

# Perform data quality analysis and quality control

Principal component analysis (PCA) can be performed to visualize clusters in the methylation data, but also serves as a quality control procedure; outliers within a group could suggest poor data quality, batch effects, mislabeled samples, or uninformative groupings.

- Select **PCA Scatter Plot** from the *QA/QC* section of the *Illumina BeadArray Methylation* workflow to bring up a *Scatter Plot* tab
- Select **2. Cell Type** for *Color by*
- Select **3. Gender** for *Size by*
- Select (🔄) to enable *Rotate Mode*
- **Left click and drag** to rotate the plot and view different angles (Figure 1)

Each dot of the plot is a single sample and represents the average methylation status across all CpG loci. Two of the LCLs samples do not cluster with the others, but we will not exclude them for this tutorial.

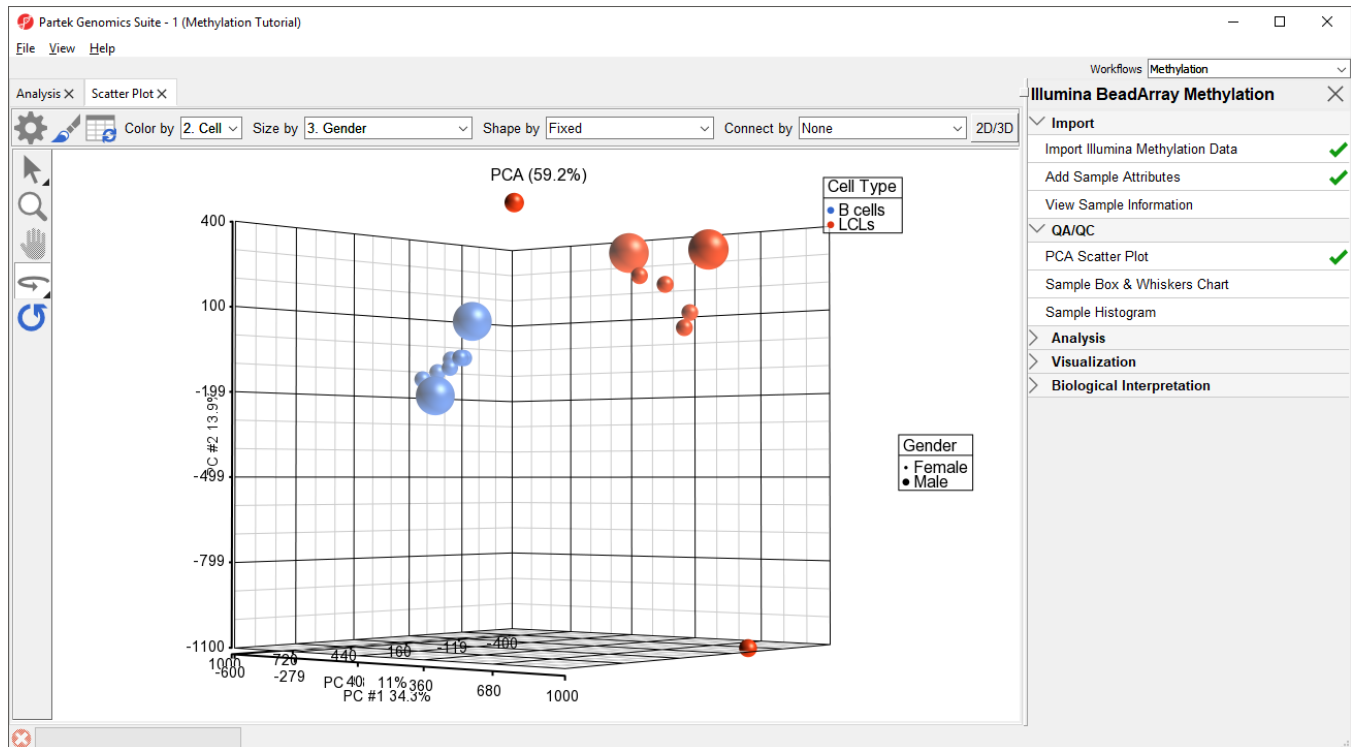


Figure 1. Principal components analysis (PCA) showing methylation profiles of the study samples. Each sample is represented by a dot, the axes are first three PCs, the number in parenthesis indicate the fraction of variance explained by each PC. The number at the top is the variance explained by the first three PCs. The samples are colored by levels of 2. Cell Type

Next, distribution of beta values across the samples can also be inspected by a box-and-whiskers plot.

- Select **Sample Box and Whiskers Chart** from the *QA/QC* section of the *Illumina BeadArray Methylation* workflow to bring up a *Box and Whiskers* tab

Each box-and-whisker is a sample and the y-axis shows beta-value ranges. Samples in this data set seem reasonably uniform (Figure 2).

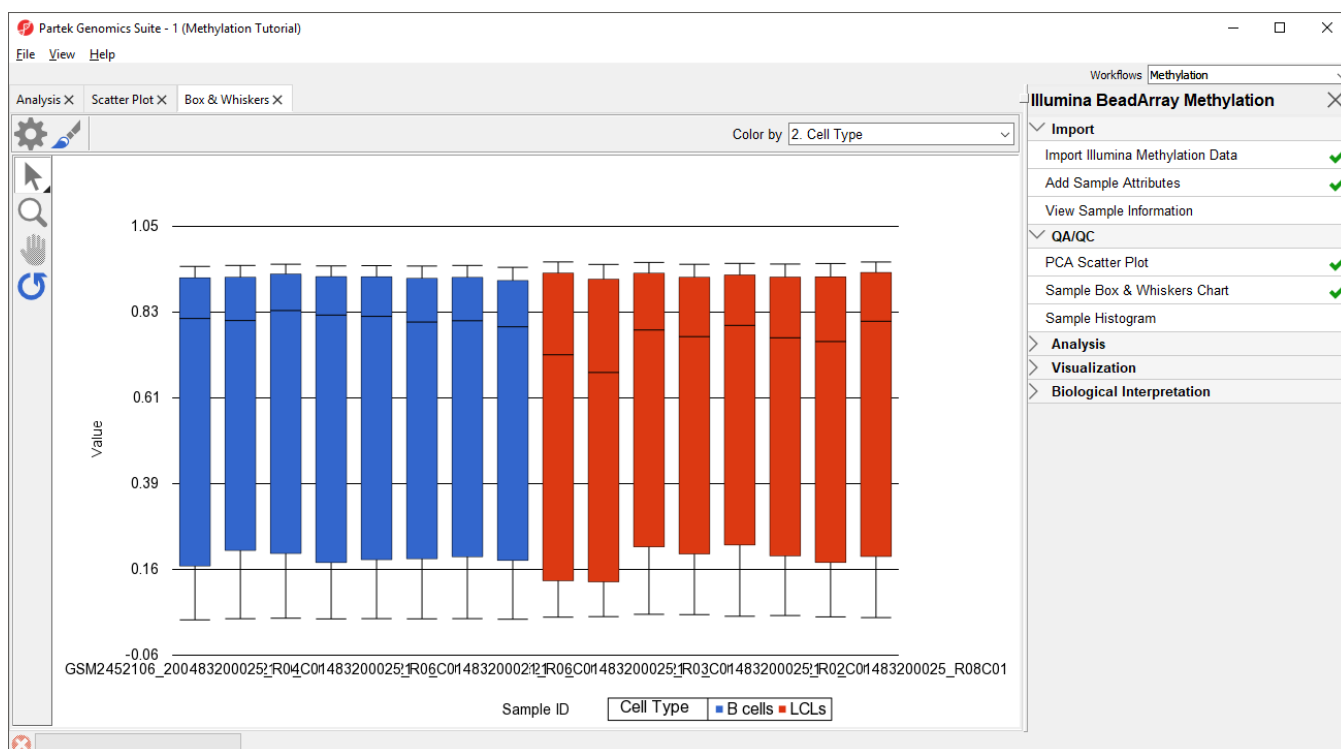


Figure 2. Box and whiskers plot showing distribution of M-values (y-axis) across the study samples (x-axis). Samples are colored by a categorical attribute (Cell Type). The middle line is the median, box represents the upper and the lower quartile, while the whiskers correspond to the 90th and 10th percentile of the data

An alternative way to take a look at the distribution of beta-values is a histogram.

- Select **Sample Histogram** from the *QA/QC* section of the *Illumina BeadArray Methylation* workflow to bring up a *Histogram* tab

Again, no sample in the tutorial data set stands out (Figure 3).

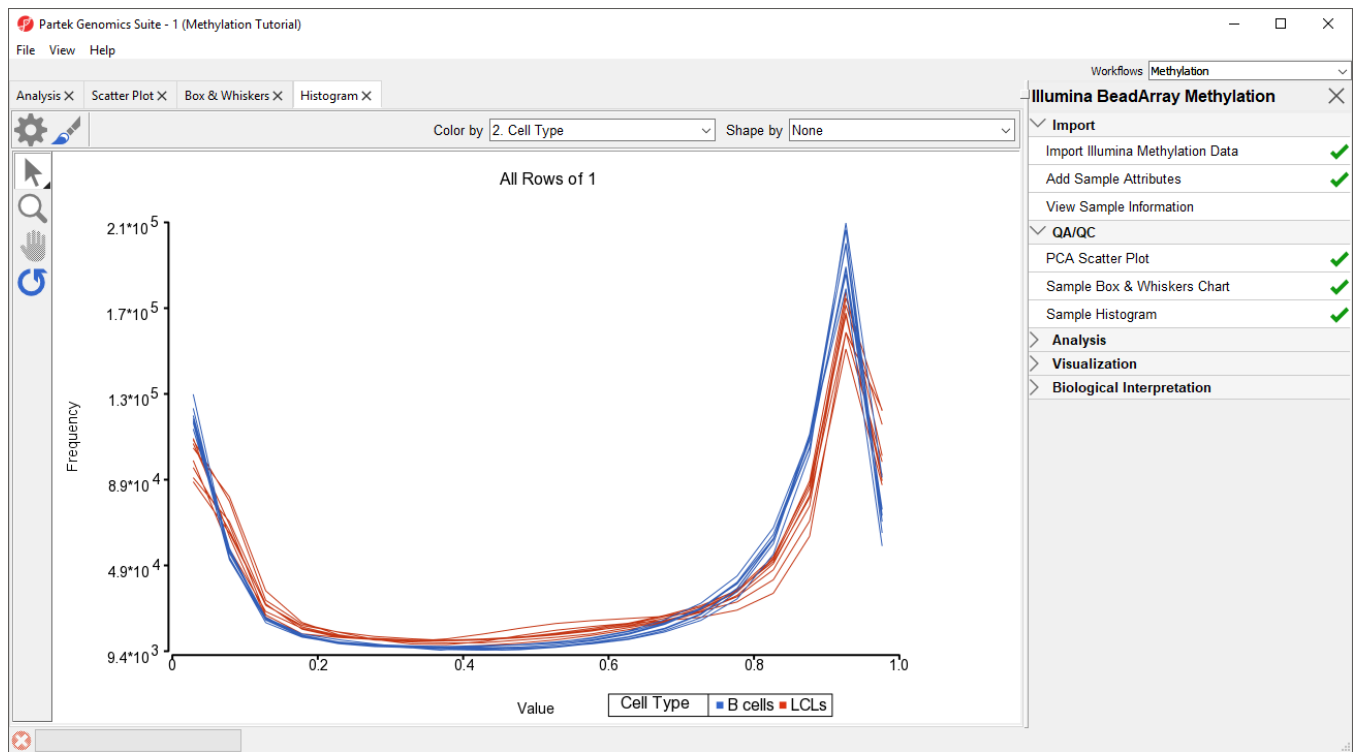


Figure 3. Sample histogram. Each sample is a line, beta values are on the horizontal axis and their frequencies on the vertical axis. Two peaks correspond to two probe types (I and II) present on the MethylationEPIC array. Sample colors correspond to a categorical attribute (Cell Type)

[« Annotate samples](#) [Detect differentially methylated loci](#) »

## 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: 36 rates