

Cancer Genomics - Analysis Plan

Courtney Vaughn, Isaac Robson, Kentaro Hoffman, and John Sperger

March 19th, 2018

0.1 Introduction

Serous Ovarian Cancer is serious.

Past work by The Cancer Genome Atlas Research Network “delineated four ovarian cancer transcriptional subtypes, three microRNA subtypes, four promoter methylation subtypes and a transcriptional signature associated with survival duration” [3].

The clinical data includes both overall survival and progression-free survival.

As an introduction, the data we have relates to patient response to Platinum-based chemotherapies (survival time with treatment, but all patients are treated basically the same way, longer survival time means the patients responded better to the drug). Platinums are some of the most commonly used chemotherapies and are used on most solid tumors. They are unfortunately very toxic though and can cause kidney failure, deafness, etc... Platinum based chemotherapies work by damaging DNA. When DNA is replicated, the replication machinery can recognize damage and will signal to the cell that there is stress so the cell goes into programmed cell death (apoptosis). The damage can also be repaired or it can be ignored and the DNA can be replicated over it (often leading to mutations). DNA is double stranded and made up of nucleic acids (also referred to as bases), most platinum induced damage links two bases on the same strand of DNA (intrastrand crosslinks); this damage can be repaired by a process known as Nucleotide Excision Repair. A small amount of the damage (about 1%) links bases between two strands (interstrand crosslinks). There is a lot of talk in the field about how the latter damage may be the more toxic damage. Interstrand crosslinks are probably repaired using double strand break methods. As of right now, the only definite known mechanism of resistance to platinums are drug transporters (less drug in cell=less damage= less cell death) so whichever question we choose we should probably control for this somehow. I'll talk to Katie about it once we pick a question.

0.2 Data & Pre-processing

Restrict the sample to only those with clinical data

There are 21,000 unique genes expressed in the RNA-Seq data. Looking at a few other papers, it seems like they come up with a candidate list of say 500

Data Type	Cases
Clinical	587
DNA Methylation	602
Copy Number Variation	573
RNA-Seq	376

Table 1: TCGA Ovarian Data Summary

to a few thousand and then run clustering with that candidate list instead of all the genes.

0.3 Planned Analysis 1 - Hierarchical Clustering

First going to do hierarchical clustering with the (? DNA Methylation, RNA-seq?) data [2]

Purpose: Sanity check ourselves and make sure we can recreate the clusters in the Nature paper [3]

0.4 Planned Analysis 2 - Mining differential correlation

Differential Correlation Mining (DCM) is a method for identifying sets of variables where the average pairwise correlation between variables in a set is higher under one sample condition than the other developed by Kelly Bodwin, Kai Zhang, and Andrew Nobel [1]. We will split the DNA Methylation data into clusters based on the subgroups identified in Analysis 1 and run DCM to identify sets of differentially correlated genes across subtypes. After identifying these sets of differentially correlated genes, Would be neat to see if the gene sets identified by running DCM on the methylation data can predict differences in mRNA expression level between subtypes. My very hazy understanding and comically simplified understanding is DNA Methylation -> mRNA -> protein stuff

Limitation: We don't have healthy cell data, and so DCM can only tell us new information about already discovered clusters. Ideally, we'd have cases and controls and run DCM on case vs control to try and see if there are different potential clusterings.

0.5 How does expression of DNA damage response factors influence survival?

For this, we can make clusters based on RNA levels of genes involved in DNA damage response and then look to see if there is a difference in survival between our clusters to determine if there is a role for repair/ other damage response pathways in drug effectiveness. If there is, then we can follow up to see which repair pathways are most important to drug response. This is an interesting

question since people often assume repair is a mechanism of resistance to platinum but there really isn't any strong evidence for this.

0.6 Does the immune system play a role in response to platinum? and if so, which immune pathways improve/reduce survival?

We know that there are synergistic effects between platinum and immune inhibitors but the relationship between these has not been fleshed out fully. We can cluster samples based on immune response related genes and then see if there is a difference in survival based on these. The follow-up we can look at which pathways may be associated with decreased response to platinum and see if there are drugs that modulate those. The original TCGA paper did define a subtype of ovarian cancer as "immune responsive" so we'll have to make sure we aren't just recreating that subtype.

0.7 Can we reproduce microRNA signature as predictors of platinum response?

There have been a number of microRNAs implicated in platinum resistance and a few microRNA signatures for predicting response to platinum in other cancers. We can use these lists to see if similar microRNA signatures could predict survival in ovarian as well. (Lots of people are interested in microRNAs right now so understanding their role in treatment response is a hot topic now. Given that they already looked for which microRNAs correlated with survival in the original ovarian cancer TCGA paper, I'm not sure if we can actually find anything new with this, but I wanted to put it on the list since it is a hot topic)

References

- [1] Kelly Bodwin, Kai Zhang, and Andrew Nobel. A testing-based approach to the discovery of differentially correlated variable sets, 2016.
- [2] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863–14868, December 1998.
- [3] Cancer Genome Atlas Research Network et al. Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353):609, 2011.