Genomic markers for decision making: what is preventing us from using markers?

Vicky M. Coyle and Patrick G. Johnston

Abstract | The advent of novel genomic technologies that enable the evaluation of genomic alterations on a genome-wide scale has significantly altered the field of genomic marker research in solid tumors. Researchers have moved away from the traditional model of identifying a particular genomic alteration and evaluating the association between this finding and a clinical outcome measure to a new approach involving the identification and measurement of multiple genomic markers simultaneously within clinical studies. This in turn has presented additional challenges in considering the use of genomic markers in oncology, such as clinical study design, reproducibility and interpretation and reporting of results. This Review will explore these challenges, focusing on microarray-based gene-expression profiling, and highlights some common failings in study design that have impacted on the use of putative genomic markers in the clinic. Despite these rapid technological advances there is still a paucity of genomic markers in routine clinical use at present. A rational and focused approach to the evaluation and validation of genomic markers is needed, whereby analytically validated markers are investigated in clinical studies that are adequately powered and have pre-defined patient populations and study endpoints. Furthermore, novel adaptive clinical trial designs, incorporating putative genomic markers into prospective clinical trials, will enable the evaluation of these markers in a rigorous and timely fashion. Such approaches have the potential to facilitate the implementation of such markers into routine clinical practice and consequently enable the rational and tailored use of cancer therapies for individual patients.

Coyle, V. M. & Johnston, P. G. *Nat. Rev. Clin. Oncol.* **7**, 90–97 (2010); published online 15 December 2009; doi:10.1038/nrclinonc.2009.214

Introduction

Over the past two decades there has been extensive research evaluating the clinical usefulness of a range of genomic markers in predicting patient outcome or response to treatment in solid tumors. Typically, this approach has involved the identification of a particular genomic alteration with a subsequent evaluation of the association between this finding and a clinical outcome measure. However, the clinical impact of many of these studies has been limited by inconsistencies in study design and results, and despite substantial research effort very few putative genomic markers have been implemented into routine clinical use. Concurrently, genomic technologies have been evolving rapidly, with the advent of high-throughput genomic technologies, such as microarray-based gene-expression profiling. These techniques have revolutionized genomic research, and have allowed the simultaneous analysis of multiple markers on a genome-wide scale. Researchers have rapidly adopted such techniques into both basic scientific and translational research, and in doing so, have moved away from the traditional model of investigating putative genomic markers identified from focused research programs to the identification and measurement of multiple genomic markers simultaneously within clinical studies. This progress has in turn presented additional challenges when considering the use of genomic markers in oncology such as clinical study design, reproducibility and interpretation of results. This Review will explore these challenges, focusing on the most mature high-throughput genomic technology, microarray-based gene-expression profiling, and will highlight some common failings in study design that have impacted on the clinical use of putative genomic markers.

Single genomic markers

Although many studies have evaluated potential genomic markers in a variety of cancer types, the vast majority of these markers have not been proven to be useful in the clinic. For example, in almost 1,000 studies evaluating biomarkers in breast cancer, only a handful of markers, including estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), have been implemented into routine clinical use worldwide.[1] Of the more recently evaluated prognostic markers in this disease, it is sets of multiple genomic markers that have received FDA clearance (MammaPrint, Agendia) or have been recommended as clinical decision making tools (Oncotype DX, Genomic Health).[2]

Centre for Cancer Research and Cell Biology, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK (**V. M. Coyle**, **P. G. Johnston**).

Correspondence to: P. G. Johnston p.johnston@qub.ac.uk

Competing interests
P. G. Johnston declares associations with the following companies: Almac, Fusion Antibodies, GlaxoSmithKine. See the article online for full details of the relationship. V. M. Coyle declares no competing interests.

Human cancers are complex diseases characterized by cellular and molecular heterogeneity. In tumors that have similar clinical and pathological characteristics, significant variability in patient outcomes has been observed.[3] Therefore, a 'single disease, single genomic marker' approach fails to take account of this heterogeneity and consequently single genomic markers are often found to be inadequate as biomarkers in clinical studies. Moreover, with the traditional approach of investigating individual genes as potential therapeutic targets, and assessing the presence of mutations or nucleotide polymorphisms in these genes or changes in their expression as a marker of treatment response, the effects of other genomic alterations on pharmacological response has often been missed.[4] Essentially, the 'wrong' genomic marker may have been evaluated. For example, using EGFR expression as a marker of response to EGFR targeted therapies, a number of studies failed to identify an association between EGFR expression and response to cetuximab in colorectal cancer,[5,6] indeed some EGFR-negative patients seemed to benefit from cetuximab.[7] Although these findings may partly be a reflection of variations in methodology and study design, new research has shown that the presence of mutations in the *KRAS* gene have been consistently demonstrated to be associated with resistance to cetuximab[8,9] and other EGFR-targeted therapies, such as panitumumab.[10] Such a finding is in keeping with the recent proposal that expression of downstream effector molecules may potentially be more relevant than expression of the receptor itself. This proposal is in light of our increased understanding of the effects of alterations in individual genes on the signaling networks that drive tumor growth and progression.[11]

Key points

- Despite extensive research, relatively few genomic markers have been implemented into routine clinical use, often because of failings in clinical study design
- The traditional 'single disease, single genomic marker' approach does not consider tumor heterogeneity and consequently single genomic markers are often found to be inadequate biomarkers in clinical studies
- The introduction of new high-throughput genomic technologies has enabled the simultaneous measurement of multiple genomic alterations, and has revolutionized the field of genomic marker research in oncology
- High-throughput technologies have presented additional challenges to considering the routine clinical use of putative genomic markers
- Putative genomic markers should undergo extensive validation before they are implemented into routine clinical practice
- Novel adaptive clinical trial designs that incorporate genomic markers into prospective studies will enable the evaluation of markers in a rigorous, timely fashion and facilitate their implementation into clinical practice

Box 1 | Challenges incorporating microarray technology

Technical
Array platform
Reproducibility
Specimen considerations

Data analysis
Quality control assessments
Data analysis methodology

Interpretation of results
Assigning cut-off thresholds
Validation of results

Microarray-based profiling

Microarray-based gene-expression profiling enables the rapid, simultaneous measurement of gene expression on a genome-wide scale and consequently this technology represents an exciting development in genomic marker research that has been enthusiastically embraced by researchers worldwide. This Review will focus on this technology, not only because is it the most mature of the high-throughput technologies, but it has also begun to be evaluated as a tool for decision-making in routine clinical practice. It is this experience that highlights the particular challenges arising from using these technologies in genomic marker research, and unless these challenges can be overcome, incorporation of these technologies into routine clinical use will be limited. A number of specific challenges apparent from microarray studies are listed in Box 1, and several of these important issues will be discussed in more detail.

Technical considerations of specimens

Tissue specimens often have a heterogeneous composition, including varying proportions of malignant cells, stromal elements, vasculature or other cells such as inflammatory cells. Researchers have attempted to control for this variability in several ways: use of specimens of similar composition, aiming for all specimens to contain an equivalent proportion of tumor cells, and including patient-matched controls or increasing the number of specimens.[12] Controlling for these parameters is not always possible; for example, when specimens are small in size or fragmented such as diagnostic or pre-treatment biopsies. Alternatively, tumor cells can be enriched from a tumor specimen by laser capture microdissection,[13] and successful gene-expression profiling using this technique has been reported in a range of malignancies,[14] including prostate carcinoma, colorectal carcinoma and breast carcinoma.

Microarray profiling technology has primarily relied upon tissue samples that are freshly frozen, which has limited availability, mainly due to the cost and feasibility of collecting and storing large numbers of these specimens. Moreover, the processing of these specimens is critical, as the technical success of these experiments depends upon the availability of adequate amounts of high-quality RNA from the starting material. Ideally, tumor specimens should be frozen in liquid nitrogen within 30 minutes of surgical excision and stored at a minimum temperature of –80 °C to prevent degradation of the nucleic acids.[15] A recent pilot study that assessed data from six European centers for gene-expression profiling studies to investigate the feasibility of obtaining fresh frozen tumor tissue reported that appropriate tumor specimens were obtained from 60 of the 64 patients, and that 46 of

these 60 specimens contained sufficient tumor content and yielded sufficient RNA for successful transcriptional profiling.[16] However, this required a coordinated effort from all disciplines to obtain an appropriate fresh frozen tumor specimen within one hour of surgery.

To obtain fresh frozen tumor tissue is clearly challenging to achieve in the routine clinical setting, where specimens are typically formalin fixed at the time of surgery for subsequent histological examination. RNA extracted from formalin-fixed paraffin-embedded (FFPE) tissue is often of poor quality because of RNA fragmentation and crosslinking of RNA to DNA and protein. Large tumor specimens may also be inadequately fixed, which adversely affects RNA viability.[17] This problem can be overcome by PCR-based analysis of gene-expression by choosing small fragments for detection; consistent and reproducible gene-expression data has been reported using this approach.[18]

RNA degradation has proven more challenging in microarray-based expression profiling, firstly due to technical considerations, and secondly, the potential impact of information 'lost' during the fixation, embedding and storage processes. Novel microarray technologies and tissue processing protocols specifically designed to address these problems are in development as a first step to utilize FFPE specimens routinely for microarray analysis.[19] These include modifications in both RNA extraction processes,[20] and microarray platforms, for example, the cDNA-mediated annealing, selection, extension and ligation (DASL) assay (Illumina Inc.), a hybrid between microarray and reverse transcription-PCR technology.[21] Other assays include the CupPrint assay, an oligonucleotide array specifically designed to classify tumors of unknown primary to a histogenetic primary origin using FFPE specimens[22] and the Disease Specific Arrays (Almac Diagnostics) in which the disease-specific probe sets have been designed specifically to the 3' extremities of the relevant transcripts facilitating profiling of RNA extracted from FFPE tissues.[23]

In an attempt to identify the transcriptional information lost when profiling FFPE specimens, Scicchitano and coauthors compared the transcriptional profiles of fresh frozen and FFPE pelleted human bone marrow stromal cell specimens profiled on Affymetrix HGU133 Plus 2.0 arrays.[24] Both sets of specimens demonstrated comparable performance at array level with similar proportions of present calls and only small differences in 3':5'ratios were observed. Although only 23% of the genes that were identified as differentially expressed from the frozen specimens were identified as differentially expressed from the FFPE specimens, pathway analysis demonstrated that the majority of relevant gene functions and key signaling pathways identified from the frozen samples could be detected from the FFPE specimens.[24] Similarly, Srivastava *et al.*[25] found that key pathways involved in prostate cancer development and progression were identified in both FFPE and fresh frozen prostate cancer sets profiled on the Affymetrix platform despite differences in the numbers of differentially expressed genes identified in the analysis of these datasets.[25] Such studies suggest that meaningful data can be obtained from microarray studies using FFPE material; however, the rapid processing and fixation times used in these studies may not reflect the processing and prolonged storage of specimens performed in routine clinical practice. Hoshida *et al.*[26] have suggested that over 90% of archived FFPE tissues can be used successfully for gene-expression profiling, as demonstrated by reporting the identification of a genomic classifier predictive of survival in hepatocellular carcinoma. If subsequent studies confirm the feasibility of using FFPE tissue obtained by routine clinical practices for gene-expression studies, this represents a significant advance in genomic marker research considering the huge potential tissue resources of archived diagnostic specimens and banked specimens from clinical studies.

Comparison of primary and metastatic tumor

The ability to investigate gene expression on a genome-wide scale has the potential to offer significant insights into tumor progression; for example, by comparing transcriptional profiles of primary and metastatic tumor. Such an approach has been investigated in colorectal cancer, where several studies have used DNA microarray profiling to compare gene expression in colorectal tumors of various stages. From these studies, panels of genes differentially expressed between normal mucosa and colorectal tumor,[27,28] between stage II and stage III colorectal tumors,[29,30] and between colorectal primary tumors and metastatic tumors[31–33] have been identified.

In another study, Koehler and coauthors did not identify a gene-expression signature specific to metastases, but instead identified genes that were differentially expressed between low-stage groups (up to pT3 tumors) and high-stage groups (T4 tumors or extensive nodal involvement or metastatic disease). The results from this study indicate that changes in gene expression may occur relatively early in tumor development.[34] Thus, it is likely that distinct gene signatures for tumors of a different stage may be identifiable, but these may not correspond to conventional clinicopathological subgroups. Moreover, most studies investigating gene signatures between primary and metastatic samples are generally small, include heterogenous patient populations and have significant variations in study design and methodology, which limits the impact of their findings.

Obtaining appropriate tumor specimens for microarray studies is clearly a major challenge. In particular, obtaining metastatic tumor specimens for genomic analysis in relation to treatment response prediction may be difficult or require patients to undergo invasive procedures to obtain small tumor biopsies, as in the case of patients with colorectal cancer presenting with unresectable hepatic recurrence. In this setting, the ability to use archived colorectal primary tumor as a surrogate model for measuring gene expression would be highly valuable. The reports of differential gene expression between primary and metastatic tumor outlined above[31–33] might suggest that analysis of primary tumor would not be useful in predicting response to chemotherapy in advanced disease. Only a few small studies have specifically addressed the

ability of primary colorectal tumor specimens to act as a surrogate for metastatic tumors with regard to predicting response to treatment.

Matsuyama *et al.*[35] investigated the expression of a panel of 81 candidate genes associated with resistance to 5-fluorouracil (5-FU) chemotherapy identified from a previous *in vitro* microarray study and genes encoding 5-FU-related enzymes in resected colorectal primary tumors to determine if they could predict response to 5-FU chemotherapy from synchronous liver metastases. From this analysis, a three-gene expression model was developed that could accurately predict treatment response. In a small study of 21 tumor samples, a global profiling approach was used to identify a classifier predictive of response to irinotecan and 5-FU chemotherapy in patients with advanced colorectal cancer. Of note, patients enrolled in this study were those who presented with advanced colorectal cancer with synchronous unresectable liver metastases, and the specimen was taken from the primary tumor rather than the metastatic lesion. Genes that were differentially expressed were initially selected using Statistical Analysis of Microarrays (SAM), and a 14-gene predictive classifier constructed using the Support Vector Machine algorithm. This approach demonstrated a predictive accuracy of 95% when assessed by cross-validation.[36] This was a small study, however, and this 14-gene signature needs further testing in an independent validation set. If this gene classifier can be successfully validated then the potential of obtaining tissue that will be of use for genomic analysis in advanced colorectal cancer is greatly increased.

Box 2 | Objectives of microarray studies[64]

Class discovery
Identification of unknown patterns or trends in the microarray data; for example, the identification of groups of samples with similar gene-expression profiles across all expressed genes

Class comparison
Identification of genes that are differentially expressed between samples belonging to pre-defined classes; for example, different tissue types, different tumor subtypes or response to treatment

Class prediction
Involves construction of a classifier, a mathematical function that outputs a prediction of the class of an unknown sample, by the following steps:

- Identify the genes whose expression is used in the prediction
- Specify the form of the mathematical function used to translate gene expression to the class indicator
- Specify any mathematical parameters used in the prediction such as weighting intensity methods or threshold cut-offs

Study design considerations

Microarray studies can be broadly categorized into three types of studies: class discovery, class comparison and class prediction studies (Box 2), and multiple data analysis methods have been used in interpreting microarray data in these studies.[37]

Controlling for multiple testing

Typically, identification of differentially expressed genes between the classes of interest, for example, treatment response or non-response, is a common premise of both class comparison and class prediction studies. However, the identification of these genes is itself a challenge, as traditional statistical regression modeling has generally used data sets containing smaller numbers of variables than cases, in contrast to most microarray data sets that have significantly more variables than cases.[38] Early microarray studies used a fold-change cut-off defined by the researcher, for example two fold, to identify differentially expressed genes between classes. This approach did not consider the amount of variability in gene expression and, therefore, false positives were produced in the gene list generated. Subsequent studies have identified differentially expressed genes using statistical tests of significance such as t-tests, which assign *P*-values to genes based on their ability to distinguish between sample classes. Although widely used, such tests can also be limited by the problem of multiple testing because the number of specimens is generally much smaller than the numbers of genes and is usually too few to adequately control for false-positive results. For example, with a typical statistical significance level of $P<0.05$, 50 false-positive genes might be expected to occur for every 1,000 genes analyzed. Using this approach, these tests might be better regarded as a means of prioritizing genes for further analysis.[39] Alternatively, methods such as the Bonferroni correction or SAM algorithm have been developed in an attempt to overcome the multiple testing problem and control the false discovery rate.[40] For example, the SAM algorithm identifies gene expression changes that are of statistical significance using calculations similar to a *t*-test for specific genes for random permutations of the data set. Significant genes are those demonstrating a test statistic better than would be expected by chance. This method provides correction for multiple testing and an estimate of the false discovery rate. To date, many of the studies reported that identify genes significantly associated with a clinical outcome measure, have not addressed the issue of false positives in their data analysis. In their review of 23 such studies published in 2004, Dupuy and Simon identified nine studies where methods to control the false discovery rate were absent or inadequate.[41] Results from these studies must be interpreted with caution, particularly if researchers attempt to identify novel prognostic or predictive genomic markers or potential therapeutic targets. This is because a significant proportion of the differentially expressed genes identified are unlikely to be biologically or clinically relevant.

Sample size calculations

To date, many of the gene-expression profiling studies reported in the literature have included small numbers of patients, and there has been a general lack of formal sample size calculations reported in the literature. In

predictive microarray studies, the main challenge in determining sample size is the complexity of the power calculation, which is different to calculating the power of a traditional clinical trial to test the null hypothesis. A recent study attempted to address this problem by identifying the minimum information required to determine the sample size for a training set and by demonstrating error rates from statistical modeling of prediction rules.[42] The authors have also provided a web-based interface for sample size calculation for a binary classifier based on the following factors: the largest standardized fold change between the classes, the number of genes or features on the array and the population prevalence in the largest group.[43] They have suggested that this methodology can be used for a retrospective determination of the appropriateness of the size of training set used, and have applied this approach to previously reported microarray data sets. The results suggest that the training set sizes in some of these studies may have been inadequate to generate meaningful results, although the methodology has been intentionally developed to generate conservative sample size estimates, particularly in easy classification situations with diverse classes. This accessible interface for sample size calculation will certainly help researchers to assess whether reported microarray studies have been adequately powered to generate meaningful results, and to enable the design of more-robust studies in the future.

Data analysis and validation of results

There are generally three steps in constructing a class predictor—a mathematical function that outputs a prediction of the class of an unknown sample—from microarray data: gene selection, classifier construction and estimation of predictive accuracy. Firstly, genes that are differentially expressed must be identified between the classes of interest; for example, treatment response or non-response, as previously discussed. Secondly, the pre-defined learning algorithm uses these genes to generate the class predictor, from data obtained from the training set. To date, there has been no consensus on the 'best' methods of gene selection and classifier construction, and studies have demonstrated that identification of the best predictor is a somewhat arbitrary process as different gene lists and classification methods can produce variations in performance within a given data set.[44,45] Moreover, the major flaw identified in a critical appraisal of published microarray studies was a biased estimation of the prediction accuracy rather than the methods used in classifier construction.[41]

Assessment of the performance of a predictor is ideally done by applying the predictor to an independent data set (the validation set) as predictors that perform well on the original data set may perform poorly on an independent data set.[1] Testing a predictor in an independent test set, however, is often difficult to achieve due to the frequently limited supply of suitable patient specimens, and researchers may prefer to use all available specimens as a training set in view of the observation that classification error rates decrease as sample size increases.[46] In this setting, cross-validation approaches are frequently used to estimate the accuracy of a predictor using the data from the training set. To avoid a biased estimation of predictive accuracy, each stage of the construction of a class predictor should be included in the cross-validation process. Hence for each training set/test set, a predictor should be constructed using only the training set data, and evaluated using the test set data. Failure to perform adequate cross-validation processes can result in an overestimation of predictive accuracy by 'overfitting' the data, particularly when a predictor is derived using all the data and then tested on a subset of the data.[38,47] Although rigorous cross-validation processes help to reduce the error associated with internal validation of the performance of a predictor using only the original data, external validation of this predictor using truly independent data (ideally in a separate prospective study) is required to assess the potential clinical utility of a proposed predictor.[48,49]

In addition to determining the performance of a given predictive classifier, external validation studies should also serve to address the reproducibility of the assay in an independent laboratory setting, as this is of key importance when considering the potential future clinical implementation of this novel technology. Generation of consistent gene-expression data from comparable samples processed in different laboratories has been found to be feasible in small, focused studies. To achieve this goal, development of a common laboratory protocol and demonstration of reproducibility by means of a pilot study has been recommended.[50]

Interpretation of results

Gene-expression profiling studies often generate large amounts of data that may be subjected to multiple different data analysis processes, even within the same study. These differences in data analysis coupled with variation in microarray platforms used for transcriptional profiling and heterogeneity in the patient populations included have made interpretation of results and comparison between studies difficult.[39] This has also limited the ability to perform meta-analysis of published data to aid decisions regarding the utility of putative gene signatures. Guidelines for reporting studies of prognostic markers (REMARK, Reporting recommendations for tumor marker prognostic studies[51]) and for microarray studies (MIAME, Minimum Information about a Microarray Experiment[52]), in addition to the CONSORT Standard for reporting clinical studies,[53] should assist the reporting of the study design, patient characteristics, methodology, and results of genomic marker studies in a transparent way, and facilitate the comparison of results between studies.

Regulatory hurdles

The traditional model of biomarker development has described five phases (Box 3) from the initial exploratory studies addressing the proposed research question, through to establishing the validity of the research assay and subsequent clinical validation studies. This is a lengthy process, and in the case of microarray-based technologies it may be difficult to deliver considering the difficulties in obtaining adequate numbers of appropriate tumor samples that adequately represent the heterogenous

Box 3 | Phases of biomarker study design[65]

Exploratory studies
Identification of potential biomarkers from pre-clinical or clinical studies

Clinical assay development and validation
Development and optimization of the clinical assay and assessment of its reproducibility

Retrospective longitudinal studies
Demonstration of the relationship between biomarker expression and clinical outcome, which also includes definition of criteria for 'positive' results

Prospective studies
Determination of the operating characteristics of the biomarker based on testing in a relevant population

Randomized controlled trials
Impact of the biomarker in the relevant population such as the evaluation of prognostic markers in the individual treatment arms of a randomized trial

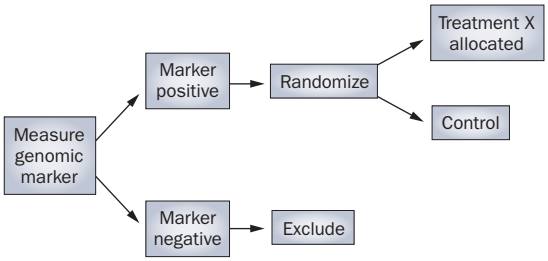

Figure 1 | 'Marker enrichment' clinical trial design. Expression of a known validated genomic marker is used to identify eligible patients for a randomized controlled trial of Treatment X. Marker expression is measured in all eligible patients, but only 'marker positive' patients are included in the treatment randomization process. 'Marker negative' patients are excluded from this study.[61]

patient population under investigation. Moreover, concerns have been raised regarding the potential use of this technology in routine clinical practice given the lack of standardization of microarray studies in the literature. As outlined earlier, a number of small studies have addressed the reproducibility of this methodology, but large-scale studies have yet to fully address these logistical issues. The Microarray Quality Control (MAQC) study is a project led by the FDA and involves academic, government and commercial institutions that are aiming to address these concerns regarding reliability, reproducibility and quality. The first phase (MAQC-I) included researchers from 51 different organizations, and evaluated the gene-expression data generated from transcriptional profiling of pooled samples derived from two commercially available reference RNA samples using seven different microarray platforms at multiple sites. Overall, the data generated was found to reflect underlying biology rather than technological variability, with the researchers able to demonstrate both inter-platform and inter-laboratory reproducibility and technical reliability.[54,55]

The second phase of the project (MAQC-II) will attempt to address the scientific issues around the development and validation of predictive classifiers, in particular the application of this approach in daily clinical management decisions. Collaborators will attempt to validate procedures and models within data sets, across data sets and prospectively with independent samples using independent analysis of microarray data sets.[56] It is likely that until this has been achieved and a consensus reached regarding the procedures necessary to develop a valid predictive model, skepticism regarding the routine clinical use of genomic technologies will remain, and integration of these technologies into routine clinical practice will remain limited.

Researchers have rapidly adopted microarray technology into translational research, and given initial promising results, a number of academic institutions in partnership with commercial organizations have been driving forward the implementation of this technology into clinical practice. Indeed, despite the criticisms of the published literature, microarray technology has been identified as a core technology for the advancement of medicinal product development and individualized medicine by the FDA's Critical Path Initiative,[57] and this organization has also issued guidelines regarding the submission of pharmacogenomic data in drug development and medical diagnostics to facilitate this process.[58]

The implementation of genomic technology into clinical decision making is most advanced in breast cancer, where several prognostic gene signatures have been made commercially available, and achieved FDA approval for clinical use (MammaPrint, Agendia) or have been endorsed in international guidelines of breast cancer management (Oncotype DX, Genomic Health).[2,59] Notably, this process of FDA approval and endorsement has occurred without large-scale prospective validation studies of these signatures. Indeed, prospective studies using these commercial prognostic signatures (the TAILORx and MINDACT trials) are still in the process of recruiting patients. Moreover, the widespread application of these prognostic signatures has occurred despite findings from the prospective validation RASTER study of the 70 gene 'MammaPrint' signature, in which samples from over one quarter of eligible patients were excluded due to technical failures.[60]

Finally, given the rapid pace of genomics research and the focus on developing targeted therapies that can be individualized to patients in a rational manner, to date, the evaluation of genomic markers in clinical studies seems a relatively inefficient process. This has prompted calls for changes in clinical trial design both in therapeutic studies with translational components, and in primary biomarker studies. In the case of a predictive classifier that has been previously developed and validated, a 'marker enrichment' design allocates treatment only to 'marker positive' patients (Figure 1). Alternatively, all patients are randomized to receive treatment but are tested for marker expression and subgroup analysis performed based on marker expression on completion of the study (Figure 2), in accordance with a pre-defined analysis plan.[61]

The classifier or marker in question for genomic markers have often not been validated, or may be one of

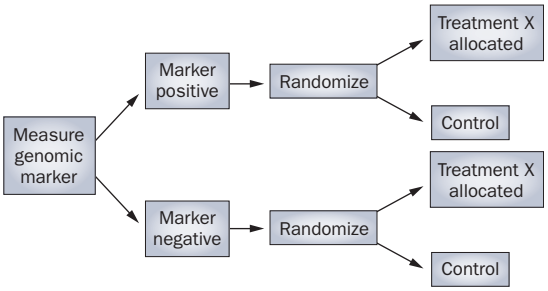

Figure 2 | 'Planned analysis' clinical trial design. Expression of a known validated genomic marker is measured in all patients eligible for a randomized controlled trial of Treatment X. Both 'marker positive' and 'marker negative' patients are included and randomized to receive Treatment X or control. Data is analysed in accordance with a pre-defined analysis plan.[61]

multiple potential classifiers before considering the trial. In this case, Freidlin and Simon proposed a novel model of adaptive trial design that within a single clinical trial can provide both development of the predictive classifier and evaluation of treatment effects in subgroups based on classifier expression.[62] Simulation studies evaluating this novel adaptive design relative to traditional study design have demonstrated that biomarkers or genomic signatures of drug sensitivity can be successfully incorporated into prospective randomized phase III trials, without impacting on the ability to detect an overall effect in the broad population. Moreover, in situations where the drug effect is restricted to a low proportion of the population the adaptive design demonstrated improvements in efficiency and reduced likelihood of concluding the treatment was ineffective. However, the authors acknowledge that using this adaptive approach may require an increase in the overall sample size, particularly in situations in which there is a small difference in treatment effect between the sensitive subgroup of patients and the non-sensitive subgroup of patients.[62,63]

Conclusions

Over the past two decades, there have been major advances in our understanding of tumor biology with the concomitant development of novel targeted therapeutic agents. However, the investigation and validation of clinically useful genomic markers accompanying these research developments has tended to be less focused, often occurring as retrospective subgroup analyses from larger trials or data used from series of tumor specimens available for research purposes, often obtained from heterogenous patient cohorts. Consequently, these studies have often been relatively underpowered with poorly defined study endpoints, inclusion criteria and protocols. Moreover, the allocation of cancer therapies to patients in both clinical trials and in the routine management of individual patients continues to be based predominantly on observations of population risk, rather than on the expression of validated biomarkers. This traditional approach fails to take account of the molecular heterogeneity of tumors and may account for the variations in outcomes observed among patients with tumors that have similar clinicopathological features, such as tumor stage or grade. This tumor heterogeneity has actually been highlighted by the new high-throughput genomic technologies, with 'class discovery' gene-expression profiling studies suggesting new subgroups within tumor types.

Despite these rapid technological advances there is still a paucity of genomic markers in routine clinical use at present. A rational and focused approach to the evaluation and validation of genomic markers is needed, whereby analytically validated markers are investigated in clinical studies that are adequately powered and have pre-defined patient populations and study endpoints. Furthermore, novel adaptive clinical trial designs, incorporating putative genomic markers into prospective clinical trials, will enable the evaluation of these markers in a rigorous and timely fashion, while maintaining an accurate assessment of treatment effect. Such approaches have the potential to facilitate the implementation of such markers into routine clinical practice and consequently enable the rational and tailored use of cancer therapies for individual patients.

Review criteria

Information was obtained by searching the PubMed database for articles published before 30 April 2009. The search terms used were "molecular markers" in association with "cancer", "implementation", "gene expression profiling", "microarrays" and "tumor markers". Only articles published in English were considered and primary sources have been quoted where possible. Full articles were obtained and references were checked for additional material when appropriate. The following online resources were accessed: http://linus.nci.nih.gov/brb/samplesize/; the FDA's Critical Path Initiative (http://www.fda.gov/oc/initiatives/criticalpath/) and 'Guidance for Industry: Pharmacogenomic Data Submissions' (http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126957.pdf).

1. Pusztai, L. *et al.* Clinical application of cDNA microarrays in oncology. *Oncologist* **8**, 252–258 (2003).
2. Harris, L. *et al.* American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J. Clin. Oncol.* **25**, 5287–5312 (2007).
3. Sotiriou, C. & Pusztai, L. Gene-expression signatures in breast cancer. *N. Engl. J. Med.* **360**, 790–800 (2009).
4. Workman, P. & de Bono, J. Targeted therapeutics for cancer treatment: major progress towards personalised molecular medicine. *Curr. Opin. Pharmacol.* **8**, 359–362 (2008).
5. Cunningham, D. *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N. Engl. J. Med.* **351**, 337–345 (2004).
6. Saltz, L. B. *et al.* Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. *J. Clin. Oncol.* **22**, 1201–1208 (2004).
7. Chung, K. Y. *et al.* Cetuximab shows activity in colorectal cancer patients with tumors that do not express the epidermal growth factor receptor by immunohistochemistry. *J. Clin. Oncol.* **23**, 1803–1810 (2005).
8. Khambata-Ford, S. *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.* **25**, 3230–3237 (2007).
9. Lievre, A. *et al. KRAS* mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J. Clin. Oncol.* **26**, 374–379 (2008).

10. Amado, R. G. *et al.* Wild-type *KRAS* is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 1626–1634 (2008).
11. Spano, J. P. *et al.* Potential predictive markers of response to EGFR-targeted therapies in colorectal cancer. *Crit. Rev. Oncol. Hematol*. **66**, 21–30 (2008).
12. Churchill, G. A. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32**, 490–495 (2002).
13. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
14. Luzzi, V. I., Holtschlag, V. & Watson, M. A. Gene expression profiling of primary tumor cell populations using laser capture microdissection, RNA transcript amplification, and GeneChip microarrays. *Methods Mol. Biol*. **293**, 187–207 (2005).
15. Ramaswamy, S. & Golub, T. R. DNA microarrays in clinical oncology. *J. Clin. Oncol*. **20**, 1932–1941 (2002).
16. Mook, S. *et al.* Daily clinical practice of fresh tumour tissue freezing and gene expression profiling; logistics pilot study preceding the MINDACT trial. *Eur. J. Cancer* **45**, 1201–1208 (2009).
17. Masuda, N. *et al.* Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res*. **27**, 4436–4443 (1999).
18. Oberli, A. *et al.* Expression profiling with RNA from formalin-fixed, paraffin-embedded material. *BMC Med. Genomics* **1**, 9 (2008).
19. Lassmann, S. *et al.* A novel approach for reliable microarray analysis of microdissected tumor cells from formalin-fixed and paraffin-embedded colorectal cancer resection specimens. *J. Mol. Med.* **87**, 211–224 (2009).
20. Chen, J., Byrne, G. E. Jr & Lossos, I. S. Optimization of RNA extraction from formalin-fixed, paraffin-embedded lymphoid tissues. *Diagn. Mol. Pathol.* **16**, 61–72 (2007).
21. Bibikova, M. *et al.* Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am. J. Pathol*. **165**, 1799–1807 (2004).
22. Horlings, H. M. *et al.* Gene expression profiling to identify the histogenetic origin of metastatic adenocarcinomas of unknown primary. *J. Clin. Oncol.* **26**, 4435–4441 (2008).
23. Farragher, S. M. *et al.* RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochem. Cell. Biol*. **130**, 435–445 (2008).
24. Scicchitano, M. S. *et al.* Preliminary comparison of quantity, quality, and microarray performance of RNA extracted from formalin-fixed, paraffin-embedded, and unfixed frozen tissue samples. *J. Histochem. Cytochem.* **54**, 1229–1237 (2006).
25. Srivastava, P. K. *et al.* A cut-off based approach for gene expression analysis of formalin-fixed and paraffin-embedded tissue samples. *Genomics* **91**, 522–529 (2008).
26. Hoshida, Y. *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).
27. Bertucci, F. *et al.* Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* **23**, 1377–1391 (2004).
28. Croner, R. S. *et al.* Common denominator genes that distinguish colorectal carcinoma from normal mucosa. *Int. J. Colorectal Dis*. **20**, 353–362 (2005).
29. Friederichs, 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*. **20**, 391–402 (2005).
30. Groene, J. *et al.* Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III. *Int. J. Cancer* **119**, 1829–1836 (2006).
31. Agrawal, D. *et al.* Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J. Natl Cancer Inst.* **94**, 513–521 (2002).
32. Ki, D. H. *et al.* Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int. J. Cancer* **121**, 2005–2012 (2007).
33. Koh, K. H. *et al.* Differential gene expression profiles of metastases in paired primary and metastatic colorectal carcinomas. *Oncology* **75**, 92–101 (2008).
34. Koehler, A. *et al.* Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. *J. Pathol.* **204**, 65–74 (2004).
35. Matsuyama, R. *et al.* Predicting 5-fluorouracil chemosensitivity of liver metastases from colorectal cancer using primary tumor specimens: three-gene expression model predicts clinical response. *Int. J. Cancer* **119**, 406–413 (2006).
36. Del Rio, M. *et al.* Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. *J. Clin. Oncol.* **25**, 773–780 (2007).
37. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
38. Simon, R. *et al.* Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl Cancer Inst.* **95**, 14–18 (2003).
39. Quackenbush, J. Microarray analysis and tumor classification. *N. Engl. J. Med.* **354**, 2463–2472 (2006).
40. Eschrich, S. & Yeatman, T. J. DNA microarrays and data analysis: an overview. *Surgery* **136**, 500–503 (2004).
41. Dupuy, A. & Simon, R. M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl Cancer Inst.* **99**, 147–157 (2007).
42. Dobbin, K. K., Zhao, Y. & Simon, R. M. How large a training set is needed to develop a classifier for microarray data? *Clin. Cancer Res.* **14**, 108–114 (2008).
43. National Institutes of Health. *Sample size calculation* [online] http://linus.nci.nih.gov/brb/samplesize/ (2009).
44. Wang, Y. *et al.* Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J. Clin. Oncol*. **22**, 1564–1571 (2004).
45. Hess, K. R. *et al.* Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J. Clin. Oncol*. **24**, 4236–4244 (2006).
46. Pusztai, L. & Hess, K. R. Clinical trial design for microarray predictive marker discovery and assessment. *Ann. Oncol.* **15**, 1731–1737 (2004).
47. Taylor, J. M., Ankerst, D. P. & Andridge, R. R. Validation of biomarker-based risk prediction models. *Clin. Cancer Res.* **14**, 5977–5983 (2008).
48. Simon, R. & Wang, S. J. Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenomics J*. **6**, 166–173 (2006).
49. George, S. L. Statistical issues in translational cancer research. *Clin. Cancer Res.* **14**, 5954–5958 (2008).
50. Dobbin, K. K. *et al.* Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin. Cancer Res.* **11**, 565–572 (2005).
51. McShane, L. M. *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK). *Nat. Clin. Pract. Oncol.* **2**, 416–422 (2005).
52. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
53. Moher, D., Schulz, K. F. & Altman, D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* **285**, 1987–1991 (2001).
54. Canales, R. D. *et al.* Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat. Biotechnol.* **24**, 1115–1122 (2006).
55. Shi, L. *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol*. **24**, 1151–1161 (2006).
56. Shi, L. *et al.* Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Curr. Opin. Biotechnol.* **19**, 10–18 (2008).
57. FDA US Food and Drug Administration. *Critical Path Initiative* [online] http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm (2009).
58. FDA US Food and Drug Administration. *Guidance for Industry: Pharmacogenomic Data Submissions* [online] http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126957.pdf (2009).
59. Geyer, F. C. & Reis-Filho, J. S. Microarray-based gene expression profiling as a clinical tool for breast cancer management: are we there yet? *Int. J. Surg. Pathol.* **17**, 285–302 (2008).
60. Bueno-de-Mesquita, J. M. *et al.* Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol.* **8**, 1079–1087 (2007).
61. Simon, R. The use of genomics in clinical trial design. *Clin. Cancer Res.* **14**, 5984–5993 (2008).
62. Freidlin, B. & Simon, R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin. Cancer Res.* **11**, 7872–7878 (2005).
63. Jiang, W., Freidlin, B. & Simon, R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl Cancer Inst.* **99**, 1036–1043 (2007).
64. Simon, R. Challenges of microarray data and the evaluation of gene expression profile signatures. *Cancer Invest.* **26**, 327–332 (2008).
65. Ransohoff, D. F. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *J. Clin. Epidemiol*. **60**, 1205–1219 (2007).