Assessing Sequence and Microarray Data Quality: Commands (Based on BaRC SOP)

Microarray Data: *Example of commands using the packages simpleaffy and arrayQualityMetrics from BioConductor*

```
library("simpleaffy")
library("affyPLM")  #for NUSE
# Read cel files from directory
data = ReadAffy()
# Create affy QA matrix
data.qc = qc(data)
# percent present
pp = percent.present(data.qc)
# RLE and NUSE
plmStruc = fitPLM(data)
RLE(plmStruct, type="stats")
NUSE(plmStruct, type="stats")

#arrayQualityMetrics (Affy data)
library(arrayQualityMetrics)
CELs = ReadAffy()
eset = rma(CELs)
arrayQualityMetrics(eset, outdir="QC", force=TRUE)

#arrayQualityMetrics (Agilent Data)
library(arrayQualityMetrics)
scanFiles = dir(pattern = ".*.txt$")
maData = read.maimages(scanFiles, source="agilent")
arrayQualityMetrics(expressionset=maData, outdir="QC", force
 = TRUE, do.logtransform = TRUE)
```

Sequence Data: *Example of commands using FastX and FastQC*

```
#FastQC
fastqc s_1_sequence.txt s_2_sequence.txt

#FastX Toolkit
#quality_stats
fastx_quality_stats -i s_1_1_sequence.txt -o s_1_1_sequence.stats
#Nucleotide Distribution:
fastx_nucleotide_distribution_graph.sh -i s_1_1_sequence.stats  -o
   s_1_1_sequence.stats.nuc.png -t "s_1_1_sequence.stats Nucleotide
   Distribution"
# boxplot:
fastq_quality_boxplot_graph.sh -i s_1_1_sequence.stats –o
   s_1_1_sequence.stats.quality.png -t "s_1_1_sequence.stats Quality
   Scores"
```

The above commands are run on tak. If you're running the commands on your own computer, make sure the programs/packages are installed.