

## Optional: Use MethylationEPIC for CNV analysis

Although the 450K and MethylationEPIC arrays were initially designed to analyze DNA methylation, they are essentially a dense SNP array and can be used for copy number analysis (Feber *et al.* 2014). The probe intensity data is easily parsed from the idat files by using the *Additional Probe Data Spreadsheet Selection* dialog (Figure 1) when importing the raw data. Examining the raw intensity data can also be useful for QA/QC purposes.

Follow the steps for importing Illumina methylation data detailed in [Import and normalize methylation data](#) until you reach the *Import Illumina iDAT Data* dialog with *Manifest File* and *Output File* panels (Figure 1).

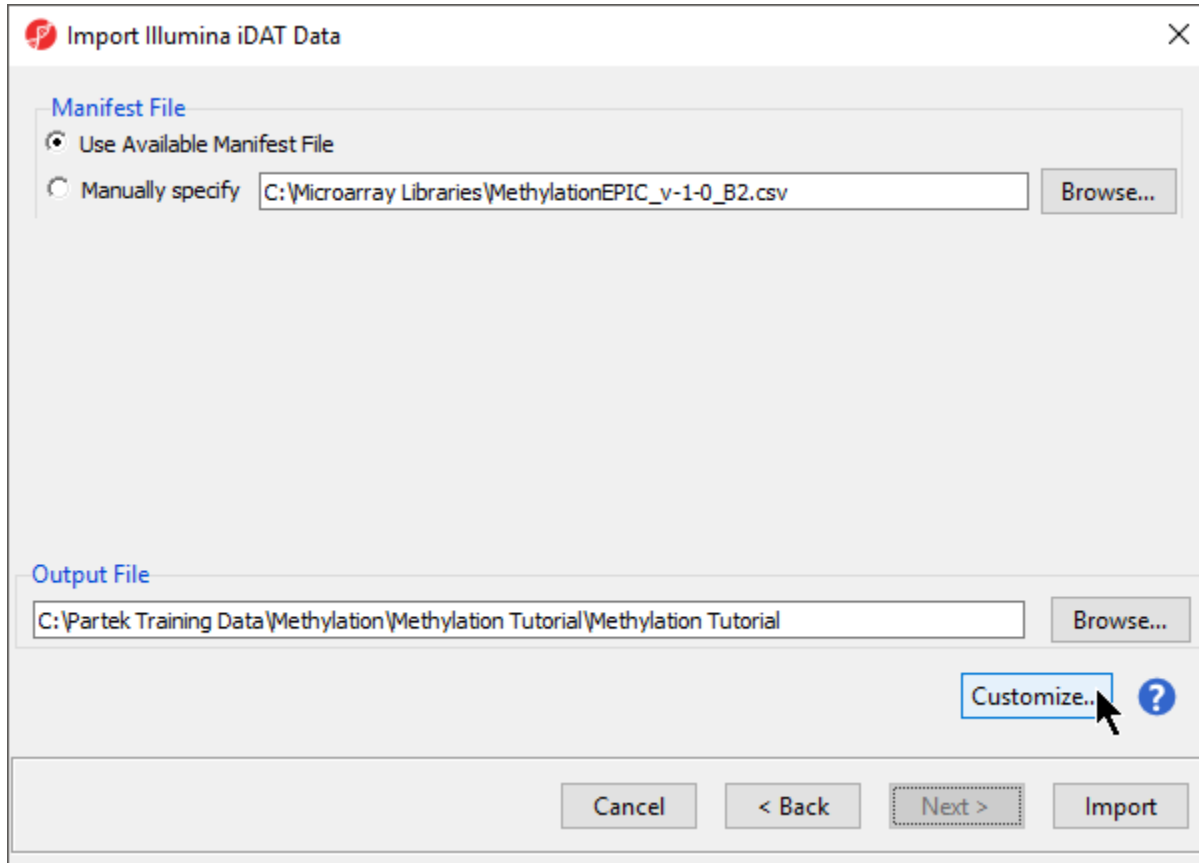


Figure 1. Customizing output during data import

- Select **Customize...** to open the *Advanced Import Options* dialog
- Choose **No normalization** in the *Normalization* section of the *Algorithm* tab
- Select the *Outputs* tab (Figure 2)

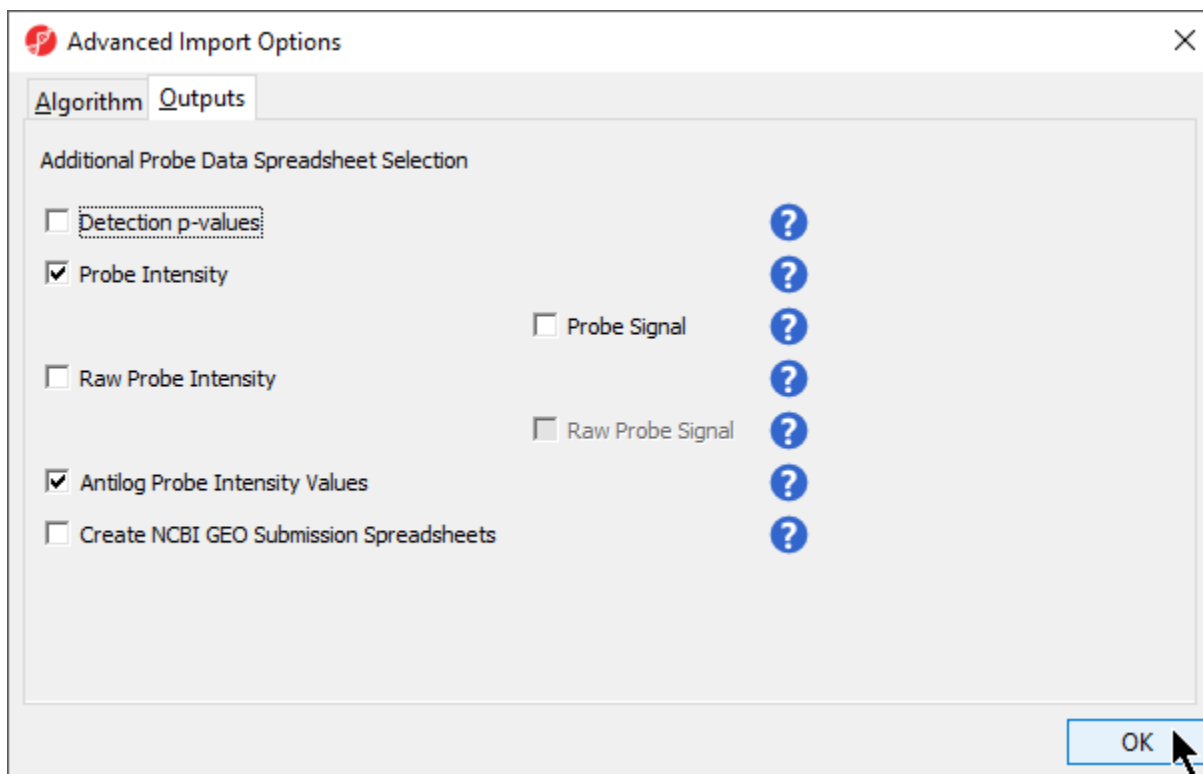


Figure 2. Selecting additional probe data to include during data import

Information about the different output options can be found by selecting the adjacent ( ? ) icon.

**Detection p-values.** This is the confidence score that the signal of a probe was significantly higher than the background defined by negative control probes. Selecting this checkbox produces a spreadsheet ending with '\_detectionp' in addition to the spreadsheet containing beta values. Each row of the \_detectionp spreadsheet will be a different sample and the sample names will end in '\_detectionp'. This spreadsheet can be used to filter out probes that do not show signal above background.

**Probe Intensity.** This is the sum of the methylated and unmethylated intensities per probe. Selecting this checkbox produces a spreadsheet ending with '\_probe' in addition to the spreadsheet containing beta values. Each row of the \_probe spreadsheet will be a different sample and the file names will also end in '\_probe.' The probe intensity values will be log2 transformed by default (note that the beta values are not log2 transformed).

**Probe Signal.** This option will become available if *Probe Intensity* is selected. Selecting this checkbox produces a spreadsheet ending with '\_probe.' The methylated and unmethylated intensities are shown on separate rows for each sample, in addition to the summed values. The sample names will end in '\_meth' or '\_unmeth' for methylated and unmethylated values, respectively. The probe intensity values will be log2 transformed by default.

**Raw Probe Intensity.** This is the sum of the raw red and green signal intensities per probe. Selecting this checkbox produces a spreadsheet ending with '\_raw' in addition to the spreadsheet containing beta values. Each row of the spreadsheet will be a different sample and the file names will also end in '\_raw.' The raw probe intensity values will be log2 transformed by default.

*Raw Probe Signal.* This option will become available if *Raw Probe Intensity* is selected. Selecting this checkbox produces a spreadsheet ending with '\_raw.' The red and green intensities will be shown on separate rows for each sample, in addition to the summed values. The sample names will end in '\_red' or '\_green' for red and green values, respectively. The raw probe intensity values will be log2 transformed by default.

*Antilog Probe Intensity Values.* Selecting this checkbox will show the probe intensity data without log2 transformation.

*Create NCBI GEO Submission Spreadsheets.* Generates matrix processed and matrix signal intensities spreadsheets for GEO submission.

How you proceed depends on your study design. Here is an example series of steps to prepare the tutorial data set for copy number analysis:

- Select **Probe Intensity** and **Antilog Probe Intensity Values** (Figure 2)
- Select **OK** to close the *Advanced Import Options* dialog
- Select **Import** to import the data and perform the selected normalization method
- Select the (**\_probe**) spreadsheet from the spreadsheet tree
- Delete any samples with **\_detectionp** names
- Create sample attributes and assign samples to the groups as described in [Annotate samples](#)
- Select **Transform** from the main toolbar
- Select **Normalize to baseline**

Configure the *Normalize to Baseline 1* dialog as shown (Figure 3)

- Select **Use control set from this spreadsheet**
- Set *Control Category* to **B cells**
- Select **Ratio to baseline** from the *Normalization Method* section
- Select **After ratio apply log base 2**
- Select **New Spreadsheet** from the *Configure Output* section

**Normalize to Baseline 1**

**Choose control samples to be used as baseline**

☐ Use all samples from this spreadsheet

☒ Use control set from this spreadsheet

Column that contains the treatment variable: 2. Cell Type    Control category: B cells

☐ Use matched pairs from this spreadsheet

Column that contains the treatment variable: 2. Cell Type    Control category: B cells

Column that matches the pairs: None

Estimate baseline using: Mean

**Normalization Method**

☒ Ratio to baseline    ☒ After ratio apply log base 2

☐ Subtract baseline

**Configure Output**

☐ Remove control samples after normalization

☐ Same spreadsheet    ☒ New spreadsheet

Result file: Methylation\_cnv\_probe\_normalized.txt    Browse...

OK    Cancel

Figure 3. Configuring normalize to baseline

- Select **OK** to generate the spreadsheet

This spreadsheet contains copy number values per probe in log2 space (i.e. diploid = 0). Prior to performing copy number analysis, you can normalize for local GC abundance.

- Select **Transform**
- Select **Adjust Based on Local GC Content...**
- Click **OK** to run *Local GC Adjustment* (Figure 4)

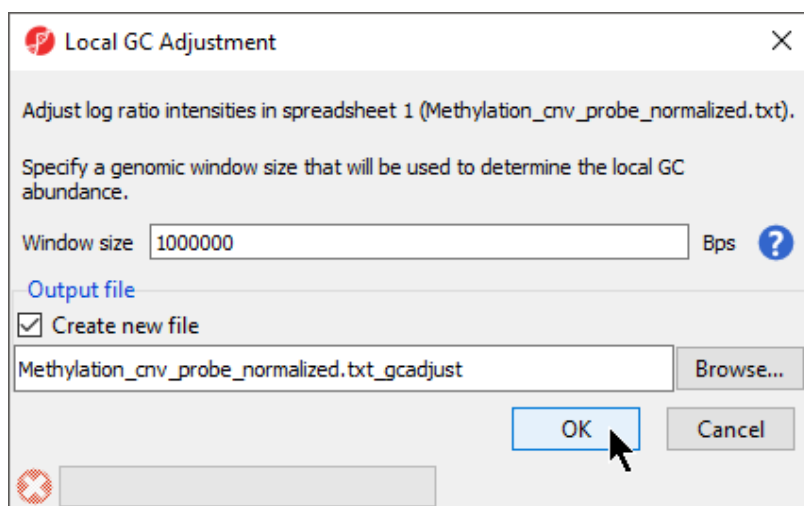


Figure 4. Adjusting for local GC content

The GC adjusted spreadsheet is the starting spreadsheet for copy number analysis. You can now switch over to the **Copy number workflow**, skip the **Create copy number** step, and begin with the **Detect amplifications and deletions** step. Consult the user's guide for the copy number workflow for subsequent steps.

## References

Feber A, Guilhamon P, Lechner M, et al. Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biology*. 2014;15(2):R30. doi:10.1186/gb-2014-15-2-r30.

## Additional Assistance

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.



★

Your Rating: ☆☆☆☆☆ Results: ★★★★★ 33 rates