

Syllabus

PAOKC-cursus Klinische Chemie en Laboratoriumgeneeskunde

Tumormarkers bij solide tumoren

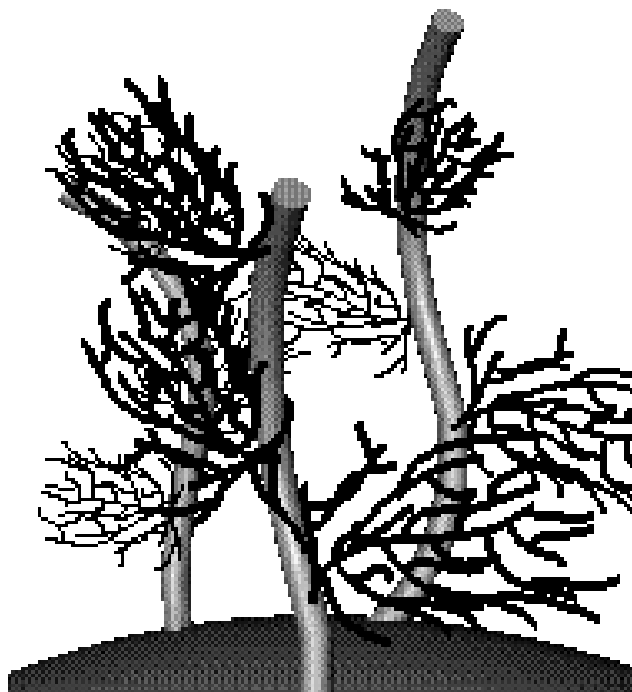


Werkgroep Tumormarkers

Nederlandse Vereniging
voor Klinische Chemie
en Laboratoriumgeneeskunde

Vrijdag 29 september 2006

Beatrixgebouw, Jaarbeurs, Utrecht



Programma

09.15 - 09.55	Ontvangst en koffie
	OCHTENDPROGRAMMA <i>Voorzitter: Prof.dr. C.G.J. Sweep</i>
09.55 – 10.00	Opening <i>Prof.dr. C.G.J. Sweep</i>
10.00 - 10.45	Practice guidelines for tumour marker use in the clinic <i>Dr. C.M. Sturgeon</i>
10.45 - 11.30	Epidemiologie van tumormarkers: screening, diagnose en prognose <i>Dr. E.W. Steyerberg</i>
11.30 - 12.15	Microarrays bij bepaling prognose in borstkanker <i>Dr. J.A. Foekens</i>
12.15 - 13.15	Lunch
	MIDDAGPROGRAMMA <i>Voorzitter: Dr. J. ten Kate</i>
13.15 - 14.00	Interpretatie van tumormarkers door de clinicus <i>Prof.dr. Th. Wobbes</i>
14.00 – 14.30	Thyreoglobuline als tumormarker <i>Dr. H.A. Ross</i>
14.30-15.00	Pauze
15.00-15.45	Proteomics: de stand van zaken <i>Prof.dr. M.P. van Dieijen-Visser</i>
15.45-16.30	Serum markers in breast cancer: are they of value, will they get better? <i>Prof.dr. M.J. Duffy</i>
16.30	Afsluiting

Sprekers

Dr. Catherine M. Sturgeon

Dept. of Clinical Biochemistry
Royal Infirmary of Edinburgh
51 Little France Crescent
Edinburgh EH16 4SA, UK
Tel.: +44 131 242 6871
Fax: + 44 131 242 6882
cs@csturgeon.net

Dr. Ewout W. Steyerberg

Department of Public Health
Erasmus MC, kamer AE-236
Postbus 2040
3000 CA Rotterdam
Tel.: 010 463 8470
Fax: 010 408 9449
e.steyerberg@erasmusmc.nl

Dr. John A. Foekens

Afdeling Interne Oncologie
Erasmus MC, kamer BE-426
Postbus 2040
3000 CA Rotterdam
Tel.: 010 408 8369
j.foekens@erasmusmc.nl

Prof.dr. Theo Wobbes

Afdeling Heelkunde
UMC St. Radboud
Postbus 9101
6500 HB Nijmegen
Tel.: 024 361 9908
t.wobbes@chir.umcn.nl

Dr. H. Alec Ross

Afdeling Chemische Endocrinologie
UMC St Radboud
Postbus 9101
6500 HB Nijmegen
Tel.: 024 361 4276
a.ross@ace.umcn.nl

Prof.dr. Marja P. van Diejen-Visser
Afdeling Klinische Chemie
Academisch Ziekenhuis Maastricht
Postbus 5800
6202 AZ Maastricht
Tel.: 043 – 387 6695
diejen@klinchem.azm.nl

Prof.dr. Michael J. Duffy
Dept. of Nuclear Medicine
St. Vincent's University Hospital
Dublin 4, Ireland
Tel.: +353 1 2094607
Fax: +353 1 2696018
micheal.j.duffy@ucd.ie

Organisatie

Organisatiecommissie

Namens de Werkgroep Tumormarkers:

Dr. Joke G. Boonstra

Erasmus MC
Rotterdam

Prof.dr. Fred C. G.J. Sweep

UMC St Radboud
Nijmegen

Dr. Joop ten Kate

Maaslandziekenhuis
Sittard

Namens de PAOKC commissie:

Dr. Nasser E. Ajubi

Medisch Centrum Leeuwarden
Leeuwarden

Inhoud

	Pagina
Practice guidelines for tumor marker use in the clinic <i>C. Sturgeon</i>	7
Epidemiology of tumor markers: screening, diagnosis and prognosis <i>E.W. Steyerberg</i>	13
Microarrays by bepaling prognose in borstkanker <i>J.A. Foekens</i>	21
Interpretatie van tumormarkers door de clinicus <i>Th. Wobbes</i>	29
Thyreoglobuline als tumormarker: wat is detecteerbaar? <i>H.A. Ross</i>	33
Application of seldi-tof-ms in protein profiling: state of the art <i>M.P. van Dieijen-Visser</i>	37
Serum markers in breast cancer: are they of value?, will they get better? <i>M.J. Duffy</i>	50

PRACTICE GUIDELINES FOR TUMOR MARKER USE IN THE CLINIC

C.M. Sturgeon, clinical researcher

Increasing pressure to provide health care based on “best practice” has stimulated local, national and international groups to develop guidelines in a number of clinical areas, particularly in cancer medicine, where diagnostic procedures are often invasive and therapy expensive.¹ The broad recommendations of the American Society of Clinical Oncology (ASCO)² have recently been complemented by more detailed guidelines from the National Academy of Clinical Biochemistry (NACB)³ in the United States and the European Group for Tumour Markers (EGTM)⁴ in Europe. Based on expert opinion and published reports, these guidelines include recommendations about which tumor markers are likely to be most helpful in given clinical circumstances. The NACB and EGTM guidelines also highlight requirements and pitfalls of tumor marker measurements in the pre-analytical, analytical and post-analytical phases.

Recommendations relating to the pre-analytical phase

The pre-analytical phase is in many respects the most important phase of laboratory analysis. Errors in the pre-analytical phase reportedly occur up to ten times as often as in the analytical phase, are often difficult to identify and if unrecognised can seriously compromise patient care. Care and attention to detail when requesting a tumor marker measurement is therefore essential.

Selecting the most appropriate tumor marker.

Tumor markers may be regarded as surrogate indicators that can increase or decrease a clinician’s suspicion that a future clinically important event will or will not occur, and/or that a specific treatment will reduce its likelihood.³ Their value is in permitting an assessment of risk that should enable therapy to be offered to those patients most likely to benefit, while reducing exposure to toxicities for those who would not benefit. Selection of the most appropriate marker should therefore take heed both of the clinical question – whether risk assessment, screening, diagnosis, prognosis, prediction or monitoring – and of the reliability of the separation in outcomes for marker positive and marker negative patients.

Unfortunately there are as yet relatively few well-designed and validated prospective studies for individual tumor markers. Many of the studies reported in the literature have

not been validated thoroughly (i.e. in the same assay, using the same cut-off limits and types of patients), while in others statistical significance ($p < 0.05$) in outcomes of two groups separated by marker results has incorrectly been regarded as evidence of clinical utility, which is not always the case. With these *caveats* in mind, it is nevertheless possible to make some recommendations about the most appropriate markers in given clinical situations, as summarised in Table 1 for some major cancer sites.³ [It should be noted that these recommendations apply to requests for tumor marker measurement in routine clinical practice and not necessarily to their use in clinical trials where other considerations may apply.]

Optimising specimen collection.

There is no strong evidence of diurnal variation for most markers, so specimens can be taken at any time of day. While the advent of the electronic health record should make it possible to link the ordering process with advice about relevant pre-analytical concerns, pending such developments it is important for both clinical and laboratory staff to be aware of clinical interventions and conditions which may transiently increase tumor markers concentrations. For example, blood for PSA should be taken before any manipulation of the prostate, and blood for CA125 should not be taken during menstruation, which may increase the serum concentration two to three-fold. PSA may also be increased markedly in men with urinary tract infections and prostatitis while CA125 may be mildly elevated in endometriosis and the first two trimesters of pregnancy and markedly raised in any patient with benign ascites. CA19.9 may be significantly increased in patients with cholestasis and patients in this category should be noted on the clinical report. The effect of medication and other treatment should also be considered: 5 α -reductase inhibitors cause a median decrease in PSA concentration of approximately 50%. Transient increases in tumor marker concentrations can also occur following chemotherapy. Cannabis may increase serum levels of hCG while smoking may slightly increase apparent CEA levels in some immunoassays.

The laboratory should provide clear and readily available advice about the appropriate tube type for each test, so as to ensure that manufacturers' instructions are always followed. Standardised conditions of specimen collection and fixation are crucial for immuno-histochemical analyses. Although most tumor markers are reasonably stable, serum or plasma should be separated and stored appropriately as soon as possible. At high ambient temperature the potential influence of transit time on analyte results should be considered. As for other analytes, the majority of pre-analytical errors for tumor markers will be simple

specimen handling errors – hemolyzed specimens, insufficient specimens, and incorrect specimens – and their occurrence should be minimized by adherence to good laboratory practice. Extra vigilance is required in ensuring that correct tumor marker results are reported, since reporting erroneous results is more likely to cause patients undue alarm than is the case for many other laboratory tests.

Recommendations relating to the analytical phase

In the analytical phase it is essential for satisfactory measurement of any analyte that laboratories ensure that methods used, whether immunoassay or immunohistochemical, are well validated and that their performance is carefully monitored. Implementation of rigorous Internal Quality Control (IQC) procedures and participation in well-designed Proficiency Testing [External Quality Assessment (EQA)] programs should help to ensure that methods are performing according to specification. The NACB has made recommendations for IQC and EQA provision that are generally applicable to all analytes, several factors being especially relevant for tumor marker measurement.^{3,4}

Excellent precision and reproducibility (intra-assay variability <5%; inter-assay variability <10%) is essential particularly near critical clinical decision points, for example in screening programs (e.g. when using PSA to select asymptomatic patients for biopsy) or where chemotherapy may be instituted on the basis of a rising tumor marker level in the absence of other scan evidence (e.g. when monitoring testicular cancer patients with AFP and hCG). Long-term assay stability, which can be readily assessed by proficiency testing schemes, should also be demonstrated since tumor markers are often measured for cancer patients over months or years.

Such long-term monitoring presents major analytical challenges as patients may be treated in different hospitals using different methods and laboratories may change tumor marker methods during the relevant time period. Data from proficiency testing schemes confirm that there are still significant between-method differences in results, with coefficients of variation in excess of 20% observed for some tumor markers. Poor calibration and differences in the specificity of antibodies used as well as method design all contribute to this variation. A number of international initiatives to address these issues are currently in progress. For the present, considerable care must be taken when changing tumor marker methods or when interpreting cumulated results for individual patients that have been obtained in different methods.

Table 1. Summary or current NACB recommendation for the use of tumor markers in specific malignancies.³

	Screening / early detection	Diagnosis / case-finding	Staging / prognosis	Detecting recurrence	Monitoring therapy
Testicular tumors	No tumor markers recommended	AFP, hCG, LDH	AFP, hCG, LDH	AFP, hCG, LDH	AFP, hCG, LDH
Prostate cancer	PSA, cPSA, %fPSA [With DRE]	PSA, cPSA, %fPSA [With DRE]	PSA, cPSA [With DRE & biopsy Gleason Grade]	PSA, cPSA	PSA, cPSA
Colorectal cancer	FOB [In subjects >50 years old; Genetic testing in high risk subjects]	No tumor markers recommended	CEA	CEA	CEA
Liver cancer	AFP [In high risk subjects]	AFP	AFP	AFP	AFP
Ovarian cancer	CA125 [Only in combination with TVUS for early detection in hereditary syndromes]	CA125 [Post-menopausal women only]	CA125	CA125	CA125
Breast cancer	No tumor markers recommended	No tumor markers recommended	ER, PR, HER-2, uPA, PAI-1	No tumor markers recommended	CA 15-3, CEA [Monitoring advanced disease]
Gastric cancer	No tumor markers recommended	No tumor markers recommended	No tumor markers recommended	No tumor markers recommended	No tumor markers recommended
Pancreatic cancer	No tumor markers recommended	CA19-9 [If used, only with CT or EUS and in an appropriate clinical context]	CA19-9	No tumor markers recommended	CA19-9 [During palliative therapy with imaging tests or after potentially curative surgery]
Cervical cancer	No tumor markers recommended	SCC [Possibly in squamous cell cervical carcinoma]	SCC [Possibly in squamous cell cervical carcinoma]	SCC [Possibly in squamous cell cervical carcinoma]	SCC [Possibly in squamous cell cervical carcinoma]

AFP, α -fetoprotein; CEA, carcinoembryonic antigen; CT, computed tomography; DRE, digital rectal examination; ER, estrogen receptor; EUS, examination under ultrasound; FOB, fecal occult blood; hCG,

human chorionic gonadotropin; LDH, lactate dehydrogenase; MIA, melanoma inhibiting activity; PAI-1, plasminogen activator inhibitor-1; PR, progesterone receptor; PSA, prostate specific antigen; cPSA, complexed PSA; %fPSA, % of free (uncomplexed) PSA to total (complexed + free) PSA; SCC, squamous cell carcinoma antigen; TVUS, transvaginal ultrasound; uPA, urokinase plasminogen activator

Differences in method design contribute not only to the numerical differences in results observed, but also influence method robustness to clinically relevant interferences. These include cross-reactions with closely related molecules, the high dose hook effect and interference from heterophilic or human anti-mouse antibodies. Maintaining vigilant awareness of the potential for such interference is very important and requires good understanding of the characteristics of the assays being used. Regular dialogue between laboratory and clinical staff should be actively promoted, since early discussion and investigation of any results that are not in accord with the clinical picture is likely to facilitate early identification of erroneous results caused by interference.

Recommendations relating to the post-analytical phase

In the post-analytical phase, dialogue can be encouraged by provision of brief clinical information by clinicians when requesting tumor marker measurements. It is recommended that laboratory reports for tumor markers include fully cumulated results, an appropriate reference interval and the name of the assay method used, together with an indication of whether any change in marker level is significant and, if appropriate, whether any change of method is likely to have affected interpretation of the trend in marker level.^{3,4}

Laboratories should also be actively involved in on-going audit of the clinical utility of the tumor marker results they provide and of their influence on clinical outcome. Proficiency testing schemes can also contribute by undertaking occasional surveys to compare practice in participating laboratories, highlighting differences in reference intervals, reporting practice and interpretation of clinical results. Such surveys can also provide an indication of how effectively guidelines such as those outlined here are being implemented in routine practice.

References

1. Sturgeon CM. Practice guidelines for tumor marker use in the clinic. *Clin Chem* 2002; 48: 1151-1159
2. Bast RC, Radvin P, Hayes DF, Bates S, Fritsche H, Jessup JM, *et al.* 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001; 19: 1865-78

3. NACB Practice Guidelines and Recommendations for use of tumor markers in the clinic. Available in electronic form at <http://www.aacc.org/NR/rdonlyres>
4. European Group on Tumor Markers (EGTM) Consensus Recommendations. Available in electronic form at <http://egtm.web.med.uni-muenchen.de/index2.html>

EPIDEMIOLOGY OF TUMOR MARKERS: SCREENING, DIAGNOSIS, AND PROGNOSIS

E.W. Steyerberg, epidemioloog

Epidemiology includes two main areas of research: explanatory research and prediction research. Explanatory research includes searching for causes of diseases (etiology), and the effect of treatments (therapeutic research). An example of an etiologic question is “Does smoking cause colorectal cancer?”; a therapeutic question is “Does high-dose chemotherapy increase survival in poor prognosis testicular cancer patients?”.

Tumor markers are especially relevant for prediction research, which includes questions regarding diagnosis (including early diagnosis of disease, i.e. screening), and prognosis (the outcome of a disease process). After presenting various introductory examples, some epidemiological principles are discussed, followed by challenges in the scientific development of clinically useful new tumor markers.

Tumor markers in screening, diagnosis and prognosis

With screening (early diagnosis), we aim to detect a disease early in its course, before it would manifest itself clinically by signs and symptoms of the patient. If detected earlier, we expect that the disease can be treated better, such that long-term outcomes are improved. Screening can be done in asymptomatic subjects in the population, for example, the prostate-specific antigen (PSA) level in the blood can be used to detect prostate cancer. Subjects with high PSA are evaluated further with biopsy to diagnose presence or absence of cancer. Screening can also be done after treatment of cancer to detect recurrent disease during follow-up. For example, we can test PSA levels after radical prostatectomy to monitor patients. Or we use carcinoembryonic antigen (CEA) to detect relapse of colorectal cancer.

Tumor markers may also assist in diagnosing disease, for example in the case of unknown primary tumors. Also, marker decline is a sign of response to treatment. A complete remission of disease often requires normalization of markers if these were initially elevated, in combination with other criteria, such as CT scan assessments. Other examples of diagnosing disease are the prediction of relevant cancer. In prostate cancer, we can consider small, confined cancers without poorly differentiated characteristics as probably indolent. These cancers do not require radical prostatectomy. The likelihood of an indolent

cancer can be estimated by a combination of PSA levels, prostate volume, and biopsy features ¹. Similarly, patients with metastatic testicular cancer often have small, residual masses after treatment with chemotherapy. If these masses are necrotic, surgical resection is not necessary. The likelihood of a necrotic residual mass can be estimated by a combination of tumor marker levels (alpha-fetoprotein, AFP, and the beta subunit of human chorionic gonadotropin, b-HCG), CT scan measurements (pre- and postchemotherapy size), and initial histology (presence of teratoma elements) ^{2,3}.

For prognosis, tumor markers are often important predictors, since they reflect the extent and aggressiveness of the cancer. A good prognosis can often be based on the combination of patient, tumor, and treatment characteristics. Solely using extent of disease as with TNM staging can often be improved by considering more predictive characteristics ⁴. For example, the survival of testicular cancer patients can be predicted based on AFP and HCG levels, combined with extent of disease characteristics ⁵. Prognostic classifications can be devised, which are helpful for informing patients, decision-making on treatment, and in medical research, e.g. for stratification in randomized clinical trials.

Table 1 Diagnostic and prognostic epidemiologic research with examples of tumor markers

Area	Question	Example
Screening	Can we detect the disease early in its course, such that it can be treated with better prognosis?	Does PSA testing lead to early detection and better treatment of prostate cancer? Does CEA testing lead to early detection and better treatment of recurrent colorectal cancer, after curative treatment?
Diagnosis	Can we make a diagnosis, which guides treatment choice?	Does this patient have a complete remission of his/her disease, including normalization of markers? Is the cancer relevant to treat, or is the marker profile that favorable that the disease may be treated conservatively?
Prognosis	What is the likely outcome of the disease?	Given AFP and HCG levels and other characteristics, what is the 5-year survival of this testicular cancer patient?

Epidemiological aspects of screening

Screening requires a number of conditions to be fulfilled. First, we require that the tumor marker can separate subjects without from those with the cancer. This is similar for diagnostic studies. For example, a high PSA is found among patients with prostate cancer, but also among some subjects without cancer. The latter are false-positive classifications by PSA testing. A low PSA is found among most subjects without prostate cancer, but also among some with prostate cancer (false-negatives). The characteristics of a screening test can be summarized in sensitivity and specificity (Table 2).

Table 2 Sensitivity and specificity of a tumor marker in screening or diagnosing cancer. Sensitivity, or true positive rate, is defined as $a/(a+c)$. Specificity, or true negative rate, is defined as $d/(b+d)$.

	Cancer	No cancer
Tumor maker positive	a	b
Tumor maker negative	c	d

A good test has a high sensitivity and high specificity. Especially a high specificity is important in screening settings, to prevent the finding of many false-positive results among the many subjects without cancer (cell d in Table 2). The trade-off between sensitivity and specificity can well be visualized in a receiver operating characteristic (ROC) curve. We plot the true-positive rate (sensitivity) against the false-positive rate ($1 - \text{specificity}$), for consecutive cut-points of the tumor marker (Fig 1). The area under the ROC curve is a measure for the overall diagnostic performance of the test.

In screening, we moreover need to have high specificity at a time before the cancer is clinically diagnosed. A careful selection of cases and controls is hence important. Often, diagnostic qualities are less at longer time before diagnosing the cancer. See Figure 1 for an illustration of PSA and prostate cancer ⁶.

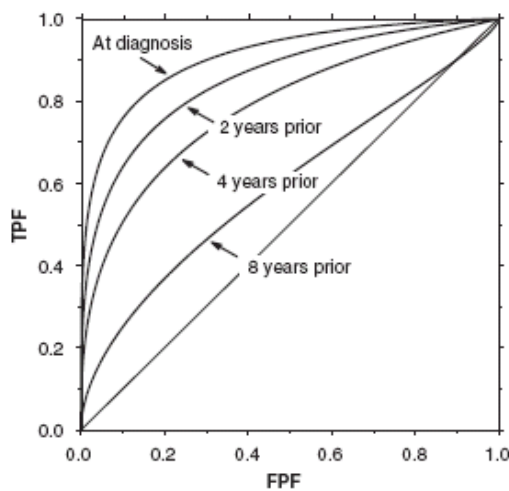


Figure 1 Time-dependent ROC curves for PSA in screening for prostate cancer. TPF: true-positive fraction; FPF: false-positive fraction. From Pepe, 2005 ⁶.

A further requirement of a useful screening test is that the detected disease has a better prognosis than when detected later. Lead time bias and length bias are among the major problems in the evaluation of effectiveness of treatment in screen-detected cancers.

Lead time bias refers to the apparent increase in survival time as calculated from date of diagnosis, while no true increase is caused by the earlier detection of the cancer. By

screening, we intend to diagnose a cancer earlier than it would be without screening. Without screening, the disease may be discovered later once symptoms appear. Even if in both cases a person will die at the same time, because we diagnosed the disease early with screening, the survival time since diagnosis is longer with screening. No additional life years have been gained, while we may have added anxiety as the patient lives with knowledge of the disease for longer.

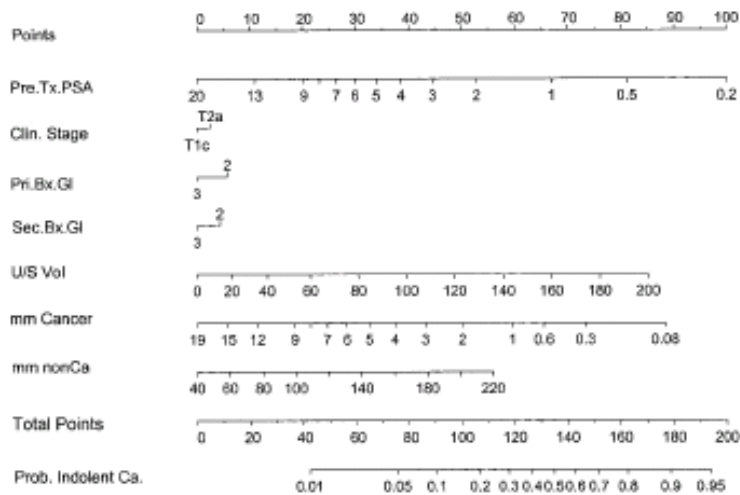
Length bias relates to the type of cancers that are detected by screening. For many cancers, fast growing tumours have worst prognosis. Screening is more likely to catch the slower growing cancers, which have a higher survival rate.

Before a screening program is implemented, it should hence be ensured that putting it in place would do more good than harm. A cross-sectional study showing reasonable sensitivity and specificity is not enough. The best studies for assessing whether a screening test will increase a population's health are rigorous randomized controlled trials. Indeed, for some cancers randomized trials have been performed or are underway (e.g. breast cancer, prostate, lung).

Epidemiological studies of diagnosis

As for screening, we require that the tumor marker can separate subjects without from those with the cancer. However, setting a diagnosis is usually not possible with a single characteristic, and has a sequential nature, starting with simple tests. Moreover, some diagnoses are probabilistic in nature, i.e. that we can give a probability of a certain condition rather than 100% certainty. An example is the qualification of a screen-detected prostate cancer as probably indolent. PSA levels, prostate volume, and biopsy features were combined in a logistic regression model to estimate this probability¹. This model was presented as a nomogram, such that it would be relatively easy to apply by physicians. A nomogram is a graphical presentation of the model (Fig 2); alternatives include tables of predicted probabilities according to predictive characteristics, and score charts.

The nomogram from Fig 2 was validated recently in patients from the European screening trial on prostate cancer (ERSPC), and found systematically invalid⁷. The probability of indolent cancer was around 50% in the ERSPC, while the average predicted probability was 20%. An updated version of the model was constructed, which may guide decision making in screen-detected prostate cancer after further validation.



Instructions for Physician: Locate the patient's PSA on the **PreTx PSA** axis. Draw a line straight upwards to the **Points** axis to determine how many points towards having an indolent cancer the patient receives for his PSA. Repeat this process for the remaining axes, each time drawing straight upward to the **Points** axis. Sum the points achieved for each predictor and locate this sum on the **Total Points** axis. Draw a line straight down to find the patient's probability of having indolent cancer.

Instruction to Patient: "Mr. X, if we had 100 men exactly like you, we would expect <predicted probability from nomogram * 100 > to have indolent cancer."

Figure 2 Nomogram to predict the probability of indolent cancer, as published by Kattan et al¹. Pre.Tx., pretreatment; Clin., clinical; Pri.Bx.Gl, primary biopsy Gleason score; Sec.Bx.Gl, secondary biopsy Gleason score; U/S, ultrasound; Prob., probability.

Epidemiological studies of prognosis

For prognostic studies, we usually consider a cohort of patients, and follow them in time for the occurrence of a relevant medical outcome. In cancer, common outcome measures are recurrence of disease, and (cancer-specific) survival. Tumor marker values can be related to these outcomes in statistical analyses. A simple cross-table can be constructed for high and low tumor marker values versus the outcome, when we know the outcome for all patients. For example, we may want to study 1 year survival and have at least 1 year follow-up for all patients. Survival analysis is required when the outcome is not observed for all patients, e.g. 5-year survival. Kaplan-Meier curves can then be constructed, which consider patients who did not die as censored observations. They are considered in the analysis until the end of their follow-up period (Fig 3).

Combinations of characteristics can be considered in the Cox regression model. In such a regression model, we do not have to categorize tumor marker values, but can analyze them as continuous variables. For example, we can consider the log(PSA) in predicting survival of prostate cancer patients. Predictions from such models usually need validation in new patients before we can use them in clinical practice, especially when the number of events

in the analysis was relatively small, and new markers were considered in the model rather than well-established markers.

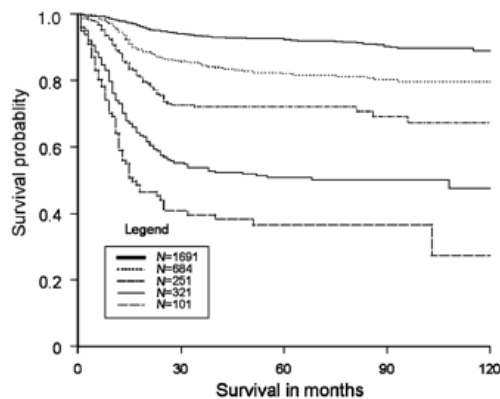


Figure 3 Survival curves for patients with nonseminomatous testicular cancer, as published by Van Dijk *et al*⁸. Five groups were created based on the IGCC classification, which includes tumor markers AFP and HCG⁵.

Developing useful new tumor markers

The development of clinically useful new tumor markers poses a number of challenges. Technological advancements allow for an increasing number of markers to be studied, including biochemical and genetic markers. Standardization of the tests and reproducibility of results are minimum requirements before thinking about wider applications of new markers in medicine. Case-control studies can subsequently be conducted using ‘convenient samples’, e.g. from stored tissue, but better with population based samples⁶.

Specific problems arise in the current evaluation of new markers⁹. First, a great number of markers are potential candidates for use in screening, diagnosis, or prognosis. This leads to a multiple testing problem. If we evaluate 1000 markers, we expect that 50 of them reach statistical significance with the classical $p < 0.05$ criterion, even if none of them is truly associated with cancer. Genome-wide searches include even more potential markers. Several approaches can be followed to address the multiple testing problem, including Bonferroni’s correction, but such approaches limit the statistical power of a study, i.e. the probability of finding a significant result if a true association was present. Power in markers studies is typically already quite limited because of small sample size, e.g. less than 100 cases, and less than 100 controls. Initiatives have been proposed for analyses with larger sample sizes¹⁰. Further, results of marker evaluations may only be reported if statistically significant¹¹. The effects are then overestimated, since non-significant results are by definition closer to ‘no effect’, and are not reported (‘publication bias’)¹².

A specific issue is how we deal with continuous values of a marker. It may be convenient to search for a dichotomization as positive versus negative. Searching for an optimal cut-off is fraught with statistical problems; the association will be exaggerated¹³. In general, dichotomization is discouraged; continuous versions of a marker contain often much more information. In addition to the linear and logarithmic transformation, spline functions may well be used to study diagnostic and prognostic relationships¹⁴.

Finally, new markers should be judged for their incremental value over classical markers and other diagnostic or prognostic characteristics. This can well be studied with regression models, where first a model is made with the traditional predictors, and next a model with these traditional predictors plus the new marker. If the latter model is promising, it should be validated in new patients to assess generalizability. In many clinical areas, the choice of more modern statistical methods, e.g. a classification tree or neural network instead of a regression model, did not improve the quality of predictions¹⁵.

Summary points

- Tumor markers have an important role in screening, diagnosis, and prognosis of cancer
- Sensitivity and specificity are important characteristics of a screening or diagnostic test; they can jointly be considered in a ROC curve
- Lead time bias and length bias are among the major problems in the evaluation of screening programs; randomized controlled trials are the ultimate way to evaluate benefits and harms of screening
- Regression models may combine tumor markers with other characteristics to estimate a probability of a diagnosis (e.g. indolent cancer) or a prognostic outcome (e.g. survival). These models require validation before widespread use is possible
- The development of clinically useful new tumor markers is associated with several problems, including multiple testing, limited power because of small sample sizes, publication bias, and having to deal with continuous marker values. The incremental value should be assessed over traditional predictors, with rigorous validation of initially (too?) promising results.

References

1. Kattan MW, Eastham JA, Wheeler TM, et al. Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. *J Urol* 2003; 170:1792-7.

2. Steyerberg EW, Keizer HJ, Fossa SD, et al. Prediction of residual retroperitoneal mass histology after chemotherapy for metastatic nonseminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. *J Clin Oncol* 1995; 13:1177-87.
3. Steyerberg EW, Keizer HJ, Habbema JD. Prediction models for the histology of residual masses after chemotherapy for metastatic testicular cancer. ReHiT Study Group. *Int J Cancer* 1999; 83:856-9.
4. Kattan MW. Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. *Curr Opin Urol* 2003; 13:111-6.
5. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997; 15:594-603.
6. Pepe MS. Evaluating technologies for classification and prediction in medicine. *Stat Med* 2005; 24:3687-96.
7. Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schröder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol* 2007; Jan.
8. van Dijk MR, Steyerberg EW, Stenning SP, Dusseldorp E, Habbema JD. Survival of patients with nonseminomatous germ cell cancer: a review of the IGCC classification by Cox regression and recursive partitioning. *Br J Cancer* 2004; 90:1176-83.
9. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; 2:e124.
10. Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006; 38:3-5.
11. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 2005; 97:1043-55.
12. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; 58:543-9.
13. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005; 97:1180-4.
14. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer, 2001:xxii, 568.
15. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. A comparison of statistical learning methods on the Gusto database. *Stat Med* 1998; 17:2501-8.

MICROARRAYS BIJ BEPALING PROGNOSE IN BORSTKANKER

J.A. Foekens, biochemicus

Achtergrond

Klinische en histopathologische eigenschappen van de borsttumor bepalen voornamelijk de keuze van therapie. In het huidige beleid wordt bij vrouwen met primaire borstkanker (tumor in de borst) de therapiekeuze (al dan niet adjuvante systemische hormonale- en/of chemotherapie en/of immunotherapie) gemaakt op grond van de leeftijd en menopausale status van de patiënt, het al of niet aanwezig zijn van tumorcellen in de lymfeklieren, en op basis van kenmerken van de tumor (tumorgrootte, differentiatiegraad, steroidhormoon[ER/PgR]-receptor status en, sinds kort, ook de HER2-neu status). Voor vrouwen met reeds aantoonbaar uitgezaaide ziekte wordt voor het geven van een bepaalde vorm van systemische therapie voornamelijk rekening gehouden met de klinische symptomen, locatie van de metastase(n), leeftijd van de patiënt en de ER/PgR/HER2-neu status van de tumor. De behandeling is palliatief en de patiënten zullen uiteindelijk overlijden ten gevolge van het resistent worden van de metastasen tegen alle vormen van therapie.

In veruit de meeste gevallen wordt na de chirurgische verwijdering van de primaire tumor een adjuvante systemische therapie gegeven om de kans op het later optreden van metastasen te verminderen. Ook vrouwen van wie tijdens de primaire chirurgie de lymfeklieren nog niet aangedaan zijn met tumorcellen (N0 patiënten) en derhalve een relatief goede prognose hebben, krijgen voor het overgrote deel een adjuvante systemische therapie. Dit ondanks het feit dat van deze groep van patiënten met een goede prognose het merendeel al genezen is door de regionale behandeling met chirurgie, al of niet aangevuld met radiotherapie. Bovendien is een adjuvante systemische behandeling met de bestaande medicijnen nog lang geen garantie dat de tumor later niet terugkomt. De afgelopen 20 jaar is er daarom veel onderzoek verricht naar celbiologische factoren die de prognose van de patiënt en de response op therapie kunnen voorspellen. Dit type onderzoek is erg bewerkelijk en in de meeste gevallen werden er slechts één of enkele factoren tegelijk onderzocht. Het onderzoek naar celbiologische karakterisering van tumoren maakt op het ogenblik een stormachtige ontwikkeling door. Dit is in de eerste plaats te danken aan het ophelderen van de genetische code aan het eind van de vorige eeuw. Gebaseerd op deze informatie zijn technieken ontwikkeld om de mRNA-expressie van tienduizenden genen

tegelijk te kunnen bestuderen. Aangezien een patiënt met een gemetastaseerde borstkanker niet meer te genezen is door het uiteindelijk resistent worden van celklonen binnen de metastasen tegen alle huidige beschikbare medicijnen, is het van groot belang om te kunnen begrijpen waarom de ene tumor wel kan uitzaaien en de andere niet. Verder is het van belang om op voorhand te weten of een tumor zal reageren op een bepaalde vorm van systemische behandeling. In ons onderzoek in het Erasmus MC in Rotterdam hebben wij met behulp van genexpressiearrays honderden borsttumoren onderzocht met betrekking tot het vermogen om uit te kunnen zaaien en hebben we de effectiviteit bepaald van specifieke vormen van therapie. Bovendien is een genprofiel ontwikkeld dat kan voorspelen of een tumor uit zal zaaien naar een bepaald orgaan, met name naar de bot.

Primaire borstkanker

De groep van lymfeklier-negatieve patiënten beslaat tegenwoordig ongeveer 60-65% van alle borstkankerpatiënten die jaarlijks in Nederland geopereerd worden. Omdat deze patiënten een relatief goede prognose hebben, kregen ze in ons land tot voor kort over het algemeen geen adjuvante systemische therapie. Echter, indien deze patiënten niet aanvullend behandeld worden met adjuvante therapie, ontwikkelen er zich in ongeveer 30-40% van de patiënten op termijn één of meerdere metastasen. De bijdrage van adjuvante systemische therapie in deze groep van patiënten varieert van 5 tot 10% absoluut overlevingsvoordeel (1,2). Indien echter al deze patiënten behandeld zouden worden met adjuvante systemische therapie, hetgeen het geval is voor ca. 85-90% van de patiënten indien de St. Gallen (3,4) of de US National Institutes of Health (5) consensus richtlijnen gehanteerd zouden worden, wordt het merendeel onterecht met vaak toxische systemische therapie behandeld. Het is derhalve van belang om de patiënten die geen aanvullende behandeling nodig hebben omdat ze al genezen zijn door de regionale behandeling (chirurgie met of zonder aanvullende radiotherapie) op voorhand te kunnen identificeren om hen de belasting van vooral chemotherapie niet te hoeven laten ondergaan. Patiënten die wel een hoog risico hebben voor het krijgen van metastasen kunnen, afhankelijk van het celbiologische phenotype van de tumor (bijvoorbeeld ER/PgR-status, HER2/neu-status), behandeld worden met een specifieke vorm van adjuvante systemische therapie.

Patiënten bij wie meerdere lymfeklieren reeds aangedaan zijn met tumorcellen ten tijde van de chirurgische verwijdering van de primaire tumor uit de borst (lymfeklier-positieve patiënten) hebben een slechte prognose. Bij deze patiënten bestaat er, afhankelijk van het

aantal aangedane lymfeklieren, een grote kans (60-80%) dat de ziekte terugkomt op andere plaatsen in het lichaam. Deze patiënten die een hoog risico hebben om later metastasen te ontwikkelen, krijgen kort na de operatie bijna allemaal een vorm van adjuvante systemische therapie. Op deze manier hoopt men de mogelijk aanwezige micro-metastasen uit te roeien.

Gemetastaseerde ziekte en therapie-resistentie

Ongeveer drie-kwart van de primaire borsttumoren zijn hormoonreceptor-positief. Patiënten met uitgezaaide ziekte zullen op een hormonale behandeling in 50-60% van de gevallen gunstig reageren met een complete of partiële remissie of een landurige stabiele ziekte. Voor patiënten met een ER/PgR-negatieve tumor is dit minder dan 10%. Patiënten met metastasen op afstand reageren over het algemeen (in 50-60% van de gevallen) gunstig op chemotherapie alvorens de tumor uiteindelijk resistent wordt tegen systemische therapie. Er is dus een duidelijke klinische behoefte om vast te kunnen stellen welke patiënten wel en welke niet gunstig zullen reageren op een specifieke vorm van therapie.

Prognostisch genexpressieprofiel

Bij het verkrijgen van het Rotterdamse prognostische profiel hebben wij het RNA van 286 primaire borsttumoren geanalyseerd op Affymetrix U133a chips, dat informatie bevat over 22.000 messenger-RNAs. Geen van de patiënten, allen lymfeklier-negatief en behandeld tussen 1980 en 1995, had enige adjuvante systemische therapie gekregen. Dit maakte het mogelijk om de zuivere prognostische waarde van genexpressie te bestuderen, zonder dat de resultaten beïnvloed konden worden door effecten van systemische therapie. Het profiel dat ontwikkeld werd bestaat in totaal uit 76 genen, 60 voor de ER-positieve tumoren, 16 voor de ER-negatieve tumoren. In de validatiegroep bleek de sterke prognostische waarde van het profiel bevestigd te kunnen worden (6). Met een nauwkeurigheid van 93% kon het 76-genprofiel voorspellen welke patiënten wel of niet een uitzaaiing kregen binnen vijf jaar na chirurgie. In de analyse voor metastase-vrije overleving bleken patiënten met een hoog-risico 76-genprofiel een hazard ratio (HR) te hebben van 5,7. Dit betekent dat ze een bijna zes maal zo grote kans hebben om een afstandsmetastase te krijgen vergeleken met patiënten van wie de tumor een laag-risico 76-genprofiel heeft. De prognostische waarde van het 76-genprofiel was ook sterk aanwezig in subgroepen van pre- (HR = 9,6) en

postmenopausale patiënten (HR = 4,0). In een aanvullende studie werd de prognostische waarde van het 76-genprofiel bevestigd in 180 lymfeklier-negatieve patiënten van vier Europese ziekenhuizen (7). Geen van de patiënten in deze multicentrische studie had enige adjuvante systemische therapie ontvangen. Het RNA werd geanalyseerd op speciaal gefabriceerde Affymetrixchips waarop alleen de informatie van de 76-genen uit het Rotterdamse profiel en een hoeveelheid controle-genen waren gespot. In deze volledig onafhankelijke validatieset bleek het 76-genprofiel opnieuw een sterke prognostische factor. De gevonden HR voor metastase-vrije overleving was 7,4. Voor pre- en postmenopausale patiënten werden HRs gevonden van respectievelijk 4,8 en 9,8. In de multivariate analyse voor metastase-vrije overleving was het 76-genprofiel, idem als in de vorige studie (6), weer de enige statistisch significante factor met een HR van 11,4. Deze resultaten bevestigen dat het 76-genprofiel bruikbaar is voor alle lymfeklier-negatieve patiënten, ongeacht leeftijd, menopausale status, tumorgrootte en tumorgraad.

Indien het Rotterdamse 76-genprofiel vergeleken wordt met het eerder gepubliceerde Amsterdamse 70-genprofiel (8, 9), is het belangrijkste verschil dat het Rotterdamse profiel ook bruikbaar is bij patiënten ouder dan 55 jaar, terwijl het Amsterdamse profiel werd ontwikkeld voor alleen patiënten jonger dan 55 jaar. Van het Amsterdamse 70-genprofiel en het Rotterdamse 76-genprofiel waren maar 3 genen identiek. Dit is niet echt verrassend te noemen omdat er van de 70 genen van het Amsterdamse profiel, dat verkregen werd met de Agilent array, er slechts 48 aanwezig zijn op de Affymetrix U133a array die gebruikt werd voor het maken van het Rotterdamse 76-genprofiel. Aan de andere kant waren er van de 76 genen uit het Rotterdamse profiel slechts 38 aanwezig op de array die gebruikt werd voor het maken van het Amsterdamse 70-genprofiel. Een ander belangrijke oorzaak die de verschillen in de beide profielen kunnen verklaren is ongetwijfeld gelegen in het feit dat in het Amsterdamse onderzoek geen onderscheid werd gemaakt tussen ER-positieve en ER-negatieve tumoren, terwijl in het Rotterdamse onderzoek aparte profielen werden gemaakt voor de ER-positieve subgroep (60 genen) en de ER-negatieve subgroep van tumoren (16 genen). Niettegenstaande de verschillen in beide profielen, zijn er veel overeenkomstige signaalpaden vertegenwoordigd in beide profielen, maar toch waren er ook hier duidelijk merkbare verschillen (10).

Botmetastaseprofiel

Het bot is de meest frequente plaats waar de tumor naar uitzaait in o.a. patiënten met borstkanker. Borstkanker kan resulteren in osteolytische botlesies and het is aannemelijk dat het micromilieu in het bot er voor zorgt dat de circulerende kankercellen kunnen hechten en uitgroeien tot macrometastasen. Het is van belang om te kunnen voorspellen of de tumor uit zal zaaien naar de bot of andere organen om een eventuele extra adjuvante therapie te geven dat specifiek de homing and uitgroei van de uitzaaiingen in betreffende organen kan voorkomen. In dit kader zouden bisphosphonaten mogelijk botmetastasen kunnen voorkomen (11). Van de 107 patiënten van wie the locatie van de metastasen bekend was onze genexpressie studie (6), werd onderzocht of een genprofiel kon voorspellen dat de tumor naar de bot zou gaan uitzaaien. Een panel van 69 genen bleek differentiëel tot expressie te komen tussen tumoren die wel en die niet naar de botten uitzaaiden (12). De twee meest discriminerende genen waren trefoil factor-1 (TFF1) en TFF3. TFF1 is een oude bekende die voorheen beschreven werd als het oestrogen-gereguleerde pS2-eiwit dat een gunstige prognostische en predictieve waarde bleek te hebben (13-16). Bovendien bleek dat het fibroblastgroeifactor-signaalpad betrokken is bij uitzaaiing van borstkankercellen naar de botten. Een 31-genprofiel dat vervolgens ontwikkeld werd in de trainingset, toonde bij validatie een sensitiviteit van 100% en een specificiteit van 50%. Met andere woorden, met het 31-genprofiel konden alle patiënten die een uitzaaiing naar het bot kregen geïdentificeerd worden, terwijl tumoren die niet naar de botten uitzaaiden in de helft van de gevallen toch een positief 31-genprofiel bleken te hebben (12). Indien deze resultaten bevestigd kunnen worden in een grotere en onafhankelijke groep van patiënten, zou overwogen kunnen worden om borstkankerpatiënten met een positief 31-genprofiel van de primaire tumor een extra behandeling te geven bovenop de reeds geïndiceerde adjuvante systemische therapie voor de hoog-risico patiënten.

Therapie-resistentie genexpressieprofielen

Met betrekking tot de effectiviteit van tamoxifen behandeling voor ER-positieve patiënten hebben we de expressie van het RNA van 112 ER-positieve primaire tumoren bestudeerd van patiënten van wie bekend was dat ze later metastasen ontwikkeld hadden en hiervoor behandeld waren met tamoxifen als eerstelijns enkelvoudige therapie. In een trainingset

van 46 tumoren kwamen 81 genen differentiëel tot expressie tussen tumoren die wel en die niet resistent waren tegen de hormonale therapie. De genen waren voornamelijk betrokken bij oestrogeenwerking, apoptose, extracellulaire matrix-aanmaak en immuunrespons. Uitgaande van deze 81 genen werd een 44-genprofiel gemaakt dat bij validatie op een onafhankelijke groep van 66 tumoren een sterke predictieve waarde te zien gaf. Het 44-genprofiel voorspelde beter welke patiënt een tamoxifen-resistente tumor had dan de combinatie van de traditionele factoren inclusief leeftijd en menopausale status, ziektevrij interval, hoeveelheid ER en PgR, en de plaats van de uitzaaiing (17).

Onlangs hebben we ook een genexpressieprofiel ontwikkeld dat kan voorspellen welke tumoren niet zullen reageren op conventionele chemotherapie. Hiervoor werden primaire tumoren onderzocht van 156 patiënten die chemotherapie (103 x FAC, 53 x CMF) hadden ontvangen als eerstelijns behandeling voor uitgezaaide ziekte. Van de chemotherapie-resistente profielen die voor alle patiënten, en apart ook voor de ER-positieve en ER-negatieve subgroepen ontwikkeld werden, was alleen een 31-genprofiel voor de ER-negatieve subgroep voldoende robuust dat het bevestigd kon worden in een onafhankelijke testset. Met een gevoeligheid van 91% en een specificiteit van 56% kon het 31-genprofiel voorspellen welke patiënten niet op chemotherapie zullen reageren (18).

Toekomstverwachting

Genprofielen zijn veelbelovend bij het vaststellen van de prognose van patiënten met primaire en gemetastaseerde borstkanker. Echter, geen van de huidig beschikbare expressieprofielen zijn zodanig gevalideerd dat ze in de praktijk toegepast kunnen worden. Desondanks zijn er al twee prognostische testen commercieel beschikbaar. Agendia verkoopt een test voor het bepalen van het Amsterdamse prognostische 70-genprofiel en Genomic Health verkoopt een test voor het bepalen van een prognostisch 21-genprofiel. Dit laatste genprofiel werd ontwikkeld voor patiënten die adjuvante hormonale therapie krijgen (19). In deze studie werd gebruik gemaakt van 21 (inclusief 5 controle genen) vooraf sterk geselecteerde genen, die door de onderzoekers echter niet verkregen waren door een genomwijde screening zoals wij in het NKI (8) en Erasmus MC (6) wel gedaan hebben. De Amerikaanse onderzoekers vermelden echter niet hoe de selectie van de genen heeft plaats gevonden. Bovendien is er nog geen onafhankelijke validatiestudie gepubliceerd.

Het ligt in de lijn der verwachting dat genprofielen binnen afzienbare tijd ingepast gaan worden in consensus aangaande diagnose en behandeling van borstkanker. Het is nog niet duidelijk welk detectie-platform het uiteindelijk zal gaan worden. Een alternatief voor microarrays is om voor elke klinische vraagstelling een multiple quantitative PCR-assay te gebruiken zoals onlangs ontwikkeld voor patiënten die adjuvante hormonale therapie krijgen [19]. Het voordeel van een multiplex PCR-assay is dat het uitgevoerd kan worden op paraffine-materiaal waardoor de logistiek een stuk gemakkelijker is.

Referenties

1. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998;352:930-942.
2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of randomised trials. *Lancet* 1998;351:1451-1467.
3. Goldhirsch A, Wood C, Gelber RD, Coates AS, Thürlimann B, Senn HJ. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003;21:3357-3365.
4. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thürlimann B, Senn HJ & Panel Members. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005;16:1569-1583.
5. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ, Deshler A, Fulton S, Hendricks CB, Kemeny M, Kornblith AB, Markman M, Mayer R, Roter D. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, Nov 1-3, 2000. *J Natl Cancer Inst* 2001;93:979-989.
6. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EMJJ, Atkins D, Foekens JA. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671-679.
7. Foekens JA, Atkins D, Zhang Y, Sweep CGJ, Harbeck N, Paradiso A, Cufer T, Sieuwerts AM, Talantov D, Timmermans M, Span PN, Tjan-Heijnen VCG, Mangia A, Zito AF, Specht K, Hoefler H, Meijer-van Gelder ME, Golouh R, Schittulli F, Schmitt M, Beex LVAM, Klijn JGM, Wang Y. Multicenter validation of a gene expression based prognostic signature in lymph-node-negative primary breast cancer. *J Clin Oncol* 2006;24:1665-1671.
8. Van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-536.
9. Van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bertelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999-2009.
10. Wang Y, Klijn JGM, Zhang Y, Atkins D, Foekens JA. A 76-gene signature to predict distant metastasis in lymph-node negative breast cancer. *Am J Oncol Rev* 2005;4:448-454.

11. Clézardin P, Ebetino FH, Fournier GJ. Bisphosphonates and cancer-induced bone disease: beyond their antiresorptive activity. *Cancer Res* 2005;65:4971-4974.
12. Smid M, Wang Y, Klijn JGM, Sieuwerts AM, Zhang Y, Atkins D, Martens JWM, Foekens JA. Genes associated with breast cancer metastasising to bone. *J Clin Oncol* 2006;24:2261-2267.
13. Rio MC, Bellocq JP, Gairard B, Rasmussen UB, Krust A, Koehl C, Calderoli H, Schiff V, Renaud R, Chambon P. Specific expression of the pS2 gene in subclasses of breast cancers in comparison with expression of the estrogen and progesterone receptors and the oncogene ERBB2. *Proc Natl Acad Sci U S A* 1987;84:9243-9247.
14. Foekens JA, Rio M-C, Seguin P, van Putten W, Fauque J, Nap M, Klijn JGM, Chambon P. Prediction of relapse and survival in breast cancer patients by pS2 protein status. *Cancer Res* 50: 3832-3837 (1990).
15. Foekens JA, van Putten WLJ, Portengen, H, de Koning YWCM, Thirion B, Alexieva-Figusch J, Klijn JGM. Prognostic value of PS2 and cathepsin D in 710 human primary breast tumors: multivariate analysis. *J Clin Oncol* 1993;11: 899-908.
16. Foekens JA, Portengen H, Look MP, van Putten WLJ, Thirion B, Bontenbal M, Klijn JGM. Relationship of PS2 with response to tamoxifen therapy in patients with recurrent breast cancer. *Br J Cancer* 1994;70:1217-1223.
17. Jansen MPH, Foekens JA, van Staveren IL, Dirkszwager-Kiel MM, Ritsier K, Look MP, Meijer-van Gelder ME, Portengen H, Dorsers LCJ, Klijn JGM, Berns EMJJ. Molecular classification of tamoxifen resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 2005;23:732-640.
18. Martens JWM, Smid M, Sieuwerts AM, Klijn JGM, Foekens JA. Identification of a gene signature that predicts failure of conventional first line chemotherapy in estrogen receptor-negative recurrent breast cancer patients. San Antonio Breast Cancer Conference, Abstract (December 2006).
19. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, baehner EL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-2826.

INTERPRETATIE VAN TUMORMARKERS DOOR DE CLINICUS

Th. Wobbes, chirurg

De clinicus laat zich bij de behandeling van de patiënt met kanker leiden door een reeks factoren, idealiter op een zodanige wijze dat, gegeven de uitgangssituatie, de patiënt er uiteindelijk het meest bij gebaat is. In de praktijk komt dit neer op een juiste stadiëring van de ziekte en een behandeling die bij het stadium van die ziekte past. Vaak wordt de ziekte van de patiënt letterlijk in beeld gebracht met allerlei diagnostiek. Het bloedonderzoek speelt in de stadiëringfase slechts bij een beperkt aantal tumoren een rol, vooral in de fase dat we nog met een primaire tumor te maken hebben zonder metastasen op afstand. En met het bloedonderzoek wordt in dit geval het onderzoek in ruime zin bedoeld. Een tumor is een lichaamseigen groeisel, dat zich meestal pas in het bloed manifesteert indien er organen worden aangetast die specifieke pathologische parameters geven, zoals bijvoorbeeld in geval van metastasering in de lever. Toch is er een aantal tumoren die de eigenschap hebben zogenaamde merkstoffen in het serum tot expressie te brengen om van hun aanwezigheid blijk te geven, ook al in een relatief vroeg stadium van de ziekte. De vraag is wat de clinicus met deze tumormerkstoffen in de dagelijkse praktijk kan. In tabel 1 is overzichtelijk weergegeven wat idealiter de betekenis zou kunnen zijn.

De klinische toepassing op dit moment van in het serum aanwezige tumormerkstoffen is sterk afhankelijk van het type tumor. Voor het vaststellen van de kans op de ontwikkeling van ziekte hebben serumtumormerkstoffen geen betekenis. Als screeningsparameter zou het prostaatspecifiek antigeen (psa) van betekenis kunnen zijn. Het psa wordt al bijna 20 jaar bepaald maar heeft door de beperkte specificiteit ook negatieve kanten, die vaak uit het oog worden verloren. De merkstof is populair, in die zin dat het op een bepaalde leeftijd naast de cholesterolspiegel en het eeg in het standaardpakket van screening lijkt thuis te horen.

Mannen van middelbare leeftijd wordt geadviseerd hun serum-psa-concentratie te kennen, zodat bij enige stijging van de serumspiegel direct diagnostiek en eventuele behandeling van het prostaatcarcinoom kan worden ingesteld. Maar een verhoogde serum-psa-concentratie is niet bewijzend voor prostaatkanker en niet alle prostaatkarcinomen geven verhoogde psa-concentraties, waarmee meteen het probleem van de tumormerkstoffen voor het detecteren van ziekte wordt aangegeven. (1) Onterechte bezorgdheid of zorgeloosheid,

angst en overbehandeling zijn begeleidende negatieve kanten aan het weten van de persoonlijke serum-psa-concentratie en daarmee ook aan screening op prostaatkanker. Een belangrijk probleem is dat de psa-waarde niet discrimineert voor relatief goedaardige en agressievere vormen van prostaatcarcinoom. Overigens zijn dit niet alleen de problemen van de tumormerkstoffen. Het gebrek aan sensitiviteit en specificiteit en dan met name als een prognostische factor geldt in meer of mindere mate voor alle vormen van diagnostiek.

Tabel 1. Klinische toepassingen van tumormerkstoffen. (4)

Vaststellen risico om ziekte te ontwikkelen

Screening

Stellen van diagnose

- Onderscheiden benigne/maligne
- Vaststellen type tumor

Vaststellen prognose

- Primaire ziekte: vaststellen kans op recidief
- Uitgezaaide ziekte: vaststellen progressie
- Voorspellen van prognose

Voorspellen respons op behandeling:

- Hormonale maatregel
- Chemotherapie
- Andere (“targeted”) therapieën

Monitoren ziekte

- Primaire ziekte: recidief voorspellen
 - Uitgezaaide ziekte: volgen aantoonbare ziekte
-

Voor het stellen van de diagnose hebben serumtumormerkstoffen slechts een beperkte aanvullende betekenis. De belangrijkste betekenis is gelegen in de ondersteuning van de diagnose voor zover deze niet histologische of cytologisch gesteld kan worden.

Bij een (radiologische) zwelling in de pancreasregio kan een sterk verhoogd CA19.9 een aanwijzing zijn dat er sprake is van een pancreascarcinoom. Maar men dient zich te realiseren dat ook de radiologische verschijnselen niet bewijzend zijn voor een carcinoom en dat de specificiteit ervan even beperkt als die van de tumormerkstof. Toch kunnen de combinatie van klinische en radiologische verschijnselen met een sterk verhoogde CA19.9 waarde de diagnose pancreascarcinoom waarschijnlijk maken. In dit opzicht kunnen de gevonden serumafwijkingen bijdragen tot het stellen van de uiteindelijke diagnose, waarbij de histologie per saldo de enige waarheid vormt, maar niet zelden moeizaam te verkrijgen. Er bestaat al een jarenlange discussie over de betekenis van tumormerkstoffen in de follow-up na een primaire behandeling van een tumor (monitoring). De eerste vraag die gesteld moet worden of follow-up van belang is, in die zin dat vroege detectie van recidief ziekte leidt tot betere overleving. Het moet gezegd worden dat dit zeker niet geldt voor de patiënt met een mammacarcinoom. In dit opzicht heeft het gebruik van CA15.3 als tumormerkstof dan ook geen betekenis. (2) Anders is het voor patiënten met een colon- of rectumcarcinoom, die wel degelijk voordeel kunnen hebben van intensieve follow-up. De absolute overlevingswinst is bijna 10% indien het mogelijk is een recidief chirurgisch te verwijderen. In de follow-up van deze patiëntengroep kan de serum-cea-concentratie een belangrijke rol spelen, als eerste indicator voor recidief ziekte. (3)

De betekenis van een tumormerkstof in de follow-up van een maligniteit is dus afhankelijk van het biologisch gedrag van de tumor. De biologie van het colon- en rectumcarcinoom laat in een aantal gevallen chirurgische interventie toe met als gevolg genezing, die van het mammacarcinoom vrijwel nooit. Dit principe geldt uiteraard voor alle tumoren die worden vervolgd met een tumormerkstof. In dit opzicht zijn ook de mogelijkheden van het CA125 bij patiënten met een ovariumcarcinoom beperkt. Het ontbreken van adequate behandeling, indien zich een recidief openbaart, beperkt de toepassing van tumormerkstoffen in de follow-up. Ik laat daarbij de psychologische effecten bij de patiënt van het weten van een afwijkende serumspiegel, zonder de mogelijkheid van een effectieve behandeling, buiten beschouwing.

De vraag is wat de clinicus op dit moment als de meest ideale eigenschap van een tumormerkstof zou willen zien. Deze eigenschap is gelegen in de mogelijkheid om, los van allerlei pathologische variabelen, een duidelijke aanwijzing te geven voor de prognose. Voor lokale resectie is in principe deze mogelijke eigenschap niet zo relevant. Juist voor de aanvullende behandeling zou een aanwijzing voor een slechtere prognose van belang kunnen zijn, omdat daarmee een subgroep van patiënten zou kunnen worden gedefinieerd

die belang zou kunnen hebben voor een aanvullende behandeling met bijvoorbeeld chemotherapie. Dit speelt bijvoorbeeld sterk bij patiënten met een colon- of rectumcarcinoom. Het is duidelijk dat het cea bij deze tumor in deze hele problematiek geen rol speelt. Er moet gezocht worden naar andere parameters, in het serum of mogelijk zelfs in het tumorweefsel aanwezig, die een verschuiving teweeg zouden kunnen brengen in de indicatiestelling voor aanvullende behandelingen, ten einde de prognose te kunnen verbeteren.

Literatuur

1. Schenk-Braat EAM, Bangma CH. De zoektocht naar betere markers voor prostaat­kanker dan prostaatspecifiek antigeen. Ned Tijdschr Geneeskd 2006; 150: 1286-1290.
2. Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value? Clin Chem 2006 52: 345-351.
3. Renehan AG, O'Dwyer ST, Whynes DK. Cost effectiveness analysis of intensive versus conventional follow up after curative resection for colorectal cancer. Brit Med J 2004; 328: 81-85.
4. Stearns V, Yamauchi H, Hayes DF. Circulating tumor markers in breast cancer: accepted utilities

THYROGLOBULINE ALS TUMORMARKER: WAT IS DETECTEERBAAR?

H.A. Ross, analytisch chemicus

Bij de follow-up van patiënten die behandeld zijn voor gedifferentieerd schildklier-carcinoom (DTC) is de meting van serum Thyreoglobuline (Tg) niet meer weg te denken. Gebleken is, dat de sensitiviteit van de Tg meting voor het detecteren van recidieven en metastasen van DTC, met name tijdens hoog TSH, superieur is aan totale lichaams-scintigrafie met een speurdosis ^{131}I . De behandeling van DTC is gericht op volledige ablatie van al het schild-klierweefsel, omdat bij kleine resten geen onderscheid gemaakt kan worden tussen maligne en normaal schildklierweefsel. De zekerheid dat bij een negatieve uitslag geen schildklier-weefsel meer aanwezig is, neemt toe naarmate de analytische gevoeligheid van de bepaling van Tg beter is. Hier spelen fabrikanten van Tg-kits op in door te streven naar een zo hoog mogelijke analytische gevoeligheid. Vanwege het grote belang hiervan is het zinvol om bij de karakterisering en kwantificering van deze gevoeligheid uitvoeriger stil te staan. Naar analogie van het gebruik bij 1e, 2e en 3e generatie TSH metingen wordt het begrip “functionele sensitiviteit” gehanteerd, een term die door Spencer werd gelanceerd, voor de concentratie die met een tussen-run variatie van 20% gemeten kan worden. Het is gebruikelijk om geen waarden beneden deze concentratie (Functionele Detectiegrens, FDG) te rapporteren. Een Tg meetresultaat boven deze grens wijst dan op een recidief of dreigend recidief en een waarde daaronder wordt als “voorlopig ok” beschouwd. Bij deze benadering vallen kritische kanttekeningen te plaatsen, zowel van uit het oogpunt van de onderliggende gedachte als vanuit de beperkingen die de praktijk van immunoassays kenmerken. Als vooralsnog aangenomen wordt, dat we beschikken over een Tg immunoassay (het immunometrische type, met twee antilichamen) waarin geen positief of negatief blanco-effect (bias) optreedt, dan geven monsters waarin geen Tg aanwezig is een meetsignaal dat gelijk is aan dat van de nul-kalibrator, dwz. gemiddeld, want iedere meting is behept met een toevallige fout. Deze toevallige fout, gekarakteriseerd door een SD-waarde, bepaalt hoe groot de kans is dat een nul-monster een signaal boven of beneden een bepaalde waarde geeft. Gemiddeld zal van een aantal metingen aan nul-monsters de helft een hoger signaal geven dan de nul-kalibrator, en de andere helft een lager signaal, afgelezen derhalve als virtuele negatieve concentratie. Ongeveer 2,5% van de metingen zal boven de 2 SD grens uitkomen, en 0,14% boven een 3 SD grens. Wordt een waarde van $>3\text{SD}$ gevonden, dan is de kans, dat

deze meting afkomstig is van een monster zonder Tg, te verwaarlozen. Er is in dit geval dus Tg gedetecteerd. Afhankelijk van hoe groot men de kans op een vals positief resultaat wil hebben kan gekozen worden voor 1,6 SD (5%), 2 SD (2,5%) of 3 SD (0,14%). Deze grens is de analytische detectiegrens. De variatiecoëfficiënt (CV) van een meting op het niveau van 3 SD is per definitie 33%, aangenomen dat in het lage meetgebied de SD constant is. De concentratie waarbij de CV 20% is, de functionele detectiegrens dus, ligt derhalve bij 5 SD. De kans dat een nulmonster boven deze grens komt is natuurlijk helemaal nihil. Daarentegen zal een monster waarin de Tg concentratie gelijk is aan 5 SD met een waarschijnlijkheid van 50% een meetresultaat geven dat lager ligt dan deze grens. Een monster moet een werkelijk Tg gehalte van 6,3 SD hebben, wil het met een waarschijnlijkheid van 90% een meetresultaat boven de FDG opleveren (“limit of quantification”). Door uitkomsten beneden de FDG niet te rapporteren hebben monsters met een werkelijk Tg gehalte van 3,7 SD slechts een kans van 10% om te worden gerapporteerd. Op deze wijze wordt dus een hoge diagnostische specificiteit zeer ten koste van diagnostische gevoeligheid gecreëerd. Bij het leggen van beslisgrenzen dienen altijd de consequenties van vals positieve en vals negatieve resultaten te worden afgewogen. In dit geval lijkt het aan de aandacht te zijn ontsnapt. De rapporterings-grens kan een heel eind omlaag gebracht worden zonder noemenswaardige kans op vals positieven. Bij een grens van 3 SD is deze kans nog maar 0,14%, terwijl een werkelijk gehalte van 4,3 SD met een waarschijnlijkheid van 90% boven deze grens uitkomt. De conclusie uit deze redenering is, dat geheel onnodig de diagnostische gevoeligheid met een factor 1,5 verslechterd wordt door de FDG als laagste rapporteringswaarde te hanteren.

In werkelijkheid bestaat echter wel degelijk kans op vals positieve resultaten. De oorzaak hiervan ligt in de onvolkomenheden van de immunoassays zelf. Hoewel het mogelijk is zeer lage SD's te realiseren, zijn er factoren in het spel die de mogelijkheden die dit zou bieden, geheel teniet doen. In volledige afwezigheid van antigeen produceert een immunometrische bepaling altijd nog een blanco signaal, dat veroorzaakt wordt door niet verwijderde resten gelabeld antilichaam, of instrumentele oorzaken. De monster- en kalibratormatrix kan hierop van invloed zijn, zodat positieve of negatieve blanco effecten optreden.

We vergeleken 4 verschillende Tg bepalingmethoden (CIS-IRMA, BRAHMS Tg-Plus IRMA, DPC Tg op Immulite 2000 “LIEMA” en Nichols Tg op Advantage “ILMA”) bij 53 voor DTC behandelde patiënten waarbij de Tg waarden beneden de eerder met onze routine bepaling (CIS-ELSA) vastgestelde klinische beslisgrens van 3 pmol/l (2 ng/ml)

lagen. Tg werd gemeten vóór en na toediening van r-TSH (Thyrogen). Sera van 24 patiënten waarvan het serum Tg niet steeg na r-TSH en beneden de FDG van 5 verschillende Tg-bepalingsmethoden bleef, werden als “Tg-negatief” bestempeld. Met één van de immunoassays (ILMA) werd gemiddeld $0,31 \pm 0,074$ pmol/l gevonden en in een andere (LIEMA) $-0,10 \pm 0,077$ pmol/l. Beide gemiddelden waren significant verschillend van 0. Met de ILMA werd ook geen enkele maal een signaal kleiner dan 0 gevonden, iets wat gezien het eerdere betoog noodzakelijk is, als er daadwerkelijk geen Tg aanwezig is. Het ligt voor de hand aan te nemen dat deze bepaling een positieve bias van 0,31 pmol/l heeft. Als de sera van deze patiënten geen Tg bevatten, is de analytische variatie (SD) gelijk aan de biologische variatie. Deze bedraagt dus 0,074 pmol/l. De FDG zou dus volgens de definitie $5 \times 0,074 = 0,37$ pmol/l zijn. Volgens de fabrikant gold een FDG van 0,45 pmol/l, dus niet daarmee in tegenspraak, zeker niet wanneer bedacht wordt dat de tussen-run variatie altijd wat hoger is dan binnen-run. Deze FDG ligt echter aanmerkelijk dichter dan 5 SD bij het gemiddelde van Tg-negatieve monsters, en wel 1,9 SD. De kans dat een meting aan een Tg-negatief monster een waarde $> \text{FDG}$ geeft is dus niet meer helemaal te verwaarlozen, maar toch ongeveer 2,5%. Het gemiddelde + 5SD ligt bij 0,68 pmol/l. Deze waarde moet dus eigenlijk als de echte FDG beschouwd worden, hoewel de CV op dat niveau nog maar 15% is. Naast de 24 patiënten waarbij geen van de methoden een waarde $> \text{FDG}$ gaven, waren er 9 patiënten die met de andere vier bepalingmethoden met en zonder r-TSH beneden de corresponderende FDG bleven en met de ILMA tussen 0,45 en 0,68, en daarom dus eigenlijk tot de groep Tg-negatieven behoren. De aldus uitgebreide groep van 33 Tg-negatieven geeft gemiddeld $0,35 \pm 0,102$ pmol/l, zodat de FDG van 0,45 op gemiddelde + 1SD ligt. Hiervoor is de overschrijdingskans 16%, hetgeen gelijk is aan de verwachte frequentie van vals positieven. Bij 9 van de 33 patiënten waren één of beide waarden $> 0,45$, in totaal 13 metingen van de 66, dus 20%.

Uit het voorgaande blijkt dat een FDG die gebaseerd is op het precisieprofiel in het geval van positieve bias tot een aanzienlijke kans op vals positieve resultaten kan leiden. Als deze FDG vervolgens wordt toegepast om Tg-negatieve monsters te identificeren en daar een betrouwbaarheidsinterval voor te berekenen, wordt weer een fout gemaakt. Uit het precisieprofiel kan niet worden afgeleid of er sprake is van bias, dat kan alleen op basis van de ligging van de ruwe meetsignalen t.o.v. de 0-kalibrator. Verhoging van de analytische sensitiviteit is zinloos als bias niet kan worden uitgesloten. Indien bias kan

worden uitgesloten, leidt het hanteren van de FDG volgens de gangbare definitie tot een te lage diagnostische gevoeligheid.

Referenties

1. Mazzaferri EL, Robbins RJ, Spencer CA, Braverman LE, Pacini F, Wartofsky L et al. A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. J Clin Endocrinol Metab 2003;88:1433-41
2. Spencer CA, Takeuchi M and Kazarosyan M. Current status and performance goals for serum thyrotropin (TSH) assays. Clin Chem 1996 ;42 :141-145
3. Demers LM, Spencer CA 7-13-2002 Laboratory support for the diagnosis and monitoring of thyroid disease <http://www.nacb.org/imp/imp/thyroidLMPG.stm>
4. Currie LA et al. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC recommendations 1995) Pure & Appl Chem 1995;67:1699-1723

Nawoord

De ILMA is om redenen die geheel buiten deze discussie liggen, uit de handel genomen. Dit wil geenszins zeggen dat het geschetste probleem uit de wereld is. Bias kan bij nog te verschijnen methoden optreden, en de ervaring leert dat ook bij bestaande methoden met wisseling van reagens lot de bias ineens kan veranderen. Het is geboden om dit dmv. de interne QC kritisch te volgen.

APPLICATION OF SELDI-TOF-MS IN PROTEIN PROFILING: STATE OF THE ART

M. P. van Dieijen-Visser, J.A.P. Bons, D. de Boer, and K.W.H. Wodzig, klinisch chemici

Abstract

Serum protein profiling by Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (SELDI-TOF-MS) appears to be an important diagnostic tool for a whole range of diseases. Sensitivities and specificities obtained with this new technology often seem superior to those obtained with current biomarkers. However, reproducibility and standardization are still problematic. This presentation explains the SELDI-TOF-MS technique and some important aspects for proteomics studies are discussed, like pre- and post-analytical aspects and quality control procedures.

Introduction

The field of proteomics has developed rapidly in recent years. Until the mid-1990s mainly individual genes and proteins were studied. The point of proteomics is to characterize the behaviour of the system as a whole, rather than the behaviour of any single component. Proteomics is in fact the comprehensive study of the proteome: all the proteins in either a cell, tissue, organ, or organism. The proteome is dynamic and in constant flux due to a combination of factors. These factors include posttranslational modifications and functional regulation of gene expression [1]. Moreover, in proteomics protein identification is not necessarily performed by complete sequence analysis, but can also be performed by partial sequence analysis with the aid of database matching tools.

To further improve and coordinate proteomics research, the international human proteome organization HUPO (www.hupo.org) has been founded in 2001, aiming to define and promote proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques and training to better understand human disease.

Proteomic analysis requires the combination of various technologies, including biochemistry, mass spectrometry and bioinformatics. Important techniques for expression analysis of proteins are two-dimensional electrophoresis (2-DE) combined with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF-MS), and/or Liquid Chromatography Tandem Mass Spectrometry (LC/MS/MS). This paper is focused on a novel approach the SELDI-TOF-MS technique. Important issues in

SELDI-TOF-MS analyses are protein profiling and detection of new biomarkers for different diseases. Although promising, the fact is that until now no widely applicable new approaches to patient diagnosis and therapy have become available. Research is still focused on improvement of reproducibility and sensitivity of the different techniques.

SELDI-TOF-MS

There have been many reports on the application of SELDI-TOF-MS technology since its first introduction in 1993 by Hutchens and Yip [2]. SELDI-TOF-MS is an approach that tries to overcome the requirements for purification and separation of proteins prior to mass spectrometry analysis [3]. It is a novel approach to biomarker discovery that combines two powerful techniques: chromatography and mass spectrometry. One of the key features of SELDI-TOF-MS is its ability to provide a rapid protein expression profile from a variety of biological and clinical samples [4]. It consists of selective protein extraction and retention on chromatographic chip surfaces and their subsequent analysis by a simple laser desorption ionization mass spectrometer [5]. It differs in several aspects from conventional MALDI-TOF-MS. For MALDI-TOF-MS, analytes are directly spotted onto a plate. This is usually a metal plate. The applied samples are usually tryptic digests from proteins separated by 2-DE, although proteins purified by other separation methods are also compatible with the method. Before deposition of the analytes, the energy absorbing matrix (EAM) is placed on the plate or mixed in with the sample. The matrix will absorb energy from the laser causing the analytes to be ionized by MALDI-TOF-MS [6].

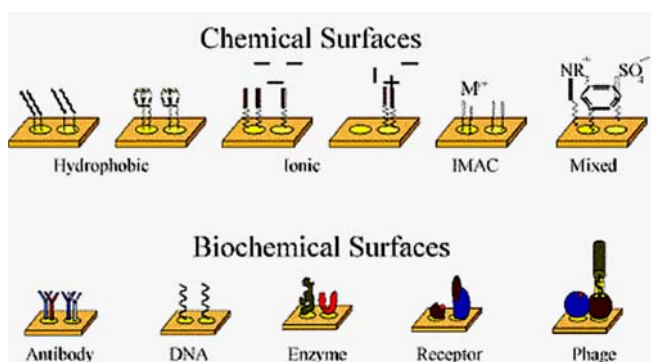


Figure 1.
The different types of ProteinChip arrays.
The chemical surfaces are chromatographic ProteinChip arrays with hydrophobic, cationic, anionic, metal ions for immobilized metal affinity binding (IMAC) or hydrophilic spots. The biochemical surfaces are designed for coupling of biomolecules in antibody-antigen assays, DNA-protein binding experiment, coupling of enzymes, receptor-ligand interaction and for coupling of phages.

ProteinChip arrays

For the SELDI-TOF-MS technique different ProteinChip arrays (CIPHERGEN Biosystems Inc.) can be used. The chromatographic surfaces that make up the various ProteinChip arrays are uniquely designed to retain proteins from a complex sample mixture according to specific properties such as hydrophobicity, charge [4] (Figure 1). The procedure for detecting protein biomarkers is very simple. A few microliters of the sample are dispensed onto the ProteinChip surface under specific binding conditions that determine which proteins will be retained by the surface. Protein specificity is achieved through the application of a series of washes with an appropriate solvent or buffer designed to elute unbound proteins and interfering substances, such as salts, detergents, lipids. Only proteins actively interacting with the spot surfaces are analyzed in the Protein Biosystem series instrument (CIPHERGEN Biosystems Inc.), because all other components are washed off in advance [7].

By choosing different ProteinChip arrays with array-specific surface components, different proteins will be analyzed depending on the chip characteristics. In fact the interaction of the analyte and the chip introduces a purification step. After addition of sample and washing buffers, the EAM is applied to the ProteinChip array. The EAM will facilitate desorption and ionization in the PBS series instrument.

Desorption/Ionization process

After introducing the ProteinChip array into the ProteinChip Reader, a laser beam is directed onto the sample on the spot. Upon laser activation, the sample becomes irradiated and the desorption and ionization proceeds to liberate gaseous ions from the ProteinChip arrays. These gaseous ions enter the TOF-MS region of the instrument, which measures the mass-to-charge ratio (m/z) of molecular ions of each protein, based on its velocity through a vacuum chamber [4]. The time-of-flight corresponds inversely to the m/z value. As a first result, the molecules in the sample are represented in a graph with the m/z value on the x-axis and the corresponding signal intensity on the y-axis [7]. (Figure 2).

Advantages SELDI-TOF-MS

SELDI-TOF-MS has several advantages over other methods such as 2-DE combined with MALDI-TOF-MS and/or LC/MS/MS. SELDI-TOF-MS has a much higher throughput

capability, requires significantly lower amounts of the sample, has small range sensitivity, offers higher resolution at low mass ranges, and is easy to use [8]. SELDI-TOF-MS can

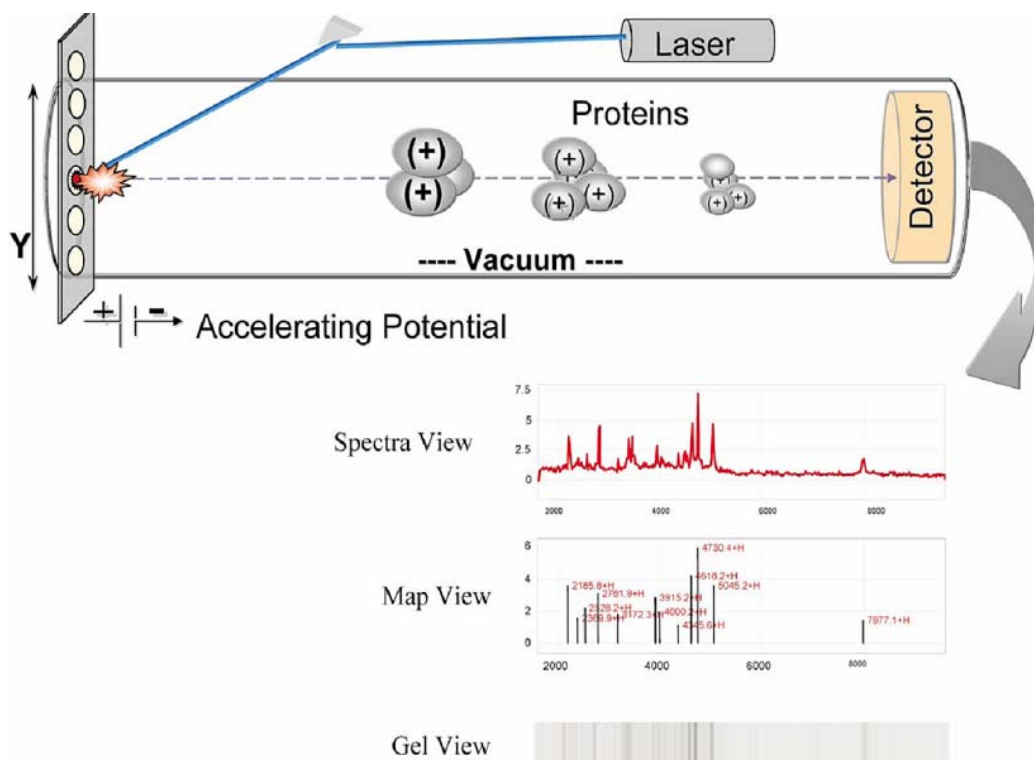


Figure 2. The ions of the molecules in the sample are represented in a spectra, map and gel view with the mass-to-charge ratio (m/z) on the x-axis and the corresponding signal intensity on the y-axis.

effectively resolve polypeptides and peptides smaller than 20 kDa [9]. The 2-DE approach, where proteins are at first separated by their isoelectric point and subsequently by their molecular weight, was developed to increase the resolving power for the analysis of complex protein mixtures. Whereas the enhanced resolution of 2-DE gels contributed greatly to our understanding of the wide variety of proteins in a given sample, it still includes the disadvantage of the sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) method of giving preference to the most abundant proteins. In addition, proteins in the peptide range as well as those of high hydrophobicity or of extreme isoelectric points are typically neglected, resulting again in a loss of potentially interesting proteins [10]. Moreover, 2-DE is labour intensive, time consuming, and difficult to standardize between laboratories [4].

The high throughput ability of the SELDI-TOF-MS system allows hundreds of samples to be screened for disease biomarker identification in a relatively short time period, providing investigators the opportunity to compare patient-to-patient variability [4]. SELDI-TOF-MS is a recently established improvement on some of the concepts of

MALDI-TOF-MS. ProteinChip arrays allow researchers to purify and detect a subset of proteins in the sample at the same time by using a variety of surface chemistries such as classic chromatographic surfaces (e.g., cation/anion exchanges) and biologically activated surfaces to capture specific molecular counterparts. This benefit is effective to especially the biological samples such as body fluids and conditioned medium containing a variety of proteins [11].

Biomarker Discovery

The true scientific goal of serum proteomic pattern analysis is in fact biomarker discovery. Since the study by Petricoin et al. [12] on proteomic patterns to detect ovarian cancer, the use of SELDI-TOF-MS protein profiling as a diagnostic tool, has become an important subject of investigation [13]. Until now, this approach has been suggested for different diseases, like ovarian [12, 14-18], prostate [9, 19-22] and lung [23] cancer, but also for inflammatory diseases [24, 25]. Although the first papers on SELDI-TOF-MS seemed very promising, until now it did not yet result in widely applicable new approaches or diagnostic tools. It is with no doubt a promising technique, but further improvement of both reproducibility and sensitivity is required, before the process of translating new markers into clinical laboratory tests can be further developed [26]. The vast majority of the currently available data on biomarker detection in cancer, have been produced with SELDI-TOF-MS. As was recently critically discussed by Diamandis [27], two types of data have been reported in the literature, 1) discriminating peaks of unknown identity (increased or decreased) between normal individuals and patients with cancer; and 2) data in which at least some of the peaks have been positively identified. Table 1 shows differences found in prostate and ovarium carcinoma, even in studies using comparable experimental conditions. Some important pre-analytical and analytical aspects will be discussed here.

Pre-analytical aspects

Protein profiling can only become a reliable diagnostic tool when it fulfils the criteria for reproducibility and standardization that are generally accepted for diagnostic tests in clinical chemistry. Therefore, some essential aspects to improve reproducibility and standardization of SELFI-TOF-MS will be discussed here.

Storage effects

To avoid pre-analytical errors, sample collection for proteomic analysis should be accurately described and standardized. Effects of sample storage and the consequences of differences in sample preparation are highly underestimated. We recently compared protein profiles of

Table 1. Summary of previous reports on prostate and ovarian cancer with a proteomics approach using SELDI-TOF-MS.

Prostate cancer	Chip	m/z values (Da)
Author		
Adam [20]	IMAC3-Cu	4475, 5074, 5382, 7024, 7820, 8141, 9179, 9507, 9656
Banez [19]	IMAC3-Cu	IMAC: 3960, 4469, 9713, 10266, 22832
	WCX2	WCX2 : 3972, 8226, 13952, 16087, 25167, 33270
Qu [9]	IMAC3-Cu	PCA vs non/cancer: 9655, 9720, 6542, 6797, 6949, 7024, 8067, 8356, 3963, 4079, 7885, 6991 BPH vs HC: 7820, 4580, 7844, 4071, 7054, 5298, 3486, 6099, 8943
Li [22]	IMAC3-Ni	2680, 10300, 17900
Petricoin [21]	C16 hydro-phobic interaction	2029, 2367, 2582, 3080, 4819, 5439, 18220
Ovarian cancer	Chip	m/z values (Da)
Author		
Petricoin [12]	C16 hydro-phobic interaction	543, 989, 2111, 2251, 2465
Zhang [15]	IMAC3-Cu	3272 (pH 9 fraction) 12828, 28043 (pH 4 fraction)
Rai [18]	IMAC3-Ni	8600, 9200, 19800, 39800, 54000, 60.000, 79.000
Kozak [14]	SAX2	3100, 13900, 21000, 79000, 106700
Ye [17]	IMAC3-Cu	11723
Vlahou [16]	SAX2	SAX2 4400, 21500
	IMAC	IMAC 5540, 6650, 11700

freshly frozen serum samples with frequently thawed serum samples. The number of freeze-thaw cycles should be perfectly standardized and comparable for control and disease population. It is to be expected that especially in the earlier protein profiling studies archived samples were used for which conditions of control and patient populations were not fully identical. It has now become apparent that both the number of freeze-thaw cycles, freezing temperature and storage time should at least be identical for both study and control population. This can easily be overcome in prospective studies by dividing the samples in aliquots before storage [28].

Serum or plasma

Until now, insufficient information is available to decide whether serum or plasma should be preferred in proteomic studies. Most studies have used serum, but further research on this topic is required [28]. It is generally assumed that, more peaks can lead to more significant differences between populations as was the case in our study. Theoretically, however, plasma with protease inhibitors contains more intact proteins not attacked by proteolytic enzymes. Further examinations on the differences between serum and plasma are required.

Sample preparation

Samples can be denatured with urea/CHAPS [9, 14, 16-20, 22], but can also be fractionated with anion exchange chromatography [15]. Denaturing conditions allow protein-protein interaction disruption before analysis by SELDI-TOF-MS. With fractionation by anion exchange chromatography the highly abundant proteins such as albumin and immunoglobulins (60-80% of total serum protein content), which can interfere with the resolution and sensitivity of the proteome profiling techniques, will be visible in specific fractions. Linke et al. [29] illustrated that fractionation greatly increases the number of peptide and protein ion signals that can be observed by SELDI-TOF-MS, when compared to both unfractionated (only denatured) as well as albumin-depleted samples. By using different denaturing steps or using fractionated samples, other significant peaks resulting in different biomarkers can be detected.

Sampling time

We suggest, according to World Health Organization (WHO), anticoagulants in diagnostic laboratory investigations to use a clotting time of 30 minutes at room temperature, spinning for 15 minutes at a minimum speed of 1500g at 4°C and storage of the samples in aliquots within 1 hour at -80 °C after blood collection. Obviously the consequences of differences in sample characteristics within a study population, but also between study and control population, like for instance use of fasting or non-fasting samples, age-matching of the samples should always be properly standardized.

Patient population

The number of patients and healthy controls in the training and validation sets is very important because the reliability of the results improves with increasing numbers. A clear description of the training and validation population is essential, like the severity of disease. Because SELDI-TOF-MS fingerprinting probably measures peptides present in high abundance in serum (e.g. mg/L to g/L range) the molecules, which are detected, probably originate from common disease mechanisms or general protection mechanisms, i.e. epiphenomena of the diseases, such as acute phase response, cachexia etc. It is clear that the robustness of the technology should be validated by comparing patient groups with comparable disease mechanisms. Method validation should therefore be extended not only to healthy controls, but also to diseases with comparable generalized disease conditions (infection, cachexia etc).

Post-analytical aspects

Bioinformatics and biostatistics

Peak detection, laser settings and data analysis software affect the ultimate m/z values found. Different multivariate analysis software can be used to classify different groups. Biomarkers Patterns (Ciphergen Biosystems Inc.) is a decision tree algorithm which is very often used in protein profiling studies. The decision trees can be based on the intensity, S/N ratio or Area Under Curve (AUC). Propeak, Classification and Regression Tree (CART), AdaBoost, and principal component analysis (PCA) are other examples of multivariate analysis software programs which can be used to classify the different groups. Some groups develop their own statistical software program by combining more

multivariate analysis techniques. It is hard to compare the results of studies when all these different kind of software programs are used to classify groups.

In a recent review we showed that apart from the pre-analytical strategy, the post-analytical strategy also has an enormous impact on the final results. By comparing previous reports on prostate and ovarian cancer we showed large differences in m/z values of the biomarkers presented in the different studies, even in studies with comparable patient populations [28]. It should be noted that careful and precise selection of the peak labelling settings and normalization of peak intensities are considered critical for biomarker identification and for the efficient and reliable performance of any learning algorithm used in conjunction with the SELDI-TOF-MS system [16].

Quality control

As mentioned before, the effect of pre- and post-analytical variables on protein profiling needs further and more systematic investigation. Therefore, a stringent standardized protocol is needed, not only for pre- and post-analytical aspects, but also for calibration and quality control (QC) performance. The results of the quality control procedure are described in a previous report of Bons et al. [30].

Recently Plebani et al. [31] indicated that only few studies published carefully described their quality control procedures incorporated in proteomic experimental protocols. We developed a well-defined protocol for calibration of the Protein Biosystems IIC (PBS IIC) instrument, to implement QC samples with independent certified standards and to determine acceptance criteria for quality control. Because the QC samples are spotted on a NP20 array, which is a normal phase array, without washing or selective binding steps, only the MALDI-TOF-MS part of the PBS IIC instrument is checked. Stable instrument performance over time is a prerequisite before any proteomic experiments should be performed. The QC procedure described acts prospectively by checking the calibration every week in contrast to some other studies, where QC samples are included in the profiling studies and quality control thus acts retrospectively or where no quality control procedure is performed at all.

Data analysis was performed with in house developed software (ShewhartPlots), which was based on the Shewhart control chart principle [32]. M/z values, intensities, signal-to-noise (S/N) ratios and peak resolutions are imported in this software. Two dimensional Youden plots are made by drawing insulin (x-axis) and apomyoglobin (y-axis) in one plot

for all parameters and three dimensional Youden plots are made by drawing insulin (x-axis), apomyoglobin (y-axis) and albumin (z-axis) in one plot for all parameters. Analysis are only performed if minimal quality requirements are fulfilled.

Reproducibility

We showed that variations in the signal of the QC samples can be caused by pipetting variability in the handling of the QC sample, spot and chip variability, crystallization of the EAM and laser detector variability over time.

The reproducibility of serum protein profiling by SELDI-TOF-MS was investigated by spotting one QC sample, consisting of insulin and apomyoglobin on 2 or 4 NP20 chips. Coefficient of variation (CV) values from approximately 10 to 40% were achieved for intensities and signal-to-noise (S/N) ratios. The pooled CV value for the mass accuracy was below 0.1%. The CV values for intensities, S/N ratios and mass accuracy described in the study of Bons et al. were comparable with the CV values reported by Semmes et al. [33]. Semmes et al. performed across-laboratory measurement of three m/z peaks in a standard pooled serum. This resulted in a 0.1% CV for mass accuracy. The CVs for signal to noise ratio's were 34-40% and the variations in the intensities of the three peaks for all laboratories were 15-36%.

Lee et al. [34] also indicated that it is hard to reproduce experiments. They investigated renal cell carcinoma and included samples from patients with renal cell carcinoma, patients with benign urological diseases and healthy controls in the training set. An initial blind group of samples was used to test the models. Sensitivities and specificities of 81.3-83.3% were achieved. However, subsequent testing 10 months later with a different blind group of samples resulted in much lower sensitivities and specificities (41.0-76.6%).

Potential sources of variability that arise during SELDI-TOF-MS profiling include spot-to-spot variation of chip surfaces, laser detector variability over time, pipetting variability [35] and the crystallization process of the EAM [36, 37]. We demonstrated that the reproducibility of the crystallisation process can be increased by using an incubator with a constant temperature of 28 °C and a constant atmospheric humidity of 45%. The same QC sample (insulin and apomyoglobin) as described above was used and CV values of 4 to maximal 25% were achieved for intensities and S/N ratios.

Conclusions

Any new technology, particularly one being presented as a potential clinically used diagnostic tool, requires stringent quality control to evaluate analytical performance over time. Instrument performance, however, must be compared not only during one experiment, but also over the course of time. We recently defined a standard protocol for calibration and acceptance criteria for the independent certified QC samples were established [30]. Stringent QC as indicated above prevents unreliable data acquisition from the very start.

By introducing standard protocols and strict quality control, the analytical variation of protein profiling experiments can be significantly reduced. However, further optimization of SELDI-TOF-MS is required in order to become a reliable technique for biomarker detection and only if reproducibility of SELDI-TOF-MS protein profiling is significantly improved it can become a valuable diagnostic tool in different diseases.

References

1. Srinivas PR, Srivastava S, Hanash S, Wright GL, Jr. Proteomics in Early Detection of Cancer. *Clin Chem* 2001;47(10):1901-11.
2. Hutchens TW, Yip T. New Desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 1993;7:576-80.
3. Bischoff R, Luider TM. Methodological advances in the discovery of protein and peptide disease markers. *J Chromatogr B Analyt Technol Biomed Life Sci* 2004;803(1):27-40.
4. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS Approach to Proteomics: Protein Profiling and Biomarker Identification. *Biochemical and Biophysical Research Communications* 2002;292(3):587-92.
5. Caputo E, Moharram R, Martin BM. Methods for on-chip protein analysis. *Anal Biochem* 2003;321(1):116-24.
6. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev* 2001;101(2):269-95.
7. Wiesner A. Detection of Tumor Markers with ProteinChip(R) Technology. *Curr Pharm Biotechnol* 2004;5(1):45-67.
8. Shiwa M, Nishimura Y, Wakatabe R, Fukawa A, Arikuni H, Ota H, et al. Rapid discovery and identification of a tissue-specific tumor biomarker from 39 human cancer cell lines using the SELDI ProteinChip platform. *Biochem Biophys Res Commun* 2003;309(1):18-25.
9. Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin Chem* 2002;48(10):1835-43.
10. Gygi SP, Aebersold R. Mass spectrometry and proteomics. *Curr Opin Chem Biol* 2000;4(5):489-94.

11. Furuta M, Shiraishi T, Okamoto H, Mineta T, Tabuchi K, Shiwa M. Identification of pleiotrophin in conditioned medium secreted from neural stem cells by SELDI-TOF and SELDI-tandem mass spectrometry. *Brain Res Dev Brain Res* 2004;152(2):189-97.
12. Petricoin III EF, Mills GB, Kohn EC, Liotta LA. Proteomic patterns in serum and identification of ovarian cancer. *The Lancet* 2002;360(9327):170-1.
13. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359(9306):572-7.
14. Kozak KR, Amneus MW, Pusey SM, Su F, Luong MN, Luong SA, et al. Identification of biomarkers for ovarian cancer using strong anion-exchange ProteinChips: potential use in diagnosis and prognosis. *Proc Natl Acad Sci U S A* 2003;100(21):12343-8.
15. Zhang Z, Bast RC, Jr., Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 2004;64(16):5882-90.
16. Vlahou A, Schorge JO, Gregory BW, Coleman RL. Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data. *J Biomed Biotechnol* 2003;2003(5):308-14.
17. Ye B, Cramer DW, Skates SJ, Gygi SP, Pratomo V, Fu L, et al. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin Cancer Res* 2003;9(8):2904-11.
18. Rai AJ, Zhang Z, Rosenzweig J, Shih Ie M, Pham T, Fung ET, et al. Proteomic approaches to tumor marker discovery. *Arch Pathol Lab Med* 2002;126(12):1518-26.
19. Banez LL, Prasanna P, Sun L, Ali A, Zou Z, Adam BL, et al. Diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol* 2003;170(2 Pt 1):442-6.
20. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62(13):3609-14.
21. Petricoin EF, 3rd, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002;94(20):1576-8.
22. Li J, White N, Zhang Z, Rosenzweig J, Mangold LA, Partin AW, et al. Detection of prostate cancer using serum proteomics pattern in a histologically confirmed population. *J Urol* 2004;171(5):1782-7.
23. Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Discovery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. *Lung Cancer* 2003;40(3):267-79.
24. Poon TC, Hui AY, Chan HL, Ang IL, Chow SM, Wong N, et al. Prediction of Liver Fibrosis and Cirrhosis in Chronic Hepatitis B Infection by Serum Proteomic Fingerprinting: A Pilot Study. *Clin Chem* 2004;51:328-35.
25. Zhu XD, Zhang WH, Li CL, Xu Y, Liang WJ, Tien P. New serum biomarkers for detection of HBV-induced liver cirrhosis using SELDI protein chip technology. *World J Gastroenterol* 2004;10(16):2327-9.
26. Hortin GL, Jortani SA, Ritchie JC, Jr., Valdes R, Jr., Chan DW. Proteomics: A New Diagnostic Frontier. *Clin Chem* 2006.
27. Diamandis EP. Serum Proteomic Profiling by Matrix-Assisted Laser Desorption-Ionization Time-of-Flight Mass Spectrometry for Cancer Diagnosis: Next Steps. *Cancer Res* 2006;66(11):5540-1.

28. Bons JA, Wodzig WK, van Dieijen-Visser MP. Protein profiling as a diagnostic tool in clinical chemistry: a review. *Clin Chem Lab Med* 2005;43(12):1281-90.
29. Linke T, Ross AC, Harrison EH. Profiling of rat plasma by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry, a novel tool for biomarker discovery in nutrition research. *J Chromatogr A* 2004;1043(1):65-71.
30. Bons JA, de Boer D, van Dieijen-Visser MP, Wodzig WK. Standardization of calibration and quality control using surface enhanced laser desorption ionization-time of flight-mass spectrometry. *Clin Chim Acta* 2006;366:249-56.
31. Plebani M. Proteomics: the next revolution in laboratory medicine? *Clin Chim Acta* 2005;357(2):113-22.
32. Westgard JO, Groth T, Aronsson T, de Verdier CH. Combined Shewhart-cusum control chart for improved quality control in clinical chemistry. *Clin Chem* 1977;23(10):1881-7.
33. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, et al. Evaluation of Serum Protein Profiling by Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry for the Detection of Prostate Cancer: I. Assessment of Platform Reproducibility. *Clin Chem* 2005;51(1):102-12.
34. Lee SW, Lee KI, Kim JY. Revealing urologic diseases by proteomic techniques. *J Chromatogr B Analyt Technol Biomed Life Sci* 2005;815(1-2):203-13.
35. Koopmann J, Zhang Z, White N, Rosenzweig J, Fedarko N, Jagannath S, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res* 2004;10(3):860-8.
36. Jock CA, Paulauskis JD, Baker D, Olle E, Bleavins MR, Johnson KJ, et al. Influence of matrix application timing on spectral reproducibility and quality in SELDI-TOF-MS. *Biotechniques* 2004;37(1):30-2, 4.
37. Cordingley HC, Roberts SL, Tooke P, Armitage JR, Lane PW, Wu W, et al. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. *Biotechniques* 2003;34(2):364-5, 8-73.

SERUM MARKERS IN BREAST CANCER: ARE THEY OF VALUE?, WILL THEY GET BETTER?

M.J. Duffy, clinical chemist

Available serum markers for breast cancer include CA 15-3, BR 27.29 (also known as CA27.29), CEA, tissue polypeptide antigen (TPA), tissue polypeptide specific antigen (TPS) and the shed form of HER-2. Of these, the most widely used are CA 15-3 and CEA (for review, see refs 1,2). The aim of this presentation is to discuss the present and likely future use of serum markers in breast cancer.

Screening/Aiding Early Diagnosis

Lack of sensitivity and specificity preclude the use of all existing serum markers for the early detection of breast cancer. Women with apparently localized breast cancer who present with a high preoperative marker level (e.g., 5-10 times the upper limit of normal) are likely to have advanced disease (3) and should undergo appropriate investigations to diagnose or exclude this possibility.

Determining Prognosis

A number of studies have shown that elevated preoperative levels of either CA 15-3 or CEA are associated with poor outcome in patients with breast cancer (1). For example, in our study on 600 newly diagnosed breast cancer patients, the prognostic impact of preoperative CA 15-3 levels was independent of tumor size and lymph node status (4). Importantly, the prognostic value of CA 15-3 was also observed in lymph node-negative patients, the subgroup of breast cancer patients in which new prognostic factors are most urgently needed.

Compared to tissue prognostic factors, serum markers have a number of advantages. Firstly, unlike tumor tissue which requires either biopsy or surgery, blood can be obtained with minimal inconvenience. Secondly, automated, relatively cheap and standardized assays are available for serum markers. Thirdly, serum-based markers can be determined in patients with small tumors including those with in situ cancers. For tissue-based markers, especially if freshly-frozen tissue is necessary, patients with very small tumors cannot be assessed.

Surveillance Following Surgery

Following surgery for breast cancer, it is now common practice to follow-up patients on a regular basis with clinical examination, radiology and tumor marker determinations. This practice is based on the belief that the early detection of recurrent or metastatic disease improves the chance of cure or results in an improved survival. From a biological point of view, it might be expected that the early detection of recurrent disease followed by the initiation of therapy would improve outcome compared with starting therapy when recurrence/metastasis is clinically evident. There is however little evidence available to support this hypothesis.

Serial determinations of markers such as CA 15-3 and CEA have the potential to detect recurrent breast cancer in asymptomatic women with median lead-times of 4-5 months (1,3). Since it is unclear whether knowing this lead-time enhances outcome, guidelines vary in their recommendations regarding the use of tumor markers in postoperative surveillance in breast cancer. For example, the American Society of Oncology, the European Society of Medical Oncology and the European Society of Mastology recommend that serum markers should not be used in the routine surveillance of patients following primary treatment for breast cancer (3,5,6,7). In contrast, the European Group on Tumor Markers (EGTM) and the National Academy of Clinical Biochemistry (NACB) recommend the use of markers during follow-up (8,9).

Monitoring Therapy in Advanced Disease

Following the commencement of therapy for advanced disease, it is important to know as soon as possible if the patient is responding to the treatment. If the patient is benefiting, clearly treatment should be continued. If on the other hand, treatment is not effective, an alternative therapy might be given. If an alternative therapy is unavailable, these patients could be willing to participate in clinical trials or they could decide to avoid further therapy.

A convenient and relatively inexpensive approach for helping to establish response is by measuring serum markers such as CA 15-3 or CEA. Generally, decreasing marker levels correlate with tumor response while increasing markers levels correlate with tumor regression (10-12). According to the EGTM guidelines (8), markers should be measured prior to every chemotherapy course and at least three monthly intervals for patients receiving hormone therapy. The EGTM defines an increase in marker concentration of at

least 25% to be significant (8). It is recommended that such an increase be confirmed with a second specimen obtained within a month. If the increase is confirmed, this provides evidence of progressive disease. Similarly, a confirmed decrease in serum levels of more than 50% was stated to be consistent with tumor regression (8).

In contrast to the EGTM recommendations, the ASCO guidelines state that neither CA 15-3 nor CEA should be routinely used for monitoring therapy in patients with advanced breast cancer (17,23). However, this panel also stated “that in exceptional circumstances such as the presence of osseous metastasis, which are difficult to evaluate clinically, the marker level may be able to support the clinical estimate of disease status. However, the marker cannot in any situation stand alone to define response to treatment” (3,6).

The European Society of Mastology (EUSOMA) also recommend against the general use of serum markers for monitoring therapy in advanced breast cancer (7). However, as with the ASCO guidelines, the EUSOMA guidelines stated that “in the absence of evaluable disease, increase in tumor marker accompanied by an increase in symptoms (e.g., bone pain) should be taken as indicating disease progression”. Also, according to these guidelines, “an increase in serum markers without symptoms of progression should prompt a complete work-up to investigate for progression of known disease sites or appearance of new sites” (7).

Potential New Markers for Breast Cancer

A desirable property of a serum marker is organ-specificity. None of the available serum markers for breast cancer is breast-specific as all can be elevated in serum from patients with most types of adenocarcinoma, especially in patients with advanced disease (1).

In recent years however, a number of proteins have been described that are expressed almost exclusively in breast tissue including breast cancer. These include, mammaglobin A (10-13), lipophilin B (14-15), NY-BR-1 (16,17), B726P (18), and small breast epithelial mucin (SBEM) (19). The challenge now is to devise sensitive and specific assays for measuring these proteins in serum and then evaluate their clinical value in breast cancer.

Will Serum Markers for Breast Cancer Get Better?

As mentioned above, the main problem with all existing serum markers for breast cancer is lack of sensitivity for early disease and lack of specificity for breast cancer. Clearly, new markers must offer improved sensitivity and specificity. One of the most promising approaches in this respect is the use of proteomics. In recent years a number, a number of

preliminary reports claimed to be able to detect breast cancer with sensitivities and specificities of 85-95% (for review, see ref. 20). These findings however, will require extensive validation before they can be used clinically. Finally, if the relatively breast-specific proteins mentioned above could be detected in serum, they also have the potential to provide a new generation of serum markers for breast cancer.

References

1. Duffy MJ. Serum tumor markers in breast cancer: are they of clinical value. *Clin Chem* 2006;52:345-351.
2. Cheung K, Graves CRL, Robertson JFR. Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Rev.* 2000;26:91-102.
3. Anonymous. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. *J Clin Oncol* 1996; 14: 2843-2877.
4. Duffy MJ, Duggan C, Keane R, et al. High preoperative CA 15-3 concentrations predict adverse outcome in node-negative and node-positive breast cancer: study of 600 patients with histologically confirmed breast cancer. *Clin Chem* 2004;50:559-563.
5. Bast RC, Ravdin P, Hayes DF, et al. 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;19:1865-1878.
6. Blamey RW. Guidelines on endocrine therapy of breast cancer, EUSOMA. *Eur J Cancer* 2002;38:615-634.
7. Pestalozzi BC, Luporsi-Gely E, Jost LM, et al ESMO Minimum Clinical Recommendations for diagnosis, adjuvant treatment and follow-up of primary breast cancer. 2005;16 (Suppl 1):7-9
8. Molina R, Barak V, van Dalen A, Duffy MJ, et al. Tumor markers in breast cancer: European Group of Tumor Markers (EGTM) recommendations. *Tumor Biol* 2005;26:281-293.
9. Fleisher M, Dnistrian AM, Sturgeon CM, et al. Practice guidelines and recommendations for use of tumor markers in the clinic. In: Diamindis EP, Fritsche H, Scharwitz MK, Chan DW, eds, *Tumor markers, physiology, pathobiology, technology and clinical applications*, Chicago: AACC Press 2002:33-63.
10. Robertson JFR, Jaeger W, Syzmendera JJ, et al. The objective measurement of remission and progression in metastatic breast cancer by use of serum tumor markers. *Eur J Cancer* 1999;35:47-53.
11. Van Dalen A, Heering KJ, Barak V et al. Treatment response in metastatic breast cancer: a multicenter study comparing UICC criteria and tumor marker changes. *The Breast* 1996;5:82-88.
12. Kurebayashi J, Nishimura R, Tanaka K, et al. Significance of serum tumor markers in monitoring advanced breast cancer patients treated with systemic therapy: a prospective study. *Breast Cancer* 2004;11:389-395.
13. Watson MA, Dintzis S, Darrow CM, et al. Mammaglobin expression in primary, metastatic, and occult breast cancer. *Cancer Res* 1999;59:3028-31.
14. Jiang Y, Harlocker SL, Molesh DA, et al. Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. *Oncogene* 2002;21:2270-82.
15. O'Brien N, Maguire TM, O'Donovan N, et al. Mammaglobin a: a promising marker for breast cancer. *Clin Chem* 2002;48:1362-4.

16. Colpitts TL, Billing-Medel P, Friedman P, et al. Mammaglobin is found in breast tissue as a complex with BU101. *Biochemistry* 2001;40:11048-11059.
17. Houghton RL, Dillon DC, Molesh DA, et al. Transcriptional complementarity in breast cancer: application to detection of circulating tumor cells. *Mol Diagn* 2001;6:79-91.
18. Miksicek, R J, Y. Myal Y, Watson PH, et al. Identification of a novel breast- and salivary gland-specific, mucin-like gene strongly expressed in normal and tumor human mammary epithelium. *Cancer Res* 2002;62:2736-40.
19. Bertucci F, Birnbaum D, Goncalves A. Proteomics of breast cancer: principles and clinical applications. *Mol Cell Proteomics* 2006, in press
20. Colpitts TL, Billing P, Granados. E, et al. Identification and immunohistochemical characterization of a mucin-like glycoprotein expressed in early breast cancer. *Tumor Biol* 2002;23:263-278.