Genomic approach to translational studies in colorectal cancer

Dane Cheasley, Robert N. Jorissen, Sheng Liu, Chin Wee Tan, Christopher Love, Michelle Palmieri, Oliver M. Sieber

1Systems Biology and Personalised Medicine Division, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; 2Faculty of Medicine, Dentistry and Health Sciences, Department of Medical Biology, University of Melbourne, Parkville, VIC, Australia; 3Structural Biology Division, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

*These authors contributed equally to this work.

Abstract: Colorectal cancer (CRC) is a leading cause of cancer-related mortality worldwide. The pathogenesis of CRC is complex with molecular subtypes defined by pathways of sequential (epi-) genetic alterations and different forms of genomic instability. The management of CRC has evolved considerably over the past decade, consisting of surgery and treatment with radiation, chemotherapy or molecularly targeted agents. Population screening is progressively being introduced for early detection of disease. Decisions to use a particular treatment are principally related to primary site, TNM stage and comorbidity of the patient. Advances in the development of microarray and next-generation sequencing (NGS) technologies have enabled the possibility of personalized medicine for CRC, revolutionizing our knowledge of tumor biology, identifying prognostic and predictive biomarkers, and contributing new tools for diagnosis and surveillance. Genomic studies have identified new cancer driver genes and druggable targets, revealed substantial inter- and intra-tumor (epi-) genetic heterogeneity, and highlighted the importance of cell-of-origin, differentiation hierarchy, phenotypic plasticity and stromal contribution for tumor clinical behavior. Here, we review results of recent translational studies of the CRC genome, transcriptome, methylome and miRNAome, with a focus on tumor classification, diagnostic, prognostic and predictive findings.

Keywords: Colorectal cancer (CRC); genomic; methylation; microRNA; transcriptome

Introduction

Colorectal cancer (CRC) is the third most common malignant neoplasm worldwide and a leading cause of cancer-related morbidity and mortality (1). Metastatic CRC (also called stage IV or advanced CRC) is the principal cause of death, but if cancer is detected at early stages curative treatment is often possible. Surgery is the primary form of treatment and results in cure for ~60% of patients with localized (stage I-III) disease (2,3). However, recurrence following surgery remains a major problem, and patients with lymph-node positive stage III and high-risk stage II disease are offered fluoropyrimidine-based adjuvant chemotherapy (5-fluorouracil, capecitabine) with or without oxaliplatin. Rectal cancers may also receive pre-operative chemoradiation. Recurrent tumor can develop in the bowel or at distant sites including the liver, lung, peritoneum, brain and bone (2,3). Guidelines for postsurgery surveillance recommend a combination of clinical assessment, serum carcinoembryonic antigen (CEA) testing, colonoscopy and computed tomography (CT) scanning (4-6). In current practice, many CRC patients receive adjuvant therapy unnecessarily, either because they were cured by surgery alone, or because they will relapse despite treatment. Conversely, some stage II patients with low-risk clinicopathological features who are currently not considered for adjuvant therapy do relapse and might benefit from therapy.
Outcomes from metastatic CRC remain poor, with a 5-year survival rate of less than 20% (7). Curative surgery is only rarely possible in these patients, but an increase in therapeutic options has resulted in an improvement of median overall survival to ~24 months. Approved agents include standard chemotherapeutics (5-fluorouracil, capecitabine, oxaliplatin, irinotecan) and targeted therapies directed against the epidermal growth factor receptor (EGFR) (cetuximab, panitumumab) or angiogenesis (bevacizumab, afibercept, regorafenib). Although these treatments have prolonged the lives of patients with metastatic CRC, clinical responses are limited to a subset of individuals and are generally short-lived with most tumors developing resistance within a few months. Significant side effects and costs are associated with these treatments, and identification of individuals who are likely to derive the greatest benefit remains a major challenge.

Many patients with CRC will remain asymptomatic until the development of late-stage disease, where symptoms may include abdominal pain, changes in bowel habit and the presence of blood in stool. The principal method adopted by national CRC screening programs for early disease detection is the fecal occult blood test (FOBT), targeted at high-risk age groups with follow-up by colonoscopy (8,9). FOBT screening is cost effective, but tests suffer from limited sensitivity and specificity. A further challenge is population participation for stool-based diagnostics.

Advances in the development of microarray and next-generation sequencing (NGS) technologies have enabled global studies of CRC genomes, methylomes, as well as coding and non-coding transcriptomes (Figure 1). Integrated omics data have led to the identification of new cancer genes and pathways, and have improved our understanding of tumor biology and molecular subtypes. Translational genomics studies have revealed clinically relevant biomarkers for improving CRC diagnosis, surveillance, prediction of prognosis and therapy response. In addition, such studies have identified new druggable targets, opening up novel therapeutic opportunities. Here, we summarize pertinent results of CRC genomics studies to date, with an emphasis on tumor classification, diagnosis, prognostication and prediction of therapy benefit.

**The CRC genome**

Comparative genomic hybridization (CGH) arrays, single nucleotide polymorphism (SNP) arrays and more recently NGS approaches have provided fundamental insights into the complex landscapes of CRC mutations, DNA copy number alterations and chromosomal rearrangements. Sjöblom et al. and Wood et al. first used classic PCR-based Sanger sequencing for exome-wide profiling of CRC mutations, identifying well-known, high-frequency mutated genes such as **APC**, **KRAS**, **PIK3CA**, **SMAD4**, **TP53** and **FBXW7** as ‘gene mountains’, and describing a large number of ‘gene hills’ that were mutated at low frequency (10,11).
These pioneering studies were followed by integrated whole-exome NGS and DNA copy-number studies by The Cancer Genome Atlas (TCGA) Network, presenting a detailed survey of the genomic profiles on over 270 sporadic CRCs (12). Approximately 15% of CRCs were found to exhibit hypermutation with two distinct mutation patterns: microsatellite instability (MSI) in three-quarters of cases, characterized by increased insertions, deletions and single nucleotide substitutions, usually with hypermethylation and MLH1 silencing, and a nucleotide substitution hypermutator phenotype (NSHP) in one-quarter of cases, associated with mutations in polymerase ε (POLE). Twenty-four genes were highlighted as significantly mutated, targeting the WNT, RTK/RAS, PI3K, TGF-β and TP53 pathways in both non-hypermutated and hypermutated tumors, but with different genetic alterations between these CRC subtypes. Non-hypermutated tumors showed common mutations in APC, TP53, KRAS, PIK3CA, FBXW7, SMAD4, TCF7L2, NRAS, CTNNB1, SMAD2, FAM123B, SOX9, ATM and ARID1A, while hypermutated tumors showed frequent alterations in ACVR2A, APC, TGFB2, BRAF, MSH3, MSH6, SLC9A9 and TCF7L2. At the chromosomal level, non-hypermutated tumors tended to be near-diploid, while hypermutated tumors tended to be near-aneuploid. Consistent with previous CGH and SNP array studies (13-18), the most commonly deleted chromosome arms were 8p, 15q, 17p (including TP53) and 18q (including SMAD4), and the most commonly gained regions were chromosome 7, 8q (including MYC), 13 and 20q. Recurrent copy-number alterations included potentially drug-targetable amplifications of ERBB2 and IGF2. Low prevalence chromosomal translocations were detected between NAV2 and the WNT pathway member TCF7L1 using whole-genome sequencing on a subset of samples. A similar genomic study on 74 primary colon tumors reported highly concordant results, and also identified recurrent fusion transcripts involving R-spondin family members (EIF3E-RSPO2 and PTPRK-RSPO3) that were shown to contribute to activation of oncogenic WNT/β-catenin signaling (19,20). Additional low prevalence translocations identified by whole-genome or targeted NGS studies in CRC include C2orf44-ALK, VT11A-TCF7L2 and LACTB2-NCOA2 (19,21,22). Recent NGS studies have provided additional details on the mutation spectra of colorectal adenomas, MSI and microsatellite stable (MSS) carcinomas (23-29). Mutational heterogeneity has been investigated between primary cancers and matched metastases indicating high genomic concordance, with a thick common trunk and smaller genomic branches (30-33). Some evidence exists for intra-tumor mutational heterogeneity, but data on this are still emerging (33). CGH array studies have proposed a refined classification of non-hypermutated CRCs into chromosomally stable (CSS) and chromosomal instability (CIN) groups (17,34). However, these groups have not as yet been systematically investigated for specific mutation signatures.

Non-invasive analysis of circulating tumor DNA (ctDNA) is an emerging genomics-tool that is actively being developed to improve CRC diagnosis and post-surgery surveillance. It is based on the detection of tumor specific single-base substitutions or larger somatic structural variations (SSVs) in DNA fragments that are released by tumors into plasma. Assays are typically designed against either point mutations in hotspot genes or patient-specific SSVs (35-40). Hotspot mutations can be utilized in both the diagnostic and surveillance setting, but these may only identify a subset of patients and have limited specificity. Application of patient-specific SSVs is restricted to the surveillance setting, requiring low coverage whole-genome sequencing and/or microarray analysis of resected tumor for assay design, yet highly-specific tests can in principle be produced for all individuals. Several reports have shown that assays against point mutations in hotspot mutated genes like KRAS, BRAF and PIK3CA can identify ctDNA fragments in plasma and serum in ~70% of patients with CRC (38,39). Recently, a clinical pipeline for identification of patient-specific SSVs for post-surgery CRC surveillance has been presented, demonstrating sensitive temporal assessment of disease status, response to surgical and oncological intervention, and early detection of recurrence (40). Recommending the use of at least three SSVs per patient to counter observed primary-metastasis genetic heterogeneity, this approach achieved sensitivity and specificity of 100% for detecting relapse, with a 2-15 (mean 10) months lead time compared to conventional follow-up.

Stool-based diagnostic tests have also been successfully tested for detection of mutations in high-frequency mutated CRC genes, including APC, KRAS and TP53 (41-48) such as the clinically used PreGen-Plus™ kit (49). Additionally, studies have evaluated long fragment DNA from exfoliated cancer cells in stool as diagnostic marker, with modest sensitivity and specificity (50-52).

Genomic instability phenotypes of CRC, MSI and CIN, have been demonstrated to be predictive of good and poor prognosis, respectively (53,54). The extent of CIN may provide additional prognostic value (55,56). Several CGH
array studies have attempted to define particular regions of chromosomal gain or loss related to tumor progression and outcome (14,16,57-68). Perhaps the strongest data exist for loss of chromosome arms 4q and 18q and inferior survival, but whether these relationships are independent of global CIN status remains uncertain (68,69). Recently, different types of CIN, such as genome-doubling and chromothripsis, have been suggested to be adversely related with patient outcome (70,71).

Targeted gene sequencing studies to develop integrated mutation signatures for CRC prognostication are only beginning to emerge. A recent study evaluating 187 recurrent and pathway-related genes in 160 patients with stage I-IV CRCs, has proposed a five-gene-signature (CDH10, COL6A3, SMAD4, TMEM132D, VCAN) for stratifying patients by outcome independent of TNM status (72).

Genomic approaches are gradually being applied for identification of molecular markers of therapy benefit. To date, unbiased exome mutation and DNA copy number studies have focused on cancer cell lines in the context of high-throughput drug screens. However, only small numbers of CRC cell lines have been included in such screens thus limiting the power of these studies to identify robust biomarker-drug response associations (73-75). In patients with metastatic CRC, several targeted gene mutation and copy-number analyses have investigated resistance to treatment with monoclonal antibodies targeting EGFR. These studies have largely considered “rational” candidate genes indicated by previous focused studies. For example, Peeters et al. evaluated cancer resistance to panitumumab using massively parallel multigene tumor sequencing of KRAS, NRAS, BRAF, PIK3CA, PTEN, TP53, EGFR, AKT1 and CTNNB1. As found in other reports (76,77), wild-type KRAS, NRAS and BRAF status were associated with longer progression-free survival (78). Giardiello et al. reported a similar targeted NGS study interrogating 22 genes in patients treated with FOLFIRI plus cetuximab, reporting worse outcome in cases with KRAS, NRAS, BRAF, or PIK3CA mutations (79). The potential of ctDNA analysis for monitoring intrinsic and acquired resistance to anti-EGFR antibody therapy has been successfully demonstrated, applying both targeted mutation and SSV analysis (38,80-82). Limited data suggest that tumor DNA copy number profiles may correlate with outcome in advanced CRC patients treated with fluoropyrimidine-based regimens. In particular, chromosomal losses of 18q, 17p11.2-p13.2 and gains of 20p13-q13.3 have been associated with response to the FU + irinotecan (FOLFIRI) and capcitabine + irinotecan (CAPIRI) (83,84).

Limited data exist for rectal cancer response to preoperative chemoradiation. A study by Chen et al. highlighted loss of chromosome 4 as associated with the risk of lymph node metastasis (85). Similarly, Grade et al. suggested that pre-therapeutic evaluation of gains of chromosomal regions 7q32-q36 and 7q11-q31, and amplifications of 20q11-q13 may predict responsiveness to chemoradiotherapy (86).

The CRC transcriptome

Analysis of the protein-coding CRC transcriptome using microarray platforms has provided a framework for classification of CRC subtypes and prediction of cancer outcomes and therapy benefit. These signatures are generally derived from the analysis of resected tumor specimens with limited micro-dissection and capture neoplastic, stromal and immune components.

Several classification schemes for CRC have been proposed based on unsupervised clustering of tumor gene expression data (87-91). Approaches to tumor categorization have included hierarchical clustering, non-negative matrix factorization and the clustering of meta-genes (medians of groups of genes with correlated expression). Although these classification schemes differ in the number and detail of the subtypes proposed, ranging from three to six groups, major themes are separation into classes differentiated by MSI and CIN status, tumor location, and expression of epithelial versus mesenchymal markers (Table 1). One study has aligned their classification with different types of precursor lesions, classic versus serrated adenoma (87), while another has connected their classes with the cell types in colorectal crypts, stem cell, transit-amplifying cell, goblet cell and enterocyte (88). Recently, two studies have demonstrated major contributions of stromal cells in tumor groups with increased mesenchymal marker expression, rather than tumor cells undergoing epithelial-to-mesenchymal transition as was originally proposed (92,93).

mRNA extracted from blood and stool have been considered as biomarker analysis for diagnosis of CRC (94-96). Several groups have used expression microarrays on blood from patients with CRC and healthy controls to identify an initial set of candidate diagnostic mRNAs followed by further refinement of candidates using RT-PCR (96,97). Other groups have screened normal and tumor tissue to find differentially expressed candidate genes, from which a refined set was obtained upon
follow-up using mRNA extracted from stool-derived colonocytes (98) or blood (99). A related approach has been to examine previously described candidate genes with reported high levels of expression in tumour or patient blood (e.g., COX2, MMP7 and CEA) (100-102).

One commercial blood-based diagnostic test for CRC, ColonSentry®, which uses a 7-gene mRNA signature is currently available (103,104) (Table 2). Validation in two cohorts yielded sensitivity and specificity of 72-82% and 64-70%, respectively (103,111). However, research into using mRNA for CRC diagnosis appears to be waning relative to approaches which utilise aberrant DNA methylation or miRNAs discussed below.

Multiple studies (126-139) have searched for gene expression signatures for predicting risk of tumor recurrence following surgical resection of the primary tumor (112,120,125,140-155). Early studies often had modest sample sizes and relied on cross-validation to assess performance of their signatures, while later studies evaluated larger sample sizes and included independent patient cohorts for signature assessment. A survey of 31 gene signatures demonstrated little overlap in the component genes (156), and only modest prognostic performance when assessed in independent datasets (156). Recognized reasons for these

<table>
<thead>
<tr>
<th>Classification scheme</th>
<th>Class</th>
<th>MSI, CIN</th>
<th>CIMP, BRAF</th>
<th>Prognosis (RFS, OS)</th>
<th>Molecular phenotype</th>
<th>Similarity to precursor polyp</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSS/conventional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Sousa E Melo CCS1-CIN</td>
<td>MSS, CIN</td>
<td>CIMP0, BRAFwt</td>
<td>Intermediate</td>
<td>Epithelial</td>
<td>Tubular</td>
<td>Left</td>
<td></td>
</tr>
<tr>
<td>Roepman B-type</td>
<td>MSS</td>
<td>BRAFwt</td>
<td>Intermediate</td>
<td>Epithelial</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Sadanandam Transit-amplifying</td>
<td>MSS, CIN</td>
<td>CIMP0, BRAFwt</td>
<td>Mixed</td>
<td>Epithelial</td>
<td></td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>Marisa C1</td>
<td>MSS, CIN</td>
<td>CIMP0, BRAFwt</td>
<td>Intermediate</td>
<td>Epithelial</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Marisa C5</td>
<td>MSS, CIN</td>
<td>CIMP0, BRAFwt</td>
<td>Intermediate</td>
<td>Not serrated</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Budinska B</td>
<td>MSS, CIN</td>
<td>CIMP0↑, BRAFwt</td>
<td>Good</td>
<td>Epithelial</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Sadanandam Enterocyte</td>
<td>MSI, MSS</td>
<td>BRAFwt</td>
<td>Intermediate</td>
<td>Epithelial</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Budinska E (mixed)</td>
<td>MSS, CIN</td>
<td>CIMP0↑, BRAFwt</td>
<td>Poor</td>
<td>Epithelial, inflammatory</td>
<td></td>
<td>Left</td>
<td></td>
</tr>
<tr>
<td>Budinska A</td>
<td>MSS, CSS</td>
<td>CIMP0↑, BRAFwt</td>
<td>Good</td>
<td>Epithelial</td>
<td></td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>MSI-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Sousa E Melo CCS2-MSI</td>
<td>MSI</td>
<td>CIMP, BRAFmt↑</td>
<td>Good</td>
<td>Inflammatory</td>
<td></td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Sadanandam Inflammatory</td>
<td>MSI</td>
<td>BRAFmt↑</td>
<td>Intermediate</td>
<td>Inflammatory</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>Marisa C2</td>
<td>MSI↑, CSS↑</td>
<td>CIMP↑, BRAFmt↑</td>
<td>Intermediate</td>
<td>Serrated</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>Budinska C</td>
<td>MSI, CSS↑</td>
<td>CIMP↑, BRAFmt↑</td>
<td>Intermediate</td>
<td>Inflammatory</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>Roepman A-type</td>
<td>MSI↑</td>
<td>BRAFmt↑</td>
<td>Good</td>
<td>Epithelial</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>Sadanandam Goblet-like</td>
<td>MSI</td>
<td>BRAFmt↑</td>
<td>Good</td>
<td>Epithelial</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>Marisa C3</td>
<td>CSS↑</td>
<td>CIMP↑, BRAFwt</td>
<td>Intermediate</td>
<td>Serrated</td>
<td></td>
<td>Right ↑</td>
<td></td>
</tr>
<tr>
<td>MSS/serrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Sousa E Melo CCS3-serrated</td>
<td>MSS↑</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td>Serrated</td>
<td></td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>Roepman C-type</td>
<td>MSS/MSI</td>
<td>BRAFmt↑</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Sadanandam Stem-like</td>
<td>MSS</td>
<td>BRAFmt↑</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Marisa C4</td>
<td>MSS, CIN↑</td>
<td>CIMP↑, BRAFwt↑</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td></td>
<td>Both</td>
<td></td>
</tr>
<tr>
<td>Marisa C6</td>
<td>MSS, CIN</td>
<td>CIMP0↑, BRAFwt</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
<tr>
<td>Budinska D</td>
<td>MSI↑, CIN↑</td>
<td>CIMP0*↑, BRAFmt↑</td>
<td>Poor</td>
<td>Mesenchymal</td>
<td></td>
<td>Left ↑</td>
<td></td>
</tr>
</tbody>
</table>

*CIMP status assigned using a microarray expression signature rather than a panel of methylation markers.
findings are technical differences in sample preparation and microarray processing, cohort heterogeneity and gene selection methods. These challenges can be overcome using rigorously controlled assay conditions. Accordingly, four prognostic gene expression signatures have been translated into clinical use following extensive validation on external cohorts: Oncotype DX® (Colon), a 12-gene RT-PCR based assay (112-119,157), ColoPrint®, an 18 gene microarray-based assay (120-122), OncoDefender™, a 5 gene RT-PCR based assay (123,124), and GeneFx Colon®, a 634 gene microarray-based assay (125) (Table 2). Gene expression based CRC classification schemes may also have prognostic potential, with MSI-associated classes showing good prognosis and serrated/mesenchymal classes exhibiting poor prognosis (87-91).

Transcriptome analyses have also been attempted for predicting response to chemotherapy and radiotherapy for CRC. Perhaps the most studied scenario has been that of pre-operative chemoradiation in rectal cancer patients, utilizing pre-treatment biopsies (158-165). These transcriptomic studies typically involve smaller training sets (n<100) than those for prognosis signatures and generally lack external validation. Classifier genes show little overlap, and an evaluation of three reported signatures found poor performance in an external dataset (166). One recent review concluded that an optimal gene signature for prediction of chemoradiotherapy in rectal cancer patients has not yet been found (167). A small number of studies have used transcriptomic data to generate models of 5-FU-based chemotherapy benefit in patients with advanced CRC (168-171). These studies are limited by small numbers of patients and a lack of validation in large patient cohorts. In general, gene signatures developed for predicting risk of tumor recurrence following surgical resection of the primary tumor have not been shown to exhibit predictive value for 5-FU based adjuvant chemotherapy.

Table 2 Diagnostic and prognostic tests for colorectal cancer in which genomic methods were used as part of development and/or implementation of clinical test

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Assay(s)</th>
<th>Source</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ColoVantage® Plasma</td>
<td>Methylation status of SEPT9 DNA promoter region</td>
<td>RT-PCR</td>
<td>blood plasma</td>
<td>(105)</td>
</tr>
<tr>
<td>Epi proColon®</td>
<td>Methylation status of SEPT9 DNA promoter region</td>
<td>RT-PCR</td>
<td>blood plasma</td>
<td>(106-108)</td>
</tr>
<tr>
<td>RealTime mS9™</td>
<td>Methylation status of SEPT9 DNA promoter region</td>
<td>RT-PCR</td>
<td>blood plasma</td>
<td>(109)</td>
</tr>
<tr>
<td>ColoSure™</td>
<td>Methylation of VIM from DNA</td>
<td>RT-PCR</td>
<td>Stool</td>
<td>(110)</td>
</tr>
<tr>
<td>PreGen-Plus™</td>
<td>21 point mutations in KRAS, APC, TP53; a MSI marker (BAT-28) and a DNA integrity marker</td>
<td>Capillary electrophoretograms</td>
<td>Stool</td>
<td>(49)</td>
</tr>
<tr>
<td>ColoGuard®</td>
<td>KRAS mutations, VIM, NDRG4 and BMP3 methylation (plus ACTB reference), plus presence of haemoglobin</td>
<td>RT-qPCR</td>
<td>Stool</td>
<td>(48)</td>
</tr>
<tr>
<td>ColonSentry®</td>
<td>Expression levels of seven gene biomarkers</td>
<td>RT-PCR</td>
<td>Blood plasma</td>
<td>(103,111)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prognostic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oncotype DX® Colon</td>
<td>Expression levels of 12 genes (7 cancer related genes and 5 reference genes)</td>
<td>RT-qPCR</td>
<td>FFPE primary tumour</td>
<td>(112-119)</td>
</tr>
<tr>
<td>ColoPrint®</td>
<td>Expression levels of 18 genes</td>
<td>Microarray</td>
<td>Fresh-frozen primary tumour</td>
<td>(120-122)</td>
</tr>
<tr>
<td>OncoDefender™</td>
<td>Expression levels of 5 genes</td>
<td>RT-PCR</td>
<td>FFPE primary tumour</td>
<td>(123,124)</td>
</tr>
<tr>
<td>GeneFx Colon® (formerly ColDx)</td>
<td>Expression levels of 634 genes</td>
<td>Microarray</td>
<td>FFPE primary tumour</td>
<td>(125)</td>
</tr>
</tbody>
</table>

RT-PCR, quantitative real-time PCR.
benefit, although in two studies benefit was suggested to be limited to the poor prognosis groups (148,149). The application of transcriptome approaches to targeted biological therapies is an emerging field (172-176). Several of the CRC classification schemes have been suggested to have predictive value for 5-FU-based chemotherapy or radiotherapy, but with apparently conflicting results (Table 1). For example, the mesenchymal “stem cell-like” class of Sadanandam et al. (88) was found to be sensitive to FOLFIRI and radiotherapy, while the mesenchymal “C-type” class of Roepman was associated with 5FU-resistance (91).

The CRC methylome

DNA methylation of cytosines in the context of CpG dinucleotides is a central mechanism of epigenetic control, with essential roles in the maintenance of genome integrity, genomic imprinting, transcriptional regulation, and developmental processes. Multiple approaches for genome-wide studies of DNA methylation patterns have been developed, generally combining DNA analysis by microarrays or NGS with one of three techniques to convert DNA methylation patterns into DNA sequence information or library enrichment: endonuclease digestion, affinity enrichment and bisulphite conversion (177,178).

Genome-wide methyleome analyses have highlighted extensive disruption of DNA methylation in CRC. Tumors are typically characterized by global loss of methylation (hypermethylation), predominantly in repetitive sequences, and focal gain in methylation (hypermethylation) in CpG islands, the latter often occurring simultaneously within defined megabase regions (179-181). Hypermethylation within CpG islands is associated with transcriptional silencing of tumor suppressor genes, whilst hypomethylation within gene bodies can affect transcriptional elongation or alternative promoter usage and cause aberrant transcription of oncogenes (182-196). Global loss of methylation may trigger cancer genomic instability and activation of transposons and genes within regions of repetitive sequence (186-188). Both hypo- and hypermethylation occur early in tumorigenesis (189-196), and the average CRC genome carries thousands of methylation changes with marked impact on the cellular transcriptional program (197,198).

Studies have identified a subset of CRCs that exhibit particularly widespread promoter hypermethylation, referred to as the CpG island methylator phenotype (CIMP) (199,200). CIMP is observed in ~30% of CRCs, and presence and extent of CIMP have been used to classify CRC into three major subgroups, CIMP high (CIMP-H), CIMP low (CIMP-L) and non-CIMP (CIMP-0), with distinct clinical and molecular features (201,202). CIMP-H is associated with proximal tumor location, female gender, BRAFV600E mutation, MLH1 methylation and MSI; CIMP-L is characterized by proximal tumor location and KRAS mutation, while CIMP-0 is associated with distal tumor location, TP53 mutation and CIN (202-205).

Aberrant DNA methylation patterns are attractive tumor biomarkers because of their high frequency in neoplasms, and the detection of methylation in DNA isolated from stool and/or blood has emerged as a promising approach for early diagnosis and surveillance of CRC (206,207). Microarray based studies of hypermethylated CpG sites in CRC and benign adenomas have revealed a large number of tumor-specific candidate detection markers (195,208-210). Translation of these candidates into blood- or stool-based diagnostic tests is actively being pursued by academia and industry, involving method development, validation of specificity against normal tissues and other pathologies, and evaluation of performance against routine clinical assays (FOBT, CEA). A recent study evaluating circulating DNA detection of HLTF and HPP1 hypermethylation in addition to CEA serum measurements showed that combination of all three markers outperformed each assay on its own (211). In a related study, Lange et al. suggested that blood-based detection of THBD and C9orf50 hypermethylation outperformed CEA (212). Two diagnostic tests have already been introduced into clinical practice, including a blood-based PCR test for methylated septin-9 (Epi proColon®, ColoVantage® Plasma, RealTime mS9™ kit) (105-109,213-215), and a stool-based test for methylated vimentin (ColoGuard® assay, ColoSure™ assay) (48,110) (Table 2).

The association between CIMP and risk of CRC recurrence has been analyzed extensively, but results remain inconclusive. Several studies indicate CIMP-H as a poor prognostic factor in MSS but not MSI tumors (216,217), while CIMP-L has been suggested to be an indicator of poor outcome regardless of MSI (216,218,219). An association between CIMP and shortened survival was also reported in advanced CRC patients, among whom the contribution of MSI is relatively limited (220). However, there is evidence that the adverse effects associated with CIMP status may be attributable to BRAF mutation (205,221,222). Global hypomethylation as measured by analysis of LINE-1 elements has also been associated with poor outcomes, but
data are limited (223,224). Several studies have investigated small numbers of methylated candidate loci not included in CIMP marker panels identifying some evidence for prognostic associations (210,225-228), but no genome-wide methylome studies have been reported.

Epigenetic signatures are increasingly being considered in the context of response to chemotherapeutic and target agents. Recently, Ha et al. correlated genome-wide methylation array data with histopathological rectal tumor regression grade, highlighting hypomethylation of KLHL34 as a candidate predictive marker for sensitivity to preoperative chemoradiation therapy (229). Miyaki et al. related DEXI hypermethylation and transcriptional silencing, identified by genome-wide methylation sensitive amplified fragment length polymorphism (MS-AFLP) analysis, to resistance of camptothecin 11 (CPT-11) based chemotherapy via inhibition of apoptosis (230). Integrating gene expression microarray analysis and methylation-specific PCR, Tan et al. identified PPP2R2B hypermethylation and transcriptional silencing as a modulator of PDK1-directed Myc signaling and rapamycin sensitivity in CRC (231). CIMP status has been assessed in the context of 5-FU-based adjuvant chemotherapy, but results have not been conclusive. Some investigators have found that 5-FU treatment increases survival in patients with CIMP-H CRC (205,232,233), but others have not replicated this finding (234).

The CRC miRNAome

MicroRNAs (miRNAs) are short (19 to 25 nucleotides), double-stranded, non-protein coding RNAs, that regulate expression of complementary mRNAs at the post-transcriptional level by inducing mRNA degradation or blocking translation into protein. Abnormal miRNA expression profiles are related to clinical and biological behavior of tumors and, given their high stability, have been investigated as robust diagnostic, surveillance, prognostic and predictive biomarkers in cancer tissues and body fluids from cancer patients (235,236). Genomic approaches have mainly utilized qRT-PCR and microarray technologies.

To date, multiple studies have reported unsupervised principle component or cluster analyses of miRNA expression data to classify CRC. Oberg et al. analyzed 315 normal colonic mucosa, tubulovillous adenoma, MSS/ proficient mismatch repair (pMMR) sporadic carcinoma, and MSI/deficient mismatch repair (dMMR) sporadic and inherited carcinoma samples using microarrays (237). Unsupervised analysis demonstrated that normal colon tissue, adenomas, MSS/pMMR carcinomas and MSI/dMMR carcinomas were clearly discernible. Consistent with these data, several other studies analyzing MSS/pMMR and MSI/dMMR cancers also found miRNA expression differences between these tumor groups (238-241). One report suggested that Lynch syndrome tumors may display a different miRNA profile as compared to sporadic MSI tumors (242), but this was not noted by Oberg et al. (237). However, overlap between MSI/dMMR associated genes identified across studies is limited. One supervised analysis has suggested miRNA expression differences by CRC location, CIMP, KRAS and TP53 status although this has not been replicated (241). A recent microarray study on 1,141 CRC cases, analyzing 121 miRNAs previously reported with advanced tumor stage and/or survival, verified stage associations for five miRNAs (hsa-miR-145-5p, hsa-miR-31-5p, hsa-miR-200b-3p, hsa-miR-215 and hsa-miR-451a) (243).

miRNA signatures in the blood or stool of CRC patients have been evaluated as an alternative to FOBT testing for CRC diagnosis. Multiple studies have used separate discovery and validation cohorts to derive diagnostic miRNA blood/stool profiles using qRT-PCR panels or microarrays (244-255). Proposed classifiers comprise 1 to 21 miRNAs, with sensitivities of 34-85% and specificities of 68-97% reported across studies. In particular, up-regulation of miRNAs miR21 and miRNA92/miRNA92a have been highlighted in several blood- and stool-based studies for CRC diagnosis (236). To date, signatures have not been validated in independent follow-up reports or been rigorously compared against FOBT testing. Despite these caveats, miRNA signatures show promise as non-invasive CRC biomarkers.

Tumor miRNA signatures have been studied to predict prognosis using qRT-PCR, microarrays and NGS approaches. Several prognostic miRNA signatures for stage I-IV CRC patients have been proposed with little overlap between classifiers (243,256,257). One international study identified a 2-miRNAs classifier for predicting recurrence risk in MSS stage II-III CRC using NGS (258). The most comprehensive discovery and validation study to date has been reported by Zhang et al. (259). Using microarrays, a panel of 35 miRNAs was identified as differentially expressed between 40 paired stage II colon cancer tumors and adjacent normal tissues, and validated in independent samples from 138 patients. Based on these candidate genes, a six-miRNA prognostic classifier (miR-21-5p, miR-20a-
DNA sequencing and copy number studies in CRC have validated established genetic pathways of tumorigenesis and mutator phenotypes, while highlighting extensive mutational heterogeneity and identifying novel cancer gene candidates. Gene expression and DNA methylation data have demonstrated widespread deregulation of the CRC epigenome and indicate the importance of the cell-of-origin, retention of differentiation hierarchy and tumor stroma for CRC molecular classification (87,92,279,280).

Advances in genomics have begun to contribute new tools for clinical diagnosis and management of CRC. Blood- and stool-based tumor DNA sequencing, miRNA detection and DNA methylation assays are being developed for improved population screening, to facilitate surveillance of tumor recurrence and for dynamic monitoring of cancer response to therapy (207,281,282). Direct genomic and transcriptomic analyses of patient tumors are being pursued to provide prognostic and predictive information about the course of disease and benefit of treatment, with the current standard of care already involving assessment of KRAS mutation prior to treatment of metastatic CRC with anti-EGFR antibody therapy (283). Germline pharmacogenomic variation, which we did not consider in this review, further has the potential to predict patient treatment tolerance in order to avoid deleterious side effects (284).

Challenges for translation of genomic-based CRC biomarkers include the need for well-defined clinically characterized cohorts and for standardization regarding specimen collection, handling, and storage. Biomarker translation may further be improved through integration with functional genomics approaches to establish mechanistic rather than correlative links with tumor biology (285). Besides inter-tumor molecular heterogeneity, intra-tumor molecular heterogeneity poses a major hurdle to the translation of genomics findings and remains to be fully elucidated. Efforts focusing on molecular profiling of tumor regional heterogeneity and (epi-) genomic variation between metastatic deposits are ongoing. Besides clonal heterogeneity, hierarchical organization and phenotypic plasticity may play clinically important roles and will be subject of future genomic studies (286,287).

The application of genomic approaches, in particular whole exome sequencing, presents issues beyond the assessment of molecular alterations related to the patient's original presentation of CRC. Given the comprehensive nature of these tests, incidental findings on clinically relevant variants in genes with no relationship to the primary diagnosis may be made. This raises questions as to
whether such findings should be reported back to patients, what method of reporting should be used, and when to disclose these results (288-290).

The revolutionary advances in genomic technologies are enabling the possibility of personalized medicine for CRC. Evolving platforms such as NGS and high-density microarrays are starting to bring precision genomic profiling to the clinic at a reasonable cost. Ongoing innovations in existing applications and clinical informatics algorithms, as well as the many emerging technologies, will continue to advance translational cancer genomics and ultimately contribute to improving patient outcomes.

Acknowledgements

Funding: This work was supported by the Ludwig Institute for Cancer Research, an Australian Rotary Health/District 9780 PhD Scholarship to MP and a National Health and Medical Research Council of Australia (NHMRC) R.D. Wright Biomedical Career Development Fellowship to OMS (APP1062226).

Disclosure: The authors declare no conflict of interest.

References


© Translational Cancer Research. All rights reserved. www.thetcr.org Transl Cancer Res 2015;4(3):235-255


node-negative, invasive colorectal carcinoma. Cancer 2012;118:5234-44.


282. Ganepola GAP, Nizin J, Rutledge JR, et al. Use of blood-based biomarkers for early diagnosis and
surveillance of colorectal cancer. World J Gastrointest Oncol 2014;6:83-97.


Cite this article as: Cheasley D, Jorissen RN, Liu S, Tan CW, Love C, Palmieri M, Sieber OM. Genomic approach to translational studies in colorectal cancer. Transl Cancer Res 2015;4(3):235-255. doi: 10.3978/j.issn.2218-676X.2015.05.02