

# Proteomics Analysis of Conditioned Media from Three Breast Cancer Cell Lines

A MINE FOR BIOMARKERS AND THERAPEUTIC TARGETS<sup>†§</sup>

Vathany Kulasingam<sup>‡§</sup> and Eleftherios P. Diamandis<sup>†§¶||</sup>

A “bottom-up” proteomics approach and a two-dimensional (strong cation exchange followed by reversed-phase) LC-MS/MS strategy on a linear ion trap (LTQ) were utilized to identify and compare expressions of extracellular and membrane-bound proteins in the conditioned media of three breast cell lines (MCF-10A, BT474, and MDA-MB-468). Proteomics analysis of the media identified in excess of 600, 500, and 700 proteins in MCF-10A, BT474, and MDA-MB-468, respectively. We successfully identified the internal control proteins, kallikreins 5, 6, and 10 (ranging in concentration from 2 to 50  $\mu\text{g}/\text{liter}$ ) in MDA-MB-468 conditioned medium as validated by ELISA and confidently identified Her-2/neu in BT474 cells. Subcellular localization was determined based on Genome Ontology terms for all the 1,139 proteins of which 34% were classified as extracellular and membrane-bound. Proteomics analysis of MDA-MB-468 cell lysate demonstrated that only 5% of all identified proteins were extracellular. This confirmed our hypothesis that examining the CM of cell lines, as opposed to the cell lysates, leads to a significant enrichment in secreted proteins. Tissue specificity, functional classifications, and spectral counting were performed. Elafin, a protease inhibitor, identified in the conditioned media of BT474 and MDA-MB-468 and the three kallikreins (KLK5, KLK6, and KLK10) were validated using an immunoassay on various serum and biological samples. Some of the secreted proteins identified have established roles in breast cancer development (cell growth, differentiation, and metastasis) and/or are linked to early onset breast cancer. Our approach to mining for low abundance molecules could identify proteins in various stages of breast cancer development. Many of the identified proteins are potentially useful to investigate as circulating serum breast cancer biomarkers. *Molecular & Cellular Proteomics* 6:1997–2011, 2007.

Breast cancer is a leading cause of death among women with solid tumors in North America (1). It is a disease of the

middle and late ages of life as 75% of breast cancer is diagnosed in women over the age of 50 (2). Although breast cancer is less common at a young age, younger women tend to have a more aggressive form of the disease than older women. The 5-year survival rate is close to 97% when the cancer is confined to the breast (2). However, when breast cancer has metastasized at the time of diagnosis, the 5-year survival rate is  $\sim 23\%$ . Gene expression patterns have been used to classify breast tumors into clinically relevant subgroups (luminal A, luminal B, basal, ERBB2-overexpressing, and normal-like) (3, 4). In general, the luminal subtypes are estrogen receptor (ER)<sup>1</sup>-positive and grow slowly, whereas basal types lack ER and are usually high grade cancers that grow rapidly. Recently the molecular taxonomy has been confirmed by protein expression profiling (5, 6).

Aberrant secretion or shedding of proteins is commonly associated with disease, including cancer. The pathogenic signaling pathways involved during the process of cancer initiation and progression are not confined to the cancer cell itself but are rather extended to the tumor-host interface (7). It is a dynamic environment in which fluctuating information flows between the tumor cells and the normal host tissue. Therefore, it is conceivable that either the tumor itself or its microenvironment could be sources for biomarkers that would ultimately be shed into the serum proteome, allowing for early disease detection (8), monitoring therapeutic efficacy, or understanding the biology of the disease (9). Given that  $\sim 20$ – $25\%$  of all cellular proteins are secreted, it is reasonable to hypothesize that proteins or their fragments originating from cancer cells or their microenvironment may eventually enter the circulation (10).

Accordingly one of the best ways to diagnose cancer early or to predict therapeutic response is to use serum or tissue biomarkers. Carcinoembryonic antigen and carbohydrate antigen 15-3 (CA 15-3) are the most commonly used tumor markers for breast cancer. Their levels in serum are related to tumor size and nodal involvement and are recommended for

From the <sup>‡</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1X5, Canada, <sup>§</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada, and <sup>¶||</sup>Department of Clinical Biochemistry, University Health Network and Toronto Medical Laboratories, Toronto, Ontario M5G 1X5, Canada

Received, December 8, 2006, and in revised form, July 25, 2007  
Published, MCP Papers in Press, July 25, 2007, DOI 10.1074/mcp.M600465-MCP200

<sup>1</sup> The abbreviations used are: ER, estrogen receptor; CA, carbohydrate antigen; CM, conditioned media; 2-D, two-dimensional; ECD, extracellular ligand binding domain; GO, Genome Ontology; KLK, kallikrein; LDH, lactate dehydrogenase; NAF, nipple aspirate fluid; SFM, serum-free media; TIF, tumor interstitial fluid; DMEM, Dulbecco's modified Eagle's medium; IPI, International Protein Index; IPA, Ingenuity Pathways Analysis.

monitoring therapy of advanced breast cancer or recurrence but are not suitable for population screening (due to low diagnostic sensitivity and specificity) (11–13). Currently mammography remains the cornerstone of breast cancer screening despite its disadvantages such as high false positive and negative rates, hazardous exposure, and patient discomfort (14, 15). In addition, for women under the age of 40, mammographic screening yields a poor sensitivity of 33% (16, 17). Recent technological advances in proteomics have opened up new and exciting avenues for the discovery of biomarkers or for characterization of molecules involved in cancer initiation and progression.

A number of different proteomics-based approaches have been utilized to discover and characterize disease-specific molecules. Fluid found within the ductal and lobular system of the breast can be extracted through the nipple using an aspiration device to obtain nipple aspirate fluid (NAF) (18). Non-pregnant and non-lactating women continuously secrete and reabsorb this fluid (19). Because of the complex nature of biological fluids relevant to breast cancer, only a handful of high abundance proteins have been identified in NAF, illustrating the need to find another source to mine for the initial biomarker discovery (20–22).

A number of studies have used a cell culture model system in which the cells were grown in serum-free media to perform proteomics analysis (23–28). The clinical relevance of using a cell culture model to understand biological processes and functions has been examined. Using DNA microarrays, the molecular subtypes of 31 breast cell lines yielded two discriminating clusters corresponding to luminal cell lines and basal/mesenchymal cell lines (29). The basal subtype was further subdivided into Basal A and Basal B; this subdivision was not observed in primary tumors. Also recently, it was found that cell lines display the same heterogeneity in copy number and expression abnormalities as the primary tumors (30).

In this study, we report a shotgun proteomics approach to sample the conditioned media of three human breast cell lines (MCF-10A, BT474, and MDA-MB-468). MCF-10A, a Basal B subtype with intact p53, was derived by spontaneous immortalization of breast epithelial cells from a patient with fibrocystic disease, and it has been used extensively as a normal control in breast cancer studies (31). These cells do not survive when implanted subcutaneously into immunodeficient mice (31). BT474, a luminal subtype obtained from a stage II localized solid tumor, is positive for ER and progesterone receptor (50–60% of all breast cancer cases) (32). This cell line also displays amplification of Her-2/neu or ERBB2 (30% of all breast cancer cases) (33). Her-2/neu is a cell membrane surface-bound tyrosine kinase involved in signal transduction, leading to cell growth and differentiation. Its overexpression is associated with a high risk of relapse and death (33) and is the target of the therapeutic monoclonal antibody Herceptin (34). Finally MDA-MB-468, a Basal A-like subtype obtained from a pleural effusion of a stage IV patient (35), is ER- and progesterone

receptor-negative (15–25% of breast cancer) and phosphatase and tensin homolog-negative (30% of breast cancer) (36, 37).

These cell lines were cultured in serum-free media (SFM) to ensure that the collected conditioned media (CM) contain no other extraneous proteins except for the secreted or shed proteins from the cancer cells. By collecting and concentrating large volumes of CM produced from cell lines representing seminormal (MCF-10A), non-invasive (BT474), and metastatic origins (MDA-MB-468), the secreted and shed proteins would accumulate in the CM, thereby facilitating their identification through MS. Our comparative proteomics analysis of the CM of MCF-10A, BT474, and MDA-MB-468 identified over 600, 500, and 700 proteins, respectively. A large portion of the proteins was present in all three cell lines; however, a significant portion contained proteins that were unique to each of the lines. Among these were our internal control proteins, human kallikreins 5, 6, and 10, that were identified by MS and ELISA in MDA-MB-468 cells at a concentration ranging from 2 to 50  $\mu\text{g/liter}$ . Members of the human kallikrein family (KLKs) have been implicated in the process of carcinogenesis, and the application of kallikreins as biomarkers for diagnosis and prognosis is currently being investigated. Kallikreins are secreted enzymes that encode for trypsin-like or chymotrypsin-like serine proteases (38). Prostate-specific antigen (KLK3), belonging to the family of human tissue kallikreins, and human kallikrein 2 (KLK2) currently have important clinical applications as prostate cancer biomarkers (39). In addition to the control proteins, various proteases, receptors, protease inhibitors, cytokines, and growth factors were identified. Cellular localization, biological function, and Unigene analyses were performed for the shortened list of candidates consisting of extracellular, membrane, and unclassified proteins. A significant degree of overlap was observed among the proteins identified in this study using a cell culture model and other studies using relevant biological fluids such as NAF and tumor interstitial fluid (TIF). The expression of four candidate molecules was examined in biological fluids, tissues, serum, and breast cytosols. Finally spectral counting analysis revealed promising molecules to investigate further for both understanding the disease and as potential biomarkers for breast cancer.

#### MATERIALS AND METHODS

**Cell Lines**—The breast epithelial cell line MCF-10A and the breast cancer cell lines BT-474 and MDA-MB-468 were purchased from the American Type Culture Collection (ATCC), Manassas, VA. MCF-10A was maintained in Dulbecco's modified Eagle's medium and F-12 medium (DMEM/F-12) supplemented with 8% fetal bovine serum, epidermal growth factor (20 ng/ml), hydrocortisone (0.5  $\mu\text{g/ml}$ ), cholera toxin (100 ng/ml), and insulin (10  $\mu\text{g/ml}$ ). BT-474 and MDA-MB-468 were maintained in phenol red-free RPMI 1640 culture medium (Invitrogen) supplemented with 8% fetal bovine serum. All cells were cultured in a humidified incubator at 37 °C and 5%  $\text{CO}_2$  in tissue culture T 75-cm<sup>2</sup> flasks.

**Cell Culture**—Approximately  $30 \times 10^6$  cells were seeded individually into six 175-cm<sup>2</sup> tissue culture flasks per cell line. After 2 days, the

RPMI 1640 or DMEM/F-12 media were discarded, and the cells were rinsed twice with  $1 \times$  PBS. Following this, 30 ml of chemically defined Chinese hamster ovary serum-free medium (Invitrogen) supplemented with glutamine (8 mM) (Invitrogen) were added, and the flasks were incubated for an additional 24 h. The CM were collected and spun down to remove cellular debris. CM were then frozen at  $-80^{\circ}\text{C}$  until further use. A 1-ml aliquot was taken at the time of harvest to measure for total protein (Bradford assay), lactate dehydrogenase (LDH), KLK5, KLK6, and KLK10 via ELISA. The adhered cells were trypsinized and counted using a hemocytometer. This procedure was repeated several times for reproducibility. In addition, 30 ml of the culture media (RPMI 1640 and DMEM/F-12) were subjected to the same conditions as above with no cells added and used for comparison. For the MDA-MB-468 cell lysate experiment, at the end of 24 h in SFM, the adhered cells were lysed using a French press (Thermo Electron) in which the cells are sheared by forcing them through a narrow space. Total protein was measured, and 400  $\mu\text{g}$  of protein from the lysate were added to 60 ml of chemically defined Chinese hamster ovary medium and processed in the same manner as the CM. The cell lysate experiment was performed in duplicate.

**Sample Preparation**—Two 30-ml CM aliquots were combined (60 ml) for each cell line, creating three biological replicates per cell line, and dialyzed using a 3.5-kDa molecular mass cutoff membrane. The CM were dialyzed in 5 liters of 1 mM ammonium bicarbonate solution overnight at  $4^{\circ}\text{C}$  with two buffer changes. The dialyzed CM were poured equally into two 50-ml conical tubes. The CM were frozen and lyophilized to dryness. The lyophilized sample was denatured using 8 M urea and reduced with DTT (final concentration, 13 mM; Sigma). Following reduction, the sample was alkylated with 500 mM iodoacetamide (Sigma) and desalted using a NAP5 column (GE Healthcare). The sample was lyophilized and trypsin (Promega)-digested (1:50, trypsin:protein concentration) overnight in a  $37^{\circ}\text{C}$  waterbath. Following this, the peptides were lyophilized to dryness.

**Strong Cation Exchange Liquid Chromatography**—The trypsin-digested dry sample was resuspended in 120  $\mu\text{l}$  of mobile phase A (0.26 M formic acid in 10% acetonitrile). The sample was directly loaded onto a PolySULFOETHYL A<sup>TM</sup> column (The Nest Group, Inc.) containing a hydrophilic, anionic polymer (poly-2-sulfoethyl aspartamide). A 200-Å pore size column with a diameter of 5  $\mu\text{m}$  was used. A 1-h fractionation procedure was performed using an HPLC system (Agilent 1100). A linear gradient of 0.26 M formic acid in 10% acetonitrile as the running buffer and 1 M ammonium formate added as the elution buffer was used. The eluent was monitored at a wavelength of 280 nm. Forty fractions, 200  $\mu\text{l}$  each, were collected every minute after the start of the elution gradient. These 40 fractions were pooled into eight combined fractions (each pool consisting of five fractions) and lyophilized to  $\sim 200 \mu\text{l}$ .

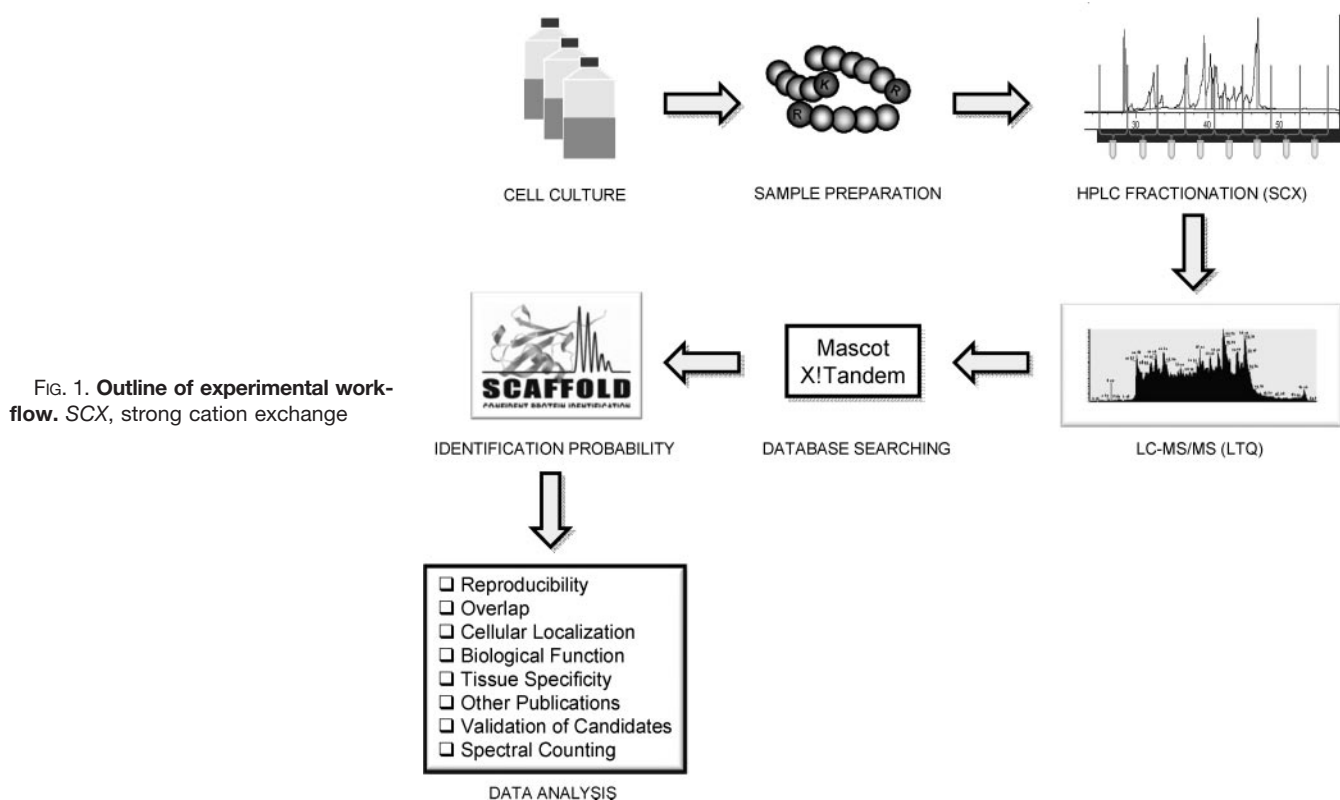
**Mass Spectrometry (LC-MS/MS)**—The eight pooled fractions per replicate per cell line were loaded into a ZipTip<sub>C18</sub> pipette tip (Millipore; catalogue number ZTC18S096) and eluted in 4  $\mu\text{l}$  of 68% ACN made up of Buffer A (95% water, 0.1% formic acid, 5% ACN, 0.02% TFA) and Buffer B (90% ACN, 0.1% formic acid, 10% water, 0.02% TFA). 80  $\mu\text{l}$  of Buffer A were added, and 40  $\mu\text{l}$  were injected onto a 2-cm C<sub>18</sub> trap column (inner diameter, 200  $\mu\text{m}$ ). The peptides were eluted from the trap column onto a resolving 5-cm analytical C<sub>18</sub> column (inner diameter, 75  $\mu\text{m}$ ) with an 8- $\mu\text{m}$  tip (New Objective). The LC setup was coupled on line to a 2-D linear ion trap (LTQ, Thermo Inc.) mass spectrometer using a nano-ESI source in data-dependent mode. Each pooled fraction was run on a 120-min gradient. The eluted peptides were subjected to MS/MS. DTAs were created using the Mascot Daemon (version 2.16) and extract\_msn. The parameters for DTA creation were: minimum mass, 300 Da; maximum mass, 4000 Da; automatic precursor charge selection; minimum peaks, 10 per MS/MS scan for acquisition; and minimum scans per group, 1.

**Data Analysis**—The resulting raw mass spectra from each pooled fraction were analyzed using Mascot (Matrix Science, London, UK; version 2.1.03) and X!Tandem (Global Proteome Machine Manager, version 2.0.0.4) search engines on the non-redundant International Protein Index (IPI) human database version 3.16 ( $>62,000$  entries). Up to one missed cleavage was allowed, and searches were performed with fixed carbamidomethylation of cysteines and variable oxidation of methionine residues. A fragment tolerance of 0.4 Da and a parent tolerance of 3.0 Da were used for both search engines with trypsin as the digestion enzyme. This operation resulted in eight DAT files (Mascot) and eight XML files (X!Tandem) for each replicate sample per cell line. Scaffold (version Scaffold-01\_05\_19, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the PeptideProphet algorithm (40). Protein identifications were accepted if they could be established at greater than 80.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the ProteinProphet algorithm (41). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The DAT and XML files for each cell line plus their respective negative control files (RPMI 1640 or DMEM culture media only) were inputted into Scaffold to cross-validate Mascot and X!Tandem data files. Each replicate sample was designated as one biological sample containing both DAT and XML files in Scaffold and searched with MudPIT (multidimensional protein identification technology) option clicked. The results obtained from Scaffold were processed using an in-house-developed program that generated the protein overlaps between samples. Each protein identification was assigned a cellular localization based on information available from Swiss-Prot, Genome Ontology (GO), Human Protein Reference Database, and other publicly available databases. To calculate the false positive error rate, the individual fractions were analyzed using the "sequence-reversed" decoy IPI human version 3.16 database by Mascot and X!Tandem, and data analysis was performed as mentioned above.

**Ingenuity Pathways Analysis (IPA)**—Extracellular, membrane-bound, and unclassified proteins were evaluated by Ingenuity Pathways Analysis software to identify global functions of the proteins. This software uses a knowledge base derived from the literature to relate gene products to each other based on their interaction and function. The list of proteins and their corresponding IPI human identification numbers were uploaded as an Excel spreadsheet file onto the Ingenuity software (Ingenuity Systems). Ingenuity then used these proteins and their identifiers to navigate the curated literature database. The biological functions assigned to each network were ranked according to the significance of that biological function to the network. A Fischer's exact test was used to calculate a  $p$  value. A detailed description of IPA can be found on the Ingenuity Systems website.

**Spectral Counting**—Using the number of total spectra output from Scaffold, we identified the differentially expressed proteins using spectral counting. Common peptides among proteins were grouped, proteins containing more than 10% of their total spectra from negative control samples were removed, and one Excel file containing total proteins identified and their presence (defined by spectral counts) in the three cell lines was generated. A normalization criterion was applied to normalize the spectral counts so that the values of the total spectral counts per sample were similar. An average of the spectral counts was generated for each cell line (based on the triplicate samples). The sum of the three variances for the cell lines, an indicator of the variance within each cell line, was calculated. The variance of the average spectral counts for each cell line revealed the variability between the cell lines. Analysis of variance (Fisher test) was performed to obtain the ratio of the "between sample variance" to the





“within sample variance.” Apparent -fold changes were calculated when possible.

**Total Protein Assay and LDH Measurements**—Total protein was quantitated in the CM using a Coomassie (Bradford) protein assay reagent (Pierce). All samples were loaded in triplicates on a microtiter plate, and protein concentrations were estimated by reference to absorbances obtained for a series of BSA standard protein dilutions. LDH, an intracellular enzyme that if found in the CM is an indicator of cell death, was measured using an enzymatic assay based on lactate to pyruvate conversion and parallel production of NADH from NAD. The production of NADH was measured by spectrophotometry at 340 nm using an automated method (Roche Applied Science Modular system).

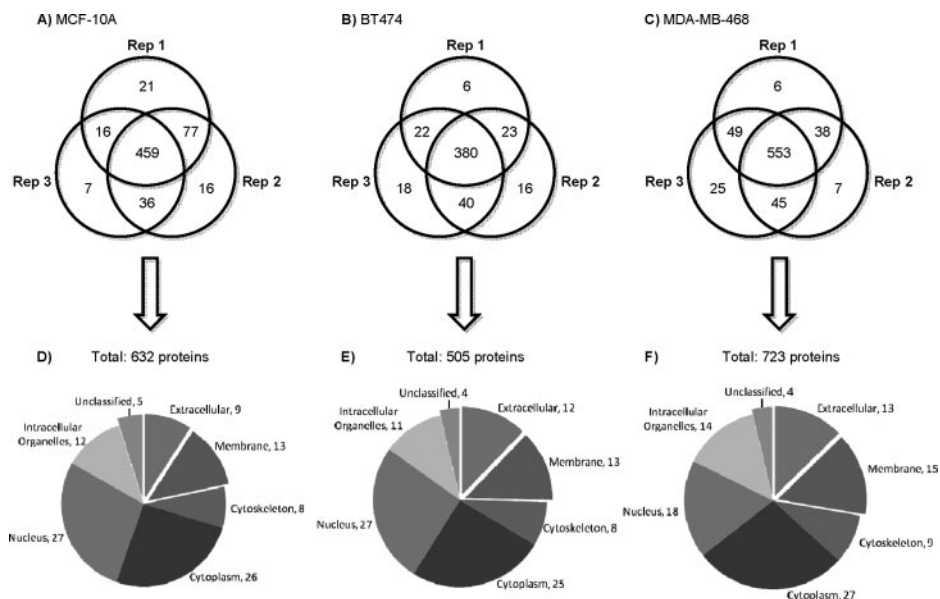
**Quantification of Elafin and KLK5, -6, and -10 by ELISA**—Elafin sandwich ELISA kit, purchased from Hycult biotechnology, was used to measure levels of human elafin in serum, pooled biological samples, and pooled tissue lysates. The assay was performed according to the manufacturer’s instructions. The concentration of KLK5, -6, and -10 was quantified with KLK5, -6, or -10-specific non-competitive immunoassays developed in our laboratory (42–44). For more details, see the cited literature.

**Biological Samples Examined**—The following 10 biological fluids were examined: amniotic fluid, ascites, breast cyst fluid, cerebral spinal fluid, follicular fluid, milk, NAF, urine, saliva, and seminal plasma. Ten samples were combined for each fluid to generate a pooled sample. The following 27 human tissues were examined (the number indicates the number of samples pooled): adrenal, 5; aorta, 3; bladder, 3; bone, 4; colon, 5; esophagus, 4; heart, 5; kidney, 5; liver, 5; lung, 4; lymph node, 3; muscles, 3; pancreas, 5; skin, 4; small intestine, 4; spinal cord, 3; spleen, 5; stomach, 4; thyroid, 3; trachea, 4; ureter, 4; prostate, 4; testis, 4; breast, 4; fallopian tube, 4; uterus, 4; and ovary, 2.

## RESULTS

**Optimization of Cell Culture**—MCF-10A, BT474, and MDA-MB-468 cells were grown in SFM to ensure that the conditioned media contained no other exogenous proteins. Seeding density, incubation time in SFM, volume of media used, type of SFM, and type of tissue culture flasks were all variables that were explored thoroughly to select the most optimal conditions for growth. Refer to “Materials and Methods” for details on the optimal conditions selected for the cell lines. Approximately 20, 33, and 40 × 10<sup>6</sup> cells were found to be attached and alive at the end of the experiment with 25, 10, and 15 μg of total protein/ml in MCF-10A, BT474, and MDA-MB-468, respectively (Supplemental Fig. 1, A and B). Furthermore to minimize cell death and maximize secreted protein concentration in the CM, LDH levels, which represent the amount of cell death occurring in cell culture, were also measured (Supplemental Fig. 1C). Known amounts of all three types of cells were lysed individually, and their corresponding LDH levels were graphed to create an LDH standard curve (Supplemental Fig. 1D). Using the LDH standard graph, the value of LDH found in the 24-h CM, and the total number of cells alive at the end of harvest, it was estimated that ~6–7% cell death was occurring in the CM of the cells. Finally to demonstrate the accumulation of extracellular proteins in our optimized cell culture model system, our internal control proteins KLK5, KLK6, and KLK10 were quantified in MDA-MB-

**FIG. 2. Number of proteins identified in CM by LC-MS/MS for the three cell lines and their cellular localization.** A, MCF-10A; B, BT474; C, MDA-MB-468. Three independent replicates (Rep 1–3) were generated and processed in the same manner for each cell line. Good reproducibility is shown as 73–76% overlap is observed among the cell lines. D, MCF-10A; E, BT474; F, MDA-MB-468. A non-redundant list of proteins was created after Mascot and XITandem searching, and each protein was classified by its cellular location.



468 using ELISAs (Supplemental Fig. 1E). The remaining two cell lines do not secrete any kallikreins unless they are hormonally stimulated. Kallikrein 5 was the most abundant kallikrein expressed in this cell line (50  $\mu\text{g}/\text{liter}$ ) followed by KLK10 ( $\sim 3.5$   $\mu\text{g}/\text{liter}$ ) and KLK6 ( $\sim 2$   $\mu\text{g}/\text{liter}$ ).

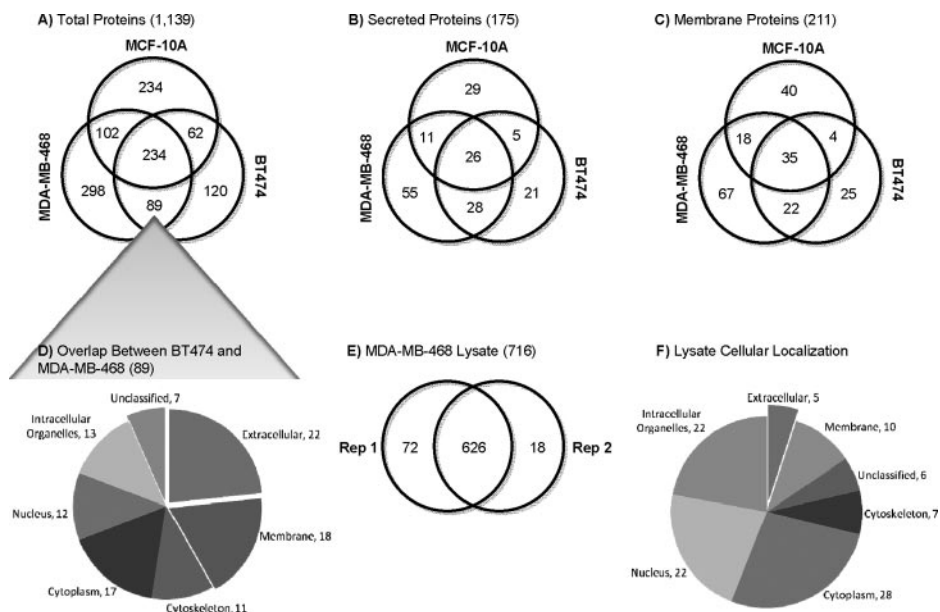
**Identification of Proteins by Mass Spectrometry**—The workflow and experimental design performed in this study are shown in Fig. 1. Approximately 35–40 proteins were identified in the negative control samples, which were made up of the culture media only (Supplemental Table 1A). Many of the proteins in this list originated from fetal bovine serum used to initially culture the cells. These proteins were deleted from the list of total proteins identified in the CM and were not considered further. In MCF-10A, we identified 632 proteins (Fig. 2A). Of these, 459 were identified in all three replicates, yielding a protein identification reproducibility of 73%. Furthermore a total of 505 proteins were identified in BT474 (Fig. 2B). For this cell line, 380 proteins were common to all three replicates (75% reproducibility). Finally 723 proteins were identified in MDA-MB-468 (Fig. 2C) of which 553 were identified in all three replicates (76% reproducibility). In general, using the workflow presented here, we typically achieved technical reproducibility (same sample injected twice into the LTQ) of  $\sim 90\%$  (data not shown). The total number of proteins identified per number of unique peptides per cell line is shown in Table I. Many of the proteins identified contained two or more unique peptide hits. Supplemental Table 2 contains detailed information on all of the proteins identified for each of the cell lines, including number of unique peptides identified per protein, peptide sequences, precursor ion mass, and charge states. In addition, eight, three, and five proteins were identified in MCF-10A, BT474, and MDA-MB-468 cells, respectively, using a non-sense database, yielding a false positive rate of  $\sim 1\%$  (see Supplemental Table 1B) (45).

**TABLE I**  
Total number of proteins identified per number of unique peptides

No. of peptides identified	MCF-10A	BT474	MDA-MB-468
1	120	124	155
2	172	137	211 (KLK6)
3	95	77	111 (KLK5, -10)
4	63	41 (Her-2/neu)	67
$\geq 5$	182	126	179
Total	632	505	723

**Identification of Internal Control Proteins in CM by MS**—One of the advantages of our approach to biomarker discovery was the presence of endogenous internal control proteins in the CM of MDA-MB-468. We identified KLK5, -6, and -10 in the CM of MDA-MB-468 in all three replicates for this cell line by MS. Supplemental Fig. 2 illustrates the sequences of KLK5, -6, and -10 and the peptides that were successfully identified by MS. Three unique peptides were identified for KLK5, two unique peptides were identified for KLK6, and three unique peptides were identified for KLK10 (Supplemental Fig. 2, A–C). Furthermore the BT474 cell line is well characterized to exhibit amplification of Her-2/neu (also known as ERBB2). The Her-2/neu protein consists of a cysteine-rich extracellular ligand binding domain (ECD), a short transmembrane domain, and a cytoplasmic protein-tyrosine kinase domain (46). ECD/Her-2 can be released by proteolytic cleavage from the full-length Her-2 receptor and detected in serum. In the CM of BT474, we successfully identified receptor tyrosine-protein kinase ERBB2 precursor with four unique peptides (Supplemental Fig. 2D). All of the peptides identified fall into the ECD portion (amino acids 23–652) of this 1255-amino acid protein. However, we did not identify CA 15-3 in the

**FIG. 3. Overlap of proteins in CM and proteome of MDA-MB-468 cell lysate by LC-MS/MS.** A, B, and C, overlap of proteins (total number in parentheses) between the three cell lines. D, pie chart showing the subcellular localization of 89 proteins uniquely identified in BT474 and MDA-MB-468, two cancer cell lines (not including MCF-10A proteins). 40% of the proteins are classified as extracellular and membrane-bound. E, two independent replicates (Rep 1 and Rep 2) for the lysate were generated and processed in the same manner. Good reproducibility is shown (87% overlap). F, GO cellular localization for the 716 proteins identified in the lysate. 5% are classified as being extracellular.



conditioned media of the three cell lines by mass spectrometry. To validate this observation, using a commercially available ELISA kit (Elecsys CA 15-3 Immunoassay; Roche Diagnostics GmbH), we measured the levels of this glycoprotein in the same CM of the cell lines. Very little (<3 units/ml) to no CA 15-3 was measured by ELISA in the CM (Supplemental Fig. 3).

**Cellular Localization of Identified Proteins**—Fig. 2, D–F, display the cellular distribution of proteins identified in the conditioned media of MCF-10A (D), BT474 (E), and MDA-MB-468 (F). 22% of the proteins identified in MCF-10A CM were classified as being extracellular and membrane-bound. For BT474 and MDA-MB-468, the percentages were 25 and 28%, respectively. A large portion of proteins identified were classified as intracellular (>50%), whereas 4–5% remained unclassified.

**Overlap of Proteins between the Three Cell Lines**—The proteins identified among the three cell lines were analyzed for overlapping members (Fig. 3). A significant portion (234 or 20%) of the 1,139 proteins was identified in all three cell lines (Fig. 3A). Fig. 3, B and C, show the overlap among the 175 extracellular proteins and the 211 membrane-bound proteins, respectively. Combined together, extracellular and membrane proteins accounted for 34% of all proteins identified. MDA-MB-468 displayed the greatest number of extracellular and membrane proteins, presumably illustrating that cancer cells secrete and/or express an increased amount of these proteins. In accordance with this postulation, cellular localization analysis of the overlap between BT474 and MDA-MB-468 yielded the greatest percentage (40%) of secreted and membrane proteins (Fig. 3D).

**Cell Lysate Proteome**—One of the major challenges in the analysis of secreted proteins is distinguishing between proteins that are targeted to the extracellular space *versus* those that arise as low level contaminants due to cell death during routine cell culture. To address this, we performed a cell

lysate proteome experiment to examine whether our approach was enriching for secreted proteins. A total of 716 proteins were identified in MDA-MB-468 cell lysate after removal of the negative control proteins (culture media only). Supplemental Table 3A shows all proteins identified. 87% protein identification reproducibility was observed among the two replicates (Fig. 3E). 5% of the total proteome was classified as extracellular/secreted (Fig. 3F). In the CM of MDA-MB-468, 13% of the total proteins identified were classified as being secreted (Fig. 2F). The internal control secreted proteins, kallikreins 5, 6, and 10, were not identified in the cell lysate. Of the secreted proteins identified in the cell lysate, 30 were also identified in the CM for this cell line, whereas 19 proteins were unique to the lysate. Supplemental Table 3B shows the secreted proteins that were found in both the MDA-MB-468 lysate and CM as well as the extracellular proteins found only in the lysate and CM, respectively.

**Tissue-specific Expression and Biological Function Analysis**—Unigene analysis of 422 extracellular, membrane-bound, and unclassified proteins (“shortened list of candidates”), using the Expressed Sequence Tag ProfileViewer, was utilized to identify genes that were relatively breast-specific. Of these genes, five were found to be relatively specific to normal and/or cancerous breast (SCGB1D2, SBEM, TFF1, DCD, and CALML5). Literature mining on these proteins showed that one of the proteins had previously been evaluated in serum of breast cancer patients (trefoil factor 1). As well, using the IPA software, we classified the shortened list of proteins by biological function. The top 15 functions are displayed in Fig. 4A. The top functions were cellular movement followed by cell-to-cell signaling and interaction. Finally these proteins were searched with the Human Plasma Proteome database to decipher whether they have been identified in plasma. 104 of 422 proteins were identified in human plasma.



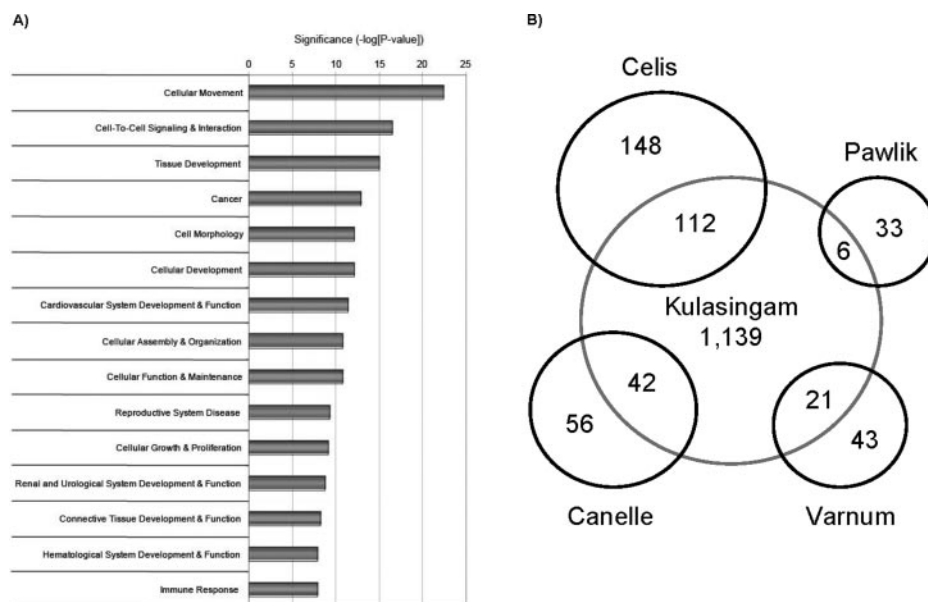


FIG. 4. **Biological function analyses and overlap among other publications.** A, the top 15 functions for the 422 extracellular, membrane-bound, and unclassified proteins are shown as determined by IPA. The y axis shows the negative log of  $p$  value. B, the overlap of proteins identified in the current conditioned media cell culture study (Kulasingham) and four other proteomics studies (Celis, Ref. 49; Pawlik, Ref. 47; Canelle, Ref. 25; Varnum, Ref. 48).

**Comparison of Proteins Identified from CM with Other Publications**—Using the ICAT technology, Pawlik *et al.* (47) quantified tumor-specific proteins in NAF. In total, of the 39 proteins that were differentially expressed in tumor-bearing versus disease-free breasts, six were also found among the 1,139 proteins in our CM. In addition, Varnum *et al.* (48) identified 64 proteins in NAF using MS of which 15 were previously reported to be altered in patients with breast cancer. From the 64 proteins, 21 were also found in our proteomics study. More importantly, of the 15 proteins previously reported to be altered in serum or tumor from women with breast cancer, 10 were found in our CM proteome. A proteomics study involving MCF-7 targeted membrane-associated breast cancer proteins as potential biomarkers (25). Using 2-D electrophoresis and MS analysis of the protein spots, Canelle *et al.* (25) identified 98 proteins. Among the 98 proteins, 42 of them were also found in our study. In addition, Celis *et al.* (49) mined another source of biomarkers through the investigation of TIF that perfuses the breast tumor microenvironment. Given the importance of the tumor-host interface and the increasing appreciation of the role that the microenvironment plays in cancer initiation and progression, we compared our list of proteins identified through a cell culture model system with the proteins excreted by the cells that are within their native cancer microenvironment. The authors identified ~260 proteins using 2-D gel electrophoresis, immunoblotting, and mass spectrometry of which 112 were also identified in our study (43%). Fig. 4B summarizes the overlaps observed among the other publications and the data presented here. Supplemental Table 4 contains all of the proteins identified by us and the four previous studies. Finally the lung, along with the bone, is one of the most frequent sites of breast cancer metastasis. A set of 54 genes that mediate breast cancer metastasis to the lungs have been identified (50).

Given that MDA-MB-468 cells were collected from pleural effusion, we compared the list of proteins identified from MDA-MB-468 conditioned media with the 54 candidate lung metastasis genes. Seven genes were found in the CM of our cell line (*KYNU*, *TNC*, *ROBO1*, *FSCN1*, *MAN1A1*, *LTBP1*, and *GSM*). Interestingly none of the genes that overlapped between the lung and bone metastasis signatures were identified in MDA-MB-468.

**Spectral Counting and Differential Expression of Proteins Identified**—An alternative way to decipher protein abundance is to perform multidimensional scaling for all nine experiments (each cell line in triplicate) using spectral counts. Refer to “Materials and Methods” for details on how the analysis was conducted. The Venn diagram in Supplemental Fig. 4 displays the overlaps of proteins among the cell lines based on spectral count analysis and their cellular localization. The top ~100 extracellular and membrane-bound proteins obtained from spectral counting analysis are shown in Table II. The variability within the replicates (within variance) and between the three cell lines (between variance) are highlighted along with the F ratio. Apparent -fold changes were calculated where possible. A numerical value is indicated in places where both cell lines/conditions being examined contain a normalized spectral count greater than zero. In the event of a comparison where one of the conditions/cell lines had a spectral count of zero, an expression is given (e.g. BT >> MCF; indicating that the spectral count for BT474 was greater than MCF10A). Cells that are gray indicate a negative -fold change, whereas cells that are white indicate a positive (numerical value indicated) or no -fold change (cells are blank). In addition, in the first column displaying the -fold change between BT474/MCF-10A, cells in white highlight the proteins that have a higher spectral count in BT474 compared with MCF-10A, whereas cells in black highlight proteins that have a lower spectral

TABLE II  
Top 100 differentially expressed extracellular and membrane proteins

Apparent -fold changes obtained from spectral counting were calculated where possible. A numerical value is indicated in places where both cell lines/conditions being examined contain a normalized spectral count greater than zero. An expression value is given (e.g. BT ≥ MCF, indicating that the spectral count for BT474 was greater than MCF10A) when one of the conditions exhibited a spectral count of zero. Cells that are gray indicate a negative -fold change, whereas cells that are white indicate a positive (numerical value indicated) or no -fold change (cells are blank). DBI, diazepam-binding inhibitor; LISCH, lipolysis-stimulated lipoprotein receptor.

Protein name	Cellular Localization	Within variance	Between variance	Fischer Test	FOLD CHANGE	FOLD CHANGE	FOLD CHANGE
					BT474/MCF10A	MDA468/MCF10A	MDA468/BT474
Ly6/PLAUR domain-containing protein 3 precursor	membrane	0	16	77	BT>>MCF		BT>>MDA
Thrombospondin-1 precursor	extracellular	34	1917	57	4.7	116.6	24.8
4F2 cell-surface antigen heavy chain	membrane	1	50	47		MDA>>MCF	BT<<MDA
Tumor-associated calcium signal transducer 2 precursor	cytoplasm, membrane	1	27	39		MDA>>MCF	BT<<MDA
Receptor tyrosine-protein kinase erbB-2 precursor	membrane	2	81	37	BT>>MCF		BT>>MDA
Small breast epithelial mucin precursor	extracellular	6	193	33	BT>>MCF		BT>>MDA
Cadherin-13 precursor	membrane	2	47	29	BT<<MCF	MDA<<MCF	
Insulin-like growth factor-binding protein 5 precursor	extracellular	422	10626	25	BT>>MCF	MDA>>MCF	-59.5
Splice Isoform 1 of Tubulointerstitial nephritis antigen-like precursor	extracellular	2	55	24		MDA>>MCF	BT<<MDA
Splice Isoform Long of Laminin gamma-2 chain precursor	extracellular	583	12683	22	BT<<MCF	MDA<<MCF	
Splice Isoform C of Fibulin-1 precursor	extracellular	14	266	19		MDA>>MCF	BT<<MDA
Laminin beta-3 chain precursor	extracellular	22	411	18	BT<<MCF	MDA<<MCF	
Insulin-like growth factor-binding protein 6 precursor	extracellular	4	75	18	BT<<MCF	MDA<<MCF	
Insulin-like growth factor-binding protein 2 precursor	extracellular	58	983	17	BT>>MCF	MDA>>MCF	-2.6
Splice Isoform APP770 of Amyloid beta A4 protein precursor (Fragment)	extracellular, membrane	368	6094	17	5.4	13.5	2.5
DBI protein	mitochondrion, extracellular, cytoplasm	12	176	14	1.5	-2.6	-3.8
Galectin-3-binding protein precursor	extracellular, membrane	157	2138	14	1.1	4.3	4
Cystatin C precursor	extracellular	95	1257	13	3.4	MDA<<MCF	BT>>MDA
Midkine precursor	extracellular	60	786	13	BT>>MCF	MDA>>MCF	-8.8
Insulin-like growth factor-binding protein 3 precursor	extracellular	48	630	13		MDA>>MCF	BT<<MDA
laminin alpha 3 subunit isoform 1	extracellular	73	879	12	BT<<MCF	MDA<<MCF	
Hippocalcin-like protein 1	cytoplasm, membrane	1	8	12	BT<<MCF	MDA<<MCF	
Insulin-like growth factor-binding protein 7 precursor	extracellular	14	150	11	BT<<MCF	MDA<<MCF	
Splice Isoform Long of Sodium/potassium-transporting ATPase alpha-1 chain precursor	membrane	4	39	10		MDA>>MCF	BT<<MDA
Transforming growth factor-beta-induced protein ig-h3 precursor	extracellular	49	487	10	BT<<MCF	-16.7	BT<<MDA

count in BT474 compared with MCF-10A. A similar color coding scheme applies to the other two columns comparing the different cell lines/conditions within each column. Known breast cancer biomarkers such as Her-2/neu are among the top five proteins identified by this unbiased method of analysis. Furthermore 23 proteins previously associated with cancer (as determined by Ingenuity Biomarkers Comparison Analysis software) were found among the top ~100 extracellular and membrane proteins including epidermal growth factor receptor and various insulin-like growth factor-binding proteins (IGFBP-2, -3, -5, and -6). Supplemental Table 5 contains all of the 1,062 proteins on which this analysis was performed. In addition, Supplemental Table 6 contains the overlaps of the proteins among the cell lines.

**Validation of Elafin and Kallikreins**—Proteases and their inhibitors play an important role in cancer metastasis and angiogenesis (51). Elafin, a secreted epithelial proteinase inhibitor, also referred to as skin-derived antileukoprotein-

ase or elastase-specific inhibitor, belongs to the Trappin gene family (52). Elafin has been found to be expressed in normal mammary epithelial cells but is down-regulated in most breast tumor cell lines (53). We identified elafin in BT474 and MDA-MB-468 with the latter having more unique peptides for the protein. Using a commercially available sandwich immunoassay for elafin, we examined a number of different biological samples for its expression. Serum from women with different levels of CA 15-3 was measured. Typically women with CA 15-3 levels of <30 units/ml are considered in the "normal" range (13). Examining sera from women with <30 units/ml, >30 but <100 units/ml, and >100 units/ml CA 15-3 for elafin showed no significant differences among the groups (Fig. 5A). In addition, the median values for elafin in normal and tumor breast cytosols showed that tumor cytosols had higher levels of this inhibitor than normal (Fig. 5B). However, the range in normal breast cytosols among the 15 samples examined varied



TABLE II—continued

Splice Isoform 1 of Epidermal growth factor receptor precursor	extracellular, nucleus, cytoplasm, endosome, membrane	220	2117	10		MDA>>MCF	BT<<MDA
Ribonuclease 6 precursor	extracellular	6	54	9	BT>>MCF	MDA>>MCF	-12.3
Receptor-type tyrosine-protein phosphatase kappa precursor	membrane	25	219	9	32.1	13	-2.5
Stanniocalcin-1 precursor	extracellular	28	203	7		MDA>>MCF	BT<<MDA
Alpha-1-antitrypsin precursor	extracellular	15	98	7	BT>>MCF		BT>>MDA
Carboxypeptidase A4 precursor	extracellular	13	88	7		MDA>>MCF	BT<<MDA
Splice Isoform 1 of Soluble calcium-activated nucleotidase 1	endoplasmic reticulum, membrane	56	372	7	BT>>MCF		BT>>MDA
Cargo selection protein TIP47	cytoplasm, endosome, Golgi apparatus, membrane	3	21	7	BT<<MCF	-16.7	BT<<MDA
Splice Isoform Long of Spectrin beta chain, brain 1	cytoskeleton, membrane,	2	12	6		MDA>>MCF	BT<<MDA
Syndecan-4 precursor	membrane	3	18	6	BT>>MCF	MDA>>MCF	-2.4
Transcobalamin-1 precursor	extracellular	87	544	6	BT>>MCF		BT>>MDA
Lutheran blood group glycoprotein precursor	membrane	66	380	6	BT>>MCF	MDA>>MCF	-4.4
Ras GTPase-activating-like protein IQGAP1	cytoskeleton, membrane,	6	34	6	BT<<MCF	15.7	BT<<MDA
Stress-70 protein, mitochondrial precursor	cytoplasm, mitochondrion, membrane	15	80	5	BT<<MCF	-10.6	BT<<MDA
Latent transforming growth factor beta-binding protein, isoform 1L precursor	extracellular	141	744	5	BT<<MCF	-88.4	BT<<MDA
E-cadherin	membrane	167	876	5	BT>>MCF	MDA>>MCF	-17.4
Ezrin	cytoplasm, cytoskeleton, membrane	5	26	5	BT>>MCF	MDA>>MCF	3.2
Splice Isoform 2B of Desmocollin-2 precursor	cytoskeleton, membrane	6	29	5	15.1	1.5	-10.4
G protein-coupled receptor 56 isoform b	membrane	2	8	5		MDA>>MCF	BT<<MDA
Transforming protein RhoA precursor	membrane, cytoskeleton, membrane,	1	5	5		MDA>>MCF	BT<<MDA
Fractalkine precursor	extracellular, membrane	22	95	4	BT>>MCF	MDA>>MCF	-4.4
Splice Isoform 1 of Fibronectin precursor	extracellular	493	2138	4	-17.6	-1.6	11.3
CD44 antigen isoform 4 precursor	membrane	897	3669	4	-1.1	14.7	16.1
Polymeric-immunoglobulin receptor precursor	membrane	16	65	4	BT>>MCF		BT>>MDA
Splice Isoform 1 of Connective tissue growth factor precursor	extracellular, membrane	6	24	4		MDA>>MCF	BT<<MDA

considerably. The levels of elafin in a variety of pooled biological samples showed that milk contained this protein at 18  $\mu\text{g/liter}$  followed by urine at 7  $\mu\text{g/liter}$ , follicular fluid, and seminal plasma (Fig. 5C). These were biological samples from normal individuals and hence do not correlate to diseased phenotype. Finally we checked the levels of elafin in normal pooled tissue lysates. After correction of total protein content in the samples, colon, small intestine, and ureter contained the highest levels. Among the top 10 expressing tissues, breast was ranked as seventh (Fig. 5D). In addition, using the same sample sets, we measured the levels of kallikreins 5, 6, and 10. The serum samples did not show any biomarker potential. Kallikrein 6 levels in breast cytosols demonstrated some differentiation potential between normal and tumor. Interestingly in pooled biological samples, KLK5 was found to be expressed in all of the relevant breast fluids (milk, breast cyst fluid, and NAF) (Supplemental Figs. 5 and 6).

## DISCUSSION

In this study, a shotgun proteomics strategy was utilized to sample the conditioned media of three breast cell lines, MCF-10A, BT474, and MDA-MB-468. By searching with both Mascot and X!Tandem, we successfully identified over 1,100 proteins in the CM of all three cell lines combined, which, to our knowledge, is one of the largest repositories of proteins identified for breast cancer. Studies have shown that by combining results from multiple search engines, a better, more confident protein identification is made because different search engines are based on different algorithms and scoring (54, 55). Two different methods of determining relative abundance were used: protein identification and spectral counts. Although spectral counting as an index of protein abundance is appealing, there are a number of different ways to analyze a dataset such as counting the spectra and adjusting by the length of the protein (normalized spectral abundance factor) (56), counting peptides (not individual spectra) and adjusting for the number of tryptic peptides in the protein (protein

TABLE II—continued

Splice Isoform 1 of Neuronal cell adhesion molecule precursor	membrane	13	52	4	BT>>MCF		BT>>MDA
Clusterin precursor	extracellular	499	1977	4	-1.8	-4.2	-2.4
GARS protein	extracellular, cytoplasm,	5	18	4		MDA>>MCF	BT<<MDA
Splice Isoform 2 of Basigin precursor	membrane	9	34	4	BT<<MCF	-2.6	BT<<MDA
Matrilysin precursor	extracellular	205	786	4		MDA>>MCF	BT<<MDA
Alpha-1-antichymotrypsin precursor	extracellular	159	607	4	BT>>MCF	MDA>>MCF	28.2
Splice Isoform 2 of Retinoic acid receptor responder protein 1	membrane	1209	4407	4	46.2	264.3	5.7
Elafin precursor	extracellular	63	231	4	BT>>MCF	MDA>>MCF	11.8
Complement component 1, Q subcomponent-binding protein, mitochondrial precursor	mitochondrion, membrane,	20	73	4	BT<<MCF	MDA<<MCF	
quiescin Q6 isoform a	membrane, extracellular	119	420	4	9.1	9.5	1
Splice Isoform 1 of Choline transporter-like protein 2	membrane	0	1	3		MDA>>MCF	BT<<MDA
Extracellular superoxide dismutase [Cu-Zn] precursor	extracellular	18	61	3	BT>>MCF	MDA>>MCF	-15.1
Cystatin SN precursor	cytoplasm, extracellular	190	629	3	BT>>MCF		BT>>MDA
Insulin-like growth factor-binding protein 4 precursor	extracellular	199	639	3	BT>>MCF	MDA>>MCF	-4.4
LanC-like protein 2	nucleus, cytoplasm, membrane,	11	35	3		MDA>>MCF	BT<<MDA
28 kDa heat- and acid-stable phosphoprotein	extracellular	17	54	3	-1.2	-8.2	-7
CD166 antigen precursor	membrane	146	449	3	44.4	12.5	-3.6
urokinase plasminogen activator preproprotein	extracellular	37	112	3	BT<<MCF	MDA<<MCF	
CDNA FLJ16791 fis, clone TBAES2004105, weakly similar to ADAM-TS 6 (Fragment)	extracellular	6	16	3	BT>>MCF		BT>>MDA
olfactomedin 4 precursor	extracellular	24	69	3		MDA>>MCF	BT<<MDA
Zinc-alpha-2-glycoprotein precursor	extracellular	2608	7282	3	148.7	255.9	1.7
Moesin	cytoplasm, cytoskeleton, membrane	121	338	3	BT<<MCF	2.7	BT<<MDA
Splice Isoform 1 of Filamin-B	cytoskeleton, membrane	243	670	3	-2.3	-2.2	1.1
Calmodulin	cytoplasm, membrane,	153	417	3	2.1	-1.7	-3.5
LISCH protein isoform 2	membrane	48	128	3	BT>>MCF	MDA>>MCF	1.7
Kallikrein-6 precursor	extracellular, cytoplasm,	2	4	3		MDA>>MCF	BT<<MDA
Splice Isoform 1 of Complement factor B precursor (Fragment)	extracellular	10	27	3		MDA>>MCF	BT<<MDA
Neutrophil gelatinase-associated lipocalin precursor	extracellular, cytoplasm,	367	971	3	BT>>MCF	MDA>>MCF	-1.7
Alcadein beta	membrane	2	5	3	BT>>MCF		BT>>MDA
mesothelin isoform 1 preproprotein	membrane	2	6	3	BT<<MCF	MDA<<MCF	

abundance index) (57), calculating a function of protein abundance index called exponentially modified protein abundance index (58), counting 1 if any spectra matched a peptide and assigning 0 otherwise (SASPECT (significant analysis of peptide counts)), or merely counting the spectra. Currently there is no consensus as to which approach to use. In this study, we used both protein identification and spectral counts to determine protein abundance.

We specifically examined MCF-10A, BT474, and MDA-MB-468 because they represent progression phases to breast cancer. We observed a higher cell death in MCF-10A, which was expected because MCF-10A is considered to be a normal breast epithelial cell line that does not have the advantage of growing uncontrollably as do the cancer cell lines in SFM (Supplemental Fig. 1). It was our aim that examining the proteins that are unique to each of the cell lines might shed light into both the pathways leading to breast cancer development and to the discovery of biomarkers for breast cancer. Toward this aim, using the Ingenuity Pathways Analysis, the shortened list of candidates was filtered by extracting only

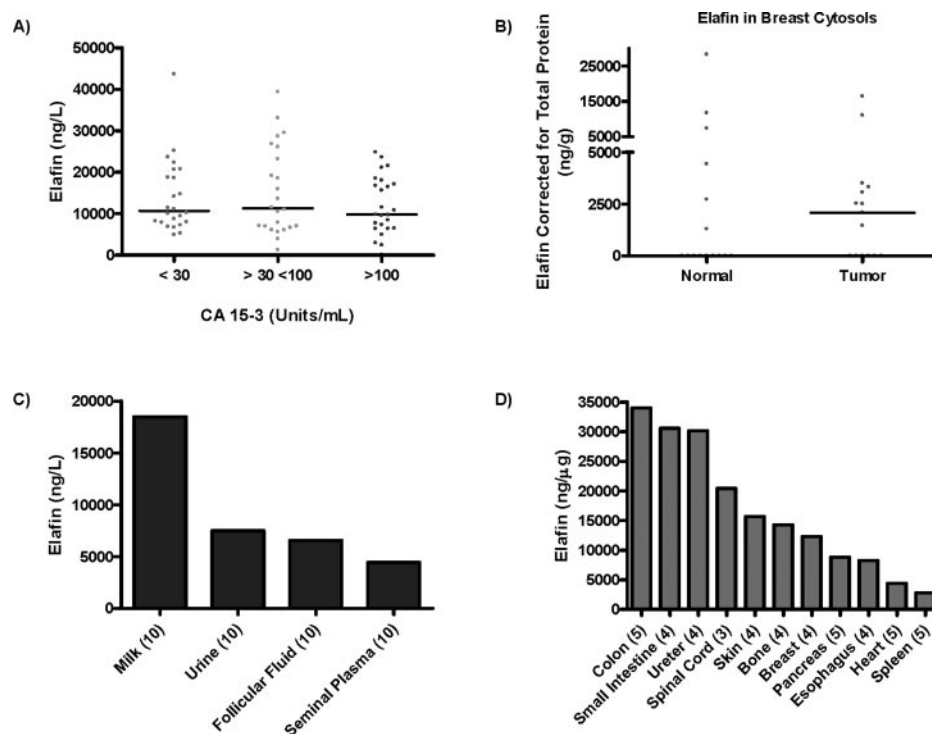
those genes that were identified in literature to be present in human and associated with cancer. Approximately 100 of 422 extracellular and membrane genes were identified to meet this criterion. The top 14 canonical pathways that demonstrated a relationship to the 100 genes filtered as well as known genes linked to breast cancer are highlighted in Fig. 6. Dysregulation of signaling pathways plays an important role in cancer initiation and progression. The proteins identified in this study may be useful for further investigation.

One of the advantages of our high throughput qualitative comparative proteomics analysis was the presence of internal controls in the CM of MDA-MB-468. KLK5 expression, as assessed by quantitative RT-PCR, has been implicated as an independent and unfavorable prognostic marker for breast carcinoma (59). The clinical utility of KLK6 as a breast cancer marker has not been determined, whereas KLK10 levels have been shown to predict response to tamoxifen therapy (60). Based on experience with other currently used biomarkers, the expected concentration of new cancer markers in serum should be in the low  $\mu\text{g/liter}$  range. Through our strategy, we

TABLE II—continued

Splice Isoform Beta of Poliovirus receptor precursor	extracellular, cytoplasm, membrane,	5	14	3		MDA>>MCF	BT<<MDA
Kallikrein-5 precursor	extracellular	9	21	3		MDA>>MCF	BT<<MDA
Membrane-associated progesterone receptor component 1	endoplasmic reticulum, membrane	4	9	2		MDA>>MCF	BT<<MDA
Splice Isoform A of Transforming growth factor beta-2 precursor	extracellular	49	118	2		MDA>>MCF	BT<<MDA
biotinidase precursor	extracellular	2	5	2	BT>>MCF		BT>>MDA
Splice Isoform B of Fibulin-1 precursor	extracellular	2	5	2	BT>>MCF		BT>>MDA
Hypothetical protein LRRC59	membrane	19	43	2	5	1.3	-3.8
CREG1 protein precursor	extracellular	19	43	2	BT<<MCF	2.7	BT<<MDA
ladinin 1	extracellular	17	38	2	BT<<MCF	MDA<<MCF	
slit-like 2	membrane	2	5	2		MDA>>MCF	BT<<MDA
Dystroglycan precursor	extracellular, cytoskeleton, membrane	93	207	2	4.5	5.9	1.3
Leucine-rich alpha-2-glycoprotein precursor	extracellular, membrane,	10	23	2		MDA>>MCF	BT<<MDA
Protein disulfide-isomerase precursor	extracellular, endoplasmic reticulum, membrane,	83	183	2	BT>>MCF	MDA>>MCF	-3.1

FIG. 5. Levels of elafin in biological samples. A, serum from women with known CA 15-3 levels was analyzed for elafin expression by ELISA (median values are shown). B, normal and tumor breast cytosols as well as various pooled biological samples (C) and pooled tissues (D) were assayed. The number of samples used to generate a pooled sample is indicated in parentheses. No expression was observed in other biological samples and tissues examined (see "Materials and Methods" for a complete list of all samples surveyed).



successfully identified all three control proteins by MS, thus supporting the notion that we are enriching for, and surveying deep enough, into the low abundance proteins to mine for other candidate biomarkers for breast cancer. Furthermore the fact that we successfully identified the ECD portion of Her-2/neu in BT474 further supports our hypothesis that candidate biomarkers can be discovered using the CM of breast cancer cell lines. High levels of Her-2 in serum correlate with poor prognosis in patients with breast cancer (61). In 2000, the Food and Drug Administration cleared the serum Her-2/neu test, the first Food and Drug Administration-cleared blood

test for measuring circulating levels of Her-2 in the follow-up and monitoring of patients with metastatic breast cancer. The assay, ADVIA Centaur HER-2/neu Assay (Oncogene Science Biomarker Group), is a sandwich immunoassay that uses two monoclonal antibodies that are specific for unique epitopes on the extracellular domain of the Her-2 oncoprotein. The assay has a reference limit of 15  $\mu\text{g/liter}$  with a sensitivity of 0.1  $\mu\text{g/liter}$  (62).

Particular emphasis was placed on extracellular, membrane-bound, and unclassified proteins because these proteins have the highest chance of being found in the circulation



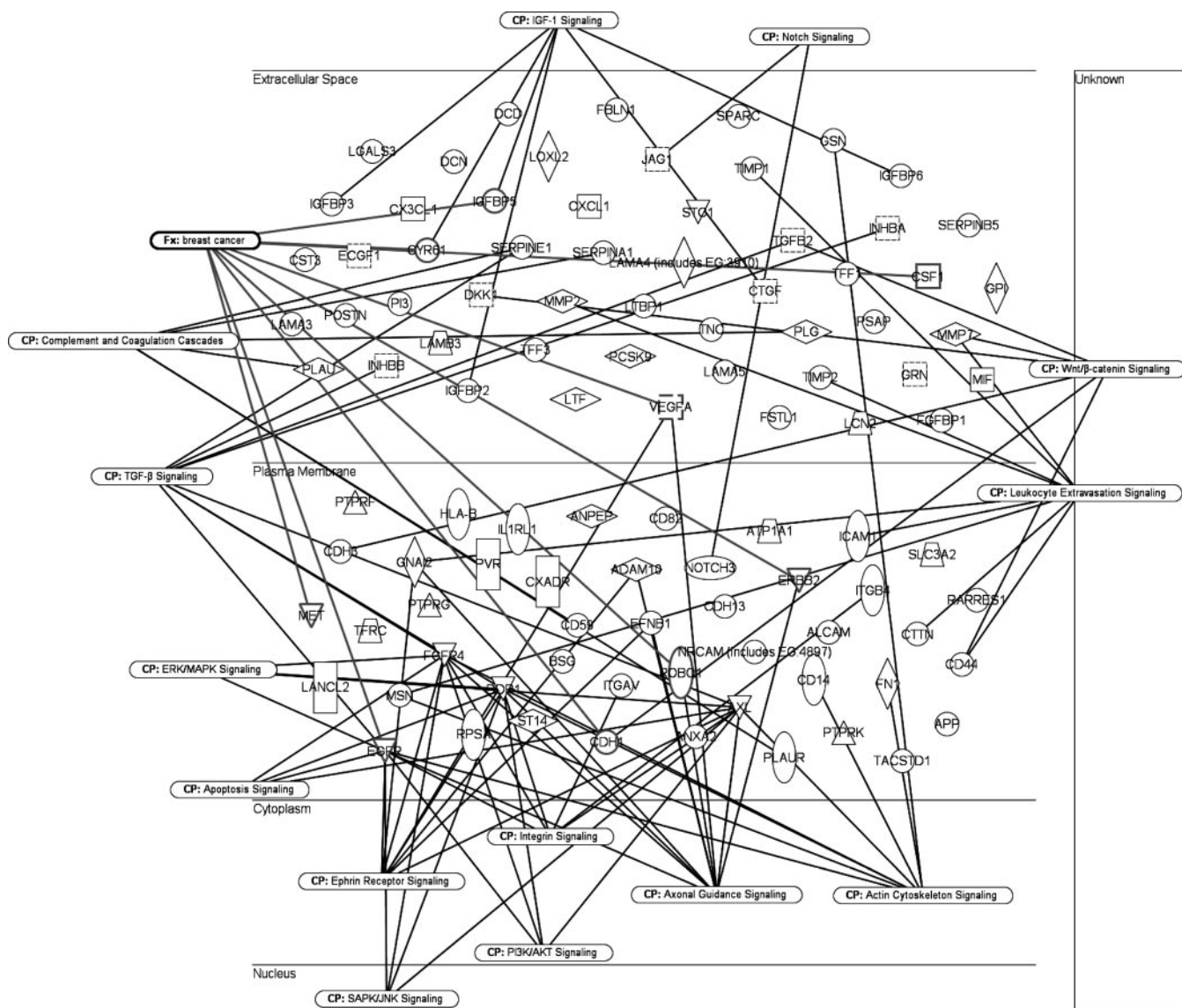


FIG. 6. **Canonical pathways and known genes linked to breast cancer.** Using Ingenuity Pathways Analysis, ~100 genes were identified in the literature as being reported in humans and cancer from the shortened list of candidates. The top 14 canonical pathways have been mapped onto the 100 genes along with genes linked to breast cancer including *IGFBP5*, *VEGF*, and *ERBB2*. CP, canonical pathway; Fx, function.

and thus serving as cancer biomarkers or as important molecules involved in cancer progression. More than 34% of these proteins were classified as being extracellular and membrane-bound. Among the known and novel proteins released by breast cancer cells, we identified various proteases, receptors, protease inhibitors, and cytokines. All experiments were performed in triplicates with excellent reproducibility between runs. Due to the inherent nature of mass spectrometers, not all peptides were ionized in each run, and consequently, different peptides were selected for ionization and detected. This selective ionization can account for the 75% reproducibility in our biological triplicates. As well, the various steps during sample preparation, including  $C_{18}$  extraction of the fractions, cannot be dismissed as an important contribut-

ing factor to the variations observed. In addition, we validated the ability of elafin to discriminate between normal and tumor breast cytosols. Currently no conclusion can be drawn without examining more cytosols. Although we did not observe circulating biomarker potential in elafin to discriminate normal from diseased individuals, to our knowledge, we are the first to report on the expression of this protein in breast cytosols, biological fluids, and tissue lysates using an immunoassay. Elafin was found to be expressed in normal breast tissue.

Although the objective of this study was to identify secreted and membrane-bound proteins that have the potential to be cleaved and thus found in circulation, the identified proteins included many intracellular proteins, including ones classified by GO as nuclear and cytoplasmic. During the cell culture

process, a portion of the cell population will die, resulting in the release of intracellular proteins into the media. Despite optimizing cell culture conditions to minimize cell death, the identification of intracellular proteins in the CM is inevitable because of the high sensitivity of MS-based techniques utilized in this study. Martin *et al.* (23) examined the CM of a prostate cancer cell line and found a very similar GO distribution to the one we present here: that more than 50% of the proteins identified were intracellular. Therefore, one of the major challenges in the analysis of secreted proteins is distinguishing between proteins that are targeted to the extracellular space *versus* those that arise as low level contaminants due to normal cell death in routine cell culture. To address this, we performed a cell lysate proteome experiment to demonstrate conclusively that through our approach we are significantly enriching for secreted proteins. Furthermore a recent study examining the cell lysate proteome of a human mammary epithelial cell line found that 2% of the entire proteome was classified as extracellular, which was consistent with our findings (63).

Recently the analysis of thousands of genes in breast and colorectal cancers has shown that individual tumors can accumulate an average of 90 mutant genes (64). The authors identified 189 previously unknown genes that were mutated at a high frequency. From these genes, six were found in this study (filamin-B, spectrin  $\alpha$  chain, gelsolin, extracellular sulfatase Sulf-2, neuronal cell adhesion molecule, and polypeptide *N*-acetylgalactosaminyltransferase 5). Furthermore it is particularly important that many of the proteins identified by other groups using relevant biological fluids such as NAF and TIF were also present in our analysis (see Supplemental Table 8). Many of the proteins identified by Pawlik *et al.* (47) were highly abundant serum proteins such as albumin, transferrin, and various immunoglobulins, all of which were not identified in our serum-free media. As well, the top biological functions of the candidate molecules appear well suited to cancer initiation and progression such as those involved in cellular movement and cell-to-cell signaling. Finally a large portion of the shortened list of proteins identified was also found in the Human Plasma Proteome database. This finding was not unexpected as it served to highlight the fact that many of the proteins identified in our study had previously been found in plasma. But it also demonstrated that many more proteins have yet to be identified in plasma. There can be a number of reasons why they have not been identified, one of which is the fact that their concentration in plasma is too low to measure by current technologies, and thus other means of initially identifying them and then developing a specific and sensitive immunoassay are critical.

Our group has previously published data on the conditioned media of a prostate cancer cell line, PC3(AR)6, using a roller bottle cell culture method (65). Through this approach, the authors identified 262 proteins in the CM. The workflow presented in the current study has significant differences and

improvements compared with our previous work. At the tissue culture step, we cultured three cell lines in triplicate *versus* only one cell line studied. We also optimized our cell culture conditions (seeding density, incubation time, and volume of media used) to minimize cell death and maximize secreted protein content. Different methods of fractionation of the peptides *versus* protein fractionation and a more robust and sensitive mass spectrometer were utilized in this study compared with our previous work. Finally the bioinformatics analysis has also been significantly improved upon from before as we used two different search engines and incorporated protein and peptide probability calculations into our final list of proteins. As a result, in the current study, we identified over 1,000 proteins using the conditioned media approach.

Current therapies for advanced cancers are elusive. Novel breast cancer biomarkers that can be effective early in the course of the disease have the potential to reduce morbidity and mortality as well as receive a higher compliance rate by patients for undergoing screening. However, there is a growing consensus that panels of markers may be able to supply the specificity and sensitivity that individual markers lack. A number of studies have demonstrated that this is indeed true. Although a biomarker should be detected in serum using an immunoassay (ELISA), developing such an assay for multiple potential novel biomarkers is very labor-intensive (66). The vast majority of the proteins identified in this study (extracellular, membrane, and unclassified) do not have commercially available ELISA kits. Alternatively to decipher whether the candidates are present in serum, multiple reaction monitoring mass spectrometry technology can be performed. Using the latter technology, it is possible, in a single experiment, to detect and quantify specific peptides (representing specific proteins) in biological fluids of patients with breast cancer to determine whether the protein has biomarker potential. A number of studies have shown the feasibility of such an approach (67–69).

**Acknowledgments**—We thank Christopher R Smith for expert technical advice and support in mass spectrometry. We are grateful to Adrian Pasculescu, David Nguyen, and Mark Turner (Proteome Software Inc.) for bioinformatics support and Paul Taylor for many helpful discussions relating to this project.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

|| To whom correspondence should be addressed: Dept. of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Ave., Toronto, Ontario M5G 1X5, Canada. Tel.: 416-586-8443; Fax: 416-58-8628; E-mail: ediamandis@mtsina.on.ca.

#### REFERENCES

1. van Diest, P. J., van der, W. E., and Baak, J. P. (2004) Prognostic value of proliferation in invasive breast cancer: a review. *J. Clin. Pathol.* **57**, 675–681

2. Jemal, A., Tiwari, R. C., Murray, T., Ghafoor, A., Samuels, A., Ward, E., Feuer, E. J., and Thun, M. J. (2004) Cancer statistics, 2004. *CA Cancer J. Clin.* **54**, 8–29
3. Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C. M., Lonning, P. E., Brown, P. O., Borresen-Dale, A. L., and Botstein, D. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8418–8423
4. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de, R. M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein, L. 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
5. Abd El-Rehim, D. M., Ball, G., Pinder, S. E., Rakha, E., Paish, C., Robertson, J. F., Macmillan, D., Blamey, R. W., and Ellis, I. O. (2005) High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int. J. Cancer* **116**, 340–350
6. Jacquemier, J., Ginestier, C., Rougemont, J., Bardou, V. J., Charafe-Jauffret, E., Geneix, J., Adelaide, J., Koki, A., Houvenaeghel, G., Hassoun, J., Maraninchi, D., Viens, P., Birnbaum, D., and Bertucci, F. (2005) Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. *Cancer Res.* **65**, 767–779
7. Liotta, L. A., and Kohn, E. C. (2001) The microenvironment of the tumour-host interface. *Nature* **411**, 375–379
8. Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G., and Hartwell, L. (2003) The case for early detection. *Nat. Rev. Cancer* **3**, 243–252
9. Petricoin, E. F., and Liotta, L. A. (2003) Clinical applications of proteomics. *J. Nutr.* **133**, 2476S–2484S
10. Liotta, L. A., Ferrari, M., and Petricoin, E. (2003) Clinical proteomics: written in blood. *Nature* **425**, 905
11. Lumachi, F., and Basso, S. M. (2004) Serum tumor markers in patients with breast cancer. *Expert Rev. Anticancer Ther.* **4**, 921–931
12. Khatcheressian, J. L., Wolff, A. C., Smith, T. J., Grunfeld, E., Muss, H. B., Vogel, V. G., Halberg, F., Somerfield, M. R., and Davidson, N. E. (2006) American society of clinical oncology 2006 update of the breast cancer follow-up and management guidelines in the adjuvant setting. *J. Clin. Oncol.* **24**, 5091–5097
13. Diamandis, E. P., Fritsche, H. A., Lilja, H., Chan, D. W., and Schwartz, M. K. (2002) *Tumor Markers: Physiology, Pathobiology, Technology, and Clinical Applications*, AACC (American Association for Clinical Chemistry) Press, Washington, D. C.
14. Leitch, A. M. (1995) Controversies in breast cancer screening. *Cancer* **76**, 2064–2069
15. Esserman, M. R., Cowley, H., Eberle, C., Kirkpatrick, A., Chang, S., Berbaum, K., and Gale, A. (2002) Improving the accuracy of mammography: volume and outcome relationships. *J. Natl. Cancer Inst.* **94**, 369–375
16. Mincey, B. A., and Perez, E. A. (2004) Advances in screening, diagnosis, and treatment of breast cancer. *Mayo Clin. Proc.* **79**, 810–816
17. Antman, K., and Shea, S. (1999) Screening mammography under age 50. *J. Am. Med. Assoc.* **281**, 1470–1472
18. Shao, Z. M., and Nguyen, M. (2001) Nipple aspiration in diagnosis of breast cancer. *Semin. Surg. Oncol.* **20**, 175–180
19. Wrensch, M. R., Petrakis, N. L., Gruenke, L. D., Ernster, V. L., Miike, R., King, E. B., and Hauck, W. W. (1990) Factors associated with obtaining nipple aspirate fluid: analysis of 1428 women and literature review. *Breast Cancer Res. Treat.* **15**, 39–51
20. Malatesta, M., Mannello, F., Bianchi, G., Sebastiani, M., and Gazzanelli, G. (2000) Biochemical and ultrastructural features of human milk and nipple aspirate fluids. *J. Clin. Lab. Anal.* **14**, 330–335
21. Klein, P., Glaser, E., Grogan, L., Keane, M., Lipkowitz, S., Soballe, P., Brooks, L., Jenkins, J., Steinberg, S. M., DeMarini, D. M., and Kirsch, I. (2001) Biomarker assays in nipple aspirate fluid. *Breast J.* **7**, 378–387
22. Alexander, H., Stegner, A. L., Wagner-Mann, C., Du Bois, G. C., Alexander, S., and Sauter, E. R. (2004) Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clin. Cancer Res.* **10**, 7500–7510
23. Martin, D. B., Gifford, D. R., Wright, M. E., Keller, A., Yi, E., Goodlett, D. R., Aebersold, R., and Nelson, P. S. (2004) Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer Res.* **64**, 347–355
24. Mbeunkui, F., Fodstad, O., and Pannell, L. K. (2006) Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. *J. Proteome Res.* **5**, 899–906
25. Canelle, L., Bousquet, J., Pionneau, C., Hardouin, J., Choquet-Kastylevsky, G., Joubert-Caron, R., and Caron, M. (2006) A proteomic approach to investigate potential biomarkers directed against membrane-associated breast cancer proteins. *Electrophoresis* **27**, 1609–1616
26. Xiang, R., Shi, Y., Dillon, D. A., Negin, B., Horvath, C., and Wilkins, J. A. (2004) 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474. *J. Proteome Res.* **3**, 1278–1283
27. Adam, P. J., Boyd, R., Tyson, K. L., Fletcher, G. C., Stamps, A., Hudson, L., Poyser, H. R., Redpath, N., Griffiths, M., Steers, G., Harris, A. L., Patel, S., Berry, J., Loader, J. A., Townsend, R. R., Daviet, L., Legrain, P., Parekh, R., and Terrett, J. A. (2003) Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. *J. Biol. Chem.* **278**, 6482–6489
28. Patwardhan, A. J., Strittmatter, E. F., Camp, D. G., Smith, R. D., and Pallavicini, M. G. (2005) Comparison of normal and breast cancer cell lines using proteome, genome, and interactome data. *J. Proteome Res.* **4**, 1952–1960
29. Charafe-Jauffret, E., Ginestier, C., Monville, F., Finetti, P., Adelaide, J., Cervera, N., Fekairi, S., Xerri, L., Jacquemier, J., Birnbaum, D., and Bertucci, F. (2006) Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **25**, 2273–2284
30. Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J. P., Tong, F., Speed, T., Spellman, P. T., DeVries, S., Lapuk, A., Wang, N. J., Kuo, W. L., Stilwell, J. L., Pinkel, D., Albertson, D. G., Waldman, F. M., McCormick, F., Dickson, R. B., Johnson, M. D., Lippman, M., Ethier, S., Gazdar, A., and Gray, J. W. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527
31. Soule, H. D., Maloney, T. M., Wolman, S. R., Peterson, W. D., Jr., Brenz, R., McGrath, C. M., Russo, J., Pauley, R. J., Jones, R. F., and Brooks, S. C. (1990) Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **50**, 6075–6086
32. Lasfargues, E. Y., Coutinho, W. G., and Redfield, E. S. (1978) Isolation of two human tumor epithelial cell lines from solid breast carcinomas. *J. Natl. Cancer Inst.* **61**, 967–978
33. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182
34. Shak, S. (1999) Overview of the trastuzumab (Herceptin) anti-HER2 monoclonal antibody clinical program in HER2-overexpressing metastatic breast cancer. Herceptin Multinational Investigator Study Group. *Semin. Oncol.* **26**, 71–77
35. Cailleau, R., Olive, M., and Cruciger, Q. V. (1978) Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* **14**, 911–915
36. She, Q. B., Solit, D., Basso, A., and Moasser, M. M. (2003) Resistance to gefitinib in PTEN-null HER-overexpressing tumor cells can be overcome through restoration of PTEN function or pharmacologic modulation of constitutive phosphatidylinositol 3'-kinase/Akt pathway signaling. *Clin. Cancer Res.* **9**, 4340–4346
37. Panigrahi, A. R., Pinder, S. E., Chan, S. Y., Paish, E. C., Robertson, J. F., and Ellis, I. O. (2004) The role of PTEN and its signalling pathways, including AKT, in breast cancer; an assessment of relationships with other prognostic factors and with outcome. *J. Pathol.* **204**, 93–100
38. Borgono, C. A., and Diamandis, E. P. (2004) The emerging roles of human tissue kallikreins in cancer. *Nat. Rev. Cancer* **4**, 876–890
39. Rittenhouse, H. G., Finlay, J. A., Mikolajczyk, S. D., and Partin, A. W. (1998) Human Kallikrein 2 (hK2) and prostate-specific antigen (PSA): two closely related, but distinct, kallikreins in the prostate. *Crit. Rev. Clin. Lab. Sci.* **35**, 275–368
40. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
41. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical



- model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
42. Luo, L. Y., Grass, L., Howarth, D. J., Thibault, P., Ong, H., and Diamandis, E. P. (2001) Immunofluorometric assay of human kallikrein 10 and its identification in biological fluids and tissues. *Clin. Chem.* **47**, 237–246
  43. Yousef, G. M., Polymeris, M. E., Grass, L., Soosaipillai, A., Chan, P. C., Scorilas, A., Borgono, C., Harbeck, N., Schmalfeldt, B., Dorn, J., Schmitt, M., and Diamandis, E. P. (2003) Human kallikrein 5: a potential novel serum biomarker for breast and ovarian cancer. *Cancer Res.* **63**, 3958–3965
  44. Diamandis, E. P., Scorilas, A., Fracchioli, S., Van Gramberen, M., De Bruijn, H., Henrik, A., Soosaipillai, A., Grass, L., Yousef, G. M., Stenman, U. H., Massobrio, M., Van Der Zee, A. G., Vergote, I., and Katsaros, D. (2003) Human kallikrein 6 (hK6): a new potential serum biomarker for diagnosis and prognosis of ovarian carcinoma. *J. Clin. Oncol.* **21**, 1035–1043
  45. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675
  46. Pinkas-Kramarski, R., Alroy, I., and Yarden, Y. (1997) ErbB receptors and EGF-like ligands: cell lineage determination and oncogenesis through combinatorial signaling. *J. Mammary Gland Biol. Neoplasia* **2**, 97–107
  47. Pawlik, T. M., Hawke, D. H., Liu, Y., Krishnamurthy, S., Fritsche, H., Hunt, K. K., and Kueter, H. M. (2006) Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC Cancer* **6**, 68
  48. Varnum, S. M., Covington, C. C., Woodbury, R. L., Petritis, K., Kangas, L. J., Abdullah, M. S., Pounds, J. G., Smith, R. D., and Zangar, R. C. (2003) Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. *Breast Cancer Res. Treat.* **80**, 87–97
  49. Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M., Ambartsumian, N., Sandelin, K., Rank, F., and Gromova, I. (2004) Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol. Cell. Proteomics* **3**, 327–344
  50. Minn, A. J., Gupta, G. P., Siegel, P. M., Bos, P. D., Shu, W., Giri, D. D., Viale, A., Olshen, A. B., Gerald, W. L., and Massague, J. (2005) Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524
  51. Liotta, L. A., Steeg, P. S., and Stetler-Stevenson, W. G. (1991) Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell* **64**, 327–336
  52. Schalkwijk, J., Wiedow, O., and Hirose, S. (1999) The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. *Biochem. J.* **340**, 569–577
  53. Zhang, M., Zou, Z., Maass, N., and Sager, R. (1995) Differential expression of elafin in human normal mammary epithelial cells and carcinomas is regulated at the transcriptional level. *Cancer Res.* **55**, 2537–2541
  54. Searle, B. C., Brundage, J. M., and Turner, M. (2005) Improving sensitivity by combining results from multiple MS/MS search methodologies with the scaffold computer algorithm, in *Human Proteome Organisation (HUPO), 4th Annual World Congress, Munich, Germany, August 28, 2005*, Abstr. WP299, HUPO, Montreal, Quebec, Canada
  55. Domon, B., and Aebersold, R. (2006) Challenges and opportunities in proteomics data analysis. *Mol. Cell. Proteomics* **5**, 1921–1926
  56. Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., and Washburn, M. P. (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**, 2339–2347
  57. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245
  58. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005) Exponentially modified protein abundance index (em-PAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265–1272
  59. Yousef, G. M., Scorilas, A., Kyriakopoulou, L. G., Rendl, L., Diamandis, M., Ponzone, R., Biglia, N., Giai, M., Roagna, R., Sismondi, P., and Diamandis, E. P. (2002) Human kallikrein gene 5 (KLK5) expression by quantitative PCR: an independent indicator of poor prognosis in breast cancer. *Clin. Chem.* **48**, 1241–1250
  60. Luo, L. Y., Diamandis, E. P., Look, M. P., Soosaipillai, A. P., and Foekens, J. A. (2002) Higher expression of human kallikrein 10 in breast cancer tissue predicts tamoxifen resistance. *Br. J. Cancer* **86**, 1790–1796
  61. Lipton, A., Ali, S. M., Leitzel, K., Demers, L., Chinchilli, V., Engle, L., Harvey, H. A., Brady, C., Nalin, C. M., Dugan, M., Carney, W., and Allard, J. (2002) Elevated serum Her-2/neu level predicts decreased response to hormone therapy in metastatic breast cancer. *J. Clin. Oncol.* **20**, 1467–1472
  62. Esteve, F. J., Valero, V., Booser, D., Guerra, L. T., Murray, J. L., Pusztai, L., Cristofanilli, M., Arun, B., Esmaili, B., Fritsche, H. A., Sneige, N., Smith, T. L., and Hortobagyi, G. N. (2002) Phase II study of weekly docetaxel and trastuzumab for patients with HER-2-overexpressing metastatic breast cancer. *J. Clin. Oncol.* **20**, 1800–1808
  63. Jacobs, J. M., Mottaz, H. M., Yu, L. R., Anderson, D. J., Moore, R. J., Chen, W. N., Auberry, K. J., Strittmatter, E. F., Monroe, M. E., Thrall, B. D., Camp, D. G., and Smith, R. D. (2004) Multidimensional proteome analysis of human mammary epithelial cells. *J. Proteome Res.* **3**, 68–75
  64. Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274
  65. Sardana, G., Marshall, J., and Diamandis, E. P. (2007) Discovery of candidate tumor markers for prostate cancer via proteomic analysis of cell culture-conditioned medium. *Clin. Chem.* **53**, 429–437
  66. Zolg, W. (2006) The proteomic search for diagnostic biomarkers: lost in translation? *Mol. Cell. Proteomics* **5**, 1720–1726
  67. Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* **5**, 573–588
  68. Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., and Pearson, T. W. (2004) Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). *J. Proteome Res.* **3**, 235–244
  69. Barnidge, D. R., Goodmanson, M. K., Klee, G. G., and Muddiman, D. C. (2004) Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution mass spectrometry. *J. Proteome Res.* **3**, 644–652

## SUPPLEMENTARY FIGURES

### **Figure 1: Cell number, protein, LDH and kallikrein levels in 24 hour conditioned media of cell**

**lines.** (A) Total number of cells alive during harvest of the CM; (B) Total protein concentration of CM (C) LDH levels, indicating approximately 6-7% cell death; (D) Standard LDH curve (n = 2); (E) KLK5, KLK6 and KLK10 levels in MDA-MB-468. This procedure was repeated several times for reproducibility.

### **Figure 2: Identification of Positive Internal Control Proteins by MS.** MDA-MB-468 expresses

human kallikreins 5, 6 and 10 in concentrations ranging from 2-50µg/L, as measured and validated by ELISA. (A) Kallikrein 5 (B) Kallikrein 6 (C) Kallikrein 10 showing the sequences of the respective proteins and highlighted in yellow are the peptides identified. BT474 is known to display amplification of Her-2/neu, a well established breast cancer biomarker. (D) Her-2/neu or receptor tyrosine-protein kinase erbB-2 precursor, showing the sequence of the protein and its peptide coverage in the ECD domain.

**Figure 3: Levels of CA 15-3 in the Conditioned Media of Cell Lines** (measured by Elecsys CA 15-3 Immunoassay; Roche Diagnostics GmbH).

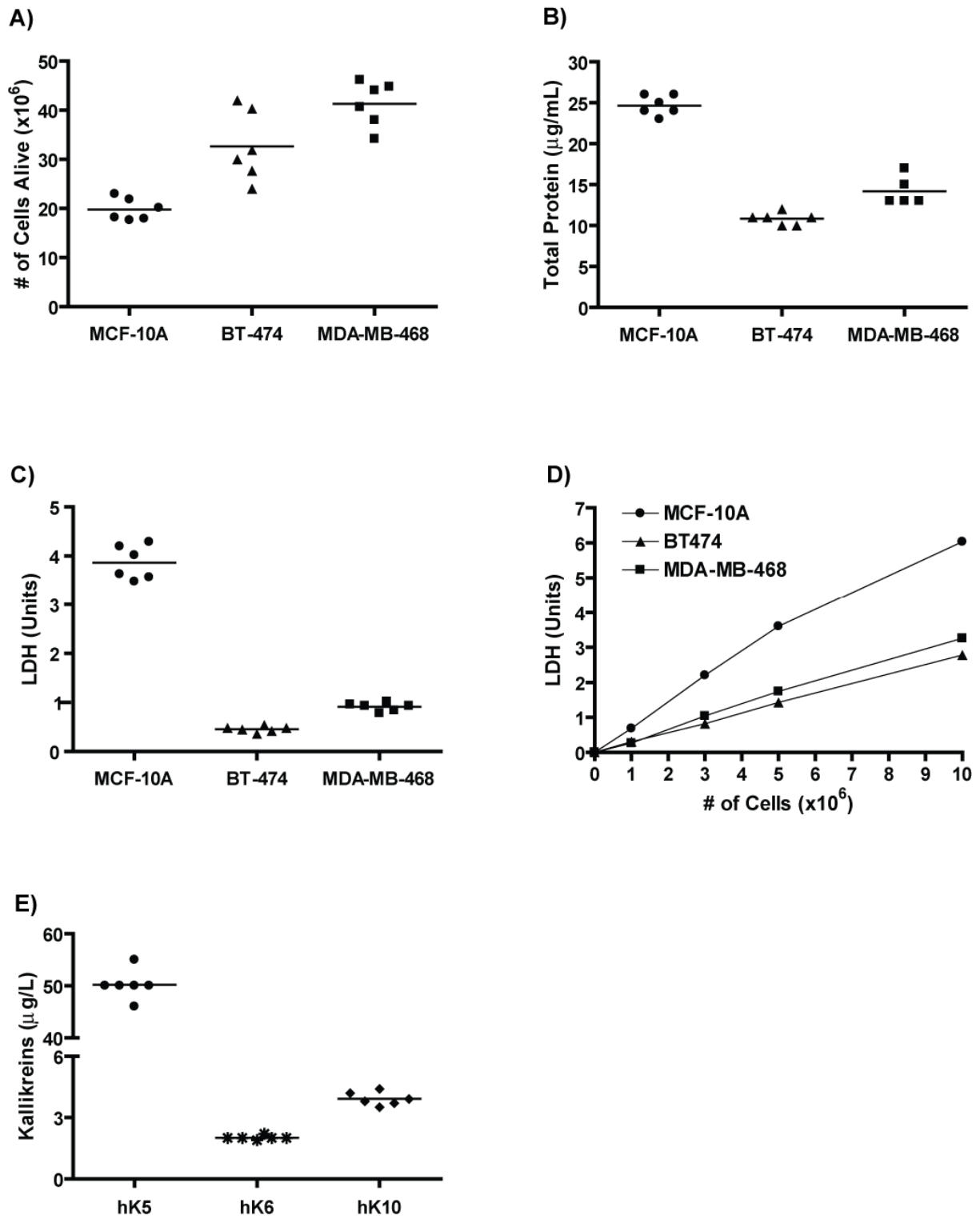
**Figure 4: Overlap of proteins between cell lines and cellular localization using spectral counts.**

**Figure 5: Levels of KLK5, KLK6 and KLK10 in serum.** Serum from normal female (15), normal male (15) and females with high CA 15-3 levels (17) were analyzed for KLK expression by ELISA (median values shown). For the serum samples for which CA 15-3 levels were known, their values were plotted against each other. No correlation was observed.

**Figure 6: Levels of KLK5, KLK6 and KLK10 in biological samples.** Normal and tumor breast cytosols as well as various pooled biological samples were assayed for kallikrein expression by ELISA (median values shown). The number of samples used to generate a pooled sample is indicated in brackets.



Supplementary Figure 1



## Supplementary Figure 2

A)

IPI00032464 (100%), 32019.9 Da

Kallikrein-5 precursor

3 unique peptides, 6 unique spectra, 7 total spectra, 46/293 amino acids (16% coverage)

MATARPPWMW	VLCALITALL	LGVTEHVLAN	NDVSCDHPSN	TVPSGSNQDL
GAGAGEDARS	DDSSSRING	SDCDMHTQPW	QAALLLRPNQ	LYCGAVLVHP
QWLLTAAHCR	KKVFRVR	<b>LGH</b>	<b>YSLSPVYESG</b>	<b>QQMFQGVKSI</b>
SNDLMLIKLN	RRIRPTK	<b>DVR</b>	<b>PINVSSHCPS</b>	<b>AGTKCLVSGW</b>
<b>PK</b> VLQCLNIS	VLSQKRCEDA	YPRQIDDTMF	CAGDKAGRDS	CQGDSSGGPVV
CNGSLQGLVS	WGDYPCARP	RPGVYTNLCK	FTKWIQETIQ	ANS

B)

IPI00023845 (100%), 26855.6 Da

Kallikrein-6 precursor

2 unique peptides, 3 unique spectra, 3 total spectra, 30/244 amino acids (12% coverage)

MKKLMVVLSS	IAAAWAEEN	KL VHGGPCDK	TSHPYQAALY	TSGHLLCGGV
L I HPLWVLT	AHCK	<b>KPNLQV</b>	<b>FLGKHNLQR</b>	<b>ESSQEQSSVV</b>
<b>ASHDQDMLL</b>	<b>RLARPAKLSE</b>	<b>LIQPLPLERD</b>	<b>CSANTTTSCHI</b>	<b>RAVIHPDYDA</b>
PRVRALQLPY	HLVSREECEH	AYPGQITQNM	LCAGDEKYGK	LGWGKTADGD
L VCGDHLRGL	VSNGNIPCGS	KEKPGVYTNV	CRYTNWIQKT	DSCQDSSGGP
				IQAK

C)

IPI00305266 (100%), 30137.6 Da

Kallikrein-10 precursor

3 unique peptides, 4 unique spectra, 4 total spectra, 37/276 amino acids (13% coverage)

MRAPHLHLSA	ASGARALAKL	LPLLMAQLWA	AEAALLPQND	TRLDP EAYGA
PCARGSQPWQ	VSLFNGLSFH	CAGVLVDQSW	VLTAHHCNGK	PLWAR <b>VGGDH</b>
<b>LLLLQGEOQLR</b>	RTTRS VVHPK	<b>YHQGSGPILP</b>	<b>RRTDEHDLML</b>	<b>LKLARPVVP</b>
PRVRALQLPY	RCAQPGDQCC	VAGWGTTAAR	RVKYNKGLTC	SSITILSPKE
CEVFYPGVVT	NNMICAGLDR	GQDPCQSDSG	GPLVCDETLQ	GILSWGVPYC
GSAQHPAVYT	QICKYMSWIN	KVIRSN		

D)

IPI00300384 (100%), 137910.5 Da

Receptor tyrosine-protein kinase erbB-2 precursor

4 unique peptides, 5 unique spectra, 6 total spectra, 53/1255 amino acids (4% coverage)

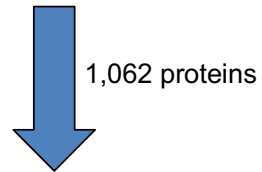
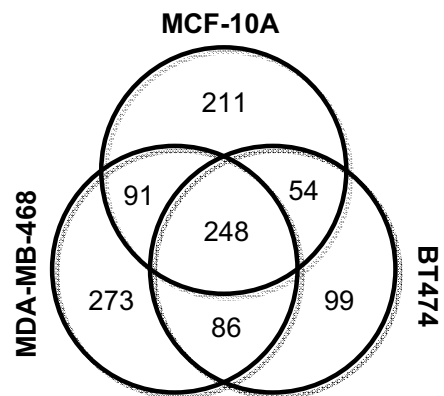
MELAAALCRWG	LLLALLPPGA	ASTQVCTGTD	MKLR	<b>LPASPE</b>	<b>THLDMRLHLY</b>
QGCQVVGQNL	ELTYLPTNAS	LSFLQDIQEV	QGYVLIAHNQ	VROVPLQRLR	
IVRGTLQFED	NYALAVLDNG	DPLNNTTPVT	GASPGGLREL	QLRSLTEILK	
GGVLIQRNPQ	<b>LCYQDTILWK</b>	DIFHK	<b>NNQLA</b>	<b>LTLIDTNRSR</b>	ACHPCSPMCK
GSR <b>CWGESSE</b>	<b>DCQSLTRTVC</b>	AGGCARCKGP	LPTDCCHEQC	AAGCTGPKHS	
DCLACLHFNH	SGICELHCPA	LVTYNTDTFE	SMPNPEGRTY	FGASCVTACP	
YNYLSTDVGS	CTLVCPHNLQ	EVTAE DGTQR	CEKCSKPCAR	VCYGLGMEHL	
REVRAVTSAN	IQEFAGCKKI	FGSLAFLPES	FDGDPASNTA	PLQPEQLQVF	
ETLEEITGYL	YISAWPDSL	DLSVFQNLQV	IRGRILHNGA	YSLTLQGLGI	
SWLGLRSLRE	LGSGLALIH	NTHLCFVHTV	PWDQLFRNPH	QALLHTANRP	
EDECVGEGLA	CHQLCARGHC	WGPPTQCVN	CSQFLRGQEC	VEECCRVLQGL	
PREYVNARHC	LPCHPQCQPQ	NGSVTCFGPE	ADQCVCACHY	KDPPFCVARC	
PSGVKPDLSY	MPIWKFPDEE	GACQPCPINC	THSCVDLDDK	GCPAEQRASP	
LTSIISAVVG	ILLVVVLGVV	FGILIKRRQQ	KIRKYTMRR	LQETELVEPL	
TPSGAMPNQA	QMRILKETEL	RKVKVLGSGA	FGTVYKGIWI	PDGENVKIPV	
AIKVLRENTS	PKANKEILDE	AYVMAGVGS	YVSRLGLICL	TSTVQLVTQL	
MPYGCLLDHV	RENRRGLGSQ	DLLNWCMI	IAKMSYLEDVR	LVHRDLAARN	
VLVKSPNHVK	ITDFGLARLL	DIDETEHYAD	GGKVPKWMMA	LESILRRRFT	
HQSDVWSYGV	TVWELMTFGA	KPYDGIIPARE	IPDLLEKGER	LPQPPICTID	
VYIMVVKCWM	IDSECRPRFR	ELVSEFSRMA	RDPPRFVVIQ	NEDLGPASPL	
DSTFYRSLLE	DDDMGDLVDA	EEYLVPPQGG	FCDPAPAGAG	GMVHHRHRSS	
STRSGGGDLT	LGLEPSEEEA	PRSLPAPESE	AGSDVDFDGL	GMGAACKGLQS	
LPTHDPSPPLQ	RYSEDPTVPL	PSETDGYVAP	LTCSPQPEYV	NQDPVRPQPP	
SPREGPLPAA	RPAGATLERP	KTLSPGKNGV	VKDVFVAFGGA	VENPEYLTPO	
GGAAPQPHPP	PAFSPAFDNL	YYWDQDPPER	GAPPSTFKGT	PTAENPEYLG	
INDPV					

**Supplementary Figure 3**

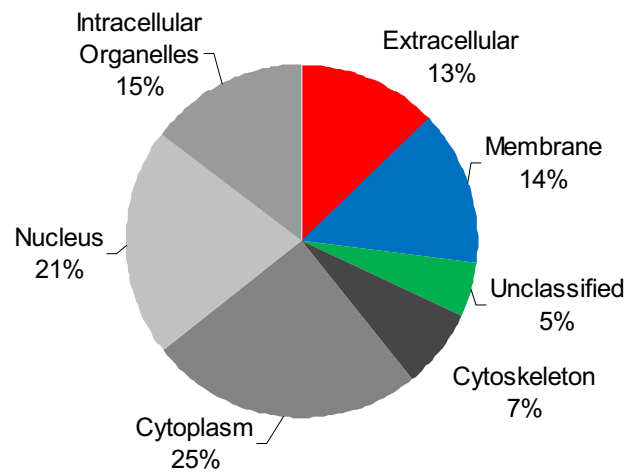
Cell Line	CA 15-3 Levels (Units/mL)
MCF-10A Rep 1	2.57
MCF-10A Rep 2	2.24
MCF-10A Rep 3	1.92
BT474 Rep 1	0
BT474 Rep 2	0
BT474 Rep 3	0
MDA-MB-468 Rep 1	0
MDA-MB-468 Rep 2	0
MDA-MB-468 Rep 3	0

**Supplementary Figure 4**

**A)**

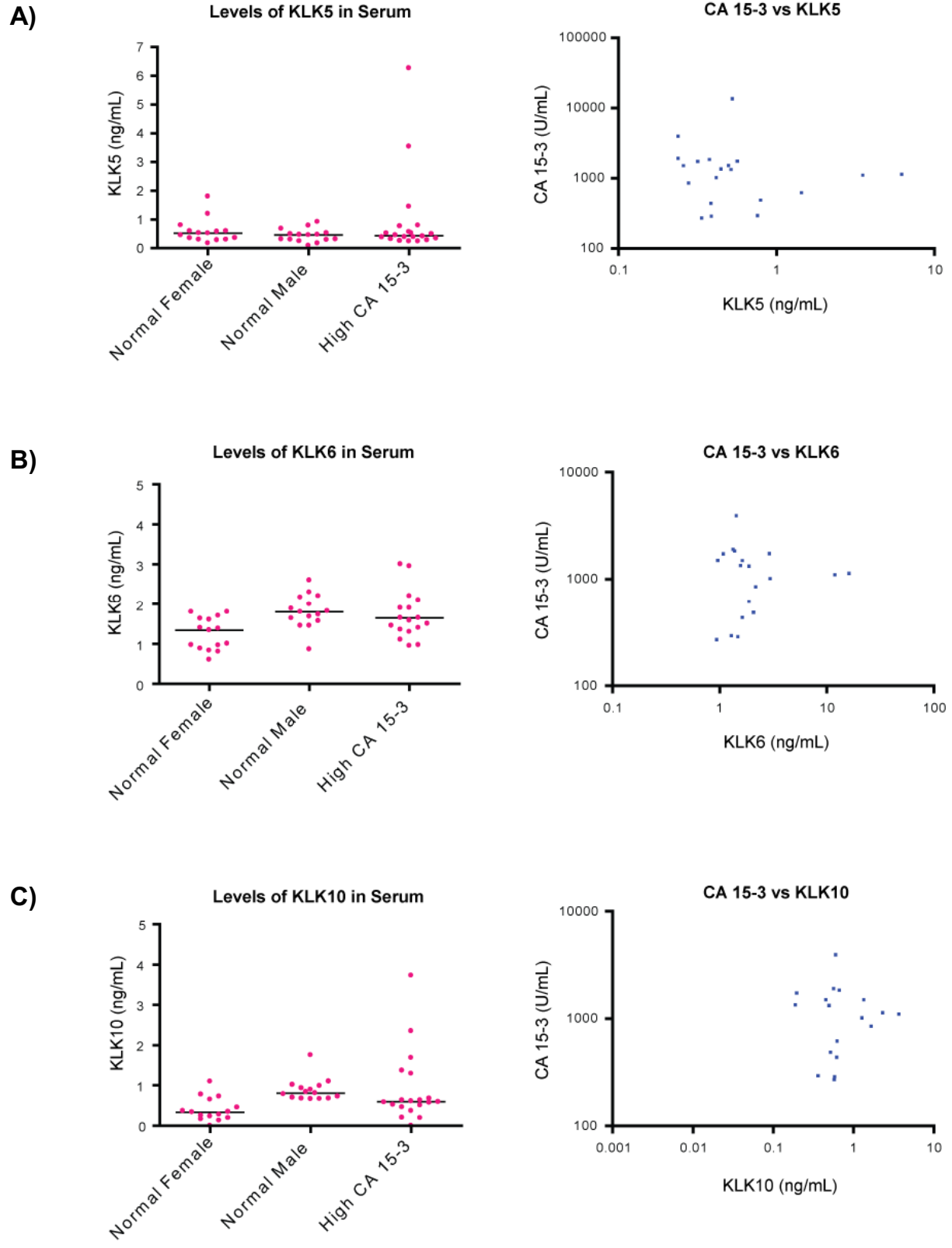


**B)**



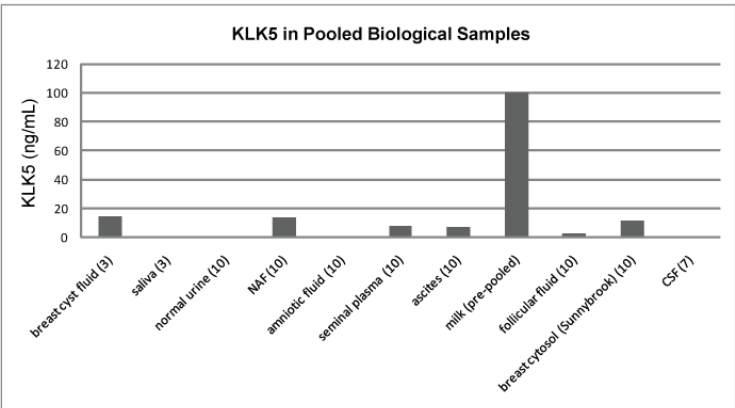
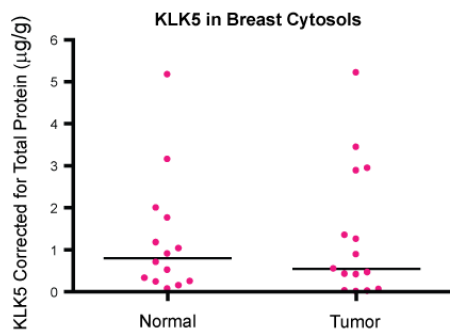


Supplementary Figure 5

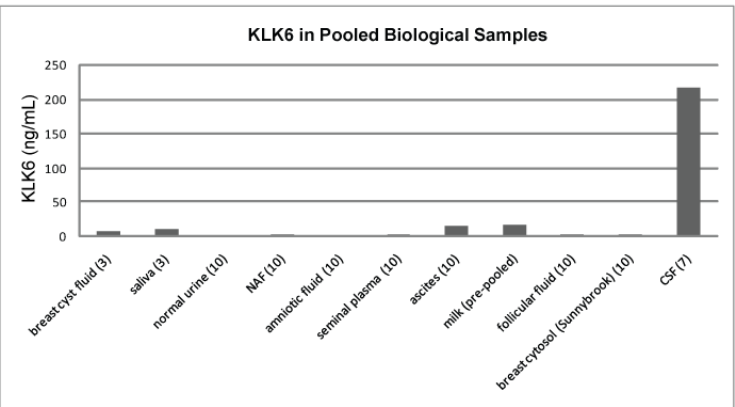
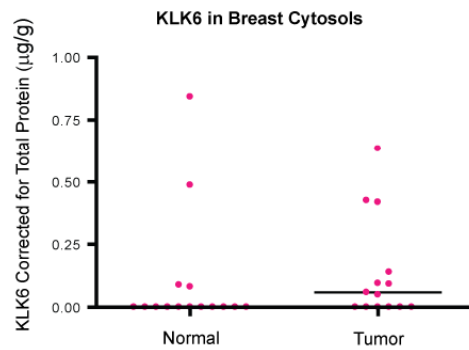


Supplementary Figure 6

A)



B)



C)

