

IRON HACK

#MassiveSeq

#MassiveSeq



Team Members (in alphabetical order)

Swamy Rakesh Adapa

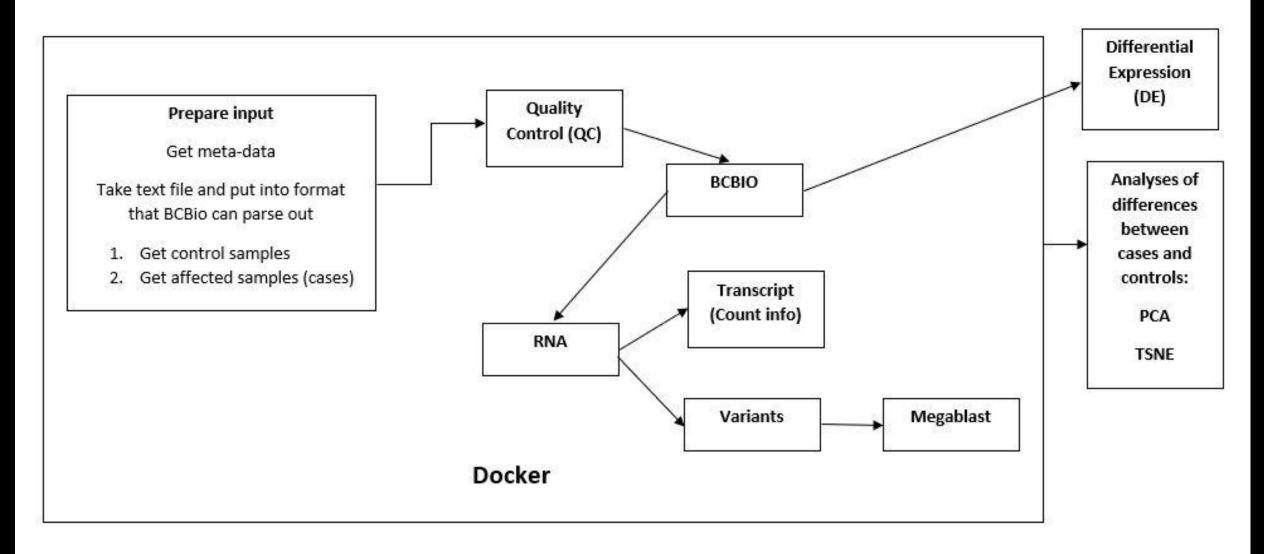
Linh M. Duong

Doug Franz

Shivani Kapadia

Thomas Keller

#MassiveSeq Flow Diagram for Iron Hack



Multiply by n

Pipeline

- Using SnakeMake to run a batch of jobs on the clusters (automate process) to get through 99 samples
- Intermediate step run every sample through
 - Last step is process with FeatureCounts (raw counts) to get to differential expression analyses

Post-processing

- Use R package in postprocessing
 - Do one initial pass of meta-analyses with Meta-Seek package
 - Come back and look at other ways of classifying case and control samples by binning genes based on those classes

Expression Analyses

- Compare cases and controls among 4 datasets
 - What are variations between datasets?
 - Make our own methods or use package already established?

Future Directions

- Expanding out the GATK variant calling
- Make system fully parallel where we can run all samples at once
 - Right <u>now</u> we can only run 12 samples at a time

Limitations

- Reinforcing the between study batch effects
 - If we took all the data at once and look at clustering on their own, different conclusions may be reached