Breast Cancer Proteomics – Differences in Protein Expression between Estrogen Receptor-Positive and -Negative Tumors Identified by Tandem Mass Tag Technology

Eugen Ruckhäberle[a]   Thomas Karn[a]   Lars Hanker[a]   Josef Schwarz[b]   Peter Schulz-Knappe[b]
Karsten Kuhn[b]   Gitte Böhm[b]   Stefan Selzer[b]   Erhard Neukum[b]   Knut Engels[c]   Uwe Holtrich[a]
Manfred Kaufmann[a]   Achim Rody[a]

[a] Department of Obstetrics and Gynecology, Goethe University Frankfurt,
[b] Proteome Sciences, Frankfurt,
[c] Department of Pathology, Goethe University Frankfurt, Germany

Key Words

Schlüsselwörter

Summary

Background: Proteomic analysis has become an effective tool in breast cancer research. In this study, we applied the new gel-free tandem mass tag (TMT) reference method for the first time in breast cancer. **Materials and Methods:** Proteomic analysis was used to compare 10 estrogen receptor (ER)-positive and 10 ER-negative samples. The results of the proteomic approach were validated by Western blot, immunohistochemistry and gene expression analysis. **Results:** 17 proteins with significant differences in expression were identified. 13 proteins were overexpressed in ER-negative tumors and 4 were overexpressed in ER-positive samples. All these proteins were characterized by relatively high cellular abundance. **Conclusions:** Our results demonstrate that the gel-free TMT approach allows the quantification of differences in protein expression levels. Further improvement of the sensitivity by subfractionation of the tissue should allow also the identification of low-abundance proteins and might lead to the use of this method in breast cancer research.

Zusammenfassung

Hintergrund: Proteomanalysen haben sich zu einem effektiven Werkzeug in der Krebsforschung entwickelt. In dieser Arbeit setzen wir eine neue, gelfreie Methode, die sogenannte Tandem-Mass-Tag (TMT)-Referenzmethode, zur Charakterisierung von Mammakarzinomsubklassen ein. **Material und Methoden:** Proteomanalysen wurden für einen Vergleich von 10 Mammakarzinomproben mit positivem und 10 Proben mit negativem Östrogenrezeptorstatus verwendet. Die erhaltenen Ergebnisse wurden anschließend mittels Western-Blot, Immunhistochemie und Genexpressionsanalysen validiert. **Ergebnisse:** 17 Proteine mit signifikanten Unterschieden in ihrer Expressionshöhe wurden identifiziert. 13 der Proteine zeigten eine stärkere Expression bei Östrogenrezeptor-negativen Tumoren und 4 eine bei Östrogenrezeptor-positiven Proben. Bei allen identifizierten Markern handelte es sich um Proteine mit relativ hoher Expressionsstärke in der Zelle. **Schlussfolgerungen:** Unsere Ergebnisse zeigen, dass der gelfreie TMT-Ansatz eine Quantifizierung von Unterschieden in der Expressionshöhe von Proteinen beim Mammakarzinom erlaubt. Eine weitere Verbesserung der Empfindlichkeit durch Subfraktionierung des Gewebes sollte auch die Identifikation von in geringer Menge vorkommenden Proteinen erlauben und könnte zum Einsatz dieser Methode in der Brustkrebsforschung führen.

Introduction

Breast cancer is the most frequent cancer in women worldwide. Proteomics could open up new ways of screening [1–4], prediction of therapeutic success and screening for metastases in breast cancer. Proteomic analysis offers several advantages compared to genomics [5]. For example, the expression of mRNA does not necessarily lead to translation of the corresponding proteins, or proteins can undergo a vast number of posttranslational modifications which affect their stability, localization, interaction and function.

The common protein expression profiling (expression proteomics) approach relies on the separation of the proteome using two-dimensional gel electrophoresis (2-DE), staining of the separated proteins with silver or an appropriate dye, comparing the gel images using specific software packages, and identification of the proteins of interest by matrix-assisted laser desorption/ionization (MALDI) peptide mass fingerprinting [6]. A major drawback of 2-DE-based proteomics may be that proteins with extreme biochemical properties like membrane proteins are more difficult to detect [5]. Also, the technical variations of the workflows used for 2-DE-polyacrylamid gel electrophoresis (PAGE) make the search for smaller changes in expression very challenging [7]. Over the last years, gel-free proteomic workflows have been developed, which are mainly based on chromatographic separation after tryptic protein digestion and the use of mass spectrometry (MS) for the identification and quantification of hundreds of proteins from minute amounts of sample [6, 8]. A distinct advantage of these workflows is in the virtually complete proteome capture since almost all proteins of every cellular compartment are made accessible. In a special refinement, such gel-free proteomic approaches today make use of specific protein labeling of individual patient samples, with subsequent mixing of the samples. Thus, technical variations in workflows should be minimized and therefore small changes in protein expression can be analyzed. Recently, isobaric label reagents, such as tandem mass tags (TMT®; Proteome Sciences) and isobaric tags for relative and absolute quantification (iTRAQ™; Applied Biosystems), have been developed for MS-based protein detection and quantification in complex mixtures [9–11]. Two different sets of tags of the TMT label reagents are available, which share the chemical structure but differ in the number of incorporated heavy isotopes; i.e., the TMTduplex allows for the routine investigation of 2 different samples in parallel and the TMTsixplex allows for the simultaneous analysis of 6 samples. This multiplexing feature is particularly useful when more than 2 sample sets are compared or when biological systems are investigated over multiple time points. The TMT label attaches to amino groups of peptides (both N-terminal α-amino groups and ε-amino groups of lysine residues) that are generated by tryptic digestion of a protein mixture.

In this study, we applied for the first time a TMT reference material-based workflow to breast cancer specimens in order to screen for differentially expressed proteins. We demonstrate the feasibility of this TMT reference workflow. Furthermore, we validate our results by Western blot, immunohistochemistry, and microarray analyses of breast cancer samples.

Materials and Methods

Tissue samples of invasive breast cancer cases were obtained with institutional review board approval and informed consent from consecutive patients undergoing surgical resection between December 1996 and July 2007 at the Department of Gynecology and Obstetrics at the Goethe University in Frankfurt. Protein preparation from tumor tissues, liquid chromatography (LC)-MS and peptide identification was done as described previously [9]. Validation of the LC-MS results was done by immunohistochemistry and Western blot analyses using an anti-glutathione S-transferase P1 (GSTP1) antibody (HPA019869, Sigma). Isolation of RNA and expression profiling using Affymetrix Human Genome U133A microarrays was performed as described previously [12–14]. All reported p values are two sided, and p values of less than 0.05 were considered to indicate a significant result. All statistical analyses were performed using SPSS 15.0 (SPSS Inc., Chicago, IL, USA). A detailed description of all experimental procedures is given in the supplementary material.

Results

We used the two isobaric TMTduplex reagents TMT^2–126 and TMT^2–127 (supplementary figure 1A, B), since their performance was demonstrated quite recently [9]. Our goal was to screen estrogen receptor (ER)-positive versus -negative breast cancer specimens for differentially expressed proteins. The rationale of isobaric label reagents is the simultaneous generation of both identification and quantification data whilst maintaining optimal quantitative control by comparing all samples to the same spiked reference. We used a reference sample from equal aliquots of all study samples.

Proteomic analysis could be performed with all (10) samples of the ER-negative sample group. The clinical characteristics of the analyzed breast cancers are presented in supplementary table 3. From the ER-positive sample collective, one sample was excluded because of low quality. For an internal standardization, each individual sample was spiked with the same amount of the reference sample. This reference was generated as an equal mixture of all individual samples. Supplementary figure 2 exemplifies the results of LC-MS/MS of 2 ER-positive and 2 ER-negative samples.

Overall, 375 proteins could be identified over the 19 investigated samples, with a range from 150 to 206 proteins identified per individual sample (table 1). From the 375 proteins, 149 were identified and quantified in at least 5 individual samples per group, which made them accessible for statistical analysis. Based on the protein expression levels, 3 groups could be identified. For 18 proteins, significant (p ≤ 0.05) differences in expression levels were found between the ER-pos-

Ruckhäberle/Karn/Hanker/Schwarz/
Schulz-Knappe/Kuhn/Böhm/Selzer/Neukum/
Engels/Holtrich/Kaufmann/Rody

Table 1. Numbers of identified
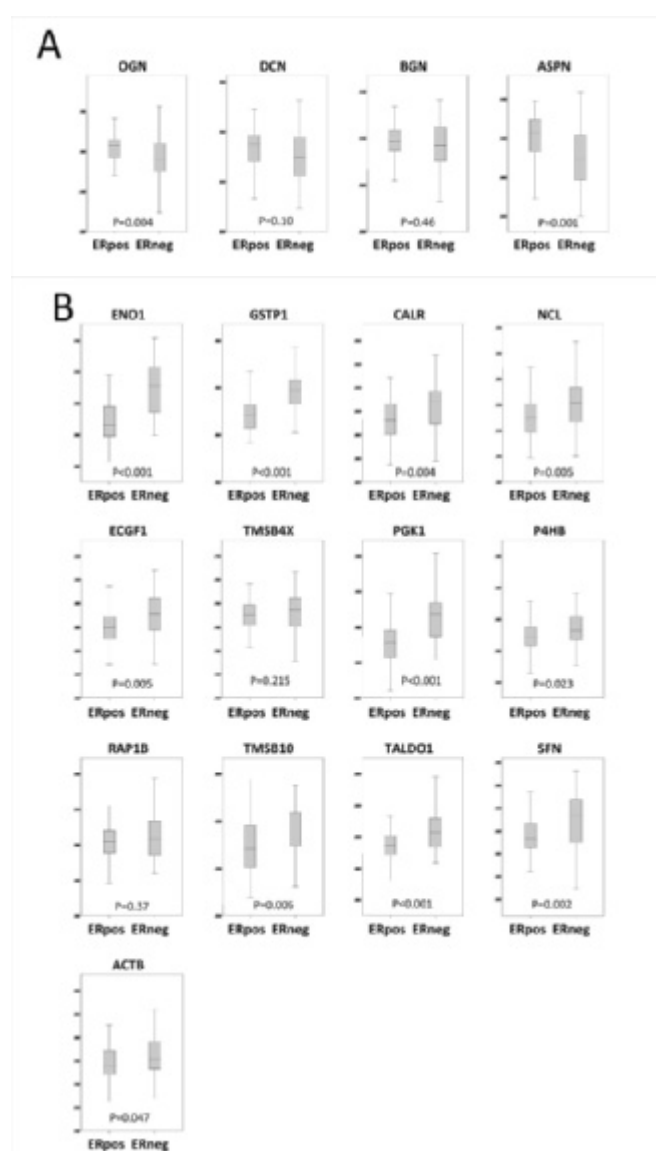
	Proteins, n
Overall identified in 19 samples	375
Range	150–260
Identified in at least 5 samples per group	149
Differentially expressed	32
Statistically significant (p < 0.05)	17

itive and the ER-negative group (supplementary table 4). The remaining proteins exhibited similar expression levels in both groups.

Within the group with significantly differentially expressed proteins, 13 proteins were higher expressed in the ER-negative subgroup (supplementary table 4). These proteins are two isoforms of α-enolase, GSTP, calreticulin, nucleolin, thymidine phosphorylase, thymosin-like 3, phosphoglycerat kinase-1, protein disulfide-isomerase, ras-related protein rap-1b, thymosin beta-10, transaldolase 1, 14–3–3 protein sigma, and actin cytoplasmic. The remaining four proteins, decorin, mimecan, biglycan, and asporin (ASPN) protein, show higher expression levels in the ER-positive subgroup. We next verified the results of LC-MS by immunohistochemistry and Western blot using a monoclonal antibody raised against GSTP1. Both immunohistochemistry and Western blot analysis confirmed the LC-MS results (supplementary figures 3 and 4). In addition, we validated our LC-MS data on the mRNA level using microarray analysis. We used microarray data from a cohort of 119 primary breast cancer patients we have described previously [12, 13]. Clinical characteristics of the investigated breast cancers are described in supplementary table 2. The gene expression data from all 17 proteins identified above as differentially expressed between ER-positive and ER-negative samples were compared. As shown in figure 1, 2 out of 4 proteins (mimecan, ASPN) that were found at higher levels in the ER-positive subgroup also displayed higher mRNA expression in the analyzed dataset (fig. 1A). Decorin und biglycan failed to show differences between ER-positive and -negative samples on the mRNA level. Furthermore, we confirmed the higher expression of 12 out of 13 proteins in ER-negative samples on the mRNA level (fig. 1B). Only thymosine beta 4 (TMSB4X) and RAP1B did not show a significant result.

Discussion

In the present work, isobaric label reagents, namely the TMT-duplex reagents, were applied in order to search for differences in protein expression between ER-positive and ER-negative breast cancer by LC-MS/MS. Isobaric label reagents show several features that make them very appropriate for discovery purposes in biomarker search studies: The labeling of proteins with chemical tags in general allows for mixing,

Fig. 1. Validation of proteomic data by microarray analysis. The mRNA expression of the 17 proteins that were found by proteome analysis to be differentially expressed between ER-positive and ER-negative tumors was studied using a microarray dataset from 119 breast cancer samples. Box plots comparing the normalized expression values of ER-positive and ER-negative samples are shown for those 4 markers that were found to be higher expressed in ER-positive samples (A) and those 13 markers found to be higher expressed in ER-negative samples (B); p values according to Mann-Whitney U-test are given for each gene.

reducing analytical deviations from run to run. The use of isobaric label reagents in particular allows for a larger number of samples to be analyzed. By including all study samples in the reference, it is made sure that every single protein present in any sample is represented as reference point in all individual analysis runs.

Gene expression analysis has become an established clinical predictive tools to identify tumor subtypes, to identify patients showing poor/good prognosis and to identify patients likely to have disease recurrence [15]. We validated the data from our

proteomic analyses on the mRNA level using microarray analysis. We compared the gene expression of proteins identified as differentially expressed between ER-positive and ER-negative subgroups. In our dataset of 119 clinical breast cancers, we confirmed higher expression of mimecan and ASPN protein in the ER-positive subgroup and 12 out of 13 identified proteins in the ER-negative subgroup (fig. 1). Several microarray studies investigated differences between ER-positive and -negative tumors on the level of gene expression [15, 16]. Even though differences exist between those reports, they all agree on some top markers that are highly overexpressed in ER-positive samples, e.g. estrogen receptor one (ESR1), GATA 3 and SLC39A6 (LIV1), FOXA1, IGFBP2, and cyclin D1. In contrast, EGFR, keratin 7, cathepsin C, cadherin 3, and lactate dehydrogenase B featured higher expression in microarray analysis of ER-negative samples. However, a large number of markers that were consistently found in microarray studies by various authors were missing in our proteomic approach.

Our results demonstrate the principal feasibility of the TMTduplex label reagents in breast cancer research. We found significant differences between ER-positive and -negative samples. However, our quantitative profiling of the approximately top 200 proteins delivered mainly mid- to high-abundant proteins from the entire tissue sample, including tumor cells, non-tumor cells and interstitial fluids. Subfractionation strategies, e.g. by laser capture microdissection or by biochemical separation, would also be possible using the TMT reference concept. Based on the obvious lack of analytical sensitivity, we were not able to demonstrate differences in expression of lower-abundant proteins, including the ER itself. By subfractionating samples, we expect to quantify many more proteins, thereby improving the method. Furthermore, it should be noted that our pilot study on a very small sample cohort was not planned to identify predictors of the ER status of the tumor. Thus, no analyses on specificity and sensitivity were performed. Moreover, validation studies on larger cohorts are needed to analyze the biological value of the detected differences.

Conclusions

Our results demonstrate that the gel-free TMTduplex approach allows the quantification of differences in protein expression levels. Further improvement of the sensitivity by subfractionation of the tissue should allow also the identification of low-abundance proteins and might lead to the use of this method in breast cancer research.

Supplemental Material and Methods

Please refer to *www.karger.com/doi=10.1159/000272241*.

Conflict of Interest

There are no conflicts of interest to declare.

References

1 Callesen AK, Mohammed S, Bunkenborg J, Kruse TA, Cold S, Mogensen O, Christensen R, Vach W, Jørgensen PE, Jensen ON: Serum protein profiling by miniaturized solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom 2005;19:1578–1586.
2 Celis JE, Gromova I, Gromov P, Moreira JM, Cabezón T, Friis E, Rank F: Molecular pathology of breast apocrine carcinomas: a protein expression signature specific for benign apocrine metaplasia. FEBS Lett 2006;580:2935–2944.
3 Datta S: Classification of breast cancer versus normal samples from mass spectrometry profiles using linear discriminant analysis of important features selected by random forest. Stat Appl Genet Mol Biol 2008;7:Article 7.
4 Gonçalves A, Charafe-Jauffret E, Bertucci F, Audebert S, Toiron Y, Esterni B, Monville F, Tarpin C, Jacquemier J, Houvenaeghel G, Chabannon C, Extra JM, Viens P, Borg JP, Birnbaum D: Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. Mol Cell Proteomics 2008;7:1420–1433.
5 Laronga C, Drake R: Proteomic approach to breast cancer. Cancer Control 2007;14:360–368.
6 Aebersold R, Mann M: Mass spectrometry-based proteomics. Nature 2003;422:198–207.

7 Gevaert K, Vandekerckhove J: Protein identification methods in proteomics. Electrophoresis 2000;21:1145–1154.
8 Molloy MP, Brzezinski EE, Hang J, McDowell MT, VanBogelen RA: Overcoming technical variation and biological variation in quantitative proteomics. Proteomics 2003;3:1912–1919.
9 van Ulsen P, Kuhn K, Prinz T, Legner H, Schmid P, Baumann C, Tommassen J: Identification of proteins of *Neisseria meningitidis* induced under iron-limiting conditions using the isobaric tandem mass tag (TMT) labeling approach. Proteomics 2009;9:1771–1781.
10 Dayon L, Hainard A, Licker V, Turck N: Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 2008;80:2921–2931.
11 Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ: Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 2004;3:1154–1169.
12 Rody A, Holtrich U, Gaetje R, Gehrmann M, Engels K, von Minckwitz G, Loibl S, Diallo-Danebrock R, Ruckhäberle E, Metzler D, Ahr A, Solbach C, Karn T,

Kaufmann M: Poor outcome in estrogen receptor-positive breast cancers predicted by loss of plexin B1. Clin Cancer Res 2007;13:1115–1122.
13 Ruckhäberle E, Rody A, Engels K, Gaetje R, von Minckwitz G, Schiffmann S, Grösch S, Geisslinger G, Holtrich U, Karn T, Kaufmann M: Microarray analysis of altered sphingolipid metabolism reveals prognostic significance of sphingosine kinase 1 in breast cancer. Breast Cancer Res Treat 2008;112:41–52.
14 Karn T, Metzler D, Ruckhäberle E, Hanker L, Gätje R, Solbach C, Ahr A, Schmidt M, Holtrich U, Kaufmann M, Rody A: Data driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. Breast Cancer Res Treat 2009, in press.
15 Smith DD, Saetrom P, Snøve O Jr, Lundberg C, Rivas GE, Glackin C, Larson GP: Meta-analysis of breast cancer microarray studies in conjunction with conserved cis-elements suggest patterns for coordinate regulation. BMC Bioinformatics 2008;9:63.
16 Schneider J, Ruschhaupt M, Buness A, Asslaber M, Regitnig P, Zatloukal K, Schippinger W, Ploner F, Poustka A, Sültmann H: Identification and meta-analysis of a small gene expression signature for the diagnosis of estrogen receptor status in invasive ductal breast cancer. Int J Cancer 2006;119:2974–2979.

Ruckhäberle/Karn/Hanker/Schwarz/
Schulz-Knappe/Kuhn/Böhm/Selzer/Neukum/
Engels/Holtrich/Kaufmann/Rody