test summary report

2024-11-08

so you've run a CRISPR functional screen..but what does it mean??

this will be some fun text about the crispr functional screen (but they should) already know what's up

summary stats table

this my attempt at pulling a table in. can read in .tsv/.csv files via R or python (I'm using R here). ive hidden the code chunk so you can only see the output.

Table 1: sample counts per analysis step

Sample ID	Total Reads Sequenced	Reads with Vector	Reads without Vector	Reads Mapped to sgRNA	Total sgRNA Represented	sgRNA above 10 Reads
transE-High_S47_L005	1263892	445377	4036.75	30431955 (89.29%)	18894	18514
transE-Low_S45_L005	992104	444769	5474.50	29817404 (88.71%)	18161	17895
transE-Medium_S46_L005	1434710	553272	7389.50	37816504 (88.86%)	20015	19383
transE-Pre_S48_L005	1163340	523980	4476.75	35301646 (88.86%)	18711	18420

Table 2: summary table explained

Column	Explanation
Total Reads Sequenced	Total number of reads sequenced in your sample.
Reads with Vector	Total number of reads containing the vector sequence. This is the number of reads used for alignment.
Reads without Vector	Total number of reads where the vector sequence was not detected and therefore not considered for sgRNA mapping. These reads are not included in the alignment step.
Reads Mapped to sgRNA	Total number of reads that contained vector sequence that had a read mapping to a sgRNA at the set mismatch rate of 0.
Total sgRNA Represented	The total number of unique guides with at least 1 read count detected. Percentage is based upon fasta of guide library used as the reference.
sgRNA above 10 Reads	The total number of unique guides with at least 10 read counts detected. Percentage is based upon fasta of guide library used as the reference.

what we do

(workflow overview)

1. **FASTQC and MultiQC**
   Starting out, we check the quality of the FASTQs we recieve along with the sequences themselves.

   Key considerations:

   - Are the sequences properly positioned on the sense strand or are they reverse complemented?
   - Is the sgRNA position staggered or not staggered? Staggering sgRNAs increases library complexity and provides the sequencer with greater diversity.
   - What is the plasmid/vector sequences on the 5' end? This is library specific and will change depending on which sgRNA library you prepped your samples with.

2. **Cutadapt**
   We then use Cutadapt to trim the plasmid/vector sequence from the 5' end (this is the same sequence that we identified in the previous step based on the library used).

3. **Bowtie1**
   We use Bowtie1 to build an index of sgRNA library sequences to align to your FASTA sequences. In this same step we perform the alignment.

4. **BBMap**
   BBMap can be used to count the number of reads per sample that aligned to the sgRNA index on per sequence and gene basis.

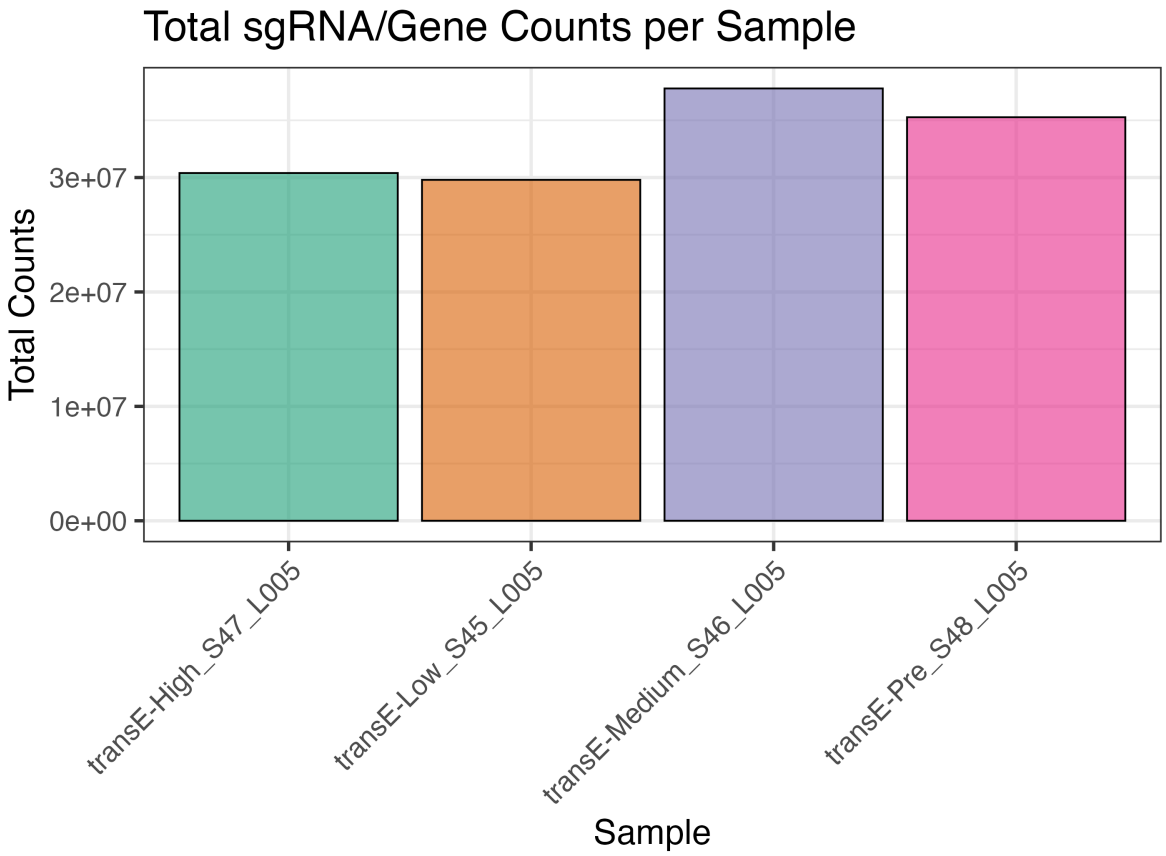
5. **R Analysis**
   We then move the counts of reads per sample to R and look at total/transformed reads per sample, sample PCA clustering, and correlation analysis. This is to verify that the results look similar across samples in different biological groups and that your set of experiments was successful.

plots

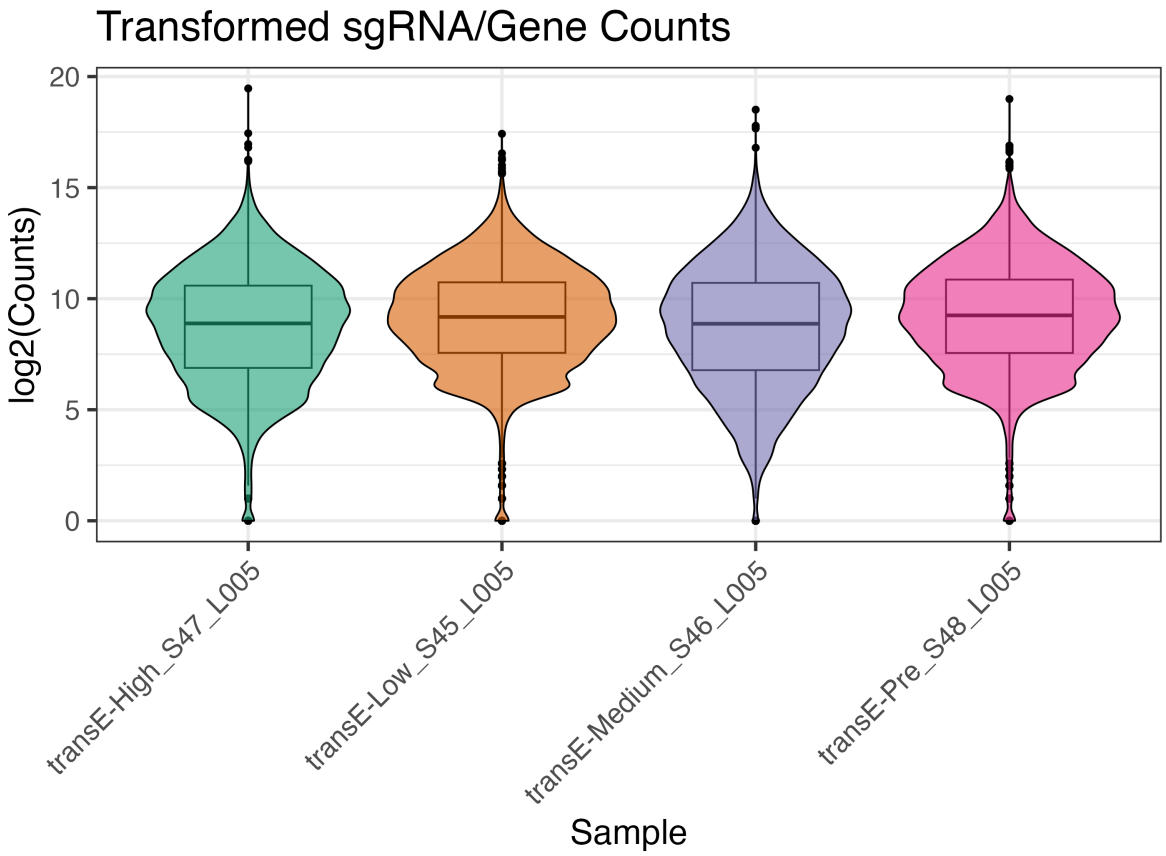
this is my attempt at pulling a plot in. update: plots need to be in .png format to render ! saved all plots as a .png as well and put them in `report/plots` for report generation.

total and transformed counts

here's info on why we looked at total and transformed counts sgRNA guide sequences overall (there isn't a separate one for individual genes since they both add up to the same number).



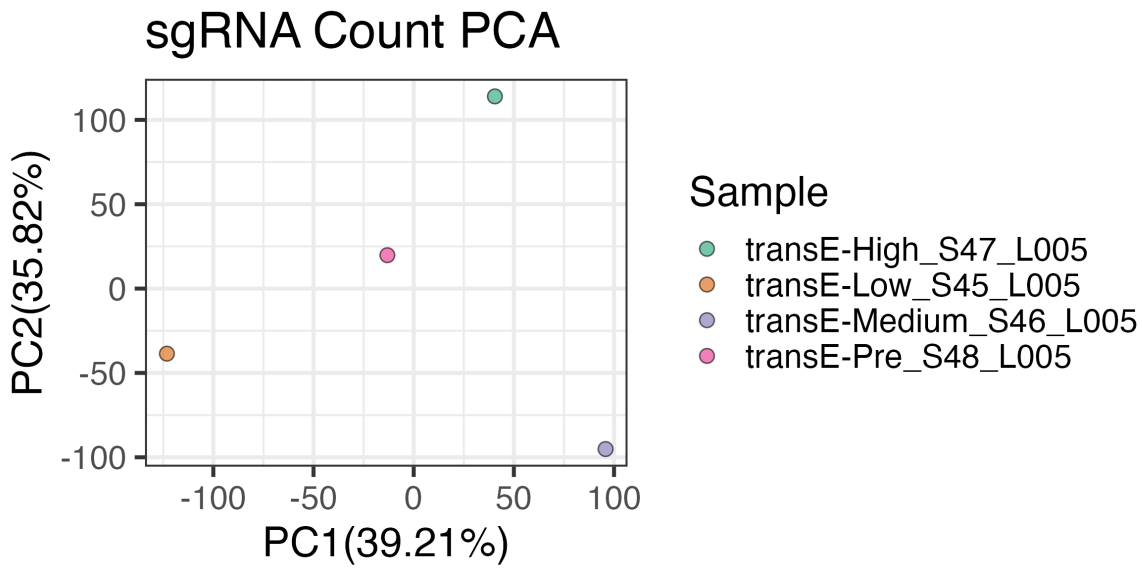
total counts of guide sequences per biological group



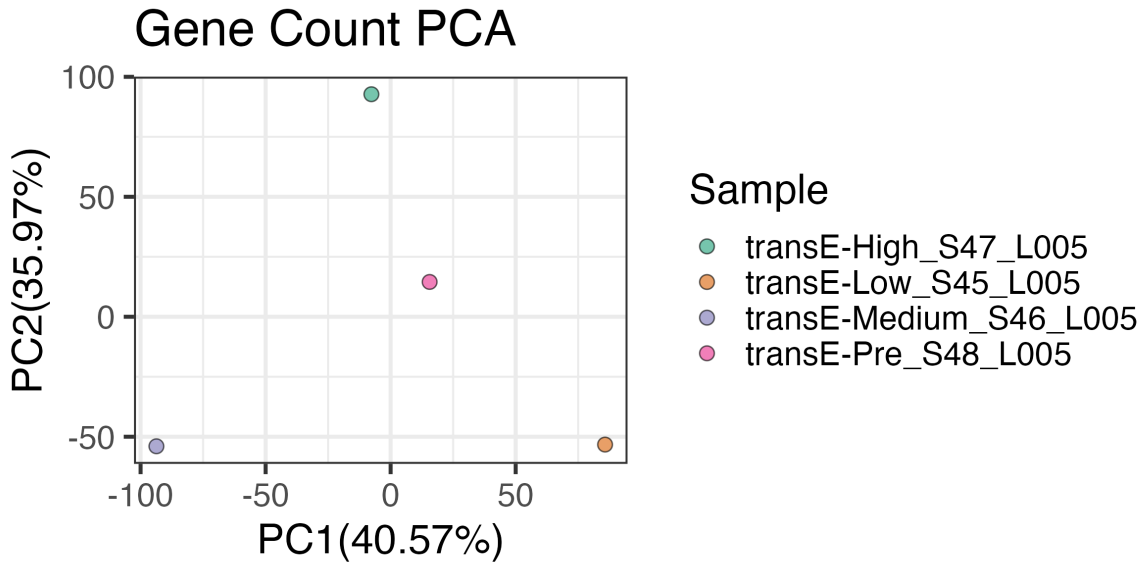
log2() transformed counts of guide sequences per biological group **(why are we doing this??)**

pca analysis

here's a little info on the pca analysis done and the two outputs! looking at the clustering of biological replicates via principal components analysis. we should expect replicates in the same biological group to cluster together (this is not a good example).



we did this using all of the guide RNA sequences

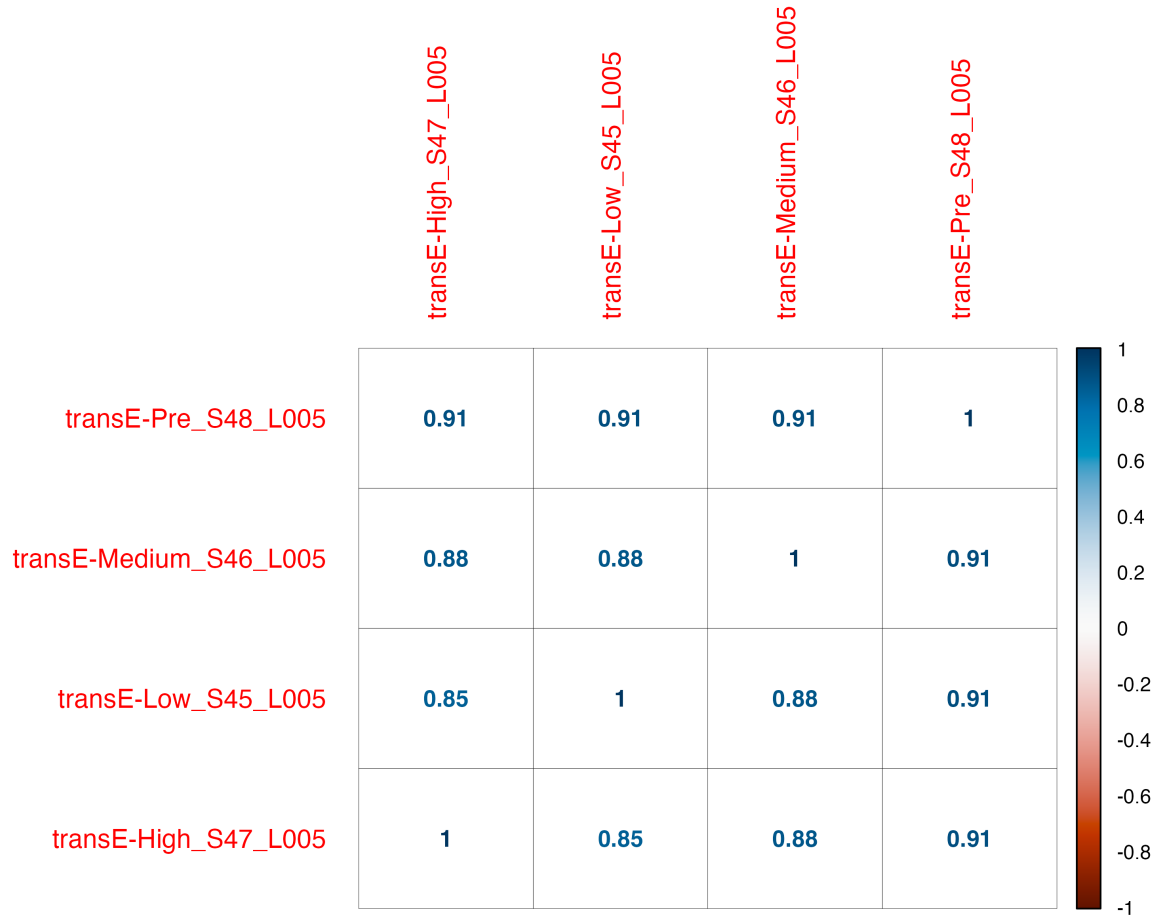


and the invidiual genes

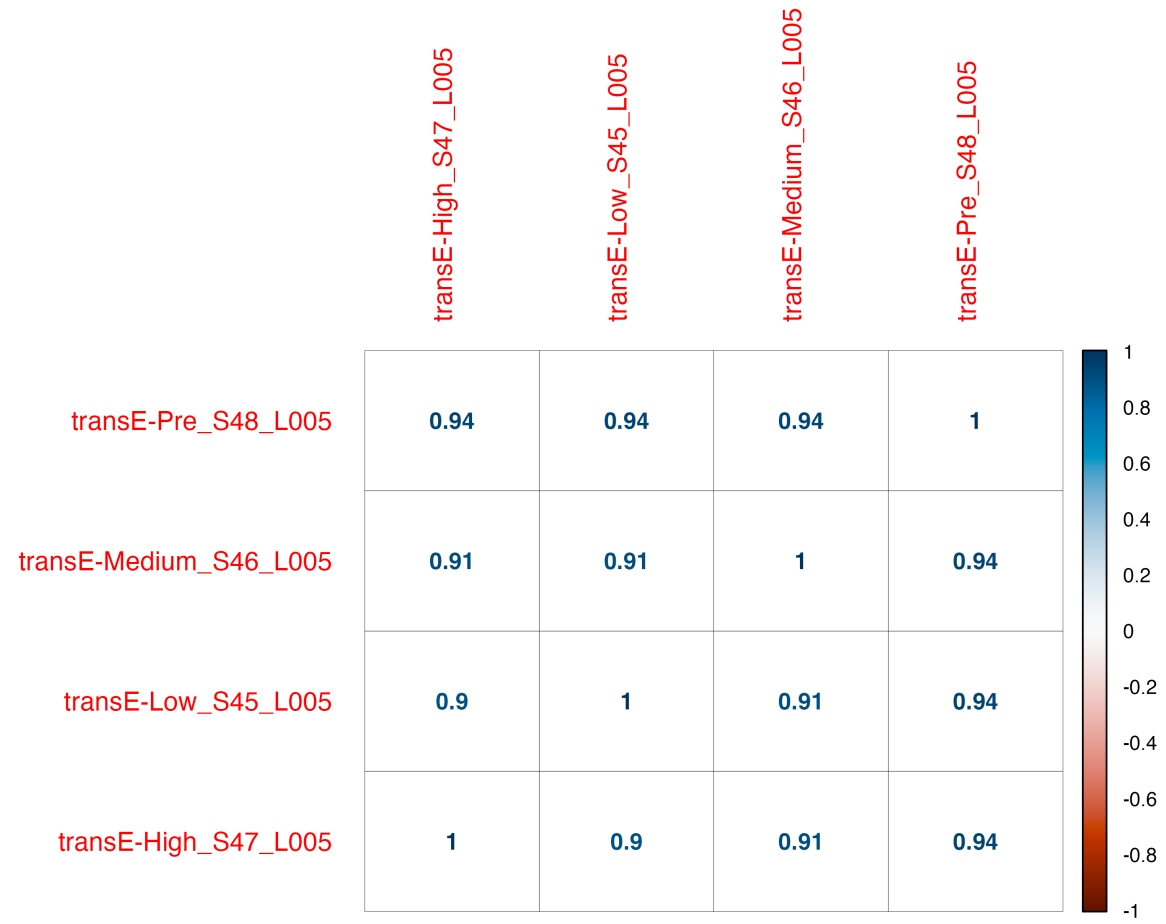
correlation analysis

here's info on why we did correlation analysis on the different biological groups! we wanted to make sure that all replicates in a biological group correlate more closely with each other than

other biological groups (i.e. we want a correlation value closer to 1). positive correlations will be deeper blue and negative correlations will be deeper red.



we did this for all guide RNA sequences



and all individual genes

contact information

If you have any questions regarding your results, our analysis, or ways to further process your data please feel free to contact us! As always, if we have done any analysis for you that ends up in a publication please consider including us in the author list.

Madi Apgar, MS: madison.apgar@cuanschutz.edu
Tonya Brunetti, PhD: tonya.brunetti@cuanschutz.edu

This workflow is publically available as a GitHub repository.