

User's Guide to Fingerprint Assignment (Main Processing Tool)

In this step, the background subtracted data is used to generate relative values (a ratio of each probe / reference probe) and then fingerprints are assigned based on the binding of each probe to a set of control clones.

Special case: multiple reference probes

If you have multiple reference probes, each associated with a different series of differential probes, you should analyze each subset of data (i.e. a reference probe plus its associated differential probes) separately in separate files at first, then combine the data for all probes in an additional file toward the end. You may combine your data after running the ControlAvg macro for each subset, or after generating fingerprints for each subset. In order to combine your data you must ensure that your clones are listed in the same order in each file.

To assign fingerprints for a set of probes:

Open the **Main Processing Tool** file (filename = WholeProcess). Save it immediately under another file name, and periodically save the file under this new name as you work on it. Next you will have to enter or delete data on various sheets of this file, then run the relevant macros to assign fingerprints.

Notes, Oldstuff and Readme sheets:

Ignore the “Notes” and “OldStuff” sheets, but the “Readme” sheet is relevant.

Preliminaries sheet:

Most of what is on this page can be ignored since the programs no longer use any data except the following:

- The number of control clones: this should ordinarily be 384.
- The number of probes to be analyzed, excluding the reference probe.
- The cost function estimator, or “rho” value.
 - o This is used to help set the boundaries of N assignments when the Reduced_Unequal_Variance macro is used later on. As the value of rho increases, the probability of making a 0/1 mismatch in fingerprints decreases, but the probability of assigning Ns increases. If there are too many Ns, there is not enough information in fingerprints to cluster them.
 - o We currently recommend a rho value of 3.4 for fungal OFRG probe sets and 3.0 for bacterial OFRG probe sets, but the optimal rho may vary between sets of hybridizations, depending on your results. The most accurate fingerprints will be those with the highest possible rho that have an acceptable number of fingerprints. You may wish to figure out the percentage of Ns various values of rho will produce for your fingerprints, then choose the highest value of rho that produces at maximum maybe 5-8% Ns.

1sthyb and 2ndhyb sheets:

- On the “1sthyb” sheet, clear the sheet then paste the output of the Background subtraction macro: there should be the following columns: Gene ID (column A), Flag (Column C), Reference Probe (Column D), then your other probes in the desired order (Columns E-onward). Delete the empty column between the Flag and Reference probe columns. Make sure that the name of each probe is located immediately above the rows of data (there should not be blank spaces between the data and the probe name), and that the word “Probe” is found in the row above the name of each differential probe. The following is an example of how to format the data so that it can be processed:

	Column A	B	C	D	E	F
Row 4					Probe	Probe
5	Gene_ID		Flag	Ref2	F1	F2
6	_MBBBA_Plate_1-A1		0	47469.21	93.1963	785.8782
7	_MBBBA_Plate_1-A2		0	40087.54	90.9657	831.3289

- **For the Flag column: this should contain the number 2 for each clone that is missing in at least one of the Reference probe hybridizations. Make sure you have a correct Flag column to enter before starting the analysis. Use the Flag Column Tool to generate a correct Flag column.**
- The page “2ndhyb” is there in case you are making duplicate hybridizations and want to take an averaged result. If you are, paste the second set of results in. If not, paste in the identical information you put into “1sthyb”. One thing to note: the reference probe data is taken only from the “1sthyb” sheet with these macros. If you have multiple reference probe data points, you’ll have to modify the “Avg” macro so that it takes the reference probe data from the “2ndhyb” sheet also. We recommend that you calculate an averaged log-transformed result for each clone as $(\ln((\text{hyb1}/\text{ref1})+(\text{hyb2}/\text{ref2}))/2)$ where hyb1 and ref1 are the differential probe and reference probe data points from “1sthyb” and hyb2 and ref2 are the differential probe and reference probe data points from “2ndhyb”.

Avg, AvgControlClones, AvgRatio, Reduced, Reduced_Unequal_Variance sheets:

- These should all be blank. Check these and delete any information you find on them. Be sure to delete background colours from the AvgControlClones page by selecting all (Ctrl + A), going to Format -> Cells -> Patterns and selecting “No Color” and clicking “OK”.

MinMax Result, VerificationResult, VerificationTheoretical, VerificationTheoretical0 sheets:

- Ignore all of these. They’re for something other than assigning fingerprints.

Theoretical sheet:

- The page “theoretical” contains fingerprints for your control clones. Make sure that you have the correct theoretical fingerprints on this page: you must have (1) the correct number of control clones **in the same order as they are listed in the Background Subtract output**, and (2) theoretical fingerprints for the same probes listed (one per column) from left to right **in the same order** as the probes in the “1sthyb” and “2ndhyb” sheets.
- To obtain the data needed for the “theoretical” page: you will need two data files in FASTA format. One file will contain the complete, unambiguous insert sequences for the control clones, with the clones in the same order as they appear in the “1sthyb” and “2ndhyb” sheets. The other file will contain the sequences of the oligonucleotide probes to be analyzed, with the probes in the same order as the probe data columns in the “1sthyb” and “2ndhyb” sheets. Use the **Theoretical Fingerprint Generator** to generate the theoretical fingerprints from these files that you need to cut and paste into the “theoretical” page.

To calculate average log-transformed ratios:

- Macro-> Macros-> Avg. Wait until the macro stops running, which may take a few minutes. You should see output on the “Avg” sheet, consisting of an averaged value for the two values on the “1sthyb” and “2ndhyb” sheets. There should also be output on the “AvgRatio” sheet, consisting of the natural log of the value on the “Avg” sheet divided by the reference probe value on the “1sthyb” sheet.

To calculate the control clone means and variances that are used for calculating boundaries of N for each probe:

- You must do this after running the Avg macro.
- Macro-> Macros -> Control Avg. You must have the “theoretical” page filled in before running this macro, or you will get no output.
- Wait until the Control Avg macro stops running. The format of the data is as follows: the first row has the probe number or name in it, and remaining rows contain the binary fingerprint values for the control clones, which must be in the same order as the data in “1sthyb” and “2ndhyb”.
- You should see output on the “AvgControlClones” sheet, consisting of the natural log-ratio values for each of the control clones (except those for which there was no reference probe binding, which are excluded) and each of the probes. The colours indicate whether the clone is supposed to be theoretically positive (green) or negative (black). Omitted values where the ratio value was less than or equal to zero (and hence, cannot be log transformed) are highlighted in red. These values are also omitted from the analysis.
- If you have multiple reference probes, you should have one file for each reference probe and its associated differential probes. For each of these files, the clones need to be listed in the same order or you cannot continue. You can combine the

data for all the probes at this point if you wish. Do this by copying and pasting columns containing data for each differential probe from all the AvgControlClones sheets as well as the AvgRatio sheets into the appropriate sheet in one of the files. When you are finished pasting make sure to save the file under a new name. At this point you do not need to copy data from any other sheets, since the Reduced_Unequal_Variance macro should only use data from the AvgControlClones and AvgRatio sheets. Be sure to adjust the number of probes you are analyzing on the Preliminaries page before you run the Reduced_Unequal_Variance macro, though.

To assign fingerprints:

- You need to have a value of rho entered at this point, or the macro will not run.
- Macro->Macros->Reduced_Unequal_Variance. Wait for the macro to finish. If you have a lot of data this will take some time.
- You should see output on the “Reduced_Unequal_Variance” page consisting of each clone name followed by fingerprint digits in the form of 0,1 and N values for each probe. Be sure to save the file at this point.
- If you choose to combine your data at this point, you can simply cut and paste columns containing fingerprint data for each probe into a new Excel worksheet. Be sure to leave one blank column between the clone names and the probe fingerprint information to make it easier to generate a fingerprint file later (see User’s Guide to GCPAT).

Other macros in the file:

Reduce: assigns N boundaries based on minimum positive and maximum negative values seen among the control clones, and outputs this information to the “Reduced” page. We do not recommend using this method, as it tends to assign far too many Ns.

Pick_Rho and Pick_Rho_MBR: We figured out that these macros don’t actually pick optimal values of rho, so don’t use them.