**Running differential gene expression analysis on bulk-RNASeq data.**

Read counts for every feature + sample were derived via

featureCounts -T 16 -p -a </scratch/jm/tcs.gtf> -o /scratch/kreitzer/ [\*bam]

The resulting count-matrix was transferred (local machine) and analyzed with DESeq2 [doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8)]. The entire script for descriptive and exploratory analysis is hosted on: https://github.com/chris-kreitzer/Twist/blob/main/Scripts/DGE.bulkRNA.R

1. Naïve look at the raw count data for *twist* across samples

Chart, bar chart

Description automatically generated

This is the most basic summary (and not good visualization!!); however, the mean (upper bound of bar chart) across the five groups (summary stats among 3 replicates).

If we are looking into variance-stabilized-transformed counts (absolutely necessary when dealing with this kind of data; provided by DESeq2 package) –



* WT4d (n=3) still appears to produce the least *twist* quantity
* TwistHead (n = 3) stands out (highest *twist* quantity)
* Quantities of Bubble, Twi4d and WThead are similar

**Questions:**

* We do expect *twist* transcripts, even in CRISPR-mutant samples (Bubble, Twi4d and TwiHead); however, is there a naïve explanation why Twihead stands-out (e.g. compared to WT4d); can this be explained by developmental stage, tissue samples, etc.

1. Inter- and Intra-sample comparability via clustering and PCA

PCA on vst-data (variance-stabilized transformation)



Clustering via (distance matrix via euclidean distance)

Timeline

Description automatically generated

* Inter-sample comparability (n=3 per group) (gene expression pattern) looks really good (both PCA and clustering)
* Intra-sample comparability: I am not sure, but it seems (appear ??) that maybe Twi4d (=mutant) and WThead may be mixed (confused, etc.) – not sure, but the pattern observed above looks a bit like a confusion…?

1. Differential Gene Expression:

I did DGE-analysis. However, since the intra-sample comparability above (see plots) is not entirely sure to me I will reluctantly provide some results.

* There are 325 genes differentially expressed between mutants and wild-types (Lfc-Threshold @1 and FDR < 0.1)
* *twist* is not listed as differential expressed among the ‘condition’ (mutant vs wt) – but makes sense, if we look into the raw (naïve) count-charts above.
* **NV2.11253** = gene with greatest log-fold change between mut and wt group

![Chart, scatter chart

Description automatically generated]()

@another example: NV2.618



Happy to discuss more on that (especially PCA plot) and further approach;

Any feedback is highly appreciated :)