Journal of Cellular Biochemistry 17G:173–185 (1993)

# Deranged Activity of the CD44 Gene and Other Loci as Biomarkers for Progression to Metastatic Malignancy

**David Tarin, MA, BSc, DM, MRCPath and Yasuhiro Matsumura, MD**

Nuffield Department of Pathology, Oxford University, John Radcliffe Hospital, Headington, Oxford, England OX3 9DU

**Abstract** About one in three people in modern industrialised countries die of the consequences of malignant tumours or are found to carry an unsuspected one at the time of autopsy. Early resection of such lesions and appropriate adjuvant therapy is very effective in curing the disease. There is therefore a strong clinical incentive to find effective methods of early diagnosis, assessment of prognosis and treatment of neoplastic lesions and research on this topic is directed at a numerically significant medical problem.

Recently it has been found that many human tumours show severe abnormalities in the expression of the CD44 gene which increase with progression to metastatic malignancy. By alternative splicing mechanisms this gene codes for a family of heavily glycosylated cell surface proteins involved in many important cellular activities. In neoplasia there is gross overexpression of various products of the gene associated with disorderly splicing, which can be detected in clinical samples with the sensitive technique of reverse transcription-polymerase chain reaction (RT-PCR). These disturbances begin early in the neoplastic process and can be detected in very small biopsy samples. It has also been shown that it is possible to achieve non-invasive diagnosis of malignancy by analysis of CD44 expression in exfoliated cells in body fluids and waste products. The potential significance of these observations for early diagnosis of symptomatic cancer and for screening of the population for asymptomatic lesions are readily seen and await further investigation.

Separate work in our laboratory has succeeded in DNA-mediated transfer of metastatic capability through two rounds of transfection into non-metastatic tumour cells and a metastasis-associated human DNA fragment has been recovered from the transfectants and sequenced. Using primers designed to anneal to a coding region identified by computer analysis within the novel sequence, it has been shown with RT-PCR that it is heavily expressed in metastatic cancer tissues, but not in corresponding normal ones. This could be of value in assessing the prognosis of patients using small biopsy samples from their primary tumours and the potential of this sequence for such purposes and for possible therapeutic intervention is currently being explored.

Recent work in several laboratories has shown that elevated expression of certain other specific growth factor genes, including c-*met* and EGFR, correlates with metastatic capability. Combined evaluation of such markers in further studies will in time give useful information on the prognosis of individual patients to guide therapeutic decisions and the implications of these recent advances for clinical practice and future research are discussed below. 

Address correspondence to David Tarin, MA, BSc, DM, MRCPath, Nuffield Department of Pathology, Oxford University, John Radcliffe Hospital, Headington, Oxford, England OX3 9DU.

**Key words:** Assessment of prognosis, CD44, early tumour diagnosis, MAGNA sequence, metastasis, tumour progression

In wealthy industrial nations, where good nutrition, public health programmes and modern medicines have minimised the effects of malnutrition and infectious diseases on the length and quality of life, cancer is a major cause of morbidity and mortality. The magnitude of its impact on human health is highlighted by reference to published mortality statistics [1], which show that approximately 50% of deaths annually in the UK are due to cardiovascular disease and, of the remainder, about half (*i.e.*, 25% of the total) to cancer. Most of these patients have disseminated metastases. In addition to this, approximately 5% of patients dying of other causes have an undetected malignant tumour, often metastatic, at autopsy. Also, many more patients with localised cancer are effectively treated and cured. Thus, well over one in three of the population develop a malignant neoplasm at some time in life and about one in four die as a direct result of this parasite which shares near-complete genetic identity with the host. Although there is considerable improvement in cancer treatment, resulting from advances in chemotherapy, immunotherapy and radiotherapy, there is still insufficient reduction in overall cancer mortality and morbidity.

It is now widely recognised that the majority of cancers can be cured by surgical resection if diagnosed at an early stage [2] and this has led to the introduction of cancer screening programmes in many countries. Additionally it is becoming realised that identification of premalignant states and of genetic susceptibilities to certain types of tumours can lead to the development of medical strategies which prevent or substantially delay the onset of frank neoplasia. Therefore, there is considerable incentive for investment of time, effort and money in trying to identify means by which cancers can be detected early in their life history, when they might still be localised and easily resected or destroyed. Radiological diagnostic methods, including contrast studies, computerised tomography, angiography and magnetic resonance imaging are suitable for confirming cancer but not for mass screening, because they are time consuming, expensive and often invasive procedures. Current non-invasive investigations, such as examination of the stool for occult blood [3] and microscopical evaluation of urine cytology [4], have limitations such as lack of specificity or sampling difficulties and are, in some cases, labour intensive. Radioimmunoassay for tumour markers, such as CEA, is not suitable for detecting cancer in its early stages, because the sensitivity or specificity of the ones so far reported are insufficient for identification of very small tumours, but such measures can sometimes be good for judging cancer prognosis or for monitoring patients for increasing tumour burden or tumour recurrence [5]. These circumstances continue to drive a search for biomarkers which could be helpful in diagnosing early tumours and in assessing their metastatic potential.

## EARLY TUMOUR DIAGNOSIS AS A PRACTICAL MEASURE FOR CIRCUMVENTING METASTASIS AND OTHER SYSTEMIC SEQUELAE OF NEOPLASIA

Recent studies have provided evidence that the activity of the CD44 gene is severely deranged in many types of tumours [6–9]. This gene normally codes for a family of heavily glycosylated cell surface proteins, the several isoforms of which exercise many important cellular functions [10]. In cancer there is chaotic overproduction of many unusual mRNA transcripts relative to the picture seen in corresponding normal or non-neoplastic tissues. As the disorder is present in early tumours [11] and can be identified in very small samples using amplification techniques, it seems a promising candidate to study as a possible marker for early diagnosis and for monitoring patients for local and distant recurrent disease [6].

The first observation that there is gross overexpression of numerous splice variants from this gene in human tumour tissue was serendipitously made in the course of studying metastatic cells to test the hypothesis [12] that such cells are inappropriately activating genetic programmes for white blood cell (especially lymphocyte)

recirculatory traffic. An earlier study [13] in this laboratory had provided evidence that the integrin VLA-4, commonly seen on trafficking lymphocytes, is upregulated on metastatic but not on non-metastatic tumour cells. It was therefore decided to examine CD44 expression in such cell lines, because it had been identified as a lymphocyte homing molecule, aiding these white cells to traffic through lymph nodes. The results with human *cell lines* were disappointing, there being no clear correlation with metastatic capability [Matsumura Y, Bao L, Fowler W, and Tarin D, unpublished observation]. However, a recent report [14] of an association between abnormalities in expression of the CD44 gene and metastatic behaviour of tumour cell lines injected into nude mice prompted us to pursue the subject further and to investigate CD44 activity in *fresh human tumour biopsies.* This led to the discovery that there is gross overexpression and abnormal splicing of messenger RNA transcripts from this gene in cancerous tissue samples, relative to samples from corresponding normal tissues or to ones affected by non-neoplastic diseases [6].

The precise functions of the products of this gene are still unclear but there is evidence that some of them are involved in attachment of cells to each other and to the surrounding matrix [15]. The gene has a standard region [16] which appears to be constitutively expressed in most tissues and a group of exons which can be variably transcribed and spliced into the standard region [17–19] according to the prevailing requirements of the cell. The standard portion codes for a 90 kD protein with a cytoplasmic tail, a transmembrane section and an extracellular domain which is heavily glycosylated [16]. To date, one other isoform has been characterised and this contains 135 extra amino acids inserted into the proximal extramembrane domain resulting in a molecule 180 kD in size which is more heavily glycosylated [17]. Immunocytochemical studies show it to be present on many epithelial cells and it is known as the epithelial variant. The detailed functions of molecules containing regions coded for by the other exons shown in Figure 1 are so far unknown but evidence has already accrued that the products of this gene are involved in an impressive variety of functions including:

- adhesion to other cells and to surrounding mesenchymal components including hyaluran, chondroitin sulphate, fibronectin and collagen
- lymphocyte "homing" to mesenteric lymphoid tissue
- T-lymphocyte activation
- signal transduction; alternative splicing of exons coding for the cytoplasmic tail of the molecule is believed to provide forms with different transduction roles which include influencing G-protein status within the cell and stimulating it to secrete specific growth factors
- hyaluran binding and internalisation, as part of the turnover of this component of mesenchymal tissues
- drug uptake and sensitivity, suggesting that some form of the molecule may regulate a membrane channel.

It is not yet known which isoforms are associated with each of these reported activities but the diversity of processes with which this family of molecules appears to be linked in a number of different cell lineages suggests it has important coordinating or regulatory functions.

The technique used for our initial analysis of CD44 gene activity [6] involved cDNA synthesis by reverse transcription (RT) of mRNA extracted from the tumour tissue, followed by its amplification with the polymerase chain reaction (PCR), using primers designed to anneal to specific portions of the gene. This method (*i.e.*, RT-PCR) provides exceptional power and flexibility for the detection of abnormal gene expression in minute samples of tissue, or even a few cells. The differences seen between matched normal and tumour tissue samples from a given patient were obvious at a glance (Fig. 1). When the PCR products were separated by electrophoresis, blotted and probed with a probe for one of the exons in the variably expressed region of the gene, nine or more bands including a doublet at 1500 base pairs (bp) and 1650 bp, each corresponding to a different sized transcript, were seen in all tumours but not in any normal tissue examined. Thus it is the severely deranged *pattern* of CD44 activity rather than any individual band that is seen as tumour specific. The cumulative results of studies on solid tissue samples to date are that cancers affecting the breast, colon, bladder, stomach, prostate and thyroid in 47 different patients had significant overexpression of various exons of CD44 but corresponding normal tissues from

FIG 1

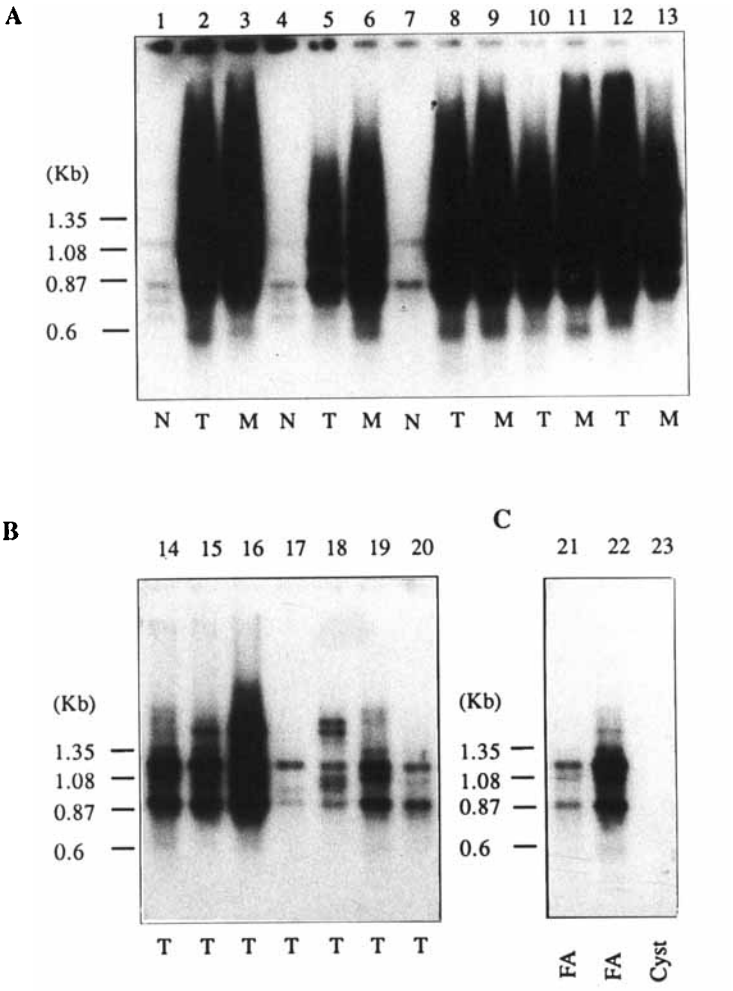


**Fig. 1. Autoradiogram of PCR products from breast tissue samples probed with oligonucleotide E4 (see Fig. 4).** *Panel A:* malignant primary breast carcinomas with their metastases. Tracks 1, 2 and 3: patient B1; tracks 4, 5, and 6: patient B2; tracks 7, 8 and 9: patient B3; tracks 10 and 11: patient B4; tracks 12 and 13: patient B5. It can be seen that compared to normal breast tissue, primary breast carcinomas and their metastatic deposits overexpress several splice-variants. Note the doublet (arrows) at 1500 bp and 1650 bp best seen in track 5. This is present in all tumours and metastases but is fogged in the other tracks by this time of exposure. It is not detectable in any normal samples even at much longer exposure times (23 hours). *Panel B:* Breast carcinomas with no clinical evidence of metastasis. Tracks 14–20 are from patients B15–B21. The tumours all overexpress several variants, but show less bands and the signal intensity is less, except track 16 (patient B17)—see text. The 1500/1650 bp doublet (arrow) is easily recognisable in tracks 15, 16 and 18 at this length of exposure and became detectable in all other tumour-containing tracks on longer exposure. The illustration, however, shows only the shorter exposure, to avoid fogging the tracks which have stronger signals. *Panel C:* Fibroadenomas (FA) and fibrocystic disease of the breast (Cyst). Tracks 21 and 22, containing the benign tumour samples (samples B22 and 23), express more than the non-neoplastic sample (fibrocystic disease) in track 23 (sample B24). Reproduced from [6] with permission from the publisher.

***Notation: N = normal, T = primary tumour, M = metastasis***

TABLE I.

| Type of tissue | Number of patients/volunteers | Number showing increased splice variants |
|---|---|---|
| *Neoplastic* | *47* | *46* |
| Breast cancer | 21 | 21 |
| Colon cancer | 13 | 13 |
| Bladder cancer | 6 | 6 |
| Stomach cancer | 1 | 1 |
| Thyroid cancer | 1 | 1 |
| Fibroadenoma | 2 | 2 |
| Prostate cancer | 3 | 2 |
| *Non-neoplastic* | *39* | *0* |
| Normal breast | 9 | 0 |
| Cystic disease of breast | 1 | 0 |
| Normal colon | 9 | 0 |
| Crohn's disease | 1 | 0 |
| Ulcerative colitis | 1 | 0 |
| Appendicitis | 1 | 0 |
| Normal bladder | 4 | 0 |
| PBL | 10 | 0 |
| Bone marrow | 3 | 0 |

Reproduced from [39] with permission from the publisher

39 people did not (Table I). These consistent and clear differences between normal and tumour tissue are viewed as giving as a possible method for early cancer diagnosis.

A cautionary point should however be noted by those becoming interested in this field. This is that mRNA is notoriously susceptible to degradation by any ribonucleases in the sample and, if extraction is not commenced *immediately* after thawing a specimen, variant CD44 molecules, being larger than "standard" ones, are more susceptible and are not detected, giving an incorrect negative result. In our further experience of detecting tumour cells with this method we have made observations which would be useful to others investigating its diagnostic potential. The main requirement is to minimise false negative results by ensuring that high quality mRNA is routinely obtained and by using internal standards in every reaction to monitor the PCR amplification step. False positives, providing they are not too frequent, are not a serious problem, because they can be detected by replicate assays on the same or further samples and by reference to other clinical data.

We have explored the procedures needed to ensure the routine RT-PCR detection of abnormal CD44 gene activity in small clinical samples containing tumour cells. If a tissue sample is divided into aliquots, half of which are frozen in liquid nitrogen immediately and the remainder of which are left on the bench at ambient temperature, one can show how the ability to detect CD44 splice variants declines with time and with mode of specimen handling. Fresh samples submitted to mRNA extraction within half an hour of excision give the most reliable results and there is a gradual decline in quality over the next few hours (Fig. 2) if the fresh tissue is left on the bench. If the sample is first snap frozen, the results obtained when RNA is extracted immediately after thawing are satisfactory, but decline very rapidly, beginning within 15 minutes, the larger variant transcripts being lost first and ultimately even the standard form (Fig. 3). We have also found that if snap-frozen cell and tissue samples are stored at –70°C the results decline after 4 weeks even if the mRNA is extracted immediately after thawing. It would seem therefore that degradation of RNA by ribonucle-

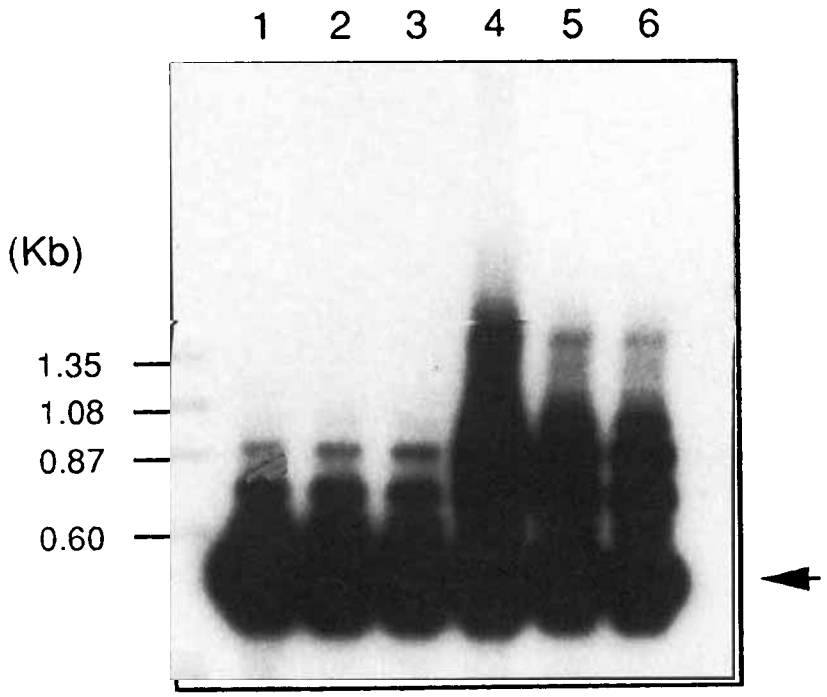


**Fig. 2. Autoradiogram of RT-PCR—Effects of delay in processing a fresh sample.** RT and PCR were performed as described previously [6]. Hybridisation was done with probe S1 which anneals to the standard part of the CD44 mRNA transcript and identifies several alternatively spliced transcripts, as well as the standard form (arrow).

Lanes 1–3 represent normal colon mucosa from a patient. Lanes 4–6 represent colon cancer tissue from the same patient.

Samples analysed in lanes 1 and 4, lanes 2 and 5 and lanes 3 and 6 were left at room temperature for 40 minutes, 3 hours and 5 hours, respectively, after resection.

ases released from cells ruptured during freezing continues even at this temperature, although at a slower rate. This knowledge is critically important to effective use of this potential new diagnostic method. Finally it is important to emphasize the importance of using internal standards in assaying CD44 activity. One way to do this is to confirm amplification of the standard form of CD44 in the same clinical sample. This monitors the PCR process and contributes to ensuring that differences between benign and malignant samples are not artifacts. A further important precaution is to monitor the reliability of amplification of the target DNA segment by the PCR reaction under these conditions: a simultaneous serial dilution study using an external standard consisting of a plasmid containing the segment of interest to confirm that the procedure is operating quantitatively.

As explained previously [6], the sizes of the PCR products depend upon the primers chosen and the types of variant mRNA species present in the tissue. If the cells produce only the standard form of CD44, primers P1 and P4 produce a single band of 482 bp. This corresponds to the distance between the sites to which the primers anneal on the mRNA transcript from the standard part of the gene. Variant mRNA molecules result in larger PCR products, depending on the number and size of the extra exon transcripts fitted into the insertion site (see Fig. 4).

In our earlier study [6] describing abnormalities in CD44 expression in tumours it was noted that one of the PCR products consistently obtained, when amplifying tumour samples with primers P1 and P4, was larger (1650 bp) than the predicted maximum size of 1500 bp, which would result from inclusion of all known exons at the insertion site [18]. The conclusion was that there had to be a further unidentified exon in the variably expressed region of the gene. By further studies using suitably chosen combinations of primers we found this exon to lie 5' to all the other variably expressed ones and thus to be

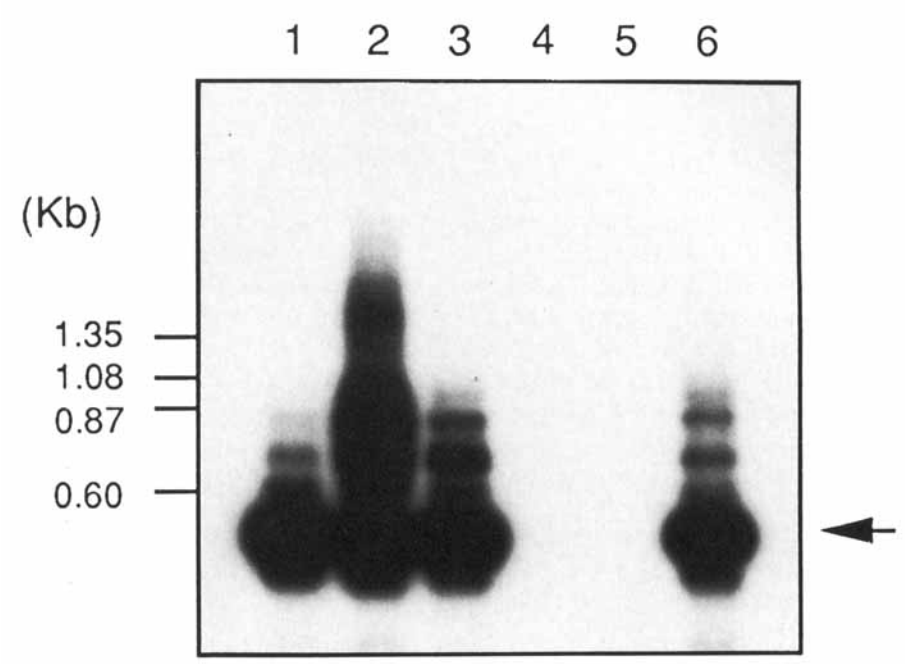


**Fig. 3. Effects of delay in processing a frozen sample after thawing.** All tissue was frozen immediately after resection and RNA was extracted as scheduled below:
Lane 1: normal colon mucosa immediately after thawing
Lane 2: colon cancer tissue, from the same patient, immediately after thawing
Lane 3: colon cancer tissue, from the same patient, 15 minutes after thawing
Lane 4: colon cancer tissue, from the same patient, one day after thawing
Lane 5: No DNA
Lane 6: normal thyroid tissue
The arrow denotes the position of the amplified standard CD44 transcript.

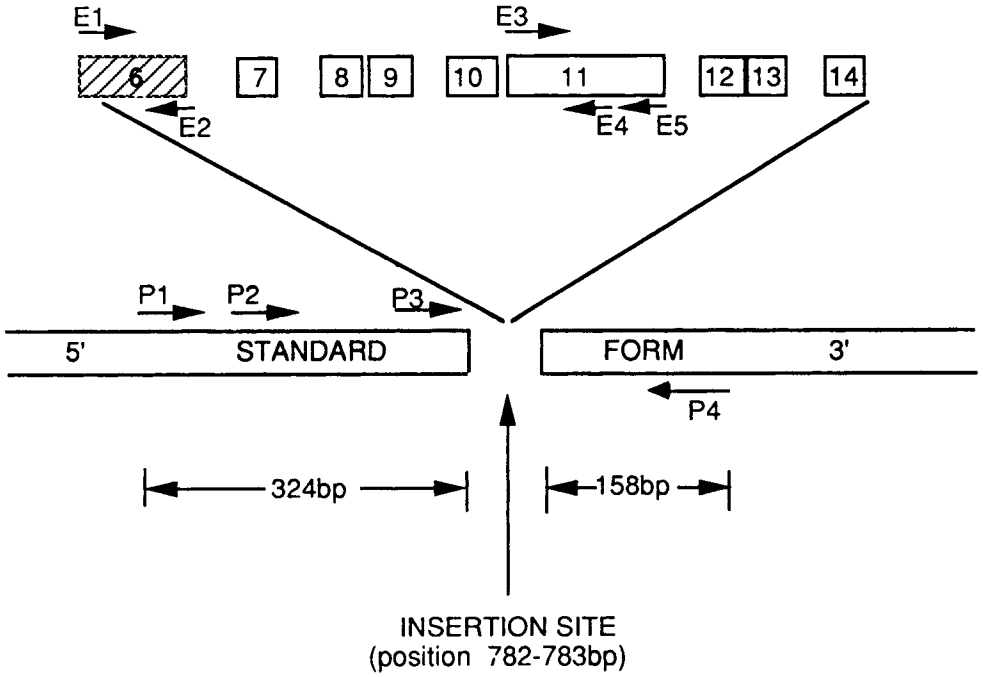


**Fig. 4. Map of the CD44 gene to show positions of probes and primers.**

between them and the 5' section of the standard portion of the gene (see Fig. 4). The new exon was cloned and sequenced and the details are published elsewhere [11]. It is 129 bp long and the predicted 43 amino acid peptide for which it codes contains 34% of serine and threonine residues and two potential O-glycosylation sites. This sequence is identical to that recently isolated from a genomic library of the HT-29 colon cancer cell line and published by Screaton *et al.* [20], during the conduct of our present study. In accordance with their system of numbering of CD44 exons, based upon analysis of genomic DNA, we shall refer to it as exon 6.

We next decided to examine the feasibility of detecting occasional tumour cells exfoliated into body fluids and effluents, obtainable non-invasively. We chose to start with the detection of malignant cells in the cell pellet sedimented from the urine of patients with bladder cancer by identification of abnormal CD44 activity. In this study, amplification was performed across a shorter section of the CD44 cDNA in each sample to increase sensitivity and specificity. Primers were designed to amplify across segments which included the new exon, on the grounds that this is present in the largest transcripts typical of neoplasia. If the tumour cells in the urine were to be expressing all the exons from 6 to 11, it was predicted that the PCR products derived from this protocol using primers E1 and E5 should include a 735 bp band. Detection of this was enhanced by blotting and probing with a labelled oligonucleotide (E4) designed to anneal to the amplified transcripts.

Urine specimens from 90 individuals were examined; 44 from patients with biopsy-proven bladder carcinoma and 46 from asymptomatic volunteers or patients with non-neoplastic conditions. The overall results of this preliminary study showed that samples from 91% (40/44) of tumour-bearing patients were positive and that 83% (38/46) of those from people in the non-neoplastic control group were negative [11]. These observations therefore confirm earlier scientific reports of unusual activity of the CD44 locus in neoplasia and malignancy. They also indicate that techniques for the analysis of such activity can enable non-invasive investigation of patients for primary or recurrent bladder cancer and for tumours of other organs, by detection of neoplastic cells shed into body fluids. For instance, we have recently successfully applied it to the non-invasive detection of exfoliated colorectal cancer cells in stools [Matsumura and Tarin, unpublished observations]. The presence and intensity of the band is not simply related to the total number of cells present. From the present work and from our earlier studies on calibration of the sensitivity of tumour cell detection with this marker [6], we believe that tumour cells produce orders of magnitude more of the unusual transcripts from the CD44 locus than non-neoplastic cells. Even a sprinkling of tumour cells in a sample can therefore be detected under optimal conditions and the potential utility of this maker in clinical diagnosis, screening and monitoring for tumour recurrence is now being evaluated in larger studies. Reports from other laboratories have provided evidence supporting the above description of disturbed CD44 expression in tumour tissue at both the mRNA and protein level [7–9]. The latter involved immunocytochemical studies of tissues stained with monoclonal antibodies to some CD44 variants and further investigation of the possible diagnostic benefits of working with such antibodies in various types of assays is needed.

Recently, Sidransky *et al.* [21] reported identification of p53 gene mutations in exfoliated cancer cells in urine in three patients. However, in studies on solid tissue samples they found mutations in these genes are present in only 61% of bladder tumours. Haliassos *et al.* [22] sought evidence of H-*ras* gene mutations in the urine of 21 patients with bladder cancer and reported mutations at codon 12 in 10 (47%) of them. Such DNA-based diagnostic methods are useful for epidemiological and family studies on cancer aetiology because they detect predisposition and risk, but RNA and protein-based methods are more suited to the screening, diagnosis and clinical assessment of patients who currently have neoplastic disease, because they evaluate disturbances in the functional activity of relevant genes. Whilst protein-based methods using antibodies are obviously easier and more robust, because mRNA is so readily degraded by ubiquitous ribonucleases, the RT-PCR method has unrivalled sensitivity and power. The abundance of the expression of the marker gene will dictate whether RT-PCR or ELISA is more appropriate for a given clinical setting.

A different approach to tumour diagnosis by

analysis of gene expression which is being explored by Folkman and colleagues [23] for bladder cancer detection using urine samples, deserves consideration. This short report describes detection of tumour angiogenesis factor by radioimmunoassay as being an effective method. The production of angiogenesis factors has been demonstrated by Folkman's group in many previous studies to be an important requirement for recruitment of a vascular supply for growing tumours without which they cannot grow beyond a few millimetres in diameter. The idea is therefore very original and interesting but at present there is insufficient information to judge whether, in the absence of amplification it will be possible to reproducibly detect small quantities of such signal molecules in large volumes of fluid.

Further work is needed before the relative practical merits of all of these novel methods for early tumour detection can be evaluated.

## NEW METHODS FOR ASSESSMENT OF METASTATIC POTENTIAL AND PROGNOSIS

Prognostic evaluation of patients who already have established tumours but no current clinical evidence of metastatic spread, is another important area of cancer management in which progress is taking place. Chemoprevention of this aspect of malignant disease has substantial appeal, even if it is of subordinate priority to attempts to block the whole process of neoplasia. Recent advances include the recognition of some potentially useful markers of metastatic progression and some possible insights into the mechanisms involved and these are reviewed below.

In work recently conducted in this laboratory it has been found that, if one is sufficiently persistent, it is feasible to transfer metastatic capability from metastatic tumour cells to non-metastatic tumour cells, by transfection (Fig. 5) with genomic DNA from the metastatic population [24]. On inoculation into nude mice the transfected cells make many metastatic deposits in various organs. The new phenotype is stable through many cell generations and can be transferred again in a second round of transfection, using DNA from metastases formed by the primary transfectants which we have introduced into fresh cells of the non-metastatic mouse cell line. Subsequently, we demonstrated that concomitant transfer of the donor DNA (of human origin) through both rounds of transfection, can be detected by several convergent lines of evidence, including Southern blotting, *Alu*-PCR and *in situ* hybridisation [25] using human *Alu*-specific probes with appropriate controls. Still more recently we have isolated clones containing human DNA, from the transfected metastatic cells by making a genomic library of their DNA, in cosmids and screening it with human *Alu*-specific probes. From one of the bacterial clones so identified we have subcloned a 2.9 kb DNA fragment (designated MAGNA, an acronym for Metastasis Associated Gene or Nucleic Acid) that hybridises specifically to Southern blots of human DNA to identify a sharp homologous band suggestive of a sequence present in single or low copy number. This indicates that the homology is not due to multiple iterative sequences, which would have been expected to produce a smear. (It should be mentioned that, to visualise the band, non-specific cross hybridisation of *Alu* repeats in the probe to counterparts in the target human DNA, was blocked with excess unlabelled *Alu* DNA prepared by PCR).

We have now sequenced this genomic subclone, and comparison with GenBank/EMBL DataBank indicates that it contains a novel piece of DNA that has not previously been recorded. Initial computer analysis has indicated that the sequence contains sections with characteristics signifying high probability that they are coding regions. The 2.9 kb fragment also contains numerous *Alu* repeat elements of the specific type seen in higher primates, confirming its human origin. Preliminary data, obtained with Northern blotting and RT-PCR techniques indicate one of these regions to be differentially transcribed in metastatic versus benign neoplasms. The implications of this for prognostic assessment of patients with localised, early invasive, tumours and the possibilities for understanding underlying mechanisms of metastasis are now being explored.

In 1988, Steeg and colleagues [26] cloned and sequenced a separate gene which appears to be relevant to the metastatic process and may have diagnostic potential. This gene, designated *nm*23, was isolated by differential hybridisation from a non-metastatic mouse cell line and is conserved in human cells. Studies to date show a promising correlation between reduced expression of the gene and metastasis in breast cancer patients

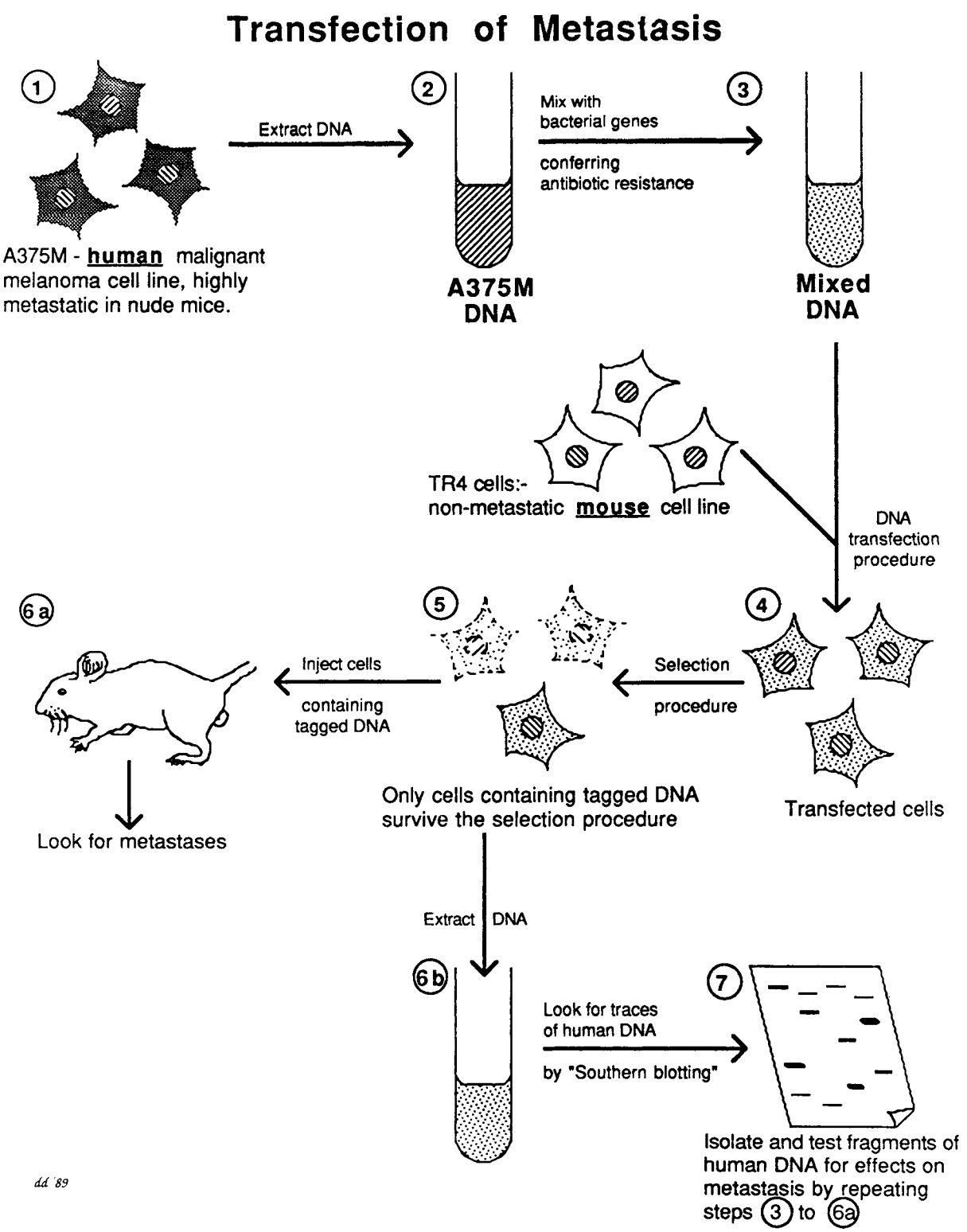


**Fig. 5. Diagram of the experiment to transfer metastatic activity with total genomic DNA.** Reproduced from [39] with permission from the publisher.

[27,28], but the results in colon carcinoma [29,30] are so far contradictory. However, the gene might prove to be a useful marker in certain types of cancer and further results are awaited with interest. Functional relationships between the *nm*23 gene, which appears to have metastasis-inhibitory activity in some circumstances and the sequence we have just isolated, which seems to be positively associated with clinical expression of the phenotype, remain to be determined, but these issues do not directly affect the analysis of their clinical utility.

Very recently results from three laboratories [31–33] have indicated that counting the number of capillaries in the most vascularised area of a tumour might give a useful pragmatic indication of patient survival or of tumour recurrence in breast cancer and non-small cell lung cancer. The underlying basis for this clinical evaluation is the well established knowledge that vascularisation is essential for tumours to grow beyond a few millimetres in size and for dissemination of tumours by blood and lymphatic routes.

Other variables such as tumour size, oestrogen receptor status and level of expression of certain other markers, such as the epidermal growth factor receptor (EGFR) and the *neu* oncogene have been reported [34] to be useful extra discriminators for judging which patients with *node-negative* breast cancer at the time of diagnosis, might have earlier progression of disease, shorter survival and poorer response to hormones. The work of Tahara and colleagues [35,36] illustrating the association of increased expression of certain cellular growth factor-related genes including EGFR, *erb*B-2 and c-*met* with poor prognosis in patients with stomach cancer is particularly notable. Such stratification is clearly beneficial and enables therapy to be optimised for patients in good and bad prognosis categories.

So far, the associations with prognosis for all of these markers are based on statistical correlations which are highly significant but do not specifically indicate *individual* prognosis. Research therefore continues for some marker that is invariably associated with good or bad prognosis, so that therapy can be still more accurately and appropriately adjusted to the particular patients needs. Further work, both on the markers currently under investigation, and on the underlying genetic basis of metastasis recently initiated in some laboratories (see above), is hopefully going to increase our skill in assessment of the requirements of individual patients.

Meanwhile, the subject of how to treat disseminated metastases requires similar intensive scrutiny.

## OVERVIEW

The aims of defining biomarkers which herald incipient neoplasia (or metastatic dissemination) and of finding chemopreventive agents which will block or delay tumour progression as proposed by Kelloff *et al.* [37], are novel, ambitious and worthwhile. At present there are several interesting candidates for such markers, some of which are briefly reviewed above, but more work is needed before it will be possible to judge which, if any, of those so far found, are of the most practical value. Even though the task may appear daunting, those who wish to ascertain if the aim can be realised will perhaps be inspired and sustained by the insight of the ancient Chinese philosopher Lao Tzu (604–531 BC) [38], who counselled that a journey of a thousand miles must begin with a single step.

## ACKNOWLEDGMENT

The authors are grateful to Mrs. L. Summerville for expert help in assembly of this manuscript.

## REFERENCES

1. Office of Population Censuses and Surveys, Mortality statistics: Cause 1989. Series DH2 No. 16, Her Majesty's Stationery Office, London.
2. Kantoff PW. (1990) Bladder cancer. Curr Probl Cancer 14:233–292.
3. Winawer SJ, Schottenfield D, Frehinger BS. (1991) Colorectal cancer screening. J Natl Cancer Inst 83: 243–253.
4. Matzkin H, Moinuddin SM, Soloway MS. (1992) Value of urine cytology versus bladder washing in bladder cancer. Urology 39:201–203.
5. Hedin A, Carlsson L, Berglund A, Hammarstrom S. (1983) A monoclonal antibody-enzyme immunoassay for serum carcinoembryonic antigen with increased specificity for carcinomas. Proc Natl Acad Sci USA 80:3470–3474.
6. Matsumura Y, Tarin D. (1992) Significance of CD44 gene products for cancer diagnosis and disease evaluation. Lancet 340:1053–1058.
7. Tanabe KK, Ellis LM, Saya H. (1993) Expression of CD44R1 adhesion molecule in colon carcinomas and

metastases. Lancet 341:725–726.
8. Finn L, Becich M, Dougherty G, Finley G, Meisler A, Cooper D. (1993) CD44 alternative splicing in human colon carcinomas. Lab Invest 68:49A.
9. Heider K-H, Hofmann M, Hors E, van den Berg F, Ponta H, Herrlich P, Pals ST. (1993) A human homologue of the rat metastasis-associated variant of CD44 is expressed in colorectal carcinomas and adenomatous polyps. J Cell Biol 120:227–233.
10. Haynes BF, Telen MJ, Hale LP, Denning SM. (1989) CD44—A molecule involved in leukocyte adherence and T-cell activation. Immunol Today 10:423–428.
11. Matsumura Y, Hanbury D, Smith J, Tarin D. (1993) Non-invasive detection of malignancy by testing for presence of unusual CD44 gene activity in exfoliated cancer cells. Br Med J (in press).
12. Tarin D. (1985) Clinical and experimental studies on the biology of metastasis. Biochim. Biophys Acta 780: 227–235.
13. Bao L, Pigott R, Matsumura Y, Baban D, Tarin D. (1993) Correlation of VLA-4 integrin expression with metastatic potential in various human tumour cell lines. Differentiation 52:239–246.
14. Gunthert U, Hofmann M, Rudy W, Reber S, Zoller M, Haubmann I, Matzku S, Wenzel A, Ponta H, Herrlich P. (1991) A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell 65:13–24.
15. Picker LJ, De los Toyos J, Telen MJ, Haynes BF, Butcher EC. (1989) Monoclonal antibodies against the CD44 [*In(Lu)*-related p80], and Pgp-1 antigens in man recognize the Hermes class of lymphocyte homing receptors. J Immunol 142:2046–2051.
16. Stamenkovic I, Amiot M, Pesando JM, Seed B. (1989) A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. Cell 56:1057–1062.
17. Stamenkovic I, Aruffo A, Amiot M, Seed B. (1991) The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. EMBO J 10:343–348.
18. Hofmann M, Tudy W, Zoller M, Tolg C, Ponta H, Herrlich P, Gunthert U. (1991) CD44 splice variants confer metastatic behaviour in rats: Homologous sequences are expressed in human tumour cell lines. Cancer Res 51:5292–5297.
19. Jackson DG, Buckley J, Bell JI. (1992) Multiple variants of the human lymphocyte homing receptor CD44 generated by insertions at a single site in the extracellular domain. J Biol Chem 267:4732–4739.
20. Screaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U, Bell JI. (1992) Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. Proc Natl Acad Sci USA 89:12160–12164.
21. Sidransky D, von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, Paul M, Green P, Hamilton SR, Frost P, Vogelstein B. (1991) Identification of p53 gene mutations in bladder cancers and urine samples. Science 252:706–709.
22. Haliassos A, Liloglou T, Likourinas M, Doumas C, Ricci N, Spandidos DA. (1992) H-*ras* oncogene mutations in the urine of patients with bladder tumors: Description of a non-invasive method for the detection of neoplasia. Int J Oncol 1:731–734.
23. Nguyen M, Watanabe H, Budson AE, Richie JP, Folkman J. (1993) Elevated levels of the angiogenic peptide basic fibroblast growth factor in urine of bladder cancer patients. J Natl Cancer Inst 85:241–242.
24. Tarin D. (1988) Molecular genetics of metastasis. In Whelan J, Bock GR (eds): "Ciba Foundation Symposium on Metastasis." London: John Wiley & Sons Ltd, pp 149–169.
25. Hayle AJ, Darling DL, Taylor AR, Tarin D. (1993) Transfection of metastatic capability with total genomic DNA from metastatic tumour cell lines. Differentiation 54:177.
26. Steeg PS, Bevilacqua G, Kopper L, Thorgeirsson UP, Talmadge JE, Liotta LA, Sobel ME. (1988) Evidence for a novel gene associated with low tumor metastatic potential. J Natl Cancer Inst 80:200–204.
27. Barnes R, Masood S, Barker E, Rosengard AM, Coggin DL, Crowell T, Richter King C, Porter-Jordan K, Wargotz ES, Liotta LA, Steeg PS. (1991) Low *nm*23 protein expression in infiltrating ductal breast carcinomas correlates with reduced patient survival. Am J Pathol 139:245–250.
28. Hennessy C, Henry JA, May FEB, Westley BR, Angus B, Lennard TWJ. (1991) Expression of the antimetastatic gene *nm*23 in human breast cancer: An association with good prognosis. J Natl Cancer Inst 83:281–285.
29. Cohn KH, Wang F, Desoto-Lapaix F, Solomon WB, Patterson LG, Arnold MR, Weimar J, Feldman JG, Levy AT, Leone A, Steeg PS. (1991) Association of *nm*23-H1 allelic deletions with distant metastases in colorectal carcinoma. Lancet 338:722–724.
30. Haunt M, Steeg PS, Willson JKV, Markowitz SD. (1991) Induction of *nm*23 gene expression in human colonic neoplasms and equal expression in colon tumors of high and low metastatic potential. J Natl Cancer Inst 83:712–716.
31. Macchiarini P, Gontanini G, Hardin MJ, Squartini F, Angeletti CA. (1992) Relation of neovascularisation to metastasis of non-small-cell lung cancer. Lancet 340: 145–146.
32. Bosari S, Lee AK, DeLellis RA, Wiley BD, Heatley GJ, Silverman ML. (1992) Microvessel quantitation and prognosis in invasive breast carcinoma. Hum Pathol 23:755–761.
33. Horak E, Leek R, Klenk N, Le-Jeune S, Smith K, Stuart N, Greenall M, Stepniewska K, Harris AL. (1992) Angiogenesis, assessed by platelet/endothelial cell adhesion molecule antibodies, as indicator of node metastases and survival in breast cancer. Lancet 340:1120–1124.
34. Harris AL, Nicholson S, Sainsbury JRC, Farndon J, Wright C. (1989) Epidermal growth factor receptors

in breast cancer: Association with early relapse and death, poor response to hormones and interactions with *neu*. J Steroid Biochem 34:123–131.
35. Tahara E. (1993) Molecular mechanisms of stomach carcinogenesis. J Cancer Res Clin Oncol 119:265–272.
36. Kuniyasu H, Yasui W, Yokazaki H, Kitadi Y, Tahara E. (1993) Aberrant expression of c-*met* mRNA in human gastric carcinomas. Int J Cancer 55:1–4.
37. Kelloff GJ, Malone WF, Boone CW, Steele VE, Doody LA. (1992) Intermediate biomarkers of precancer and their application in chemoprevention. J Cell Biochem 16G (Suppl):5–21.
38. Lao Tzu. (1992) In: The Way of Lao Tzu; 64. Cited by Bartlett's Familiar Quotations, 16th ed. Boston: Little, Brown & Co.
39. Tarin D, Matsumura Y. (1993) Advances in the study of tumour invasion and metastasis. J Clin Pathol (in press).