## CIDA Omics Data Additional Needs Report

# Preprocessing & QC Report

### Overview

Overview on type of data and groups (e.g. looking at gene expression differences between treatment and control via RNA-Seq technology)

### Raw Data

* How was the data delivered (e.g. fastq files or idat files)
* Where it is saved (which server)

### Processing (if needed)

This is very dependent on the previous step. Sometimes the data is raw and sometimes it is aligned or already normalized.

* Write up any code that was used outside of R
* RNA-seq example from fastq files:
  + Trim reads
  + FastQC
  + Map to genome
  + Quantitate using RSEM

### Diagnostic Plots

* PCA plots
* Dendrogram
* RLE

### Normalization

* Method used for normalizing data
* Possible transformation (e.g. rlog)

### Diagnostic Plots

* Compare the pre and post diagnostic plots

### Outlier Detection

* Sample level
* Feature level
* NOTE: depending on technology, this step looking for outliers might be in a different place.
* Summary
* Summary of final dataset
* Where it is located at

# Analysis Report

### Overview

Overview on type of data and groups (e.g. looking at gene expression differences between treatment and control via RNA-Seq technology, where the data that is actually used for the analysis is and where the preprocessing and QC report is)

* Table 1 of phenotypes
  + Many omics are just a 2 group comparison with no covariates, so in this case just reporting the total n’s for each group
* Methods of analyses going to perform (differential expression, network analysis, etc.)

### Candidate Summary

Summary of candidate features

* Total number of candidate genes
* If multiple tests, how do the candidates compare in each test

### Enrichment Analysis

* Get a systematic view of what is happening in your model
* Specify which ontology databases, which candidates inputted (probe ID, gene symbol…), was a background used.
* Specify program (EnrichR, Panther…)

### Future Work

* Plans for validation
  + Public datasets
  + In house: rtPCR or pyrosequencing

### References