Hi Nick,

I would like to go ahead and make some final decisions to classify hits from the shRNA screens using the analyses you made from the Proton runs. It will be some time before we can compare the Illumina data to the Proton data.

The endpoint is to have a list of genes that are most likely hits. In addition, I would like to generate a list of genes that are least likely to be hits, so they can serve as negative controls in the future. I know you are probably busy, but in terms of stuff you are helping me with this would be high priority because I need to help get Runqiang going on to the next set of wet-bench experiments. Here is the criteria for prioritizing the hits:

Hits that affect Q3 to Q1,2,4 transition:

1. DESseq analysis (exp56\_de\_14\_1421352958-V2) and select hits (at gene level) as any gene with:
   1. 2 or more shRNAs have absolute Wald stat > 0.6 with the same sign
   2. 1 shRNA with Wald stat absolute value > 2.0 and has 2 or more additional shRNAs of any value with the same sign.
   3. Make a separate list of ambiguous hits in which 2 shRNAs meet the criteria in (a) but also have 2 shRNAs with the opposite sign.
   4. select negative control hits as genes in which all four shRNAs have Wald stat absolute values of <0.6
2. Quantile norm analysis (exp56\_de\_quantile\_norm\_11) and select hits (at gene level) as any gene with:
   1. 2 or more shRNAs with an absolute log2 fold change value  > 1.0 with the same sign
   2. 1 shRNA with absolute log2 fold change value > 4.0 and has 2 or more additional shRNAs of any value with the same sign.
   3. Make a separate list of ambiguous hits in which 2 shRNAs meet the criteria in (a) but also have 2 shRNAs with the opposite sign.
   4. select negative control hits as genes in which all four shRNAs have absolute log2 fold change values of < 1.0
3. Select hit list based on GSEA analysis as any gene with an Enrichment score, absolute value, of > 0.75.

Hits that affect Q4 to Q1 differentially:

1. DESseq analysis (exp56\_de\_14\_1421352958-V2) and select hits (at gene level) as any gene with:
   1. 2 or more shRNAs have Wald stat > 0.6 absolute value with the same sign
   2. 1 shRNA with Wald stat > 2.0 absolute value and has 2 or more additional shRNAs with the same sign, of any value.
   3. Make a separate list of “ambiguous hits” in which 2 shRNAs meet the criteria in (a) but also have 2 shRNAs with Wald stat > 0.6 of the opposite sign.
   4. select negative control hits as genes in which all four shRNAs have Wald stat absolute values of <0.6
2. Use your Quantile norm analysis (exp56\_de\_quantile\_norm\_11) and select hits (at gene level) as any gene with:
   1. 2 or more shRNAs with an log2 fold change > 1.0 absolute value, with the same sign
   2. 1 shRNA with log2 fold change > 4.0 absolute value and has 2 or more additional shRNAs with the same sign, of any value.
   3. Make a separate list of ambiguous hits in which 2 shRNAs meet the criteria in (a) but also have 2 shRNAs with the opposite sign.
   4. select negative control hits as genes in which all four shRNAs have absolute log2 fold change values of < 1.0
3. Select hit list based on GSEA analysis as any gene with an Enrichment score, absolute value, of > 0.75.

Once you have the lists, I would like to assess overlap between all three approaches at the gene level. Which genes did all three methods classify as hits? So I imagine a Venn Diagram for the three approaches, and then use a Fisher’s exact to test significance of the overlap. I would do this for genes that positively correlate with Q1 and those that positively correlate with Q4; use the same approach for the Q3 and Q1,2,4 analyses.

DEseq

QN

GSEA