available at www.sciencedirect.com

www.elsevier.com/locate/molonc

Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: Clinical implications

*Mika Suzuki, David Tarin**

Rebecca and John Moores UCSD Comprehensive Cancer Center, Department of Pathology, University of California, San Diego, 3855 Health Sciences Drive MC0803, La Jolla, CA 92093-0803, USA

ARTICLE INFO

Article history:
Received 14 February 2007
Received in revised form
26 March 2007
Accepted 28 March 2007
Available online 7 April 2007

Keywords:
Breast cancer
Clonal heterogeneity
Gene expression
Metastasis
Tumor progression

ABSTRACT

The genetic program that drives tumor metastasis and the mode and timing of its initiation are of great practical significance to clinical management. Modern technical advances open new opportunities for gaining useful relevant information. Gene expression profiles of histologically-verified viable tissue from lymph node metastases were compared with those of matched primary breast cancers from 10 different patients, among samples from over 400 cases, using high-throughput oligonucleotide arrays comprising probes for 22,000 genes. It was observed that metastases have very similar expression signatures to their parent tumors. However, detailed computational analysis revealed that a small number of genes were consistently differentially expressed between 100% of tumors and metastases, suggesting that these are mechanistically important. Lists of such candidate genes, of potential clinical interest, are provided. We interpret these results in the framework of a meta-analysis of previous investigations by others and ourselves and of existing clinical knowledge on the behavior of human tumors. The collective data show that metastases resemble their primary tumors but the signatures obtained in different studies are not sufficiently reproducible or informative to be prognostically useful, although they do give valuable insights into the pathogenesis and biology of human tumor metastasis. The findings indicate that the genetic program encoding metastasis is implemented progressively over time although, occasionally, this evolution can occur rapidly, early in the life of the neoplasm. The important clinical significance of this deduction is that, in most patients, early detection provides time for appropriate therapeutic intervention to be effective in obstructing metastasis.

1. Introduction

One of the major urgent needs in modern cancer research is to understand the molecular mechanisms driving tumor cells to metastasize, including how and when these are activated (Paget, 1889; Tarin et al., 1984; Tarin and Price, 1981). Accordingly, we have conducted a detailed study of gene expression patterns of primary tumors and their metastases seeking clinically meaningful patterns. We focused our enquiry on matched primary breast cancers and their metastases, to reduce background noise from genetic variations between different patients. Additionally, we used only lymph node metastases, so that variations in gene expression due to organ site would be excluded, thereby benefiting the search for genes

* *Corresponding author*. Tel.: +1 858 822 2081.
E-mail address: dtarin@ucsd.edu (D. Tarin).

1574-7891/$ – see front matter © 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.molonc.2007.03.005

involved in metastasis. In this article we interpret the findings in terms of their wider meaning for the biology and pathogenesis of cancer metastasis in humans.

There have been only four other original studies of gene expression in matched metastases and primary tumors from human subjects (Weigelt et al., 2003, 2005a; Hao et al., 2004; Feng et al., 2006). (NB: Two additional papers have appeared from these groups re-analysing the same data (Weigelt et al., 2005b; Lahdesmaki et al., 2004).) The first compared breast cancer primaries with their metastases in several different organs including a single lymph node metastasis (Weigelt et al., 2003). The other three (Hao et al., 2004; Weigelt et al., 2005a; Feng et al., 2006) compared matched primary breast carcinomas with their lymph node metastases. All three were conducted with custom cDNA arrays, used simple histology (in contrast to our methods) for sample selection and tested results for 5 (or less Feng et al., 2006) genes with another method. The current work progresses from these studies by (i) performing rigorous, multi-observer, histopathological quality control of all tumor tissue used for analysis, (ii) using HG-U133A oligonucleotide arrays interrogating approximately 22,000 genes simultaneously, (iii) validating the results for a panel of 24 genes with quantitative PCR on RNA from every one of the same samples, (iv) comparing the findings with data from (a) a parallel study on xenografted breast cancer and its metastatic deposits and (b) a meta-analysis of previous clinical studies. (v) Identifying genes which are differentially expressed to a statistically highly significant degree in 100% of cases or 90% of cases and conducting correlation coefficient analysis on the remaining genes to show that they are indeed very similarly expressed in primaries and metastases. The results provide identities of differentially expressed genes in primary and secondary breast cancers and strongly support the conclusion that acquisition of metastatic competence occurs progressively over time and is only rarely an early event in neoplastic progression. This current work therefore rebuts some recent claims (Weigelt et al., 2003, 2005a; Ramaswamy et al., 2003; Bernards and Weinberg, 2002; van't Veer et al., 2002) asserting that early conversion is the rule and leads to clinically important conclusions regarding the emergence of the metastatic phenotype in human neoplasms.

2. Results

2.1. Histopathological data on clinical samples

Seven of the 10 primary tumors selected for analysis were ductal carcinomas, two were lobular carcinomas and one was a mixed ductal and lobular carcinoma. The histologic grade of the ductal carcinomas, evaluated according to criteria described by Bloom and Richardson (1957), are provided in Table 1. Representative photomicrographs (Figure 1A, B) of frozen sections of one of the primary tumors and its matching lymph node metastasis, demonstrate that the samples selected for study were histopathologically confirmed high quality, viable samples of infiltrating breast carcinomas and metastases containing abundant tumor cells.

Table 1 – Clinical information

Sample number	Pathological diagnosis	Grade	% Of tumor tissue	
			Primary tumor	Metastases
1	Lobular carcinoma	(–)	73.75	76.25
2	Ductal carcinoma	G3	83.5	87.5
3	Ductal carcinoma	G1	71.25	85
4	Ductal carcinoma	G2	70	80
5	Ductal carcinoma	G2	72.5	83.75
6	Ductal carcinoma	G3	73.75	73.3
7	Ductal carcinoma	G3	83.75	91.25
8	Ductal carcinoma	G3	80	80
9	Lobular carcinoma	(–)	83.75	81.25
10	Lobular and ductal carcinoma (Mix)	(–)	82.5	85

Pathological diagnosis, histologic grade and percentage of tumor tissue in each sample. There was no significant difference between primary tumors and metastases in proportion of tumor tissue (PT vs. LN; $p = 0.0645$).

The proportions of tumor and stroma in each specimen are recorded in Table 1. There was no normal breast or lymph node tissue and the only non-malignant component was tumor stroma, which includes capillaries, lymphatics, macrophages, etc. All samples yielded sufficient high quality RNA for array hybridization.

2.2. Microarray analysis

The numbers of differentially expressed genes identified by each of the sets of screening filters (p-value combined with fc) shown in Table 2 were relatively small compared with the overall number of genes interrogated. Even at the lowest level of stringency used, $p < 0.05$ and fc > 1.2, only 155 genes ($98 + 57$) were differentially expressed between primary tumor and metastases. Setting the filters at $p < 0.05$ and fc > 1.5, the list amounted to 84 genes (51 up-regulated in primaries + 33 up-regulated in metastases versus their parent tumors) and Table 3 presents the top 20 candidates, ranked by fc. (The full list is provided in Table S1 in the supplementary material at the journal website.) This result indicates that the expression signatures of the primaries and their metastases are very similar. However, it is important to note that when the number of highly statistically significant differences between groups of observations is small it does not necessarily mean that the two groups are very similar. For this to be concluded one needs to demonstrate that the correlation coefficient between them approaches a value of 1. The correlation coefficient (R) between levels of expression of the 22,000 genes in primaries and metastases in this study was 0.991 ($R^2 = 0.98$). This measure of similarity has not been undertaken by previous studies. (A separate analysis of 96 transcription factor expression we conducted, using commercial membrane arrays, also showed that the profiles of expression of these proteins in primary and matched metastatic tumors are extremely similar – Suzuki, M., Montel, V., Tarin, D., unpublished observations.) However, Table 2 shows that some (19) ($16 + 3$) of these genes are highly reproducibly ($p < 0.01$) differentially

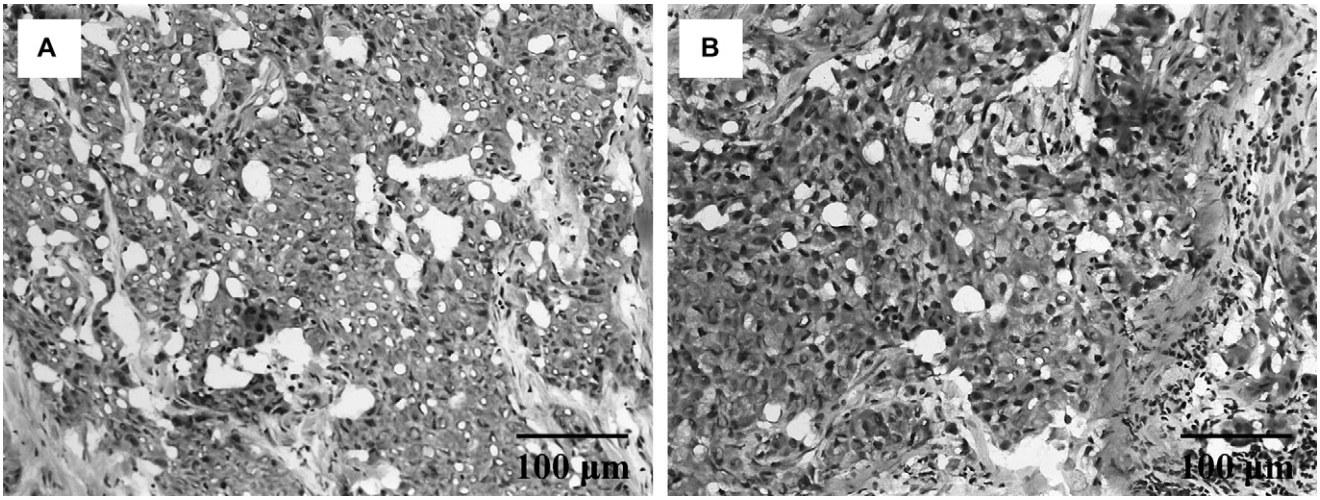

Figure 1 – Photomicrographs of sections of tumor and metastasis. Hematoxylin and eosin staining of a representative primary tumor (A) and corresponding lymph node metastasis (B) to illustrate high quality of tissue and matching tumour/stromal cell content. Original magnification, 200×.

(fc > 1.5) expressed between metastases and primary tumor samples, which represents a reliable change in magnitude of expression of at least 50%. We next examined the magnitude and direction of differential expression of each of the 84 genes selected by the filter combination $p < 0.05$ and fc > 1.5, in each of the 10 pairs of tumor samples. This revealed that 12 genes were consistently differentially expressed in the same trend (i.e. up or down relative to its matched sample from the same patient) in all 10 patients and for 9 genes the p-value was <0.01. An additional 15 genes were differentially regulated consistently in 9 out of 10 patients. Table 4 a and b provides the identities of these 27 (i.e. 12 + 15) of these genes. (The sets of genes listed in these tables do not overlap because the probability values change when the number of patients in the group differs.) It can be seen that several of them, such as genes for metalloproteinases, collagens, and sparc/osteonectin, have been implicated in previous studies on metastasis. Some, such as the genes for metalloproteinases and collagens, are up-regulated in the primary tumors relative to the metastases, implying that they may be of more importance in the early stages of metastasis. Also, prostaglandin D2 synthase has been identified in a previous study of differential gene expression in human prostate cancer and confirmed by immunohistochemistry (Stuart et al., 2004). For all of these significant genes, the degree of differential expression is not huge. The range of fc extends from 1.51 to 3.68, but the differential expression of these genes is still consistently detectable in a background of data from ~22,000 other genes, although the samples come from patients of completely different genetic backgrounds.

Table 2 – Numbers of genes up-regulated in (a) primary tumors and (b) lymph node metastases at different filter settings for p-value and fold change

	$p < 0.05$	$p < 0.01$	$p < 0.005$	$p < 0.001$
(a)				
fc > 1.2	98	22	10	1
fc > 1.5	51	16	8	1
fc > 2.0	19	11	5	1
fc > 3.0	4	3	1	0
(b)				
fc > 1.2	57	3	0	0
fc > 1.5	33	3	0	0
fc > 2.0	22	0	0	0
fc > 3.0	6	0	0	0

2.3. *Validation of microarray results*

To validate the data from microarray analysis, transcript levels for 24 genes were re-examined by quantitative real-time reverse transcriptase PCR (qRT-PCR), using the RNA from the same tissue samples. Six of these genes also belonged to the set of 27 genes (Table 4), which were differentially regulated in 9 out of 10 patients (see above). Absolute intensity values and fc ratios for each of these genes, calculated from the microarray data, were compared with those calculated from the qRT-PCR measurements (Table 5). The results confirmed that for 19 (79.1%) genes the qRT-PCR measurements were in the same trend (i.e. up or down regulated relative to its paired sample) as those obtained by the microarray analysis. For some genes the levels of expression measured by microarrays were approximately the same in tumors and metastases (i.e. fc ~ 1) and the qRT-PCR measurements validated this result. For some other genes, the actual magnitude of expression and fold change measured by real-time quantitative PCR was significantly greater, indicating that the dynamic range of the chip is limited for these genes and can be saturated (Table 5). Overall, the microarray analyses were very effective in providing high-throughput genome-wide screening for differential gene expression between the primary tumors and metastases, but somewhat underestimated the degree of change.

Table 3 – Lists of differentially expressed genes ($p < 0.05$, fc > 1.5)

Accession No.	Gene	Fold change	p-Value
(a) Top 20 genes up-regulated in the primary tumor (vs. the LN metastases) ranked according to fc			
NM_001854.1	Collagen, type XI, alpha 1	3.68	0.00634
BF062629	Ras-induced senescence 1	3.65	0.013503
AW665892	Microfibril-associated glycoprotein-2	3.37	0.010502
NM_017680.1	Asporin (LRR class 1)	3.19	0.005181
NM_002427.2	Matrix metalloproteinase 13 (collagenase 3)	2.95	0.014246
NM_002421.2	Matrix metalloproteinase 1 (interstitial collagenase)	2.84	0.009708
BC002690.1	Keratin 14	2.77	0.008984
NM_002048.1	Growth arrest-specific 1	2.73	0.002825
AF231124.1	Sparc/osteonectin	2.55	0.000432
AI885290	Spondin 1, (f-spondin) extracellular matrix protein	2.45	0.028658
NM_002423.2	Matrix metalloproteinase 7 (matrilysin, uterine)	2.36	0.025449
NM_006207.1	Platelet-derived growth factor receptor-like	2.35	0.013912
NM_013989.1	Deiodinase, iodothyronine, type II	2.33	0.00426
NM_004791.1	Integrin, beta-like 1 (with EGF-like repeat domains)	2.22	0.005629
NM_001609.1	Acyl-Coenzyme A dehydrogenase, short/branched chain	2.2	0.036216
NM_002422.2	Matrix metalloproteinase 3 (stromelysin 1, progelatinase)	2.1	0.001903
AU147799	LOC151996	2	0.041053
AI806793	Collagen, type VIII, alpha 2	1.91	0.029079
NM_001548.1	Interferon-induced protein with tetratricopeptide repeats 1	1.87	0.027199
NM_006486.1	Fibulin 1	1.86	0.013926
(b) Top 20 genes up-regulated in the LN metastases (vs. the primary tumor) ranked according to fc			
NM_006419.1	Chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)	3.63	0.044477
BC001872.1	Immunoglobulin heavy constant mu	3.42	0.01889
NM_000587.1	Complement component 7	2.82	0.013163
U88321.1	Chemokine (C-C motif) ligand 19	2.71	0.039997
D84140.1	Immunoglobulin light chain lambda variable region	2.62	0.01445
NM_002341.1	Lymphotoxin beta (TNF superfamily, member 3)	2.52	0.045825
NM_016459.1	Proapoptotic caspase adaptor protein	2.43	0.031119
M15564.1	T cell receptor beta locus	2.31	0.025533
N90866	CDW52 antigen (CAMPATH-1 antigen)	2.15	0.043219
NM_003385.1	Visinin-like 1	1.89	0.039138
NM_000954.1	Prostaglandin D2 synthase 21kDa (brain)	1.89	0.008787
AW241715	Ubiquitination factor E4B (UFD2 homolog, yeast)	1.78	0.04823
NM_003933.2	BAI1-associated protein 3	1.69	0.032552
BC002807.1	Membrane-spanning 4-domains, subfamily A, member 1	1.67	0.031056
NM_018468.1	Presenilin enhancer 2	1.67	0.028974
L07335.1	SRY (sex determining region Y)-box 2	1.64	0.047188
AI719730	Guanylate cyclase 1, soluble, alpha 3	1.64	0.037663
D83778.1	KIAA0194 protein	1.63	0.0376
NM_007286.1	Synaptopodin	1.51	0.02217
NM_003057.1	Solute carrier family 22 (organic cation transporter), member 1	1.49	0.048529

3. Discussion

Studies aiming to identify candidate genes responsible for the phenomenon of tumor metastasis are difficult to conduct, especially using human tumor samples. First, it is a kinetic event, the timing of which is unpredictable and this complicates the collection of clinical samples at appropriate stages. Secondly, it is essential to use tissue samples from patients who have not already been treated with chemotherapy or radiotherapy, to ensure that the analysis and interpretation of the results is not confounded by changes in gene expression induced by the drugs or radiation. Thirdly, to obtain a comprehensive picture, it is important to compare gene expression profiles of a specific primary tumor type (e.g. breast) with metastases in a given target organ (e.g. lymph node) to improve identification of mechanistically significant candidates and avoid organ-related variations. Furthermore, in order to minimize the background noise from genetic variations among unrelated patients, it is necessary to compare metastases with matched primary tumors from the same persons. All of these considerations severely restrict the quantity of material available for study. The current work has satisfied all of these demanding criteria, and the resulting collection of pathological specimens was also subjected to rigorous quality control by serial histological sectioning to ensure that they all contained sufficient tumor tissue to provide meaningful information.

These new data show that the expression signatures of orthotopic cancers and their lymphatic metastases are very similar to each other, but that there are a few highly statistically significant differences (Table 2). Previous reports by others in this field have not reinforced their conclusions of

Table 4 – Lists of genes which have the same trend in 10/10 or 9/10 paired samples and satisfy criteria of at least fc > 1.5, $p < 0.05$)

Accession No.	Gene	Fold change	p-value
(a) List of genes which were up-regulated in primary tumors or lymph node metastases in all 10 matched paired samples			
Up-regulated in primary tumor			
NM_001854.1	Collagen, type XI, alpha 1	3.68	0.00634
BF062629	Ras-induced senescence 1	3.65	0.013503
NM_017680.1	Asporin (LRR class 1)	3.19	0.005181
NM_002048.1	Growth arrest-specific 1	2.73	0.002825
AF231124.1	Sparc/osteonectin	2.55	0.000432
NM_013989.1	Deiodinase, iodothyronine, type II	2.33	0.00426
NM_002422.2	Matrix metalloproteinase 3 (stromelysin 1, progelatinase)	2.1	0.001903
Z95331	Fibulin 1	1.77	0.00783
NM_000129.2	Coagulation factor XIII, A1 polypeptide	1.69	0.018433
AA292373	Collagen, type VI, alpha 1	1.51	0.001636
Up-regulated in lymph node metastases			
BC002807.1	Membrane-spanning 4-domains, subfamily A, member 1	1.67	0.031056
U46752.1	Oxidative stress induced like	1.95	0.009057
(b) List of genes which were up-regulated in primary tumors or lymph node metastases in 9 out of 10 matched paired samples			
Up-regulated in primary tumor			
AI885290	Spondin 1, (f-spondin) extracellular matrix protein	2.45	0.028658
NM_002423.2	Matrix metalloproteinase 7 (matrilysin, uterine)	2.36	0.025449
NM_004791.1	Integrin, beta-like 1 (with EGF-like repeat domains)	2.22	0.005629
AU147399	Caveolin 1, caveolae protein, 22 kDa	1.62	0.026766
NM_004530.1	Matrix metalloproteinase 2	1.58	0.006568
M92934.1	Connective tissue growth factor	1.57	0.037712
NM_004684.1	SPARC-like 1 (mast9, hevin)	1.55	0.002578
NM_002345.1	Lumican	1.55	0.045494
AI572079	Snail homolog 2 (Drosophila)	1.55	0.031732
AI762174	Zinc finger protein 42	1.54	0.013105
NM_001038.1	Sodium channel, nonvoltage-gated 1 alpha	1.52	0.04749
Up-regulated in lymph node metastases			
AL559122	T cell receptor beta locus	1.62	0.034135
BC005939.1	Prostaglandin D2 synthase 21kDa (brain)	1.65	0.009066
NM_018468.1	MDS033	1.67	0.028974
NM_000587.1	Complement component 7	2.82	0.013163

similarities in expression between metastases and their primaries by correlation coefficient analysis. Hence there could have been differences in the expression profiles, which failed to reach the level of statistical significance. (Briefly stated, absence of *significant* differences does not prove close similarity.) Some of these genes were significantly differentially regulated in 10 of 10, or 9 of 10 of the paired samples, indicating potential mechanistic importance. However, it is also possible that some of these consistent differences were due to the different organs in which the primaries and metastases are growing. Although great care was taken to remove as much non-malignant host cells and tissue from the tumors as possible, it is important to recognize that tumors are not just balls of malignant cells, but are complex mixtures of tissues including host cell populations and the composite gene expression signature of all these components may be meaningful to tumor behavior. Therefore, although microarrays are useful screening tools, the genes which statistically relate to metastasis will need to be tested extensively in clinical studies analyzing prognostic value and in experiments in animals. This will involve correlations of individual genes with clinical outcome and manipulations to assess the effects of altering the expression of candidate genes by drugs, siRNA and other emerging methods.

Detailed comparisons of the lists of differentially regulated genes published by other laboratories (Hao et al., 2004; Feng et al., 2006; Ramaswamy et al., 2003; van't Veer et al., 2002; Wang et al., 2005) with each other and with our own lists above, revealed very few overlaps in identity between any of them, even though four of those studies also analyzed matched primary and metastatic samples from breast cancer patients (Weigelt et al., 2003; Hao et al., 2004; Feng et al., 2006; Weigelt et al., 2005b). For example, there was no overlap gene between our lists and the 70 gene set of van't Veer et al. (2002), only one overlap with the 17 predictive gene set of Ramaswamy et al. (2003) (collagen 1α2), 3 gene overlaps with the differentially expressed gene lists of Hao et al. (2004) (CD52, integrin β1-like, MMP-2) and no overlap with the 76 gene set of (Wang et al. (2005). As probability values were not provided for the differences claimed in those studies we adopted a value of 2 fc as appropriately cautious for these comparisons. The more recent study by Feng et al. (2006) provided 10 overlaps, 9 genes up-regulated in the primary tumors and 1 in the metastases but, as the authors adopted a general cut-off of 1.5 fc in 14 of 26 cases to identify candidates and did not give details of the fc of individual genes, we are unable to compare their results with ours, in which all candidates were differentially expressed to fc 2.0 and to at least $p < 0.05$ *in 9/10 or 10/10 cases*. It should be noted that almost all these previous studies were done with custom (i.e. non-commercial) cDNA microarrays on glass slides, or on smaller silicon-based oligonucleotide microarrays, which did not include probes for all genes.

Table 5 – Validation of selected genes measured by real-time qPCR

Gene	qPCR fc	Affy fc	qPCR fc
MMP-1	−16.88	−2.84	−302.48
MMP-3	−13.39	−2.1	−4.47
MMP-13	−16.96	−2.95	−207.03
Snail homolog 2	−2.25	−1.55	−1.52
PDGFR	−2.7	−2.35	***9.52***
COL 1A2	−4.07	−1.45	−45.9
COL 11A1	−6.8	−3.68	PT only
ABCG2	−1.55	−1.18	***3.01***
CCL 19	***2.18***	***2.71***	***3.44***
CXCL 13	***11.31***	***3.63***	−3.93
TRB	***1.82***	***1.62***	−1.06
MMP-7	−60.47	***2.36***	−1.46
TGFB-1	−1.28	1.01	−3.72
COL 9A1	−2.46	1.01	PT only
PTGDS	−1.6	***1.89***	1.4
Basigin	−1.26	***1.35***	***1.92***
Integrin beta-like 1	***4.28***	−2.22	−1.84
Testican	1.08	−2.55	PT only
HIF-1 (TF)	−1.18	−1.09	−2.24
E4F-1 (TF)	−1.14	−1.01	−1.73
CETP (TF)	1.09	1.11	***52.53***
Oct #1 (TF)	1.21	−1.01	***2.67***
RREB (TF)	***1.53***	1.04	***5.32***
MZF-1 (TF)	−1.11	1.16	***8.81***

Values indicate fold changes in the corresponding comparison. Negative signs indicate that expression was higher in the primary tumor than in metastases. Bold italicized values were down regulated in primary tumors. ''PT only'' means that expression was measurable only in the primary tumor and a value for fold change cannot be obtained. PT; primary tumor, LN; lymph node.

Also, comparisons of lists in previously published articles with each other showed equivalent or less overlapping genes than the comparisons discussed above.

The difficulties and uncertainties of cross-platform comparisons of data obtained in published studies from different laboratories are well known. Even so, the major discrepancies between the lists which have been reported are surprising and disconcerting. Factors which can contribute in varying degree to these discrepancies among different studies include (a) genetic variation among human individuals, (b) different study designs, for example relating gene expression signatures to prognosis, or to survival or to the presence of metastases, (c) different types of microarrays and methods to analyze the large data sets obtained, (d) the failure of most of the studies to validate even a sample of the results with a different and quantitative method, (e) lack of procedures for quality control of tissue samples, particularly the tumor cell content and (f) the relatively small numbers of samples used in all of these investigations, compared to the large number of genes being analyzed (Jenssen and Hovig, 2005).

Studies using training and test sets of data and/or permutation analysis for evaluating the prognostic significance of their metastasis signature genes have attempted to overcome these difficulties but, as remarked by Jenssen and Hovig (2005), the absence of independent corroboration of the findings of each of the reports by other investigations makes it impossible to decide which of the published signatures might be useful to guide clinical therapeutic decisions, even in clinical trials. We, therefore, concur that such practical applications are premature, although some clinical trials have already begun (Hampton, 2004; Kallioniemi, 2004). Despite these considerations, we believe that rigorously designed and conducted oligonucleotide microarray studies on human tumor tissue samples are useful *screening* procedures that help select candidate genes from the full genetic repertoire, for further analysis. We provide lists of manageable numbers of stringently selected genes (Table 3 and 4) for investigation in prospective clinical studies, where one can look for consistent association of markers with metastasis, or in model systems, where one can intervene to test the effects of altering expression (Montel et al., 2004, 2005a).

The close resemblance of gene expression patterns of clinical primary tumors and their metastases, seen in this and other studies, mirrors the similarities seen in our parallel studies on xenografted human breast cancers (Montel et al., 20005b, 2006; Urquidi et al., 2002) and holds significant implications for the pathogenesis of cancer metastasis. The xenograft data offer a unique vantage point, because, in stark contrast to the clinical tumors and patients, the metastatic human breast cancer line M4A4 is clonal in origin and the animals in which it was disseminating were genetically closely related. Accordingly, the resemblance between primary and secondary tumors in the completely unrelated patients in our current study and in most other published reports provide persuasive evidence that the fundamental biological processes which shape the emergence of the metastatic phenotype have some underlying homologies, although the exact identities of the genes picked out from the large data sets by current computational methods differ in each study.

Microarray data have been interpreted, by some recent studies (Ramaswamy et al., 2003; Bernards and Weinberg, 2002; van de Vijver et al., 2002), to indicate that the metastatic proficiency of a tumor is pre-programmed from its beginning. This concept was based upon the reasoning that the similarity of primary and metastatic tumor expression signatures results from activation of the metastatic genetic program in the earliest progenitors, leading to overgrowth and dominance of the primary and secondary lesions by a phenotypically ''pure'' tumor cell population, with a uniform signature. The hypothesis also reasoned that late emergence of metastatic clones would result in divergent expression patterns between primaries and metastases, because of masking of metastatic signatures in the primary, by persisting non-metastatic clones, not yet eliminated.

However, recent studies of clonal cell lines derived from a late-stage human carcinoma (Urquidi et al., 2002; Bao et al., 1993) have provided direct proof that individual cancer cells, co-existing within a given tumor, differ greatly in metastatic capability and that some of them are indolent or non-metastatic, confirming results from earlier work (Fidler and Kripke, 1977; Kripke et al., 1978; Nicolson et al., 1978; Poste et al., 1981) with a variety of murine tumors. Furthermore, it has recently been shown (Montel et al., 20005b, in press) that the expression signatures of tumors derived from cloned weakly/non-metastatic human cell lines and from their

isogenic metastatic counterparts from the same patient have several differences, although those of metastases and their corresponding primaries are very similar.

Together, these facts establish that metastatic primary cancers, including human examples, are not ''pure'' entities and can still, in later stages, contain many tumor cells of negligible or low metastatic proficiency with different expression profiles. This conclusion is corroborated by cell mixing experiments using genetically marked metastatic and non-metastatic clones (Moffett et al., 1992; Baban et al., 1993), which revealed that the resulting tumors remained mixed populations and that there was no general trend towards development of clonal purity within tumors. Additionally, it has recently been shown that, as the metastatic proficiency of the cell population increases, its expression signature changes concomitantly (Montel et al., 20005b, 2006), demonstrating conclusively that the malignant phenotype and its molecular signature are not predetermined and static but continue to evolve in a tumor throughout its life history.

It is possible that this evolution sometimes occurs very rapidly in the early life of the tumor, but clinical data indicate that this is uncommon. Clinical and pathological observations (Tarin, 1992) show that although occasional human tumors metastasize vigorously when they are small or even undetectable, most do so later, when they are larger, explaining why surgical excision of smaller lesions is often curative. This well recognized relationship between size and metastatic spread is utilized as an important criterion in the TNM classification of tumor stage. We conclude that this collective evidence effectively rebuts the contention advanced by some groups (Weigelt et al., 2003, 2005a; Ramaswamy et al., 2003; Bernards and Weinberg, 2002) on the basis of microarray analysis that metastatic tumors are composed of uniform cell populations which possess this capability from their inception and renders it invalid as a generalization. Incidentally, it should be stated that this notion, that all the tumor cells in a malignant tumor are equally metastatic and acquired this ability from inception, conflicts with clinical common sense, because it is evident that if it were so, there would be no remaining primary tumor, which is not in agreement with observed facts. The data provided in the current communication now provide firm evidence for rejecting the idea.

How can these facts be reconciled with the microarray findings of close similarities between the expression profiles of primary tumors and their metastases? The available data are most consistent with the coexistence of a number of cell clones within the primary, each possessing randomly different parts of the gene expression pattern required to accomplish metastasis, co-evolving to metastatic status, but collectively possessing an ''average'' signature typical of distant metastases. Additionally, the balance of laboratory and clinical evidence now indicates that metastasis by (i) genetic evolution and (ii) an early triggering of metastatic proficiency are not mutually exclusive pathogenetic options, but differing states which relate more to the rate of implementation of the program. This conclusion has considerable practical importance because it reinforces clinical observations that in most patients with small tumors, metastasis is not inevitable and there is often still time for appropriate therapeutic intervention to prevent it.

4. Materials and methods

4.1. *Procurement and selection of clinical breast cancer specimens*

All samples were obtained after informed consent, according to institutional rules, snap frozen and stored in liquid nitrogen immediately after removal. Ten breast cancer samples with their paired lymph node metastases (total of 20 samples) were selected from over 400 specimens from human breast cancer patients in our tumor bank by careful screening of clinical criteria and histopathological features. The chosen samples were confirmed, by frozen sections at multiple levels, to contain sufficient tumor tissue throughout the chosen sample for molecular analysis and to be free of necrotic and fibrous material. Areas of tissue blocks which contained fat, normal breast, lymphocyte aggregates or other inflammatory cells were removed.

It is important to understand that studies on *matched* primary tumors and their metastases from the same patients are rare because of the difficulties of obtaining suitable specimens. This fact also severely limits the numbers of samples assembled for each investigation. Restricted availability of samples results from several factors which we shall briefly review. Modern clinical practice aims to remove tumors before they have disseminated. Therefore, eventual metastases arising after resection of the primary tumor may be removed at different times in different hospitals or not removed at all, if the prognosis is grave. Several other clinical priorities compound these difficulties as follows: if the metastases are small they must be processed intact for pathology to confirm their metastatic identity, leaving none for research. Patients may sometimes deny consent and ethical as well as regulatory issues present additional difficulties. In other cases the tissue sampled is mainly necrotic or benign fibrous or adipose tissue and not tumor. In some of our cases, the patient had tumors in both breasts at different times and we could not compare the metastases with all the metachronous primaries. Consequently, as explained in Section 4, we identified only 30 cases with matched tumor and lymph node specimens in our bank out of over 400 breast cancer cases collected over 15 years. Twenty of these proved to have unsatisfactory material from either the primary or the metastasis, leaving 10 high quality cases. Notably, the few previously published studies had similar small numbers. Hao et al. (2004) had 10 lymph node metastases and Feng et al. (2006) had 26 lymph node metastases. Weigelt et al. (2003, 2005a) had 9 metastases from many different organs, only one being from a lymph node. All of these studies therefore had numbers in the same range as our own report and it is useful to assemble information from different sources.

4.2. *Tissue analysis and RNA preparation*

RNA was extracted from 30 to 40 serial 15 μm cryosections of each tissue sample, which were sufficient to yield 10 μg or

more of total RNA. The first 10 μm frozen section from the block face and a section every 150 μm thereafter were stained with hematoxylin and eosin for pathological assessment. Four observers examined each section and independently assessed the percentage area occupied by the malignant cells. The results were pooled for each piece of tissue and averaged to obtain final values. Only tissue sections with 70% or more tumor tissue were used for analysis (Table 1), to ensure that the results were related to tumor gene expression patterns. Total RNA isolated from each tissue sample using RNeasy Kits (Qiagen, Valencia, CA) was quantitated spectrophotometerically (Eppendorf, Westbury, NY). Its quality was assessed by agarose gel (1%) electrophoresis and by a Biogem analyzer (Agilent, Palo Alto, CA).

4.3. *Amplification and genechip hybridization*

cRNA prepared and labeled in the UCSD Cancer Center Microarray facility, using standard Affymetrix microarray protocols, from each of the tumor and lymph node samples were then hybridized individually to 20 separate human HG-U133A GeneChip oligonucleotide arrays (Affymetrix, Santa Clara, CA). After washing, the arrays were scanned at 560 nm using an argon-ion confocal laser as the excitation source and the intensities of emissions from the probe sets on each chip were analyzed computationally.

4.4. *Microarray data analysis*

The DAT files containing the scanned images of each microarray were individually inspected for quality control and digitized by Microarray Analysis Suite 5.0 (Affymetrix). The resultant CEL files containing the raw signal intensities were read and analyzed in dChip software (Li and Wong, 2001a,b) Briefly, each microarray was normalized against a common baseline array using the ''invariant probe set'' method. After normalization, the model-based expression index of each gene was then calculated according to the PM-MM model.

To identify genes that were differentially expressed between any two groups of arrays, an initial screening filter consisting of the following criteria was applied: (i) a fold change (fc) larger than 1.2, (ii) a two-tailed p-value (paired t-test) smaller than 0.05 and (iii) a minimal difference of 100 between the group mean normalized expression intensities. Further iterations of screening with criteria of $p < 0.01$, 0.005, 0.001 and $fc > 1.5$, 2.0 and 3.0 were also conducted to evaluate differential expression between the paired samples. As we consider reproducibility to be more important than fold change as a sorting criterion, the resulting lists of candidate genes were first selected according to p-value and then ranked by fc. Tables of the numbers of genes satisfying these criteria were then assembled as shown (Table 2 a and b).

4.5. *Quantitative real-time reverse transcriptase PCR (qRT-PCR)*

mRNA from the 10 samples in each group (primary tumors or metastases) was analyzed by qRT-PCR. It was reverse transcribed with the Ambion Retroscript cDNA synthesis system using standard protocols. The amplification reactions were conducted in 96-well plates as described previously (Montel et al., 2005b). Each sample was analyzed in triplicate, and the expression of each gene was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression as an internal standard. Microarray data showed that this gene was uniformly expressed across all samples.

Acknowledgments

This work was supported by the Loppicola Foundation whose support is gratefully acknowledged. We thank Dr Linda Wasserman for discussions and help with statistical analysis. We also thank Ms. E Mose for technical assistance, the UCSD Cancer Center Microarray facility for GeneChip hybridizations and the Molecular Pathology Shared Resource for assistance with qPCR analysis.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molonc.2007.03.005.

REFERENCES

Baban, D., Matsumura, Y., Kocialkowski, S., Tarin, D., 1993. Studies on relationships between metastatic and non-metastatic tumor cell populations using lineages labeled with dominant selectable genetic markers. Int. J. Dev. Biol. 37 (1), 237–243.

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 (3), 239–246.

Bernards, R., Weinberg, R.A., 2002. A progression puzzle. Nature 418 (6900), 823.

Bloom, H.J., Richardson, W.W., 1957. Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. Br. J. Cancer 11 (3), 359–377.

Feng, Y., Sun, B., Li, X., et al., 2006. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. Breast Cancer Res. Treat, doi:10.1007/s10549-006-9385-7. Available from: www.springerlink.com (PMID number 17123152).

Fidler, I.J., Kripke, M.L., 1977. Metastasis results from preexisting variant cells within a malignant tumor. Science 197 (4306), 893–895.

Hampton, T., 2004. Breast cancer gene chip study under way: can new technology help predict treatment success? JAMA 291 (24), 2927–2930.

Hao, X., Sun, B., Hu, L., et al., 2004. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. Cancer 100 (6), 1110–1122.

Jenssen, T.K., Hovig, E., 2005. Gene-expression profiling in breast cancer. Lancet 365 (9460), 634–635.

Kallioniemi, O., 2004. Medicine: profile of a tumour. Nature 428 (6981), 379–382.

Kripke, M.L., Gruys, E., Fidler, I.J., 1978. Metastatic heterogeneity of cells from an ultraviolet light-induced murine fibrosarcoma of recent origin. Cancer Res. 38 (9), 2962–2967.

Lahdesmaki, H., Hao, X., Sun, B., et al., 2004. Distinguishing key biological pathways between primary breast cancers and their lymph node metastases by gene function-based clustering analysis. Int. J. Oncol. 24 (6), 1589–1596.

Li, C., Wong, W.H., 2001a. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. USA 98 (1), 31–36.

Li, C., Wong, W.H., 2001b. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol. 2 (RESEARCH 0032, Epub).

Moffett, B.F., Baban, D., Bao, L., Tarin, D., 1992. Fate of clonal lineages during neoplasia and metastasis studied with an incorporated genetic marker. Cancer Res. 52 (7), 1737–1743.

Montel, V., Pestonjamasp, K., Mose, E., Tarin, D., 2005a. Tumor-host interactions contribute to the elevated expression level of alpha1-antichymotrypsin in metastatic breast tumor xenografts. Differentiation 73 (2–3), 88–98.

Montel, V., Huang, T.Y., Mose, E., Pestonjamasp, K., Tarin, D., 2005b. Expression profiling of primary tumors and matched lymphatic and lung metastases in a xenogeneic breast cancer model. Am. J. Pathol. 166 (5), 1565–1579.

Montel, V., Kleeman, J., Agarwal, D., Spinella, D., Kawai, K., Tarin, D., 2004. Altered metastatic behavior of human breast cancer cells after experimental manipulation of matrix metalloproteinase 8 gene expression. Cancer Res. 64 (5), 1687–1694.

Montel, V., Mose, E.S., Tarin, D, 2006. Tumor–stromal interactions reciprocally modulate gene expression patterns during carcinogenesis and metastasis. Int. J. Cancer 119, 251–263.

Nicolson, G.L., Brunson, K.W., Fidler, I.J., 1978. Specificity of arrest, survival, and growth of selected metastatic variant cell lines. Cancer Res. 38 (11 Pt 2), 4105–4111.

Paget, S., 1889. The distribution of secondary growths in cancer of the breast. Lancet 1, 571–573.

Poste, G., Doll, J., Fidler, I.J., 1981. Interactions among clonal subpopulations affect stability of the metastatic phenotype in polyclonal populations of B16 melanoma cells. Proc. Natl. Acad. Sci. USA 78 (10), 6226–6230.

Ramaswamy, S., Ross, K.N., Lander, E.S., Golub, T.R., 2003. A molecular signature of metastasis in primary solid tumors. Nat. Genet. 33 (1), 49–54.

Stuart, R.O., Wachsman, W., Berry, C.C., et al., 2004. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc. Natl. Acad. Sci. USA 101 (2), 615–620.

Tarin, D., 1992. Tumor Metastasis. In: McGee, J., Isaacson, P.G., Wright, N. (Eds.), Principles of Pathology. Oxford Textbook of Pathology, vol. 1. Oxford University Press, Oxford, pp. 607–633.

Tarin, D., Price, J.E., 1981. Influence of microenvironment and vascular anatomy on ''metastatic'' colonization potential of mammary tumors. Cancer Res. 41 (9 Pt 1), 3604–3609.

Tarin, D., Price, J.E., Kettlewell, M.G., Souter, R.G., Vass, A.C., Crossley, B., 1984. Mechanisms of human tumor metastasis studied in patients with peritoneovenous shunts. Cancer Res. 44 (8), 3584–3592.

Urquidi, V., Sloan, D., Kawai, K., et al., 2002. Contrasting expression of thrombospondin-1 and osteopontin correlates with absence or presence of metastatic phenotype in an isogenic model of spontaneous human breast cancer metastasis. Clin. Cancer Res. 8 (1), 61–74.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., et al., 2002. A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347 (25), 1999–2009.

van't Veer, L.J., Dai, H., van de Vijver, M.J., et al., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536.

Wang, Y., Klijn, J.G., Zhang, Y., et al., 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365 (9460), 671–679.

Weigelt, B., Glas, A.M., Wessels, L.F., Witteveen, A.T., Peterse, J.L., van't Veer, L.J., 2003. Gene expression profiles of primary breast tumors maintained in distant metastases. Proc. Natl. Acad. Sci. USA 100 (26), 15901–15905.

Weigelt, B., Wessels, L.F., Bosma, A.J., et al., 2005a. No common denominator for breast cancer lymph node metastasis. Br. J. Cancer 93 (8), 924–932.

Weigelt, B., Hu, Z., He, X., et al., 2005b. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res. 65 (20), 9155–9158.