Proteomic portrait of human breast cancer progression identifies novel prognostic markers

Tamar Geiger¹, Stephen F. Madden², William M. Gallagher³, Juergen Cox¹ and Matthias Mann¹

¹-Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

²- National Institute for Cellular Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland

³- UCD school of Biomolecular and Biomedical Science, UCD Conway Institute, Belfield, Dublin 4, Ireland

- Current address: Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Contact: M.M. mmann@biochem.mpg.de

Running title
Proteomics of breast cancer progression

Precis

In performing the deepest proteomic analysis of breast cancer progression to date this study identifies novel prognostic markers for overall survival which function in metabolic and secretory processes.

Keywords
Proteomics, mass spectrometry, breast cancer, biomarker, SILAC

**Summary**

Breast cancer is the second leading cause of cancer death for women in the United States. Of the different subtypes, estrogen receptor negative (ER-) tumors, which are ErbB2+ or triple-negative, carry a relatively poor prognosis. In this study we used system-wide analysis of breast cancer proteomes to identify proteins that are associated with the progression of ER- tumors. Our two-step approach included an initial deep analysis of cultured cells that were obtained from tumors of defined breast cancer stages, followed by a validation set using human breast tumors. Using high-resolution mass spectrometry and quantification by Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC), we identified 8750 proteins and quantified 7800 of them. A stage-specific signature was extracted and validated by mass spectrometry and immunohistochemistry on tissue microarrays. Overall, the proteomics signature reflected both a global loss of tissue architecture and a number of metabolic changes in the transformed cells. Proteomic analysis also identified high levels of IDH2 and CRABP2 and low levels of SEC14L2 to be prognostic markers for overall breast cancer survival. Together, our findings suggest that global proteomic analysis provides information regarding the protein changes specific to ER- breast tumor progression as well as important prognostic information.

**Introduction**

The transformation of a normal somatic cell into an immortalized and later into an invasive cancer cell involves dramatic phenotypic changes including changes in cell proliferation rate, motility, metabolism, genomic stability and survival. Patient prognosis largely depends on the cancer stage, which is assessed by tumor size, number of involved lymph nodes and metastases (TNM stage) (1-3). To date, critical
molecular events in breast cancer progression that have dramatic effects on disease outcome are still poorly characterized. Identification of such molecular changes can be achieved by system-wide and unbiased approaches.

In the past decade numerous studies analyzed transcriptomes of breast cancer cell lines and tumor samples using cDNA arrays. Such transcriptomic analyses classified breast tumors into five distinct subtypes: Basal-like tumors, luminal A, luminal B, ErbB2-over-expressing and normal breast-like tumors (4-6). Luminal A tumors are estrogen receptor-positive (ER+) and have better prognosis, whereas the basal-like and ErbB2+ are ER- and have poorer prognosis (6). Similar to the tumor transcriptome studies, classification has also been performed on breast cancer cell lines (7-9). As in the tumor samples, there is a discrimination between basal and luminal cells, with small differences in the sub-classification.

Analysis of the proteins rather than the mRNA levels may reflect the functional phenotype of the cells more directly. Moreover, some proportion of the changes at the genome as well as at the transcriptome levels are eliminated by higher regulatory mechanisms (10). However, measuring the proteome is technologically much more challenging than transcriptome analysis. As a result, in contrast to the extensive mRNA work, proteomics studies so far have usually analyzed only few samples, with limited proteome coverage and often with inaccurate quantification and therefore they did not provide a true global view of the system (11). For instance, two-dimensional gel electrophoresis has been used in cancer proteomics, but this technique enabled analysis of only the most abundant proteins, and generally with low quantitative accuracy. Mass spectrometry-based proteomics, particularly in a high resolution and quantitative format, has developed rapidly over the last few years (12). Hybrid mass spectrometers - such as the linear ion trap-Orbitrap - combine high resolution, high mass accuracy and high peptide sequencing speed (13 ). Together with innovations in sample preparation and computational proteomics, these technologies can enable confident peptide and protein identification
and quantification at a large scale. In our laboratory these advances allow routine coverage of 5000-7000 proteins from mammalian cells (14). We hypothesized that the combination of confident identification, high proteome coverage and accurate quantification using Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) (15), could make proteomics applicable to system-wide analysis of cancer proteomes, and could highlight proteins and processes that are altered during cancer progression.

We took a two-step approach starting with deep analysis of the proteomes of cultured cells that were isolated from tumors of different stages using state of the art mass spectrometric technology. The analysis of cell lines eliminates the effects of diverse cell populations in the tissue that may mask the changes in the cancer cells and the controlled growth environment of cell lines further reduces the variability compared to human tissue samples. These advantages enable conclusions from a smaller number of samples, and are therefore more suitable for proteomics, which still has limited throughput.

We extracted a stage-specific proteomic signature and validated the results using a directed mass spectrometric approach and IHC using tumor arrays. Examination of the signature proteins in gene expression studies of large patient cohorts identified IDH2 and CRABP2 as markers of poor prognosis and SEC14L2 as a marker of good prognosis.

**Materials and Methods**

**Cell culture and SILAC labeling**

Human mammary epithelial cells (HMEC) were obtained from Lonza and from the European Collection of Cell Cultures (ECACC); HMT-3522-S1 (16), MFM223 (17) were obtained from the ECACC; HCC202 and HCC2218 cells (18) were obtained from the American Type Culture Collection (ATCC); HCC1599, HCC1143, HCC1937 (18) and MCF7 cells were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ); MCF10a (19) and MDA-MB-453 (20) were kindly provided by Axel Ullrich
(Max-Planck Institute of Biochemistry, Martinsried) and were tested negative for mycoplasma contamination using DNA tests. Cell lines were purchased from the cell banks during 2008 and were authenticated by them using DNA tests and micorobiological cultures. Cells were used 1-6 months after arrival to the laboratory. HMEC cells were cultured in mammary epithelial cell growth medium (ECACC); HCC1599, HCC1143, HCC1937, HCC202, HCC2218 were grown in RPMI supplemented with 10% FBS; MCF10a cells were cultured in DMEM/F12 supplemented with 5% horse serum, 20 ng/ml EGF, 10 µg/ml insulin, 0.5 µg/ml hydrocortisone and 0.1 µg/ml cholera toxin. HMT-3522-S1 cells were cultured in DMEM/F12 supplemented with 250 ng/ml insulin, 10 µg/ml transferrin, 0.1 µM sodium selenite, 0.1 nM 17 beta-estradiol, 5 µg/ml ovine prolactin, 0.5 µg/ml hydrocortisone and 10 ng/ml EGF; MDA-MB-453 cells were cultured in L-15 supplemented with 10% FBS; MFM223 cells were grown in MEM supplemented with 10% FBS. All cells were cultured with Pen/Strep and under 5% CO2, except for MDA-MB-453 which were cultured under 0% CO2.

MCF7 cells were SILAC labeled with Arg10 and Lys8 by culturing them for 8 doublings in the SILAC medium to reach complete labeling. For proteomic analysis each of the cell lines was analyzed in three biological replicates. The first two replicates were lysed with modified RIPA buffer (50 mM Tris HCl pH7.4, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.25% sodium deoxycholate and protease inhibitors) at 4°C. For lysis of the cells of the third replicate we used an improved method that enables better yield of membrane proteins, with a buffer containing 4% SDS, 100 mM Tris-HCl pH 7.6 and 100 mM DTT. For tumor analysis we used the super-SILAC mix as described previously (21).

**Tissue sample preparation**

Normal and tumor tissues were kindly provided by René Bernards (NKI, Amsterdam). Analysis of the samples followed an informed consent approved by the local ethics committee. Tissue slices from snap frozen tissue samples were lysed with 4% SDS, 100 mM Tris-HCl pH 7.6 and 100 mM DTT.
Trypsin digestion

Each of the non-labeled samples (HMEC, MCF10a, HMT-3522-S1, HCC1937, HCC1143, HCC1599, HCC202 HCC2218, MFM223 and MDA-MB-453) was mixed with SILAC-labeled MCF7 cells at a 1:1 protein ratio. For tissue analysis the super-SILAC mix was combined with equal protein amount of each of the tissue samples. Two methods were used for trypsin digest: In-solution digestion was used for the first two replicates of the cell line analysis, where cells were lysed with RIPA buffer and Filter Aided Sample Preparation (FASP) (14) was used for the third cell line replicate and for the tissue/super-SILAC digestion, when lysis was done with SDS-based buffer.

Peptide fractionation

In the cell line experiments peptides were separated using an Agilent 3100 OFFGEL fractionator (Agilent,G3100AA) as described previously (22). Prior to LC-MS analysis peptides were concentrated and desalted on C18 StageTips. The tumor/super-SILAC peptides were separated by strong anion exchange in a StageTip format as described previously (23). Peptides were separated to six fractions with buffers of different pH values. Peptides were concentrated and purified on C18 StageTips prior to LC-MS analysis.

LC-MS analysis

For the cell-line analyses, peptides were separated by reverse-phase chromatography using a nanoflow HPLC system (Thermo Fisher Scientific) with a 90 min linear gradient of water/acetonitrile. HPLC was coupled on-line to an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). Fragmentation of the top five peptides in each scan was performed by collision-induced dissociation. For the tumor sample analyses, peptides were eluted with a 190 min linear gradient and the MS analysis was performed on an LTQ-Orbitrap Velos instrument with MS/MS selecting the top 10 precursor m/z values from an inclusion
list. The m/z values that were used in the inclusion list are given in Table S1. Peptides were fragmented by higher energy collisional dissociation (HCD).

Data analysis

Raw MS files from the LTQ-Orbitrap were analyzed by MaxQuant (version 1.1.1.9)(24). MS/MS spectra were searched against the decoy IPI-human database version 3.68 containing both forward and reverse protein sequences by the Andromeda search engine (25). For identification, the false discovery rate (FDR) was set to 0.01 on the protein and on the peptide levels. Complete protein and peptides lists are given as Supplementary Table 2.

Statistical analysis

All the statistical analyses of the MaxQuant protein tables were performed with the Perseus program (J.C. in preparation). For hierarchical clustering we filtered the data and kept proteins with a minimum of five ratio values from the 11 cell lines. Logarithmized ratios towards the internal standard were z-scored and clustered using Euclidean distances between averages. For ANOVA test experimental systems were grouped according to their stage, and the statistical test was performed with FDR=0.05 and $S_0=1$ (26). The $S_0$ factor was described by Tusher et al. for t-test and was here generalized for ANOVA test. Fisher exact tests were performed with a Benjamini-Hochberg FDR threshold of 0.02.

Tumor arrays

Breast cancer tissue arrays were obtained from Pantomics, Inc. and consisted of 75 breast tumor and normal breast tissues in duplicates. Primary antibodies: anti-IDH2, anti-CRABP2 and anti-ANX3 were kindly provided by the Human Protein Atlas. We performed semi-quantitative scoring of the intensity of staining using four values (0-3) for negative, low, medium and high staining intensities.

Gene expression data
8 dataset comprising 1467 samples were downloaded from the Gene Expression Omnibus and from the Stanford microarray database. 20 samples lacking clinical information were removed. Where raw data was not available the published normalized data was used. The Affymetrix datasets were quantile normalized and dual channel platforms were loess normalized (a locally weighted polynomial regression method) (27, 28). Probes were mapped to Entrez gene IDs to gene center the data (29). Entrez gene IDs for genes of interest were obtained from the gene database at NCBI. All calculations were carried out in the R statistical environment (30).

**Survival Analysis**

For each of the 52 genes median mRNA expression levels were used to determine high and low expression groups within each of the 8 individual datasets. The survival curve was based on Kaplan-Meier estimates and the log-rank p-value is shown for difference in survival. The p-values are adjusted for multiple testing using a Bonferroni correction. Cox regression analysis was used to calculate hazard ratios. The R package survival was used for all calculations and to plot the Kaplan-Meier survival curves.

**Results**

**A cell culture model for breast cancer development**

To characterize the differences in the proteomes of ductal carcinomas of various stages, we assembled a panel of cell lines that were isolated from human tumors with a defined TNM stage. We aimed to identify molecular markers and cellular processes characteristic of specific stages in the transformation process rather than of a particular cell line or an individual patient. We therefore included two to three cell lines from each stage, derived from the tumors of different cancer patients that were not previously treated. As control cells that represent the healthy tissue we used primary mammary epithelial cells (HMEC) from two different sources. Pre-malignant cells were represented by MCF10a and HMT-3522-S1.
cells, stage II tumors by HCC1143 and HCC1937 cells, stage III tumors by HCC202, HCC2218 and HCC1599 cells and metastatic cells from pleural-effusions by MFM223 and MDA-MB-453 cells (Fig 1). Together, this panel of cells models the transformation process toward development of ER- tumors that are basal-A, luminal or ErbB2 over-expressing. In microarray studies HCC1143, HCC1937 and HCC1599 have been classified as basal-like tumor cells, HCC202, HCC2218, MDA-MB-453 as luminal and ErbB2 over-expressing (8) whereas the classification of MFM223 was unknown.

First we tested whether the cell lines retained their original in vivo phenotype with respect to their ability to grow in anchorage independent conditions. A colony formation assay showed that the tumorigenic potential indeed increased with the stages (Figure 1). These results demonstrate that the metastatic state is evident in the primary tumors, and that these characteristics are maintained in the cell lines in vitro. Thus this cellular model represents crucial aspects in the development of the transformed phenotype and can therefore serve as the basis of proteomic profiling.

MS-based proteomic analysis of breast cancer cell lines

We performed a SILAC-based proteomic analysis to accurately quantify the proteomes of each of the cell lines. We SILAC-labeled MCF7 cells that served as a ‘spike-in’ standard (31). Briefly, we cultured the MCF7 cells with ‘heavy’ lysine and arginine, and the lysates of the ‘heavy’ MCF7 cells were mixed with the lysates of each of these cell lines prior to trypsin digestion. Peptides were fractionated by isoelectric-focusing and analyzed with a high-resolution mass spectrometer (LTQ-Orbitrap). The spike-in approach allowed culturing the experimental cell lines described above under their standard conditions (Figure S1; see Materials and Methods) followed by relative quantification against the common SILAC standard.

Analysis of biological triplicates of all the samples identified a total of 8750 proteins and quantified 7800 of them, the latter of which were used for all subsequent analysis (Table S2). Of the quantified proteins, approximately half did not vary significantly in any of the cell lines (53%; ANOVA test for comparison of
triplicates at a 5% False Discovery Rate, see Materials and Methods). These proteins are enriched for basic cellular processes, such as the basal transcription machinery and chromatin assembly and may be considered the ‘household proteome’ (Table S3). Expression levels of proteins involved in many basic cellular functions, such as metabolic processes, protein expression and cell adhesion did change drastically as described below, reflecting the pronounced cellular differences between the cells.

As positive controls we found in HMEC high levels of known myoepithelial markers: keratins 5, 6 and 14 and caldesmon, a regulator of actomyosin contractility (32-34). These markers were lower in the cancer cells, most dramatically in Stage III and in the cells from pleural effusions (median 25-fold lower compared to the myoepithelial cells; Figure 2A). CD44, a cell surface adhesion molecule, has been reported as a marker of breast cancer stem-cells (35). However, it was also previously found to be highly expressed in myoepithelial cells and basal tumors compared to luminal cells (9). In agreement with that observation our proteomic data showed high expression of this marker only in the cells previously reported as basal (7, 8) (Figure 2B). We further identified the expected up-regulation of the DNA repair protein Poly ADP ribose polymerase (PARP) in late transformation stage (36), and high expression level of ErbB2 in HCC2218, HCC202 and MDA-MB-453, which have the corresponding gene amplification and are known to over-express this receptor (8).

**Clustering analysis distinguishes between breast cancer subtypes**

Unsupervised hierarchical clustering of the proteomic data separated the samples into two main groups (Figure 2C). The basal cluster included the HMEC, the benign cells and the basal cancer cells from Stage II and Stage III. Within the basal cluster, cell lines perfectly segregated according to their stage. This shows that MS-based proteomics can correctly group cancer proteomes according to the stage in single cancer subtypes. The luminal cluster contained cells from Stage III and from pleural effusion metastasis, including MFM223 cells. These proteomics results show, in agreement with the above-mentioned
transcriptomic work, the overall dominance of the cancer sub-type, but additionally indicate stage-related alterations.

The hierarchical clustering of the proteins revealed three main groups; those lower in the cancer cells compared to the controls, those high in the basal tumor cells and those high in the luminal cells (Figure 2C). We performed enrichment analysis to find cellular processes significantly altered in each of the clusters. In the first one, of proteins that are low in tumors relative to control irrespective of subtype, we identified a prominent reduction in the adhesive phenotype of the cells. We extracted all cell adhesion related proteins (as annotated by GO), and indeed found that the genuine adhesion proteins in this group are reduced in the transformed cells (Figure S2A; Table S3). Among the down-regulated proteins we found α-integrins 2,3,4,5,6 and V and β-integrins 1,4,5 and 6, reflecting reduced adhesion to fibronectin, collagen and laminin (37). The reduction in integrins coincided with lower levels of the ECM proteins laminin-5 (α3β3γ2) and laminin-10/11 (α5β1/2γ1), fibronectin and collagen (COL7A1), and mediators of integrin signaling, such as ILK and α-parvin. The second cluster, of proteins that were high in the basal cells, included proteins involved in DNA replication and mitosis as well as splicing regulators. Examination of the distribution of cell cycle regulators showed that the majority (such as CDK1, CDK4, cyclin A2, cyclin B1 and the APC complex proteins) are increased in the basal cells already in the pre-malignant stage, and that they were further up-regulated in the malignant cells (Figure S2B). Regulators of splicing and spliceosomal proteins were generally higher in the transformed cells compared to the control cells, however their expression was higher in the basal cells compared to the luminal cells from the same stage (Figure S2C). The cluster of proteins that are high in the luminal cells was enriched for mitochondrial proteins as well as ER-Golgi and vesicle transport processes, reflecting the physiological characteristic of the luminal cell layer as secreting cells (Figure S2D).

Establishment of a stage-specific signature
Due to the dominant effect of the cancer subtypes it is challenging to identify proteins that can serve as stage specific markers that capture commonalities in breast cancer progression. In an attempt to find such proteins, we performed an ANOVA test on cells grouped by stage. We extracted a stage-specific signature of 52 proteins, of which 11 were up-regulated and 41 were down-regulated (FDR=0.05; Figure 3; Table S4). We divided the signature proteins into four clusters according to the pattern of their change (Figure 3). The largest cluster (22 proteins) consisted of proteins that were expressed at similar levels in the normal, pre-malignant and Stage II cells and dropped dramatically between Stage II and III. This cluster included the laminin receptors, integrins α6β4 and α6β1, as well as two laminin-5 subunits (laminin-β3 and γ2). Interestingly, unlike the other proteins in the group, integrin α6 expression increased more than 8-fold in the cells from the metastatic location and likewise expression of each of its binding partners was increased approximately 2-fold in these cells. The signature also included the β subunit of the αvβ6 fibronectin receptor, while fibronectin itself was down-regulated already in the initial step of transformation. Furthermore, this cluster included an adherens junction protein (P-Cadherin, CDH3), five actin regulators (CALD1, PDLIM5, PDLIM7, CAPG and FMNL2) and the intermediate filament protein vimentin. These results highlight the general loss of the adhesive phenotype and remodeling of cell architecture. The observed proteome changes are the molecular correlates to the detachment of the cells from the tissue of origin in the process of metastasis.

The second and third clusters include proteins that were down-regulated mainly between the pre-malignant and Stage II cells, and between the myoepithelial cells and the premalignant cells (Figure 3). They include the cell cycle regulator stratifin (SFN), which is a p53 target whose promoter region is known to be hypermethylated in breast cancers (38, 39) and four adhesion molecules (BPAG1, COL17A1, CDHF7 and CD97). We also found the membrane-bound metalloprotease MMP14 to be down-regulated in the Stage II tumor cells in our system.
The last cluster included proteins that are high in the transformed cells, mostly between the pre-malignant cells and Stage II cells, but are further increased in later stages (Figure 3). This cluster has 11 members, among them metabolic proteins (IDH2, BLVRB, UCKL1 and CRABP2), protein glycosylation (FUT8), and vesicle transport (ANX6). These and the up-regulated proteins with unknown function are potential positive breast cancer markers.

As a first assessment of the signature proteins as potential tumor markers we validated their relevance to human tumors in three ways. In the first, we used a directed mass spectrometric approach to preferentially retrieve signature proteins in the analysis of single tumor samples, and in the second, we performed IHC on tumor arrays to analyze expression of individual proteins in multiple tissue samples. Finally we validated signature proteins in a large compendium of gene expression studies.

**Validation of the protein signature in human tumor samples using a directed MS approach**

To quantify tumor samples with respect to healthy tissue, cell line based SILAC cannot be applied directly. Instead, we made use of the recently developed super-SILAC method for quantification of human breast tumor samples (21). We used the super-SILAC mix as an internal standard to quantify a tumor tissue and a normal tissue that served as control. The tumor tissue originated from ER- Stage III tumor and therefore corresponds to the Stage III tumor cell lines such as HCC1599.

To examine the expression levels of the signature proteins in tumor samples we developed a mass spectrometric method, based on an inclusion list, in which we preferentially target peptides belonging to these proteins for fragmentation and identification. With this preferential fragmentation of peptides of interest, we identified 48 of the 52 signature proteins. To compare the cell line experiments and the human tumor tissues, we normalized the cancer samples to the healthy controls in each of the experiments, HCC1599 to HMEC and the tumor tissue to the normal tissue. Of the 39 signature proteins for which we had accurate ratios compared to internal standards in all four experiments (HCC1599 and...
HMEC vs. labeled MCF7 and normal tissue and Stage III tumor tissue vs. super-SILAC mix), we found positive correlation of the cancer vs. normal ratios for 32 proteins (Figure 4). For most of the proteins the ratio between the tumor cells and the normal controls was less pronounced in the tissues compared to the cell lines. Specifically, we validated the down-regulation of ANX3, SEC14L2, all adhesion proteins and laminins, as well as the myoepithelial markers, CD109, caldesmon and caveolin. We also validated the dramatic decrease in MMP14 (greater than 30-fold in both cases). For the proteins that were highly expressed in the cancer cell lines we validated the over-expression of IDH2, CRABP2, FUT8, BLVRB, ANX6 and RBM47, suggesting six potential novel markers for breast cancer.

**Verification of signature proteins by immunohistochemistry using human tumor tissue arrays**

Three of the signature proteins, IDH2, CRABP2 and ANX3, were selected for further evaluation of their expression patterns in a larger number of tumor samples using IHC and tumor arrays. In the proteomic data the first two were high in the advanced primary tumor cells, whereas the third was down-regulated starting already in pre-malignant cells. We reacted breast cancer tissue microarrays that include 75 tumor and normal samples with the corresponding antibodies. We scored the intensity of the staining in each of the tissue sections in the array and examined the correlation between the staining intensity and the tumor TNM stage. In agreement with the previous experiments ANX3 was strongly stained in the myoepithelial layer in the normal tissue, but was low in the tumor tissues (Figure 5A). Globally, ANX3 staining negatively correlated with the tumor stage. IDH2 was completely absent in the normal tissues while it showed strong mitochondrial staining in the tumor tissues. The overall correlation was positive between the tumor and lymph node state (Figure 5B). CRABP2 was negative in the myoepithelial, but had moderate staining of the luminal cell layer in the healthy tissue, and showed strong staining of the tumor cells. For the overall correlation we used the myoepithelial cells as control, as these were the controls used in the previous experiments. This confirmed the positive correlation between the tumor
stage and the intensity of the CRABP2 staining. Thus the analysis of 75 tissue samples confirmed that ANX3 is reduced with transformation and that IDH2 and CRABP2 are potential markers of advanced breast cancers.

**Prognostic value of the signature proteins**

To evaluate the prognostic value of the signature proteins it is necessary to examine the protein expression in large patient cohorts. Since such proteomic data do not exist, we performed a meta-analysis of publicly available patient mRNA datasets. Using Kaplan-Meier analysis of overall survival (OS) with median mRNA expression levels as a cut-off, we examined the prognostic value of each of the 52 genes across 1447 samples from 8 public datasets (Table S5). The three most significant genes were SEC14L2 (adjusted p-value = 1.94e-9), CRABP2 (adjusted p-value = 9.41e-5) and IDH2 (adjusted p-value = 1.49e-4), with hazard ratios (HR) of 0.511 (Confidence Intervals (CI) = 0.4174 - 0.6259), 1.616 (CI = 1.325 - 1.972) and 1.597 (CI = 1.310 - 1.947) respectively (Figure 6A-C; Table S6). When using these three markers in combination (samples with greater than median expression of CRABP2 and IDH2 and less than median expression for SEC14L2), they had a greater effect on OS than they had individually (p=6.4e-11, HR 2.07, CI 1.66-2.59, n=1438) (Figure 6D). This is consistent with our observations that CRABP2 and IDH2 are markers of poor prognostic and SEC14L2 is a marker of good prognostic.

**Discussion**

The relevance of breast cancer cell lines to tumors has been shown in genomic and gene expression studies (7, 40). Based on these studies we here combine in-vitro and in-vivo analyses in a two-step approach. We performed a system-wide analysis in cultured cells, which provide a more homogenous cell population and controlled growth environment, and proceeded to validate the results in human tumor tissues. Our global analysis of the proteins from the cell lines quantified 7800 proteins, and is the first in-depth and quantitative proteomic study of breast cancer progression.
Collapse of the adhesive machinery in the transformed cells

Myoepithelial cells within normal tissue secrete ECM components, synthesize and maintain the basement membrane, and express high levels of adhesion proteins (33, 41, 42). Accordingly, we found high expression levels of ECM and adhesion proteins in the control cells but significantly lower expression in the transformed cells, mainly in the late-stage tumor cells. Interestingly, these proteins have been considered as basal markers (9), but our results show that they are also lost in basal Stage III tumor cells. Therefore their expression demarcates cancer stage rather than subtype.

The global proteomic profiles, as well as the signature proteins in the cell lines and tissues, reflect the general collapse of normal tissue architecture, which is a known feature of the development of carcinomas (Figure 7A). The most dramatic changes in adhesion proteins occur between Stage II and Stage III tumor cells, highlighting their importance for detachment of the tumor cells from the original tissue. We show up-regulation of integrin α6 in cells derived from pleural effusion metastases. This result agrees with previous studies that show over-expression of integrin α6 in metastatic sites of breast cancers (43, 44) and suggest a role for α6β4 and α6β1 integrins in adhesion in the metastatic location.

Novel breast cancer marker candidates reflect adhesive and metabolic changes

We found 52 potential markers of transformation; 41 of these were reduced in the transformed cells and 11 induced. The small number of commonly regulated proteins across the cancer subtypes shows that transformation likely occurs in diverse paths. Nevertheless, the existence of even a relatively small number of such proteins implies that there are commonalities in these paths. We validated these signature proteins in human tumor tissue albeit generally with lower ratios in the tissue analysis compared to the cell lines. Presumably this results from the plurality of cell types in the tissue. As a result, the proteomic differences in the tissues can be diluted compared to the cell lines. This further validates our strategy of examination of the more homogenous cell population for the study of the
changes in the cancer cells. In cases where the tissue protein is expressed mainly from non-cancer cells one may even obtain opposing results between the cultured cells and the tissues. For example, for fibronectin decreased expression by tumor cells compared to myoepithelial cells was masked in the tissue by high expression by tissue fibroblasts. To establish the significance of each of the signature proteins it would be necessary to perform broader studies using the same method on multiple tumor samples.

The top two prognostic markers in our data were retinoid-binding proteins. SEC14L2 was down-regulated and CRABP2 was up-regulated in the transformed cells. SEC14L2/TAP is a retinol or α-tocopherol binding protein that can act as a transcriptional regulator, and as a regulator of cholesterol metabolism (45). Its reduced levels were previously indicated in prostate cancer and breast cancers (46, 47). CRABP2 binds all-trans retinoic acid in the cytoplasm, and targets it to the nucleus where it binds its receptor and promotes cell differentiation. In head and neck cancers and gliomas CRABP2 levels were shown to be reduced (48, 49). In contrast to these tumor types and in support of our data, ErbB2 has been shown to induce retinoic acid resistance in breast cancer cells (50). Furthermore, CRABP2 reportedly mediates proliferative activity through retinoic acid-induced PPARβ/δ activation in the presence of another factor, FABP5 (51). In our data FABP5 is high in the basal cells and low in the luminal ones (Table S2), suggesting two distinct mechanisms by which CRABP2 may induce hyper-proliferation and possibly retinoic acid resistance.

Our results suggest IDH2 as another positive breast cancer marker. IDH2 has attracted much attention when it was identified as a proto-oncogene in glioblastomas and acute myeloid leukemia (52, 53). In those cases, mutations in IDH2 appeared in early stages of transformation and induced overproduction of 2-hydroxyglutarate thereby affecting global DNA methylation patterns (54). Furthermore, elevated IDH2 activity led to high production of NADPH, a cofactor involved in biosynthetic processes and in
control of oxidative stress. In our data the level of IDH2 increased only in late stages of transformation. Such an elevation was not seen for the mutant enzymes in other tumor types.

The proteomic signature suggests control of the cells’ oxidative state and NADPH levels as principal processes regulated with transformation (Figure 7B). The increase in IDH2 in our data coincided with increased expression of flavin reductase (BLVRB), which was included in the proteomic signature, and with glutathione reductase. These two enzymes regulate the redox state of the cells, and require NADPH for their activity. Possibly, elevated IDH2 and overproduction of NADPH enables high activity of BLVRB and glutathione reductase, thereby regulating the oxidative state in the transformed cells. NADPH can also be produced in the reaction of retinol oxidation to retinal and retinoic acid. Upon binding of retinoic acid, CRABP2 affects cell proliferation.

In conclusion we here combined an unbiased system-wide view of the proteomes of cultured cells with focused analysis of tumors. This study captured the general processes that are altered upon transformation and showed the relevance of candidate markers to the in vivo situation. This strategy is particularly suitable for proteomics, which still has relatively limited throughput, since the proteomic data can be directly translated to IHC and other protein-based analyses of tumors. Finally, our data identified CRABP2 and IDH2 as markers of poor prognostic and SEC14L2 as a marker of good prognostic, and suggest additional markers that require further evaluation.

**Acknowledgements**

We thank the members of the Department of Proteomics and Signal Transduction for helpful discussion, and especially Pawel Ostasiewicz for assistance with IHC analysis. We thank René Bernard and Jelle Wesseling from the NKI, Amsterdam for the breast tumor and healthy tissues and Axel Ullrich from the
MPI of Biochemistry, Martinsried for cell lines. We thank Mathias Uhlén and Emma Lundberg from the Human Protein Atlas for antibodies. We thank Herbert Schiller from the MPI of Biochemistry and Benny Geiger from the Weizmann Institute for critical review of the manuscript. This project was supported by the European Commission’s 7th Framework Program PROteomics SPECific ion in Time and Space (PROSPECTS, HEALTH-F4-2008-021,648). W.M.G and S.F.M. are supported by Science Foundation Ireland, Strategic Research Cluster award to Molecular Therapeutics for Cancer Ireland (award 08/SRC/B1410).

References

51. Schug TT, Berry DC, Shaw NS, Travis SN, Noy N. Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. Cell. 2007;129:723-33.
**Figure legends**

**Figure 1: A cellular model of breast cancer progression.** A. A panel of eleven cell lines, which were isolated from human tumors with a defined stage, and from normal or pre-malignant tissues. Colony formation assays shows increased tumorigenic potential in the higher tumor stages. B. Description of breast cancer cell lines, including their stage and subtype. References are given in brackets.

**Figure 2: Global proteomics results.** A. Proteomics data identifies down-regulation of known myoepithelial markers in the transformed cells. B. Quantification of known breast cancer markers C. Hierarchical clustering segregates the cell lines to basal and luminal subtypes, and in a second level, segregates according to the stage. Proteins are segregated to three clusters. The names of the cells lines are color-coded according to their stage.

**Figure 3: Stage-specific proteomic signature.** A 52-protein signature was extracted using ANOVA test (FDR=0.05, S0=1). Hierarchical clustering divided the signature to four clusters as indicated by the colored bars.

**Figure 4: Directed MS analysis of signature protein in tumor sample using super-SILAC.** Scatter plot of the cell lines vs. and tumor data. Signature proteins were quantified in ER- Stage III tumor relative to healthy tissue using super-SILAC as a spike-in standard. These ratios were compared to those of Stage III tumor cell line relative to HMEC. **Figure 5: Validation of MS data using IHC on tumor arrays.** Breast cancer tissue arrays were reacted with anti-ANX3, anti-IDH2 or anti-CRABP2 antibodies. A. Representative healthy and Stage III invasive ductal carcinomas (IDC) tumor tissues are shown. (scale-bar=150 µm) B. The staining intensity of each tissue was scored and the compared to the tumor, lymph node and metastatic state of the tumors. Heat-map shows the correlation between the staining scores and the clinical information.
Figure 6: Prognostic role of CRABP2, IDH2 and SEC14L2 at the mRNA level in breast cancer. Kaplan-Meier estimates of survival demonstrating low expression of SEC14L2 (A), high expression of IDH2 (B) and CRABP2 (C) are associated with reduced overall survival in breast cancer, based on mRNA expression levels of 1447 samples from 8 different datasets. The log-rank P value is shown for difference in survival and Cox regression analysis was used to calculate hazard ratios (HR). The p-values are adjusted for multiple testing. D. Kaplan-Meier estimate of survival demonstrating that these markers in combination have a more pronounced effect on survival.

Figure 7: Network changes in the breast cancer signature. A. Cell adhesion network from the protein signature. The protein network was built using the STRING database. All proteins were down-regulated, the majority in Stage III (dark green), MMP14 and COL17A1 in Stage II, and FN1, SERPINE1 and DST in the premalignant stage (Light green). B. Suggested involvement of the signature proteins in control of NADPH metabolism and the oxidative state of the cells. Up-regulated proteins are marked in red and down-regulated protein in green.
Figure 1

A. Normal breast duct

Pre-malignant

IDC Stage II

IDC Stage III

Metastasis-
Pleural effusion

B. Cell line Description Subtype

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Description</th>
<th>Subtype</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMEC1</td>
<td>Human Mammary epithelial cells</td>
<td>Healthy (basal)</td>
</tr>
<tr>
<td>HMEC2</td>
<td>Human Mammary epithelial cells</td>
<td>Healthy (basal)</td>
</tr>
<tr>
<td>MCF10a (20)</td>
<td>Non tumorigenic mammary epithelial cells- benign tumor</td>
<td>Non-tumorigenic (basal)</td>
</tr>
<tr>
<td>HMT-3522-S1 (17)</td>
<td>Non tumorigenic mammary epithelial cells- benign tumor</td>
<td>Non-tumorigenic (basal)</td>
</tr>
<tr>
<td>HCC1143 (19)</td>
<td>Triple negative Stage II</td>
<td>Basal-like</td>
</tr>
<tr>
<td>HCC1937 (19)</td>
<td>Triple negative Stage II BRCA1 mutation</td>
<td>Basal-like</td>
</tr>
<tr>
<td>HCC2218 (19)</td>
<td>ErbB2+ Stage III</td>
<td>Luminal</td>
</tr>
<tr>
<td>HCC1599 (19)</td>
<td>Triple negative Stage III</td>
<td>Basal-like</td>
</tr>
<tr>
<td>HCC202 (19)</td>
<td>ErbB2+ Stage III</td>
<td>Luminal</td>
</tr>
<tr>
<td>MFM223 (18)</td>
<td>Triple negative Stage III pleural effusion androgen sensitive</td>
<td>Luminal</td>
</tr>
<tr>
<td>MDAMB453 (21)</td>
<td>ErbB2+ Stage III pleural effusion</td>
<td>Luminal</td>
</tr>
</tbody>
</table>
Figure 2

A. Proteins with Log2Ratio changes in various cell lines:
- Keratin5
- Keratin14
- Keratin6B
- Caldesmon

B. Other proteins with Log2Ratio changes:
- ErbB2
- PARP1
- CD44

C. Heatmap showing expression levels of selected genes across different cell lines.
Figure 4

The figure shows a scatter plot of Log2 normalized proteomic change for two cell lines, Tumor and HCC1599. The x-axis represents Log2 normalized proteomic change for HCC1599, while the y-axis represents Log2 normalized proteomic change for Tumor. Various proteins are marked on the plot, categorized into metabolism, cell adhesion, and actin regulators.

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Downloaded from cancerres.aacrjournals.org on April 14, 2017. © 2012 American Association for Cancer Research.
Figure 5

A. Normal tissue vs Stage III IDC for ANX3, IDH2, CRABP2 proteins.

B. Heatmap showing correlation between CRABP2, IDH2, and ANX3 in different tissue stages (N, T, M) with a color scale ranging from -0.4 to 0.4.
Figure 6

A.

High SEC14L2  
n=724

Low SEC14L2  
n=723

HR=0.511  
p=1.94e-9

B.

Low IDH2  
n=726

High IDH2  
n=721

HR=1.597  
p=1.49e-4

C.

Low CRABP2  
n=718

High CRABP2  
n=720

HR=1.616  
p=9.41e-5

D.

High CRABP2  
High IDH2

Low SEC14L2  
n=228

HR=2.07  
p=6.382e-11

Overall survival

Time (months)

Overall survival

Time (months)

Overall survival

Time (months)

Overall survival

Time (months)
Proteomic portrait of human breast cancer progression identifies novel prognostic markers

Tamar Geiger, Stephen F Madden, William M Gallagher, et al.

Cancer Res Published OnlineFirst March 13, 2012.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-3711

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2012/03/15/0008-5472.CAN-11-3711.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.