

Integrative Analysis of Genomics and Proteomics Data on Clinical Breast Cancer Tissue Specimens Extracted with Acid Guanidinium Thiocyanate–Phenol–Chloroform

René B. H. Braakman,^{*,†,||} Karel Bezstarosti,[‡] Anieta M. Sieuwerts,^{†,||} Vanja de Weerd,[†] Anne M. van Galen,[†] Christoph Stingl,[§] Theo M. Luider,[§] Mieke A. M. Timmermans,[†] Marcel Smid,[†] John W. M. Martens,^{†,||} John A. Foekens,^{†,||} Jeroen A. A. Demmers,[‡] and Arzu Umar^{*,†,||}

[†]Department of Medical Oncology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Be401, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

[‡]Proteomics Center, Erasmus MC, University Medical Center Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

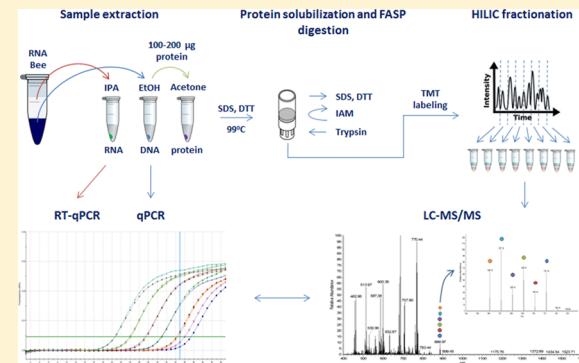
[§]Department of Neurology, Erasmus MC, University Medical Center Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

^{||}Postgraduate School of Molecular Medicine, Erasmus MC, University Medical Center Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Supporting Information

ABSTRACT: Acid guanidinium thiocyanate, phenol, and chloroform extraction (AGPC) is a commonly used procedure to extract RNA from fresh frozen tissues and cell lines. In addition, DNA and proteins can be recovered, which makes AGPC an attractive source for integrative analysis on tissues of which little material is available, such as clinical specimens. Despite this potential, AGPC has only scarcely been used for proteomic analysis, mainly due to difficulties in extracting proteins. We have used a quantitative mass spectrometry method to show that proteins can readily be recovered from AGPC extracted tissues with high recovery and repeatability, which allows this method to be used for global proteomic profiling. Protein expression data for a selected number of clinically relevant markers, of which transcript and protein levels are known to be correlated, were in agreement with genomic and transcriptomic data obtained from the same AGPC lysate. Furthermore, global proteomic profiling successfully discriminated breast tumor tissues according to their clinical subtype. Lastly, a reference gene set of differentially expressed transcripts was strongly enriched in the differentially abundant proteins in our cohort. AGPC lysates are therefore well suited for comparative protein and integrative analyses.

KEYWORDS: breast cancer, sample preparation, RNA Bee, acid guanidinium phenol-chloroform, proteomics, genomics



INTRODUCTION

In-depth characterization and quantification of proteins can currently be performed on a wide range of clinical materials, such as fresh frozen¹ and formalin-fixed paraffin-embedded material.² While methods are efficient, requiring little sample, the material available for research is not always of sufficient quantity for multiple studies, such as integrative analysis of DNA, RNA, protein, and metabolites. Alternative sources are therefore still desired. A different and often more readily available alternative is lysates from tissues extracted with acid guanidinium thiocyanate, phenol, and chloroform (AGPC, commercially sold as RNA Bee, Trizol, etc.), a commonly used procedure to isolate RNA from cell lines and fresh frozen tissues.^{3,4} This liquid–liquid extraction results in a water phase, containing RNA, and an organic phase, containing DNA and

protein. As DNA and proteins can be recovered from the organic phase,⁵ this allows integration of data obtained through gene expression and proteomics analyses on fully identical starting material. Archived organic phases are therefore an interesting alternative when the source material is no longer available for DNA or protein extraction. AGPC-extracted clinical specimens are increasingly used for comparative protein analysis, but these are mostly limited to immunoblotting, 2D electrophoresis (2-DE) or 2D difference in gel electrophoresis (2D-DIGE) applications.^{6–10} Shotgun mass-spectrometry-based protein analysis from AGPC organic phases has only scarcely been reported,^{11–14} mainly as a result of difficulties

Received: December 1, 2014

Published: January 22, 2015

associated with recovering proteins from the organic fraction.^{5,15–18} For protein recovery, usually a precipitating agent, such as acetone, is added to the phenolic phase, and proteins are pelleted through centrifugation. The resulting protein pellet is large, dense and difficult to dissolve completely. Others have therefore attempted to improve dissolution of the pellet, either by focusing on the precipitation method,¹⁶ the solvent and method to dissolve the pellet,^{17–23} or by avoiding precipitation altogether,⁵ with varying success. Nanoscale liquid-chromatography-based mass-spectrometry-based proteomics approaches, however, require input material in the submicrogram region. The protein pellet corresponding to this amount is smaller and will more readily dissolve in a buffer with a strong detergent, such as sodium dodecyl sulfate (SDS). While SDS is incompatible with digestion and downstream proteomics analysis,^{24–26} robust methods have been developed that exchange SDS with a buffer that is compatible with downstream processing, such as filter-aided sample preparation (FASP).²⁷ In the filtering step, next to SDS, other low-molecular-weight compounds that may interfere with digestion are removed prior to digestion. FASP therefore matches well with protein digestion of AGCP extracted tissues. Together, this enables quantitative protein profiling of AGPC organic fractions using shotgun proteomics.

In this study we used a tandem mass tag (TMT) multiplexed quantitative shotgun proteomics method to compare protein recovery and proteome coverage from tissues extracted with RNA Bee, a commercially available AGPC, against a matching tissue lysate prepared in SDS. Furthermore, we performed comparative analysis on RNA-Bee-extracted breast cancer specimens of different molecular subtypes and compared protein abundance data of three clinically used markers, estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), with corresponding transcript and copy number data. We have previously shown that mRNA expression of these markers is correlated with their corresponding protein abundance,^{28,29} which allowed us to directly evaluate the applicability of the method by comparing the mRNA and protein levels of these markers in AGCP lysates. Our results show that overall protein recovery is largely similar compared with a tissue lysate, with the exception of extracellular matrix proteins. Furthermore, different molecular subtypes were readily discriminated based on protein abundance data, and protein abundance, transcriptomic, and copy number variation data were in agreement for the three tested markers.

MATERIALS AND METHODS

Patients and Tumor Tissue

Snap-frozen breast tumor tissues ($n = 11$) were used from our liquid N₂ bio bank, which were selected based on invasive tumor percentage (nucleus count) and presence or absence of ER (ESR1 transcript level), PR (PgR transcript level), and HER2 (ERBB2 transcript level), the status of which was previously determined by reverse-transcriptase quantitative polymerase chain reaction (RT-qPCR), as previously described.^{29,30} One tumor tissue was used for comparison of proteins extracted from RNA Bee versus SDS lysates ($n = 1$) (experiment 1, Supplementary Table 1 in the SI); 10 tumor tissues were used for comparison between molecular subtypes, of which 4 overexpressed HER2, with negative/low ER and PR expression (HER2); five were positive for ER and PR, with

negative/low HER2 expression (ERPOS); and one did not express ER, PR or HER2 (ERNEG) (experiment 2, Supplementary Table 1 in the SI). This study was approved by the Medical Ethics Committee of the Erasmus MC Rotterdam, The Netherlands (MEC 02.953), and was performed in accordance to the Code of Conduct of the Federation of Medical Scientific Societies in The Netherlands (<http://www.federa.org/>).

RNA Extraction, cDNA Synthesis, and RT-qPCR

Total RNA was isolated from each sample using RNA Bee (TelStat, Friendswood) according to the manufacturer's protocol and as previously described.³¹ In brief, 30 μm deep frozen tissue sections (approximately 50–100 mg) were lysed in 1 mL of RNA Bee reagent, followed by phase separation by the addition of chloroform, isopropanol precipitation of the RNA present in the upper water phase, and ethanol washes of the resulting RNA pellet. The pellet was dissolved in 50 μL of RNase-DNase-free water and, together with the ~600 μL organic phase containing DNA and protein, stored at –80 °C for downstream processing.

DNA Extraction from Tissues and RNA Bee Remains, and Measuring ERBB2 Gene Amplification

Genomic DNA (gDNA) was extracted with the NucleoSpin Tissue kit (Macherey-Nagel, BIOKE, Leiden, NL) from 2 to 5 \times 20 μm sections cut in between the sections used for the RNA isolation. In addition, gDNA was precipitated from the organic phase left after RNA Bee isolation by the addition of 100% ethanol in a 1:2 (EtOH/organic phase) ratio, followed by centrifugation at 1500g for 5 min at 4 °C. Resulting DNA was processed with the ZymoBead Genomic DNA kit (Zymo Research Corporation, BaseClear, Leiden, NL) according the manufacturer's instructions. All fractions were stored at –80 °C for downstream processing. Quality and quantity were assessed by Nanodrop, the Quant-iT PicoGreen dsDNA kit (Life Technologies, Bleiswijk, NL), and agarose gel electrophoresis.

Next, ERBB2 and EFTUD2 were quantified in real time in an MX3000 PCR apparatus (Agilent, Amsterdam, NL) with previously published FAM and VIC labeled Taqman assays (Applied Biosystems, Life Technologies), respectively.³² The ratio of ERBB2 over EFTUD2 as chromosome 17 reference was used to establish the ERBB2 amplification status. For this, gDNA isolated from a panel of 18 breast cancer cell lines with known ERBB2 amplification status³³ and profiled similarly as previously described was used to determine the cutoff for ERBB2 amplification.

Protein Extraction

Following DNA extraction, an aliquot of 15, 30, or 50 μL of the remaining phenol–ethanol supernatant was used for RNA Bee protein extraction. Proteins were precipitated by the addition of three volumes of acetone and incubation at room temperature for 5 min. Samples were then centrifuged at 14 000g for 15 min at 4 °C. The pellet was washed twice with 95% ethanol and redissolved in 25 μL of 4% SDS and 100 mM dithiothreitol (DTT) in 100 mM Tris/HCl pH 8 by heating to 99 °C for 15 min. A matching SDS tissue lysate was prepared from the same tumor tissues using identical buffers and conditions as used to dissolve the RNA Bee protein pellets.

Protein Digestion

Samples were digested using FASP according to Wisniewski et al.,²⁷ with minor modifications. Protein samples were diluted to 0.5% SDS by the addition of 8 M urea in 100 mM Tris/HCl

pH 8 and loaded on a 30 kDa molecular weight cutoff filter. Samples were spun at 14 000g for 15 min. 200 μ L of 8 M urea in 100 mM Tris/HCl pH 8 was added to the proteins on the filter and the filter was spun down once more. Proteins were then alkylated by the addition of 50 μ L of 50 mM iodoacetamide in 8 M urea to the filters and incubation at room temperature for 20 min. Filters were spun down at 14 000g for 10 min and washed twice by the addition of 100 μ L of 8 M urea in 100 mM Tris/HCl pH 8 and centrifugation for 10 min at 14 000g. Filters were then washed with 50 mM triethylammonium bicarbonate (TEAB) twice. One μ g endoproteinase Lys-C and 2 μ g trypsin were added and samples were digested overnight in a wet chamber at 37 °C. Peptides were recovered by centrifugation for 10 min at 14 000g, followed by a wash with 50 μ L of TEAB and another 10 min of centrifugation at 14 000g. Peptide concentration of FASP-digested samples was measured by UV absorbance at 280 nm on a Nanodrop spectrophotometer (Thermo Scientific); 1.1 AU was estimated to correspond to a 1 mg/mL peptide concentration.

Sample Labeling

Samples were labeled with TMT reagents, according to the instructions of the manufacturer. Briefly: TMT reagents (0.8 mg) were dissolved in 41 μ L of anhydrous acetonitrile. For comparison between SDS and RNA Bee lysates (experiment 2), 75 μ L of the digest was diluted in 100 μ L in 50 mM TEAB and added to the reagents as follows: 30 μ L of SDS lysate (TMT 126); 50 μ L of RNA Bee lysate, replicate 1 (TMT 127); 30 μ L of RNA Bee lysate (TMT 128); 50 μ L of RNA Bee lysate, replicate 2 (TMT 129); 15 μ L of RNA Bee lysate (TMT 130); 50 μ L of RNA Bee lysate, replicate 3 (TMT 131). For comparison between subtypes, samples were divided over two six-plex experiments (experiments 2a and 2b). A reference was prepared by pooling 6 μ g peptides of each sample. For experiment 2a, a volume corresponding to 30 μ g of peptides was labeled as follows: pool (TMT 126), S001HER2 (TMT 127), S002HER2 (TMT 128), S003HER2 (TMT 129), S004ERPOS (TMT 130), and S005ERPOS (TMT 131). For experiment 2b, a volume corresponding to 30 μ g of peptides was labeled as follows: pool (TMT 126), S006HER2 (TMT 127), S007HER2 (TMT 128), S008ERPOS (TMT 129), S009ERPOS (TMT 130), and S010ERPOS (TMT 131). Labeling overview and detailed sample information for experiments 1 and 2 are given in Supplemental Table 1 in the SI.

Following incubation at room temperature for 1 h, the reaction was quenched with 2 μ L of 5% hydroxylamine. TMT-labeled samples were combined at a 1:1:1:1:1:1 ratio. Samples were lyophilized and subjected to C18 solid-phase extraction (SPE) (Sep-Pak, Waters).

Sample Fractionation

Samples were fractionated with HILIC on an Agilent 1100 HPLC system (Agilent, Santa Clara, CA) using a 5 μ m particle 4.6 \times 250 mm TSKgel amide-80 column (Tosoh Biosciences, Tokyo, Japan). A total of 180 (experiment 2) and 340 μ g (experiment 1) of desalted peptides were loaded onto the column in 80% acetonitrile. Next, peptides were eluted with a nonlinear gradient from 80% solvent B (100% acetonitrile) to 0% solvent B with a flow of 1 mL/min. Solvent A consisted of 20 mM ammonium formate in water. Sixteen 6 mL fractions were collected during the gradient. Fractions were lyophilized and pooled into eight fractions.

LC–MS

Measurements were performed with a nano liquid chromatography (LC) system (nanoAcquity UPLC (Waters) coupled online to a hybrid Quadrupole/Orbitrap mass spectrometer equipped with a nanospray source (Q Exactive, Thermo Fisher Scientific, San Jose, CA). Peptides were loaded onto a Waters Symmetry C18 180 μ m \times 20 mm, 5 μ m particle, and 100 Å pore size trap column at 1 μ L/min for 10 min; then, the trap column was switched online with a BEH130, 75 μ m \times 250 mm, 1.7 μ m particle, and 130 Å pore size nanoAcquity UPLC column (Waters), and peptides were eluted using a nonlinear gradient from 2 to 30% solvent B in 120 min, where solvent A consisted of 0.1% formic acid in HPLC water and solvent B consisted of 100% acetonitrile and 0.1% formic acid in HPLC water. All LC solvents were purchased from Biosolve, Valkenswaard, The Netherlands. The column flow rate was set to 300 nL/min. For MS detection, a data-dependent acquisition method was used. A high-resolution survey scan from 400–1600 Th was detected in the Orbitrap (target of automatic gain control = 10⁶, maximum injection time = 50 ms, resolution = 70 000 at 200 Th). On the basis of this full scan the 15 most intensive ions were consecutively isolated using a 3.0 Th window, fragmented by higher energy collisionally activated dissociation (HCD, applying a stepped 32 \pm 10% normalized collision energy) and detected in the Orbitrap (AGC target set to 10⁵ ions, maximum injection time = 60 ms, resolution = 17 500 at 200 Th). Precursor masses that were selected for MS/MS once were excluded for MS/MS fragmentation for 10 s.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium³⁴ via the PRIDE partner repository with the data set identifier PXD001533.

Data Analysis

Data were processed in the trans proteomic pipeline³⁵ (version 4.7.1). Raw data files were converted to mzML and mgf with msconvert³⁶ using vendor centroiding. Mass spectra were searched with Comet,³⁷ Myrimatch³⁸ (mzML), and OMSSA³⁹ (mgf) against a concatenated target-decoy database (Uniprot-Swiss-Prot 2014–4 *Homo sapiens* canonical reference proteome). In all search engines, the search enzyme was strict trypsin (no P-rule) with two missed cleavages allowed and full enzyme specificity. The precursor tolerance was set to \pm 1.5 Da and the fragment tolerance to 0.02 Da. Oxidation of methionine was set as a variable modification and carbamido-methylation of cysteine as a fixed modification. Peptide-to-spectra matches (PSMs) search engine scores were converted to probabilities and validated with PeptideProphet⁴⁰ in the semiparametric mode with accurate mass binning enabled, calculation of decoy probabilities, and a minimum probability threshold of 0. Charge states 1+, 5+, 6+, and 7+ were ignored. PeptideProphet results for each search engine were combined and further refined by iProphet⁴¹ with all available models enabled. Proteins were then inferred by ProteinProphet⁴² using a PSM probability cutoff corresponding with a PSM FDR of 1%. The protein groups were sorted on protein probability, and a cutoff was chosen based on a decoy estimated protein FDR of 1%.

Reporter ions were integrated by TagQuant.⁴³ Data were first denormalized by multiplying with the ion injection time, followed by correction of isotopic interferences. For comparison between subtypes, loading differences across reporter ion channels were corrected using the total reporter abundance for

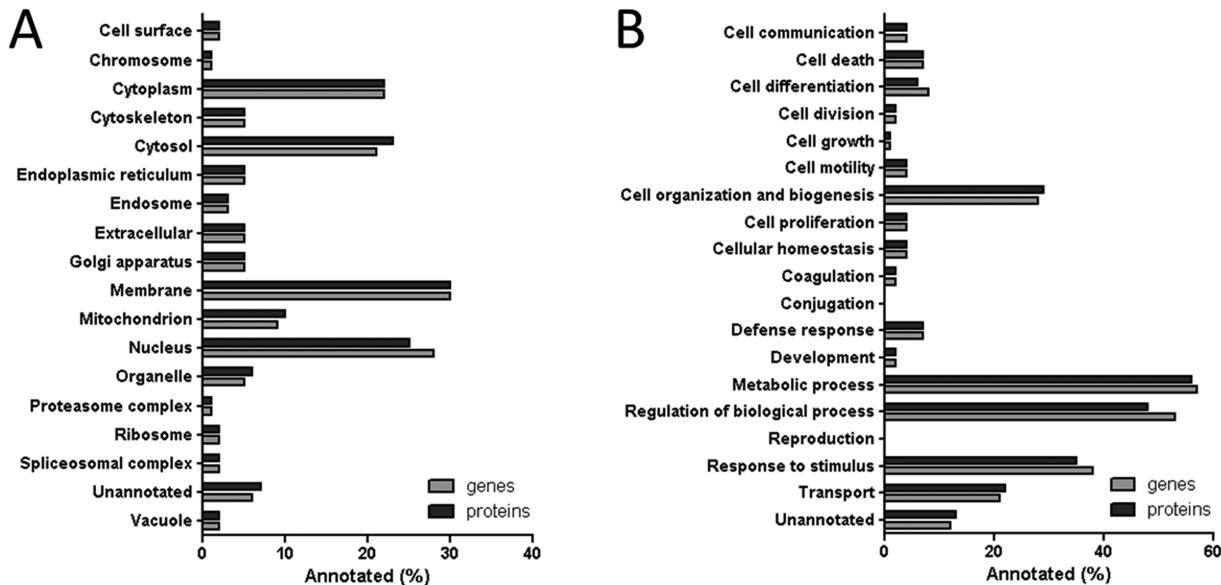


Figure 1. Comparison of RNA Bee extracted proteins with reference breast tumor transcripts. Comparison of gene ontology cellular compartment (GOCC) (A) and gene ontology biological process (GOBP) annotations of identified proteins (B) with annotations of an equal number of the most abundant transcripts identified in a cohort of publicly available data of breast tumors.

each channel as a normalization factor. Peptide abundances were obtained by summing reporter ion intensities of the top three most abundant PSMs. Additional filtering for precursor ion purity was not used, as this only modestly improves accuracy.^{44–46} Finally, protein abundances were obtained by summing razor and unique peptide abundances.

Protein annotation and testing for differences were performed in Perseus (www.biochem.mpg.de/mann/tools/) using a two-sided Welch test (*t* test assuming unequal variances). False discovery rates were estimated based on permutations.⁴⁷ Annotation enrichment analysis was performed in Perseus⁴⁸ or DAVID.^{49,50} A gene set consisting of 450 differentially abundant genes between HER2+/ER- and ER+ subtypes was added to the default Gene Set Enrichment Analysis sets that were used in the 1D enrichment analysis by Perseus. This gene set was obtained using in-house data (*n* = 344) with an additional *n* = 523 breast cancer cases from public repositories (NCBI's Gene Expression Omnibus). All selected cases were lymph-node-negative and did not receive adjuvant hormonal or chemotherapy. Cel files from GEO accession codes GSE2034, GSE5327, GSE11121, GSE2990, and GSE7390 were combined in a single data set, which was normalized using fRMA (median polished). Cutoff points for ER and HER2 positivity were selected based on the distribution of the data for these markers. Next, data were analyzed using BRB-array tools⁵¹ to identify differentially expressed genes; a permutation-based estimated FDR was used to correct for multiple testing. The top 450 most significantly different transcripts were selected as reference gene set.

Hierarchical cluster analysis was performed in the program Cluster⁵² using correlation as a distance metric and centroid linkage as agglomeration method. Resulting data were visualized in JAVA Treeview.⁵³ All other graphs were prepared in Microsoft Excel or GraphPad Prism 5.

RESULTS

Quantitative Analysis of Differences between RNA Bee and SDS Lysates

To characterize differences in proteins recovered from between RNA Bee versus SDS lysates, we selected a breast tumor tissue with >70% nuclei of invasive tumor cells, of which 50 mg tissue was lysed with 1 mL of RNA Bee or 1 mL of 4% SDS and 100 mM DTT. Protein extraction and digestion was performed with a 15, 30, and 50 μ L aliquot of the protein fraction, with the highest volume in triplicate. A corresponding aliquot of the SDS lysate from the same tissue was digested for reference. To rule out bias in colorimetric assays due to trace amounts of phenol or other interfering compounds, recovery and linearity were determined after digestion using peptide UV absorbance at 280 nm. Response was linear ($R^2 = 0.997$) in the range of 15–50 μ L of RNA Bee protein fraction (an estimated 60–200 μ g protein), resulting in approximately 30–100 μ g recovered peptides, respectively (Supplemental Figure 1A in the Supporting Information). These results correspond to what would typically be recovered from a tissue lysate assuming 8–10% of protein weight after lysis and a recovery of 50% or better after FASP. The average peptide recovery from RNA Bee extractions was 86% compared with the recovery from the SDS lysate, with a CV of 9% (*n* = 3, Supplemental Figure 1B in the Supporting Information).

A total of 5585 protein groups were confidently identified and quantified. To determine bias in identified proteins, we compared gene ontology cellular compartment (GOCC) and biological process (GOBP) annotations for identified proteins in our set to the 5500 most abundant transcripts in a cohort of 867 breast cancer patients. The relative number of proteins in each category was nearly identical to those predicted from gene expression data, in both cellular location and biological process, indicating that no particular group of proteins was lost during the processing of the samples with RNA Bee (Figure 1A,B). Next, we compared quantitative data of proteins detected in RNA Bee organic fractions with those detected in the SDS lysate. Protein abundances in RNA Bee fractions were very

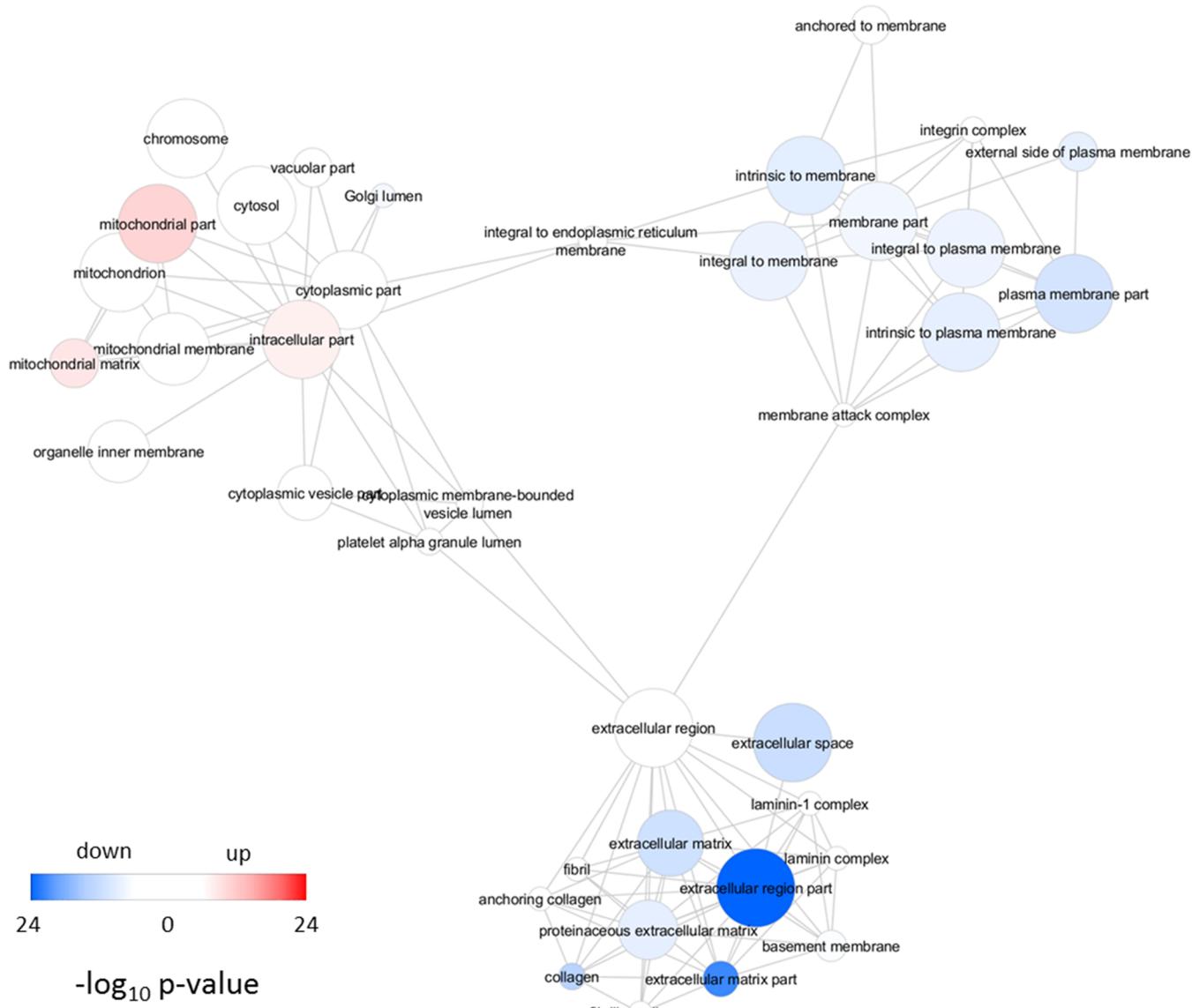


Figure 2. Enriched GOCC terms in 1D enrichment analysis of RNA Bee/SDS protein abundance ratios. (A) Identified proteins in RNA Bee extracted tissues sorted by the \log_2 ratio against a matching tissue lysate prepared in a SDS/DTT lysis buffer. (B) Enriched terms were clustered and visualized in EnrichmentMap. Red: over-represented in RNA Bee, blue: under-represented in RNA Bee. Color intensity indicates significance of the enriched term.

similar to those in the SDS lysate. 94% of all quantified proteins in RNA Bee extracted tissues had an abundance that did not differ by more than a factor of 1.25 compared with the matching SDS tissue lysate (Supplemental Table 2 and Supplemental Figure 2A in the Supporting Information). Nearly all of the proteins with a 1.5-fold change had a lower recovery in RNA Bee: 310 lower versus 9 higher. Proteins that had a lower recovery represented mainly insoluble extracellular matrix proteins and to a lesser extent membrane proteins, whereas mitochondrial proteins had a higher recovery in RNA Bee (Figure 2, Supplemental Figure 2B–D and Supplemental Table 3 in the Supporting Information).

Our results show that proteins can be recovered from RNA Bee organic fractions, with a high repeatability and similarity to a tissue lysate prepared with SDS. This suggests that RNA Bee protein fractions can be used for comparative genomic and proteomic analysis in clinical specimens.

Comparison of Proteins Extracted from RNA Bee Lysates from Breast Tumor Tissues of Different Molecular Subtypes

As a proof-of-principle, we selected 10 RNA Bee protein fractions of breast cancer tissues with different molecular subtypes for proteomics profiling. In total, 5733 protein groups were identified, of which 4967 were quantified in all samples, including ER, PR and HER2, the markers that were used for selection (Figure 3A,B). Hierarchical cluster analysis of the protein data resulted in two clusters that separated the ERPOS (ER+/PR+/HER2-) group from the HER2POS (ER-/PR-/HER2+) and ERNEG (ER-/PR-/HER2-) groups (Figure 3A). In several samples separate and large protein clusters were formed that in most cases could be linked to histological or molecular features, such as amount of stroma or infiltrate, or to breast cancer molecular subtype (Figure 3A and Supplemental Tables 1 (histological features) and 4 (enrichment analysis) in the Supporting Information). Samples S005 and S009 (>50%

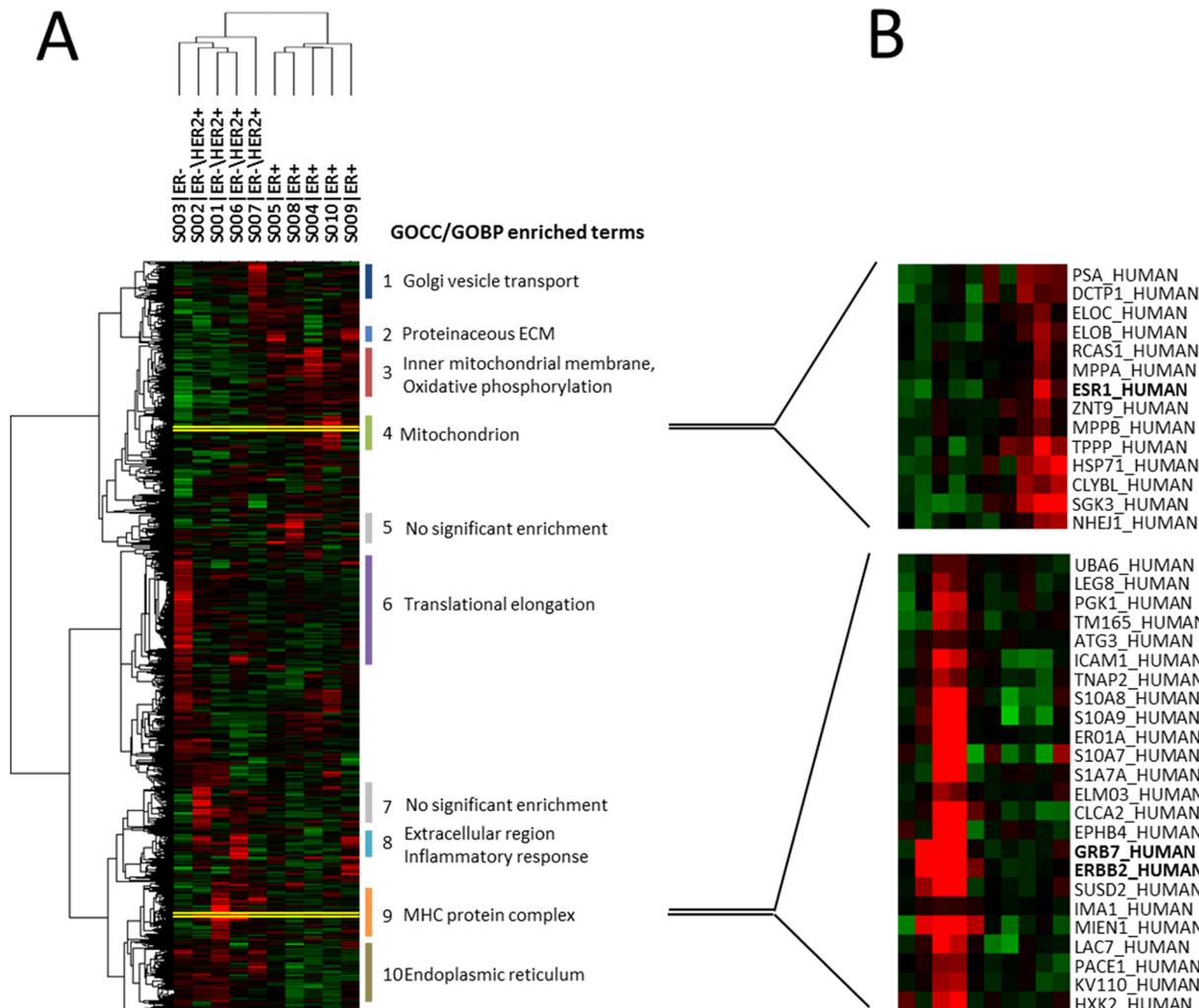


Figure 3. Hierarchical clustering of proteins detected in ten breast tumor tissues of different clinical subtypes. (A) Heatmap of clustered proteins. Two dominant clusters were formed, which separated the samples according to estrogen receptor (ER) status. Within each sample distinct protein clusters were formed that were annotated in DAVID. (B) Zoomed human epidermal growth factor receptor 2 (ERBB2_HUMAN) and estrogen receptor (ESR1_HUMAN) cluster as indicated in panel A in yellow.

stromal area) had a relatively high expression of ECM and proteinaceous ECM proteins ($p = 1.13 \times 10^{-34}$ and $p = 7.31 \times 10^{-15}$, cluster 8 and 2, respectively). Samples S001, S002, and S006 (10% - 30% infiltrating immune cells, all ER negative) had a high expression of proteins involved in inflammatory response (cluster 8, $p = 6.21 \times 10^{-19}$). Overall, ER-negative and HER2-positive tumors had higher abundances of endoplasmic reticulum proteins (cluster 10, $p = 9.55 \times 10^{-8}$). ER-positive tumors, on the other hand, had a higher expression of mitochondrial proteins involved in oxidative phosphorylation (cluster 3, $p = 1.02 \times 10^{-18}$). Next, we compared protein expression of ER, PR, and HER2 between subtypes. A clear distinction was found in levels of these proteins between the ERPOS, HER2POS, and ERNEG groups (Figure 3B). Standardized levels of ER, PR, and HER2 agreed with qPCR data from matching RNA Bee fractions, and the observed protein abundances allowed accurate discrimination between different molecular subtypes (Figure 4A–C). Elevated expression of HER2 in breast cancer is usually the result of gene amplification. For HER2, results were therefore further verified on the DNA level using both DNA extracted from the same RNA Bee fraction as well as extracted DNA from fresh

frozen sections using NucleoSpin. Results for HER2 agreed on DNA, RNA, and protein levels, and all four HER2 amplified tumors were correctly classified, in both RNA Bee and NucleoSpin lysates (Figure 4D). Finally, we compared differentially abundant proteins between HER2+ and ER+ groups with the top 450 differential probes in a reference set of public data of a cohort of 867 breast cancer patients (Figure 5). The small sample size resulted in a high false discovery rate; only four differentially abundant proteins passed an estimated 5% FDR threshold (Supplemental Table 5 in the Supporting Information). However, proteins that overlapped with the selected transcripts were highly enriched in the most differentially abundant proteins ($p = 3 \times 10^{-35}$ and $p = 3 \times 10^{-31}$ for down- and up regulated transcripts in the HER2POS group, respectively) and were nearly exclusively expressed in the same way as the transcripts (Figure 7, Supplemental Table 6 in the Supporting Information). These results indicate that comparative analysis of RNA Bee extracted tissues is feasible, provided that stringent sample inclusion criteria are used or cohorts are sufficiently large to overcome tissue heterogeneity.

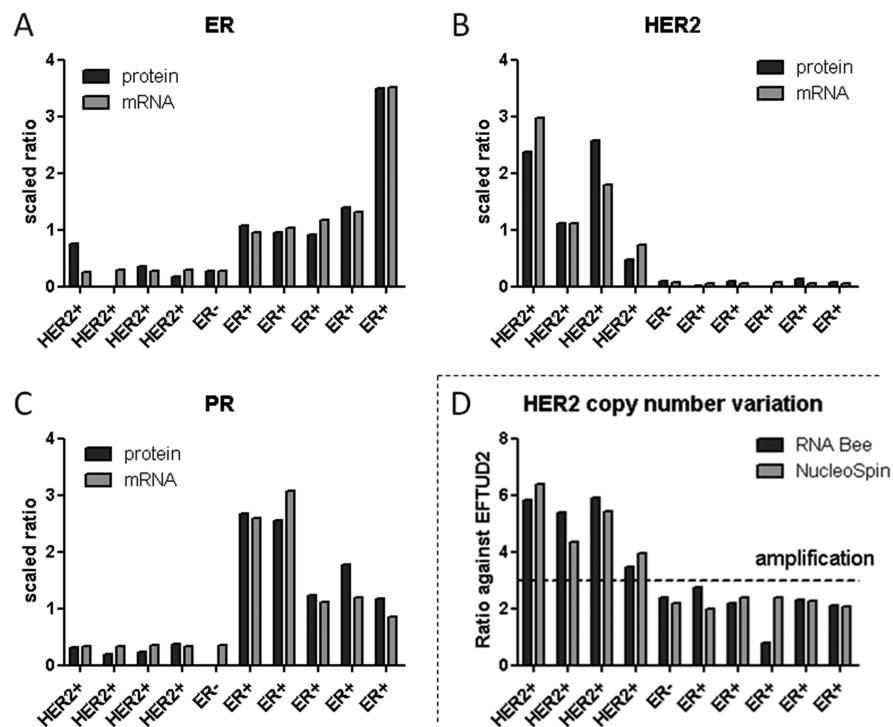


Figure 4. Comparison of protein and mRNA abundance and DNA copy number variation of selected markers. (A–C) Protein ratio's (dark bars) and mRNA abundances (light bars) for ER, PR, and HER2, respectively. Data were scaled to unit variance and centered on the lowest observed value for comparative purposes. (D) Gene amplification data for HER2 were compared between RNA Bee extracted DNA (dark bars) and NucleoSpin extracted DNA (light bars) using HER2/EFTUD2 ratios as determined by qPCR. A ratio greater than 3 was considered as gene amplification (dashed line).

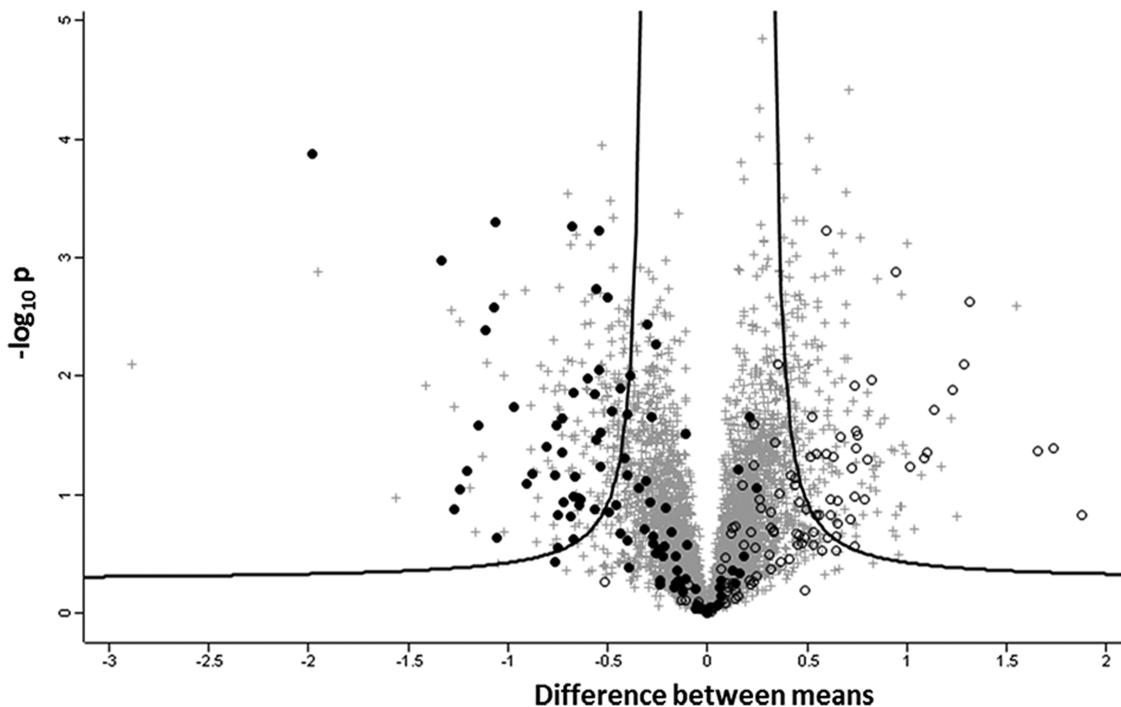


Figure 5. Overlap of differentially abundant proteins with the top 450 gene transcripts obtained from a public data set. (A) Welch test was performed between HER2+ ($n = 4$) and ER+ samples ($n = 5$). A volcano plot was made from all proteins based on the difference between groups (x axis) and their $-\log_{10} p$ value (y axis). Solid line indicates differential abundant proteins that passed an estimated 30% FDR, with an artificially introduced within group standard deviation $s_0 = 0.5$. Proteins are indicated with +, proteins of which the gene transcript was differentially abundant are indicated as ● (up in ER+) or ○ (down in ER+).

■ DISCUSSION

In this work we have shown that proteins can readily be isolated from RNA Bee extracted tissues, with high repeatability and recovery. Identified proteins were similar to proteins obtained from a tissue lysate in a more common lysis buffer, which corresponds to the finding of others.^{11,14,17,18,23} Our quantitative mass spectrometry approach further showed that the recovery of individual proteins was also similar, with nearly 95% of all quantified proteins in RNA Bee within a factor 1.25 of the SDS lysate. Even when considering that ratios might have been underestimated as a result of precursor ion interference,⁴⁴ these differences are sufficiently small to readily allow detection of these proteins in RNA Bee lysates.

Proteins with lower ratios in RNA Bee were mainly less soluble extracellular matrix proteins and, to a lesser extent, membrane proteins. Losses of these proteins may have occurred during the sequential DNA and protein precipitation steps. Because we only used a minor volume of the DNA and protein containing fraction, losses of proteins related to coprecipitation with DNA may be avoided by performing DNA precipitation on a separate aliquot. More ideally, precipitation is avoided altogether, for example, by directly loading the samples on a MWCO filter and directly performing cleanup, buffer exchange, and digestion. Unfortunately, filters typically used with FASP provide low protein adsorption and are not compatible with aggressive solvents such as phenol and chloroform. Nevertheless, while recovery of ECM proteins was lower, tumor tissues with varying stromal content were readily discriminated based on expression of these proteins. Loss of these proteins therefore appears to be limited and quantitative; that is, samples with varying levels of these proteins are accurately discriminated.

Differences in abundances for the clinically relevant proteins ER, PR, and HER2 corresponded well with those of the matching transcripts and could readily classify the samples according to their clinical subtypes. The selection criteria for this study was based on the percentage of tumor nuclei relative to the total nuclei, an appropriate criteria for gene expression studies but less so for proteomics studies because this ignores the amount of extracellular matrix. Despite this possible confounder, we were still able to accurately discriminate between subtypes based on the abundance of these selected proteins. Additionally, cluster analysis of protein abundances of all identified proteins grouped samples according to ER status of the samples. Overall, higher levels of proteins involved in immune regulation and endoplasmic reticulum stress were found in ER-negative tumors, and higher levels of mitochondrial proteins involved in oxidative phosphorylation were found in ER-positive tumors. However, heterogeneity resulted in elevated estimated false discovery rates when testing between subtypes. Few significantly differentially abundant proteins passed a 5% FDR threshold estimated based on permutation of the data, indicating the need of larger cohorts to overcome the heterogeneous nature of invasive breast tumors. Despite this heterogeneity, high overlap was found between differentially abundant proteins with a benchmark gene set consisting of differentially abundant transcripts related to ER or HER2 expression. Regulation of these proteins and genes between molecular subtypes was largely uniform. Together with the individual results for ER, PR, and HER2, this indicates that comparative analysis is possible on the protein level, provided that more strict inclusion criteria are applied or more samples

are included to reduce the effects of heterogeneity and increase the power of the analysis.

In conclusion, the organic fraction of tissues extracted with AGPC provides a rich source of proteins that largely resemble those obtained from a lysate prepared in a common lysis buffer, such as SDS. Protein extraction from AGPC is rapid and results in sufficient material to fractionate and quantitatively label the samples, providing high-throughput and extensive protein coverage. Additionally, as shown by others, post-translational modifications, such as phosphorylation, are preserved in samples extracted with AGPC.¹⁸ For samples that have been processed for RNA extraction, but for which the original specimen is no longer available, AGPC organic fractions are therefore a valuable alternative for protein extraction. Notably for clinical specimens, where the available material is often limited and priority is usually given to genomic analyses, our method provides an opportunity to obtain comprehensive protein information and additionally readily allows integrative genomic and proteomic analyses.

■ ASSOCIATED CONTENT

S Supporting Information

Supplemental Figure 1. Recovery, repeatability, and linearity of protein extraction determined by UV absorption. Supplemental Figure 2. Waterfall plots of selected enriched GO-terms. Supplemental Table 1. Labeling overview and histological and molecular features of TMT-labeled samples. Supplemental Table 2. Recovery of proteins in RNA Bee lysates compared to SDS lysates. Supplemental Table 3. Enrichment analysis on protein ratios between RNA Bee and SDS lysates. Supplemental Table 4. Enriched terms in selected dominant protein clusters in tumors with different molecular subtypes. Supplemental Table 5. Differentially abundant proteins between ER+ and HER2+/ER- molecular subtypes. Supplemental Table 6. Enrichment analysis on differentially abundant proteins between ER+ and HER2+/ER- molecular subtypes. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*A.U.: E-mail: a.umar@erasmusmc.nl. Phone: +31-10-7044373.

*R.B.H.B.: E-mail: r.braakman@erasmusmc.nl.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was financially supported through the Center for Translational Molecular Medicine, Breast CARE (03O-104-06).

■ REFERENCES

- (1) Liu, N. Q.; Dekker, L. J. M.; Stingl, C.; Güzel, C.; De Marchi, T.; Martens, J. W. M.; Foekens, J. A.; Luider, T. M.; Umar, A. Quantitative proteomic analysis of microdissected breast cancer tissues: comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *J. Proteome Res.* 2013, 12, 4627–4641.
- (2) Wiśniewski, J. R.; Duś, K.; Mann, M. Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics. Clin. Appl.* 2013, 7, 225–233.

- (3) Chomczynski, P.; Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **1987**, *162*, 156–159.
- (4) Chomczynski, P.; Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.* **2006**, *1*, 581–585.
- (5) Hummon, A. B.; Lim, S. R.; Difilippantonio, M. J.; Ried, T. Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques* **2007**, *42* (467–470), 472.
- (6) Zhao, R.; Han, C.; Eisenhauer, E.; Kroger, J.; Zhao, W.; Yu, J.; Selvendiran, K.; Liu, X.; Wani, A. A.; Wang, Q.-E. DNA damage-binding complex recruits HDAC1 to repress Bcl-2 transcription in human ovarian cancer cells. *Mol. Cancer Res.* **2014**, *12*, 370–380.
- (7) Trudu, M.; Janas, S.; Lanzani, C.; Debaix, H.; Schaeffer, C.; Ikehata, M.; Citterio, L.; Demarets, S.; Trevisani, F.; Ristagno, G.; et al. Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat. Med.* **2013**, *19*, 1655–1660.
- (8) Grieb, B.; Engler, G.; Sharott, A.; von Nicolai, C.; Streichert, T.; Papageorgiou, I.; Schulte, A.; Westphal, M.; Lamszus, K.; Engel, A. K.; et al. High-frequency stimulation of the subthalamic nucleus counteracts cortical expression of major histocompatibility complex genes in a rat model of Parkinson's disease. *PLoS One* **2014**, *9*, e91663.
- (9) Garnier, M.; Carrier, G.; Rogniaux, H.; Nicolau, E.; Bougaran, G.; Saint-Jean, B.; Cadoret, J. P. Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of *Tisochrysis lutea*. *J. Proteomics* **2014**, *105*, 107–120.
- (10) Butt, R. H.; Pfeifer, T. A.; Delaney, A.; Grigliatti, T. A.; Tetzlaff, W. G.; Coorssen, J. R. Enabling coupled quantitative genomics and proteomics analyses from rat spinal cord samples. *Mol. Cell. Proteomics* **2007**, *6*, 1574–1588.
- (11) Kline, K. G.; Frewen, B.; Bristow, M. R.; MacCoss, M. J.; Wu, C. C. High quality catalog of proteotypic peptides from human heart. *J. Proteome Res.* **2008**, *7*, S055–S061.
- (12) Kline, K. G.; Wu, C. C. MudPIT analysis: application to human heart tissue. *Methods Mol. Biol.* **2009**, *528*, 281–293.
- (13) Ralston-Hooper, K. J.; Turner, M. E.; Soderblom, E. J.; Villeneuve, D.; Ankley, G. T.; Moseley, M. A.; Hoke, R. A.; Ferguson, P. L. Application of a label-free, gel-free quantitative proteomics method for ecotoxicological studies of small fish species. *Environ. Sci. Technol.* **2013**, *47*, 1091–1100.
- (14) Pop, C.; Ameling, S.; Empen, K.; Rüdebusch, J.; Dhople, V. M.; Felix, S. B.; Loghin, F.; Völker, U.; Hammer, E. Proteome analysis of heart biopsies using a TRIzol-based protein extraction. *Clin. Chim. Acta* **2015**, *438*, 246–247.
- (15) Likhite, N.; Warawdekar, U. M. A unique method for isolation and solubilization of proteins after extraction of RNA from tumor tissue using trizol. *J. Biomol. Technol.* **2011**, *22*, 37–44.
- (16) Chey, S.; Claus, C.; Liebert, U. G. Improved method for simultaneous isolation of proteins and nucleic acids. *Anal. Biochem.* **2011**, *411*, 164–166.
- (17) Yamaguchi, H.; Hasegawa, K.; Esumi, M. Protein from the fraction remaining after RNA extraction is useful for proteomics but care must be exercised in its application. *Exp. Mol. Pathol.* **2013**, *95*, 46–50.
- (18) Simões, A. E. S.; Pereira, D. M.; Amaral, J. D.; Nunes, A. F.; Gomes, S. E.; Rodrigues, P. M.; Lo, A. C.; D'Hooge, R.; Steer, C. J.; Thibodeau, S. N.; et al. Efficient recovery of proteins from multiple source samples after TRIzol(®) or TRIzol(®)LS RNA extraction and long-term storage. *BMC Genomics* **2013**, *14*, 181.
- (19) Kirkland, P. A.; Busby, J.; Stevens, S.; Maupin-Furlow, J. A. Trizol-based method for sample preparation and isoelectric focusing of halophilic proteins. *Anal. Biochem.* **2006**, *351*, 254–259.
- (20) Young, C.; Truman, P. Proteins isolated with TRIzol are compatible with two-dimensional electrophoresis and mass spectrometry analyses. *Anal. Biochem.* **2012**, *421*, 330–332.
- (21) Nolan, R. L.; Teller, J. K. Diethylamine extraction of proteins and peptides isolated with a mono-phasic solution of phenol and guanidine isothiocyanate. *J. Biochem. Biophys. Methods* **2006**, *68*, 127–131.
- (22) Man, T.-K.; Li, Y.; Dang, T. A.; Shen, J.; Perlaky, L.; Lau, C. C. Optimising the use of TRIzol-extracted proteins in surface enhanced laser desorption/ ionization (SELDI) analysis. *Proteome Sci.* **2006**, *4*, 3.
- (23) Lee, F. W.-F.; Lo, S. C.-L. The use of Trizol reagent (phenol/guanidine isothiocyanate) for producing high quality two-dimensional gel electrophoreograms (2-DE) of dinoflagellates. *J. Microbiol. Methods* **2008**, *73*, 26–32.
- (24) Rundlett, K. L.; Armstrong, D. W. Mechanism of signal suppression by anionic surfactants in capillary electrophoresis-electrospray ionization mass spectrometry. *Anal. Chem.* **1996**, *68*, 3493–3497.
- (25) Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* **2010**, *9*, 2863–2870.
- (26) Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H. A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J. Proteome Res.* **2010**, *9*, 5422–5437.
- (27) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362.
- (28) Sieuwerts, A. M.; Meijer-van Gelder, M. E.; Timmermans, M.; Trapman, A. M. A. C.; Garcia, R. R.; Arnold, M.; Goedheer, A. J. W.; Portengen, H.; Klijn, J. G. M.; Foekens, J. A. How ADAM-9 and ADAM-11 differentially predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin. Cancer Res.* **2005**, *11*, 7311–7321.
- (29) Van Agthoven, T.; Sieuwerts, A. M.; Meijer-van Gelder, M. E.; Loo, M. P.; Smid, M.; Veldscholte, J.; Sleijfer, S.; Foekens, J. A.; Dorssers, L. C. J. Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J. Clin. Oncol.* **2009**, *27*, 542–549.
- (30) Sieuwerts, A. M.; Usher, P. A.; Meijer-van Gelder, M. E.; Timmermans, M.; Martens, J. W. M.; Brünner, N.; Klijn, J. G. M.; Offenberg, H.; Foekens, J. A. Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer. *Clin. Chem.* **2007**, *53*, 1280–1288.
- (31) Sieuwerts, A. M.; Meijer-van Gelder, M. E.; Timmermans, M.; Trapman, A. M. A. C.; Garcia, R. R.; Arnold, M.; Goedheer, A. J. W.; Portengen, H.; Klijn, J. G. M.; Foekens, J. A. How ADAM-9 and ADAM-11 differentially predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. *Clin. Cancer Res.* **2005**, *11*, 7311–7321.
- (32) Garcia-Murillas, I.; Lambros, M.; Turner, N. C. Determination of HER2 amplification status on tumour DNA by digital PCR. *PLoS One* **2013**, *8*, e83409.
- (33) Hollestelle, A.; Nagel, J. H. A.; Smid, M.; Lam, S.; Elstrodt, F.; Wasielewski, M.; Ng, S. S.; French, P. J.; Peeters, J. K.; Rozendaal, M. J.; et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res. Treat.* **2010**, *121*, 53–64.
- (34) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226.
- (35) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; et al. A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **2010**, *10*, 1150–1159.
- (36) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534–2536.

- (37) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13*, 22–24.
- (38) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **2007**, *6*, 654–661.
- (39) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958–964.
- (40) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74*, 5383–5392.
- (41) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol. Cell. Proteomics* **2011**, *10*, M111.007690.
- (42) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658.
- (43) Keller, A.; Eng, J.; Zhang, N.; Li, X.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **2005**, *1*, 2005.0017.
- (44) Altelaar, A. F. M.; Frese, C. K.; Preisinger, C.; Hennrich, M. L.; Schram, A. W.; Timmers, H. T. M.; Heck, A. J. R. Mohammed, S. Benchmarking stable isotope labeling based quantitative proteomics. *J. Proteomics* **2013**, *88*, 14–26.
- (45) Mertins, P.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Patel, J.; Ong, S.; Jaffe, J. D.; Carr, S. A. iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol. Cell. Proteomics* **2012**, *11*, M111.014423.
- (46) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **2011**, *8*, 937–940.
- (47) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5116–5121.
- (48) Cox, J.; Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **2012**, *13* (Suppl 1), S12.
- (49) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57.
- (50) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13.
- (51) Simon, R.; Lam, A.; Li, M.-C.; Ngan, M.; Menenzes, S.; Zhao, Y. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inf.* **2007**, *3*, 11–17.
- (52) Eisen, M. B.; Spellman, P. T.; Brown, P. O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14863–14868.
- (53) Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **2004**, *20*, 3246–3248.