Cancer Treatment Reviews 35 (2009) 201–209

Contents lists available at ScienceDirect

Cancer Treatment Reviews

journal homepage: www.elsevierhealth.com/journals/ctrv

HOT TOPIC

Gene expression profiling in colorectal cancer using microarray technologies: Results and perspectives

Margherita Nannini [a,*], Maria Abbondanza Pantaleo [a], Alessandra Maleddu [a], Annalisa Astolfi [b], Serena Formica [b], Guido Biasco [a,c]

[a] Department of Hematology and Oncology sciences "L. A. Seragnoli", Sant'Orsola-Malpighi Hospital, University of Bologna, Via Massarenti 9, 40138 Bologna, Italy
[b] Pediatric Oncology and Hematology, University of Bologna, Bologna, Italy
[c] Interdepartmental Centre for Cancer Research "G. Prodi", University of Bologna, Bologna, Italy

ARTICLE INFO

Article history:
Received 16 September 2008
Received in revised form 17 October 2008
Accepted 17 October 2008

Keywords:
Colorectal cancer
Gene expression profiling
Microarray
Affymetrix
Liver metastases

SUMMARY

Nowadays molecular biology represents one of the most interesting topics in medical oncology, because it provides a global and detailed view on the molecular changes involved in tumour progression, leading to a better understanding of the carcinogenesis process, to discovering new prognostic markers and novel therapeutic targets.

The gene expression profiling analysis with microarray technology has shown a great potential in cancer research and in medical oncology, mapping simultaneously the expression of thousands of genes in a single tumour sample and giving a measurement of articulated genes expression patterns.

Colorectal cancer represents a wide and exciting area of research for molecular biology, due to the growing need of a molecular classification as well as prognostic and predictive molecular factors that may guide oncologists in patient's clinical management.

The aim of this review is to analyze the state of art of gene expression profile in colorectal cancer using microarrays technologies and to explore some perspectives in this research field.

© 2008 Elsevier Ltd. All rights reserved.

Introduction

Nowadays molecular biology represents one of the most interesting topics in medical oncology, because it provides a global and detailed view on the molecular changes involved in tumour progression, leading to a better understanding of the carcinogenesis process, to discovering new prognostic markers and novel therapeutic targets. Despite of clinical and pathological parameters are available for the classification and prognostic stratification of cancer, they may be inadequate in everyday practice due to the great biologic and genetic heterogeneity of this multiform disease.

Moreover in the era of patient-tailored therapies, a more comprehensive knowledge of downstream signalling pathways is strictly necessary for the discovery of novel tumour targets as well as for the development of new biological drugs.

In this scenario one of the most innovative and promising fields of research is gene expression profiling analysis with microarray technologies which allows to simultaneously map the expression of thousands of genes in a single tumour sample and gives a measurement of articulated genes expression patterns.

Briefly, microarrays are microscope glass slides containing thousands of addressable genes that are used as probes to quantify the relative amount of RNA transcripts extracted from fresh tissue on the basis of the fluorescent signal produced by the labelled cDNA bound to the microarrays. Two different colours are used to visualize the difference between up- and down-regulation of RNA gene expression. The most currently used microarray platforms are *spotted cDNA microarrays* and *high-density oligonucleotide microarrays*. They are both effective and differ in technological and analytical features. Advantages and limits are listed on Table 1.

In comparison with traditional methods, such as Northern Blotting and Southern Blotting, the microarray technologies have the higher density of probes used that leads to a greater specificity, the miniaturization that allows to increase the number of genes studied simultaneously and the faster hybridization kinetic. Furthermore the commonly used Affymetrix GeneChip system, the main oligonucleotide microarray platform, allows recovering samples after hybridization to a chip and consequently to perform serial hybridization reactions to different arrays, when available biological material is limited. However this technique is not oversimplified and has also some limitations related to sample collection, RNA amplification and data analysis with multiple testing procedures, leading inevitably to a biological and technical variability of microarray experiments (Table 2).[1,2]

* Corresponding author. Tel.: +39 051 6363680; fax: +39 051 6364037.
E-mail address: maggie.nannini@gmail.com (M. Nannini).

0305-7372/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ctrv.2008.10.006

Table 1
cDNA microarrays versus high-density oligonucleotide microarrays.

	cDNA microarray	Oligonucleotide microarray
Probes	PCR products of cDNA clones Advantages: – They can be made by individual investigators. – Easily customizable. – They do not require *a priori* knowledge of cDNA sequence. Limits: – Difficulty of making high-quality arrays	Oligonucleotides Advantages: – They can be made by individual investigators – Greater specificity – Uniformity of probe length – Ability to discern splice variants Limits: – Sequence availability (solved by sequencing of human genome)
Fluorescent labelling	Test sample and reference are labelled with two fluorophores (Cy3 and Cy5)	Single color fluorescent label (biotin)
Readout of gene expression	Two-color readout	Single-color readout

The simultaneous analysis of thousands of genes makes microarrays a promising tool in medical oncology with great potential clinical applications: from molecular diagnosis to molecular classification of cancers, from patients stratification to prognosis prediction, from new targets discovery to tumour response prediction (Fig. 1).[3]

The most studied cancer with microarrays technologies is breast cancer that represents a model for gene expression profiling experiments of solid tumours. Perou et al. provided a distinctive molecular portrait of 65 breast adenocarcinomas, from 42 different individuals, using complementary microarrays, suggesting that tumours could be classified into subtypes distinguished by pervasive differences in their gene expression patterns.[4] Subsequently they found a correlation between those subtypes and clinical outcome, suggesting that gene expression patterns of tumours have both a taxonomic and prognostic value.[5] Some *class discovery* studies have demonstrated that distinctive genetic signatures have the capacity to discriminate breast tumours on the basis of histological type, estrogen receptor status and lymph node status and may also distinguished sporadic forms from hereditary ones.[6–8]

Furthermore some *class prediction* studies found a correlation between gene expression patterns and clinical outcome.[9–12] Finally some studies have shown that gene expression profiling may also predict treatment response in breast cancer.[13–16]

The aim of this review is to analyze the state of art of gene expression profile in colorectal cancer (CRC) using microarrays technologies and to explore its perspectives in this research field.

Potential clinical applications

CRC represents an interesting field of microarrays research for several reasons:

(1) it is considered a biological model of tumorigenesis, because clinical progression from adenoma to early stage carcinoma until advanced stage carcinoma seems to parallel distinctive molecular alterations.[17]
(2) traditional clinical and pathological parameters are not always sufficient to discriminate high risk from low risk CRC and validated molecular markers with prognostic value are still not available.
(3) several cytotoxic and biological drugs are widely used in everyday clinical practice, but as of now the data on molecular markers with predictive value of tumour response are controversial.[18]

Consequently the potential clinical applications of microarray technologies in CRC can be divided in the following subgroups, that we analysed separately (Table 3):

(1) studies on carcinogenesis process,
(2) studies on prognosis prediction,
(3) studies on treatment response prediction.

Studies on carcinogenesis process

The application of microarrays technologies on carcinogenesis studies purposes to identify specific alterations on gene expression according to tumour development and to diagnose and classify tumours on the basis of molecular features. Studies of *class comparison* between normal mucosa, adenoma and carcinoma or between primary tumour and metastases, as well as between left-side and right-side tumours are performed, in order to discover distinctive genetic signatures belonging to each (Table 3).

The studies conducted with this aim are several and are summarized on Table 4.

In 1999, Alon and colleagues reported a clustered data set of 2000 genes able to separate 22 normal and 40 tumour colon tissues with the highest minimal intensity across samples.[19] Furthermore removing the 1500-genes with the most significance differences between tumour and normal tissue, such as muscle-specific genes, a clustering algorithm with a 500-gene set still effectively separates two different tissue samples.[19] Notterman et al. detected 4000 genes with a 4–10 fold difference in expression intensity between tumour and normal tissue ($p < 0.001$), discovering three

Table 2
Main limitations of microarrays experiment.

	Description	Limitations
Samples collection	– 10–40 mg of RNA (100 mm^3 piece of fresh tissue) – Snap-freezing in liquid nitrogen and storing at −80 °C	– Need of fresh tissue only – Modification of mRNA even a few minutes after surgical procedures – Variability in sample preparation – Biological variability among tumours (different cellular composition, genetic heterogeneity)
Technical procedures	– Choice of platform – Hybridization of probes	– Differences between individual constructed spotted arrays – Quality of home-made arrays – Non-specific cross-hybridization of probes
Data analysis	– Unsupervised analysis: multiple tumour samples are clustered into groups based on similarity of their gene expression profiles (*class discovery*) – Supervised analysis: multiple tumour samples from different known classes are used as train model to classify unknown samples (*class prediction*) – Statistical analysis: multiple testing procedures	– Clustering may not necessarily reflect interesting clinical or biologic differences – Frequent absence of overlapping between unsupervised and supervised analysis

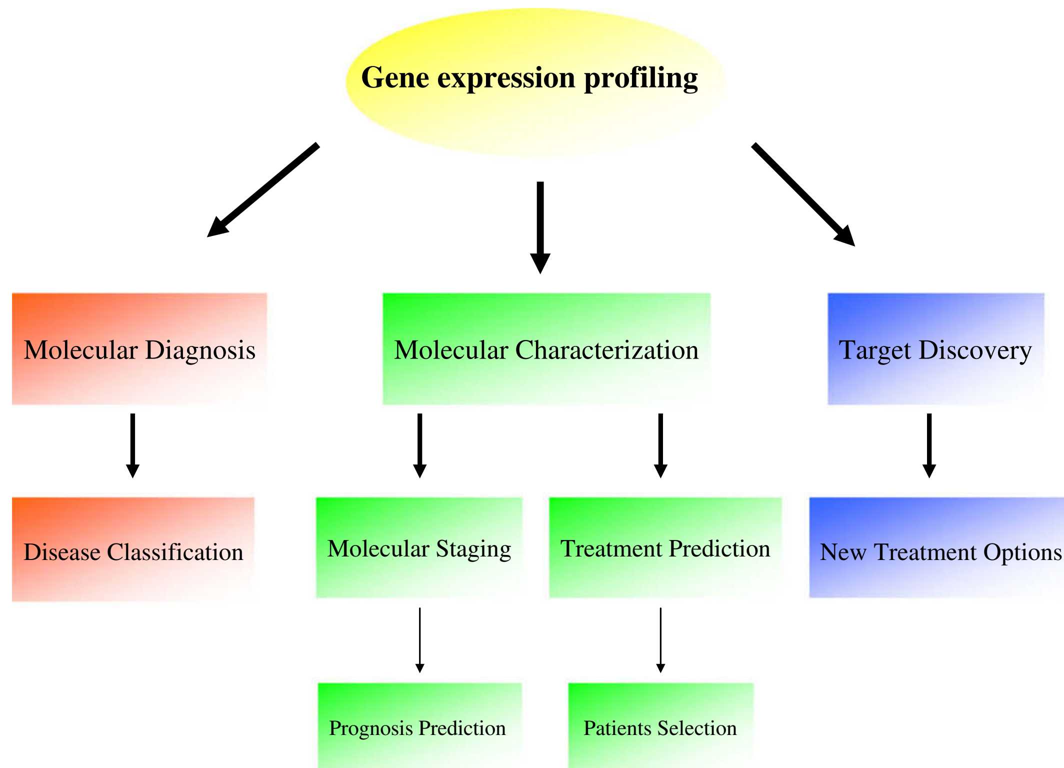

Fig. 1. Potential clinical application of gene expression profiling in medical oncology.

Table 3
Microarray in colorectal cancer: current potential clinical applications, aims and designs of studies.

Field of research	Aims of study	Design of study
Carcinogenesis process	– Identification of gene expression alterations according to tumour development – Molecular classification	*Class comparison:* – Comparison between normal colonic mucosa, adenoma and carcinomas – Comparison between primary tumours and metastases – Comparison between left-side tumours and right-side tumours
Prognosis prediction	– Identification of gene expression alterations according to tumour stage – Identification of gene expression patterns associated with metastatic potential of primary tumour – Molecular stratification according to clinical outcome	*Class comparison*: Comparison between primary tumour of different stage, metastatic vs. non metastatic primary tumours, primary tumours vs. metastases ↓ *Class prediction* Supervised analysis
Treatment response prediction	– Identification of gene expression patterns according to treatment response	*Class comparison*: Comparison between responders vs. non responders ↓ *Class prediction* Supervised analysis

main clusters of genes that seem to differentiate adenomas, carcinomas and their matched normal samples.[20] Subsequent studies reported other sets of genes ranging from 50 to 2632 with a different cut-off of expression intensity value that were differentially expressed between cancer and normal tissue and therefore potentially involved in the development of colorectal carcinogenesis.[21–34] Whereas Frederiksen et al. were not able to separate stage I and IV CRC into distinct clusters and Kwong et al. failed to discriminate stages II, III and IV CRC at the molecular level, other authors reported stage-specific gene expression patterns.[26,31,35,36] Friederichs et al. found specific sets of genes differently expressed between locally restricted CRC and normal mucosa, between lymphatically metastasized CRC and locally restricted CRC, between systematically metastasized CRC and lymphatically metastasized CRC and finally between systematically metastasized CRC and locally restricted CRC.[32]

In order to clarify the molecular modifications underlying the development of metastases, some studies compared the gene expression profile of primary tumours with their corresponding metastases.[24,28,37–41] Agrawal et al. reported that among all genes associated with disease progress, osteopontin expression seemed to be the leading candidate: in fact whereas osteopontin expression was not detectable in normal mucosa and adenoma specimens, increased osteopontin expression from invasive cancer to metastatic primary tumour and liver metastases was found.[24] In particular osteopontin expression in liver metastases was 10–20-fold over adenomas and normal mucosa, making it a candidate tumour marker with clinical utility.[24] Several genes that may play an important role in the progression of metastases in CRC were found, especially genes involved in cell proliferation, invasion, adhesion and tissue remodelling and most studies reported peculiar genetic signatures of metastatic tissue in comparison with primary tumour.[37–41] On the contrary, Koehler et al. compared the expression patterns of six tumours and corresponding metastases from the same patients and found that the expression levels of most differentially expressed genes varied moderately, suggesting that

Table 4
Studies on carcinogenesis process.

First authors	Years	Platform	Samples	Design of study	Results	Validation techniques	Ref.
Alon	1999	Affymetrix	40 Carcinomas 22 Normal colon	Carcinoma vs. normal mucosa	Identification of 2000 discriminating genes	–	[19]
Notterman	2001	Affymetrix	10 Carcinomas 4 Adenomas 22 Normal colon	Carcinoma vs. adenoma vs. normal mucosa	Identification of 4000 discriminating genes	RT-PCR	[20]
Kitahara	2001	cDNA	8 Carcinomas 8 Corresponding normal mucosa	Carcinoma vs. normal mucosa	Identification of 235 discriminating genes	RT-PCR	[21]
Lin	2002	cDNA	9 Adenomas 11 Carcinomas 20 Corresponding normal mucosa	Adenoma vs. carcinoma	Identification of 50 discriminating genes	RT-PCR	[22]
Zou	2002	cDNA	9 Carcinomas 8 Normal mucosa	Carcinoma vs. normal mucosa	Identification of 250 discriminating genes	RT-PCR	[23]
Agrawal	2002	Affymetrix	60 Primary tumours of multiple stages 10 Normal mucosa	Carcinoma vs. liver metastases vs. normal mucosa	Identification of 339 genes of tumour progression	Northern blot	[24]
Williams	2003	cDNA	20 Carcinomas 20 Corresponding normal mucosa	Carcinoma vs. normal mucosa	Identification of 2632 discriminating genes	RT-PCR	[25]
Frederiksen	2003	Affymetrix	20 Primary CRC of multiple stages 5 Normal mucosa	Molecular classification	Identification of 74 and 124 discriminating genes depending on gene selection method	–	[26]
Kwon	2004	cDNA	12 Carcinomas 12 Corresponding normal mucosa	Carcinoma vs. normal mucosa	Identification of 112 discriminating genes	RT-PCR	[27]
Koehler	2004	cDNA	25 Primary tumours of multiple stages 25 Corresponding normal mucosa 14 Liver metastases	Primary tumour vs. normal mucosa Primary tumour vs. metastases	Identification of 40 discriminating genes No one gene differentially expressed	RT-PCR	[28]
Bertucci	2004	cDNA	22 Carcinomas 23 Normal mucosa	Carcinoma vs. normal mucosa Right side vs. left side primary tumour	Identification of 245 discriminating genes Identification of 46 discriminating genes	IHC	[29]
Croner	2005	Affymetrix	30 Carcinomas 30 Corresponding normal mucosa	Carcinoma vs. normal mucosa	Identification of 451 discriminating genes	Previous 5 published studies as training test	[30]
Kwon	2005	cDNA	28 Primary tumours of multiple stage 10 Normal mucosa 10 Liver metastases	Carcinoma vs. normal mucosa Primary tumour vs. metastases	Identification of 70 discriminating genes Identification of 151 discriminating genes	2D-PAGE Mass spectrometry	[31]
Friederichs	2005	Affymetrix	25 Primary tumours of multiple stages 6 Normal mucosa	Carcinoma vs. normal mucosa Molecular classification	Identification of 1995 discriminating genes Identification of 1680 discriminating genes between stage II CRC and normal mucosa Identification of 9 discriminating genes between stage II and III CRC Identification of 50 discriminating genes between stage II and IV CRC Identification of 19 discriminating genes between stage III and IV CRC	–	[32]
Bianchini	2007	cDNA	25 Primary tumours 13 Normal mucosa	Carcinoma vs. normal mucosa	Identification of 584 discriminating genes	RT-PCR WB IHC	[33]
Grade	2007	Oligonucleotide arrays	73 Primary tumours 30 Normal mucosa	Carcinoma vs. normal mucosa	Identification of 17 discriminating genes	RT-PCR	[34]
Groene	2006	Affymetrix	36 Primary tumours	Molecular classification of stage II and stage III CRC	Identification of 45 genes discriminating	RT-PCR	[35]
Croner	2008	Affymetrix	80 Primary tumours of different stage	Molecular classification	Identification of 50 discriminating genes	Independent clinical study	[36]
Yanagawa	2001	cDNA	10 Primary tumours 10 Corresponding metastatic lesions	Primary tumour vs. metastases	Identification of 47 genes differently expressed	RT-PCR	[37]
Li	2004	cDNA	14 Primary mCRC 11 Primary non mCRC	Primary tumour vs. metastases	Identification of 429 discriminating genes	RT-PCR	[38]
Kleivi	2007	cDNA	18 Primary tumours 4 Liver metastases 4 Carcinomatoses	Primary tumour vs. metastases	Identification of 89 discriminating genes	RT-PCR	[39]
Ki	2007	cDNA	27 Primary tumours 27 Liver metastases	Primary tumour vs. metastases	Identification of 46 discriminating genes	RT-PCR	[40]
Lin	2007	cDNA	48 Primary tumours 28 Liver metastases	Primary tumour vs. metastases	Identification of 778 discriminating genes	RT-PCR	[41]

Table 4 (*continued*)

First authors	Years	Platform	Samples	Design of study	Results	Validation techniques	Ref.
Komuro	2005	cDNA	89 primary tumours	Right side vs. left side primary tumour	Identification of 191 discriminating genes	–	[42]
Birkenkamp-Demtroder	2005	Affymetrix	25 Primary tumours	Right side vs. left side normal mucosa	Identification of 160 discriminating genes	RT-PCR	[43]
				Right side primary tumour vs. right side normal mucosa	Identification of 118 discriminating genes		
			20 Normal mucosa	Left side primary tumour vs. left side normal mucosa	Identification of 186 discriminating genes		
				Right side vs. left side primary tumour	Identification of 44 discriminating genes		
Kita	2006	Affymetrix	12 Flat adenomas	Flat adenomas vs. normal mucosa	Identification of 180 discriminating genes	RT-PCR	[44]
			12 Adjacent normal mucosa	Left-side flat adenomas vs. left-side normal mucosa	Identification of 49 discriminating genes		
				Right-side flat adenomas vs. right-side normal mucosa	Identification of 89 discriminating genes		
Kim	2008	Oligonucleotide microarrays	5 Serrated adenomas and their corresponding normal mucosa	Serrated adenomas vs. normal mucosa	Identification of 124 discriminating genes	RT-PCR	[45]

many expression modifications occur during an early phase of the carcinogenesis process and only few alterations are associated with metastatic progression.[28]

Some studies have also investigated differences in gene expression between CRC of the right side and left side, due to their epidemiological, morphological and pathogenetic diversity and found distinct profiles according to the anatomical stratification.[29,42,43]

Finally two recent studies reported significant differences in gene expression profile between both flat and serrated adenoma and normal mucosa, suggesting that different mechanisms of development of these precancerous lesions may exist.[44,45]

Although a plethora of data on gene expression modifications that occur during the carcinogenesis process has been reached, no reliable biomarkers or signatures useful in clinical practise for molecular diagnosis and classification of CRC has been found. The lack of consistent overlap of results may be explained by some technical reasons, including different platforms used, different tissue collection methods and different statistic and analysis methods.

However a recent meta-analysis on 25 studies on CRC gene expression profiling published between 2000 and 2007 observed that, despite of some incongruence, the overall overlap in cancer versus normal mucosa comparison and in adenoma versus normal mucosa were statistically significant ($p < 0.0001$).[46]

In conclusion, the main considerations emerging from these studies are the following: first of all, the carcinogenesis process develops across genetic modifications paralleling clinical progression; secondly, most genes involved during carcinogenesis process are implicated in cell proliferation, migration and adhesion, but some genes with unknown functions still exist and need further genomic and proteomic investigation; third, distinctive genetic signatures of every stage of tumour development may exist but more homogeneous studies are required in order to obtain a reliable molecular classification useful in clinical practice; finally the identification of genes differently expressed during tumour progression may lead to a stage-specific tumour treatment as well as to the discovery of novel therapeutic targets.

Studies on prognosis prediction

Studies on prognosis prediction performed with microarrays technologies aim to identify specific alterations to the gene expression profile that may be useful to discriminate high risk from low risk CRC, to provide a molecular stratification according to the clinical outcome and to predict the metastatic potential of the primary tumour. They are defined as "studies of *class prediction*", based on supervised analyses using the already known genes differently expressed between metastatic and non metastatic primary tumours and between primary tumours and metastases as train model to discriminate patients with different prognosis (Table 3).

Several studies have been conducted with these purposes and they are summarized on Table 5.

Bertucci et al. found 194 discriminating genes differently expressed in primary tumours with and without metastases able to divide patients with significantly different 5-years survival ($p = 0.005$) [29].

The controversial data on the benefit of adjuvant chemotherapy in stage II CRC led to the identification of molecular prognostic factors, that may identify stage II CRC patients who develop disease recurrence and may benefit by adjuvant treatment. Wang et al. studied the gene expression profile of this subset of patients and, using two supervised class prediction approaches of analysis, they identified a 23-genes signature that may predict recurrence in stage II CRC with 78% accuracy.[47]

The development of a "molecular prognostic system" represents a big challenge in the modern era of oncology. Some studies tried to identify gene expression patterns capable to discriminate good from poor prognosis in stage II and stage III CRC better than pathological staging system.[48–53] Eschrich et al. proposed a molecular staging, based on 43 genes that was 90% accurate in predicting 36-months overall survival in 78 patients with stage II and stage III CRC.[48] Arango et al. identified a 17-gene signature that divides stage III CRC into two groups with significant different disease-free survival ($p < 0.0001$) after surgery.[49] Barrier et al. studied a small series of stage II and III CRC patients, who developed or did not develop recurrence, in order to identify a prognostic pattern of expression in tumours and normal mucosa discriminating high from low risk patients.[50,51] They found a 30-gene tumour-based predictor and a 70-gene normal mucosa-based predictor with a 78% and 83% accuracy respectively, suggesting that also non neoplastic mucosa may have a prognostic value due to its strictly interaction with tumour tissue.[51] The accuracy of those two prognostic signatures were confirmed in subsequent studies by the same authors.[52,53] A prognostic pattern of expression discriminating stage III and IV CRC patients with favourable and unfavourable clinical course may exist: in a pool of 136 differently expressed genes, Cavalieri et al. identified 8 genes significantly associated with survival.[54] Furthermore using a pathway analysis on the microarray gene expression data, 48 pathways differently regulated in the favourable and unfavourable prognosis groups were found.[54]

Table 5
Studies on prognosis prediction.

First authors	Years	Platform	Samples	Aim of study	Results	Validation techniques	Ref.
Bertucci	2004	cDNA	22 Primary tumours 23 Normal mucosa	Prognosis prediction	Identification of 194 gene prognostic signature	IHC	[29]
Wang	2004	Affymetrix	74 Primary tumours (stage II)	Prognosis prediction of stage II CRC	Identification of a 23-gene prognostic signature	–	[47]
Eschrich	2005	cDNA	75 Primary tumours of different stages	Prognosis prediction of stage II and III CRC	Identification of 43 gene prognostic signature	–	[48]
Arango	2005	Affymetrix	281 Primary tumours (stage III)	Prognosis prediction of stage III CRC		RT-PCR	[49]
Barrier	2005	Affymetrix	12 Primary tumours 12 Corresponding normal mucosa	Prognosis prediction	Identification of 47 gene prognostic signature	–	[50]
Barrier	2005	Affymetrix	18 Primary tumours (stage II and III) 18 Corresponding normal mucosa	Prognosis prediction of stage III CRC	Identification of 30- gene tumour based predictor and 70-gene normal mucosa based predictor	–	[51]
Barrier	2006	Affymetrix	50 Primary tumours (stage II)	Prognosis prediction of stage II CRC	Identification of 30-gene prognostic signature	–	[52]
Barrier	2007	Affymetrix	24 Primary tumours (stage II) 24 Corresponding normal mucosa	Prognosis prediction of stage II CRC	Identification of 70-gene prognostic signature	–	[53]
Cavalieri	2007	cDNA	19 Primary tumours (stage III and IV)	Prognosis prediction of stage III and IV CRC	Identification of 8 gene prognostic signature	RT-PCR	[54]
D'Arrigo	2005	cDNA	10 Primary metastatic tumours 10 Primary non metastatic tumours	Prediction of metastatic potential	Identification of 29 gene prognostic signature	RT-PCR	[55]
Yamasaki	2007	cDNA	58 Primary tumours of different stage 34 Liver metastases	Prediction of metastatic potential	Identification of 119 gene prognostic signature	–	[56]

Table 6
Studies on treatment response prediction.

First authors	Years	Platform	Samples	Aim of study	Results	Validation techniques	Ref.
Mariadason	2003	cDNA	*In vitro*	Prediction of response to 5-FU Prediction of response to CPT-11	Identification of 50 predictive genes Identification of 149 predictive genes	–	[57]
Shimizu	2005	cDNA	*In vitro*	Prediction of response to 5-FU	Identification of 81 predictive genes	–	[58]
Arango	2004	cDNA	*In vitro*	Prediction of response to oxaliplatin	Identification of 80 predictive genes	–	[59]
Del Rio	2007	Affymetrix	21 Primary advanced tumours	Prediction of response to FOLFIRI	Identification of 80 predictive genes	–	[60]
Khambata-Ford	2007	Affymetrix	80 Metastatic lesions	Prediction of response to Cetuximab	Epiregulin and amphiregulin expression	RT-PCR	[61]
Ghadimi	2005	cDNA	30 Locally advanced rectal cancer	Prediction of response to preoperative chemoradiotherapy	Identification of 54 predictive genes	–	[62]
Kim	2007	Affymetrix	31 Locally advanced rectal cancer	Prediction of response to preoperative chemoradiotherapy	Identification of 261 predictive genes	–	[63]
Rimkus	2008	cDNA	43 Locally advanced rectal cancer	Prediction of response to preoperative chemoradiotherapy	Identification of 42 predictive genes	–	[64]

Finally two studies aimed to identify specific expression patterns predicting the metastatic potential of primary tumour.[55,56] D'Arrigo et al. found 37 discriminating genes between 10 radically resected primary CRC from patients who did not develop recurrence within 5-year follow-up and 10 primary CRC from patients with synchronous metastases and suggested that 29 of these genes could be a distinct metastatic fingerprint that may predict the risk of distance relapse.[55] More recently Yamasaki et al. performed both hierarchical clustering and supervised analysis of 28 primary CRC using 119 genes differently expressed between synchronously or metachronously metastasized CRC and liver metastases, that were considered closely associated with metastatic potential.[56] On the basis of the supervised analysis, the expression profile of these genes divided tumours into two classes, localized and metastasized, according to their final metastatic status.[56]

In summary a molecular classification is now necessary for better identifying high risk CRC patients, who require more aggressive treatments or a more rigorous follow up. Because the biological aggressiveness may be related to a different genetic background, it is a reasonable supposition that specific prognostic signatures may exist apart from clinical stage.

Therefore gene expression profile may represent a useful tool for the stratification of patients according to their expression pattern. Until now primary tumour was the main source of information and the field most widely studied. Tumour microenvironment, such as adjacent normal mucosa, may influence and reflect some tumour biological features due to their strictly interaction, therefore its gene expression profile may also be informative. Finally metastatic tissue, although it was less investigated, probably represents the richer field of information and further studies are

required in order to improve the knowledge on the molecular background of metastases.

Studies on treatment response prediction

While gene expression profiling with microarray technologies has been widely applied to CRC for diagnosis, classification and prognosis prediction based on molecular patterns of expression, its application to response prediction to medical treatment is still lacking reliable results due to few currently available studies[57–63] (Table 6).

They are defined as "studies of *class prediction*", based on supervised analyses using already known genes differently expressed between responder and non responder tumour cell lines or patients as train model to discriminate subgroups with different chemosensitivity (Table 3).

On the contrary to what is available in breast cancer, most of the studies were performed only in colon cancer cell lines.[57–59] In a panel of 30 colon carcinoma cell lines Mariadason et al. identified 420 genes correlated with response to 5-fluorouracil (5-FU) and involved in two main biological processes, DNA replication and repair and protein processing/targeting.[57] The predictive value of 50 genes best correlated with 5-FU response was subsequently validated using a "leave one out" cross validation approach and it was higher than the traditional markers, such as thymidylate synthase, thymidine phosphorylase, mismatch repair and p53 status ($p = 0.008$).[57] Furthermore they also found that 149 genes best-correlated with CPT-11-induced apoptosis significantly predicted response of colon cancer cell lines to this agent.[57] An other study reported a genetic signature correlated to *in vitro* 5-FU sensitivity and demonstrated its predictive value.[58] D'Arango et al. found that the 80 genes best correlated with oxaliplatin-induced apoptosis on 30 colon cancer cell lines accurately predicted the response to this drug ($p = 0.002$).[59]

Only two recent studies with microarrays providing a response predictor classifier in CRC were performed in patients.[60,61] In the first one, Del Rio et al. analyzed gene expression profile of 21 primary advanced CRC tissues, in order to identify an expression pattern that could predict response to leucovorin, fluorouracil and irinotecan (FOLFIRI) as first-line treatment: 14 genes were found expressed differently between responders and non responders and were able to predict treatment response with 95% accuracy.[60] In the same year Khambata-Ford et al. investigated gene expression pattern of metastatic biopsies of 80 advanced CRC patients treated with cetuximab to identify genes whose expression correlates with best clinical response.[61] They found that, among 629 genes expressed differently between 25 patients with disease control and 55 non responders, the top candidate markers based on lowest p value were epiregulin and amphiregulin, both ligands for epidermal growth factor receptor (EGFr), suggesting that these markers could select patients for cetuximab therapy.[61]

Finally some studies evaluated the ability of gene expression profiling for predicting response of advanced rectal cancer (RC) to preoperative chemoradiotherapy.[62–64] Ghadimi et al. analyzed gene expression signatures of biopsies from 30 locally advanced RC and found 54 genes differentially expressed between responders and non responders ($p < 0.001$). Using a leave-one-out cross-validation, the estimate of response prediction accuracy was 83%, with a sensitivity of 78% and a specificity of 86%.[62] Kim et al. reported 261 genes differentiated between 20 partial response and 11 complete response patients affected by locally advanced RC treated with preoperative chemoradiotherapy.[63] By a supervised analysis a gene set comprising the top-ranked 95 genes predicted complete and partial response with an 84% accuracy.[63] Similarly a recent study identified a gene expression signature of 42 genes that was able to distinguish responder from non responder locally advanced RC patients with a 71% accuracy.[64]

Several cytotoxic and cytostatic agents are now available for treatment of CRC both in adjuvant and in advanced setting, but the role of predictive factors of tumour response is still controversial.[18]

The interaction between tumour and host genetic background may partly explain the different drug sensitivity of the same disease. Thus it is reasonable that there could be specific gene signatures capable of predicting tumour response.

Gene expression profiling with microarray technologies may represent a useful tool for selecting patients that may benefit from different treatments. In fact a molecular portrait of each patient may allow choosing the most effective therapy from among all approved agents leading to a personalized cancer treatment. However the current data are still insufficient to be translated into clinical practice for two reasons: first of all, the number of studies is too small to guarantee the reliability of predictors; secondly, most of these studies were performed in cancer cell lines, but a patient-tailored therapy needs *in vivo* evaluations assessed in an appropriate host-tumour tissue context. Therefore further large-scale studies are required and an *in vivo* validation of results obtained in an *in vitro* approach is strictly necessary.

Future perspectives

Gene expression profile with microarrays technologies is an innovative and promising area for CRC research by offering a different and deeper point of view of this complex disease.

Until now the study of carcinogenesis process and the identification of gene signatures with prognostic or predictive value represented the main applicative fields of this technique which led to improving the knowledge on molecular mechanisms of tumour development as well as identifying gene patterns strictly correlated with clinical features such as outcome or drug sensitivity.

Despite of these promising results, the translation of microarrays analysis into everyday clinical practice is still far for several reasons: (1) the lack of comparison and overlap of results obtained from each single study due to technique-related variability of sample collections and preparation, type of platform used and data analysis; (2) the lack of large-scale studies due to the small number of available patient samples and (3) the difficulty to understand and select which data would be informative and useful for a reliable clinical application.

Therefore solving these limitations should be the main goal in the future, in order to convert only speculative data into useful clinical information.

First of all, standardized methodological and analytical protocols are required for achieving data as homogeneous as possible so they can be easily compared. Novel approaches to isolating integer and non contaminated tumour tissue, such as laser capture microdissection (LCM) should be investigated. Furthermore all data should be stored in sharing databases available to every investigators. Secondly tissue banks, possibly containing both primary tumour and metastatic specimens, in association with suitable clinical information, may be a solution for the lack of fresh tissue samples available that limits large-population studies. In fact, the availability of biological material represents one of the most important challenges of molecular biology because it allows the performance of serial and different genomic and proteomic analyses on the same sample leading to a global molecular overview of analyzed tissue. Furthermore it could be challenging to investigate the feasibility of microarray analysis from formalin-fixed paraffin-embedded tissue in large series because until now few data on the sensitivity of this approach are available.[65] Thirdly, for extracting

real informative data, small sets of most significant genes should be selected among the pool of all differentially expressed genes and then retrospectively or prospectively validated by conventional quantitative assays, such as reverse-transcriptase polymerase chain reaction (RT-PCR).

It may be relevant to identify specific signatures of metastases in order to personalized treatment patients. In particular, because the liver is the most frequent site of metastases in CRC, liver metastases represent a research field of great interest, that was less investigated and thus less known biologically. In fact, most studies have analyzed metastatic tissue in comparison with primary tumour to better understand the development process of the tumour as well as to identify gene expression patterns predictive of metastatic potential. However little information about its molecular background is available. Many genes codifying for proteins involved in cell adhesion, migration, angiogenesis and proliferation have been linked to the development of colorectal liver metastases, but a real genetic signature for metastatic tissue has still not been defined.[66] Varghese et al. found a site-specific segregation of genes between liver and peritoneal metastases, identifying 23 pathways expressed differently.[67] Recently we studied gene expression profile of synchronous and metachronous liver metastases using Affymetrix platform and identified two different signatures: EGFr pathway was upregulated in metachronous lesions whereas the pathway mainly related to angiogenesis was upregulated in synchronous lesions.[68] These preliminary results, both validated on proteomic assay, suggest that a genetic signature of liver metastases may exist and it can be the basis of treatment choice. Furthermore they sustain that advanced CRC may be a multiform disease, whose medical treatment may be differentiated on the basis of their different molecular features.

Finally it can be particularly challenging to discover novel molecular targets for the development of new biological drugs. The wide list of genes identified in all previous described studies could represent a source of potential therapeutic targets because most of these genes are involved in key-mechanisms of tumour development, from cell proliferation and differentiation to cell survival. As Weir et al. suggested, systematic understanding of the molecular basis of every type of cancer requires al least three steps: comprehensive characterization of recurrent genomic aberrations, elucidation of their biological role in cancer pathogenesis and evaluation of their utility for diagnostics, prognostics and therapeutics .[69] Therefore a comprehensive whole proteomic and genomic characterization, including newer technologies such as microarrays, comparative genomic hybridization, SNPs arrays and miRNA, is strictly necessary to discover novel therapeutic targets and extend the number of treatment options for CRC.

Conclusions

Nowadays the molecular biology represents an endless source of data that may be filtered, elaborated and transformed into useful clinical information. The comprehensive knowledge on molecular background of cancers should be considered one of the main challenges of medical oncology in the future.

In this way CRC represents a wide area of research for molecular biology, due to the growing need of a molecular classification as well as prognostic and predictive molecular factors that may guide the oncologists in patient's clinical management.

Gene expression profile with microarray technologies has been shown to have several potential clinical applications in CRC, ranging from study on mechanisms involved in tumour development, to the identification of gene signatures with prognostic and predictive value and finally to the discovery of novel tumour targets.

In these last years many results have been reached, suggesting that a whole genome approach may significantly change the management of CRC patients, but their translation in clinical practice is still far and requires further steps forward in cancer research.

References

1. Duftner N, Larkins-Ford J, Legendre M, Hofmann HA. Efficacy of RNA amplification is dependent on sequence characteristics: implications for gene expression profiling using a cDNA microarray. *Genomics* 2008;**91**:108–17.
2. Jung SH, Bang H, Young S. Sample size calculation for multiple testing in microarray data analysis. *Biostatistics* 2005;**6**:157–69.
3. Wadlow R, Ramaswamy S. DNA Microarrays in clinical cancer research. *Curr Mol Med* 2005;**5**:111–20.
4. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;**406**:747–52.
5. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001;**98**:10869–74.
6. West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA* 2001;**98**:11462–7.
7. Zhao H, Langerød A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol Biol Cell* 2004;**15**: 2523–36.
8. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *New Engl J Med* 2001;**344**:539–48.
9. van't veer L, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.
10. van de Vijver MJ, He Y, van't Veer L, et al. A gene expression signature as a predictor of survival in breast cancer. *New Engl J Med* 2002;**347**:1999–2009.
11. Glinsky GV, Higashiyama T, Glinskii AB. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res* 2004;**10**:2272–83.
12. Wang Y, Klijn JG, Zhang Y, et al. Gene expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;**365**:671–9.
13. Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;**362**:362–9.
14. Ayers M, Symmans WF, Stec J, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004;**22**:2284–93.
15. Dressman HK, Hans C, Bild A, et al. Gene expression profiles of multiple breast cancer phenotypes and response to neoadjuvant chemotherapy. *Clin Cancer Res* 2006;**12**:819–26.
16. Chang JC, Wooten EC, Tsimelzon A, et al. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol* 2005;**23**:1169–77.
17. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. *New Engl J Med* 1988;**319**:525–32.
18. Spano JP, Milano G, Vignot S, Khayat D. Potential predictive markers of response to EGFR-targeted therapies in colorectal cancer. *Crit Rev Oncol Hematol* 2008;**66**:21–30.
19. 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* 1999;**96**:6745–50.
20. Notterman DA, Alon U, Sierk AJ, et al. Trascriptional gene expression profiles of colorectal adenoma, adenocarcinoma and normal tissue examined by oligonucleotide arrays. *Cancer Res* 2001;**61**:3124–30.
21. 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* 2001;**61**:3544–9.
22. 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* 2002;**21**:4120–8.
23. Zou TT, Selaru FM, Xu Y, et al. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* 2002;**21**:4855–62.
24. Agrawal D, Chen T, Irby R, et al. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 2002;**94**:513–21.
25. Williams NS, Gaynor RB, Scoggin S, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin Cancer Res* 2003;**9**:931–46.
26. Frederiksen CM, Knudsen S, Laurberg S, Ørntoft TF. Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 2003;**129**:263–71.

27. Kwon HC, Kim SH, Roh MS, et al. Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum* 2004;**47**:141–52.
28. Koheler A, Bataille F, Schmid C, et al. Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. *J Pathol* 2004;**204**:65–74.
29. Bertucci F, Salas S, Eysteries S, et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histological parameters. *Oncogene* 2004;**23**:1377–91.
30. Croner RS, Foertsch T, Brueckl WM, et al. Common denominator genes that distinguish colorectal carcinoma from normal mucosa. *Int J Colorectal Dis* 2005;**20**:353–62.
31. Kwong KY, Bloom GC, Yang I, et al. Synchronous global assessment of gene and protein expression in colorectal cancer progression. *Genomics* 2005;**86**:142–58.
32. Friederichs J, Rosenberg R, Mages J, et al. Gene expression profiles of different clinical stages of colorectal carcinoma: toward a molecular genetic understanding of tumor progression. *Int J Colorectal Dis* 2005;**20**:391–402.
33. Bianchini M, Levy E, Zucchini C, et al. Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. *Int J Oncol* 2006;**29**:83–94.
34. Grade M, Hörmann P, Becker S, 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* 2007;**67**:41–56.
35. Groene J, Mansmann U, Meister R, et al. Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III. *Int J Cancer* 2006;**119**:1829–36.
36. Croner RS, Förtsch T, Brückl WM, et al. Molecular signature for lymphatic metastasis in colorectal carcinomas. *Ann Surg* 2008;**247**:803–10.
37. 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* 2001;**5**:395–401.
38. Li M, Lin Y-M, Hasegawa S, et al. Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. *Int J Oncol* 2004;**24**:305–12.
39. Kleivi K, Lind GE, Diep CB, et al. Gene expression profiles of primary colorectal carcinomas, liver metastases, and carcinomatoses. *Mol Cancer* 2007;**6**(2).
40. Ki DH, Jeung HC, Park CH, et al. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int J Cancer* 2007;**121**:2005–12.
41. Lin HM, Chatterjee A, Lin YH, et al. Genome wide expression profiling identifies genes associated with colorectal liver metastasis. *Oncol Rep* 2007;**17**:1541–9.
42. Komuro K, Tada M, Tamoto E, et al. Right- and left-sided colorectal cancers display distinct expression profiles and the anatomical stratification allows a high accuracy prediction of lymph node metastasis. *J Surg Res* 2005;**124**:216–24.
43. Birkenkamp-Demtroder K, Olesen SH, Sørensen FB, et al. Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid. *Gut* 2005;**54**:374–84.
44. Kita H, Hikichi Y, Hikami K, et al. Differential gene expression between flat adenoma and normal mucosa in the colon in a microarray analysis. *J Gastroenterol* 2006;**41**:1053–63.
45. Kim K, Park U, Wang J, et al. Gene profiling of colonic serrated adenomas by using oligonucleotide microarray. *Int J Colorectal Dis* 2008;**23**:569–80.
46. Chan SK, Griffith OL, Tai IT, Jones SJ. Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers. *Cancer Epidem Biomar Prev* 2008;**17**:543–52.
47. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Duke's B Colon Cancer. *J Clin Oncol* 2004;**22**:1564–71.
48. Eschrich S, Yang I, Bloom G, et al. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 2005;**23**:3526–35.
49. Arango D, Laiho P, Kokko A, et al. Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* 2005;**129**:874–84.
50. Barrier A, Boelle PY, Lemoine A, et al. Gene expression profiling of nonneoplastic mucosa may predict clinical outcome of colon cancer patients. *Dis Colon Rectum* 2005;**48**:2238–48.
51. Barrier A, Lemoine A, Boelle PY, et al. Colon cancer prognosis prediction by gene expression profiling. *Oncogene* 2005;**24**:6155–64.
52. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 2006;**24**:4685–91.
53. Barrier A, Roser F, Boëlle PY, et al. Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene* 2007;**26**:2642–8.
54. Cavalieri D, Dolara P, Mini E, et al. Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. *Oncol Res* 2007;**16**:535–48.
55. D'Arrigo A, Belluco C, Ambrosi A, et al. Metastatic transcriptional pattern revealed by gene expression profiling in primary colorectal carcinoma. *Int J Cancer* 2005;**115**:256–62.
56. Yamasaki M, Takemasa I, Komori T, et al. The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. *Int J Oncol* 2007;**30**:129–38.
57. Mariadason JM, Arango D, Shi Q, et al. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res* 2003;**63**:8791–812.
58. Shimizu D, Ishikawa T, Ichikawa Y, et al. Prediction of chemosensitivity of colorectal cancer to 5-fluorouracil by gene expression profiling with cDNA microarrays. *Int J Oncol* 2005;**27**:371–6.
59. Arango D, Wilson AJ, Shi Q, et al. Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *Br J Cancer* 2004;**91**:1931–46.
60. Del Rio M, Molina F, Bascoul-Mollevi C, 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* 2007;**25**:773–80.
61. Khambata-Ford S, Garrett CR, Meropol NJ, et al. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol* 2007;**25**:3230–7.
62. Ghadimi BM, Grade M, Difilippantonio MJ, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. *J Clin Oncol* 2005;**23**:1826–38.
63. Kim IJ, Lim SB, Kang HC, et al. Microarray gene expression profiling for predicting complete response to preoperative chemoradiotherapy in patients with advanced rectal cancer. *Dis Colon Rectum* 2007;**50**:1342–53.
64. Rimkus C, Friederichs J, Boulesteix AL, et al. Microarray-based prediction of tumor response to neoadjuvant radiochemotherapy of patients with locally advanced rectal cancer. *Clin Gastroenterol Hepatol* 2008;**6**:53–61.
65. Linton KM, Hey Y, Saunders E, et al. Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours. *Br J Cancer* 2008;**98**:1403–14.
66. Nadal C, Maurel J, Gascon P. Is there a genetic signature for liver metastasis in colorectal cancer? *World J Gastroenterol* 2007;**13**:5832–44.
67. Varghese S, Burness M, Xu H, Beresnev T, Pingpank J, Alexander HR. Site-specific gene expression profiles and novel molecular prognostic factors in patients with lower gastrointestinal adenocarcinoma diffusely metastatic to liver or peritoneum. *Ann Surg Oncol* 2007;**14**:3460–71.
68. Pantaleo MA, Astolfi A, Nannini M, Ercolani, et al. Gene expression profiling differences between synchronous and metachronous liver metastases of colorectal cancer. AACR-NCI-EORTC conference on molecular targets and cancer therapeutics. San Francisco; October 2007.
69. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;**450**:893–8.