https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core.html

RPPA Dataset Report

**Control**

**and**

**Control**

**48 samples probed with 501 antibodies**

**RPPA Set198**

**May 19 2023**

Processed by: Yan, Qinghua, Xing

**Your samples are arranged in RPPA CORE 04242023\_198 (Set198).**

Please refer to this set number and the PI or submitter if you have any questions regarding your samples or data from this set.

*The Functional Proteomics RPPA Core is supported by MD Anderson Cancer Center Support Grant # 5 P30 CA016672-40.*

***Please cite this grant in the Acknowledgements section in your publications.***

Table of Contents

The dataset for each report consists of an Excel file and 2 pdf files containing clustered heatmaps. For details regarding the Excel file and heatmaps, please refer to the sections indicated below.

|  |  |
| --- | --- |
| The RPPA Procedure | 3 |
| The Excel File | 4 |
| The Antibody Labeling Key | 5 |
| The Heatmaps | 6 |
| The RPPA Antibody List | 6 |
| About Sample Quality | 7 |

The RPPA Procedure

1. **Sample preparation and spotting**. Protein extracts were diluted 2-fold in five serial dilutions (undiluted to 1:16) and spotted onto nitrocellulose-coated slides. These extracts are arranged in a 4x12-pin-11x11 subarray that accommodate 1056 serially diluted samples plus replicate vertical controls on each slide.
2. **Protein detection of samples.** The protein of interest was detected by probing with a specific antibody, amplifying the signal via a tyramide amplification system, and visualized via DAB colorimetric reaction. Each slide was probed with one antibody. The detection system used was a GenPoint-based staining kit from Agilent.
3. **Digital imaging of slide and quantification of spots.** Digital images of slides were obtained by scanning on a Huron TissueScope scanner producing 16-bit TIFF files. Spot intensities from the TIFF files were determined via Array-Pro Analyzer software.
4. **Determination of relative protein levels.** Mean net intensities of spots from all 1056 samples were used for curve-fitting for each slide (RPPA Super Position and Concentration Evaluation, aka SPACE). Relative protein level for each sample was determined by converging its sample dilution series to one point (EC50) and interpolating in RPPASPACE.
* *RPPASPACE was developed by the UT MD Anderson Cancer Center Department of Bioinformatics and Computational Biology.*
1. **Producing the RPPA dataset.** Samples were grouped by datasets into separate reports then normalized for loading and Set differences then median centered for heatmaps.

Some antibodies were removed by sample type as follows:

* Cell line: negative controls were removed
* Human tissue: negative controls and “tissue reactive” antibodies were removed (for details about “tissue reactive” antibodies, see The Antibody Labeling Key )
* Mouse or rat tissue: negative controls, “tissue reactive” antibodies, and all antibodies produced in mouse and rat were removed

If a report contains more than one sample type, the report is treated as the sample type requiring fewer antibodies to prevent introduction of spurious results.

* Example: if a report contains cell line and human tissue samples, the report will have negative control and “tissue reactive” antibodies removed.

In Set198, there were 4 “tissue reactive” antibodies and 68 antibodies of mouse or rat origin.

**A note about the RPPA platform:** we consider RPPA a screening or exploratory procedure and recommend that any RPPA result of interest be confirmed by an orthologous approach.

The Excel File

The Excel file contains the numerical values of your dataset and consists of five main worksheets as indicated by the labels on the worksheet tabs: (1) L2 (log\_2), (2) L3 (log\_2), (3) L4 (log\_2), (4) L4 (linear), and (5) L4 (CHM).

* + - 1. **L2 (log\_2):** Refers to Level 2 (L2) data, which is raw RPPA data in log2 values without any normalization that is obtained directly from RPPASPACE.

Note each worksheet has identical headings (rows 2-10). The following headings are for RPPA Core cataloging purposes that aid in processing reports in the automated RPPA system: Category\_#, Slide ID, and Antigen ID.

* + - 1. **L3 (log\_2):** Refers to Level 3 (L3) data, which is Level 2 data that has been normalized for sample loading differences by bidirectionally median centering values by antibody then by sample.
			2. **L4 (log\_2):** Refers to Level 4 (L4) data, which is Level 3 data adjusted for Set-to-Set differences (batch effects) by normalizing identical control samples in this set with an invariant control sample set and applying the adjustment (differences in means × inverse of standard deviation ratio) to each corresponding data point.
			3. **L4 (linear):** L4 (log\_2) data that has been converted from log2 to linear values. These values can be used to construct bar graphs. Note that L4 (linear) values are already normalized; no further conversions are necessary.
			4. **L4 (CHM):** L4 (log\_2) data that has been median centered by antibodies. This data corresponds to the clustered heatmaps (CHMs) produced in the report.

In addition to the data worksheets, there are two worksheets displaying quality results for antibodies and samples:

* + - 1. **Antibody QC Scores:** Slide staining (antibody) quality is determined using an algorithm that considers various aspects of the slide, including signal, background, and expected vs observed dilution results. This is expressed as a probability with 0 being worst and 1 being best. We include only slides with a QC Score of at least 0.8, which we determined to be the minimum value for acceptable slide quality.
			2. **Sample QC metrics:** **Total protein content** reflects the average signal of a sample across all antibodies. It is derived from normalizing all 1056 samples on a slide. Lower numbers reflect lower average signal; higher numbers reflect higher average signal; values close to 1 indicate little adjustment was needed for sample normalization. Log2 values less than -3 are considered too low to be reliable; these samples are flagged as red highlights. We recommend censoring them from further analyses.

The Antibody Labeling Key

In your Excel file, you will see antibody labels such as “14-3-3-beta-R-V.”

The key for antibody labeling is as follows: “**Antibody Name-Species-Validation Status**.”

The single-letter designations of the antibody species are as follows:

* **R** = produced in rabbit
* **M** = produced in mouse
* **G** = produced in goat
* **T**= produced in rat

The single-letter designations of the RPPA validation status are as follows:

* **V** = Validated; antibody performs well in all available assays
* **C** = Use with Caution; antibody provides high-quality information under most circumstances, but discrepancies were detected under certain conditions
* **Q** = Antibodies that detect “tissue reactive” components in addition to its specific protein and are used for Quality Control. The unidentified “tissue reactive” components are observed only in certain tissue samples. “Tissue reactive” antibodies are also referred to as “Red Blob” antibodies.
* **E** = Under Evaluation; antibody is currently being reevaluated

In the above example, “14-3-3-beta” is the antibody name; “R” indicates the antibody was produced in rabbit; “V” indicates that the antibody is validated for RPPA.

The Heatmaps

Heatmaps are visual representations of your data and plotted from the L4 (CHM) values in the Excel file. In these heatmaps, red, black, and green indicate samples that have high, intermediate, and low levels of the specific antibody, respectively.

Heatmaps also show the extent of similarity between antibody staining patterns and sample expression patterns. The more similar the pattern, the closer antibodies and samples will cluster next to each other and the shorter the lengths of the branches (nodes) in the clustering tree.

Clustering was determined via Ward’s method for dissimilarity and 1-Pearson’s correlation for linkage. The pdf files for the heatmaps contain data that are arranged in two ways:

* **Unsupervised hierarchical clustering:** clustering is determined by the level of similarity between both samples and antibodies and occurs in both directions
* **Supervised clustering with samples in order:** antibodies are clustered based on similarity while samples are maintained in the order submitted in iLab

Please note the RPPA Core is not equipped to provide analysis beyond the level of Excel data and heatmaps. We will refer you to other sources for additional support whenever possible.

* *Heatmaps were developed by the UT MD Anderson Cancer Center Department of Bioinformatics and Computational Biology, In Silico Solutions, Santeon, and SRA International.*
* *Heatmap development was supported in part by the US 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 RPPA Antibody List

Because RPPA can be viewed as a high-throughput dot blot, antibody quality is of utmost importance. MD Anderson’s RPPA Core is committed to providing high-quality data and extensively validates antibodies applicable to our RPPA application; information is available at our website: <https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core/antibody-information-and-protocols.html>.

We continue to evaluate new antibodies as well as established antibodies as their behavior can change over time. Additionally, we continue to reassess our quality control metrics for RPPA antibodies. Therefore, the status of some antibodies may change. Current and past RPPA Antibody Lists are updated as needed and posted on our website.

Any major concerns are updated by email; however, we recommend that you visit our webpage to review the RPPA Antibody List and validation status of each antibody for your RPPA set. If your RPPA set includes 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.

About Sample Quality

We demonstrated that phosphorylation of certain proteins, especially in the EGFR/MAPK signaling module, can be altered in tissues by cold ischemia during sample processing. Although the global proteome is remarkably stable, we recommend that every effort be made to limit the time from tissue extraction to sample preservation to reduce preanalytical variables.

Further information regarding phosphoproteins susceptible to tissue sample processing is available in the following:

* 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-micro dissected human breast cancers. Clinical Proteom 6:129-151, 2010. PMID 21691416.
* 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.