TO PROCESS AFFYMETRIX CEL DATA:

To process data, we first need to load data into Nexus. You can load data directly by loading input data files via the file chooser (see DATA LOADING VIA THE ONE-CLICK METHOD below) or you can load your data by creating and loading a sample descriptor (see DATA LOADING USING A SAMPLE DESCRIPTOR below). The one-click data loading method is the best way to load data for most cases. Only in rare situations a user will need to use the sample descriptor method to load data. With the one-click method, you will select files via a file chooser and will assign any factors (clinical annotations) after the data is loaded. With the sample descriptor method, you will specify the input file location and assign factors to your samples by specifying them in the sample descriptor. When data is loaded factors will already be associated and will appear in the **Data Set** tab. Choose one of the two methods below to load your data.

DATA LOADING VIA THE ONE-CLICK METHOD

1. Run Nexus and click the **Create a New Project** or **Open Existing Project** button. If creating a new project, in the resulting dialog, specify a name for your project and choose a genome and then the appropriate genome build based on the mapping information for your probes. If adding additional samples to an existing project (of the same genome build), select your project folder in the resulting dialog.

2. Click on the **Load** button and select **Load Data** in the **Data Set** tab. Select **Affymetrix CEL** from the drop down in the resulting **Add Sample Data** window. A number of options appear below the checkbox **Set default processing settings based on the following**. This "guide" assists in adjusting the parameters in the **Settings** and is based on the type of data you are loading. If you do not want to use this guide, uncheck the box (all the options will be grayed out) and ***skip to Step 6***. If you use the "guide", making selections here adjusts parameters in the "Analysis" section as well as "Systematic Correction" section of the **Settings** window. The Help buttons to the left of the options bring up pop ups with additional information on what changes the selection will make to the settings parameters.

3. Select one of the radio buttons depending on the type of data you are loading: **For mosaic samples such as cancer (possible increased false-positive rate)** or **For non-mosaic samples**. The selection here adjusts the thresholds used to make calls (e.g. High Gain, Gain, Loss, Big Loss).

4. Adjust the **Stringency** slider to one of 3 positions: **Lenient**, **Average**, or **Stringent**. This adjusts the sensitivity of the algorithm in making calls and sets the appropriate value for the **Significance Threshold** setting. "Lenient" will generate more calls and "Stringent", less calls (more specificity).

5. Choose if you want to apply systematic correction by checking or unchecking the checkbox for **Perform Systematic Correction**. This corrects the wavy pattern often present in the probe distribution. Since the correction is based on probes used, the specific array is needed. Select the array from the **Array Types** dropdown menu. Arrays listed here are based on the files and folders available in the **SystematicCorrection** directory of the **Organism->genome build** folders. If your array type is not listed in the dropdown, please contact BioDiscovery Support (support@biodiscovery.com) to obtain the necessary files.

6. Click on **Select Files** to select your input files via the file chooser and click **Done.** Your samples will now be listed in the **Data Set** tab. If you want to assign any factors to your samples, you can do so with the selection

in the **Add Factors** button dropdown. With **Add Factor,** you will manually enter factor values for each sample by typing them individually or by selecting column cells and copying and pasting values from a spreadsheet (see the user manual for "Allow column selection"). With **Load Factors**, you will select a tab delimited text file containing sample names and any factors assigned to them. A column called **Sample Name** containing names matching the sample names in the project is required and any number of factor columns with factor names as column headers (DO NOT prefix the column header with **Factor:**). See *Loading Factors* in the *Data Set Tab* section of the user's manual.

7. If you used the guide, files are automatically loaded and processed after you selected files and clicked **Done** in the previous step. If you did not use the guide, click **View** to process the data and see the results. If you get an error stating that a reference file cannot be found, please contact support@biodiscovery.com so that we can provide you with the appropriate file for your genome.

DATA LOADING USING A SAMPLE DESCRIPTOR

1. You need to create a sample descriptor file (a tab delimited text file) to load data into Nexus. In the T**emplates** folder of the Nexus installation directory, you will find several arrays types that Nexus is able to process. Open the folder titled **Affymetrix** and select the **AffyCEL-Template.txt** template file for your sample descriptor. It may be easiest to edit this file by opening it in a spreadsheet application such as Excel. If you have technical replicates, please refer to the document called *Processing Replicate Data* to see how to specify this in the sample descriptor.

2. In the first column, **Sample Name**, specify a unique name for each sample that you are loading (e.g. sample1, sample2, etc.). Note that you cannot use the following special characters /\:*?"<>| within the sample name. In the second column, **File**, specify the location of your input data file. You can use either the full path (e.g. C:\Nexus Projects\Brain Tumor \sample1.cnchp) or a relative path (relative to where you will save this sample descriptor file). If the descriptor is in the same folder as the sample data files then you can specify just the name (e.g. sample1.cnchp) without any path qualifiers. Those are the only required columns. You can add an unlimited number of additional columns to specify Factors (clinical annotations) for each sample. The column header is **Factor:** followed by a name for the Factor. Save this file with a new name and make sure to save it as a tab-delimited text file.

   Example of a sample descriptor for Affymetrix CEL input files:

Data Type:	Affymetrix CEL	
Sample Name	File	Factor:Tumor Type
sample1	sample1.CEL	GBM
sample2	sample2.CEL	AA
sample3	sample3.CEL	GBM

3. Run Nexus and click the **Create a New Project** or **Open Existing Project** button. If creating a new project, in the resulting dialog, specify a name for your project and choose a genome and then the appropriate genome build based on the mapping information for your probes. If adding additional samples to an existing project (of the same genome build), select your project folder in the resulting dialog.

4. On the next screen, click the **Load** button, select **Load Descriptor** from the dropdown and choose the sample descriptor file you saved in step 2 above. You will see all your samples loaded into Nexus.

5. Click **View** to process the data and see the results.