Meta-data Analysis as a Strategy to Evaluate Individual and Common Features of Proteomic Changes in Breast Cancer

OLENA ZAKHARCHENKO[1], CHRISTINA GREENWOOD[2], ANNA LEWANDOWSKA[1,3], ULF HELLMAN[4], LOUISE ALLDRIDGE[2,5] and SERHIY SOUCHELNYTSKYI[1]

[1]*Department of Oncology-Pathology, Karolinska Biomics Center, Karolinska Institute, Stockholm, Sweden;*
[2]*Helen Rollason Research Laboratory, Anglia Ruskin University, Chelmsford, United Kingdom;*
[3]*Faculty of Biology, University of Warmia and Mazury, Olsztyn, Poland;*
[4]*Ludwig Institute for Cancer Research Ltd., Uppsala University, Uppsala, Sweden;*
[5]*Griffith University, School of Medicine, Gold Coast, Australia*

Abstract. *Background: Individual differences among breast tumours in patients is a significant challenge for the treatment of breast cancer. This study reports a strategy to assess these individual differences and the common regulatory mechanisms that may underlie breast tumourigenesis. Materials and Methods: The two-step strategy was based firstly on a full-scale proteomics analysis of individual cases, and secondly on the analysis of common features of the individual proteome-centred networks (meta-data). Results: Proteomic profiling of human invasive ductal carcinoma tumours was performed and each case was analysed individually. Analysis of primary datasets for common cancer-related proteins identified keratins. Analysis of individual networks built with identified proteins predicted features and regulatory mechanisms involved in each individual case. Validation of these findings by immunohistochemistry confirmed the predicted deregulation of expression of CK2α, PDGFRα, PYK and p53 proteins. Conclusion: Meta-data analysis allowed efficient evaluation of both individual and common features of the breast cancer proteome.*

Variability of breast cancer is manifested on various levels, from histological appearance to molecular mechanisms (1-3). This variability calls for individual assessment of each patient so that the best treatment is provided. Currently, the selection of treatment for breast cancer is based on clinical data, histopathological examinations and some molecular markers. Size and location of a tumour, lymph node status and presence of distal metastases are at the core of clinical evaluation (2, 3). Histopathological examination of a biopsy or a resected tumour provides important information about types and differentiation status of cells in a tumour. Expression of HER2/neu, oestrogen and progesterone receptors is often used in determining appropriate treatment. In some clinics, expression of p53 and VEGF receptor, vascularisation level, areas of inflammation and structure of the tumour stroma are considered (2-4). Molecular diagnostics of breast cancer commenced with the introduction of mRNA expression arrays (5). Molecular characterization of the types of breast tumours, which is different from the tumour grade system based on clinical data, was a strong confirmation of the variability of breast tumours at the molecular level (6). mRNA expression studies have since provided signatures to discriminate patients with a worse prognosis and/or development of an aggressive tumour type, *e.g.* MammaPrint and OncoPrint (3-7). However, their areas of application are limited, and the array-based tools still have to prove their clinical value.

Molecular diagnostics is of high importance, as it has the potential to detect novel drug targets. However, practically all reports of molecular markers have been focused on the identification of common features in the studied cohorts (2-7). Such an approach tends to disregard any individual-specific features. It also tends to minimise the insight into regulatory mechanisms which may be affected in a majority of samples *via* different components in different individuals. As a consequence, markers of differentiation of human breast epithelial cells are mostly keratins, however, there are a number of signalling mechanisms that have been shown to regulate the differentiation of cells (4, 8). It is believed that these signalling mechanisms are not considered because different molecules may control them in different tumours, and therefore it is challenging to obtain a list of the same signalling proteins in a large cohort. Therefore, although the same regulatory processes may be affected, in many cases they may be missed due to the variability of involved signalling components.

Correspondence to: Serhiy Souchelnytskyi, Z5:01, Karolinska University Hospital, Solna, SE-17176, Stockholm, Sweden. Tel: +46 851775167, Fax: +46 851771000, e-mail: serhiy.souchelnytskyi@ki.se

Supplementary data may be requested from the authors.

Key Words: Proteomics, breast cancer, signalling, meta-data analysis.

Proteins offer a rich source of markers for diagnostics, prediction and monitoring of cancer treatment (9, 10). The importance of such proteins is emphasised by the fact that all anti cancer drugs act on or *via* proteins (11). Therefore, proteomic profiling of breast tumours has been approached extensively. Proteins extracted from tumours, microdissected cells or tumour interstitial fluids have all been studied (12-19). The main methodological approaches used in such studies are two-dimensional gel electrophoresis (2D-GE), peptide-based shotgun mass spectrometry techniques and various arrays (10-20). 2D-GE is currently the only technology which allows the separation of hundreds of full-length proteins (21). As practically all proteins *in vivo* have post-translational modifications, the use of full-length proteins, as analytes, is essential for high quality proteome studies.

Attempts to develop a general 'one-fit-for-all' proteomic profile of breast tumourigenesis have delivered lists of mostly high-abundance and structural proteins, *e.g.* keratins (9, 10, 22). This is mainly due to the averaging of primary datasets and the inability to interpret individual differences. However, the combination of proteomics technologies, systems biology tools and modern molecular and cell biology in the field of cancer studies provides a platform for achieving a new depth in tumour profiling. This study shows that a complete analysis of individual cases followed by comparison of identified protein-dependent networks is informative in gaining insight into the molecular mechanisms that may be present either in all cases or only in an individual patient.

Materials and Methods

Clinical samples and their preparation. Clinical samples were collected at Broomfield Hospital (Chelmsford, UK), under Ethical Permit 04/Q0303/28, issued by the North and Mid Essex Local Research Ethics Committee (Harlow, UK), immediately upon surgery and stored on wet ice before being dissected by a pathologist. Samples of breast epithelial tissue were snap-frozen immediately in liquid nitrogen for use in the proteomics analysis. Samples for immunohistopathological diagnostics were collected and embedded in paraffin blocks in the Department of Histopathology, Broomfield Hospital (UK) before being sectioned onto glass slides. For the proteomics study, tissue was extracted directly in a buffer for isoelectrofocusing (8 M Urea, 2.5% CHAPS, 50 mM DTT, IPGbuffer with pH 3-10, traces of bromphenol blue), with mechanical disintegration with glass beads at room temperature. Extracts were centrifuged at 13,000 rpm (15,000 ×*g*) for 15 min, and supernatants were used for 2D-GE.

Proteome profiling. Proteome profiling, 2D-GE, gel image analysis and MALDI TOF mass spectrometry were used, as described previously (23). In brief, first-dimension isoelectrofocusing (IEF) was performed using IPGDry strips (linear, pH 3-10, 18-cm long) in an IPGPhor instrument (Amersham Biosciences, Uppsala, Sweden), using the following protocol: 10 h, rehydration of strips with a sample; 2 h, 50 V; 1 h, 500 V; 1 h, 1000 V; and 10 h, 5.000 V. The second-dimension SDS-PAGE was performed in an Ettan Dalt Six electrophoresis system (Amersham Biosciences), using the following protocol: 0.5 W/gel, 15 min; 1 W/gel, 30 min; 10 W/gel, to completion of the run (45,000 Vh). Three to four 10% SDS-PAGE gels were generated for each sample, depending on the quantity of extracted proteins. Generated gels were stained with 0.1% silver nitrate. Protein spots were analysed using Image Master Platinum version 6.0 software (GE Healthcare, Uppsala, Sweden). Statistical significance of the reproducibility of spot expression in 2D gels and differences in expression were evaluated by using the Image Master 2D Platinum Version 6.0 statistical package (GE Healthcare). Proteins from 2D gels which were shown to have either a unique expression pattern or exhibited changes by more than a 50% increase or decrease in expression between tumour and histologically healthy adjacent tissue were considered for identification. The Student's *t*-test was used to determine the statistical significance of the observed changes ($p<0.05$).

Protein identification. Protein spots were excised from the gels, destained and subjected to in-gel digestion with trypsin (modified, sequence grade porcine; Promega, USA), as described previously (23). Tryptic peptides were concentrated and desalted in ZipTip's μC18 (Millipore, Billerica, USA). Peptides were eluted with 65% acetonitrile, containing the matrix α-cyano-4-hydroxycinnamic acid, and applied directly onto the metal target and analyzed by MALDI TOF MS on a Bruker Ultraflex instrument (Bruker Daltonics, Bremen, Germany). Embedded software (FlexAnalysis; Bruker Daltonics) was used to collect and process mass spectra. Peptide spectra were internally calibrated using autolytic peptides from the trypsin (842.51, 1045.56 and 2211.10 Da). To identify proteins, searches in the NCBI nr (2010/05/10) RefSeq sequence database (NCBI, Bethedsa, MD, USA; http://www.ncbi.nlm.nih.gov) were performed using the ProFound search engine (http://65.219.84.5/ service/prowl/profound.html). One miscut, alkylation and partial oxidation of methionine were allowed. Search parameters were set to 'no limitations of pI', 'Mr', 'tolerance less than 0.1 Da', and 'mammalian' for species search. Significance of the identification was evaluated according to the probability value ('Z') and sequence coverage.

Systemic analysis. Protein names were translated into gene ontology (GO) terms (http://www.geneontology.org). Functional and pathway analysis was performed using Ingenuity Pathway Analysis (IPA), a tool for description of networks and signalling pathways (http://www.ingenuity.com). IPA operates with a proprietary database and considers only those experimental data which have been evaluated by independent researchers. This ensures that only confirmed results are taken into consideration for building a network. Experimental results which have not been reported by multiple laboratories or may have controversial interpretations were not considered for analysis. Such stringent selection of experimental data was required to exclude false-positive relations. Fischer's exact test was used to calculate a *p*-value determining the network connectivity.

Immunohistochemical study. BRC961 USBiomax breast cancer arrays (US Biomax Inc., Rockville, MD, USA) were used to evaluate the expression of CK2α, PDGFRα, PYK and p53. Each array slide contained 35 cases of malignant tumours, three cases of hyperplasia, five cases of benign tumours and three cases of non-neoplastic tissues (supplementary data are available from the authors). Arrays were stained with anti-CK2α (H-286; sc-9030), anti-PDGFRα (C-20; sc-338), anti-PYK (H-102; sc-9019) and anti p53 (DO-1; sc-126) (all

Table I. *Clinical and pathological description of cases subjected to proteomic profiling in the present study.*

Case number[a]	Histopathological diagnostics	Grade[c]	ER[d]	PR[d]	HER2/neu[d]	Size total (max diam; mm)[e]	Lymph node positive[e]	Lymph node examined total[e]	Lymphovascular invasion[e]
#1	IDC DCIS	2	+	–	+	25	0	0	Yes
#6	IDC DCIS	3	–	–	–	20	0	6	No
#37	IDC	3	+	–	+	28	0	0	Yes
#40	IDC	2	+	+	n/a	14	0	9	Yes
#45	IDC	3	–	–	+	27	0	4	n/a
#47	IDC DCIS	2	+	–	–	42	0	5	n/a

[a]Identification number of the cases; [b]IDC – invasive ductal carcinoma; DCIS – ductal carcinoma *in situ*. IDC DCIS – IDC with inclusions of DCIS; [c]Grade of tumours; [d]ER, PgR and HER2/neu status was evaluated by immunohistochemistry; [e]Size of tumours and invasive areas were measured by a pathologist upon pathological examination. The number of positive lymph nodes indicates the number of lymph nodes with detected metastasis. Lymphovascular invasion in tumours was evaluated upon histopathological analysis; n/a, not available.

from Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA). All antibodies were used at a dilution 1:50 (v/v), according to the supplier's recommendations. Anti-Smad2 *C*-terminal phosphorylation (pS2) antibodies were described previously (23). pS2 antibodies were used at dilution 1:25 (v/v). Antigen retrieval was performed using DakoCytomation target retrieval solution high pH (DAKO, Carpinteria, CA, USA). The slides were stained with VECTASTAIN Elite ABC kits (Vector Laboratories Inc., Burlingame, CA, USA) following the manufacturer's instruction, and mounted with Fluoromount G (Southern Biotechnology, Birmingham, AL, USA). The stained tissues were photographed using a Leica DFC camera and images were acquired with Leica QWin Standard software (Leica Microsystems Imaging Solutions Ltd, Cambridge, UK). Intensity of staining was evaluated as absent ((–); no stained cells), weak ((+); <10% of stained cells), moderate ((++); 10% to 50% of cells stained) and strong ((+++); >50% of stained cells). The staining was evaluated in malignant (epithelial) cells of tumours and epithelial cells of healthy tissues.

Results

Generation of individual proteome profiles. Previous studies which highlighted the significant variability between clinical samples (1-19) prompted the present study to develop a new strategy based on a proteomics study of each clinical breast cancer case separately before attempting to find changes common for all cases in a studied cohort of patients. Every breast tumour tissue sample (case) was subjected to a complete proteomic analysis which included proteomic profiling, identification of proteins from the individual tumours and functional clustering and network building. The aim of using this method was to identify and predict regulatory mechanisms affected by identified proteins specific to each breast tumour case. Consequently, sets of proteins and predicted regulatory mechanisms affected in individual cases were compared against all studied cases (Figure 1A).

All the studied cases were described as invasive ductal carcinomas (IDCs); cases #1, #6 and #47 were described as IDCs with elements of ductal carcinoma *in situ* (DCIS). Tumour grades were 2 (3 cases) or 3 (3 cases) and no lymph metastases were observed in the examined cases (Table I). The aim of this selection of cases was to focus on IDCs. As expected, the histopathological evaluation of sections of tumours showed some variability in the histology of the samples, although more than 50% of the cellular component was composed of malignant cells (Figure 1B). Differences in the presence of epithelial cells and stromal elements were observed. These histological differences may reflect molecular variability between the cases. It was hypothesised that a proteomic analysis of each case separately would expose individual features of the cases. Therefore, proteomic analysis was performed for each case separately, *i.e.* each tumour was compared to the corresponding adjacent histologically healthy tissue.

2D gels were generated for each of the studied cases, as a set of a tumourous and a corresponding histologically healthy adjacent tissue. The overall pattern of protein migration in 2D gels was similar for all cases, with the majority of proteins having a tendency to shift into the area of gels corresponding to pI below 7.5 (Figure 2, supplementary data are available from the authors). A similar distribution of proteins in 2D gels was observed previously in studies of breast tumours (12, 15, 16, 18). Despite the similarity in the overall protein patterns of all cases, there was variability in the expression of tumour-related proteins. For example, 46 protein spots were detected for the case #47, while for the case #37 there were 180 tumour-related protein spots. The numbers of identified proteins which changed their expression are indicated in Table II. The lists of the identified proteins for each of the cases are given in supplementary data are available from the authors. It should be noted that a number of proteins were identified in multiple spots. In these cases, there were between 5 to 11 proteins identified in multiple spots. The number of spots for a unique protein ranged from 2 to 15. This confirmed the importance of studying full-length proteins without prior digestion to peptides.

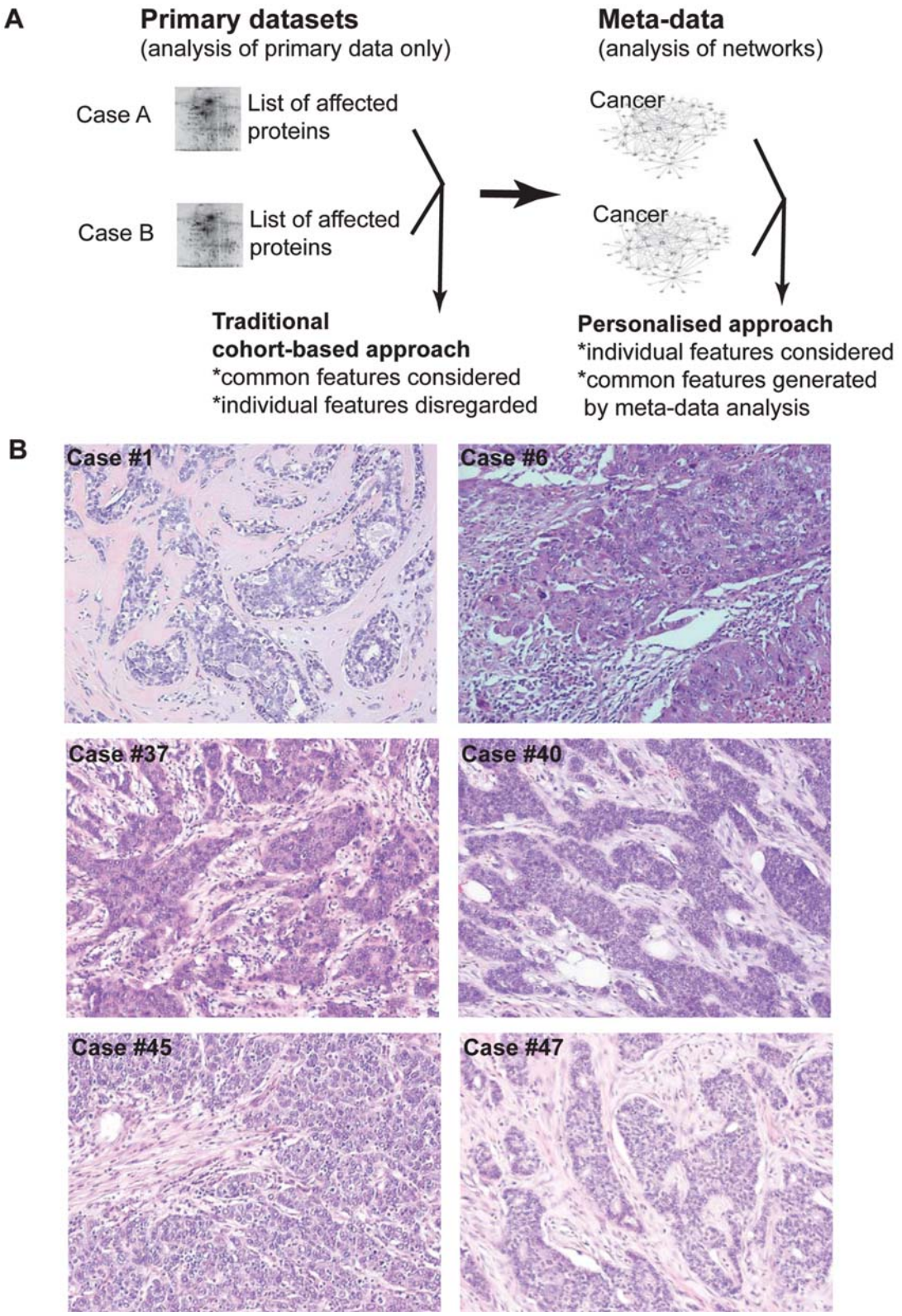

Figure 1. *Presentation of studied cases. A: The workflow of a traditional cohort-based analysis and the proposed approach. Only two cases are shown as examples. The workflow can be applied to unlimited number of patients. B: Haematoxylin-eosin-stained images of tumour sections representing cases used in this proteomics study (magnification: ×100). The cases shown are described in the Results section.*

Table II. *Summary of detection of protein spots and identified proteins.*

Case number[a]	Total number of affected protein spots[b]	Number of affected spots with identified proteins[c]	Proteins identified as being up-regulated		Total number of uniquely identified proteins[c]
			Tumours[c]	Normal tissues[c]	
#1	114	81	79	3	44
#6	141	100	86	24	54
#37	180	131	122	9	69
#40	116	52	29	23	31
#45	44	26	18	8	19
#47	46	46	38	8	34

[a]Case number is annotated in the text (Table I); [b]numbers of proteins spots were obtained following gel image analysis; [c]numbers of uniquely identified proteins. These numbers are lower than the numbers of spots with identified proteins due to identification of some of the proteins in multiple spots.

The key aspect of the strategy presented here is a full-scale proteomics study of each case separately, before making an analysis of the common and individual features. The results of individual proteomic profiling are briefly described below and detailed information may be requested from the authors (supplementary data).

Case #1. One hundred and fourteen protein spots were detected as showing changes in expression levels in the tumour tissue compared to the histologically healthy adjacent tissue. (Table II). Forty-four unique proteins in 81 spots were identified. Among them, CK2α, BRCA1, vimentin and annexin A2 were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving interferon β1, IL8, Erk1/2, Jnk, p53, ApoA1, CSF2 and BRCA1. The network formed by the tumour-related identified proteins for this case included 51 components.

Case #6. One hundred and forty-one protein spots were detected as showing changes in the expression levels in the tumour tissue compared to the histologically healthy adjacent tissue (Table II). Fifty-four unique proteins in 100 protein spots were identified. Among them, CK2α, PDGFRα, phospholipase C and protein tyrosine phosphatase 14 were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving TGFβ, TNF, insulin, TP73, JNK, Jun and HNF. The network formed by the tumour-related identified proteins included 147 components.

Case #37. One hundred and eighty protein spots were detected as showing changes in the expression levels in the tumour tissue compared to the histologically healthy adjacent tissue (Table II). Sixty-nine unique proteins were identified in 131 protein spots. Among them, CK2α, GDF2, RB binding protein 7, vimentin and annexin A2 were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving TP53, Fos, NFkB, ERK1/2, PDGF, TGFβ, TNF, insulin, PKC, HNF and AKT. The network formed by the tumour-related identified proteins included 144 components.

Case #40. One hundred and sixteen protein spots were detected as showing changes in expression levels in the tumour tissue compared to the histologically healthy adjacent tissue (Table II). Thirty-one unique proteins were identified in 52 protein spots. Among them, annexin A2 and phospholipase A2 activating protein were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving PDGF, MYC, TNF and HNF. The network formed by the tumour-related identified proteins included 122 components.

Case #45. Forty-four protein spots were detected as showing changes in expression levels in the tumour tissue compared to the histologically healthy adjacent tissue (Table II). Nineteen unique proteins were identified in 26 protein spots. Among them, CK2α, steroid 21-monooxygenase, annexin A2 and apolipoprotein A-IV precursor were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving β-estradiol, IL2, Il4, GRB2 and interferon γ. The network formed by the tumour-related identified proteins included 69 components.

Case #47. Forty-six protein spots were detected as showing changes in expression levels in the tumour tissue compared to the histologically healthy adjacent tissue (Table II). Thirty-three unique proteins were identified in 46 spots. Among them, ribosomal protein S6 kinase, protein kinase A anchor protein 2 and obscurin were identified. Systemic analysis of the identified proteins suggested changes in the regulatory processes involving TGFβ, TNF, Myc, interferon γ and CDK inhibitor p16. The network formed by the tumour-related identified proteins included 40 components.

Thus, proteomic profiling showed individual differences between cases, as primary datasets and as a prediction of interacting networks.

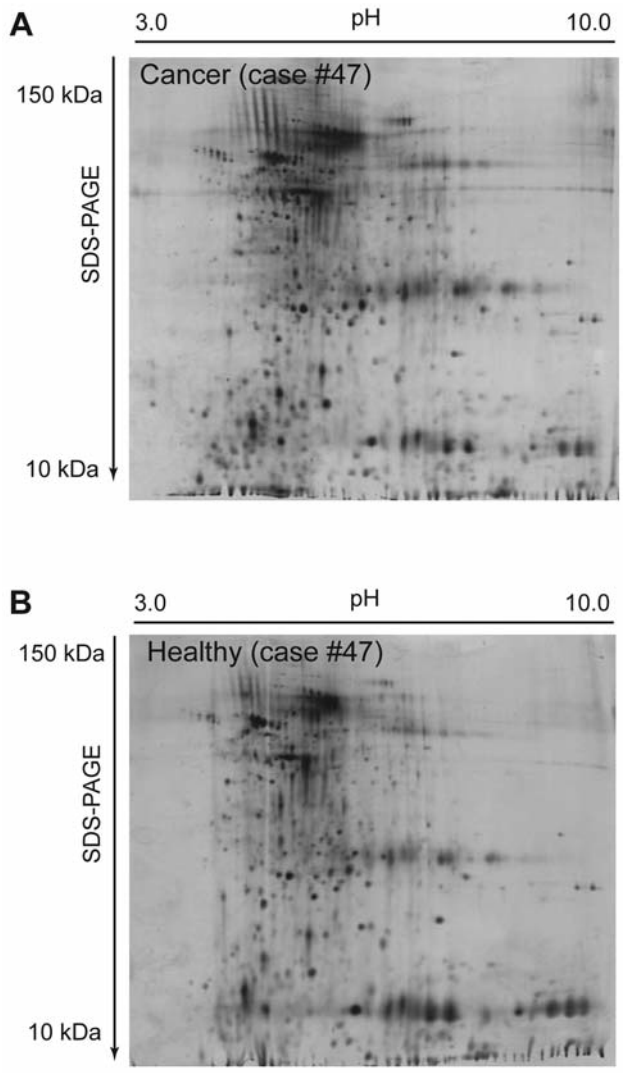

Figure 2. *Representative images of 2D gels. The images show separation of proteins extracted from tumour (A) and from histologically healthy adjacent tissue (B). The images represent gels generated with the samples of case #47. For each image, pH gradient, direction of SDS-PAGE and migration positions of molecular mass markers are shown. Images of annotated gels of all cases and lists of identified proteins are given in Supplementary data.*

Table III. *Summary of immunohistochemical detection of CK2α, PDGFRα, PYK and p53 in human breast cancer tissue microarray (healthy tissues and malignant tumours).*

CK2α staining	(–)	(+)	(++)	(+++)
Healthy (3)	2	1		
Malignant tumours (34)	2	5	9	18
PDGFRα staining	**(–)**	**(+)**	**(++)**	**(+++)**
Healthy (3)				3
Malignant tumours (35)			10	25
PYK staining	**(–)**	**(+)**	**(++)**	**(+++)**
Healthy (3)		2	1	
Malignant tumours (35)		4	22	9
p53 staining	**(–)**	**(+)**	**(++)**	**(+++)**
Healthy (3)		1	2	
Malignant tumours (35)	6	8	11	10

Immunohistochemical analysis and the grading for staining are described in the Materials and Methods section.

Generation of a common profile of deregulated signalling mechanisms. Analysis of primary proteomics datasets showed that different keratins were the only proteins common for all cases (keratins were common in 5 of 6 and 6 of 6 cases; Figure 3A). When the cut-off frequency of protein detection was decreased to 4 cases out of 6, TNF and TGFβ signalling were represented (Figure 3A). This was in contrast to a number of proteins with proven roles in intracellular signalling and tumourigenesis that were identified as cancer-related in each individual case.

The differences between the lists of identified proteins with altered expression levels in tumours, as compared to adjacent histologically healthy tissues, may be interpreted as a representation of the high variability between the cases. However, many of the regulatory mechanisms in a cell may employ different proteins to achieve the same impact on cellular functions, such as proliferation or death. Therefore, the regulatory mechanisms which may be deregulated in the tumour samples of the present study were investigated through the building of networks based on the identified proteins from each case. To predict which pathways may be involved, highly connected hubs were analysed in the individual networks. This analysis showed that TGFβ, TNF, mitogenic (EGF, PDGF, FGF) and interleukin (IL1, IL2, IL4, IL6 or IL8)-related signalling responses are overrepresented (Figure 3B). Various components of the generic MAP kinase cascade were also represented. This finding is in line with reports showing the involvement of known predicted mechanisms in tumourigenesis, such as proliferation, death, invasiveness, angiogenesis, stroma development and corruption of the immune surveillance (24). Therefore, despite differences in the primary datasets, the approach described here shows that there are significant similarities in the predicted signalling mechanisms deregulated in individual tumours.

Another important conclusion from this type of analysis was that the employed strategy allows the prediction of mechanisms which may have a more significant impact on tumourigenesis in each specific case (Figure 3C). As an example, the deregulation of BRCA1-dependent signalling

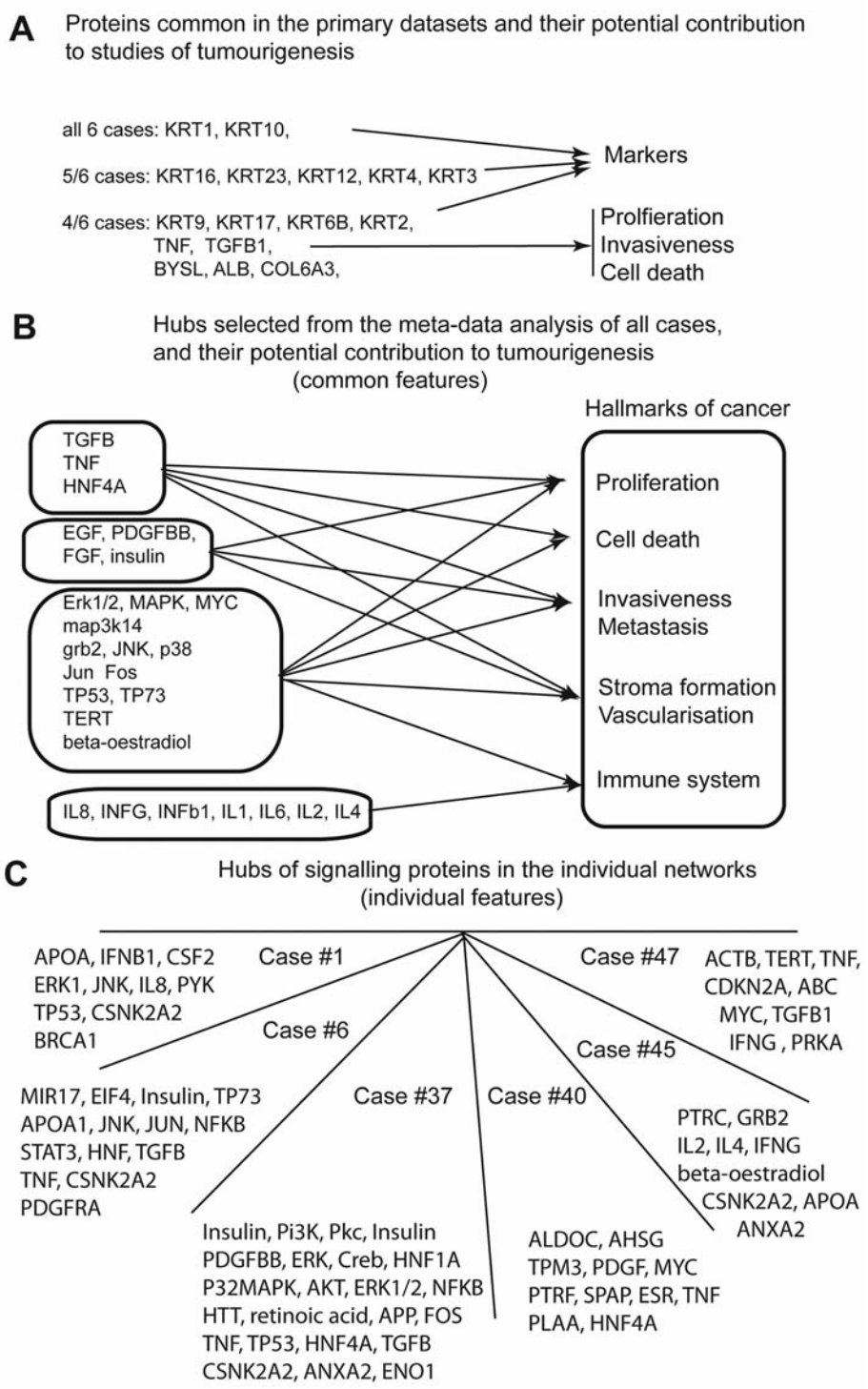

Figure 3. *Prediction of common and case-specific proteins. Proteins identified as common for all cases in primary datasets (A), hubs of networks as frequently affected in many cases (B) and as case-representative (C), are shown. A: Following a cohort-based approach, common proteins in the primary datasets were determined. The frequency of identification of proteins in each case is indicated. An impact on tumourigenesis was predicted by IPA and by the review of published reports (A, B). B: Hubs selected upon analysis of meta-data (network-based information) are shown in 4 groups. The main impacts of each group are indicated by arrows. It should be noted that all crucial for tumourigenesis regulatory mechanisms are represented, e.g. cell proliferation, cell death, metastasis, regulation of stroma and immune system. C: Selected hubs representing regulatory mechanisms in each studied case are shown. The annotation of proteins and hubs is in GO terms. The networks of each case and the networks of common primary dataset- and network-selected molecules are given in supplementary data.*

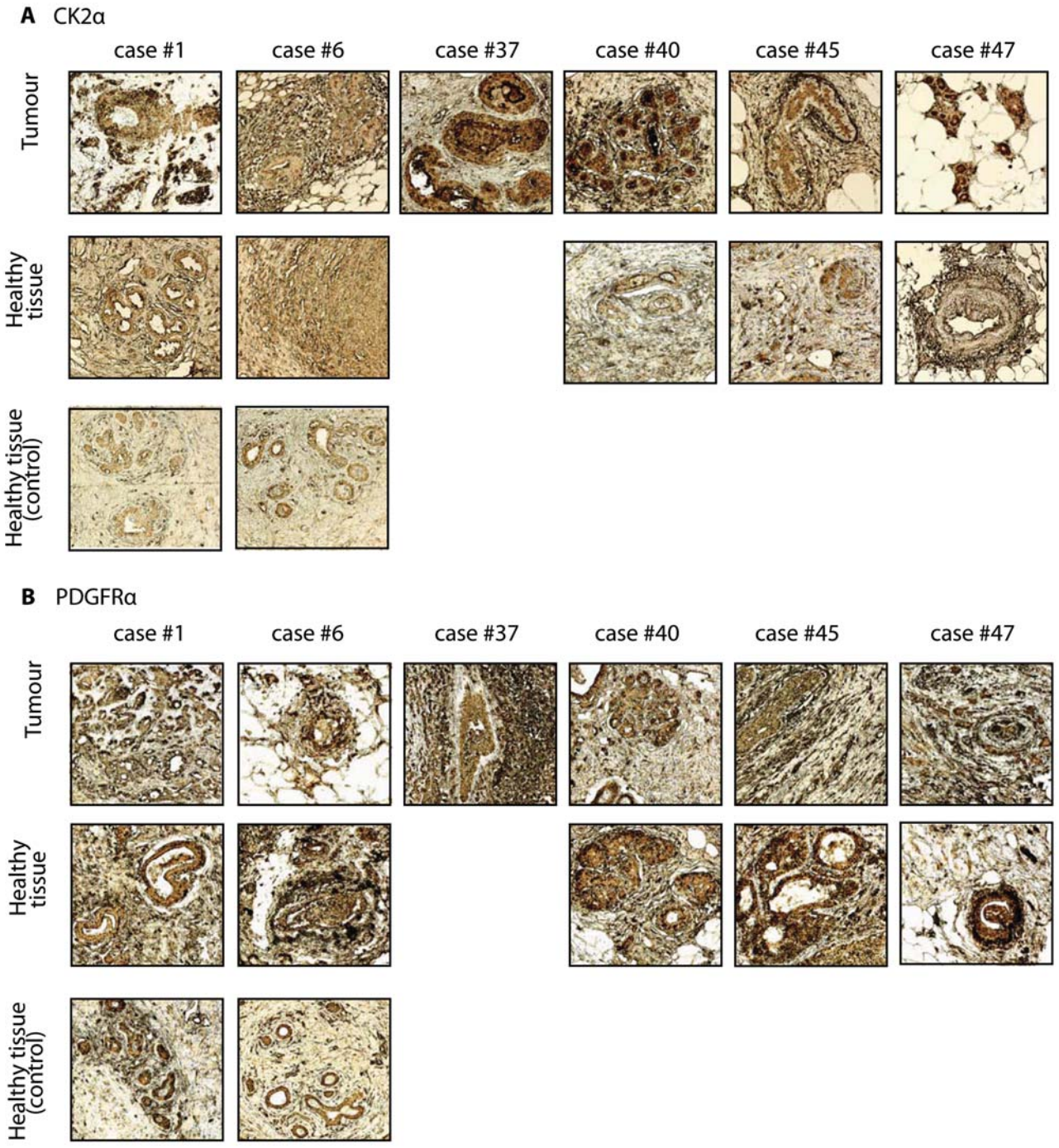

Figure 4. *Continued*

was suggested in the tumour of case #1. For the case #6, the status of TP73 may have a role in the growth of this tumour. For the case #47, areas of DCIS were observed in addition to IDC and for this case, telomerase reverse transcriptase (TERT) was predicted as a highly connected hub, indicating changes relevant to early stages of tumourigenesis. Other examples of proteins with predicted impact on tumourigenesis in individual cases were CK2α, pyruvate kinase M1/M2 (PYK), p53 and TGFβ and PDGF signalling (Figure 3C). The methodology described here allows the generation of predictions based on

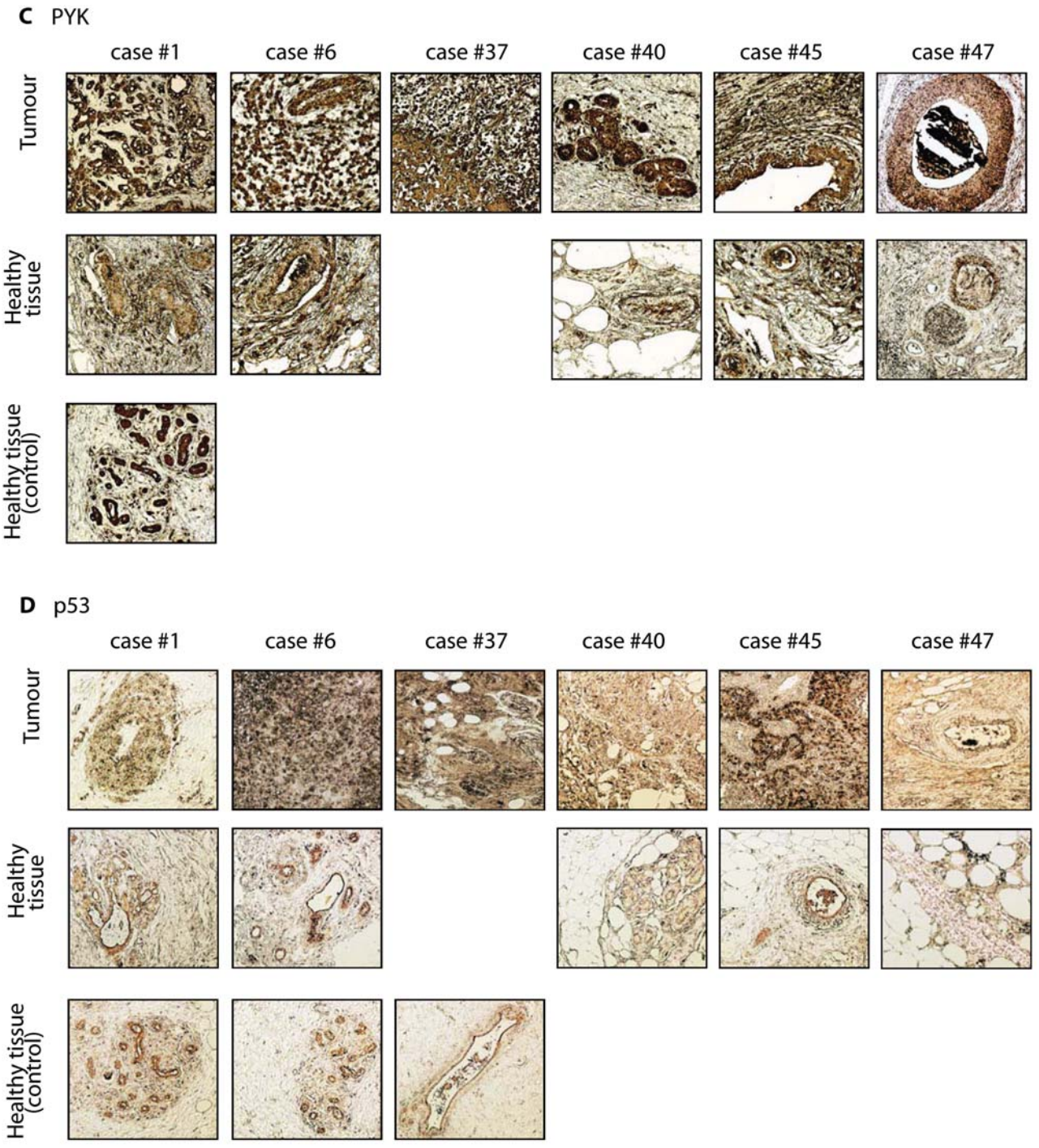

Figure 4. *Continued*

targets that are deregulated in individual tumours. This approach is crucial for gaining a greater understanding of the underlying mechanisms in individual tumours. Furthermore, this information may be essential in developing a more personalised regime of treatment options for patients.

Validation of common and individual features of tumours by immunohistochemistry. To validate the 2D-GE based findings immunohistochemistry (IHC) was performed on the samples of cases subjected to the initial proteomic profiling (Figure 4) and through the use of a tissue microarray (TMA) (Figure 5).

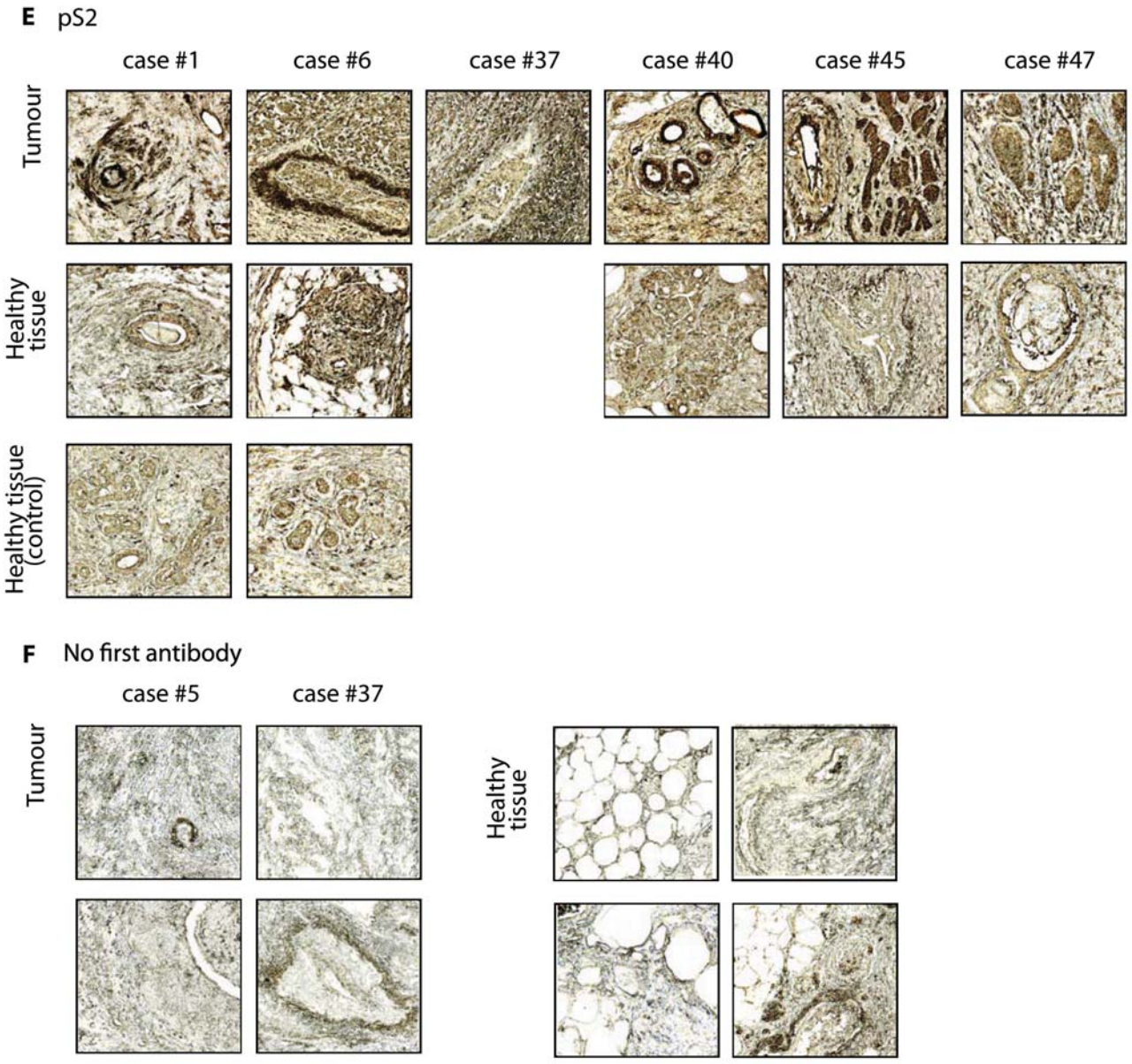

Figure 4. *Expression of CK2α and PDGFRα in the studied cases. Expression of CK2α (A) and PDGFRα (B), PYK (C), p53 (D) and pS2 (E) in tumour and healthy adjacent tissues (healthy tissue) was monitored by IHC. Healthy breast tissue from non-cancerous patients (healthy tissue control) was also stained. Control of the staining, without primary antibodies, is shown in (F). Brown colour indicates positive staining. Case #37 did not have adjacent histologically healthy tissue available for IHC. Representative images are shown (magnification: ×50).*

The TMA contained 35 cases of malignant tumours, 3 cases of fibroadenomas, 6 cases of non-malignant conditions (*e.g.* hyperplasia) and 3 cases of healthy breast tissues (supplementary data). In contrast to immunoblotting of extracts from whole tumour or tissue, IHC allows the evaluation of the expression of proteins in different cell types. Therefore, IHC is a good methodological approach to demonstrate whether the findings observed from the proteomic profiling are specific to the malignant cells of tumours.

For validation, the levels of expression of CK2α, PDGFRα, PYK, p53 and TGFβ receptor-induced phosphorylation of Smad2 protein were assessed (Figure 4). These proteins were selected due to their identification in the proteome profiling primary datasets, by network analysis and their potential involvement in breast tumourigenesis (Figure 3; supplementary data) (25-30). These proteins are known to regulate tumourigenesis-related processes, but they are not accepted in clinic as markers. One of the reasons may be that

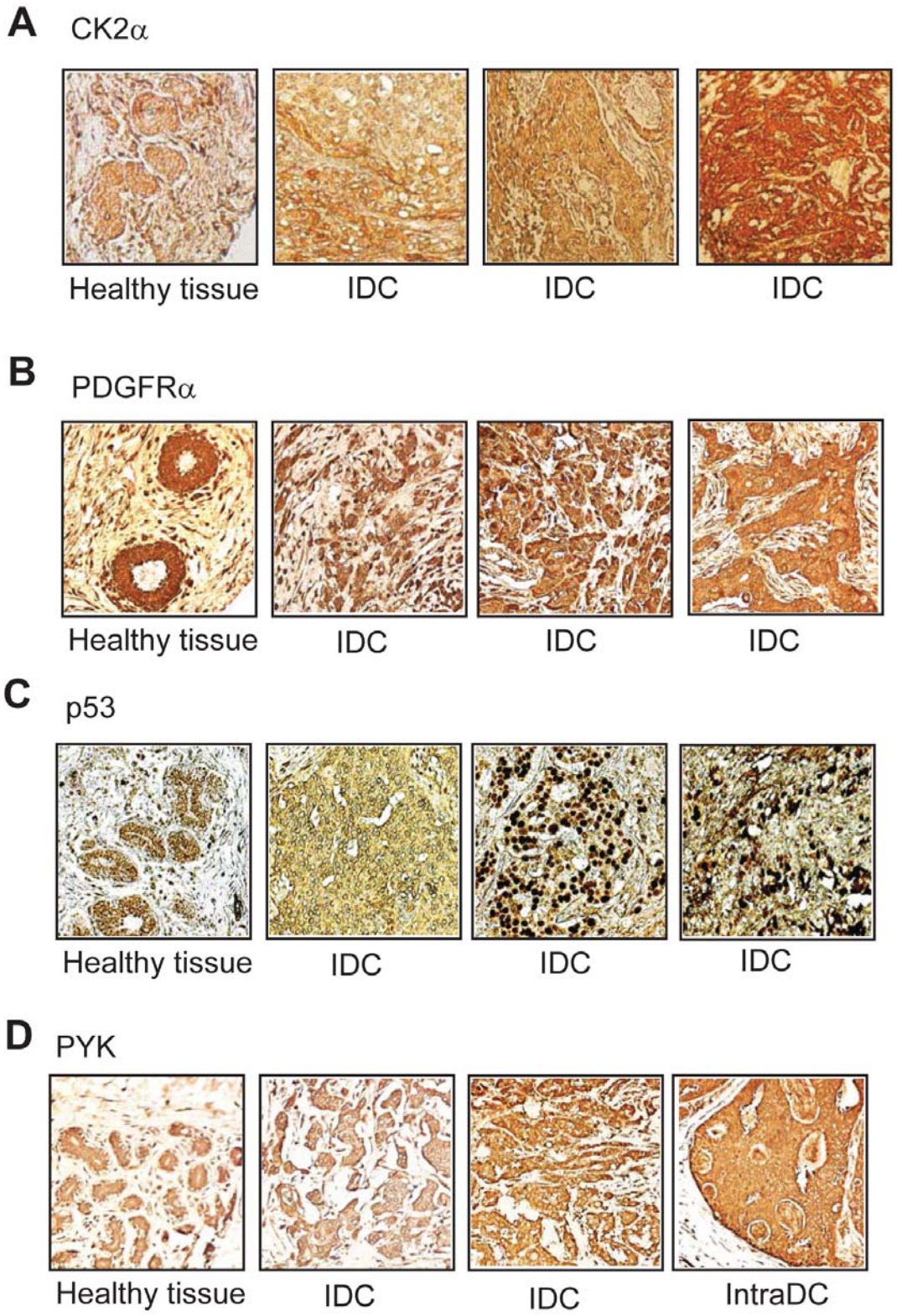

Figure 5. *Expression of CK2α, PDGFRα, PYK and p53 in human breast tumours. The expression of CK2α (A), PDGFRα (B), PYK (C) and p53 (D) in IDC and healthy tissues is shown. TMAs of human breast malignant and benign tumours, and healthy tissues were stained with respective antibodies. Representative images are shown, where brown colour indicates positive staining (magnification: ×50). Case #37 did not have enough histologically healthy tissue for IHC.*

their correlation to tumourigenesis may not be high in a large cohort study, but may be highly relevant for the individual cases. Therefore, the expression of these proteins is expected to alter in cancer, but with significant variability between individual cases.

IHC staining of sections of the studied cases with anti-CK2α, anti-PDGFRα, anti-PYK, anti-p53 and anti-phoshorylated Smad2 antibodies confirmed the proteomics and network analysis results. Notably, the expression of CK2α was enhanced in all tumours; however it showed varying levels in staining between the individual cases, with an increased expression in tumour cells (Figure 4A). PDGFRα expression also showed variable staining among the cases, with a significant staining of epithelial cells in histologically healthy adjacent tissues. Compared to histologically healthy tissues, PDGFRα staining was similar or less pronounced in tumour cells, although the total PDGFRα signal was enhanced in tumour sections (Figure 4B). PYK staining was increased in tumours, as compared to adjacent histologically healthy tissues (Figure 4C). p53 staining also showed tumour-related changes, with moderate (cases #1, #40 and #47) to strong (cases #6 and #45; Figure 4D) signal increase. TGFβ signalling was identified by the two-step strategy as deregulated in the studied cases. IHC showed that the activated *C*-terminal phosphorylation of Smad2 is enhanced in tumour cells, as compared to histologically healthy adjacent tissue (Figure 4E) Therefore the IHC staining of the individual cases confirmed the deregulation of the identified and predicted proteins, and showed them to be relevant to breast tumour tissues.

To explore whether the observed deregulation of specific proteins would be observed similarly in new cases of breast cancer, IHC staining was performed on a TMA set of human breast cancers with focus on healthy tissues and malignant tumours, *e.g.* IDC (Figure 5). Results of IHC staining of non-malignant cases in TMA are mentioned in supplementary data. IHC staining of the TMAs showed that CK2α expression is increased in almost half of IDC cases, as compared to weak or no expression in benign neoplasias and healthy tissues (Table III; Figure 5A). Thus, the deregulation of expression of CK2α may be characteristic for a proportion of tumours. Case-to-case variability in staining for PDGFRα and PYK was also observed (Figure 5B, C; Table III). An evaluation of IHC staining for PDGFRα and PYK based on staining intensities showed tumour-related changes in less than 30% of cases (Table III). IHC staining for p53 showed that in IDC cases the expression level of p53 is also de-regulated. Notably, a moderate expression of p53 was observed in healthy tissues, while in IDC there were cases with no detectable p53 (6 cases) and cases with strong expression (10 cases) In a cohort-based study, the levels of changes observed for CK2α, PDGFRα, PYK and p53 would not be considered as representative for the whole cohort, despite the fact that these changes may be relevant for individual patients. The relevance to individual patients is even more pronounced as the studied proteins are potent regulators of cellular functions and are known to affect tumourigenesis. Therefore, the results suggest that many of the changes in regulatory processes may not be random events but characteristic for the development of breast tumours in individual patients. Identification of such individual traits in tumour development would be beneficial for the individualisation of anticancer treatment.

Discussion

Studies of genome, transcriptome and proteome changes in human breast cancer have delivered a number of markers for detection, selection of treatment and prognosis (2-6, 31). The main trait of previously reported studies is a search for 'common for all cases' markers, which would have acceptable sensitivity and specificity. The drawback of this approach is that individual differences in primary data would be lost, and only common features would be considered. This cohort-based approach does not take into consideration systemic properties of cellular functions. Multiplicity of ways to control cellular functions is the basic principle of cell physiology, and it is ensured by a similar impact of different proteins on an identical signalling mechanism (32). In its application to cancer, this means that even if different sets of proteins would be identified as cancer-specific in different tumours, they may reflect de-regulations of the same cellular functions. This similarity will be visible only if a systemic analysis is performed with primary datasets, and then meta-data compared. In addition, systemic analysis of individual cases allows identification of proteins and signalling pathways specific for that patient. This was a pilot study which used a two-step approach to identify breast cancer markers; firstly using an individual proteome profiling and systemic analysis, followed by a case-specific meta-data analysis for all cases. This is the first report of such a two-step approach in the search of breast cancer related markers for their potential use in the management of breast cancer.

Proteomic profiling of breast tumours and cultured cells established from human breast epithelial cells have delivered lists of potentially cancer-specific proteins (2-8, 10). However, comparison of these lists showed that common proteins were predominantly of high abundance, *e.g.* keratins. At the top of the list of common cancer-related proteins were also keratins (Figure 3A). This is in contradiction with results of molecular studies of breast carcinogenesis, when a number of involved signalling pathways have been described (4, 9, 10, 24, 30). Proteins directly involved in these pathways have been seldom proposed as markers, with the exception of HER2/neu, oestrogen and progesterone receptors, p53, BRCA1 and

BRCA2 (1-6). Recent reports indicated that even these molecular markers are not always efficient predictors, probably due to their mutations and intracellular compensatory mechanisms (1-4). Studies of signalling pathways involved in breast tumourigenesis indicated that the possible reason for such a disproportion in output between signalling and marker studies may be in the multiplicity of cellular regulatory mechanisms. When the potential functional impact of components identified by systemic analysis of individual cases was analysed, it was found that practically all cancer hallmarks were represented (Figure 3B). This confirms that the described approach allowed gaining a more comprehensive overview of molecular changes in tumour proteomes, as compared to conclusions based on primary datasets only.

Furthermore, the developed approach allowed for the identification of regulatory mechanisms specific for individual patients (Figure 3C). Meta-data analysis predicted changes in regulatory processes which otherwise would not be detectable by a direct analysis of only identified proteins. Knowledge of these mechanisms is important for the selection of patient treatment, as it provides information about the status of potential drug targets. The IHC validation study confirmed that the observed changes in the proteome profiles are not random events, but may be specific for a subset of tumours. The size of such subsets would not be large, with up to 10% or 50% of all cases. However, as these changes may be relevant to an individual patient, to know these unique specifics would be of great importance when designing anticancer treatment regimes. The developed two-step methodology with the analysis of meta-data was a pilot study established to evaluate the feasibility of this approach. Further studies with a large cohort of patients are required to enable the introduction of this approach into the clinical practice.

This pilot study proposed that a two-step strategy in the analysis of proteomic profiles of human breast tumours is more informative in providing insight into affected molecular mechanisms than an analysis of only primary datasets. The first step was a full-scale proteomic profiling of each case separately. The second step was a comparison of meta-data from all cases. The findings from the analysis of the proteomic profiling and validation experiments reinforced the value of such a two-step approach for the development of more personalised medicinal regimes

Acknowledgements

We are grateful to the patients that consented to donate their tissues and to the surgeons (Professor Paul Sauven and Dr. Simon Smith) and the pathologist (Dr. Khalid-Al Janabi) for their expert help. This work was supported by grants from the Swedish Cancer Research foundation, the Swedish Research Council, RTN EpiPlast Carcinoma, the Swedish Institute, KI/KBC to S.S, and by the Helen Rollason Charity support.

References

1 Weigelt B and Reis-Filho JS: Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat Rev Clin Oncol *6*: 718-730, 2009.

2 Kataja V and Castiglione M: ESMO Guidelines Working Group: Primary breast cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol *20(S4)*: 10-14, 2009.

3 Cardoso F and Castiglione M: ESMO Guidelines Working Group: Locally recurrent or metastatic breast cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol *20(S4)*: 15-18, 2009.

4 Moulder S and Hortobagyi GN: Advances in the treatment of breast cancer. Clin Pharmacol Ther *83*: 26-36, 2008.

5 Pfeffer U, Romeo F, Noonan DM and Albini A: Prediction of breast cancer metastasis by genomic profiling: where do we stand? Clin Exp Metastasis *26*: 547-558, 2009.

6 Zoon CK, Starker EQ, Wilson AM, Emmert-Buck MR, Libutti SK and Tangrea MA: Current molecular diagnostics of breast cancer and the potential incorporation of microRNA. Expert Rev Mol Diagn *9*: 455-467, 2009.

7 Sauer G, Schneiderhan-Marra N, Kazmaier C, Hutzel K, Koretz K, Muche R, Kreienberg R, Joos T and Deissler H: Prediction of nodal involvement in breast cancer based on multiparametric protein analyses from preoperative core needle biopsies of the primary lesion. Clin Cancer Res *14*: 3345-3353, 2008.

8 Steinman S, Wang J, Bourne P, Yang Q and Tang P: Expression of cytokeratin markers, ER-alpha, PR, HER-2/neu, and EGFR in pure ductal carcinoma *in situ* (DCIS) and DCIS with co-existing invasive ductal carcinoma (IDC) of the breast. Ann Clin Lab Sci *37*: 127-134, 2007.

9 Hanash S: Disease proteomics. Nature *422*: 226-232, 2003.

10 Hondermarck H, Tastet C, El Yazidi-Belkoura I, Toillon RA and Le Bourhis X: Proteomics of breast cancer: the quest for markers and therapeutic targets. J Proteome Res *7*: 1403-1411, 2008.

11 Alvarez RH, Valero V and Hortobagyi GN: Emerging targeted therapies for breast cancer. J Clin Oncol *28*: 3366-3379, 2010.

12 Li J, Gromov P, Gromova I, Moreira JM, Timmermans-Wielenga V, Rank F, Wang K, Li S, Li H, Wiuf C, Yang H, Zhang X, Bolund L and Celis JE: Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. Proteomics *8*: 5038-5052, 2008.

13 Gromov P, Gromova I, Bunkenborg J, Cabezon T, Moreira JM, Timmermans-Wielenga V, Roepstorff P, Rank F and Celis JE: Up-regulated proteins in the fluid bathing the tumour cell microenvironment as potential serological markers for early detection of cancer of the breast. Mol Oncol *4*: 65-89, 2010.

14 Johann DJ, Rodriguez-Canales J, Mukherjee S, Prieto DA, Hanson JC, Emmert-Buck M and Blonder J: Approaching solid tumour heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. J Proteome Res *8*: 2310-2318, 2009.

15 Jia M, Souchelnytskyi N, Hellman U, O'Hare M, Jat P and Souchelnytskyi S: Proteome profiling of immortalization-to-senescence transition of human breast epithelial cells identified MAP2K3 as a senescence-promoting protein which is down-regulated in human breast cancer. Proteomics Clin Appl *4*: doi: 10.1002/prca.201000006, 2010.

16 Celis JE, Gromova I, Cabezón T, Gromov P, Shen T, Timmermans-Wielenga V, Rank F and Moreira JM: Identification of a subset of breast carcinomas characterized by expression of cytokeratin 15: relationship between CK15+ progenitor/amplified cells and pre-malignant lesions and invasive disease. Mol Oncol *1*: 321-349, 2007.
17 Nimeus E, Malmström J, Johnsson A, Marko-Varga G and Fernö M: Proteomics analysis identifies candidate proteins associated with distant reccurences in breast cancer after adjuvant chemotherapy. J Pharmaceut Biomed Anal *43*: 1086-1093, 2007.
18 Metodiev M and Alldridge L: Phosphoproteomics: a possible route to novel biomarkers of breast cancer. Proteomics Clin Appl *2*: 181-194, 2008.
19 Agarwal R, Gonzalez-Angulo AM, Myhre S, Carey M, Lee JS, Overgaard J, Alsner J, Stemke-Hale K, Lluch A, Neve RM, Kuo WL, Sorlie T, Sahin A, Valero V, Keyomarsi K, Gray JW, Borresen-Dale AL, Mills GB and Hennessy BT: Integrative analysis of cyclin protein levels identifies cyclin B1 as a classifier and predictor of outcomes in breast cancer. Clin Cancer Res *15*: 3654-3662, 2009.
20 Conrotto P and Souchelnytskyi S: Proteomic approaches in biological and medical sciences: principles and applications. Exp Oncol *30*: 171-180, 2008.
21 Weiss W and Görg A: High-resolution two-dimensional electrophoresis. Methods Mol Biol *564*: 13-32, 2009.
22 Goodison S and Urquidi V: Breast tumour metastasis: analysis via proteomic profiling. Expert Rev Proteomics *5*: 457-467, 2008.
23 Stasyk T, Dubrovska A, Lomnytska M, Yakymovych I, Wernstedt C, Heldin CH, Hellman U and Souchelnytskyi S: Phosphoproteome profiling of transforming growth factor (TGF)-beta signalling: abrogation of TGFbeta1-dependent phosphorylation of transcription factor-II-I (TFII-I) enhances cooperation of TFII-I and Smad3 in transcription. Mol Biol Cell *16*: 4765-4780, 2005.
24 Hanahan D and Weinberg RA: The hallmarks of cancer. Cell *100*: 57-70, 2000.
25 Romieu-Mourez R, Landesman-Bollag E, Seldin DC and Sonenshein GE: Protein kinase CK2 promotes aberrant activation of nuclear factor-kappaB, transformed phenotype, and survival of breast cancer cells. Cancer Res *62*: 6770-6778, 2002.
26 Carvalho I, Milanezi F, Martins A, Reis RM and Schmitt F: Overexpression of platelet-derived growth factor receptor alpha in breast cancer is associated with tumour progression. Breast Cancer Res *7*: R788-795, 2005.
27 Jerry DJ, Dunphy KA and Hagen MJ: Estrogens, regulation of p53 and breast cancer risk: a balancing act. Cell Mol Life Sci *67*: 1017-1023, 2010.
28 Mazurek S: Pyruvate kinase type M2: a key regulator of the budget metabolic system in tumour cells. Int J Biochem Cell Biol doi:10.1016/j.biocel.2010.02.005, 2010.
29 Barcellos-Hoff MH and Akhurst RJ: Transforming growth factor-beta in breast cancer: too much, too late. Breast Cancer Res *11*: 202, 2009.
30 Souchelnytskyi S: Proteomics of TGF-beta signalling and its impact on breast cancer. Expert Rev Prot *2*: 925-935, 2005.
31 Lomnytska M, Dubrovska A, Hellman U, Volodko N and Souchelnytskyi S: Increased expression of cSHMT, Tbx3 and utrophin in plasma of ovarian and breast cancer patients. Int J Cancer *118*: 412-421, 2006.
32 Kitano H: Cancer as a robust system: implications for anticancer therapy. Nat Rev Cancer *4*: 227-235, 2004.

Received September 17, 2010
Revised October 7, 2010
Accepted October 8, 2010