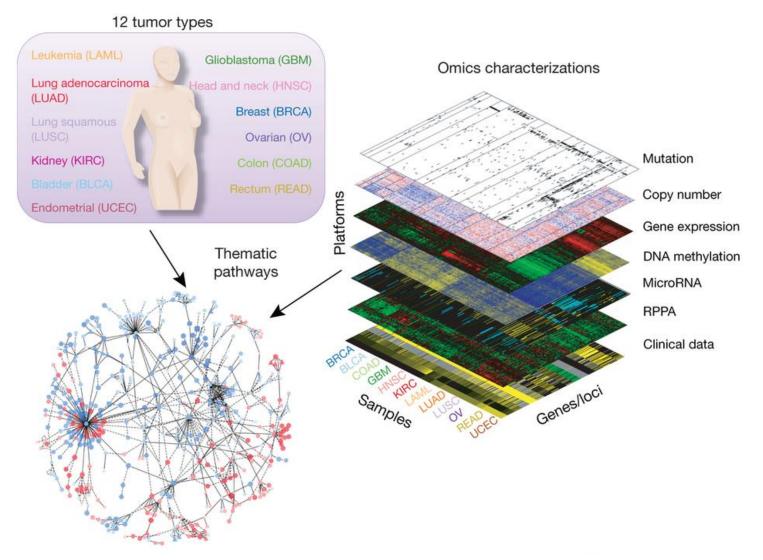
# The DDRD assay in lung cancer

Dr Reza Rafiee
Queen's University Belfast,
Centre for Cancer Research and Cell Biology
June 2017

# Part 1 RNA-Seq gene expression

## The Cancer Genome Atlas (TCGA)



 The TCGA Data Portal is no longer operational and all TCGA data now resides at the <u>Genomic Data Commons</u>.

## Genomic Data Commons (GDC)

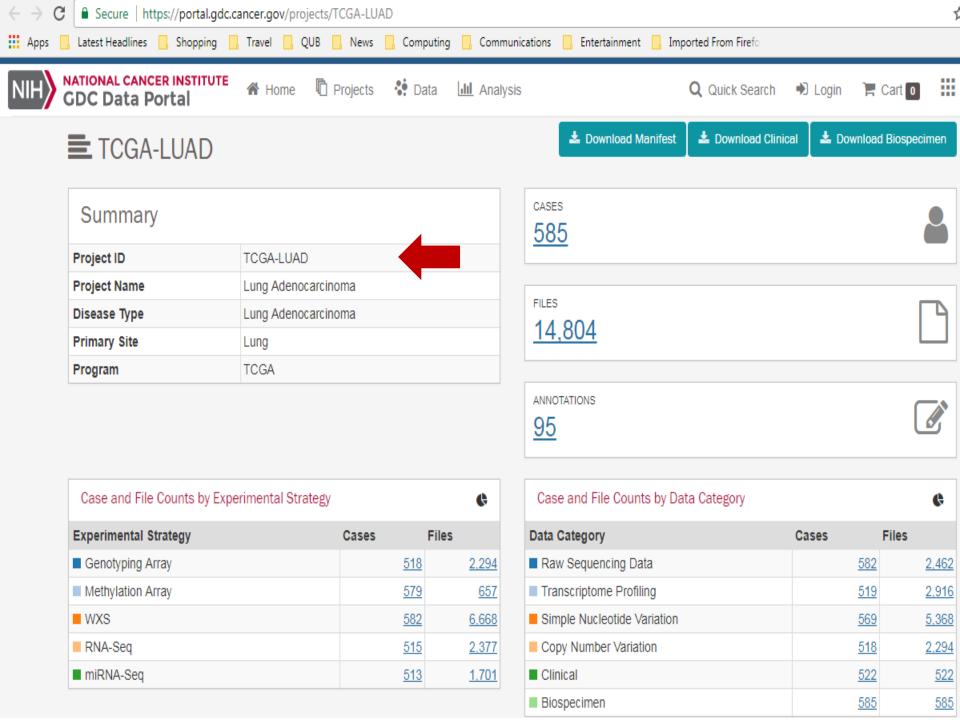
 The National Cancer Institute's (NCI's) Genomic Data Commons (GDC) is a data sharing platform that promotes precision medicine in oncology. It is not just a database or a tool; it is an expandable knowledge network supporting the import and standardization of genomic and clinical data from cancer research programs.

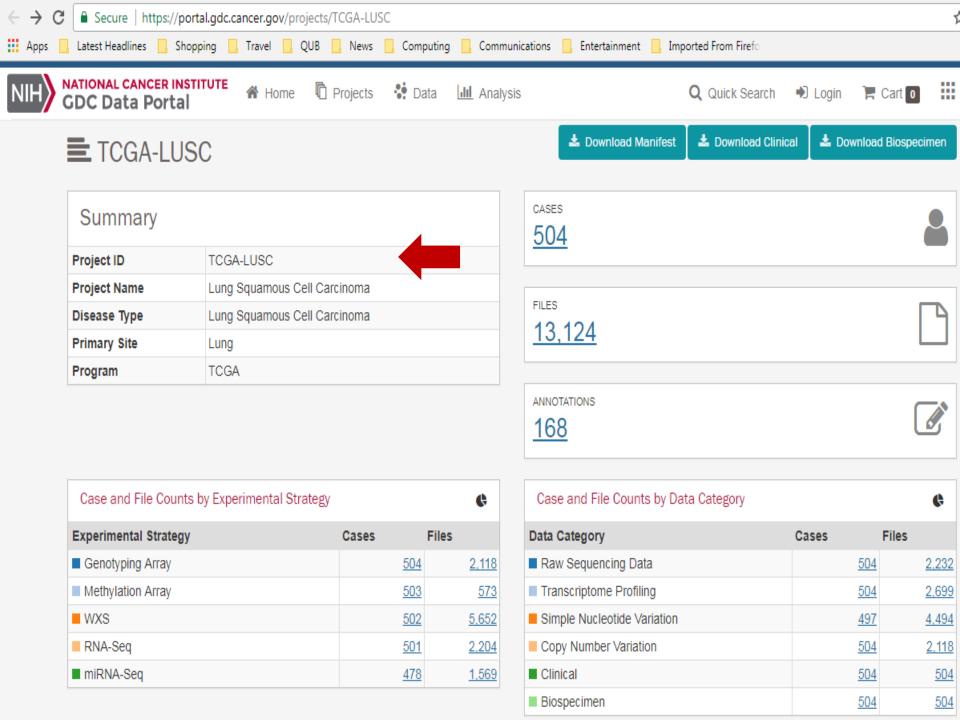
#### How to retrieve data from GDC

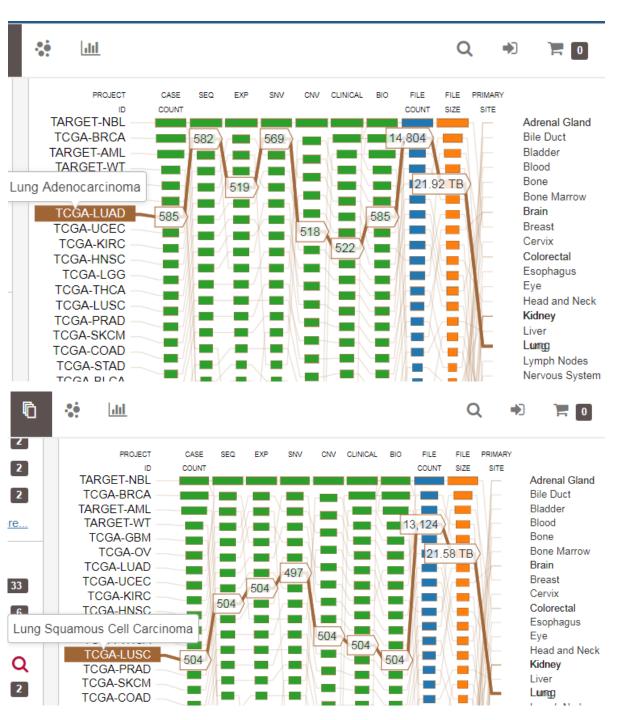
- "GenomicDataCommons" and "TCGAbiolinks" libraries
  - 1) facilitate the TCGA open-access data retrieval,
  - 2) prepare the data using the appropriate <u>pre-processing</u> <u>strategies</u>,
  - and allow the user <u>to download</u> a specific version of the data and thus to easily reproduce earlier research results. In more detail, the package provides multiple methods for <u>analysis</u> (e.g., <u>differential expression analysis</u>, <u>identifying differentially methylated regions</u>) and methods for <u>visualization</u> (e.g., survival plots, volcano plots, starburst plots) in order <u>to develop complete analysis pipelines</u>.

#### Data.type parameters

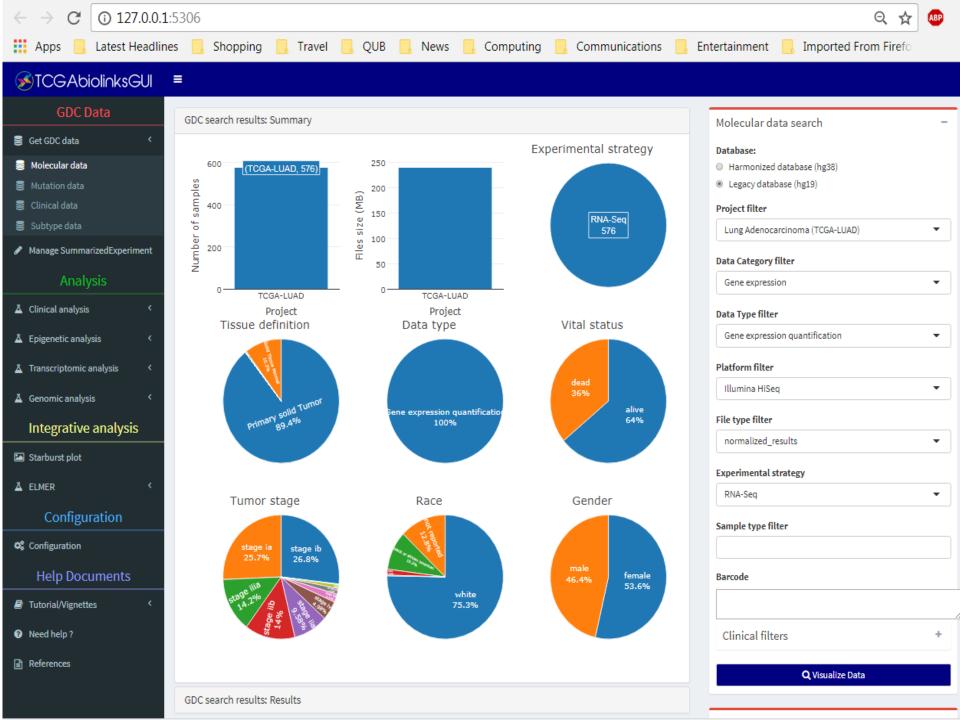
```
sort(legacy.data.type) |
                                                        |Genotypes |
                                                        Intensities
ABI sequence trace |
                                                        Intensities Log2Ratio
| Aligned reads |
                                                        | Isoform expression quantification |
| Auxiliary test |
                                                        |LOH |
|Biospecimen data |
                                                        | Methylation array QC metrics |
|Biospecimen Supplement |
                                                        | Methylation beta value |
|Bisulfite sequence alignment |
                                                        | Methylation percentage |
                                                        | Microsatellite instability |
CGH array QC |
                                                        |miRNA gene quantification |
|Clinical data |
                                                        |miRNA isoform quantification |
| Clinical Supplement |
                                                        |Normalized copy numbers |
Copy number estimate
                                                        |Normalized intensities |
Copy number QC metrics |
                                                        |Pathology report |
Copy number segmentation |
                                                        | Protein expression quantification |
                                                        |Raw intensities |
|Copy number variation |
                                                        |Sequencing tag |
|Coverage WIG |
                                                        |Sequencing tag counts |
| Diagnostic image |
                                                        |Simple nucleotide variation |
Exon junction quantification
                                                        |Simple somatic mutation |
|Exon quantification |
                                                        |Structural variation |
Gene expression quantification |
                                                        |Tissue slide image |
                                                        |Unaligned reads |
```

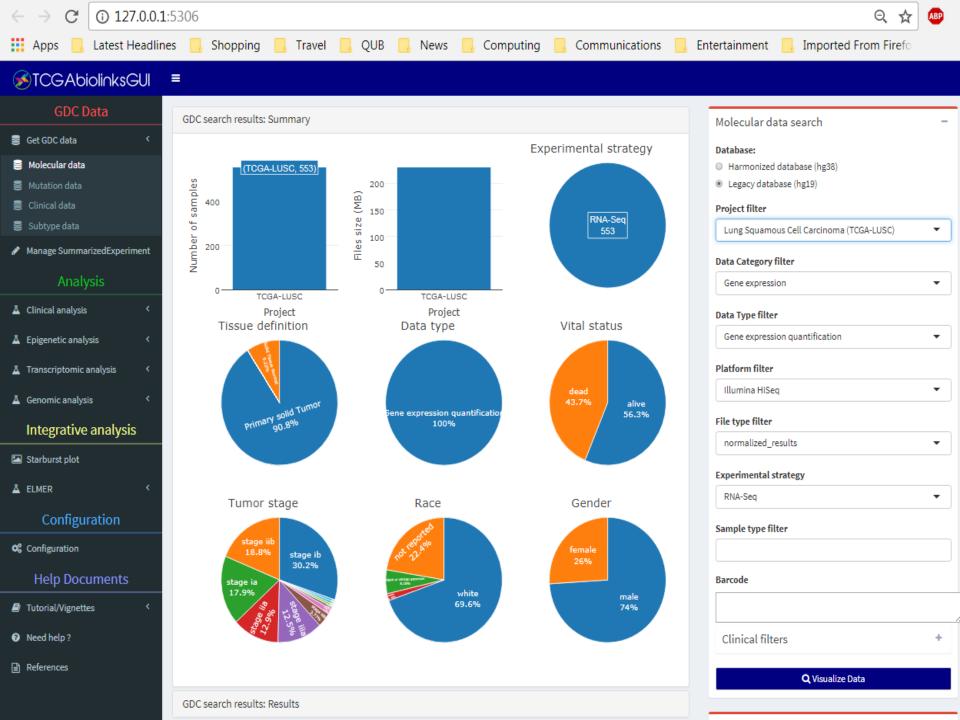






#### **GDC** portal





#### Summary: gene expression

(normalized based on FPKM)

# of samples	LUAD	LUSC
GDC portal	585	504
Downloaded (me)	576	553
Almac (Provided clinical & DDRD scores)	517	501

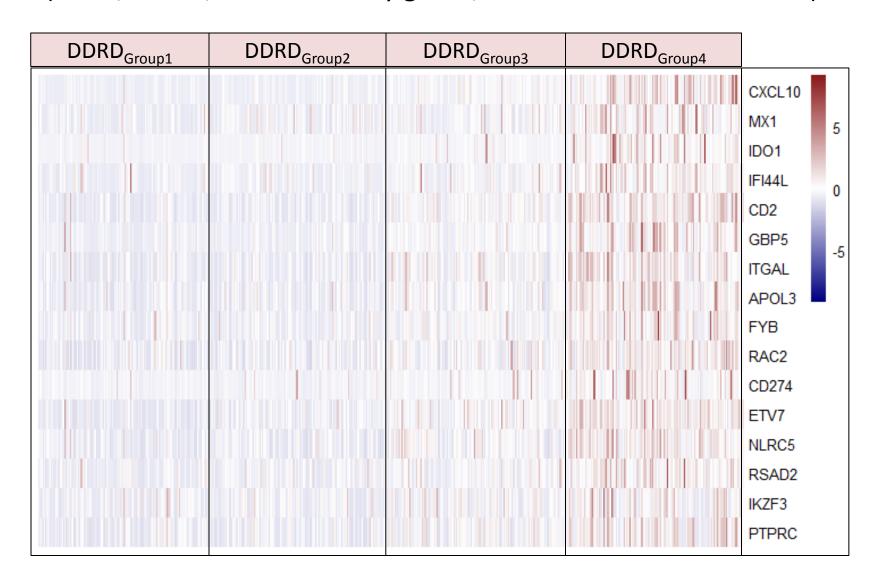
project = "TCGA-LUAD/LUSC",
legacy = TRUE,
data.category = "Gene expression",
data.type = "Gene expression quantification",
platform = "Illumina HiSeq",
file.type = "normalized\_results",
experimental.strategy = "RNA-Seq"



3 genes of DDRD assay are NOT available in the gene list of samples: AC138128.1, OR2I1P and AL137218.1

#### RNA-Seq gene expression

(LUAD, n=517, 16 DDRD assay genes, normalized based on FPKM)



#### RNA-Seq gene expression

(LUAD, n=517, 25 DDRD assay genes, normalized based on FPKM)

