



Nucleic acid quantification and disease outcome prediction in colorectal cancer

Stephen A Bustin

University of London,
Institute of Cell and
Molecular Science,
Barts and the London,
Queen Mary's School of
Medicine and Dentistry,
London, UK

*Centre for Academic Surgery,
3rd Floor Alexandra Wing,
The Royal London Hospital,
London E1 1BB, UK
Tel.: +44 020 7377 7000
Ext. 2431;
E-mail: s.a.bustin@
qmul.ac.uk

* Address for correspondence

Histopathological stage at diagnosis remains the most important prognostic determinant for colorectal cancer. However, conventional staging is unable to predict disease outcome accurately for each individual patient. This results in considerable prognostic heterogeneity within a given tumor stage and is of particular relevance for a subgroup of patients with stage II disease that would benefit from adjuvant therapy. The recent advances in functional genomics are beginning to have a significant impact on clinical oncology, and there is widespread interest in using molecular techniques for clinical applications. These have focused on two approaches: the use of polymerase chain reaction (PCR)-based methods for the detection of occult disease in lymph nodes, bone marrow and blood and the use of microarrays for the expression profiling of primary tumors. The aim is to develop molecular classifiers that will allow the prediction of disease outcome, thus matching patients with individualized treatment. Despite the obvious attractions of these approaches, there have been significant technical, biological and analytical problems in their translation into clinically relevant practice. This is particularly true for colorectal cancer, the second most common cancer in the western world. Nevertheless, progress is being made and the improved awareness and appreciation of those difficulties is beginning to generate results that should prove useful for clinical oncology.

Colorectal cancer, the second most common cancer in the western world [1], is best considered not as a homogeneous disease, but as covering a spectrum in terms of its molecular properties and its clinical and pathological diversity. Unlike other cancers, such as those of the lung and breast, the size of the primary tumor is often not a primary prognostic indicator. Tumors have a tendency to grow slowly, enlarging gradually until they eventually penetrate the bowel wall. Surgical resection of the primary cancer continues to be the only curative option for colorectal cancer. The probability of a cure following surgery is directly related to the stage of the cancer, and lymph node (LN) status is the most influential prognostic marker with respect to disease-free and overall survival. However, even after an apparent curative surgical intervention, approximately 50% of patients experience metastatic relapse that in most cases eventually leads to death. There is considerable prognostic heterogeneity within each tumor stage [2]: up to a third of LN-negative (N_0) patients experience treatment failure and die from disseminated disease, whereas around a third of stage III ($T_{1-4}N_{1-2}M_0$) patients are cured [3].

There are many approaches to improving the accuracy of clinical staging. One interesting approach utilizes artificial neural networks (ANNs), trained and validated using various

clinicopathological factors. ANNs are computer-based algorithms that can be used to discover complex relations within data sets. They permit the recognition of patterns in complex biological data sets that cannot be detected with conventional linear statistical analysis by a process of error minimization through learning from experience. Although ANNs can increase the accuracy of colon cancer classification and survival prediction when compared with other statistical or clinicopathological method [4], this improvement is still not able to predict survival for each individual patient. Furthermore, ANNs are considered to be 'black boxes' with lengthy development and optimization times, as well as huge computational requirements and, since model development is empirical, ANNs provide low decision insight. Therefore, while histopathological staging procedures and extensions such as ANN are useful in defining the extent of and prognosis for colorectal cancer and placing patients into risk groups, they are unable to accurately identify the risk of treatment failure for every individual patient. Consequently, there continues to be widespread interest in using molecular techniques in clinical diagnostics.

The challenge

There are various approaches employing molecular techniques that can be used to predict disease outcome for colorectal cancer patients. One

Keywords: cancer, colon, metastasis, microarray, occult disease, polymerase chain reaction, reverse transcription, staging

future
medicine

involves attempting to detect circulating tumor cells in blood before surgery, coupled with a molecular analysis of the LN after surgery. This is often referred to as molecular staging and aims to augment histopathological staging with additional, patient-specific information. Molecular staging can be followed up with monitoring the appearance of circulating tumor cells in blood over time after surgery, with the aim of detecting the appearance of metastatic recurrence at the earliest possible time. Another approach tries to associate a tumor's metastatic potential with the genotypic or expression characteristics of the primary tumor. All these methods have as their ultimate goal the identification of colorectal cancer patients with actual residual disease, rather than of patients belonging to a group with a statistical risk of harboring metastases, and hence in need of adjuvant treatment. At the same time, these investigations may also result in the identification of reliable markers capable of distinguishing between subgroups of patients that respond to different treatment regimens. This review deals with the first aim, the second one having been dealt with recently [5,6].

Two, not necessarily exclusive, molecular approaches are used to address these challenges. The first is based on the assumption that treatment failure is due to the presence of occult disease, sometimes referred to as 'micrometastasis', which is defined as metastatic cancer of limited disease burden that cannot be detected by standard histopathological techniques. Occult disease may be revealed by the detection of characteristic mutations and/or the expression of tissue-specific genes. Since their detection requires extremely sensitive and specific techniques, there has been widespread interest in polymerase chain reaction (PCR)-based methods capable of detecting such targets [7]. These involve conventional gel-based PCR methods, as well as quantitative methods such as real-time PCR, standardized competitive reverse transcription (StaRT)-PCR and linear-after-the-exponential (LATE)-PCR. Today, such methods provide the most sensitive and specific means of detecting occult disease.

A second approach arises from the recognition that the metastatic potential of a cancer may be encoded in the bulk of the primary tumor [8] and that characteristics present in primary tumors may predict the course of a patient's disease. Consequently, this approach aims to analyze the primary tumor for microsatellite [9] and ploidy [10] status, and increasingly, its gene

expression profile [11]. Expression profiling, in particular, is expected to lead to the identification of cellular and serum markers for colorectal neoplasia. Ideally, such markers would have several characteristics: first, they should be expressed at high levels in most tumors and at greatly reduced levels in normal tissues, thus producing high sensitivity and specificity. Second, any elevated expression should be an early event and remain at high levels during the neoplastic transformation process. Third, markers should be secreted to facilitate their detection. The identification of such markers, together with any correlation between specific types of genetic changes and relevant clinical information should be able to distinguish clinical prognostic groups.

PCR-based detection of occult disease

This approach has the longest track-record, and there are numerous publications identifying a huge range of potential prognostic markers [12,13]. Occult disease can be detected by identifying its DNA (PCR) or its mRNA (reverse transcription [RT]-PCR) in LN, blood or bone marrow. The targeting of DNA has the advantage of allowing the detection of tumor-specific mutations, thus providing direct evidence for the presence of tumor-derived DNA. The targeting of RNA aims to detect colon tissue-specific mRNA in a background where its expression would not be expected to occur. However, while simple in principle, there are several limitations associated with the detection of occult disease. First, whilst it would be ideal to be able to detect cancer-specific mutations, there are no known mutations that occur consistently in colorectal cancers. Second, even when the same p53 or K-ras mutations are identified in both primary cancers and LN, a significant number of patients do not develop distant recurrences [14,15]. One reason for this may be that the detection of DNA may simply be a reflection of tumor burden [16,17], a problem that is not solved by switching to the detection of tumor-cell-derived RNA, as even the presence of RNA is not a reliable indicator of the presence of viable cells [18]. Third, there are some concerns regarding the lack of truly tissue-specific markers. For example, the widely used epithelial tissue-specific marker cytokeratin 20 (ck20) is expressed in the blood [19–22] and LN [23] of healthy controls, with no association between ck20 mRNA levels and distant recurrence or survival [24,25]. Similar reservations have been recorded for the other commonly used tissue-specific markers [12]. Fourth, conventional RT-PCR assays can be very unreliable [26,27].

Today there are three PCR-based methods in common use for the detection of cancer cells from blood, LN and bone marrow. All three are capable of quantification, thus adding an additional factor, to the parameters potentially useful for developing a practical, clinical diagnostic assay.

Real-time PCR

The most widely used technology involves the application of real-time PCR (qPCR) assays [28], which represent a significant advance on conventional PCR-based methods [29]. Quantification is based on the detection of a fluorescent reporter molecule, and the increase in the amount of fluorescence as PCR products accumulate with each cycle of amplification. There are three main chemistries in general use:

- Intercalating dyes, such as SYBR®-Green (Invitrogen, CA, USA), which fluoresce upon light excitation when bound to double stranded DNA. These are cheap, easily added to legacy assays and amplification products can be verified by the use of melt curves. On the other hand, they can lack specificity and fluorescence varies with amplicon length.
- Primers linked to fluorophores, for example, Lux™ primers (Invitrogen) or Plexor™ (Promega, CA, USA). Fluorescence detection depends on the primers binding to, and being extended during the PCR reaction. These are cheap, and amplification products can be verified by melt curves. However, specificity depends on the primers and company-specific design software needs to be used for optimal performance.
- Hybridization-probe-based methods, for example, TaqMan® (Applied Biosystems) or molecular beacons. These are the most specific methods, as products are only detected if the probes hybridize to the appropriate amplification products. There are many variations on this theme, with melt curve analysis possible for some chemistries. Their main disadvantages are cost, complexity and occasional fragility of probe synthesis and potential problems associated with the fact that probe-based assays do not report primer dimers that can interfere with the efficiency of the amplification reaction.

If real-time PCR is preceded by a RT step to quantitate mRNA or viral RNA targets, this method is often termed qRT-PCR or, alternatively RT-qPCR. Not surprisingly, whilst simple in concept, this technology has its own

problems [30] and its appropriate application depends on optimizing and validating each individual step that makes up a qPCR assay [31]. Proper optimization of pre-assay [32] and assay [33] conditions, its validation [34] and standardization [35–37] are crucial to meeting the quality criteria expected of any laboratory test applied in clinical medicine [12]. Arguably the most important overall parameter concerns assay and data standardization. A recent survey of qRT-PCR protocols and analysis procedures revealed huge variability, even between diagnostic laboratories [38]. Differences in the pre-analytical steps, such as blood sample processing [39,40] and RNA isolation [41] lead to heterogeneous results even if evaluated in the same laboratory. Similarly, the differences in analytical variables that affect the accuracy of conventional PCR assays [42] affect both qPCR assays [43,44]. Standardized data are necessary when making interexperimental and interlaboratory comparison of gene expression measurements, and it is near impossible to develop a meaningful gene expression database with nonstandardized data.

StaRT-PCR

The problem of data standardization is addressed by an alternative method, termed StaRT-PCR [45]. StaRT-PCR is a quantitative method for measuring multigene [46] expression using standardized mixtures of competitive templates that include defined reference templates for β -actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Since these reference competitive templates are included in each reaction, any expression value can be related to the expression of one or both of these two genes. The inclusion of the reference genes in each mixture also allows for direct comparison of values obtained with different mixtures, or different experiments. The competitive templates for each gene are commercially designed, manufactured and validated. Each gene expression result is reported as the number of molecules of mRNA for gene 'X' per 10^6 molecules of reference gene. Serial dilutions of the standardized mixes allow quantitative measurements over a 6.5 log range. All data are directly comparable as data are measured using standardized lots of the internal standards. With such standardized data, expression levels of genes in diseased vs normal tissue can be compared across experiments and other laboratories. In addition, an internal standard inherently corrects for any inhibitors that may be present in tissue extracts, a problem that requires additional assays with real-time PCR analysis. The

establishment of a gene expression database that is annotated with respect to phenotype (e.g., responsiveness to particular chemotherapeutic agents) makes it possible to compare and evaluate the utility of gene expression data from new clinical specimens [47]. StaRT-PCR has been used to develop a method that is 100% accurate for diagnosing lung cancer in cytological specimen, improving on the 80% sensitivity rate currently obtained for morphological criteria [48]. Results obtained using StaRT-PCR correlate well with those obtained using TaqMan assays [49], and this technology may be better for high-throughput quantitative transcript measurements.

LATE-PCR

A third alternative is LATE-PCR, which is particularly useful when targeting low copy number targets [50]. This method is similar to asymmetric PCR, in that the two PCR primers are used at different concentrations. However, LATE-PCR takes into account the difference in melting temperature (T_m) of two primers present at different concentrations, thus improving both efficiency and specificity. LATE-PCR also makes it possible to use lower temperature detection, since the probe does not need to compete with hybridization and extension of the limiting primer during the early, exponential phase of the reaction. Hybridization of probe and target is unimpeded once the limiting primer is depleted and can be done either by lowering the annealing temperature at that point, or by introducing a low-temperature detection step between the extension and melting steps. Probes with lower melting temperatures are easier to design, more allele discriminating, and have lower background fluorescence. Moreover, as the probe dissociates from its target strand well below the extension temperature of the reaction, sufficient probe can be added to the reaction to measure all product strands without inhibiting the amplification reaction. Consequently, LATE-PCR increases the signal strength and allele discrimination capability of fluorescent probes and reduces variability among replicate samples and provides a means for improving the accuracy of single-cell genetic diagnosis [51]. Therefore, it is suited for detecting and/or quantitating nucleic acids from the very limited amounts of tumor material available for analysis following a resection.

Usefulness of PCR-based methods

The prognostic value of detecting disease-associated mRNA from circulating colorectal cancer cells in the blood, LN or bone marrow of

patients remains unclear. This is due to conceptual reservations, as well as technical limitations that interfere with the diagnostic specificity of qRT-PCR assays. The lack of clinical validation of RT-PCR-based assays in prospective studies suggests that although RT-PCR assays can identify LN harboring colon-derived cells that are not detected by conventional histopathology, these are not likely to progress to distant metastases. This implies a conceptual flaw, based on a simplistic view of the biology and kinetics of tumor cell traffic through the lymphatic and systemic circulation and subsequent metastasis development [12]. Critically, RT-PCR detection strategies do not provide any information about the metastatic potential of the cells they detect. Even unequivocal detection of tumor cells by mutant allele-specific PCR fails to predict relapse within 5 years of surgery in between 27 [14] and 70% [15] of patients. Indeed, a patient in the LN-negative group also died. In general, when RT-PCR assays are used to detect 'tissue-specific' mRNA expression, tumor recurrence is observed in only 14–50% of patients whose LN are RT-PCR positive. Together with the results described above, this suggests, as has been previously asserted [52], that while RT-PCR may be detecting cancer cells, the majority of these may not have any metastatic potential. Hence their detection may be no more prognostic for the individual patient than conventional staging.

Microarrays

DNA microarrays rely on nucleic acid hybridization and the use of nucleic acid polymers, immobilized on a solid surface, as probes for complementary nucleic acid target sequences. They can be used to monitor simultaneously the expression of thousands of genes from human tumor samples and can be applied to large numbers of samples in parallel. There are two platforms, based on cDNA or oligonucleotide microarrays. cDNA microarrays consist of cDNA clone inserts representing genes of interest spotted onto a solid support. They are commonly queried with cDNA derived from experimental and reference RNA samples that have been differentially labeled with two fluorophores to allow for the quantification of differential gene expression, and expression values are reported as ratios between two fluorescent values. Oligonucleotide arrays use a single fluorescent label and experimental RNA is amplified, biotinylated for detection, hybridized and detected through the binding of a fluorescent streptavidin-bound compound [53].

There have been many attempts to find novel markers to identify patients with the potential for colon cancer progression [54]. The concept of disease classifiers based on differential gene expression supposes that metastatic potential is associated with the expression levels of a given set of genes. Hence the characterization of global mRNA patterns is expected to reveal clues about regulatory mechanisms, biochemical pathways and broader cellular function and so advance the understanding of disease pathology and increase the accuracy of staging. Furthermore, transcriptome analysis could lead to the identification of key markers among the complex network of gene products involved in metastasis and their association with previously undetectable features of the molecular basis of individual tumor characteristics.

Profiling of colorectal cancers

Since the first report comparing expression profiles of tumor and normal colon tissue [55], there have been several studies analyzing expression profiles from different disease stages. Two have analyzed molecular differences between adenomas and carcinomas and identified 1800 [56] and 50 [57] discriminating genes capable of distinguishing the two stages. Another study reported the construction of a ‘colonochip’, a microarray specifying some 4600 genes that are expressed in colorectal cancer, normal colonic mucosa and liver metastatic cancer tissues and its use to identify expression profiles associated with colorectal tumor progress. Expression profiling of highly metastatic cell lines relative to their poorly metastatic parental cell line has identified metastasis-associated expression pattern changes in 170 genes [58]. Another study compared the expression profiles of LN-positive and LN-negative cancers using a 4608 cDNA array and identified 60 genes that could correctly place 10/12 cancers into their appropriate groups [59]. Yet another study collected primary and metastatic tissue from patients undergoing simultaneous primary resection from the colon and metastatic resection from the liver [60]. Samples were isolated using laser capture microdissection (LCM) and analyzed using a 9121-gene cDNA microarray. A total of 40 genes were upregulated significantly, while seven genes were downregulated significantly in metastatic as compared with primary tumor.

A recent report addresses the issue of prognosis and suggests that expression profiling of colorectal cancers is able to distinguish clinically relevant subgroups and metastatic versus

non-metastatic tumors [61]. The authors identified a 200-odd gene set that divided patients with significantly different 5-year survival rates. Importantly, most non-metastatic tumors that clustered with metastatic cases eventually developed metastasis, confirming the notion that metastatic potential can be predicted from the transcriptome of the primary tumor. Discriminator genes were associated with various cellular processes (e.g., cell-cycle regulation, cell adhesion and angiogenesis). However, in common with other findings they do not point to an obvious discriminatory mechanism for metastasis. Although many of the genes identifying LN metastasis and predicting distant metastasis are the same, there are differences. This is a reflection of the clinical observation that the two are not perfectly correlated and emphasizes expression of the underlying biological differences. Another recent study examined the expression profile from 74 patients with stage II colorectal cancer and identified a 23-gene signature profile that predicted recurrence. The signature was validated in 36 independent patients with an overall accuracy of 78%: 13 out of 18 patients with relapse and 15 of 18 disease-free patients were correctly predicted, representing an odds ratio of 13 [11]. A third report used a 43-gene signature to predict 3-year survival with a 90% accuracy (93% sensitivity and 84% specificity) in predicting 36-month overall survival in 78 patients and, most importantly, was able to discriminate a survival difference in an independent test set [62]. In another study, a 20-gene expression profile that predicted distant recurrence was obtained from the primary tumors of stage III patients that had not received any adjuvant therapy [63]. Although the ability to predict a poor clinical prognosis does not necessarily indicate a benefit from an intervention, it certainly seems capable of defining a group of patients that will not require any adjuvant treatment.

Usefulness of microarray-based methods

The identification of genes associated with metastasis demonstrates the feasibility of combining large-scale molecular analysis of expression profiles with classic morphological and clinical methods of staging and grading cancer for better outcome prediction. The studies above suggest that global expression profiling is beginning to identify patterns of gene expression that appear to correlate with metastatic potential and may identify molecular markers for colon cancer prognosis.

However, there are a number of limitations that result in uncertainty concerning the reliability of DNA microarray measurements. First, since the hybridization conditions are variable for different genes, accurate measurements of absolute expression levels and the reliable detection of low-abundance genes are currently beyond the reach of microarray technology. Consequently, although the direction of change indicated by microarray experiments is usually reliable, the magnitude of gene expression changes are much less so [54]. However, it is likely that technological advances in extraction, hybridization and detection methods will improve the measurement potential of microarrays. Second, differential gene expression and prognostic value of expression patterns are not necessarily synonymous and there is little consistency between studies used to predict clinical outcome [64]. A recent tissue microarray study was unable to show any association between the expression profiles of several cell cycle regulatory or proliferation-related markers previously correlated with prognostic relevance and disease-free survival in node-negative rectal cancers treated by surgery alone [65]. It is apparent that single microarray data are prone to false findings [66] and that different studies result in different gene lists; indeed different gene selection methods can lead to strikingly different gene lists from the same experiment [67]. This may be caused by technical variability, where the methods by which the classifiers are extracted from the microarray data generate variable results. In this case the use of appropriate 'models of consistency' and standard operating procedures may resolve the underlying biological phenomena of the experiments. Furthermore, the establishment of gene coexpression networks for functionally related genes is also likely to improve the biological validity of microarray data. A recent study combined independent datasets on different types of cancer to explore transcriptional changes in terms of gene interactions rather than at the level of individual genes [68]. Two distinct networks were able to detect biological changes and identify gene groups whose co-regulation might contribute to malignant transformation.

On the other hand, recent evidence implies a significant role in the metastatic process for germline polymorphisms. In practice this means that malignancy is influenced not just by environmental stimuli, and multiple genetic and epigenetic events that arise within the malignant epithelium, but also by the genetic background

of the host [69]. In addition, there is an important contribution from the tumor microenvironment and metastasis can be modified by stromal events. Consequently, it is plausible that multiple signatures could be a consequence of the multiple pathways through which colorectal metastases arise [70], as demonstrated by the fact that LN and liver metastases from the same patient do not always show the same genetic aberrations [71]. This implies that the accuracy of multigenic classifiers may be improved if their selection is based on better understanding of the underlying tumor biology. The finding of a molecular signature in primary cancers predictive of metastasis [8] suggests that the metastatic potential of many cancers is encoded in the bulk of the primary tumor and raises the expectation that this will result in a prognostic application. Nonetheless, it is notable that colorectal cancers were not included in the demonstration of clinical utility of this metastasis-associated signature, although another recent study does describe such a signature for stage III colorectal cancer patients [63]. Interestingly, this study also suggests that detection of the expression of a single mRNA, the ras homolog gene family member A (RHOA), can identify a subset of patients that may benefit from adjuvant chemotherapy. However, this depended on selecting appropriate cut-off levels for mRNA levels, with the associated problems this entails [12].

Cancer analysis by expression profiling cannot afford to ignore the molecular complexity of colorectal cancer. There is extensive intratumor genetic heterogeneity in colorectal cancers, with between two and six different clonal genotypes per tumor [72,73], with obvious implications for expression heterogeneity. Furthermore, gene expression profiles of colorectal cancers are affected by environmental conditions, such as differences in diet or the preparation of the bowel prior to surgery. It must also be remembered that colorectal cancer biopsies are composed of epithelial cells, as well as numerous other cell types. Variable proportions of these nonepithelial and non-malignant cells could lead to inconsistent and biologically inaccurate gene expression patterns. Analysis of microdissected cell populations partly addresses these problems, and has been applied to the analysis of profiles associated with colorectal cancer development [74]. However, analysis of microdissected cell populations has its own disadvantages; a significant proportion of metastasis-associated signatures appear to be derived from the nonepithelial component of the tumor and

microdissection would result in the loss of that component [8]. The case for considering expression profiles of both epithelial and stromal cells is further strengthened by the extraordinary plasticity of cells: signals arising in a tumor can reprogram both normal and cancer cells. Cancer cells may be reprogrammed to take on the structure and function of vascular cells in a process called vascular mimicry [75] and tumors can organize regional fibroblasts and endothelial cells into a collaborating metabolic domain that is able to sustain cancer cell survival [76]. All these factors may generate noise that could mask expression patterns relevant for outcome prediction.

Conclusion & expert commentary

Colorectal cancer is a complex disease arising from the interplay of genetic and environmental factors. The achievements of genome sequencing projects and technological advances in functional genomics have facilitated the transition to large-scale systemic approaches to studying this disease and, undoubtedly, the quantification of mRNA by RT-PCR methods and expression profiling by microarrays has provided important insights into its biology. Less clear is whether these technologies are more accurate at predicting disease outcome for individual patients at risk of treatment failure following supposedly curative surgery. First, they generate yet more information without any clear evidence that this is any more relevant for patient management. The results obtained from using RT-PCR to quantitate ‘tissue-specific’ markers, in particular, are not compelling. Second, there are numerous technical issues that need to be addressed and it is essential that technical variability is minimized by using standard operating procedures that implement standards at each step of the procedures, especially with regard to careful quality control of experiments with associated quality metrics. Again, the biological relevance of data obtained using real-time PCR, the most widely used RT-PCR technology, is obscured by its lack of any meaningful standardization. An alternative PCR-based method, such as STaRT-PCR may well be a more useful tool due to its very standardized protocol. Third, larger studies with appropriate clinical design, adjustment for known predictors and correct validation independent data sets are essential before clinical decisions can be based on results obtained from either technology. Ultimately, any set of molecular

marker will be validated if it is able to predict the risk of distant recurrence accurately for every individual patient. Whether this will turn out to be a blood or LN-based marker, a multigene mRNA expression profile obtained from primary tumors, or protein/peptide or metabolome profiles obtained from primary tumor, LN or blood, remains an issue open for debate. However, there can be no doubt that in the not-too-distant future there will be a more accurate classification system that will allow accurate prognostic stratification of individual patients and the administration of molecular therapeutics that show optimal activity in particular subgroups of patients.

Outlook

Host genetics plays an increasingly recognized critical role in determining tumor behavior as well as patient response to adjuvant therapies. Consequently, new molecular therapeutics are likely to find optimal activity in particular, well-defined subgroups of patients. Furthermore, whereas there are numerous tumor-type specific changes on the path to malignancy and metastasis, it appears that there may be fewer, and common changes in the cancer microenvironment, raising the hope that therapeutic targeting of these events could be generally applicable [77]. Interestingly, if polymorphisms, rather than oncogenic events, induce the metastasis-predictive gene signatures, then it may not be necessary to obtain tumor biopsies for prognosis. Instead, normal colonic biopsies or even more easily accessible tissues, for example, blood, could be useful for identifying patients at risk before they develop tumors. This issues are well worth examining and it is essential to obtain epidemiological evidence linking human polymorphisms with metastasis. In any event, the molecular networks involved in regulating the metastatic cascade are slowly being untangled and molecular medicine, by helping with more accurate prognostic stratification of colorectal cancer patients, will play a leading role in the move toward individualized, genome-based, combinatorial cancer therapy. In combination with an increasing array of additional tools, such as tissue-specific genetic knockouts, siRNA-mediated knockdown, the development of protein, tissue and interaction arrays, these serve to increase our understanding of the metastatic process and the role played by the tumor microenvironment.

Highlights

- Colorectal cancer describes a disease spectrum characterized by genetic and epigenetic heterogeneity.
- There is a clear association between different subtypes of colorectal cancer and patient prognosis.
- Molecular techniques such as polymerase chain reaction (PCR) and microarray analysis are being used in attempts to improve the accuracy of conventional histopathological staging for individual patients.
- While numerous individual markers and several expression profiles have been reported as being independent predictors of disease outcome, none have been universally validated.
- The acceptance of any markers into clinical practice must be supported by reproducible validation using standardized assay conditions or diagnostic criteria in multiple independent large sets of samples, as well as by different laboratories.
- The identification of the role germline polymorphisms have in cancer progression suggests it will be possible to match therapy to individuals, based on their susceptibilities and response profiles.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

- Boyle P, Ferlay J: Cancer incidence and mortality in Europe, 2004. *Ann Oncol.* 16, 481–488 (2005).
- Jemal A, Murray T, Ward E *et al.*: Cancer statistics, 2005. *CA Cancer J. Clin.* 55, 10–30 (2005).
- Hermanek P: pTNM and residual tumor classifications: problems of assessment and prognostic significance. *World J. Surg.* 19, 184–190 (1995).
- Ahmed FE: Artificial neural networks for diagnosis and survival prediction in colon cancer. *Mol. Cancer* 4, 29 (2005).
- Ahmed FE: Molecular markers that predict response to colon cancer therapy. *Expert Rev. Mol. Diagn.* 5, 353–375 (2005).
- Ahmed FE: Development of novel diagnostic and prognostic molecular markers for sporadic colon cancer. *Expert Rev. Mol. Diagn.* 5, 337–352 (2005).
- Bustin SA, Dorudi S: Molecular assessment of tumor stage and disease recurrence using PCR-based assays. *Mol. Med. Today* 4, 389–396 (1998).
- Ramaswamy S, Ross KN, Lander ES, Golub TR: A molecular signature of metastasis in primary solid tumors. *Nature Genet.* 33, 49–54 (2003).
- Describes the identification of an expression signature in the early stages of several solid tumor types that support the idea that some primary tumors are preconfigured to metastasize, and that this propensity is detectable at the time of initial diagnosis.
- Ribic CM, Sargent DJ, Moore MJ *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).
- Hawkins NJ, Tomlinson I, Meagher A, Ward RL: Microsatellite-stable diploid carcinoma: a biologically distinct and aggressive subset of sporadic colorectal cancer. *Br. J. Cancer* 84, 232–236 (2001).
- Wang Y, Jatkoe T, Zhang Y *et al.*: Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J. Clin. Oncol.* 22, 1564–1571 (2004).
- Identification of a set of prognostic markers for disease recurrence in lymph node negative cancers.
- Bustin SA, Mueller R: Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin. Sci. (Lond.)* 109, 365–379 (2005).
- Bustin SA, Mueller R: Real-time reverse transcription PCR and the detection of occult disease in colorectal cancer. *Mol. Aspects Med.* 27, 192–223 (2006).
- Hayashi N, Ito I, Yanagisawa A *et al.*: Genetic diagnosis of lymph-node metastasis in colorectal cancer. *Lancet* 345, 1257–1259 (1995).
- The first description of the use of a polymerase chain reaction (PCR)-based assay to try and improve staging accuracy in colorectal cancer.
- Clarke GA, Ryan E, Crowe JP, O'Keane JC, MacMathuna P: Tumor-derived mutated K-ras codon 12 expression in regional lymph nodes of stage II colorectal cancer patients is not associated with increased risk of cancer-related death. *Int. J. Colorectal Dis.* 16, 108–111 (2001).
- Anker P, Lefort F, Vasioukhin V *et al.*: K-ras mutations are found in DNA extracted from the plasma of patients with colorectal cancer. *Gastroenterology* 112, 1114–1120 (1997).
- Lecomte T, Berger A, Zinzindohoue F *et al.*: Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis. *Int. J. Cancer* 100, 542–548 (2002).
- Ng EK, Tsui NB, Lam NY *et al.*: Presence of filterable and nonfilterable mRNA in the plasma of cancer patients and healthy individuals. *Clin. Chem.* 48, 1212–1217 (2002).
- Bustin SA, Gyselman VG, Williams NS, Dorudi S: Detection of cytokeratins 19/20 and guanylyl cyclase C in peripheral blood of colorectal cancer patients. *Br. J. Cancer* 79, 1813–1820 (1999).
- Wyld DK, Selby P, Perren TJ *et al.*: Detection of colorectal cancer cells in peripheral blood by reverse-transcriptase polymerase chain reaction for cytokeratin 20. *Int. J. Cancer* 79, 288–293 (1998).
- Champelovier P, Mongelard F, Seigneurin D: CK20 gene expression: technical limits for the detection of circulating tumor cells. *Anticancer Res.* 19, 2073–2078 (1999).
- Wharton RQ, Jonas SK, Glover C *et al.*: Increased detection of circulating tumor cells in the blood of colorectal carcinoma patients using two reverse transcription-PCR assays and multiple blood samples. *Clin. Cancer Res.* 5, 4158–4163 (1999).
- Bustin SA, Gyselman VG, Siddiqi S, Dorudi S: Cytokeratin 20 is not a tissue-specific marker for the detection of malignant epithelial cells in the blood of colorectal cancer patients. *Int. J. Surg. Invest.* 2, 49–57 (2000).
- Bustin SA, Siddiqi S, Ahmed S, Hands R, Dorudi S: Quantification of cytokeratin 20, carcinoembryonic antigen and guanylyl cyclase C mRNA levels in lymph nodes may not predict treatment failure in colorectal cancer patients. *Int. J. Cancer* 108, 412–417 (2004).
- Oberg AN, Lindmark GE, Israelsson AC, Hammarstrom SG, Hammarstrom ML: Detection of occult tumor cells in lymph nodes of colorectal cancer patients using real-time quantitative RT-PCR for CEA and CK20 mRNAs. *Int. J. Cancer* 111, 101–110 (2004).
- Keilholz U, Willhauck M, Rimoldi D *et al.*: Reliability of reverse transcription-polymerase chain reaction (RT-PCR)-based

- assays for the detection of circulating tumor cells: a quality-assurance initiative of the EORTC Melanoma Cooperative Group. *Eur. J. Cancer* 34, 750–753 (1998).
- **Demonstration of the lack of reliability of conventional reverse transcriptase (RT-PCR)-based assay.**
27. Lacroix J, Doeberitz MK: Technical aspects of minimal residual disease detection in carcinoma patients. *Semin. Surg. Oncol.* 20, 252–264 (2001).
 28. Heid CA, Stevens J, Livak KJ, Williams PM: Real time quantitative PCR. *Genome Res.* 6, 986–994 (1996).
 29. Bustin SA: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25, 169–193 (2000).
 30. Bustin SA, Nolan T: Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* 15, 155–166 (2004).
 31. Bustin SA: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 29, 23–39 (2002).
 32. Muller MC, Hordt T, Paschka P *et al.*: Standardization of preanalytical factors for minimal residual disease analysis in chronic myelogenous leukemia. *Acta Haematol.* 112, 30–33 (2004).
 33. Bustin SA: *A-Z of quantitative PCR*. IUL Press, La Jolla, CA, USA (2004).
 34. Afzal MA, Osterhaus AD, Cosby SL *et al.*: Comparative evaluation of measles virus-specific RT-PCR methods through an international collaborative study. *J. Med. Virol.* 70, 171–176 (2003).
 35. Niesters HG: Quantitation of viral load using real-time amplification techniques. *Methods* 25, 419–429 (2001).
 36. Gabert J, Beillard E, van der Velden VH *et al.*: Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia – a Europe Against Cancer program. *Leukemia* 17, 2318–2357 (2003).
 37. Niesters HG: Molecular and diagnostic clinical virology in real time. *Clin. Microbiol. Infect.* 10, 5–11 (2004).
 38. Bustin SA: Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences. *Expert Rev. Mol. Diagn.* 5, 493–498 (2005).
 39. Ladanyi A, Soong R, Tabiti K, Molnar B, Tulassay Z: Quantitative reverse transcription-PCR comparison of tumor cell enrichment methods. *Clin. Chem.* 47, 1860–1863 (2001).
 40. Vlems F, Soong R, Diepstra H *et al.*: Effect of blood sample handling and reverse transcriptase-polymerase chain reaction assay sensitivity on detection of CK20 expression in healthy donor blood. *Diagn. Mol. Pathol.* 11, 90–97 (2002).
 41. Max N, Willhauck M, Wolf K *et al.*: Reliability of PCR-based detection of occult tumor cells: lessons from real-time RT-PCR. *Melanoma Res.* 11, 371–378 (2001).
 42. Zippelius A, Lutterbuse R, Riethmuller G, Pantel K: Analytical variables of reverse transcription-polymerase chain reaction-based detection of disseminated prostate cancer cells. *Clin. Cancer Res.* 6, 2741–2750 (2000).
 43. Vlems FA, Diepstra JH, Cornelissen IM *et al.*: Limitations of cytokeratin 20 RT-PCR to detect disseminated tumor cells in blood and bone marrow of patients with colorectal cancer: expression in controls and downregulation in tumor tissue. *Mol. Pathol.* 55, 156–163 (2002).
 44. Raggi CC, Verderio P, Pazzagli M *et al.*: An Italian program of external quality control for quantitative assays based on real-time PCR with Taq-Man probes. *Clin. Chem. Lab. Med.* 43, 542–548 (2005).
 - **Demonstration of the need for strict quality control of real-time PCR assays.**
 45. Willey JC, Crawford EL, Jackson CM *et al.*: Expression measurement of many genes simultaneously by quantitative RT-PCR using standardized mixtures of competitive templates. *Am. J. Respir. Cell Mol. Biol.* 19, 6–17 (1998).
 - **Application of a fully standardized PCR technology to expression quantification.**
 46. Crawford EL, Warner KA, Khuder SA, Zahorchak RJ, Willey JC: Multiplex standardized RT-PCR for expression analysis of many genes in small samples. *Biochem. Biophys. Res. Commun.* 293, 509–516 (2002).
 47. Willey JC, Crawford EL, Knight CR *et al.*: Standardized RT-PCR and the standardized expression measurement center. *Methods Mol. Biol.* 258, 13–41 (2004).
 48. Willey JC, Knight CR, Crawford EL *et al.*: Use of standardized reverse transcription-polymerase chain reaction and the standardized expression measurement center in multi-institutional trials to develop meaningful lung cancer classification based on molecular genetic criteria. *Chest* 125, 155S–156S (2004).
 49. Pagliarulo V, George B, Beil SJ *et al.*: Sensitivity and reproducibility of standardized-competitive RT-PCR for transcript quantification and its comparison with real time RT-PCR. *Mol. Cancer* 3, 5 (2004).
 50. Sanchez JA, Pierce KE, Rice JE, Wangh LJ: Linear-after-the-exponential (LATE)-PCR: an advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proc. Natl Acad. Sci. USA* 101, 1933–1938 (2004).
 51. Hartshorn C, Anshelevich A, Wangh LJ: Rapid, single-tube method for quantitative preparation and analysis of RNA and DNA in samples as small as one cell. *BMC Biotechnol.* 5, 2 (2005).
 52. Zetter BR: The cellular basis of site-specific tumor metastasis. *N. Engl. J. Med.* 322, 605–612 (1990).
 53. Ramaswamy S, Golub TR: DNA microarrays in clinical oncology. *J. Clin. Oncol.* 20, 1932–1941 (2002).
 54. Shih W, Chetty R, Tsao MS: Expression profiling by microarrays in colorectal cancer. *Oncol. Rep.* 13, 517–524 (2005).
 55. Alon U, Barkai N, Notterman DA *et al.*: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* 96, 6745–6750 (1999).
 - **The first report comparing expression profiles in normal colon and colon cancer.**
 56. Notterman DA, Alon U, Sierk AJ, Levine AJ: Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res.* 61, 3124–3130 (2001).
 57. Lin YM, Furukawa Y, Tsunoda T *et al.*: Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* 21, 4120–4128 (2002).
 58. Hegde P, Qi R, Gaspard R *et al.*: Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res.* 61, 7792–7797 (2001).
 59. Kwon HC, Kim SH, Roh MS *et al.*: Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum* 47, 141–152 (2004).
 60. Yanagawa R, Furukawa Y, Tsunoda T *et al.*: Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 3, 395–401 (2001).
 61. Bertucci F, Salas S, Eysteries S *et al.*: Gene expression profiling of colon cancer by DNA microarrays and correlation with

- histoclinical parameters. *Oncogene* 23, 1377–1391 (2004).
62. Eschrich S, Yang I, Bloom G *et al.*: Molecular staging for survival prediction of colorectal cancer patients. *J. Clin. Oncol.* 23, 3526–3535 (2005).
63. Arango D, Laiho P, Kokko A *et al.*: Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* 129, 874–884 (2005).
- **Identification of a set of prognostic markers for disease recurrence in lymph node (LN)-positive tumors.**
64. Bustin SA, Dorudi S: The value of microarray techniques for quantitative gene profiling in molecular diagnostics. *Trends Mol. Med.* 8, 269–272 (2002).
65. Hoos A, Nissan A, Stojadinovic A *et al.*: Tissue microarray molecular profiling of early, node-negative adenocarcinoma of the rectum: a comprehensive analysis. *Clin Cancer Res.* 8, 3841–3849 (2002).
- **Describes the problems involved with trying to associate expression profiles with prognosis.**
66. Choi JK, Choi JY, Kim DG *et al.*: Integrative analysis of multiple gene expression profiles applied to liver cancer study. *FEBS Lett.* 565, 93–100 (2004).
67. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA: Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4, R70 (2003).
68. Choi JK, Yu U, Yoo OJ, Kim S: Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* 21, 4348–4355 (2005).
69. Hunter K: Host genetics influence tumor metastasis. *Nature Rev. Cancer* 6, 141–146 (2006).
- **Lucid and thought-provoking review describing the critical role played by the host genetic background in the development of metastases.**
70. Boedefeld WM 2nd, Bland KI, Heslin MJ: Recent insights into angiogenesis, apoptosis, invasion, and metastasis in colorectal carcinoma. *Ann. Surg. Oncol.* 10, 839–851 (2003).
71. Al-Mulla F, Keith WN, Pickford IR, Going JJ, Birnie GD: Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer* 24, 306–314 (1999).
72. Baisse B, Bouzourene H, Saraga EP, Bosman FT, Benhattar J: Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int. J. Cancer* 93, 346–352 (2001).
- **Molecular description of colorectal cancer heterogeneity.**
73. Losi L, Baisse B, Bouzourene H, Benhattar J: Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *Carcinogenesis* 26, 916–922 (2005).
74. Kitahara O, Furukawa Y, Tanaka T *et al.*: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res.* 61, 3544–3549 (2001).
- **Demonstration of the use of laser capture microdissection for identifying epithelium-specific expression patterns.**
75. Hendrix MJ, Seftor EA, Meltzer PS *et al.*: Expression and functional significance of VE-cadherin in aggressive human melanoma cells: role in vasculogenic mimicry. *Proc. Natl Acad. Sci. USA* 98, 8018–8023 (2001).
76. Koukourakis MI, Giatromanolaki A, Harris AL, Sivridis E: Comparison of metabolic pathways between cancer cells and stromal cells in colorectal carcinomas: a metabolic survival role for tumor-associated stroma. *Cancer Res.* 66, 632–637 (2006).
77. Joyce JA: Therapeutic targeting of the tumor microenvironment. *Cancer Cell* 7, 513–520 (2005).