By Virginia Rich, graduate student in the DeLong Lab, vrich@mit.edu
Last modified 4/26/08

# Extracting data from array tiff files in GenePix 6.1

Note: some of these features are not available in GenePix 6.0.
Presumes some prior experience with GenePix.

1. Make a master Settings file:

    a. Open first array image, (= a tiff file).

    b. Load the appropriate array list for this array, (= a gal file).

    c. In block mode, select all blocks and align overall array.

    d. Double-click on any block, and adjust spot diameter (with "apply to all" checked off) so that it is a little smaller than the average spot size on your actual arrays (in my case, 150um).

    e. Align each block individually, more precisely.

    f. Adjust the auto-alignment specs:

    Set **Options Box** (Alt + I), **Alignment Tab** as follows:
    Click on "Find circular features"
    Click on "Resize features during alignment"
    between e.g. 70% and 150%
    Click on "Limit feature movement during alignment" to e.g. 40um
    Toggle for unfound features to be "Unflagged"
    (No CPI threshold – default)
    (Check Align Blocks, estimate warping and rotation – default)
    (Automatic Image Reg Max translation 10 – default)
    (No sub-pixel reg. allowed – default)

    g. Then save all of this by going to Save Settings, which will create a gps file. This will be your master Settings file to use on the other arrays *from the same print & hyb date*.

2. <u>Auto-align the array grids for each array scan:</u>

    a. Click on **Batch Analysis** tab in the main GenePix window.

    b. Click "Add" and select your tiff files to process

    c. Select all tiff files and click Add gps file, and choose the master gps file for this set (the file you just created above).

    d. Uncheck "Analyze", leaving "**Align**" checked only.

e. Click on "Configure Alignment" box, and within it check “Find Array” and “Align Features” ONLY - UNcheck “Find Blocks”. (This is because on our high-density array, several of the blocks are very close together and so block-finding gets confused. With the master gps file tailored to each array printrun, block-finding isn’t needed anyway for good gridding if the other two alignment types are used.)

f. Click "Go", with “All at once” checked. Depending on how many files you have, this can take several hours.

3. Check the new alignments.
   a. Use the results browser window (this will come up automatically during batch the alignment. If you close it accidentally, you can reopen it by clicking within the Batch Analysis tab, click on the lower right-hand array-like icon), to check the new gps alignment files it created for each tiff. Clicking on a gps file within the browser box will take you to the Image tab of the main GenePix window, and will load the tiff and its associated newly-created gps file.
   b. Manually inspect EACH gps and tiff pair: manually adjust stray features, and flag areas of surface PLL peeling, or excessive background, as “Bad”, to be discounted from further analysis.
   c. Save each gps file using the same name as before – replace the previous version.

4. Extract the data (you can do this immediately or at a later time):

   a. In the **Batch Analysis** tab, delete all files.

   b. Click on “Add”, then select all tiff files AND their associated gps files at once and click OK – this will link each file correctly.

   c. Check “Analysis” and uncheck “Alignment”.

   d. Click “Go” and “All at once”. Again, this may take a while depending on how many files you’re doing.

   e. You may wish to also use a **flag feature query** – e.g., to automatically flag as bad all spots in areas of background peeling. To set this up, you must first go into the Results Tab and Click on Flag Features, and make a new query to suit your needs. E.g., I created a query called “test” which for most of my arrays successfully flagged many of my missing features, using the following syntax:

      [B532] <= 100 AND [B635] <= 100

This had the effect of removing features along the edges of the array where the PLL coating may have peeled away. Also, if there are large interior peeled sections, it removed those. HOWEVER, it was still unable to find smaller patches or scratches, or features on the edge of a patch that should have been flagged because either part of the feature itself was peeled/scratched or part of its local background was.

In addition, the "test" query I create works for most but not all of my slides – if some have unusually low background, then it artificially flags my data as "bad" even though the spots are there. So you must tailor this to your particular slides based on their background, and also judge whether to used it based on the homogeneity of your background among your slide set. For me, it actually wasn't worth using the auto-filter query since I was manually flagging each gps file for a variety of things anyway before extracting the data.

Another way to do it, computationally longer but perhaps easier depending on your particular slides, would be to have it autoalign and then immediatately analyze, using a "flag features as bad" query. THEN go through each results file, and you'll see which features have been autoflagged on the corresponding gps that loads. You can do additional flagging at that stage, re-save over the gps files, and re-run the analysis to save over the previous analysis files. Clunky, but may be worth it based on your particular specs.

Also, you CAN get a lot more sophisticated with your queries. For example, in order to auto-flag features (spots) that are partially peeled, or at the edge of a peeled region, or have some other aberration, you might like to is have a query that says e.g. "if e.g. >20% of the pixels in the feature, or background, are less than e.g. 100, flag as Bad". I asked Sandra Lew about it, and she said it's something you should be able to do with VBScript in the Results Tab under the Flag Features button. She recommends going to the GenePix Help, where there are chapters on scripting under the Index Tab, describing some commonly-used functions.

5. So, that's it - that's the full pipeline I used to get my results.

Currently I extract all the data the software will give me, in case I ever want to go back to a parameter I don't currently use but which ends up being important, but you can decrease the columns of data that you get if you so desire.

By Virginia Rich, graduate student in the DeLong Lab, vrich@mit.edu
Last modified 4/26/08

## GenePix contact people:

The software technical details guru:
**Sandra Lew,** Sandra.Lew@moldev.com

The woman who updated our hardware and software: **Yvonne FitzGerald,** yvonne.fitzgerald@moldev.com
She recommends that if we have any further problems with software crashes (which we had for a while after she first installed 6.0 and 6.1 on our new machine, before she uninstalled both and reinstalled 6.1), then we make the noise to get a formal field engineer out here to look at the problem, because she says we've exhausted her knowledge, so if the reinstall didn't work then someone else needs to have a go. Since we just bought a new machine (new computer + software package, in March 2008) we ARE under warranty for some period, but I don't know how long – so if the problems re-occur then it would be wise to get them seen to asap.

Our local GenePix sales rep, who has given us a loaner computer when the last one broke: **David Micha**, david.micha@moldev.com