1. **更改genome ref(illumina hg38 => UCSC直接下載hg38)**

Illumina whole genome seq (2015/8/18建立後未更新)

Illumina genes annotation(2015/8/18建立後未更新)

UCSC whole genome seq (上次更新2014/1/27)

UCSC genes annotation (上次更新2017/5/6)

1. **genome seq**

illumina hg38 whole genome seq 與 UCSC hg38 whole genome seq 完全相同

1. **annotation**

**優:**

\* UCSC較頻繁更新Annotation

\* Illumina整理之UCSC hg38 : 1065949筆註解

UCSC直接下載之hg38 : 1142786 筆註解

(+76837)

**缺:**

\* UCSC下載之hg38 annotation無gene name

\* 一個isoform(對到同個gene)出現在多個chromosome則RSEM無法運作



情況一 : 一個isoform 在同個 chromosome有不同tss\_id

情況二 : 一個isoform在不同chromosome 有不同tss\_id

(一個gene有多個tss)

\* 需整合UCSC、NCBI、ensembl，仿照illumina格式製作annotation

1. 至ncbiRefSeqCurated.txt 搜尋gene name

2.

a. 透過cuffdiff (compare the transcriptomes assembled)計算transcription start site(tss\_id)

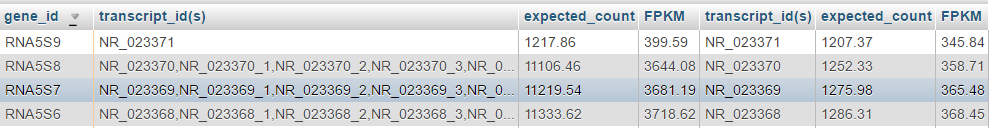
b. 由ensembl下載tss

=>篩選gene(gene\_id同時存在HGNC與NCBI)

=>ensembl transcript id 轉換 NCBI transcript id

3. 透過tss\_id將相同isoform進行區分(\_1,\_2,\_3…)

**解決 : 使用Illumina(UCSC、NCBI、ensembl) or ensembl annotation**



1. **TCGA**

TCGA僅提供DATA下載，若要進行分析須至相同機構的外部網站

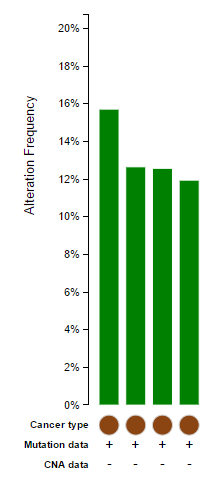
**Analysis**

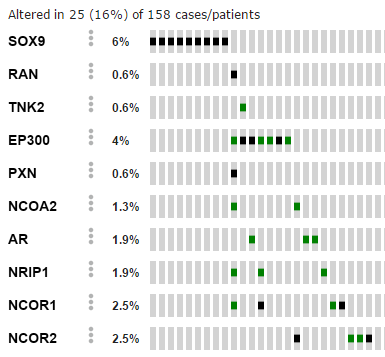
1. cBioPortal

cBioPortal : 25個部位 33種癌症

(TCGA : 29個部位 39種癌症(TCGA : 33 , TARGET : 6) )

**Overview:**





Alteration Frequency = (gene發生突變的case數/總case數)

**Mutations:**

每個基因在哪個sample發生突變的詳細資訊

**Expression:**

**Download:**

網站相關資訊

**Bookmark:**

**OncoPrint**

每個基因在case/patient 是否突變

**Mutual Exclusivity(所選的studies中x個sample中突變之間互相比)**

先將基因分群，再兩兩比較為排斥或相關(<)

(相同基因在normal和tumor的相關程度)

Mutual exclusivity : 相互排斥

Co-occurrence : 相互相關

p-value:宣稱此關係為正確的錯誤機率

Odds ratio(勝算比) : 此事件發生機率/此事件不發生機率

(兩個基因的改變程度)

Log odds ratio > 0   : Association towards co-occurrence  
Log odds ratio <= 0 : Association towards mutual exclusivity  
p-Value < 0.05         : Significant association

(OR:在相同model之間進行統計測試)

(LR:在不同model之間進行統計測試)

**Plots**

**Mutations**

每個geme在哪個case突變的詳細資訊

**Enrichments(所選的studies中x個sample中每個gene的資訊)**

Gene :所選的studies中x個sample中的基因

Cytoband : 染色體區段

Percentage of alteration : 此基因在突變group與非突變group比例

Log Ratio : Log2 based ratio of (pct in altered / pct in unaltered)

p-value : 宣稱此事件為正確的錯誤機率

q-value : 宣稱此事件為錯誤的正確機率

Direction/Tendency :

Log ratio > 0   : Co-occurrence  
Log ratio <= 0 : Mutual exclusivity  
p-Value < 0.05         : Significant association

**Download**

每個基因突變率下載

**Bookmark**

1. Firehose

**Missense mutation**

This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.

**Nonsense mutation**

A nonsense mutation is also a change in one DNA base pair. Instead of substituting one amino acid for another, however, the altered DNA sequence prematurely signals the cell to stop building a protein. This type of mutation results in a shortened protein that may function improperly or not at all.

\* Illumina之hg38 annotation有325個isoform(4162筆註解)在NCBI genbank不存在(移除or更新or被取代)

(Ex. NM\_183245，NM\_002702，NR\_051962，NR\_033992)

UCSC之hg38 annotation有186個isoform(2586筆註解)在NCBI genbank不存在(移除or更新or被取代)

(ex. NM\_004745，NM\_001302797，NR\_075100，NR\_033971)