Downloading API's: cgdsr Package

Jessica Wheeler

MSDS 6306-401

- Library(cgdsr)
- ?`cgdsr-package`

- The **Cancer Genomic Data Server (CGDS)** web service interface provides direct programmatic access to all genomic data stored within the server. This package provides a basic set of R functions for querying the CGDS hosted by the Computational Biology Center at Memorial-Sloan-Kettering Cancer Center (MSKCC).

▶ The library can issue the following types of queries:

▶ *getCancerStudies()*: What cancer studies are hosted on the server? For example TCGA Glioblastoma or TCGA Ovarian cancer.

▶ *getGeneticProfiles()*: What genetic profile types are available for cancer study X? For example mRNA expression or copy number alterations.

▶ *getCaseLists()*: what case sets are available for cancer study X? For example all samples or only samples corresponding to a given cancer subtype.

▶ *getProfileData()*: Retrieve slices of genomic data. For example, a client can retrieve all mutation data from PTEN and EGFR in TCGA glioblastoma.

▶ *getClinicalData()*: Retrieve clinical data (e.g. patient survival time and age) for a given case list.

Example

- #create a CGDS object
- **mycgds <- CGDS(http://www.cbioportal.org/)**

- # Get list of cancer studies at server
- **getCancerStudies(mycgds)**
  - Shows a list of study names and if you scroll down, shows titles of the studies, all numbered.
  - I chose #27: coadread_genentech, Colorectal Adenocarcinoma (Genentech, Nature 2012)

- # Get available case lists (collection of samples) for a given cancer study
- **mycancerstudy <- getCancerStudies(mycgds)[27,1]**
  - "coadread_genentech"
- **mycaselist <- getCaseLists(mycgds,mycancerstudy)[1,1]**
  - > mycaselist
  - [1] "coadread_genentech_all"

- **#Changed caselist argument to see complete list:**
  - mycaselist <- getCaseLists(mycgds, mycancerstudy)
  - case_list_id case_list_name   case_list_description1 coadread_genentech_all   All Tumors   All tumor samples (72 samples)
  - coadread_genentech_sequenced Sequenced Tumors All (Next-Gen) sequenced samples (72 samples) cancer_study_id 49
  - Also a bunch of case_ids listed below.

- **#Get Clinical data for the case list:**

```
> mycaselist <- getCaseLists(mycgds, mycancerstudy)[1,1]
> myclinicaldata = getClinicalData(mycgds,mycaselist)
> dim(myclinicaldata)
[1] 72  6
> head(myclinicaldata)
          DIAGNOSIS MSI_STATUS STR_STATUS TISSUE_SOURCE_SITE TUMOR_CONTENT TUMOR_TYPE
X587390 Adenocarcinoma        MSS        0/5              Tumor           50        colon
X587294 Adenocarcinoma        MSS        0/5              Tumor           80        colon
X587256 Adenocarcinoma        MSI        5/5              Tumor           85        colon
X587378 Adenocarcinoma        MSS        0/5              Tumor           90        colon
X587254 Adenocarcinoma        MSS        0/5              Tumor           80        colon
X587298 Adenocarcinoma        MSS        0/5              Tumor           80        colon
```