

# FINAL ASSIGNMENT

FELIPE NARVAEZ

2023-11-10

```
## ----message=FALSE, warning=FALSE, include=FALSE-----  
-----
```

```
library(TCGAbiolinks)  
library(SummarizedExperiment)
```

```
## Loading required package: MatrixGenerics  
## Loading required package: matrixStats  
## Warning: package 'matrixStats' was built under R version 4.3.2  
##
```

```
## Attaching package: 'MatrixGenerics'
```

```
## The following objects are masked from 'package:matrixStats':
```

```
##  
##      colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,  
##      colCounts, colCummaxs, colCummins, colCumprods, colCumsums,  
##      colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,  
##      colMads, colMaxs, colMeans2, colMedians, colMins,  
colOrderStats,  
##      colProds, colQuantiles, colRanges, colRanks, colSdDiffs,  
colSds,  
##      colSums2, colTabulates, colVarDiffs, colVars,  
colWeightedMads,  
##      colWeightedMeans, colWeightedMedians, colWeightedSds,  
##      colWeightedVars, rowAlls, rowAnyNAs, rowAnys,  
rowAvgsPerColSet,  
##      rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,  
##      rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,  
##      rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians,  
rowMins,  
##      rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,  
##      rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs,  
rowVars,  
##      rowWeightedMads, rowWeightedMeans, rowWeightedMedians,  
##      rowWeightedSds, rowWeightedVars
```

```
## Loading required package: GenomicRanges
## Loading required package: stats4
## Loading required package: BiocGenerics
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
```

```
##
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
##      anyDuplicated, aperm, append, as.data.frame, basename,
cbind,
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter,
Find,
##      get, grep, grepl, intersect, is.unsorted, lapply, Map,
mapply,
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff,
sort,
##      table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Loading required package: S4Vectors
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:utils':
```

```
##
##      findMatches
```

```
## The following objects are masked from 'package:base':
```

```
##
##      expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##
```

```
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':
```

```
##  
## windows
```

```
## Loading required package: GenomeInfoDb  
## Warning: package 'GenomeInfoDb' was built under R version 4.3.2  
## Loading required package: Biobase  
## Welcome to Bioconductor
```

```
##  
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages  
## 'citation("pkgname")'.
```

```
##
```

```
## Attaching package: 'Biobase'
```

```
## The following object is masked from 'package:MatrixGenerics':
```

```
##  
## rowMedians
```

```
## The following objects are masked from 'package:matrixStats':
```

```
##  
## anyMissing, rowMedians
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:Biobase':
```

```
##  
## combine
```

```
## The following objects are masked from 'package:GenomicRanges':
```

```
##  
## intersect, setdiff, union
```

```
## The following object is masked from 'package:GenomeInfoDb':
```

```
##
## intersect
```

```
## The following objects are masked from 'package:IRanges':
```

```
##
## collapse, desc, intersect, setdiff, slice, union
```

```
## The following objects are masked from 'package:S4Vectors':
```

```
##
## first, intersect, rename, setdiff, setequal, union
```

```
## The following objects are masked from 'package:BiocGenerics':
```

```
##
## combine, intersect, setdiff, union
```

```
## The following object is masked from 'package:matrixStats':
```

```
##
## count
```

```
## The following objects are masked from 'package:stats':
```

```
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
## intersect, setdiff, setequal, union
```

```
library(DT)
```

```
## Warning: package 'DT' was built under R version 4.3.2
```

```
## ----BCR_Biotab, results='hide', echo=TRUE, message=FALSE,
warning=FALSE-----
```

```
query <- GDCquery(
  project = "TCGA-ACC",
  data.category = "Clinical",
  data.type = "Clinical Supplement",
  data.format = "BCR Biotab"
)
```

```
## -----
```

```
## o GDCquery: Searching in GDC database
```

```
## -----
## Genome of reference: hg38
## -----
## oo Accessing GDC. This might take a while...
## -----
## ooo Project: TCGA-ACC
## -----
## oo Filtering results
## -----
## ooo By data.format
## ooo By data.type
## -----
## oo Checking data
## -----
## ooo Checking if there are duplicated cases
## Warning: There are more than one file for the same case. Please
verify query results. You can use the command
View(getResults(query)) in rstudio
## ooo Checking if there are results for the query
## -----
## o Preparing output
## -----
```

```
GDCdownload(query)
```

```
## Downloading data for project TCGA-ACC
## Of the 7 files for download 7 already exist.
## All samples have been already downloaded
```

```
clinical.BCRtab.all <- GDCprepare(query)
```

```
##
```

```
|
|
| 0%
|
```

```

|=====
| 14%
|
|=====
| 29%
|
|=====
| 43%
|
|=====
| 57%
|
|=====
| 71%
|
|=====
| 86%

```

```
## New names:
```

```
## • `metastatic_tumor_site` -> `metastatic_tumor_site...38`
## • `metastatic_tumor_site` -> `metastatic_tumor_site...39`
```

```
##
```

```
|
```

```

|=====
====| 100%

```

```
names(clinical.BCRtab.all)
```

```
## [1] "clinical_nte_acc"          "clinical_omf_v4.0_acc"
## [3] "clinical_drug_acc"         "clinical_radiation_acc"
## [5] "clinical_follow_up_v4.0_acc"
      "clinical_follow_up_v4.0_nte_acc"
## [7] "clinical_patient_acc"
```

```
## ----echo=TRUE, message=FALSE, warning=FALSE-----
-----
```

```

clinical.BCRtab.all$clinical_drug_acc %>%
  head %>%
  DT::datatable(options = list(scrollIX = TRUE, keys = TRUE))

```

```
## ----results = "hide",cache=TRUE, message=FALSE-----  
-----
```

```
library(TCGAbiolinks)  
query <- GDCquery(  
  project = "TCGA-ACC",  
  data.category = "Clinical",  
  data.type = "Clinical Supplement",  
  data.format = "BCR Biotab"  
)
```

```
## -----  
## o GDCquery: Searching in GDC database  
## -----  
## Genome of reference: hg38  
## -----  
## oo Accessing GDC. This might take a while...  
## -----  
## ooo Project: TCGA-ACC  
## -----  
## oo Filtering results  
## -----  
## ooo By data.format  
## ooo By data.type  
## -----  
## oo Checking data  
## -----  
## ooo Checking if there are duplicated cases  
## Warning: There are more than one file for the same case. Please  
verify query results. You can use the command  
View(getResults(query)) in rstudio  
## ooo Checking if there are results for the query  
## -----  
## o Preparing output  
## -----
```

```
GDCdownload(query)
```

```
## Downloading data for project TCGA-ACC
```

```
## Of the 7 files for download 7 already exist.
```

```
## All samples have been already downloaded
```

```
clinical_tab_all <- GDCprepare(query)
```

```
##
```

```
|
|
| 0%
|
|=====
| 14%
|
|=====
| 29%
|
|=====
| 43%
|
|=====
| 57%
|
|=====
| 71%
|
|=====
| 86%
```

```
## New names:
```

```
## • `metastatic_tumor_site` -> `metastatic_tumor_site...38`
```

```
## • `metastatic_tumor_site` -> `metastatic_tumor_site...39`
```

```
##
```

```
|
|=====
====| 100%
```

```
## -----
-----
```

```
# All available tables
```



```
names(clinical_tab_all)
```

```
## [1] "clinical_nte_acc" "clinical_omf_v4.0_acc"
## [3] "clinical_drug_acc" "clinical_radiation_acc"
## [5] "clinical_follow_up_v4.0_acc"
"clinical_follow_up_v4.0_nte_acc"
## [7] "clinical_patient_acc"
```

```
# columns from clinical_patient
```

```
dplyr::glimpse(clinical_tab_all$clinical_patient_acc)
```

```
## Rows: 94
```

```
## Columns: 84
## $ bcr_patient_uuid <chr>
"bcr_patient_uuid", "CDE..."
## $ bcr_patient_barcode <chr>
"bcr_patient_barcode", "..."
## $ form_completion_date <chr>
"form_completion_date", ...
## $ prospective_collection <chr>
"tissue_prospective_coll..."
## $ retrospective_collection <chr>
"tissue_retrospective_co..."
## $ gender <chr> "gender",
"CDE_ID:220060..."
## $ race <chr> "race",
"CDE_ID:2192199"..."
## $ ethnicity <chr>
"ethnicity", "CDE_ID:219..."
## $ history_other_malignancy <chr>
"other_dx", "CDE_ID:3382..."
## $ history_neoadjuvant_treatment <chr>
"history_of_neoadjuvant_..."
## $ tumor_status <chr>
"person_neoplasm_cancer_..."
## $ vital_status <chr>
"vital_status", "CDE_ID:..."
## $ radiation_treatment_adjuvant <chr>
"radiation_therapy", "CD..."
## $ pharmaceutical_tx_adjuvant <chr>
"postoperative_rx_tx", "..."
## $ pharm_tx_mitotane_indicator <chr>
"mitotane_therapy", "CDE..."
```

```

## $ pharm_tx_mitotane_adjuvant <chr>
"mitotane_therapy_adjuva...
## $ pharm_tx_mitotane_therapeutic_levels <chr>
"therapeutic_mitotane_le...
## $ pharm_tx_mitotane_therapeutic_at_rec <chr>
"therapeutic_mitotane_lv...
## $ pharm_tx_mitotane_for_macro_disease <chr>
"mitotane_therapy_for_ma...
## $ pharm_tx_mitotane_therapeutic_macro <chr>
"therapeutic_mitotane_lv...
## $ pharm_tx_mitotane_therapeutic_at_prog <chr>
"therapeutic_mitotane_lv...
## $ clinical_status_within_3_mths_surgery <chr>
"post_surgical_procedure...
## $ treatment_outcome_first_course <chr>
"primary_therapy_outcome...
## $ laterality <chr>
"laterality", "CDE_ID:82...
## $ histologic_diagnosis <chr>
"histological_type", "CD...
## $ initial_pathologic_dx_year <chr>
"year_of_initial_patholo...
## $ ct_scan_preop_indicator <chr> "ct_scan",
"CDE_ID:35348...
## $ ct_scan_preop_results <chr>
"ct_scan_findings", "CDE...
## $ lymph_nodes_examined <chr>
"primary_lymph_node_pres...
## $ lymph_nodes_examined_count <chr>
"lymph_node_examined_cou...
## $ lymph_nodes_examined_he_count <chr>
"number_of_lymphnodes_po...
## $ weiss_score_overall <chr>
"weiss_score", "CDE_ID:3...
## $ mitoses_per_50_hpf <chr>
"mitoses_count", "CDE_ID...
## $ ajcc_pathologic_tumor_stage <chr>
"pathologic_stage", "CDE...
## $ residual_tumor <chr>
"residual_tumor", "CDE_I...
## $ metastatic_dx_confirmed_by <chr>
"metastatic_neoplasm_con...
## $ metastatic_dx_confirmed_by_other <chr>
"metastatic_neoplasm_con...
## $ metastatic_tumor_site...38 <chr>
"metastatic_neoplasm_ini...

```

```

## $ metastatic_tumor_site...39 <chr>
"distant_metastasis_anat...
## $ history_adrenal_hormone_excess <chr>
"excess_adrenal_hormone_...
## $ history_basis_adrenal_hormone_dx <chr>
"excess_adrenal_hormone_...
## $ molecular_studies_others_performed <chr>
"germline_testing_perfor...
## $ new_tumor_event_dx_indicator <chr>
"new_tumor_event_after_i...
## $ age_at_initial_pathologic_diagnosis <chr>
"age_at_initial_patholog...
## $ atypical_mitotic_figures <chr>
"atypical_mitotic_figure...
## $ clinical_M <chr>
"clinical_M", "CDE_ID:34...
## $ clinical_N <chr>
"clinical_N", "CDE_ID:34...
## $ clinical_T <chr>
"clinical_T", "CDE_ID:34...
## $ clinical_stage <chr>
"clinical_stage", "CDE_I...
## $ cytoplasm_presence_less_than_equal_25_percent <chr>
"cytoplasm_presence_less...
## $ days_to_birth <chr>
"days_to_birth", "CDE_ID...
## $ days_to_death <chr>
"days_to_death", "CDE_ID...
## $ days_to_initial_pathologic_diagnosis <chr>
"days_to_initial_patholo...
## $ days_to_last_followup <chr>
"days_to_last_followup",...
## $ diffuse_architecture <chr>
"diffuse_architecture", ...
## $ disease_code <chr>
"disease_code", "CDE_ID:...
## $ extranodal_involvement <chr>
"extranodal_involvement"...
## $ icd_10 <chr> "icd_10",
"CDE_ID:322628...
## $ icd_o_3_histology <chr>
"icd_o_3_histology", "CD...
## $ icd_o_3_site <chr>
"icd_o_3_site", "CDE_ID:...
## $ informed_consent_verified <chr>
"informed_consent_verifi...

```

## \$ invasion_of_tumor_capsule	<chr>
"invasion_of_tumor_capsu...	
## \$ mitotic_rate	<chr>
"mitotic_rate", "CDE_ID:...	
## \$ necrosis	<chr>
"necrosis", "CDE_ID:3648...	
## \$ nuclear_grade_III_IV	<chr>
"nuclear_grade_III_IV", ...	
## \$ pathologic_M	<chr>
"pathologic_M", "CDE_ID:...	
## \$ pathologic_N	<chr>
"pathologic_N", "CDE_ID:...	
## \$ pathologic_T	<chr>
"pathologic_T", "CDE_ID:...	
## \$ patient_id	<chr>
"patient_id", "CDE_ID:",...	
## \$ project_code	<chr>
"project_code", "CDE_ID:...	
## \$ ret	<chr> "ret",
"CDE_ID:3121628",...	
## \$ sdha	<chr> "sdha",
"CDE_ID:3121628"...	
## \$ sdhaf2_sdh5	<chr>
"sdhaf2_sdh5", "CDE_ID:3...	
## \$ sdhb	<chr> "sdhb",
"CDE_ID:3121628"...	
## \$ sdhc	<chr> "sdhc",
"CDE_ID:3121628"...	
## \$ sdhd	<chr> "sdhd",
"CDE_ID:3121628"...	
## \$ sinusoid_invasion	<chr>
"sinusoid_invasion", "CD...	
## \$ stage_other	<chr>
"stage_other", "CDE_ID:2...	
## \$ system_version	<chr>
"system_version", "CDE_I...	
## \$ tissue_source_site	<chr>
"tissue_source_site", "C...	
## \$ tmem127	<chr> "tmem127",
"CDE_ID:31216...	
## \$ tumor_tissue_site	<chr>
"tumor_tissue_site", "CD...	
## \$ vhl	<chr> "vhl",
"CDE_ID:3121628",...	
## \$ weiss_venous_invasion	<chr>
"weiss_venous_invasion",...	

```
## ----results = "hide",cache=TRUE, message=FALSE,warning=FALSE----  
-----
```

```
# Biospecimen BCR Biotab  
query_biospecimen <- GDCquery(  
  project = "TCGA-ACC",  
  data.category = "Biospecimen",  
  data.type = "Biospecimen Supplement",  
  data.format = "BCR Biotab"  
)
```

```
## -----  
## o GDCquery: Searching in GDC database  
## -----  
## Genome of reference: hg38  
## -----  
## oo Accessing GDC. This might take a while...  
## -----  
## ooo Project: TCGA-ACC  
## -----  
## oo Filtering results  
## -----  
## ooo By data.format  
## ooo By data.type  
## -----  
## oo Checking data  
## -----  
## ooo Checking if there are duplicated cases  
## Warning: There are more than one file for the same case. Please  
verify query results. You can use the command  
View(getResults(query)) in rstudio  
## ooo Checking if there are results for the query  
## -----  
## o Preparing output  
## -----
```

```
GDCdownload(query_biospecimen)
```

```
## Downloading data for project TCGA-ACC
```

```
## GDCdownload will download 10 files. A total of 527.41 KB
```

```
## Downloading as: Tue_Nov_21_18_06_21_2023.tar.gz
```

```
##
```

```
Downloading: 16 kB
```

```
Downloading: 16 kB
```

```
Downloading: 16 kB
```

```
Downloading: 16 kB
```

```
Downloading: 16 kB
```

```
Downloading: 16 kB
```

```
Downloading: 25 kB
```

```
Downloading: 25 kB
```

```
Downloading: 25 kB
```

```
Downloading: 25 kB
```

```
Downloading: 25 kB
```

```
Downloading: 25 kB
```

```
Downloading: 33 kB
```

```
Downloading: 33 kB
```

```
Downloading: 33 kB
```

```
Downloading: 33 kB
```

```
Downloading: 41 kB
```

```
Downloading: 41 kB
```

```
Downloading: 41 kB
```

```
Downloading: 41 kB
```

```
Downloading: 49 kB
```

```
Downloading: 49 kB
```

```
Downloading: 57 kB
```

```
Downloading: 57 kB
```

```
Downloading: 57 kB
```

```
Downloading: 57 kB
```

```
Downloading: 57 kB
```

```
Downloading: 57 kB
```

```
Downloading: 74 kB
```

```
Downloading: 74 kB
```

```
Downloading: 74 kB
```

```
Downloading: 74 kB
```

```
Downloading: 90 kB
```

```
Downloading: 90 kB
```

```
Downloading: 99 kB
```

```
Downloading: 99 kB
```

```
Downloading: 99 kB
```

```
Downloading: 99 kB
```

```
biospecimen_tab_all <- GDCprepare(query_biospecimen)
```

```
##
```

```
|
|
| 0%
|
|=====
| 10%
|
|=====
| 20%
|
|=====
| 30%
|
|=====
| 40%
|
|=====
| 50%
|
|=====
| 60%
|
|=====
| 70%
|
|=====
| 80%
|
|=====
| 90%
|
|=====
====| 100%
```

```
## -----
-----
```

```
# All available tables
names(biospecimen_tab_all)
```

```
## [1] "biospecimen_shipment_portion_acc" "ssf_tumor_samples_acc"
```

```
## [3] "biospecimen_portion_acc"
"biospecimen_analyte_acc"
## [5] "biospecimen_diagnostic_slides_acc"
"biospecimen_sample_acc"
## [7] "biospecimen_protocol_acc"
"biospecimen_aliquot_acc"
## [9] "ssf_normal_controls_acc"          "biospecimen_slide_acc"
```

```
biospecimen_tab_all$biospecimen_sample_acc %>%
```

```
head %>%
DT::datatable(options = list(scrollX = TRUE, keys = TRUE))
```

```
## ----results='hide', echo=TRUE, message=FALSE, warning=FALSE-----
-----
```

```
clinical <- GDCquery_clinic(project = "TCGA-ACC", type = "clinical")
```

```
## ----echo=TRUE, message=FALSE, warning=FALSE-----
```

```
clinical %>%
  head %>%
  DT::datatable(
    filter = 'top',
    options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
    rownames = FALSE
  )
```

```
## ----eval=FALSE,results='hide', echo=TRUE, message=FALSE,
warning=FALSE-----
```

```
# clinical_beataml <- GDCquery_clinic(
#   project = "BEATAML1.0-COHORT",
#   type = "clinical"
# )
#
# clinical_cptac2 <- GDCquery_clinic(
#   project = "CPTAC-2",
#   type = "clinical"
# )
#
# clinical_genie <- GDCquery_clinic(
#   project = "GENIE-MSK",
#   type = "clinical"
# )
```

```
## ----results = 'hide',echo=TRUE, message=FALSE, warning=FALSE-----
query <- GDCquery(
```



```
project = "TCGA-COAD",
data.category = "Clinical",
data.format = "bcr xml",
barcode = c("TCGA-RU-A8FL","TCGA-AA-3972")
)
```

```
## -----
## o GDCquery: Searching in GDC database
## -----
## Genome of reference: hg38
## -----
## oo Accessing GDC. This might take a while...
## -----
## ooo Project: TCGA-COAD
## -----
## oo Filtering results
## -----
## ooo By data.format
## ooo By barcode
## -----
## oo Checking data
## -----
## ooo Checking if there are duplicated cases
## ooo Checking if there are results for the query
## -----
## o Preparing output
## -----
```

```
GDCdownload(query)
```

```
## Downloading data for project TCGA-COAD
## GDCdownload will download 2 files. A total of 129.627 KB
## Downloading as: Tue_Nov_21_18_07_23_2023.tar.gz
##
```

```
Downloading: 7.8 kB
Downloading: 7.8 kB
Downloading: 7.8 kB
Downloading: 7.8 kB
Downloading: 7.8 kB
Downloading: 7.8 kB
Downloading: 12 kB
Downloading: 12 kB
Downloading: 12 kB
Downloading: 12 kB
```

```
clinical <- GDCprepare_clinic(query, clinical.info = "patient")
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%
```

```
## To get the following information please change the clinical.info
argument
```

```
## => new_tumor_events: new_tumor_event
```

```
## => drugs: drug
```

```
## => follow_ups: follow_up
```

```
## => radiations: radiation
```

```
## Parsing follow up version: follow_up_v1.0
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%
```

```
## Adding stage event information
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%
```

```
## Updating days_to_last_followup and vital_status from follow_up
information using last entry
```

```
## Parsing follow up version: follow_up_v1.0
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%
```

```
## ----echo = TRUE, message = FALSE, warning = FALSE-----
-----
```

```
clinical %>%
  datatable(filter = 'top',
    options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
    rownames = FALSE)
```

```
## ----results = 'hide', echo=TRUE, message=FALSE, warning=FALSE---
-----
```

```
clinical.drug <- GDCprepare_clinic(query, clinical.info = "drug")
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%
```

```
## ----echo = TRUE, message = FALSE, warning = FALSE-----
-----
```

```
clinical.drug %>%
  datatable(filter = 'top',
            options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
            rownames = FALSE)
```

```
## ----results = 'hide', echo=TRUE, message=FALSE, warning=FALSE---
-----
```

```
clinical.radiation <- GDCprepare_clinic(query, clinical.info = "radiation")
```

```
##
```

```
|
|
| 0%
|
|=====
| 50%
```

```
## ----echo = TRUE, message = FALSE, warning = FALSE-----
-----
```

```
clinical.radiation %>%
  datatable(filter = 'top',
            options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
            rownames = FALSE)
```

```
## ----results = 'hide', echo=TRUE, message=FALSE, warning=FALSE---
-----
```

```
clinical.admin <- GDCprepare_clinic(query, clinical.info = "admin")
```

```
##
```

```

|
|
| 0%
|
|=====
| 50%
|
|=====
====| 100%

```

```
## ----echo = TRUE, message = FALSE, warning = FALSE-----
-----
```

```
clinical.admin %>%
  datatable(filter = 'top',
            options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
            rownames = FALSE)
```

```
## ----results = 'hide', echo=TRUE, message=FALSE, warning=FALSE---
-----
```

```
# Pathology report from harmonized portal
query_harmonized <- GDCquery(
  project = "TCGA-COAD",
  data.category = "Biospecimen",
  data.type = "Slide Image",
  experimental.strategy = "Diagnostic Slide",
  barcode = c("TCGA-RU-A8FL","TCGA-AA-3972")
)
```

```
## -----
## o GDCquery: Searching in GDC database
## -----
## Genome of reference: hg38
## -----
## oo Accessing GDC. This might take a while...
## -----
## ooo Project: TCGA-COAD
## -----
## oo Filtering results

```

```
## -----
## ooo By experimental.strategy
## ooo By data.type
## ooo By barcode
## -----
## oo Checking data
## -----
## ooo Checking if there are duplicated cases
## ooo Checking if there are results for the query
## -----
## o Preparing output
## -----
## ----echo=TRUE, message=FALSE, warning=FALSE-----
-----
```

```
query_harmonized %>%
  getResults %>%
  head %>%
  DT::datatable(options = list(scrollX = TRUE, keys = TRUE))
```

```
## ----eval = TRUE-----
-----
```

```
bar <- c(
  "TCGA-G9-6378-02A-11R-1789-07", "TCGA-CH-5767-04A-11R-1789-07",
  "TCGA-G9-6332-60A-11R-1789-07", "TCGA-G9-6336-01A-11R-1789-07",
  "TCGA-G9-6336-11A-11R-1789-07", "TCGA-G9-7336-11A-11R-1789-07",
  "TCGA-G9-7336-04A-11R-1789-07", "TCGA-G9-7336-14A-11R-1789-07",
  "TCGA-G9-7036-04A-11R-1789-07", "TCGA-G9-7036-02A-11R-1789-07",
  "TCGA-G9-7036-11A-11R-1789-07", "TCGA-G9-7036-03A-11R-1789-07",
  "TCGA-G9-7036-10A-11R-1789-07", "TCGA-BH-A1ES-10A-11R-1789-07",
  "TCGA-BH-A1F0-10A-11R-1789-07", "TCGA-BH-A0BZ-02A-11R-1789-07",
  "TCGA-B6-A0WY-04A-11R-1789-07", "TCGA-BH-A1FG-04A-11R-1789-08",
  "TCGA-D8-A1JS-04A-11R-2089-08", "TCGA-AN-A0FN-11A-11R-8789-08",
  "TCGA-AR-A2LQ-12A-11R-8799-08", "TCGA-AR-A2LH-03A-11R-1789-07",
  "TCGA-BH-A1F8-04A-11R-5789-07", "TCGA-AR-A24T-04A-55R-1789-07",
  "TCGA-AO-A0J5-05A-11R-1789-07", "TCGA-BH-A0B4-11A-12R-1789-07",
  "TCGA-B6-A1KN-60A-13R-1789-07", "TCGA-AO-A0J5-01A-11R-1789-07",
  "TCGA-AO-A0J5-01A-11R-1789-07", "TCGA-G9-6336-11A-11R-1789-07",
  "TCGA-G9-6380-11A-11R-1789-07", "TCGA-G9-6380-01A-11R-1789-07",
```

```

"TCGA-G9-6340-01A-11R-1789-07", "TCGA-G9-6340-11A-11R-1789-07"
)

S <- TCGAquery_SampleTypes(bar,"TP")
S2 <- TCGAquery_SampleTypes(bar,"NB")

# Retrieve multiple tissue types NOT FROM THE SAME PATIENTS
SS <- TCGAquery_SampleTypes(bar,c("TP","NB"))

# Retrieve multiple tissue types FROM THE SAME PATIENTS
SSS <- TCGAquery_MatchedCoupledSampleTypes(bar,c("NT","TP"))

## ----eval = FALSE-----
# # This code will get all clinical indexed data from TCGA
# library(data.table)
# library(dplyr)
# library(regexPipes)
# clinical <- TCGAbiolinks::getGDCprojects()$project_id %>%
#   regexPipes::grep("TCGA",value = TRUE) %>%
#   sort %>%
#   plyr::alply(1,GDCquery_clinic, .progress = "text") %>%
#   rbindlist
# readr::write_csv(clinical,path = paste0("all_clin_indexed.csv"))
#
# # This code will get all clinical XML data from TCGA
# getclinical <- function(proj){
#   message(proj)
#   while(1){
#     result = tryCatch({
#       query <- GDCquery(project = proj, data.category = "Clinical",data.format =
"bcr xml")
#       GDCdownload(query)
#       clinical <- GDCprepare_clinic(query, clinical.info = "patient")
#       for(i in c("admin","radiation","follow_up","drug","new_tumor_event")){
#         message(i)
#         aux <- GDCprepare_clinic(query, clinical.info = i)
#         if(is.null(aux) || nrow(aux) == 0) next
#         # add suffix manually if it already exists
#         replicated <- which(grep("bcr_patient_barcode",colnames(aux), value =
T,invert = T) %in% colnames(clinical))
#         colnames(aux)[replicated] <- paste0(colnames(aux)[replicated],".",i)
#         if(!is.null(aux)) clinical <- merge(clinical,aux,by = "bcr_patient_barcode",
all = TRUE)
#       }
#       readr::write_csv(clinical,path = paste0(proj,"_clinical_from_XML.csv")) #
Save the clinical data into a csv file

```

```

#       return(clinical)
#     }, error = function(e) {
#       message(paste0("Error clinical: ", proj))
#     })
#   }
# }
# clinical <- TCGAbiolinks::getGDCprojects()$project_id %>%
#   regexPipes::grep("TCGA",value=T) %>% sort %>%
#   plyr::alply(1,getclinical, .progress = "text") %>%
#   rbindlist(fill = TRUE) %>% setDF %>% subset(!duplicated(clinical))
#
# readr::write_csv(clinical,path = "all_clin_XML.csv")
# # result: https://drive.google.com/open?id=0B0-8N2fjttG-WWxSVE5MSGpva1U
# # Obs: this table has multiple lines for each patient, as the patient might have
# #       several followups, drug treatments,
# #       new tumor events etc...

```