CSCI 5465

Lab#7

Quincy Gu

17/11/2019

**Part I. Understanding the breast cancer study**

**Questions:**

**a). What was the purpose of each study and what experiments did the author do to answer the question they were interested in?**

For the study titled “Epigenetic portraits of human breast cancers (expression data)”, the purpose of this study is to apply the gene expressions to the different ER status of the breast cancer studies to better understand the breast cancer heterogeneity and improve the tumor taxonomy. As stated in their study page, the author used the “Infinium Methylation Platform to profile single-CpG resolution the methylomes of 119 breast tumors” and indicated that the methylation patterns may reflect the cellular origins of tumors.

For the study titled “Deep Sequence Analysis of the Relationship between Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells”, the purpose of this study is to identify differentially expressed genes that are correlated with methylation or copy number changes, and investigate how these could affect the transcriptome landscape of ER positive or negative breast cancer cells and tumors. As stated in their study web page, the author used the “deep sequencinge technology to profile the transcriptome, gene copy number, and CpG island methylation stats simultaneously in eight commonly used breast cell lines to develop a model for how these genomic features are integrated in ER positive and negative breast cancer”, and the author “sequenced 8 commonly used breast cancer cell lines for mRNA exoression, CpG methylation and DNA copy number” for this study.

**b). How many gene expression samples are in each dataset?**

There are 90 gene expression samples in the study titled with “Epigenetic portraits of human breast cancers (expression data)”.

There are 24 gene expression samples in the study titled with “Deep Sequence Analysis of the Relationship between Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells”.

**c). What two sets of tumor samples did the authors profile for gene expression?**

In the study titled with “Epigenetic portraits of human breast cancers (expression data)”, the author used the Infinium Methylation Platform to profile at single CpG resolution (over 14,000 genes interrogated) the methylomes of 119 breast tumors.

In the study titled with “Deep Sequence Analysis of the Relationship between Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells”, there are 8 commonly used breast cancer cell lines were sequenced for mRNA expression, CpG methylation and DNA copy number.

**d). How large are the raw data files for each study?**

In the study titled with “Epigenetic portraits of human breast cancers (expression data)”, the size of the raw data files is 399.8 Mb.

In the study titled with “Deep Sequence Analysis of the Relationship between Gene Expression, CpG Island Methylation, and Gene Copy Number in Breast Cancer Cells”, the size of the raw data files is 55.2 Gbp, 35.2 Gb.