

# Differential Expression of Novel Tyrosine Kinase Substrates during Breast Cancer Development<sup>\*S</sup>

Yunhao Chen<sup>‡§</sup>, Lee-Yee Choong<sup>‡§</sup>, Qingsong Lin<sup>¶</sup>, Robin Philp<sup>||</sup>, Chee-Hong Wong<sup>\*\*</sup>, Boon-Keong Ang<sup>\*\*</sup>, Yee-Ling Tan<sup>‡</sup>, Marie-Chiew-Shia Loh<sup>\*\*</sup>, Choy-Leong Hew<sup>¶</sup>, Nilesh Shah<sup>‡</sup>, Brian J. Druker<sup>‡‡</sup>, Poh-Kuan Chong<sup>‡</sup>, and Yoon-Pin Lim<sup>‡¶</sup><sup>\*\*§§</sup>

To identify novel tyrosine kinase substrates that have never been implicated in cancer, we studied the phosphoproteomic changes in the MCF10AT model of breast cancer progression using a combination of phosphotyrosyl affinity enrichment, iTRAQ<sup>TM</sup> technology, and LC-MS/MS. Using complementary MALDI- and ESI-based mass spectrometry, 57 unique proteins comprising tyrosine kinases, phosphatases, and other signaling proteins were detected to undergo differential phosphorylation during disease progression. Seven of these proteins (SPAG9, Toll-interacting protein (TOLLIP), WBP2, NSFL1C, SLC4A7, CYFIP1, and RPS2) were validated to be novel tyrosine kinase substrates. SPAG9, TOLLIP, WBP2, and NSFL1C were further proven to be authentic targets of epidermal growth factor signaling and Iressa (gefitinib). A closer examination revealed that the expression of SLC4A7, a bicarbonate transporter, was down-regulated in 64% of the 25 matched normal and tumor clinical samples. The expression of TOLLIP in clinical breast cancers was heterogeneous with 25% showing higher expression in tumor compared with normal tissues and 35% showing the reverse trend. Preliminary studies on SPAG9, on the other hand, did not show differential expression between normal and diseased states. This is the first time SLC4A7 and TOLLIP have been discovered as novel tyrosine kinase substrates that are also associated with human cancer development. Future molecular and functional studies will provide novel insights into the roles of TOLLIP and SLC4A7 in the molecular etiology of breast cancer. *Molecular & Cellular Proteomics* 6:2072–2087, 2007.

Understanding the molecular etiology of cancer is critical to its management. This could be achieved by creating a data-

base of genetic aberrations during disease initiation, maintenance, and progression. However, disease progression is practically impossible to study in an individual because of physicians' intervention. Although the use of clinical samples from patients in different stages of cancer is one option, differences in patients' genetic and environmental backgrounds complicate analysis. Consequently experimental breast cancer models are valuable and have been successfully used to identify molecular events during disease development (1–3). The MCF10AT model comprises a series of isogenic, xenograft-derived cell lines that mimic the different stages of breast cancer progression (4, 5) and has been characterized in terms of cytogenetics, apoptosis, transforming growth factor- $\beta$  signaling, and proteomics (6–12).

However, one major deficiency of most proteomics tools is their inability to detect low abundance proteins. Phosphoproteomics on the other hand offers several advantages. (i) It is a form of subproteomics that allows the detection of low abundance signaling proteins that would otherwise be masked by other major proteins in the whole proteome approach. (ii) Unlike DNA microarray, it is an open system that allows the discovery of novel phosphoproteins in the system studied. In turn, further studies would contribute to our understanding of the roles of these novel tyrosine kinase substrates in cancer biology. (iii) Because phosphoproteomics focuses on kinases and/or their substrates, it may have implication on target-directed therapeutics either on discovery or in understanding the molecular epidemiology of drug targets. For these reasons, we are interested to exploit phosphoproteomics for studying disease progression. This has never been done before.

We have previously developed two-dimensional PAGE and cleavable ICAT-based phosphoproteomics to dissect signaling pathways and map the differential phosphoprotein contents of cancer cell lines and breast tumors, respectively (13–15). In this study, we adapted a relatively newer isobaric tags for relative and absolute quantification (iTRAQ<sup>TM</sup>)<sup>1</sup> tech-

From the <sup>‡</sup>Oncology Research Institute, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117456, Singapore, <sup>¶</sup>Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 117543, Singapore, <sup>||</sup>Bio-processing Technology Institute, Agency for Science, Technology and Research, Singapore 138668, Singapore, <sup>‡‡</sup>Howard Hughes Medical Institute, Oregon Health and Science University, Portland, Oregon 97239, and <sup>\*\*</sup>Bioinformatics Institute, Agency for Science, Technology and Research, Singapore 138671, Singapore

Received, August 20, 2007, and in revised form, September 11, 2007

Published, MCP Papers in Press, September 12, 2007, DOI 10.1074/mcp.M700395-MCP200

<sup>1</sup> The abbreviations used are: iTRAQ, isobaric tags for relative and absolute quantitation; PV, pervanadate; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; FAK, focal adhesion kinase; IPI, International Protein Index; IHC, immunohistochemistry; EF, error factor; HEK, human embryonic kidney; EPHA2, Ephrin A2; ER, estrogen receptor; BIN, biological interaction network; CYFIP1, cytoplasmic FMR1 interacting protein 1 isoform 1; TOLLIP, Toll-interacting protein; VCP, transitional endoplasmic reticulum ATPase/89kDa protein.

nology to profile the tyrosine phosphorylation level of proteins in the MCF10AT model of breast cancer progression. The aims of this study were to identify novel tyrosine kinase substrates that have never been implicated in cancer development and that displayed differential expression levels in the MCF10AT model. These candidates were subsequently investigated for their involvement in tumor development in the clinical setting.

#### EXPERIMENTAL PROCEDURES

**Reagents**—PY20H mouse monoclonal anti-phosphotyrosine antibody, FAK, SHP2, Cortactin, Paxillin, plakoglobin, SHC1,  $\beta$ -catenin, EGFR, and EPS8 antibodies were from BD Biosciences. P130CAS antibody was from Neomarkers (Fremont, CA); Ephrin A2 receptor antibody was from Zymed Laboratories Inc. (Carlsbad, CA); E-cadherin antibody was from Cell Signaling Technology (Beverly, MA); SLC4A7 antibody was from Chemicon International Inc. (Temecula, CA); CYFIP antibody was from Upstate Biotechnology (Lake Placid, NY); PSMC4 antibody was from Atlas Antibodies AB (Stockholm, Sweden); TOLLIP, SPAG9, and receptor-type tyrosine-protein phosphatase kappa precursor antibodies were from Abcam (Cambridge, UK); and RPS2 antibody was from Abnova Corp. (Taipei, Taiwan). cDNAs for WBP2 and NSFL1C were from Origene (Rockville, MD). Anti-V5 tag antibodies and pcDNA<sup>TM</sup>3.1 Directional TOPO<sup>®</sup> vector were from Invitrogen. The ECL detection kit was purchased from GE Healthcare, prestained molecular weight markers were from Bio-Rad, protease inhibitor mixture were from Roche Applied Science, and polyvinylidene difluoride membranes were from Millipore (Bedford, MA). Sodium orthovanadate and 3% (0.9 M) hydrogen peroxide were purchased from Sigma-Aldrich.

**Cell Lines, Cell Stimulation, Sample Preparation, and Xenograft Study**—Xenograft-derived breast cancer cell lines (MCF10A1, MCF10AT1KCl.2, MCF10CA1h, and MCF10CA1aCl.1) were obtained from Dr. Fred Miller at the Barbara Ann Karmanos Cancer Institute (Detroit, MI). Cells were cultured as previously described elsewhere (11). Cells were incubated at 37 °C in a humidified atmosphere containing 5% CO<sub>2</sub> and were routinely serum-starved for 16 h before stimulation with 1 mM pervanadate for 15 min at 37 °C. Fresh 0.45 M stock of pervanadate was first generated by mixing an equal volume of 0.9 M sodium orthovanadate with 3% (0.9 M) H<sub>2</sub>O<sub>2</sub>. During cell stimulation, a final concentration of 1 mM pervanadate was achieved by a 1:454 dilution of the stock solution. For protein extraction, cells grown in 150-mm tissue culture dishes were rinsed with ice-cold PBS before lysis in 1 ml of non-denaturing buffer (50 mM Tris-HCl (pH 7.5), 0.5% Triton X-100, 0.5% Igepal, 150 mM NaCl, 1 mM EDTA, 50 mM NaF, 1 mM sodium orthovanadate plus protease inhibitors mixture). Protein lysates were clarified by centrifugation at 4 °C using a microcentrifuge at 14,000 rpm for 10 min. Estimation of protein concentration was performed using a bicinchoninic acid assay kit (Pierce). Immunoblotting and immunoprecipitation were performed according to previous reports (14, 16). The protocol for the xenograft study was reviewed and approved by the Institutional Animal Care and Use Committee of the National University of Singapore in compliance with international guidelines on the care and use of animals for scientific purpose. A1 (10<sup>7</sup>), 1k (5 × 10<sup>6</sup>), 1h (10<sup>6</sup>), and 1a (3 × 10<sup>5</sup>) cells in a PBS and Matrigel (BD Biosciences) mixture (1:1) were injected subcutaneously into the flanks of severe combined immunodeficiency mice and observed weekly. Histological assessment was conducted by a certified pathologist.

**LC-MS/MS of 4G10-purified Phosphoproteins: Detection and Relative Quantification**—Serum-starved cells were treated with 1 mM pervanadate for 15 min and lysed. From large scale preparations of cell lysates, one portion was used for phosphoproteomics analysis,

and the other portion was retained for subsequent validation. For the former purpose, 60 mg of total protein from each cell line were used for purification via 4G10 anti-phosphotyrosine antibody-based immunoaffinity purification. This was performed as previously reported except that the enriched phosphoproteins were eluted with a buffer that was PBS-based rather than Tris-based because the latter is incompatible with iTRAQ reagents (14). After the enriched phosphoprotein contents of four samples were determined, they were then denatured, and the cysteines were blocked as described in the iTRAQ protocol (Applied Biosystems, Foster City, CA). Each sample was then digested with trypsin provided in the reagent kit at 37 °C overnight (16 h) and labeled with the iTRAQ tags as follows: A1, 114 tag; 1k, 115 tag; 1h, 116 tag; and 1a, 117 tag. As the pervanadate (PV)-induced phosphoproteome from different cell lines had different amounts of proteins enriched ranging from 50 to 150  $\mu$ g, each sample was treated with two vials of the same isotopic tag to ensure complete labeling. The samples were then pooled and cleaned up by the cation exchange cartridge provided in the kit. The samples were desalted, lyophilized, and analyzed using ESI-LC/MS/MS and MALDI-TOF-TOF.

On-line ESI-LC/MS/MS was performed using a QSTAR-XL hybrid quadrupole time-of-flight tandem mass spectrometer (Applied Biosystems) coupled to an LC Packings (Dionex, Sunnyvale, CA) LC system comprising a FAMOS autoinjector unit, a Switchos 10-port valve unit, and an UltiMate<sup>PLUS</sup> nanoflow pumping unit. Sample was injected into a reversed-phase trapping C<sub>18</sub> peptide trapping cartridge (300  $\mu$ m × 5 mm, LC Packings) in a flow of 0.1% formic acid for 5 min at 25  $\mu$ l/min. Following the wash step the flow from the UltiMate<sup>PLUS</sup> was diverted back through the trapping cartridge at 100 nl/min using the Switchos. Peptides were eluted from the cartridge by application of a gradient from 0 to 90% acetonitrile in 0.1% formic acid over 40 min at 100 nl/min and separated by passing through a column that was packed in house and consisted of a 75- $\mu$ m × 10-cm packed volume of 5- $\mu$ m C<sub>18</sub> reversed-phase packing (Column Engineering, Ontario, Canada). Peptides eluting from the column were sprayed directly into the orifice of the mass spectrometer, which was run in information-dependent acquisition mode selecting all 2+ to 4+ charged ions with signal intensity greater than 8 counts/s over the mass range of 300–2000 amu. For CID nitrogen gas was used at a setting of 4, and the collision energy was set to automatic, allowing increased energy with increasing ion mass. Each fraction aliquot that was run was searched against the human subset of the International Protein Index (IPI) protein database (European Bioinformatics Institute, Hinxton, UK) using the Mascot (Matrix Science, London, UK) search engine. From the search results a mass exclusion list was generated for each fraction based on the peptide masses of all matching peptides or up to a maximum of 2000 masses. These masses were then excluded from MS/MS fragmentation when the remaining aliquot of each fraction was rerun. For protein identification and quantification, the complete set of data files (\*.wiff) was analyzed together using ProteinPilot<sup>TM</sup> software version 2.0 (Applied Biosystems) and searched against the human subset of the UniProt protein database (release 11, downloaded in May 2007 with a total 267,354 entries and 16,602 human entries; European Bioinformatics Institute) using the Paragon<sup>TM</sup> algorithm. The search parameters allowed for cysteine modification by methyl methanethiosulfonate, phosphorylation emphasis, and biological modifications programmed in the algorithm. The detected protein threshold (unused protscore (conf)) in the software was set to 1.3 to achieve 95% confidence. The bias correction option was not executed because this study concerned chemically induced protein phosphorylation that can take place to different extents in various cells.

For MALDI analysis, the iTRAQ labeled peptide mixture was separated using an UltiMate<sup>TM</sup> LC system (Dionex-LC Packings)

equipped with a Probot™ MALDI spotting device. The iTRAQ peptide mixture was first captured with a  $0.3 \times 1$ -mm trap column ( $3\text{-}\mu\text{m}$  C<sub>18</sub> PepMap™, 100 Å) (Dionex-LC Packings) and washed with 0.05% TFA followed by gradient separation with a  $0.2 \times 50$ -mm reversed-phase column (Monolithic PS-DVB) (Dionex-LC Packings). The mobile phase A and mobile phase B used were 98% H<sub>2</sub>O, 2% ACN with 0.05% TFA and 80% H<sub>2</sub>O, 20% ACN, and 0.04% TFA, respectively. The gradient elution step was 0–60% mobile phase B in 20 min with a flow rate of 2.7  $\mu\text{l/min}$ . The LC fractions were mixed with MALDI matrix solution (7 mg/ml  $\alpha$ -cyano-4-hydroxycinnamic acid and 130  $\mu\text{g/ml}$  ammonium citrate in 75% ACN) in a flow rate of 5.4  $\mu\text{l/min}$  through a 25-nl mixing tee (Upchurch Scientific) before spotting onto 192-well stainless steel MALDI target plates (Applied Biosystems) using a Probot microfraction collector (Dionex-LC Packings) with a speed of 5 s/well. MALDI target plates were analyzed using an ABI 4700 Proteomics Analyzer MALDI-TOF/TOF mass spectrometer (Applied Biosystems) operating in a result-independent acquisition mode. Typically 1000 shots were accumulated for each well of sample. MS/MS analyses were performed using nitrogen at collision energy of 1 kV and a collision gas pressure of  $\sim 1 \times 10^{-6}$  torr. For precursor ions with signal to noise ratio greater or equal to 100, 6000 shots were combined for each spectrum. For the precursors with signal to noise ratio between 50 and 100, 10,000 shots were accumulated. GPS Explorer™ software version 3.5 (Applied Biosystems) was used to create and search files with the Mascot search engine (version 2.1; Matrix Science) for peptide and protein identifications. IPI human database (version 3.14, released January 2006, 57,366 sequences) was used for the search and was restricted to tryptic peptides. Cysteine methanethiolation; N-terminal iTRAQ labeling; iTRAQ labeled lysine; serine, threonine, and tyrosine phosphorylation; and methionine oxidation were selected as variable modifications. One missing cleavage was allowed. Precursor error tolerance was set to 100 ppm, and MS/MS fragment error tolerance was set to 0.3 Da. Maximum peptide rank was set to 2, and minimum ion score confidence interval (peptide) was set to 95%. The iTRAQ quantification was performed using GSP Explorer software version 3.5. The normalization option was not executed because this study concerned chemically induced protein phosphorylation that can take place to different extents in various cells.

**Low Density Real Time PCR Array**—Cells were grown in a  $10\text{-cm}^2$  tissue culture dish until 80–90% confluence ( $\sim 1 \times 10^7$  exponentially dividing cells). Total RNA for each cell line was extracted using the Purescript RNA isolation kit (Gentra Qiagen, West Sussex, UK) according to the manufacturer's protocol. For first strand cDNA synthesis, 2  $\mu\text{g}$  of RNA were reverse transcribed using the following conditions:  $1 \times$  avian myeloblastosis virus reverse transcriptase reaction buffer (Promega), 30 units of avian myeloblastosis virus reverse transcriptase (Promega), 40 units RNasin® ribonuclease inhibitor (Promega), 1 mM dNTP mixture (Invitrogen), 1  $\mu\text{g}$  of oligo(dT) (Invitrogen). Each RT-PCR was performed in triplicates and pooled. The integrity of the pooled cDNA was assessed by PCR amplification with a human control gene, glyceraldehyde-3-phosphate dehydrogenase. Negative control (without cDNA template) was also included to check for contaminating cDNA and genomic DNA. Evaluation of the gene expression profile of the MCF10AT model was performed using the TaqMan® Gene Expression Assay and gene-specific primers for NSF1LC (Hs00739840\_mH), WBP2 (Hs00600857\_m1), and glyceraldehyde-3-phosphate dehydrogenase (HS99999905\_m1) from Applied Biosystems. Sample-specific PCR mixture was loaded in quadruplicates onto the microplate at 1  $\mu\text{g}$  of cDNA/well and analyzed using the ABI 7900HT system. Quantification of gene expression relative to the "normal" cell line was then done using the comparative C<sub>T</sub> method as described in the manufacturer's manual.

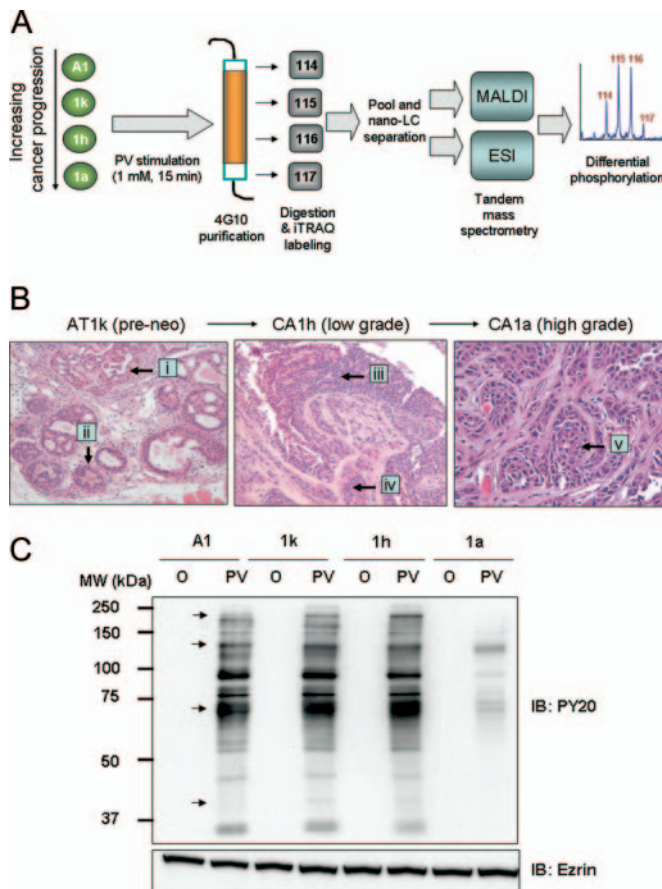
**Immunohistochemistry**—Frozen tissues from National University Hospital and Singapore General Hospital were freshly prepared for

IHC by fixing in 10% neutral buffered formalin (Sigma) for 16 h at 4 °C, subjecting to a ThermoShandon tissue processor, and embedding in paraffin. Sections were warmed in a 60 °C oven, dewaxed in three changes of xylene, and passaged through graded ethanol (100, 95, and 70%) before a final wash in double distilled H<sub>2</sub>O. For TOLLIP (Abgent), (Abcam), and SPAG9 (Abcam), antigen retrieval was performed via pressure cooking at 121 °C for 5 min in Tris-EDTA buffer, pH 9.0. For SLC4A7 (Chemicon International Inc.), antigen retrieval was performed using the Target Retrieval Solution (DakoCytomation, Glostrup, Denmark) at 95 °C for 40 min. After quenching of endogenous peroxidase activity with 3% H<sub>2</sub>O<sub>2</sub> for 10 min and blocking with BSA for 30 min, sections were incubated at 4 °C overnight with antibodies against TOLLIP at a 1:50 dilution, at a 1:50 dilution, SPAG9 at a 1:100 dilution, and SLC4A7 at a 1:50 dilution. Detection was achieved with the Envision+ horseradish peroxidase system (DakoCytomation). All slides were counterstained with Gill's hematoxylin for 1 min, dehydrated, and mounted for light microscopic evaluation. Interpretation of hematoxylin and eosin sections and analysis/scoring of IHC data were all done by the same certified pathologist to maintain consistency.

## RESULTS

**Detection and Relative Quantification of Pervanadate-induced Tyrosine Phosphorylated Proteins in MCF10AT Cell Line Model of Breast Cancer Progression**—The experimental design used in this study to detect differentially tyrosine phosphorylated proteins in the MCF10AT model is shown in Fig. 1A. The cell lines in the MCF10AT model used in this study included MCF10A1, which is modeled after normal epithelium, and MCF10AT1K.cl2, MCF10CA1h, and MCF10CA1a.cl1, which are modeled after premalignant epithelium and low grade and high grade lesions, respectively (11). They are abbreviated as A1, 1k, 1h, and 1a in this study. The tumors obtained by subcutaneous injection of 1k, 1h, and 1a cancer cells into nude mice (note that A1 is not tumorigenic) grew at different rates and were of different grades as assessed by a pathologist validating that the model was indeed reflective of disease progression (Fig. 1B and Supplemental Fig. 1). To prepare cells for phosphoproteomics analysis, cells were serum-starved overnight and treated for 15 min with 1 mM PV, a potent tyrosine phosphatase inhibitor, to enhance the presentation of tyrosine phosphorylated proteins. The lysates of the MCF10AT cells untreated or stimulated with PV were then probed with anti-phosphotyrosine antibodies to reveal the overall cellular tyrosine phosphorylation profiles (Fig. 1C, *top panel*). The intensities of a few proteins (marked by arrows) were stronger in 1k and 1h compared with A1 after PV stimulation. The amount of tyrosine phosphorylated proteins in 1a cells was remarkably low, suggesting that it is deficient in one or more tyrosine kinases. The phosphoproteins were then affinity-captured using 4G10 anti-phosphotyrosine antibodies, trypsin-digested, labeled with iTRAQ reagents (17), and analyzed using tandem mass spectrometry to determine the relative phosphoprotein levels in the MCF10AT cells. To increase the coverage of protein identifications and/or the confidence of the data generated, two separate preparations were made, and each was analyzed by ESI-LC/MS/MS and MALDI-TOF-TOF. The cells A1, 1k, 1h, and 1a were labeled with iTRAQ reagents





**FIG. 1.** A, schematic diagram showing the workflow of iTRAQ-based experiments to identify PV-induced tyrosine phosphorylation substrates in the MCF10AT model of breast cancer progression. Duplicate sets of experiments were conducted. MCF10AT cell lines were treated with 1 mM pervanadate for 15 min. The cell lysates were then separately incubated with 4G10 antibodies covalently conjugated to Sepharose beads. Enriched phosphoproteins from each cell line were then separately labeled with iTRAQ tags (114, 115, 116, and 117). The labeled samples were subsequently pooled and digested with trypsin. Labeled peptides were purified using an avidin column and analyzed with ESI- and MALDI-based mass spectrometry as described in detail under “Experimental Procedures.” B, histopathology of the tumors derived from MCF10AT cell lines 1k, 1h, and 1a cells showing preneoplastic lesion and low and high grade cancers, respectively. Note that A1 cells are not tumorigenic and therefore not shown here. Xenograft-derived tumors were processed as described under “Experimental Procedures,” and hematoxylin and eosin staining was performed. *Left panel*, AT1k-derived tumor with ductal carcinoma *in situ* (ii) and microinvasion of stroma (i); *middle panel*, invasive papillary carcinoma (iii) in CA1h tumor with stromal invasion of malignant epithelial cells (iv); *right panel*, nests and trabeculae of infiltrating malignant cells of poorly differentiated invasive ductal carcinoma (v). C, pervanadate-induced tyrosine phosphorylation in MCF10AT cell lines. Cells were serum-starved overnight and then untreated or treated with PV at 1 mM for 15 min. Proteins in lysates were then resolved and immunoblotted (IB) with anti-phosphotyrosine antibodies conjugated to horseradish peroxidase (PY20H). The level of ezrin, as detected by anti-ezrin antibodies, was used as a control for equal loading of lysates. Arrows indicate examples of protein bands that displayed differential tyrosine phosphorylation across the four cell lines. O, serum-starved cells not treated with PV.

114, 115, 116, and 117, respectively. The ratios 115:114, 116:114, and 117:114 would indicate the relative abundance of a potentially tyrosine phosphorylated protein in 1k, 1h, and 1a with respect to A1.

Table I shows the detection (gene symbol, column 2; protein name, column 3) and relative quantity (columns 6–8) of 4G10 antibody-enriched proteins in the various cell lines. The S.D. and error factor (EF) shown in columns 9–11 as confidence indicator for relative quantification were provided by ProteinPilot software for ESI-based data and by the GPS Explorer software for the MALDI-based dataset. Additional information such as the peptide sequence, *m/z* value, and sites of iTRAQ modification of proteins detected via single peptide assignment are provided in Supplemental Table 1. The protein detection (including group reporting) and quantification data from MALDI- and ESI-based LC-MS/MS analyses from which the information in Table I was derived are provided as Supplemental Table 2 (A and B) and Supplemental Table 3, respectively. Not all the proteins identified by the MALDI platform (Supplemental Table 2B) were included in Table I. Only those protein hits that had a minimal best/total ion score of 95% were selected to maintain a high confidence. For the ESI/ProteinPilot-based method, the protein threshold (unused protscore (conf)) was set to 1.3 to achieve 95% confidence. This generated Supplemental Table 3 where all the hits were of 95% confidence. Hence they were all included in Table I. In Supplemental Tables 2 and 3, protein hits were either reported as unique or as a group of proteins/isoforms (when the peptide sequence(s) did not have sufficient discriminatory power). In the latter case, a protein member, usually the one with the highest rank/score in the database search, would be listed in tables as a representative of that group.

To determine cutoff values to classify proteins as differentially expressed, we calculated the median S.D. for MALDI-based data and median EF for ESI-based data. The median worked out to be 0.27 and 1.2 for S.D. and EF, respectively. The upper (1.27 (for S.D.) and 1.2 (for EF)) and lower (0.73 (for S.D.) and 0.83 (for EF)) range were then applied to the MALDI- and ESI-based data. Proteins with iTRAQ ratio below the lower range were considered to be underexpressed, whereas those above the higher range were considered to be overexpressed. Taking a cutoff point at 30% variation is acceptable because the main variable in this study was technical variation, whereas the biological variation was minimized by sample pooling effect. Another study also reported that at least a 30% variation needed to be taken into consideration for technical variations in large scale protein identification and quantification using the iTRAQ approach (18). Although this method has its limitation in that it uses a general yardstick on every protein, it is only a first pass screening that will be followed up with validation work later. Based on the relative ratios of proteins detected in various cell lines, the proteins were categorized into seven different trend plots shown in column 1 of Table I. A total of 39 and 37 protein hits were

TABLE I

Relative quantification of 4G10 anti-phosphotyrosine antibody-enriched proteins in PV-stimulated MCF10AT breast cancer progression model

NA, not applicable. S.D. and EF were calculated from peptide detected from the MALDI and ESI platforms, respectively. Most of the single peptide quantification assignments do not have S.D./EF values except those with multiple MS<sup>2</sup>. Shaded boxes indicate common identity hits with similar expression trend detected from MALDI and ESI mass spectrometry. There are six pairs of proteins (superscript 1–6) that were identified in both MALDI and ESI platform mass spectrometry having a similar expression trend but dissimilar at one point.

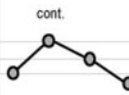




Trend	Gene symbol	Protein name	MALDI source	ESI source	AT1k/10A1 ratio	CA1h/10A1 ratio	CA1a/10A1 ratio	AT1k/10A1 SD/EF	CA1h/10A1 SD/EF	CA1a/10A1 SD/EF	Unique ITRAQ peptide
	ITGA6	Splice Isoform Alpha-6X1A of Integrin alpha-6 precursor <sup>(1)</sup>	IPI00216221		3.53	4.70	0.73	NA	NA	NA	1
	RPS27A	Ubiquitin and ribosomal protein S27a <sup>(2)</sup>	IPI00179330		1.86	2.03	0.40	NA	NA	NA	1
	TUBA4	Tubulin, alpha 4	IPI00017454		2.72	2.85	0.52	0.72	0.83	0.10	1
	RIN1	Splice Isoform RIN1-delta of Ras and Rab interactor 1	IPI00217761		3.47	4.01	0.90	NA	NA	NA	1
	DSG2	Desmoglein 2, preproprotein	IPI00028931		1.89	1.85	0.63	NA	NA	NA	1
	K-ALPHA-1	Tubulin alpha-ubiquitous chain protein	IPI00328163		2.67	2.63	0.52	0.610	0.787	0.087	2
	EPHA2	Ephrin type-A receptor 2 precursor (EC 2.7.10.1) (Tyrosine-protein kinase receptor ECK) (Epithelial cell kinase)		P29317	3.35	2.68	2.05	1.35	1.31	1.93	5
	C9orf88	Niban-like protein (Meg-3)		Q96TA1	3.79	3.03	1.02	NA	NA	NA	1
	CTNNB1	Catenin beta-1 (Beta-catenin) <sup>(3)</sup>		P35222	2.60	2.05	1.00	1.06	1.03	3.79	3
	SYNCRIP	Splice Isoform 2 of Heterogeneous nuclear ribonucleoprotein Q <sup>(4)</sup>	IPI00402182		4.69	3.05	0.95	NA	NA	NA	1
	TJP2	Tight junction protein ZO-2 (Zonula occludens 2 protein) (Zona occludens 2 protein) (Tight junction protein 2) <sup>(5)</sup>		Q9UDY2	2.65	2.13	0.85	1.12	1.09	1.14	13
	ABI1	Abi interactor 1 (Abelson interactor 1) (Abi-1) (Spectrin SH3 domain-binding protein 1) (Eps8 SH3 domain-binding protein) (Eps8-binding protein) (e3B1) (Nap1-binding protein) (Nap1BP) (Abi-binding protein 4) (AbiBP4)		Q8IZP0	2.16	1.69	0.64	4.29	2.37	9.73	2
	ACTN4	Alpha-actinin 4	IPI00013808		2.92	2.70	0.45	NA	NA	NA	1
		Alpha-actinin-4 (Non-muscle alpha-actinin 4) (F-actin cross-linking protein)		O43707	3.59	2.85	0.68	1.38	1.33	1.45	5
	BCAR1	p130 CRK-associated substrate	IPI00011998		3.01	2.37	0.54	NA	NA	NA	1
	CAV1	Caveolin-1		Q03135	1.64	1.51	0.81	NA	NA	NA	1
	CDCP1	CUB domain-containing protein 1 precursor (Transmembrane and associated with src kinases) (Membrane glycoprotein gp140) (Subtractive immunization M plus HEp3-associated 135 kDa protein) (SIMA135) (CD318 antigen)		Q9H5V8	2.18	1.71	0.70	1.062	1.071	1.051	2
	CDH1	E-cadherin	IPI00000513		2.03	1.64	0.19	0.43	0.51	0.16	3
		Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1) (CAM 120/80) (CD324 antigen) [Contains: E-Cad/CTF1; E-Cad/CTF2; E-Cad/CTF3]		P12830	2.18	1.78	0.38	1.38	1.36	1.12	2
	CTNNA1	Catenin alpha-1 (Cadherin-associated protein) (Alpha E-catenin) (NY-REN-13 antigen)		P35221	2.02	1.59	0.34	1.11	1.08	1.15	7
	CTNNB1	Splice Isoform 1 of Beta-catenin <sup>(3)</sup>	IPI00017292		2.19	1.88	0.23	0.34	0.04	0.04	2
	CTNND1	Splice Isoform 2AC of Catenin delta-1	IPI00219726		2.64	1.83	0.35	0.77	0.49	0.19	14
		Catenin delta-1 protein (p120 catenin) (Fragment)		O60716	1.96	1.60	0.47	1.13	1.09	1.05	17
	CTTN	Src substrate cortactin <sup>(6)</sup>	IPI00029601		1.74	1.60	0.44	0.49	0.23	0.03	4
	CYFIP1	Cytoplasmic FMR1 interacting protein 1 isoform 1	IPI00644231		2.39	1.94	0.38	1.03	0.42	0.23	3
	DBNL	Drebrin-like protein (SH3 domain-containing protein 7) (Drebrin-F) (Cervical SH3P7) (HPK1-interacting protein of 55 kDa) (HIP-55) (Cervical mucin-associated protein)		Q9UJU6	2.59	2.24	0.54	1.41	1.55	1.30	4
	EGFR	Splice Isoform 1 of Epidermal growth factor receptor precursor	IPI00018274		3.30	2.04	0.44	0.10	0.14	0.10	2
		Epidermal growth factor receptor precursor (EC 2.7.10.1) (Receptor tyrosine-protein kinase ErbB-1)		P00533	3.46	1.88	0.49	1.08	1.07	1.07	3
	EPS8	Epidermal growth factor receptor kinase substrate 8	IPI00290337		2.79	1.95	0.35	0.35	0.16	0.06	2
		Epidermal growth factor receptor kinase substrate 8		Q12929	2.39	1.65	0.50	NA	NA	NA	1
	FLNB	Splice Isoform 2 of Filamin-B	IPI00382696		4.04	3.12	0.50	NA	NA	NA	1
		Filamin-B (FLN-B) (Beta-filamin) (Actin-binding-like protein) (Thyroid autoantigen) (Truncated actin-binding protein) (Truncated ABP) (ABP-280 homolog) (ABP-278) (Filamin 3) (Filamin homolog 1) (Fh1)		O75369	2.86	2.28	0.56	2.02	1.82	1.36	3
	FNBP1L	Formin binding protein 1-like isoform 1/ Transducer of Cdc42-dependent actin	IPI00607808		3.05	2.40	0.39	NA	NA	NA	1
	HSPA5	78 kDa glucose-regulated protein precursor	IPI00003362		3.15	2.39	0.48	NA	NA	NA	1
	HSPA8	Splice Isoform 1 of Heat shock cognate 71 kDa protein	IPI00003865		3.82	3.29	0.64	1.16	0.65	0.07	3
		Heat shock cognate 71 kDa protein (Heat shock 70 kDa protein 8)		P11142	2.91	2.34	0.70	1.11	1.11	1.34	3
	HSPA9B	Stress-70 protein, mitochondrial precursor (75 kDa glucose-regulated protein) (GRP 75) (Peptide-binding protein 74) (PBP74) (Mortalin) (MOT)		P38646	2.63	1.97	0.54	3.39	3.29	2.41	2
	ITGA6	Integrin alpha-6 precursor (VLA-6) (CD49f antigen) [Contains: Integrin alpha-6 heavy chain; Integrin alpha-6 light chain] <sup>(1)</sup>		P23229	3.17	2.71	0.76	1.311	1.412	1.180	4
	ITGB4	Splice Isoform Beta-4C of Integrin beta-4 precursor	IPI00220031		3.80	3.38	0.60	1.43	1.54	0.10	8
		Integrin beta-4 precursor (GP150) (CD104 antigen)		P16144	2.38	2.25	0.70	1.26	1.25	1.10	13

identified via the MALDI and ESI methods, respectively. Of the total 76 proteins detected by both methods, 19 hits (25%) were common. These 19 proteins have both accession (column 5) and IPI numbers (column 4) in the table as they were detected via ESI- and MALDI-based platforms with search algorithm against different databases as mentioned under “Experimental Procedures.” Twenty hits were unique to the

MALDI-based method (column 4 only), whereas 18 were unique to the ESI-based method (column 5 only). This resulted in a total of 57 unique protein hits detected and relatively quantified. By looking at the trend plot in column 1, six of the 19 common hits (about 30%), as indicated by superscript 1–6 in column 3, exhibited dissimilarity at only one data point (either 116:114 or 117:114). Otherwise they displayed an over-



TABLE I—continued

Trend	Gene Symbol	Protein name	MALDI source	ESI source	AT1k/10A1 ratio	CA1h/10A1 ratio	CA1a/10A1 ratio	AT1k/10A1 SD/EF	CA1h/10A1 SD/EF	CA1a/10A1 SD/EF	Unique iTRAQ peptide
	JUP	Junction plakoglobin (Desmoplakin-3) (Desmoplakin III)		P14923	2.10	1.73	0.65	1.06	1.05	1.24	4
	NCKAP1	Nck-associated protein 1 (NAP 1) (p125Nap1) (Membrane-associated protein HEM-2)		Q9Y2A7	2.61	1.97	0.62	1.22	1.10	2.04	3
	NSFL1C	Splice isoform 1 of NSFL1 cofactor p47	IPI00100197		1.64	1.52	0.15	NA	NA	NA	1
	PSMC4	Splice isoform 2 of 26S protease regulatory subunit 6B	IPI00216770		3.78	3.20	0.46	NA	NA	NA	1
	PTK2	Splice isoform 1 of Focal adhesion kinase 1 (FAK1)	IPI00012885		2.48	2.19	0.44	0.85	1.35	0.22	2
		Focal adhesion kinase 1 (EC 2.7.10.2) (FADK 1) (pp125FAK) (Protein-tyrosine kinase 2)		Q05397	2.33	1.93	0.50	NA	NA	NA	1
	PTK2B	Protein tyrosine kinase 2 beta (EC 2.7.10.2) (Focal adhesion kinase 2) (FADK 2) (Proline-rich tyrosine kinase 2) (Cell adhesion kinase beta) (CAK beta) (Calcium-dependent tyrosine kinase) (CADTK) (Related adhesion focal tyrosine kinase) (RAFTK)		Q14289	1.72	1.56	0.36	NA	NA	NA	1
	PTPN11	Tyrosine-protein phosphatase non-receptor type 11 (EC 3.1.3.48) (Protein-tyrosine phosphatase 2C) (PTP-2C) (PTP-1D) (SH-PTP3) (SH-PTP2) (SHP-2) (Shp2)		Q06124	3.07	2.35	0.55	1.35	1.18	1.62	2
	PTPN18	Tyrosine-protein phosphatase non-receptor type 18 (EC 3.1.3.48) (Brain-derived phosphatase)		Q99952	2.79	2.21	0.61	NA	NA	NA	2
	PXN	Splice isoform Gamma of Paxillin	IPI00220031		2.61	2.49	0.56	NA	NA	NA	1
	Paxillin			P49023	2.86	2.31	0.49	1.21	1.10	1.17	5
	RAB7	Ras-related protein Rab-7		P51149	3.17	2.52	0.62	NA	NA	NA	1
	ROBO1	Splice isoform 2 of Roundabout homolog 1 precursor	IPI00418121		5.43	4.51	1.78	NA	NA	NA	1
	RPS2	40S ribosomal protein S2	IPI00013485		2.76	1.79	0.56	NA	NA	NA	1
	RPS27A	Ubiquitin <sup>(2)</sup>		P62988	1.59	1.45	0.49	1.39	1.28	1.15	4
	SFN	14-3-3 protein sigma	IPI00013890		2.58	1.74	0.74	NA	NA	NA	1
	SHC1	SHC (Src homology 2 domain containing) transforming protein 1	IPI00165135		2.79	2.38	0.41	0.01	0.38	0.11	1
	SLC4A7	Sodium bicarbonate cotransporter 3 (Sodium bicarbonate cotransporter 2) (Sodium bicarbonate cotransporter 2b) (Bicarbonate transporter) (Solute carrier family 4 member 7)		Q9Y6M7	3.60	2.87	0.50	1.26	1.26	1.34	3
	SPAG9	C-JUN-amino kinase-interacting protein	IPI00384432		2.41	2.06	0.33	NA	NA	NA	1
	SPTAN1	Spectrin alpha chain, brain (Spectrin, non-erythroid alpha chain) (Alpha-II spectrin) (Fodrin alpha chain)		Q13813	2.86	2.19	0.68	1.892	1.907	1.567	2
	SYNCRIP	Heterogeneous nuclear ribonucleoprotein Q (hnRNP Q) (hnRNP-Q) (Synaptotagmin-binding, cytoplasmic RNA-interacting protein) (Glycine- and tyrosine-rich RNA-binding protein) (GRY-RBP) (NS1-associated protein 1) <sup>(4)</sup>		O60506	2.83	2.16	0.58	1.28	1.31	1.23	5
	TJP1	Splice isoform Long of Tight junction protein ZO-1	IPI00216219		2.93	2.00	0.42	0.20	0.38	0.14	3
		Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona occludens 1 protein) (Tight junction protein 1)		Q07157	2.40	1.93	0.52	1.18	1.11	1.12	1
	TJP2	Tight junction protein 2 <sup>(6)</sup>	IPI00216245		3.45	2.75	0.59	0.98	0.65	0.25	6
	TOLLIP	Toll-interacting protein	IPI00100154		2.25	1.60	0.34	NA	NA	NA	1
	TUBA2	Tubulin alpha-2 chain (Alpha-tubulin 2)		Q13748	3.04	2.37	0.70	1.135	1.032	2.031	1
	TUBB	Tubulin, beta polypeptide	IPI00645452		2.30	1.99	0.54	NA	NA	NA	1
	TUBB2	Tubulin beta-2C chain	IPI00007752		3.09	2.71	0.59	NA	NA	NA	1
	TUBB3	Tubulin beta-3 chain (Tubulin beta-III) (Tubulin beta-4)		Q13509	2.81	1.97	0.56	9.044	8.903	3.804	2
	VCP	Transitional endoplasmic reticulum ATPase / 89 kDa protein	IPI00022774		3.52	3.19	0.30	1.75	1.39	0.27	14
		Transitional endoplasmic reticulum ATPase (TER ATPase) (15S Mg(2+)-ATPase p97 subunit) (Valosin-containing protein) (VCP)		P55072	2.44	2.13	0.49	1.09	1.07	1.06	44
	VIL2	Ezrin (Villin2)	IPI00479359		2.29	1.91	0.54	1.10	0.53	0.07	5
		Ezrin (p81) (Cytovillin) (Villin-2)		P15311	2.21	1.83	0.69	1.17	1.12	1.04	9
	WBP2	WW domain-binding protein 2	IPI00032050		1.86	1.68	0.23	NA	NA	NA	1
	CTTN	Src substrate cortactin (Amplixin) (Oncogene EMS1) <sup>(6)</sup>		Q14247	1.28	1.19	0.65	1.122	1.084	1.083	11
	PTPRK	Receptor-type tyrosine-protein phosphatase kappa precursor / OTTHUMP0000017181	IPI00015756		2.14	1.15	0.24	NA	NA	NA	1

all similar trend of expression during disease progression. The other 13 hits (shaded in Table I for easy reference) were observed to display a similar trend at all data points despite being analyzed by platforms and search algorithms, indicating the robustness of the methods used. In order to facilitate better understanding of the 57 unique proteins detected, they were grouped according to their reported molecular functions using the PANTHER (Protein Analysis through Evolutionary Relationships) Classification System ([www.pantherdb.org/](http://www.pantherdb.org/)). Protein classes such as kinases, phosphatases, and other signaling molecules were identified (Fig. 2A).

A considerable number of proteins detected were based on single peptide assignment (Table I, last column). This is not surprising because the study was designed to detect tyrosine phosphorylated signaling proteins that are well known to exist in low abundance. In addition, very few tyrosine phosphorylation sites were found, and this is probably due to the fact that

enrichment was performed at the phosphoprotein level instead of the phosphopeptide level. Note that due to single peptide assignment most of these proteins detected did not have S.D. or EF unless they had multiple MS<sup>2</sup> identifications. One MS<sup>2</sup> spectrum each for all proteins detected by MALDI- and ESI-based platforms is provided in Supplemental List 1. The sequence of spectra was arranged according to their chronological appearance in Table I for easy reference. As representatives, a MALDI-derived MS<sup>2</sup> spectrum of SPAG9 is shown in Fig. 2B.

**Validation of Novel Tyrosine Kinase Substrates**—The major aims in this study were (i) the identification of novel tyrosine kinase substrates and (ii) identification of novel tyrosine kinase substrates that displayed differential expression during breast cancer progression. Two key factors must be considered. First, in our previous study on EGF-stimulated A431 cells, non tyrosine phosphorylated proteins could be detected in the

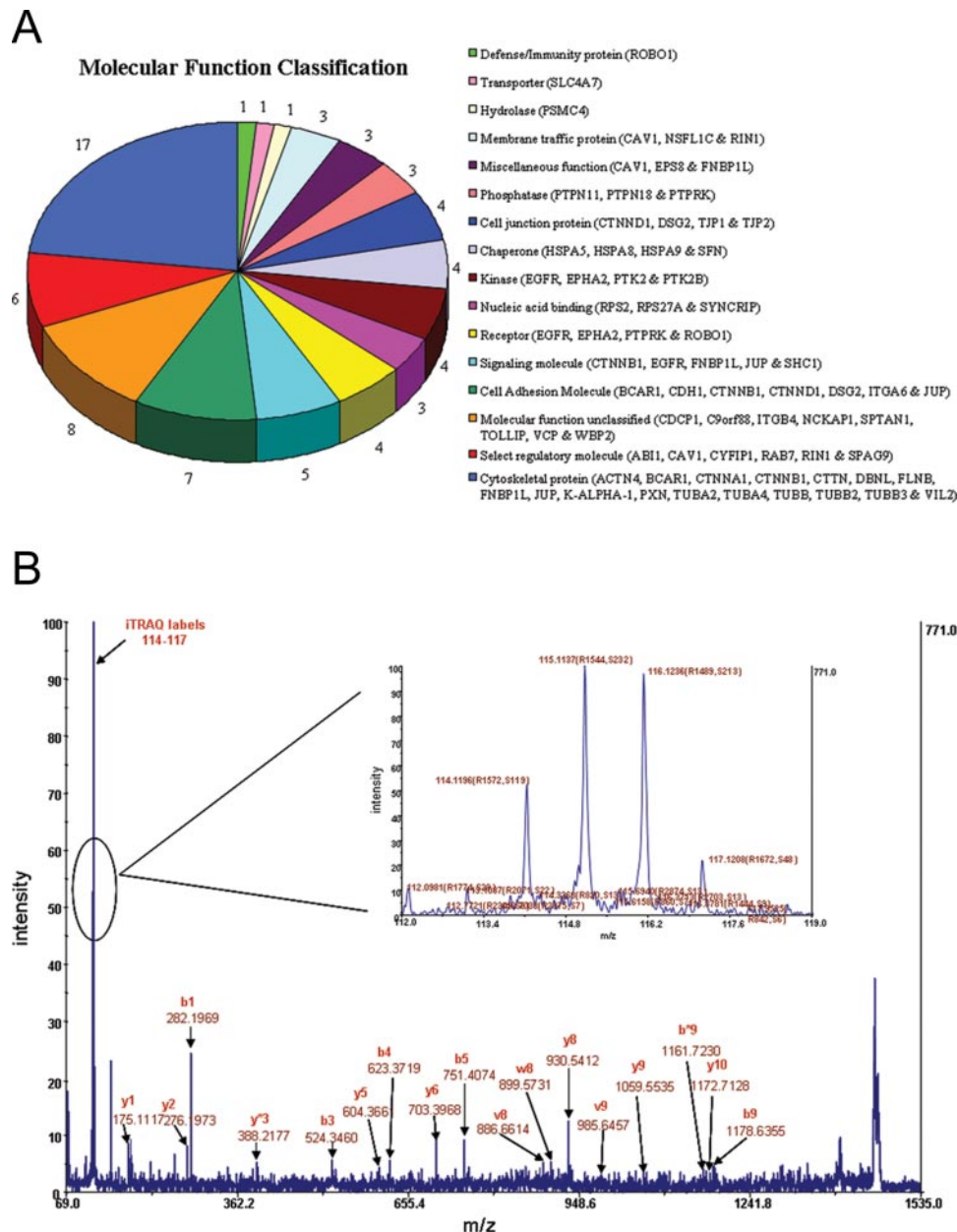
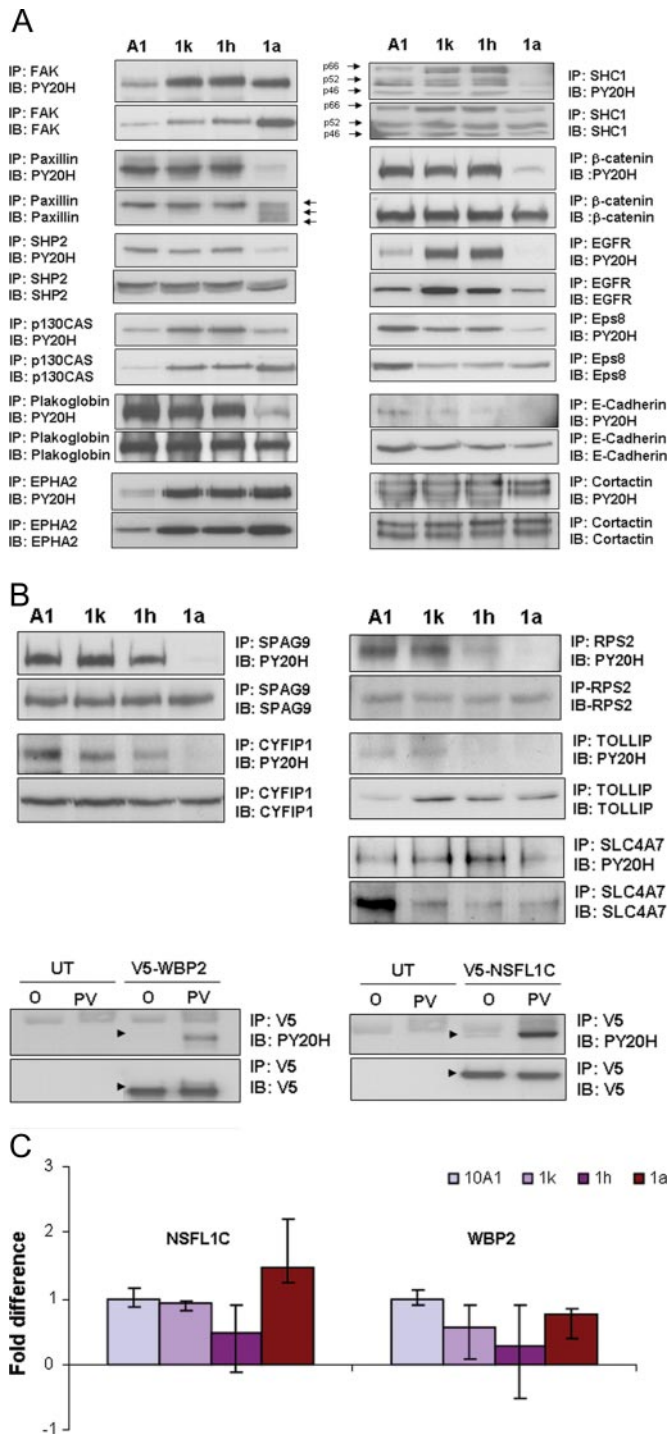


FIG. 2. *A*, classification of the proteins identified in the PV-induced phosphoproteome of the MCF10AT model into molecular functions. Phosphoproteins were categorized into different molecular functions using the PANTHER Classification System ([www.pantherdb.org/](http://www.pantherdb.org/)) analysis. Although 57 unique proteins were detected, a total of 74 assignments were obtained and sorted into 16 molecular function classifications. This is because some proteins (e.g. EGFR) may be assigned more than one molecular function. *B*, detection and relative quantification of tyrosine phosphorylation levels of proteins in PV-treated MCF10AT cells. Enriched phosphotyrosyl proteins from A1, 1k, 1h, and 1a cell lines were labeled with iTRAQ tags 114, 115, 116, and 117, respectively. Following MS<sup>2</sup> analysis, iTRAQ reporter and b and y ions of various peptides were detected. As a representative, the MALDI-based MS<sup>2</sup> spectra for the single SPAG9 peptide detected (HIEVQVAQETR, singly charged, precursor  $m/z = 1453.7902$ , iTRAQ label tagged at N terminus) is shown. The inset shows an example of the iTRAQ reporter ion sets used for relative quantification of SPAG9 across the four cell lines. *SFN*, 14-3-3 protein sigma; *JUP*, junction plakoglobin; *ITGA6*, integrin alpha-6 precursor (VLA-6); *CTNNA1*, catenin alpha-1; *CTNNB1*, splice isoform 1 of beta catenin; *DBNL*, debrin-like protein; *FLNB*, splice isoform 2 of filamin-B; *PXN*, splice isoform gamma of paxilin; *TUBB*, tublin, beta polypeptide.

4G10-purified mixture (13). These were not due to nonspecific enrichment because we have shown that all the proteins enriched by the 4G10 antibodies, regardless of whether they are tyrosine phosphorylated or non-phosphorylated proteins

associated with phosphorylated proteins, were specifically purified by the antibodies (14). This conclusion was drawn from the observation that the proteins enriched by 4G10 antibodies when A431 cells were treated with EGF were to-



**FIG. 3. Validation of known and potentially novel tyrosine phosphorylated protein identified in MCF10AT cells.** Cells were serum-starved overnight and then untreated or treated with PV at 1 mM for 15 min. Immunoprecipitations (IP) using the indicated protein-specific antibodies were performed in duplicates on the lysates. The immunoprecipitates were immunoblotted (IB) with anti-phosphotyrosine antibodies conjugated to horseradish peroxidase (PY20H) and protein-specific antibodies to reveal the relative tyrosine phosphorylation and expression levels, respectively, across the four MCF10AT cell lines. A, validation of known tyrosine phosphorylated proteins. B,

tally absent when the same EGF-treated cells were first pre-treated with Iressa, a selective EGFR inhibitor. Hence it is possible that not all the proteins listed in Table I are authentic tyrosine phosphorylated proteins. Second, the amount of phosphorylated proteins at steady state is maintained by the action of kinases and phosphatases. Pervanadate treatment blocks the action of tyrosine phosphatases thus favoring kinase activity and shifting the equilibrium toward phosphorylated proteins. The differential amount of tyrosine phosphorylated proteins induced by PV treatment of various cell lines could therefore be due to (i) different expression level of the tyrosine kinase substrates or (ii) different amounts of tyrosine kinases and/or phosphatases. To investigate the likely mechanism behind the observed differential protein phosphorylations during breast cancer progression and to determine the authenticity of tyrosine phosphorylation of potentially novel phosphoproteins, immunoprecipitations were performed on 12 well known tyrosine phosphorylated proteins and seven proteins of interest. Four of the seven proteins (SPAG9, CYFIP1, RPS2, and TOLLIP) are potentially novel tyrosine kinase substrates as revealed by extensive literature search in PubMed using the many aliases listed in Genecards®, examination of published works (19–26), and interrogation of other websites e.g. phosphorylation site database ([www.phosphida.com/](http://www.phosphida.com/)) and Phosphosite® (Cell Signaling Technology). Three of the seven proteins (WBP2, SLC4A7, and NSFL1C) have very recently been identified as potential novel tyrosine kinase substrates by others but were never validated (19, 20, 25, 26). More importantly, they have never been investigated for potential roles in cancer development, which we will describe in subsequent sections.

For consistency, immunoprecipitation was conducted on the same lysates used to generate the phosphoproteomics

validation of proteins that have never been reported or validated to be tyrosine phosphorylated. For SPAG9, CYFIP1, RPS2, TOLLIP, and SLC4A7 proteins, analyses were performed as described above. In the case of WBP2 and NSFL1C where commercially antibodies were not available, their cDNAs were cloned into pcDNA3.1 Directional TOPO vector that allowed expressed proteins to be tagged with a V5 sequence at the C terminus. NSFL1C or WBP2 cDNA was then transfected into HEK293T cells. After 72 h, HEK293T cells were untreated or treated with 1 mM PV for 15 min and lysed. Immunoprecipitation of NSFL1C and WBP2 was carried out in duplicate sets using anti-V5 antibodies, and each set was immunoblotted with either PY20H or anti-V5 antibodies. O, serum-starved cells not treated with PV; UT, untransfected cells. C, measurement of NSFL1C and WBP2 transcripts in MCF10AT cells using real time PCR. Following extraction of total RNA, real time PCR was conducted using gene-specific probes for NSFL1C and WBP2 as described in detail under “Experimental Procedures.” The relative quantity minimum/maximum range (bar) was calculated based on built-in software in the ABI 7900HT machine. It provides a range where the “true” gene expression value lies based on the 95% confidence interval. We classified a gene as significantly up-regulated when the minimum value in one condition was higher than the maximum value of the same gene in another condition. The reverse was true.



data in Table I. Two sets of immunoprecipitates were prepared. One set of immunoprecipitate was probed with anti-phosphotyrosine PY20 antibodies, and the other set was probed with protein-specific antibodies. The dataset for the 12 well known tyrosine phosphorylated proteins is shown in Fig. 3A. All 12 proteins, namely FAK, Paxillin, SHP2, p130CAS, plakoglobin, EPHA2, SHC,  $\beta$ -catenin, EGFR, EPS8, E-cadherin, and Cortactin, were confirmed to be tyrosine phosphorylated. There are no commercially available antibodies for two of the seven potentially novel phosphoproteins. They are WBP2 and NSFL1C. Consequently we subcloned the WBP2 and NSFL1C cDNA into pcDNA3.1 Directional TOPO vector that allowed expressed proteins to be tagged with a V5 sequence at the C terminus. NSFL1C or WBP2 cDNA was then transfected into HEK293T cells. After 72 h, HEK293T cells were untreated or treated with 1 mM PV for 15 min and lysed. Immunoprecipitation of NSFL1C and WBP2 was carried out in duplicate sets using anti-V5 antibodies, and each set was immunoblotted with either PY20H or anti-V5 antibodies. The immunoprecipitation results of WBP2, NSFL1C, and five other potentially novel tyrosine kinase substrates are presented in Fig. 3B. In summary, all seven proteins were validated to be authentic tyrosine phosphorylated proteins. No NSFL1C or WBP2 protein was immunoprecipitated by anti-V5 antibodies in the untransfected cells, and very little or no phosphotyrosine signal was observed in these untreated or treated cells. In contrast, NSFL1C and WBP2 proteins in transfected HEK293T cells were enriched by anti-V5 antibodies and were clearly tyrosine phosphorylated in PV-stimulated but not in untreated HEK293T cells.

From Fig. 3, A and B, we observed that the tyrosine phosphorylation levels of EGFR, EPS8, SHC (p66), and EPHA2 correlated with protein expression levels across all four cell lines. In contrast, there was no or limited correlation between tyrosine phosphorylation and expression trends for the rest of the proteins, indicating that the differential tyrosine phosphorylations observed for the rest of the proteins were likely due to aberrant tyrosine kinase or phosphatase activity rather than their expression levels. This is conceivable because EGFR, FAK, and EPHA2 tyrosine kinases were found to be aberrantly expressed in the MCF10AT model (Fig. 3A).

**Novel Tyrosine Kinase Substrates and Their Expression during Breast Cancer Progression**—From Fig. 3A, immunoblotting of the immunoprecipitates of the 12 well known phosphorylated signaling proteins with protein-specific antibodies revealed several proteins that displayed differential expression during breast cancer progression. They include FAK, p130CAS, EPHA2, EGFR, EPS8, and E-cadherin. Epithelial-mesenchymal transition is a cell transformation process that is widely accepted to be part of cancer development. This process is marked by loss of E-cadherin, an epithelial cell marker, and increased expression of Vimentin, a marker for cells of mesenchymal origin. Indeed we were able to observe

diminished E-cadherin expression (Fig. 3A) and increased expression of Vimentin in the MCF10AT model (Supplemental Fig. 2). In addition, we detected increasing Vimentin tyrosine phosphorylation from A1 through 1a cells (Supplemental Fig. 2A), and this supports the detection of enhanced tyrosine phosphorylated Vimentin in breast tumors compared with normal tissues that we made a couple of years ago (15). FAK tyrosine kinase and EPHA2 and EGFR receptor tyrosine kinases are popularly implicated in various cancers (27). Other proteins such as p130CAS have been found to be aberrantly expressed in human or metastatic cancers (28). Collectively these observations reflect the robustness of the model and analytical systems used in detecting molecular changes during cancer development.

Among the novel tyrosine kinase substrates for which we have antibodies, only SLC4A7 (decreased) and TOLLIP (increased) displayed differential expression during breast cancer progression (Fig. 3B). As we did not have antibodies against WBP2 and NSFL1C, we resorted to real time PCR. Fig. 3C shows the mRNA levels of NSFL1C and WBP2 across the four cell lines in the breast cancer progression model. These were done in four replicates, and the bars represent the minimum and maximum values, *i.e.* the true mRNA level could be as low or as high as these values. For our classification, a gene was considered overexpressed if its minimum value in one condition was higher than the maximum value of the same gene in another condition. The reverse was true. Based on this system, there was an increase and decrease of NSFL1C and WBP2 mRNA, respectively, when normal (A1) and high grade cancer (1a) cells were compared. Whether the gene products or proteins display a similar trend would require further validation using antibodies, which are not available commercially at this moment. Interestingly NSFLC1 is a cofactor of VCP, which was also found to be differentially tyrosine phosphorylated in our study (Table I).

**SLC4A7 and TOLLIP Expression in Clinical Breast Samples**—Table II shows the expression data of the seven novel tyrosine kinase substrates obtained in this study as well as the published information on the subcellular localization, function, and potential involvement in cancer development of the seven novel phosphoproteins to provide an overview of the candidate proteins with respect to cancer/tumor biology. Despite the fact that the MCF10AT model has been used by various groups, a major limitation of such *in vitro*/animal model is that frequently these systems lack the physiological context present in the human body. Consequently we selected SLC4A7 and TOLLIP for further studies in clinical human samples because they represented novel tyrosine kinase substrates that were also observed to be differentially expressed during breast cancer progression. Although SPAG9 was not found to be significantly differentially expressed in the MCF10AT model, it also was chosen for a closer examination because of a study that reported amplification of the 17q12–23 region (where SPAG9 resides) in breast cancer (29). NSFL1C has

TABLE II

List of validated novel tyrosine phosphorylated proteins complete with reported subcellular location and function

Differential protein expression for the novel proteins detected in this study is compared with reported potential involvement in human cancers. JNK, c-JUN N-terminal kinase; IL, interleukin.

No.	Novel tyrosine phosphoprotein	Gene symbol	Subcellular location	Function	Potential involvement in human cancers (published) <sup>a</sup>	Ref.	Expression across MCF10AT model in current study (method used)
1	Cytoplasmic FMR1-interacting protein 1 isoform 1	<i>CYFIP1</i>	Cytoplasm; perinuclear region	Formation of membrane ruffles and lamellipodia protrusions and in axon outgrowth. Binds to F-actin and profilin.	Ovarian cancer (overexpressed; associated with size, stage, and disease).	45	No change
2	Toll-interacting protein	<i>TOLLIP</i>	Cytoplasm; transmembrane protein	Involved in IL-1 and Toll-like receptor signaling pathways. Inhibits cell activation by microbial products. Signal transducer activity.	— <sup>b</sup>	—	Increased (immunoblotting)
3	Splice isoform 1 of NSFL1 cofactor p47	<i>NSFL1C</i>	Predominantly nucleus; small proportion is cytoplasmic, associated with Golgi stacks	Reduce ATPase activity of VCP. Required for fragmentation during mitosis and for VCP-mediated reassembly after mitosis of the Golgi stacks. May have a role in VCP-mediated formation of transitional endoplasmic reticulum.	—	—	Increased (real time PCR)
4	c-JUN N-terminal kinase-interacting protein	<i>SPAG9</i>	Cytoplasm; perinuclear region	Mediates JNK signaling. Kinesin binding.	Primary breast cancer (amplification of overexpressed gene). Epithelial ovarian cancer (highly expressed; especially in early stage).	29 46	No change NA
5	40 S ribosomal protein S2	<i>RPS2</i>	Intracellular	Protein biosynthesis/translation. RNA binding. Structural constituent of ribosome.	High expression in mouse liver tumors and proliferating hepatocytes <i>in vitro</i> .	47	No change
6	WW domain-binding protein 2	<i>WBP2</i>	Unknown	Protein binding. WW-binding motif mediates interaction with NEDD4 by similarity.	—	—	Decreased (real time PCR)
7	Sodium bicarbonate cotransporter 3	<i>SLC4A7</i>	Multipass membrane protein in basolateral membrane	Regulates intracellular pH.	—	—	Decreased (immunoblotting)

<sup>a</sup>Searched using exhaustive literature search ([www.ncbi.nlm.nih.gov/entrez/query.fcgi](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)).

<sup>b</sup>—, the protein has never been reported to be implicated in human cancer.

been validated to be a novel phosphotyrosyl protein and shown by real time PCR to be overexpressed in breast cancer cells. However, it could not be evaluated further because of the lack of antibodies. A preliminary immunohistochemistry study on SPAG9, TOLLIP, and SLC4A7 was conducted on

five matched normal and breast tumors samples of Asian origin. Consistent with the results obtained from the MCF10AT model, differential expressions of TOLLIP and SLC4A7 but not SPAG9 were apparent between normal and tumor cells (only data for SPAG9 are shown in Supplemental

TABLE III  
Summary of IHC results on ErbB2, ER, TOLLIP, and SLC4A7 in clinical breast samples

N, normal; IDC, invasive ductal carcinoma.

Lesions	ErbB2 expression (percent frequency)	ER expression (percent frequency)	TOLLIP expression (percent frequency)	SLC4A7 expression (percent frequency)
N > IDC	9/22 (41%)	4/19 (21%)	7/20 (35%)	14/22 (64%)
N < IDC	13/22 (59%)	6/19 (32%)	5/20 (25%)	3/22 (14%)
N = IDC (positive)	0/22 (0%)	0/19 (0%)	7/20 (35%)	5/22 (23%)
N = IDC (negative)	0/22 (0%)	9/19 (47%)	1/20 (5%)	0/22 (0%)

Fig. 3). Subsequently we conducted IHC of SLC4A7 and TOLLIP on an additional 20 matched clinical samples.

In addition to SLC4A7 and TOLLIP, IHC on two other molecules highly relevant to clinical breast cancer, namely ER and ErbB2, was also performed for comparison. The complete raw IHC results are provided as Supplemental Table 4 and summarized in Table III. In 64% (14 of 22) of all cases, there was more SLC4A7 expression in normal compared with tumor sections. Note that three matched cases could not be analyzed because of the lack of ductal components in the normal section. On the other hand, 14% displayed more SLC4A7 in tumor compared with normal tissues, whereas 23% did not show differential expression between normal and diseased samples. This demonstrated that loss of SLC4A7 expression occurred in the majority of breast cancer cases studied. Representative data on the IHC of SLC4A7 showing more expression in normal ducts compared with tumor section are shown in Fig. 4, *top panels*. The differential expression of TOLLIP among normal and tumor sections was more heterogeneous with 25 and 35% showing elevated and diminished expression, respectively, in tumor compared with normal tissues. No differential expression was observed in the rest of the cases. Representative data showing more expression of TOLLIP in tumor compared with normal breast and vice versa are shown in Fig. 4, *middle and bottom panels*, respectively. From the IHC data, SLC4A7 was clearly localized in the cell membrane (especially in the tumor histological section), whereas TOLLIP was predominantly cytoplasmic. Both observations are consistent with the literature (Table II). ErbB2 was overexpressed in close to 60% (13 of 22) of the cases when matched normal and tumor samples were analyzed (Table III). In invasive carcinomas alone, ErbB2 expression could be detected (a score of 1+ and above) in 82.6% (19 of 23) of the cases (Supplemental Table 4). If we define a score of 2+ or more as overexpression, then ErbB2 was overexpressed in 30% (7 of 23) of all tumor cases. This is consistent with the 20–25% reported originally by Slamon *et al.* (30). Similar to existing reports, 53% of all cases analyzed here were positive (as defined by 1+ and above) for ER expression (31) (Table III). These data support the robustness of the data obtained.

*SPAG9, NSFL1C, WBP2, and TOLLIP Are Validated as Novel Tyrosine Phosphorylation Targets of EGF Signaling*—One striking observation that was consistent between the data in Table I and Fig. 3 was the marked deprivation of

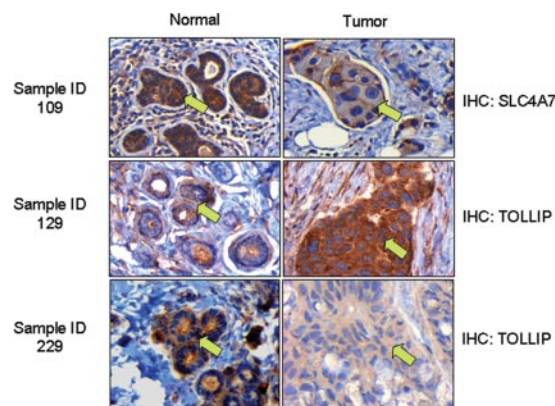


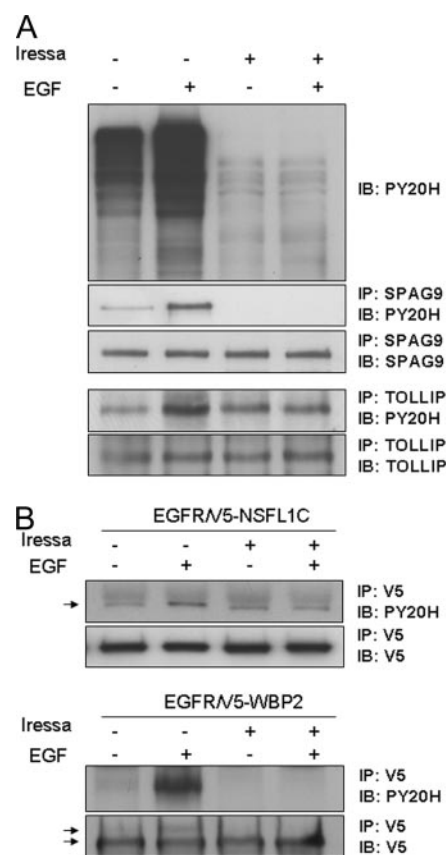
FIG. 4. Immunohistochemistry of SLC4A7 and TOLLIP in clinical breast samples. *Top panels*, SLC4A7 expression in matched normal and tumor breast samples from patient 109. The arrow in the *left panel* indicates a very strong SLC4A7 stain in normal acinar of lobules in the breast, whereas the arrow in the *right panel* points to invasive tumor cells showing a weaker signal. *Middle panels*, TOLLIP expression in matched normal and tumor breast samples from patient 129. Here the TOLLIP stain in the invasive ductal carcinoma cells shows a much stronger signal compared with the normal ductal cells. *Bottom panels*, TOLLIP expression in matched normal and tumor breast samples from patient 229. In this case, the signal for TOLLIP in the normal ductal cells is stronger than that in the invasive ductal carcinoma cells.

almost all tyrosine phosphorylation proteins examined (except FAK and EPHA2) in 1a cells. Coincidentally EGFR expression exhibited a trend that bore similarity to the tyrosine phosphorylation pattern of these proteins. In other words, loss of EGFR in late stage 1a cancer cells correlated with the diminished tyrosine phosphorylation levels of a considerable number of proteins detected here. Note that the expression of EGFR has since been further examined in close to 100 clinical breast samples. In 93% of the cases, there was more EGFR in normal epithelial cells than breast tumor cells suggesting that loss of EGFR during breast cancer progression is clinically relevant (48). Hence we postulated that EGFR might be the enzyme for the novel kinase substrates identified. To test this hypothesis, the tyrosine phosphorylation levels of these novel substrates were examined when cells were untreated or stimulated with EGF. To increase the specificity of the data, cells were either untreated or pretreated with Iressa, a highly selective EGFR inhibitor, at 10  $\mu$ M for 1 h prior to stimulation with EGF. Immunoprecipitations of SPAG9, CYFIP1, RPS2,



TOLLIP, and SLC4A7 were conducted in A431 cells untreated or treated with EGF at 50 ng/ml for 30 min. Immunoprecipitations were also performed for NSFL1C and WBP2 from EGFR-cotransfected HEK293T cells untreated or treated with EGF at 50 ng/ml for 2 min. The EGF time point selected for both cells had been optimized previously (data not shown). Immunoprecipitates were separately probed with anti-phosphotyrosine and protein-specific antibodies to reveal the tyrosine phosphorylation and expression levels of the proteins, respectively. Fig. 5A, *top panel*, shows the induction and abrogation of overall tyrosine phosphorylation in EGF-treated A431 cells. Immunoprecipitates of SPAG9 and TOLLIP showed a weak basal phosphotyrosine signal in unstimulated cells that increased upon EGF stimulation (Fig. 5A, *bottom panels*). Abrogation of the EGF-induced increase in tyrosine phosphorylation of these proteins by Iressa implied that these are authentic and specific targets of EGF signaling. Similarly induction of NSFL1C and WBP2 tyrosine phosphorylation by EGF in transfected HEK293T cells was also specifically inhibited by Iressa pretreatment (Fig. 5B). Note that the tyrosine phosphorylation levels of TOLLIP and NSFL1C were not completely abolished by Iressa suggesting that both proteins were constitutively phosphorylated probably by another tyrosine kinase in unstimulated cells. Hence SPAG9, TOLLIP, NSFL1C, and WBP2 were all authentic tyrosine phosphorylation targets of EGF signaling. On the other hand, tyrosine phosphorylation of CYFIP1, SLC4A7, and RPS2 was not induced by EGF treatment suggesting that they are probably downstream substrates of other growth factor signaling (data not shown).

Table I only shows a list of proteins that were induced to undergo tyrosine phosphorylation by pervanadate treatment. It lacks the biochemical context. To create significance out of otherwise static phosphoproteomics data, we constructed a biological interaction network (BIN) of the proteins identified in the PV-induced phosphotyrosine proteome (Fig. 6). This was generated with PathwayAssist from Ariadne Genomics. Databases that were used include the Kyoto Encyclopedia of Genes and Genomes pathway database (32–34), Biomolecular Interaction Network Database (35), and ResnetCore from Ariadne Genomics. We also incorporated information from the breast cancer category of the PathArt database which is proprietary to Jubilant Biosys. The proteins that could be networked were linked by various relationships such as protein interactions, modifications including phosphorylation, and regulation of expression. These relationships are color-coded, and the legends are provided next to the map. However, this does not mean that all the interactions took place within a single spatial and temporal situation, but it suggested strongly that the majority of the proteins identified in this study were integral parts of protein complexes. Two proteins with the most numbers of connectivity were EGFR and FAK1/PTK2, suggesting that these proteins represent central nodes in the mapped phosphoproteome.



**FIG. 5. Identifying novel tyrosine phosphorylation targets of EGF signaling.** A431 cells were serum-starved overnight and then untreated or treated with EGF at 50 ng/ml for 30 min. Where indicated, cells were also pretreated or not pretreated with Iressa at 10  $\mu$ M for 1 h. **A**, *top panel*, whole cell lysates were probed with anti-phosphotyrosine antibodies to show the induction and abrogation of tyrosine phosphorylation signals in EGF-treated cells. *Bottom panels*, immunoprecipitations (IP) using the indicated protein-specific antibodies were performed in duplicates on the lysates. The immunoprecipitates were subsequently immunoblotted (IB) with anti-phosphotyrosine antibodies conjugated to horseradish peroxidase (PY20H) and protein-specific antibodies to reveal the relative tyrosine phosphorylation and expression levels, respectively. **B**, WBP2 and NSFL1C cDNAs were cloned into pcDNA3.1 Directional TOPO vector that allowed expressed proteins to be tagged with a V5 sequence at the C terminus. NSFL1C or WBP2 cDNA was then co-transfected with EGFR into HEK293T cells. After 72 h, HEK293T cells were untreated or treated with 50 ng/ml for 2 min and lysed. Immunoprecipitations of NSFL1C and WBP2 were carried out in duplicate sets using anti-V5 antibodies. The immunoprecipitates were subsequently immunoblotted with anti-phosphotyrosine antibodies conjugated to horseradish peroxidase (PY20H) and V5 antibodies to reveal the relative tyrosine phosphorylation and expression levels, respectively. The *double arrows* in the *bottom panel* point to the native and possibly phosphorylated form of WBP2. The *single arrow* points to the position of tyrosine-phosphorylated NSFL1C.

Several proteins remained orphans because there is insufficient information in the database to link them to other proteins in the BIN. With the validation of SPAG9, NSFL1C, WBP2, and TOLLIP as authentic tyrosine kinase substrates

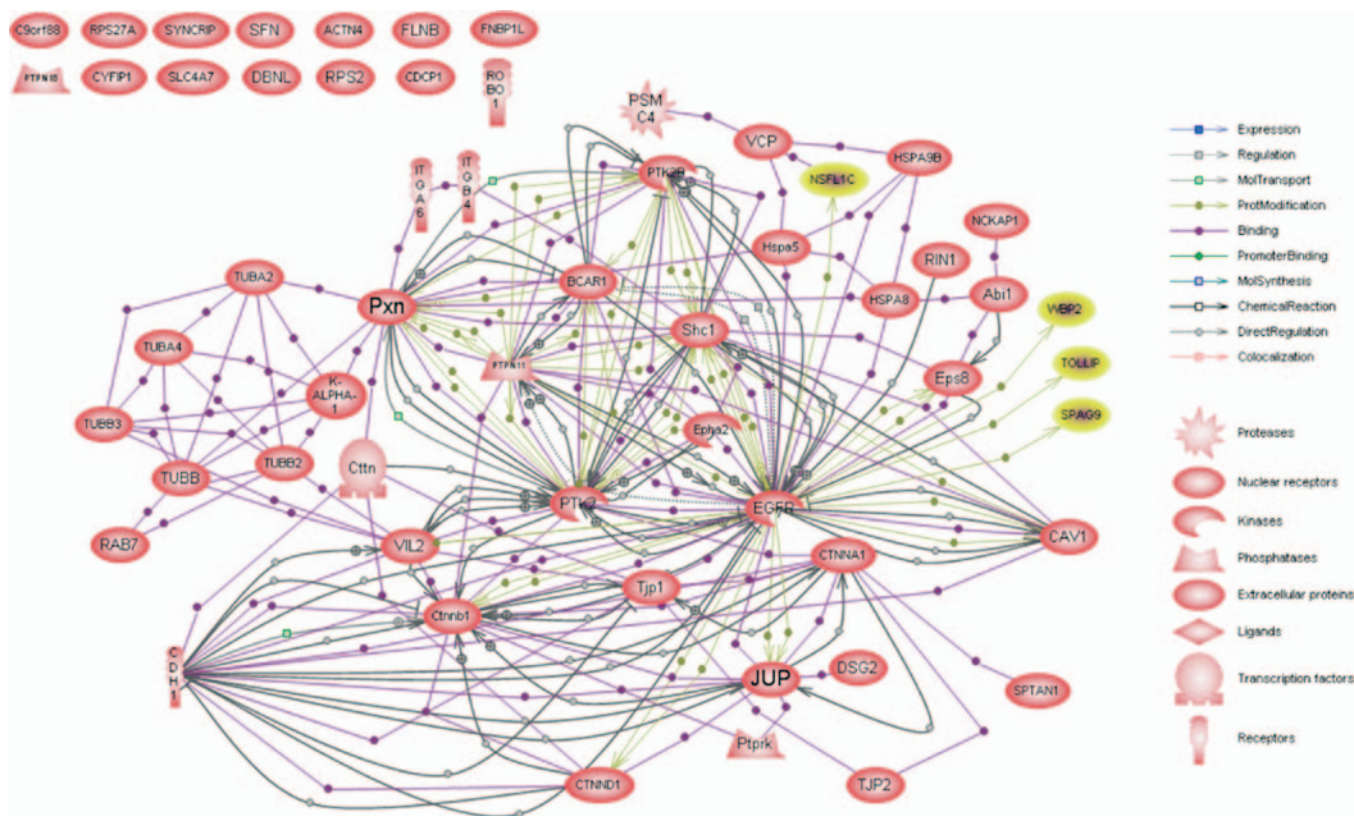


FIG. 6. **Biological interaction networking of tyrosine phosphorylated proteins identified in pervanadate-induced phosphotyrosine proteome in MCF10AT cells.** Proteins identified in Table I were imported into PathwayAssist, and an interaction map was generated with information from ResnetCore of Ariadne Genomics, Kyoto Encyclopedia of Genes and Genomes pathway database, Biomolecular Interaction Network Database, and PathArt breast cancer category from Jubilant Biosys. Each *node* represents either a protein entity or a control mechanism of the interaction. The legend of the interaction network is summarized on the *right* of the figure. SPAG9, TOLLIP, NSFL1C, and WBP2, which were identified as novel tyrosine phosphorylation substrates of EGF signaling, are *highlighted* on the map. The biological network can be accessed on the Internet (see Footnote 2). *SFN*, 14-3-3 protein sigma; *FLNB*, splice isoform 2 of filamin B; *DBNL*, drebin-like protein; *TUBB*, tubulin, beta polypeptide; *CTTN*, catenin; *PXN*, splice isoform gamma of paxilin; *JUP*, junction plakoglobin.

in EGF signaling, these proteins could now be formally linked to the EGFR node in the BIN.<sup>2</sup>

#### DISCUSSION

In this study, we used phosphoproteomics and the MCF10AT model as the biological and analytical systems, respectively, to identify expression changes in novel tyrosine kinase substrates during breast cancer progression. The xenograft-derived MCF10AT model comprises isogenic derivatives of the MCF10A mammary epithelial cell line that represent different stages of breast cancer progression (5, 36). The

MCF10AT cells were derived from MCF10A normal mammary epithelial cells transfected with T24 constitutively active RAS mutant. Although RAS mutations are rare in breast cancer, it is commonly activated in breast cancer by overexpressed growth factor receptors, which signal through RAS (37). Hence this model is likely to reflect a subset of human breast cancers and their progression. Indeed many observations obtained in this study supported this notion. They include the differential expression of known phosphotyrosyl proteins such as E-cadherin and Vimentin during epithelial-mesenchymal transition and of FAK, EPHA2, and p130CAS during cancer development. In addition, the loss of EGFR during breast cancer progression observed in this dataset has subsequently been validated in a separate study where close to 100 clinical breast samples were analyzed. About 93% of all cases expressed much less EGFR in tumor compared with normal tissues (48).

In a considerable number of proteins detected in this study, the relative levels of tyrosine phosphorylation as revealed by the protein-specific immunoprecipitation method in Fig. 3, A

<sup>2</sup> The BIN created in this study will be available on the Internet for public access at the time of publication of this study. The URL can be obtained by contacting the corresponding author. The web page contains hyperlinks through which the information on the proteins and the nature of their interactions with each other are provided. Where available, the references including a description of the experimental findings on which the BIN was based are also included. Such protein networking information should be useful for formulating testable hypotheses to understand the function of novel tyrosine kinase substrates.

and B, were dissimilar (except for EGFR, p130CAS, and p66 isoform of SHC) from that detected by the phospho-ITRAQ method in Table I. Note that datasets from both methods were derived from at least two experiments, and consistent results were obtained within each method. Such discrepancies were also observed in other phosphoproteomics studies. For example, in the study by Blagoev *et al.* (19), the degree of tyrosine phosphorylation of breast cancer anti-estrogen resistance and ARP3 at time points 0 and 20 min was not consistent between the phosphoproteomics and the immunochemistry methods. Such irregularity could be attributed to the fact that phosphoproteomics approaches using generic anti-phosphotyrosine antibodies for purification at the protein level are influenced by protein-protein interactions. In contrast, immunoprecipitation using protein-specific antibodies in principle enriches the entire population of the protein of interest and presents them for a more accurate determination of phosphorylation level following immunoblotting with anti-phosphotyrosine antibodies. With respect to this, phosphoproteomics involving purification at the peptide level as conducted by other groups (20, 22, 26) might circumvent the above mentioned limitation and offer a more accurate quantification of the phosphorylation levels of proteins. However, the potential difficulty of ionizing negatively charged phosphopeptides as is commonly encountered may present another challenge to accurate quantification.

Although several groups have conducted phosphoproteomics using various strategies, and some potentially novel phosphoproteins have been identified, very few have been validated (19–26). In contrast, this study identified and validated seven proteins, SPAG9, CYFIP1, RPS2, TOLLIP, SLC4A7, WBP2, and NSFLC1, to be authentic tyrosine kinase substrates. In addition, SPAG9, WBP2, TOLLIP, and NSFL1C were demonstrated to be authentic tyrosine phosphorylation targets of EGF signaling. Coupled to the three proteins that we discovered in our previous study (13), namely Endofin, DCBLD2, and CEP68, a total of seven novel phosphotyrosyl targets of EGF signaling have been discovered by our group. Further studies on these proteins will eventually lead to new insights on EGF signal transduction. Four of the seven novel tyrosine kinases, NSFL1C, WBP2, TOLLIP, and SLC4A7, displayed differential expression in the MCF10AT model at the protein (TOLLIP and SLC4A7) or mRNA (WBP2 and NSFL1C) level. Differential expression of TOLLIP and SLC4A7 was subsequently validated in clinical breast cancer samples. Although the sample size ( $n = 25$ ) tested was not large, it suffices in validating the *in vitro* observations made in this study. Consistent with the MCF10AT model, more than 2/3 of breast cancer cases analyzed displayed a diminished level of SLC4A7 compared with normal tissues. In contrast, only 25% of the cases revealed increased expression of TOLLIP when normal cells become cancerous. Although this supports the observation made in the MCF10AT model, the role of TOLLIP in breast cancer development is molecularly heterogeneous

because 35% of the cases showed the reverse trend. The reason for the divergence of phenotypic expression is not clear, although it is likely to be influenced by other molecular determinants, e.g. p53, leading to polyclonal expansion and heterogeneity in cancer cell populations. Further investigations with a larger sample size and correlation with clinical information will be conducted in a separate study to determine the potentially novel clinical utility of these candidate proteins for diagnosis and/or prognosis.

Little information is known about the TOLLIP protein except that it is a negative regulator of NF- $\kappa$ B and Toll signaling pathway during inflammation and innate immune response (38, 39). It has also been implicated in the trafficking of ubiquitinated proteins and IL-1 receptor (40, 41). The association of chronic inflammation and cancer has long been observed. Moreover inflammatory cytokines and chemokines such as tumor necrosis factor have been implicated in cancer progression and chemoresistance (42). Hence it is conceivable that TOLLIP might regulate cancer cell development. The role of SLC4A7 in cancer biology is less obvious. It is a transporter for bicarbonate ( $\text{HCO}_3^-$ ), which is a by-product of energy production.  $\text{HCO}_3^-/\text{CO}_2$  forms the major pH buffering system of our bodies. SLC4A7 probably facilitates  $\text{HCO}_3^-$  influx and efflux leading to cellular alkalization and acidification, respectively. Interestingly magnetic resonance has shown that although the intracellular compartment of tumor is kept at neutral pH the extracellular compartment is almost always more acidic by about 0.2 pH units (43). The scientific reason for the more acidic microenvironment in tumor is not known although excessive production of lactic acid by tumors during hypoxia and the greater distance required for the  $\text{H}^+$  to diffuse to capillaries in tumor compared with normal tissue leading to increased acidity have been suggested. It has also been hypothesized that hypoxia and acidity may contribute to cancer progression by providing a selection pressure for active selection of cancer phenotype, chemoresistance, and metastatic behavior (44). In fact, one of the popular alternative therapies for cancer but also one of the least understood concepts of nutrition is alkaline therapy through diet. Our observation on the loss of SLC4A7 bicarbonate transporter in the majority of the breast cancer cases generates a testable hypothesis that loss of SLC4A7 by malignant cells may contribute to acidity in the microenvironment that confers growth and development advantage to the cancer cells but that is unsuitable for normal or more differentiated cells.

This is the first time SLC4A7 and TOLLIP have been validated as novel tyrosine kinase substrates that are also implicated in human cancer development. Discovery of these phosphoproteins might not have been possible via a conventional proteomics approach. Our study therefore demonstrated the power of phosphoproteomics and model systems used here in identifying novel tyrosine phosphorylated proteins as potential agents in cancer initiation and progression. Further molecular and functional studies will certainly provide



novel insights into the roles of TOLLIP and SLC4A7 in the molecular etiology of cancer. Moreover detection of aberrant expression of TOLLIP and SLC4A7 in preneoplastic lesions suggests that they represent potential biomarkers that could complement mammography and histopathology, which are thwarted with uncertainties, for screening and early detection of breast cancer.

\* This work was supported by funding from the Singapore Cancer Syndicate; Biomedical Research Council; Agency for Science, Technology and Research (A\*STAR); and the Office of Life Sciences, National University of Singapore. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

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

§ Both authors contributed equally to this work.

§§ To whom all correspondence should be addressed: Oncology Research Inst., Centre for Life Sciences, 28 Medical Dr., 02-14C, Singapore 117456, Singapore. Tel.: 65-65161313; Fax: 65-68739664; E-mail: nmilyp@nus.edu.sg.

## REFERENCES

- Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. (2000) Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532–535
- Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordon-Cardo, C., Guise, T. A., and Massague, J. (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R. A. (2004) Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939
- Dawson, P. J., Wolman, S. R., Tait, L., Heppner, G. H., and Miller, F. R. (1996) MCF10AT: a model for the evolution of cancer from proliferative breast disease. *Am. J. Pathol.* **148**, 313–319
- Miller, F. R. (2000) Xenograft models of premalignant breast disease. *J. Mammary Gland Biol. Neoplasia* **5**, 379–391
- Chong, B. E., Hamler, R. L., Lubman, D. M., Ethier, S. P., Rosenspire, A. J., and Miller, F. R. (2001) Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with mass spectrometric analysis. *Anal. Chem.* **73**, 1219–1227
- Chong, B. E., Lubman, D. M., Miller, F. R., and Rosenspire, A. J. (1999) Rapid screening of protein profiles of human breast cancer cell lines using non-porous reversed-phase high performance liquid chromatography separation with matrix-assisted laser desorption/ionization time-of-flight mass spectral analysis. *Rapid Commun. Mass Spectrom.* **13**, 1808–1812
- Chong, B. E., Lubman, D. M., Rosenspire, A., and Miller, F. (1998) Protein profiles and identification of high performance liquid chromatography isolated proteins of cancer cell lines using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **12**, 1986–1993
- Starcevic, S. L., Diotte, N. M., Zukowski, K. L., Cameron, M. J., and Novak, R. F. (2003) Oxidative DNA damage and repair in a cell lineage model of human proliferative breast disease (PBD). *Toxicol. Sci.* **75**, 74–81
- Starcevic, S. L., Elferink, C., and Novak, R. F. (2001) Progressive resistance to apoptosis in a cell lineage model of human proliferative breast disease. *J. Natl. Cancer Inst.* **93**, 776–782
- Tang, B., Vu, M., Booker, T., Santner, S. J., Miller, F. R., Anver, M. R., and Wakefield, L. M. (2003) TGF- $\beta$  switches from tumor suppressor to pro-metastatic factor in a model of breast cancer progression. *J. Clin. Invest.* **112**, 1116–1124
- Worsham, M. J., Pals, G., Schouten, J. P., Miller, F., Tiwari, N., van Spaendonck, R., and Wolman, S. R. (2006) High-resolution mapping of molecular events associated with immortalization, transformation, and progression to breast cancer in the MCF10 model. *Breast Cancer Res. Treat.* **96**, 177–186
- Chen, Y., Low, T. Y., Choong, L. Y., Ray, R. S., Tan, Y. L., Toy, W., Lin, Q., Ang, B. K., Wong, C. H., Lim, S., Li, B., Hew, C. L., Sze, N. S., Druker, B. J., and Lim, Y. P. (2007) Phosphoproteomics identified Endofin, DCBLD2, and KIAA0582 as novel tyrosine phosphorylation targets of EGF signaling and Iressa in human cancer cells. *Proteomics* **7**, 2384–2397
- Lim, Y. P., Diong, L. S., Qi, R., Druker, B. J., and Epstein, R. J. (2003) Phosphoproteomic fingerprinting of epidermal growth factor signaling and anticancer drug action in human tumor cells. *Mol. Cancer Ther.* **2**, 1369–1377
- Lim, Y. P., Wong, C. Y., Ooi, L. L., Druker, B. J., and Epstein, R. J. (2004) Selective tyrosine hyperphosphorylation of cytoskeletal and stress proteins in primary human breast cancers: implications for adjuvant use of kinase-inhibitory drugs. *Clin. Cancer Res.* **10**, 3980–3987
- Lim, Y. P., Low, B. C., Lim, J., Wong, E. S., and Guy, G. R. (1999) Association of atypical protein kinase C isotypes with the docker protein FRS2 in fibroblast growth factor signaling. *J. Biol. Chem.* **274**, 19025–19034
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169
- Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2006) Technical, experimental and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* **6**, 821–827
- Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22**, 1139–1145
- Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648
- Thelemann, A., Petti, F., Griffin, G., Iwata, K., Hunt, T., Settinar, T., Fenyo, D., Gibson, N., and Haley, J. D. (2005) Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. *Mol. Cell. Proteomics* **4**, 356–376
- Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., and White, F. M. (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell. Proteomics* **4**, 1240–1250
- Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., and Peters, E. C. (2004) Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **76**, 2763–2772
- Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23**, 94–101
- Tao, W. A., Wollschied, B., O'Brien, R., Eng, J. K., Li, X. J., Bodenmiller, B., Watts, J. D., Hood, L., and Aebersold, R. (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat. Methods* **2**, 591–598
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5860–5865
- Blume-Jensen, P., and Hunter, T. (2001) Oncogenic kinase signalling. *Nature* **411**, 355–365
- Cabodi, S., Tinnirello, A., Di Stefano, P., Bisaro, B., Ambrosino, E., Castellan, I., Sapino, A., Arisio, R., Cavallo, F., Forni, G., Glukhova, M., Silengo, L., Altruda, F., Turco, E., Tarone, G., and Defilippi, P. (2006) p130Cas as a new regulator of mammary epithelial cell proliferation, survival, and HER2-neu oncogene-dependent breast tumorigenesis. *Cancer Res.* **66**, 4672–4680
- Willis, S., Hutchins, A. M., Hammet, F., Ciciulla, J., Soo, W. K., White, D.,

- van der Spek, P., Henderson, M. A., Gish, K., Venter, D. J., and Armes, J. E. (2003) Detailed gene copy number and RNA expression analysis of the 17q12–23 region in primary breast cancers. *Genes Chromosomes Cancer* **36**, 382–392
30. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177–182
31. Fernandopulle, S. M., Cher-Siangang, P., and Tan, P. H. (2006) Breast carcinoma in women 35 years and younger: a pathological study. *Pathology* **38**, 219–222
32. Kanehisa, M. (1997) A database for post-genome analysis. *Trends Genet.* **13**, 375–376
33. Kanehisa, M., and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30
34. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**, D354–D357
35. Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., and Hogue, C. W. (2001) BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **29**, 242–245
36. Santner, S. J., Dawson, P. J., Tait, L., Soule, H. D., Eliason, J., Mohamed, A. N., Wolman, S. R., Heppner, G. H., and Miller, F. R. (2001) Malignant MCF10CA1 cell lines derived from premalignant human breast epithelial MCF10AT cells. *Breast Cancer Res. Treat.* **65**, 101–110
37. Malaney, S., and Daly, R. J. (2001) The ras signaling pathway in mammary tumorigenesis and metastasis. *J. Mammary Gland Biol. Neoplasia* **6**, 101–113
38. Didierlaurent, A., Brissoni, B., Velin, D., Aebi, N., Tardivel, A., Kaslin, E., Sirard, J. C., Angelov, G., Tschopp, J., and Burns, K. (2006) Tollip regulates proinflammatory responses to interleukin-1 and lipopolysaccharide. *Mol. Cell. Biol.* **26**, 735–742
39. Burns, K., Clatworthy, J., Martin, L., Martinon, F., Plumpton, C., Maschera, B., Lewis, A., Ray, K., Tschopp, J., and Volpe, F. (2000) Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nat. Cell Biol.* **2**, 346–351
40. Brissoni, B., Agostini, L., Kropf, M., Martinon, F., Swoboda, V., Lippens, S., Everett, H., Aebi, N., Janssens, S., Meylan, E., Felberbaum-Corti, M., Hirling, H., Gruenberg, J., Tschopp, J., and Burns, K. (2006) Intracellular trafficking of interleukin-1 receptor I requires Tollip. *Curr. Biol.* **16**, 2265–2270
41. Katoh, Y., Imakagura, H., Futatsumori, M., and Nakayama, K. (2006) Recruitment of clathrin onto endosomes by the Tom1-Tollip complex. *Biochem. Biophys. Res. Commun.* **341**, 143–149
42. Chen, R., Alvero, A. B., Silasi, D. A., and Mor, G. (2007) Inflammation, cancer and chemoresistance: taking advantage of the toll-like receptor signaling pathway. *Am. J. Reprod. Immunol.* **57**, 93–107
43. Stubbs, M., McSheehy, P. M., and Griffiths, J. R. (1999) Causes and consequences of acidic pH in tumors: a magnetic resonance study. *Adv. Enzyme Regul.* **39**, 13–30
44. De Milito, A., and Fais, S. (2005) Tumor acidity, chemoresistance and proton pump inhibitors. *Future Oncol.* **1**, 779–786
45. Leoutsakou, T., Talieri, M., and Scorilas, A. (2006) Expression analysis and prognostic significance of the SRA1 gene, in ovarian cancer. *Biochem. Biophys. Res. Commun.* **344**, 667–674
46. Garg, M., Chaurasiya, D., Rana, R., Jagadish, N., Kanojia, D., Dudha, N., Kamran, N., Salhan, S., Bhatnagar, A., Suri, S., Gupta, A., and Suri, A. (2007) Sperm-associated antigen 9, a novel cancer testis antigen, is a potential target for immunotherapy in epithelial ovarian cancer. *Clin. Cancer Res.* **13**, 1421–1428
47. Kowalczyk, P., Woszczynski, M., and Ostrowski, J. (2002) Increased expression of ribosomal protein S2 in liver tumors, posthepatomized livers, and proliferating hepatocytes in vitro. *Acta Biochim. Pol.* **49**, 615–624
48. Choong, L. Y., Lim, S., Loh, M. C. S., Man, X., Chen, Y., Toy, W., Tan, Y. L., Pang, M., Chen, C. S., Poonepalli, A., Hande, P. M., Tan, P. H., Salto-Tellez, M., Wong, C. Y., Shah, N., Druker, B. J., and Lim, Y. P. (2007) Progressive loss of EGFR in a sub-population of breast cancers—implications in target-directed therapeutics. *Mol. Cancer Ther.* **6**, 1–15