Grading Breast Cancer Tissues Using Molecular Portraits*Ⓢ

Niclas Olsson‡, Petter Carlsson‡, Peter James‡, Karin Hansson‡, Sofia Waldemarson‡, Per Malmström§¶, Mårten Fernö¶, Lisa Ryden‖, Christer Wingren‡‡‡, and Carl A. K. Borrebaeck‡‡‡**

Tumor progression and prognosis in breast cancer patients are difficult to assess using current clinical and laboratory parameters, where a pathological grading is indicative of tumor aggressiveness. This grading is based on assessments of nuclear grade, tubule formation, and mitotic rate. We report here the first protein signatures associated with histological grades of breast cancer, determined using a novel affinity proteomics approach. We profiled 52 breast cancer tissue samples by combining nine antibodies and label-free LC-MS/MS, which generated detailed quantified proteomic maps representing 1,388 proteins. The results showed that we could define in-depth molecular portraits of histologically graded breast cancer tumors. Consequently, a 49-plex candidate tissue protein signature was defined that discriminated between histological grades 1, 2, and 3 of breast cancer tumors with high accuracy. Highly biologically relevant proteins were identified, and the differentially expressed proteins indicated further support for the current hypothesis regarding remodeling of the tumor microenvironment during tumor progression. The protein signature was corroborated using meta-analysis of transcriptional profiling data from an independent patient cohort. In addition, the potential for using the markers to estimate the likelihood of long-term metastasis-free survival was also indicated. Taken together, these molecular portraits could pave the way for improved classification and prognostication of breast cancer. *Molecular & Cellular Proteomics 12: 10.1074/mcp.M113.030379, 3612–3623, 2013.*

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among women, accounting for 23% of the total cancer cases and 14% of cancer-related deaths (1). Traditional clinicopathological parameters such as histological grading, tumor size, age, lymph node involvement, and hormonal receptor status are used to determine prognosis and treatment decisions (2–6). Histological grading, one of the most commonly used prognostic factors, is a combined score based on microscopic evaluation of the morphological and cytological features of tumor cells that reflects the aggressiveness of a tumor. This combined score is then used to stratify breast cancer tumors into three grades: grade 1, slow growing and well differentiated; grade 2, moderately differentiated; and grade 3, highly proliferative and poorly differentiated (2). However, the clinical value of histological grades for patient prognosis has been questioned, mainly reflecting the current challenges associated with traditional grading of tumors (7, 8). Furthermore, 30% to 60% of tumors are classified as histological grade 2, which represents a heterogeneous patient cohort and has proven to be less informative for clinical decision making (9). Clearly, traditional clinical parameters are still not sufficient for adequate prognosis and risk-group discrimination or for therapy selection. As a result, many patients will be overtreated or treated with a therapy that will not offer any benefits. Molecular grading of tumors could be clinically valuable, if the grading could be performed using an objective, high-performing classifier. Thus, a deeper molecular understanding of breast cancer biology and tumor progression, in combination with improved ways to individualize prognosis and treatment decisions, is required in order to further advance treatment outcomes (10, 11).

To date, a set of genomic efforts have generated molecular signatures for the subgrouping of breast cancer types (12–14), as well as for breast cancer prognostics and risk stratification (15–17). In addition, proteomic findings have been anticipated to accelerate the translation of key discoveries into clinical practice (18). In this context, classical mass-spectrometry-based proteomics have generated valuable inventories of breast cancer proteomes, although using mainly cell lines and only a few breast cancer tissue samples (19–24). More recently, affinity proteomics has delivered the first multiplexed serum portraits for the diagnosis of breast cancer and for predicting the risk of tumor recurrence (25, 26). However, generating detailed protein expression profiles in a sensitive and reproducible manner, using large cohorts of complex

From the ‡Department of Immunotechnology and CREATE HEALTH, Lund University, Medicon Village, SE-223 81 Lund, Sweden; §Department of Oncology, Skåne University Hospital, SE-221 85 Lund, Sweden; ¶Department of Oncology, Clinical Sciences, Lund University, SE-221 85 Lund, Sweden; ‖Department of Surgery, Clinical Sciences, Lund University, SE-221 85 Lund, Sweden; **Department of Surgery, Skåne University Hospital, SE-221 85 Lund, Sweden

proteomes such as tissue extracts, remains a challenge when using either classical proteomic technologies or affinity proteomics. To resolve these issues, we recently developed the global proteome survey (GPS)[1] technology platform (27), combining the best features of affinity proteomics (large-scale, multiplexed proteome analysis based on the use of antibodies or other specific reagents (28)) and MS. GPS is best suited for discovery endeavors aiming to reproducibly decipher crude proteomes in a sensitive and quantitative manner (29, 30).

In this first study of breast tumors, we delineated in-depth molecular portraits associated with histologically graded breast cancer tissues using GPS. For this purpose, 52 selected breast cancer tissue proteomes were profiled, representing one of the largest label-free LC-MS/MS-based breast cancer tissue studies. The protein expression profiles subsequently were validated using an orthogonal method. In the longer term, these tissue protein portraits might pave the way for improved classification and prognostication of breast cancer patients, and potentially even aid in defining candidate targets for therapy.

EXPERIMENTAL PROCEDURES

Clinical Samples—This study was approved by the regional ethics review board at Lund University, Sweden. Fifty-two breast cancer patients (stages I and II) were recruited from the Department of Oncology (Skane University Hospital, Lund, Sweden). Freshly frozen breast tumor tissues were stored at −80 °C until analysis. Full clinical records were accessible for 50 of the reevaluated tissue samples, including tumor size, steroid receptor status, and lymph node involvement (Table I and supplemental Table S1). The two additional tumors were not primary tumors and consequently were included only for peptide identification purposes. The samples were subdivided via careful pathologic evaluation at the Department of Pathology (Skane University Hospital) based on Nottingham histological grades 1 ($n = 9$), 2 ($n = 17$), and 3 ($n = 24$). Furthermore, 66% of the tumors were estrogen receptor (ER) positive and progesterone receptor (PR) positive. Both the ER-positive and PR-positive tumors were found in all histological grades, with over 40% of ER-positive tumors being grade 3 tumors. The ER-negative tumors were found only in histological grades 2 and 3. Forty-six of the specimens had a defined HER2 status, and all HER2-positive tumors (10%) were grade 3 tumors (Table I and supplemental Table S1). In addition, 41 of the tumors had a defined Ki67 status, and 17 of the samples were defined as Ki67-positive (supplemental Table S1).

Preparation of Trypsin-digested Human Breast Cancer Tissue Samples—Protein was extracted from the breast cancer tissue pieces and stored at −80 °C until use. Briefly, tissue pieces (about 50 mg/sample) were homogenized in Teflon containers, pre-cooled in liquid

---

[1] The abbreviations used are: AUC, area under the curve; CDK1, cyclin-dependent kinase 1; CIMS, context independent motif specific; ECM, extracellular matrix; ER, estrogen receptor; FDR, false discovery rate; GOBO, gene expression-based outcome for breast cancer online; GPS, global proteome survey; IPA, Ingenuity Systems Pathway Analysis; MCM, minichromosome maintenance complex component; PR, progesterone receptor; ROC, receiver operating characteristic; SVM, support vector machine; TGFβ1, transforming growth factor β.

TABLE I
Patient demographics and clinical parameters

Parameter	Histological grade 1[a]	Histological grade 2[a]	Histological grade 3[a]
Number of patients[b]	9	17	24
Age (years)	55.8 (11.9)	45.9 (4.0)	45.8 (5.2)
Tumor size (mm)	24.3 (5.9)	21.9 (10.9)	29.5 (9.2)
ER+/ER−[c]	9/0	14/3	10/14
PgR+/PgR−	9/0	13/4	11/13
Lymph node+/lymph node−	5/4	14/3	10/14
HER2+/HER2−[d]	0/9	0/16[e]	5/16

[a] Values in parentheses represent standard deviation.
[b] For two patients, clinical parameters were not received.
[c] ER and PgR were analyzed in cytosol samples with ligand binding assays (LBAs) or enzyme immunoassays (EIAs) as previously described (11). Samples with a receptor content greater than or equal to 10 (LBA) or 25 (EIA) fmol/mg protein were classified as ER or PgR positive, and samples with values below these levels were ER or PgR negative.
[d] All patients with fluorescence *in situ* hybridization (FISH) amplified tumors and all patients with an immunohistochemical 3+ where FISH could not be evaluated were considered HER2+.
[e] In cases where the sum is less than the number in the group, patient data are missing.

nitrogen by fixating the bomb in a shaker for two 30-s periods with quick cooling in liquid nitrogen between the two shaking rounds. The homogenized tissue powder was collected in lysis buffer (2 mg tissue/30 μl buffer) containing 8 M urea, 30 mM Tris, 5 mM magnesium acetate, and 4% (w/v) CHAPS (pH 8.5). The tubes were briefly vortexed and incubated on ice for 40 min, with brief vortexing of the sample every 5 min. After incubation, the samples were centrifuged at 13,000 rpm, and the supernatant was transferred to new tubes and subjected to a second centrifugation. The buffer was exchanged for 0.15 M HEPES, 0.5 M urea (pH 8.0) using Zeba desalting spin columns (Pierce, Rockford, IL) before the protein concentration was determined using a Total Protein Kit, Micro Lowry (Sigma, St. Louis, MO). Finally, the samples were aliquoted and stored at −80 °C until further use. The protein extracts were thawed, reduced, alkylated, and trypsin digested. First, SDS and *tris*(2-carboxyethyl)phosphine-HCl (Thermo Scientific, Rockford, IL) were added to final concentrations of 0.02% (w/v) and 5 mM, respectively, and the samples were reduced for 60 min at 56 °C. The samples were cooled down to room temperature before iodoacetamide was added to 10 mM, and then the samples were alkylated for 30 min at room temperature. Next, sequencing-grade modified trypsin (Promega, Madison, WI) was added at 20 μg per milligram of protein for 16 h at 37 °C. In order to ensure complete digestion, a second aliquot of trypsin (10 μg per milligram of protein) was added and the tubes were incubated for an additional 3 h at 37 °C. Finally, the digested samples were aliquoted and stored at −80 °C until further use. In addition, a separate pooled sample, generated by combining 5-μl aliquots from all digested samples, was prepared and stored at −80 °C until further use. In order to increase the potential tentative proteome coverage, the two samples for which limited clinical data were at hand (supplemental Table S1) were still analyzed individually and included in the pooled sample.

Production and Coupling of CIMS-scFv Antibodies to Magnetic Beads—Nine CIMS scFv antibodies (clones 1-B03, 15-A06, 17-C08, 17-E02, 31−001-D01, 32–3A-G03, 33–3C-A09, 33–3D-F06, and 34−3A-D10) directed against seven short C-terminal amino acid peptide motifs (supplemental Table S2) were used as affinity probes. The selection of the antibodies was based on the criterion of a limited set

of high-performing CIMS antibodies. To accomplish this, we selected nine CIMS antibodies proven in earlier studies to obtain reasonably large, wide (deep), sensitive, and quantitatively reproducible proteome coverage (29, 30). Of note, these binders and their motif specificities were not specifically chosen to address a specific indication such as breast cancer or for targeting a specific subset of proteins. The specificity and dissociation constants (low micromolar range) for eight of the CIMS antibodies have recently been determined (29, 31). The antibodies were produced in 100-ml *E. coli* cultures and purified using affinity chromatography on Ni^{2+}-nitrilotriacetic acid agarose (Qiagen, Hilden, Germany). Bound molecules were eluted with 250 mM imidazole, dialyzed against PBS (pH 7.4) for 72 h, and then stored at + 4 °C until use. The protein concentration was determined by measuring the absorbance at 280 nm. The integrity and purity of the scFv antibodies were evaluated via 10% SDS-PAGE (Invitrogen, Carlsbad, CA). The purified scFvs were individually coupled to magnetic beads (M-270 carboxylic acid activated, Invitrogen Dynal, Oslo, Norway) as previously described (29). Briefly, batches of 180 to 250 μg purified scFv were covalently coupled (EDC-NHS chemistry N-Ethyl-N′-(3-dimethylaminopropy) carbodiimide (Sigma- Aldrich, St. Louis, MO) and Sulfo-NHS (Thermo Scientific, Rockford, IL.)). to ~9 mg (300 μl) of magnetic beads and stored in 0.005% (v/v) Tween-20 in PBS at 4 °C until further use. A batch of blank beads was also generated (*i.e.* beads generated with the coupling protocol but without the addition of scFv).

Label-free Quantitative GPS Experiments—Four different pools (CIMS-binder mixes 1 to 4) of conjugated beads were made by mixing equal amounts of two or three different binders as follows: mix 1, CIMS-33–3D-F06 and CIMS-33–3C-A09; mix 2, CIMS-17-C08 and CIMS-17-E02; mix 3, CIMS-15-A06 and CIMS-34–3A-D10; and mix 4, CIMS-1-B03, CIMS-32–3A-G03, and CIMS-31– 001-D01 (supplemental Table S2). For each capture, 50 μl of the pooled bead solution was used, and the scFv beads were never reused. The beads were prewashed with 350 μl PBS prior to being exposed to a tryptic sample digest in a final volume of 35 μl (diluted with PBS and the addition of phenylmethylsulfonyl fluoride to a final concentration of 1 mM) and then incubated with the beads for 20 min with gentle mixing. Next, the tubes were placed on a magnet, the supernatant was removed, and the beads were washed with 100 and 90 μl PBS, respectively (the beads were transferred to new tubes between washing steps, and the total washing time was 5 min). Finally, the beads were incubated with 9.5 μl of a 5% (v/v) acetic acid solution for 2 min in order to elute captured peptides. The eluate was then used directly for mass spectrometry analysis without any additional clean up.

An electrospray ionization LTQ-Orbitrap XL mass spectrometer (Thermo Electron, Bremen, Germany) interfaced with an Eksigent nanoLC 2DTM Plus HPLC system (Eksigent Technologies, Dublin, CA) was used for all samples. The autosampler injected 6 μl of the GPS-generated eluates. A blank LC-MS/MS run was used between analyzed samples. Peptides were loaded with a constant flow rate of 15 μl/min onto a pre-column (PepMap 100, C18, 5 μm, 5 mm × 0.3 mm, LC Packings, Amsterdam, The Netherlands). The peptides were subsequently separated on a 10-μm fused silica emitter, 75 μm × 16 cm (PicoTipTM Emitter, New Objective, Inc., Woburn, MA), packed in-house with Reprosil-Pur C18-AQ resin (3 μm; Dr. Maisch, GmbH, Ammerbuch-Entringen, Germany). Peptides were eluted with a 35-min linear gradient of 3% to 35% (v/v) acetonitrile in water containing 0.1% (v/v) formic acid, with a flow rate of 300 nl/min. The LTQ-Orbitrap was operated in data-dependent mode to automatically switch between Orbitrap-MS (from *m/z* 400 to 2000) and LTQ-MS/MS acquisition. Four MS/MS spectra were acquired in the linear ion trap per each Fourier transform MS scan, which was acquired at 60,000 full width at half-maximum nominal resolution using the lock mass option (*m/z* 445.1200257) for internal calibration. The dynamic exclu-

sion list was restricted to 500 entries using a repeat count of two with a repeat duration of 20 s and with a maximum retention period of 120 s. Precursor ion charge state screening was enabled to select for ions with at least two charges and rejecting ions with undetermined charge states. The normalized collision energy was set at 35%, and one microscan was acquired for each spectrum. The complete study was run using 26 days of MS-instrumentation time, divided into four blocks of 6.5 days each (one CIMS-binder mix/block). All samples were individually analyzed one time per CIMS-binder mix. In addition, triplicate captures of selected samples were performed within each block as back-to-back LC-MS/MS runs. The reference sample was repeatedly analyzed over time within and between the four blocks (supplemental Fig. S1). A total of 238 LC-MS/MS runs were performed. Blank beads (*i.e.* beads without any conjugated antibody) were exposed to the pooled digest to allow us to evaluate potential bead background binding peptides. Based on the low number of identified background binding peptides from two blank bead "captures," all generated data were left unfiltered unless otherwise noted.

Protein Identification and Quantification—The generated data were first analyzed using Proteios (32) for generating identifications using both Mascot and X!Tandem. Briefly, all files were processed and converted into mzML and mgf format using the Proteios software environment (v 2.17) platform, and the following search parameters were used for Mascot and X!Tandem: enzyme, trypsin; one missed cleavage; fixed modification, carbamidomethyl (C); variable modification, methionine oxidation (O). In addition, a variable N-acetyl was allowed for searches performed in X!Tandem. A peptide mass tolerance of 3 ppm and a fragment mass tolerance of 0.5 Da were used, and searches were performed against a forward and a reverse combined database (*Homo sapiens*, Swiss-Prot, August 2011, resulting in a total of 71,324 database entries). The automated database searches in both Mascot and X!Tandem and consequent combination (with a false discovery rate (FDR) of 0.01) were used (estimated on the basis of the number of identified reverse hits) for generating peptide identifications. The search results from both Mascot and X!Tandem were combined at the peptide-spectrum match level when calculating peptide-level FDRs. All peptide identifications passing the FDR combined threshold were kept. For details regarding the Proteios software equipment see Hakkinen et al. (32). Protein identifications derived from Proteios were generated by finding protein groups for peptides that passed the peptide combined FDR cutoff and then were further filtered based on a protein FDR of 0.01. This could be done via a search in the target-decoy database, and the decoys were kept in the combined hits report to then set the protein FDR. The proteins were assembled per sample, and the "Occam's razor" approach was used when calculating protein groups. A spectral library was generated that can be directly uploaded in Skyline (33) for viewing of all fragment ion spectra (see supplemental data).

Because the Proteios software environment at the time of analysis offered no quantitative label-free plug-in analyzing modules (development is in progress), Progenesis-LC-MS software (v 4.0, Nonlinear Dynamics, Newcastle upon Tyne, UK) was used for generating all quantitative values. Briefly, the raw data files were converted to mzXML using the ProteoWizard software package prior to application of the Progenesis-LC-MS software. The built-in feature-finding tool, Mascot search tool, and combined fractions tool (CIMS-binder mixes 1, 2, 3, and 4) with default settings and minimal input were used. To obtain optimal feature alignment, the first injection run of the pooled sample for each CIMS-binder mix (supplemental Fig. S1) was used as a reference alignment file, except for CIMS mix 3 runs, in which the halfway pool run was used as the reference alignment file. Features aligned and detected with retention times between 10 and 50 min for CIMS-binder mixes 1 and 2 and between 10 and 49 min for CIMS-binder mixes 3 and 4 were included for quantification. Due to limita-

tions of the Progenesis-LC-MS software, the identification was limited to only Mascot searches, meaning that no X!Tandem-generated peptide identifications from Proteios were included for downstream quantitative analysis. The same database (*Homo sapiens*, Swiss-Prot, August 2011, a forward and a reverse combined database) and search parameters as mentioned above were used, and a cutoff FDR value of 0.01 was applied. Furthermore, the default protein options for protein grouping and protein quantitation within the Progenesis-LC-MS software were used (*i.e.* quantitate from nonconflicting features and group similar proteins). All values that were reported from Progenesis-LC-MS as between 0 and 1 were set to 1. The generated normalized abundance values were then used (log2 values) for statistical and bioinformatics analysis. For details of all protein identifications and protein quantifications, see the supplemental data.

Statistical and Bioinformatical Analysis—Qlucore Omics Explorer v 2.2 (Qlucore AB, Lund, Sweden) was used for identifying significantly up- or down-regulated proteins ($p < 0.01$) using a one-way analysis of variance (ANOVA). The q-values were generated based on the Benjamini–Hochberg method (34). Principal component analysis plots and heat maps were generated in Qlucore or Matrix2png (35). The support vector machine (SVM) is a learning method (36) that was used to classify the samples using a leave-one-out cross-validation procedure, and the analyses were performed on both unfiltered and *p*-value-filtered data. A receiver operating characteristic (ROC) curve (37), constructed using the SVM decision values and the area under the curve (AUC), was used as a measurement of the performance of the classifier.

Hence, differentially expressed proteins identified from three-group comparisons were fed into the SVM for two group comparisons. We selected this as our initial strategy because the sample cohort for grade 1 in particular was small so that the cohort could be divided into a training set and a test set (which would be the preferred approach). However, in order to assess the effect and evaluate the potential risk of overestimation of ROC–AUC values when using the entire sample cohort for deriving significant analytes prior to an SVM ROC analysis, we randomly divided our samples into 10 different training (two-thirds of grades 1, 2, and 3) and test sets (one-third of grades 1, 2, and 3). This was done to demonstrate the range of AUC values obtained when using only two-thirds of all samples to derive a candidate panel, using an ANOVA (three-group, $p < 0.01$ in Qlucore) comparison as before. Subsequently, using the training set and the candidate marker panel, an SVM was performed. The SVM was then frozen and applied to the independent test set (one-third of the samples), generating an ROC AUC value (26). This procedure was then repeated nine additional times using the next randomly generated training and test sets and candidate marker panels. Furthermore, Ingenuity Systems Pathway Analysis (IPA) (v 11904312) was used for the significantly differentially expressed proteins in order to extract information such as protein localization, potential network interactions, transcription factor associations, and association with tumorigenesis. The experimentally derived protein signatures were finally validated at the mRNA level using the GOBO search tool against large cohorts of published gene expression data of defined breast cancer tissues (38), including clinical parameters such as histological grades 1, 2, and 3 or ER status. The validation cohort was composed of 1,881 mRNA samples (based on 11 public datasets), of which 1,411 had assigned histological grades, including grade 1 ($n = 239$), grade 2 ($n = 677$), and grade 3 ($n = 495$). 1,620 tumors had assigned ER status, including ER-positive ($n = 1,225$) and ER-negative tumors ($n = 395$). The GOBO tool calculates for each tumor (in the database) an activity value of the eight gene modules (emulating breast-cancer-specific biological processes) assessed. The average expression level can then be determined for a gene set submitted to GOBO. Consequently, for each gene set and each of the modules, the Spearman correlation can be determined over all tumors. A p(ANOVA) value is calculated and reported for all the modules, and this simply tests the null hypothesis that a gene set has the same association to all modules. For specific details regarding the underlying calculations and activity values of the different modules emulating breast-cancer-specific biological processes, see Refs. 38 and 39.

RESULTS

Protein expression profiles of 52 breast cancer tissue extracts were deciphered, including identification and quantification. Tissue biomarker signatures and individual proteins, reflecting either histological grade or tumor progression, could be delineated. An overall workflow outlining the experimental design is shown in supplemental Fig. S1.

Protein Coverage, Dynamic Range, and Assay Performance—Using the GPS technology based on nine antibodies, a total of 2,140 protein groups were identified (Figs. 1A–1C). The identification reproducibility was high, resulting in a 54.7% peptide overlap (supplemental Fig. S2A). In comparison, the reference sample, which was repeatedly analyzed throughout the entire project, showed a 43.9% peptide identification overlap (supplemental Fig. S2B). Of the identified proteins, a total of 1,388 were quantified (supplemental Fig. S3) and subsequently used in the search for tumor-associated markers. The total median coefficient of variation for quantification for one sample (ID 7267) was 10.8% (supplemental Fig. S4A), whereas the corresponding total median coefficient of variation for the reference sample was 22.8% (supplemental Fig. S4B). Notably, about 38% (833 peptides) of the quantified peptides had not previously been reported in the PeptideAtlas (Fig. 1D) (40), indicating substantial novel coverage. This was further highlighted by the fact that a significant portion of the detected peptides were shorter than those previously reported, with a median length of 9 *versus* 11 amino acids commonly found in PeptideAtlas (Fig. 1E).

The distribution of measured $\log_2$-MS intensity normalized abundances for all quantified proteins was assessed and indicated a dynamic range of $\sim 10^6$ (supplemental Fig. S3A). The in-depth coverage generated by the GPS technology was further illustrated by the fact that peptides both frequently and rarely reported in PeptideAtlas were detected (Fig. 1F). The detected proteins were then grouped by major biological process and were found to cover several groups (supplemental Fig. S3B). Of note, proteins associated with processes such as translation (*e.g.* 60S ribosomal protein) were found to display a higher overall abundance than other proteins involved in, for example, mitosis (*e.g.* CDK1), demonstrating the capability to provide deep and reproducible coverage of both high- and low-abundance proteins in breast tissue.

Protein Expression Profiles Reflecting Histological Grades 1 to 3—We first examined whether a tissue protein signature reflecting histological grade could be deciphered. Using a multivariate analysis (three-group comparison), we identified 49 significantly ($p < 0.01$, q-value < 0.25) differentially expressed proteins among the grade 1, grade 2, and grade 3
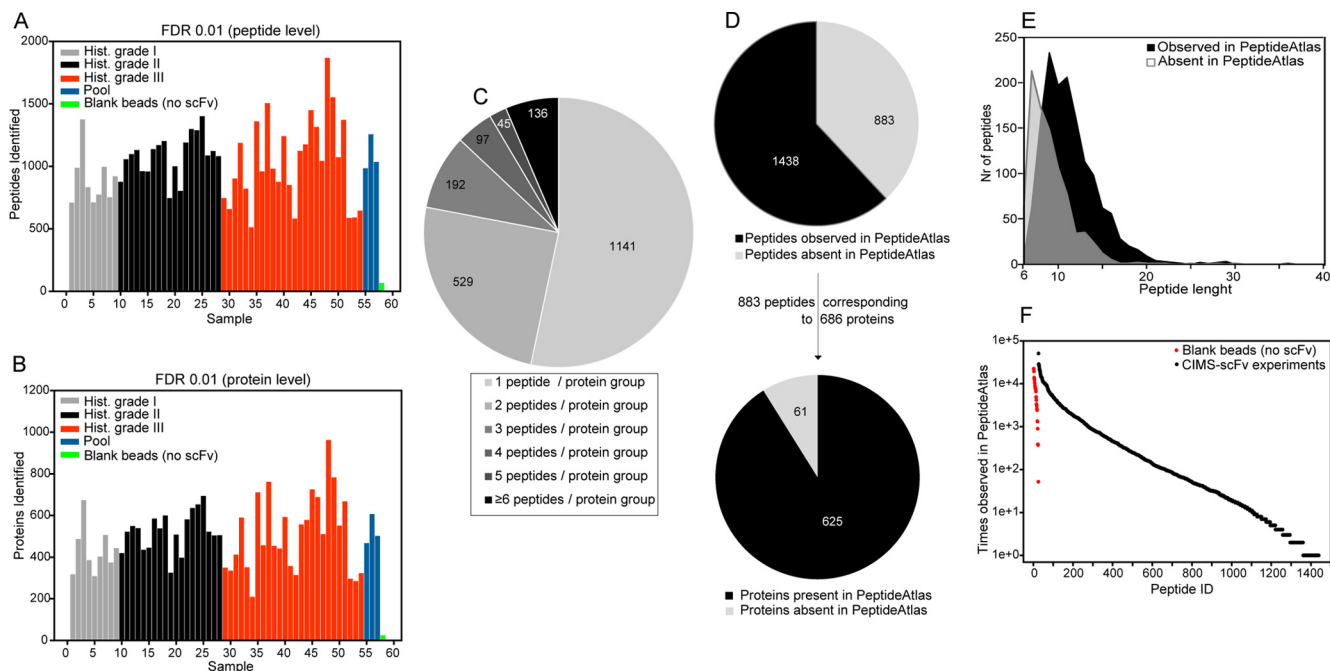
FIG. 1. **Peptide and protein statistics.** *A*, total number of unique peptide sequences identified per sample (using Mascot + X!Tandem; FDR of 0.01). Bars ordered according to supplemental Table S1. Data for samples analyzed in triplicate also displayed (bars 26–28, 52–54, and 55–57). *B*, total number of assembled protein groups identified per sample (FDR of 0.01, set at protein level, using Mascot + X!Tandem). Bars ordered according to supplemental Table S1. Data for samples analyzed in triplicate also displayed (bars 26–28, 52–54, and 55–57). *C*, number of unique peptides per protein group (FDR of 0.01, set at protein level, using Mascot + X!Tandem) resulting in a total protein coverage of 2,140 protein groups in the entire study. (Data based on all samples and runs, including replicates, pool runs, and samples with missing clinical parameters.) *D*, evaluation of quantified peptides (Progenesis LC-MS software, limited to Mascot scored peptides using an FDR of 0.01) against the PeptideAtlas (version 2011–08 Ens62, human). In addition, for peptides not present in the PeptideAtlas, a second comparison was performed in order to evaluate whether the corresponding protein had been reported. In cases of multiple protein accessions, all were assessed. *E*, comparison of peptide length. *F*, observed peptide frequency in PeptideAtlas.

cohorts (supplemental Table S3). Based on this protein signature, principal component analysis plots showed that histological grade 1 and grade 3 tumors could be well separated, whereas histological grade 2 tumors appeared to be more heterogeneous (Fig. 2), indicating a further subclassification of grade 2 into two subgroups. A clear trend involving both up- and down-regulated proteins could be observed with increasing histological grade. For example, cyclin-dependent kinase 1 (CDK1), minichromosome maintenance complex component 3 (MCM3), DNA replication licensing factor MCM7, ATP-citrate synthase, polyadenylate-binding protein 4, and 6-phosphofructokinase type C displayed an increasing trend and were most up-regulated in grade 3 tumors (Fig. 2 and supplemental Fig. S5). In contrast, analytes such as asporin, spondin, keratocan, chymase, and olfactomedin-like protein 3 displayed higher expression levels in histological grade 1 than in grade 3 tumors (Fig. 2 and supplemental Fig. S6).

We then examined whether the 49 *p*-value-filtered (*p* < 0.01) proteins could be used to classify the tissues based on histological grade. To this end, we ran a leave-one-out cross-validation with the SVM and collected the decision values for all samples. We adopted this approach because the number of samples, especially of grade 1, was too low for the samples

to be split into a test set and a training set. The prediction values were then used to construct an ROC curve, and the AUC values were calculated (Fig. 2). The results showed that the histological grade tumor subgroups could be well separated (AUC = 0.75 to 0.93), although grade 2 again displayed a more heterogeneous pattern.

To examine the robustness of the classification, we then re-analyzed the data after having split the sample set into a random training set (two-thirds of the samples) and test set (one-third of the samples). This process was repeated 10 times, and consequently 10 frozen SVMs were used to test the deciphered biomarker signatures on the test sets. The results showed that the median AUC value was 0.86 for the classification of grade 1 *versus* grade 3 tumors (Fig. 2). The median AUC value for the classification of grade 1 *versus* grade 2 was 0.67, and for the classification of grade 2 *versus* grade 3 tumors the median AUC value was 0.65 (Fig. 2). Hence, consistent results were obtained also when using more stringent data analysis approaches.

After ER status had been eliminated as a potential confounding variable, 27 analytes were still present, of which 25 overlapped with the original 49, and the AUC was 0.90 instead of 0.93 for grade 1 *versus* grade 3 (supplemental Table S3). As
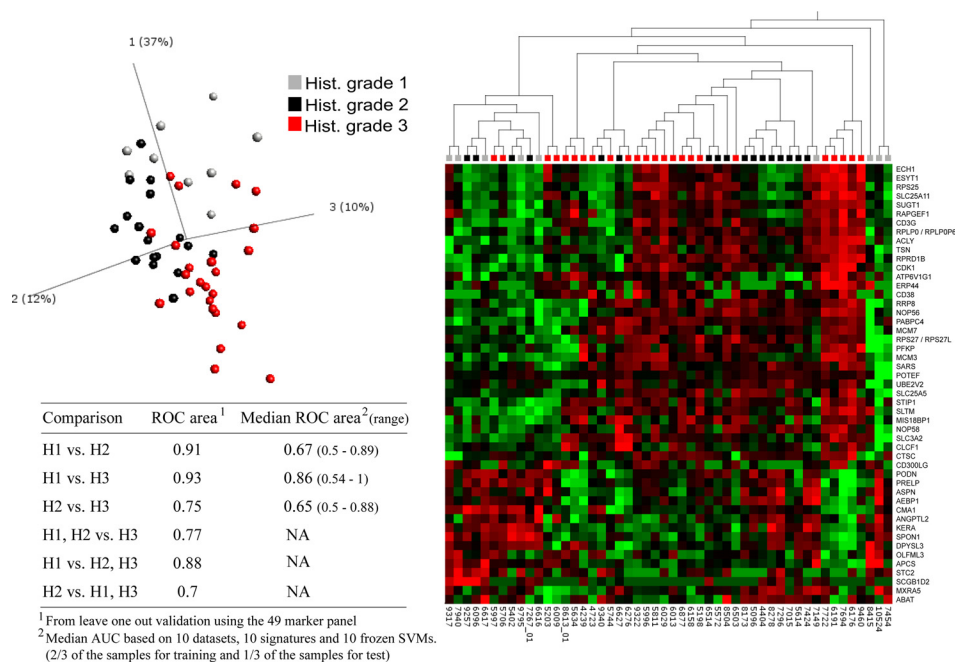
Comparison	ROC area[1]	Median ROC area[2] (range)
H1 vs. H2	0.91	0.67 (0.5 - 0.89)
H1 vs. H3	0.93	0.86 (0.54 - 1)
H2 vs. H3	0.75	0.65 (0.5 - 0.88)
H1, H2 vs. H3	0.77	NA
H1 vs. H2, H3	0.88	NA
H2 vs. H1, H3	0.7	NA

[1] From leave one out validation using the 49 marker panel
[2] Median AUC based on 10 datasets, 10 signatures and 10 frozen SVMs.
(2/3 of the samples for training and 1/3 of the samples for test)

FIG. 2. **Significantly differentially expressed proteins based on histological grade.** Differentially expressed analytes are shown in heatmaps (red, up-regulated; green, down-regulated). Principal component analysis plot and associated heatmap of histological grade 1, grade 2, and grade 3 samples (data filtered on variance 0.2, p value < 0.01, q-value < 0.25). Results from a leave-one-out validation approach (using SVM) demonstrated with ROC-AUC values based on using all 49 proteins are presented. In addition, the median ROC-AUC values (for 10 random training and test sets) reported when using two-thirds of our samples to first derive significant proteins (using an ANOVA (three-group, $p < 0.01$ in Qlucore)) were then fed to an SVM, and the test set was tested (one-third of the samples) using the frozen SVM.

could be expected, similar results were observed whether ER status or PR status was eliminated as a confounding variable, because the distributions of these two clinical parameters followed each other closely among the tumor samples (supplemental Table S3). When the HER2 status was eliminated as a potential confounding variable, it resulted in an overlap of 35 proteins ($p < 0.01$) with the 49-protein panel (supplemental Table S3).

Next, we investigated the effect of using a two-group comparison instead of a multivariate approach to define differentially expressed markers between individual grades (supplemental Fig. S7). As might be expected, data showed that the classification of the individual histological subgroups was improved, as judged by the AUC values (AUC = 0.91 to 0.92). Focusing on histological grade 1 *versus* grade 3, 50 significantly ($p < 0.01$) differentially expressed proteins were identified (supplemental Table S4), and 31 proteins overlapped with the previous 49-marker signature (see Fig. 2 and supplemental Fig. S7C) and resulted in an overlap of 19 proteins with the 50-marker panel when ER status was eliminated as a potential confounding factor (supplemental Table S4). When histological grade 2 was mapped onto the frozen 50-protein comparison of grade 1 *versus* grade 3, it again displayed large heterogeneity and partly overlapped with both cohorts (see Fig. 2 and supplemental Fig. S7D). When an unsupervised clustering analysis was used on only grade 2 samples, a subdivision into two subgroups was evident (Appendix, Fig.

S8A). We therefore split grade 2 into two groups, 2a ($n = 11$) and 2b ($n = 6$), and showed that 2a was closer associated with grade 1, whereas 2b resembled grade 3 (supplemental Fig. S8B), which indicated that molecular diagnostics might further refine the classification owing to its higher resolution.

Biological Relevance of Proteins Associated with Histological Grades 1–3—The biological relevance of the 49-tissue protein signature differentiating histological grades 1 to 3 was then examined. To this end, the cellular localization of each individual protein was mapped using the IPA software (Fig. 3), and network-associated functions and potential relationships were investigated (supplemental Fig. S9). A trend of down-regulated (extracellular matrix (ECM)) and up-regulated proteins (plasma membrane, cytoplasm, and nucleus) going from grade 1 to grade 3 was identified. Of note, the top-ranked network had the highest expression in grade 3 and was found to be associated with DNA replication, recombination, and repair; cell cycle; and free radical scavenging. The second highest ranked network contained protein associated with tissue structural proteins and was most abundant in grade 1 tumors. For example, a majority of ECM proteins were found in this network, and several were directly or indirectly associated with transforming growth factor β (TGFβ1) (supplemental Fig. S9B). In the top-ranked network, several proteins were directly or indirectly associated with NF-kB and VEGF (supplemental Fig. S9A). In addition, the relationship between the 49-analyte signature and the transcription factor network was
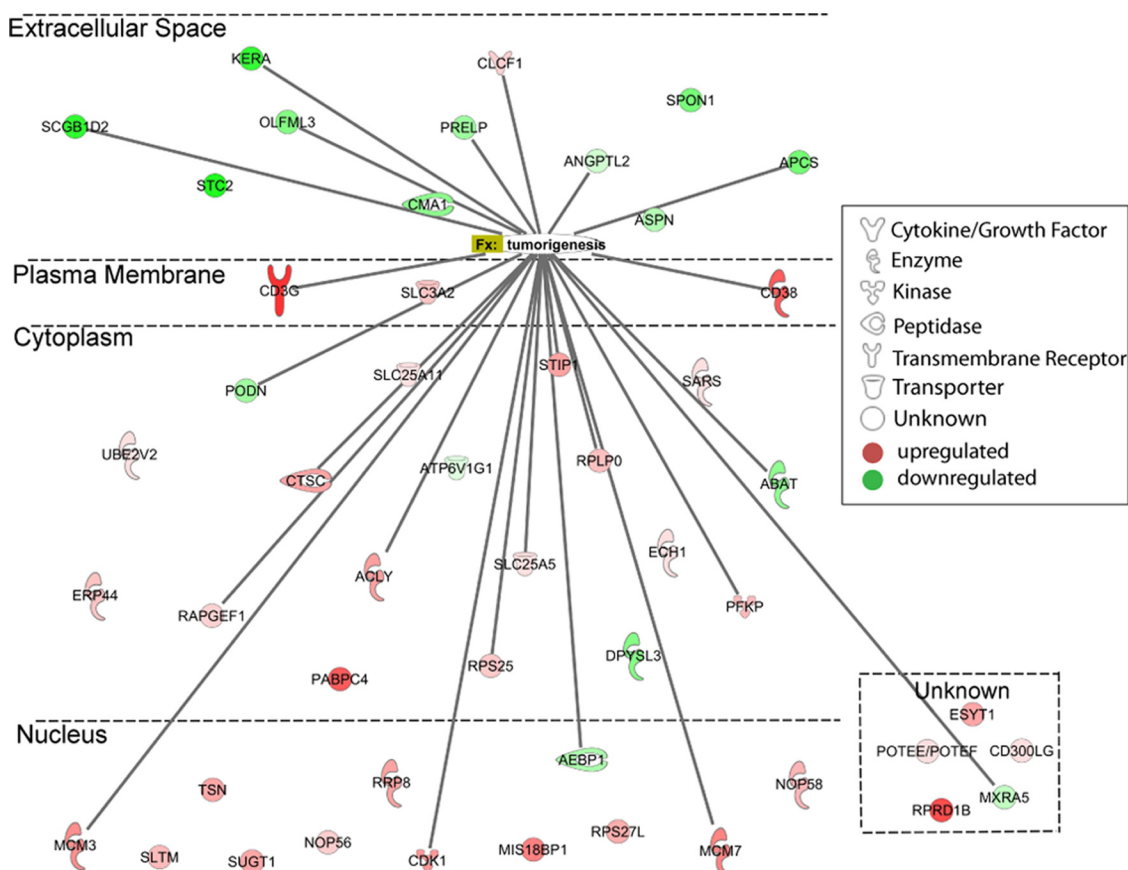
FIG. 3. **Biological relevance of differentially expressed analytes in the three histologically graded tumor types determined using IPA.** The 49 proteins identified as significantly differentially expressed proteins in the three tumor cohorts mapped to their cellular localization. Log$_2$-ratio (median grade 3/median grade 1); red denotes up-regulation, and green denotes down-regulation. Proteins with known associations to tumorigenesis are indicated.

also assessed using IPA (supplemental Fig. S9C), and Rb and E2F2 were found to be among the top associated transcription regulators (supplemental Fig. S9C).

Validation of Candidate Signature Using Independent Data— In order to validate the 49-tissue protein signature discriminating histological grades 1 to 3, protein data were compared with independent publicly available orthogonal mRNA profiling data of breast cancer. The validation cohort was composed of 1,881 mRNA samples (based on 11 public datasets), of which 1,411 had an assigned histological grade of grade 1 ($n = 239$), grade 2 ($n = 677$), or grade 3 ($n = 495$). Forty-two of the 49 analytes could be mapped to the gene expression data using Gene Entrez ID and were subsequently used in the validation test (supplemental Table S5). The 42 analytes were then split into two groups based on the observed down- (15 analytes) or up-regulated (27 analytes) protein expression profile for grade 3 *versus* grade 1 and compared with the corresponding mRNA expression profiles (Fig. 4). The protein expression profiles of both down-regulated (*e.g.* spondin 1 and keratocan) (Fig. 4A, supplemental Figs. S6, S10I, and S10J) and up-regulated proteins (*e.g.* CDK1 and MCM3) (Fig. 4B, supplemental Figs. S5, S10A, and S10B) were found

to corroborate well with the mRNA expression levels in the majority of cases. In one case (serum amyloid P component) the protein expression profile decreased in grade 3 (supplemental Fig. S6); this did not correlate with the mRNA profile, which was unchanged (supplemental Fig. S10G). Interestingly, the up-regulated protein markers in grade 3 were found to display mRNA profiles with a high correlation to checkpoint and M-phase gene modules (Fig. 4A), whereas the group of down-regulated protein markers displayed mRNA profiles with high correlation to the stromal gene module (Fig. 4B).

Assessing Distant Metastasis-free Survival—Finally, we examined whether the 49-tissue protein signature reflecting histological grade also could be used to assess distant metastasis-free survival, using the same publicly available gene expression dataset. Forty-two of 49 analytes could be mapped to 1,379 mRNA samples with 10-year end-point survival data. The markers were split into two groups, reflecting down-regulated ($n = 15$) and up-regulated ($n = 27$) markers in grade 3 *versus* grade 1, and Kaplan–Meier analyses were then performed to calculate distant metastasis-free survival with a 10-year end point by stratifying the gene expression data into three quantiles (low, intermediate, and high) based on the

FIG. 4. **Validation of protein expression profiles using an orthogonal method.** To this end, mRNA expression profiles based on data from 1,411 histologically graded tumor samples were used. 42 of 49 differentially expressed proteins among histological grades 1, 2, and 3 were successfully mapped (using Gene Entrez ID) into the GOBO database. *A*, mRNA expression profiles for proteins found to display decreased protein expression in histological grade 3 tumors (median ratio compared with histological grade 1); 15 (out of 16 total) analytes could be mapped with the GOBO tool. In addition, the correlation of the 15 genes to different gene set module expression patterns is indicated. Gray dots indicate actual correlation values. *B*, mRNA expression profiles for proteins found to display increased expression in histological grade 3 tumors (compared with histological grade 1); 27 (of 33) could be mapped with the GOBO tool. In addition, the correlation of the 27 genes to different gene set module expression patterns is indicated. Gray dots indicate actual correlation values.
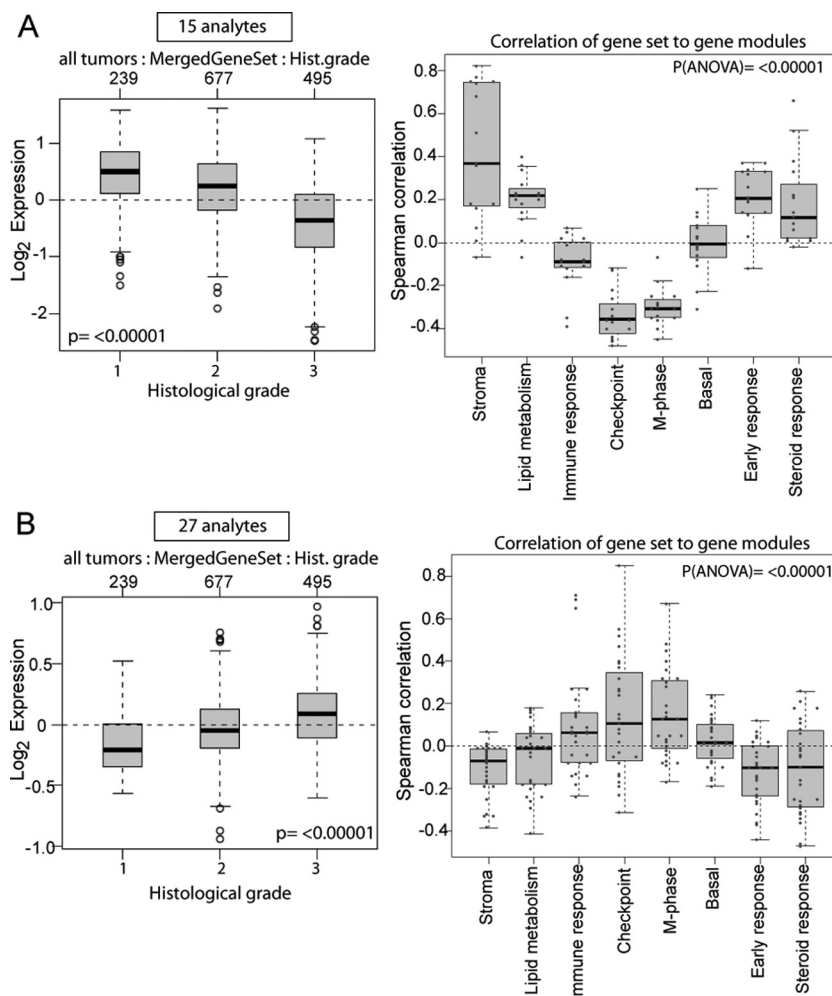
expression levels of these up- and down-regulated genes (supplemental Fig. S11). The data implied that in particular the cohort of down-regulated genes (mainly ECM-associated) might be useful in predicting distant metastasis-free survival. In fact, this might even be accomplished by targeting single down-regulated (*e.g.* keratocan and olfactomedin-like protein 3) or up-regulated (*e.g.* CDK1) proteins—that is, low levels of olfactomedin-like protein 3 and high levels of CDK1 increase the risk for distant metastasis.

DISCUSSION

Breast cancer grading has important clinical relevance, because it influences the therapy selected for each patient. However, the histological grading today depends on a microscopic analysis of cellular morphology and is, like most visual analyses, subject to more operator-dependent variation than molecular diagnostics. Consequently, we set out to design an improved grading analysis based on protein expression portraits reflecting each grade, and to identify trends in expression patterns associated with disease aggressiveness. Furthermore, information on pathways regulated in association with tumor grade also has the potential to provide insight into

the mechanisms underlying tumor progression, as well as an improved understanding of the features of histological grade that influence prognosis.

By combining label-free LC-MS/MS with an affinity proteomic step, using only nine so-called CIMS antibodies, we were able to identify the first tissue protein signature associated with tumor grade and disease progression in breast cancer. This was accomplished by profiling 52 breast cancer tissues and generating detailed, quantified proteomic maps of 1,388 proteins using the recently developed GPS technology (27, 29, 30). The proteome coverage for the GPS platform is limited by the range of specificities of the nine CIMS antibodies, as well as by conventional MS-MS-related limitations (*e.g.* MS time, LC gradient design, and limit of detection). In order to optimize the MS time, we applied a limited set of CIMS antibodies and used short LC gradients. Hence, a highly streamlined workflow and reasonable sample throughput were achieved. In comparison, performing the same study using a conventional MS-based approached would have been challenging considering the number of fractions (obtained via prefractionation) per sample and instrument time that would have been needed in order to handle the complexity of breast

tissue proteomes (24). This was exemplified by Geiger *et al.,* (24) who separated peptides via strong anion exchange into six fractions that were then concentrated and purified using C$_{18}$ StageTips and subjected to LC-MS/MS analysis with a long LC gradient. Furthermore, relative to GPS, an antibody-based microarray approach offers a substantially higher throughput and can easily handle hundreds of samples within a short period of time (41). However, the microarray platform does not offer a discovery mode as the GPS methodology does, because the array approach needs one antibody per protein of interest with a pre-defined specificity.

The GPS allowed the identification of a 49-plex tissue protein signature differentiating histological grade 1, 2, and 3 breast cancer tumors with high specificity and sensitivity. Furthermore, using only two-thirds of the samples as a training set for deriving candidate markers and testing using frozen SVMs still resulted in high median ROC-AUC values, particularly for histological grade 1 *versus* grade 3 (0.86). The 49-plex molecular protein fingerprint supported the current view that grade 1 and grade 3 tumors were more distinct, whereas grade 2 tumors were more heterogeneous (9). In fact, the data indicated that grade 2 could be further subdivided into grades 2a and 2b.

The biological relevance of the identified signature was manifest in comparison with the characteristic cancer hallmarks (42), and a clear trend was observed in protein expression from grade 1 to grade 3 breast tumors. Grade 1 expressed stromal and ECM-associated protein, indicating a more structurally conserved tumor, whereas grade three tumors appeared to have lost those properties and, in addition, expressed higher levels of proteins involved in cell proliferation and mitosis. This observation coincides well with the concept that histological grade is based on mitotic index, nuclear pleomorphism, and differentiation (2). Proliferation has also been recognized as one of the key prognostic factors in breast cancer and has been found to be one of the major components of several prognostic gene expression signatures (43, 44). In this study, CDK1, MCM3, and MCM7 were among the markers found to increase in expression from histological grade 1 tumors to grade 3 tumors (Fig. 2 and supplemental Fig. S5). CDK1 is a key player in cell cycle regulation (45), and it was recently demonstrated that the depletion of CDK1 compromises the ability of cells to repair DNA by means of homologous recombination. Consequently, as reduced CDK1 expression impairs BRCA1 function and DNA repair, the inhibition of CDK1 represents a potential strategy for expanding the utility of poly-ADP ribose polymerase inhibitors to BRCA-competent cancers (46). MCM2 and MCM7 have been shown to play a role in both initiation and elongation phases of eukaryotic DNA replication (47, 48). The overexpression of MCM3 has been identified in primary cancer tissues, including carcinomas of breast, colon, kidney, cervix, and stomach, as well as in a number of cancer cell lines, implying a role for MCM3 in tumorigenesis (48). Note-

worthy, another minichromosome maintenance complex (MCM6) is one of the 70 genes (15) included in the current MammaPrint® test. In addition to the above mentioned nucleus-associated proteins, stress-induced-phosphoprotein 1 and polyadenylate-binding protein 4, both localized in the cytoplasm, also displayed increased protein expression going from histological grade 1 to grade 3. Stress-induced-phosphoprotein 1 has been shown to be secreted by ovarian cancer cells into their environment and is functional in promoting cell proliferation (49). The polyadenylate-binding proteins, aside from binding to poly(A) sequences, have critical roles in RNA processing and can be shuttled from the nucleus to the cytoplasm with mRNAs, increase eIF4F assembly to caps, aid in the recruitment of ribosomal subunits to 5′ UTRs, and increase the reuse of translational machinery after polypeptide synthesis (50). Thus, the polyadenylate-binding proteins are also directly or indirectly involved in cell proliferation and were most highly expressed in grade 3 tumors.

One of the primary metabolic changes associated with proliferating tumor cells is the induction of aerobic glycolysis. Phosphofructokinase has been demonstrated to play a crucial role in glycolytic activities and cell proliferation in breast cancer, and it is another potential target in designing selective breast cancer chemotherapy (51). Phosphofructokinase displayed a trend of increased expression from grade 1 to grade 3. Two additional proteins with potential importance for tumor cells adopting metabolic changes that displayed significantly increased expression trends were ACLY and ADP/ATP translocase 2 (SLC25A5), both localized in the cytoplasm.

Notably, all the ECM-associated proteins, such as asporin, keratocan, spondin 1, chymase 1, olfactomedin-like protein 3, and stanniocalcin-2, were found to display a trend of decreasing protein expression, with the lowest levels in the grade 3 tumors (Fig. 2 and Fig. 3), again indicating disseminating tumor–stromal interactions, as ECM goes through drastic changes and collapses during tumor dissemination. The ECM of breast cancers is considered abnormal and is believed to promote tumor progression (52), and it regulates gene expression and phenotype through adhesion-mediated signaling (53). Bergamaschi and colleagues defined several ECM signatures based on gene expression profiling and suggested that primary breast tumors could be classified based upon the ECM composition (54). Thus, the importance of the ECM and stromal characteristics, and their capability to provide relevant information about breast carcinomas, is now further indicated by the present proteomic data. Asporin has been proven to be associated with the cartilage matrix (55), and we noted a clear decrease in expression when we compared grades 1 and 3. Notably, one of the top-ranked networks from the IPA analysis included most of the differentially expressed ECM-localized proteins, and TGFβ-1 was reported to be directly or indirectly associated with several of these extracellular proteins (supplemental Fig. S9*B*). TGFβ-1 regulates tumor growth either through mechanisms that function within

the cell itself or through host–tumor cell interactions. It has been shown that factors in the tumor microenvironment, such as the ECM, influence the ability of TGFβ-1 to promote or suppress carcinoma progression and metastasis (56). Consequently, the systematic inhibition of TGFβ-1 signaling pathways is being considered as an attractive therapeutic intervention strategy when targeting cancer (56).

Importantly, the protein expression reflecting histological grade was validated, using an independent, large dataset and an orthogonal method (mRNA expression), with the GOBO tool (38). Groups of up- and down-regulated proteins were evaluated based on their correlation to known gene set modules, as it is often the functional processes captured by a gene signature, and not the individual genes, that are important (44). The significant correlation to the gene set modules for stroma, cellular checkpoint, and early response are in particular noteworthy (Fig. 4). Furthermore, when we assessed the distant metastasis-free survival length as the end point, using the proteomic-derived signature, data clearly indicated a worse clinical outcome, in particular when we used the down-regulated ECM proteins. Although mRNA data for the GPS-profiled patient tumors were not present or possible to obtain because of the amount of patient tumor material, it should be noted that the mRNA profiling data within GOBO is based on 11 international, large-scale studies, indicating the robustness of the mRNA data trends. Thus, the independent mRNA validations added support for the candidate protein signatures and their potential capability for breast tumor grading. Briefly, proteins are normally more stable than RNA, which is reflected in the standard clinical handling of tumors, ranging from immediate freezing in liquid nitrogen to incubation on ice for 20 to 30 min, potentially showing significant changes in mRNA levels but fewer (small) changes in protein levels (57). In contrast, in discovery studies, the techniques for mRNA analysis are more mature than multiplexed protein analysis, which in the end might facilitate clinical implementation. However, we believe that the stability of the protein biomarkers is an advantage in a clinical setting where routine protein analysis can be readily implemented using standard ELISA or selected reaction monitoring, increasing the general applicability in nonspecialized hospitals.

In summary, we have defined proteins associated with histological breast cancer grades, including a subclassification of histological grade 2, using a combination of affinity and MS-based proteomics. Several proteins associated with tumor aggressiveness also were identified, and cell proliferation appeared to be one of the main driving mechanisms associated with histological grades 2b and 3 of breast cancer tumors. Furthermore, the identification of several ECM-related proteins displaying reduced expression levels in grades 2b and 3, potentially facilitating remodeling and collapse of the ECM, was equally significant. The reported candidate tissue protein signatures, reflecting histological grade, and the clinical relevance of this will be explored in larger, independent patient cohorts in order to generate pre-validation and validation data, respectively. In this process, the length of the signature will be optimized to yield the best discrimination with the shortest signatures, using a backward elimination approach (41). It could then be envisioned that a selected reaction monitoring or ELISA-based readout might be the desired way to translate current findings into the clinic. In the longer term, these tissue molecular portraits could pave the way for improved histological grading and prognostication.

‡‡ To whom correspondence should be addressed: Carl Borrebaeck, Department of Immunotechnology and CREATE Health, Lund University, Medicon Village, SE-223 81 Lund, Sweden, Tel.: 46-46-2229613, Fax: 46-46-2224200, E-mail: carl.borrebaeck@immun.lth.se; Christer Wingren, Department of Immunotechnology and CREATE Health, Lund University, Medicon Village, SE-223 81 Lund, Sweden, Tel.: 46-46-2224323, Fax: 46-46-2224200, E-mail: christer.wingren@immun.lth.se.

Conflict of interest statement: C.A.K.B. and C.W. are inventors on a pending patent application for the grading of tumors based on the findings presented in this study.

REFERENCES

1. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011) Global cancer statistics. *CA Cancer J. Clin.* **61,** 69–90
2. Elston, C. W., and Ellis, I. O. (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19,** 403–410
3. Hondermarck, H., Tastet, C., El Yazidi-Belkoura, I., Toillon, R. A., and Le Bourhis, X. (2008) Proteomics of breast cancer: the quest for markers and therapeutic targets. *J. Proteome Res.* **7,** 1403–1411
4. Ciocca, D. R., and Elledge, R. (2000) Molecular markers for predicting response to tamoxifen in breast cancer patients. *Endocrine* **13,** 1–10
5. Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344,** 783–792
6. Hudis, C. A. (2007) Trastuzumab—mechanism of action and use in clinical practice. *N. Engl. J. Med.* **357,** 39–51
7. Robbins, P., Pinder, S., de Klerk, N., Dawkins, H., Harvey, J., Sterrett, G., Ellis, I., and Elston, C. (1995) Histological grading of breast carcinomas: a study of interobserver agreement. *Hum. Pathol.* **26,** 873–879
8. Frierson, H. F., Jr., Wolber, R. A., Berean, K. W., Franquemont, D. W., Gaffey, M. J., Boyd, J. C., and Wilbur, D. C. (1995) Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am. J. Clin. Pathol.* **103,** 195–198
9. Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., Desmedt, C., Larsimont, D., Cardoso, F., Peterse, H., Nuyten, D., Buyse, M., Van de Vijver, M. J., Bergh, J., Piccart, M., and Delorenzi, M. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98,** 262–272

10. Dowsett, M., Goldhirsch, A., Hayes, D. F., Senn, H. J., Wood, W., and Viale, G. (2007) International Web-based consultation on priorities for translational breast cancer research. *Breast Cancer Res.* **9,** R81

11. Ferno, M., Stal, O., Baldetorp, B., Hatschek, T., Kallstrom, A. C., Malmstrom, P., Nordenskjold, B., and Ryden, S. (2000) Results of two or five years of adjuvant tamoxifen correlated to steroid receptor and S-phase levels. South Sweden Breast Cancer Group, and South-East Sweden Breast Cancer Group. *Breast Cancer Res. Treat.* **59,** 69–76

12. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature* **406,** 747–752

13. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10869–10874

14. Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., Wong, J. E., Liu, E. T., Bergh, J., Kuznetsov, V. A., and Miller, L. D. (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66,** 10292–10301

15. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536

16. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009

17. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J., and Wolmark, N. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817–2826

18. Hanash, S. (2003) Disease proteomics. *Nature* **422,** 226–232

19. Gong, Y., Wang, N., Wu, F., Cass, C. E., Damaraju, S., Mackey, J. R., and Li, L. (2008) Proteome profile of human breast cancer tissue generated by LC-ESI-MS/MS combined with sequential protein precipitation and solubilization. *J. Proteome Res.* **7,** 3583–3590

20. Bouchal, P., Roumeliotis, T., Hrstka, R., Nenutil, R., Vojtesek, B., and Garbis, S. D. (2009) Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J. Proteome Res.* **8,** 362–373

21. Sutton, C. W., Rustogi, N., Gurkan, C., Scally, A., Loizidou, M. A., Hadjisavvas, A., and Kyriacou, K. (2010) Quantitative proteomic profiling of matched normal and tumor breast tissues. *J. Proteome Res.* **9,** 3891–3902

22. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7,** 383–385

23. Strande, V., Canelle, L., Tastet, C., Burlet-Schiltz, O., Monsarrat, B., and Hondermarck, H. (2009) The proteome of the human breast cancer cell line MDA-MB-231: analysis by LTQ-Orbitrap mass spectrometry. *Proteomics Clin. Appl.* **3,** 41–50

24. Geiger, T., Madden, S. F., Gallagher, W. M., Cox, J., and Mann, M. (2012) Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Res.* **2012,** 2428–2439

25. Carlsson, A., Wingren, C., Ingvarsson, J., Ellmark, P., Baldertorp, B., Ferno, M., Olsson, H., and Borrebaeck, C. A. (2008) Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur. J. Cancer* **44,** 472–480

26. Carlsson, A., Wingren, C., Kristensson, M., Rose, C., Ferno, M., Olsson, H., Jernstrom, H., Ek, S., Gustavsson, E., Ingvar, C., Ohlsson, M., Peterson, C., and Borrebaeck, C. A. (2011) Molecular serum portraits in patients

27. with primary breast cancer predict the development of distant metastases. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 14252–14257

27. Wingren, C., James, P., and Borrebaeck, C. A. (2009) Strategy for surveying the proteome using affinity proteomics and mass spectrometry. *Proteomics* **9,** 1511–1517

28. Stoevesandt, O., and Taussig, M. J. (2012) Affinity proteomics: the role of specific binding reagents in human proteome analysis. *Expert Rev. Proteomics* **9,** 401–414

29. Olsson, N., Wingren, C., Mattsson, M., James, P., O'Connell, D., Nilsson, F., Cahill, D. J., and Borrebaeck, C. A. (2011) Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Mol. Cell. Proteomics* **10,** M110.003962

30. Olsson, N., James, P., Borrebaeck, C. A., and Wingren, C. (2012) Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry. *Mol. Cell. Proteomics* **11,** 342–354

31. Olsson, N., Wallin, S., James, P., Borrebaeck, C. A., and Wingren, C. (2012) Epitope-specificity of recombinant antibodies reveals promiscuous peptide-binding properties. *Protein Sci.* **21,** 1897–1910

32. Hakkinen, J., Vincic, G., Mansson, O., Warell, K., and Levander, F. (2009) The Proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *J. Proteome Res.* **8,** 3037–3043

33. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968

34. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57,** 289–300

35. Pavlidis, P., and Noble, W. S. (2003) Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19,** 295–296

36. Cortes, C., and Vapnik, V. (1995) Support-vector networks. *Machine Learning* **20,** 273–297

37. Lasko, T. A., Bhagwat, J. G., Zou, K. H., and Ohno-Machado, L. (2005) The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.* **38,** 404–415

38. Ringner, M., Fredlund, E., Hakkinen, J., Borg, A., and Staaf, J. (2011) GOBO: gene expression-based outcome for breast cancer online. *PloS One* **6,** e17911

39. Fredlund, E., Staaf, J., Rantala, J. K., Kallioniemi, O., Borg, A., and Ringner, M. (2012) The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res.* **14,** R113

40. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9,** 429–434

41. Wingren, C., Sandstrom, A., Segersvard, R., Carlsson, A., Andersson, R., Lohr, M., and Borrebaeck, C. A. (2012) Identification of serum biomarker signatures associated with pancreatic cancer. *Cancer Res.* **72,** 2481–2490

42. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144,** 646–674

43. Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008) Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin. Cancer Res.* **14,** 5158–5165

44. Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F., Goldstein, D. R., Piccart, M., and Delorenzi, M. (2008) Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res.* **10,** R65

45. Malumbres, M., and Barbacid, M. (2009) Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* **9,** 153–166

46. Johnson, N., Li, Y. C., Walton, Z. E., Cheng, K. A., Li, D., Rodig, S. J., Moreau, L. A., Unitt, C., Bronson, R. T., Thomas, H. D., Newell, D. R., D'Andrea, A. D., Curtin, N. J., Wong, K. K., and Shapiro, G. I. (2011) Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. *Nat. Med.* **17,** 875–882

47. Cortez, D., Glick, G., and Elledge, S. J. (2004) Minichromosome mainte-

nance proteins are direct targets of the ATM and ATR checkpoint kinases. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 10078–10083

48. Ha, S. A., Shin, S. M., Namkoong, H., Lee, H., Cho, G. W., Hur, S. Y., Kim, T. E., and Kim, J. W. (2004) Cancer-associated expression of minichromosome maintenance 3 gene in several human cancers and its involvement in tumorigenesis. *Clin. Cancer Res.* **10,** 8386–8395

49. Wang, T. H., Chao, A., Tsai, C. L., Chang, C. L., Chen, S. H., Lee, Y. S., Chen, J. K., Lin, Y. J., Chang, P. Y., Wang, C. J., Chao, A. S., Chang, S. D., Chang, T. C., Lai, C. H., and Wang, H. S. (2010) Stress-induced phosphoprotein 1 as a secreted biomarker for human ovarian cancer promotes cancer cell proliferation. *Mol. Cell. Proteomics* **9,** 1873–1884

50. Mangus, D. A., Evans, M. C., and Jacobson, A. (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol.* **4,** 223

51. Moon, J. S., Kim, H. E., Koh, E., Park, S. H., Jin, W. J., Park, B. W., Park, S. W., and Kim, K. S. (2011) Kruppel-like factor 4 (KLF4) activates the transcription of the gene for the platelet isoform of phosphofructokinase (PFKP) in breast cancer. *J. Biol. Chem.* **286,** 23808–23816

52. Place, A. E., Jin Huh, S., and Polyak, K. (2011) The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Res.* **13,** 227

53. Fata, J. E., Werb, Z., and Bissell, M. J. (2004) Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast Cancer Res.* **6,** 1–11

54. Bergamaschi, A., Tagliabue, E., Sorlie, T., Naume, B., Triulzi, T., Orlandi, R., Russnes, H. G., Nesland, J. M., Tammi, R., Auvinen, P., Kosma, V. M., Menard, S., and Borresen-Dale, A. L. (2008) Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J. Pathol.* **214,** 357–367

55. Turashvili, G., Bouchal, J., Baumforth, K., Wei, W., Dziechciarkova, M., Ehrmann, J., Klein, J., Fridman, E., Skarda, J., Srovnal, J., Hajduch, M., Murray, P., and Kolar, Z. (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* **7,** 55

56. Bierie, B., and Moses, H. L. (2006) Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nat. Rev. Cancer* **6,** 506–520

57. Skold, K., Alm, H., and Scholz, B. (2013) The impact of biosampling procedures on molecular data interpretation. *Mol. Cell. Proteomics* **12,** 1489–1501