REVIEW

Reverse phase protein arrays in signaling pathways: a data integration perspective

Chad J Creighton[1,3]
Shixia Huang[2,3]

[1]Department of Medicine,
[2]Department of Molecular and
Cellular Biology, [3]Dan L Duncan
Cancer Center, Baylor College
of Medicine, Houston, TX, USA

Abstract: The reverse phase protein array (RPPA) data platform provides expression data for a prespecified set of proteins, across a set of tissue or cell line samples. Being able to measure either total proteins or posttranslationally modified proteins, even ones present at lower abundances, RPPA represents an excellent way to capture the state of key signaling transduction pathways in normal or diseased cells. RPPA data can be combined with those of other molecular profiling platforms, in order to obtain a more complete molecular picture of the cell. This review offers perspective on the use of RPPA as a component of integrative molecular analysis, using recent case examples from The Cancer Genome Altas consortium, showing how RPPA may provide additional insight into cancer besides what other data platforms may provide. There also exists a clear need for effective visualization approaches to RPPA-based proteomic results; this was highlighted by the recent challenge, put forth by the HPN-DREAM consortium, to develop visualization methods for a highly complex RPPA dataset involving many cancer cell lines, stimuli, and inhibitors applied over time course. In this review, we put forth a number of general guidelines for effective visualization of complex molecular datasets, namely, showing the data, ordering data elements deliberately, enabling generalization, focusing on relevant specifics, and putting things into context. We give examples of how these principles can be utilized in visualizing the intrinsic subtypes of breast cancer and in meaningfully displaying the entire HPN-DREAM RPPA dataset within a single page.

Keywords: RPPA, proteomics, molecular profiling, integrative analysis, breast cancer, TCGA

Introduction

Human diseases such as cancer can be incredibly complex at the molecular level, where a good understanding is needed for the signaling pathways involved. Cancer itself may initiate from DNA damage or aberrant DNA methylation affecting a key gene or set of genes, but the end result is cells showing widespread deregulation of signaling pathways and gene transcription. By incorporating multiple levels of molecular data on the diseased state of the cell, a more complete picture may emerge. With the advent of DNA microarray technologies,[1] it became possible for us to profile the mRNA expression of thousands of genes in a single experiment.[2] However, it quickly became apparent that gene transcriptional changes would represent just one level of the overall picture, as these are one step removed from signal transduction pathways.[3] Proteomic profiling would therefore provide another important level. In particular, the reverse phase protein array (RPPA) data platform provides relative abundances for a set of key proteins (either total proteins or posttranslationally modified proteins),[4] and this platform is establishing itself as a valuable research tool in human diseases.

The RPPA technology is a type of protein microarray which is the derivative of two technological advances: gene expression microarrays,[1] which print DNA molecules on a

Correspondence: Chad J Creighton;
Shixia Huang
Dan L Duncan Cancer Center, Baylor
College of Medicine, One Baylor Plaza
MS305 Houston, TX 77030, USA
Email creighto@bcm.edu;
shixiah@bcm.edu

glass slide, and immunoassays,[5,6] which enable the detection of protein expression through antibody and antigen interaction. MacBeath and Schireiber[7] were the first to develop protein microarray. They used a high-precision robot to print recombinant proteins on to glass slides, and used them for high-throughput detection of protein–protein interaction. In 2001, Brown and colleagues who invented the gene expression microarray,[1] reported another protein array, an antibody microarray[8] that contained hundreds of antibodies printed onto glass slides for measuring the abundance of many specific proteins in complex biological samples.

In 2001, Paweletz et al[9] reported a new variation of protein microarray, in which tissue lysates, rather than recombinant proteins or antibodies, were spotted onto slides. They named this array "reverse phase protein microarray" in contrast to the "forward phase" antibody arrays which spot antibodies onto a slide. Other names in the literature include "lysate microarray",[10] "reverse-phase lysate microarray",[11] "reverse phase protein lysate microarrays",[4] and "reverse phase protein array".[12] Since 2011, annual RPPA workshops/conferences have been held to provide a platform for scientific communications and exchanging ideas about technical developments.[13] In its annual meeting in Paris, France, in October 2014, the standardized nomenclature "RPPA" was recognized. Therefore, in this review, we will use the current term "RPPA".

The purpose of this review is to shed light on the important role that RPPA data may play in integrative molecular analyses. Here, integrative analysis may involve effective combination of results from multiple data platforms including RPPA, as well as incorporating our prior knowledge of biological systems into the interpretation of molecular-based results. For examples, we will focus here on recent cancer-related studies and datasets, in particular those initiated by The Cancer Genome Altas (TCGA).[14–21] In addition, overall approaches for more effectively visualizing RPPA results will be discussed.

Profiling of signaling pathway using the RPPA platform

The RPPA platform involves micro-blots of protein lysates from multiple samples of tissues, cell lines, or bodily fluids (such as serum, cerebrospinal fluid, urine, saliva) on a single array, with each sample represented by at least one spot.[4,22] Each array is then incubated with one specific antibody, in order to detect the relative expression of the corresponding protein across many samples simultaneously. Multiplexing using different antibodies on multiple arrays of the same set of lysates can be carried out to measure many proteins in

a high-throughput manner. This platform has the capacity to allow hundreds and thousands of samples to be assayed. However, it requires highly specific antibodies (ie, a single specific band or predominant band to be observed by Western blot analysis) and the corresponding RPPA validation (ie, RPPA expression level correlating with Western blot).[12] Due to the stringent validation process, currently, most RPPA datasets often include on the order of 150–300 antibodies for total or various modified proteins. For a specific study, the set of proteins profiled can be tailored to focus on particular pathways (eg, the PI3K pathway involving multiple signaling components) or cell functions (eg, apoptosis or invasion), but RPPA also allows for more exploratory analysis leading to unexpected pathway associations.

Developing antibodies and validating them for use with RPPA are laborious processes and are currently the bottleneck of this technology. Usually, the antibody validation workflow starts with obtaining a commercially available antibody that shows a specific band or bands, followed by Western blot performed in RPPA laboratories to confirm its specificity using proper positive and negative control lysates. Antibodies with multiple nonspecific bands are excluded from further testing, while antibodies showing specificity by Western blot will then be tested further by RPPA profiling of cell lines and other types of samples. Antibodies against phosphorylated proteins need to show specificity against samples stimulated (eg, using growth factors) or inhibited (eg, using targeted inhibition) to generate phosphorylated or non-phosphorylated forms of a protein.[12] Antibodies for both Western blot and RPPA usually need titration for optimal results; one may typically start with the company's recommended dilution for Western blot and adjust for better results if necessary. For RPPA, an antibody concentration four times higher than what would be required for Western blot is often used. For antibodies that require longer incubation or exposure times for Western blot, even higher concentrations would be needed for RPPA. We and others have found that ~50% of the commercially available antibodies (those showing a single band by Western blot according to the company's data sheets) can be validated to generate reliable RPPA data.[4]

Through DNA microarrays and (more recently) RNA sequencing, it has long been possible to comprehensively profile all mRNAs in a single experiment, where the number of data points in an mRNA expression profile (~20 K genes) would be far greater than that for an RPPA proteomic profile (usually fewer than 300 proteins). However, in terms of actual information, RPPA data and mRNA data would be highly complementary to each other. Multiple studies have

shown that mRNA levels do not necessarily correlate with their corresponding protein levels.[23,24] In principle, pathways work by protein signaling transduction events that eventually lead to changes in gene transcription.[3] Posttranslational modifications may not be reflected at the gene transcription level. Gene transcription data may therefore inform on the downstream effects of deregulated signaling, but the pathways occurring upstream of an observed transcriptional pattern may best be discerned using proteomics. There are analytical techniques for defining gene transcriptional signatures of deregulated pathways, but the caveat with these analyses is that they can represent indirect targets of aberrant pathway signaling, where different pathways may converge on a similar set of transcriptional targets.[25]

The RPPA platform offers a number of unique benefits compared to other proteomics approaches.[22] The use of a highly specific antibody in its optimal reaction condition ensures high specificity and sensitivity. Sample handling and preparations are straightforward and simple. Rapid protein extraction and denaturation prevent degradation and preserve proteins and phosphorylated proteins, which are often labile. Large numbers of samples spotted on the same slides allow for easier and more reliable normalization, comparison, and data analyses. In addition, RPPA allows for the identification and profiling of target proteins and signaling pathways in small amounts of samples, such as biopsy specimens, tissues from laser capture microdissection (LCM), and fluorescence-activated cell sorting (FACS)-sorted minor cell populations, such as stem cells or cancer stem cells. For assaying 200–300 proteins with different abundance levels, 4–35 µg total protein would be standard when using a volume range from 20–35 µL at a concentration of 0.2–1.0 µg/µL. (The required volume needed is mostly due to evaporation rather than protein deposition on to slides, though even lower amounts of starting material may be accommodated in some cases.) The amount of protein needed per spot is very small, at the level of nanograms, where for specific proteins, the detection is at picograms to femtograms levels per spot.[26]

In contrast to RPPA (which profiles a smaller, predefined set of proteins), there are other proteomic technologies that are more global in nature, seeking to profile as many proteins as possible.[27] Mass spectrometry-based approaches can potentially profile larger numbers of proteins (up to thousands of proteins in practice), including those representing potentially unanticipated proteins that may not be represented in RPPA datasets. However, one main challenge with mass spectrometry is resolving all the proteins within one sample. In whole proteomic profiling, the most abundant

proteins – often uninteresting from the standpoint of the biological questions at hand, eg, actins – will compete for detection with less abundant, but more interesting, proteins, such as signaling molecules.[28] In addition, mass spectrometry can be resource-intensive from an informatics standpoint.[29] In contrast, the RPPA platform is higher throughput in terms of numbers of samples and can be analyzed with smaller sample aliquots.

RPPA requires only one primary antibody for each target protein or its modified form, in comparison to some other immunoassays such as sandwich enzyme-linked immunosorbent assay (ELISA), which requires two primary antibodies against the same proteins.[9] Therefore, RPPA does not have an unusually high demand for specific antibodies for a given protein or phosphorylated protein. Its quantitation, sensitivity, and multiplexing capacity also largely exceeds what can be typically achieved by Western blotting and immunohistochemistry.[9] Initiatives such as the Human Protein Atlas provide lists of validated antibodies for screening studies.[30]

RPPA as a clinical proteomic platform

Conventional characterizations of cancer (including histology, tumor size, tumor grade, tumor differentiation, invasion, status of local and distant metastasis, cytogenetic analysis, and immunohistochemical staining of protein markers) do not usually detect the oncogenic signaling pathways that drive cancer growth and thus fail to identify the prime targets for intervention. The earlier-discussed benefits of RPPA, especially the ability to quantify multiple phosphorylated proteins and to probe pathway activity in very small amounts of tissue samples, make RPPA a suitable platform for patient-tailored therapy or precision medicine. For example, while standard clinical assays (immunohistochemistry and fluorescence in situ hybridization) can detect HER2 total proteins or their gene copy numbers in breast cancer, these assays do not measure HER2 protein activity and the activity of the HER2-regulated signaling pathways, which are the better indicator of the likelihood that anti-HER2 therapeutics may be effective. In 2013, the first commercial RPPA assay – the TheraLink HER Family Assay – was introduced by Theranostics Health[31] to quantify not only the HER2 total protein and its two heterodimerizing partners and family members (EGFR and HER3), but also the specific autophorylation sites on these receptor tyrosine kinases that are indicators of activation levels. Importantly, this assay also monitors the levels of their key downstream pathways including the Akt/mTOR

pathway (p-AKT, p-mTOR, p-S6, and p-4E-BP1), the MAPK pathway (p-Mek1/2, and p-Erk1/2), and the Jak/Stat pathway (p-Jak2 and p-Stat3), against which there are also US Food and Drug Administration (FDA)-approved therapeutics. This assay is now allowed by some insurance companies. Using this assay, some of the triple-negative breast cancers, which are generally not treated with targeted therapy, have been found to show enhanced activity of EGFR and HER3 as well as the PI3K-Akt pathway,[32] and to respond to combinatorial therapy targeting these three components in a preclinical setting. Therefore, an RPPA assay may provide actionable information for therapeutic selections.

RPPA, coupled with LCM,[33] has the potential to survey oncogenic pathways in selected cell compartments of clinical samples.[22] This is very important since cancer tissue is often an admixture of different and interacting cell sub populations (different subsets and subclones of tumor cells and the admixed stromal cells), and the cell subpopulation of therapeutic interest – such as cancer stem cells for some cancers – may constitute only a minor fraction of the cancer mass. For example, Wulfkuhle et al[34] used LCM-coupled RPPA and found metastasis-specific changes that occurred within a new microenvironment, but this change was not detected when whole section lysates were assayed.[35] Therefore, LCM-RPPA may further advance personalized therapy.

Added value of RPPA to integrative, multi-platform analysis

In studying diseases such as cancer, the value of integrative molecular analysis, incorporating multiple levels of data, has been well established. One well-known example involving data integration included one of the first studies (by Perou et al)[36] to profile breast cancers at the mRNA expression level, where mRNA data were integrated with data from immunohistochemical staining for key molecular markers of treatment response (namely, ER and HER2). This integrative analysis resulted in intrinsic molecular subtypes of human breast cancer being defined – these being eventually known as Luminal A, Luminal B, HER2-enriched, basal-like, and normal-like. Subsequent studies have defined a core set of 50 mRNAs, also known as the "PAM50" gene set,[37] which can be used to distinguish these intrinsic subtypes. Here, we represent these breast tumor subtypes and associated PAM50 genes in Figure 1, using data generated by TCGA, including data from RPPA and from RNA sequencing. As the RPPA dataset for TCGA includes on the order of 180 protein features, not all of these would be represented in the original PAM50 gene set, though a number of key proteins are (including ER, PR, HER2, EGFR, Bcl-2, CCNE1, and CCNB1). As indicated in Figure 1, RPPA features may include phosphorylated as well as total forms of a given protein. In general, Figure 1 shows good correlations between
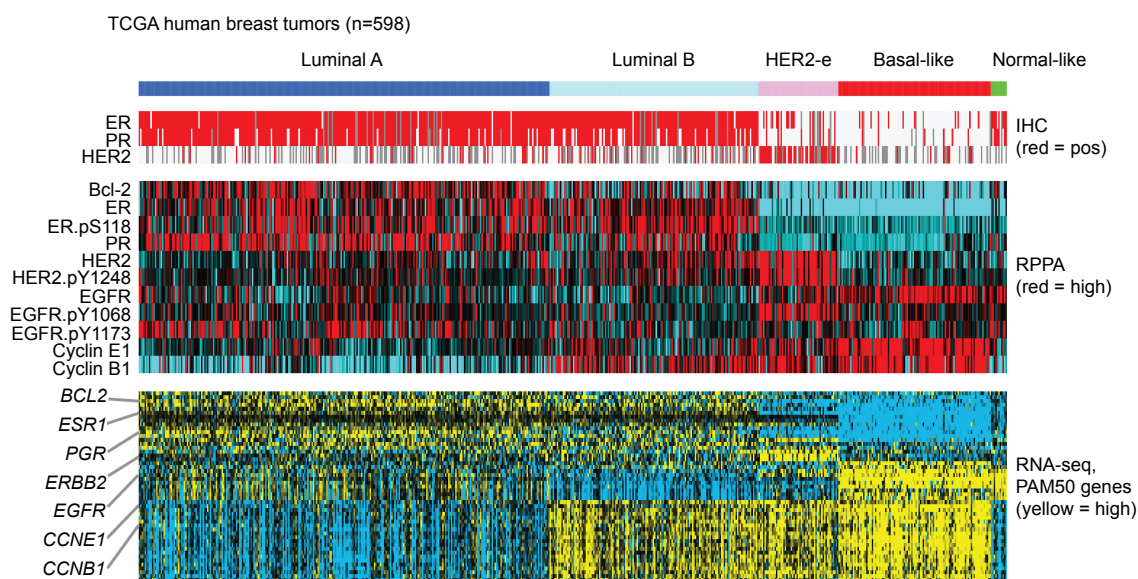

Figure 1 Proteomic and transcriptomic patterns associated with the intrinsic molecular subtypes of human breast cancer.
Notes: Data on 598 human breast tumors are from TCGA[15] (RPPA data from The Cancer Proteome Atlas dataset).[19] Using the PAM50 gene set,[37] tumors were previously classified by intrinsic molecular subtype (Luminal A, Luminal B, HER2-enriched, basal-like, and normal-like).[15] The mRNA heat map features the PAM50 genes (used to classify breast cancer subtype), while the RPPA heat map features protein equivalents of the PAM50 genes (where available). This analysis uses publicly available data but is original to this review article.
Abbreviations: TCGA, The Cancer Genome Atlas; RPPA, reverse phase protein array; HER2-e, HER2-enriched; IHC, immunohistochemistry; pos, positive; ER, estrogen receptor; PR, progesterone receptor; EGFR, epidermal growth factor receptor; RNA-seq, RNA sequencing.

these select proteins and the genes, where the proteins also appear differentially expressed among the tumor subtypes originally defined using mRNA data.

Over the last decade, the development of additional data platforms to globally profile the cell at various molecular levels – including DNA mutation, DNA copy, DNA methylation, microRNAs, and proteins – has offered us challenges and opportunities to integrate these various data types in meaningful ways. TCGA is an ambitious project currently ongoing to comprehensively profile more than 10,000 cancers of various histological subtypes, using all of the above platforms. In recent years, TCGA consortium members have carried out numerous comprehensive molecular analysis studies focusing on a specific disease,[15–17,21,38] as well as numerous "pan-cancer" studies that make observations that cut across different diseases.[14,18,19] TCGA datasets include an extensive RPPA dataset, the most recently published version being comprised of 3,467 patient samples from eleven tumor subtypes, using 181 high-quality antibody features targeting 128 total proteins and 53 post translationally modified proteins.[19] Additional RPPA data generation is ongoing, for additional cancer subtypes currently under study by TCGA.

In a number of these TCGA-initiated studies, the inclusion of RPPA data into multi-platform analyses has led to key insights, just a few examples of which are noted here (these examples involving analysis work by this review's leading author). In breast cancer,[15] a set of proteins with core roles in the PI3K pathway were examined by RPPA (including phospho-protein levels of Akt, mTOR, GSK3, S6K, and S6, and total levels of pathway inhibitors PTEN and INPP4B), and it was found that the pathway as a whole appeared more active in the basal-like subtype of breast cancer; this finding was corroborated in the corresponding mRNA data, by examining PI3K-associated transcriptional signatures. In clear cell renal cell carcinoma,[16] molecular correlates of patient survival were defined for each of four different data platforms (RPPA, RNA-sequencing, microRNA-sequencing, and DNA methylation arrays). By RPPA, top survival correlates included AMPK and ACC, which were oppositely correlated to each other, thereby suggesting a metabolic shift of the cell from oxidative phosphorylation to aerobic glycolysis. The RPPA-based observations led to a focused analysis of metabolism involving all platforms, which further supported a type of glycolytic shift being associated with more aggressive kidney cancer.

In TCGA's recent lung adenocarcinoma study,[17] lung tumors could be separated into three main groups on the basis of RPPA and mutation data: 1) those tumors with the PI3K-Akt branch of mTOR pathway appearing activated (either PIK3CA activating mutation or high p-Akt), 2) those tumors with LKB1-AMPK branch inactivated (either STK11 mutation or low combined levels of LKB1/p-AMPK), and 3) those tumors unaligned with the above. The RPPA data were also used to define an mTOR pathway proteomic signature, which was the average of the phosphor-proteomic forms of 4E-BP1, 70S6K, and S6. In principle, mTOR signaling may be activated, by either Akt (eg, via PI3K) or inactivation of AMPK (eg, via STK11/LKB1 loss), and in fact those tumors that showed alterations in either mTOR-associated pathway branches, as defined earlier, did show increased mTOR pathway activity by RPPA. This finding illustrates the need for incorporating prior biological knowledge into pathway analysis of RPPA data, where pathways may not always behave in a linear fashion. In addition, the analysis demonstrated many cases that showed aberrant phospho-Akt or loss of LKB1 at the protein level, without an associated genetic driver, illustrating that RPPA data hold additional information on pathway activities that may not be fully captured by mutation analysis alone or by our current understanding of potential driver alterations for these key pathways.

Need for effective visualization approaches to RPPA data

As molecular datasets become more rich and complex, a challenge that presents itself is that of making results from integrative analyses understandable to everyone, of which effective data visualization would be a key component. While statisticians and computer scientists may often express results in terms of statistical P-values, statistical significance may not necessarily translate into biological significance. Almost any pattern that is of biological significance could be shown as such by some visual presentation of the data, thereby allowing the results to be even more apparent and more readily accepted by others. With RPPA and other molecular data platforms, there is an obvious need for better software tools to allow researchers, who may not have the benefit of a strong computational background, to be able to access and visualize multidimensional molecular datasets (Oncomine[39] or CBioPortal[40] being good examples of these types of tools). At the same time, we can also seek to better utilize the printed page, which represents a static view of the data but one that can also be readily digested and shared with others.

In our own analysis work, we have formulated for ourselves some general guidelines for effective data visualization of genomic datasets as follows: 1) Show the data: the visualization should strive to show the actual data underlying

a pattern of interest; the more people can "see" the associated patterns for themselves, the more the reported trends in the data may become more concrete in people's minds and readily accepted. 2) Order elements deliberately: the ordering of the data elements should be deliberate, to provide optimal viewing of the primary pattern meant to be visualized. 3) Enable generalizations: the visualization should provide a global view of the data, allowing one to make generalizations about the system under study. 4) Focus on relevant specifics: as well as allowing one to follow the overall trends, the visualization should put emphasis on specific data features of particular relevance. 5) Put things in context: as much as possible, the data elements should be annotated, in order to put them into some meaningful context for the benefit of viewers having knowledge of the domain. In the following, we illustrate these guidelines, using concrete examples.

In the seminal paper by Perou et al[36] who first identified the intrinsic molecular subtypes of breast cancer, a key figure presented the results of unsupervised clustering of gene expression profiling with associated heat map. These subtypes are represented in Figure 1, using heat maps of TCGA data from human breast tumors. While our above visualization guidelines were not explicitly stated in the Perou et al paper, these guidelines were in fact put to good use in Perou et al's presentation of the unsupervised clustering heat map.

Show the data: while real data are never perfect and can even be somewhat "messy" (which could be considered the case for the earlier cDNA microarrays in particular), showing a heat map of the differentially expressed genes further reinforces the notion of distinct subtypes of breast cancer (as is also the case for Figure 1).

Order elements deliberately: for the Perou et al paper, a computer algorithm grouped the samples and genes based on distinctive patterns, thereby defining the breast cancer subtypes and the genes that underlie these subtypes.

Enable generalizations: from all of the individual data points derived from all of the samples and the thousands of genes profiled, Perou et al arrive at four or five basic subtypes of breast cancer, which represents a powerful generalization for this disease.

Focus on relevant specifics: two proteins, ER and HER2 (these genes being noted both in Perou et al's figure and in our Figure 1), are of particular relevance to the intrinsic subtypes, as these represent known biomarkers of treatment response, thereby grounding these subtypes in reality.

Put things in context: ER and HER2 received particular focus in the Perou study, due to their previously established roles in the biology of breast cancer, thereby providing

meaningful context as to the biology underlying the intrinsic subtypes.

Developing effective data visualization approaches for molecular profiling datasets continues to be an active area of research, one notable example being the recent visualization "sub-challenge" put forth by the HPN-DREAM consortium,[41] as part of its overarching challenge of network inference in breast cancer using RPPA data. The goal of HPN-DREAM's visualization sub-challenge was for participants to devise novel approaches to represent a complex RPPA dataset, involving ~45 phosphoproteins being profiled for four different breast cancer cell lines, grown under conditions of eight different ligand stimuli, which stimuli groups were further divided into treatments by either one of three inhibitors or dimethyl sulfoxide (DMSO) vehicle control, with profiles taken for each treatment over seven different time points. In all, this RPPA dataset represented over 48,000 data points over multiple cell lines, stimuli conditions, inhibitor treatments and times, and the obvious challenge here was to find ways to meaningfully present all of these data.

Figure 2 provides a visualization of the above RPPA dataset, which was originally submitted by this review's lead author, as part of the HPN-DREAM challenge. This visualization makes use of our above guidelines for present complex molecular datasets.

Show the data: using a heat map to display the protein expression patterns allows for compact presentation of all the individual data points.

Order elements deliberately: the ordering of protein features in the heat map is deliberately chosen, grouping them by biologically meaningful protein class (PI3K, apoptosis, cell cycle, MAPK, etc). The ordering of sample profiles in the heat map is likewise deliberate; this would represent a critical step in our defining what patterns might be readily viewed. The sample profiles are grouped first by cell line, then by stimulus, then by inhibitor, then by time; in this way, the viewer can readily look up how a particular stimulus and inhibitor treatment impacted protein expression.

Enable generalizations: the single page presentation provides a global view of all the data, allowing the viewer to observe overall trends and to make generalizations. Overall trends that are apparent in the figure include AKT inhibitors activating p-Akt in all cell lines, while suppressing PI(3)K activity (as measured by downstream effectors S6K, S6, and 4EBP1) in the UACC812 and BT549 cell lines. Other patterns are discernible here, such as HGF-induced MET, or EGF- induced EGFR/Her2, in both BT549 and BT20 cell lines.
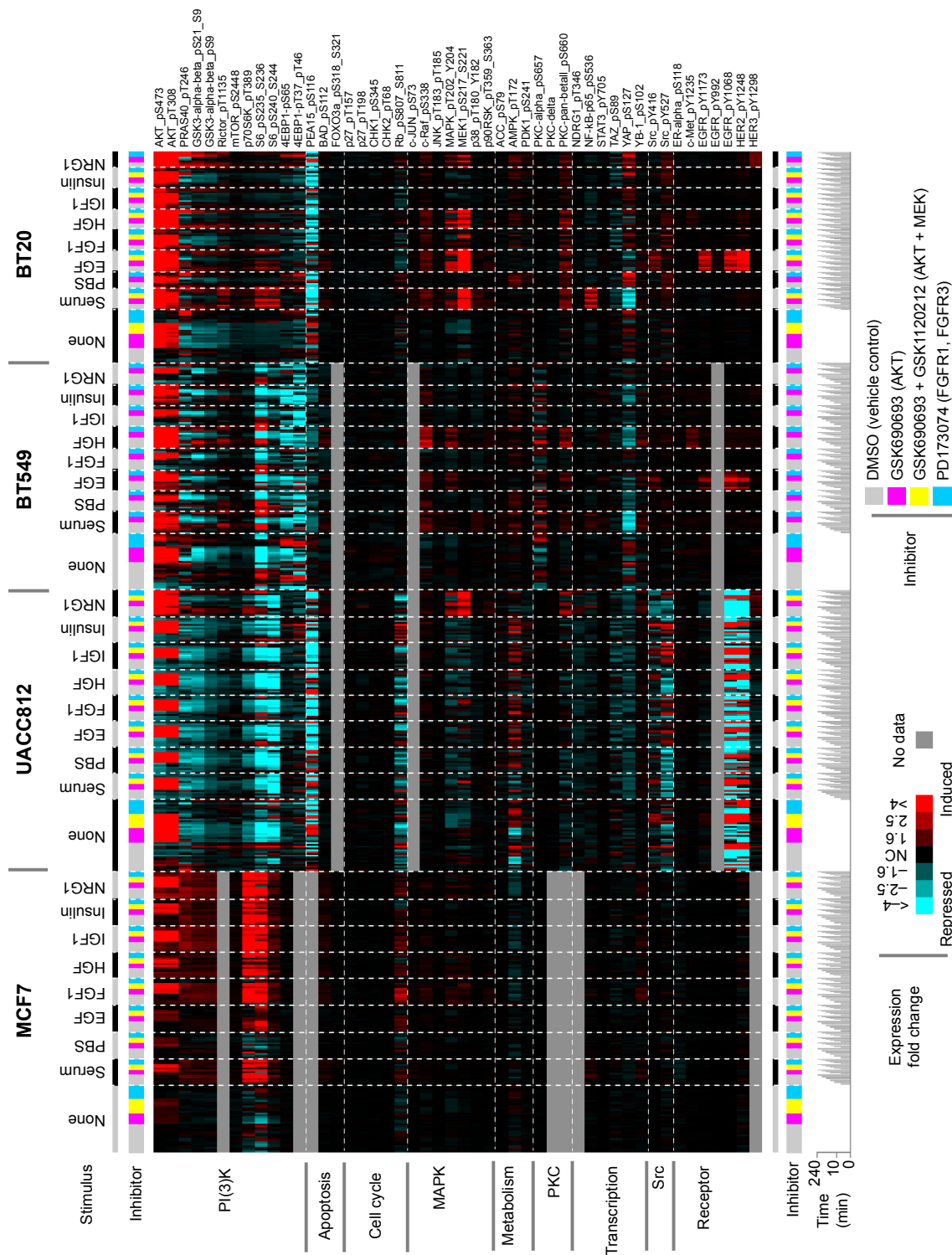
Figure 2 Example of how RPPA data representing multiple cell types, treatment conditions, and time points may be visually presented.

Notes: Heat map graphically shows expression changes for 47 proteins, in response to treatment by various stimuli and inhibitors, across four different breast cancer cell lines. For each cell line, the expression fold changes (relative to no treatment at time 0) are shown (red, induction; blue, repression). Each profiled sample was treated with or without a specific stimulus (serum, PBS, EGF, FGF1, HGF, insulin, NRG1) and with or without a specific inhibitor (AKT inhibitor, AKT + MEK inhibitors, FGFRI/FGFR3 inhibitor). For each stimulus–inhibitor combination, treatment times varied from 5 minutes to 4 hours, as indicated by the time plot along the bottom. The present visualization was rendered, using Microsoft Excel to center the expression values and to sort the data elements, JavaTreeview[5] to generate the heat map images, and Adobe Illustrator to assemble and annotate the pieces. This visualization was previously entered as part of the HPN-DREAM breast cancer network inference challenge (sub-challenge 3: visualization),[41] but not previously published in an article. This visualization has been posted on the Internet,[42] but not previously published in an article.

Abbreviations: RPPA, reverse phase protein array; PBS, phosphate buffered saline; DMSO, dimethyl sulfoxide; EGF, epidermal growth factor; FGF1, fibroblast growth factor 1; HGF, hepatocyte growth factor; NRG1, neuregulin 1.

Focus on relevant specifics: the visualization allows one to readily look up patterns for specific stimulus–inhibitor combinations. While all of the proteins represented in the dataset could be presented here, another dataset with a larger number of features might require one to focus on the features most relevant to the question at hand.

Put things in context: pathway knowledge provides the context that guides the ordering of the protein features, allowing the viewer to scan for a particular pathway or functional group.

The visualization guidelines described earlier are meant to be generalizable and not necessarily limited to heat maps. For example, many of the TCGA studies (eg, the clear cell kidney cancer study mentioned earlier, highlighting the role of glycolytic shift in more aggressive cancers) have made effective use of pathway diagrams. A pathway diagram, eg, one where the nodes may represent genes or proteins and would provide information on genomic alteration or differential expression, can also illustrate our general guidelines.

Show the data: all of the relevant components of the pathway should be considered (eg, not limiting ourselves to only components that would appear altered in a preconceived direction).

Order elements deliberately: the ordering of elements in the diagram is dictated by our prior knowledge of the pathway flow.

Enable generalizations: when considering the entire pathway, sub-pathways within the larger pathway may be shown to broadly change in a given direction.

Focus on relevant specifics: critically altered nodes in the pathway may also be identifiable.

Put things in context: our prior knowledge of the pathway, a product of decades of cumulative research, provides a meaningful context and framework for the data. In addition to what would be represented with the examples discussed here, other visualization approaches could be explored as well. In fact, there were many other creative entries submitted by others to the HPN-DREAM visualization challenge,[42] some of which may not be able to be fully captured on the static page and may require the development of new software tools, but which can help stimulate additional thinking in this important area.

Conclusion and future perspective

The recent explosion of molecular data, made possible by the wider availability of new technologies to comprehensively profile the cell, has enabled us to live in interesting times. Rather than being daunted by all these data, we should be excited at the potential for discovery. More and more, RPPA is establishing itself as a core data platform, which may be used in conjunction with other data platforms, for examining signaling pathways as they may change between diseased and healthy cells. Integration with other data platforms will be important in our maximizing the potential of RPPA as a research tool, along with the importance of having good knowledge of molecular biology and an effective grasp of effective visualization techniques. In addition, one area that holds great potential is the clinical use of RPPA, eg, for personalizing therapeutics, clinical diagnosis, and drug discovery,[43,44] which can extend to human diseases beyond cancer.

Additional efforts are needed to facilitate more widespread use of RPPA, in both areas of research and clinical practice. Technical improvements regarding the reagents used in RPPA data generation would be of great potential benefit, including improvements in labeling chemistry to allow for higher sensitivity, as well as an expansion of validated antibodies for use with the platform. Improved software tools for automated image analysis could save considerable technician time and thereby lower the costs of implementation. With regard to the clinical setting, automation of an analysis workflow to integrate genomics data and RPPA would greatly aid future applications in personalized medicine.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270(5235):467–470.
2. Lashkari DA, DeRisi JL, McCusker JH, et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A*. 1997;94(24):13057–13062.
3. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100(1):57–70.
4. Spurrier B, Ramalingam S, Nishizuka S. Reverse-phase protein lysate microarrays for cell signaling analysis. *Nat Protoc*. 2008;3(11): 1796–1808.
5. Ekins RP. Multi-analyte immunoassay. *J Pharm Biomed Anal*. 1989;7(2): 155–168.
6. Ekins RP, Chu FW. Multianalyte microspot immunoassay – microanalytical "compact disk" of the future. *Clin Chem*. 1991;37(11):1955–1967.

7. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science*. 2000;289(5485): 1760–1763.

8. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol*. 2001;2(2):RESEARCH0004.

9. 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*. 2001;20(16): 1981–1989.

10. Mircean C, Shmulevich I, Cogdell D, et al. Robust estimation of protein expression ratios with lysate microarray technology. *Bioinformatics*. 2005;21(9):1935–1942.

11. Romeo MJ, Wunderlich J, Ngo L, Rosenberg SA, Steinberg SM, Berman DM. Measuring tissue-based biomarkers by immunochromatography coupled with reverse-phase lysate microarray. *Clin Cancer Res*. 2006;12(8):2463–2467.

12. Tibes R, Qiu Y, Lu Y, et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther*. 2006; 5(10):2512–2521.

13. Akbani R, Becker KF, Carragher N, et al. Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society. *Mol Cell Proteomics*. 2014;13(7):1625–1643.

14. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113–1120.

15. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.

16. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013;499(7456):43–49.

17. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511): 543–550.

18. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929–944.

19. Akbani R, Ng PK, Werner HM, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun*. 2014;5:3887.

20. Brennan CW, Verhaak RG, McKenna A, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155(2):462–477.

21. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497(7447):67–73.

22. Mueller C, Liotta LA, Espina V. Reverse phase protein microarrays advance to use in clinical trials. *Mol Oncol*. 2010;4(6):461–481.

23. Chen G, Gharib TG, Huang CC, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics*. 2002;1(5): 304–313.

24. Rogers S, Girolami M, Kolch W, et al. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*. 2008;24(24):2894–2900.

25. Creighton CJ. Multiple oncogenic pathway signatures show coordinate expression patterns in human prostate tumors. *PloS One*. 2008;3(3): e1816.

26. Sheehan KM, Calvert VS, Kay EW, et al. Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. *Mol Cell Proteomics*. 2005;4(4): 346–355.

27. Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res*. 2004;3(2):179–196.

28. Guo S, Zou J, Wang G. Advances in the proteomic discovery of novel therapeutic targets in cancer. *Drug Des Devel Ther*. 2013;7: 1259–1271.

29. Li YF, Radivojac P. Computational approaches to protein inference in shotgun proteomics. *BMC Bioinformatics*. 2012;13 Suppl 16:S4.

30. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010;28(12):1248–1250.

31. Theranostics Health [webpage on the Internet]. Available from: http://www.theranosticshealth.com. Accessed May 1, 2015.

32. Tao JJ, Castel P, Radosevic-Robin N, et al. Antagonism of EGFR and HER3 enhances the response to inhibitors of the PI3K-Akt pathway in triple-negative breast cancer. *Sci Signal*. 2014;7(318):ra29.

33. Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. *Science*. 1996;274(5289):998–1001.

34. Wulfkuhle JD, Speer R, Pierobon M, et al. Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy. *J Proteome Res*. 2008;7(4):1508–1517.

35. Mueller C, deCarvalho AC, Mikkelsen T, et al. Glioblastoma cell enrichment is critical for analysis of phosphorylated drug targets and proteomic-genomic correlations. *Cancer Res*. 2014;74(3):818–828.

36. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.

37. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009; 27(8):1160–1167.

38. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014; 507(7492):315–322.

39. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6(1):1–6.

40. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–404.

41. Synapse [homepage on the Internet]. HPN-DREAM breast cancer network inference challenge. Synapse. Available from: https://www.synapse.org/#!Synapse:syn1720047. Accessed September 11, 2014.

42. Synapse [homepage on the Internet]. HPN-DREAM visualization challenge. Synapse. Available from: https://www.synapse.org/#!Synapse:syn2274074. Accessed May 1, 2015.

43. Gallagher RI, Espina V. Reverse phase protein arrays: mapping the path towards personalized medicine. *Mol Diagn Ther*. 2014;18(6): 619–630.

44. Masuda M, Yamada T. Signaling pathway profiling by reverse-phase protein array for personalized cancer medicine. *Biochim Biophys Acta*. 2015;1845(6):651–657.

45. Saldanha AJ. Java Treeview – extensible visualization of microarray data. *Bioinformatics*. 2004;20(17):3246–3248.