RNA-seq and Adaptive Shrinkage

David Gerard and Mengyin Lu

September, 2016

Department of Human Genetics
University of Chicago
Boss: Matthew Stephens

RNA-seq

What is RNA-seq?

1. Each cell in our body contains a population of RNA fragments.
2. These RNA fragments are the output ("expression") of genes.
3. RNA-seq uses next-gen sequencing to measure the relative expression of genes within a sample.
4. Relative differences in the expression levels of these RNA fragments between groups tell us interesting things:
   - 4.1 Which genes are influenced by a drug?
   - 4.2 How do cancer patients differ from non-cancer patients?
   - 4.3 Which genes are important in liver tissue versus muscle tissue?

Many Problems

1. Count Data: A new type of data that requires new normalization procedures/pipelines.
2. Small Sample Sizes: Low power. Hard to detect differences between groups.
3. Unobserved Variables: Can ruin analyses. Can make results uninterpretable.
   - 3.1 Which lab/technician processed a sample?
   - 3.2 The ancestry of the sample.
   - 3.3 Subject attributes such as age or sex.

Our Work

1. Count Data: Develop optimal pipelines for analysis.
   - 1.1 Which published methods work in practice?
   - 1.2 Develop benchmarks to compare pipelines.
2. Small Sample Sizes:
   - 2.1 Harness the power of ASH.
   - 2.2 Borrow strength between variables.
3. Unobserved Variables:
   - 3.1 Integrate "factor-augmented regression" models with the shrinkage ideas from ASH.
   - 3.2 Modify known approaches to optimize the performance of ASH.