DGE:

1. Th17 vs Naive
2. (HAR Th17 5h vs naive) vs (Th17 5h vs naive)

and then

      1. Do GSEA using the hallmark and c7 (immunological signatures) on the DE genes.

Overlap:

* Th17 vs naïve
* harmine vs naïve
* harmine vs Th17

Scatterplots:

* log2 FC in (Th17 vs naïve) against log2 FC in (harmine vs naïve).

4-way DGE is Bernard’s main concern.

Figure that communicates that a subset of the Th17 genes are dysregulated by harmine (I envision first focusing on genes altered by Th17, that are less/conversely regulated with harmine, but I’m not sure whether to include upfront genes that were NOT altered by Th17, but are not altered with harmine, which a simple differential-of-differential may not discriminate between?) and then some pathway analysis of those genes

Q1: I already looked at the subset with |FC|>1.5 in the NoHAR vs naive comparison - this is an easy thing to do in Excel since you included all pertinent data (thanks!), so no need to spend your time on it. This brought the list of counter-regulated genes to a much smaller set of 31 which is a reasonable start, but perhaps not the end of the analysis. This brought up the classic concern of how we think about small changes that may/may not be relevant, hence thinking about the best way to ID genes not regulated in opposite directions.

Q2: Yes - I think I broadly agree with the plan. I’ve definitely seen this done before. My naive question was the following. Since the plot is logFC vs logFC, this does not include probablility/variation between samples. In other words, if I think about the plot as a clock, points in 3-6 o’clock and 9-12 o’clock are counter-regulated. A point far out (highly expressed) at say 2.30 is very tempting to call and would be called by using a fitted parallel line, but if the position at 2.30 was because of 1-2 “off” samples in one group would this be better reflected by using the DEG comparison?

I don’t really know the answer, so this is really a question to learn about best practices. To be honest, I think the emerging theme is that we are going to ID that a harmine only dysregulates a subset of Th17 genes, so I’d like to use best practice to ID that subset and move on. Ie a descriptive answer.

If I understand correctly, and I’m trying to figure out what the figure looks like.

We could have a logic plot that says

“Th17 vs naive” = XX DEGs (Th17 signature)

Of these XX DEGs, search for overlap with HAR vs NOHAR @ 5h DEGs

YY counterregulated (chi sq p value) -> list + GSEA

This could also be illustrated as highlighted points on the scatterplot you’ve already generated

Run GSEA for c7 modules

-limit to just the 5h timepoint for now.

Annotate scatterplot with

-Har vs NoHar DGE

-top 50 off-diags

Venn

-(Th17 vs Naïve) vs (Har vs NoHar @ 5h)

List of counter-regulated genes, GSEA.

Enrichr

-counter-regulated genes

-differently-regulated genes

Trp metabolism pathway in Th17/Treg

-Kegg?

clusterProfiler dotplot