Reading Assignment 1

Anish Shah

January 26, 2019

Question 1

Read the article entitled "Micro array analysis and tumor classification" by John Quackenbush, available on the Canvas.

- Describe briefly the usefulness of this new technology in the scientific context. Specifically, what scientific questions have been addressed using microarray technology?
    - This technology seems to focus on the genetic phenotype over a clinical stage, which helps to classify subtypes of cancer. It also could allow better and more targeted therapy at appropriate subgroups. The classification of peripheral versus germinal B cell malignancy is an example of this. Golub et al found a difference in profiles of ALL and AML.
- What are the goals of these experiments in the statistical context? State how these goals are different from the goals of classical studies (i.e., clinical trials).
    - Hierarchical clustering and k-means clustering are two common examples. They also have to handle repeat hypothesis testing.
- State the names of common bioinformatics tools that are used in analyzing the data specified in this article.
    - The goals are to assess high dimensional data, and to identify features or clusters that are somehow important.

Question 2

Read the article "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling" by Alizadeh et. al available on the Canvas.

- State the scientific question of interest.

  - Globally, they asked is there a distinct molecular subtype of DLBCL that associates with certain worse phenotypes of NHL? If so, are there specific gene expressions that are associated with subtypes of NHL?
  - Specifically, they examined whether they could generate a molecular portrait of subtypes of DLBCL, identify distinct types DLBCL, and find correlations with pathophysiology of DLBCL development, all using gene expression profiling.

- Describe the structure of the data.

  - They created DNA microarrays ("lymphochip") that had genes preferentially expressed in lymphoid cells and genes that were known to be important in cancer. They selected ~18k cDNA segments. This led to ~1.8 million gene expression measurements in 96 samples using the lymphochips.

- State the bioinformatics tool used in this article.

  - They used hierarchical clustering algorithm to group genes on basis of similarity in the patterns of expression. This was applied to both axes using weighted pair-group methods with centroid average.

- Describe briefly the statistical analysis.

  - They statistically analyzed differences in outcomes based on their clustering of B-cell gene expression types. They examined this using survival data analysis.

- Describe briefly the main results.

  - They found that germinal centre B-like DLBCL had better survival than those with B-like DLBCL (an additiona gene expression type).

- Propose a future study to confirm these results.

  - For the future, I would do a prospective cohort analysis grouping cohorts based on these subtypes.