



2016 Colorectal Cancer: Global view

Integration of genome scale data for identifying new players in colorectal cancer

Viktorija Sokolova, Elisabetta Crippa, Manuela Gariboldi

Viktorija Sokolova, Elisabetta Crippa, Manuela Gariboldi, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milano, Italy

Viktorija Sokolova, Elisabetta Crippa, Manuela Gariboldi, Molecular Genetics of Cancer, Fondazione Istituto FIRC di Oncologia Molecolare, 20139 Milano, Italy

Author contributions: Sokolova V and Gariboldi M wrote the manuscript; Crippa E contributed to writing the manuscript and collected the bibliography.

Supported by Associazione Italiana per la Ricerca sul Cancro, Grants No. 10529 and No. 12162; and by funds obtained through an Italian law that allows taxpayers to allocate 0.5% share of their income tax contribution to a research institution of their choice.

Conflict-of-interest statement: No potential conflicts of interest.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Correspondence to: Manuela Gariboldi, PhD, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, via G. Venezian 1, 20133 Milan, Italy. manuela.gariboldi@istitutotumori.mi.it
Telephone: +39-2-23902042
Fax: +39-2-23903141

Received: June 22, 2015

Peer-review started: June 26, 2015

First decision: September 11, 2015

Revised: October 13, 2015

Accepted: November 9, 2015

Article in press: November 9, 2015

Published online: January 14, 2016

Abstract

Colorectal cancers (CRCs) display a wide variety of genomic aberrations that may be either causally linked to their development and progression, or might serve as biomarkers for their presence. Recent advances in rapid high-throughput genetic and genomic analysis have helped to identify a plethora of alterations that can potentially serve as new cancer biomarkers, and thus help to improve CRC diagnosis, prognosis, and treatment. Each distinct data type (copy number variations, gene and microRNAs expression, CpG island methylation) provides an investigator with a different, partially independent, and complementary view of the entire genome. However, elucidation of gene function will require more information than can be provided by analyzing a single type of data. The integration of knowledge obtained from different sources is becoming increasingly essential for obtaining an interdisciplinary view of large amounts of information, and also for cross-validating experimental results. The integration of numerous types of genetic and genomic data derived from public sources, and *via* the use of *ad-hoc* bioinformatics tools and statistical methods facilitates the discovery and validation of novel, informative biomarkers. This combinatory approach will also enable researchers to more accurately and comprehensively understand the associations between different biologic pathways, mechanisms, and phenomena, and gain new insights into the etiology of CRC.

Key words: Colorectal cancer; Copy number variations; Gene expression; miRNA expression; Methylome; Data integration

© The Author(s) 2016. Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: The development of colorectal cancer (CRC) is driven by the accumulation of various genetic and epigenetic alterations, which have been only partially identified. The increasing financial affordability of high-throughput genome-wide assays has enabled the comprehensive analysis of genomic, transcriptomic, and epigenetic data obtained by analyzing the same biologic samples, and thereby facilitated the identification of new molecular players in CRC. An integrative approach that considers all of these multiple factors provides for better results when seeking to identify genes or microRNAs related to new interactions or biomarkers that might improve CRC diagnosis, prognosis, and treatment.

Sokolova V, Crippa E, Gariboldi M. Integration of genome scale data for identifying new players in colorectal cancer. *World J Gastroenterol* 2016; 22(2): 534-545 Available from: URL: <http://www.wjgnet.com/1007-9327/full/v22/i2/534.htm> DOI: <http://dx.doi.org/10.3748/wjg.v22.i2.534>

INTRODUCTION

Due to its incidence of one million new cases and mortality rate of > 500000 deaths per year, colorectal cancer (CRC) is now the third most common type of cancer, and the third leading cause of cancer-related death worldwide^[1]. CRC can be classified on the basis of its clinical, pathologic, and genetic characteristics, and is commonly described as a progressive malignant transformation of the normal colonic epithelium to invasive adenocarcinoma due to an accumulation of acquired genetic and epigenetic aberrations^[2].

At least three different pathogenetic mechanisms have been proposed to explain the development of CRC. Chromosomal instability (CIN) is thought to account for 85% of all cases, while microsatellite instability (MSI) and the presence of a CpG island methylator phenotype (CIMP) may account for a majority of the remaining cases. The CIN pathway involves the sequential deregulation of tumor suppressors and oncogenes, and includes mutation of the *APC* gene and/or the loss of chromosome 5q where it maps. CIN can also refer to mutation of the *KRAS* oncogene, the loss of chromosomal arm 18q, and deletion of chromosome 17p, which harbors the tumor suppressor gene *TP53*^[3].

MSI results from loss of the DNA mismatch repair system. This loss destabilizes repetitive units of DNA (DNA microsatellites), resulting in the generation of inactivating frameshift mutations in the coding sequences of tumor suppressor genes. Tumors that display MSI are divided into two subtypes: MSI-H (instability in > 30% of microsatellites examined) and MSS/MSI-L (instability in < 30% of microsatellites examined)^[4]. The CIMP pathway is characterized

by hypermethylation of DNA located in CpG islands (specific regulatory sites enriched in CpG motifs) found in the promoter regions of tumor suppressor genes. Such hypermethylation suppresses gene transcription^[5]. CRCs with a CIMP display unique epigenetic phenotypes, and well-defined clinical, pathologic, and molecular profiles^[6].

Recent studies of these distinct pathways have highlighted the heterogeneity of CRC, and their results have suggested that all CRCs cannot be fully explained by the initial model proposed by Fearon *et al.*^[2]. The results of recent large-scale sequencing studies have shown that CRC development requires the participation of numerous critical "driver" genes and "passenger" genetic alterations, many of which remain to be identified^[7-9]. Identification of these additional genetic alterations should allow physicians to better characterize and identify the various clinical stages of CRC, utilize such aberrations as biomarkers for the early diagnosis and prognosis of CRC, or to develop treatments for CRC patients.

In recent years, thousands of non-coding RNAs [*e.g.*, microRNAs (miRNAs), long non-coding RNAs, and competitive endogenous RNAs] have been identified as key regulators of various cellular processes that control tumor initiation and progression^[10,11]. As the most studied type of non-coding RNA, miRNAs are thought to regulate the post-transcriptional expression of tumor suppressor genes and oncogenes involved in CRC development. Moreover, due to their high chemical stability, miRNAs represent potential biomarkers for use in diagnosing and monitoring human cancers^[12]. Additionally, certain miRNAs may be associated with a patient's response to treatment^[13].

The various genomic alterations described above can all contribute to helping investigators understand the vast landscape of CRC. Furthermore, the ability to simultaneously measure and integrate the effects of such alterations should assist in identifying previously unknown molecular changes in key genomic factors, and contribute to a better understanding of how such changes interact with each other to induce the development of CRC. A combination of these data might also be used to help predict disease risk or patient outcomes. The continuous improvement of high-throughput screening and sequencing technologies has made it possible to use the same biologic samples to gather data for several different projects. For instance, The Cancer Genome Atlas (TCGA)^[14] was compiled with the goal of first profiling and then integrating the genomic changes that occur in cancers, including CRC^[15]. Although several limitations can hamper such an analysis (*e.g.*, patterns of missing data and noise across different data types), numerous bioinformatics tools and statistical frameworks are now available for integrating multiple genomic features found in the same sample, and then assist in investigating their related biologic pathways or gene sets^[16].

Here, we reviewed the most relevant publications that have integrated various types of molecular data related to CRC, with the goal of identifying new biomarkers for CRC detection and progression, or targets that may assist in dissecting the mechanisms involved in CRC development.

INTEGRATION OF COPY NUMBER VARIATION AND GENE EXPRESSION DATA

Several studies have shown that a gain or loss in the number of copies of DNA segments affects the expression of genes as well as miRNAs positioned within it. Moreover, such changes are known to affect cancer-related biologic processes^[17]. Although not all genes with increased copy number are overexpressed^[18-20], those that display a strong positive correlation between expression and copy number may play important roles in cancer progression. However, copy number variations (CNVs) alone cannot fully explain the altered expression levels of all genes, because changes in gene expression are also determined by complex mechanisms that regulate gene transcription. Nevertheless, integrating the results of gene expression analyses with genomic profiling results in an efficient approach for discovering novel cancer-related genes.

Numerous studies on gene CNVs in samples of CRC tissue have yielded similar results regarding the focal genomic regions of gains and losses, and confirmed the high prevalence of gains on chromosomes 8q, 13, and 20q, and losses on chromosomes 8p, 17p, and 18q^[15,18,21-23].

The most recent published studies that integrated data concerning gene expression and CNV are shown in Table 1. Ali Hassan *et al.*^[24] found that increased gene expression was correlated with significantly increased copy numbers of genes that were mostly located on chromosome 20q12, where eight highly expressed genes were identified. Moreover, three of those genes (*TOP1*, *PLCG1*, and *PTPRT*) were related to CRC. The highest number of copy losses was observed on chromosome 8p23.2 and was correlated with reduced expression of *CSMD1* and *DLC1*. A mapping on KEGG pathways for genes showing an association between changes in their copy number and expression level highlighted their involvement in processes related to the cell cycle.

Loo *et al.*^[25] reported a correlation between gene expression and CNVs in 23% (356/1573) of the differentially expressed genes they analyzed. The most significant correlation between genomic alterations and changes in gene expression was found on chromosome 20q (20q11-13), where several overexpressed genes (*AURKA*, *AHCY*, *POFUT1*, *RPN2*, *TH1L*, and *PRPF6*) were amplified. However, an opposite pattern was detected on 8p, where the tumor suppressor genes

MTUS1 and *PPP2CB* are located. Moreover, several of the identified genes have known involvement in the Wnt signaling pathway, which plays a role in CRC progression.

Yoshida *et al.*^[26] integrated the results of CNV and gene expression studies with the clinical characteristics of CRC cases they had analyzed. Their comparison of tumor and normal tissue samples revealed copy number gains on chromosomes 7, 8q, 13, and 20q, and losses on chromosomes 8p, 17p, and 18. A further analysis based on tumor stage revealed that *UGT2B28* was downregulated and lost on chromosome 4q13.2 during the early stages of CRC (T₁₋₂), while *LOC440995* (3q29), *CXCL6*, and *CXCL3* (both on chromosome 4q21), and *SULT1B1* (4q13.3) distinguished T₁₋₂₋₃ cases from T₄ cases. Furthermore, copy numbers of *RALBP1*, *TYMS*, *RAB12*, and *RNMT*, (all mapping to chromosome 18p11) were higher in lymph node-negative cases, *ARHGDI1B* (12p12) was absent in metastatic cases, while *S100A2* (1q21), *ABHD2* (15q26), *OIT3* (10q22), and *ABHD12* (20p11) showed different copy numbers in cases of recurrent vs non-recurrent disease. Many of the genes identified were known to be associated with CRC, suggesting the validity of using data integration as a strategy for identifying new biomarkers for CRC or targets for its treatment.

Kikuchi *et al.*^[27] used a combination of copy number and mRNA expression data to demonstrate the clinical relevance of cancer-related gene protein expression. Those investigators searched for potential therapeutic targets or clinical biomarkers for advanced CRC, and then integrated their results with CNV and gene expression data obtained from a subgroup of patients with distant metastases. They found that 51 genes had both an elevated copy number and expression level. Among the three most highly expressed genes (*NUCKS1*, *SYNPR*, and *TMEM63A*), *NUCKS1* (1q.32) is known to be overexpressed in several cancer types^[28]. The investigators found that *NUCKS1* protein levels were upregulated in patients with distant metastasis, and associated with an invasive and metastatic tumor phenotype. Such findings suggest *NUCKS1* as a potential biomarker for predicting CRC recurrence following colorectal surgery, and a novel target for CRC treatment.

Reid *et al.*^[18] identified 412 genes whose expression was correlated with CNVs, among which 80% and 20% mapped in gained and lost regions of chromosomes, respectively. Chromosomal arms 20q and 13q contained the highest numbers of genes whose expression correlated with copy number (182 and 118, respectively). Newly identified and possible CRC-related genes found in that study were *PLCG1* on 20q, *DBC1* on 8q21, and *NDGR1* on 8p24.

Those investigators also analyzed combined data regarding the correlation between CNVs and gene expression, mutations of *APC*, *KRAS*, and *TP53*, 18q

Table 1 Relevant studies that integrated copy number variations and gene expression data when studying colorectal cancer

Ref.	Case series		Chr 20q (gain)	Chr 8p (loss)	Additional genes/chromosomes
	CNV	mRNA			
Ali Hassan <i>et al</i> ^[24]	64 couples (4 A, 33 B, 27 C ¹)	15/64 couples	<i>20q12</i> <i>PTPRT</i> ² <i>CHD6</i> <i>EMILIN3</i> <i>LPIN3</i> <i>PLCG1</i> ² <i>TOP1</i> ² <i>ZHX3</i> <i>MAFB</i>	8p23.2 <i>CSMD1</i> <i>DLC1</i>	Gains: 8q21-22, 8q24, 13q21, 13q34, 20p11-13, 20q11-13
Loo <i>et al</i> ^[25]	40 couples	40 couples	<i>20q11-13</i> <i>AHCY</i> <i>POFUT1 RPN2</i> <i>THIL</i> <i>PRPF6</i> <i>AURKA</i>	8p <i>FBXO25</i> <i>MTUST2</i> <i>EPHX2</i> <i>KIF13B</i> <i>PPP2CB</i> ²	Gains: 20pter-p12 (<i>PCNA</i>), 20p13 (<i>CDC25B</i>), 13q34 (<i>CUL4A</i>)
Yoshida <i>et al</i> ^[26]	70 couples	70 couples			Gains: 1q21 (<i>S100A2</i>), 18p11 (<i>RALBP1</i> , <i>TYMS</i> , <i>RAB12</i> , <i>RNMT</i>), 15q26 (<i>ABHD2</i>), 20p11 (<i>ABHD12</i>) Losses: 3q29 (<i>LOC440995</i>), 4q13 (<i>UGT2B28</i> , <i>SULT1B1</i>), 4q21 (<i>CXCL6</i> , <i>CXCL3</i>), 10q22 (<i>OIT3</i>), 12p12 (<i>ARHGDI3</i>)
Kikuchi <i>et al</i> ^[27]	122 couples (18 I, 42 II, 37 III, 25 IV ³)	115/122 couples (16 I, 41 II, 35 III, 23 IV ³)			Gains: 1q32.1 (<i>NUCKS1</i>), 1q42 (<i>TMEM63A</i>), 3p14.2 (<i>SYNPR</i>)
Reid <i>et al</i> ^[18]	48 couples (1 I, 10 II, 10 III, 23 IV ³ , 4 N.I.)	36/48 couples (1 I, 5 II, 8 III, 19 IV ³ , 4 N.I.)	<i>20q11-13</i> <i>PLCG1</i> <i>ADRM1</i> <i>JAG1</i> <i>AURKA</i> <i>C20ORF20</i> <i>C20ORF24 TCFL5</i> <i>THIL</i> <i>AHCY</i> <i>TGIF2</i>	8p21-24 <i>BIN</i> <i>DBC1</i> <i>TNFRSF10A</i> <i>TNFRSF10B</i> <i>EXTL3</i>	Gains: 7p12 (<i>EGFR</i>), 8q24 (<i>NDRG1</i>), 13q (<i>TFDP1</i> , <i>CDK8</i> , <i>GAS6</i> , <i>SPATA13</i>) Losses: 6p21, 8p, 18q (<i>MAPRE2</i> , <i>INO80C</i> , <i>ARKL1</i>), 20p (<i>FKBP12</i>)

¹Duke's; ²CRC-related; ³Stage. Chr: Chromosome; CNV: Copy number variations; CRC: Colorectal cancer; Couple: Tumor tissue matched with its normal colorectal mucosa. Genes identified in more than one study are shown in bold font.

loss of heterozygosity, and patient survival. Their results showed that chromosomal losses were frequent in wild-type *TP53* patients, while *TP53*-mutant patients showed pronounced gains on chromosome 20q. Chromosomal alterations were rarely present in *TP53*-mutant cases without a 20q gain. These findings suggest that CRC can develop *via* two alternative routes: one mainly involving CIN, and the other involving the combined effects of having mutant *TP53* plus a 20q gain. The simultaneous presence of *TP53* mutations and a 20q gain might be sufficient to deregulate specific molecular pathways responsible for CRC progression. Finally, 34 genes located on chromosomes 7p, 8p, 13q, 18q, and 20 were found to be associated with overall survival.

Results from the above-mentioned studies support the hypothesis that genomic aberrations that result in CNVs lead to the deregulation of normal gene expression and directly affect critical cellular functions related to CRC tumorigenesis. Therefore, integrative approaches that utilize both gene CNV and expression data may enable the identification of markers for early detection of cancer and favor the development of new molecular agents for chemoprevention and

chemotherapy.

INTEGRATED ANALYSIS OF miRNA AND mRNA EXPRESSION DATA

Numerous miRNA microarray profiling studies^[29-31] and next-generation sequencing^[32-34] analyses have identified miRNAs that are differentially expressed in samples of CRC tissue and adjacent non-cancerous tissue. Specialized bioinformatic tools that predict the complementarity between a miRNA seed sequence and the 3' untranslated region of its target mRNA^[35-37] can be used to combine these data with those obtained from mRNA expression profiling studies, and thus identify miRNA/mRNA pairs with opposite expression patterns. The predicted interaction between a miRNA binding site and the 3' untranslated region of the complementary gene is usually functionally verified by using appropriate luciferase activity reporter vectors in an *in vitro* model^[38]. However, such experiments can only provide information concerning possible miRNA/mRNA interactions; furthermore, the results of functional validation experiments do not apply to

Table 2 Studies that integrated gene and miRNA expression data when studying colorectal cancer

Ref.	Case series		Modulated genes	Modulated miRNA	miRNA/mRNA pairs with opposite expression	
	mRNA	miRNA			Predicted	Confirmed
Fu <i>et al</i> ^[39]	8 couples (1 I, 4 II, 3 III ¹)	8 couples (1 I, 4 II, 3 III ¹)	2916 N vs T	32 N vs T	72 miRNA/mRNA	mir-29a/KLF4, miR-224/SFRP2
Vishnubalaji <i>et al</i> ^[40]	13 couples (5 II, 8 III ¹)	13 couples (5 II, 8 III ¹)	3175 N vs T	103 N vs T	794 miRNA (downregulated)/ mRNA	miR-26a-5p/EZH2 let-7b-5p/EZH2
Pizzini <i>et al</i> ^[41]	80 samples (23 NOR, 30 CRC, 27 liver metastasis)	78 samples (23 NOR, 31 CRC, 24 liver metastasis)	12748 N vs T	N vs T	3078 miRNA/ mRNA	miR182/ENTPD5, miR-145/c-Myc
Lanza <i>et al</i> ^[42]	39 couples (23 MSS, 16 MSI-H)	39 couples (23 MSS, 16 MSI-H)	72 MSI-H vs MSS colon cancers	14 MSI-H vs MSS colon cancers		Predictor composed of 27 elements (genes and miRNAs) for distinguishing MSI-H vs MSS
Gattoliat <i>et al</i> ^[44]	9 NOR, 37 CRA, 9 CRC	5 NOR, 28 CRA, 15 CRC		1 CRC vs NOR, 2 CRA vs NOR, 5 CRA and CRC vs NOR		miR-21/PDCD4, miR-21/ MARCKS, miR-200b/ZEB2, miR-15b/BCL2, miR-16/ BCL2, miR-21/BCL2 miR-1/MET
Reid <i>et al</i> ^[45]	7 public gene expression datasets CRC vs N samples	40 couples (10 A, 10 B, 10 C, 10 D)	7629 N vs T	70 N vs T		
Ling <i>et al</i> ^[47]	SW480 cells after miR-224 overexpression (84 metastasis related genes analyzed)	4 CRC with metastasis, 8 CRC without metastasis, SW480, SW620	13 metastasis- related genes down- regulated	4 up in primary metastatic CRCs vs early CRC stages	miR-224/SMAD4 miR-224/CDH1	miR-224/SMAD4

¹Stage. Couple: Tumor tissue matched with its normal colorectal mucosa; CRA: Colorectal adenoma; CRC: Colorectal cancer; NOR: Normal colorectal mucosa.

clinical samples. The regulatory miRNA/mRNA pairs vary in different diseases, and their expression profiles can vary with the stage of a disease. Currently, there is no efficient method for stratifying miRNA/mRNA interactions that reflects the clinical characteristics of a tissue specimen. However, the integration of molecular and bioinformatic tools represents a promising approach for fully understanding the miRNA regulatory mechanisms that underlie CRC development.

As summarized in Table 2, Fu *et al*^[39] and Vishnubalaji *et al*^[40] applied genome-wide mRNA and miRNA microarray expression profiling techniques to the same samples for purposes of identifying CRC-specific miRNA/gene pairs with potential diagnostic, prognostic, or therapeutic roles. Pizzini *et al*^[41] performed similar studies with samples of metastatic tissue to investigate how miRNA/mRNA changes affected CRC progression. Finally, Lanza *et al*^[42] focused on base-pair differences between specific CRC subgroups (MSI and MSS).

Fu *et al*^[39] identified 72 predicted miRNA/mRNA pairs and found that a large number of genes were participants in the Wnt signaling pathway, which is crucial for the initiation of CRC development. The most relevant pairs were validated in a study conducted using 40 additional, matched CRC tissue samples. The highest negative correlation was found between miR-224/SFRP2 and miR-29a/KLF4, which provided new information useful for elucidating CRC tumorigenesis. Moreover, Wang *et al*^[43] analyzed those data using a multiple linear regression model and identified three

additional pairs: miR-16/BCL2, miR-567/SMAD4, and miR-142-5p/MSH6.

In an attempt to find putative tumor suppressor miRNAs, Vishnubalaji *et al*^[40] identified 794 pairs that showed opposite expression patterns between normal and tumor samples. Those studies focused on the interactions between *EZH2*, a gene frequently overexpressed in cancer, and its negatively correlated miRNAs (miR-26a-5p and let-7b-5p). Pharmacologic inhibition of *EZH2* in CRC cell lines was found to markedly reduce both cell proliferation and migration, while *in vitro* silencing of *EZH2* and overexpression of both miR-26a-5p and let-7b-5p decreased cell viability. Those study results suggest that miR-26a-5p and let-7b-5p play prominent roles in regulating *EZH2* expression in CRC.

Pizzini *et al*^[41] investigated changes in miRNA and mRNA expression in samples of normal colonic mucosa, primary CRC tissue, and liver tumor metastases, and found that 95% of miRNAs and 93% of genes that were deregulated in CRC samples when compared to samples of matched, normal tissue remained invariant after metastasis had occurred. Only five miRNAs (miR-146a, miR-15a, miR-15b, miR-196a, and miR-708) were deregulated during the "tumor-to-metastasis" transition period. The data regarding miRNAs and genes with opposite expression patterns were integrated to define putative post-transcriptional regulatory networks. The tumor vs normal network comparison included two components of six upregulated and 17 downregulated miRNAs, together with

their putative target genes. In comparison, a network constructed using the five miRNAs identified in the metastasis vs tumor comparison plus their target genes was smaller, and contained five unrelated components. The results of that study suggest that significant changes in the transcriptome mostly occur during the early stages of CRC progression. Moreover, opposite expression patterns for miR-145 and its target gene *MYC* were confirmed, and a proposed interaction between miR-182 and *ENTPD5*, a gene involved in energy metabolism, was functionally validated. Finally, an early survival analysis indicated that miR-10b expression (modulated between tumor and metastatic tissues) was inversely correlated with survival in stage IV CRC patients.

Lanza *et al.*^[42] added miRNA and mRNA profiling data to molecular signature data that could distinguish between MSI-H and MSS CRCs. Those investigators then used prediction algorithms that combined information regarding the 14 differentially expressed miRNAs and 72 deregulated genes identified in MSI-H and MSS-type tumors to construct a combinative predictor containing 27 elements (miRNAs and genes) that was more sensitive than the predictor lists that contained only one type of element. The addition of miRNAs to the molecular predictor improved their categorization and represented a novel approach for elucidating the respective molecular features of the two CRC subgroups. The combinative predictor includes various members of the miR-17-92 family, which is a class of miRNAs with proven oncogenic characteristics that are probably involved in the molecular mechanisms that distinguish MSS from MSI colon cancers.

Several studies have combined global mRNA and miRNA microarray expression data derived from two independent cohorts of patients, and in some cases also integrated publically available datasets (Table 2). Gattolliat *et al.*^[44] combined miRNA and gene expression results obtained from two independent sets of normal mucosa (NOR), colorectal adenoma (CRA), and CRC tissue samples. They found different levels of miR-320b expression in CRC tissue when compared to NOR tissue, as well as differences in miR-15b and miR-16 expression in CRA tissue vs NOR tissue. Their data also showed that miR-21, miR-24, miR-145, miR-150, and miR-378 were deregulated in samples of both CRA and CRC tissue when compared to samples of NOR tissue. When the expression of these miRNAs were compared with expression of the genes predicted as their putative targets, 30 pairs with opposite expression patterns were identified, including *PDCD4* and *MARCKS* for miR-21, *ZEB2* for miR-200b, and *BCL2* for miR-15b, miR-16, and miR-21.

Reid *et al.*^[45] identified 23 miRNAs that were differentially expressed in matched samples of CRC

and normal tissue, and then searched seven public gene expression datasets, plus an independent cohort of patients from the same group, to identify putative target genes of these miRNAs. The selected pairs were mapped on the KEGG pathway, and many were found to be included in CRC-related molecular pathways. One candidate pair (miR-1 and the *MET* oncogene) was functionally validated in studies that used *in vitro* models of CRC. miR-1 overexpression was found to diminish *MET* levels and result in reduced cell proliferation, migration, and invasion, suggesting a prominent role for miR-1 in CRC progression. As miRNA expression can also be affected by genomic alterations in regions in which they are positioned, the miRNA profiles were integrated with data obtained from a whole-genome copy-number analysis^[18]. The results showed that chromosomal regions that frequently gained copy numbers contained upregulated miRNAs, whereas regions which lost copy numbers contained both up- and downregulated miRNAs. The role of miR-20a, which was overexpressed and localized in an amplified region, was functionally investigated in CRC cell lines, where studies showed that it interfered with transforming growth factor- β -induced growth arrest^[46].

Ling *et al.*^[47] conducted whole genome miRNA expression studies using samples of primary CRCs with or without metastasis, and also two cell lines: one derived from a primary CRC lesion (SW480), and the other derived from a metastasis located in a lymph node (SW620). The results were validated in a large international patient cohort, and also by using data obtained from TCGA^[14]. The investigators identified four miRNAs (miR-141, miR-181b, miR-221, and miR-224) that were upregulated in primary CRCs with metastatic dissemination at the early stages. The highest levels of expression were observed in the metastasis-related cell line SW620. Only miR-224 overexpression induced the migration of CRC cells, and expression of metastasis-related genes was analyzed following their insertion into SW480 cells. Two target genes (*CDH1* and *SMAD4*) showed reduced expression, and a functional analysis demonstrated that *SMAD4* mediated the miR-224-induced prometastatic effect. Furthermore, elevated miR-224 expression was shown to correlate with survival in CRC patients, indicating a prominent role for miR-224 as a specific diagnostic marker for CRC.

While all of the above-mentioned studies identified several CRC-related miRNAs with potential value as diagnostic or prognostic markers, and/or therapeutic targets, only a few of the identified miRNA/gene target pairs have been functionally analyzed. This limits our understanding of their role in CRC, as the effects of miRNA deregulation are complex and impact the modulation of entire pathways rather than just a single gene.

Table 3 Relevant studies that integrated methylation and gene expression profiles

Ref.	Case series		Genes
	Methylation	mRNA	
Hinoue <i>et al</i> ^[56]	29 NOR, 125 CRC	19 couples	464 genes downregulated CIMP-H <i>vs</i> normal samples, 112 of them (24%) exhibit promoter hypermethylation
Szmida <i>et al</i> ^[57]	12 couples	19 couples (from Hinoue <i>et al</i> ^[56] , 2012)	12 genes downregulated and hypermethylated in non-CIMP tumors 4 ErbB-associated genes (<i>PIK3CD</i> , <i>PKCB</i> , <i>ERBB4</i> , <i>PAK7</i>) differentially methylated in CRC
Wang <i>et al</i> ^[58]	42 NOR, 231 CRC	26 NOR, 231 CRC	118 methylation-perturbed genes whose expression is affected by the highly variable DNA methylation sites

Couple: Tumor tissue matched with its normal colorectal mucosa; CRC: Colorectal cancer; NOR: Normal colorectal mucosa.

INTEGRATION OF METHYLATION AND GENE EXPRESSION PROFILES

Hypermethylation of DNA segments located in promoter CpG islands is the most studied epigenetic alteration involved in the transcriptional repression commonly observed in various cancer types^[6,48]. In CRC, aberrant DNA methylation in CpG islands occurs during the early stages of an oncogenic transformation process, and can be detected in aberrant crypt foci, which are the earliest detectable oncogenic changes in colonic mucosa^[49-51]. The methylation of DNA in several genes, including *APC*, *p16INK4a*, and *TIMP3*, and its significance in CRC have previously been reported^[52,53], and represents a potential biomarker for use in the early detection of CRC^[54]. The development of array and sequencing-based high-throughput assay techniques now permits the profiling of genome-scale DNA methylation (methylomes), and within the same tumor type, has enabled the characterization of different subgroups that display heterogeneous DNA methylation^[55]. Integration of these data with analyses of gene expression can assist in identifying new candidate diagnostic biomarkers that become methylated during the early stages of oncogenesis.

Three recent studies (Table 3) have provided new insights into the role of the CpG island hypermethylation in the regulation of gene expression. The first study integrated gene expression and methylation data obtained from studies that used tissue samples from the same CRC cases, even if only a small number of matched normal and tumor tissue samples were available^[56]. Szmida *et al*^[57] and Wang *et al*^[58] used these data to implement their methylation analyses, with the purpose of validating the results they obtained by integrating methylation and gene expression data retrieved from the TCGA database. All of their studies identified genes whose expression was affected by DNA hypermethylation. Furthermore, two of the studies identified genes involved in the same pathway (ErbB-signaling pathway), and thus highlighted the pivotal role played by DNA methylation in CRC development.

Hinoue *et al*^[56] performed a model-based cluster analysis to identify four distinct methylation-based

subgroups of CRC patients with specific genetic and clinical features. They found that the CIMP was correlated with a high frequency of cancer-specific DNA hypermethylation (CIMP-H), a high incidence of the *BRAF*^{V600E} mutation, and high rates of *KRAS* mutation in a CIMP-low subgroup. Next, the non-CIMP tumors were separated into two subgroups: one with a high frequency of *TP53* mutations in the distal colon, and the other with low incidences of gene mutation and cancer-specific DNA hypermethylation, but significantly enriched for rectal tumors. Those investigators also identified a panel of five genes that could specifically distinguish CIMP from non-CIMP tumors, and also another panel that was highly specific for CIMP-H tumors. Gene expression profiling studies revealed that about 7% of the promoter DNA methylations observed in CIMP-H tumors were linked to the downregulation of 112 genes. Those downregulated genes represented 25% of the genes with lower expression in CIMP-H tumors compared with their expression in adjacent normal tissue. Intriguingly, twelve genes were also downregulated and hypermethylated in non-CIMP tumors. This result is interesting because *SFRP1* and *SFRP2* are negative regulators of Wnt signaling.

Szmida *et al*^[57] integrated their previously published genome-wide methylation data^[59] with results from the gene expression and methylation profiling studies conducted by Hinoue *et al*^[56], and identified four ErbB-associated genes (*PIK3CD*, *PKCB*, *ERBB4*, and *PAK7*) that were differentially methylated in CRC. In particular, *PKCB* hypermethylation was correlated with the presence of a *KRAS* mutation, and hypermethylation of *ERBB4* was linked with highly methylated epigenotypes HME and MSI with the presence of mutated *BRAF*. Methylation appeared to only impact the modulation of *PKCB* expression that was significantly downregulated in CRCs following methylation of its promoter. *PKCB* is a component of the vascular endothelial growth factor signaling pathway and regulates cell proliferation and survival processes that promote tumor angiogenesis. Indeed, therapies that target the vascular endothelial growth factor pathway are currently in clinical studies for treatment of late-stage CRC^[60].

Wang *et al*^[58] combined the gene methylation and expression profiles typical of CRC as retrieved from the

TCGA database^[14] and identified highly variable DNA methylation sites, as well as genes whose expression was affected by a tumor's highly variable DNA methylation status, that were named methylation-perturbed genes (MP). Those results showed that the number of MP genes was significantly lower in samples of CRC tissue as compared to normal tissue. The genes were then clustered based on connectivity between their expression levels and subgrouped according to their MP status. The number of coexpressed partners of MP genes was significantly lower in samples of CRC tissue when compared to samples of normal tissue, and also when compared to the number of non-MP genes in both CRC and normal samples. Interestingly, the lost coexpression partners were often members of cancer pathways, such as the ErbB and mitogen-activated protein kinase signaling pathway. The loss of coexpression connectivity mediated by methylation heterogeneity as described in this study might play an important role in CRC development.

Similar to the observations regarding protein-coding genes, miRNAs can also be silenced by hypermethylation of a CpG island promoter region^[61]. Sometimes this can occur indirectly when miRNAs are transcribed from intronic regions of coding genes that are controlled by hypermethylation^[62]. While no comprehensive analysis and integration of miRNA and methylation data obtained from studies with CRC tissue have yet been performed, several methylation-sensitive miRNAs including miR-9, miR-129, and miR-137^[62], miR-34b/c^[63], miR-200^[64], miR-342^[65], and miR-345^[66] have been identified.

DATA INTEGRATION BY COMBINING PUBLIC DATASETS: THE TCGA NETWORK

The information most valuable for better understanding CRC development and progression can be derived from studies in which tissue samples are characterized for various molecular changes that define the disease. However, most research groups who have combined genetic and genomic data conducted a maximum of two whole genome analyses on the same samples. Furthermore, the experiments were frequently implemented based on data retrieved from publically available datasets containing information regarding CNVs, gene and miRNA expression, and methylation (e.g., GEO^[67] and ArrayExpress^[68]). The TCGA research program^[14] has been of great benefit to investigators seeking to combine multiple data types in hopes of better understanding disease processes. TCGA was launched by the National Institutes of Health in 2006 for the purpose of comprehensively characterizing the genomic and molecular features of cancer by using high-throughput genome, transcriptome, and epigenome analysis techniques. These techniques

include gene and miRNA expression profiling, CNV and genome-wide DNA methylation profiling, single-nucleotide polymorphism genotyping, and exon sequencing performed on thousands of samples obtained from about 20 different tumor types. Numerous robust studies have been conducted using different types of data retrieved from TCGA, and several projects have also been conducted by the TCGA research program itself on the different cancer types analyzed. As for CRC, an analysis of 224 paired tumor and normal tissue samples performed using different platforms (exome sequencing, DNA copy number, promoter methylation, mRNA and microRNA expression, and whole-genome sequencing for 97 samples) identified several specific characteristics of this tumor type^[47]. The CRCs were divided into categories of hypermutated (16%) and non-hypermutated (84%), and 24 genes in each category were found to be significantly mutated. It was interesting that, although colon and rectal cancers have different characteristics, they showed similar patterns of genomic alterations. An analysis of CNVs revealed amplifications of *ERBB2* and *IGF2* and fusion of *NAV2* with *TCF7L1*, which is a member of the Wnt pathway. An integration of CNV, expression, and methylation data revealed that all cancers showed changes in genes known to be involved in *MYC* transcription, suggesting an important role for *MYC* in CRC. Numerous molecular signatures were linked to tumor aggressiveness, and in particular, two chromosomal regions (20q13.12 and 22q12.3) that showed amplifications of genes linked to tumor aggression. To better interpret the effects of genomic abnormalities in terms of cancer biology, genomic data have also been integrated with proteomic data obtained from the same tumors analyzed by TCGA, but generated by the Clinical Proteomic Tumor Analysis Consortium^[69]. This combined proteomic and genomic data can be used to link genotypes with phenotypes and assist in prioritizing genes that merit further examination. Results from examinations of CRC tissue revealed a weak correlation between levels of mRNA and proteins produced by genes in CNV regions, and only the 20q amplification was found to be associated with the largest global changes in both mRNA and protein levels. The identified candidate genes were *HNF4A*, which codes for a transcription factor that plays a key role in normal gastrointestinal development, *TOMM34*, which is involved in the growth of CRC cells, and *SRC*, which encodes a non-receptor tyrosine kinase implicated in several human cancers, including CRC.

TCGA and Clinical Proteomic Tumor Analysis Consortium data are available to the scientific community, and several research groups are currently using TCGA data for their studies. Additionally, some investigators have expanded their own datasets or combined them with datasets for different categories of genetic aberrations^[47,58]. Other investigators have

utilized specific analytical methods to obtain more in-depth analyses that have enabled the identification of additional genes associated with specific features of CRCs. For example, Ashktorab *et al.*^[54] and Zhu *et al.*^[70] used the elastic-net regression method to perform a supervised analysis that integrated multiple types of genomic data, and then compared the results with the clinical stage of CRC to identify genes associated with advanced CRC. They found that the tumor suppressor gene *WRN* exhibited the highest number of genomic variations (CNVs, expression changes, and methylation) that could be used to delineate advanced CRC.

More simple and immediate uses of these data include the development of portals that collect all CRC-related - omics data, and tools that facilitate data visualization or allow an investigator to perform queries and integrate different types of information^[70].

Finally, RNA-Seq and miRNA-Seq profiles obtained from TCGA project have been used to identify numerous non-coding RNAs (miRNAs, long non-coding RNAs, and competitive endogenous RNAs) in multiple types of cancer. These RNAs are capable of binding to each other, to mRNA, or even to proteins, and regulate their expression. These capabilities make them promising sources of possible new cancer biomarkers or targets that can be used to study cancer development^[71]. Similar to what has been done using genetic and genomic data, several groups have recently began to integrate non-coding RNA, RNA, and protein data obtained from TCGA for the purpose of constructing experimentally supported networks of RNA-RNA and protein-RNA interactions that may become deregulated in different types of cancers^[72]. Although these data are derived from 14 different cancer types, they represent an additional source of information that can be accessed to elucidate disease processes and identify new targets for the treatment of CRC.

CONCLUSION

The development of new high-throughput screening and sequencing technologies has significantly increased the amount of information available concerning the cancer genome and transcriptome. A combinative approach that integrates genetic and genomic data obtained from the same samples provides a more realistic view of the biologic system being analyzed and assists in identifying new therapeutic targets and disease biomarkers. In fact, it can compensate for missing or unreliable information regarding any single data type and better explain the genetic complexities and basic biologic pathways involved in cancer. Additionally, identification of the same gene or biologic pathway in various datasets, and also in different studies, supports its involvement in a disease. This was observed for the four genes that were identified after

integrating CNV and gene expression data obtained from three different studies, *PLCG1*^[18,24], *AHCY*, *TH1L*, and *AURKA*^[18,24,25], all mapping to 20q11-13. Amplification of the 20q11-13 region has been linked to processes that facilitate the progression of adenoma to carcinoma^[73]. *AURKA*, which has previously been associated with gains in 20q11-13, is known to affect cell migration^[74] and may synergistically act with the other three genes (*PLCG1*, *AHCY*, and *TH1L*) to promote CRC progression.

As for miRNA and epigenetic profiles, their integration with gene expression data provides a more comprehensive picture of the regulatory networks involved in cancer, and has confirmed that the Wnt and ErbB pathways play major roles in CRC development.

In summary, the integrative approaches described in this review will ultimately provide investigators and physicians with a more accurate and detailed picture of the complex molecular characteristics of cancers. Furthermore, they should provide new insights that will allow us to better predict cancer, develop a prognosis for cancer patients, and identify new treatments.

REFERENCES

- 1 **Ferlay J**, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893-2917 [PMID: 21351269 DOI: 10.1002/ijc.25516]
- 2 **Fearon ER**, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; **61**: 759-767 [PMID: 2188735]
- 3 **Grady WM**. Genomic instability and colon cancer. *Cancer Metastasis Rev* 2004; **23**: 11-27 [PMID: 15000146]
- 4 **Boland CR**, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* 2010; **138**: 2073-2087.e3 [PMID: 20420947 DOI: 10.1053/j.gastro.2009.12.064]
- 5 **Toyota M**, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 1999; **96**: 8681-8686 [PMID: 10411935]
- 6 **Esteller M**. Epigenetics in cancer. *N Engl J Med* 2008; **358**: 1148-1159 [PMID: 18337604 DOI: 10.1056/NEJMra072067]
- 7 **Sjöblom T**, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006; **314**: 268-274 [PMID: 16959974 DOI: 10.1126/science.1133427]
- 8 **Vogelstein B**, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science* 2013; **339**: 1546-1558 [PMID: 23539594 DOI: 10.1126/science.1235122]
- 9 **Wood LD**, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. *Science* 2007; **318**: 1108-1113 [PMID: 17932254 DOI: 10.1126/science.1145720]
- 10 **Batista PJ**, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell* 2013; **152**: 1298-1307 [PMID: 23498938 DOI: 10.1016/j.cell.2013.02.012]

- 11 **Yates LA**, Norbury CJ, Gilbert RJ. The long and short of microRNA. *Cell* 2013; **153**: 516-519 [PMID: 23622238 DOI: 10.1016/j.cell.2013.04.003]
- 12 **Calin GA**, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer* 2006; **6**: 857-866 [PMID: 17060945 DOI: 10.1038/nrc1997]
- 13 **Hong L**, Yang Z, Ma J, Fan D. Function of miRNA in controlling drug resistance of human cancers. *Curr Drug Targets* 2013; **14**: 1118-1127 [PMID: 23834156]
- 14 The Cancer Genome Atlas. Available from: URL: <http://cancergenome.nih.gov/>
- 15 **Cancer Genome Atlas Network**. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; **487**: 330-337 [PMID: 22810696 DOI: 10.1038/nature11252]
- 16 **Ritchie MD**, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* 2015; **16**: 85-97 [PMID: 25582081 DOI: 10.1038/nrg3868]
- 17 **Stranger BE**, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007; **315**: 848-853 [PMID: 17289997 DOI: 10.1126/science.1136678]
- 18 **Reid JF**, Gariboldi M, Sokolova V, Capobianco P, Lampis A, Perrone F, Signoroni S, Costa A, Leo E, Pilotti S, Pierotti MA. Integrative approach for prioritizing cancer genes in sporadic colon cancer. *Genes Chromosomes Cancer* 2009; **48**: 953-962 [PMID: 19672874 DOI: 10.1002/gcc.20697]
- 19 **Staub E**, Gröne J, Mennerich D, Röpcke S, Klamann I, Hinzmann B, Castanos-Velez E, Mann B, Pilarsky C, Brümmendorf T, Weber B, Buhr HJ, Rosenthal A. A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer. *Mol Cancer* 2006; **5**: 37 [PMID: 16982006 DOI: 10.1186/1476-4598-5-37]
- 20 **Tsafirir D**, Bacolod M, Selvanayagam Z, Tsafirir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, Gerald WL, Paty PB, Domany E, Notterman DA. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 2006; **66**: 2129-2137 [PMID: 16489013 DOI: 10.1158/0008-5472.CAN-05-2569]
- 21 **Alcock HE**, Stephenson TJ, Royds JA, Hammond DW. Analysis of colorectal tumor progression by microdissection and comparative genomic hybridization. *Genes Chromosomes Cancer* 2003; **37**: 369-380 [PMID: 12800148 DOI: 10.1002/gcc.10201]
- 22 **Lassmann S**, Weis R, Makowiec F, Roth J, Danciu M, Hopt U, Werner M. Array CGH identifies distinct DNA copy number profiles of oncogenes and tumor suppressor genes in chromosomal and microsatellite-unstable sporadic colorectal carcinomas. *J Mol Med (Berl)* 2007; **85**: 293-304 [PMID: 17143621 DOI: 10.1007/s00109-006-0126-5]
- 23 **Nakao M**, Kawauchi S, Furuya T, Uchiyama T, Adachi J, Okada T, Ikemoto K, Oga A, Sasaki K. Identification of DNA copy number aberrations associated with metastases of colorectal cancer using array CGH profiles. *Cancer Genet Cytogenet* 2009; **188**: 70-76 [PMID: 19100508 DOI: 10.1016/j.cancergencyto.2008.09.013]
- 24 **Ali Hassan NZ**, Mokhtar NM, Kok Sin T, Mohamed Rose I, Sagap I, Harun R, Jamal R. Integrated analysis of copy number variation and genome-wide expression profiling in colorectal cancer tissues. *PLoS One* 2014; **9**: e92553 [PMID: 24694993 DOI: 10.1371/journal.pone.0092553]
- 25 **Loo LW**, Tiirikainen M, Cheng I, Lum-Jones A, Seifried A, Church JM, Gryfe R, Weisenberger DJ, Lindor NM, Gallinger S, Haile RW, Duggan DJ, Thibodeau SN, Casey G, Le Marchand L. Integrated analysis of genome-wide copy number alterations and gene expression in microsatellite stable, CpG island methylator phenotype-negative colon cancer. *Genes Chromosomes Cancer* 2013; **52**: 450-466 [PMID: 23341073 DOI: 10.1002/gcc.22043]
- 26 **Yoshida T**, Kobayashi T, Itoda M, Muto T, Miyaguchi K, Mogushi K, Shoji S, Shimokawa K, Iida S, Uetake H, Ishikawa T, Sugihara K, Mizushima H, Tanaka H. Clinical omics analysis of colorectal cancer incorporating copy number aberrations and gene expression data. *Cancer Inform* 2010; **9**: 147-161 [PMID: 20706620]
- 27 **Kikuchi A**, Ishikawa T, Mogushi K, Ishiguro M, Iida S, Mizushima H, Uetake H, Tanaka H, Sugihara K. Identification of NUCKS1 as a colorectal cancer prognostic marker through integrated expression and copy number analysis. *Int J Cancer* 2013; **132**: 2295-2302 [PMID: 23065711 DOI: 10.1002/ijc.27911]
- 28 **Gu L**, Xia B, Zhong L, Ma Y, Liu L, Yang L, Lou G. NUCKS1 overexpression is a novel biomarker for recurrence-free survival in cervical squamous cell carcinoma. *Tumour Biol* 2014; **35**: 7831-7836 [PMID: 24819170 DOI: 10.1007/s13277-014-2035-5]
- 29 **Lu J**, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. *Nature* 2005; **435**: 834-838 [PMID: 15944708 DOI: 10.1038/nature03702]
- 30 **Schetter AJ**, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, Yuen ST, Chan TL, Kwong DL, Au GK, Liu CG, Calin GA, Croce CM, Harris CC. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* 2008; **299**: 425-436 [PMID: 18230780 DOI: 10.1001/jama.299.4.425]
- 31 **Volinia S**, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 2006; **103**: 2257-2261 [PMID: 16461460 DOI: 10.1073/pnas.0510565103]
- 32 **Cummins JM**, He Y, Leary RJ, Pagliarini R, Diaz LA, Sjoblom T, Barad O, Bentwich Z, Szafrańska AE, Labourier E, Raymond CK, Roberts BS, Juhl H, Kinzler KW, Vogelstein B, Velculescu VE. The colorectal microRNAome. *Proc Natl Acad Sci USA* 2006; **103**: 3687-3692 [PMID: 16505370 DOI: 10.1073/pnas.0511155103]
- 33 **Hamfjord J**, Stangeland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T, Kure EH. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. *PLoS One* 2012; **7**: e34150 [PMID: 22529906 DOI: 10.1371/journal.pone.0034150]
- 34 **Schee K**, Lorenz S, Worren MM, Günther CC, Holden M, Hovig E, Fodstad O, Meza-Zepeda LA, Flatmark K. Deep Sequencing the MicroRNA Transcriptome in Colorectal Cancer. *PLoS One* 2013; **8**: e66165 [PMID: 23824282 DOI: 10.1371/journal.pone.0066165]
- 35 Target Scan Human. Available from: URL: <http://www.targetscan.org>
- 36 PicTar. Available from: URL: <http://pictar.mdc-berlin.de/>
- 37 miRwalk. Available from: URL: <http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/>
- 38 **Nicolas FE**. Experimental validation of microRNA targets using a luciferase reporter system. *Methods Mol Biol* 2011; **732**: 139-152 [PMID: 21431711 DOI: 10.1007/978-1-61779-083-6_11]
- 39 **Fu J**, Tang W, Du P, Wang G, Chen W, Li J, Zhu Y, Gao J, Cui L. Identifying microRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. *BMC Syst Biol* 2012; **6**: 68 [PMID: 22703586 DOI: 10.1186/1752-0509-6-68]
- 40 **Vishnubalaji R**, Hamam R, Abdulla MH, Mohammed MA, Kassem M, Al-Obeed O, Aldahmash A, Alajez NM. Genome-wide mRNA and miRNA expression profiling reveal multiple regulatory networks in colorectal cancer. *Cell Death Dis* 2015; **6**: e1614 [PMID: 25611389 DOI: 10.1038/cddis.2014.556]
- 41 **Pizzini S**, Bisognin A, Mandruzzato S, Biasiolo M, Faccioli A, Perilli L, Rossi E, Esposito G, Rugge M, Pilati P, Mocellin S, Nitti D, Bortoluzzi S, Zanovello P. Impact of microRNAs on regulatory networks and pathways in human colorectal carcinogenesis and development of metastasis. *BMC Genomics* 2013; **14**: 589 [PMID: 23987127 DOI: 10.1186/1471-2164-14-589]
- 42 **Lanza G**, Ferracin M, Gafà R, Veronese A, Spizzo R, Piciorri F, Liu CG, Calin GA, Croce CM, Negrini M. mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. *Mol Cancer* 2007; **6**: 54 [PMID: 17716371 DOI: 10.1186/1476-4598-6-54]

- 43 **Wang F**, Wong SC, Chan LW, Cho WC, Yip SP, Yung BY. Multiple regression analysis of mRNA-miRNA associations in colorectal cancer pathway. *Biomed Res Int* 2014; **2014**: 676724 [PMID: 24895601 DOI: 10.1155/2014/676724]
- 44 **Gattolliat CH**, Uguen A, Pesson M, Trillet K, Simon B, Doucet L, Robaszekiewicz M, Corcos L. MicroRNA and targeted mRNA expression profiling analysis in human colorectal adenomas and adenocarcinomas. *Eur J Cancer* 2015; **51**: 409-420 [PMID: 25586944 DOI: 10.1016/j.ejca.2014.12.007]
- 45 **Reid JF**, Sokolova V, Zoni E, Lampis A, Pizzamiglio S, Bertan C, Zanutto S, Perrone F, Camerini T, Gallino G, Verderio P, Leo E, Pilotti S, Gariboldi M, Pierotti MA. miRNA profiling in colorectal cancer highlights miR-1 involvement in MET-dependent proliferation. *Mol Cancer Res* 2012; **10**: 504-515 [PMID: 22343615 DOI: 10.1158/1541-7786.MCR-11-0342]
- 46 **Sokolova V**, Fiorino A, Zoni E, Crippa E, Reid JF, Gariboldi M, Pierotti MA. The Effects of miR-20a on p21: Two Mechanisms Blocking Growth Arrest in TGF- β -Responsive Colon Carcinoma. *J Cell Physiol* 2015; **230**: 3105-3114 [PMID: 26012475 DOI: 10.1002/jcp.25051]
- 47 **Ling H**, Pickard K, Ivan C, Isella C, Ikuro M, Mitter R, Spizzo R, Bullock MD, Braicu C, Pileczki V, Vincent K, Pichler M, Stiegelbauer V, Hoefler G, Almeida MI, Hsiao A, Zhang X, Primrose JN, Packham GK, Liu K, Bojja K, Gafà R, Xiao L, Rossi S, Song JH, Vannini I, Fanini F, Kopetz S, Zweidler-McKay P, Wang X, Ionescu C, Irimie A, Fabbri M, Lanza G, Hamilton SR, Berindan-Neagoe I, Medico E, Mirnezami AH, Calin GA, Nicoloso MS. The clinical and biological significance of MIR-224 expression in colorectal cancer metastasis. *Gut* 2015; Epub ahead of print [PMID: 25804630 DOI: 10.1136/gutjnl-2015-309372]
- 48 **Dawson MA**, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012; **150**: 12-27 [PMID: 22770212 DOI: 10.1016/j.cell.2012.06.013]
- 49 **Baylin SB**, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* 2006; **6**: 107-116 [PMID: 16491070 DOI: 10.1038/nrc1799]
- 50 **Chan AO**, Broaddus RR, Houlihan PS, Issa JP, Hamilton SR, Rashid A. CpG island methylation in aberrant crypt foci of the colorectum. *Am J Pathol* 2002; **160**: 1823-1830 [PMID: 12000733 DOI: 10.1016/S0002-9440(10)61128-5]
- 51 **Chan TA**, Glockner S, Yi JM, Chen W, Van Neste L, Cope L, Herman JG, Velculescu V, Schuebel KE, Ahuja N, Baylin SB. Convergence of mutation and epigenetic alterations identifies common genes in cancer that predict for poor prognosis. *PLoS Med* 2008; **5**: e114 [PMID: 18507500 DOI: 10.1371/journal.pmed.0050114]
- 52 **Easwaran H**, Baylin SB. Epigenetic abnormalities in cancer find a "home on the range". *Cancer Cell* 2013; **23**: 1-3 [PMID: 23328477 DOI: 10.1016/j.ccr.2012.12.018]
- 53 **Pandiyani K**, You JS, Yang X, Dai C, Zhou XJ, Baylin SB, Jones PA, Liang G. Functional DNA demethylation is accompanied by chromatin accessibility. *Nucleic Acids Res* 2013; **41**: 3973-3985 [PMID: 23408854 DOI: 10.1093/nar/gkt077]
- 54 **Ashktorab H**, Rahi H, Wansley D, Varma S, Shokrani B, Lee E, Darempouran M, Laiyemo A, Goel A, Carethers JM, Brim H. Toward a comprehensive and systematic methylome signature in colorectal cancers. *Epigenetics* 2013; **8**: 807-815 [PMID: 23975090 DOI: 10.4161/epi.25497]
- 55 **Ang PW**, Loh M, Liem N, Lim PL, Grieu F, Vaithilingam A, Platell C, Yong WP, Iacopetta B, Soong R. Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer* 2010; **10**: 227 [PMID: 20492682 DOI: 10.1186/1471-2407-10-227]
- 56 **Hinoue T**, Weisenberger DJ, Lange CP, Shen H, Byun HM, Van Den Berg D, Malik S, Pan F, Noushmehr H, van Dijk CM, Tollenaar RA, Laird PW. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* 2012; **22**: 271-282 [PMID: 21659424 DOI: 10.1101/gr.117523.110]
- 57 **Szmida E**, Karpinski P, Leszczynski P, Sedziak T, Kielan W, Ostasiewicz P, Sasiadek MM. Aberrant methylation of ERBB pathway genes in sporadic colorectal cancer. *J Appl Genet* 2015; **56**: 185-192 [PMID: 25366420 DOI: 10.1007/s13353-014-0253-6]
- 58 **Wang Q**, Jia P, Cheng F, Zhao Z. Heterogeneous DNA methylation contributes to tumorigenesis through inducing the loss of coexpression connectivity in colorectal cancer. *Genes Chromosomes Cancer* 2015; **54**: 110-121 [PMID: 25407423 DOI: 10.1002/gcc.22224]
- 59 **Laczmanska I**, Karpinski P, Bebenek M, Sedziak T, Ramsey D, Szmida E, Sasiadek MM. Protein tyrosine phosphatase receptor-like genes are frequently hypermethylated in sporadic colorectal cancer. *J Hum Genet* 2013; **58**: 11-15 [PMID: 23096495 DOI: 10.1038/jhg.2012.119]
- 60 **Rousseau B**, Chibaudel B, Bachet JB, Larsen AK, Tournigand C, Louvet C, André T, de Gramont A. Stage II and stage III colon cancer: treatment advances and future directions. *Cancer J* 2010; **16**: 202-209 [PMID: 20526097 DOI: 10.1097/PP0.0b013e3181ddc5bf]
- 61 **Lopez-Serra P**, Esteller M. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene* 2012; **31**: 1609-1622 [PMID: 21860412 DOI: 10.1038/onc.2011.354]
- 62 **Bandres E**, Agirre X, Bitarte N, Ramirez N, Zarate R, Roman-Gomez J, Prosper F, Garcia-Foncillas J. Epigenetic regulation of microRNA expression in colorectal cancer. *Int J Cancer* 2009; **125**: 2737-2743 [PMID: 19521961 DOI: 10.1002/ijc.24638]
- 63 **Kalimutho M**, Di Cecilia S, Del Vecchio Blanco G, Roviello F, Sileri P, Cretella M, Formosa A, Corso G, Marrelli D, Pallone F, Federici G, Bernardini S. Epigenetically silenced miR-34b/c as a novel faecal-based screening marker for colorectal cancer. *Br J Cancer* 2011; **104**: 1770-1778 [PMID: 21610744 DOI: 10.1038/bjc.2011.82]
- 64 **Davalos V**, Moutinho C, Villanueva A, Boque R, Silva P, Carneiro F, Esteller M. Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene* 2012; **31**: 2062-2074 [PMID: 21874049 DOI: 10.1038/onc.2011.383]
- 65 **Grady WM**, Parkin RK, Mitchell PS, Lee JH, Kim YH, Tsuchiya KD, Washington MK, Paraskeva C, Willson JK, Kaz AM, Kroh EM, Allen A, Fritz BR, Markowitz SD, Tewari M. Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. *Oncogene* 2008; **27**: 3880-3888 [PMID: 18264139 DOI: 10.1038/onc.2008.10]
- 66 **Tang JT**, Wang JL, Du W, Hong J, Zhao SL, Wang YC, Xiong H, Chen HM, Fang JY. MicroRNA 345, a methylation-sensitive microRNA is involved in cell proliferation and invasion in human colorectal cancer. *Carcinogenesis* 2011; **32**: 1207-1215 [PMID: 21665895 DOI: 10.1093/carcin/bgr114]
- 67 Gene Expression Omnibus. Available from: URL: <http://www.ncbi.nlm.nih.gov/geo/>
- 68 ArrayExpress-functional genomics data. Available from: URL: <http://www.ebi.ac.uk/arrayexpress>
- 69 **Zhang B**, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, Kinsinger CR, Rivers RC, Rodriguez H, Townsend RR, Ellis MJ, Carr SA, Tabb DL, Coffey RJ, Slebos RJ, Liebler DC. Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014; **513**: 382-387 [PMID: 25043054 DOI: 10.1038/nature13438]
- 70 **Zhu J**, Shi Z, Wang J, Zhang B. Empowering biologists with multi-omics data: colorectal cancer as a paradigm. *Bioinformatics* 2015; **31**: 1436-1443 [PMID: 25527095 DOI: 10.1093/bioinformatics/btu834]
- 71 **Weinstein JN**, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013; **45**: 1113-1120 [PMID: 24071849 DOI: 10.1038/ng.2764]
- 72 **Li JH**, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014; **42**: D92-D97 [PMID: 24297251 DOI: 10.1093/nar/gkt1248]
- 73 **Carvalho B**, Postma C, Mongera S, Hoppmans E, Diskin S, van

de Wiel MA, van Criekinge W, Thas O, Matthäi A, Cuesta MA, Terhaar Sive Droste JS, Craanen M, Schröck E, Ylstra B, Meijer GA. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* 2009; **58**: 79-89 [PMID: 18829976 DOI: 10.1136/gut.2007.143065]

74 **Sillars-Hardebol AH**, Carvalho B, Tijssen M, Beliën JA, de Wit M, Delis-van Diemen PM, Pontén F, van de Wiel MA, Fijneman RJ, Meijer GA. TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression. *Gut* 2012; **61**: 1568-1575 [PMID: 22207630 DOI: 10.1136/gutjnl-2011-301153]

P- Reviewer: Huang ZH, Meshikhes AW
S- Editor: Qi Y **L- Editor:** A **E- Editor:** Liu XM





Published by **Baishideng Publishing Group Inc**

8226 Regency Drive, Pleasanton, CA 94588, USA

Telephone: +1-925-223-8242

Fax: +1-925-223-8243

E-mail: bpgoffice@wjgnet.com

Help Desk: <http://www.wjgnet.com/esps/helpdesk.aspx>

<http://www.wjgnet.com>



ISSN 1007-9327



9 771007 932045