

# Candidate Serological Biomarkers for Cancer Identified from the Secretomes of 23 Cancer Cell Lines and the Human Protein Atlas\*<sup>§</sup>

Chih-Ching Wu<sup>‡§</sup>, Chia-Wei Hsu<sup>¶</sup>, Chi-De Chen<sup>¶</sup>, Chia-Jung Yu<sup>¶||</sup>, Kai-Ping Chang<sup>\*\*</sup>, Dar-In Tai<sup>‡‡</sup>, Hao-Ping Liu<sup>‡</sup>, Wen-Hui Su<sup>‡</sup>, Yu-Sun Chang<sup>¶¶</sup>, and Jau-Song Yu<sup>¶||§§</sup>

Although cancer cell secretome profiling is a promising strategy used to identify potential body fluid-accessible cancer biomarkers, questions remain regarding the depth to which the cancer cell secretome can be mined and the efficiency with which researchers can select useful candidates from the growing list of identified proteins. Therefore, we analyzed the secretomes of 23 human cancer cell lines derived from 11 cancer types using one-dimensional SDS-PAGE and nano-LC-MS/MS performed on an LTQ-Orbitrap mass spectrometer to generate a more comprehensive cancer cell secretome. A total of 31,180 proteins was detected, accounting for 4,584 non-redundant proteins, with an average of 1,300 proteins identified per cell line. Using protein secretion-predictive algorithms, 55.8% of the proteins appeared to be released or shed from cells. The identified proteins were selected as potential marker candidates according to three strategies: (i) proteins apparently secreted by one cancer type but not by others (cancer type-specific marker candidates), (ii) proteins released by most cancer cell lines (pan-cancer marker candidates), and (iii) proteins putatively linked to cancer-relevant pathways. We then examined protein expression profiles in the Human Protein Atlas to identify biomarker candidates that were simultaneously detected in the secretomes and highly expressed in cancer tissues. This analysis yielded 6–137 marker candidates selective for each tumor type and 94 potential pan-cancer markers. Among these, we selectively validated monocyte differentiation antigen CD14 (for liver cancer), stromal cell-derived factor 1 (for lung cancer), and cathepsin L1 and interferon-induced 17-kDa protein (for nasopharyngeal carcinoma) as potential serological cancer markers. In summary, the proteins identified from the secretomes of 23 cancer cell lines and the Human Protein Atlas represent a focused reservoir of potential cancer biomarkers. *Molecular & Cellular Proteomics* 9:1100–1117, 2010.

From the <sup>‡</sup>Molecular Medicine Research Center, <sup>¶</sup>Graduate Institute of Biomedical Sciences, and <sup>||</sup>Department of Cell and Molecular Biology, Chang Gung University and Departments of <sup>\*\*</sup>Otolaryngology-Head and Neck Surgery and <sup>‡‡</sup>Hepatogastroenterology, Chang Gung Memorial Hospital, Tao-Yuan 333, Taiwan

Received, August 26, 2009, and in revised form, November 23, 2009

Published, MCP Papers in Press, February 1, 2010, DOI 10.1074/mcp.M900398-MCP200

Cancer is a major cause of mortality worldwide, accounting for 10 million new cases and more than 6 million deaths per year. In developing countries, cancer is the second most common cause of death, accounting for 23–25% of the overall mortality rate (1). Notwithstanding improvements in diagnostic imaging technologies and medical treatments, the long term survival of most cancer patients is poor. Cancer therapy is often challenging because the majority of cancers are initially diagnosed in their advanced stages. For example, the 5-year survival rate for patients with HNC<sup>1</sup> is less than 50%. More than 50% of all HNC patients have advanced disease at the time of diagnosis (2, 3). Enormous effort has been devoted to screening and characterizing cancer markers for the early detection of cancer. Thus far, these markers include carcino-embryonic antigen, prostate-specific antigen,  $\alpha$ -fetoprotein, CA 125, CA 15-3, and CA 19-9. Unfortunately, most biomarkers have limited specificity, sensitivity, or both (4). Thus, there is a growing consensus that marker panels, which are more sensitive and specific than individual markers, would increase the efficacy and accuracy of early stage cancer detection (4–8). The development of novel and useful biomarker panels is therefore an urgent need in the field of cancer management.

Proteomics technology platforms are promising tools for the discovery of new cancer biomarkers (9). Over the past decade, serum and plasma have been the major targets of proteomics studies aimed at identifying potential cancer biomarkers (10–13). However, the progress of these studies has been hampered by the complex nature of serum/plasma samples and the large dynamic range between the concentrations of different proteins (14). As cancer biomarkers are likely to be

<sup>1</sup> The abbreviations used are: HNC, head and neck cancer; BIGH3, transforming growth factor- $\beta$ -induced protein ig-h3; CRC, colorectal carcinoma; emPAI, exponentially modified protein abundance index; GeLC-MS/MS, one-dimensional SDS-PAGE in conjunction with nano-LC-MS/MS; GO, gene ontology; HCC, hepatocellular carcinoma; HPA, Human Protein Atlas; IPI, International Protein Index; NPC, nasopharyngeal carcinoma; PAI-1, plasminogen activator inhibitor 1; TMHMM, transmembrane hidden Markov model; TPP, Trans-Proteomic Pipeline; FDR, false discovery rate; IHC, immunohistochemical; Mac-2BP, Mac-2-binding protein; ALCAM, activated leukocyte cell adhesion; MIF, macrophage migration inhibitory factor; TARC, thymus and activation-regulated chemokine.

present in low amounts in blood samples, the direct isolation of these markers from plasma and serum samples requires a labor-intensive process involving the depletion of abundant proteins and extensive protein fractionation prior to mass spectrometric analysis (15–18). Alternatively, the secretome, or group of proteins secreted by cancer cells (19), can be analyzed to identify circulating molecules present at elevated levels in serum or plasma samples from cancer patients. These proteins have the potential to act as cancer-derived marker candidates, which are distinct from host-responsive marker candidates. We, along with other groups, have demonstrated the efficacy of secretome-based strategies in a variety of cancer types, including NPC (20), breast cancer (21, 22), lung cancer (23, 24), CRC (25, 26), oral cancer (27), prostate cancer (28, 29), ovarian cancer (30), and Hodgkin lymphoma (31). In these studies, proteins secreted from cancer cells into serum-free media were resolved by one- or two-dimensional gels followed by in-gel tryptic digestion and analysis via MALDI-TOF MS or LC-MS/MS. Alternatively, the proteins were trypsin-digested in solution and analyzed by LC-MS/MS. In general, more proteins were detected in the secretome using the LC-MS/MS method than the MALDI-TOF MS method. Advanced protein separation and identification technologies have made it possible to detect more proteins in the secretomes of cancer cells, thereby facilitating the discovery of cancer biomarkers.

Although the cancer cell secretomes of various tumor types have been individually analyzed by different groups using distinct protocols, few studies have used the same protocol to compare cancer cell secretomes derived from different tumor types. We previously assessed the secretomes of 21 cancer cell lines derived from 12 cancer types (*i.e.* consisting of 795 protein identities and 325 non-redundant proteins) by one-dimensional gel and MALDI-TOF MS (25). Our preliminary findings revealed that different cell lines have distinct secreted protein profiles and that several putative biomarkers, such as Mac-2BP (20, 26, 27, 29) and cathepsin D (21, 23, 32), present in the secretome of a given cancer cell type are commonly shared among different cancers. These observations suggest that an in-depth comparison of secretomes derived from different tumor types may identify marker candidates common to most cancers as well as markers for specific cancer types. As an increasing number of proteins are identified in the secretomes of various cancer cell lines, scientists are faced with the challenge of quickly and efficiently narrowing down the list to candidates with higher chances of success during validation testing with precious clinical specimens.

In the present study, we applied one-dimensional SDS-PAGE in conjunction with nano-LC-MS/MS (GeLC-MS/MS) (33, 34) to analyze the conditioned media of 23 cancer cell lines derived from 11 cancer types, including NPC, breast cancer, bladder cancer, cervical cancer, CRC, epidermoid carcinoma, liver cancer, lung cancer, T cell lymphoma, oral cancer, and pancreatic cancer. Within this data set, 4,584

non-redundant proteins were identified from a total of 23 cell lines, yielding an average of ~1,300 proteins per cell line. Potential marker candidates were identified via the comparative analysis of different cell line secretomes and by putative linkages to cancer-relevant pathways. The selected proteins were further compared with the HPA (35) to generate a focused data set of proteins that are secreted or released, cancer type-specific, and highly expressed in human cancer tissues. Finally, we selectively validated four proteins as potential serological cancer markers using blood samples from cancer patients.

#### EXPERIMENTAL PROCEDURES

**Cell Culture**—Colorectal carcinoma cells (Colo205, ATCC CCL-222; SW480, ATCC CCL-228; and SW620, ATCC CCL-227), acute T cell leukemia cells (Jurkat, ATCC TIB-152), bladder cancer cells (U1 and U4), oral cancer cells (OEC-M1; SCC-4, ATCC CRL-1624), and lung cancer cells (CL<sub>1-0</sub> and CL<sub>1-5</sub>) were maintained in RPMI 1640 medium with 10% fetal bovine serum and antibiotics at 37 °C in an atmosphere of 5% CO<sub>2</sub>. Nasopharyngeal carcinoma cells (NPC-TW02, NPC-TW04, and NPC-BM1), HCC cells (SK-Hep-1, ATCC HTB-52; Hep G2, ATCC HB-8065; and Hep 3B, ATCC HB-8064), cervical carcinoma cells (C-33A, ATCC HTB-31), HeLa cells (ATCC CCL-2), epidermoid carcinoma cells (A431, ATCC CRL-1555), pancreatic carcinoma cells (PANC-1, ATCC CRL-1469; MIA PaCa-2, ATCC CRL-1420), and breast cancer cells (MCF-7, ATCC HTB-22; and MDA-MB435S, ATCC HTB-129) were grown in Dulbecco's modified Eagle's medium with 10% fetal bovine serum and antibiotics at 37 °C in an atmosphere of 5% CO<sub>2</sub>. All cell culture reagents were obtained from Invitrogen. The NPC-TW02 and NPC-TW04 cell lines were kindly provided by Dr. C.-T. Lin (Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan). The NPC-BM1 cell line was obtained from Dr. S.-K. Liao (Graduate Institute of Clinical Medicine, Chang Gung University, Tao-Yuan, Taiwan). The CL<sub>1-0</sub> and CL<sub>1-5</sub> cell lines were kindly provided by Dr. P.-C. Yang (Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan). The U1 and U4 cells were obtained from Dr. Y.-M. Yung (Department of Physiology and Pharmacology, Chang Gung University, Tao-Yuan, Taiwan).

**Preparation of Conditioned Media and Cell Extracts**—Conditioned media from the various cancer cell lines were collected and processed as described previously (25). Briefly, cancer cells were grown to confluence in 15-cm tissue culture dishes. The cells were washed with serum-free media and incubated in serum-free media for 24 h. The supernatants were then harvested and centrifuged to remove the intact cells followed by centrifugation in Amicon Ultra-15 tubes (molecular mass cutoff, 5,000 Da; Millipore, Billerica, MA) to concentrate and desalt the supernatants. Cells remaining on the dishes were washed twice with PBS and lysed in hypotonic buffer (10 mM Tris, pH 7.4, 1 mM EDTA, 1 mM EGTA, 50 mM NaCl, 50 mM NaF, 20 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 1 mM PMSF, 1 mM benzamidine, 0.5 μg/ml leupeptin, and 1% Triton X-100) on ice for 15 min. The cell lysates were collected, sonicated on ice, and centrifuged at 10,000 × *g* for 25 min at 4 °C. The resulting supernatants were used as cell extracts (20). The protein concentrations of the various samples were determined using the BCA protein assay reagent from Pierce.

**One-dimensional SDS-PAGE and In-gel Protein Digestion**—In preparation for secretome analyses, proteins (50 μg) were applied to 8–14% gradient gels for SDS-PAGE. After staining by 0.5% Coomassie Brilliant Blue G-250 (AppliChem GmbH, Darmstadt, Germany), the gel lane was cut into 70 pieces and subjected to in-gel tryptic digestion as described by Wu *et al.* (25). Briefly, the gel pieces were

destained in 10% methanol (Mallinckrodt Baker), dehydrated in acetonitrile (Mallinckrodt Baker), and dried using a SpeedVac. The proteins were reduced with 25 mM  $\text{NH}_4\text{HCO}_3$  containing 10 mM dithiothreitol (Biosynth AG) at 60 °C for 30 min and alkylated with 55 mM iodoacetamide (Amersham Biosciences) at room temperature for 30 min. After reduction and alkylation, proteins were digested via overnight incubation with sequencing grade modified porcine trypsin (20  $\mu\text{g}/\text{ml}$ ) (Promega, Madison, WI) at 37 °C. Peptides were extracted using acetonitrile and dried in a SpeedVac.

**Reverse-phase Liquid Chromatography/Tandem Mass Spectrometry**—To analyze the cancer cell secretomes, each peptide mixture was reconstituted in HPLC buffer A (0.1% formic acid; Sigma), loaded across a trap column (Zorbax 300SB-C18, 0.3  $\times$  5 mm; Agilent Technologies, Wilmington, DE) at a flow rate of 0.2  $\mu\text{l}/\text{min}$  in HPLC buffer A, and separated on a resolving 10-cm analytical  $\text{C}_{18}$  column (inner diameter, 75  $\mu\text{m}$ ) using a 15- $\mu\text{m}$  tip (New Objective, Woburn, MA). The peptides were eluted using a linear gradient of 0–10% HPLC buffer B (*i.e.* 99.9% ACN containing 0.1% formic acid) for 3 min, 10–30% buffer B for 35 min, 30–35% buffer B for 4 min, 35–50% buffer B for 1 min, 50–95% buffer B for 1 min, and 95% buffer B for 8 min with a flow rate of 0.25  $\mu\text{l}/\text{min}$  across the analytical column.

The LC apparatus was coupled with a two-dimensional linear ion trap mass spectrometer (LTQ-Orbitrap, Thermo Fisher, San Jose, CA), which was operated using Xcalibur 2.0 software (Thermo Fisher). Intact peptides were detected in the Orbitrap at a resolution of 30,000. Internal calibration was performed using the ion signal of  $(\text{Si}(\text{CH}_3)_2\text{O})_6\text{H}^+$  at  $m/z$  445.120025 as a lock mass (36). We used a data-dependent procedure that alternated between one MS scan and six MS/MS scans for the six most abundant precursor ions in the MS survey scan. The  $m/z$  values selected for MS/MS were dynamically excluded for 180 s. The electrospray voltage applied was 1.8 kV. Both MS and MS/MS spectra were acquired using the one microscan with a maximum fill time of 1,000 and 100 ms for MS and MS/MS analyses, respectively. Automatic gain control was used to prevent overfilling of the ion trap, and  $5 \times 10^4$  ions were accumulated in the ion trap for the generation of MS/MS spectra. The  $m/z$  scan range for MS scans was 350–2,000 Da.

**Bioinformatics**—The resulting MS/MS spectra were used to search the non-redundant IPI human sequence database Version 3.26 (released February 2007; 67,665 sequences; 28,353,548 residues) of the European Bioinformatics Institute using the SEQUEST algorithm (Thermo Fisher). Up to two missed cleavages were allowed, and searches were performed with variable oxidation of methionine residues (16 Da) and fixed modification for carbamidomethylcysteines (57 Da). A fragment ion mass tolerance of 0.5 Da and a parent ion mass tolerance of 10 ppm were used for the search engine with trypsin as the digestion enzyme. The random sequence database was used to estimate false-positive rates for peptide matches, and the false-positive rate for the peptide sequence matches using the criteria was estimated to be <1% via random database searching.

Protein identities were validated using the open source TPP software (Version 3.3). The SEQUEST search resulted in a DTA file. The raw data and DTA files containing information about identified peptides were then processed and analyzed in the TPP. The TPP software includes a peptide probability score program, PeptideProphet, that aids in the assignment of peptide MS spectra (37), as well as a ProteinProphet program that assigns and groups peptides to a unique protein or a protein family if the peptide is shared among several isoforms (38). ProteinProphet allows for the filtering of large scale data sets with assessment of predictable sensitivity and false-positive identification error rates. We used PeptideProphet and ProteinProphet probability scores  $\geq 0.95$  to ensure an overall false-positive rate below 0.5%. Furthermore, proteins with single peptide identities were excluded from this study. Information about the

PeptideProphet and ProteinProphet programs can be obtained from the Seattle Proteome Center at Institute for Systems Biology.

We used the SignalP program with hidden Markov models to predict the presence of secretory signal peptide sequences (39, 40). In addition, we used the SecretomeP program to predict non-signal peptide-triggered protein secretion (41) and the TMHMM to predict transmembrane helices in proteins (42). The identified proteins were further analyzed using ProteinCenter (Proxeon Bioinformatics, Odense, Denmark), a proteomics data mining and management software, to compare cell line secretomes with each other, functionally categorize the identified proteins, and calculate the emPAI (43, 44).

**Hierarchical Clustering**—The emPAI values of identified proteins were imported into Microsoft Excel. If a protein was identified in one cell line but not the other, half the minimum emPAI value from the data set was assigned to that protein to facilitate visualization and comparison. All values were then transformed to Z scores, a commonly used normalization method for microarray data (45). The Z scores were calculated as  $Z = (X - \mu_x)/\sigma_x$  where  $X$  is the individual emPAI value,  $\mu_x$  is the mean of emPAI values for a identified protein across cell lines, and  $\sigma_x$  is the standard deviation associated with  $\mu_x$ . A spreadsheet containing the Z scores was uploaded to the Partek® Genome Suite (Partek Inc., St. Louis, MO) and analyzed using a two-way hierarchical clustering algorithm according to Pearson distance and Ward's aggregation method. Cell lines and proteins were organized into mock phylogenetic trees (dendrograms) with the cell lines shown along the  $x$  axis and the proteins along the  $y$  axis.

**Network Analysis**—Proteins selected from the clustering analysis were converted into gene symbols and uploaded into MetaCore (GeneGo, St. Joseph, MI) for biological network building. MetaCore consists of curated protein interaction networks based on manually annotated and regularly updated databases. The databases describe millions of relationships between proteins according to publications on proteins and small molecules. The relationships include direct protein interactions, transcriptional regulation, binding, enzyme-substrate interactions, and other structural or functional relationships. The “shortest paths” and “analyze network” algorithms were used to map the hypothetical networks of uploaded proteins. The relevant pathway maps were then prioritized based on their statistical significance with respect to the uploaded data sets.

**Patient Population and Clinical Specimens**—Plasma samples were collected from 45 healthy controls (*i.e.* 32 men and 13 women ranging in age from 43 to 77 years; mean, 62.2 years), 44 patients with liver cancer (*i.e.* 32 men and 13 women ranging in age from 44 to 77 years; mean, 63.4 years), and 44 patients with lung cancer (*i.e.* 28 men and 16 women ranging in age from 32 to 88 years; mean, 64.7 years). Serum samples were collected from 45 healthy controls (*i.e.* 26 men and 19 women ranging in age from 21 to 72 years; mean, 48.2 years) and 45 NPC patients (*i.e.* 32 men and 13 women ranging in age from 16 to 79 years; mean, 47.2 years). All the blood samples were collected at Chang Gung Memorial Hospital. The study protocol was approved by the Medical Ethics and Human Clinical Trial Committee at Chang Gung Memorial Hospital. All patients entered in the study signed an informed consent.

Blood samples were collected from the patients preoperatively following a standardized protocol. Plasma and serum samples were prepared by collecting blood in EDTA and empty tubes, respectively, and left at room temperature (for a maximum of 30 min) until centrifugation. Plasma samples were centrifuged at  $2,000 \times g$  for 10 min at room temperature to pellet the cells. Serum samples were centrifuged at  $1,500 \times g$  for 10 min at 4 °C. After centrifugation, samples were divided into 1.0-ml aliquots in sterile cryotubes and immediately frozen at  $-80$  °C for storage until ELISAs. The samples had only undergone one freeze/thaw cycle before the measurements were conducted.

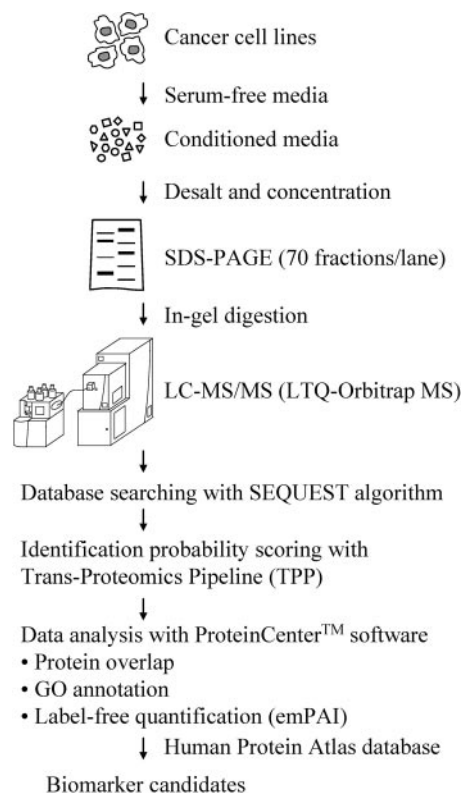
**Western Blot Analysis**—The prepared samples (20  $\mu\text{g}$  of protein) were separated by SDS-PAGE, transferred to PVDF membranes (Millipore), and probed with various antibodies (*i.e.* anti-fascin, anti-BIGH3, anti-PAI-1 (Santa Cruz Biotechnology), and anti- $\beta$ -tubulin (MDBio, Taipei, Taiwan)) as described previously (20, 27). Polyclonal antibodies specific to prosaposin were produced in rabbits using recombinant proteins, and these antibodies were affinity-purified as described in the supplemental Materials and Methods. Proteins of interest were detected with alkaline phosphatase-conjugated goat anti-rabbit IgG antibodies (Santa Cruz Biotechnology) and visualized using the CDP-Star<sup>TM</sup> chemiluminescent substrate (Roche Applied Science) according to the manufacturer's protocol.

**ELISA**—The concentrations of four candidate proteins were measured by ELISA in the blood samples of healthy controls and cancer patients. The concentrations of CD14 (ELISA kit from R&D Systems, Minneapolis, MN), stromal cell-derived factor 1 (SDF-1) (ELISA kit from R&D Systems), and cathepsin L1 (ELISA kit from Bender Med-Systems) were measured according to their respective manufacturer's instructions. An ELISA developed in house was used to measure ISG15 as described in the supplemental Materials and Methods.

**Statistical Analysis**—For the analysis of ELISA results, continuous measures were summarized using means, standard deviations, medians, and interquartile ranges. Differences between controls and cancer patients in blood concentrations of CD14, SDF-1, cathepsin L1, and ISG15 were performed using the nonparametric Mann-Whitney *U* test. Statistical analyses were conducted using SPSS software (Version 13.0, SPSS Inc., Chicago, IL). Two-tailed *p* values of 0.05 or less were considered significant.

## RESULTS

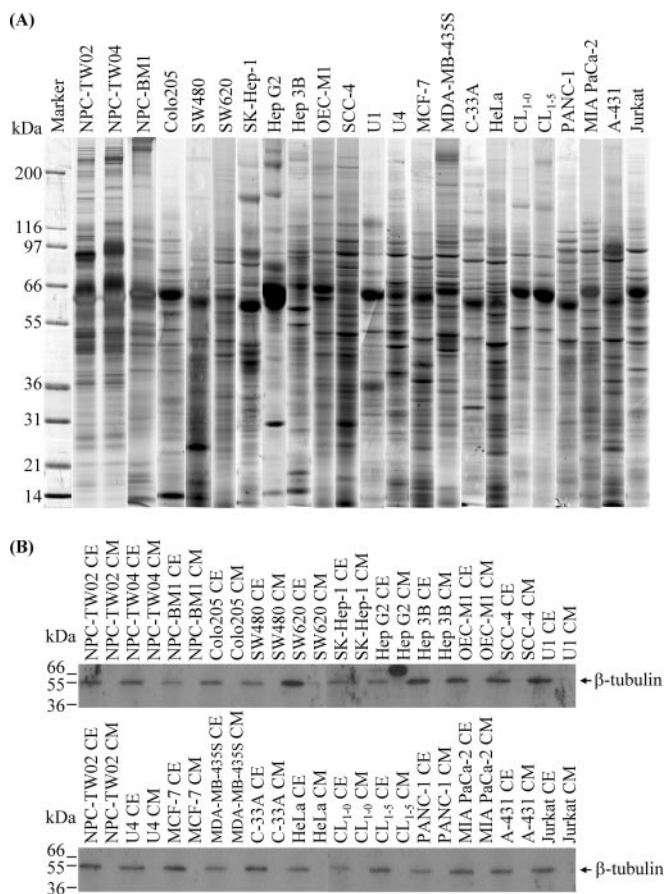
**Secretomes of 23 Cancer Cell Lines**—We previously examined the secretomes of 21 cancer cell lines using one-dimensional SDS gels and MALDI-TOF MS and identified an average of 38 secreted proteins in each cell line (25). More recently, we identified 1,096 and 1,830 proteins in the secretomes of two lung cancer cell lines (CL<sub>1-0</sub> and CL<sub>1-5</sub>, respectively) using the GeLC-MS/MS strategy (46). To facilitate the secretome-based discovery of cancer biomarkers, we have now used the GeLC-MS/MS strategy to perform an in-depth analysis of conditioned media from 21 cancer cell lines derived from 10 cancer types, including NPC (NPC-TW02, NPC-TW04, and NPC-BM1), colon cancer (Colo205, SW480, and SW620), HCC (SK-Hep-1, Hep G2, and Hep 3B), oral cancer (OEC-M1 and SCC-4), bladder cancer (U1 and U4), breast cancer (MCF7 and MDA-MB-435S), cervical cancer (C-33A and HeLa), pancreatic cancer (PANC-1 and MIA-PaCa-2), epidermoid carcinoma (A431), and T cell lymphoma (Jurkat). A schematic diagram of the procedure is shown in Fig. 1. Proteins (50  $\mu\text{g}$ ) in the conditioned media collected from cells cultured in serum-free media for 24 h were resolved by SDS-PAGE, visualized by Coomassie Blue staining, consecutively sliced into 70 pieces, digested individually with trypsin, and analyzed by LC-ESI-MS/MS on an LTQ-Orbitrap MS. The protein patterns observed in conditioned media from 21 cancer cell lines and two lung cancer cell lines are shown in Fig. 2A. As a quality control, we performed Western blot to examine the distribution of  $\beta$ -tubulin, an abundant cytosolic protein, between the total cell lysates and conditioned media.



**FIG. 1. Cancer cell secretome profiling via analysis of conditioned media from cancer cell lines.** The strategy involves the proteomic profiling of cancer cell-conditioned media, secretome analysis, and subsequent preliminary validation using the HPA.

$\beta$ -Tubulin was clearly detected in the total cell extracts but not in the conditioned media (Fig. 2B). We also found that serum starvation for 24 h had little effect on cell viability compared with cells cultured in the presence of 10% serum (supplemental Fig. 1). These observations collectively indicate that recovery of proteins in the conditioned media was not due to cell death.

The resulting MS/MS spectra were used in a search of the non-redundant IPI human sequence database (Version 3.26) using the SEQUEST algorithm. The search results were analyzed using the open source TPP software (Version 3.3) with stringent criteria regarding peptide probability ( $\geq 0.95$ ) and protein probability ( $\geq 0.95$ ) (38). Proteins present in the two lung cancer cell lines (46) as well as proteins identified by this analysis are summarized in Table I. Emphasis was placed on proteins identified by multiple (*i.e.* at least two) peptides because the chances of false-positive results decrease exponentially with each additional peptide identified (47). After setting the cutoffs, an average of 1,356 proteins per cell line were examined, and a total of 31,180 protein identities, accounting for 4,584 non-redundant proteins, were detected (Table I and supplemental Table 1). In addition to identifying proteins in each cancer cell secretome, we applied the emPAI (43) to estimate the abundance of each protein in the secretome of each cell line (supplemental Table 1). The false dis-



**FIG. 2. SDS-PAGE analysis of conditioned media harvested from cancer cells.** A, conditioned media from cancer cells were collected and processed as described under “Experimental Procedures.” Proteins (50  $\mu$ g) were resolved on 8–14% gradient SDS gels and stained with Coomassie Blue. Protein bands were excised for further analysis. B, proteins (40  $\mu$ g) in the conditioned media (CM) and cell extracts (CE) were analyzed via Western blotting using an anti- $\beta$ -tubulin antibody.

covery rate (FDR) of peptide detection was empirically determined by searching the data set against a random IPI Human database (Version 3.26) using the same search parameters and TPP cutoffs. The FDRs determined for each cell line are shown in Table I; all were  $<1\%$ .

**Distribution and Ontology Analysis of Identified Proteins—**The identified proteins were further analyzed using bioinformatics programs designed to predict protein secretion pathways (Table II and supplemental Table 2). Among the 4,584 non-redundant proteins identified, the SignalP program predicted that 998 proteins were secreted in the classical secretory pathway (*i.e.* the endoplasmic reticulum/Golgi-dependent pathway; SignalP probability  $\geq 0.90$ ) based on the presence of a signal peptide (39, 40). The SecretomeP program predicted that 1,438 proteins were released via the nonclassical secretory pathway (SignalP probability  $< 0.90$  and SecretomeP score  $\geq 0.50$ ) (41). In addition, the TMHMM determined that 121 integral membrane proteins were not

**TABLE I**  
Number of identified proteins in secretome of each cancer cell line

Cancer type and cell line	No. of identified proteins <sup>a</sup>	FDR <sup>b</sup>
		%
<b>NPC</b>		
NPC-TW02	841	0.99
NPC-TW04	1,067	0.47
NPC-BM1	1,129	0.72
<b>CRC</b>		
Colo205	1,322	0.39
SW480	857	0.41
SW620	1,440	0.61
<b>HCC</b>		
SK-Hep-1	954	0.42
HepG2	775	0.62
Hep-3B	1,819	0.31
<b>Oral cancer</b>		
OEC-M1	1,343	0.59
SCC-4	1,506	0.71
<b>Bladder cancer</b>		
U1	1,071	0.51
U4	2,142	0.50
<b>Breast cancer</b>		
MCF-7	1,318	0.81
MDA-MB-4355	1,780	0.34
<b>Cervix cancer</b>		
C-33A	1,284	0.92
HeLa	1,223	0.82
<b>Lung adenocarcinoma</b>		
CL <sub>1-0</sub>	1,096	0.70
CL <sub>1-5</sub>	1,830	0.26
<b>Pancreatic ductal adenocarcinoma</b>		
PANC-1	1,035	0.37
MIA-PaCa-2	1,609	0.31
<b>Epidermoid carcinoma</b>		
A-431	1,458	0.51
<b>T cell lymphoma</b>		
Jurkat	2,281	0.41
All 23 cell lines	4,584	

<sup>a</sup> The number of identified proteins that contained at least two peptide hits.

<sup>b</sup> The FDR was calculated from the spectra assigned to random database over normal database.

secreted via the classical or nonclassical secretion pathways (42). The predicted secretion pathways of the proteins in each cell line are summarized in Table II and supplemental Table 1. Collectively, these analyses predicted that 55.8% (2,557 of 4,584) of the identified proteins were released into the conditioned media of cultured cancer cells via different mechanisms. It should be noted that many chemokines, cytokines, and growth factors, which are known as very low abundance secreted proteins, could be detected in the secretomes of various cancer cell lines (supplemental Table 3), thereby demonstrating the sensitivity of the GeLC-MS/MS strategy. To evaluate the effectiveness of this protocol with regard to secretome analysis, we analyzed proteins extracted from lysates of NPC-TW04 and A431 cells that remained on culture dishes after the removal of conditioned media. The results

TABLE II  
 Predicted secretion pathways of proteins identified in conditioned media from 23 cancer cell lines

Cell line	No. of identified proteins				Percentage of predicted secreted proteins
	Classical secretion <sup>a</sup>	Nonclassical secretion <sup>b</sup>	Membrane protein <sup>c</sup>	Others <sup>d</sup>	
NPC-TW02	255	235	14	337	59.9
NPC-TW04	284	304	16	463	56.6
NPC-BM1	292	324	17	496	56.1
Colo205	295	430	34	563	57.4
SW480	213	250	17	377	56.0
SW620	229	468	16	727	49.5
SK-Hep-1	269	293	16	376	60.6
HepG2	267	206	12	290	62.6
Hep-3B	383	573	39	824	54.7
OEC-M1	333	422	29	559	58.4
SCC-4	209	487	27	783	48.0
U1	280	305	18	468	56.3
U4	364	687	44	1,047	51.1
MCF-7	219	390	23	686	48.0
MDA-MB-435S	253	589	17	921	48.3
C-33A	285	384	25	590	54.0
HeLa	243	350	9	621	49.2
CL <sub>1-0</sub>	174	369	14	539	50.8
CL <sub>1-5</sub>	364	621	41	804	56.1
PANC-1	224	319	19	473	54.3
MIA-PaCa-2	266	522	20	801	50.2
A-431	299	476	23	660	54.7
Jurkat	195	796	21	1,269	44.4
All 23 cell lines	998	1,438	121	2,027	55.8

<sup>a</sup> Proteins predicted by the SignalP program to be secreted via the classical secretion pathway (SignalP probability  $\geq 0.90$ ).

<sup>b</sup> Proteins predicted to be secreted by the nonclassical secretion pathway using SignalP and SecretomeP (SignalP probability  $< 0.90$  and SecretomeP score  $\geq 0.50$ ).

<sup>c</sup> Proteins predicted by the TMHMM to form integral membrane proteins that were not predicted to be secreted via the classical or nonclassical secretion pathways.

<sup>d</sup> Proteins that could not be classified as classical secreted, nonclassical secreted, or integral membrane proteins.

showed that only 34.0% (415 of 1,219) and 33.8% (395 of 1,169) of the proteins in NPC-TW04 and A431 cells, respectively, were predicted to be secreted (data not shown). Of the 4,584 proteins identified in this report, 1,241 (27.1%) were found in the Human Plasma Proteome Project database (48) (supplemental Table 2).

ProteinCenter software was used to predict the functions of the 4,584 identified proteins based on universal GO annotation terms. These proteins were linked to at least one annotation term within the GO molecular function and biological process categories, respectively. As shown in Fig. 3A, the top three most common molecular functions were protein binding (63.4%), catalytic activity (61.3%), and metal ion binding (30.6%). The major biological process categories included metabolic processes (73.8%) and regulation of biological processes (34.5%) followed by cell organization (33.7%) and cell communication (26.8%) (Fig. 3B). The results of our GO analysis of identified proteins in the molecular function and biological process categories are shown in supplemental Tables 4 and 5, respectively.

**Overlap of Identified Proteins between All Cell Lines Examined**—The proteins identified among the 23 cell lines were analyzed for overlapping members (Table III and supplemental Table 2). One hundred and seventy-two proteins (3.8% of the 4,584 proteins) were detected in all cancer cell

secretomes. About 23.0% of the 4,584 proteins were detected in more than half ( $\geq 12$ ) of the cell lines, and 35.1% were found in 3–11 cell lines. Nearly one-third (*i.e.* 29.3%) of the 4,584 proteins were uniquely detected in the secretome of one of the 23 cell lines, and 12.6% (576 proteins) were identified in two of the 23 cell lines.

To reduce the number of potential tumor marker candidates, we combined proteins identified in the secretomes of cell lines from each cancer type to form a list of non-redundant proteins for each cancer type. These 11 lists were used to assess the overlap in identified proteins (Table IV and supplemental Table 6). A significant portion (36.3%) of the proteins were found in more than half (at least six) of the cancer types; 33.6% (1,539 proteins) were detected in two to five cancer types, and 30.1% (1,381 proteins) were detected in a single cancer type. Taken together, these data reveal that cell lines from different tumor origins secrete/release hundreds of common proteins and that cancer cell lines can also secrete/release proteins unique to a specific cancer type.

**Evaluation of Potential Cancer-specific Biomarkers**—Detection of proteins that are uniquely released by each cancer type might facilitate the discovery of biomarkers for individual cancers. Thus, we focused our attention on the 1,381 proteins that were uniquely detected in the secretomes of a specific

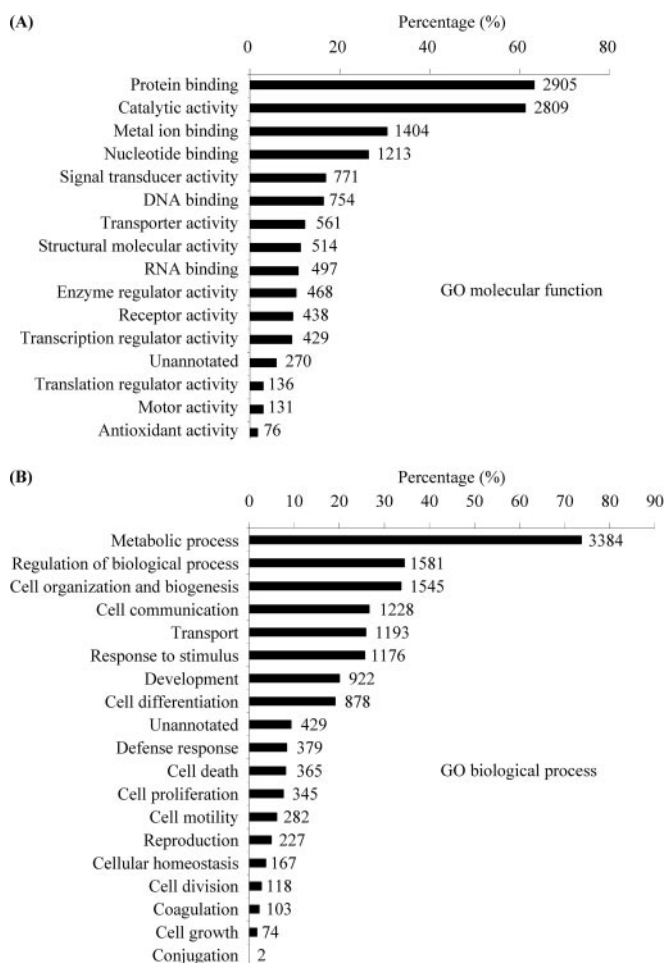


FIG. 3. **Functional classification of proteins identified in conditioned media using ProteinCenter software based on universal GO annotation terms.** The proteins were linked to at least one annotation term within the GO molecular function (A) and biological process (B) categories. The numbers represent the proteins annotated as each GO term.

cancer type. To efficiently narrow down our candidate list of potential cancer-specific biomarkers, we consulted the HPA. This database contains the immunohistochemical (IHC) staining profiles of numerous proteins in a variety of cancerous and non-cancerous tissues based on more than 8,800 antibodies (35). We searched all 1,381 proteins in the HPA database and selected those whose expression has been examined in corresponding cancer tissues from a small number of patients. The IHC staining profiles of corresponding non-cancer tissues in the HPA were also analyzed, although only three or fewer than three biopsies were available (supplemental Table 7). We found that 603 of 1,381 proteins have been examined in their corresponding tumor tissues (Table V). Among these, 77.8% (469) of the proteins were detected in more than 50% of the tumor tissue sections (Table V). The IHC staining results for the 603 proteins and their corresponding cancer types from the HPA database are summarized in Table V and supplemental Table 7.

TABLE III  
Proteins detected in secretomes of 23 cancer cell lines

Detected in no. of cell lines	No. of proteins	Percentage	emPAI <sup>a</sup> (mean ± S.D.)
		%	
23	172	3.8	2.35 ± 1.98
22	102	2.2	1.61 ± 1.18
21	89	1.9	1.29 ± 1.07
20	79	1.7	1.09 ± 0.96
19	73	1.6	1.03 ± 1.23
18	80	1.7	1.02 ± 0.96
17	79	1.7	1.04 ± 0.89
16	68	1.5	0.99 ± 0.85
15	76	1.7	1.43 ± 3.67
14	79	1.7	1.03 ± 1.19
13	74	1.6	1.12 ± 2.26
12	83	1.8	0.85 ± 0.69
11	118	2.6	0.93 ± 0.92
10	91	2.0	1.16 ± 1.00
9	114	2.5	0.89 ± 1.31
8	144	3.1	0.89 ± 1.52
7	148	3.2	0.89 ± 0.87
6	163	3.6	0.88 ± 1.01
5	203	4.4	0.71 ± 0.87
4	269	5.9	0.68 ± 0.85
3	361	7.8	0.78 ± 1.15
2	576	12.6	0.68 ± 0.88
1	1,343	29.3	0.74 ± 1.72

<sup>a</sup> The average emPAI values of all identified proteins. The average emPAI value for each protein was determined by adding the emPAI values of each identified protein and dividing the sum by the number of cell lines in which the protein was detected. The average emPAI value shown here was determined by adding the average emPAI values of each protein and dividing the sum by the protein numbers.

TABLE IV  
Proteins detected in secretomes of cell lines derived from 11 cancer types

Detected no. of cancer types	No. of proteins	%
11	599	13.1
10	267	5.8
9	193	4.2
8	207	4.5
7	179	3.9
6	219	4.8
5	262	5.7
4	283	6.2
3	394	8.6
2	600	13.1
1	1,381	30.1

The following examples illustrate the ability of our analyses to identify multiple marker candidates that warrant further validation (Table VI). Among the 40 proteins detected in most CRC tissues (Table V), cell surface A33 antigen was found to be mainly negative in other cancer types, whereas neutral amino acid transporter A, isoform CSBP1 of mitogen-activated protein kinase 14, and bone morphogenetic protein 4 were overexpressed in CRC relative to other cancers. Among

TABLE V  
 Expression profiles of cancer type-specific secreted/released proteins in HPA

Cancer type	No. of unique proteins	No. of unique proteins available in HPA <sup>a</sup>	No. of unique proteins detected in over half of tumor tissues <sup>b</sup>
Nasopharyngeal carcinoma	28	9	6
Colorectal carcinoma	109	50	40
Hepatocellular carcinoma	196	88	70
Oral cancer	76	34	28
Bladder cancer	147	72	55
Breast cancer	153	60	47
Cervix cancer	85	33	23
Lung adenocarcinoma	109	53	36
Pancreatic ductal carcinoma	46	21	18
Epidermoid carcinoma	21	9	9
T cell lymphoma	411	174	137
All 11 cancer types	1,381	603	469

<sup>a</sup> Proteins that were solely detected in the secretome of a specific cancer type were immunohistochemically examined in the corresponding cancer tissue sections of the HPA.

<sup>b</sup> Proteins that were solely detected in the secretome of a specific cancer type were immunohistochemically detected in more than half of the tumor tissue sections examined in the HPA.

 TABLE VI  
 Partial list of expression profiles of cancer type-specific marker candidates in HPA

Accession no.	Protein name	HPA Ab <sup>a</sup>	IHC staining <sup>b</sup>				Positive rate %
			3+	2+	+	-	
Colon cancer							
IPI00293853.3	Cell surface A33 antigen	HPA018858	5	5	1	0	100.0
IPI00015476.1	Neutral amino acid transporter A	CAB002757	6	5	1	0	100.0
IPI00221141.3	Isoform CSBP1 of mitogen-activated protein kinase 14	CAB010285	0	5	6	1	91.7
IPI00292640.3	Bone morphogenetic protein 4	CAB006863	3	1	3	4	63.6
Liver cancer							
IPI00216133.5	Bile salt sulfotransferase	CAB018755	8	0	2	1	90.9
IPI00295363.2	Ornithine carbamoyltransferase	HPA000243	8	1	0	2	81.8
IPI00029260.2	Monocyte differentiation antigen CD14	HPA002127	0	5	5	0	100.0
IPI00011155.5	Isoform 1 of asialoglycoprotein receptor 2	HPA015998	6	0	1	5	58.3
IPI00027481.1	Multidrug resistance protein 1	CAB001716	1	6	3	0	100.0
IPI00021817.1	Vitamin K-dependent protein C	HPA005550	0	5	2	3	70.0
Bladder cancer							
IPI00024035.1	Isoform 1 of cadherin-6	HPA007456	0	5	6	0	100.0
IPI00020944.1	Squalene synthetase	HPA008874	4	5	2	1	91.7
IPI00028635.4	Ribophorin II	HPA008297	11	0	0	0	100.0
IPI00305286.1	15-Hydroxyprostaglandin dehydrogenase	HPA004919	8	3	1	0	100.0
Breast cancer							
IPI00029819.3	Neurogenic locus notch homolog protein 3	CAB005393	3	6	3	0	100.0
IPI00022283.1	Trefoil factor 1	HPA003425	11	0	0	1	91.7
Lung cancer							
IPI00216304.1	Stromal cell-derived factor 1	CAB017564	0	4	3	3	70.0

<sup>a</sup> Identification number of antibodies in the HPA.

<sup>b</sup> Immunohistochemical staining scores in the HPA: 3+, strong staining; 2+, moderate staining; +, weak staining; -, negative staining.

the 70 proteins positively stained in most HCC tissues (Table V), bile salt sulfotransferase, ornithine carbamoyltransferase, monocyte differentiation antigen CD14, and isoform 1 of asialoglycoprotein receptor 2 were less immunoreactive in tissues of other cancer types, whereas multidrug resistance protein 1 and vitamin K-dependent protein C were overexpressed in HCC *versus* other cancers. Compared with other cancers, bladder cancer tissues reacted more strongly with proteins such as cadherin-6, squalene synthetase, ribophorin II, and 15-hydroxyprostaglandin dehydrogenase. The levels of neu-

rogenic locus notch homolog protein 3 and trefoil factor 1 were higher in breast cancer tissues *versus* tissues of other cancers. Compared with non-tumor lung tissues, stromal cell-derived factor 1 (CXCL12) reacted more strongly with lung cancer tissues.

**Evaluation of Potential Pan-cancer Biomarkers**—The identification of proteins that have a high frequency of detection in the conditioned media of different cell lines clearly revealed proteins commonly released by various cancer cells. A total of 172 proteins (3.8%) were found in the conditioned media of all



cell lines examined (Table III and supplemental Table 8). To evaluate the potential of these proteins to serve as pan-cancer marker candidates, we evaluated their expression in the tumor tissues of nine cancer types in the HPA database, including breast, cervix, colon, head and neck, liver, lung, pancreas, skin, and bladder cancers (35). In the HPA database (Version 5.0), 114 of the 172 proteins had been analyzed by IHC staining (supplemental Table 8). Among the proteins detected in more than half of the tumor tissue sections, 70.2% (80 of 114) of the proteins were observed in all nine tumor types, and 12.3% (14 of 114) of the proteins were detected in eight of nine cancer types (supplemental Table 8). Moreover, 45 proteins were detected in human plasma samples as documented in the Human Plasma Proteome Project (supplemental Table 8), and eight of 45 proteins showed negative or weak staining in over half of the nine corresponding normal tissue types (Table VII). These observations suggest that secreted proteins common to multiple cancer cell lines are potential pan-cancer markers.

**Hierarchical Clustering Analysis for Pathway-based Biomarker Searches**—A new approach toward biomarker discovery was recently proposed wherein pathways are monitored and targeted rather than individual proteins (49, 50). Many secreted proteins appear to play important roles in the cancer microenvironment (51); thus, we attempted to cluster proteins according to their abundance in the conditioned media from each cancer cell line in an effort to identify potential pathways involved in the regulation of cancer microenvironments. Toward this end, we calculated the empAI values of all proteins identified in the conditioned media of 23 cell lines, transformed these values to Z scores, and analyzed these data via unsupervised hierarchical classification as described under “Experimental Procedures.” To examine the ability of empAI-based Z scores to calculate the relative abundance of proteins in the conditioned media, we compared the Z score values of four selected targets (*i.e.* BIGH3, fascin, PAI-1, and prosaposin) with their corresponding signal intensities as determined by Western blot analyses of conditioned media (supplemental Fig. 2). There was a significant correlation between empAI-based Z scores and Western blot-based Z scores, suggesting that empAI-based Z scores can be used to estimate the relative abundance of proteins in conditioned media.

When proteins detected in the conditioned media were clustered according to Z scores, the three NPC cell lines and the two lung cancer cell lines clustered together. However, the other cell lines could not be categorized by tissue type (Fig. 4A and supplemental Fig. 3). We further selected the 79 proteins with the most different features used to distinguish the NPC cell lines (Fig. 4B and supplemental Table 9).

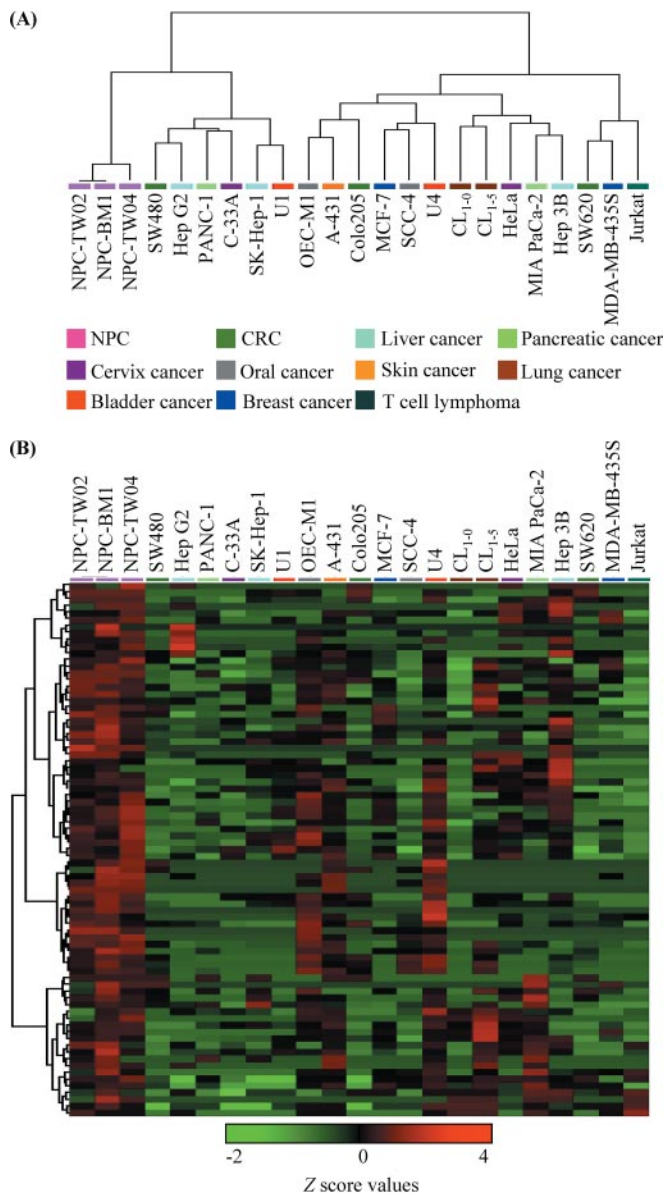
We then used MetaCore software to build biological networks and analyze the possible biological linkages between these 79 proteins. When the shortest paths algorithm was used to map the shortest interaction path and obtain a global view, 36 proteins were brought together in the net-

TABLE VII  
Partial list of proteins identified in secretome of all examined cell lines and their expression profiles in HPA

Accession no.	Description	HPA Ab <sup>a</sup>	IHC results (no. of positive staining tumor tissues/no. of total staining tumor tissues/intensity in corresponding normal tissues) <sup>b</sup>								
			Breast cancer	Cervix cancer	Colon cancer	Head and neck cancer	Liver cancer	Lung cancer	Pancreatic cancer	Skin cancer	Bladder cancer
IP100219018.7	Glyceraldehyde-3-phosphate dehydrogenase	CAB005197	12/12/+	10/10/+	12/12/+	4/4/+	11/11/2+	11/11/2+	12/12/+	10/10/+	11/11/2+
IP100006608.1	Isoform APP770 of amyloid $\beta$ A4 protein	HPA001462	12/12/2+	11/11/2+	11/11/2+	3/3/-	8/12/-	11/12/+	12/12/+	11/11/+	11/11/+
IP100000874.1	Peroxiredoxin-1	CAB004682	12/12/2+	12/12/+	11/12/2+	3/3/2+	9/10/+	12/12/+	11/12/+	12/12/+	12/12/2+
IP100000875.6	Elongation factor 1- $\gamma$	CAB004969	12/12/+	10/11/+	9/11/+	4/4/+	6/10/2+	9/11/2+	7/8/2+	8/11/+	12/12/2+
IP100299086.3	Syntenin-1	CAB012245	10/12/+	9/12/-	12/12/3+	4/4/-	10/12/2+	9/10/+	9/11/+	7/12/2+	9/12/2+
IP100032293.1	Cystatin C precursor	HPA013143	11/12/-	11/12/+	12/12/3+	3/4/2+	6/11/+	11/12/-	11/12/2+	9/11/+	12/12/2+
IP100027497.5	Glucose-6-phosphate isomerase	HPA024305	12/12/+	11/12/NA	12/12/2+	3/3/+	11/12/2+	9/10/-	8/11/-	6/10/+	11/11/2+
IP100027235.1	Isoform 1 of attractin precursor	HPA008853	11/12/2+	9/12/+	10/12/2+	4/4/2+	10/12/+	12/12/+	10/12/-	5/12/+	11/12/2+

<sup>a</sup> Identification number of antibodies in the HPA.

<sup>b</sup> Immunohistochemical staining scores in the HPA: 3+, strong staining; 2+, moderate staining; +, weak staining; -, negative staining; NA, not available.



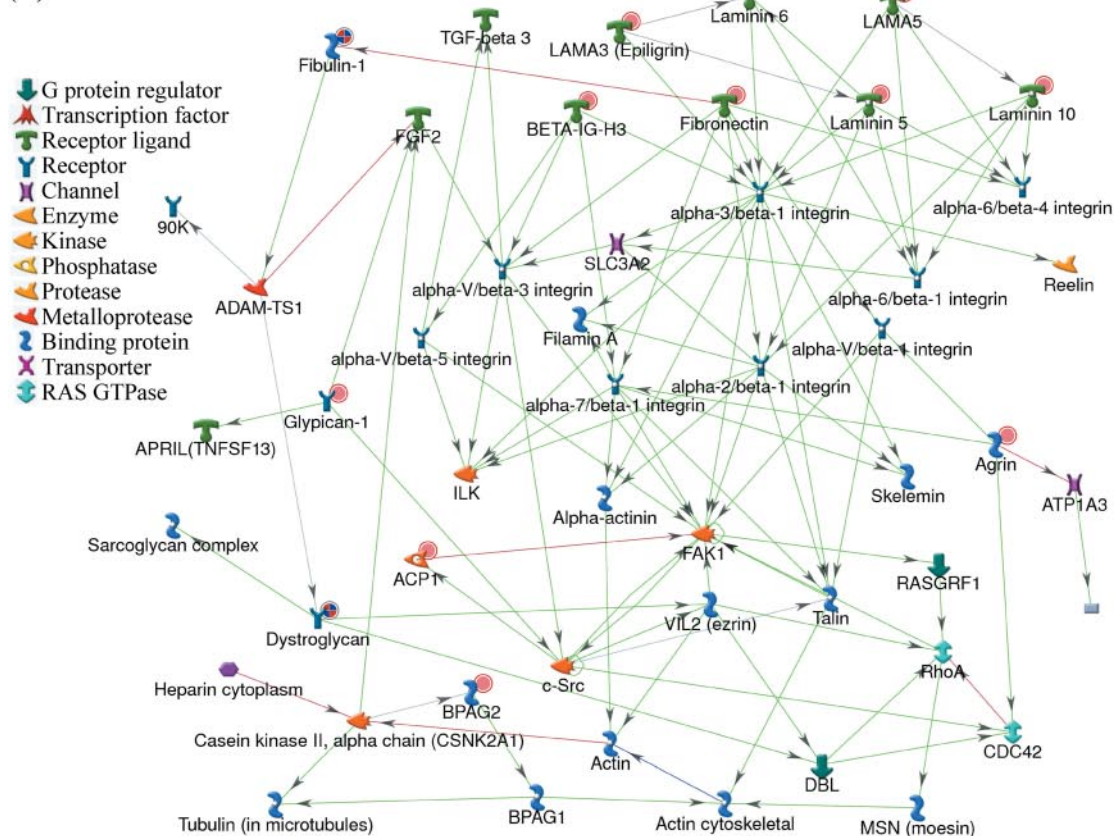
**FIG. 4. Hierarchical clustering of cancer cell lines by secreted proteins.** The emPAI values of all identified proteins were transformed to Z scores and analyzed via unsupervised hierarchical classification. *A*, hierarchical classification according to a distance tree constructed from all identified proteins. *B*, clustering analysis of the cell lines by 79 selected proteins, which possessed unique features used to sort the NPC cell line from others. Cell lines are shown in columns, and proteins are shown in rows. The heat map scale of Z scores ranges from  $-2$  (green) to  $4$  (red) with a midpoint of  $0$  (black).

works (supplemental Fig. 4). The networks shown in supplemental Fig. 4 demonstrate the enormous number of complex interactions between the 36 identified proteins and various intracellular signaling proteins. The 79 proteins were analyzed using the MetaCore analyze network algorithm to further explore their involvement in various biological processes. This analysis revealed a significant number of networks involved in cell adhesion and migration (Fig. 5A;  $p = 2.5 \times$

$10^{-26}$ ) and immune system regulation (Fig. 5B;  $p = 5.7 \times 10^{-22}$ ). The 24 proteins involved in both networks are listed in Table VIII. Among them, fibronectin is a potential NPC serum biomarker (20), laminin subunit  $\gamma$ -1 is overexpressed in NPC via the down-regulation of *mir-29c* microRNA (52), and cathepsin L is highly expressed in NPC, and its overexpression correlates with lymph node metastasis and distant metastasis (53). These observations support the feasibility of a pathway-based search strategy for biomarker discovery and furthermore suggest that the 22 additional proteins in the networks described above are potential NPC biomarkers that warrant further investigation.

*Validation of Monocyte Differentiation Antigen CD14, Stromal Cell-derived Factor 1, Cathepsin L1, and Interferon-induced 17-kDa Protein as Potential Serological Cancer Biomarkers*—To determine the clinical relevance of the results described above, we used ELISA to detect the levels of a potential liver cancer marker known as monocyte differentiation antigen CD14 (Table VI), a potential lung cancer marker known as SDF-1 (or CXCL12) (Table VI), and two potential NPC markers (*i.e.* cathepsin L1 and interferon-induced 17-kDa protein (ISG15)) (Table VIII) in serum or plasma samples from cancer patients and healthy controls. The CD14 and SDF-1 markers were selected based on a combined analysis of secretomes from 23 cell lines and the HPA, whereas cathepsin L1 and ISG15 were selected via the pathway-based strategy. In our data set, CD14 was selectively detected in the secretome of HepG2 cell line on the basis of three tryptic peptides (*i.e.* AFPALTSLDLSDNPGLGER, LTVGAAQVPAQLLVGALR, and TGTMPPLPLEATGLALSSLR). In the HPA database, expression of CD14 (detected using the HPA002127 antibody) was observed in different cell types in a variety of normal human tissues but not in bile duct cells or hepatocytes in normal liver tissue. Interestingly, positive CD14 staining was observed at a much higher rate (*i.e.* nine of 11 times) in HCC specimens than in the other 19 cancer types according to data obtained from the HPA database (supplemental Fig. 5). The SDF-1 marker was selectively detected in the CL<sub>1-5</sub> secretome based on the presence of two tryptic peptides (FFESHVAR and ILNTPNCALQIVAR). In the HPA database, SDF-1 expression (detected using the CAB017564 antibody) was found in different cell types in a variety of normal human tissues, including macrophages in the lung, but was not found in lung alveolar cells. Positive staining of SDF-1 was observed in many cancer types, including seven of 10 lung cancer specimens, according to data obtained from the HPA database (supplemental Fig. 6). As shown in Fig. 6A, the plasma levels of CD14 were statistically higher in patients with liver cancer ( $n = 44$ ) than in healthy controls ( $n = 45$ ) ( $1.83 \pm 0.36$  and  $1.43 \pm 0.29$   $\mu\text{g/ml}$ , respectively;  $p = 0.0117$ ). The SDF-1 plasma levels were significantly higher in patients with lung cancer ( $n = 44$ ) than in healthy controls ( $n = 44$ ) ( $4.01 \pm 1.57$  and  $3.11 \pm 0.56$   $\text{ng/ml}$ , respectively;  $p = 0.0007$ ) (Fig. 6B). Finally, the serum levels of

(A)



(B)

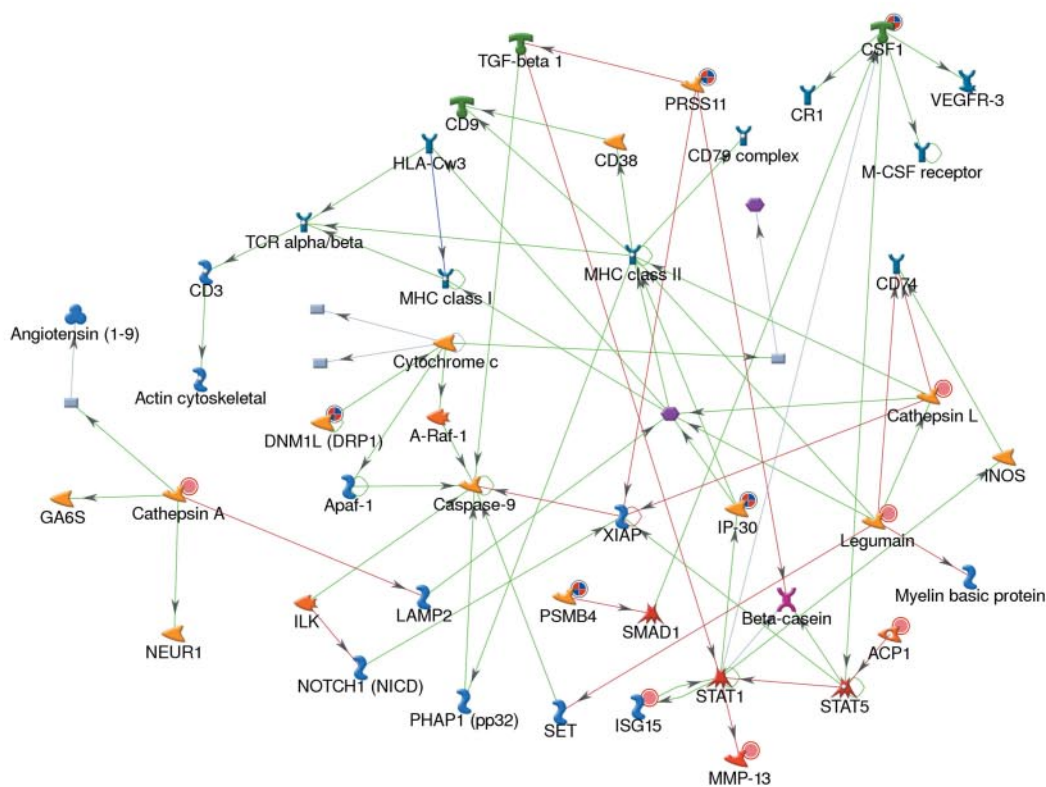


TABLE VIII  
Proteins involved in cell adhesion and immune system regulation

Accession no.	Gene symbol	Protein name
Cell adhesion <sup>a</sup>		
IPI00218847.5	<i>ACP1</i>	Low molecular weight phosphotyrosine protein phosphatase
IPI00374563.3	<i>AGRN</i>	Agtrin
IPI00398272.4	<i>COL17A1</i>	Collagen $\alpha$ -1(XVII) chain
IPI00028911.1	<i>DAG1</i>	Dystroglycan
IPI00218803.2	<i>FBLN1</i>	Fibulin-1
IPI00414283.3	<i>FN1</i>	Fibronectin
IPI00015688.1	<i>GPC1</i>	Glypican-1
IPI00377045.3	<i>LAMA3</i>	Laminin subunit $\alpha$ -3
IPI00783665.2	<i>LAMA5</i>	Laminin subunit $\alpha$ -5
IPI00013976.3	<i>LAMB1</i>	Laminin subunit $\beta$ -1
IPI00299404.1	<i>LAMB3</i>	Laminin subunit $\beta$ -3
IPI00298281.3	<i>LAMC1</i>	Laminin subunit $\gamma$ -1
IPI00015117.2	<i>LAMC2</i>	Laminin subunit $\gamma$ -2
IPI00018219.1	<i>TGFB1</i>	Transforming growth factor- $\beta$ -induced protein ig-h3
Immune system regulation <sup>a</sup>		
IPI00218847.5	<i>ACP1</i>	Low molecular weight phosphotyrosine protein phosphatase
IPI00015881.1	<i>CSF1</i>	Macrophage colony-stimulating factor 1
IPI00021794.8	<i>CTSA</i>	Lysosomal protective protein
IPI00012887.1	<i>CTSL1</i>	Cathepsin L1
IPI00037283.3	<i>DNM1L</i>	Dynamin-1-like protein
IPI00003176.1	<i>HTRA1</i>	Serine protease HTRA1
IPI00007853.1	<i>IFI30</i>	$\gamma$ -Interferon-inducible lysosomal thiol reductase
IPI00375631.6	<i>ISG15</i>	Interferon-induced 17-kDa protein
IPI00293303.1	<i>LGMM</i>	Legumain
IPI00021738.1	<i>MMP13</i>	Collagenase 3
IPI00555956.2	<i>PSMB4</i>	Proteasome subunit $\beta$ type-4

<sup>a</sup> Networks were constructed using the analyze network algorithm in MetaCore.

cathepsin L1 (Fig. 6C) and ISG15 (Fig. 6D) were elevated in patients with NPC relative to healthy controls ( $6.87 \pm 3.30$  and  $5.46 \pm 1.36$  ng/ml, respectively, for cathepsin L1;  $p = 0.0106$ ; and  $3.53 \pm 3.91$  and  $1.90 \pm 2.16$  ng/ml, respectively, for ISG15;  $p = 0.0184$ ).

#### DISCUSSION

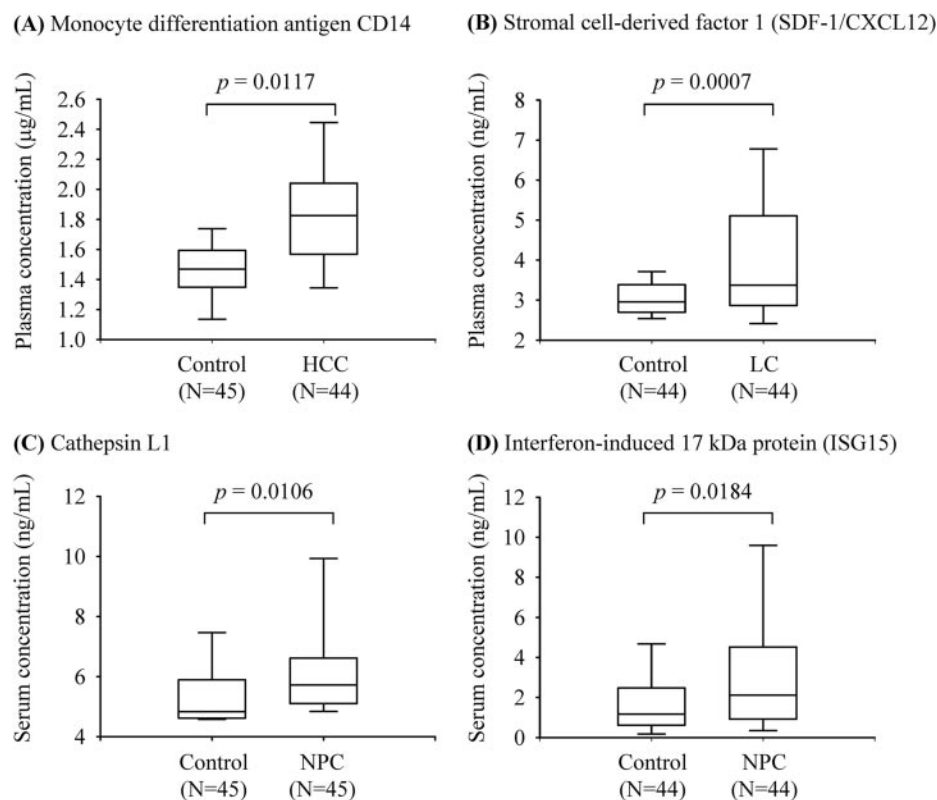
Although much time and effort have been devoted to the study of molecular alterations in cancer, early detection remains one of the most promising approaches to reducing the growing cancer burden. Thus, biomarkers capable of early detection will play a key role in the management and control of most, if not all, cancers in the future (54, 55). Cancer biomarkers should be

measurable in bodily fluids, especially in blood samples, to allow for the screening of large populations (11, 56). As numerous proteins have been found to exhibit altered levels in various cancer tissues, recent efforts have focused on compiling lists of general and cancer-specific cancer biomarker candidates, focusing on those with a higher chance of detection in bodily fluids (4, 57). Thus, the present study sought to conduct an in-depth analysis of the secretomes of 23 cancer cell lines and the HPA in an effort to construct a focused data set of serological cancer biomarker candidates.

Proteins detected in cancer cell secretomes include growth factors, proteases, cell motility factors, cytokines, chemokines, and/or cell surface receptors. These proteins play

**FIG. 5. Biological network analysis of NPC-related proteins.** The proteins in Fig. 4B were uploaded to the MetaCore mapping tool. The biological networks were generated using the analyze network algorithm. Two prominent networks involved in cell adhesion (A) and immune system regulation (B) were identified from the protein list. The *concentric circles* denote uploaded proteins. *Nodes* represent proteins with *shapes* representing functional class. *Lines between the nodes* indicate direct protein-protein interactions. *Green, red, and gray lines* represent stimulatory, inhibitory, or unspecified interactions, respectively. 90K, tumor-associated antigen 90K/Mac-2 binding protein; ACP1, low molecular weight phosphotyrosine protein phosphatase; APRIL/TNFSF13, a proliferation-inducing ligand/tumor necrosis factor ligand superfamily member 13; BPAG1/2, bullous pemphigoid antigen 1/2; CR1, complement component (3b/4b) receptor 1; CSF1, colony stimulating factor 1; DBL, dachaete beadex lethal; DNM1L/DRP1, dynamin 1-like protein/dynamin related protein 1; FAK1, focal adhesion kinase 1; FGF2, fibroblast growth factor 2; GA6S, galactosamine (N-acetyl)-6-sulfate sulfatase; ILK, integrin-linked protein kinase; IP-30, interferon, gamma-inducible protein 30; LAMA2/3/5, laminin subunit  $\alpha$ -2/3/5; M-CSF, macrophage colony-stimulating factor; MMP-13, collagenase 3; PSMB4, proteasome subunit  $\beta$  type-4; RASGRF1, ras protein-specific guanine nucleotide-releasing factor 1; SLC3A2, solute carrier family 3, member 2; STAT1/5, signal transducer and activator of transcription 1/5; VEGFR3, vascular endothelial growth factor receptor 3; XIAP, X-linked inhibitor of apoptosis.

**FIG. 6. Validation of CD14, SDF-1, cathepsin L1, and ISG15 in serum/plasma samples.** The plasma levels of CD14 (A) and SDF-1 (B) in healthy controls (Control), liver cancer patients (HCC), and lung cancer patients (LC) were measured by sandwich ELISA. The serum levels of cathepsin L1 (C) and ISG15 (D) in healthy controls (Control) and NPC patients (NPC) were detected by sandwich ELISA. Data are presented as the upper and lower quartiles and range (box), the median value (horizontal line), and the middle 90% distribution (dashed line).



pivotal roles in tumor progression, invasion, metastasis, and/or angiogenesis by regulating cell-to-cell and cell-to-extracellular matrix interactions. More importantly, these secreted proteins are likely to enter bodily fluids and could therefore be measured using non-invasive diagnostic tests. Thus, the analysis of cancer cell secretomes derived *in vitro* may help identify potential serum tumor biomarkers (58, 59). Although the secretomes of cell lines derived from breast, ovarian, prostate, pancreatic, and lung cancers have been profiled individually (21–24, 28–30, 32, 51), no studies have compared the secretomes of cell lines from various cancer types or developed in-depth profiles of these proteins. We previously examined cancer cell secretomes using SDS-PAGE and MALDI-TOF MS (25). Now, the current study examines the secretomes of 23 cancer cell lines using the GeLC-MS/MS approach. We have generated, for the first time, a data set consisting of 31,180 protein identities, corresponding to 4,584 non-redundant proteins, with an approximate average of 1,300 proteins from each cell line and an error rate less than 1% (Table I and supplemental Table 1). We used this data set to compare the secretomes of cell lines from various cancer types, which were generated and analyzed using the same experimental conditions. Our analysis revealed numerous proteins that are commonly or selectively secreted by cell lines derived from different cancer types. Therefore, this valuable reference data set may facilitate the discovery of pan-cancer biomarkers and cancer type-specific biomarkers using a secretome-based strategy.

We and other groups have performed extensive studies to identify cancer-derived, body fluid-accessible marker candidates in conditioned media from various cancer cell lines. However, it has become apparent that not all proteins found in cancer cell secretomes are useful tumor-associated markers. Thus, researchers need a way to determine which candidates should be further validated as useful body fluid-accessible markers. Theoretically, the most likely candidates are proteins that are released by cancer cells and are overexpressed at a high frequency in tumor tissues. This notion is supported by recent immunohistochemical studies showing that numerous proteins are secreted by cancer cell lines and are overexpressed in cancerous tissues and that the serum and plasma levels of these proteins are indeed elevated in cancer patients (20, 23, 25–27, 31). These proteins include PAI-1 and Mac-2BP in patients with NPC (20), cathepsin D in patients with lung cancer (23), Mac-2BP and CRMP-2 (collapsin response mediator protein-2) in patients with CRC (25, 26), Mac-2BP in patients with oral cancer (27), and ALCAM, CD44, IL1R2, MIF, and TARC in patients with Hodgkin lymphoma (31). Studies that explore the tissue expression of proteins secreted by cell lines of different cancer types may therefore allow for the efficient prioritization of targets for subsequent validation.

The HPA is the first publicly available database containing information about protein expression in standardized normal human tissues, cancers, and *in vitro* cultured cell lines (60, 61). Much of these data have been analyzed via tissue mi-

croarrays (62), and the individual images have been annotated by certified pathologists. We searched the HPA database for cancer cell-secreted/released proteins of interest using an arbitrary selection criteria (*i.e.* proteins expressed in more than 50% of the tumor tissue sections examined) and thereby performed an *in silico* validation of the expression of hundreds of target proteins in various human cancers. Ultimately, we generated a list of ~470 serological cancer biomarker candidates from 11 cancer types for further validation (Table V and supplemental Table 7). In the present study, we confirmed the significantly elevated plasma levels of two such candidates (*i.e.* CD14 and SDF-1/CXCL12) in HCC and lung cancer patients, respectively (Fig. 6). The CD14 protein, a glycosylphosphatidylinositol-anchored glycoprotein of 50–55 kDa, was constitutively expressed on the surface of myeloid cells and acts as a pattern recognition receptor that plays an important role in innate immunity. A second soluble form of CD14 (sCD14) lacking the glycosylphosphatidylinositol tail was abundant in serum (63, 64). In addition to myeloid cells, recent studies have shown that CD14 is also expressed by many non-myeloid cells (65), including the hepatoma cell line Hep G2 (66, 67). Although many studies have examined the role of CD14 in infection- or immunity-related diseases, little is known about the possible significance of this protein in the development of cancer. To our knowledge, only one previous study reported higher preoperative serum levels of sCD14 in patients with epithelial ovarian cancer than in patients with benign ovarian disease (68). The SDF-1/CXCL12 protein is a member of the chemokine family that acts as a ligand for the CXCR4 receptor and plays multiple roles in tumor pathogenesis (69, 70). Although SDF-1/CXCL12 is overexpressed in numerous cancers, few studies have investigated SDF-1/CXCL12 levels in blood samples obtained from cancer patients (71–73), and the clinical significance of SDF-1/CXCL12 blood levels in most cancers is largely unknown. Our integrative analysis of cancer cell secretomes and the HPA reveals, for the first time, that CD14 and SDF-1/CXCL12 are potential plasma makers for HCC and lung cancer, respectively.

In addition to the 469 candidates selected from 11 cancer types, the present study also highlighted ~80 protein candidates as pan-cancer serological marker candidates. The expression of these marker candidates in specimens from nine cancer types has been validated in the HPA (supplemental Table 8). These candidates are involved in many biological processes, including metabolic processes (mainly the glycolytic enzymes), cell motility (cytoskeleton-related regulatory proteins), protein folding (molecular chaperones), proteolytic systems (proteasome and proteases), and protein synthesis. It has long been recognized that these biological processes are dysregulated in many cancers as is the expression of proteins and enzymes involved in these processes (74). However, few studies have systemically determined whether these proteins can be secreted by different types of cancer cells and whether these proteins are potentially useful serological mark-

ers for cancer. We addressed the former question and found that more than 170 proteins were released by all 23 cancer cell lines in serum-free conditions. It is possible that these proteins were detected in all cell lines because they are simply more abundant in the conditioned media. This notion is supported by our findings that most, if not all, of the 170 proteins were identified on the basis of multiple peptides and have significantly higher emPAI values than do proteins that were only detected in some cell lines (Table III and supplemental Tables 1 and 8). Another important question is whether proteins that are released by multiple cell lines and are highly expressed in most cancer types are, in fact, suitable serological pan-cancer marker candidates that should be further validated. Some of the target proteins on our list have been previously described as potential serological markers for various cancers. For example, elevated serum or plasma levels of galectin-3-binding protein (*i.e.* Mac-2BP) (75) have been reported in six cancer types, including breast cancer, HCC, lymphoma, NPC, CRC, and oral cancer (20, 26, 27, 76–78). Similarly, higher serum or plasma levels of cathepsin D have been detected in nine cancer types, including breast cancer, HCC, HNC, prostate cancer, glioma, CRC, stomach cancer, pancreatic cancer, and lung cancer (23, 79–85).

One of major challenges in the fields of tumor marker discovery and cancer biology is the formation of biological hypotheses about numerous marker candidates and the design of effective follow-up experiments (49). We evaluated our short list of secreted regulatory proteins for NPC using emPAI-based, label-free quantification and hierarchical clustering analyses in an effort to put our secretome data into a biological context (Fig. 4 and supplemental Table 9). It should be noted that the three NPC cell lines examined in this study originated from distinct NPC types. Specifically, NPC-TW02 and NPC-TW04 were derived from a keratinizing carcinoma and undifferentiated carcinoma, respectively (86), whereas NPC-BM1 was derived from a bone marrow biopsy of a patient with nonkeratinizing NPC (87). However, the three NPC cell lines could be clustered together with the proteins listed in supplemental Table 8, suggesting that these proteins and pathways may play an important role in NPC initiation or progression. We performed an additional pathway analysis by integrating our secretome data into a cellular signaling context. This analysis suggested that cell adhesion/migration and immune system regulation are among the most differentially regulated biological processes in NPC as compared with other cancers (Fig. 5 and supplemental Fig. 4). Metastasis has occurred in 7% of NPC patients by the time of initial diagnosis, and over 20% of patients with NPC develop metastasis after treatment (88, 89). In addition, immune suppression and evasion via the inactivation of tumor-infiltrating lymphocytes and imbalances in regulatory and effector T cells have been proposed in NPC patients (90, 91). We selected two targets for validation (*i.e.* cathepsin L1 and ISG15) that are known to be involved in the regulation of the immune system (Fig. 5B).

Cathepsin L1 (also known as cathepsin L, which is distinct from cathepsin L2 and cathepsin V) is a lysosomal cysteine protease that can degrade the components of extracellular matrices and basement membranes (92). This protease is overexpressed in a variety of cancer tissues (93). Cathepsin L1 is often overexpressed in metastatic cervical lymph node samples from patients with NPC, which correlates with lymph node metastasis and distant metastasis (53). The second target for validation, ISG15 (*i.e.* an interferon-stimulated gene also known as ubiquitin cross-reactive protein), plays a critical role in the interferon-mediated immune response against antiviral infection (94, 95). ISG15 was recently identified as a novel tumor marker candidate in bladder, breast, and oral cancers (96–98). However, we are the first to examine the serum levels of cathepsin L1 and ISG15 in patients with NPC, and our results show that the levels of these two proteins are higher in patients with NPC than in healthy controls. Taken together, these observations suggest that an emPAI-based, label-free quantification and hierarchical clustering analyses of multiple cancer cell secretomes, in conjunction with biological pathway analyses, are an effective strategy for discovering dysregulated pathway-based serological marker candidates in cancers such as NPC.

To date, ~8,800 antibodies have been applied against the human proteome to examine the expression of corresponding proteins in the HPA. There are still many human proteins whose expression in cancer tissues has not been systematically examined via immunohistochemical analyses and therefore could not be analyzed using the methods described here. Expression of these proteins in cancerous *versus* non-cancerous tissues may be alternatively evaluated via the analysis of numerous cDNA microarray data sets available in the public domain, and the results could then be integrated with the cancer cell secretome data set to identify potential serological marker candidates. The feasibility of this notion is supported by our recent identification of Mac-2BP as a CRC plasma biomarker based on its apparent secretion by CRC cell lines and elevated transcriptional levels in a public array-based analysis of CRC tissues (26). In addition, we performed an in-house cDNA microarray analysis to demonstrate that the mRNA levels of cystatin A, manganese-superoxide dismutase, and MMP2 were higher in NPC tissues than in adjacent non-tumor tissues. We furthermore showed that these proteins were released by NPC cell lines and proposed that they serve as a potential serum biomarker panel for NPC (99). As immunohistochemical staining signals can vary with the use of different antibodies, our findings suggest that the potential serological markers identified using the strategy described here might complement the list of potential markers derived from the combined analysis of cancer cell secretomes and cancer tissue transcriptomes.

In conclusion, we have performed an in-depth comparative analysis of the secretomes of 23 cancer cell lines and have integrated our data with the HPA to generate a list of potential body fluid-accessible cancer biomarker candidates. The val-

idation of potential cancer biomarkers requires a large cohort of well defined, high quality clinical specimens, a process that slows the discovery of clinically useful biomarkers (54, 55). Therefore, efficient strategies for narrowing down the list of proteins that are dysregulated in cancers would greatly reduce the costs of manpower, reagents, and precious clinical specimens. The potential cancer biomarkers described here could complement or serve as a useful alternative to previously constructed data sets (4, 57) for the more rapid validation of clinically useful biomarkers.

\* This work was supported by Ministry of Education Grants EMRPD150231 and EMRPD170191 (to Chang Gung University), Chang Gung Memorial Hospital Grants CMRPD160096 and CMRPD160097, and National Science Council of Taiwan, Republic of China Grant NSC96-2320-B-182-031-MY3.

§ This article contains supplemental Materials and Methods, Figs. 1–6, and Tables 1–9.

§ Both authors contributed equally to this work.

§§ To whom correspondence should be addressed: Dept. of Cell and Molecular Biology, Chang Gung University, Tao-Yuan 333, Taiwan. Tel.: 886-3-2118800 (ext. 5171); Fax: 886-3-2118891; E-mail: yusong@mail.cgu.edu.tw.

#### REFERENCES

- Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., and Thun, M. J. (2008) Cancer statistics, 2008. *CA Cancer J. Clin.* **58**, 71–96
- Bettendorf, O., Piffkò, J., and Bänkfalvi, A. (2004) Prognostic and predictive factors in oral squamous cell cancer: important tools for planning individual therapy? *Oral. Oncol.* **40**, 110–119
- Chen, Y. J., Chang, J. T., Liao, C. T., Wang, H. M., Yen, T. C., Chiu, C. C., Lu, Y. C., Li, H. F., and Cheng, A. J. (2008) Head and neck cancer in the betel quid chewing area: recent advances in molecular carcinogenesis. *Cancer Sci.* **99**, 1507–1514
- Polanski, M., and Anderson, N. L. (2007) A list of candidate cancer biomarkers for targeted proteomics. *Biomark. Insights* **1**, 1–48
- Conrads, T. P., Zhou, M., Petricoin, E. F., 3rd, Liotta, L., and Veenstra, T. D. (2003) Cancer diagnosis using proteomic patterns. *Expert Rev. Mol. Diagn.* **3**, 411–420
- Gorelik, E., Landsittel, D. P., Marrangoni, A. M., Modugno, F., Velikokhatnaya, L., Winans, M. T., Bigbee, W. L., Herberman, R. B., and Lokshin, A. E. (2005) Multiplexed immunobead-based cytokine profiling for early detection of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **14**, 981–987
- Mor, G., Visintin, I., Lai, Y., Zhao, H., Schwartz, P., Rutherford, T., Yue, L., Bray-Ward, P., and Ward, D. C. (2005) Serum protein markers for early detection of ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7677–7682
- Xiao, T., Ying, W., Li, L., Hu, Z., Ma, Y., Jiao, L., Ma, J., Cai, Y., Lin, D., Guo, S., Han, N., Di, X., Li, M., Zhang, D., Su, K., Yuan, J., Zheng, H., Gao, M., He, J., Shi, S., Li, W., Xu, N., Zhang, H., Liu, Y., Zhang, K., Gao, Y., Qian, X., and Cheng, S. (2005) An approach to studying lung cancer-related proteins in human blood. *Mol. Cell. Proteomics* **4**, 1480–1486
- Kulasingam, V., and Diamandis, E. P. (2008) Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat. Clin. Pract. Oncol.* **5**, 588–599
- Good, D. M., Thongboonkerd, V., Novak, J., Bascands, J. L., Schanstra, J. P., Coon, J. J., Dominiczak, A., and Mischak, H. (2007) Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *J. Proteome Res.* **6**, 4549–4555
- Hanash, S. M., Pitteri, S. J., and Faca, V. M. (2008) Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579
- Reymond, M. A., and Schlegel, W. (2007) Proteomics in cancer. *Adv. Clin. Chem.* **44**, 103–142
- Schmidt, A., and Aebersold, R. (2006) High-accuracy proteome maps of human body fluids. *Genome Biol.* **7**, 242

14. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
15. Tang, H. Y., Ali-Khan, N., Echan, L. A., Levenkova, N., Rux, J. J., and Speicher, D. W. (2005) A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes. *Proteomics* **5**, 3329–3342
16. Wu, C. C., Peng, P. H., Chang, Y. T., Huang, Y. S., Chang, K. P., Hao, S. P., Tsang, N. M., Yeh, C. T., Chang, Y. S., and Yu, J. S. (2008) Identification of potential serum markers for nasopharyngeal carcinoma from a xenografted mouse model using Cy-dye labeling combined with three-dimensional fractionation. *Proteomics* **8**, 3605–3620
17. Yocum, A. K., Yu, K., Oe, T., and Blair, I. A. (2005) Effect of immunoaffinity depletion of human serum during proteomic investigations. *J. Proteome Res.* **4**, 1722–1731
18. Yu, K. H., Barry, C. G., Austin, D., Busch, C. M., Sangar, V., Rustgi, A. K., and Blair, I. A. (2009) Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery. *J. Proteome Res.* **8**, 1565–1576
19. Volmer, M. W., Radacz, Y., Hahn, S. A., Klein-Scory, S., Stühler, K., Zapatka, M., Schmiegel, W., Meyer, H. E., and Schwarte-Waldhoff, I. (2004) Tumor suppressor Smad4 mediates downregulation of the anti-adhesive invasion-promoting matricellular protein SPARC: Landscaping activity of Smad4 as revealed by a “secretome” analysis. *Proteomics* **4**, 1324–1334
20. Wu, C. C., Chien, K. Y., Tsang, N. M., Chang, K. P., Hao, S. P., Tsao, C. H., Chang, Y. S., and Yu, J. S. (2005) Cancer cell-secreted proteomes as a basis for searching potential tumor markers: nasopharyngeal carcinoma as a model. *Proteomics* **5**, 3173–3182
21. Kulasingam, V., and Diamandis, E. P. (2007) Proteomics analysis of conditioned media from three breast cancer cell lines: a mine for biomarkers and therapeutic targets. *Mol. Cell. Proteomics* **6**, 1997–2011
22. Kulasingam, V., and Diamandis, E. P. (2008) Tissue culture-based breast cancer biomarker discovery platform. *Int. J. Cancer* **123**, 2007–2012
23. Lou, X., Xiao, T., Zhao, K., Wang, H., Zheng, H., Lin, D., Lu, Y., Gao, Y., Cheng, S., Liu, S., and Xu, N. (2007) Cathepsin D is secreted from M-BE cells: its potential role as a biomarker of lung cancer. *J. Proteome Res.* **6**, 1083–1092
24. Planque, C., Kulasingam, V., Smith, C. R., Reckamp, K., Goodglick, L., and Diamandis, E. P. (2009) Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol. Cell. Proteomics* **8**, 2746–2758
25. Wu, C. C., Chen, H. C., Chen, S. J., Liu, H. P., Hsieh, Y. Y., Yu, C. J., Tang, R., Hsieh, L. L., Yu, J. S., and Chang, Y. S. (2008) Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics* **8**, 316–332
26. Wu, C. C., Huang, Y. S., Lee, L. Y., Liang, Y., Tang, R. P., Chang, Y. S., Hsieh, L. L., and Yu, J. S. (2008) Overexpression and elevated plasma level of tumor-associated antigen 90K/Mac-2 binding protein in colorectal carcinoma. *Proteomics Clin. Appl.* **2**, 1586–1595
27. Weng, L. P., Wu, C. C., Hsu, B. L., Chi, L. M., Liang, Y., Tseng, C. P., Hsieh, L. L., and Yu, J. S. (2008) Secretome-based identification of Mac-2 binding protein as a potential oral cancer marker involved in cell growth and motility. *J. Proteome Res.* **7**, 3765–3775
28. Sardana, G., Jung, K., Stephan, C., and Diamandis, E. P. (2008) Proteomic analysis of conditioned media from the PC3, LNCaP, and 22Rv1 prostate cancer cell lines: discovery and validation of candidate prostate cancer biomarkers. *J. Proteome Res.* **7**, 3329–3338
29. 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
30. Gunawardana, C. G., Kuk, C., Smith, C. R., Batruch, I., Soosaipillai, A., and Diamandis, E. P. (2009) Comprehensive analysis of conditioned media from ovarian cancer cell lines identifies novel candidate markers of epithelial ovarian cancer. *J. Proteome Res.* **8**, 4705–4713
31. Ma, Y., Visser, L., Roelofsens, H., de Vries, M., Diepstra, A., van Imhoff, G., van der Wal, T., Luinge, M., Alvarez-Llamas, G., Vos, H., Poppema, S., Vonk, R., and van den Berg, A. (2008) Proteomics analysis of Hodgkin lymphoma: identification of new players involved in the cross-talk between HRS cells and infiltrating lymphocytes. *Blood* **111**, 2339–2346
32. Gronborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., Goggins, M. G., Maitra, A., and Pandey, A. (2006) Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol. Cell. Proteomics* **5**, 157–171
33. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1–11
34. Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G., and Mann, M. (2002) Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542
35. Björling, E., Lindskog, C., Oksvold, P., Linné, J., Kampf, C., Hober, S., Uhlén, M., and Pontén, F. (2008) A web-based tool for in silico biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. *Mol. Cell. Proteomics* **7**, 825–844
36. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021
37. 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
38. 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
39. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795
40. Nielsen, H., and Krogh, A. (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 122–130
41. Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., and Brunak, S. (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **17**, 349–356
42. Möller, S., Croning, M. D., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653
43. 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
44. Zhang, Y., Zhang, Y., Adachi, J., Olsen, J. V., Shi, R., de Souza, G., Pasini, E., Foster, L. J., Macek, B., Zougman, A., Kumar, C., Wisniewski, J. R., Jun, W., and Mann, M. (2007) MAPU: Max-Planck Unified database of organellar, cellular, tissue and body fluid proteomes. *Nucleic Acids Res.* **35**, D771–D779
45. Cheadle, C., Vawter, M. P., Freed, W. J., and Becker, K. G. (2003) Analysis of microarray data using Z score transformation. *J. Mol. Diagn.* **5**, 73–81
46. Wang, C. L., Wang, C. I., Liao, P. C., Chen, C. D., Liang, Y., Chuang, W. Y., Tsai, Y. H., Chen, H. C., Chang, Y. S., Yu, J. S., Wu, C. C., and Yu, C. J. (2009) Discovery of retinoblastoma-associated binding protein 46 as a novel prognostic marker for distant metastasis in non-small cell lung cancer by combined analysis of cancer cell secretome and pleural effusion proteome. *J. Proteome Res.* **8**, 4428–4440
47. MacCoss, M. J., Wu, C. C., and Yates, J. R., 3rd (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–5599
48. Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226–3245



49. Lawlor, K., Nazarian, A., Lacomis, L., Tempst, P., and Villanueva, J. (2009) Pathway-Based Biomarker Search by High-Throughput Proteomics Profiling of Secretomes. *J. Proteome Res.* **8**, 1489–1503
50. Sawyers, C. L. (2008) The cancer biomarker problem. *Nature* **452**, 548–552
51. Faça, V. M., Ventura, A. P., Fitzgibbon, M. P., Pereira-Faça, S. R., Pitteri, S. J., Green, A. E., Ireton, R. C., Zhang, Q., Wang, H., O'Brian, K. C., Drescher, C. W., Schummer, M., McIntosh, M. W., Knudsen, B. S., and Hanash, S. M. (2008) Proteomic analysis of ovarian cancer cells reveals dynamic processes of protein secretion and shedding of extra-cellular domains. *PLoS One* **3**, e2425
52. Sengupta, S., den Boon, J. A., Chen, I. H., Newton, M. A., Stanhope, S. A., Cheng, Y. J., Chen, C. J., Hildesheim, A., Sugden, B., and Ahlquist, P. (2008) MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5874–5878
53. Xu, X., Yuan, G., Liu, W., Zhang, Y., and Chen, W. (2009) Expression of cathepsin L in nasopharyngeal carcinoma and its clinical significance. *Exp. Oncol.* **31**, 102–105
54. 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
55. Hartwell, L., Mankoff, D., Paulovich, A., Ramsey, S., and Swisher, E. (2006) Cancer biomarkers: a systems approach. *Nat. Biotechnol.* **24**, 905–908
56. Veenstra, T. D., Conrads, T. P., Hood, B. L., Avellino, A. M., Ellenbogen, R. G., and Morrison, R. S. (2005) Biomarkers: mining the biofluid proteome. *Mol. Cell. Proteomics* **4**, 409–418
57. Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S. B., Telikicherla, D., Selvan, L. D., Goel, R., Mathivanan, S., Marimuthu, A., Kashyap, M., Vizza, R. F., Mayer, R. J., Decaprio, J. A., Srivastava, S., Hanash, S. M., Hruban, R. H., and Pandey, A. (2009) A compendium of potential biomarkers of pancreatic cancer. *PLoS Med.* **6**, e1000046
58. Hathout, Y. (2007) Approaches to the study of the cell secretome. *Expert Rev. Proteomics* **4**, 239–248
59. Xue, H., Lu, B., and Lai, M. (2008) The cancer secretome: a reservoir of biomarkers. *J. Transl. Med.* **6**, 52
60. Andersson, A. C., Strömberg, S., Bäckvall, H., Kampf, C., Uhlen, M., Wester, K., and Pontén, F. (2006) Analysis of protein expression in cell microarrays: a tool for antibody-based proteomics. *J. Histochem. Cytochem.* **54**, 1413–1423
61. Kampf, C., Andersson, A. C., Wester, K., Bjorling, E., Uhlen, M., and Pontén, F. (2004) Antibody-based tissue profiling as a tool for clinical proteomics. *Clin. Proteomics* **1**, 285–299
62. Kononen, J., Bubendorf, L., Kallioniemi, A., Bärklund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M. J., Sauter, G., and Kallioniemi, O. P. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **4**, 844–847
63. Haziot, A., Chen, S., Ferrero, E., Low, M. G., Silber, R., and Goyert, S. M. (1988) The monocyte differentiation antigen, CD14, is anchored to the cell membrane by a phosphatidylinositol linkage. *J. Immunol.* **141**, 547–552
64. Ziegler-Heitbrock, H. W., and Ulevitch, R. J. (1993) CD14: cell surface receptor and differentiation marker. *Immunol. Today* **14**, 121–125
65. Jersmann, H. P. (2005) Time to abandon dogma: CD14 is expressed by non-myeloid lineage cells. *Immunol. Cell Biol.* **83**, 462–467
66. Bas, S., Gauthier, B. R., Spenato, U., Stingelin, S., and Gabay, C. (2004) CD14 is an acute-phase protein. *J. Immunol.* **172**, 4470–4479
67. Nanbo, A., Nishimura, H., Muta, T., and Nagasawa, S. (1999) Lipopolysaccharide stimulates HepG2 human hepatoma cells in the presence of lipopolysaccharide-binding protein via CD14. *Eur. J. Biochem.* **260**, 183–191
68. Gadducci, A., Ferdeghini, M., Castellani, C., Annicchiarico, C., Galletti, O., Prontera, C., Bianchi, R., and Facchini, V. (1995) Serum levels of tumor necrosis factor (TNF), soluble receptors for TNF (55- and 75-kDa sTNFr), and soluble CD14 (sCD14) in epithelial ovarian cancer. *Gynecol. Oncol.* **58**, 184–188
69. Kryczek, I., Wei, S., Keller, E., Liu, R., and Zou, W. (2007) Stroma-derived factor (SDF-1/CXCL12) and human tumor pathogenesis. *Am. J. Physiol. Cell Physiol.* **292**, C987–C995
70. Ratajczak, M. Z., Zuba-Surma, E., Kucia, M., Reza, R., Wojakowski, W., and Ratajczak, J. (2006) The pleiotropic effects of the SDF-1-CXCR4 axis in organogenesis, regeneration and tumorigenesis. *Leukemia* **20**, 1915–1924
71. Dimberg, J., Hugander, A., Löfgren, S., and Wågsäter, D. (2007) Polymorphism and circulating levels of the chemokine CXCL12 in colorectal cancer patients. *Int. J. Mol. Med.* **19**, 11–15
72. Hassan, S., Baccarelli, A., Salvucci, O., and Basik, M. (2008) Plasma stromal cell-derived factor-1: host derived marker predictive of distant metastasis in breast cancer. *Clin. Cancer Res.* **14**, 446–454
73. Zannettino, A. C., Farrugia, A. N., Kortesis, A., Manavis, J., To, L. B., Martin, S. K., Diamond, P., Tamamura, H., Lapidot, T., Fujii, N., and Gronthos, S. (2005) Elevated serum levels of stromal-derived factor-1alpha are associated with increased osteoclast activity and osteolytic bone disease in multiple myeloma patients. *Cancer Res.* **65**, 1700–1709
74. Hanahan, D., and Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* **100**, 57–70
75. Koths, K., Taylor, E., Halenbeck, R., Casipit, C., and Wang, A. (1993) Cloning and characterization of a human Mac-2-binding protein, a new member of the superfamily defined by the macrophage scavenger receptor cysteine-rich domain. *J. Biol. Chem.* **268**, 14245–14249
76. Fornarini, B., D'Ambrosio, C., Natoli, C., Tinari, N., Silingardi, V., and Iacobelli, S. (2000) Adhesion to 90K (Mac-2 BP) as a mechanism for lymphoma drug resistance in vivo. *Blood* **96**, 3282–3285
77. Iacobelli, S., Sismondì, P., Gai, M., D'Egidio, M., Tinari, N., Amatetti, C., Di Stefano, P., and Natoli, C. (1994) Prognostic value of a novel circulating serum 90K antigen in breast cancer. *Br. J. Cancer* **69**, 172–176
78. Iacovazzi, P. A., Trisolini, A., Barletta, D., Elba, S., Manghisi, O. G., and Correale, M. (2001) Serum 90K/MAC-2BP glycoprotein in patients with liver cirrhosis and hepatocellular carcinoma: a comparison with alpha-fetoprotein. *Clin. Chem. Lab. Med.* **39**, 961–965
79. Fukuda, M. E., Iwadate, Y., Machida, T., Hiwasa, T., Nimura, Y., Nagai, Y., Takiguchi, M., Tanzawa, H., Yamaura, A., and Seki, N. (2005) Cathepsin D is a potential serum marker for poor prognosis in glioma patients. *Cancer Res.* **65**, 5190–5194
80. Hara, I., Miyake, H., Yamanaka, K., Hara, S., and Kamidono, S. (2002) Serum cathepsin D and its density in men with prostate cancer as new predictors of disease progression. *Oncol. Rep.* **9**, 1379–1383
81. Leto, G., Tumminello, F. M., Pizzolanti, G., Montalto, G., Soresi, M., Ruggeri, I., and Gebbia, N. (1996) Cathepsin D serum mass concentrations in patients with hepatocellular carcinoma and/or liver cirrhosis. *Eur. J. Clin. Chem. Clin. Biochem.* **34**, 555–560
82. Schultz, D. C., Bazel, S., Wright, L. M., Tucker, S., Lange, M. K., Tachovsky, T., Longo, S., Niedbala, S., and Alhadeff, J. A. (1994) Western blotting and enzymatic activity analysis of cathepsin D in breast tissue and sera of patients with breast cancer and benign breast disease and of normal controls. *Cancer Res.* **54**, 48–54
83. Skrzydlewska, E., Sulkowska, M., Wincewicz, A., Koda, M., and Sulkowski, S. (2005) Evaluation of serum cathepsin B and D in relation to clinico-pathological staging of colorectal cancer. *World J. Gastroenterol.* **11**, 4225–4229
84. Strojjan, P., Budihna, M., Smid, L., Vrhovec, I., and Skrk, J. (1998) Cathepsin D in tissue and serum of patients with squamous cell carcinoma of the head and neck. *Cancer Lett.* **130**, 49–56
85. Szajda, S. D., Snarska, J., Roszkowska-Jakimiec, W., Waszkiewicz, N., Siedlecka, K., Zwierz, K., and Krupkowska, A. (2006) Activity of cathepsin D in the blood serum and urine of patients with cancer of the stomach, pancreas and liver. *Pol. Arch. Med. Wewn.* **116**, 1150–1154
86. Lin, C. T., Chan, W. Y., Chen, W., Huang, H. M., Wu, H. C., Hsu, M. M., Chuang, S. M., and Wang, C. C. (1993) Characterization of seven newly established nasopharyngeal carcinoma cell lines. *Lab. Invest.* **68**, 716–727
87. Liao, S. K., Perng, Y. P., Shen, Y. C., Chung, P. J., Chang, Y. S., and Wang, C. H. (1998) Chromosomal abnormalities of a new nasopharyngeal carcinoma cell line (NPC-BM1) derived from a bone marrow metastatic lesion. *Cancer Genet. Cytogenet.* **103**, 52–58
88. Fandi, A., Altun, M., Azli, N., Armand, J. P., and Cvitkovic, E. (1994) Nasopharyngeal cancer: epidemiology, staging, and treatment. *Semin. Oncol.* **21**, 382–397
89. Sham, J. S., and Choy, D. (1990) Prognostic factors of nasopharyngeal carcinoma: a review of 759 patients. *Br. J. Radiol.* **63**, 51–58
90. Fogg, M. H., Wirth, L. J., Posner, M., and Wang, F. (2009) Decreased EBNA-1-specific CD8+ T cells in patients with Epstein-Barr virus-asso-

- ciated nasopharyngeal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3318–3323
91. Yip, W. K., Abdullah, M. A., Yusoff, S. M., and Seow, H. F. (2009) Increase in tumour-infiltrating lymphocytes with regulatory T cell immunophenotypes and reduced zeta-chain expression in nasopharyngeal carcinoma patients. *Clin. Exp. Immunol.* **155**, 412–422
92. Mason, R. W., Johnson, D. A., Barrett, A. J., and Chapman, H. A. (1986) Elastolytic activity of human cathepsin L. *Biochem. J.* **233**, 925–927
93. Chauhan, S. S., Goldstein, L. J., and Gottesman, M. M. (1991) Expression of cathepsin L in human tumors. *Cancer Res.* **51**, 1478–1481
94. Andersen, J. B., and Hassel, B. A. (2006) The interferon regulated ubiquitin-like protein, ISG15, in tumorigenesis: friend or foe? *Cytokine Growth Factor Rev.* **17**, 411–421
95. Pitha-Rowe, I. F., and Pitha, P. M. (2007) Viral defense, carcinogenesis and ISG15: novel roles for an old ISG. *Cytokine Growth Factor Rev.* **18**, 409–417
96. Andersen, J. B., Aaboe, M., Borden, E. C., Goloubeva, O. G., Hassel, B. A., and Orntoft, T. F. (2006) Stage-associated overexpression of the ubiquitin-like protein, ISG15, in bladder cancer. *Br. J. Cancer* **94**, 1465–1471
97. Chi, L. M., Lee, C. W., Chang, K. P., Hao, S. P., Lee, H. M., Liang, Y., Hsueh, C., Yu, C. J., Lee, I. N., Chang, Y. J., Lee, S. Y., Yeh, Y. M., Chang, Y. S., Chien, K. Y., and Yu, J. S. (2009) Enhanced interferon signaling pathway in oral cancer revealed by quantitative proteome analysis of microdissected specimens using <sup>16</sup>O/<sup>18</sup>O labeling and integrated two-dimensional LC-ESI-MALDI tandem MS. *Mol. Cell. Proteomics* **8**, 1453–1474
98. Desai, S. D., Haas, A. L., Wood, L. M., Tsai, Y. C., Pestka, S., Rubin, E. H., Saleem, A., Nur-E-Kamal, A., and Liu, L. F. (2006) Elevated expression of ISG15 in tumor cells interferes with the ubiquitin/26S proteasome pathway. *Cancer Res.* **66**, 921–928
99. Chang, K. P., Wu, C. C., Chen, H. C., Chen, S. J., Peng, P. H., Tsang, N. M., Lee, L. Y., Liu, S. C., Liang, Y., Lee, Y. S., Hao, S. P., Chang, Y. S., and Yu, J. S. (2010) Identification of candidate nasopharyngeal carcinoma serum biomarkers by cancer cell secretome and tissue transcriptome analysis: Potential usage of cystatin A for predicting nodal stage and poor prognosis. *Proteomics* in press