BioTechniques Rapid Dispatches

Supplementary Material For:

Improved data normalization methods for reverse phase protein microarray analysis of complex biological samples

Antonella Chiechi[1,2], Claudius Mueller[2], Kevin M. Boehm[2], Alessandra Romano[2], Maria Serena Benassi[1], Piero Picci[1], Lance A. Liotta[2], and Virginia Espina[2]

[1]Laboratory of Experimental Oncology, Istituto Ortopedico Rizzoli, Bologna, Italy, and [2]Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA, USA

Materials and methods

Reverse phase protein microarray controls

A non-cellular protein control was prepared from 1mg/mL stock solution of BSA (BSA) (Fraction V) (Fisher Scientific, Fair Lawn, NJ, USA) in Dulbecco's Phosphate Buffered Saline 1X (Invitrogen, Carlsbad, CA, USA).

HeLa, HeLa + Pervanadate (BD, Franklin Lakes, NJ, USA), Jurkat Untreated, and Jurkat + CalyculinA commercial cell lysates (Cell Signaling Technology, Danvers, MA, USA) were printed on the reverse phase protein microarrays (RPMA) as quality controls to ensure deposition of sample during the array printing process and to provide evidence of the appropriated immunostaining and detection processes.

Human tissue sample collection and processing

Whole cell lysates were prepared from fresh frozen human breast cancer lymph node metastasis tissues collected with informed consent following an Institutional Review Board (IRB) approved protocol as previously described (1). Frozen lymph node metastasis tissue was pulverized then cells were lysed with protein extraction buffer: 45% T-PER (Pierce, Rockford, IL), 45% Novex Tris-Glycine SDS Sample Buffer (2X) (Invitrogen), 10% TCEP Bond Breaker (Pierce) and heated at 100°C for 5 min. Surgical tissue specimens, from liver carcinoma, ovarian carcinoma, uterus, lymph node metastasis and melanoma were collected with informed consent following an IRB-approved protocol. For this current study tissue samples were microdissected (Arcturus^{XT}, Applied Biosystems, Carlsbad, CA, USA **[Q10]**), and lysed in protein extraction buffer, heated at 100°C for 5 min. Samples were stored at -80°C prior to RPMA printing.

The 10 bone metastasis tissue samples, collected at the Istituto Ortopedico Rizzoli in Bologna, derived from different primary tumors: 2 from thyroid carcinoma, 2 from lung carcinoma, 1 from melanoma, 1 from bladder carcinoma, 1 from uterine carcinoma, 1 from renal carcinoma, 1 from soft tissue sarcoma, and 1 from an undifferentiated tumor.

Sixteen fresh-frozen, clear renal cell carcinoma specimens were obtained from the tissue bio-bank of Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan (Italy). Surgical specimens were collected with informed consent following an IRB-approved protocol and frozen at -80°C within 5 min of surgical removal. Tumor and surrounding adjacent renal cells were procured by laser capture microdissection using an Arcturus PixCell IIe instrument (Applied Biosystems). Microdissected samples were lysed as described above.

Fifteen bone marrow aspirate core biopsies were collected from patients diagnosed with multiple myeloma with informed consent following an IRB-approved protocol. Samples were obtained by trephine needle biopsy from the posterior iliac crest. Core biopsy samples were placed in protein extraction buffer and lysed using Adaptive Focus Acoustic technology (Covaris, Woburn, MA, USA) at 20% duty factor, 275 pick incident power and 200 cycles per burst for 90 s.

DNA isolation/extraction

5µL NaCl 2M were added to 50µL of sample lysate, followed by 165µL of 100% ethanol. The samples were incubated for 15 min on ice and spun at 14,000× *g* for 30 min at 4°C. The supernatant fraction, containing protein free DNA, was removed and frozen at -80°C until analysis.

Antibody validation

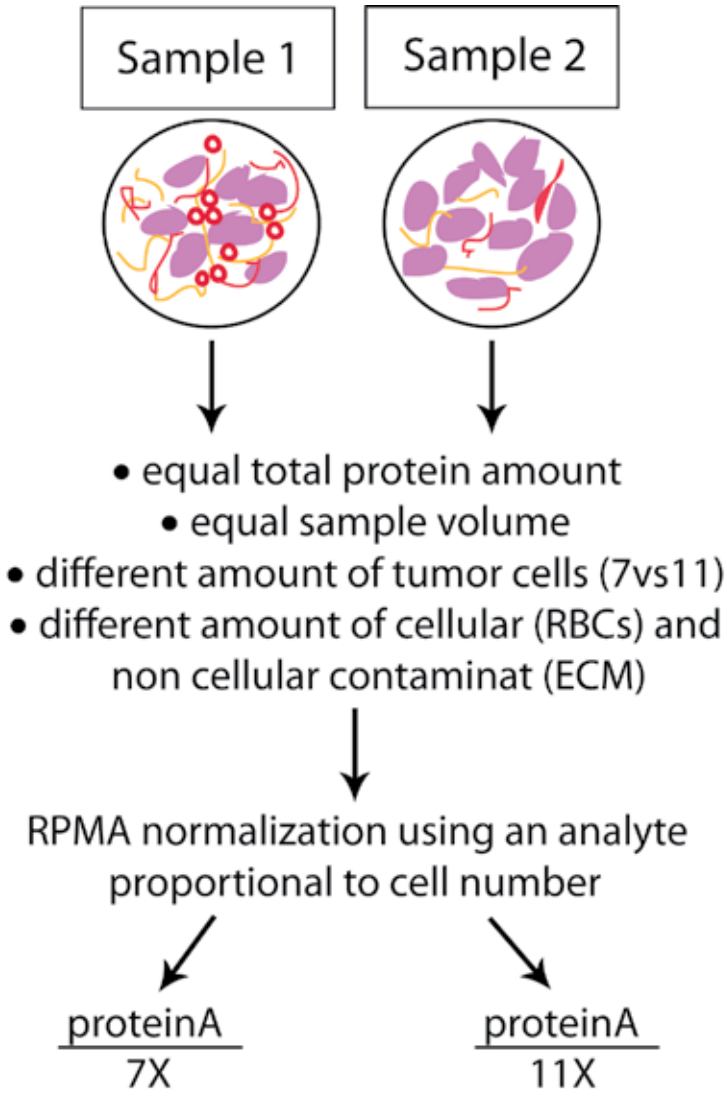
Primary antibodies were validated prior to use by immunoblotting with complex cellular lysates (commercial cell lysates or human tissue lysates). Criteria for antibody validation were *(i)* a single band at the correct molecular weight, or *(ii)* if two bands were present 80% of the signal

Supplementary Table 1. Summary of analyte stability in 8 different data sets per geNorm and NormFinder.

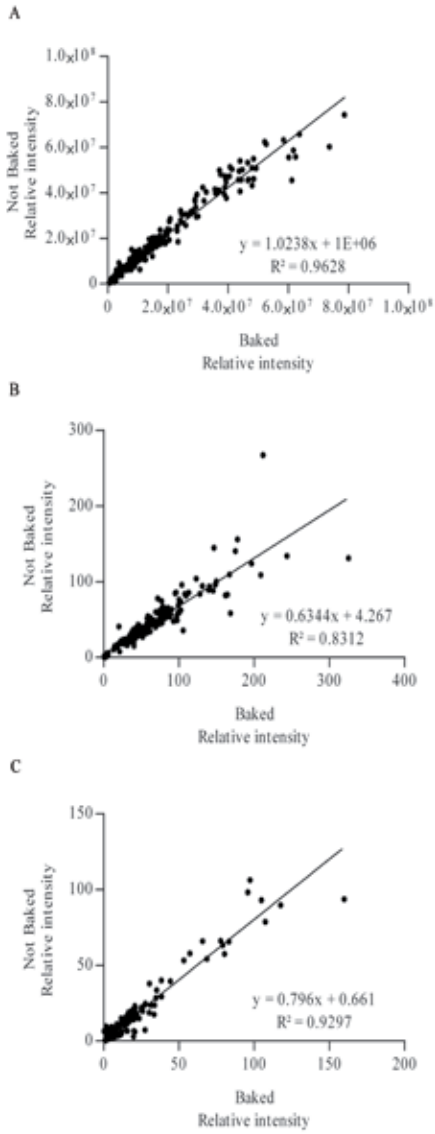
	Liver carcinoma		Bone metastasis		Melanoma		Carcinoma of the ovary		Lymph node metastasis		Renal carcinoma		Normal uterus		Normal muscle	
	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder	geNorm	NormFinder
1	TP	MRPL11 (a)	ssDNA	ssDNA	α/β-tubulin	α/β-tubulin	MRPL11	MRPL11	RPL13a	RPL13a	RPL13a	RPL13a	RPL13a	RPL13a	ssDNA	ssDNA
2	RPL13a	ssDNA (a)	MRPL11	TP	β-actin	β-actin	β-actin	α/β-tubulin	β-actin	β-actin	ssDNA	ssDNA	β-actin	β-actin	MRPL11	MRPL11
3	MRPL11	RPL13a (b)	β-actin	MRPL11	RPL13a	RPL13a	RPL13a	TP	TP	TP	α/β-tubulin	α/β-tubulin	α/β-tubulin	α/β-tubulin	RPL13a	RPL13a
4	ssDNA	TP (b)	RPL13a	RPL13a	MRPL11	GAPDH	α/β-tubulin	β-actin	α/β-tubulin	α/β-tubulin	MRPL11	MRPL11	ssDNA	MRPL11	TP	TP
5	β-actin	β-actin	TP	β-actin	GAPDH	MRPL11	TP	RPL13a	MRPL11	MRPL11	β-actin	β-actin	MRPL11	ssDNA	α/β-tubulin	α/β-tubulin
6	α/β-tubulin	α/β-tubulin	GAPDH	GAPDH	TP	TP	ssDNA	ssDNA	ssDNA	ssDNA	TP	TP	TP	TP	β-actin	β-actin
7	GAPDH	GAPDH	α/β-tubulin	α/β-tubulin	ssDNA	ssDNA	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH	GAPDH

Stability classification of the seven normalization analytes in each of the eight studied data sets. Analytes are listed from the most stable (top row) to the least stable (bottom row) as per geNorm and NormFinder algorithms.

(a) ranked equally most stable (b) ranked equally as the second most stable



Supplementary Figure S1. Cellular and noncellular contaminants bias RPMA normalization. The figure depicts an example of different amounts of contaminants in samples with equal amounts of total protein and equal volumes. The correct normalization, proportional to the number of cells in the sample, will eliminate bias due to contaminants.



Supplementary Figure S2. Antigenicity and immunostaining is unaffected by heating (baking) reverse phase protein microarrays after printing. The staining intensity of baked arrays (x axis) are plotted against unbaked arrays (y axis). Regression coefficients and curve slopes attest to the equivalency of antigenicity after baking for (A) total protein (R2=0.9628, slope=1.0238), (B) phospho-Akt Ser473 (R2=0.8312, slope=0.6344), and (C) β-actin (R2=0.9297, slope=0.796).

must have been at the correct molecular weight.

RPMA Analysis Suite

RPMA Analysis Suite (RAS) (http://capmm.gmu.edu/rpma-analysis-suite) was used for data reduction. Replicate staining intensity values were averaged and the CV between replicates was calculated. The replicate CV limit can be set by the user depending on the particular experimental conditions. Non-specific staining due to the secondary antibody was corrected by subtracting the spot intensity values for the array probed with antibody diluent and secondary antibody from the arrays stained with protein specific antibodies (primary antibodies) and secondary antibody. The corrected intensity values for each microarray spot were divided by the corresponding intensity value of the selected normalization factor. A single dilution was selected that was in the linear region of the antibody dose-response curve for all the samples. Normalized values for the selected dilution for each sample were finally used in the statistical analysis.

Standard operating procedure for RAS 1.10

1. Open the RAS macro template. Enable editing and enable macro content.
2. Save the RAS template with a unique name representing the project/date/user etc.
3. Setup tab: RPMA analysis details are entered here.

a) Enter the number of endpoints. This includes any potential normalizers (such as total protein concentration, cell count, amount of DNA, etc.) and negative controls. Changing this number and hitting enter will automatically adjust the data entry tabs correspondingly.

b) Enter the number of technical replicates (printed Spot Replicates).

c) Enter a value for the Background Cutoff, which is the number of SD above the substrate background signal intensity. The background cutoff is used to set the SD threshold based on the background values of the replicate spots. If the spot intensity is below × times the background SD, the

data value will be considered not above background and set to zero. Default value is 2 SD.

d) Enter a multiplier to convert all data from <1 to >1 for ease of analysis and review. The multiplier is used to increase the final values across the whole data set and make numbers easier to read. It is optional and "1" should be entered if no data multiplication is desired. Default value is 100.

e) Show intermediate results in individual tabs. If intermediate results (flag removal, filtering, averaging, CV calculations, negative control subtractions, etc.) need to be visible check this box. This will slow down the calculation. If speed is important leave the box unchecked. Default value is checked (enabled).

f) Low-value threshold. Default is zero. This value allows you to select a threshold above biological/technical "noise." The low-value threshold must be determined in advance by a manual review of the raw data values for the negative arrays and the corresponding endpoint arrays.

g) CV cutoff (above low-value threshold). Default value is 20%. This value is the allowable CV difference between technical replicates on the array for each of the primary end points. If a set of replicates has a CV greater than the specified value, a comment is returned rather than a numerical value, which indicates the replicates failed the established CV cutoff.

h) CV cutoff (below low-value threshold). Default value is 30, but the cutoff is ignored if the low value threshold is zero.

i) Filter negative control by CV. Default is checked (enabled).

j) CV cutoff (negative control). This value is the allowable CV difference between technical replicates on the negative control array (secondary antibody only array). If a set of replicates has a CV greater than the specified value, a comment is returned rather than a numerical value, which indicates the replicates failed the established CV cutoff.

4. Enter the raw spot intensity data in the "Enter_Raw_Data" tab. Use the copy, paste special, paste values function in Microsoft Excel. Enter the names of the endpoints in the appropriate cells for each column. Changing the name of endpoints or the sample information (Grid ID, Sample ID, Dilution) will be automatically reflected in all other tabs.

5. Enter any flagged spots in the "Enter_Flags" tab. Flagged spots will not be used in analysis. If nothing is entered, no spot will be flagged. Any value, in any cell, will flag that spot and remove it from downstream analysis. The sample information (Grid ID, Sample ID, Dilution) and endpoint names are entered automatically by the macro. Changing them here will have no effect downstream. If you need to change this information, please do so in the "Enter_Raw_Data" tab. It will automatically be adjusted in the "Enter_Flags" tab.

6. Match the appropriate negative controls for each endpoint in the "Enter_Neg" tab. Select the negative control for each endpoint by double-clicking on the respective cell or entering an "a" into the cell. Selections can be copied and pasted for faster entry. Please select only one negative control. If two are selected the macro will take the first one to subtract.

7. Match the appropriate normalizers for each endpoint in the "Enter_Norm" tab. Selection is done the same way as for the negative controls (see point 4). However, it is possible to select several normalizers per endpoint. In this case they will be combined into a geometric mean that is then used to normalize the respective end point.

8. Go back to the setup tab, and click on "normalize." This will calculate and display the normalized data in a separate tab as the last tab of the Excel template.

9. The "clear all" button will delete all calculated and raw data to allow reanalysis from scratch.

10. Copy the values from the "normalize" tab into a blank Excel worksheet for further bioinformatics analysis and statistical analysis.

Results and discussion

Normalization is a correction factor applied to reduce sample-to-sample variability because of variations in input amount and biochemical constituents. For example, imagine if one patient's lysate sample was derived from 4000 cells and the next patient's lysate sample of the same volume was derived from 2000 cells. In this example there will be a two-fold difference in the value of a test analyte (e.g., a cell receptor protein) between the two samples simply because of the difference in cellular content. If we divided the value of our test analyte in each sample by a reference term (such as DNA) that represented the cell content in the sample, then the normalized value would represent the concentration of the test analyte per cell in both samples. By this calculation, normalization is achieved.

New normalization analytes are proposed in this study: ssDNA and analytes selected among the constitutive cellular proteins, such as α/β-tubulin, GAPDH, MRPL11, and RPL13a. The expression level of total protein, β-actin, ssDNA, α/β-tubulin, MRPL11, RPL13a, and GAPDH was determined by RPMA in different tissue types: liver carcinoma, renal carcinoma, normal uterus, normal muscle, bone metastasis, melanoma, ovarian carcinoma, and lymph node metastasis. The stability of the normalization analytes was calculated using both the geNorm algorithm (2) and the NormFinder algorithm (3). The most and least stable analytes varied by tissue type. Based on the geNorm algorithm, RPL13a was the most stable analyte in lymph node metastasis, renal carcinoma, and normal uterus groups, while the most stable analyte was ssDNA in the normal muscle and bone metastasis group. Total protein was most stable in the liver carcinoma group, MRPL11 in the ovarian carcinoma group, and α/β-tubulin in the melanoma group (Supplementary Table 1). Running the same data by the NormFinder algorithm, results remained the same in every case but the liver carcinoma group, where the most stable analytes were MRPL11 and ssDNA (with identical scores) while total protein was the second-best normalization analyte. The discrepancy between geNorm and NormFinder in the liver samples could potentially be explained by the difference in the two algorithms. geNorm selects an optimal pair of reference analytes from the set of input analytes. geNorm assumes that the candidate molecules are not co-regulated. On the other hand, NormFinder ranks a set of candidate normalization analytes based on their expression stability (variance) in a sample set. Liver tissue is highly vascular and metabolically active. The presence of blood in the liver tissue could contribute to variation within the samples, thus allowing ssDNA to be selected by NormFinder as the most stable analyte, as we showed for bone metastasis samples contaminated with blood.

References

1. **Espina, V., K.H. Edmiston, M. Heiby, M. Pierobon, M. Sciro, B. Merritt, S. Banks, J. Deng, et al.** 2008. A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. Mol. Cell. Proteomics *7*:1998-2018.
2. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3:RESEARCH0034.
3. **Andersen, C.L., J.L. Jensen, and T.F. Ørntoft.** 2004. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. *64*:5245-5250.