

**Hands-on exercise on integrative analysis using cBioportal  
BTEP Data Integration Workshop  
June 2, 2015**

Prepared by: Parthav Jailwala, CCBR

---

**Step 1 :** Open the cBioportal webpage (<http://cbioportal.org>)

**Step 2 :** Type 'glioblastoma' in the search box for 'Select Cancer Study' and click on the 'Glioblastoma (TCGA, Cell 2013)' sample collection

**Query** **Download Data**

**Select Cancer Study:**

glioblastoma 1 study selected. Deselect all

- All Search Results (3)
  - Brain CNS (3)
    - Diffuse Glioma (3)
      - Glioblastoma (3)
        - Glioblastoma Multiforme (3)
          - Glioblastoma Multiforme (TCGA, Provisional)
          - Glioblastoma (TCGA, Nature 2008)
          - Glioblastoma (TCGA, Cell 2013)

**Step 3:** Select Mutations, Putative copy-number alterations from GISTIC and mRNA Expression z-Scores under 'Select Genomic Profiles'

**Select Genomic Profiles:**

- Mutations
- Putative copy-number alterations from GISTIC
- mRNA Expression z-Scores (RNA Seq V2 RSEM)
  - Enter a z-score threshold z: 2.0
- protein/phosphoprotein level (RPPA)

**Step 4:** Select 'All Complete Tumors (291)' under 'Select Patient/Case Set'. Then in the box for 'Enter Gene Set', type in these 6 genes in order : IDH1 EGFR PTEN CDKN2A CDK4 RB1 and click 'Submit' at the bottom of the page

**Select Patient/Case Set:** All Complete Tumors (291)

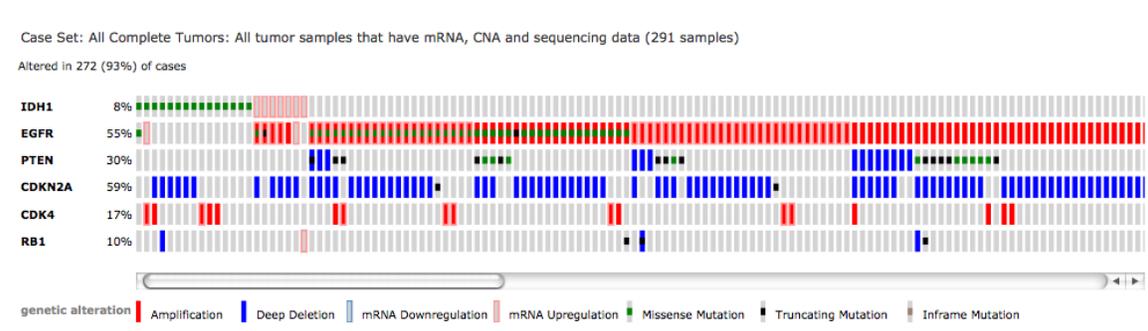
**Enter Gene Set:** Advanced: Onco Query Language (OQL)

User-defined List

IDH1 EGFR PTEN CDKN2A CDK4 RB1

All gene symbols are valid.

**Step 5:** You will now see a new page with 11 tabs as shown in this snapshot, with the first ‘OncoPrint’ tab visible. There is one row for each gene and the columns are the samples. Using glyphs and color-codes, we can now visually inspect various alterations (mRNA expression, Copy-number alteration and somatic mutations) across these genes.



From the OncoPrint, 271 cases (93%) have an alteration in at least one of the six genes, with the frequency of alteration in each of the 6 selected genes shown. For IDH1, we have missense mutations and a few mRNA upregulation events. For EGFR we have amplifications, alongwith missense mutations in the amplified samples. For PTEN, we have a mixture of deep deletions, missense mutations and truncating mutations. For *CDKN2A*, most of the alterations are deep deletions, and there are a few mutations. The alterations in *CDK4* are amplifications and mRNA upregulation. Events associated with *RB1* included a deletion, mRNA downregulation, mRNA upregulation and several truncating mutations. EGFR shows a lot of amplifications, with some mRNA upregulation as well as missense mutations. The alterations in CDKN2A, CDK4 and RB1 are distributed in a nearly mutually exclusive way across samples, which can be statistically analyzed and visualized with the Mutual Exclusivity tab.

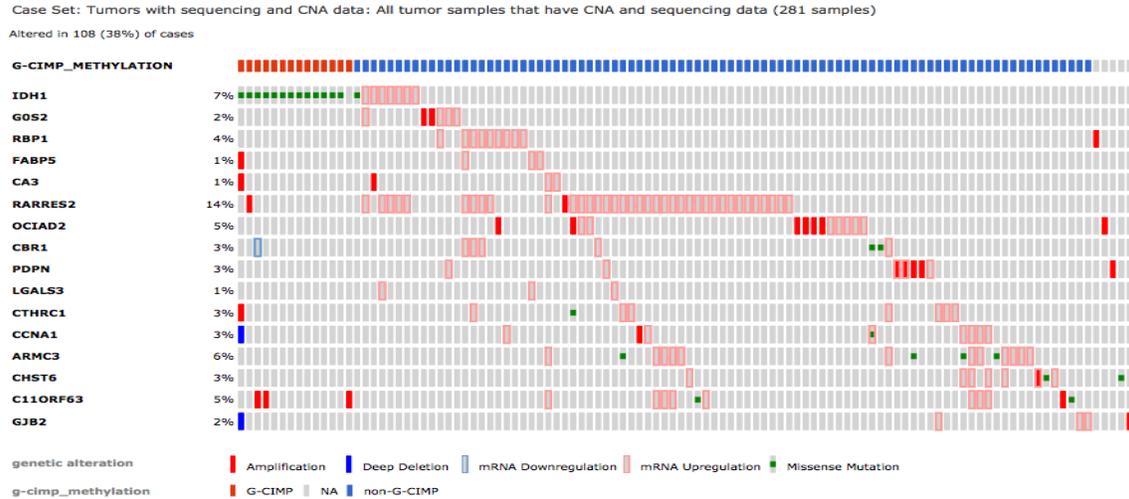
**Step 6:** Mouse-over just on the top-right of the gene-tracks to see the control panel. Select G-CIMP\_METHYLATION track and the ‘Expression subtype’ track using the ‘+’ drop-down list of clinical variables. Then Sort by ‘clinical data first’. Also click to show legend for clinical data, remove unaltered samples and remove whitespace between samples. You should get the view as shown below. We can now observe that all the samples with IDH1 mutation also harbor the G-CIMP methylator phenotype



**Step 7:** Now we will try to visualize the genes that show the hypermethylated G-CIMP phenotype. These are both downregulated at the mRNA expression level and hypermethylated. For this, I have extracted the top 15 genes from a table in a manuscript that established the G-CIMP methylated sub-group (Cancer Cell,

2010 May 18; 17(5): 510-522). We will input IDH1 and these 15 genes as the input gene-list for generating this Oncoprint.

There is a text file available that has the list of these 16 input gene-symbols. You can copy the gene-list and paste it into cBioportal. For reference, these genes are: IDH1 GOS2 RBP1 FABP5 CA3 RARRES2 OCIAD2 CBR1 PDPN LGALS3 CTHRC1 CCNA1 ARMC3 CHST6 C11orf63 GJB2  
 In the Oncoprint, select the ‘G-CIMP\_METHYLATION’ track to be overlaid with the gene-tracks. We can now see that all the IDH1 mutant samples and the G-CIMP positive samples do not show any mRNA upregulation of these 15 genes, whereas the G-CIMP negative samples do show some mRNA up-regulation.



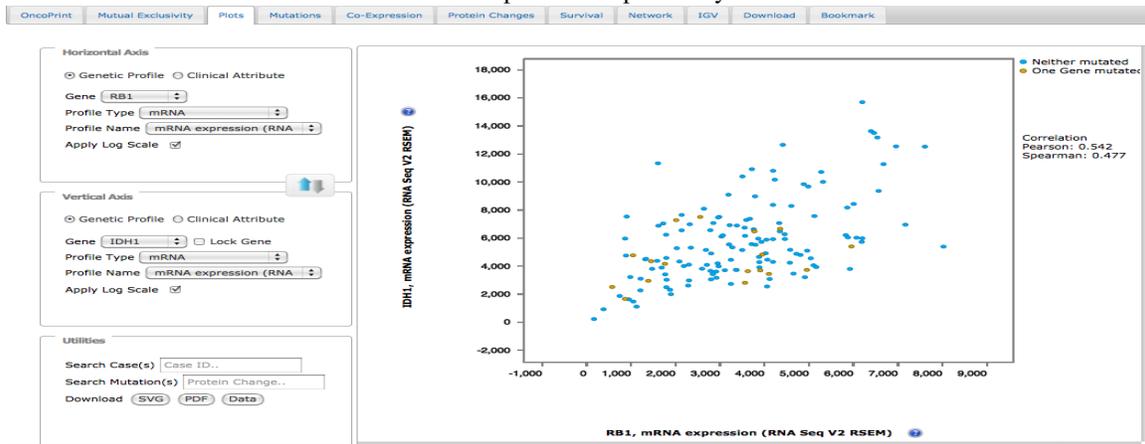
**Step 8:** For this step, we go back to the 6 genes as input (IDH1 EGFR PTEN CDKN2A CDK4 RB1 ). Click on the ‘Mutual Exclusivity’ tab. This will show 11 gene-pairs with mutually exclusive alterations, of which 5 are significant; 4 gene-pairs with co-occurring alterations, of which 1 is significant.

The query contains **11** gene pairs with mutually exclusive alterations (**5** significant), and **4** gene pairs with co-occurrent alterations (**1** significant).

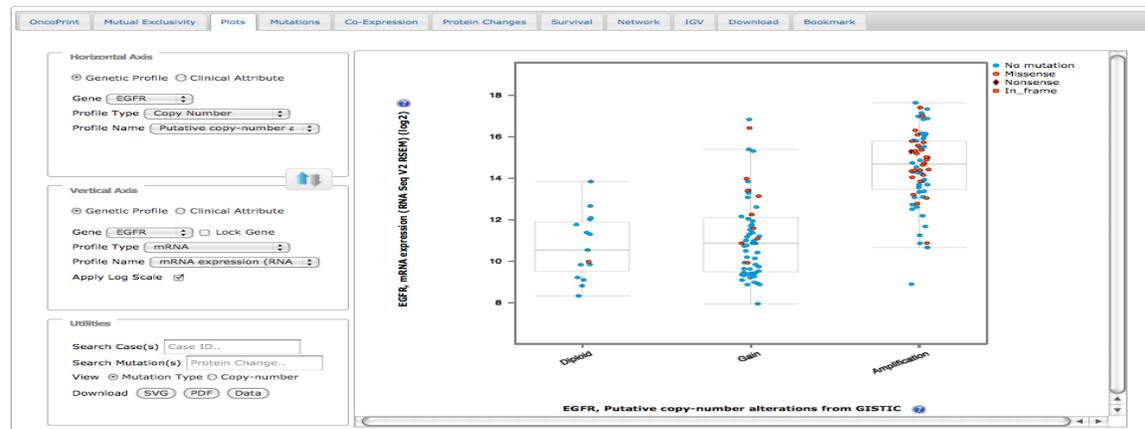
<input checked="" type="checkbox"/> Mutual exclusivity <input checked="" type="checkbox"/> Co-occurrence <input type="checkbox"/> Significant pairs		Search Gene		
Gene A	Gene B	p-Value	Log Odds Ratio	Association
IDH1	PTEN	<0.001	<-3	Tendency towards mutual exclusivity <b>Significant</b>
EGFR	RB1	<0.001	-1.661	Tendency towards mutual exclusivity <b>Significant</b>
CDKN2A	CDK4	<0.001	-2.067	Tendency towards mutual exclusivity <b>Significant</b>
CDKN2A	RB1	<0.001	-1.620	Tendency towards mutual exclusivity <b>Significant</b>
CDK4	RB1	0.004	<-3	Tendency towards mutual exclusivity <b>Significant</b>
EGFR	CDKN2A	0.007	0.625	Tendency towards co-occurrence <b>Significant</b>
IDH1	EGFR	0.051	-0.836	Tendency towards mutual exclusivity
EGFR	PTEN	0.141	-0.308	Tendency towards mutual exclusivity
PTEN	CDKN2A	0.183	0.272	Tendency towards co-occurrence
PTEN	RB1	0.188	0.447	Tendency towards co-occurrence
IDH1	CDKN2A	0.247	-0.399	Tendency towards mutual exclusivity
IDH1	CDK4	0.304	0.408	Tendency towards co-occurrence
EGFR	CDK4	0.422	-0.110	Tendency towards mutual exclusivity
PTEN	CDK4	0.463	-0.096	Tendency towards mutual exclusivity
IDH1	RB1	0.644	-0.068	Tendency towards mutual exclusivity

Using the same query used for describing OncoPrints, the mutual exclusivity analysis shows that events in these 6 genes tended to occur in a mutually exclusive way, and the pattern was statistically significant (for IDH1 & PTEN, EGFR & RB1, *CDKN2A* and *CDK4*, for *CDKN2A* and *RB1*, and for *CDK4* and *RB1*). The exclusivity between *CDKN2A*, *RB1* and *CDK4* fits with what is known about RB signaling in GBM, which can be deactivated by inactivation of *RB1* itself (through mutation or deletion), by activation of *CDK4* (a CDK that inhibits *RB1* activity) through amplification, or by inactivation of the CDK inhibitor p16, which is encoded by *CDKN2A*, through deletion or mutation. Thus, a single alteration in one of these genes is sufficient to deactivate the pathway, and this is what the mutual exclusivity analysis showed. Further, there is strong evidence that *IDH1* mutations confer the G-CIMP methylation phenotype and in almost all cases where *IDH1* is mutated, *EGFR* has a normal copy-number.

**Step 9:** On the 'Plots' tab, select *RB1*-mRNA on the Horizontal axis and *IDH1*-mRNA on the Vertical Axis. We can observe that *RB1* and *IDH1* mRNA expression is positively correlated



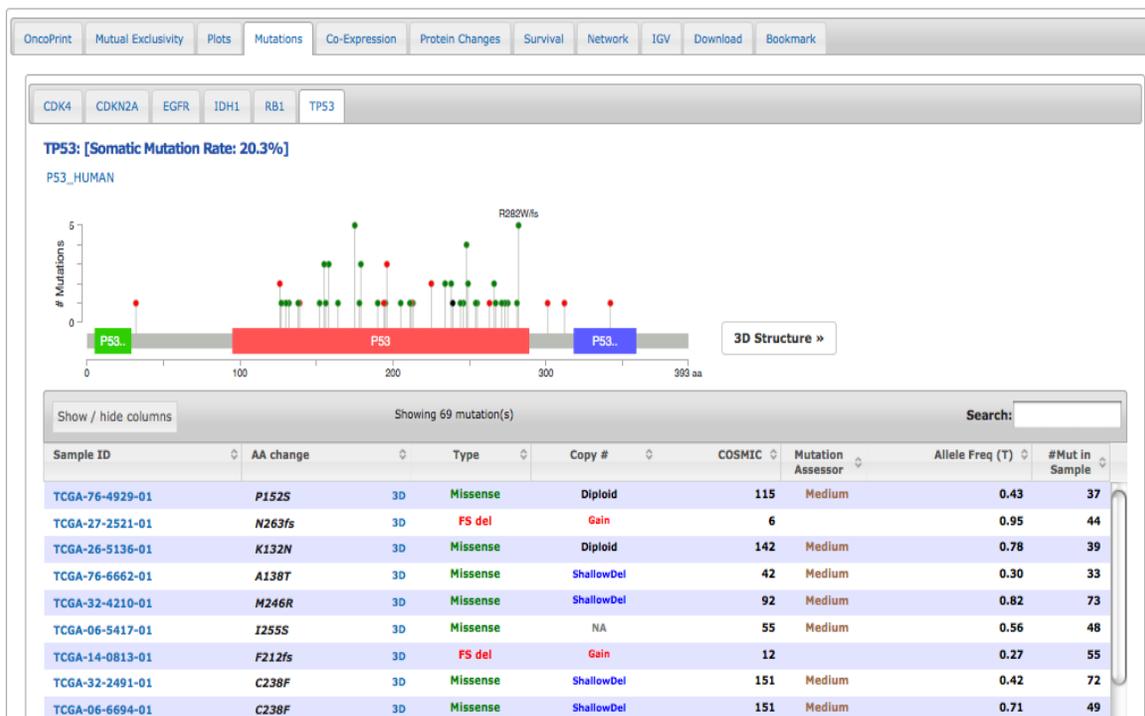
**Step 10:** Select *EGFR*-Copy number on Horizontal axis and *EGFR*-mRNA on Vertical axis. We can see the positive correlation between *EGFR* amplification and mRNA expression. Also, samples having *EGFR* amplification and high mRNA expression also tend to have missense mutations



**Step 11:** Select EGFR-Copy number on Horizontal axis and CDKN2A-Copy number on Vertical axis. This shows a 4x4 table with the number of samples for each of the 16 combination of events between these two genes. Interestingly, we now observe that when EGFR is amplified or has a gain, CDKN2A tends to have deep deletion event



**Step 12:** On the 'Mutations' tab, you can visualize the mutations on each of the 6 query genes within important functional domains across the gene, via 'Lollipop plots'. For example, this view of TP53 mutations show a number of missense mutations (green) as well as deletion events (red).



**Step 13:** On the 'Network' tab, we can visualize an interaction network connecting the 6 query genes (dark red circles with black outline) as well as other genes having interactions with these genes. A number of layout, overlay & filtering options allow us to customize this network and zoom in into the interactions of interest.

