A streamlined platform for high-content functional proteomics of primary human specimens

Nadim Jessani[1,4,5], Sherry Niessen[1,5], BinQing Q Wei[1,5], Monica Nicolau[2], Mark Humphrey[1], Youngran Ji[2], Wonshik Han[3], Dong-Young Noh[3], John R Yates III[1], Stefanie S Jeffrey[2] & Benjamin F Cravatt[1]

Achieving information content of satisfactory breadth and depth remains a formidable challenge for proteomics. This problem is particularly relevant to the study of primary human specimens, such as tumor biopsies, which are heterogeneous and of finite quantity. Here we present a functional proteomics strategy that unites the activity-based protein profiling and multidimensional protein identification technologies (ABPP-MudPIT) for the streamlined analysis of human samples. This convergent platform involves a rapid initial phase, in which enzyme activity signatures are generated for functional classification of samples, followed by in-depth analysis of representative members from each class. Using this two-tiered approach, we identified more than 50 enzyme activities in human breast tumors, nearly a third of which represent previously uncharacterized proteins. Comparison with cDNA microarrays revealed enzymes whose activity, but not mRNA expression, depicted tumor class, underscoring the power of ABPP-MudPIT for the discovery of new markers of human disease that may evade detection by other molecular profiling methods.

Postgenomic research promises to deliver humankind into the age of molecular medicine, where diseases are diagnosed and treated through the detection and pharmacological targeting of pathology-linked proteins[1]. The success of this biomedical vision hinges on the rapid translation of genomic information into the discovery of new protein biomarkers and therapeutic targets. Proteomics aims to facilitate this process by developing new methods for the parallel analysis of many proteins in samples of high biological complexity[2].
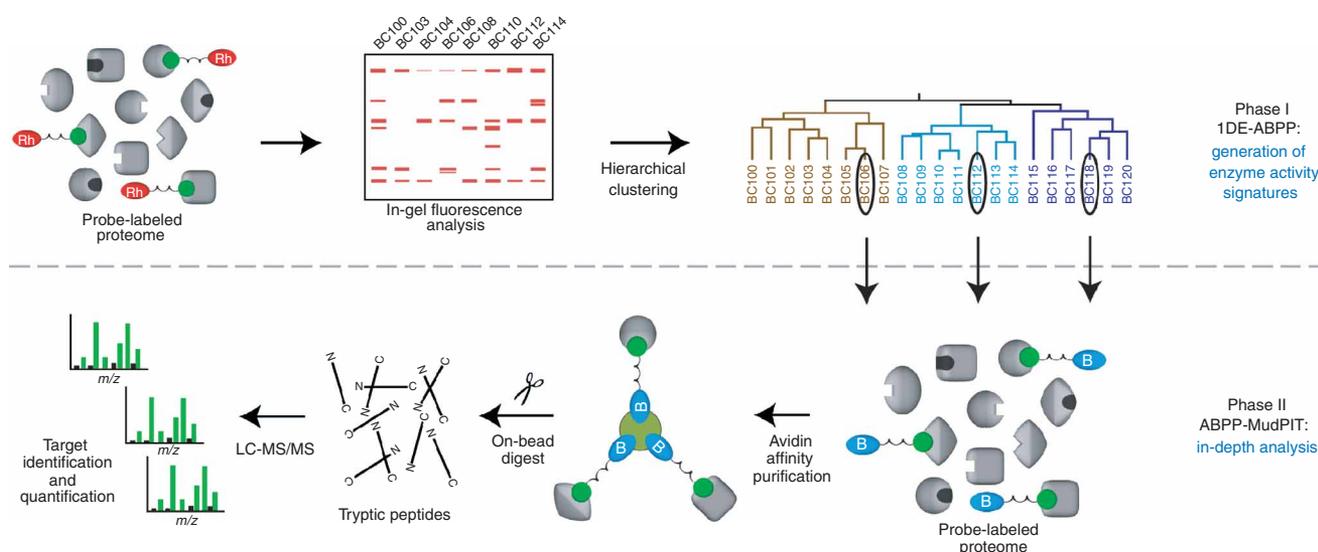
To date, several strategies have emerged for the large-scale discovery of differentially expressed proteins in cells, tissues and fluids, including two-dimensional gel electrophoresis (2DE)[3], liquid chromatography–mass spectrometry (LC-MS; for example, isotope-coded affinity tagging (ICAT)[4] and multidimensional-protein identification technology (MudPIT)[5]) and MS-based proteomic pattern analysis[6,7]. A survey of the data acquired using each of these approaches reveals an intriguing paradox. Multidimensional separation methods have remarkable sensitivity and resolution, facilitating the identification of low-abundance proteins in complex samples[8,9]. These techniques, however, are time-consuming (many hours per sample), difficult to perform in parallel and require large quantities of proteome (∼1 mg or greater) for optimal analysis. Alternative proteomic approaches have emerged that have greater throughput (matrix-assisted laser desorption ionization (MALDI) imaging[6] and surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF)[7] spectrometry), albeit at the expense of sensitivity. Furthermore, each of these methods, by focusing on measurements of protein expression, provides only an indirect readout of protein activity, which is regulated by a myriad of post-translational events *in vivo*[10].

The troublesome trade-off between the breadth and depth of information procurable by different techniques has frustrated efforts to implement an efficient, systematic strategy for high-content proteomics. These concerns are particularly relevant to the study of primary human specimens, which are often heterogeneous and of limited quantity[11], and therefore mandate interrogation by proteomic methods that can achieve both high throughput (to deal with sample heterogeneity) and sensitivity (to identify rarely expressed and post-translationally regulated proteins in samples of finite amount).

To address these important issues, we have introduced a chemical technology, referred to as activity-based protein profiling (ABPP)[12,13] that uses active site–directed probes to read out the functional state of many enzymes directly in whole proteomes. ABPP probes selectively label active enzymes, but not their inactive (for example, zymogen or inhibitor-bound) forms, facilitating the characterization of changes in enzyme activity that occur without corresponding alterations in protein or transcript expression[14,15]. Additionally, because ABPP probes label enzymes based on shared catalytic properties rather than mere expression level, they provide exceptional access to low-abundance portions of the proteome, which can be read out even in simple formats like one-dimensional gel electrophoresis (1DE)[14]. Using such 1DE-ABPP methods, hundreds of proteomes can be analyzed per day by a single academic group. These features have allowed the routine application

Figure 1 | Integration of the ABPP and MudPIT methods for high-content functional proteomics of primary human specimens. A two-tiered platform is presented. In the first phase, proteomic samples are treated with a rhodamine (Rh)-tagged ABPP probe and resolved by 1DE to provide enzyme activity signatures (visualized by in-gel fluorescence scanning). Hierarchical clustering of these molecular profiles is then used for the functional classification of samples. In the second phase, representative members of each sample class are treated with a biotinylated ABPP probe (B) and the labeled proteins are enriched by binding to avidin-conjugated beads and subjected to on-bead trypsin digestion. The resulting tryptic peptide mixtures are analyzed by multidimensional LC-MS (MudPIT), and the levels of specific enzyme activities are estimated by spectral counting[16,17].

of ABPP to many biological samples in parallel, permitting, for example, the classification of human cancer cell lines into phenotypically relevant groups based on their shared enzyme activity profiles[14]. Nonetheless, despite these valuable attributes, the inherent resolution and sensitivity limits of 1DE have, to date, precluded comprehensive access to many of the enzyme activities present in individual proteomic samples.

Here we present an advanced platform that unites the ABPP and MudPIT technologies, providing a general and efficient route for achieving both breadth and depth in functional proteome analysis. We apply this integrated approach to more than 30 primary human breast tumor and normal breast specimens, and find several enzyme activities that are elevated in specific breast tumor classes.

RESULTS

A tiered platform for high-content functional proteomics

For the efficient analysis of large numbers of primary human tissue specimens, we envisaged a two-tiered platform (**Fig. 1**). In a rapid initial phase, 1DE-ABPP would be used to generate enzyme activity signatures for the functional classification of proteomic samples (phase I). ABPP-MudPIT would then be performed on representative members of each class to provide an in-depth analysis of the enzyme activities present in these samples (phase II). The success of this integrated strategy hinges on several key issues. First, gel-based enzyme activity profiles need to provide sufficient information to permit proper biological classification of human samples. Additionally, this initial phase of proteome analysis should apply to any tissue specimen, even those of very limited quantity. In the second phase, ABPP-MudPIT experiments need to be quantitative, such that the relative levels of enzyme activities can be determined with confidence. Inspired by recent evidence suggesting a linear relationship between a protein's abundance and its degree of sampling during an LC–tandem MS (LC-MS/MS) run[16,17], we proposed that

spectral counting of probe-enriched enzymes might address this final issue. We set out to test this integrated ABPP-MudPIT platform by comparatively profiling a set of 33 primary human breast tumor and normal breast tissue biopsies. Clinical parameters such as estrogen receptor (ER) and progesterone receptor (PR) status were determined for tumor samples to assist in the interpretation of proteomic profiles (see **Supplementary Table 1** online for a list of clinical data).

Phase I: rapid analysis of breast tumors by 1DE-ABPP

For tumor profiling experiments, we used fluorophosphonate (FP)-based ABPP probes that target the serine hydrolase superfamily[14,18] (see **Supplementary Fig. 1** online for structures of FP probes). Serine hydrolases are a large and diverse class of enzymes that comprise about 1% of the human proteome, including several enzymes implicated in cancer[19–21].

We carried out the phase I analysis using 5–10-μm frozen tissue sections, portions of which were also allocated for cDNA microarray experiments to allow a direct comparison of transcript and enzyme activity profiles. Homogenized tissue sections (soluble and membrane; 12 μg protein/sample) were treated with a rhodamine-tagged FP probe (FP-rhodamine (2 μM); **Supplementary Fig. 1**) and analyzed by 1DE-ABPP. Equivalent probe-labeled samples were also treated with PNGaseF to generate 'deglycosylated' proteomes, which, when analyzed in comparison to their native counterparts, enhances the detection of some enzymes[14].

Gel profiles of membrane enzyme activities from representative breast tissue samples are shown in **Figure 2a**, where brackets highlight examples of enzymes elevated in tumors (see **Supplementary Fig. 2** online for a complete set of tissue profiles). 1DE analyses typically resolved about 15 probe-labeled enzyme activities per tumor sample, which generated a data set that was too large for manual analysis (total number of enzyme activities, ∼2,000:

15 enzymes × 2 proteomic fractions (membrane and soluble) × 2 glycosylation states (glycosylated and deglycosylated) × 33 samples). Therefore, a suite of computational programs was developed to automate the alignment and quantification of 1DE data. Because our goal was to rapidly compare proteomes before the time-consuming step of target identification, each probe-labeled enzyme was identified solely by its predicted molecular mass (see **Supplementary Fig. 3** online for a representative gel showing molecular mass annotations for probe-labeled enzymes). The relative activity levels of each enzyme were then calculated across all proteomic samples, and the resulting normalized data set was analyzed with an unsupervised hierarchical clustering algorithm.

Breast tissue specimens clustered into five major groups based on their membrane enzyme activity profiles (**Fig. 2b**), at least three of which represented clinically relevant breast tissue subtypes: normal

breast, ER(+)/PR(+) breast cancer and ER(−)/PR(−) breast cancer. The other two groups contained a mixture of tumors that were positive for either ER or PR, as well as some double-positive and double-negative tumors. In contrast, and consistent with previous findings from ABPP of human cancer cell lines[14], serine hydrolase activity profiles of the soluble proteome did not effectively differentiate the breast tissue specimens and, therefore, were not examined further. A comparison of the membrane ABPP dendrogram with one generated by cDNA microarray analysis (Y.J., M.N. and S.S.J., unpublished data) revealed a strong correlation (**Fig. 2c**). For example, six of the eight tumors found in the ABPP ER(−)/PR(−) group were also part of the ER(−)/PR(−) group defined by cDNA microarrays. The ABPP-derived ER(+)/PR(+) tumor group also had five specimens in common with the cDNA microarray ER(+)/PR(+) group. Finally, two
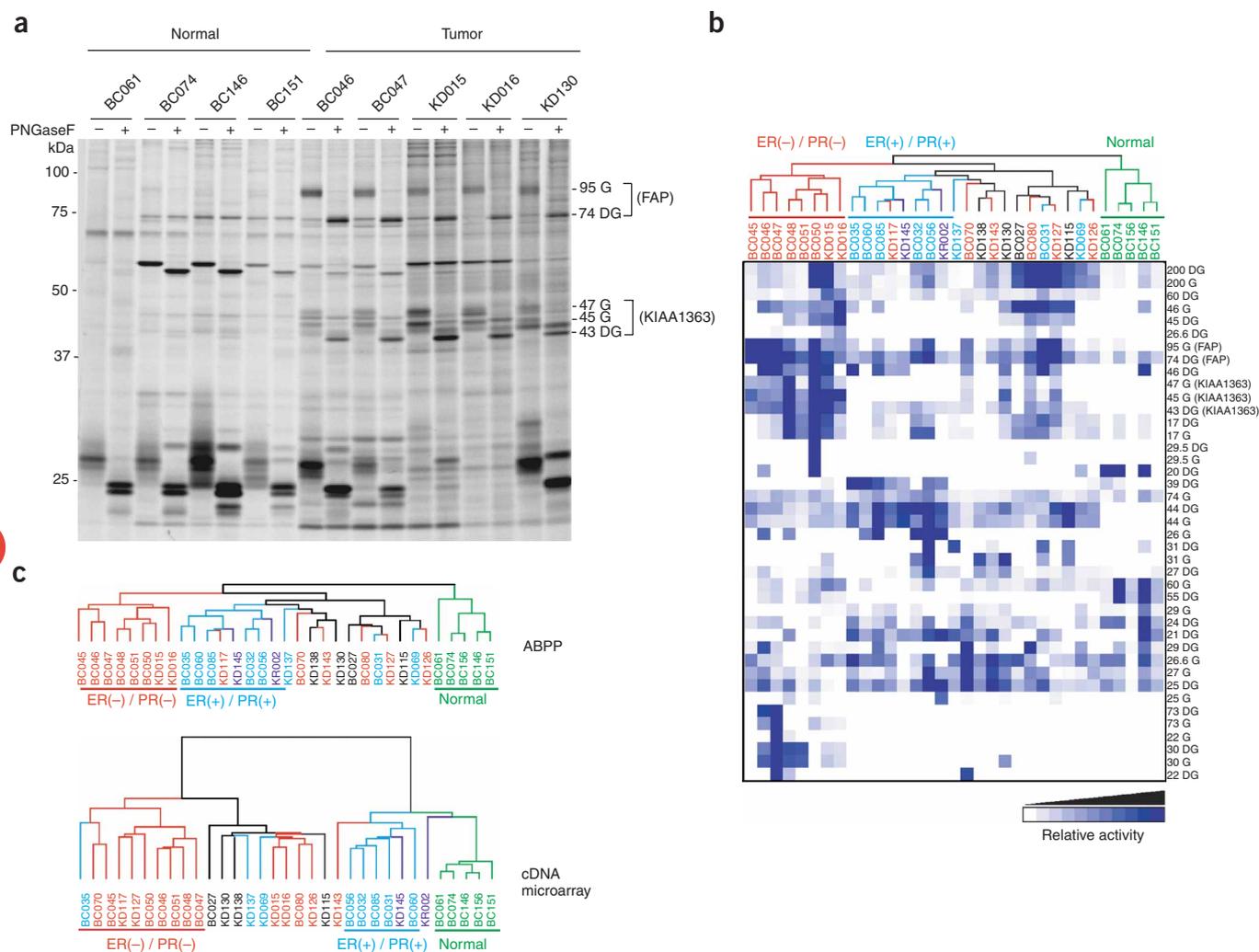
Figure 2 | Enzyme activity signatures of primary human breast tumors and normal breast tissue. (**a**) Representative gel showing the membrane serine hydrolase activity profiles of breast tissue specimens treated with FP-rhodamine before and after deglycosylation with PNGaseF (fluorescent image shown in grayscale). Representative enzyme activities elevated in tumor specimens are indicated. See **Supplementary Figure 2** for a complete set of enzyme activity profiles of the breast tissue specimens. (**b**) Hierarchical clustering analysis of membrane enzyme activity profiles. The intensity of blue color scales directly with the relative activity of each enzyme among the specimens. Enzyme activities are designated by their predicted molecular mass (kDa) and their presence in glycosylated (G) or deglycosylated (DG) samples. Sample names and dendrogram are color-coded by ER/PR status; double-positive (light blue), double-negative (red), mixed positive and negative (black), and unknown (dark blue). Normal samples are in green. (**c**) Comparison of the classifications of breast specimens by enzyme activity (top) and gene expression (bottom) profiles.

Table 1 | Enzymes with differential activity among breast tumor samples, as determined by ABPP-MudPIT[a]

	Spectral counts									Relative activity		
	Normal (N)			ER(−)/PR(−) (DN)			ER(+)/PR(+) (DP)					
	BC151	BC156	Average	BC45	BC48	Average	BC35	BC56	Average	DN/N	DN/DP	DP/N
KIAA1363[b]	10	11	10	53	105	79	14	37	25	7.9	3.2	2.5
FAP[b]	7	5	6	60	86	73	13	17	15	12	4.9	2.5
PAF-AH2[b]	5	0	2	18	18	18	5	2	3	9	6	1.5
CE1[c]	571	1,273	922	30	61	45	64	81	72	0.05	0.6	0.08
HSL[c]	125	113	119	5	15	10	30	0	15	0.08	0.7	0.1
LPL[c]	22	16	19	0	0	0	0	0	0	<0.05	–	<0.05
DPPIV[c]	15	12	13	0	2	1	3	0	1	0.08	1	0.08
Thrombin[c]	17	38	27	0	0	0	0	0	0	<0.04	–	<0.04

[a]Enzymes that had greater than threefold differences in activity in one tissue class relative to the other two classes are listed. Average values are rounded down to the nearest integer value. LPL, lipoprotein lipase. For a complete list of enzyme activities identified by ABPP-MudPIT, see **Supplementary Table 2**. [b]Enzymes elevated in ER(−)/PR(−) tumors. [c]Enzymes elevated in normal breast. DN, double negative; DP, double positive.

tumors of unknown ER and PR status (KR002 and KD145) had profiles similar to ER(+)/PR(+) tumors in both ABPP and cDNA microarray experiments, suggesting that they may belong to this breast cancer class.

Results from this initial phase indicated that 1DE-ABPP could generate enzyme activity profiles sufficiently rich in information content to classify primary human tumor specimens into biologically relevant subtypes, and do so using only minute quantities of proteome (12–25 µg protein/sample). We next transitioned to the second phase of the ABPP-MudPIT platform, in which individual members of each breast cancer class were selected for in-depth proteomic profiling.

### Phase II: in-depth analysis of tumors by ABPP-MudPIT

For the second phase, two tumor specimens were selected from each of the following tissue classes: normal breast, ER(+)/PR(+) tumors and ER(−)/PR(−) tumors. These specimens were chosen because they were available in sufficient quantity for in-depth analysis (0.75 mg protein/sample) using an advanced gel-free version of ABPP that integrates this method with MudPIT. The ABPP-MudPIT approach involves first treatment of proteomes with a biotinylated activity-based probe (for example, FP-biotin; **Supplementary Fig. 1**), and then enrichment of probe-labeled

proteins using avidin-conjugated beads, on-bead trypsin digestion and multidimensional LC-MS/MS analysis of the resulting tryptic peptide mixture (**Fig. 1**, lower scheme). We expected ABPP-MudPIT to have several advantages over gel-based methods for in-depth proteome analysis, including enhanced resolution (owing to multidimensional separation), sensitivity (owing to affinity enrichment of probe-labeled targets) and coupled target detection and identification. Additionally, we considered that the relative quantity of enzyme activities could be estimated by ABPP-MudPIT using spectral counting methods[16,17].

Using ABPP-MudPIT, we identified over 50 serine hydrolase activities in membrane proteomes, the vast majority of which were not observed in control reactions, in which the FP-biotin probe was left out. These enzymes included proteases, lipases, esterases and at least 15 uncharacterized hydrolases (see **Supplementary Table 2** online for a full list of enzymes). In addition to this diverse collection of serine hydrolases, several proteasome subunits (β1, 4–7 and 9) were also identified in probe-treated, but not in control samples, suggesting that this class of amino-terminal threonine proteases may also be susceptible to specific labeling by FP reagents.

For comparative quantitation, the spectral counts for each hydrolase were averaged for the two samples from each breast cancer class, and only those enzymes that had greater than a threefold difference in spectra number ('activity') in one tissue class relative to the other two were considered of potential interest. Based on these criteria, several enzymes were identified that had altered levels of activity among the breast cancer specimens (**Table 1**). For example, three enzyme activities, fibroblast activation protein (FAP or seprase), KIAA1363 and platelet-activating factor acetylhydrolase 2 (PAF-AH2) were elevated in ER(−)/PR(−) tumors compared to either ER(+)/PR(+) tumors or normal breast tissue. Conversely, multiple enzyme activities were higher in normal breast, including thrombin, dipeptidylpeptidase IV (DPPIV) and hormone-sensitive lipase (HSL). Finally, two hydrolase activities (fatty acid synthase and carboxylesterase 1 (CE1))
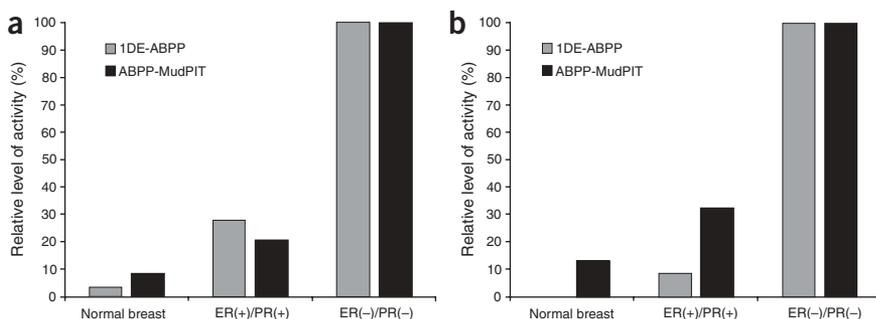
Figure 3 | Comparison of enzyme activity levels measured by 1DE-ABPP and MudPIT-ABPP. (**a,b**) Similar relative levels of FAP (**a**) and KIAA1363 (**b**) were estimated by 1DE-ABPP and MudPIT-ABPP in the normal breast, ER(+)/PR(+) tumor and ER(−)/PR(−) tumor samples. Bars represent the average activity levels of the two samples examined for each breast tissue class: BC151, BC156 (normal); BC35, BC56 (ER(+)/PR(+)); BC45, BC48 (ER(−)/PR(−)).
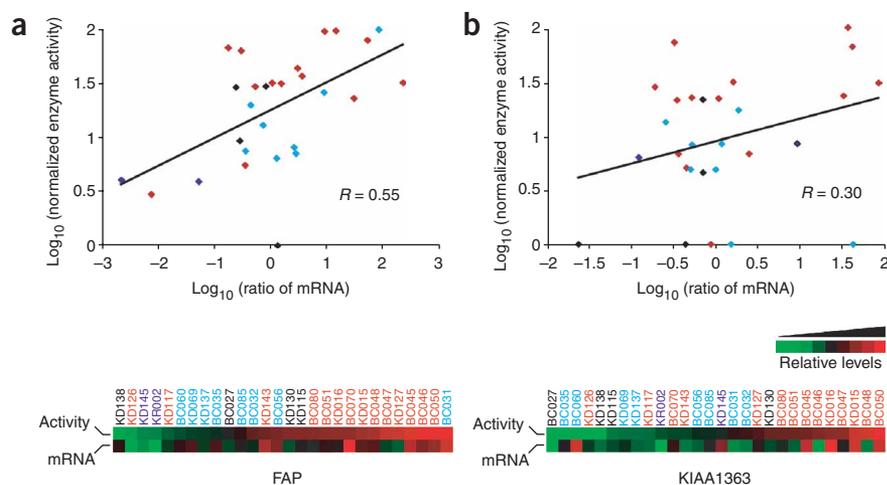
Figure 4 | Comparison of activity and mRNA expression levels for FAP and KIAA1363. (**a**) A significant correlation between FAP activity and mRNA levels was observed across the breast tumor samples (Pearson's correlation coefficent $R = 0.55$, $P < 0.001$). (**b**) Activity and mRNA levels for KIAA1363 were not significantly correlated across the breast tumor samples ($R = 0.30$, $P > 0.10$). The mRNA expression data represent the logarithm values of signals relative to a reference standard (Universal Human Reference RNA, Stratagene). For the heat maps, the logarithm values of both relative enzyme activity (0–100%) and mRNA expression were centered at their respective mean and scaled to between -1 and 1 to enhance visualization.

method provided a similar estimate of the relative activity levels of these hydrolases across breast tissue specimens (**Fig. 3**). Considering that signals for most of the other differentially expressed enzyme activities shown in **Table 1** were not discernible by 1DE-ABPP, these data highlight the impressive increase in information content garnered by higher-resolution ABPP-Mud-PIT experiments.

We next compared the relative activity levels of KIAA1363 and FAP to their gene expression profiles as measured by cDNA microarrays across the entire set of breast tumor samples. We observed a good correlation for FAP: ER(−)/PR(−) tumors had both high activity and transcript levels of this protease (**Fig. 4a**). In contrast, the activity and transcript profiles for KIAA1363 were largely uncorrelated (**Fig. 4b**), suggesting that this enzyme activity may be regulated by post-transcriptional mechanisms in breast cancer. Finally, a survey of the gene expression patterns of enzyme activities enriched in normal breast indicated that HSL and CE1, but not thrombin, had good correlation between their respective transcript and activity levels across the set of specimens analyzed by ABPP-MudPIT (**Supplementary Fig. 4** online).

DISCUSSION

Here we have presented a two-tiered functional proteomic strategy that integrates the ABPP and MudPIT methods for the streamlined analysis of primary human specimens. First-phase enzyme activity signatures allowed the logical selection of representative members of breast tumor classes for in-depth proteomic analysis. Notably, these 1DE-ABPP studies required minimal quantities of sample (12–25 μg of protein per sample), indicating that they should be applicable to a wide range of primary human specimens. In the future, other high-throughput and sample-conserving techniques could be considered for phase I analysis, including a recently described capillary electrophoresis approach for ABPP[22].

The second phase exploited the exceptional sensitivity and resolution of multidimensional LC-MS/MS (MudPIT), while at the same time preserving the high-content functional information procured by ABPP. ABPP-MudPIT identified more than 50 enzyme activities in breast tumor samples, including three enzymes that were highly elevated in ER(−)/PR(−) cancers: FAP/seprase, a cell surface-expressed protease of the dipeptidylpeptidase (DPP) clan[23]; PAF-AH2, a lipase that degrades the endogenous signaling molecule PAF[24]; and KIAA1363, an integral membrane hydrolase of unknown function. Comparisons of ABPP and cDNA microarray results revealed that, for certain enzymes (for example, KIAA1363) activity and mRNA levels were largely uncorrelated. These findings underscore the versatility of ABPP-MudPIT, which can be applied to any proteomic fraction to discover disease-associated enzyme activities that may evade detection by other molecular profiling methods.

were identified in both probe-treated and control reactions, but their spectral counts were approximately tenfold higher in the former samples (**Supplementary Table 2**). As these enzymes also had the highest overall spectral counts in probe-treated samples, they likely represent abundant active proteins specifically labeled by FP probes.

The reproducibility of ABPP-MudPIT was evaluated by analyzing a single ER(−)/PR(−) tumor (BC48) in triplicate and averaging the data sets. The activities of more than 80% of the enzymes identified in this tumor (34 of 41) were measured with a standard error of the means (s.e.m.) of <30% (**Supplementary Table 3** online). If the lowest-abundance targets were excluded (that is, enzymes with an average spectrum count of less than ten), then greater than 90% of the remaining enzyme activities (23 of 24) were quantified with a s.e.m. of <30%, and greater than 80% (21 of 24) were quantified with a s.e.m. of <20%. These results indicate that ABPP-MudPIT has good reproducibility and is suitable for the comparative quantitation of enzyme activities between samples, especially if interpretations are restricted to relatively large–fold differences in activity (for example, threefold or greater).

We sought to corroborate our spectral counting data with an independent method. Re-examination of the 1DE-ABPP data identified a set of three glycosylated enzyme activities that were highly elevated in the ER(−)/PR(−) breast tumors (95, 47 and 45 kDa). These targets were enriched using FP-biotin and avidin chromatography, separated by SDS-PAGE, digested in-gel with trypsin and identified by LC-MS/MS analysis as FAP (95 kDa) and two forms of KIAA1363 (47 and 45 kDa). The two forms of KIAA1363 reflected different glycosylation states, as they were converted into a single lower molecular mass species upon treatment with PNGaseF (43 kDa, **Fig. 2a**). Thus, both 1DE-ABPP and MudPIT-ABPP identified FAP and KIAA1363 as enzyme activities elevated in ER(−)/PR(−) breast tumors. Notably, each proteomic

The discovery of a set of enzyme activities elevated in ER(−)/PR(−) tumors is notable because these tumors are generally regarded as the most aggressive form of breast cancer and are often refractory to conventional treatment[25]. FAP, which is elevated in expression in several human cancers[26–28], has recently been shown to promote breast tumor growth in animal models[20]. Less is understood about the roles PAF-AH2 and KIAA1363 might have in cancer pathogenesis, although we have found that the latter enzyme activity is elevated in invasive cancer lines from several different tumor types[14]. Testing the function of new cancer-associated enzymes like KIAA1363 will likely require the development of selective inhibitors for these proteins, a goal that may be facilitated by competitive ABPP methods[29].

ABPP-MudPIT has some limitations that should be noted. First, the second phase of analysis still requires relatively large quantities of proteome. We were able to circumvent this problem because sufficient amounts of representative members of each tumor class were available for MudPIT analysis. In cases in which none of the individual human specimens are of adequate quantity, members of each sample class could be pooled prior to analysis by ABPP-MudPIT. Such an approach might offer the additional advantage of enriching for enzyme activities shared among related samples, while diminishing the signal intensities of proteins with large sample-to-sample variation. A second shortcoming of ABPP-MudPIT is that neither of its core technologies appears suitable for direct implementation in the clinic. Emerging antibody micro-array formats for ABPP[30] may eventually address this issue. Finally, in this study, we have focused on the proteomic analysis of one enzyme class, the serine hydrolases. As ABPP probes are now available for more than a dozen enzyme families[31–36], it is important that ABPP-MudPIT also accommodate the analysis of these additional enzymes without a significant increase in sample requirement. We anticipate that this objective may be accomplished in the second phase of analysis, in which multiple probes could be added to individual proteomic samples to profile several enzyme classes concurrently. In this manner, ABPP-MudPIT should approach the ultimate goal of providing a universal platform for the comprehensive analysis of enzyme activities in any primary human specimen.

METHODS

Patients and tumor specimens. A total of 28 breast tumors (all ductal) and 5 normal breast tissue specimens were obtained from 33 individuals. Specimens came from either the Stanford Hospital (designated BC) or Seoul National University College of Medicine (designated KD). Cases were accrued in accordance with local institutional review board guidelines. The distribution of cases according to patient source, lymph node status, tumor grade, patient age, the expression of hormone receptors (ER and PR) and other prognostic markers, are listed in **Supplementary Table 1**. For proteomic studies, tumor specimens were obtained as 5–10-μm frozen sections (∼4–6 sections per tumor) which provided, on average, ∼60–100 μg of total protein per sample. Some tumor specimens, as well as normal breast tissue, were also obtained as larger frozen blocks of tissue that provided greater quantities of protein (∼1–3 mg). All samples were frozen in either liquid nitrogen or on dry ice within 20 min after devascularization and stored at −80 °C before proteomic or genomic analysis.

1DE-ABPP of breast tissue specimens. Tumor and normal breast specimens were processed and analyzed by 1DE-ABPP as described in **Supplementary Methods**. To increase the throughput of gel data analysis, an automated procedure was developed using in-house computer programs and commercial software (see **Supplementary Methods** for more details). Fluorescence intensity signals (integrated optical density; IOD) for each probe-labeled protein were compared across all samples; the highest intensity value was defined as 100%, and intensities of this enzyme activity observed in the other samples were normalized to the maximum. Signals with a maximum intensity of less than twofold above the background (IOD = 800) were discarded as noise. Hierarchical clustering of the resulting data set was performed using averaged linkage method with Pearson correlation coefficient as the similarity metric[37], and the results visualized using TreeView software (http://rana.lbl.gov/EisenSoftware.htm). The clustering was performed unweighted (that is, all enzyme activities receive equal weight in the computation).

ABPP-MudPIT of breast tissue specimens. Membrane proteomes (∼0.5 mg/ml, 0.77 mg total protein per sample) were treated with FP-biotin (5 μM) for 2 h at 20–25 °C. Proteomes were then solubilized with 1% Triton X-100 by rotating at 4 °C for 1 h. Enrichment of FP-labeled proteins was achieved using avidin-conjugated beads as previously described[38] and processed for MudPIT analysis as described in **Supplementary Methods**.

cDNA microarray analysis of breast tissue specimens. Gene expression analysis was performed as described previously[24,39] using 42,000-feature cDNA microarrays that contained approximately 24,000 clones that corresponded to approximately 13,000 distinct Unigene clusters. Detailed protocols are available (**Supplementary Methods** and http://www.stanford.edu/group/sjeffreylab/). A more complete analysis of the cDNA microarray results will be reported elsewhere.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

1. Kramer, R. & Cohen, D. Functional genomics to new drug targets. *Nat. Rev. Drug Discov.* **3**, 965–972 (2004).
2. Patterson, S.D. & Aebersold, R. Proteomics: the first decade and beyond. *Nat. Genet.* **33**, 311–323 (2003).
3. Patton, W.F., Schulenberg, B. & Steinberg, T.H. Two-dimensional electrophoresis: better than a poke in the ICAT? *Curr. Opin. Biotechnol.* **13**, 321–328 (2002).
4. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
5. Washburn, M.P., Wolters, D. & Yates, J.R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247 (2001).
6. Chaurand, P., Sanders, M.E., Jensen, R.A. & Caprioli, R.M. Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. *Am. J. Pathol.* **165**, 1057–1068 (2004).
7. Petricoin, E.F. & Liotta, L.A. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr. Opin. Biotechnol.* **15**, 24–30 (2004).

8. Durr, E. *et al.* Direct proteomic mapping of the lung microvascular endothelial cell surface *in vivo* and in cell culture. *Nat. Biotechnol.* **22**, 985–992 (2004).

9. Gygi, S.P., Rist, B., Griffin, T.J., Eng, J. & Aebersold, R. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J. Proteome Res.* **1**, 47–54 (2002).

10. Kobe, B. & Kemp, B.E. Active site-directed protein regulation. *Nature* **402**, 373–376 (1999).

11. Zangar, R.C., Varnum, S.M., Covington, C.Y. & Smith, R.D. A rational approach for discovering and validating cancer markers in very small samples using mass spectrometry and ELISA microarrays. *Dis. Markers* **20**, 135–148 (2004).

12. Jessani, N. & Cravatt, B.F. The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **8**, 54–59 (2004).

13. Speers, A.E. & Cravatt, B.F. Chemical strategies for activity-based proteomics. *ChemBioChem.* **5**, 41–47 (2004).

14. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B.F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer invasiveness. *Proc. Natl. Acad. Sci. USA* **99**, 10335–10340 (2002).

15. Jessani, N. *et al.* Carcinoma and stromal enzyme activity profiles associated with breast tumor growth *in vivo*. *Proc. Natl. Acad. Sci. USA* **101**, 13756–13761 (2004).

16. Pang, J.X., Ginanni, N., Dongre, A.R., Hefta, S.A. & Opitek, G.J. Biomarker discovery in urine by proteomics. *J. Proteome Res.* **1**, 161–169 (2002).

17. Liu, H., Sadygov, R.G. & Yates, J.R., III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).

18. Liu, Y., Patricelli, M.P. & Cravatt, B.F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**, 14694–14699 (1999).

19. Andreasen, P.A., Kjoller, L., Christensen, L. & Duffy, M.J. The urokinase-type plasminogen activator system in cancer metastasis: a review. *Int. J. Cancer* **72**, 1–22 (1997).

20. Huang, Y., Wang, S. & Kelly, T. Seprase promotes rapid tumor growth and increased microvessel density in a mouse model of human breast cancer. *Cancer Res.* **64**, 2712–2716 (2004).

21. Gallardo-Williams, M.T., Maronpot, R.R., Wine, R.N., Brunssen, S.H. & Chapin, R.E. Inhibition of the enzymatic activity of prostate-specific antigen by boric acid and 3-nitrophenyl boronic acid. *Prostate* **54**, 44–49 (2003).

22. Okerberg, E.S. *et al.* High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl. Acad. Sci. USA* **102**, 4996–5001 (2005).

23. Chen, W.T., Kelly, T. & Ghersi, G. DPPIV, seprase, and related serine peptidases in multiple cellular functions. *Curr. Top. Dev. Biol.* **54**, 207–232 (2003).

24. Arai, H. Platelet-activating factor acetylhydrolase. *Prostaglandins Other Lipid Mediat.* **68–69**, 83–94 (2002).

25. Sorlie, T. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).

26. Huber, M.A. *et al.* Fibroblast activation protein: differential expression and serine protease activity in reactive stromal fibroblasts of melanocytic skin tumors. *J. Invest. Dermatol.* **120**, 182–188 (2003).

27. Scanlan, M.J. *et al.* Molecular cloning of fibroblast activation protein alpha, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers. *Proc. Natl. Acad. Sci. USA* **91**, 5657–5661 (1994).

28. Kelly, T., Kechelava, S., Rozypal, T.L., West, K.W. & Korourian, S. Seprase, a membrane-bound protease, is overexpressed by invasive ductal carcinoma cells of human breast cancers. *Mod. Pathol.* **11**, 855–863 (1998).

29. Leung, D., Hardouin, C., Boger, D.L. & Cravatt, B.F. Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* **21**, 687–691 (2003).

30. Sieber, S.A., Mondala, T.S., Head, S.R. & Cravatt, B.F. Microarray platform for profiling enzyme activities in complex proteomes. *J. Am. Chem. Soc.* **126**, 15640–15641 (2004).

31. Adam, G.C., Sorensen, E.J. & Cravatt, B.F. Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* **20**, 805–809 (2002).

32. Greenbaum, D. *et al.* Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1**, 60–68 (2002).

33. Barglow, K.T. & Cravatt, B.F. Discovering disease-associated enzymes by proteome reactivity profiling. *Chem. Biol.* **11**, 1523–1531 (2004).

34. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B.F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA* **101**, 10000–10005 (2004).

35. Chan, E.W., Chattopadhaya, S., Panicker, R.C., Huang, X. & Yao, S.Q. Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J. Am. Chem. Soc.* **126**, 14435–14446 (2004).

36. Liu, Y. *et al.* Wortmannin, a widely used phosphoinositide 3-kinase inhibitor, also potently inhibits mammalian polo-like kinase. *Chem. Biol.* **12**, 99–107 (2005).

37. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).

38. Kidd, D., Liu, Y. & Cravatt, B.F. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005–4015 (2001).

39. Zhao, H. *et al.* Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell* **15**, 2523–2536 (2004).