Genetic prognostic and predictive markers in colorectal cancer

Axel Walther[*‡§], *Elaine Johnstone*[‖], *Charles Swanton*[*§], *Rachel Midgley*[‖], *Ian Tomlinson*[‡] *and David Kerr*[‖¶]

Abstract | Despite many studies of the likely survival outcome of individual patients with colorectal cancer, our knowledge of this subject remains poor. Until recently, we had virtually no understanding of individual responses to therapy, but the discovery of the KRAS mutation as a marker of probable failure of epidermal growth factor receptor (EGFR)-targeted therapy is a first step in the tailoring of treatment to the individual. With the application of molecular analyses, as well as the ability to perform high-throughput screens, there has been an explosive increase in the number of markers thought to be associated with prognosis and treatment outcome in this disease. In this Review, we attempt to summarize the sometimes confusing findings, and critically assess those markers already in the public domain.

Great strides have recently been made towards improving the survival outcome of patients with early and advanced colorectal cancer (CRC), but it remains a major health burden with over one million cases worldwide and a disease-specific mortality of approximately 33% in the developed world[1].

Despite the overall improvements in CRC therapy, our understanding of why individual patients respond to therapy and others do not, and why some patients relapse, whereas others do not, remains poor. Consequently, there is a large, unknown subset of patients who receive treatment from which they do not benefit. In the metastatic setting, this failure to benefit from treatment is easy to judge by demonstrating tumour progression using clinical assessment and radiological imaging. In the adjuvant setting, many patients must be treated, with significant attendant toxicity[2], so that a few might benefit, as there are clearly patients who would not have relapsed even without adjuvant therapy[3]. Understanding the reasons for treatment failure and developing an ability to predict those who would benefit the most (and least) remain important aims in the management of CRC.

Currently, the gold standard for prognostication remains clinicopathological staging. The recommended staging system is the TNM classification, although systems based on that proposed by Dukes (Dukes staging system) in 1932 (REF. 4) are also used. Histopathological examination of tumour material can help to define prognosis further[5], using lymphovascular invasion, resection margins and tumour grade. In addition, some clinical parameters independently influence outcome: obstruction and perforation at presentation[6], performance status[7] and pre-operative carcinoembryonic antigen levels[8], which probably reflect tumour burden at surgery. Although clinicopathological staging separates patients into groups with distinct outcomes, importantly it offers little information about response to treatment in individual patients.

Several protein and genetic markers have been described in an attempt to refine prognostic information and predict the benefit derived from systemic treatment. This could help to avoid the toxicities associated with systemic therapy in those patients who will not benefit from this treatment. None of these markers is in routine clinical use[9], but progress has been made in other tumour types towards such goals. In breast cancer, for example, hormone receptor status is associated with prognosis. There is also a clearer understanding of which patients benefit from trastuzumab therapy, and a gene expression signature conveying worse prognosis[10] has been approved by the US Food and Drug Administration to support clinical decision making[11]. Although the CRC community has lagged behind other cancers somewhat, there is now a growing knowledge base of analogous determinants and an increasing realization that well-designed trials are required to plot an approval pathway for the regulatory authorities.

In this Review we examine the range of prognostic markers and predictive markers in CRC, assess the clinical usefulness of those already published and discuss future strategies for discovery, with a focus on DNA- and RNA-based markers.

*Department of Medicine, Royal Marsden Hospital, Downs Road, Sutton, SM2 5PT, UK.
‡Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK.
§London Research Institute, Cancer Research UK, 44 Lincoln's Inn Fields, London, WC2A 3PX, UK.
‖Department of Clinical Pharmacology, University of Oxford, Old Road Campus, Oxford, OX3 7DQ, UK.
¶Sidra Medical and Research Center, Qatar Foundation, PO BOX 26999 Doha, Qatar.
Correspondence to D.K.
e-mail: dkerr@sidra.org
doi:10.1038/nrc2645
Published online 18 June 2009

At a glance

- The most studied markers of colorectal cancer prognosis and response to therapy are somatic mutations in *KRAS*, adenomatous polyposis coli (*APC*) and *TP53*. With the exception of *KRAS* mutations and their association with clinical resistance to epidermal growth factor receptor (EGFR)-specific antibody therapy, there is no compelling evidence that these markers have a role in clinical decision making.
- Chromosomal instability is associated with a worse prognosis, and microsatellite instability with a better prognosis.
- Germline polymorphisms have been described in the metabolic pathways of chemotherapeutic agents used in colorectal cancer — for example, 5-fluorouracil (5-FU) and irinotecan — which correlate with the degree of toxicity.
- High-throughput expression and genotyping arrays are starting to generate novel markers and gene signatures that may be of use in the management of colorectal cancer. At present, these are not sufficiently validated to be clinically useful.
- Linking the collection of tissue and germline DNA to well-designed clinical trials will increase our understanding of the mechanisms of poor prognosis, and with it our capacity to identify novel biomarkers.

Hypothesis-driven markers

Currently, the most studied markers are somatic (acquired) markers for which some biological rationale exists for a potential effect on cancer outcome. These are changes associated with tumour progression in the adenoma–carcinoma sequence model proposed by Vogelstein *et al.*[12] (FIG. 1) or the observed type of genomic instability (chromosomal or microsatellite). Germline (inherited) changes have mostly been studied in pharmacological pathways involved in the metabolism and mechanism of action of 5-fluorouracil (5-FU), the therapeutic mainstay for CRC. These studies are all hypothesis-driven, with often limited a priori evidence to suggest a link to prognosis.

KRAS. Many studies have evaluated *KRAS* mutations in exon 2 (codons 12 and 13), and to a lesser extent in exon 3 (codon 61), for their association with CRC outcome. *KRAS* mutations are an early event in the adenoma–carcinoma sequence, although they are demonstrable in only approximately one-third of CRCs, with a large majority located in codon 12 (REF. 13). Mutations in each of the three codons compromise the ability of GTPase-activating proteins to effect the inactivating hydrolysis of Ras-bound GTP to GDP[14]. Other mutations are uncommon as they result in lower constitutive Ras signalling than mutations in codons 12, 13 and 61 and so are selected against in the tumour[15].

Although the largest international effort to combine data from different groups, the RASCAL collaborative, found that *KRAS* mutations generally confer a worse prognosis[13], the subsequent analysis (which included a further 700 patients) found that only the glycine to valine substitution at codon 12 was associated with a poorer prognosis, and this was found only in American Joint Committee on Cancer (AJCC staging system) stage III patients[16]. Other large studies have found no association[17–20], and there is currently no convincing evidence to suggest that *KRAS* mutations are independent prognostic factors in CRC.

More recently, however, *KRAS* mutation status has been established as a predictive marker for treatment with epidermal growth factor receptor (EGFR) inhibitors. This interaction was initially observed with small-molecule inhibitors of EGFR in non-small-cell lung cancer[21]. However, several studies in CRC have now shown that, owing to the convergence of the EGFR and KRAS pathways (FIG. 2), patients with stage IV *KRAS*-mutant CRC being treated with the EGFR-specific antibodies cetuximab[22–26] and panitumumab[27] derive considerably less benefit than patients with wild-type *KRAS* tumours. Further, in 80 *KRAS* wild-type patients treated with cetuximab, 11 patients with the *BRAF* V600E mutation (previously named V599E[28]) did not respond to the treatment[29]. *BRAF* acts downstream from *KRAS*, and mutations in the two genes seem to be mutually exclusive[30]; therefore, *BRAF* could be a locus for a second hit affecting the same pathway, and both genes are important determinants of resistance to EGFR-specific therapies.

KRAS is almost an ideal predictive biomarker: mutations are limited to a small part of the gene and are easily detected, the negative predictive value is high (99% of patients with mutated *KRAS* do not respond to EGFR inhibition[22]) and the effects of the mutations are based on a plausible biological rationale. Taking *KRAS* as an example further demonstrates how our evolving knowledge of cancer biology can refine treatment strategies and interventions, but also highlights the difficulties that the regulatory agencies will face when evaluating EGFR-specific or other targeted agents in response to retrospective clinical studies.

Adenomatous polyposis coli and β-catenin. Adenomatous polyposis coli (*APC*) acts to promote the degradation of β-catenin and so limits the transcription of Wnt target genes involved in regulating the cell cycle. This pathway is integral to colorectal tumorigenesis, and more than 90% of patients have alterations that affect it[31]. Given the frequency of changes, it is not surprising that neither APC nor β-catenin is a useful prognostic marker able to differentiate between patients. Although the specific type of APC mutation could hold prognostic information — for example, mutations that abolish β-catenin-binding sites of APC may be associated with poorer prognosis[32] — addressing specific mutations in clinical practice could be technically difficult because of the large number of APC mutations already described. Testing for general overexpression of β-catenin does not seem to be useful, but determining the cellular location of overexpressed β-catenin may hold prognostic information[33,34], even if this overexpression is likely to be a surrogate marker for a genetic change elsewhere in the β-catenin degradation pathway. Changes in APC and β-catenin are therefore insufficiently validated in patients and currently have no role in clinical practice.

TP53. Loss of heterozygosity at chromosome 17p frequently occurs in CRC, and many studies have focused on the 17p region containing the *TP53* tumour suppressor gene. *TP53* has been investigated as both a prognostic factor and a predictor of response to therapy, with conflicting results. The methods used to assess the mutation of *TP53* vary greatly between studies, as do study designs

Dukes staging system
A staging system based on the depth of invasion of the primary tumour and the presence of lymph node metastasis. Originally described only for rectal cancer, it did not include distant metastasis or unresectable tumours, although both have been addressed in modifications since.

Prognostic marker
A marker that provides information about the natural history of the disease.

Predictive marker
A marker that provides information about the likelihood of response to a treatment.

AJCC staging system
Stage groupings based on the TNM system: depth of primary tumour invasion (T stage), presence and number of lymph node metastases (N stage) and presence of distant metastasis (M stage).

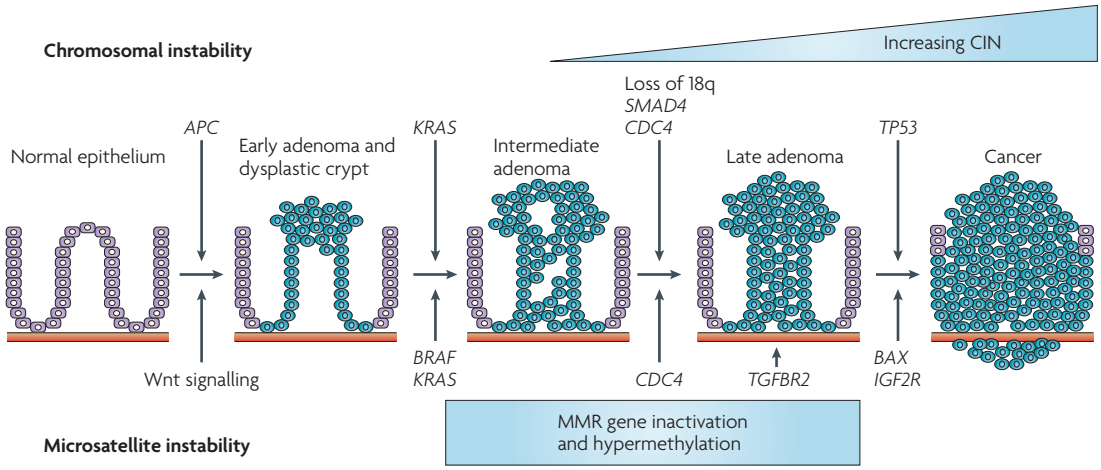

Figure 1 | **Adenoma–carcinoma sequence model for chromosomal instability in colorectal cancer.** This model is likely to be an oversimplification, but it aligns observed clinicopathological changes with genetic abnormalities in the progression of chromosomally unstable colorectal cancer (CRC)[12] (the gatekeeper pathway involving genes that regulate cell growth[134]). The initial step in tumorigenesis is that of adenoma formation, associated with loss of adenomatous polyposis coli (*APC*). Larger adenomas and early carcinomas acquire mutations in the small GTPase *KRAS*, followed by loss of chromosome 18q with *SMAD4*, which is downstream of transforming growth factor-β (TGFβ), and mutations in *TP53* in frank carcinoma. Microsatellite instability (MSI^+) CRCs, characterized by a deficiency of the mismatch repair system that leads to slippage in microsatellites (the caretaker pathway involving genes that maintain genomic stability[134]), only carry the above changes infrequently; therefore, development of CRC must involve different, but analogous, genetic changes to those described in chromosomal instability ($CIN)^+$ CRC. MSI is uncommon in adenomata[135], and the initial step is thought to involve alteration in Wnt signalling[136], possibly involving axin[137]. Mutations in *BRAF*, common in MSI^+ CRC, are likely to occur in the place of *KRAS* mutations[30], although the latter do occur in a minority of cases. Mismatch repair (MMR) deficiency in sporadic CRC occurs predominantly by downregulation of MLH1 through promoter methylation[138], and MSI status is increased by positive selection of tumour cells with mutated microsatellites in MSH3 and MSH6 (REF. 139). Further positive selection occurs for mutations affecting microsatellites in TGFβ receptor 2 (*TGFBR2*)[140], insulin-like growth factor 2 receptor (*IGF2R*)[141] and *BAX*[142], which in turn provides a *TP53*-independent mechanism of progression to carcinoma[142]. *FBXW7* (F box and WD40 domain protein) inactivation may precede *TP53* mutation[143], leading to increasing CIN[144], although it is not always associated with CIN and may also have a role in the MSI pathway[145].

in terms of reported clinicopathological data and patient selection, making it difficult to draw firm conclusions about the prognostic value of p53 (REFS 35,36).

Loss of 18q. Deletion of the long arm of chromosome 18 is the most common cytogenetic abnormality in CRC and has been associated with poor prognosis[37]. However, as above, this is not a uniform finding[38,39]. Many studies have investigated the genes found in this region as prognostic markers, in particular deleted in colorectal carcinoma (*DCC*) without demonstrating a clear link to prognosis[40]. *SMAD4*, a known CRC predisposition gene that is also located on 18q, is a member of the transforming growth factor-β (TGFβ) signalling pathway, and decreased *SMAD4* mRNA levels seem to be associated with a worse prognosis[41] and poorer response to 5-FU[42]. However, not all studies have found a relationship between the loss of 18q and *SMAD4* expression[43]; therefore, these studies have failed to definitively link any gene on 18q to prognosis. Furthermore, it seems likely that loss of 18q is a marker of chromosomal instability (CIN)[44] and therefore not an independent prognostic marker, whereas the other common mutation in the TGFβ pathway in patients with CRC (mutation of TGFβ receptor 2) is almost exclusively linked to microsatellite instability (MSI)[45].

Overall, somatic markers have proved to be of limited prognostic value in patients with CRC. This might reflect the fact that many of these changes are associated with molecular phenotype, and most (for example, mutations in *KRAS*, *APC* and *TP53*, and loss of 18q) are much less common in patients with MSI^+ CRC[45,46]. Therefore, any somatic marker studied must be evaluated in the light of possible confounding by genomic instability.

Genomic and epigenomic instability

Two major types of genomic instability are recognized as alternative mechanisms of colorectal carcinogenesis. The more common, CIN (or aneuploidy), is present in approximately 65–70% of CRCs and is poorly defined as the presence of multiple structural or numerical chromosome changes in tumour cells. Around 15% of CRCs have a near-diploid chromosome set and MSI, defined as a tumour having instability in at least two of five standard microsatellite markers[47]. The prognostic value of CIN and MSI is no longer in question: both have been subject to large meta-analyses[48,49], which unequivocally established that patients with CIN^+ disease have a poorer prognosis (hazard ratio (HR) for death = 1.45) and patients with MSI^+ CRC have a better prognosis (HR = 0.65) than patients with CIN^- and MSI^- CRC,

respectively. Strikingly, despite some studies not reporting significant associations, the direction of the effect is the same in virtually all contributory data.

Less clear is the prognostic relationship between CIN and MSI. The traditional view that CIN and MSI are mutually exclusive, and that all CRCs are either one or the other, is probably incorrect (FIG. 3). Although CIN and MSI may carry separate prognostic information, the only published study to date to assess the effect of both MSI and CIN in multivariate analysis did not find that the effect of MSI on survival was independent to that of CIN[50].

CRCs also harbour epigenomic instability, either as global hypomethylation or as the CpG island methylator phenotype (CIMP), defined as methylation at three or more specific marker loci[51]. CIMP has considerable overlap with MSI and seems to be associated with MSI in patients who do not have germline mutations in mismatch repair (MMR) genes, which are characteristic of hereditary non-polyposis colon cancer (also known as Lynch syndrome)[51]. In addition, *BRAF* V600E mutations are strongly associated with MSI$^+$ and CIMP$^+$ disease. Patients with this phenotype have been variably termed CIMP-high (CIMP-H)[20,52] or CIMP 1 (REF. 53) and have a better prognosis[20]. A further group of patients with a CIMP$^+$ phenotype is associated with *KRAS* mutations and has been termed CIMP-low[54] (CIMP-L) or CIMP 2 in response to the (not universally reported[52]) lower levels of methylation[53]. By contrast, MSI$^-$ CRCs, which are CIMP$^+$ and harbour *BRAF* V600E mutations, are associated with a worse prognosis[55], and increasing levels of CIMP (negative to low to high) confer worsening prognosis in MSI$^-$ CRC[52]. Despite the finding in one study that MSI-associated prognostic information was not independent of CIMP status[20], the relative contributions of CIN, MSI and CIMP to survival outcome need to be further investigated to understand the effects and interactions of these variables.

Decreased methylation in the form of global genomic hypomethylation is also involved in colorectal carcinogenesis[56] and is associated with CIN[57,58], representing the opposite end of the methylation spectrum to CIMP$^+$ MSI$^+$ CRC. CRC with genomic hypomethylation was associated with worse prognosis in one study, but this analysis did not include CIN status[57] and, again, the interplay of the various types of genomic and epigenomic instability is not clear.

The consequences of each type of instability at the gene level remain poorly characterized, but it is now reasonable to use CIN and MSI status in clinical trials to stratify patients. CIMP and global hypomethylation should be evaluated further in retrospective series that are also determining CIN, MSI and *BRAF* status.

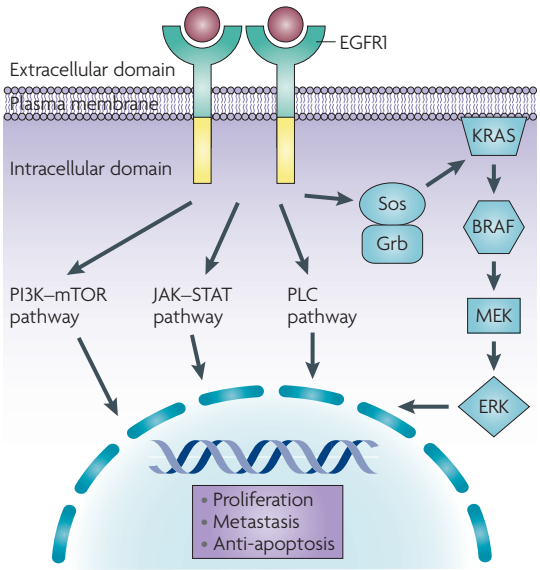

Figure 2 | **The epidermal growth factor receptor signalling pathway.** On ligand binding, the epidermal growth factor receptor (EGFR) type 1 homodimerises, leading to the activation of the intracellular kinase domain. Through the small adaptor proteins Sos and Grb, the KRAS signalling cascade is activated, leading to increased proliferation. Part of the KRAS pathway is BRAF, which explains why non-constitutively activated KRAS and BRAF are necessary for EGFR blockade to work. The initial assumption that EGFR overexpression, as is common in colorectal cancer[146], is also required is challenged by a lack of correlation between EGFR expression and EGFR-specific antibody response[147] and by the observation that patients with irinotecan-refractory, EGFR-negative colorectal cancer can still respond to irinotecan and cetuximab[148]. JAK, Janus kinase; PLC, phospholipase C; STAT, signal transducer and activator of transcription.

MSI, CIN and drug sensitivity

Since the initial publication of a detrimental effect of adjuvant 5-FU therapy in loco-regional MSI$^+$ CRC[59], further data have been published by the same group confirming their earlier findings[60]. The clinical data are supported by *in vitro* evidence showing that a functioning MMR system is required for cytotoxicity in response to 5-FU incorporation into DNA[61], in addition to its effects mediated by inhibition of thymidylate synthetase (see below). By contrast, other retrospective data suggest that patients with MSI$^+$ CRC do benefit from adjuvant 5-FU administration[62]. This is supported by the observation that in stage IV CRC, 5-FU treatment is effective in MSI$^+$ CRC[63]. The conflicting data demonstrate the difficulty in examining uncommon subgroups. However, the evidence for a detrimental effect of 5-FU treatment is sufficiently strong for a clinical trial (E5202) to be undertaken to compare MSI$^+$ CRC without adjuvant chemotherapy to MSI$^-$ CRC with adjuvant chemotherapy. This clinical trial, which is now underway, aims to prospectively resolve these conflicting data.

In addition, there are data to suggest that MSI$^+$ CRCs are more sensitive to irinotecan-based regimens in some clinical series[64,65] and cell line studies[66]. The increased survival following irinotecan treatment in MSI$^+$ CRC was also observed in the adjuvant setting[64], in which irinotecan had failed to benefit unselected patients[67], making it an attractive proposition for prospective study in this setting.

CIN could also be a negative predictive marker for the response to taxanes, drugs that showed limited activity in CRC during their initial development. An intact spindle

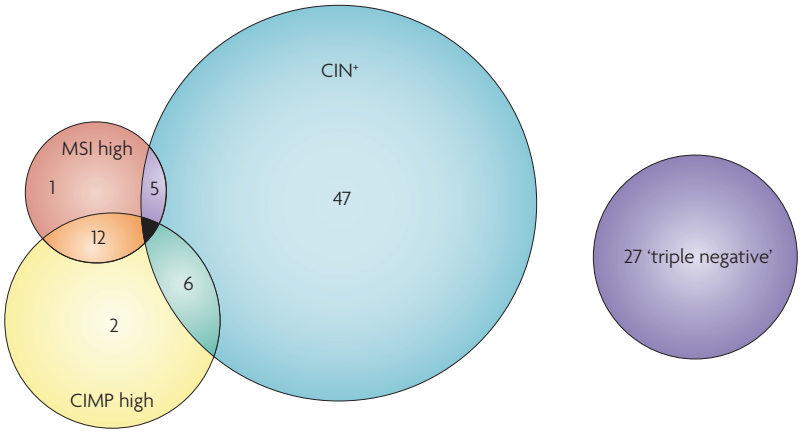

Figure 3 | **Venn diagram of chromosomal instability, microsatellite instability and CpG island methylator phenotype.** Approximately 17% of patients display microsatellite instability (MSI)[49], approximately 60% chromosomal instability (CIN)[48] and approximately 20% CpG island methylation phenotype (CIMP)[20,51,52,149]. Around one-quarter of MSI$^+$ colorectal cancers (CRCs) are CIN$^+$ (REF. 50). Only one study has addressed the intersection of all three forms of instability[150]. The best estimate of the distribution of the CIMP phenotype relative to CIN and MSI is that it accounts for most of MSI$^+$, CIN$^-$ CRC (the sporadic MSI$^+$ CRC)[51] and approximately 12% of all CRC. CIN is probably not associated with CIMP, and CIMP is therefore proportionally spread over the CIN$^+$ and CIN$^-$ groups. CIN$^+$ MSI$^+$ CRC might also be CIMP$^+$, but there is no evidence in the literature for this, which probably reflects the small size of this group[150]. The subset of CRC that displays no genomic instability might be smaller than shown here as flow cytometry (used to identify CIN) is relatively insensitive and a proportion of these cancers might display minor chromosomal abnormalities. It is likely that many of the 'triple negative' CRCs would become overtly CIN$^+$, but changes in CIMP status over time cannot be ruled out at present. All numbers are percentages of the overall number of patients.

assembly checkpoint is required for taxane sensitivity. In other words, only diploid cells segregate chromosomes normally and are sensitive to paclitaxel[68,69], a hypothesis that is currently being tested in a Phase II clinical trial in our institutions.

Pharmacogenetics of 5-FU and capecitabine

Inherited polymorphisms have the potential to greatly affect treatment: for conventional chemotherapy agents with a narrow therapeutic window, subtle genomic changes can modulate drug-specific pharmacokinetics or pharmacodynamics and substantially affect individual responses and toxicity after chemotherapy. Almost all chemotherapy regimes for CRC incorporate 5-FU or its oral prodrug, capecitabine, and the ability to tailor therapy to take into account variation in their metabolism has the potential to translate into beneficial effects for a large patient population (FIG. 4).

Thymidylate synthetase. Thymidylate synthetase (TS; also known as TYMS) is thought to be the dominant target for the active principle of 5-FU, fluorodeoxyuridine monophosphate (5-FdUMP). A meta-analysis of TS expression suggested that higher expression of TS is associated with a poorer overall survival rate[70]. Two main genetic determinants of TS expression have been described: a variable number tandem repeat polymorphism in the *TS* promoter–enhancer region (TSE) and a 6 base pair (6 bp) insertion and deletion polymorphism in the 3′-untranslated region of *TS*. The tandem repeat sequence is 28 bp, and is usually present as either two or three repeats, although four and more repeats have been described[71]. An increase in the number of repeats leads to increased efficiency of mRNA translation and expression[72]. In addition, the G allele of a G–C single nucleotide polymorphism (SNP) when present in the second repeat of a three repeat sequence correlates with even higher mRNA expression through conservation of an upstream stimulatory factor binding site in which the SNP lies[73]. The 6 bp deletion polymorphism in the 3′-untranslated region alters mRNA stability[74] and is associated with low mRNA expression[75]. The high expression variants have been associated with decreased survival in patients treated with 5-FU[71,76]. These three markers may in combination predict which patients have an increased risk of recurrence in stage II and stage III colon cancer[77], reflecting a biologically aggressive phenotype that would also be resistant to 5-FU-based chemotherapy.

Dehydropyrimidine dehydrogenase. More than 80% of 5-FU is catabolised by dehydropyrimidine dehydrogenase (DPD; also known as DPYD) but levels of DPD activity can vary widely between individuals: 3–5% of the population is partially, and 0.2% completely, DPD deficient. There are more than 30 polymorphisms that cause DPD deficiency[78], leading to severe, sometimes life-threatening, toxicity after 5-FU treatment[79]. The lower toxicity associated with modern infusional or oral 5-FU-based regimens makes it impractical to screen the entire population for 30 polymorphisms a priori. Despite the obvious and logical effect on toxicity, its prognostic or predictive value is less clear: some studies report an association of mRNA expression levels with prognosis, whereas others do not[80], possibly reflecting differences in the clinicians' responses to the encountered toxicity. Despite extensive investigation, the pharmacogenetic basis of varied DPD activity remains to be fully elucidated, although advances in high-throughput sequencing techniques may make the pre-treatment prediction of 5-FU toxicity achievable[81].

Methylenetetrahydrofolate reductase. Reduced methylenetetrahydrofolate reductase (MTHFR) activity creates variation in folate pools, indirectly increasing sensitivity to 5-FU. Two common polymorphisms in *MTHFR* lower enzyme activity[82]: the C677T polymorphism leads to a change of alanine to valine at position 222 and A1298C results in a glutamine to alanine change at position 429. Increased response to 5-FU treatment has been associated with the 677 T allele[83,84] and, to a lesser extent, the 1298 C allele. Recent studies suggest that these polymorphisms affect capecitabine toxicity[85], as well as the efficacy of 5-FU[86]. However, clinical data do not unequivocally support the influence of the *MTHFR* genotype on 5-FU responsiveness, toxicity and patient clinical outcome[87].

Pharmacogenetics of oxaliplatin and irinotecan

Oxaliplatin. There is some evidence to suggest that genetic polymorphisms in the genes that encode detoxifying enzymes and DNA repair proteins have an important role in determining the treatment response to the

Quantitative marker
A continuous marker that can change its value many fold; for example, gene expression, which increases power for detection with fewer samples, provided there is no noise in the signal.

Discrete marker
A marker that can take one of several forms — for example, a single nucleotide polymorphism — and which may therefore require more samples than a quantitative marker for the same power, but is often more robust in its determination.

DNA binding agent oxaliplatin. Decreased sensitivity to platinum agents has been attributed to diminished cellular drug accumulation, increased intracellular drug detoxification and increased DNA repair[88].

Glutathione-*S*-transferases (GSTs) are phase II detoxification enzymes that target a wide variety of drugs for excretion by conjugation with glutathione. There are several isoenzymes and polymorphisms in this class of enzymes, with varying specificities, activities and tissue localizations[89]. The isoenzyme GSTP1 is the primary enzyme for the detoxification of platinum derivatives. Two missense polymorphisms in *GSTP1* that result in Ile105Val and Ala114Val lead to decreased GSTP1 activity and predict neuropathy after oxaliplatin treatment[90]. There is no evidence for any isoenzymes or alleles (including null alleles) being associated with prognosis.

The primary anti-tumour mechanism of platinum derivatives is the formation of DNA adducts, which interfere with DNA replication and require the activity of DNA repair enzymes to avoid cell death. Several polymorphisms in different DNA repair enzymes have been shown to correlate with function[91]; however, association studies with outcome seem to be regimen and cancer type specific. Of six commonly studied functional polymorphisms in four DNA repair genes (*ERCC1*, *ERCC2*, *XRCC1* and *XRCC3*) only *ERCC1* Asn118Asn and *ERCC2* Lys751Gln were associated with overall survival in colorectal cancer in one report[92], but not in another report[93].

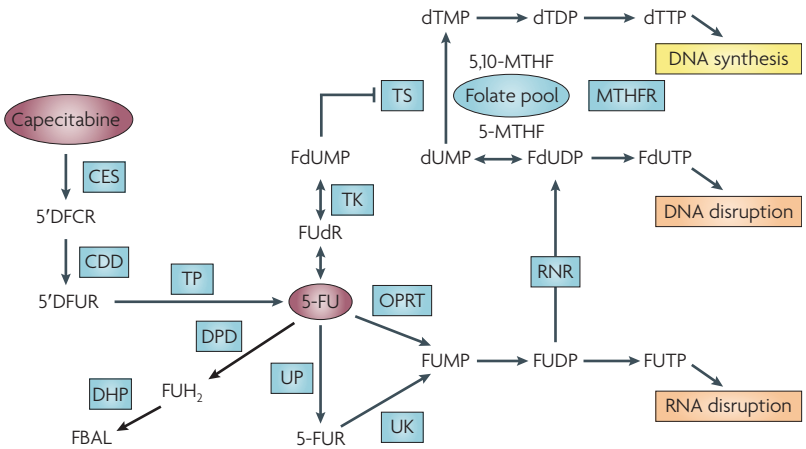

Figure 4 | **Pathways that affect 5-fluorouracil efficacy.** The anti-tumour effect of 5-fluorouracil (5-FU), mediated by thymidylate synthetase (TS) inhibition and its effect on RNA, is well established[151]. Incorporation of 5-FU derivatives into DNA can lead to cytotoxicity, but this may require an intact mismatch repair system to detect the incorporated fluorodeoxyuracil triphosphate (FdUTP)[61]. Variation in the enzymes that mediate the incorporation into RNA and DNA, the conversion of the oral prodrug capecitabine to 5-FU, or metabolism to inactive breakdown products can alter the intracellular 5-FU concentrations and cytotoxicity, leading to altered anti-tumour activity or systemic toxicity. For example, TS expression levels are associated with drug efficacy: colorectal cancer with high levels of TS seems to have a poorer overall survival than tumours with low TS expression[70]. In many instances, it is not clear what drives the changes in expression levels, although it is likely that underlying germline genetic variation or change may be responsible for a proportion of the observed variation. 5′DFCR, 5′deoxy-5-fluorocytidine; 5′DFUR, 3′deoxy-5-fluorouridine; 5-FUR, 5-fluorouridine; CDD, cytosine deaminase; CES, carboxylesterase; DHP, dihydropyrimidinase; DPD, dihydropyrimidine dehydrogenase; FBAL, fluoro-β-alanine; FUH_2, dihydro-5-fluorouracil; MTHFR, methylenetetrahydrofolate reductase; OPRT, uridine monophosphate synthetase; RNR, ribonucleotide reductase; TK, thymidine kinase; TP, thymidine phosphorylase; UK, uridine-cytidine kinase 2; UP, uridine phosphorylase 1.

Irinotecan. The active metabolite of the topoisomerase I inhibitor irinotecan, SN-38, is conjugated and detoxified primarily by UDP-glucuronosyltransferase (UGT1A1). The number of TA repeats in the TATA element in *UGT1A1* correlates with reduced enzyme expression and activity[94]. Individuals who are homozygous for the 7-repeat allele, also known as the UGT1A1*28 allele, commonly suffer dose-limiting neutropenia through decreased degradation and clearance of SN-38 (REF. 95). Initially, several groups reported this association, but their findings were heterogeneous in significance and effect size. A recent meta-analysis established that the incidence of toxicity in UGT1A1*28 patients was positively correlated with the dose used[96]. A commercial genetic test was approved by the US Food and Drug Administration in 2005 to make dosing recommendations based on *UGT1A1* genotypic results and avoid life-threatening neutropenia, a first for any chemotherapy agent. The routine use of this test by oncologists has been limited in clinical practice because haematological toxicity has decreased with the lower irinotecan doses used in combination regimens with 5-FU. The prognostic impact of UGT1A1*28 has not been established[97].

Unbiased high-throughput screening

High-throughput arrays for evaluating mRNA expression levels or single nucleotide polymorphisms have opened new avenues of marker discovery by moving from hypothesis-driven, targeted research to unbiased screening of the whole genetic spectrum. RNA-based research is aimed at quantitative markers, whereas DNA-based techniques analyse discrete markers. The bulk of biomarker research has focused on somatic changes, but germline variation can equally have an impact on prognosis and response to therapy, although currently only the few pharmacogenetic markers mentioned in the previous section have been described. For prognostic purposes, gene expression analysis is classically carried out on tumour tissue, although comparative analyses with normal surrounding mucosa have been published, whereas genome-wide DNA screens have focused on the germline, particularly in the search for disease susceptibility loci.

Gene expression signatures. Gene expression analysis holds great promise for the understanding of the functional differences between tumour tissue and normal tissue, and with it the hope of developing meaningful signatures that stratify patients beyond pathological staging. In CRC, much effort has been expended to repeat the perceived success in breast cancer by analysing both tumour tissue[98–106] and the surrounding mucosa[107–110]. Similar analysis has also been applied to the study of microRNA (miRNA) arrays in tumour tissue[111–113].

Test set
The set of patients in whom a hypothesis is generated, which is then validated in the validation set. The test set is necessary only in unbiased screening approaches, as hypothesis-driven discovery already has a hypothesis to test.

Validation set
The set of patients in whom the hypotheses generated in the test set are studied further, to prove or refute the validity of the initial hypothesis. This step is vital in unbiased approaches as otherwise the test set could merely report chance findings.

Linkage disequilibrium
Deviation from the association that would be expected between two genetic markers if they were subject to random recombination during meiosis, the likelihood of which (without linkage disequilibrium) is a function of the distance between the two markers.

The first published signature in CRC by Wang *et al.* identified a 23-gene prognostic signature based on 31 relapses in 74 stage II patients. This was subsequently subjected to validation both independently and by the same group. On the basis of 25 relapses in 50 stage II patients, the independent group found a positive predictive value of 67% (REF. 102), only marginally better than chance, and therefore proposed a separate set of 30 genes. Wang *et al.* found that using only 7 genes of the original test set performed better in their validation set[105], which this time was based on 123 patients and further validated in 104 patients. Similarly to other studies describing signatures based on differentially expressed genes (range 7 to 72), the overlap between the 30- and 7-gene signatures, even when in the same pathological stage, has been poor. One of the more convincing studies examined miRNA expression using 74 patients of all stages in the testing set and 110 patients in the validation set, and suggested a single miRNA (miR-21) as a prognostic marker. Again, this has not yet been validated further.

The reasons for the failure to find robust multi-gene signatures include that the studies were statistically underpowered, looking at sample sets as small as 20 patients but analysing the expression of tens of thousands of genes; that validation was not always carried out; and that correction for other prognostic variables was omitted from the main analysis, leading to overfitting of the prognostic model with poor reproducibility and a high false-discovery rate[114]. This problem is not unique to CRC, nor has the quality of studies improved with increased familiarity with the technology[115–117]. Concerns have been raised for the first such signature, which was described in breast cancer, over the inclusion of the training set in the initial validation set[118] and the use of patient numbers that were insufficient to derive a robust signature. Many different signatures could be derived from the same data, depending on the clustering parameters[119], and a sample set of at least 3,000 patients may be necessary to identify a unique, robust signature[120]. Efforts are now underway to generate a gene expression signature from fixed paraffin-embedded tumour tissue using quantitative real-time PCR[121]. The initial signature was generated in a set of over 500 samples and will be validated in a large (n = 2,000) cohort of patients with colon cancer enrolled in the QUASAR trial[122].

Meta-analysis is commonly carried out for other markers combining data sets to boost numbers (small data sets with patient numbers in the twenties and thirties to several hundred or one thousand patients[48,49]), but this is difficult and methodologically unreliable as expression signatures differ in experimental protocol in each study[123], leading to a high false-positive and false-negative discovery rate[124]. Given the undoubted potential, there must be a concerted international effort to find sample sets of sufficient size to allow the definition of a robust, reproducible signature in CRC as "small sample sizes might actually hinder the identification of truly important genes"[REF. 125], and the emphasis on the detection of bias and chance and the validation of the findings remain as important as ever[126].

Genome-wide association studies. High-density SNP array-based genome-wide association studies (GWAS) have been successful in identifying susceptibility loci for CRC[127,128], prostate[129] and breast cancer[130], as well as non-malignant disease[131,132]. They have relied on the premise that many diseases have a heritable component and that traditional linkage analysis using a collection of affected families will only identify rare, high-penetrance susceptibility alleles. GWAS attempt to define common alleles that confer only a moderate risk for the disease in question.

It is likely that host factors (the genetic make-up of normal tissue) can have an influence on outcome, either as prognostic or predictive markers. Genetic variation associated with genes involved in cell adhesion and motility may make early metastasis more likely, and cell cycle checkpoint-associated variation may make the cell more tolerant to genomic abnormality and thus more resistant to chemotherapy. Pharmacogenomic variation is likely to have a role in both host response (toxicity) and tumour response to chemotherapy. This is discussed above, but GWAS have the potential to uncover new variants affecting the metabolism and efficacy of anticancer agents.

Until all genetic variation has been described and can be tested simultaneously, GWAS necessarily rely on the association of a tested marker with the true or causal variant affecting outcome, although there is no inherent reason why the tested marker could not be the true variant. This means that all identified markers need to be followed up irrespectively of biological rationale, although if the marker identified is not in complete linkage disequilibrium with the true variant, replication would be more difficult. Once identified, however, routine clinical use would be easy, as the necessary DNA could easily be extracted from a standard blood sample and genotyping is robust.

It is possible to apply the same techniques to tumour DNA, although prognostic markers may arise from mutations in the malignant tissue and may not be captured by the SNPs analysed if these do not predispose to the mutation in the first place. Furthermore, the SNPs would not capture epigenetic events, which undoubtedly have a role in tumorigenesis and can also affect outcome. In malignant tissue, however, SNP typing platforms would give information about copy number changes (deletions, insertions and loss of homozygosity) that can affect gene expression (functional aneuploidy[133]).

To date, no GWAS looking at outcome, either in untreated or treated patients, have been published for any cancer. The main obstacle is that for the analysis of several hundred thousand SNPs, to retain a degree of statistical power, sample sets need to be large. Although it is easy to collect these sample sets for the analysis of risk, outcome requires follow-up for at least 3 years to assess progression-free survival with top-quality data in a clinical trial setting if at all possible.

GWAS of outcome in CRC are now underway, and one study has completed the initial screening phase of genotyping and testing the SNPs on the array for

their association with outcome in 950 patients. The 40 SNPs most strongly associated have been validated in further, independent sample sets of similar sizes. Although there is likely to be a (substantial) proportion of SNPs that do not replicate, several of the SNPs taken forwards do look promising on the basis of the biological function of the genes in which they reside (kinases involved in cell cycle progression or tumour invasion) or the functional class of these genes, with 10 of the top 40 genes involved in cell adhesion and motility. However, it is important to test all significant SNPs from the screening phase, including those in regions without any genes in the vicinity, to ensure an unbiased approach and avoid moving into the hypothesis-driven search for markers described in the first section of this Review.

Prospective clinical evaluation

As stated previously, it is possible to test and validate CRC biomarkers using archived tissue and DNA, ideally linked to a statistically informative clinical trial with a control group of sufficient size to allow random separation of material into test and validation sets. This allows internal validation of the biomarker with outcome (progression-free being preferred to overall survival) using retrospective data. This may be sufficient to assess the statistical underpinning of the biomarker, but does not provide any formal test of its clinical utility. This is best achieved in prospective clinical trials designed to measure improvements in outcome and health economic parameters; however, different trial approaches are required for prognostic and predictive tests.

Assuming that the prognostic marker would separate the population of patients who are potential candidates for post-operative adjuvant chemotherapy into three statistically distinct groups, patients with a low risk of recurrence would receive no further therapy (they would be deemed cured by surgery alone), those with an intermediate risk would be randomized to conventional chemotherapy or control, and patients at the highest risk of relapse would be entered into a trial comparing conventional chemotherapy with the next generation combination.

If a predictive marker defines a subpopulation of patients who are particularly sensitive to a specific therapeutic intervention, the first investigative step would be to demonstrate the predictive value prospectively in small Phase II clinical studies: patients would be stratified according to marker status. Assuming a 30% response rate in those deemed to be sensitive and ≤10% in those who are not sensitive, 20–40 patients in the former and 15–20 patients in the latter group would give initial confidence in the power of the predictive biomarker to justify a randomized Phase III trial. A trial schema for the Phase III trial that could be used to demonstrate the clinical use of a predictive biomarker is outlined in FIG. 5.

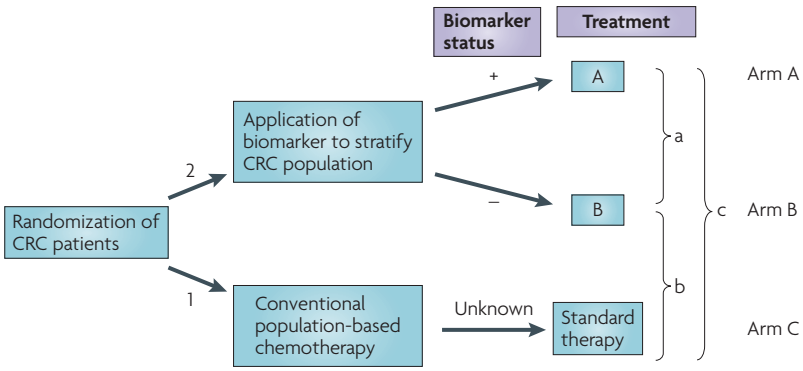

Figure 5 | **Trial schema to demonstrate the use of a predictive biomarker.** Patients with early- or late-stage colorectal cancer (CRC) could be randomized (in a 2 to 1 ratio) to stratification by the biomarker or to receive unselected or population-based treatment. All three groups of patients would receive the same drug (or drugs), giving the potential to test three hypotheses and link the clinical outcome data (disease or progression-free survival depending on the stage) to health and economic parameters (that treatment A equals treatment B equals standard therapy). On the basis of this several hypotheses can be tested. The first is that the biomarker-positive group has a superior clinical outcome compared with the biomarker-negative group following treatment with identical chemotherapy. The second hypothesis is that the conventional population-based group has a superior clinical outcome compared with the biomarker-negative group following treatment with identical chemotherapy. The third hypothesis that the biomarker-positive group has a superior clinical outcome compared with the conventional population-based group following treatment with identical chemotherapy. If sufficient evidence has been accumulated for a particular predictive marker, treatments A and B can differ to test the hypothesis that distinct treatments result in the same outcomes depending on marker status. A current clinical trial (E5202) in stage II CRC uses the latter approach: patients are screened for microsatellite instability (MSI) and loss of 18q (as a marker of chromosomal instability), and those who are MSI^+ (irrespective of 18q status) and those who have not lost 18q do not receive adjuvant chemotherapy (arm B), whereas MSI^- patients with loss of chromosome 18q receive FOLFOX (an infusion of 5-FU, leucovorin and oxaliplatin) chemotherapy (arm A). This trial does not have an arm C, but adds a further randomization to arm A: patients are given either FOLFOX alone or FOLFOX in combination with bevacizumab.

Conclusion

The encouraging improvement in patient outcome over the past 20 years has been followed by a plethora of markers of prognosis and response to anticancer therapy, most of which fail to demonstrate clinical utility. Many more DNA- and RNA-based markers of prognosis than those mentioned in this Review have already been described, often only once and in a small series. Clearly there has to be considerably better validation before any can be thought of as being associated with prognosis or the response to therapy and therefore incorporated into clinical decision-making algorithms. With an increasing understanding of tumour biology, more hypothesis-driven markers will be evaluated, and unbiased screening will generate more hypotheses to be tested. Even for more widely studied markers, there is often little consensus about their true clinical value: the prognostic effect of the molecular phenotype, initially described many years ago, is only now entering prospective clinical trials.

Currently, the only marker with sufficient evidence to justify routine clinical assessment is *KRAS* mutational analysis and selection for EGFR-specific therapy. It is clear that a much greater degree of cooperation is required between basic and clinical scientists to bring sufficient rigour to bear on this field, designing trials with enough statistical power to provide results that will compel clinical engagement.

1. Wolpin, B. M. & Mayer, R. J. Systemic treatment of colorectal cancer. *Gastroenterology* **134**, 1296–1310 (2008).
2. Midgley, R. & Kerr, D. J. Capecitabine: have we got the dose right? *Nature Clin. Pract. Oncol.* **6**, 17–24 (2009).
3. Midgley, R. S., Yanagisawa, Y. & Kerr, D. J. Evolution of nonsurgical therapy for colorectal cancer. *Nature Clin. Pract. Gastroenterol. Hepatol.* **6**, 108–120 (2009).
4. Dukes, C. The classification of cancer of the rectum. *J. Pathol. Bacteriol.* **35**, 323 (1932).
5. Compton, C. C. *et al.* Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch. Pathol. Lab. Med.* **124**, 979–994 (2000).
6. Steinberg, S. M., Barkin, J. S., Kaplan, R. S. & Stablein, D. M. Prognostic indicators of colon tumors. The Gastrointestinal Tumor Study Group experience. *Cancer* **57**, 1866–1870 (1986).
7. Kemeny, N. & Braun, D. W. Jr. Prognostic factors in advanced colorectal carcinoma. Importance of lactic dehydrogenase level, performance status, and white blood cell count. *Am. J. Med.* **74**, 786–794 (1983).
8. Wanebo, H. J., Rao, B. & Pinski, C. M. Preoperative carcinoembryonic antigen level as a prognostic indicator in colorectal cancer. *N. Engl. J. Med.* **299**, 448–451 (1978).
9. Locker, G. Y. *et al.* ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–5327 (2006).
10. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
The first prognostic gene expression signature in any of the common cancers, although it has since come under criticism.
11. FDA Clears Breast Cancer Specific Molecular Prognostic Test. *US Food and Drug Administration* [online], http://www.fda.gov/bbs/topics/NEWS/2007/NEW01555.html (2009).
12. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–532 (1988).
The classical model of colorectal carcinogenesis. Although increasing numbers of exceptions and additions are becoming apparent, this model describes the common mutations found in CRC.
13. Andreyev, H. J., Norman, A. R., Cunningham, D., Oates, J. R. & Clarke, P. A. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J. Natl Cancer Inst.* **90**, 675–684 (1998).
14. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nature Rev. Cancer* **3**, 11–22 (2003).
15. Stites, E. C., Trampont, P. C., Ma, Z. & Ravichandran, K. S. Network analysis of oncogenic Ras activation in cancer. *Science* **318**, 463–467 (2007).
16. Andreyev, H. J. *et al.* Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study. *Br. J. Cancer* **85**, 692–696 (2001).
17. Samowitz, W. S. *et al.* Relationship of Ki-ras mutations in colon cancers to tumor location, stage, and survival: a population-based study. *Cancer Epidemiol. Biomarkers Prev.* **9**, 1193–1197 (2000).
18. Ince, W. L. *et al.* Association of k-ras, b-raf, and p53 status with the treatment effect of bevacizumab. *J. Natl Cancer Inst.* **97**, 981–989 (2005).
19. Wang, C. *et al.* Prognostic significance of microsatellite instability and Ki-ras mutation type in stage II colorectal cancer. *Oncology* **64**, 259–265 (2003).
20. Ogino, S. *et al.* CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* **58**, 90–96 (2008).
21. Eberhard, D. A. *et al.* Mutations in the epidermal growth factor receptor and in *KRAS* are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. *J. Clin. Oncol.* **23**, 5900–5909 (2005).
22. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
23. Bokemeyer, C. *et al.* KRAS status and efficacy of first-line treatment of patients with metastatic colorectal cancer (mCRC) with FOLFOX with or without cetuximab: the OPUS experience. *J. Clin. Oncol.* **26**, Abstr. 4000 (2008).
24. Tejpar, S. *et al.* Relationship of efficacy with KRAS status (wild type versus mutant) in patients with irinotecan-refractory metastatic colorectal cancer (mCRC), treated with irinotecan (q2w) and escalating doses of cetuximab (q1w): the EVEREST experience. *J. Clin. Oncol.* **26**, Abstr. 4001 (2008).
25. Van Cutsem, E., Lang, I. & D'haens, G. KRAS status and efficacy in the first-line treatment of patients with metastatic colorectal cancer (mCrc) treated with FOLFIRI with or without cetuximab: the CRYSTAL experience. *J. Clin. Oncol.* **26**, Abstr. 2 (2008).
26. Van Cutsem, E. *et al.* Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 1408–1417 (2009).
27. Amado, R. G. *et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 1626–1634 (2008).
28. BRAF V599E becomes V600E. *Catalogue Of Somatic Mutations In Cancer - (COSMIC)* [online], http://www.sanger.ac.uk/genetics/CGP/cosmic/News/#20040203163548 (2004).
29. Di Nicolantonio, F. *et al.* Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 5705–5712 (2008).
30. Rajagopalan, H. *et al.* Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**, 934 (2002).
Demonstrates for the first time that *KRAS* and *BRAF* mutations are mutually exclusive, with the important implications of this finding becoming obvious only now.
31. Thorstensen, L. *et al.* Genetic and epigenetic changes of components affecting the WNT pathway in colorectal carcinomas stratified by microsatellite instability. *Neoplasia* **7**, 99–108 (2005).
32. Lovig, T. *et al.* APC and CTNNB1 mutations in a large series of sporadic colorectal carcinomas stratified by the microsatellite instability status. *Scand. J. Gastroenterol.* **37**, 1184–1193 (2002).
33. Bondi, J., Bukholm, G., Nesland, J. M. & Bukholm, I. R. Expression of non-membranous β-catenin and γ-catenin, c-Myc and cyclin D1 in relation to patient outcome in human colon adenocarcinomas. *Apmis* **112**, 49–56 (2004).
34. Hugh, T. J. *et al.* Cadherin-catenin expression in primary colorectal cancer: a survival analysis. *Br. J. Cancer* **80**, 1046–1051 (1999).
35. Russo, A. *et al.* The *TP53* colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: influence of tumor site, type of mutation, and adjuvant treatment. *J. Clin. Oncol.* **23**, 7518–7528 (2005).
36. Munro, A. J., Lain, S. & Lane, D. P. P53 abnormalities and outcomes in colorectal cancer: a systematic review. *Br. J. Cancer* **92**, 434–444 (2005).
37. Popat, S. & Houlston, R. S. A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *Eur. J. Cancer* **41**, 2060–2070 (2005).
38. Halling, K. C. *et al.* Microsatellite instability and 8p allelic imbalance in stage B2 and C colorectal cancers. *J. Natl Cancer Inst.* **91**, 1295–1303 (1999).
39. Roth, A. D. *et al.* Correlation of molecular markers in colon cancer with stage-specific prognosis: results of the translational study on the PETACC 3 - EORTC 40993-SAKK 60–00 trial. ASCO *Gastrointest. Cancers Symp. Abstr.* 288 (2009).
40. Carethers, J. M., Hawn, M. T., Greenson, J. K., Hitchcock, C. L. & Boland, C. R. Prognostic significance of allelic lost at chromosome 18q21 for stage II colorectal cancer. *Gastroenterology* **114**, 1188–1195 (1998).
41. Alhopuro, P. *et al.* SMAD4 levels and response to 5-fluorouracil in colorectal cancer. *Clin. Cancer Res.* **11**, 6311–6316 (2005).
42. Boulay, J. L. *et al.* SMAD4 is a predictive marker for 5-fluorouracil-based chemotherapy in patients with colorectal cancer. *Br. J. Cancer* **87**, 630–634 (2002).
43. Alazzouzi, H. *et al.* SMAD4 as a prognostic marker in colorectal cancer. *Clin. Cancer Res.* **11**, 2606–2611 (2005).
44. Rowan, A. *et al.* Refining molecular analysis in the pathways of colorectal carcinogenesis. *Clin. Gastroenterol. Hepatol.* **3**, 1115–1123 (2005).
45. Olschwang, S. *et al.* Alternative genetic pathways in colorectal carcinogenesis. *Proc. Natl Acad. Sci. USA* **94**, 12122–12127 (1997).
46. Jass, J. R. *et al.* Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *J. Clin. Pathol.* **52**, 455–460 (1999).
47. Boland, C. R. *et al.* A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).
48. Walther, A., Houlston, R. & Tomlinson, I. Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut* **57**, 941–950 (2008).
49. Popat, S., Hubner, R. & Houlston, R. S. Systematic review of microsatellite instability and colorectal cancer prognosis. *J. Clin. Oncol.* **23**, 609–618 (2005).
Together with reference 48, this paper underlines the importance of the molecular phenotype and illustrates the delay from novel insight to established clinical practice.
50. Sinicrope, F. A. *et al.* Prognostic impact of microsatellite instability and DNA ploidy in human colon carcinoma patients. *Gastroenterology* **131**, 729–737 (2006).
51. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genet.* **38**, 787–793 (2006).
52. Barault, L. *et al.* Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res.* **68**, 8541–8546 (2008).
53. Shen, L. *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc. Natl Acad. Sci. USA* **104**, 18654–18659 (2007).
54. Ogino, S., Kawasaki, T., Kirkner, G. J., Loda, M. & Fuchs, C. S. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J. Mol. Diagn.* **8**, 582–588 (2006).
55. Samowitz, W. S. *et al.* Poor survival associated with the BRAF V600E mutation in microsatellite-stable colon cancers. *Cancer Res.* **65**, 6063–6069 (2005).
56. Yamada, Y. *et al.* Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. *Proc. Natl Acad. Sci. USA* **102**, 13580–13585 (2005).
57. Matsuzaki, K. *et al.* The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. *Clin. Cancer Res.* **11**, 8564–8569 (2005).
58. Rodriguez, J. *et al.* Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res.* **66**, 8462–9468 (2006).
59. Ribic, C. M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N. Engl. J. Med.* **349**, 247–257 (2003).
The first description of the potential detrimental effect of adjuvant 5-FU chemotherapy in MSI-positive CRC.
60. Sargent, D. J. *et al.* Confirmation of deficient mismatch repair (dMMR) as a predictive marker for lack of benefit from 5-FU based chemotherapy in stage II and III colon cancer (CC): a pooled molecular reanalysis of randomized chemotherapy trials. *J. Clin. Oncol.* **26**, Abstr. 4008 (2008).
61. Jo, W. S. & Carethers, J. M. Chemotherapeutic implications in microsatellite unstable colorectal cancer. *Cancer Biomark.* **2**, 51–60 (2006).
62. Kim, G. P. *et al.* Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a National Cancer Institute-National Surgical Adjuvant Breast and Bowel Project Collaborative Study. *J. Clin. Oncol.* **25**, 767–772 (2007).
63. Liang, J. T. *et al.* High-frequency microsatellite instability predicts better chemosensitivity to high-dose 5-fluorouracil plus leucovorin chemotherapy for stage IV sporadic colorectal cancer after palliative bowel resection. *Int. J. Cancer* **101**, 519–525 (2002).
64. Bertagnolli, M. M. *et al.* Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *J. Clin. Oncol.* **27**, 1814–1821 (2009).
65. Fallik, D. *et al.* Microsatellite instability is a predictive factor of the tumor response to irinotecan in patients with advanced colorectal cancer. *Cancer Res.* **63**, 5738–5744 (2003).

66. Vilar, E. *et al.* Microsatellite instability due to hMLH1 deficiency is associated with increased cytotoxicity to irinotecan in human colorectal cancer cell lines. *Br. J. Cancer* **99**, 1607–1612 (2008).
67. Saltz, L. B. *et al.* Irinotecan fluorouracil plus leucovorin is not superior to fluorouracil plus leucovorin alone as adjuvant treatment for stage III colon cancer: results of CALGB 89803. *J. Clin. Oncol.* **25**, 3456–3461 (2007).
68. Swanton, C. *et al.* Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs. *Cancer Cell* **11**, 498–512 (2007).
69. Sudo, T., Nitta, M., Saya, H. & Ueno, N. T. Dependence of paclitaxel sensitivity on a functional spindle assembly checkpoint. *Cancer Res.* **64**, 2502–2508 (2004).
70. Popat, S., Matakidou, A. & Houlston, R. S. Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis. *J. Clin. Oncol.* **22**, 529–536 (2004).
71. Marsh, S., McKay, J. A., Cassidy, J. & McLeod, H. L. Polymorphism in the thymidylate synthase promoter enhancer region in colorectal cancer. *Int. J. Oncol.* **19**, 383–386 (2001).
72. Kawakami, K. *et al.* Different lengths of a polymorphic repeat sequence in the thymidylate synthase gene affect translational efficiency but not its gene expression. *Clin. Cancer Res.* **7**, 4096–4101 (2001).
73. Mandola, M. V. *et al.* A novel single nucleotide polymorphism within the 5' tandem repeat polymorphism of the thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. *Cancer Res.* **63**, 2898–2904 (2003).
74. Mandola, M. V. *et al.* A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. *Pharmacogenetics* **14**, 319–327 (2004).
75. Lenz, H. J. Pharmacogenomics in colorectal cancer. *Semin. Oncol.* **30**, 47–53 (2003).
76. Hitre, E. *et al.* Influence of thymidylate synthase gene polymorphisms on the survival of colorectal cancer patients receiving adjuvant 5-fluorouracil. *Pharmacogenet. Genomics* **15**, 723–730 (2005).
77. Lurje, G. *et al.* Thymidylate synthase haplotype is associated with tumor recurrence in stage II and stage III colon cancer. *Pharmacogenet. Genomics* **18**, 161–168 (2008).
78. Diasio, R. B. & Johnson, M. R. The role of pharmacogenetics and pharmacogenomics in cancer chemotherapy with 5-fluorouracil. *Pharmacology* **61**, 199–203 (2000).
79. Wei, X., McLeod, H. L., McMurrough, J., Gonzalez, F. J. & Fernandez-Salguero, P. Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J. Clin. Invest.* **98**, 610–615 (1996).
The first description of a germline change in the 5-FU degradation pathway as a marker of outcome, in this case toxicity.
80. van Kuilenburg, A. B. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *Eur. J. Cancer* **40**, 939–950 (2004).
81. Ezzeldin, H. H. & Diasio, R. B. Predicting fluorouracil toxicity: can we finally do it? *J. Clin. Oncol.* **26**, 2080–2082 (2008).
82. Frosst, P. *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nature Genet.* **10**, 111–113 (1995).
83. Cohen, V. *et al.* Methylenetetrahydrofolate reductase polymorphism in advanced colorectal cancer: a novel genomic predictor of clinical response to fluoropyrimidine-based chemotherapy. *Clin. Cancer Res.* **9**, 1611–1615 (2003).
84. Jakobsen, A., Nielsen, J. N., Gyldenkerne, N. & Lindeberg, J. Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphism in normal tissue as predictors of fluorouracil sensitivity. *J. Clin. Oncol.* **23**, 1365–1369 (2005).
85. Sharma, R. *et al.* Thymidylate synthase and methylenetetrahydrofolate reductase gene polymorphisms and toxicity to capecitabine in advanced colorectal cancer patients. *Clin. Cancer Res.* **14**, 817–825 (2008).
86. Etienne, M. C. *et al.* Methylenetetrahydrofolate reductase gene polymorphisms and response to fluorouracil-based treatment in advanced colorectal cancer patients. *Pharmacogenetics* **14**, 785–792 (2004).
87. Etienne-Grimaldi, M. C., Francoual, M., Formento, J. L. & Milano, G. Methylenetetrahydrofolate reductase (MTHFR) variants and fluorouracil-based treatments in colorectal cancer. *Pharmacogenomics* **8**, 1561–1566 (2007).
88. Mishima, M., Samimi, G., Kondo, A., Lin, X. & Howell, S. B. The cellular pharmacology of oxaliplatin resistance. *Eur. J. Cancer* **38**, 1405–1412 (2002).
89. McIlwain, C. C., Townsend, D. M. & Tew, K. D. Glutathione *S*-transferase polymorphisms: cancer incidence and therapy. *Oncogene* **25**, 1639–1648 (2006).
90. Lecomte, T., Landi, B., Beaune, P., Laurent-Puig, P. & Loriot, M. A. Glutathione *S*-transferase P1 polymorphism (Ile105Val) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. *Clin. Cancer Res.* **12**, 3050–3056 (2006).
91. Kweekel, D. M., Gelderblom, H. & Guchelaar, H. J. Pharmacology of oxaliplatin and the use of pharmacogenomics to individualize therapy. *Cancer Treat. Rev.* **31**, 90–105 (2005).
92. Stoehlmacher, J. *et al.* A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *Br. J. Cancer* **91**, 344–354 (2004).
93. McLeod, H. L. *et al.* Pharmacogenetic analysis of systemic toxicity and response after 5-fluorouracil (5FU)/CPT-11, 5FU/oxaliplatin (oxal), or CPT-11/oxal therapy for advanced colorectal cancer (CRC): results from an intergroup trial. *Proc. Am. Soc. Clin. Oncol.* **22**, Abstr. 1013 (2003).
94. Iyer, L. *et al.* Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. *Clin. Pharmacol. Ther.* **65**, 576–582 (1999).
95. Iyer, L. *et al.* UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J.* **2**, 43–47 (2002).
96. Hoskins, J. M., Goldberg, R. M., Qu, P., Ibrahim, J. G. & McLeod, H. L. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. *J. Natl Cancer Inst.* **99**, 1290–1295 (2007).
97. Liu, C. Y. *et al.* UGT1A1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma. *Cancer* **112**, 1932–1940 (2008).
98. Bandres, E. *et al.* A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients. *Oncol. Rep.* **17**, 1089–1094 (2007).
99. Del Rio, M. *et al.* Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. *J. Clin. Oncol.* **25**, 773–780 (2007).
100. Grade, M. *et al.* Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. *Cancer Res.* **67**, 41–56 (2007).
101. Kleivi, K. *et al.* Gene expression profiles of primary colorectal carcinomas, liver metastases, and carcinomatoses. *Mol. Cancer* **6**, 2 (2007).
102. Barrier, A. *et al.* Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J. Clin. Oncol.* **24**, 4685–4691 (2006).
103. Wang, Y. *et al.* Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J. Clin. Oncol.* **22**, 1564–1571 (2004).
The first description of a prognostic gene expression signature in CRC, highlighting both the potential of and problems with this technology.
104. Barrier, A. *et al.* Colon cancer prognosis prediction by gene expression profiling. *Oncogene* **24**, 6155–6164 (2005).
105. Jiang, Y. *et al.* Development of a clinically feasible molecular assay to predict recurrence of stage II colon cancer. *J. Mol. Diagn.* **10**, 346–354 (2008).
106. Lin, Y. H. *et al.* Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer. *Clin. Cancer Res.* **13**, 498–507 (2007).
107. Barrier, A. *et al.* Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene* **26**, 2642–2648 (2007).
108. Barrier, A. *et al.* Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. *Dis. Colon Rectum* **48**, 2238–2248 (2005).
109. Arango, D. *et al.* Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* **129**, 874–884 (2005).
110. Frederiksen, C. M., Knudsen, S., Laurberg, S. & Orntoft, T. F. Classification of Dukes' B and C colorectal cancers using expression arrays. *J. Cancer Res. Clin. Oncol.* **129**, 263–271 (2003).
111. Schepeler, T. *et al.* Diagnostic and prognostic microRNAs in stage II colon cancer. *Cancer Res.* **68**, 6416–6424 (2008).
112. Schetter, A. J. *et al.* MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA* **299**, 425–436 (2008).
113. Lanza, G. *et al.* mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. *Mol. Cancer* **6**, 54 (2007).
114. Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A. & Ploner, A. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* **21**, 3017–3024 (2005).
115. Ntzani, E. E. & Ioannidis, J. P. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* **362**, 1439–1444 (2003).
116. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488–492 (2005).
117. Dupuy, A. & Simon, R. M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl Cancer Inst.* **99**, 147–157 (2007).
This should be compulsory reading for anyone with an interest in gene expression signatures, as it underlines the pitfalls in analysing the expression data.
118. Ransohoff, D. F. Gene-expression signatures in breast cancer. *N. Engl. J. Med.* **348**, 1715–1717; author reply 1715–1717 (2003).
119. Simon, R., Radmacher, M. D., Dobbin, K. & McShane, L. M. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl Cancer Inst.* **95**, 14–18 (2003).
120. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl Acad. Sci. USA* **103**, 5923–5928 (2006).
121. Lavery, I. C. *et al.* Relationship between tumor gene expression and recurrence in an observational cohort of patients with stage II/III colon cancer treated with surgery only: quantitative RT-PCR assay of 375 genes in fixed paraffin-embedded (FPE) tissue. *ASCO Gastrointest. Cancer Symp.*, Abstr. 302 (2008).
122. Gray, R. *et al.* Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* **370**, 2020–2029 (2007).
123. Lusa, L. *et al.* Challenges in projecting clustering results across gene expression-profiling datasets. *J. Natl Cancer Inst.* **99**, 1715–1723 (2007).
124. Lyman, G. H. & Kuderer, N. M. Gene expression profile assays as predictors of recurrence-free survival in early-stage breast cancer: a metaanalysis. *Clin. Breast Cancer* **7**, 372–379 (2006).
125. Ioannidis, J. P. Microarrays and molecular research: noise discovery? *Lancet* **365**, 454–455 (2005).
126. Ransohoff, D. F. Bias as a threat to the validity of cancer molecular-marker research. *Nature Rev. Cancer* **5**, 142–149 (2005).
127. Tomlinson, I. P. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genet.* **40**, 623–630 (2008).
128. Jaeger, E. *et al.* Common genetic variants at the *CRAC1* (*HMPS*) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nature Genet.* **40**, 26–28 (2008).
129. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature Genet.* **39**, 645–649 (2007).
130. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nature Genet.* **39**, 870–874 (2007).
131. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
Well-conducted GWAS for risk of myocardial infarction, demonstrating the power of this strategy for novel marker discovery.

132. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
133. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature Genet.* **38**, 1043–1048 (2006).
134. Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763 (1997).
135. Young, J. *et al.* Genomic instability occurs in colorectal carcinomas but not in adenomas. *Hum. Mutat.* **2**, 351–354 (1993).
136. Grady, W. M. & Carethers, J. M. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).
137. Segditsas, S. & Tomlinson, I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**, 7531–7537 (2006).
138. Herman, J. G. *et al.* Incidence and functional consequences of *hMLH1* promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA* **95**, 6870–6875 (1998).
139. Malkhosyan, S., Rampino, N., Yamamoto, H. & Perucho, M. Frameshift mutator mutations. *Nature* **382**, 499–500 (1996).
140. Parsons, R. *et al.* Microsatellite instability and mutations of the transforming growth factor β type II receptor gene in colorectal cancer. *Cancer Res.* **55**, 5548–5550 (1995).
141. Souza, R. F. *et al.* Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. *Nature Genet.* **14**, 255–257 (1996).
142. Rampino, N. *et al.* Somatic frameshift mutations in the *BAX* gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**, 967–969 (1997).
143. Mao, J. H. *et al. Fbxw7/Cdc4* is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature* **432**, 775–779 (2004).
144. Rajagopalan, H. *et al.* Inactivation of *hCDC4* can cause chromosomal instability. *Nature* **428**, 77–81 (2004).
145. Kemp, Z. *et al. CDC4* mutations occur in a subset of colorectal cancers but are not predicted to cause loss of function and are not associated with chromosomal instability. *Cancer Res.* **65**, 11361–11366 (2005).
146. Salomon, D. S., Brandt, R., Ciardiello, F. & Normanno, N. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit. Rev. Oncol. Hematol.* **19**, 183–232 (1995).
147. Cunningham, D. *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **351**, 337–345 (2004).
148. Chung, K. Y. *et al.* Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. *J. Clin. Oncol.* **23**, 1803–1810 (2005).
149. Samowitz, W. S. *et al.* Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* **129**, 837–845 (2005).
150. Cheng, Y. W. *et al.* CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin. Cancer Res.* **14**, 6005–6013 (2008).
151. Parker, W. B. & Cheng, Y. C. Metabolism and mechanism of action of 5-fluorouracil. *Pharmacol. Ther.* **48**, 381–395 (1990).

DATABASES

ClinicalTrials.gov: http://clinicaltrials.gov/E5202

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
APC | *BRAF* | *DCC* | *KRAS* | *SMAD4* | *TP53*

National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/
5-FU | capecitabine | cetuximab | irinotecan | oxaliplatin | panitumumab | trastuzumab

UniProtKB: http://www.uniprot.org
β-catenin | DPD | EGFR | MTHFR | TS

ALL LINKS ARE ACTIVE IN THE ONLINE PDF