Q1. What can we find from the expression dataset comparing treatment and genotypes, and combined effect (G + tx + Gxtx)

(ILS+EtOH)

(ILS+Saline)

(ISS+EtOH)

(ISS+Saline)

Linear regression comparisons between these.

Q2. Variance is biological and technical.

Don't expect a high technical variance

{3 mice in a blender -> RNA -> library -> (flowcell1, flowcell2, possibly flowcell3)} == (9 flowcells encompassing the 4 different treatment+genotype sets) x 3 replicates

Is there variance on a gene-by-gene basis between the technical replicates?

Account for 'noise' in the input datasets.

\*notes\*

- RNASeq :

expression <- all reads mapped over a gene (approach 1, count based)

expression <- all reads over gene / (length of gene x avg reads per million) (approach 2, fpkm)

How do approach 1 and 2 differ / change output

Feature reduction?

Q3. Recombinant inbreds

(ISS x ILS)

RI1...RI30

Expect a spectrum of relationships between these samples. Are there any patterns that can be derived from exploring this dataset? For example, correlation between phenotype and genotype?

\*Pre-processing already done\*

Quality trimming

Mapping (Tophat 2)

Read counting using HTC (ignoring reads that map to multiple positions)

STEPS

1) Data preprocessing – **Jasleen, Phil, Shams (*Wendy, Rachelle*).**

**IMPORTANT: Understanding this section is vital before continuing with the next steps!!! So even if you’re not one of the primaries for this section, you must keep in mind the preprocessing that has been done, and/or can be done (even more helpful!), to improve/respond to downstream analysis.**

\*Possible avenues for pre-processing\*

Size factor correction (total mass, median, quartile)

FPKM

LFC transform -> Normalize

Hierarchical clustering

2) Variance -

- Feature reduction (Gene set size reduction) – **Rachelle, Jasleen**

- (co)-variance between technical replicates\*\* - **Wendy**

3) Differential Expression (Apply to both part 1 and part 3)

- Two way ANOVA in parents (ILS, ISS) - **Phil**

- K-means or k-medoids clustering - **Jasleen, Shams**

- DeSeq

- EdgeR

- Limma

DUTIES

Code review : Shams

Proposal writing : JaWSPR

Poster making : JaWSPR

\*\* Not a true technical replicate set. Must note this in the proposal/writeup.

Poisson distribution of RNA-seq data: https://www.biostars.org/p/84445/

DESeq and edgeR