使用TCGAbiolinks下载TCGA中结直肠癌数据

Liu huisheng

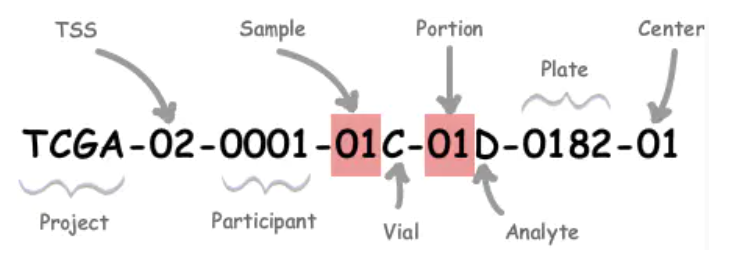
2020/9/23

TCGAbiolinks能够通过调用NCI（美国国家癌症研究所）GDC（基因组数据共享）的API来进行搜索，下载和准备相关数据以便用于下游分析（差异基因分析、甲基化区域分析）以及可视化（生存图、火山图、Starburst plot ）。

TCGA数据包括两个来源，一是，GDC Legacy Archive ：之前存储在GDC的数据，主要参考 hg19 and hg18；二是，GDC harmonized database：GDC标准协调后的数据，主要参考 hg38。

TCGA样本使用TCGA barcode 进行标识。Barcode由一串标识符组成，每一个标识符代表TCGA数据的一个元素。例如：

* Aliquot barcode: TCGA-G4-6317-02A-11D-2064-05,一串完整的TCGA barcode；
* Participant: TCGA-G4-6317，病人号;
* Sample: TCGA-G4-6317-02，样本号, 各标识符，具体含义如图1，具体可见此[链接](https://www.jianshu.com/p/5df2a498fc9a)。



**图1: TCGA barcode**

# Required libraries  
library(TCGAbiolinks)  
library(dplyr)  
library(DT)

## 参考资料

1. 官方文档：<https://www.bioconductor.org/packages/3.4/bioc/vignettes/TCGAbiolinks/inst/doc/query.html#useful_information>
2. <https://www.omicsclass.com/article/1060>
3. <https://www.omicsclass.com/article/1125>
4. Barcode:<https://www.jianshu.com/p/5df2a498fc9a>