Prognostic and predictive markers in colorectal cancer:

implications for clinical management

Only two biomarkers for colorectal cancer are currently used in the clinic. However, efforts to find genetic patterns that distinguish between tumours with good or poor prognosis, or between patients who do or don't respond to various therapies, may offer the basis for identifying subgroups of colorectal cancer similar to those now used in breast cancer.

Colorectal cancer is a very heterogeneous disease, possibly even different diseases hitting the same organ. This has huge implications for clinical practice. For example, in the adjuvant setting, our ability to accurately predict the prognosis for a patient is around 50% in stage II/III resected disease. This is the clinical reality we face every day, so we are unable to inform our patients of their prognosis with more than about 50% accuracy.

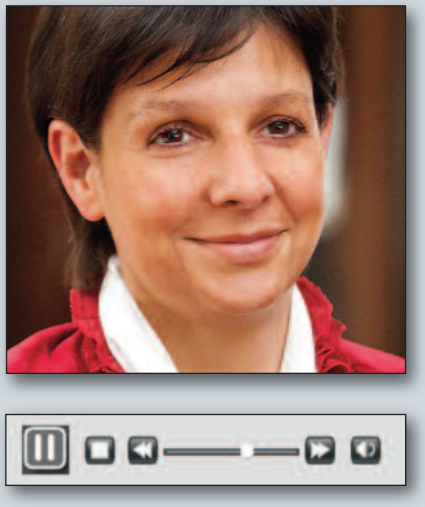
Even our best-guess models, based on traditional histopathological markers such as that lymph node metastases would be associated with a worse outcome than no lymph node metastases, are not straightforward. For example, some patients who have no lymph node metastases but have T4b tumours fare worse than patients with lymph node metastases (see table overleaf). This indicates that our current understanding of how colorectal cancer behaves in the body and metastasises is probably flawed.

Colorectal cancer is also heterogeneous in the metastatic setting. This is where drug efficacy needs to be pre-

European School of Oncology
e-grandround

The European School of Oncology presents weekly e-grandrounds which offer participants the opportunity to discuss a range of cutting-edge issues, from controversial areas and the latest scientific developments to challenging clinical cases, with leading European experts in the field. One of these is selected for publication in each issue of *Cancer World*.

In this issue, Sabine Tejpar, from the University Hospital Gasthuisberg, Leuven, Belgium, provides an update on the implications for clinical management of developments in prognostic and predictive markers for colorectal cancer (CRC). Daniel Helbling, Onkozentrum Zurich, Switzerland, poses questions arising during the e-grandround live presentation. It was summarised by Susan Mayor.



The recorded version of this and other e-grandrounds is available at www.e-eso.net

RELATION BETWEEN TUMOUR SUBSTAGE AND SURVIVAL

Category	SEER				SEER	
TN	Relative survival, 5-year (%)	SE	TNM stage, 6th ed	TNM stage, 7th ed	Observed survival, 5-year (%)	SE
T1N0	97.4	0.6	I	I	78.7	0.5
T2N0	96.8	0.6	I	I	74.3	0.4
T3N0	87.5	0.4	IIA	IIA	66.7	0.6
T4aN0	79.6	1.0	IIB	IIB	60.6	0.8
T4bN0	58.4	1.3	IIB	IIC	45.7	1.0
T1-2N1a	90.7	1.5	IIIA	IIIA	73.7	1.2
T1-2N1b	83.0	2.0	IIIA	IIIA	67.2	1.6
T1-2N2a	79.0	3.6	IIIC	IIIA/IIIB	64.7	3.0
T3N1a	74.2	0.8	IIIB	IIIB	58.2	0.6
T4aN1a	67.6	2.0	IIIB	IIIB	52.2	1.5

These relative survival figures, based on expanded SEER data and presented according to AJCC substaging for stage II and III colon cancers, indicate flaws in our current understanding of how colorectal cancer behaves

SEER – Surveillance, Epidemiology and End Results. *Source:* SB Edge, DR Byrd, CC Compton et al. (eds) (2010) AJCC Cancer Staging Manual, 7th edn. Springer, reprinted with permission © Springer 2010

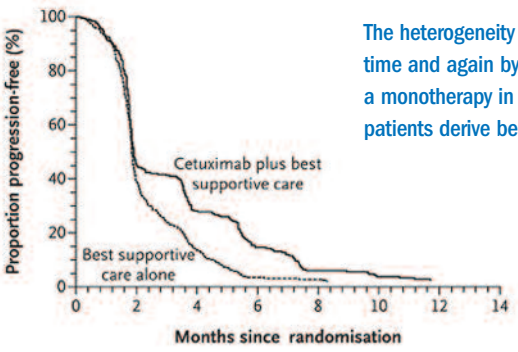
dicted to obtain the best possible outcome for the patient. With the recent drugs, not just targeted agents but also chemotherapy, we have accepted survival curves showing that drug A or B works in a subset of the population, for example cetuximab in unselected patients (see figure, below right). However, these curves also indicate a whole set of patients that do not benefit from these drugs, and we are unable to separate the patient groups, even though we use the drugs in our daily practice. We see these types of curves repeatedly for many types of drugs, both standard therapies and targeted agents, with a group of patients that benefits and a group that does not. This is because of the inherent heterogeneity of colorectal cancer, which we need to understand in order to better target therapy.

Continuing with the example of EGFR monoclonal antibodies, there are two key messages. Firstly, these drugs are remarkably effective as monotherapy. This is crucial because if a drug works as a monotherapy, it means that it addresses the biology underlying the disease. The second point is the limited groups of patients in which they work: 10% in monotherapy if patients are unselected; 25% in monotherapy if patients are *KRAS* wild type. To do a good job we must identify the subgroup upfront, and we are not yet at that stage.

Many pathways are involved in colorectal cancer, and any one of these could be affected in a particular colorectal cancer patient. This means there are many different versions of the one disease that we call colorectal cancer, but we currently have only two markers that have been more or less validated: *KRAS* and microsatellite instability (MSI). The first of these is used to predict response to EGFR targeted therapies and the second for prognosis in stage II disease.

We tend to simplify the way we look at the biology of tumours. For example, having found the role of EGFR signalling in non-small-cell lung cancer, or the role of HER2 signalling in breast cancer, we assume that these pathways act in the same ways in other diseases. However, we know that EGFR in non-small-cell lung cancer does not act in the same way as the EGFR pathway in colon cancer. Having identified a pathway, we have to look at which disease it is working in, and remember the effect of the pathway can be completely different according to the tumour type. *KRAS* mutations have different roles in pancreatic cancer, melanoma and colon cancer. This means that we have to look at tumour environment specificity for each marker.

PROGRESSION-FREE SURVIVAL FOR CETUXIMAB IN UNSELECTED PATIENTS



The heterogeneity of colorectal cancer is demonstrated time and again by graphs like this one for cetuximab as a monotherapy in unselected patients, showing some patients derive benefit while others don't

Source: DJ Jonker et al (2007) Cetuximab for the treatment of colorectal cancer. *NEJM* 357: 2040–48, reprinted with permission © Massachusetts Medical Society 2007

Tumour origin

Colorectal cancer originates from the very undifferentiated stem cell compartment in the colon. This is important for everyday functioning of the bowel, but the negative impact is that colon tumours have properties of self-renewal, de-differentiation and plasticity. This means we are faced with a very difficult disease. We are not sure which cells in the bowel give rise to the majority of tumours, and it is not something we currently take into account. There is probably a lot of refinement needed in terms of cell subtype and cell origin.

A distinction that we often forget to make, and which is very relevant, relates to the primary tumour site – between tumours originating from the right side of the colon, which is the mid-gut in embryonic origin, and those from the left side of the colon, which is the hind-gut. The mid-gut and hind-gut have different origins, driven by different genes. Tumours arising on the right side, which goes almost to the hepatic flexure, probably have inherently different biology compared to left-sided tumours.

Data reported by Arnaud Roth at ASCO two years ago showed Kaplan-Meier survival curves for patients based on the origin of their tumour (see figure, right). Patients whose tumours had a left-sided origin had better prognosis than those with tumours originating on the right. This is because the driving biology is different, with different genes in tumours originating on the left versus right.

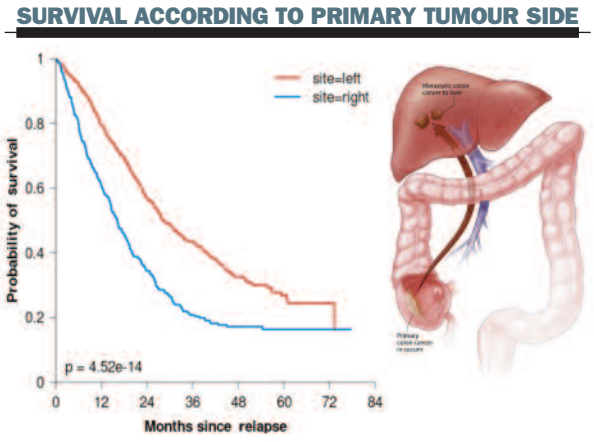
Current descriptors of CRC heterogeneity

At the moment, only the *KRAS* and MSI markers have made it into clinical practice. We are all accustomed to the Vogelgram, which suggests *APC*, *KRAS* and *TP53* are needed to drive colon cancers. However, although a very useful model, it is not clear if this is the way all colon tumours progress. Most of our mouse models develop small bowel instead of large bowel cancer, and further refinement is needed with modelling of the effect of multiple genes in dedicated models.

Furthermore, while tumour initiation is something that we should be able to understand in appropriate mouse models that assess whether particular mutations lead to tumour development, in clinical practice this is not what you are treating. In clinical practice, patients present with metastatic disease that has evolved in ways we do not yet understand and cannot yet model. Molecularly, this is probably quite far from the simple situation of tumour initiation.

Metastatic disease is several years removed and can have a lot of new alterations that would be very difficult for researchers to map. It would be very difficult to make a mouse model of the whole metastatic cascade. In addition, every time you give drugs to a patient you are probably changing the identity of the tumour, particularly with very targeted agents such as an *EGFR* inhibitor or an HGF (hepatocyte growth factor) inhibitor. This will probably remove certain cell populations and enable others to take over as part of a resistant mechanism. A static image of a patient's tumour is probably not correct and it might be that we should biopsy multiple times during treatment to check the molecular identity over time. This may explain why current biomarkers do not correlate well with outcome, because they do not reflect the actual disease in a patient.

Recent studies have demonstrated tumour plasticity. For example, a study giving a MAP kinase inhibitor to a *BRAF* mutant cell line or to a *KRAS* mutant cell line showed that the cell lines were able to escape the drug in a few months. In the *BRAF* mutation, this was achieved simply by amplifying the BRAF chromosome, and in the case of *KRAS* mutation, the KRAS chromosome (*Sci Signal* 2010; doi: 10.1126/scisignal.2001148; *Sci Signal* 2011; doi: 10.1126/scisignal2001752). So, there is a very targeted and selective way of acquiring resistance to whatever drug treatment we are giving, which I think happens frequently in patients as we treat them.



Differences in survival according to the side the tumour originates reflect a difference in tumour biology

Source: Arnaud Roth, presented at ASCO 2009

Question: *If they amplify these cells, the genes, can it not be circumvented by giving more of the drug, and increasing the dose?*
Answer: *Yes, that could be a solution if you know that it is going on, but there might be some dose-limiting toxicity. However, what really struck me in these reports was the very targeted way that the cancer uses the genetic instability that underlies all cancers to simply select cells that are resistant to the drug being used, so those cloned cells survive.*

Question: *What do you think about sequencing the whole genome for a patient?*
Answer: *There are fantastic technologies at hand and sequencing a patient's tumour at repeated time points will be feasible and cost-effective in the future. The problem is how to interpret that information: which of the markers is the important one and which therapeutic drug do you link to this?*

Our current efforts are focused on taking a step back: taking a very unbiased approach, not dividing the disease into MSI+ or MSI– or to *KRAS*+ or *KRAS*–. We should adopt a very comprehensive approach, including analysing DNA, RNA, and protein, and measuring everything without a hypothesis, and biology may become apparent in that information. Very useful information is emerging in the *Cancer Genome Atlas* on colon cancer in 2011.

We are trying to generate subgroups similar to those now used in breast cancer, which are based on gene expression and show both prognostic and predictive relevance. To gain the necessary critical mass of information, large consortia will be needed, and everyone will have to share information, including doctors, patients, and the pharmaceutical companies who often have very large series of well annotated samples from clinical trials.

There is another factor underlining why collaboration is necessary. Even if you have full sequencing for a patient and have identified all the mutations – there are 71 mutations on average for colorectal cancer (B Vogelstein, *Science* 2007, 318:1108–13) – you still do not know what these mutations mean for the patient, nor the drugs he or she will respond to, because a map of the mutations does not mean that we understand what they are doing.

The big challenge now is to get functional annotations of the mutations we see. We have identified some of the mutations, including *KRAS* and *BRAF*, and we know that there is *HER* amplification but we have no idea what they are doing. One way to do functional annotation is to use cell lines and mouse models, and this is ongoing but it is time-consuming and difficult and sometimes unproductive. Another way is to explore what these genes are doing in patients. If you have a very specific mutation in a patient, for example a deletion of *PTEN*, and look at how patients with this amplification respond to different drug treatments, you will probably be able to learn about the function of the mutation, because it will show high sensitivity or resistance. This is using the patient as the ultimate test tube, which is necessary because *in vitro* methods are not always successful.

Question: *Are these small trials, where you just test out hypotheses in certain mutations and certain drugs with a low number of patients?*
Answer: *A ballpark figure from our experience is around 60–80 patients, often in phase II trials. As long as you have a clear map of the molecular alterations you are looking at, so the biomarker is clear, and you track it throughout a trial, for example with an IGF inhibitor versus a C-MET inhibitor, and you see that the biomarker predicts something completely different in these two trials, then you have learnt something about the pathway of your biomarker.*
Question: *You just mentioned that patients have mutations in 71 genes, on average. How many pathways are relevant in colorectal cancer, if 71 genes are affected?*
Answer: *Bert Vogelstein presented a schema of all the relevant pathways at ASCO last year and ended up with about 15, including Wnt and Hedgehog (JCO 2009, 27 Suppl 15). But we can't yet put a number to this. We now have enough samples for colorectal cancer analysed worldwide to get a first grip on the subgroups; however, the static versus dynamic element probably makes this more complex.*

Clinical trials with targeted agents have been very helpful. We never really knew where to position *KRAS* in colon cancer signalling until EGFR inhibitors came along. We now know much more about *KRAS* thanks to the cetuximab and panitumumab trials. This is just one example, but many more trials with other drugs are coming through. It will be interesting to see whether other receptor tyrosine kinase inhibitors will show the same influence of *KRAS* mutations.

Biomarker development

The necessary factors for biomarker development include:

- A good understanding of what is going on in metastatic colorectal cancer
- Therapies with known targets
- Knowledge of the effect of target inhibition
- Tractable risk/benefit profile
- Biomarkers that have a large impact
- Validation

The first step is a good understanding of what is happening in the disease. We do not really yet have that. We do have some therapies with known targets, although a lot have no clear cellular anticancer mechanisms, which makes it difficult to make biomarker/therapy relationships. Validation is essential, requiring large datasets for which we have to learn to collaborate much more.

KRAS and MSI

KRAS and MSI are the first biomarkers in colorectal cancer. However, we sometimes oversimplify things. We know MSI is a marker for good prognosis and we would like to be able to use it in clinical practice, but there are some pitfalls. There is a different incidence of MSI in stage II and III tumours, as for many markers. However, many publications report stage II and III series together, or analyse the effect in a compound way. We must be very cautious and try to be as precise as

possible in studying the effect of a biomarker in a homogeneous population.

Not only does the incidence of MSI, and maybe also its prognostic value, differ between stages II and III, but the prognostic value also differs according to the presence or absence of other markers and features. The table below shows that MSI and 18qLOH behave differently as markers in stage II and III disease.

The take home message is to be very precise about the disease group you are looking at and never forget that a marker, as simple as it may seem, may have hidden complexity, such as interaction with stage or other markers.

MSI instability is a good prognostic marker in univariate analysis. The same is true for 18qLOH as a marker of poor prognosis. However, would the 18q information still matter if you knew the MSI status of your patient? In the microsatellite stable (MSS) population, which is the largest population, 18q is no longer prognostic (see figure, above right). This means 18q only gives useful information if you do not know the microsatellite status. This is just one of many examples where you might see strong effects of a marker in univariate analysis, yet it is no longer present in multivariate analysis with relevant interacting markers.

E5202 is the first trial to use risk assessment based on 18q/MSI to determine treatment in stage II colon cancer (www.clinicaltrials.gov). High-risk patients, defined as MSS and 18qLOH, are treated with chemotherapy. Low-risk patients, defined as MSI-high and MSS with no 18q, undergo only observation and no treatment. The design is flawed, however, as 18q does not matter in MSS disease, and these patients are still at high risk. This design was based on a combination of two univariate analyses that were not put into a multivariate analysis.

Another interesting study was presented by Dan Sargent at ASCO 2008. He conducted a pooled analysis of multiple trials in patients with stage II and III tumours, comparing patients treated in the adjuvant setting with those who were untreated (see figures, p 20, left, centre). Patients with high MSI who were untreated did much better than MSS patients. However, this effect completely disappeared in the treated patients, and it might be that giving 5FU to patients with high MSI is harmful because the benefit of being MSI disappears.

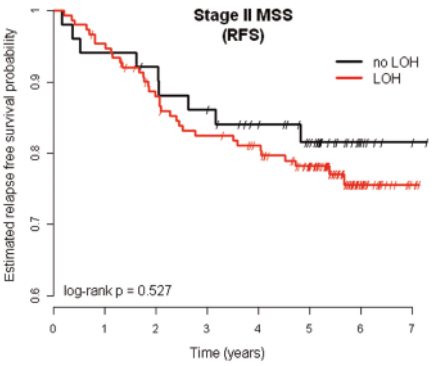
PROGNOSTIC VALUE OF MARKERS IN STAGE II AND III TUMOURS

The prognostic value of these markers (looked at in isolation – univariate analysis) varies according to the stage of disease, which has implications for how studies of biomarkers are designed and reported

Marker	Stage II (n=420)		Stage III (n=984)		Interaction
	HR	p val	HR	p val	
MSI (Hi vs Stable)	0.3	0.004	0.7	0.06	0.04
18qLOH	2.1	0.03	1	0.91	0.05
SMAD4 (any loss)	1.4	0.21	1.6	<0.0001	0.23
hTERT (High)	1.4	0.32	1.5	0.01	0.92
p53 (High)	1.0	0.98	1.3	0.03	0.37
TS (High)	0.5	0.03	0.7	0.02	0.30
KRAS (Mutated)	1.1	0.84	1.0	0.72	0.32
BRAF (Mutated)	0.9	0.90	1.2	0.28	0.38

Source: Arnaud Roth, presented at ASCO 2009

PROGNOSTIC VALUE OF 18QLOH ON MSS STAGE II DISEASE



18qLOH and microsatellite instability (MSI) status are both good prognostic markers when used alone, but for patients known to be microsatellite stable (MSS), 18qLOH loses its prognostic value

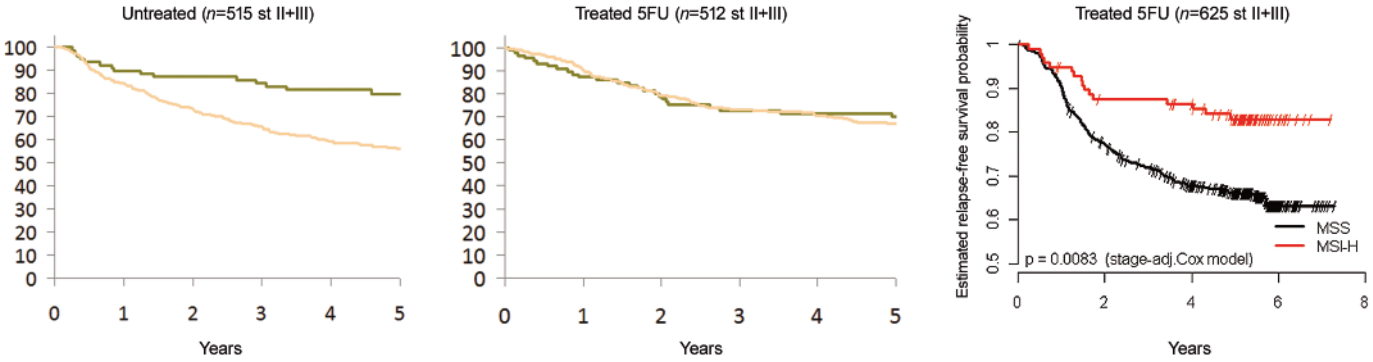
Source: Arnaud Roth, presented at ASCO 2009

In contrast, a clinical trial from our group (see figure, p 20, right) showed a very strong prognostic effect of high MSI versus MSS, unlike the Sargent data. The difference in results between similarly powered studies suggests there is still something that we are not capturing, and more and larger studies are needed.

The take home message on MSI status is that, although we would love to say that MSI is a simple marker of prognosis and response to adjuvant treatment, there are several unresolved issues, including sporadic versus hereditary MSI, the role of CIMP, the role of *BRAF* and the impact of novel therapies. We should not embrace biomarkers if they do not have clear validation for clinical practice.

Question: *In clinical practice, do you measure MSI and do you consider it in treatment decisions for stage II patients?*
Answer: *At the moment, I do not make decisions based on MSI status, although I acknowledge it is a very strong marker*

THE EFFECT OF TREATMENT BY MSI STATUS – CONFLICTING TRIAL RESULTS



A study by Dan Sargent and co-workers showed that patients with high MSI lost their survival advantage when treated with 5FU (left and centre graphs); however, in a study conducted by Sabine Tejpar and colleagues, patients with high MSI (MSI-H) responded much better to 5FU treatment than MSS patients

patients with high MSI; microsatellite stable (MSS) patients.

Sources: Dan Sargent, presented at ASCO 2008 and Sabine Tejpar, presented at ASCO 2009

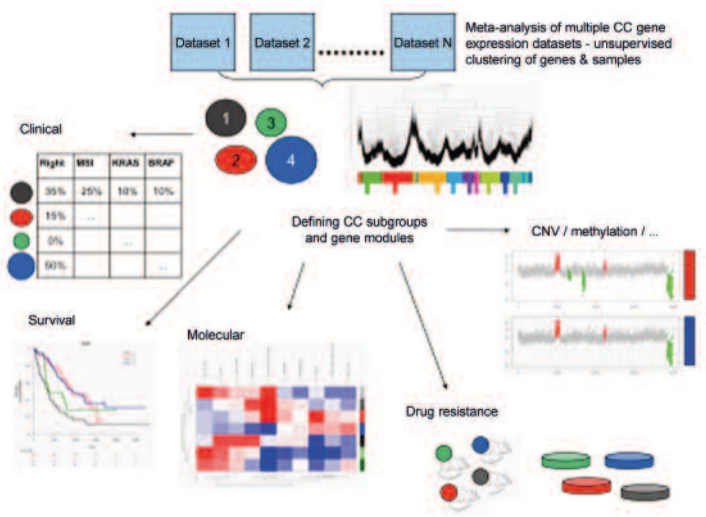
and a good basis for patient risk stratification. However, I presented data at ASCO 2010 on the uncertainty that still exists if you use MSI for treatment decisions in stage II patients. In stage II patients, we also use ASCO clinical high-risk criteria, such as fewer than 12 lymph nodes examined, T4, poor differentiation, and obstruction. MSI patients are often poorly differentiated, which is high risk, and T4, which is also high risk. So, on the one hand you have MSI telling you that this is a good prognosis patient, and on the other hand there are high-risk features telling you that this is not a good prognosis. If you combine all these factors in a multivariate risk model, you would still be wrong in a small number of cases if you used MSI as a standalone marker. A paper was recently published by Frank Sinicrope and Dan Sargent's group looking at the difference between sporadic and hereditary MSI, which reports intriguing findings that again warrant further detailed investigation into MSI as a standalone marker (JNCI 2011, 103:863–875).

Question: *Do you use clinical markers, or do you not consider any markers?*

Answer: *We use clinical markers from ASCO guidelines, as these have quite a lot of data behind them. I am not discouraging people from using MSI, but you have to be aware of the margin of error, and the need for further studies.*

PETACC3 provided a very large series of 1400 patients to look at multiple markers (*Clin Cancer Res* 2009, 15:5528–33). It showed how the integration of molecular markers often changes the view that you have based on a single marker, and the impact of integrating variables such as T stage and

SUBGROUPING BASED ON GENE EXPRESSION



This analysis of unprespecified gene expression identified four main subgroups that seem to be in agreement with other studies, but not with currently used descriptors of the disease

CC – colon cancer.

Source: Swiss Group of Bioinformatics in Lausanne, presented at AACR 2011

N stage. These are very large effects that have not been modelled sufficiently, and offer important work for the colorectal cancer community that can easily be performed over the next few years.

Eva Budinska and Mauro Delorenzi, of the group at the Swiss Institute of Bioinformatics in Lausanne, took multiple data sets and looked at gene expression in an unprespecified way, trying to identify spontaneous subgroups in the disease (see figure, p 20, lower). Results showed subgroups, in agreement with other studies. I think we are at the point of identifying the subgroups in colon cancer just as in breast cancer. This figure simplifies the subgroups into four colours, numbered 1, 2, 3, and 4 (although there were a few more).

These subgroups, which are spontaneously present in the disease, correlate poorly with current descriptors of the disease, including clinical descriptors such as stage, T or N status, and even *KRAS*, *BRAF* or MSI. This means the subgroups better describe the ongoing disease process than current markers. The existing cell lines can be compared to see whether they match the patient subgroups, as well as mouse models. This means we can now refine the tools that we use in the lab, such as cell lines, to ensure they better match the true subgroups present in tumours. Another similar study used unsupervised subgrouping analysis of colorectal cancer (*BMC Med Genomics* 2011, 4:9), and I think these studies are going to be very important over the next few years.

The same message is emerging from recent work on *KRAS*, questioning whether all *KRAS* mutations have the same effect. The table above summarises the incidence of different *KRAS* mutations. There may be differences between mutations, and maybe even between different patients with the same mutations.

NOT ALL *KRAS* MUTATIONS ARE ALIKE

KRAS mutation		Incidence (%)
Amino acid substitution	Nucleotide substitution	
Codon 12 mutations		
Aspartate (G12D)	G35A	32.5
Valine (G12V)	G35T	22.5
Cysteine (G12C)	G34T	8.8
Serine (G12S)	G3A	7.8
Alanine (G12A)	G35C	6.4
Arginine (G12R)	G34C	0.9
Codon 13 mutations		
Aspartate (G13D)	G38A	19.5
Other Mutations		1.8

Different mutations in the *KRAS* gene affect tumour behaviour in different ways

Source: N Normanno et al. (2009) *Nat Rev Clin Oncol* 6: 519–527, published with permission, © Nature 2009

In some patients a *RAS* mutation may activate the RAF MAP kinase pathway, but in other patients the same mutation may activate another pathway, such as PI3 kinase or RAL (see figure below). Just because we have *KRAS* mutants or wild types does not mean the two types have homogeneous biology.

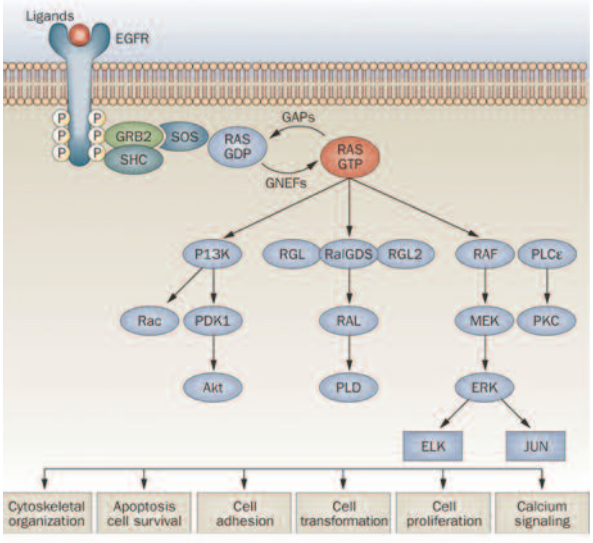
To find proof of this we looked at gene expression data in patients with *BRAF* mutations, patients with *KRAS* mutations and those with neither of these mutations (double wild type). Results showed that *BRAF*-mutant patients have some genes always on and some genes always off, while these are reversed in wild type patients (Popovici et al, manuscript in preparation). The conclusion is of very homogeneous disease in *BRAF* mutants, so this marker is indicating something useful.

However, this division of gene expression is not nearly as clear in *KRAS* mutants versus wild types. There still seem to be different groups of *KRAS* mutants, which are quite different in terms of gene expression. This indicates *KRAS* is not a marker of homogeneous disease.

The take home message is that the underlying biology is much more heterogeneous than current markers might indicate, and an unsupervised approach is necessary that does not separate patients into prespecified groups. This has important therapeutic implications. For example, treating all *KRAS* mutant patients with MAP kinase inhibitors is not going to be successful, because of heterogeneity between them.

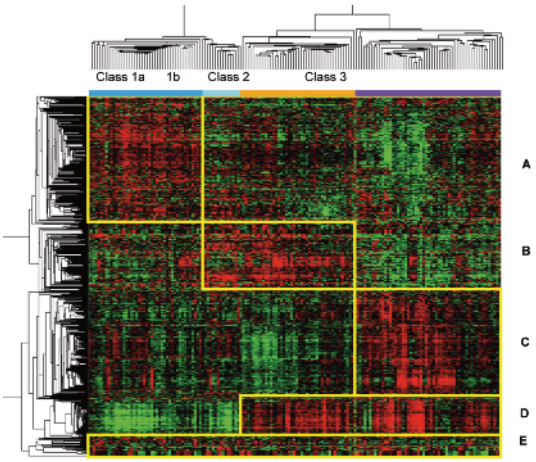
A solution to this problem is illustrated by a study performed by Shirin Khambata-Ford at Bristol Myers Squibb (the company that markets cetuximab in the US) in 2007 (*JCO* 25:3230–37). This study was very open, and was not just looking at *EGFR*

RAS MUTATION PATHWAYS



Source: N Normanno et al. (2009) *Nat Rev Clin Oncol* 6: 519-527, published with permission © Nature 2009

BIOMARKER DISCOVERY STUDY IN PATIENTS TREATED WITH CETUXIMAB



Studies like this one, which analysed how gene expression profiles differ between patients who did well on cetuximab and those who did not, will help identify markers of response

Source: S Khambata-Ford et al. (2007) *JCO* 25:3230–37, published with permission © ASCO 2007

copy number or *KRAS*. The trial biopsied liver metastases in 80 patients just before treatment with cetuximab. Full Affymetrix profiling compared gene expression in patients who did well against those who did badly, revealing the biomarkers for sensitivity to the drug (see figure above).

If we collect material in the many ongoing trials with targeted agents and analyse it in an unprespecified way, we can make a lot of progress in understanding the biology of colorectal cancer over the next few years (see figure below).

SUMMING UP

In terms of biomarker development in colorectal cancer, we have a good grasp of what is going on in metastatic disease. Therapies have been developed that have known targets and the effect of target inhibition is known. It is essential that we keep an open mind on biomarkers and critically evaluate the available information, ensuring all findings are thoroughly validated.

Question: *Looking at gene expression profiles – do you think there are three or four groups, or more?*

Answer: *The published data mentioned previously show two big groups. I believe the number is likely to be fewer than 10, but more than two. We are pleased with this number, because it comes close to something that people can use in the future. It is important to note this grouping was based only on gene expression. If you add in copy number, mutation data and microRNA, you can probably refine subgroups further.*

This is not the end of the story, but, in a similar way to breast, it is a very good start. It puts us on track for planning a clinical trial, giving an idea of benchmarks, what to power for, and how much heterogeneity to expect within the population or within the drug effect.

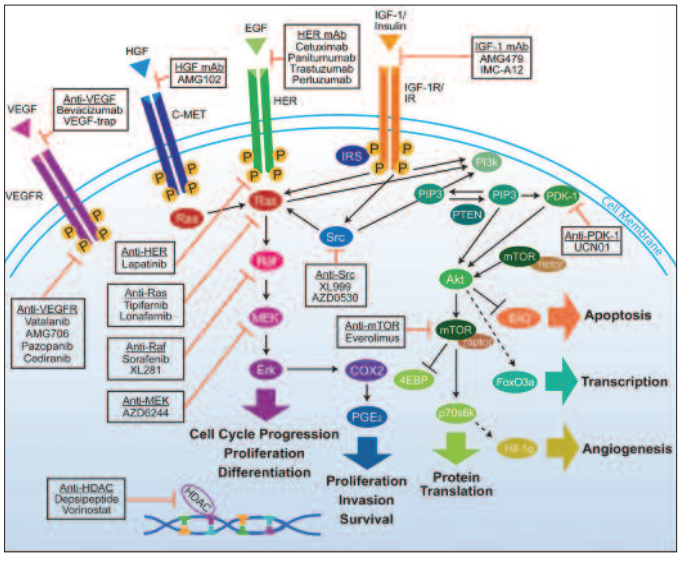
Question: *Do you think the future will be based on these different groups distinguished by gene expression, and then digging deeper by knowing more about it?*

Answer: *Yes, I hope that nature has not made every colon tumour completely different, but that there are recurring themes. The assumption is that every tumour would fit into some category and we are working hard towards getting that classification.*

Question: *What is the general methodology to adopt in biomarker studies?*

Answer: *It is important to be aware of the shortcomings of whatever assay you are using. You need large sample sizes, and you need to be sure that effects are stable and that there are no other variables that change the effect of the marker. Setting up both a discovery and one or two validation sets is very important.*

SUBGROUPING BASED ON GENE EXPRESSION



This analysis of unprespecified gene expression identified four main subgroups that seem to be in agreement with other studies, but not with currently used descriptors of the disease

S Siena et al (2009) *JNCI* 101:1308–24, published with permission © Oxford University Press 2009