The CGDS-R library

Anders Jacobsen and Augustin Luna

June 25, 2019

Contents

1 Introduction

This package provides a basic set of R functions for querying the Cancer Genomic Data Server (CGDS) hosted by the Computational Biology Center (cBio) at the Memorial Sloan-Kettering Cancer Center (MSKCC). This service is a part of the cBio Cancer Genomics Portal, `http://www.cbioportal.org/`.

In summary, 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 for PTEN and EGFR in TCGA glioblastoma.
- `getClinicalData()`: Retrieve clinical data (e.g. patient survival time and age) for a given cancer study and list of cases.

Each of these functions will be briefly described in the following sections. The last part of this document includes some concrete examples of how to access and plot the data.

The purpose of this document is to give the reader a quick overview of the `cgdsr` package. Please refer to the corresponding R manual pages for a more detailed explanation of arguments and output for each function.

2 The CGDS R interface

2.1 `CGDS()` : Create a CGDS connection object

Initially, we will establish a connection to the public CGDS server hosted by Memorial Sloan-Kettering Cancer Center. The function for creating a CGDS connection object requires the URL of the CGDS server service, in this case `http://www.cbioportal.org/`, as an argument.

```
> library(cgdsr)
> # Create CGDS object
> mycgds = CGDS("http://www.cbioportal.org/")
```

The variable `mycgds` is now a CGDS connection object pointing at the URL for the public CGDS server. This connection object must be included as an argument to all subsequent interface calls. Optionally, we can now perform a set of simple tests of the data returned from the CGDS connection object using the `test` function:

```
> # Test the CGDS endpoint URL using a few simple API tests
> test(mycgds)

getCancerStudies...  OK
getCaseLists (1/2) ...  OK
getCaseLists (2/2) ...  OK
getGeneticProfiles (1/2) ...  OK
getGeneticProfiles (2/2) ...  OK
getClinicalData (1/1) ...  OK
getProfileData (1/6) ...  OK
getProfileData (2/6) ...  OK
getProfileData (3/6) ...  OK
getProfileData (4/6) ...  OK
getProfileData (5/6) ...  OK
getProfileData (6/6) ...  OK
```

Note that the tests may not work if you are connecting to a portal other than the one in the above example. The tests can fail if the portal instance

does not contain the data that is being tested against, or if you do have have authorization to access the data that is being tested against.

A verbose option can be set for the CGDS connection object. This will cause function calls that retrieve data from cBioPortal to additionally display the programming interface URL to be displayed. This is useful for debugging and troubleshooting issues with the package.

```
> # Set verbose flag
> setVerbose(mycgds, TRUE)
```

```
[1] TRUE
```

[Optional] A data access token can be optionally attached to a CGDS connection object when it is created. This allows you to connect to cBioPortal instances that require authentication. Data access tokens (when this feature is enabled) can be created through the cbioportal website. If you attempt to access data that you are not authorized to access you will get an `Unauthorized (HTTP 401)` error. Note the public portal at http://www.cbioportal.org/ does not require authentication so you do not need a token to connect to it.

```
> # Connect to a portal instance that requires authetication
> mysecurecgds = CGDS("https://cbioportal.mskcc.org/",
+                     token="fd0522cb-7972-40d0-9d83-cb4c14e8a337")
```

2.2 `getCancerStudies()` : Retrieve a set of available cancer studies

Having created a CGDS connection object, we can now retrieve a data frame with available cancer studies using the `getCancerStudies` function:

```
> # Get list of cancer studies at server
> getCancerStudies(mycgds)[,c(1,2)]
```

```
http://www.cbioportal.org/webservice.do?cmd=getCancerStudies&
                      cancer_study_id
1                       paac_jhu_2014
2                 mel_tsam_liang_2017
3                      all_stjude_2015
4                      all_stjude_2016
5                        aml_ohsu_2018
6                        laml_tcga_pub
7         laml_tcga_pan_can_atlas_2018
8                            laml_tcga
9                        acyc_fmi_2014
10                       acyc_jhu_2016
11                       acyc_mda_2015
12                       acyc_mgh_2016
13                     acyc_mskcc_2013
14                    acyc_sanger_2013
15                     acbc_mskcc_2015
16         acc_tcga_pan_can_atlas_2018
```

```
                          acc_tcga
                       sarc_tcga_pub
                      ampca_bcm_2016
                        odg_msk_2017
                      bcc_unige_2016
               blca_mskcc_solit_2014
               blca_mskcc_solit_2012
      blca_plasmacytoid_mskcc_2016
                 blca_tcga_pub_2017
                           blca_bgi
            blca_dfarber_mskcc_2014
                       blca_tcga_pub
      blca_tcga_pan_can_atlas_2018
                           blca_tcga
       lgg_tcga_pan_can_atlas_2018
                           lgg_tcga
                      brca_metabric
                    breast_msk_2018
         brca_bccrc_xenograft_2014
                   bfn_duke_nus_2015
                          brca_bccrc
                          brca_broad
                         brca_sanger
                   brca_tcga_pub2015
                       brca_tcga_pub
      brca_tcga_pan_can_atlas_2018
                           brca_tcga
                 cellline_ccle_broad
      cesc_tcga_pan_can_atlas_2018
                           cesc_tcga
                       chol_msk_2018
                      chol_nccs_2013
                       chol_nus_2012
      chol_tcga_pan_can_atlas_2018
                           chol_tcga
                     lcll_broad_2013
                      cll_iuopa_2015
                    cllsll_icgc_2011
                  coadread_dfci_2016
                  coadread_genentech
                   coadread_tcga_pub
coadread_tcga_pan_can_atlas_2018
                       coadread_tcga
                      coadread_mskcc
                   cscc_dfarber_2015
                  cscc_hgsc_bcm_2014
                  ctcl_columbia_2015
                       pact_jhu_2011
                     desm_broad_2015
                     dlbcl_dfci_2018
```

```
                dlbc_broad_2012
                dlbcl_duke_2017
    dlbc_tcga_pan_can_atlas_2018
                nhl_bcgsc_2013
                 ucec_msk_2018
                    esca_broad
    esca_tcga_pan_can_atlas_2018
                  stes_tcga_pub
                      esca_tcga
                      escc_icgc
                  escc_ucla_2014
                 es_iocurie_2014
                    gbc_msk_2018
              gbc_shanghai_2014
                 egc_tmucih_2015
                    gct_msk_2016
               gbm_tcga_pub2013
                    gbm_tcga_pub
     gbm_tcga_pan_can_atlas_2018
                        gbm_tcga
                glioma_msk_2018
                     hnsc_broad
                       hnsc_jhu
                  hnsc_tcga_pub
    hnsc_tcga_pan_can_atlas_2018
                        hnsc_tcga
            liad_inserm_fr_2014
             hcc_mskimpact_2018
             hcc_inserm_fr_2015
     histiocytosis_cobi_msk_2019
                 all_stjude_2013
            panet_shanghai_2013
                  chol_jhu_2013
                 kich_tcga_pub
    kich_tcga_pan_can_atlas_2018
                       kich_tcga
                        kirc_bgi
                ccrcc_irc_2014
                 kirc_tcga_pub
    kirc_tcga_pan_can_atlas_2018
                       kirc_tcga
    kirp_tcga_pan_can_atlas_2018
                       kirp_tcga
          hcc_msk_venturaa_2018
                   lihc_amc_prv
                     lihc_riken
    lihc_tcga_pan_can_atlas_2018
                       lihc_tcga
                  lgg_ucsf_2014
                     luad_broad
```

```
                  luad_mskcc_2015
                    luad_tcga_pub
     luad_tcga_pan_can_atlas_2018
                         luad_tcga
                          luad_tsp
                    lusc_tcga_pub
     lusc_tcga_pan_can_atlas_2018
                         lusc_tcga
                         dlbc_tcga
                  msk_impact_2017
                 mixed_allen_2018
                      mpnst_mskcc
                 mcl_idibips_2013
                   mbn_mdacc_2013
                   mbl_broad_2012
                         mbl_icgc
                         mbl_pcgp
                mbl_sickkids_2016
               skcm_broad_dfarber
                 lgggbm_tcga_pub
     meso_tcga_pan_can_atlas_2018
                         meso_tcga
                   brca_igr_2015
                    crc_msk_2017
                    egc_msk_2017
       skcm_vanderbilt_mskcc_2015
                   skcm_ucla_2016
                        prad_mich
                   prad_su2c_2019
                   prad_su2c_2015
metastatic_solid_tumors_mich_2017
                         mm_broad
                   mds_tokyo_2011
                   cellline_nci60
                  npc_nusingapore
                     nbl_amc_2012
                   nbl_broad_2013
                nbl_ucologne_2015
                    nepc_wcm_2016
                   nhl_bcgsc_2011
                    lung_msk_2017
               nsclc_pd1_msk_2018
                 nsclc_unito_2016
                  blca_nmibc_2017
             hnsc_mdanderson_2013
                       ov_tcga_pub
       ov_tcga_pan_can_atlas_2018
                           ov_tcga
            nsclc_tcga_broad_2016
                        paad_icgc
```

```
                paad_qcmg_uq_2016
    paad_tcga_pan_can_atlas_2018
                        paad_tcga
                   paad_utsw_2015
                   panet_jhu_2011
                panet_arcnet_2017
                    thca_tcga_pub
      all_phase2_target_2018_pub
             aml_target_2018_pub
           es_dfarber_broad_2014
             nbl_target_2018_pub
             pediatric_dkfz_2017
               mixed_pipseq_2017
              rt_target_2018_pub
              wt_target_2018_pub
                    pcpg_tcga_pub
    pcpg_tcga_pan_can_atlas_2018
                        pcpg_tcga
       past_dkfz_heidelberg_2013
                 plmeso_nyu_2015
               thyroid_mskcc_2016
                 pcnsl_mayo_2015
                 prad_broad_2013
                      prad_broad
                  prad_cpcg_2017
                      prad_fhcrc
                      prad_mskcc
                 prad_mskcc_2014
                      prad_p1000
               prad_eururol_2017
                   prad_tcga_pub
    prad_tcga_pan_can_atlas_2018
                        prad_tcga
prad_mskcc_cheny1_organoids_2014
                 prad_mskcc_2017
                  hnc_mskcc_2016
               ccrcc_utokyo_2013
           nccrcc_genentech_2014
                  mrt_bcgsc_2016
                    rms_nih_2014
                     summit_2018
                      sarc_mskcc
    sarc_tcga_pan_can_atlas_2018
                       sarc_tcga
                      skcm_broad
    skcm_tcga_pan_can_atlas_2018
                       skcm_tcga
                       skcm_yale
      skcm_broad_brafresist_2012
                       scco_mskcc
```

217	sclc_clcgp
218	sclc_jhu
219	sclc_ucologne_2015
220	sclc_cancercell_gardner_2017
221	vsc_cuk_2018
222	stad_pfizer_uhongkong
223	stad_tcga_pub
224	stad_tcga_pan_can_atlas_2018
225	stad_tcga
226	stad_utokyo
227	stad_uhongkong
228	tmb_mskcc_2018
229	tgct_tcga
230	tgct_tcga_pan_can_atlas_2018
231	angs_project_painter_2018
232	brca_mbcproject_wagle_2017
233	prad_mpcproject_2018
234	lung_msk_pdx
235	tet_nci_2014
236	thym_tcga_pan_can_atlas_2018
237	thym_tcga
238	thca_tcga_pan_can_atlas_2018
239	thca_tcga
240	urcc_mskcc_2016
241	utuc_mskcc_2015
242	utuc_cornell_baylor_mdacc_2019
243	blca_cornell_2016
244	ucs_jhu_2014
245	ucs_tcga_pan_can_atlas_2018
246	ucs_tcga
247	uccc_nih_2017
248	ucec_tcga_pub
249	ucec_tcga_pan_can_atlas_2018
250	ucec_tcga
251	um_qimr_2016
252	uvm_tcga_pan_can_atlas_2018
253	uvm_tcga

	name
1	Acinar Cell Carcinoma of the Pancreas (JHU, J Pathol 2014)
2	Acral Melanoma (TGEN, Genome Res 2017)
3	Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)
4	Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)
5	Acute Myeloid Leukemia (OHSU, Nature 2018)
6	Acute Myeloid Leukemia (TCGA, NEJM 2013)
7	Acute Myeloid Leukemia (TCGA, PanCancer Atlas)
8	Acute Myeloid Leukemia (TCGA, Provisional)
9	Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)
10	Adenoid Cystic Carcinoma (JHU, Cancer Prev Res 2016)
11	Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)
12	Adenoid Cystic Carcinoma (MGH, Nat Gen 2016)

```
13                                          Adenoid Cystic Carcinoma (MSKCC, Nat Genet 2013)
14                                           Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)
15                             Adenoid Cystic Carcinoma of the Breast (MSKCC, J Pathol. 2015)
16                                          Adrenocortical Carcinoma (TCGA, PanCancer Atlas)
17                                              Adrenocortical Carcinoma (TCGA, Provisional)
18                                              Adult Soft Tissue Sarcomas (TCGA, Cell 2017)
19                         Ampullary Carcinoma (Baylor College of Medicine, Cell Reports 2016)
20  Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytoma (MSKCC, Neuro Oncol 2017)
21                                          Basal Cell Carcinoma (UNIGE, Nat Genet 2016)
22                                                 Bladder Cancer (MSKCC, Eur Urol 2014)
23                                             Bladder Cancer (MSKCC, J Clin Onco 2013)
24                                               Bladder Cancer (MSKCC, Nat Genet 2016)
25                                                    Bladder Cancer (TCGA, Cell 2017)
26                                       Bladder Urothelial Carcinoma (BGI, Nat Genet 2013)
27                               Bladder Urothelial Carcinoma (DFCI/MSKCC, Cancer Discov 2014)
28                                       Bladder Urothelial Carcinoma (TCGA, Nature 2014)
29                                    Bladder Urothelial Carcinoma (TCGA, PanCancer Atlas)
30                                        Bladder Urothelial Carcinoma (TCGA, Provisional)
31                                        Brain Lower Grade Glioma (TCGA, PanCancer Atlas)
32                                            Brain Lower Grade Glioma (TCGA, Provisional)
33                                  Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)
34                                                 Breast Cancer (MSK, Cancer Cell 2018)
35                                  Breast Cancer Xenografts (British Columbia, Nature 2015)
36                                  Breast Fibroepithelial Tumors (Duke-NUS, Nat Genet 2015)
37                                Breast Invasive Carcinoma (British Columbia, Nature 2012)
38                                     Breast Invasive Carcinoma (Broad, Nature 2012)
39                                   Breast Invasive Carcinoma (Sanger, Nature 2012)
40                                      Breast Invasive Carcinoma (TCGA, Cell 2015)
41                                    Breast Invasive Carcinoma (TCGA, Nature 2012)
42                                  Breast Invasive Carcinoma (TCGA, PanCancer Atlas)
43                                      Breast Invasive Carcinoma (TCGA, Provisional)
44                               Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012)
45                                  Cervical Squamous Cell Carcinoma (TCGA, PanCancer Atlas)
46   Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)
47                                        Cholangiocarcinoma (MSK, Clin Cancer Res 2018)
48                 Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013)
49                    Cholangiocarcinoma (National University of Singapore, Nat Genet 2012)
50                                        Cholangiocarcinoma (TCGA, PanCancer Atlas)
51                                            Cholangiocarcinoma (TCGA, Provisional)
52                                     Chronic Lymphocytic Leukemia (Broad, Cell 2013)
53                                   Chronic Lymphocytic Leukemia (IUOPA, Nature 2015)
54                                Chronic lymphocytic leukemia (ICGC, Nature Genetics 2011)
55                                      Colorectal Adenocarcinoma (DFCI, Cell Reports 2016)
56                                   Colorectal Adenocarcinoma (Genentech, Nature 2012)
57                                        Colorectal Adenocarcinoma (TCGA, Nature 2012)
58                                    Colorectal Adenocarcinoma (TCGA, PanCancer Atlas)
59                                        Colorectal Adenocarcinoma (TCGA, Provisional)
60                              Colorectal Adenocarcinoma Triplets (MSKCC, Genome Biol 2014)
61                             Cutaneous Squamous Cell Carcinoma (DFCI, Clin Cancer Res 2015)
62                        Cutaneous Squamous Cell Carcinoma (MD Anderson, Clin Cancer Res 2014)
```

63 Cutaneous T Cell Lymphoma (Columbia U, Nat Genet 2015)
64 Cystic Tumor of the Pancreas (Johns Hopkins, PNAS 2011)
65 Desmoplastic Melanoma (Broad Institute, Nat Genet 2015)
66 Diffuse Large B cell Lymphoma (DFCI, Nat Med 2018)
67 Diffuse Large B-Cell Lymphoma (Broad, PNAS 2012)
68 Diffuse Large B-Cell Lymphoma (Duke, Cell 2017)
69 Diffuse Large B-Cell Lymphoma (TCGA, PanCancer Atlas)
70 Diffuse Large B-cell Lymphoma (BCGSC, Blood 2013)
71 Endometrial Cancer (MSK, 2018)
72 Esophageal Adenocarcinoma (DFCI, Nat Genet 2013)
73 Esophageal Adenocarcinoma (TCGA, PanCancer Atlas)
74 Esophageal Carcinoma (TCGA, Nature 2017)
75 Esophageal Carcinoma (TCGA, Provisional)
76 Esophageal Squamous Cell Carcinoma (ICGC, Nature 2014)
77 Esophageal Squamous Cell Carcinoma (UCLA, Nat Genet 2014)
78 Ewing Sarcoma (Institut Curie, Cancer Discov 2014)
79 Gallbladder Cancer (MSK, Cancer 2018)
80 Gallbladder Carcinoma (Shanghai, Nat Genet 2014)
81 Gastric Adenocarcinoma (TMUCIH, PNAS 2015)
82 Germ Cell Tumors (MSKCC, J Clin Oncol 2016)
83 Glioblastoma (TCGA, Cell 2013)
84 Glioblastoma (TCGA, Nature 2008)
85 Glioblastoma Multiforme (TCGA, PanCancer Atlas)
86 Glioblastoma Multiforme (TCGA, Provisional)
87 Glioma (MSK, 2018)
88 Head and Neck Squamous Cell Carcinoma (Broad, Science 2011)
89 Head and Neck Squamous Cell Carcinoma (Johns Hopkins, Science 2011)
90 Head and Neck Squamous Cell Carcinoma (TCGA, Nature 2015)
91 Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)
92 Head and Neck Squamous Cell Carcinoma (TCGA, Provisional)
93 Hepatocellular Adenoma (INSERM, Cancer Cell 2014)
94 Hepatocellular Carcinoma (MSK, Clin Cancer Res 2018)
95 Hepatocellular Carcinomas (INSERM, Nat Genet 2015)
96 Histiocytosis Cobimetinib (MSK, Nature 2019)
97 Hypodiploid Acute Lymphoid Leukemia (St Jude, Nat Genet 2013)
98 Insulinoma (Shanghai, Nat Commun 2013)
99 Intrahepatic Cholangiocarcinoma (JHU, Nat Genet 2013)
100 Kidney Chromophobe (TCGA, Cancer Cell 2014)
101 Kidney Chromophobe (TCGA, PanCancer Atlas)
102 Kidney Chromophobe (TCGA, Provisional)
103 Kidney Renal Clear Cell Carcinoma (BGI, Nat Genet 2012)
104 Kidney Renal Clear Cell Carcinoma (IRC, Nat Genet 2014)
105 Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013)
106 Kidney Renal Clear Cell Carcinoma (TCGA, PanCancer Atlas)
107 Kidney Renal Clear Cell Carcinoma (TCGA, Provisional)
108 Kidney Renal Papillary Cell Carcinoma (TCGA, PanCancer Atlas)
109 Kidney Renal Papillary Cell Carcinoma (TCGA, Provisional)
110 Liver Hepatocellular Adenoma and Carcinomas (MSK, PLOS One 2018)
111 Liver Hepatocellular Carcinoma (AMC, Hepatology 2014)
112 Liver Hepatocellular Carcinoma (RIKEN, Nat Genet 2012)

113 Liver Hepatocellular Carcinoma (TCGA, PanCancer Atlas)
114 Liver Hepatocellular Carcinoma (TCGA, Provisional)
115 Low-Grade Gliomas (UCSF, Science 2014)
116 Lung Adenocarcinoma (Broad, Cell 2012)
117 Lung Adenocarcinoma (MSKCC, Science 2015)
118 Lung Adenocarcinoma (TCGA, Nature 2014)
119 Lung Adenocarcinoma (TCGA, PanCancer Atlas)
120 Lung Adenocarcinoma (TCGA, Provisional)
121 Lung Adenocarcinoma (TSP, Nature 2008)
122 Lung Squamous Cell Carcinoma (TCGA, Nature 2012)
123 Lung Squamous Cell Carcinoma (TCGA, PanCancer Atlas)
124 Lung Squamous Cell Carcinoma (TCGA, Provisional)
125 Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)
126 MSK-IMPACT Clinical Sequencing Cohort (MSKCC, Nat Med 2017)
127 MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)
128 Malignant Peripheral Nerve Sheath Tumor (MSKCC, Nat Genet 2014)
129 Mantle Cell Lymphoma (IDIBIPS, PNAS 2013)
130 Mature B-cell malignancies (MD Anderson Cancer Center)
131 Medulloblastoma (Broad, Nature 2012)
132 Medulloblastoma (ICGC, Nature 2012)
133 Medulloblastoma (PCGP, Nature 2012)
134 Medulloblastoma (Sickkids, Nature 2016)
135 Melanoma (Broad/Dana Farber, Nature 2012)
136 Merged Cohort of LGG and GBM (TCGA, Cell 2016)
137 Mesothelioma (TCGA, PanCancer Atlas)
138 Mesothelioma (TCGA, Provisional)
139 Metastatic Breast Cancer (INSERM, PLoS Med 2016)
140 Metastatic Colorectal Cancer (MSKCC, Cancer Cell 2018)
141 Metastatic Esophagogastric Cancer (MSKCC, Cancer Discovery 2017)
142 Metastatic Melanoma (MSKCC, JCO Precis Oncol 2017)
143 Metastatic Melanoma (UCLA, Cell 2016)
144 Metastatic Prostate Adenocarcinoma (MCTP, Nature 2012)
145 Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019)
146 Metastatic Prostate Cancer (SU2C/PCF Dream Team, Cell 2015)
147 Metastatic Solid Cancers (UMich, Nature 2017)
148 Multiple Myeloma (Broad, Cancer Cell 2014)
149 Myelodysplasia (UTokyo, Nature 2011)
150 NCI-60 Cell Lines (NCI, Cancer Res 2012)
151 Nasopharyngeal Carcinoma (Singapore, Nat Genet 2014)
152 Neuroblastoma (AMC Amsterdam, Nature 2012)
153 Neuroblastoma (Broad, Nat Genet 2013)
154 Neuroblastoma (Broad, Nature 2015)
155 Neuroendocrine Prostate Cancer (Multi-Institute, Nat Med 2016)
156 Non-Hodgkin Lymphoma (BCGSC, Nature 2011)
157 Non-Small Cell Cancer (MSKCC, Cancer Discov 2017)
158 Non-Small Cell Lung Cancer (MSKCC, J Clin Oncol 2018)
159 Non-Small Cell Lung Cancer (University of Turin, Lung Cancer 2017)
160 Nonmuscle Invasive Bladder Cancer (MSK Eur Urol 2017)
161 Oral Squamous Cell Carcinoma (MD Anderson, Cancer Discov 2013)
162 Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)

```
                                  Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas)
                                      Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)
                                                  Pan-Lung Cancer (TCGA, Nat Genet 2016)
                                          Pancreatic Adenocarcinoma (ICGC, Nature 2012)
                                          Pancreatic Adenocarcinoma (QCMG, Nature 2016)
                                       Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)
                                          Pancreatic Adenocarcinoma (TCGA, Provisional)
                                              Pancreatic Cancer (UTSW, Nat Commun 2015)
          Pancreatic Neuroendocrine Tumors (Johns Hopkins University, Science 2011)
                    Pancreatic Neuroendocrine Tumors (Multi-Institute, Nature 2017)
                                          Papillary Thyroid Carcinoma (TCGA, Cell 2014)
                             Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)
                                       Pediatric Acute Myeloid Leukemia (TARGET, 2018)
                                    Pediatric Ewing Sarcoma (DFCI, Cancer Discov 2014)
                                         Pediatric Neuroblastoma (TARGET, 2018)
                                       Pediatric Pan-Cancer (DKFZ, Nature 2017)
                                    Pediatric Pan-cancer (Columbia U, Genome Med 2016)
                                        Pediatric Rhabdoid Tumor (TARGET, 2018)
                                         Pediatric Wilms' Tumor (TARGET, 2018)
                                  Pheochromocytoma and Paraganglioma (TCGA, Cell 2017)
                               Pheochromocytoma and Paraganglioma (TCGA, PanCancer Atlas)
                                  Pheochromocytoma and Paraganglioma (TCGA, Provisional)
                                      Pilocytic Astrocytoma (ICGC, Nature Genetics 2013)
                                            Pleural Mesothelioma (NYU, Cancer Res 2015)
          Poorly-Differentiated and Anaplastic Thyroid Cancers (MSKCC, JCI 2016)
Primary Central Nervous System Lymphoma (Mayo Clinic, Clin Cancer Res 2015)
                                       Prostate Adenocarcinoma (Broad/Cornell, Cell 2013)
                                  Prostate Adenocarcinoma (Broad/Cornell, Nat Genet 2012)
                                       Prostate Adenocarcinoma (CPC-GENE, Nature 2017)
                             Prostate Adenocarcinoma (Fred Hutchinson CRC, Nat Med 2016)
                                       Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)
                                          Prostate Adenocarcinoma (MSKCC, PNAS 2014)
                               Prostate Adenocarcinoma (MSKCC/DFCI, Nature Genetics 2018)
                                   Prostate Adenocarcinoma (SMMU, Eur Urol 2017)
                                       Prostate Adenocarcinoma (TCGA, Cell 2015)
                                    Prostate Adenocarcinoma (TCGA, PanCancer Atlas)
                                       Prostate Adenocarcinoma (TCGA, Provisional)
                                  Prostate Adenocarcinoma Organoids (MSKCC, Cell 2014)
                                       Prostate Cancer (MSKCC, JCO Precis Oncol 2017)
     Recurrent and Metastatic Head & Neck Cancer (MSKCC, JAMA Oncol 2016)
                                      Renal Clear Cell Carcinoma (UTokyo, Nat Genet 2013)
                            Renal Non-Clear Cell Carcinoma (Genentech, Nat Genet 2014)
                                        Rhabdoid Cancer (BCGSC, Cancer Cell 2016)
                                       Rhabdomyosarcoma (NIH, Cancer Discov 2014)
                 SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)
                                            Sarcoma (MSKCC/Broad, Nat Genet 2010)
                                                  Sarcoma (TCGA, PanCancer Atlas)
                                                      Sarcoma (TCGA, Provisional)
                                        Skin Cutaneous Melanoma (Broad, Cell 2012)
                                     Skin Cutaneous Melanoma (TCGA, PanCancer Atlas)
```

```
                                                  Skin Cutaneous Melanoma (TCGA, Provisional)
                                                Skin Cutaneous Melanoma (Yale, Nat Genet 2012)
                                          Skin Cutaneous Melanoma(Broad, Cancer Discov 2014)
                                  Small Cell Carcinoma of the Ovary (MSKCC, Nat Genet 2014)
                                                  Small Cell Lung Cancer (CLCGP, Nat Genet 2012)
                                       Small Cell Lung Cancer (Johns Hopkins, Nat Genet 2012)
                                                Small Cell Lung Cancer (U Cologne, Nature 2015)
                                 Small-Cell Lung Cancer (Multi-Institute, Cancer Cell 2017)
                              Squamous Cell Carcinoma of the Vulva (CUK, Exp Mol Med 2018)
                                   Stomach Adenocarcinoma (Pfizer and UHK, Nat Genet 2014)
                                                  Stomach Adenocarcinoma (TCGA, Nature 2014)
                                                Stomach Adenocarcinoma (TCGA, PanCancer Atlas)
                                                  Stomach Adenocarcinoma (TCGA, Provisional)
                                               Stomach Adenocarcinoma (U Tokyo, Nat Genet 2014)
                                                  Stomach Adenocarcinoma (UHK, Nat Genet 2011)
                                                TMB and Immunotherapy (MSKCC, Nat Genet 2019)
                                               Testicular Germ Cell Cancer (TCGA, Provisional)
                                             Testicular Germ Cell Tumors (TCGA, PanCancer Atlas)
                  The Angiosarcoma Project - Count Me In (Provisional, September 2018)
                      The Metastatic Breast Cancer Project (Provisional, October 2018)
                   The Metastatic Prostate Cancer Project (Provisional, December 2018)
                                                          Thoracic PDX (MSK, Provisional)
                                               Thymic Epithelial Tumors (NCI, Nat Genet 2014)
                                                          Thymoma (TCGA, PanCancer Atlas)
                                                              Thymoma (TCGA, Provisional)
                                                    Thyroid Carcinoma (TCGA, PanCancer Atlas)
                                                       Thyroid Carcinoma (TCGA, Provisional)
                                          Unclassified Renal Cell Carcinoma (MSK, Nature 2016)
                                            Upper Tract Urothelial Cancer (MSK, Eur Urol 2015)
          Upper Tract Urothelial Carcinoma (Cornell/Baylor/MDACC, Nat Comm 2019)
                                           Urothelial Carcinoma (Cornell/Trento, Nat Gen 2016)
                                 Uterine Carcinosarcoma (Johns Hopkins, Nat Commun 2014)
                                          Uterine Carcinosarcoma (TCGA, PanCancer Atlas)
                                              Uterine Carcinosarcoma (TCGA, Provisional)
                                        Uterine Clear Cell Carcinoma (NIH, Cancer 2017)
                                Uterine Corpus Endometrial Carcinoma (TCGA, Nature 2013)
                             Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas)
                                 Uterine Corpus Endometrial Carcinoma (TCGA, Provisional)
                                                    Uveal Melanoma (QIMR, Oncotarget 2016)
                                                    Uveal Melanoma (TCGA, PanCancer Atlas)
                                                        Uveal Melanoma (TCGA, Provisional)
```

Here we are only showing the first two columns, the cancer study ID and short name, of the result data frame. There is also a third column, a longer description of the cancer study. The cancer study ID must be used in subsequent interface calls to retrieve case lists and genetic data profiles (see below).

2.3 getGeneticProfiles() : Retrieve genetic data profiles for a specific cancer study

This function queries the CGDS API and returns the available genetic profiles, e.g. mutation or copy number profiles, stored about a specific cancer study. Below we list the current genetic profiles for the TCGA glioblastoma cancer study:

```
> getGeneticProfiles(mycgds,'gbm_tcga')[,c(1:2)]

http://www.cbioportal.org/webservice.do?cmd=getGeneticProfiles&cancer_study_id=gbm_tcga
                        genetic_profile_id
1                            gbm_tcga_rppa
2                   gbm_tcga_rppa_Zscores
3                         gbm_tcga_gistic
4                       gbm_tcga_mrna_U133
5              gbm_tcga_mrna_U133_Zscores
6                           gbm_tcga_mrna
7            gbm_tcga_mrna_median_Zscores
8                gbm_tcga_rna_seq_v2_mrna
9  gbm_tcga_rna_seq_v2_mrna_median_Zscores
10                    gbm_tcga_linear_CNA
11               gbm_tcga_methylation_hm27
12              gbm_tcga_methylation_hm450
13                      gbm_tcga_mutations
                                genetic_profile_name
1                          Protein expression (RPPA)
2                Protein expression Z-scores (RPPA)
3     Putative copy-number alterations from GISTIC
4           mRNA expression (U133 microarray only)
5  mRNA Expression z-Scores (U133 microarray only)
6                      mRNA expression (microarray)
7             mRNA Expression z-Scores (microarray)
8                 mRNA expression (RNA Seq V2 RSEM)
9       mRNA Expression z-Scores (RNA Seq V2 RSEM)
10              Relative linear copy-number values
11                              Methylation (HM27)
12                             Methylation (HM450)
13                                       Mutations
```

Here we are only listing the first two columns, genetic profile ID and short name, of the resulting data frame. Please refer to the R manual pages for a more extended specification of the arguments and output.

2.4 getCaseLists() : Retrieve case lists for a specific cancer study

This function queries the CGDS API and returns available case lists for a specific cancer study. For example, within a particular study, only some cases may have sequence data, and another subset of cases may have been sequenced and

treated with a specific therapeutic protocol. Multiple case lists may be associated with each cancer study, and this method enables you to retrieve meta-data regarding all of these case lists. Below we list the current case lists for the TCGA glioblastoma cancer study:

```
> getCaseLists(mycgds,'gbm_tcga')[,c(1:2)]
```

```
http://www.cbioportal.org/webservice.do?cmd=getCaseLists&cancer_study_id=gbm_tcga
                case_list_id                                case_list_name
1               gbm_tcga_all                                   All samples
2     gbm_tcga_3way_complete                              Complete samples
3               gbm_tcga_cna                          Samples with CNA data
4   gbm_tcga_methylation_all                 Samples with methylation data
5  gbm_tcga_methylation_hm27          Samples with methylation data (HM27)
6 gbm_tcga_methylation_hm450         Samples with methylation data (HM450)
7             gbm_tcga_mrna Samples with mRNA data (Agilent microarray)
8   gbm_tcga_rna_seq_v2_mrna           Samples with mRNA data (RNA Seq V2)
9         gbm_tcga_mrna_U133      Samples with mRNA data (U133 microarray)
10           gbm_tcga_cnaseq            Samples with mutation and CNA data
11        gbm_tcga_sequenced                     Samples with mutation data
12             gbm_tcga_rppa               Samples with protein data (RPPA)
```

Here we are only listing the first two columns, case list ID and short name, of the resulting data frame. Please refer to the R manual pages for a more extended specification of the arguments and output.

2.5 `getProfileData()` : Retrieve genomic profile data for genes and genetic profiles

The function queries the CGDS API and returns data based on gene(s), genetic profile(s), and a case list. The function only allows specifying a list of genes and a single genetic profile, or oppositely a single gene and a list of genetic profiles. Importantly, the format of the output data frame depends on if a single or a list of genes was specified in the arguments. Below we are retrieving mRNA expression and copy number alteration genetic profiles for the NF1 gene in all samples of the TCGA glioblastoma cancer study:

```
> getProfileData(mycgds, "NF1", c("gbm_tcga_gistic","gbm_tcga_mrna"), "gbm_tcga_all")[c(1:
```

```
http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=NF1&genetic_profile_i
                gbm_tcga_gistic gbm_tcga_mrna
TCGA.02.0001.01              -1           NaN
TCGA.02.0003.01               0           NaN
TCGA.02.0006.01               0           NaN
TCGA.02.0007.01               0           NaN
TCGA.02.0009.01               0           NaN
```

We are here only showing the first five rows of the data frame. Entries with NaN indicate missing values. In the next example, we are retrieving mRNA expression data for the MDM2 and MDM4 genes:

```
> getProfileData(mycgds, c("MDM2","MDM4"), "gbm_tcga_mrna", "gbm_tcga_all")[c(25:30),]
```

```
http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=MDM2,MDM4&genetic_pro
                      MDM2       MDM4
TCGA.02.0060.01        NaN        NaN
TCGA.06.0875.01 -0.1220625 -0.2091818
TCGA.06.0876.01 -0.0874375 -0.6283636
TCGA.06.0877.01  0.0237500 -0.8740000
TCGA.06.0878.01  0.2522500 -0.1246364
TCGA.06.0879.01 -0.4213750 -0.6226364
```

We are again only showing the first five rows of the data frame.

2.6 `getClinicalData()` : Retrieve clinical data for a list of cases

The function queries the CGDS API and returns available clinical data (e.g. patient survival time and age) for a given case list. Results are returned in a data frame with a row for each case and a column for each clinical attribute. The available clinical attributes are:

- `overall_survival_months`: Overall survival, in months.
- `overall_survival_status`: Overall survival status, usually indicated as "LIVING" or "DECEASED".
- `disease_free_survival_months`: Disease free survival, in months.
- `disease_free_survival_status`: Disease free survival status, usually indicated as "DiseaseFree" or "Recurred/Progressed".
- `age_at_diagnosis`: Age at diagnosis.

Below we retrieve clinical data for the TCGA ovarian cancer dataset (only first five cases/rows are shown):

```
> getClinicalData(mycgds, "ova_all")[c(1:5),]

http://www.cbioportal.org/webservice.do?cmd=getClinicalData&case_set_id=ova_all
data frame with 0 columns and 5 rows
```

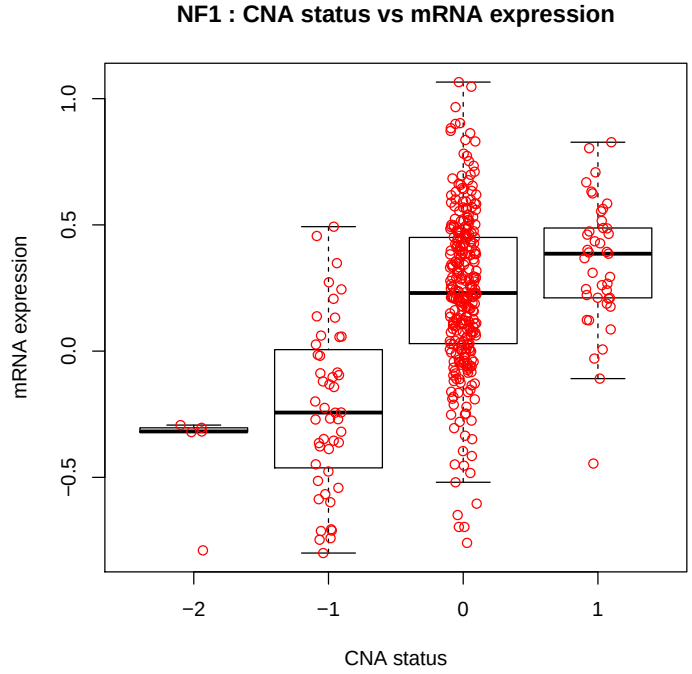
3 Examples

3.1 Example 1: Association of NF1 copy number alteration and mRNA expression in glioblastoma

As a simple example, we will generate a plot of the association between copy number alteration (CNA) status and mRNA expression change for the NF1 tumor suprpressor gene in glioblastoma. This plot is very similar to Figure 2b in the TCGA research network paper on glioblastoma (McLendon et al. 2008). The mRNA expression of NF1 has been median adjusted on the gene level (by globally subtracting the median expression level of NF1 across all samples).

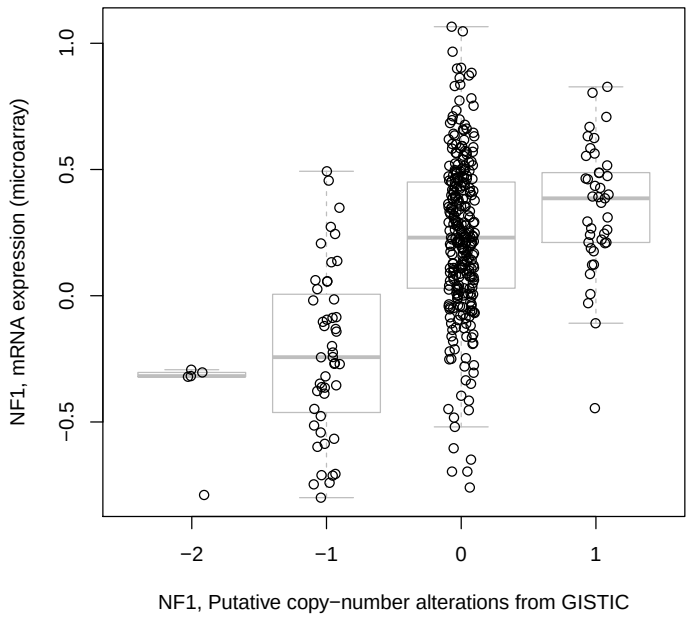
```
> df = getProfileData(mycgds, "NF1", c("gbm_tcga_gistic","gbm_tcga_mrna"), "gbm_tcga_all")
```

```
http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=NF1&genetic_profile_i

> head(df)

                 gbm_tcga_gistic gbm_tcga_mrna
TCGA.02.0001.01              -1           NaN
TCGA.02.0003.01               0           NaN
TCGA.02.0006.01               0           NaN
TCGA.02.0007.01               0           NaN
TCGA.02.0009.01               0           NaN
TCGA.02.0010.01               0           NaN

> boxplot(df[,2] ~ df[,1], main="NF1 : CNA status vs mRNA expression", xlab="CNA status", 
> stripchart(df[,2] ~ df[,1], vertical=T, add=T, method="jitter",pch=1,col='red')
```



Alternatively, the generic `cgdsr plot()` function can be used to generate a similar plot:

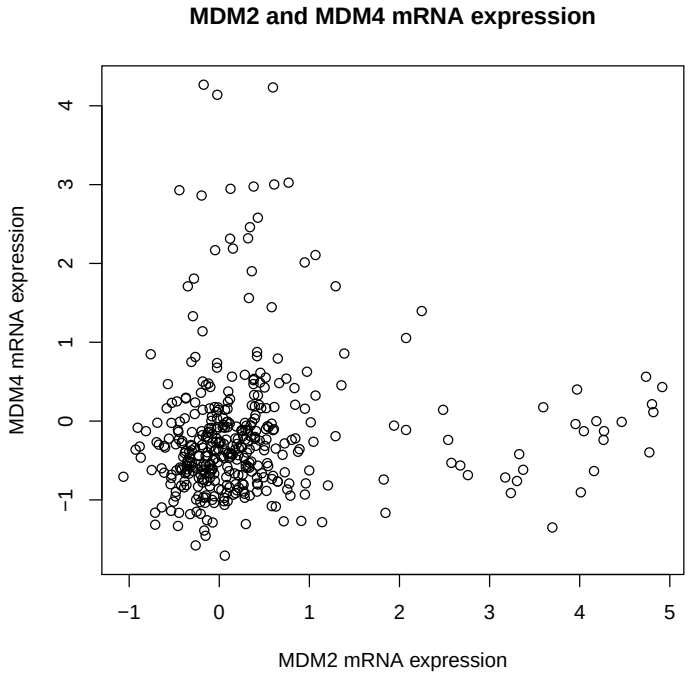
```
> plot(mycgds, "gbm_tcga", "NF1", c("gbm_tcga_gistic","gbm_tcga_mrna"), "gbm_tcga_all", sk

http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=NF1&genetic_profile_i
http://www.cbioportal.org/webservice.do?cmd=getGeneticProfiles&cancer_study_id=gbm_tcga
[1] TRUE
```



3.2 Example 2: MDM2 and MDM4 mRNA expression levels in glioblastoma

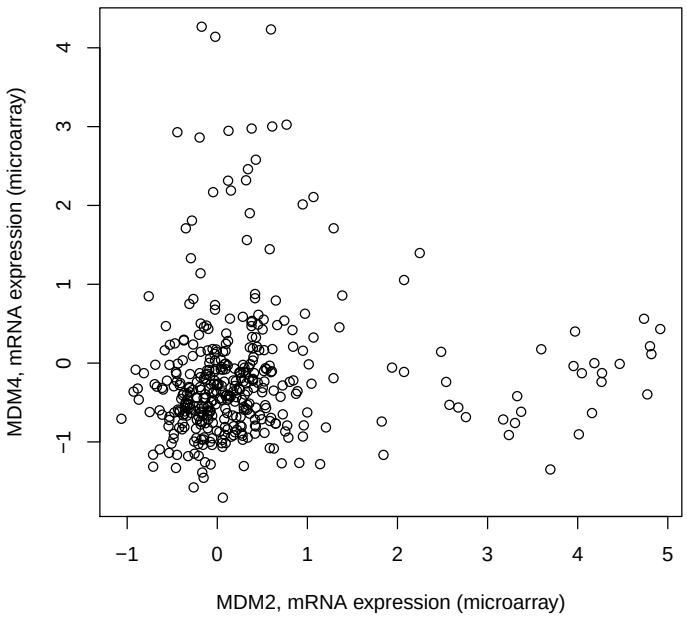
In this example, we evaluate the relationship of MDM2 and MDM4 expression levels in glioblastoma. mRNA expression levels of MDM2 and MDM4 have been median adjusted on the gene level (by globally subtracting the median expression level of the individual gene across all samples). Samples with "NaN" do not have measurements.

```
> df = getProfileData(mycgds, c("MDM2","MDM4"), "gbm_tcga_mrna", "gbm_tcga_all")

http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=MDM2,MDM4&genetic_pro

> head(df)

                 MDM2 MDM4
TCGA.02.0001.01   NaN  NaN
TCGA.02.0003.01   NaN  NaN
TCGA.02.0006.01   NaN  NaN
TCGA.02.0007.01   NaN  NaN
TCGA.02.0009.01   NaN  NaN
TCGA.02.0010.01   NaN  NaN

> plot(df, main="MDM2 and MDM4 mRNA expression", xlab="MDM2 mRNA expression", ylab="MDM4 m
```



Alternatively, the generic `cgdsr plot()` function can be used to generate a similar plot:

```
> plot(mycgds, "gbm_tcga", c("MDM2","MDM4"), "gbm_tcga_mrna" ,"gbm_tcga_all")

http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=MDM2,MDM4&genetic_pro
http://www.cbioportal.org/webservice.do?cmd=getGeneticProfiles&cancer_study_id=gbm_tcga
[1] TRUE
```



3.3 Example 3: Comparing expression of PTEN in primary and metastatic prostate cancer tumors

In this example we plot the mRNA expression levels of PTEN in primary and metastatic prostate cancer tumors.

```
> df.pri = getProfileData(mycgds, "PTEN", "prad_mskcc_mrna_median_Zscores", "prad_mskcc_pr

http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=PTEN&genetic_profile_

> head(df.pri)

            PTEN
PCA0001 9.467183
PCA0002 9.041528
PCA0003 8.511305
PCA0004      NaN
PCA0005 9.413217
PCA0006      NaN

> df.met = getProfileData(mycgds, "PTEN", "prad_mskcc_mrna_median_Zscores", "prad_mskcc_me

http://www.cbioportal.org/webservice.do?cmd=getProfileData&gene_list=PTEN&genetic_profile_

> head(df.met)
```

```
            PTEN
PCA0182 7.486938
PCA0183      NaN
PCA0184 7.578755
PCA0185      NaN
PCA0186      NaN
PCA0187 8.756132
```

```
> boxplot(list(t(df.pri),t(df.met)), main="PTEN expression in primary and metastatic tumor
> stripchart(list(t(df.pri),t(df.met)), vertical=T, add=T, method="jitter",pch=1,col='red'
```