Online tools to analyze TCGA data

Rehan Akbani, Ph.D.
UT MD Anderson Cancer Center
Some Common Analyses Types

• Subtype discovery
  – mRNA, miRNA, DNA methylation, Protein, Copy number
• Outcome prediction
• Pathway alterations
• Copy number profiles
• Cross-tumor comparisons
• Many others
• E.g. TCGA Ovarian Cancer paper

“Integrated genomic analyses of ovarian carcinoma”
Subtype discovery

1) Four subtypes found based on mRNA expression

2) Three miRNA based subtypes found
Outcome prediction

1) Outcome differences in TCGA + 3 other data sets using mRNA expression

2) Outcome differences in miRNA subtypes
Pathway alterations

1) Alterations in the RB and PI3K/RAS signaling pathways

2) BRCA1/2 alterations in the HR pathway
1) Significantly amplified (red) and deleted (blue) chromosome arms
Cross-tumor comparisons

1) Ovarian cancer vs. GBM copy number profiles
Dichotomy observed

Kidney cancer (KIRC) DNA methylation data (27k)
After removing sex chromosomes, a new dichotomy based on batch ID appears.
After removing bad probes

Chromophobes
Demo and Examples
Introduction

• Go to http://cBioPortal.org/ by Memorial Sloan Kettering Cancer Center (MSKCC)
The cBio Cancer Genomics Portal provides visualization, analysis and download of large-scale cancer genomics data sets. The portal is developed and maintained by the Computational Biology Center at Memorial Sloan-Kettering Cancer Center.

Cancer Discovery. May 2012 2; 401. [Abstract]
Workshop Exercise

Problems:

1. In what percentage of patients with GBM is the RB pathway altered?

2. Which gene in the RB pathway is most frequently altered?

3. Which is the most common type of alteration; amplification, deletion or mutation?

4. Based on the data, is CDKN2A likely to be a tumor suppressor or oncogene? How about CDK4?
Steps

1. In item 1 select “Glioblastoma (TCGA)”

2. In item 2 select the following:

3. In item 3 select “All Complete Tumors (seq, mRNA, CNA)”
Steps

4. In item 4 select “Glioblastoma: RB Pathway (7 genes)” from the drop-down list below:

The list of 7 genes will automatically be filled in

5. Hit the “Submit” button
Result

OncoPrint

Case Set: All Complete Tumors: All tumor samples that have mRNA, CNA and sequencing data (122 samples)

Altered in 87 (71%) of cases.

Total 122 cases --> altered

CDKN2A  47%
CDKN2B  44%
CDKN2C  2%
CDK4    16%
CDK5    3%
CCND2   2%
RB1     10%

Amplification  Homozygous Deletion  Mutation
Copy number alterations are putative.

Move the mouse pointer over the OncoPrint below for more details about cases and alterations.
Exercise 1

Problem 1:

1. In what percentage of patients with GBM is the RB pathway altered?
2. Which gene in the RB pathway is most frequently altered?
3. Which is the most common type of alteration; amplification, deletion or mutation?
4. Based on the data, is CDKN2A likely to be a tumor suppressor or oncogene? How about CDK4?
Result

OncoPrint: Altered in 87 (71%) of cases.
Exercise 1

Problem 1:

1. In what percentage of patients with GBM is the RB pathway altered?
2. Which gene in the RB pathway is most frequently altered?
3. Which is the most common type of alteration; amplification, deletion or mutation?
4. Based on the data, is CDKN2A likely to be a tumor suppressor or oncogene? How about CDK4?
Result

OncoPrint: Case Set: All Complete Tumors: All tumor samples that have mRNA, CNA and sequencing data (122 samples)
Altered in 87 (71%) of cases.

CDKN2A 47%
CDKN2B 44%
CDK4 16%
CDK5 3%
CCND2 2%
Rb1 10%

Amplification  |  Homozygous Deletion  |  Mutation
Copy number alterations are putative.
Exercise 1

Problem 1:
1. In what percentage of patients with GBM is the RB pathway altered?
2. Which gene in the RB pathway is most frequently altered?
3. Which is the most common type of alteration; amplification, deletion or mutation?
4. Based on the data, is CDKN2A likely to be a tumor suppressor or oncogene? How about CDK4?
Result
Problem 1:

1. In what percentage of patients with GBM is the RB pathway altered?
2. Which gene in the RB pathway is most frequently altered?
3. Which is the most common type of alteration; amplification, deletion or mutation?
4. Based on the data, is CDKN2A likely to be a tumor suppressor or oncogene? How about CDK4?
Result

OncoPrint

Case Set: All Complete Tumors: All tumor samples that have mRNA, CNA and sequencing data (122 samples)

Altered in 87 (71%) of cases.

Total 122 cases --> altered

- CDKN2A 47%
- CDKN2B 44%
- CDKN2C 3%
- CDK4 16%
- CDK5 3%
- CCND2 2%
- RB1 10%

Move the mouse pointer over the OncoPrint below for more details about cases and alterations.
Exercise

• Is there a statistically significant difference in overall and progression free survival between RB pathway altered vs. non-altered patients?
Gene Set / Pathway is altered in 75.2% of all cases.

Glioblastoma (TCGA, Nature 2008)/All Complete Tumors: (206)/Glioblastoma: RB Pathway (7 genes)/7 genes

Modify Query

Overall Survival, Logrank Test P-Value: 0.03556 [PDF]

Disease Free Survival, Logrank Test P-Value: 0.017291 [PDF]
Exercise

• Which genes are altered mutually exclusively?
Gene Set / Pathway is altered in 75.2% of all cases.

Glioblastoma (TCGA, Nature 2008)/All Complete Tumors: (206)/Glioblastoma: RB Pathway (7 genes)/7 genes

- **Modify Query**

<table>
<thead>
<tr>
<th>Gene</th>
<th>CDKN2B</th>
<th>CDKN2A</th>
<th>CDK4</th>
<th>RB1</th>
<th>CDKN2C</th>
<th>CCND2</th>
<th>CDK6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN2B</td>
<td>---</td>
<td>0</td>
<td>0</td>
<td>0.021944</td>
<td>0.058545</td>
<td>0.241304</td>
<td>0.572941</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>---</td>
<td></td>
<td>0.000006</td>
<td>0.023168</td>
<td>0.110337</td>
<td>0.348115</td>
<td>0.463442</td>
</tr>
<tr>
<td>CDK4</td>
<td>---</td>
<td></td>
<td></td>
<td>0.094052</td>
<td>0.686582</td>
<td>0.009202</td>
<td>0.052594</td>
</tr>
<tr>
<td>RB1</td>
<td>---</td>
<td></td>
<td></td>
<td>0.540223</td>
<td>0.262674</td>
<td>0.796155</td>
<td></td>
</tr>
<tr>
<td>CDKN2C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.852462</td>
<td>0.887434</td>
<td></td>
</tr>
<tr>
<td>CCND2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.057403</td>
<td></td>
</tr>
<tr>
<td>CDK6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p-values <0.05, as derived via Fisher’s Exact test are outlined in red. p-values are not adjusted for FDR.

**Legend**
- strong tendency toward mutual exclusivity
- some tendency toward mutual exclusivity
Exercise

• Find out which gene is most commonly mutated in breast cancer

• https://confluence.broadinstitute.org/display/GDAC/Home

• Find the distribution of correlations between DNA methylation and mRNA in BRCA