

ProMAT:
Protein Microarray Analysis Tool
User's Manual

Amanda M. White
Don S. Daly
Richard C. Zangar
Susan M. Varnum
Kevin K. Anderson

Pacific Northwest National Laboratory
PO Box 999
Richland, WA 99354

For information contact:
Amanda M. White (amanda.white@pnl.gov)

Contents

1 Introduction	2
2 Installing the ProMAT Tools	4
2.1 Windows	4
2.2 Macintosh	4
3 Preparing Data for Analysis	5
3.1 Array Layout	5
3.2 Experiment Information	6
4 ProMAT	9
4.1 Running ProMAT	9
4.2 ProMAT Output	11
4.3 Example	12
5 ProMAT Calibrator	14
5.1 Running ProMAT Calibrator	14
5.2 ProMAT Calibrator Output	15
6 References	17

1 Introduction

The Protein Microarray Analysis Tool (ProMAT) is an R-based statistical application for analyzing quantitative protein microarray data. The software was developed by the Statistics Group and the Cell Biology and Biochemistry Group at Pacific Northwest National Laboratory (<http://www.pnl.gov>).

The ProMAT package consists of two pieces: the original ProMAT tool which fits standard curves and estimates protein concentrations and the new ProMAT Calibrator tool which estimates and adjusts for processing and measurement effects between replicate arrays and replicate spots within arrays. The use of both tools is covered in this user guide.

ProMAT takes as input microarray data (the output of microarray image analysis) where some arrays are treated with standards of known antigen concentration and some arrays are treated with samples of unknown concentration. ProMAT fits standard curves to the standards' data to relate spot fluorescence to concentration. ProMAT also calculates and outputs confidence bounds on these standard curves. The tool then uses the standard curves to predict antigen concentrations for the unknown samples, along with prediction intervals. A description of the statistical methods implemented in this tool can be found in [1-3].

A recent addition to ProMAT is the use of multiple imager settings for imaging arrays. If arrays are scanned at multiple imager settings (e.g., laser power, PMT, etc) then a separate standard curve is created for each imager setting, and a separate concentration prediction is estimated for each spot at each imager setting. Then the estimates from each spot are averaged by weighting according to the uncertainty of each estimate.

ProMAT produces diagnostic plots to aid the user in determining where there may be problems with the data. Figure 1 shows one such diagnostic plot. The lower right panel shows an example standard curve (in black) with its prediction intervals (in blue) plotted with the standard data. The gray region is the region of the curve where concentration predictions are the "best." The histogram in the lower left panel shows the sample spot values. The upper panel gives the coefficient of variation of the standard curve. The plot quickly shows the user how well the standard curve fits the data, how wide the prediction intervals are, and how well the range of the sample data matches up to the range of the standard curve. This type of plot is produced for each antigen and is displayed in an HTML interface for easy viewing.

ProMAT Calibrator takes data from ELISA microarrays and uses replicate data both within and between arrays to estimate effects due to slide and reagent processing. ProMAT Calibrator then adjusts for these effects and produces a dataset of normalized spot intensities which may be used for data analysis. This tool also produces diagnostics showing comparing the original spot intensities to the normalized values and an HTML file showing the user inputs and results for easy browsing. The results generated by ProMAT Calibrator are output in a way that is compatible with ProMAT so that standard curves may be generated without further data manipulation. Further details on the algorithms implemented in this tool can be found in [4].

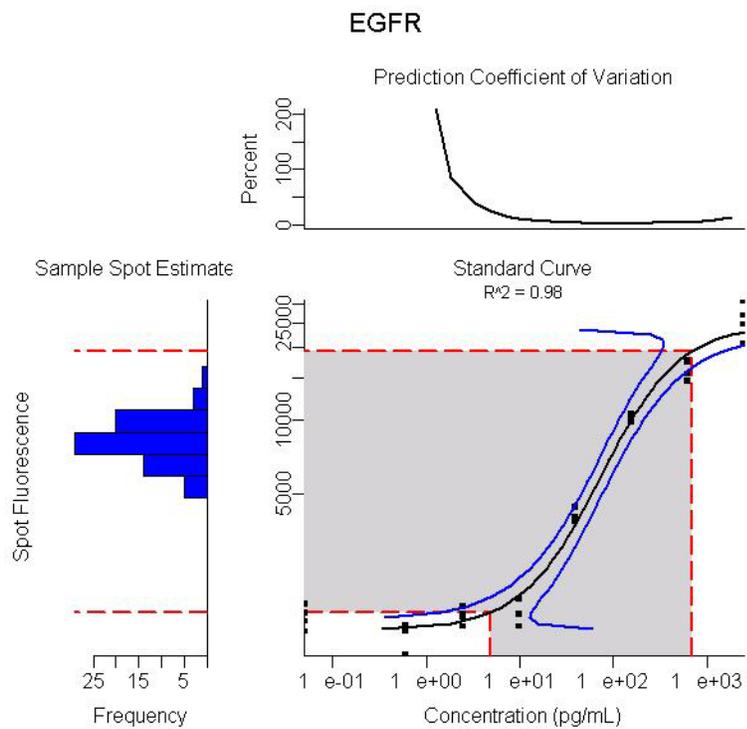


Figure 1: Example of prediction curve (black curve) with prediction bounds (blue). The histogram on the left shows the sample spot values for this antigen. The red bounds represent the region of best prediction on the curve; ideally sample spots should lie within these bounds. The curve at the top shows the percent coefficient of variation for the prediction curve.

2 Installing the ProMAT Tools

2.1 Windows

[Java version 5.0](#) or newer and [R version 2.7.0](#) or newer are required for ProMAT. These should be installed before installing ProMAT. The installer should be run by a system administrator, and on Windows 7 be sure to right click the installer and choose *Run as administrator*, even if the user is an administrator. (This is due to the way Windows 7 handles administrator privileges.) The installer provides the option to install ProMAT, ProMAT Calibrator or both. By default, files will be placed in the *C:\Program Files\ProMAT* directory; however the installer provides the option to choose another directory. (In this document, *ProMAT_ROOT* will refer to the *C:\Program Files\ProMAT* directory or installation directory chosen by the user.) By default, the installer will add shortcuts to your Start Menu and desktop. An uninstaller will also be created in *ProMAT_ROOT* and may be accessed via the Start Menu.

During the installation, the user will be prompted to choose a home directory for ProMAT. This is the directory from which you will always begin browsing for data files, so if all data files are organized under some directory, say *c:\experiment_data*, then choose this as the home directory.

2.2 Macintosh

ProMAT for Mac is distributed as a compressed archive containing the necessary Java and R files. Uncompress the archive and save anywhere on your hard drive (for the remainder of this document we will refer to the ProMAT directory of the uncompressed archive as *ProMAT_ROOT*). The user must separately download [Java version 5.0](#) or newer and [R version 2.7.0](#) or newer.

After unzipping the archive, open a Terminal window and type

```
cd [ProMAT_ROOT]
bash install_libraries.sh
```

This will install additional R libraries which are needed for ProMAT and ProMAT Calibrator (Matrix and lme4).

Run ProMAT by opening a Terminal windows and typing

```
cd [ProMAT_ROOT]
bash ProMAT.sh &
```

Use the same process to run ProMAT Calibrator, but substitute *ProMAT_Calibrator.sh* for *ProMAT.sh*.

3 Preparing Data for Analysis

ProMAT and the ProMAT Calibrator both import data from comma-delimited (.csv) text files, including ScanArray files, and GenePix (.gpr) files. (Data summary files are also generated in .csv files. These .csv files are compatible with Excel and most other spreadsheet programs.) In addition to the data files, both tools expect two additional files containing information about the array design (i.e., spot characteristics) and the array incubation and processing procedures. Therefore, 4 types of files are needed. These files are described in detail below. A summary of these files is provided here.

1. Data file: File of spot intensity values that was created from a scanned microarray image. These files can be generated by a variety of image analysis programs but must be in a .csv format for ProMAT to read.
2. Array Layout file: Defines print pattern on individual chip and the maximum concentration of each antigen standard.
3. Standards experiment information file: Defines which slides contain standards, what position (i.e., which array on the slide) has standards, and what the dilution of the standard mixture is on that array compared to the maximum value entered in the Array Layout file.
4. Samples experiment information file: Defines which slides contain samples, what samples are located at what position (i.e., which array on the slide), and what the dilution of the sample is compared to original, undiluted sample.

Make sure that none of the following characters:

, \ / : * ? “ < > | #

are used in file names or in any user-defined fields such as antigen name or slide ID in the array layout and experiment info files as this will cause an error.

3.1 Array Layout

ProMAT is designed to process data from slides with multiple identical arrays on each slide. An example of a typical slide layout is shown in Figure 2.

ProMAT needs information about the array layout to map the antigens to the correct spot intensities in the data files. This information is stored in a comma-delimited text file typically named *array_layout.csv*, although the file can be given any name. Figure 3 shows an example of an array layout file. The necessary elements of this file are:

1. Column (*spot.name* in Fig. 3) giving the antigen name or spot name. This column can be given any name and the user will enter the column name in the ProMAT window in the “Spot ID Column” field. ProMAT will create standard curves by combining data with the same antigen or spot name.
2. The maximum concentration for the standard data for this antigen (no units). **This column must be called *max.concentration*.** ProMAT assumes that multiple antigens (standards) are premixed, with varying concentrations of the individual antigens, and subsequently diluted to generate data for the standard curves. The maximum concentration column gives the antigen concentration in the standard mixture before

dilution. The relative dilution of the standard mixture or sample used for each array is entered into the experimental info file (see below).

Glass slide with 16 identical arrays

1,1	1,2
2,1	2,2
3,1	3,2
4,1	4,2
5,1	5,2
6,1	6,2
7,1	7,2
8,1	8,2

Single array with 4 identical quadrants

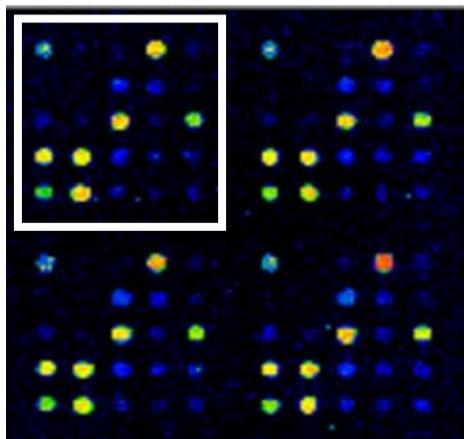


Figure 2: Example of slide with 16 identical arrays identified based on row and column positioning. Each array contains 25 spots printed once in each quadrant. Only the information for one array must be entered into the ProMAT program in the “array_layout.csv” file since they are all identical. The program will automatically replicate this information for other quadrants and other arrays.

- Columns necessary to match up the antigens to the lines of the data file. For example, if your data files have columns “Spot Row” and “Spot Column”, and these are the columns that uniquely determine the position of a specific capture antibody, then these column names should appear in your array layout file and with **the exact same name as in the data files**. You may have as few or as many columns are necessary to specify the antigen to data mapping. For example, if your arrays have the same reagent printed multiple times across a complete row, you only need a column for “Spot Row”. In this case, ProMAT will automatically assume that all columns of spots contain the identical pattern of reagents.

Any spot intensity data found in the data files that is not matched to a line in the array layout will be ignored. So if you want to run the data analysis on only a subset of the antigens, only enter those in the array layout file. For an example of an array layout file, see

[ProMAT_ROOT]\example\standards\array_layout.csv

3.2 Experiment Information

ProMAT also needs information about the incubation and processing of each array. This information is typically stored in a CSV file called *experiment_info.csv* (although it may be given any name). Separate experiment information files must be prepared for the samples (Fig. 4) and the standards (Fig. 5). These files can be given different names or, if stored in different directories, the same name. The sample file contains the information about the location (i.e., slide and array) and the dilution of individual samples. The standards file contains information

on the relative concentration of the antigen mix and the location. The experiment info files must contain the following columns:

	A	B	C	D	I
1	Spot Row	Spot Column	spot_name	max.concentration	
2	1	1	Alexa546-Ab	1	
3	1	2	PBS-2	1	
4	1	3	VEGF	2000	
5	1	4	HGF	2000	
6	1	5	bFGF	1000	
7	2	1	PBS-6	1	
8	2	2	PBS-7	1	
9	2	3	CA15-3	1000	
10	2	4	FAS-ligand	1000	
11	2	5	CEA	4000	
12	3	1	PSA	500	
13	3	2	HER-2	8000	
14	3	3	TNF-alpha	1000	
15	3	4	E-selectin	1000	
16	3	5	MMP 2	6000	
17	4	1	MMP 9	6000	
18	4	2	RANTES	2000	
19	4	3	sICAM-1	60000	
20	4	4	uPAR	10000	
21	4	5	IGF-I	2000	
22	5	1	TGF-alpha	500	
23	5	2	PDGF-AA	2000	
24	5	3	PDGF-BB	500	
25	5	4	MMP 1	2000	
26	5	5	biot-alpha-WF	1	
27					

Figure 3: Example array layout file. Array layout files can be created in Excel. To save as a CSV file, choose *File*→*Save As* and choose CSV in the *Save As Type* box.

1. Data file name. **This column must be labeled “*file.name*”.**
2. A unique slide ID (either numeric or text). A slide ID may correspond to a single file or may be associated with multiple files, if multiple images per file were taken. **This column must be named “*slide.number*”.**
3. A sample ID (either numeric or character). For standards this may be simply “standard” for all rows. This name should be unique for each sample. **This column must be called “*sample.id*”.**
4. The dilution factor. For the standard data file, this is the dilution of the standard mixture relative to the maximum concentration value entered into the array layout file (Fig. 3). For sample data, this is the dilution factor for the sample prior to incubation. This should be a number less than or equal to 1 (i.e. 0.25 is a four-fold dilution). **This column must be called “*dilution*”.**
5. If one data file contains data for multiple samples or multiple standard dilutions, columns must be included that will allow these data to be uniquely matched to the data files. For example, if each array was treated with a different dilution of the standard mixture, there should be two columns with the headings such as “*Array Row*” and “*Array Column*” (or whatever are the labels in the data files) where each array will have a different value in the “*dilution*” column. **These additional columns must be labeled exactly as they appear in the data files.**
6. If multiple scans were taken of each slide (e.g. with different imager settings) additional columns may be added to hold that information. These columns may be given any name, and the user should enter those column names in the ProMAT window in the “Imager settings column names” field so that a separate standard curve may be made for each type

of scan. If there are multiple columns of scanner settings, then enter all column headings separated by a comma (e.g., laser_power,PMT).

	A	B	C	D	E	F
1	file.name	slide number	sample.id	Array Row	Array Column	dilution
2	060803 #2_PL 70.csv	2	0a 0.9X	1	3	0.9
3	060803 #2_PL 70.csv	2	0a 0.9X	1	4	0.9
4	060803 #2_PL 70.csv	2	5a 0.9X	2	3	0.9
5	060803 #2_PL 70.csv	2	5a 0.9X	2	4	0.9
6	060803 #2_PL 70.csv	2	10a 0.9X	3	3	0.9
7	060803 #2_PL 70.csv	2	10a 0.9X	3	4	0.9
8	060803 #2_PL 70.csv	2	50a 0.9X	4	3	0.9
9	060803 #2_PL 70.csv	2	50a 0.9X	4	4	0.9
10	060803 #2_PL 70.csv	2	100a 0.9X	5	3	0.9
11	060803 #2_PL 70.csv	2	100a 0.9X	5	4	0.9
12	060803 #2_PL 70.csv	2	150a 0.9X	6	3	0.9
13	060803 #2_PL 70.csv	2	150a 0.9X	6	4	0.9
14	060803 #2_PL 70.csv	2	Media 0.9X	7	3	0.9
15	060803 #2_PL 70.csv	2	Media 0.9X	7	4	0.9
16	060803 #2_PL 70.csv	2	50b 0.9X	8	3	0.9
17	060803 #2_PL 70.csv	2	50b 0.9X	8	4	0.9
18	060803 #3_PL 70.csv	3	50b 0.9X	1	1	0.9
19	060803 #3_PL 70.csv	3	50b 0.9X	1	2	0.9
20	060803 #3_PL 70.csv	3	Media 0.9X	2	1	0.9
21	060803 #3_PL 70.csv	3	Media 0.9X	2	2	0.9
22	060803 #3_PL 70.csv	3	150a 0.9X	3	1	0.9

Figure 4: Example experiment info file where a sample is associated with a single array, thus the columns “Array Row” and “Array Column” are included to match up the sample IDs to the correct data.

The experiment info files determine which data files are loaded for analysis, so any data files not mentioned in either the standards or the samples experiment info will not be used. Likewise any data in the files not matched up (e.g., an array not specified in the experiment info) will not be used. For an example experiment information file see

[ProMAT_ROOT]\example\standards\experiment_info.csv

	A	B	C	D	E
1	file.name	sample.id	Array Row	Array Column	dilution
2	060803 #2_PL 70.csv	standard	1	1	1
3	060803 #2_PL 70.csv	standard	1	2	1
4	060803 #2_PL 70.csv	standard	2	1	0.333333333
5	060803 #2_PL 70.csv	standard	2	2	0.333333333
6	060803 #2_PL 70.csv	standard	3	1	0.111111111
7	060803 #2_PL 70.csv	standard	3	2	0.111111111
8	060803 #2_PL 70.csv	standard	4	1	0.037037037
9	060803 #2_PL 70.csv	standard	4	2	0.037037037
10	060803 #2_PL 70.csv	standard	5	1	0.012345679
11	060803 #2_PL 70.csv	standard	5	2	0.012345679
12	060803 #2_PL 70.csv	standard	6	1	0.004115226
13	060803 #2_PL 70.csv	standard	6	2	0.004115226
14	060803 #2_PL 70.csv	standard	7	1	0.001371742
15	060803 #2_PL 70.csv	standard	7	2	0.001371742
16	060803 #2_PL 70.csv	standard	8	1	0.0002
17	060803 #2_PL 70.csv	standard	8	2	0.0002
18	Copy of 060803 #2_PL 70.csv	standard	8	1	0
19	Copy of 060803 #2_PL 70.csv	standard	8	2	0
20					

Figure 5: Example experiment info file where a standard dilution is associated with a single array.

4 ProMAT

4.1 Running ProMAT

To begin the analysis, double-click the *ProMAT* icon on your desktop or Start Menu. A window (Figure 6) will appear asking the user to provide the following information.

Figure 6: ProMAT user interface

The screenshot shows the ProMAT user interface window. It is divided into several sections:

- Data:** Fields for Analysis name, Standards directory, Standards slide layout (array_layout.csv), Standards experiment info (experiment_info.csv), Samples directory, Samples slide layout (array_layout.csv), Samples experiment info (experiment_info.csv), Output directory, Spot intensity column, and Spot ID column. Each field has a corresponding 'Browse' button.
- Options:** Two dropdown menus for 'Transform to apply to spot values' and 'Transform to apply to concentration values', both set to 'Natural Log: f(x) = ln(1+x)'. A text field for 'Imager settings column names'. Checkboxes for 'Create standard curves?' and 'Identify sample outliers?'.
- Curve types:** Radio buttons for 'Logistic with Spline backup', 'Logistic curve', 'Power curve', 'Linear curve', and 'Spline curve (Monte Carlo bounds)'. 'Logistic with Spline backup' is selected.
- Method for calculating bounds:** Radio buttons for 'Analytic bounds', 'Monte Carlo bounds', and 'No bounds'. 'Analytic bounds' is selected.
- Plotting Options:** Fields for 'x-axis label for plotting' (Concentration (pg/mL)) and 'y-axis label for plotting' (Spot Fluorescence). Checkboxes for 'Keep y-axis range constant in plots?', 'Show standard data on plots?', 'Show replicate means on plots?', 'Show replicate standard deviations on plots?', 'Show standard curves in plots?', and 'Show upper and lower bounds in plots?'.

At the bottom of the window are 'Run' and 'Exit' buttons.

Data:

- **Analysis name:** the analysis name will be used to name all files created by ProMAT.
- **Standards directory:** the directory containing the standards data files
- **Standards slide layout:** array layout file for the standards
- **Standards experiment info:** experiment info file for the standards
- **Samples directory:** the directory containing the samples data files (optional: if no samples directory is provided then only standard curves will be created)
- **Samples slide layout:** array layout file for the samples
- **Samples experiment info:** experiment info file for the samples

- **Outputs directory:** the directory in which to write outputs. Since any existing files will be overwritten, a new directory needs to be created for each analysis.
- **Spot intensity column:** the *exact* column name in the data files that contains the spot intensity values to be used for curve fitting and sample concentration predictions.
- **Spot ID column:** column name of unique spot ID (e.g. *spot.name* or other unique identifier)

Curve Types:

Choose one or more of the following standard curve types to fit. ProMAT will fit each type of curve you select to each antigen and then choose the curve that best fits the data.

- **Logistic with Spline backup:** four parameter logistic model. If and only if the logistic curve fit does not converge, it returns a spline curve. (If this option is chosen in addition to other curve types and a logistic curve cannot be fit, then the spline curve will be compared to the other options.)
- **Logistic curve**
- **Power curve**
- **Linear curve**
- **Spline curve** (this requires Monte Carlo bounds)

Method for calculating bounds:

- **Analytic bounds:** use propagation of error to calculate curve bounds and prediction intervals. For spline curves Monte Carlo bounds are always created unless “**No bounds**” is chosen. Analytic bounds are much quicker to calculate, however, due to the fact that a polynomial approximation of the curve is used to estimate the bounds, the intervals diverge near curve asymptotes (e.g., the upper and lower asymptote of a four-parameter logistic curve).
- **Monte Carlo bounds:** use Monte Carlo simulation to estimate curve bounds and prediction intervals. Monte Carlo bounds give much more realistic intervals near curve asymptotes, but take more time to compute because many simulations are required for an accurate estimate.
- **No bounds:** do not calculate curve bounds or prediction intervals.

Options:

- **Transform to apply to spot values:** choice of Identity or Natural Log. In most cases the Natural Log transform should be chosen, unless the data has been previously log transformed. (See [1] for a discussion of why the data is log transformed prior to standard curve estimation and prediction of concentrations.)
- **Transform to apply to concentration values:** see previous.
- **Imager settings column names:** if arrays are imaged at multiple laser settings, list the column name(s) indicating the laser settings here. If not, leave blank. Use commas (with or without spaces) between column headings to delineate multiple columns of input.
- **Create standard curves:** to simply plot the data without fitting standard curves, uncheck this box (default is checked). If unchecked, ProMAT will simply plot the data without attempting to fit standard curves or make concentration predictions.

- **Identify sample outliers:** to identify and exclude outliers in the sample data, check this box (default is checked). To identify outliers we calculate the spot replicate variance (one value for each curve) then for each set of replicate spots, values that are more than three standard deviations from the median value are determined to be outliers.

Plotting Options:

- **x-axis label for plotting:** label that appears on the x (concentration) axis of plots. This value will define the units for the values that were entered in the “max.concentration” column of the array layout file (Fig. 2).
- **y-axis label for plotting:** label that appears on the y (spot intensity) axis of plots.
- **Keep y-axis range constant in plots:** to keep the y-axis range constant in all plots for ease of comparison, check this box (default is unchecked).
- **Show standard data on plots:** show standard data points in the plots (black points)
- **Show replicate means on plots:** show the means of replicate values (cyan points)
- **Show replicate standard deviations on plots:** show the one standard deviation above and below the replicate means
- **Show standard curves in plots:** show the fitted curves in the plots
- **Show upper and lower bounds in plots:** show the upper and lower confidence intervals in the plots

Once all required parameters are entered, click *Run* to start the analysis. ProMAT will first read the standard data and fit the standard curves, then it will read the sample data and predict the sample concentrations. If no directory of samples is provided, the program will create and export standard curves and then stop. When analysis is complete a message will appear.

If ProMAT fails to successfully complete, a pop-up window to open the log file will automatically appear if running on Windows. On Mac, a log file may be found in the results directory.

4.2 ProMAT Output

For a full analysis (standard curves estimation *and* concentration prediction) two types of diagnostic plots will be created. The first plot is a standard curve with confidence bounds on curve fitting. This gives the user an idea of the variability in standard data and how well the estimated curves fit the data. The second type of diagnostic plot consists of the standard curves displayed with prediction intervals, which give the upper and lower confidence bounds on the predicted concentrations. These plots also include a histogram of the sample data, which shows if the range of the standard curve encompasses to the sample data. Finally, the plot shows the percent coefficient of variation of the predictions, which is calculated from the distance between the concentration prediction and the prediction bounds divided by the prediction value. If the data is log transformed then this value is calculated on the log scale. An HTML file will also be created in the outputs directory to allow the user to easily browse through the second type of diagnostic plots for each antigen.

In addition to the diagnostic plots, four data tables will be created in the outputs directory: *predicted_concentrations.csv*, *standard_data.csv*, and *standard_curve_statistics.csv*.

- ***predicted_concentrations.csv***: contains predicted concentrations and confidence bounds for each sample observation, where the replicates have been pooled prior to prediction. This file includes the spot and sample IDs and dilution plus the number of spots determined to be outliers which were excluded from the concentration prediction.
- ***standard_data.csv***: contains the data used to fit the standard curves, and is provided for user reference. This file includes all the columns found in the array layout and experiment info files (e.g., sample.id, dilution, max.concentration) plus the data file names, spot intensities and the calculated actual concentration which is $\text{max.concentration} \times \text{dilution}$.
- ***standard_curve_statistics.csv***: contains information about each standard curve including equation and goodness-of-fit statistics such as R^2 and mean-squared error. See the included file “[ProMAT_ROOT]\Standard Curve Statistics.pdf” for a description of the data in this file.
- ***spot_intensity_variances.csv***: contains estimates of the variance of spot replicates for both the standards and samples data for each curve. If you choose to log transform your data prior to curve fitting, then these variances are on the log scale.

4.3 Example

In the [ProMAT_ROOT]\example directory is a set of data files consisting of both standard and sample data. The experiment info and array layout files have already been created.

Each data file contains spot intensity data for eight rows and four columns of arrays, denoted by the “Array Row” and “Array Column” headers. Each row of arrays contains four arrays of two different types (see Figure 7). Since each row of arrays is the same, the array layout file need only contain information for one row of arrays, so the array layout file contains “Array Column”, “Spot Row”, and “Spot Column” columns. This information will be matched to each row of data in the data files that contains the matching “Array Column”, “Spot Row” and “Spot Column” values. Thus, only the minimum amount of non-replicated information must be specified in the array layout file. Note also, that the array layout file is identical for the standards data and the samples data. This file also lists the maximum concentration of each antigen in the standard mixture. This information is needed to transform concentration predictions back to the original units. If this column is uniformly 1, then the predicted concentrations will be as a fraction of the standard mixture concentration of the corresponding protein.

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4
5,1	5,2	5,3	5,4
6,1	6,2	6,3	6,4
7,1	7,2	7,3	7,4
8,1	8,2	8,3	8,4

Figure 7: Slide layout for example with 8 rows and 4 columns of arrays, where array columns 1 and 3 are of one array layout and columns 2 and 4 are of another array layout.

There are two experiment info files: one for the standards and one for the samples. Each experiment info file lists the file name along with slide and sample identifiers and the dilution of the sample or standard prior to slide treatment. In this example, array is treated with a different standard or sample, so the experiment info files have columns for “Array Row” and “Array Column” to match the data to the appropriate standard or sample.

To run the example in ProMAT, enter the following values in the ProMAT window:

Field	Value
Standards directory	c:\Program Files\ProMAT\example
Standards slide layout	c:\Program Files\ProMAT\example\array_layout.csv
Standards experiment info	c:\Program Files\ProMAT\example\experiment_info_standards.csv
Samples directory	c:\Program Files\ProMAT\example
Samples slide layout	c:\Program Files\ProMAT\example\array_layout.csv
Samples experiment info	c:\Program Files\ProMAT\example\experiment_info_samples.csv
Spot intensity column	Ch1 Median
Spot ID column	spot.name
Transform to apply to spot values	Natural Log
Transform to apply to concentration values	Natural Log

Choose one or more curve types and any bounds method, and then click *Run* to start the analysis.

5 ProMAT Calibrator

5.1 Running ProMAT Calibrator

To begin the analysis, double-click the *ProMAT Calibrator* icon on your desktop or Start Menu. A window (Figure 8) will appear asking the user to provide the following information.

Data:

- **Analysis name:** the analysis name will be used to name all files created by ProMAT Calibrator
- **Standards directory:** the directory containing the standards data files
- **Standards slide layout:** array layout file for the standards
- **Standards experiment info:** experiment info file for the standards
- **Samples directory:** the directory containing the samples data files (optional: if no samples directory is provided then only standard curves will be created)
- **Samples slide layout:** array layout file for the samples
- **Samples experiment info:** experiment info file for the samples
- **Outputs directory:** the directory in which to write outputs. Since any existing files will be overwritten, a new directory needs to be created for each analysis.

Figure 8: ProMAT Calibrator user interface

The screenshot shows the ProMAT Calibrator user interface window. The window title is "ProMAT Calibrator". The interface is organized into two main sections: "Data:" and "Options:".

Data:

- Analysis name: [Text Input Field]
- Standards directory: [Text Input Field] [Browse]
- Standards slide layout: [Text Input Field] [Browse]
- Standards experiment info: [Text Input Field] [Browse]
- Samples directory: [Text Input Field] [Browse]
- Samples slide layout: [Text Input Field] [Browse]
- Samples experiment info: [Text Input Field] [Browse]
- Output directory: [Text Input Field] [Browse]
- Spot intensity column: [Text Input Field]
- Spot ID column: [Text Input Field]

Options:

- Transform to apply to spot values: [Dropdown Menu] (Selected: Natural Log: $f(x) = \ln(1+x)$)
- Imager settings column names: [Text Input Field]

At the bottom of the window, there are two buttons: "Run" and "Exit".

- **Spot intensity column:** the *exact* column name in the data files that contains the spot intensity values to be used for curve fitting and sample concentration predictions.
- **Spot ID column:** column name of unique spot ID (e.g. *spot.name* or other unique identifier)

Options:

- **Transform to apply to spot values:** choice of Identity or Natural Log. In most cases the Natural Log transform should be chosen, unless the data has been previously log transformed. (See [1] for a discussion of why the data is log transformed prior to analysis.)
- **Imager settings column names:** if arrays are imaged at multiple laser settings, list the column name(s) indicating the laser settings here. If not, leave blank. Use commas (with or without spaces) between column headings to delineate multiple columns of input.

Once all required parameters are entered, click *Run* to start the analysis. ProMAT Calibrator will first read the data and present a list of antigens to use for calibration. Choose one or more antigens (hold the Ctrl key to select multiple values) and then click OK to continue. It is expected that the antigens chosen for calibration will not vary due to treatment. This assumption is used to estimate and correct for the effects due to processing and measurement.

When analysis is complete a message will appear asking if the user wants to view the results now. If Yes is selected, an HTML page will be displayed in the default browser. This HTML page is saved in the output directory. If ProMAT Calibrator fails to finish the analysis successfully, a pop-up window will display the error message. For more information on the problem consult the log file, “*normalization.log*” in the output directory.

5.2 ProMAT Calibrator Output

ProMAT Calibrator creates several data tables and diagnostic images in the output directory. The primary data of interest is in the file “*adjusted_values.csv*” which contains the normalized values for all input data. This data table contains the metadata from the array layout and experiment info files, the original spot intensities and the spot intensities adjusted according to two models: a diagnostic model and a calibration model. The diagnostic model adjusts for chip-level effects using within-sample data comparisons, and is good for identifying systematic patterns in the data. The calibration model adjusts for chip-level effects by estimating bias and noise across all samples using one or more calibrant spots, which are selected because they are not expected to vary between samples. See [4] for further description of the calibrant spots and these models. The adjusted values are contained in the columns “*Adjusted.Intensity.Diagnostic.Model*” and “*Adjusted.Intensity.Calibration.Model*”. ProMAT Calibrator also copies the original experiment info and array layout files to the output directory, adjusting the experiment info to refer to the “[*analysis_name*]*_adjusted_values.csv*” file. This allows the user to immediately run ProMAT on the results. The output directory will also contain a file called “[*analysis_name*]*_variances.csv*” which contains the within array and between array variances (on the log scale if applicable) for each antigen/imager setting pair.

ProMAT Calibrator creates plots of the array means for the original values and adjusted values so that the effects of the calibration can be seen. These are saved in files called

“*[analysis_name]_array_means_*.jpg*”. The other diagnostic image type is a plot of spot intensity versus print order for each of the calibration spots for the original values and adjusted values. This allows the user to see trends in the calibrant data that may be interfering with the calibration. These images are named “*[analysis_name]_calibrant_*.jpg*”.

Finally, an HTML file is created in the output directory which lists the user input, the within and between array variances and shows all diagnostic images. This is created to provide an easy way to review the results. The HTML file is called “*[analysis_name]_results.html*”.

6 References

1. Daly DS, AM White, SM Varnum, KK Anderson, and RC Zangar. "Evaluating concentration estimation errors in ELISA microarray experiments." *BMC Bioinformatics* 2005, 6:17. <http://www.biomedcentral.com/1471-2105/6/17>.
2. White AM, DS Daly, SM Varnum, KK Anderson, N Bollinger, and RC Zangar. 2006. "ProMAT: Protein Microarray Analysis Tool." *Bioinformatics* 22(10):1278-1279.
3. Daly DS, KK Anderson, AM White, RM Gonzalez, SM Varnum, and RC Zangar. 2008. "Predicting protein concentrations with ELISA microarray assays, monotonic splines and Monte Carlo simulation." *Statistical Applications in Genetics and Molecular Biology* 7(1): Article 21. <http://www.bepress.com/sagmb/vol7/iss1/art21/>.
4. Zangar, RC, DS Daly, AM White, SS Servoss, RM Tan, and JR Collett. "ProMAT Calibrator: a tool for normalizing data across antibody microarrays." Submitted to *Journal of Proteome Research*.