DOI 10.1002/pmic.200800303

Proteomics 2008, *8,* 5038–5052

RESEARCH ARTICLE

Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis

Jian Li[1, 2, **], Pavel Gromov[3*], Irina Gromova[3], José M. Moreira[3], Vera Timmermans-Wielenga[4], Fritz Rank[4], Kai Wang[1, 2, 5], Shengting Li[1, 2], Heng Li[1, 2], Carsten Wiuf[5], Huanming Yang[2], Xiuqing Zhang[2***], Lars Bolund[1, 2***] and Julio E. Celis[3***]*

[1] Institute of Human Genetics, University of Aarhus, Aarhus C, Denmark
[2] Beijing Institute of Genomics, Chinese Academy of Sciences, B-6, Beijing, China
[3] Department of Proteomics in Cancer, Institute of Cancer Biology, Danish Cancer Society, Danish Centre for Translational Breast Cancer Research (DCTB), Copenhagen, Denmark
[4] Department of Pathology, The Centre of Diagnostic Investigations, Rigshospitalet afsnit 5442, Copenhagen, Denmark
[5] Bioinformatics Research Center (BiRC), University of Aarhus, Aarhus, Denmark

Received: April 4, 2008
Revised: June 10, 2008
Accepted: June 10, 2008

Axillary lymph node (ALN) status is currently used as an important clinical indicator of breast cancer prognosis. However, the molecular mechanisms underlying lymph node metastasis are poorly understood and the relationship between ALN metastasis and the primary tumor remains unclear. In an effort to reveal structural changes in the genome and related protein responses that may drive regional metastatic progression we have analyzed matched pairs of primary breast tumors and ALN metastases both at the genomic and proteomic levels using comparative genomic hybridization (CGH) array, quantitative high-resolution 2-D PAGE in combination with MS, and immunohistochemistry (IHC). Array CGH revealed a remarkable similarity in genomic aberration profiles between the matched primary tumors and the ALN metastases. Quantitative profiling of 135 known proteins also revealed striking similarities in their overall expression patterns, although we observed distinct changes in the levels of individual proteins in some sample pairs. The remarkable similarities of the overall genomic and proteomic profiles between primary tumors and matched ALN metastases are taken to suggest that, in general, key biological characteristics of the primary breast tumor are maintained in the corresponding lymph node metastases. Given that the omics-based technologies are oblivious to changes that only occur in minor cellular subsets we validated the proteomic data using IHC, which provides protein expression information with a valuable topological component. Besides confirming the omics-derived data, the IHC analysis revealed that in two cases the ALN metastases may have been derived from a distinct minor cell subpopulation present in the primary tumor rather than from the bulk of it.

Keywords:
Breast cancer / Genomics / Immunohistochemistry / Primary tumor and lymph node metastases

Correspondence: Dr. Pavel Gromov, Department of Proteomics in Cancer, Institute of Cancer Biology, Danish Cancer Society, Danish Centre for Translational Breast Cancer Research (DCTB), Strandboulevarden 49, DK-2100 Copenhagen, Denmark
E-mail: psg@cancer.dk
Fax: +45-35257721

Abbreviations: **ALN**, axillary lymph node; **BAC**, bacterial artificial chromosome; **CBS**, circular binary segmentation; **CGH**, comparative genomic hybridization; **ER**, estrogen receptor; **Her2**, epidermal growth factor receptor-2; **IHC**, immunohistochemistry; **PgR**, progesterone-receptor

1 Introduction

Secondary site metastasis is the major cause of morbidity and mortality associated with breast cancer and the lymphatic system is generally regarded as the primary route for the metastasis of many cancers [1]. In primary breast cancer,

* Jian Li and Pavel Gromov contributed equally to this work.
** Additional corresponding author: Dr. Jian Li, E-mail: jianl1977@humgen.au.dk
*** Xiuqing Zhang, Lars Bolund and Julio E. Celis are equal principal authors.

ALN involvement is an important prognostic parameter used for clinical decision-making [2]. Epidemiological studies have shown that patients who are ALN negative have better prognosis as compared with those who are ALN positive as judged by the 10-year recurrence rate, the 5-year survival, and the median survival after relapse [3–5]. Subsequent studies have further shown that the number of affected lymph nodes is an important selection criterion for adjuvant therapy [6]. However, approximately one-third of the patients with ALN-negative status develop widespread metastases, whereas about one-third of the ALN-positive patients remain free of distant metastases after 10 years [7, 8]. The lack of absolute correlation between ALN status and eventual development of distant metastasis suggests that the molecular mechanisms leading to hematogenous (distant) metastases might be distinct from those leading to lymphogenous (regional) spread of tumor cells [9].

Previous studies of matched pairs of primary tumors and lymph node metastases have shown similar phenotypes for the matched lesions as judged both by histological and proliferation studies [10–12]. Similar comparisons have also been performed at the transcriptomic level using expression microarrays [9, 13–15], but the outcome of these studies has been somewhat controversial. To complicate matters further, results from a chromosomal CGH analysis revealed differences between the primary tumors and their matched lymph node metastases, although the number of samples in this study was small and the resolution of chromosomal CGH was limited [16].

Recently, as part of a long-term ongoing effort in breast cancer we have started to apply various "omic" technologies to the study of fresh tissue biopsies followed by the integration of multiplatform datasets collected from the same patients [17–20]. Here, we report a study in which we have analyzed 29 pairs of well-matched primary tumors and ALN metastases using 1-Mb resolution whole-genome BAC CGH arrays. Thirteen of these pairs were also analyzed using quantitative high-resolution 2-D PAGE in combination with MS, and IHC for validation of results in paraffin-embedded formalin-fixed tissue sections. The combined use of genomic and proteomic methods allowed us to compare for the first time genomic aberrations and proteomic profiles for individual patients.

2 Materials and methods

2.1 Sample collection and handling

Matched pairs of primary tumors and ALN metastases were obtained from 29 patients diagnosed with high-risk breast carcinomas at the Copenhagen University Hospital during the period between February 2004 and September 2005. Clinicopathological information for the individual cases is given in Table 1. None of the patients had received any treatment prior to the sample collection. The project was approved by the Scientific and Ethical Committee of the Copenhagen and Frederiksberg Municipalities (KF 01-069/03). Following surgery, fresh tissue samples were immediately dissected into three blocks; one was placed in formalin fixative and embedded in paraffin for IHC analysis and archival use, while the others were snap-frozen in liquid nitrogen and stored at –80°C for CGH, gel-based proteomics and IHC studies (see below). Sections of the paraffin-embedded specimen from each breast tissue sample were used for IHC analysis using cytokeratin 19 (CK19) antibodies, a marker that is ubiquitously expressed by mammary epithelial cells [21]. The histological images greatly facilitated the interpretation of array CGH and protein gel data as they provided with a rough estimate of the ratio of gland cells to stromal tissue (see Section 3). In some cases, the first and last sections of frozen tissue blocks taken for CGH and/or proteomic studies were used for immunofluorescence analysis using CK19 antibodies. Both the tumor and ALN metastasis samples selected for the study contained more than 60% of tumor cells.

2.2 2-DE and image analysis

High-resolution 2-D PAGE IEF was performed as previously described [22]. Twenty to thirty, 6-μm cryostat sections of frozen specimen were resuspended in 0.1 mL lysis solution [23] and were kept at –20°C until used [19]; 40 μL was applied to the gels. After running, the gels were fixed in 7.5% acetic acid, 50% ethanol, 0.05% formalin for 1 h, stained with Sypro Ruby Protein Gel Stain (Molecular Probes) overnight, rinsed twice briefly in 7.5% acetic acid/ 10% ethanol, and destained in the same solution for 30 min. Imaging was performed using a Typhoon 9410 image system equipped with a 457-nm laser and 580-nm filter. Fluorescent intensities of the scanned 2-D images were analyzed using the PDQUEST software v7.3 from Bio-Rad. All 26 images (primary tumors and ALN from 13 patients) were analyzed in a single matchset. The albumin spot was excluded from the matchset due to intensity saturation. Well-resolved protein spots (135) were chosen for quantitation and were matched for each set of 26 images. The only criteria for choosing the protein spots were: good resolution of the area in which the protein migrates and sufficient amounts in a single spot for MS identification. No special bias was introduced concerning the selection of this group of proteins. The spot volumes were normalized to the total density of detected spots on the image.

2.3 Protein identification by MS

After scanning, the gels were placed in 7.5% acetic acid, 50% ethanol, 0.05% formalin for 1 h, washed 3×30 min in 7.5% acetic acid, 10% ethanol and stained with silver nitrate according to procedures compatible with MS [24]. Protein spots were excised from dry gels and the gel pieces were rehydrated in water. Gel pieces were detached from the cello-

Table 1. Patient/sample information

Patient number	Patient code[a)]	Method[b)]	Age	Type[c)]	Size (mm)[d)]	Grade	Her2[e)]	Her2-FISH	ALN status[f)]	ER/PgR status[g)]	
1	51	CGH/2-D/IHC	27	D	40	3	2+	1.49	N+ 4/26	ER+	PgR–
2	53	CGH/2-D/IHC	51	D	21	1	1+		N+ 2/11	ER+	PgR+
3	57	CGH/2-D/IHC	66	D	30	3	3+		N+ 27/31	ER–	PgR–
4	60	CGH/2-D/IHC	72	D	30	2	2+	1.2	N+ 3/14	ER+	PgR–
5	70	CGH/2-D/IHC	83	D	33	3	3+		N+ 3/11	ER–	PgR–
6	74	CGH/2-D/IHC	49	D	21	3	1+		N+ 3/22	ER–	PgR–
7	84	CGH/2-D/IHC	38	D	18	2	0		N+ 3/11	ER+	PgR–
8	85	CGH/2-D/IHC	80	D	30	2	2+	1.69	N+ 3/16	ER+	PgR–
9	86	CGH/2-D/IHC	77	D	110	2	1+		N+ 20/20	ER+	PgR–
10	88	CGH/2-D/IHC	83	D	35	2	2+	1.48	N+ 8/15	ER+	PgR+
11	91	CGH/2-D/IHC	59	D	30	2	3+		N+ 13/16	ER+	PgR+/–
12	93	CGH/2-D/IHC	60	D	35	2	1+		N+ 10/16	ER+	PgR+
13	94	CGH/2-D/IHC	57	D	21	3	3+		N+ 3/13	ER–	PgR–
14	78	CGH/IHC	31	D	32	3	0		N+ 14/17	ER–	PgR–
15	47	CGH	47	D	20	2	2+	2.21	N+ 4/12	ER+	PgR+
16	49	CGH	78	D	35	2	2+	1.65	N+ 6/7	ER+	PgR–
17	50	CGH	46	D	26	2	3+		N+ 2/10	ER+	PgR+
18	55	CGH	46	D	50	2	0		N+ 4/22	ER+	PgR+
19	58	CGH	38	D	20	3	3+		N+ 9/20	ER–	PgR–
20	61	CGH	99	D	40	2	2+	1.5	N+ 7/11	ER+	PgR+
21	64	CGH	52	Tu/Kr	23	1	1+		N+ 23/25	ER+	PgR+
22	69	CGH	86	D	25	2	0		N+ 3/15	ER+	PgR+
23	75	CGH	62	D	20	3			N+ 9/14	ER–	PgR–
24	77	CGH	62	L	30	2	2+	1.39	N+ 11/17	ER+	PgR–
25	82	CGH	72	D	25	2	2+	2.18	N+ 1/15	ER+	PgR+
26	83	CGH	57	D	45	2	2+	1.19	N+ 10/16	ER+	PgR+
27	90	CGH	50	D	40	3	3+		N+ 15/16	ER–	PgR–
28	95	CGH	44	L	25	1	1+		N+ 4/11	ER+	PgR+
29	98	CGH	51	D	45	2	1+		N+ 2/14	ER+	PgR+

a) The column represents internal DCTB patient anonymization number codes that we have used in all our breast cancer studies and publications.
b) Method by which the sample was analyzed.
c) D, ductal carcinoma; L, lobular carcinoma.
d) Tumor size (mm) was determined by the diameter of the primary tumor.
e) Determined using Hercep Test™ (DAKO).
f) Metastastic ALN number/total ALN number.
g) ER, estrogen receptor; PgR, progesterone receptor.

phane film and cut into 1-mm^2 pieces followed by "in-gel" digestion as previously described [17]. Samples were prepared for analysis by applying 2 μL of digested and extracted peptides on the surface of a 400/384 AnchorChip target (Bruker Daltonik), followed by co-crystallization with CHCA matrix [25]. MS was performed using a Reflex IV MALDI-TOF MS equipped with a Scout 384 ion source. All spectra were obtained in positive reflector mode with delayed extraction using an acceleration voltage of 28 kV. The resulting mass spectra were internally calibrated using the auto-digested tryptic mass values visible in all the spectra. Calibrated spectra were processed by the Xmass 5.1.1 and BioTools 2.1 software packages (Bruker Daltonik). Irrelevant masses (matrix, metal adducts, auto-digested tryptic masses as well as masses of tryptic peptides from keratins) were excluded from the analysis by manual examination of all spectra by pair wise comparison. The spectrum of interest was superimposed with the spectrum of the negative control (set of peptides from a non-stained gel piece treated in parallel) to exclude the most common contaminations. For protein identification, peptide masses were transferred to the BioTools 2.0 interface (Bruker Daltonics) to search in the National Center for Biotechnology non-redundant NCBInr (version 22.09.2007, mammalian entries) database using the MASCOT search engine (version 2.2, released 28.02.2007, Matrix Science). No restriction on the protein molecular mass and taxonomy was applied. A number of fixed (acrylamide modified cysteine, *i.e.* propionamide/carbamidomethylation) and variable modifications (methionine oxidation and protein N terminus acetylation) were included in

the search parameters. The peptide tolerance did not exceed 50 ppm and a maximum of one trypsin missed cleavage was allowed. Protein identifications were considered to be confident when the protein score of the hit exceeded the threshold significance score of 70 ($p < 0.05$) and no less than six peptides were recognized. Whenever the protein score hit was close to the threshold significance score of 70, PSD was performed as an additional means to confirm the identity of the proteins identified by PTM. The following PSD search parameters were used: peptide tolerance 50 ppm and MS/MS tolerance 1 Da without any restriction on the protein molecular mass and taxonomy. Since the amount of peptides extracted from the silver stained gels did not yield overall peak intensities high enough to allow multiple peptide sequencing (a pre-requirement for conclusive PSD analysis), the identification of proteins was never made solely based on PSD analysis. Often, the peptides identified matched equally well to multiple database entries using the non-redundant NCBInr database and that is why the second/final search was performed using the same parameters, but using the UniProtKB/Swiss-Prot 54.2 (17 252 human entries) database. If the number and the sequence of the recognized peptides were identical to the first search, the Swiss-Prot accession number was assigned to the identified protein. In a few cases where protein score hit was close to the threshold significance score of 70 the spot identity was confirmed by Western blotting using specific antibodies or by PSD analysis (Supporting Information Table 1). The information of 135 identified proteins is presented in Supporting Information Table 1.

2.4 Antibodies

Polyclonal antibodies against acyl-CoA dehydrogenase, alpha glycosidase (lysosomal) and prohibitin were prepared by Eurogentec (Seraing, Belgium). The mAb against p63 (clone 4A4), smooth muscle alpha actin (SMA; clone 1A4), androgen receptor (AR; clone AR441), estrogen receptor alpha (ER; clone 1D5), progesterone receptor (clone PgR636), cyclooxygenase 2 (COX-2), E-cadherin, Glut.1, and Ki67 (clone MIB-1) were purchased from DakoCytomation (Glostrup, Denmark). Antibodies recognizing CK 14, 17 and 19, cyclin D1 and S100 A6 were from NeoMarkers (LABVISION, CA). CK18 was obtained from Cappel (Organon Teknika, ICN). The mAb against CK 7 (clone RCK105) and 8 (clone M20) were purchased from MP Biomedicals (Irvine, CA). GATA-3 was purchased from Santa Cruz Biotechnology. The mAb against psoriasin (S100 A7) has been previously described [26]. The antibody against mts 1 (S100 A4) was obtained from Prolifia. The specificity of the antibodies was determined by 2-D PAGE immunoblotting as described elsewhere [27].

2.5 Immunohistochemistry

Following surgery, fresh tissue blocks were immediately placed in formalin fixative and paraffin embedded for archival use. Six-micrometer sections were cut from the paraffin embedded tissue blocks and mounted on Super Frost Plus slides (Menzel-Gläser, Braunschweig, Germany), baked at 60°C for 60 min, deparaffinized, and rehydrated through graded alcohol rinses [28]. Heat induced antigen retrieval was performed by immersing slides in 10 mM citrate buffer (pH 6.0) and microwaving in a 750-W microwave oven for 10 min. The slides were then cooled at room temperature for 20 min and rinsed abundantly in tap water. Nonspecific staining of slides was blocked (10% normal goat serum in PBS buffer) for 15 min, and endogenous peroxidase activity was quenched using 0.3% H_2O_2 in methanol for 30 min. Antigen was detected with a relevant primary antibody followed by a suitable secondary antibody conjugated to a peroxidase complex (HRP-conjugated goat anti-rabbit or anti-mouse antibody; DakoCytomation (Glostrup, Denmark). Finally, color development was done with 3, 3'- diaminobenzidine (Pierce, IL) as a chromogen to detect bound antibody complex. Slides were counterstained with hematoxylin. Standardization of the dilution, incubation, and development times appropriate for each antibody allowed an accurate comparison of expression levels in all cases. At least three independent staining of the samples were performed for each antibody. Sections were imaged using either a standard bright field microscope (Leica DMRB) equipped with a high-resolution digital camera (Leica DC500), or a motorized digital microscope (Leica DM6000B) controlled by Objective Imaging's Surveyor Software (Objective Imaging, UK) for automated scanning and imaging, which enables tiled mosaic image creation. Original magnification for all images was 200x.

2.6 Immunofluorescence on paraffin sections

The 5-μm sections were cut from paraffin blocks of breast tissue samples mounted on Super Frost Plus slides (Menzel-Gläser, Braunschweig, Germany), baked at 60°C for 60 min, deparaffinized, and rehydrated through graded alcohol rinses. Heat-induced antigen retrieval as well as additional steps were carried out as described above. Antigens were detected by overnight incubation at 4°C with primary antibodies at the appropriate dilution followed by detection with species-specific secondary antibodies conjugated to Alexa Fluor® 488 and Alexa Fluor® 594 (Molecular Probes, OR). Sections were imaged using confocal laser scanning microscopy (Zeiss 510LSM).

2.7 Isolation of DNA

Whole genomic DNA was isolated using a DNA isolation kit (NucleoSpin® Tissue, MACHEREY-NAGEL, France) following the manufacturer's instructions. Reference DNA was obtained from the peripheral blood of a healthy male.

2.8 Array-based CGH

Arrays representing the whole human genome with 1-Mb resolution were produced based on Linker-adaptor PCR amplification of DNA [29] from a library of mapped and vali-

dated bacterial artificial chromosome (BAC) clones obtained from the Wellcome Trust Sanger Institute. The clones were spaced at approximately 1-Mb intervals across the genome. DNA for the array elements was isolated from the BAC clones and amplified by linker-adapter PCR. PCR products were suspended in printing buffer (150 mM sodium phosphate at pH 8.5) and were spotted onto the slides ("Codelink" slide, Amersham Biosciences) using our custom-built capillary-tube based printer. Each clone was spotted on the slides in neighboring triplicate pattern. Four drosophila sequences linker adaptor PCR amplified products were printed on the slides as control using the same procedure as for human clones. The clones' annotation was based on the "Wellcome Trust Sanger Institute" published 1-Mb clone information and modified on the basis of the updated 38_36 version of the 1-Mb clone information released by "Ensembl".

2.9 Hybridization

Tumor samples and reference DNA were labeled by a random priming method with the labeling kit BioPrime® DNA labeling system (Invitrogen). Cy3-dCTP and Cy5-dCTP (Amersham Biosciences) were used for labeling tumor samples and reference, respectively. The mixture of the labeled tumor sample and reference DNA together with 40 μL (1 mg/mL) Cot-1 DNA (Invitrogen), which suppresses hybridization to repetitive sequences, were ethanol precipitated and resuspended in hybridization mix (50% formamide, 10% dextran sulfate, 2 × SSC, 1%–4% v/v SDS) in a total volume of 36 μL. After being denatured at 75°C for 15 min, the mixture of DNA was incubated at 37°C for 1.5 h to allow reassociation of the repetitive sequences. The hybridization mix was added to the array, and a cover glass (LifterSlip, Erie Scientific Company) was applied to avoid evaporation during hybridization. The array was placed on a slowly rocking table at 37°C for 48–60 h [29]. After hybridization, the slide was rapidly washed with PN buffer (0.1 M sodium phosphate; 0.1% nonidet P40; pH8). The slide was then incubated in 50% formamide/2 × SSC at 45°C for 15 min followed by a final wash at room temperature in PN buffer for an additional 15 min. Finally, the slide was dipped into 0.1 × SSC very quickly and spin-dried.

2.10 Digital image analysis

The arrays were imaged in a Charge-Coupled Device (CCD) arrayWorx[e] scanner (Applied Precision). After optimizing the exposure time, the arrays were scanned at Cy3 and Cy5 channels, respectively. The two single-channel 16-bit images were combined for analysis by the image analysis software "Tracker" (Applied Precision Company).

2.11 Statistical filter for raw data

Spots were excluded from analysis in the following situations: (i) the spots were labeled "undetected" due to low signal and/or high background with "Tracker" analysis, (ii) the clone was inconsistent compared with the updated 38_36 version of 1-Mb resolution clone set information released by "Ensembl" (data not shown), (iii) the spot's Cy5 (reference) intensity was less than twice that of the background intensity's SD, (iv) the spot's Cy5 intensity was lower than the quartile of control *Drosophila* DNA Cy5, (v) only a single spot of the triplicate was left after filtering with the above criteria, (vi) CV of the triplicate or duplicate Cy3/Cy5 ratios of spots representing one clone exceeded 0.08, (vii) clones were mapped to chromosome Y.

2.12 Determination of copy number changes

After filtering, clones representing the same DNA sequence were averaged and subjected to base 2 log transformation. Data were then sent to "DNAcopy" R/Bioconductor package, which performs Circular Binary Segmentation (CBS) [30]. The noise intensity measurements were translated into regions of equal copy number, which means missing values for clones mapping within segmented regions of equal copy number were assigned the value of the corresponding segment. If a few clones with missing values were located between segmented regions, their values were set to the normalized maximum absolute value of the two flanking segments. The output clone segments from "DNAcopy" were merged using MergeLevel procedure [31]. In this process, segmental values across the genome were merged to create a common set of copy number levels for each individual tumor sample. The segments corresponding to the copy number level with the smallest absolute median value were declared unchanged. Additionally, to account for some complex genomic aberration patterns, the normal levels of paired samples were determined by considering the paired samples together. All the segments for each sample were then normalized by subtracting their corresponding normal level values. In this way, the normal level value would be 0 (log transformation based 2 scale). Because all breast cancer patients examined in this study were female and we employed a healthy male as reference, the ratio of clones located on chrX would be 2 in theory, if no DNA copy number changed. We plotted a serial of chrX DNA copy number dosage (one copy, two copies, three copies, five copies) and the copy number changes presented a linear relationship with the corresponding ratios (log based 2 transformed). The slope was calculated to be 0.6049 (data not shown). We subtracted 0.6049 from all clones on chrX to shift them to the same level as the autosomal clones.

Cell lines with known copy number gains and losses in several chromosomes were used to establish mean log based 2 transformed values for single gain and loss: 0.1575 and −0.1815, respectively. Briefly, in order to determine the thresholds of DNA copy number aberrations we employed the normal references, 0.5 ratio (the ratio values of the clones located on chrX for the hybridization of the normal males

versus normal females), 1.5 ratio (the ratio values of the clones located on chr13 for the hybridization of trisomy 47,XX+13 *versus* the normal reference, and chr18 for the hybridization of 47,XX+18 *versus* the normal reference), 2 ratios (the ratio values of clones located on chrX for the hybridization of the normal females *versus* the normal males) as well as 2.5 ratios (the ratio values of clones located on chrX for the hybridization of 49,XXXXX *versus* 46,XX) reference. All the above references ratio values have been analyzed statistically and find out the range of percentile 5 and percentile 95 of each reference group, in which 90% ratio values distributed. Then we compared the 90% ratio values distribution from the normal people hybridizations with the 90% distributions both from the 0.5 ratio value (single loss) and 1.5 ratio value (single gain) hybridizations. There was no overlap among their 90% distribution ranges (data not shown). We believe that the variance of the normal reference from the theoretical value 0 (log transformed) is derived from the experimental random errors that were similar in all the experimental outcomes. In order to get the relative strict thresholds and to decrease the rate of false positive aberrations, we used the normal reference percentile 1 (–0.1815 (log-transformed)) and percentile 99 (0.1575 (log-transformed)) as thresholds to determine the aberrations in breast cancer samples. In short, the ratio values of clones within the percentile 1 and percentile 99 were regarded as no DNA copy number change, below the percentile 1 as loss, and above the percentile 99 as gain.

2.13 Unsupervised clustering and bootstrapping methodology

An unsupervised hierarchical clustering analysis was applied for the genomic aberrations and proteomic pattern similarities across samples using the "Cluster 3.0" software [32]. Pearson correlation (uncentered) was carried out for similarity metrics computation for array CGH data, whereas a non-parametric distance measurement algorithm, Spearman rank correlation, was used for proteomic data due to the large differences in the expression values of different functional classes of expressed proteins. Complete linkage clustering was chosen to organize samples into a tree structure. The "TreeView" software was utilized for visualization of the cluster analysis result [32].

To estimate the robustness of the clustering result, a "Bootstrapping" method was applied. Briefly, the same filter criteria as in the previous unsupervised clustering were set up. Then 1000 new trees were created on the basis of 1000 random subsets derived from the whole dataset after filtering. In the subset, we compensated for missing observations (clones or proteins) by re-sampling other observations in the dataset, thus keeping the number of observations the same as in the whole dataset. Thereafter, the frequency of each node from 1000 random trees was accounted and labeled in the corresponding node in the original tree by using the "CONSENSE" software [33]. In the consensus tree structure, every node was labeled with a number indicating the frequency of this node reoccurring in 1000 iterations.

3 Results

3.1 Genomic aberration profiles of primary breast carcinomas and ALN metastases revealed by array CGH

We identified some frequent genetic aberrations in both primary tumors and ALN metastases using 1-Mb resolution BAC clone array CGH; these included gains on 1q23.3-44 (44%) and 8q11.22-24.3 (52%), and losses on 16q21-23.2 (42%), 17p13.3-11.2 (54%) and 22q11.23-13.33 (53%), very much in line with previously published studies [34] (see Fig. 1 and Table 2). The genomic profiles for each of the samples analyzed in this study are presented in Supporting Information Fig. 1.

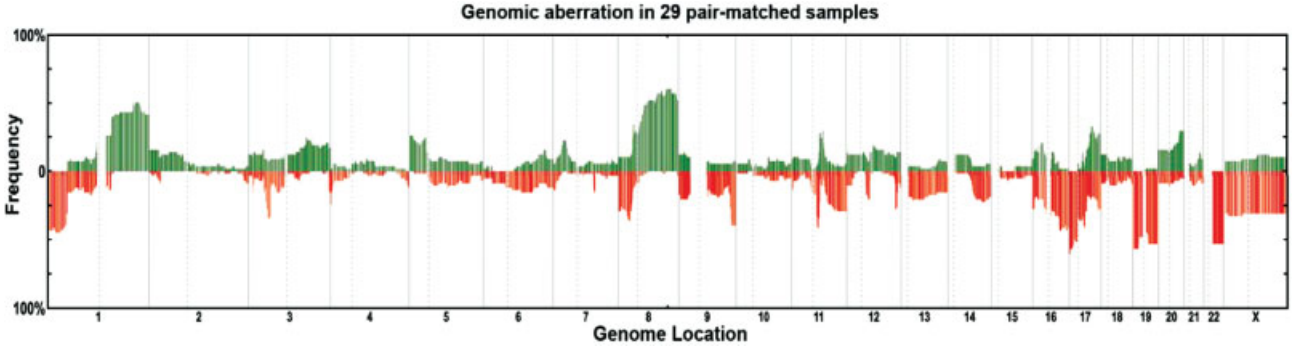

Figure 1. Genomic abnormalities in 29 pairs of matched primary breast tumors and ALN metastases revealed by array CGH. Frequencies of genome copy number gains and losses were plotted as a function of genome location with chromosomes 1pter to the left and chromosomes 22qter and X to the right. The vertical lines indicate chromosome boundaries while the dashed lines indicate centromere locations. Green and red columns indicate frequencies of tumors showing copy number gains and losses, respectively.

Table 2. The summary of the reoccurring genomic aberrations in 30, 40, 50 and 60% of the 29 pair-matched samples

Chr.	Start	End	Aberration	Percentage	Start clone	End clone	Cytogenetic loci
Reoccurring aberration in 30% of the 29 pair-matched samples							
1	5866184	47041895	Loss	42%	bA49J3	bA8J9	1p36.31-p33
1	156056126	246181016	Gain	43%	bA206M24	bA438H8	1q23.1-q44
3	47384084	52743025	Loss	34%	dJ544D10	bA447A21	3p21.31-p21.1
8	21698622	28631547	Loss	35%	bA529P14	bA356F24	8p21.3-p21.1
8	37204481	38488992	Gain	34%	bA197P20	bA350N15	8p12
8	50882181	146167102	Gain	52%	bA401N18	dJ1056B24	8q11.22-q24.3
9	128256298	139715973	Loss	39%	bA205K6	bA417A4	9q33.3-q34.3
11	60851894	67554176	Loss	38%	bA286N22	dJ901A4	11q12.2-q13.2
16	33544449	45688066	Loss	32%	bA274A17	bA283C7	16p11.2-q12.1
16	53774536	88643456	Loss	38%	bA26L21	bA533D19	16q12.2-q24.3
17	801989	39379547	Loss	45%	bA216P6	bA546M21	17p13.3-q21.31
17	55350495	58237753	Gain	32%	bA178C3	bA156L14	17q23.1-q23.2
19	183116	63516696	Loss	52%	bK3113P16	bK3138B18	19p13.3-q13.43
22	22510132	49521447	Loss	53%	bA80O7	bK799F10	22q11.23-13.33
X	4616007	152538528	Loss	31%	bA62N12	bA54I20	Xp22.32-q28
Reoccurring aberration in 40% of the 29 pair-matched samples							
1	5866184	40958746	Loss	43%	bA49J3	dJ739H11	1p36.31-p34.2
1	162445439	246181016	Gain	44%	bA541J2	bA438H8	1q23.3-q44
8	57918204	146167102	Gain	54%	bA342K10	dJ1056B24	8q12.1-q24.3
11	63543757	66035373	Loss	41%	bA424O11	bA142G8	11q13.1
16	65152761	79773882	Loss	42%	bA63M22	bA303E16	16q21-q23.2
16	82361616	88643456	Loss	42%	bA483P21	bA533D19	16q23.3-q24.3
17	801989	20230301	Loss	54%	bA216P6	bA121A13	17p13.3-p11.2
17	37192625	39091531	Loss	41%	bA156E6	bA392O1	17q21.2-q21.31
19	183116	63516696	Loss	52%	bK3113P16	bK3138B18	19p13.3-q13.43
22	22510132	49521447	Loss	53%	bA80O7	bK799F10	22q11.23-q13.33
Reoccurring aberration in 50% of the 29 pair-matched samples							
1	208589372	208727000	Gain	50%	dJ879K22	dJ879K22	1q32.2
1	213686097	222926484	Gain	50%	bA438G15	bA100E13	1q41-q42.12
8	69451201	146167102	Gain	55%	bA21C5	dJ1056B24	8q13.2-q24.3
17	801989	19748613	Loss	55%	bA216P6	bA78O7	17p13.3-p11.2
19	183116	16124561	Loss	56%	bK3113P16	bK2231E14	19p13.3-p13.12
19	37636197	63516696	Loss	53%	bK1325L16	bK3138B18	19q13.11-q13.43
22	22510132	49521447	Loss	53%	bA80O7	bK799F10	22q11.23-q13.33
Reoccurring aberration in 60% of the 29 pair-matched samples							
8	118297084	128884770	Gain	60%	bA67N21	dJ80K22	8q24.11-q24.21
17	801989	2492162	Loss	60%	bA216P6	bA135N5	17p13.3

3.2 Similarity of the genomic profiles of primary breast carcinomas and their matched ALN metastases

Primary tumors and ALN metastases from the same patient were found to harbor similar genomic aberrations (see Supporting Information Fig. 1). An unsupervised hierarchical clustering algorithm was applied to cluster the genomic profiles of 29 matched-pair samples on the basis of the similarity of their aberrations. Twenty-six out of the 29 pairs showed robust similarity in the unsupervised hierarchical clustering analysis (see Fig. 2). The only exceptions were samples from patients 58, 82 and 83. Even though these primary tumors did not converge with their corresponding ALN metastases in the unsupervised clustering analysis, the pair-matched samples presented similar genomic profiles by in large (see Supporting Information Fig. 1). To confirm the data we applied a variety of distance measurement algorithms and in all cases these yielded similar cluster results (data not shown). To evaluate the robustness of the data we further applied a bootstrapping method to create pseudo-replicate datasets by permutations. The similarity of the 26 pairs of primary tumors and their matched ALN metastases proved to be reproducible and highly in excess of stochastic effects, showing a highly converging frequency (25 pairs show 100% convergences in 1000 permutations).

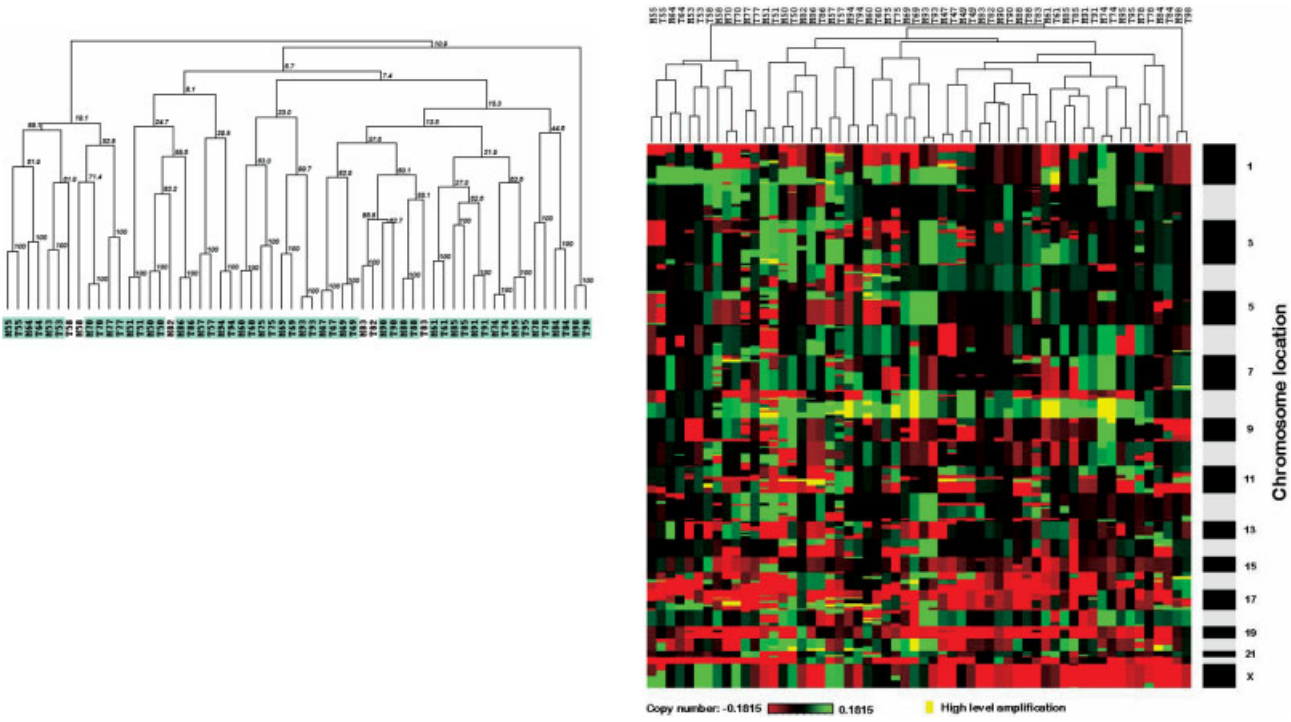

Figure 2. CGH genomic profiles of 29 pairs of primary tumors and matched ALN metastases grouped using unsupervised hierarchical clustering. Primary tumors and ALN metastases are indicated with T and M, respectively. The dendrogram represents 26 pairs of primary tumors and their matched metastases. The robustness of each fork in the dendrogram was tested by the bootstrapping method. The percentage of the recurrence in 1000 permutations is displayed for each fork (left figure). In the right figure, green indicates DNA copy number gain and red indicated loss. The bar to the right indicates chromosome location with chromosome 1pter at the top and X at the bottom. The locations of the odd-numbered chromosomes are indicated for reference.

3.3 Gel-based proteomic profiling of primary breast carcinomas and their matched ALN metastases

Thirteen pairs of primary tumors and their matched ALN were selected based on frozen tissue availability for high-resolution gel-based proteomic analysis. Representative 2-D gels of a primary tumor and its ALN metastasis (patient 51) are shown in Fig. 3 as a reference. A total of 135 well-resolved primary translation products and their variants identified by MALDI-TOF-MS were selected for matching across all 26 2-D gel images and are listed in Supporting Information Table 1 with the same numbers as those indicated in Fig. 3. Proteins were grouped in the following functional classes: cell communication/signal transduction, protein metabolism, energy pathways, cell growth/maintenance, nucleic acid metabolism and transport.

All 135 spot volumes were quantified across the 13 matched-pairs and the data are given in Supporting Information Table 2. The results showed a high similarity between the overall proteomic expression profiles of the primary tumors and the matched ALN metastases from the same patient as illustrated by the unsupervised hierarchical clustering analysis presented in Fig. 4. Similar results were obtained using other algorithms (data not shown). In spite of the striking overall similarities observed between the expression profiles, we detected a few significant changes in the quantitative level of individual proteins in the various pairs analyzed (Supporting Information Table 2). There was, however, no common protein denominator among those quantified that could distinguish the primary tumors from the ALN metastases (supervised analysis, data not shown).

As shown in Fig. 4, only M53 did not cluster with T53, a low-grade tumor (Table 1), which presented areas with ductal carcinoma *in situ* (DCIS) as judged by visual inspection of hematoxylin and eosin (HE) stained sections as well as by staining with a cytokeratin 14 antibody (compare T53 with its paired ALN metastasis in Figs. 5A and B). Additional IHC staining using a battery of 21 antibodies – against cytoskeletal proteins, receptors, proliferation markers and others – used routinely in our tumor classification programme (see Section 2) revealed that about 10% of the invasive epithelial cells in T53 were positive for the progesterone receptor (PgR) (Fig. 5C), while more than 50% of the cells in the ALN metastasis expressed PgR (Fig. 5D). As expected from the gel quantifications, the levels of staining for CK19 were very similar between the primary tumor (Fig. 5E) and the ALN metastases (Fig. 5F; see also Fig. 3).

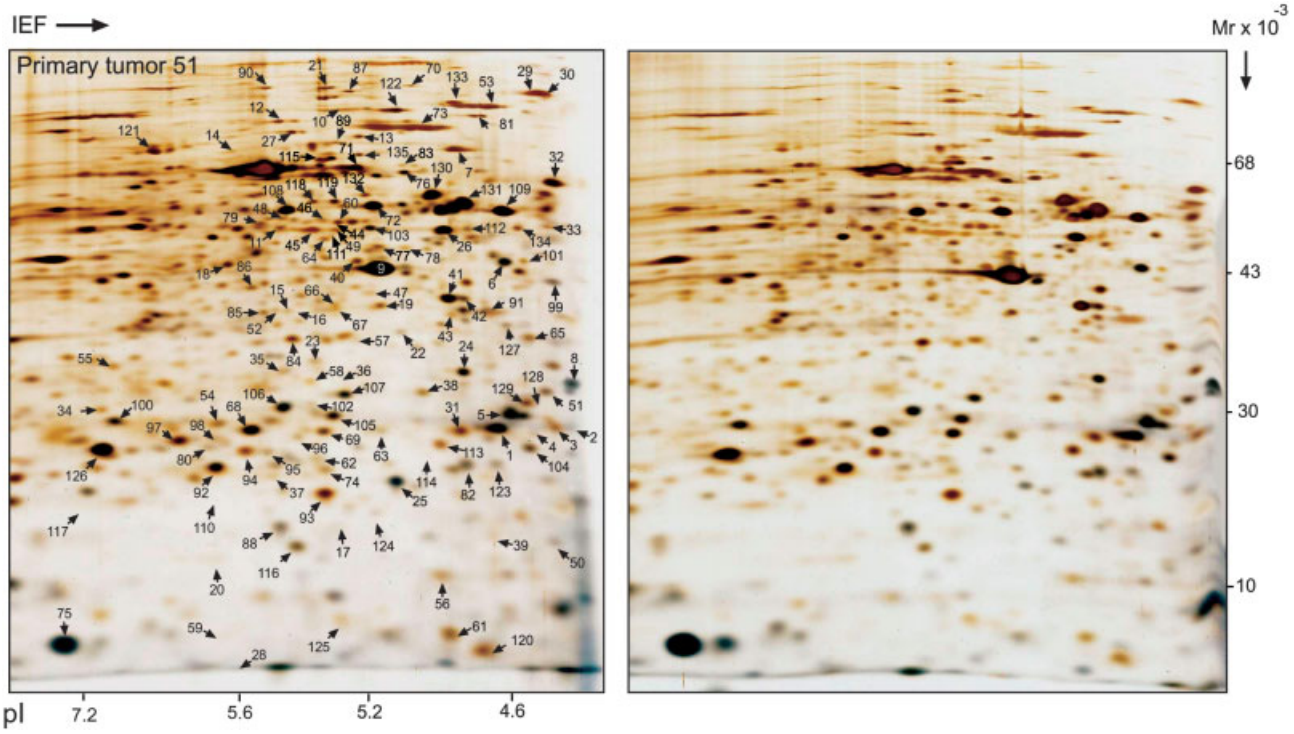

Figure 3. 2-D gel patterns of whole protein extracts from a primary breast tumor (left image) and its matched ALN metastasis (right image). Proteins were separated by IEF 2-D PAGE and stained with silver nitrate as described in Methods. The positions of the 135 proteins selected for quantification (2-D gels were stained with Sypro Ruby; images not shown) are indicated with arrows. The numbers shown in the gel image correspond to the protein numbers listed in Supporting Information Table 1.

Considering the limitations of the 2-D PAGE technology for detecting protein changes occurring in a small cell subpopulation in a tumor and the implications of the IHC findings presented above, we performed similar IHC staining of the other 12 matched-pairs analyzed by 2-D PAGE as well as of five additional tumor/ALN metastases pairs for which we did not have enough materials to perform proteomics or CGH analysis. The results showed that most individual pairs exhibited similar immunostaining patterns and staining intensities for all the antibodies analyzed (IHC staining data will be available at http://proteomics.cancer.dk), although the analysis of one pair (T78/M78), a triple negative tumor [35], suggested that the epithelial cells in the ALN metastases may have been derived from a subpopulation of cells in the primary tumor, rather than from the bulk of it. As shown in Fig. 6, the bulk of T78 was composed of epithelial cells that are negative for α-SMA (Fig. 6A), although there was a focal area in the section that contained cells expressing this protein (Fig. 6B). Analysis of the ALN metastasis showed that most of the epithelial cells expressed α-SMA (Fig. 6C); a fact that was confirmed by double immunofluorescence of sections reacted with CK19 and α-SMA antibodies (Fig. 7). Double immunofluorescence staining showed the presence of double positive cells (the yellow color in the "merge" frames indicates cells stained with both antibodies) for α-SMA and CK19 in the primary tumor samples (Fig. 7B) and the matched ALN metastases (Fig. 7C), confirming their epithelial origin. The percentages of double positive cells for α-SMA and CK19 detected by indirect immunofluorescence staining were consistent with the results obtained by IHC.

4 Discussion

Genome instability is a major hallmark of cancer [36] that is expected to play a key role in human carcinogenesis. In line with this, the recurrent aberration regions identified in this study harbor several oncogenes and tumor suppressor genes. For example, oncogenes *MMP16*, *MYC*, and *PTK2* are located on 8q11.22-24.3, tumor suppressor genes *TP53* and *NF1* are harbored on 17p13.3-11.2, and *CHEK2* and *NF2* are located on 22q11.23-13.33 (see Fig. 1 and Table 2). These regions are likely to be like "drivers" in tumor progression, while the less frequent aberration regions may behave as "passengers", *i.e.* their occurrence may be largely due to random genome instability and/or individual genetic background [37, 38]. Our studies showed that 54% of the tumors analyzed exhibited deletion of the *TP53* gene region (see Table 2), an important tumor suppressor that modulates cell cycle arrest, DNA repair, apoptosis and genome stability.

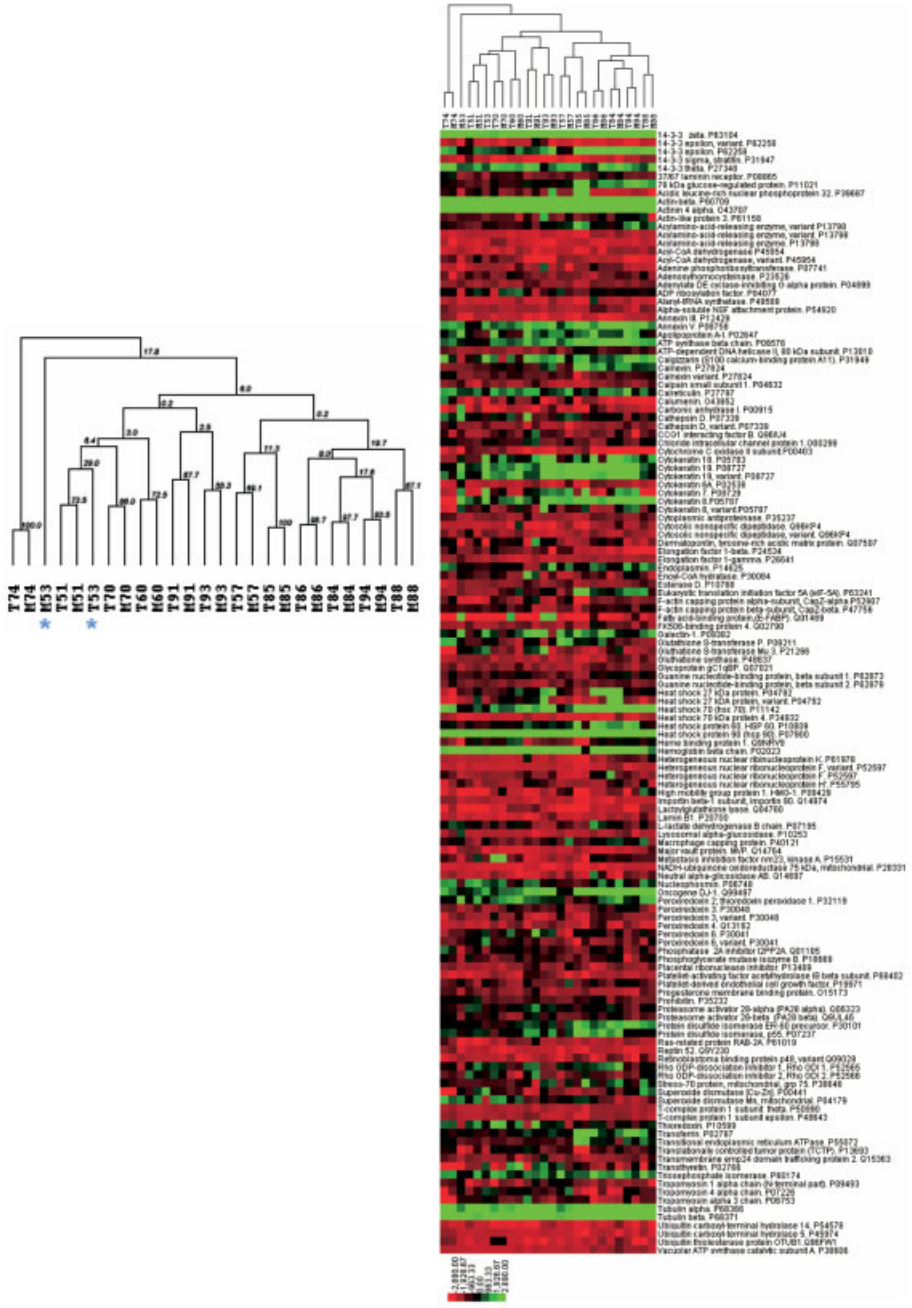

Figure 4. Dendrogram of protein expression patterns revealed by the quantitative 2-D gel analysis of 135 proteins. Proteomic profiles of 13 pairs of primary tumors and matched ALN metastases were grouped using unsupervised hierarchical clustering. Primary tumors and ALN metastases are indicated with T and M, respectively. The dendrogram shows similarities between the 12 pairs of primary tumors and their matched ALN metastases. Only T53 and M53 did not converge together (blue stars). The robustness of each fork in the dendrogram was tested by the bootstrapping method. The percentage of the recurrence in 1000 permutations is displayed for each fork (left figure).

Loss of *TP53* function, which leads to tumor progression, has been reported in many cancers [39]. The region harboring oncogene *MYC* was amplified in 52% of cases in our study. *MYC* is implicated in the control of cellular proliferation, differentiation and apoptosis [40], and its amplification has been found to be associated with progression from CIS to invasive breast cancer [41, 42].

A long-held mode of metastasis states that the primary tumor is heterogeneous and that a rare subpopulation acquires metastatic capacity at a later stage during tumorigenesis [43–45]. Bernards and Weinberg questioned this notion as they surmised that the genes specifying the final step of tumor progression/metastasis would not confer increased proliferative benefit at the primary site [46].

In breast cancer, the ALNs are often the first sites to harbor metastatic cells escaping from the primary tumor [47], and there is growing evidence that lymphageneous metastases occurs independently of distant metastasis and that the molecular mechanisms that drive these two processes are different [9]. Several groups have analyzed the gene expres-

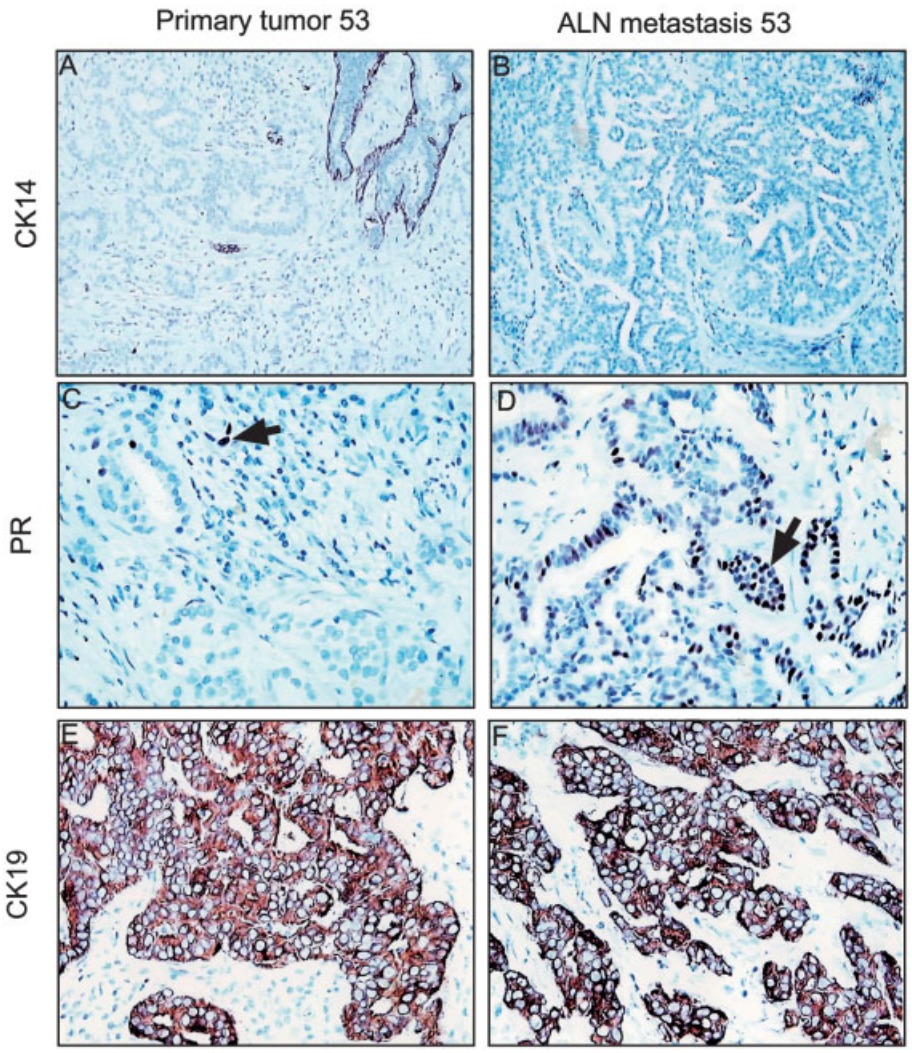

Figure 5. IHC analysis of paraffin-embedded sections of tissue samples from primary tumor and its matched ALN metastasis (patient 53). Sections were stained with antibodies against cytokeratin 14 (CK14) (A and B), progesterone receptor (PgR) (C and D) and cytokeratin 19 (CK19) (E and F). Cells positive for PgR are indicated with arrows (panels C and D).

sion profiles of primary breast carcinomas and their matched ALN metastases using cDNA microarrays and found that in general primary breast carcinomas and ALN metastases are very much alike and do not differ at the transcriptional level by a common subset of genes [9]. Contrary to these studies, however, some recent publications have reported metastasis signatures that imply that breast ALN metastases are molecularly distinct from their primary tumors [15, 48, 49].

The work we have presented here revealed a remarkable similarity both in terms genomic aberrations and overall protein expression profiles in carefully selected pairs of matched primary tumors and ALN metastases. In general, the data is in good agreement with previous published results at the transcriptional level [9]. Furthermore, our studies failed to reveal a common clone or protein denominator that could discriminate between primary tumors from the ALN metastases. To our knowledge, this is the first time that the similarity between the tumor/ALN metastasis pairs was confirmed using a combination of complementary "omic" approaches. Our studies cannot exclude, however, the possibility that the ALN metastases originate from specialized cell subpopulations having metastatic advantages as neither array CGH, expression microarrays, nor 2-D PAGE can sensitively detect all changes, especially if these subpopulations are minor and bear a similar overall genomic/transcriptomic/proteomic profiles to the majority of the cells in the tumor. The IHC results showed indeed that most individual pairs exhibited similar immunostaining patterns for all the antibodies analyzed, although in two cases (patient 53 and 78) the cells in the ALN metastasis exhibited phenotypic variations as compared to the cells present in the bulk of the primary tumor suggesting that they may either be derived from a distinct group of cells present in only some areas of the tumor or that novel phenotypic characteristics have been acquired by the cells in the ALN. One possible explanation for our observation is that the primary tumor cells changed their phenotype as a result of an epithelial-to-mesenchymal transition (EMT) event. The resulting cells are expected to

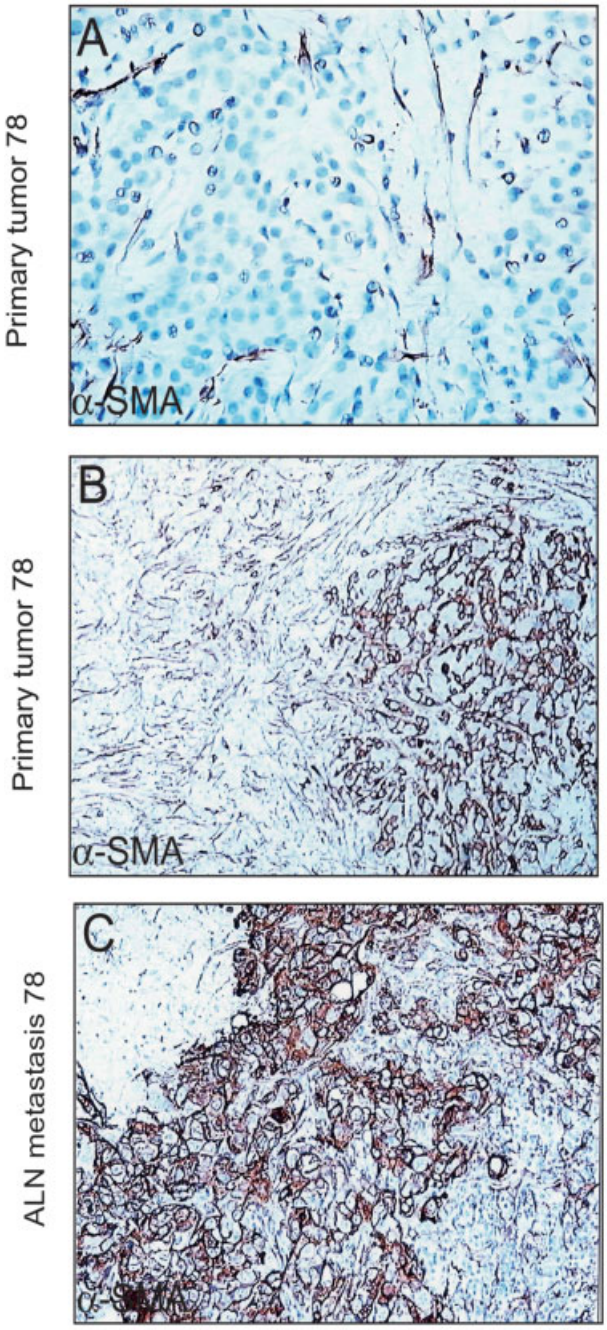

Figure 6. IHC staining of paraffin-embedded sections of tissue samples from primary tumor and its matched ALN metastasis (patient 78). Sections were stained with antibodies against α-smooth muscle actin (α-SMA). (A) primary tumor, high magnification; (B) primary tumor, low magnification); (C) ALN metastasis, low magnification.

have a greater metastatic potential and be more prone to dissemination [51, 52], which again would fit with our observation. Even though we cannot formally dismiss this possibility taking place in the two tumor/ALN metastasis pairs in question we do not think this is very likely as (i) EMT events are very rare in breast cancer [53], and (ii) we did not observe expression of mesenchymal-specific markers, such as vimentin, by cells in the tumors (data not shown) or in the ALN metastasis. Moreover, in both cases cells expressed stratifin (data not shown), an epithelial specific marker that we have previously shown to be down-regulated during EMT [54]. Thus, our findings could be interpreted in terms of coexistence in the bulk of the primary tumor of clonal subpopulations that differ in their metastatic potential [48, 49] (Fig. 8A) and/or adaptation of disseminated metastatic cells to the ALN microenvironment (Fig. 8B), which clearly imposes a great selective pressure on these cells leading to phenotypic drift [50]. Further studies will be needed to answer the question of which of these mechanisms is predominant.

Even though from our data it is not possible to conclusively exclude any of the two models of metastases currently under consideration [43, 44, 55], it seems reasonable to assume that the formation of lymphogenous metastasis depends on the accumulation of a number of seeding of cells derived from the primary tumors which in our experience are often very heterogeneous [56]. The remarkable similarities of the overall genomic and proteomic profiles between primary tumors and matched ALN metastases are taken to indicate that important biological characteristics of the primary breast tumor are maintained in the corresponding lymph node metastases. Further studies using a larger patient cohort with a more precise tumor classification and a more comprehensive battery of antibodies will allow us to gain a better understanding of tumor heterogeneity in terms of acquisition of metastatic competence and should lead to a deeper insight into the biology of regional metastasis.

We would like to thank Kitt Christensen, Hanne Nors, Michael Radich Johansen, Signe Trentemøller, Sofia Svensson, Thanja Poulsen, Bodil Schmidt and Søren Pedersen for expert technical assistance. We also thank the Wellcome Trust Sanger Institute for kindly providing the 1-Mb resolution BAC clone. This work was supported by the Danish Cancer Society through the budget of the Institute of Cancer Biology and by grants from the Danish Medical Research Council, the Natural and Medical Sciences Committee of the Danish Cancer Society, Novo Nordisk, the John and Birthe Meyer Foundation, the Solar Fonden, the Stensbygaard Fonden, the Kai Lange og Gundhild Kai Lange Fond, the will of Edith Stern and a "Race Against Breast Cancer". The support of the Marketing Department at the Danish Cancer Society is greatly appreciated. This study was also supported by a Project grant from the Hi-Tech Research and Development Program of China (2006AA02A301).

The authors have declared no conflict of interest.

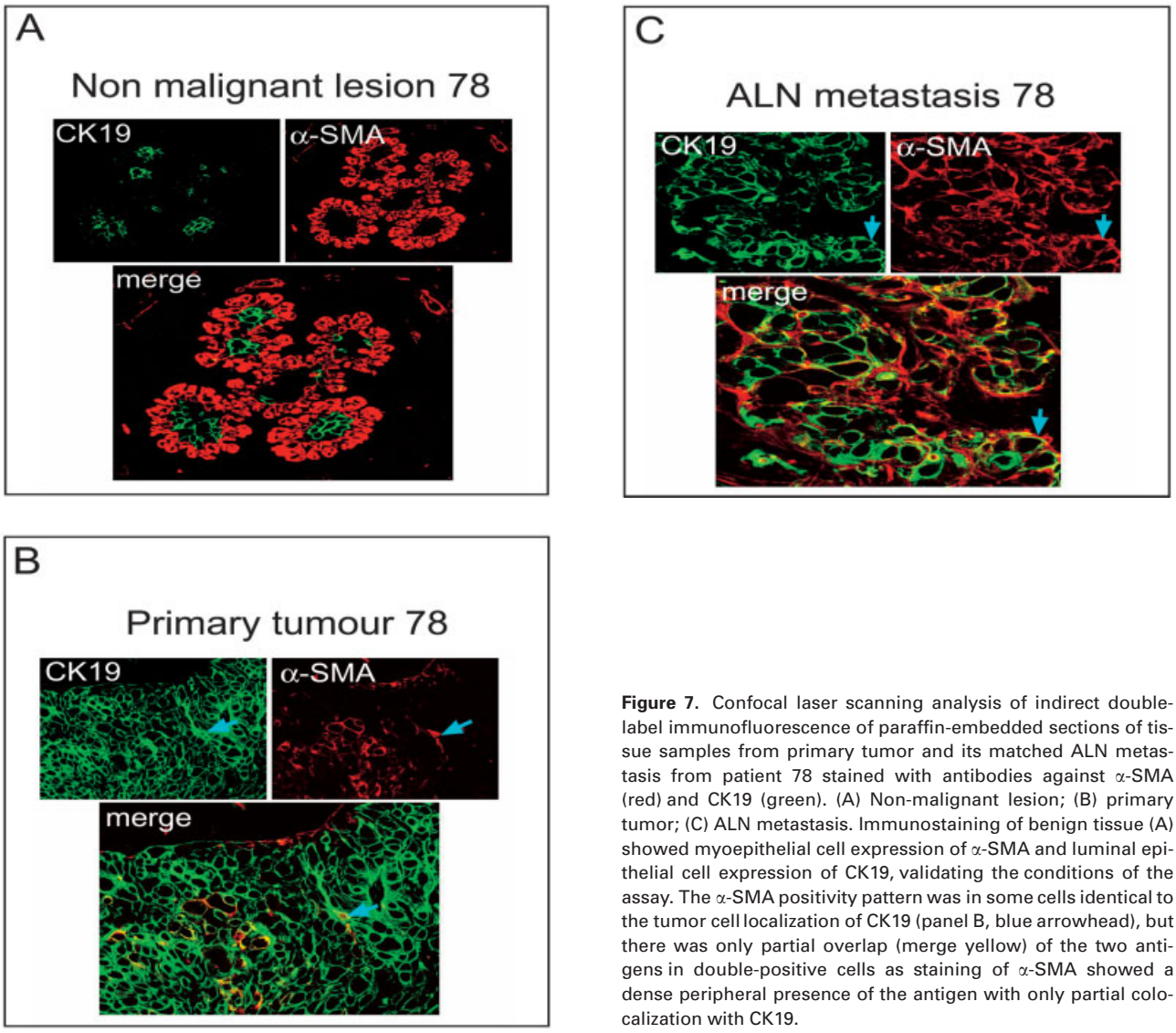

Figure 7. Confocal laser scanning analysis of indirect double-label immunofluorescence of paraffin-embedded sections of tissue samples from primary tumor and its matched ALN metastasis from patient 78 stained with antibodies against α-SMA (red) and CK19 (green). (A) Non-malignant lesion; (B) primary tumor; (C) ALN metastasis. Immunostaining of benign tissue (A) showed myoepithelial cell expression of α-SMA and luminal epithelial cell expression of CK19, validating the conditions of the assay. The α-SMA positivity pattern was in some cells identical to the tumor cell localization of CK19 (panel B, blue arrowhead), but there was only partial overlap (merge yellow) of the two antigens in double-positive cells as staining of α-SMA showed a dense peripheral presence of the antigen with only partial colocalization with CK19.

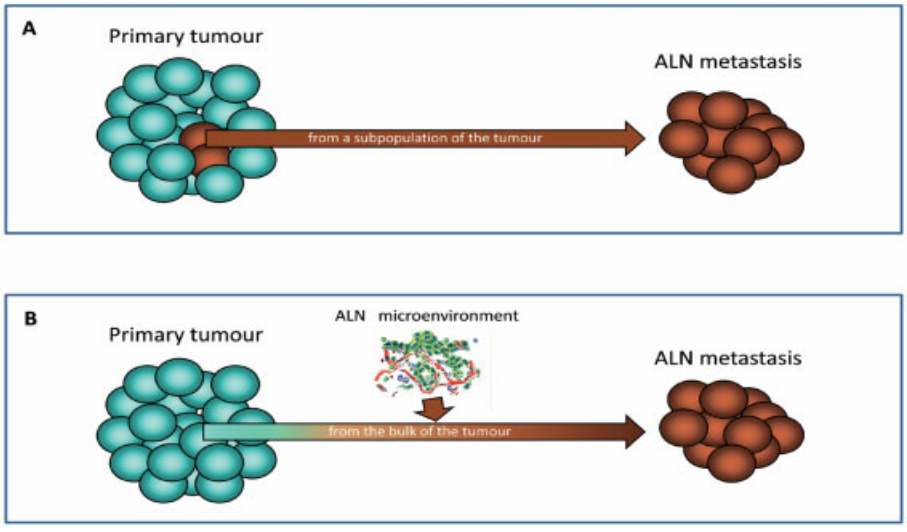

Figure 8. Putative mechanisms of origin of ALN metastasis phenotypically distinct from the primary breast tumor. (A) ALN metastasis originates from a distinct group of cells present in some areas of the tumor. The specific expression profiles for some genes/proteins observed in ALN metastasis and its matched primaries reflect primary tumor heterogeneity in terms of presence of selected subpopulations of cells that acquire metastatic advantages. (B) ALN metastasis derives from a bulk of the tumor and changes its phenotypic profile for selected genes/proteins as a result of a selective adaptation to the lymph node microenvironment.

5 References

[1] Swartz, M. A., Skobe, M., Lymphatic function, lymphangiogenesis, and cancer metastasis. *Microscopy research and technique* 2001, *55*, 92–99.

[2] Dent, D. M., Axillary lymphadenectomy for breast cancer. Paradigm shifts and pragmatic surgeons. *Arch. Surg.* 1996, *131*, 1125–1127.

[3] Mirza, A. N., Mirza, N. Q., Vlastos, G., Singletary, S. E., Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. *Annals Surg.* 2002, *235*, 10–26.

[4] Nemoto, T., Natarajan, N., Bedwani, R., Vana, J., Murphy, G. P., Breast cancer in the medial half. Results of 1978 National Survey of the American College of Surgeons. *Cancer* 1983, *51*, 1333–1338.

[5] Rack, B., Janni, W., Gerber, B., Strobl, B. *et al.*, Patients with recurrent breast cancer: does the primary axillary lymph node status predict more aggressive tumor progression? *Breast Cancer Res. Treat.* 2003, *82*, 83–92.

[6] Woodward, W. A., Vinh-Hung, V., Ueno, N. T., Cheng, Y. C. *et al.*, Prognostic value of nodal ratios in node-positive breast cancer. *J. Clin. Oncol.* 2006, *24*, 2910–2916.

[7] Rosen, P. R., Groshen, S., Saigo, P. E., Kinne, D. W., Hellman, S., A long-term follow-up study of survival in stage I (T1N0M0) and stage II (T1N1M0) breast carcinoma. *J. Clin. Oncol.* 1989, *7*, 355–366.

[8] Hellman, S., Karnofsky Memorial Lecture. Natural history of small breast cancers. *J. Clin. Oncol.* 1994, *12*, 2229–2234.

[9] Weigelt, B., Wessels, L. F., Bosma, A. J., Glas, A. M. *et al.*, No common denominator for breast cancer lymph node metastasis. *Br. J. cancer* 2005, *93*, 924–932.

[10] Daidone, M. G., Silvestrini, R., Valentinis, B., Persici, P. *et al.*, Proliferative activity of primary breast cancer and of synchronous lymph node metastases evaluated by [3H]-thymidine labelling index. *Cell Tissue Kinetics* 1990, *23*, 401–408.

[11] Feichter, G. E., Kaufmann, M., Muller, A., Haag, D. *et al.*, DNA index and cell cycle analysis of primary breast cancer and synchronous axillary lymph node metastases. *Breast Cancer Res. Treat.* 1989, *13*, 17–22.

[12] Goodson, W. H., 3rd, Ljung, B. M., Moore, D. H., 2nd, Mayall, B. *et al.*, Tumor labeling indices of primary breast cancers and their regional lymph node metastases. *Cancer* 1993, *71*, 3914–3919.

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

[14] Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M. *et al.*, Molecular portraits of human breast tumours. *Nature* 2000, *406*, 747–752.

[15] Feng, Y., Sun, B., Li, X., Zhang, L. *et al.*, Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. *Breast Cancer Res. Treat.t* 2007, *103*, 319–329.

[16] Nishizaki, T., DeVries, S., Chew, K., Goodson, W. H., 3rd *et al.*, Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes, chromosomes & cancer* 1997, *19*, 267–272.

[17] Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M. *et al.*, Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol. Cell. Proteomics* 2004, *3*, 327–344.

[18] Celis, J. E., Gromov, P., Gromova, I., Moreira, J. M. *et al.*, Integrating proteomic and functional genomic technologies in discovery-driven translational breast cancer research. *Mol. Cell. Proteomics* 2003, *2*, 369–377.

[19] Celis, J. E., Moreira, J. M., Cabezon, T., Gromov, P. *et al.*, Identification of extracellular and intracellular signaling components of the mammary adipose tissue and its interstitial fluid in high risk breast cancer patients: toward dissecting the molecular circuitry of epithelial-adipocyte stromal cell interactions. *Mol. Cell. Proteomics* 2005, *4*, 492–522.

[20] Celis, J. E., Moreira, J. M., Gromova, I., Cabezon, T. *et al.*, Towards discovery-driven translational research in breast cancer. *FEBS J.* 2005, *272*, 2–15.

[21] Moll, R., Cytokeratins as markers of differentiation in the diagnosis of epithelial tumors. *Sub-cellular Biochem.* 1998, *31*, 205–262.

[22] Celis, J. E., Trentemølle, S., Gromov, P., in: Celis, J. E., Carter, N., Hunter, T., Shotton, D. (Eds.), *Cell Biology. A Laboratory Handbook*, Academic Press, Inc, San Diego 2006, pp. 165–174.

[23] O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975, *250*, 4007–4021.

[24] Gromova, I., Celis, J. E., in: Celis, J. E., Carter, N., Hunter, T., Shotton, D. (Eds.), *Cell Biology. A Laboratory Handbook*, Academic Press, Inc, San Diego 2006, pp. 219–223.

[25] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 1996, *68*, 850–858.

[26] Ostergaard, M., Wolf, H., Orntoft, T. F., Celis, J. E., Psoriasin (S100A7): a putative urinary marker for the follow-up of patients with bladder squamous cell carcinomas. *Electrophoresis* 1999, *20*, 349–354.

[27] Celis, J. E., Gromov, P., High-resolution two-dimensional gel electrophoresis and protein identification using western blotting and ECL detection. *Exs* 2000, *88*, 55–67.

[28] Moreira, J. M., Ohlsson, G., Rank, F. E., Celis, J. E., Down-regulation of the tumor suppressor protein 14-3-3sigma is a sporadic event in cancer of the breast. *Mol. Cell. Proteomics* 2005, *4*, 555–569.

[29] Zhang, X., Snijders, A., Segraves, R., Zhang, X. *et al.*, High-resolution mapping of genotype-phenotype relationships in cri du chat syndrome using array comparative genomic hybridization. *Am. J. Hum. Genet.* 2005, *76*, 312–326.

[30] Olshen, A. B., Venkatraman, E. S., Lucito, R., Wigler, M., Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics (Oxford, England)* 2004, *5*, 557–572.

[31] Willenbrock, H., Fridlyand, J., A comparison study: applying segmentation to array CGH data for downstream analyses. *Bioinformatics (Oxford, England)* 2005, *21*, 4084–4091.

[32] de Hoon, M. J., Imoto, S., Nolan, J., Miyano, S., Open source clustering software. *Bioinformatics (Oxford, England)* 2004, *20*, 1453–1454.

[33] Felsenstein, J., PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* 1989, *5*, 164–166.

[34] Struski, S., Doco-Fenzy, M., Cornillet-Lefebvre, P., Compilation of published comparative genomic hybridization studies. *Cancer Genet. Cytogenet.* 2002, *135*, 63–90.

[35] Rakha, E. A., Putti, T. C., Abd El-Rehim, D. M., Paish, C. *et al.*, Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation. *J. Pathol.* 2006, *208*, 495–506.

[36] Albertson, D. G., Collins, C., McCormick, F., Gray, J. W., Chromosome aberrations in solid tumors. *Nat. Genet.* 2003, *34*, 369–376.

[37] Hunter, K., Welch, D. R., Liu, E. T., Genetic background is an important determinant of metastatic potential. *Nat. Genet.* 2003, *34*, 23–24; author reply 25.

[38] Hunter, K. W., Allelic diversity in the host genetic background may be an important determinant in tumor metastatic dissemination. *Cancer Lett.* 2003, *200*, 97–105.

[39] Wang, X. W., Harris, C. C., p53 tumor-suppressor gene: clues to molecular carcinogenesis. *J. Cell. Physiol.* 1997, *173*, 247–255.

[40] Henriksson, M., Luscher, B., Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv. Cancer Res.* 1996, *68*, 109–182.

[41] Courjal, F., Cuny, M., Simony-Lafontaine, J., Louason, G. *et al.*, Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.* 1997, *57*, 4360–4367.

[42] Robanus-Maandag, E. C., Bosch, C. A., Kristel, P. M., Hart, A. A. *et al.*, Association of C-MYC amplification with progression from the in situ to the invasive stage in C-MYC-amplified breast carcinomas. *J. Pathol.* 2003, *201*, 75–82.

[43] Nowell, P. C., The clonal evolution of tumor cell populations. *Science* 1976, *194*, 23–28.

[44] Fidler, I. J., Kripke, M. L., Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977, *197*, 893–895.

[45] Poste, G., Fidler, I. J., The pathogenesis of cancer metastasis. *Nature* 1980, *283*, 139–146.

[46] Bernards, R., Weinberg, R. A., A progression puzzle. *Nature* 2002, *418*, 823.

[47] Stacker, S. A., Achen, M. G., Jussila, L., Baldwin, M. E., Alitalo, K., Lymphangiogenesis and cancer metastasis. *Nat. Rev.* 2002, *2*, 10.

[48] Suzuki, M., Tarin, D., Gene expression profiling of human lymph node metastases and matched primary breast carcinomas from the same patients to identify candidate metastasis-related genes. *Mol. Oncol.* 2007, *1*, 172–180.

[49] Vecchi, M., Confalonieri, S., Nuciforo, P., Vigano, M. A. *et al.*, Breast cancer metastases are molecularly distinct from their primary tumors. *Oncogene* 2008, *27*, 2148–2158.

[50] Weinberg, R. A., Is metastasis predetermined? *Mol. Oncol.* 2007, *1*, 263–264.

[51] Hugo, H., Ackland, M. L., Blick, T., Lawrence, M. G. *et al.*, Epithelial–mesenchymal and mesenchymal–epithelial transitions in carcinoma progression. *J. Cell. Physiol.* 2007, *213*, 374–383.

[52] Huber, M. A., Kraut, N., Beug, H., Molecular requirements for epithelial-mesenchymal transition during tumor progression. *Curr. Opin. Cell Biol.* 2005, *17*, 548–558.

[53] Trimboli, A. J., Fukino, K., de Bruin, A., Wei, G. *et al.*, Direct evidence for epithelial-mesenchymal transitions in breast cancer. *Cancer Res.* 2008, *68*, 937–945.

[54] Moreira, J. M., Gromov, P., Celis, J. E., Expression of the tumor suppressor protein 14-3-3 sigma is down-regulated in invasive transitional cell carcinomas of the urinary bladder undergoing epithelial-to-mesenchymal transition. *Mol. Cell. Proteomics* 2004, *3*, 410–419.

[55] Weigelt, B., Glas, A. M., Wessels, L. F., Witteveen, A. T. *et al.*, Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 15901–15905.

[56] Celis J. E., Moreira, J.M., Gromova, I., Cabezón I. *et al.*, Characterization of breast precancerous lesions and myoepithelial hyperplasia in sclerosing adenosis with apocrine metaplasia. *Mol. Oncol.* 2007, *1*, 97–119.

