Reverse-phase protein array (RPPA) is a high-throughput antibody-based targeted proteomics platform that can quantify hundreds of proteins in thousands of samples derived from tissue or cell lysates, serum, plasma, or other body fluids. Protein samples are robotically arrayed as microspots on nitrocellulose-coated glass slides. Each slide is probed with a specific antibody that can detect levels of total protein expression or post-translational modifications, such as phosphorylation as a measure of protein activity. Here we describe workflow protocols and software tools that we have developed and optimized for RPPA in a core facility setting that includes sample preparation, microarray mapping and printing of protein samples, antibody labeling, slide scanning, image analysis, data normalization and quality control, data reporting, statistical analysis, and management of data. Our RPPA platform currently analyzes ~240 validated antibodies that primarily detect proteins in signaling pathways and cellular processes that are important in cancer biology. This is a robust technology that has proven to be of value for both validation and discovery proteomic research and integration with other omics data sets.

**KEY WORDS:** protein array analysis, post-translational, proteomics, software, validation study

**INTRODUCTION**

Proteomics is a valuable approach in cancer and other diseases to identify biomarkers for diagnosis and targeted therapy through the analysis of global protein expression and activation status of proteins. The correlation between DNA, RNA, and protein changes is only about 40%, requiring an ability to assess the functional proteome. Reverse-phase protein array (RPPA) is an antibody-based microarray technology that performs immunoassays on thousands of samples simultaneously, including tissue or cell lysates, serum, plasma, or other body fluids. As a targeted proteomic platform, RPPA is complementary to mass-spectrometry profiling and is capable of quantifying relative levels of total protein expression as well as post-translation modifications, such as phosphorylation as a measure of protein activity. RPPA was initially designed for identification of alterations in protein-signaling pathways in cancer but is easily scalable for assessment of protein pathways in other diseases and cell biology processes.

RPPA is a robust, highly reproducible proteomic technology that enables concordant interrogation of multiple protein-signaling pathways and their functional status in a microarray format. The reverse order of arraying protein targets and incubating replicate arrays, each with a single well-characterized antibody, contributes to the robustness of the assay because it enables the use of appropriate controls and conditions for each specific antibody. RPPA is a higher-throughput, lower-cost procedure compared with mass-spectrometry proteomics and is thus more amenable to rapid analysis of large numbers of experimental samples. It also has the advantage of using serum or cell/tissue lysates without prefractionation steps. RPPA can detect proteins over a wide dynamic range with small sample volumes and amounts of protein (<5 µg protein), and it is highly sensitive and capable of detecting low-abundance regulatory proteins that are often difficult to measure by mass-spectrometry profiling. Because of these properties, RPPA is utilized to validate candidate protein biomarkers or protein pathways discovered by mass spectrometry or gene-expression profiling. RPPA has also proven to be valuable as a discovery tool for identifying protein-signaling pathways in cancer progression and pathways vulnerable to therapies or responsible for resistance.

In this report we describe workflow protocols and in-house software tools for RPPA that have been developed and
successfully applied in a core facility (Fig. 1). This workflow includes procedures for sample preparation, microarray mapping and printing of protein samples, antibody labeling, slide scanning, image analysis, data normalization and quality control (QC), data reporting, statistical analysis, and management of data.

**MATERIALS AND METHODS**

**Cell lines and antibodies**

Antibodies and related information are listed in Supplemental Table S1 with catalog and Research Resource Identifier numbers. Please see our core’s website for future updates (https://www.bcm.edu/academic-centers/dan-d-duncan-comprehensive-cancer-center/research/cancer-shared-resources/reverse-phase-protein-array/antibodies-list). Most cell lines used in this study were obtained from the American Type Culture Collection (ATCC) (Manassas, VA, USA) (Supplemental Table S2). The ATCC provides Short Tandem Repeat analysis for authentication of cell lines. Other cell lines were authenticated by Short Tandem Repeat analysis by the Cytogenetics and Cell Authentication Core at MD Anderson Cancer Center (Houston, TX, USA). For preparation of protein lysates, cell lines were grown for minimal passages under conditions and media formulations specified by ATCC and tested negative for mycoplasma.

**Analysis software**

The steps in the overall RPPA workflow depicted in Fig. 1 are detailed in the following sections. To support the RPPA workflow, a software suite was developed and comprises 3 parts (outlined in Fig. 3): 1) RPPA Setup Tool: a Python application with graphical interface using PyQT5 for protein array design and layout and plate setup; 2) RPPA ImGrid Tool: a Matlab application with a graphical interface to facilitate image analysis; and 3) a suite of data analysis Python scripts designed to be run using a Linux/MacOS command-line interface and to perform data normalization, QC, data reporting, and basic statistical analysis. The software suite is available via the open-source platform github at https://github.com/coarfa/RPPAworkflowBCM.

**Step 1: antibody validation**

Because RPPA technology is highly dependent on the quality of the antibodies, a key feature is rigorous antibody validation for specificity and selectivity. Our criteria for antibody validation include an immunoblot assay that results in a single protein band (or specific multiple bands for protein isoforms) of correct molecular size expected with known positive and negative controls (cell or tissue) and an equivalent performance under RPPA assay conditions. Most of the validated antibodies are from commercial vendors as listed in Supplemental Table S1, and we have built an inventory of ~240 such antibodies. Initial selection of commercial antibodies is based on vendor data demonstrating immunoblot detection of endogenous protein as a predominant single band of the correct molecular mass in a known positive cell line or tissue. Lack of immunoblot signal or substantial reduction in a known negative cell or tissue type or after genetic knockdown or knockout is preferred if vendor data are available. We subsequently perform in-house testing and validation. Because of well-known potential variations between antibody batches, we also test replacement batches in our inventory by RPPA and immunoblot. Examples of validation with antibodies to

---

**FIGURE 1**

Overview of the RPPA workflow. After protein lysate extraction from cells or tissue, samples are transferred to 384-well microarray plates; individual slides are printed and then labeled with validated antibodies. Slides are scanned at different intensities, and protein levels are determined after normalization and QC. Normalized data are then analyzed according to individual experimental designs. Ab neg, negative antibody.
Lysine Acetyltransferase 2A (KAT2A), Sirtuin 6 (SIRT6), and DNA Methyltransferase 3 Beta (DNMT3B) by both RPPA and immunoblot analyses are shown in Fig. 2. By immunoblotting, the antibodies detected a predominant protein band of expected size in known positive cell lines and reduced or lack of reaction with cells known to have low or no expression of the protein of interest. Single bands were obtained with the antibodies to KAT2A and DNMT3B; however, SIRT6 gave a doublet band. The doublet likely represents different protein isoforms generated by alternative splicing known to produce multiple transcript variants encoding different isoforms of SIRT6 (https://www.uniprot.org/uniprot/Q8N6T7) (Fig. 2). Importantly, the specific intensity signals by RPPA analysis of same cell lines corresponded well with immunoblot results (Fig. 2).

Step 2: sample preparation

Protein lysates are prepared from cultured cells or tissues with RPPA lysis buffer containing Tissue Protein Extraction Reagent (78510; Thermo Fisher Scientific, Waltham, MA, USA) as previously described.9, 14 For cell cultures, we either use cell pellets or directly lyse cells in the cell culture dishes. We use \(5 \times 10^6\) cells prepared fresh in a 10-cm cell culture dish after removing media, which is followed by washing the cells 3 times with cold PBS, scraping or pipetting lysis buffer over cells (200 \(\mu\)l/\(5 \times 10^6\) cells), transferring the cells and buffer to a 1.5-ml tube, and lysing the cells with intermittent vortexing at 4°C for 30 min. Fresh cell pellets are harvested by trypsinization, scraping, or other method per project requirement by individual research lab. Cells are collected in a 1.5-ml tube and washed twice with cold PBS. The pellets without residual PBS are suspended in RPPA lysis buffer (200 \(\mu\)l/\(5 \times 10^6\) cells) with intermittent vortexing at 4°C for 30 min. For tissues, samples (10–15 mg) are snap-frozen on dry ice, and tissue pieces are transferred to precooled 2-ml Sample Tubes RB (990381; Qiagen, Germantown, MD, USA) with a 5-mm stainless steel bead (69989; Qiagen) in \(\sim 250 \mu\)l RPPA lysis buffer (\(\sim 1:20\) w/v). The tubes are then placed on a precooled adapter for the TissueLyser II (85300; Qiagen), and tissues are homogenized in a cold room (4°C) for 2 min at 23 Hz.

Cell culture or tissue lysates are centrifuged at 20,000 g for 15 min at 4°C, and supernatant containing soluble proteins is transferred to a new microcentrifuge tube. Centrifugation is repeated 2–3 times for cell culture samples and 3–5 times for tissue samples. Protein concentration of the soluble protein supernatant is quantified by bicinechoninic assay, with an optimal target concentration between 1.1 and 3.0 mg/ml. The lysates are diluted to a final protein concentration of 0.5 mg/ml in the volume of 120 \(\mu\)l in 1X SDS sample buffer [62.3 Tris/HCl (pH 6.8), 10% glycerol, 2% SDS, 0.01% bromophenol blue, and 2.5% \(\beta\)-mercaptoethanol (M3148; MilliporeSigma, Burlington, MA, USA)]. The samples are heated to 100°C for 8 min and centrifuged at 20,000 g for 2 min to remove any additional particulate material. Two aliquots of 50 \(\mu\)l lysate per sample are stored in a cardboard box at \(-80°C\) for RPPA (one for printing, one as a backup or for future printing). Our RPPA samples are usually thawed only 1 time and do not exceed 3 freeze/thaw cycles. The procedures and reagents above are optimal for sample microarray printing in terms of

![FIGURE 2](images/figure2.png)

Examples of antibody validation immunoblots and RPPA analysis of KAT2A (A), SIRT6 (B), and DNMT3B (C) antibodies with total protein lysates from the cell lines shown. RPPA-normalized SIs are graphed for each antibody (left half), and corresponding immunoblot analyses are shown as images (right half). SIRT6 has double bands from the 2 isoforms of the protein (B). IB, immunoblot; NE, nuclear extract.
providing appropriate viscosity and protein solubility for robotic pins, for quality of arrayed spots, and for proper range of total protein concentrations. Bromophenol blue in the SDS sample buffer can be observed under the microscope to visualize the morphology of printed spots and to check slide layout accuracy before they are processed for antibody labeling. This serves as a troubleshooting step in our QC process.

Step 3: setup and printing of RPPA slides

On the day of sample printing, 1 aliquot of every frozen protein lysate (at 0.5 mg/ml) and 1 aliquot of each control sample in SDS sample buffer are thawed at room temperature, heated at 100°C for 2 min, centrifuged briefly to bring down the volume, transferred to 384-well microarray plates (X7022UN; Molecular Devices, Sunnyvale, CA, USA), and sealed with VIEWseal plate covers (676070; Greiner Bio-One, Monroe, NC, USA). Microarray spot locations on slides and 384-well locations of protein lysates are mapped using the RPPA Setup Tool Python program (Fig. 4). A Microsoft Excel spreadsheet is prepared that contains experimental protein lysate identifiers grouped by individual core users. Experimental samples are arranged on the array, with the constraint that samples of the same user are grouped together in runs of contiguous spots and with each sample corresponding to technical triplicate spots. The RPPA Setup Tool plate and layout utility generates the design for individual plates, loads the curated spots/matrix assignment, and selects between configurations with 6, 9, or 12 plates. The RPPA Setup Tool also generates the layout design files (Fig. 4). The Aushon 2470 Arrayer (Aushon BioSystems/Quanterix, Billerica, MA, USA) with a 40-pin (185-μm) configuration is used to spot experimental samples and control lysates onto Grace Bio-Labs ONCY-TEAVID nitrocellulose-coated glass slides (GBL305177; Grace Bio-labs, Bend, OR, USA). The virtual contents of 384-well plates are then uploaded into an Aushon 2470

FIGURE 3
Overview of the RPPA workflow analysis software. I) The protein array setup uses a suite of user-friendly Python utilities for array design, array to plate layout, and per-slide antibody-based renaming. II) Image analysis is carried out using the Matlab image-processing library, including interactive utilities for slide gridding and spot annotation. III) Data analysis is carried out using Python command-line tools for normalization, QC, data reporting, and basic statistical analysis.

FIGURE 4
The RPPA Setup Tool. 1) Protein array design is created using the Array tool based on the RPPA sample list and the desired number of plates. Select Array tab (A1), upload input file (B), select number of plates (C), generate output file (D). 2) The plates and layouts design function is based on the array design and the desired number of plates. The steps are similar to those from the previous tab, with the output file from the Array tool being used as the input file for this step. The output for this tool consists of 2 comma-separated values (CSV) files: plates.csv and layouts.csv. We currently have 3 standard plate layouts: 12 × 40 spot matrix by 6 plates (up to 480 samples and controls), 18 × 40 spot matrix by 9 plates (up to 720 samples and controls), and 24 × 40 spot matrix by 12 plates (up to 960 samples and controls). The process could be further customized to include more layouts.
Arrayer software tool (Aushon BioSystems/Quanterix) to generate a GenePix Array List (GAL) file containing identifiers for each arrayed sample spot.

Depending on the overall number of experimental samples, 3 major array configurations can be generated: a 12 × 40 spot matrix by 6 plates (up to 480 samples and controls), an 18 × 40 spot matrix by 9 plates (up to 720 samples and controls), and a 24 × 40 spot matrix by 12 plates (up to 960 samples and controls). Each sample is spotted as technical triplicates, and up to 2880 spots can be printed on each slide (corresponding to a maximum of 960 experimental samples and controls) under the current layout. In addition, customized configurations for fewer or more samples are feasible with our arrayer because of the flexibility of the arrayer configuration. The first and last 10% of the rows in each configuration are allocated for control samples (total 76); therefore, the total number of experimental samples are 404, 644, and 884 for the 3 different layouts. The arrayer is designed to deposit fluid samples in microarray formats using a solid-pin printing technology, and the fluid droplet is carried on the tip of the pin. The droplet is transferred to the substrate (slide) when the pin tip gently touches the surface. Solid pins can handle a wide range of fluid viscosities that are suitable for protein lysates for RPPA. We print 350–400 replicate slides for each RPPA experiment. As a core service we collect protein analyte samples from multiple investigators to fill the majority of each slide, resulting in both an effective use of core resources and time and cost efficiency for core users. Arrayed slides are stored in a sealed plastic bag at room temperature to use within 30 d or frozen at −20°C for long-term storage and usage.

The entire printing process takes 4–5 d from thawing, denaturing protein samples, and transferring samples to 384-well plates (1 d) and then printing lysates on 350 replicate slides (3–4 d). During the printing, we remove selected slides to quality check spot printing under the microscope. Spots shifted away from the grid layout may indicate that adjustment to the robotic arrayer setup is required. Spots with abnormal morphology (such as a stripe across the spot area) may be due to dirty pins, requiring pin cleaning before continuing with the printing.

Up to 76 control samples are spotted on each slide, which serve to ensure quality of the RPPA assay and data reliability and to facilitate troubleshooting during various steps. The 4 corners of the arrayed slides are spotted with a mixture of purified mouse, rabbit, and goat IgG at 0.05 mg/ml (Fig. 5). The IgG mix serves as a positive control for secondary antibodies and the detection probe reagent as well as a guide for image gridding, positional location, and the perimeter of the entire array on the slide. Array matrix assignments are subjected to manual inspection to ensure accuracy. The control samples serve as a key troubleshooting tool to evaluate primary antibodies during the RPPA process. The IgG controls are not only used as a gridding control but also to troubleshoot problems with various reagents. For example, if no signal is detected for sample spots on an entire slide, but there are signals for the IgG controls, this would indicate a failure of the primary antibodies, whereas the other reagents are working properly.

### Step 4: antibody and total protein labeling of RPPA slides

Immunolabeling is performed with an Autolink 48 slide autostainer (Agilent Technologies, Santa Clara, CA, USA). The autostainer has a capacity for incubating a total of 30 slides; for each batch of 30 slides, we typically incubate with 29 different primary antibodies, and the 1 remaining slide is incubated without primary antibody as a negative control. The VECTASTAIN Elite ABC-HRP Kit (PK-6100; Vector Laboratories, Burlingame, CA, USA) and Tyramide Signal Amplification plus biotin kit (NEL749B001KT; Perkin-Elmer, Waltham, MA, USA) are used as part of the catalyzed signal amplification system, and fluorescent IRDye 680 Streptavidin (926-68079; Li-Cor Biosciences, Lincoln, NE, USA) is used as the probe to detect antibody binding (Fig. 5A). Each slide is pretreated with Re-Blot Plus Strong Antibody Stripping Solution (2504; MilliporeSigma) and I-Block Protein-Based Blocking Reagent (T2015; Thermo Fisher Scientific) and then incubated with a single primary antibody at room temperature for 30 min followed by a goat anti-rabbit or mouse IgG secondary antibody (NC9256157 or NC9372061; Vector Laboratories) (Fig. 5B). A predetermined optimal concentration is used for each specific primary antibody (at Step 1: antibody validation, in Fig. 1). Negative control slides are incubated with antibody diluent instead of primary antibody (Fig. 5C). Total protein for each spot is assessed by staining 1 in every 20 replicate slides with SYPRO Ruby protein fluorescence dye (S11791; Molecular Probes, Eugene, OR, USA) (Fig. 5D). SYPRO Ruby is a highly sensitive fluorescence-based dye for detecting proteins on nitrocellulose membranes. The sensitivity of this dye (0.25–1 ng protein/mm²) provides >3 orders of magnitude over a linear quantitation range. Slides are immersed in a SYPRO Ruby Protein Fixative Solution [10% v/v methanol (A412-500) and 7% v/v acetic acid (A38S-500); Thermo Fisher Scientific] at room temperature for 15 min, washed in deionized water, and immersed in SYPRO Ruby Blot stain for 15 min at room temperature in the dark. Slides are washed in deionized water twice and dried before scanning. Total protein staining is carried out on the bench in a microscope slide staining dish set (50-287-04; Thermo Fisher Scientific). The time required for total protein staining (SYPRO Ruby) is 1 d. The total time required for immunolabeling and scanning of replicate slides for 240
antibodies is 11–13 d. Immunolabeling with the autostainer for each batch of 30 slides requires 1 full day with the scanning of each batch performed overnight (Step 5: slide scanning and image analysis below). Common reagents for immunolabeling are prepared and stored in advance (the previous week), and antibodies and immunolabeling reagents are added the day of the assay run.

**Step 5: slide scanning and image analysis**

Antibody-labeled and total protein slides are imaged with a GenePixAL4400 fluorescence scanner (GENEPIX 4400; Molecular Devices). The signal of each spot is obtained from the fluorescence intensity after subtraction of the local slide background signal. For antibody labeling and corresponding negative control slides, the scanner is set at an excitation wavelength of 635 nm, 100% laser power, and emission filter 655–695 nm coupled with a neutral density filter to block light transmittance by 90% and a single PMT setting of 380 (Fig. 5). PMT settings are optimized periodically to ensure a dynamic range of detection for both low- and high-expression proteins. This scanning process usually generates ~1000 images from about 240 antibody slides plus associated negative control and total protein slides. Image analysis and processing are performed concurrently with antibody labeling of slides. Because of the ability to overlap these steps, by the time the antibody labeling is finalized, only 3–5 additional days are required to complete the image analysis.

Images from each PMT are analyzed to generate data results files [GenePix Results (GPR) files] for subsequent normalization, data processing, and statistical analysis. The third component of the RPPA Setup Tool, the ReNAME utility, uses as input the TIF image files generated for all the scanning intensities and the antibody staining list that identifies the specific antibody used for each slide (Fig. 6). Each image file is then renamed using file system safe names with the name of the antibody used to stain each slide. The RPPA ImGrid Tool, a Matlab utility with a user-friendly graphical interface, was developed to facilitate and accelerate image analysis (Fig. 7). This tool uses MATLAB...
9.8.0.1323502 (R2020a) and standard image-processing libraries. ImGrid facilitates matching of scanned and renamed images with GAL file grids, enabling a rapid visual evaluation of gridding accuracy and image quality, and serves as a troubleshooting tool for image analysis. By visual inspection, any images that were not gridded accurately are reanalyzed after adjusting the gridding coordination with the image analysis software GenePix Pro 7.2 (Fig. 7). The RPPA ImGrid Tool can operate on multiple images in bulk, wherein each image is examined carefully for background evenness across the slide, spot informality, areas of nonspecific signals, and defective fluorescent labeling. If a defect occurs with an individual spot only, it is flagged. If the image quality is widely affected over large regions or the entire slide, we determine if this is an antibody or slide failure. For a specific antibody or slide that fails quality inspection, extra replicate slides from the same RPPA run are incubated with the primary antibody of interest again (with adjustment for optimization if necessary) and processed to completion. If the repeated slide still fails, results with that antibody are removed before data reporting, and the specific antibody is re-evaluated for future RPPA runs. In a typical run of ~240 antibodies, the first-round failure rate ranges between 2 and 10 primary antibody-labeled slides, and repeat antibody labeling of replicate slides corrects the problem in the majority of cases. To further facilitate and expedite visual inspection, we developed an ancillary feature: the Spot Annotation Tool. This uses the Matlab image-processing standard library to print sample labels onto the technical triplicate sample spots on the array image (Fig. 8). The tool helps identify and track samples by their identifier names and by the name of the researchers who submitted the samples. The Spot Annotation Tool is run on selected individual image files that were flagged after the visual inspection of the RPPA ImGrid results to identify and characterize problems or misaligned spots.

**Step 6: data normalization and QC**

Data normalization is carried out within the samples of each individual user project, employing 2 computational steps implemented in the Python normalization and processing tool (Fig. 9A). First, total protein slides stained with SYPRO Ruby are used to normalize the levels of antibody labeling. To reduce technical variability, every 20th slide is stained for total protein (slide 10, 30, 50, etc.) and is used to normalize the 8–9 slides stained with specific antibodies before and after (i.e., antibody slides 2–19 are normalized by total protein on slide 10; antibody slides 20–39 are normalized by total protein on slide 30). The corresponding total protein slide for each antibody slide is encoded into an input file used by the Python data normalization tool. For every spot on the slide, image analysis determines an intensity value for...
antibody labeling at a specific PMT setting as well as an intensity value for the corresponding negative control at the same PMT setting and that of total protein by SYPRO Ruby from the corresponding normalization slide. Secondly, the raw signal intensity (SI) for each antibody is determined by subtracting the background from the antibody spot signal during the image analysis process (automatically generated and stored in the GPR file). A sample group is assigned to each RPPA core user, corresponding to their experimental project. During data normalization, the negative control SI is subtracted from the raw antibody SI for each spot and then normalized to the total protein SI within the user-defined sample group.

Specifically, for each spot, the normalized antibody SI can be expressed using the following formula:

\[ N = \frac{A - C}{T} \times M \]

wherein N is the normalized antibody SI, A is the raw antibody SI, C is the negative control SI, T is the SI of total protein, and M is the median SI of total protein from spots within the same user’s sample group. If the raw antibody SI is lower than or equal to the negative control SI, the normalized SI is set to 1; if the raw antibody SI, the negative control SI, or the total protein SI has a flag indicating the SI is problematic, the normalized antibody SI is set to not applicable (NA).
Extensive QC information is collected and evaluated for each antibody at the multiple PMT settings with the main goal of identifying and flagging potential bad antibodies. We generate QC scores to assess antibody quality by assessing the variability across the technical triplicates and by evaluating antibody SI on serial dilutions of a control cell mixture (Ctrl Cell Mix) (Fig. 9B). The coefficient of variation (CV) score (CV score) is determined for each antibody across all the technical triplicates measured. The CV score is defined as the percentage of samples on a slide (including controls and user samples) with a CV $<20\%$. To maintain flexibility, the 20% threshold for CV can be

Data normalization, QC, and report generation. A Python suite of command-line tools designed for bioinformatics/biostatistics experts is used. A) Normalization strategy. Protein raw SI for each sample and antibody is computed by subtraction of negative control and scaling to total protein. B) A quality score (QC score) is computed for each antibody based on its CV across technical replicates (CV score) and the PCC between the antibody SI and the dilution index over the Ctrl Cell Mix (Cellmix score). C) The optimal scanning intensity is determined for each antibody and slide to avoid saturation or low SI. Examples of saturated SI spots (green box, white signal) and low SI spots (blue box, dim regions) are provided. D) Reports for individual RPPA users and sample groups are generated from the master RPPA report. E) Statistical analysis can be performed across user-defined experimental groups.
changed to different values in the data QC utility. The Ctrl
Cell Mix score (Cellmixscore) is assessed using the antibody SI
over the normalized fluorescent intensity from Ctrl Cell
Mix. The Ctrl Cell Mix comprises equal amounts of lysates
from 4 sources: breast cancer cell lines MDA-MB-415 and
T47D, pervanadate-treated HeLa cells, and calyculin
A–treated Jurkat cells (see results for details). The Ctrl Cell
Mix is diluted in 8 serial dilutions and spotted in 2 different
locations on each slide. For each dilution of the Ctrl Cell
Mix, we compute the Pearson correlation coefficient (PCC)
and the corresponding P value between the antibody
intensity and the dilution factor. If the Pearson correlation
is significant and positive for either of the 2 cell mix groups
of spots, then the Cellmixscore is defined as the maximum
significant PCC across the cell mixtures normalized to 100;
otherwise, it is defined as 0. The final quality score QCscore is
defined as follows: QCscore = 0.1*CVscore + 0.9*Cellmixscore.
For antibodies that passed our previous visual inspection, we
use the QCscore for further stringent quality assessment. This
entire process takes a total of 4–5 d to complete.

Step 7: data reporting
After data normalization and QC, individual user data
reports are generated (Fig. 9D). Separate reports are
generated for each investigator’s set of experimental samples
based on sample group normalization, and a single PMT
setting is selected to produce an optimal data set across all the
antibodies (Fig. 9C). This step ensures that for each antibody
in a given sample group the signals of the total and
the phosphorylated proteins are within the linear range.
Because each experimental lysate is printed using technical
triplicates, the final report also provides technical triplicate
data for each sample on every antibody. A QC report
encompassing all antibodies is also provided so quality
assessment can be made for each antibody by the user (Fig.
9D). The final data report consists of a Microsoft Excel file
containing 4 or 7 tabs separated by the organism of the
experimental samples: 4 worksheets for human samples and
7 worksheets for mouse samples. The user reports for human
samples include: 1) data-processing methods and table of
contents for report, 2) normalized SI data for nonmouse antibodies,
3) QC data for all antibodies, and 4) raw SI data for all antibodies.
Our current antibody inventory contains 15 mouse monoclonal antibodies for which there are no high-
quality specific rabbit counterparts. When mouse tissues are
profiled, samples with high mouse IgG content can lead to
high negative control SI for the 15 mouse antibodies due to
cross-reactivity. Generating distinct data sheets for mouse
tissue samples stained with the mouse antibodies allows a
focused evaluation and an informed decision on whether to
select the mouse antibodies for further analysis. The user
reports for mouse samples include: 1) data-processing
methods and table of contents for report, 2) normalized SI
data for nonmouse antibodies, 3) QC data for nonmouse antibodies, 4) raw SI data for nonmouse antibodies, 5)
normalized SI data for the mouse antibodies, 6) QC data for
the mouse antibodies, and 7) raw SI data for the mouse antibodies.

The QC report is generated across all antibodies in
addition to the antibody-specific QCscore and its compo-
nents, the CVscore and Cellmixscore (Fig. 9B). The QC
report contains the range of each antibody captured by
descriptive statistics across the entire slide for all the
samples including slide median, slide mean, and slide
maximum (because each slide is dedicated to a single
antibody). The QC report further contains the antibody
SIs across several diverse groups of control samples that
are profiled for each RPPA experiment. The control sample groups include:
- BCC mix: lysate mixture of 39 breast cancer cell lines.
- NCI mix: lysate mixture of 60 NCI cell lines.
- Ctrl Cell Mix: lysate mixture of breast cancer cell lines
  MDA-MB-415 and T47D, pervanadate-treated HeLa cells,
  and calyculin A–treated Jurkat cells.
- HP: pervanadate-treated HeLa lysate.
- JC: calyculin A–treated Jurkat lysate.
- Mouse tissue mix: an equal mix of lysates from multiple
  mouse tissues (heart, brain, lung, intestine, gastrocne-
  mius, kidney, mammary glands, ovary, uterus, quadri-
  ceps, brown adipose tissue) with a final concentration of
  0.5 mg/ml in RPPA printing buffer. Mouse liver and
  spleen were excluded because of high background signal
  caused by tissue-specific IgG content. For experimental
  mouse liver samples, we were able to generate reliable data
  for investigators, even though the background was higher
  than other samples. We have not performed projects on
  spleen tissues.
- Buffer control: RPPA sample buffer, no protein
  lysate.

Antibody levels and variability are captured with the
following 3 metrics defined across each complete printed
slide (i.e., not limited to samples from 1 individual user):
The aggregated CV across the technical replicates.
The percentage of samples with signal >200 (an empirically
determined threshold).
The percentage of samples with SI >200 and CV <20%.
The minimum SI level and the maximum CV allowed
can be adjusted to ensure reporting and analysis flexibility.
The time to generate individual user reports along with QC
metrics takes ~2–3 d.

Step 8: data analysis
Based on the QC report, antibodies with a mean SI across
user samples <200 are filtered out. Antibodies with median
technical triplicate CV >20% within the user’s samples might also be excluded from further analysis. Statistical analysis is performed using a Python Statistical Analysis Tool to identify differentially expressed proteins between experimental groups (Fig. 9E). Principal component analysis (PCA) plots are generated for all samples and annotated with their experimental group and across all measured proteins. ANOVA is used for multiple-group analysis, whereas a t test is carried out for 2-group comparison. A heat map of all differentially expressed proteins is generated for visualization. Further pathway analysis may be carried out for functional interpretation of results.15, 19–21

RESULTS AND DISCUSSION

Antibody inventory and controls

We have built an inventory of ~240 validated antibodies for the RPPA assay as described in Step 1 and have created a spreadsheet that includes vendors and related information of antibodies (Supplemental Table S1). An updated antibody list will be maintained online at https://www.bcm.edu/academic-centers/dan-l-duncan-comprehensive-cancer-center/research/cancer-shared-resources/reverse-phase-protein-array/antibodies-list. The antibodies recognize regulatory proteins of major signaling pathways and cellular processes, with ~33% of the antibodies to specific phosphorylation sites used as an indication of functional activity. The Molecular Signatures Database collection of hallmark gene sets summarizes and represents specific well-defined biologic states or processes. Fig. 10 shows the genes (and by association, their encoded proteins) in each of the 50 hallmark pathways represented by our antibody inventory. Overall, the RPPA target protein panel includes cell cycle and cell proliferation, angiogenesis, growth factor receptor, apoptosis, cytokines, Signal Transducer and Activator of Transcription (STAT) signaling, stem cells, DNA damage response, epithelial-mesenchymal transition, stress responses, autophagy, mechanistic Target of Rapamycin (mTOR) signaling, oncogenes and tumor suppressors, inflammatory signaling pathways, transcription factors, chromatin modifier proteins, and metabolic enzymes. Out of the 50 hallmark pathways, 46 pathways are represented with antibodies on the RPPA (109 antibodies against total protein and 63 phospho-antibodies recognizing 45 distinct proteins). Note that not all our validated antibodies recognize proteins (genes) in the hallmark pathway collection.

Extensive control samples for the primary antibodies used for RPPA are spotted on all replicate slides; that includes lysates from known positive and negative cell lines and tissues, dilution curves of positive controls to determine whether antibody signals are in a dynamic range, and calibrators for phospho-specific antibodies. To identify positive controls for each antibody in our inventory, we screened >100 cell lines individually (treated or untreated with different stimuli), including lysates from the NCI-60 cancer cell lines22 and NCI-ATCC 43 breast cancer cell line (www.atcc.org). From these results, we selected a mix of 4 cell lines (T47D, MDA-MB-415, pervanadate-treated HeLa, and calyculin A–treated Jurkat cells) that represents lysates (Ctrl Cell Mix) expressing ~90% of target proteins of interest and corresponding phosphorylated forms of the
proteins. Ctrl Cell Mix is printed as a series of eight 2-fold serial dilutions on different locations of the slide for quality assessment and to generate a curve to validate the dynamic range of each antibody. Examples of serial dilutions of the Ctrl Cell Mix with 4 different antibodies (ERK1/2, Notch1, phospho–Janus kinase 1, and histone deacetylase 1) from our inventory are shown in Fig. 11A. The SIs for each antibody were dose responsive with an optimal value at the 0.25–0.5 mg/ml protein concentrations. Thus, 0.5 mg/ml was chosen as the single dilution for most experimental protein lysates (Fig. 11A). Other selected controls are spotted on every slide to ensure confidence and reliability of results, including calibrators for phosphorylated proteins to validate responsiveness of phospho-proteins to various signal transduction activators.23 Calibrator lysates are spotted from HeLa cells treated with the phospho-tyrosine phosphatase inhibitor pervanadate and from Jurkat cells treated with the serine/threonine phosphatase inhibitor calyculin A (Fig. 11B). These inhibitors promote a general increase in protein tyrosine phosphorylation and serine/threonine phosphorylation, respectively. Treated lysates are mixed at different percentages with untreated cell lysates ranging from 0 to 100% to generate a standard reference curve for activated proteins in different pathways. As shown in Fig. 11B, the intensity signals generated by the dual serine (S202)/tyrosine (Y204) phospho-ERK antibody varied with the percentage of the lysate that originated from stimulated samples. QC data are generated for each antibody used. As an example, the intensities for the total ERK1/2 antibody are shown for each positive control spotted on the slide (Fig. 11C). ERK1/2 shows a strong median and mean SI across all spotted lysate samples on the slide, the mix of 39 breast cancer cell lines (BCC mix) and the NCI-60 cancer cell lines (NCI mix), the 4-cell-line Ctrl Cell Mix, and calibrator-treated HeLa (HP) and Jurkat (JC) cells. Only very low SI was detected for ERK1/2 in the buffer control.
Effects of steroid hormones on protein-signaling pathways in breast cancer cells

As an example of the workflow process applied to an experimental system, we used RPPA to analyze the effects of steroid hormones in an estrogen receptor (ER) and progesterone receptor (PR) positive human ductal carcinoma in situ (DCIS) cell line. Human DCIS.com cells were engineered to express both ER and PR, and the cell line has been previously characterized as an in vitro model of luminal DCIS that is highly responsive to both estrogen and progesterone. ER and PR are ligand-dependent steroid receptor nuclear transcription factors that act primarily to regulate expression of gene networks. They also mediate rapid effects of steroid hormones on extranuclear protein-signaling pathways through alteration of protein kinase activity and phosphorylation of protein substrates and downstream events.

DCIS cells were treated with vehicle (ethanol) or E2 (1 nM E2) together with a progestin agonist (10 nM R5020 (17 alpha, 21-dimethyl-19-nor-pregna-4, 9-diene-3,20-dione)) for either 6 or 24 h in triplicate experimental cultures. Cell lysates were prepared and spotted on slides and processed for labeling with our validated antibody inventory using standard protocols and workflow processes.

PCA (principal component analysis) for all antibody protein targets showed that the samples grouped together based on the different treatments (Fig. 12A). After filtering out low- or no-expression proteins across all the samples (SI < 200) and high-CV samples (CV > 20%), all remaining antibody SI values were subjected to an ANOVA, with significance achieved at false discovery rate–adjusted P value <0.05. Using a fold change cutoff of >1.25× or <0.8×, we identified proteins differentially expressed between the 6-h hormone treatment vs. vehicle (no hormone) and the 24-h treatment vs. vehicle. A list of fold changes in all proteins at both times of hormone treatment is shown in Supplemental Table S3. A Venn diagram analysis accounting for protein change direction is presented in Fig. 12B, showing

![FIGURE 12](image_url)

RPPA analysis of DCIS.com ER+PR-B cells treated with E2 and R5020. DCIS.com cells engineered to express both ER and PR-B were treated with ethanol or with E2 and PR agonist R5020 for 6 or 24 h. (A) PCA shows a robust separation of the experimental groups. (B) We identified proteins commonly changed at both 6 and 24 h, as well as specific to each time point, both up and down. ANOVA was used to determine significance, achieved at false discovery rate–adjusted P value <0.05. (C) Heatmap representation of protein changes after E2 and R5020 exposure of 6 or 24 h, as compared with ethanol. Expression values were scaled by converting intensities to z-score values. (D) Antibody intensities are graphed showing increased KLF4 or c-Jun expression and decreased PR expression in response to E2+R5020. (E) An example of a hormone induced phospho-protein, p38 Thr180/Tyr182, showing that total p38 protein does not significantly change with hormone. The ratio of phospho to total protein shows a net gain of p38 Thr180/Tyr182 reactivity. (F) An example of a hormone repressed phospho-protein, AKT Ser473, demonstrating that total AKT protein does not significantly change with hormone. The ratio of phospho to total protein shows a net loss of AKT Ser473 reactivity. EtOH, ethanol; KLF4, Kruppel-like factor 4. PC, principal component.
a robust overlap between the 6-h and the 24-h protein changes. At 6 h of hormone treatment, a total of 52 protein changes were detected, with 34 exhibiting an increase and 18 exhibiting a decrease. At 24 h of treatment, 54 protein changes were detected, with 27 exhibiting an increase and 27 exhibiting a decrease. A group of 24 up-regulated proteins were common at both time points, whereas 18 down-regulated proteins were common.

A heat map of z-score normalized intensities at 6- and 24-h of hormone treatment vs. vehicle control (ethanol) grouped by unique and common up- and down-regulated proteins is shown in Fig. 12C. The presence of distinct sets of protein changes at early and later times of hormone treatment is consistent with known actions of steroid hormones. For example, the levels of the transcription factor Kruppel-like factor 4 protein increase at 6 h and further increase after 24 h of hormone treatment (Fig. 12D). Early response genes, such as c-Jun, are induced acutely within a few hours of hormone treatment and return to baseline levels by 24 h, as observed by RPPA (Fig. 12D). Additionally, many of the changes in SI are with phospho-specific antibodies, and about half of these are increased only at the early 6-h time point and not at 24 h. This is also consistent with the rapid nongenomic effects of ER and PR on phosphorylation of protein-signaling pathways that occurs within a few minutes of hormone stimulation and can be transient. ER and PR themselves are down-regulated with prolonged treatment with their respective ligands as a mechanism to desensitize tissue to continuous exposure to hormone. Significant reductions in ER and PR protein were detected at 6 h, whereas both proteins were more substantially decreased by 24 h (Fig. 12D for PR).

An important criterion for phospho-specific antibodies is to confirm that a change in SI is independent of a change in total protein and represents a true increase or decrease in the phosphorylation state of the protein of interest. As examples, to demonstrate true differences in phosphorylation, increased phosphorylation of p38 and decreased phosphorylation of AKT in response to hormone are shown in Fig. 12E, F. No significant change in the levels of the corresponding total proteins was observed. The ratio of SI for phospho-specific antibodies to antibodies against total protein reflects a true change in phosphorylation status.

RPPA, as deployed in our core, has proven to be a valuable discovery tool in cancer cell lines and animal models, particularly to identify changes in phosphorylation of key regulatory proteins as markers of alterations in the activity-signaling pathways that contribute to cancer cell biology. As with all omics platforms, RPPA in this context is a discovery and hypothesis-generating tool requiring further confirmation and guiding the prioritization of in-depth analysis of specific protein pathways. 7–9, 17, 19

In summary, RPPA is a robust targeted proteomic platform that is complementary to mass-spectrometry profiling. In this report, we describe a stepwise workflow process for RPPA in a core facility. The workflow ensures proper handling of the experimental protein samples and data files through arrays to plate script layouts and renaming scripts, quality of image analysis, QC, accuracy and flexibility of data normalization, and data processing.

ACKNOWLEDGMENTS

This work was supported in part by a Cancer Prevention and Research Institute of Texas (CPRIT) Proteomics and Metabolomics Core Facility Award RP170005 (D.P.E.), National Institutes of Health National Cancer Institute Cancer Center Support Grant P30-CA125123 (Kendall Osborne PI, D.P.E. Shared Resource Leader), and institutional support from Baylor College of Medicine Advanced Technology Cores (ATC) for the Antibody-Based Proteomic Core (S.H., Director). C.C. and S.L.G. were partially supported by National Institute of Environmental Health Sciences Grants P30-ES030285 and P42-ES027725. The authors declare no conflicts of interest.

REFERENCES

1. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–232.
2. Creighton CJ, Huang S. Reverse phase protein arrays in signaling pathways: a data integration perspective. Drug Des Devel Ther. 2015;9:3519–3527.
3. Grubb RL, Deng J, Pinto PA, et al. Pathway biomarker profiling of localized and metastatic human prostate cancer reveal metastatic and prognostic signatures. J Proteome Res. 2009;8:3044–3054.
4. Masuda M, Yamada T. Utility of reverse-phase protein array for refining precision oncology. Adv Exp Med Biol. 2019;1188: 239–249.
5. PetricoinE, Wulffkuhle J, Howard M, et al. RPPA: origins, transition to a validated clinical research tool, and next generations of the technology. Adv Exp Med Biol. 2019;1188: 1–19.
6. Pierobon M, Wulffkuhle J, Liotta LA. Petricoin lii EF. Utilization of proteomic technologies for precision oncology applications. Cancer Treat Res. 2019;178:171–187.
7. Acharya KD, Nettles SA, Sellers KJ, et al. The progesterin receptor interactome in the female mouse hypothalamus: interactions with synaptic proteins are isoform specific and ligand dependent. eNeuro. 2017;4:ENEURO.0272-17.2017.
8. Bu W, Liu Z, Jiang W, et al. Mammary precancerous stem and non-stem cells evolve into cancers of distinct subtypes. Cancer Res. 2019;79:61–71.
9. Chang CH, Zhang M, Rajapakshe K, et al. Mammary stem cells and tumor-initiating cells are more resistant to apoptosis and exhibit increased DNA repair activity in response to DNA damage. Stem Cell Rep. 2015;5:378–391.
10. Dougherty MI, Lehman CE, Spencer A, et al. PRAS40 phosphorylation correlates with insulin-like growth factor-1 receptor-induced resistance to epidermal growth factor receptor inhibition in head and neck cancer cells. Mol Cancer Res. 2020; 18:1392–1401.
11. Finnerty BM, Moore MD, Verma A, et al. UCHL1 loss alters the cell-cycle in metastatic pancreatic neuroendocrine tumors. Endocr Relat Cancer. 2019;26:411–423.
12. Fu X, Jelskohn R, Pereira R, et al. FOXA1 overexpression mediates endocrine resistance by altering the ER transcriptome.
and IL-8 expression in ER-positive breast cancer. Proc Natl Acad Sci USA. 2016;113:E6600–E6609.

13. Haricharan S, Punturi N, Singh P, et al. Loss of MutL disrupts CHK2-dependent cell-cycle control through CDK4/6 to promote intrinsic endocrine therapy resistance in primary breast cancer. Cancer Discov. 2017;7:1168–1183.

14. Holdman XB, Welte T, Rajapakse K, et al. Upregulation of EGFR signaling is correlated with tumor stroma remodeling and tumor recurrence in FGFR1-driven breast cancer. Breast Cancer Res. 2015;17:141.

15. Jayaraman P, Parikh F, Newton JM, et al. TGF-β1 programmed myeloid-derived suppressor cells (MDSC) acquire immune-stimulating and tumor killing activity capable of rejecting established tumors in combination with radiotherapy. OncolImmunology. 2018;7:e1490853.

16. Sun C, Yin J, Fang Y, et al. BRD4 inhibition is synthetic lethal with PARP inhibitors through the induction of homologous recombination deficiency. Cancer Cell. 2018;33:401–416.e8.

17. Welte T, Kim IS, Tian L, et al. Oncogenic mTOR signalling recruits myeloid-derived suppressor cells to promote tumour initiation. Nat Cell Biol. 2016;18:632–644.

18. Berggren K, Steinberg TH, Lauber WM, et al. A luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on membrane supports. Anal Biochem. 1999;276:129–143.

19. Morriss GR, Rajapakse K, Huang S, Coarfa C, Cooper TA. Mechanisms of skeletal muscle wasting in a mouse model for myotonic dystrophy type 1. Hum Mol Genet. 2018;27:2789–2804.

20. Ware MJ, Keshishian V, Law JJ, et al. Generation of an in vitro 3D PDAC stroma rich spheroid model. Biomaterials. 2016;108:129–142.

21. Ware MJ, Nguyen LP, Law JJ, et al. A new mild hyperthermia device to treat vascular involvement in cancer surgery. Sci Rep. 2017;7:11299.

22. Federici G, Gao X, Slawek J, et al. Systems analysis of the NCI-60 cancer cell lines by alignment of protein pathway activation modules with “-OMIC” data fields and therapeutic response signatures. Mol Cancer Res. 2013;11:676–685.

23. Paweletz CP, Charboneau L, Bichsel VE, et al. Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene. 2003;20:1981–1989.

24. Miller FR, Santner SJ, Tait L, Dawson PJ. MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ. J Natl Cancer Inst. 2000;92:1185–1186.

25. Villanueva H, Grimm S, Dhamne S, et al. The emerging roles of steroid hormone receptors in ductal carcinoma in situ (DCIS) of the breast. J Mammary Gland Biol Neoplasia. 2018;23:237–248.

26. Grimm SL, Hartig SM, Edwards DP. Progesterone receptor signaling mechanisms. J Mol Biol. 2016;428:3831–3849.