have reduced risk of ovarian cancer (OR = 0.986, 95% CI 0.978-0.994 per month of breastfeeding) (48).This effect of breastfeeding was also varies between histological subtypes as there was no association between breastfeeding and borderline serous or mucinous cancer (48).

Ovarian cancer was initially divided based on molecular pathways involved in the development and progression of the subtypes (49). Type I is low-grade serous, low-grade endometrioid, mucinous and clear cells. They are believed to arise from benign lesions such as ovarian inclusion cyst or endometriotic lesions. These lesions follow the stepwise pattern, whereby it evolved from the benign adenoma to borderline and finally to malignant tumours (table 1).

Type II ovarian cancer is high-grade serous, high-grade endometrioid and undifferentiated. The common mutations that are found in these subtypes are p53, BRCA1/2, PIK3CA with chromosomal instability. They normally involve the peritoneum and grow rapidly.