

Homework on differential expression

Kasper D. Hansen

Fall 2022

The data We will use data from Luperchio, Boukas et al “Leveraging the Mendelian disorders of the epigenetic machinery to systematically map functional epigenetic variation” (eLife, 2021). DOI

We will study part of the experiment where the authors compare wild-type mice to mice with a loss-of-function variant in the gene CREBBP causing Rubenstein-Taybi’s syndrome (RT or RT1).

The count data is available as an Rda file. Load it in R using

```
{r, eval=FALSE load("rubstein.rda")
```

There is data from both B-cells and T-cells and this is indicated by sample name like 8_B_CBP is sample 8 from B cells and CBP (Rubenstein-Taybi).

1. Examine the data. Subset the data to only contain the B-cells.
2. Prepare for analysis by limma-voom. Make a `DGEList` object using the function `SE2DGEList()`
3. Make a PCA or MDS plot of the data. Interpret it.
4. Filter the data for expression level.
5. Set up a design matrix.
6. Perform differential expression (model mean-variance relationship, do variance shrinkage)
7. Assess the p-value distribution
8. Interpret the results