Anil Sood

and

Katelyn Handely

8 samples probed with

485 antibodies

RPPA Set180

Mar 23 2021

**Your samples are arranged in RPPA CORE 03012021\_180 (Set180).**

**Please refer to this set number if you have any questions regarding your samples or data.**

Processed by: Xing, Qinghua, Doris, Yiling

**PROCEDURES FOR RPPA CORE 03012021\_180 (Set180).**

1. Tissue or cell lysate samples were serially diluted two-fold for 5 dilutions (undiluted, 1:2, 1:4, 1:8; 1:16) and arrayed on nitrocellulose-coated slides in an 11x11 format to produce sample spots.
2. Sample spots were then probed with antibodies by a tyramide-based signal amplification approach and visualized by DAB colorimetric reaction to produce stained slides.
3. Stained slides were scanned on a Huron TissueScope scanner to produce 16-bit tiff images.
4. Sample spots in tiff images were identified and their densities quantified by Array-Pro Analyzer.
5. Relative protein levels for each sample were determined by interpolating each dilution curve produced from the densities of the 5-dilution sample spots using a "standard curve" (SuperCurve) for each slide (antibody). SuperCurve is constructed by a script in R, written by Bioinformatics. Relative protein levels are designated as log2 values and are in the Excel worksheet labeled "RawLog2."
6. All relative protein level data points were normalized for protein loading and transformed to linear values, which are designated "Normalized Linear" in the Excel worksheet labeled “NormLinear.”
   * The linear values can be used for bar graphs or further analyses according to your study design.
   * We have recently implemented an improved normalization algorithm for protein loading correction and antibody variation adjustment. This approach is critical to provide accurate values for RPPA data merging if you have samples performed in separate RPPA sets that you wish to combine. We listed the protein loading correction factors (CF1 and CF2) in the last column on this page (NormLinear) for your reference. Each sample has its unique correction factor. CF1 values are calculated within your sample set while CF2 values are calculated within all 1056 samples on the same slide. In contrast to CF1 indicating protein loading factor among your sample set, CF2 determines protein level for each sample in the entire set of RPPA. **If CF2 is less than 0.25 or greater than 2.5, we consider these samples “outliers,”** indicating that protein concentration is much lower or much higher than the other samples. **We suggest you exclude these “outliers” from further analysis. (One exception: intrinsic protein expression patterns can skew the correction factor to some extent. In that case, you may want to include this data, but we suggest that you carefully examine the dataset.)**
7. "Normalized Linear" values were transformed to log2 values (designated "NormLog2" in worksheet) and then median-centered for hierarchical clustering analysis (designated "NormLog2\_MedianCentered" in the worksheet).
8. We included 2 heatmaps for your reference: (1) unsupervised hierarchical clustering (unsupervised for both antibodies and samples) and (2) antibody unsupervised with samples arranged in the order in which you submitted them. Please note that your RPPA data report is presented in Excel format. Heatmaps display visualization figures for your reference. Further bioinformatics analysis should be performed at your end.
9. The heatmaps were developed by the UT MD Anderson Cancer Center Department of Bioinformatics and Computational Biology, In Silico Solutions, Santeon, and SRA International. They appear as pdf files in your reports.
   * Heatmaps were previously generated in Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) as a hierarchical cluster using Pearson Correlation and a center metric. The resulting heatmaps were visualized in Java Treeview (http://jtreeview.sourceforge.net) and presented as high-resolution bmp files. This software is available for customer-generated heatmaps.
10. We stained RPPA slides for 485 unique antibodies, which were analyzed by Array-Pro Analyzer 6.3 then by SuperCurve\_1.5.0 via SuperCurveGUI\_2.1.1. We performed QC tests for each antibody staining (slide). The Probability (QC Score) values are also included in the data spreadsheet for your reference. A QC score above 0.8 indicates good antibody staining. We included only the data for the 485 individual antibodies with QC Scores higher than 0.8 in the heatmaps.
    * For all tissue samples, we remove 4 antibodies that cross-react with "damaged components" (of unknown mechanism).
    * For mouse tissue, rat tissue, or xenograft samples, we also remove the 67 individual mouse and rat antibodies used in staining.
11. The bioinformatics should be done at your end from the Excel file we provide. We recommend that you create bar graphs based on the data in the Excel file. The heatmaps are solely to provide overall patterns.

**Antibody status for RPPA**

* (V) = Validated antibody for RPPA
* (C) = Use with caution; validation in progress
* (Q) = These antibodies recognize unidentified "damaged" component(s) in addition to its specific protein. The "damaged" component(s) were observed only in certain tissue samples.
* (E) = Under evaluation
* (M) = Mouse antibody was used
* (G) = Goat antibody was used
* (R) = Rabbit antibody was used
* (T) = Rat antibody was used

**Example:** Akt\_pS473-R-V\_GBL9016996 means the antibody specifically recognizes Akt phosphorylated on serine 473. This is a rabbit antibody validated for RPPA application. GBL9016996 is the slide ID (barcode). The slide ID is not included in the final version of the heatmap.

**Your samples are arranged in RPPA CORE 03012021\_180 (Set180). Please refer to this set number if you have any questions regarding your samples or data.**

**Important Information for Excel file data (RPPA Report)**

Thank you for using the RPPA Core for your functional proteomics studies.

Your RPPA results are provided in an Excel file with multiple tabs:

1. The 1st tab (labeled “RawLog2”) provides the raw RPPA data in log2 values without any normalization.
2. The 2nd tab (labeled “NormLinear”) provides normalized RPPA data in linear values, which can be used for bar graphs or line graphs. **We have recently implemented an improved normalization algorithm for protein loading correction and antibody variation adjustment (see Step 6 in the previous “Procedures” section).** We emphasize that you exclude these “outliers” from further analysis.
3. The 3rd tab (labeled “NormLog2”) provides normalized RPPA data in log2 values for further bioinformatics analysis.
4. In order to generate heatmaps for data visualization, we median centered samples across each antibody from the normalized log2 values. The median-centered values are located on the 4th tab (labeled “NormLog2 MedianCentered”).

* Beware of outliers in the heatmaps. They can be determined by viewing the values in the last column (titled “CF2”) of the “NormLinear" worksheet in the Excel file.
* “Red” in the heatmaps indicates the values are above the median and “green” indicates the values are below the median.

Heatmap development was supported in part by the U.S. National Cancer Institute (NCI; MD Anderson TCGA Genome Data Analysis Center) grant numbers CA143883 and CA083639, the Mary K. Chapman Foundation, the Michael & Susan Dell Foundation (honoring Lorraine Dell), and MD Anderson Cancer Center Support Grant P30 CA016672 (the Bioinformatics Shared Resource).

The Functional Proteomics RPPA Core facility is supported by MD Anderson Cancer Center Support Grant # 5 P30 CA016672-40. **Publications using data generated by the Core should cite the Core grant in the Acknowledgements section.**

**Important Information Regarding the RPPA Antibody List**

Reverse phase protein arrays (RPPA) are dependent on the quality of antibodies and control metrics used for each antibody and each sample. We perform extensive evaluation of each antibody and also have a continuous reassessment program to ensure delivery of high quality data.

It is important to emphasize that although the majority of our antibodies are from commercial sources and usually are monoclonal antibodies, behavior of the antibodies can change over time and between batches. Furthermore, with ongoing use and reassessment on our RPPA platform, new information comes to light (particularly under different conditions and with different types of samples) that may result in a change in validation status and raise concerns about the utility of specific antibodies under particular conditions. We are committed to providing high quality data and will thus update the description of the antibody performance on our RPPA website: <http://www.mdanderson.org/education-and-research/resources-for-professionals/scientific-resources/core-facilities-and-services/functional-proteomics-rppa-core/index.html>.

We will also provide updates on major concerns by e-mail; however, we recommend that you frequently visit our webpage to review the antibody list and validation status of each antibody. The antibodies are labeled as follows: (1) "valid" which means they perform well in all assays available, (2) "use with caution" means that under most circumstances the antibodies provide high quality information but might not perform well under some conditions, or (3) "under evaluation" means that the antibodies are currently being evaluated or re-evaluated for performance. We would like to emphasize that RPPA is best thought of as a high-throughput screening assay. We recommend that all RPPA results be confirmed by an orthologous approach.

The antibodies listed on our website perform well in both cell lines and tissue samples based on the designations listed above. Any antibodies that are not on our current list may have challenges. If you have antibodies that are no longer listed, we recommend that you consider the data with caution, and contact the RPPA Core to determine the reason for removal from our standard list. You may also contact us on the status of any specific antibodies of interest not listed.

**Antibodies that do not pass current quality control should be deleted from previous data sets**

We have identified antibodies by which their performance no longer meet our standards as they have been shown to have liabilities under a number of conditions.

These antibodies include:

* AIB1 (BD Biosciences #611105)
* Caspase-9\_cleavedD330 (Cell Signaling Technology #9501)
* COX2 (Epitomics #2169-1)
* PTCH (SDI 2113.0002)
* TAZ\_pS89 (Santa Cruz #sc-17610)

We no longer run these antibodies in RPPA, and will continue searching for alternative sources. **If you have these antibodies in your previous data set, we recommend that you remove them from further consideration.**

**Antibodies that cross-react within EGFR family members**

As a further part of our analysis, we have found that a number of antibodies to phospho-HER2 and phospho-EGFR cross-react when the opposite molecule is present at very high levels. However, they perform well when the other receptor is not highly expressed so we will continue to provide information for these antibodies.

These antibodies include:

1. EGFR antibody from Santa Cruz (#sc-03): cross-reacts significantly with overexpressed HER2. (We have removed this antibody from our list.)
2. Phospho-EGFR Y1068 (Cell Signaling Technology, #2234) and phospho-EGFR Y992 antibodies (Cell Signaling Technology, #2235): modestly cross-react with overexpressed phosphorylated HER2. (We have removed these antibodies from our list.)
3. Phospho-HER2 Y1248 antibody from Millipore (#06-229): modestly cross-reacts with phospho-EGFR. (We have removed this antibody from our list.)

In contrast, we have demonstrated a set of non-cross-reacting antibodies against EGFR family members under our test conditions. We have included these antibodies in our revised antibody list. These antibodies include:

1. EGFR antibody from Cell Signaling Technology (#2232): does not demonstrably cross-react with HER2.
2. HER2 antibody from Lab Vision (#MS-325-P1): does not demonstrably cross-react with EGFR.
3. Phospho-EGFR Y1173 antibody from Abcam (#ab32578): does not detectably cross-react with any other EGFR family member including HER2.

**Antibodies that cross react with “damaged components” in tissue samples**

As part of our continuous reassessment process, we have identified a series of antibodies that may not perform well in tissue samples. While we do not understand the mechanism, we are concerned that these antibodies may cross-react with “damaged components” present in tissue samples. We therefore believe that tissue samples containing high levels of these components identified by these antibodies should be removed from further consideration. Consequently, we have removed these antibodies in datasets containing **tissue samples**.

These antibodies include:

* c-Met (Cell Signaling Technology #3127)
* ERCC1 (Lab Vision #MS-671-PO) (Not performed in current RPPA analysis)
* Caspase 8 (Cell Signaling Technology #9746)
* Cleaved PARP (Cell Signaling Technology #9546) (Not performed in current RPPA analysis)
* Rab25 (Covance Custom) (Not performed in current RPPA analysis)
* Rb (Cell Signaling Technology #9309)
* SETD2 (Abcam #ab69836) (Not performed in current RPPA analysis)
* Smac (Cell Signaling Technology #2954)
* Snail (Cell Signaling Technology #3895)

Please note that we designate these antibodies as “Used for QC” in our antibody list. We report these antibodies in cell line samples but not in tissue samples.

**Protein phosphorylation in tissue samples may change during sample processing**

We have evaluated and reported that phosphorylation of certain specific proteins (especially the EGFR/MAPK signaling module) in tissues can be altered by cold ischemia during sample processing while levels of total protein and the majority of phosphoproteins remain constant. Although the global proteome is remarkably stable, we still recommend that every effort be made to limit the time from tissue excision to sample preservation, aiming to reduce modifiable pre-analytical variables.

Please refer to the following references regarding phosphoproteins susceptible to tissue sample processing.

1. Hennessy, B. and Lu, Y. et al., A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. Clinical Proteom 6:129-151, 2010. PMID 21691416.
2. Mertins, P. and Yang, F. et al., Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. Molecular and Cellular Proteomics 13:1690-1704, 2014. PMID 24719451.

**In Summary**

We provide RPPA data of your 8 samples probed with 485 antibodies. The dataset is arranged in the Excel file, and the two heatmaps are provided for visualization of each dataset.

Please review your dataset with 8 samples carefully, especially on the 2nd tab (labeled “NormLinear”) of the Excel file. We highlighted these samples that had problems with “protein loading,” which we consider to be either “outliers” or having protein expression patterns that skew the correction factors. Please interpret this data with caution.