

# The CGDS-R library

Anders Jacobsen

April 11, 2017

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The CGDS R interface</b>	<b>2</b>
2.1	<code>CGDS()</code> : Create a CGDS connection object . . . . .	2
2.2	<code>getCancerStudies()</code> : Retrieve a set of available cancer studies . . . . .	3
2.3	<code>getGeneticProfiles()</code> : Retrieve genetic data profiles for a specific cancer study . . . . .	9
2.4	<code>getCaseLists()</code> : Retrieve case lists for a specific cancer study . . . . .	10
2.5	<code>getProfileData()</code> : Retrieve genomic profile data for genes and genetic profiles . . . . .	11
2.6	<code>getClinicalData()</code> : Retrieve clinical data for a list of cases . . . . .	11
<b>3</b>	<b>Examples</b>	<b>12</b>
3.1	Example 1: Association of NF1 copy number alteration and mRNA expression in glioblastoma . . . . .	12
3.2	Example 2: MDM2 and MDM4 mRNA expression levels in glioblastoma . . . . .	14
3.3	Example 3: Comparing expression of PTEN in primary and metastatic prostate cancer tumors . . . . .	16

## 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
```

## 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)]
```

```
      cancer_study_id  
1      paac_jhu_2014  
2      laml_tcga_pub  
3      laml_tcga  
4      acyc_fmi_2014  
5      acyc_mda_2015  
6      acyc_mskcc_2013  
7      acyc_sanger_2013  
8      acbc_mskcc_2015  
9      acc_tcga  
10     blca_mskcc_solit_2014  
11     blca_mskcc_solit_2012  
12     blca_plasmacytoid_mskcc_2016  
13     blca_bgi  
14     blca_dfarber_mskcc_2014  
15     blca_tcga_pub  
16     blca_tcga  
17     lgg_tcga  
18     brca_metabric  
19     brca_bccrc  
20     brca_broad  
21     brca_sanger  
22     brca_tcga_pub2015  
23     brca_tcga_pub  
24     brca_tcga  
25     brca_bccrc_xenograft_2014  
26     cellline_ccle_broad  
27     cesc_tcga  
28     chol_nccs_2013  
29     chol_nus_2012  
30     chol_tcga  
31     lc11_broad_2013  
32     c11_iuopa_2015  
33     ccrcc_utokyo_2013  
34     coadread_dfci_2016  
35     coadread_genentech  
36     coadread_tcga_pub  
37     coadread_tcga  
38     coadread_mskcc  
39     ctcl_columbia_2015  
40     csccl_dfarber_2015  
41     desm_broad_2015
```

42 dlbcbroad\_2012  
43 esca\_broad  
44 esca\_tcga  
45 escc\_icgc  
46 escc\_ucla\_2014  
47 es\_iocurie\_2014  
48 gbc\_shanghai\_2014  
49 egc\_tmucih\_2015  
50 pradcpcg\_2017  
51 gbm\_tcga\_pub2013  
52 gbm\_tcga\_pub  
53 gbm\_tcga  
54 hnscbroad  
55 hnsccju  
56 hnsctcga\_pub  
57 hnsctcga  
58 liad\_inserm\_fr\_2014  
59 all\_stjude\_2013  
60 all\_stjude\_2015  
61 panet\_shanghai\_2013  
62 chol\_jhu\_2013  
63 kich\_tcga\_pub  
64 kich\_tcga  
65 kirc\_bgi  
66 kirc\_tcga\_pub  
67 kirc\_tcga  
68 kirp\_tcga  
69 lihcamc\_prv  
70 lihccriken  
71 lihctcga  
72 lgg\_ucsf\_2014  
73 luadbroad  
74 luadmkscc\_2015  
75 luad\_tcga\_pub  
76 luad\_tcga  
77 luad\_tsp  
78 lusc\_tcga\_pub  
79 lusc\_tcga  
80 dlbc\_tcga  
81 mpnst\_mkscc  
82 plmeso\_nyu\_2015  
83 mcl\_idibips\_2013  
84 mbl\_broad\_2012  
85 mbl\_icgc  
86 mbl\_pcgp  
87 skcm\_broad\_dfarber  
88 lgggbm\_tcga\_pub  
89 meso\_tcga  
90 pradsu2c\_2015  
91 mm\_broad

92	ccrcc_irc_2014
93	brca_igr_2015
94	mds_tokyo_2011
95	cellline_nci60
96	npc_nusingapore
97	nbl_amc_2012
98	nbl_ucologne_2015
99	nepc_wcm_2016
100	hnsc_mdanderson_2013
101	ov_tcga_pub
102	ov_tcga
103	nsclc_tcga_broad_2016
104	paad_icgc
105	paad_qcmg_uq_2016
106	paad_tcga
107	paad_utsu_2015
108	panet_jhu_2011
109	thca_tcga_pub
110	es_dfarber_broad_2014
111	pcpg_tcga
112	thyroid_mskcc_2016
113	pcnsl_mayo_2015
114	prad_broad_2013
115	prad_broad
116	prad_fhcr
117	prad_mskcc
118	prad_tcga_pub
119	prad_tcga
120	prad_mskcc_2014
121	prad_mskcc_chenyl_organoids_2014
122	prad_mich
123	hnc_mskcc_2016
124	nccrcc_genentech_2014
125	rms_nih_2014
126	sarc_mskcc
127	sarc_tcga
128	skcm_broad
129	skcm_tcga
130	skcm_yale
131	scco_mskcc
132	sclc_clcgp
133	sclc_jhu
134	sclc_ucologne_2015
135	stad_pfizer_uhongkong
136	stad_tcga_pub
137	stad_tcga
138	stad_utokyo
139	stad_uhongkong
140	stes_tcga_pub
141	tgct_tcga

142 tet\_nci\_2014  
 143 thym\_tcga  
 144 thca\_tcga  
 145 urcc\_mskcc\_2016  
 146 ucs\_jhu\_2014  
 147 ucs\_tcga  
 148 ucec\_tcga\_pub  
 149 ucec\_tcga  
 150 uvm\_tcga

	name
1	Acinar Cell Carcinoma of the Pancreas (Johns Hopkins, J Pathol 2014)
2	Acute Myeloid Leukemia (TCGA, NEJM 2013)
3	Acute Myeloid Leukemia (TCGA, Provisional)
4	Adenoid Cystic Carcinoma (FMI, Am J Surg Pathl. 2014)
5	Adenoid Cystic Carcinoma (MDA, Clin Cancer Res 2015)
6	Adenoid Cystic Carcinoma (MSKCC, Nat Genet 2013)
7	Adenoid Cystic Carcinoma (Sanger/MDA, JCI 2013)
8	Adenoid Cystic Carcinoma of the Breast (MSKCC, J Pathol. 2015)
9	Adrenocortical Carcinoma (TCGA, Provisional)
10	Bladder Cancer (MSKCC, Eur Urol 2014)
11	Bladder Cancer (MSKCC, JCO 2013)
12	Bladder Cancer, Plasmacytoid Variant (MSKCC, Nat Genet 2016)
13	Bladder Urothelial Carcinoma (BGI, Nat Genet 2013)
14	Bladder Urothelial Carcinoma (Dana Farber & MSKCC, Cancer Discov 2014)
15	Bladder Urothelial Carcinoma (TCGA, Nature 2014)
16	Bladder Urothelial Carcinoma (TCGA, Provisional)
17	Brain Lower Grade Glioma (TCGA, Provisional)
18	Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)
19	Breast Invasive Carcinoma (British Columbia, Nature 2012)
20	Breast Invasive Carcinoma (Broad, Nature 2012)
21	Breast Invasive Carcinoma (Sanger, Nature 2012)
22	Breast Invasive Carcinoma (TCGA, Cell 2015)
23	Breast Invasive Carcinoma (TCGA, Nature 2012)
24	Breast Invasive Carcinoma (TCGA, Provisional)
25	Breast cancer patient xenografts (British Columbia, Nature 2014)
26	Cancer Cell Line Encyclopedia (Novartis/Broad, Nature 2012)
27	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Provisional)
28	Cholangiocarcinoma (National Cancer Centre of Singapore, Nat Genet 2013)
29	Cholangiocarcinoma (National University of Singapore, Nat Genet 2012)
30	Cholangiocarcinoma (TCGA, Provisional)
31	Chronic Lymphocytic Leukemia (Broad, Cell 2013)
32	Chronic Lymphocytic Leukemia (IUOPA, Nature 2015)
33	Clear Cell Renal Cell Carcinoma (U Tokyo, Nat Genet 2013)
34	Colorectal Adenocarcinoma (DFCI, Cell Reports 2016)
35	Colorectal Adenocarcinoma (Genentech, Nature 2012)
36	Colorectal Adenocarcinoma (TCGA, Nature 2012)
37	Colorectal Adenocarcinoma (TCGA, Provisional)
38	Colorectal Adenocarcinoma Triplets (MSKCC, Genome Biol 2014)
39	Cutaneous T Cell Lymphoma (Columbia U, Nat Genet 2015)
40	Cutaneous squamous cell carcinoma (DFCI, Clin Cancer Res 2015)

41                                    Desmoplastic Melanoma (Broad Institute, Nat Genet 2015)  
 42                                    Diffuse Large B-Cell Lymphoma (Broad, PNAS 2012)  
 43                                    Esophageal Adenocarcinoma (Broad, Nat Genet 2013)  
 44                                    Esophageal Carcinoma (TCGA, Provisional)  
 45                                    Esophageal Squamous Cell Carcinoma (ICGC, Nature 2014)  
 46                                    Esophageal Squamous Cell Carcinoma (UCLA, Nat Genet 2014)  
 47                                    Ewing Sarcoma (Institut Cuire, Cancer Discov 2014)  
 48                                    Gallbladder Carcinoma (Shanghai, Nat Genet 2014)  
 49                                    Gastric Adenocarcinoma (TMUCIH, PNAS 2015)  
 50                                    Genomic Hallmarks of Prostate Adenocarcinoma (CPC-GENE, Nature 2017)  
 51                                    Glioblastoma (TCGA, Cell 2013)  
 52                                    Glioblastoma (TCGA, Nature 2008)  
 53                                    Glioblastoma Multiforme (TCGA, Provisional)  
 54                                    Head and Neck Squamous Cell Carcinoma (Broad, Science 2011)  
 55                                    Head and Neck Squamous Cell Carcinoma (Johns Hopkins, Science 2011)  
 56                                    Head and Neck Squamous Cell Carcinoma (TCGA, Nature 2015)  
 57                                    Head and Neck Squamous Cell Carcinoma (TCGA, Provisional)  
 58                                    Hepatocellular Adenoma (Inserm, Cancer Cell 2014)  
 59                                    Hypodiploid Acute Lymphoid Leukemia (St Jude, Nat Genet 2013)  
 60                                    Infant MLL-Rearranged Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)  
 61                                    Insulinoma (Shanghai, Nat Commun 2013)  
 62                                    Intrahepatic Cholangiocarcinoma (Johns Hopkins University, Nat Genet 2013)  
 63                                    Kidney Chromophobe (TCGA, Cancer Cell 2014)  
 64                                    Kidney Chromophobe (TCGA, Provisional)  
 65                                    Kidney Renal Clear Cell Carcinoma (BGI, Nat Genet 2012)  
 66                                    Kidney Renal Clear Cell Carcinoma (TCGA, Nature 2013)  
 67                                    Kidney Renal Clear Cell Carcinoma (TCGA, Provisional)  
 68                                    Kidney Renal Papillary Cell Carcinoma (TCGA, Provisional)  
 69                                    Liver Hepatocellular Carcinoma (AMC, Hepatology 2014)  
 70                                    Liver Hepatocellular Carcinoma (RIKEN, Nat Genet 2012)  
 71                                    Liver Hepatocellular Carcinoma (TCGA, Provisional)  
 72                                    Low-Grade Gliomas (UCSF, Science 2014)  
 73                                    Lung Adenocarcinoma (Broad, Cell 2012)  
 74                                    Lung Adenocarcinoma (MSKCC 2015)  
 75                                    Lung Adenocarcinoma (TCGA, Nature 2014)  
 76                                    Lung Adenocarcinoma (TCGA, Provisional)  
 77                                    Lung Adenocarcinoma (TSP, Nature 2008)  
 78                                    Lung Squamous Cell Carcinoma (TCGA, Nature 2012)  
 79                                    Lung Squamous Cell Carcinoma (TCGA, Provisional)  
 80                                    Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (TCGA, Provisional)  
 81                                    Malignant Peripheral Nerve Sheath Tumor (MSKCC, Nat Genet 2014)  
 82                                    Malignant Pleural Mesothelioma (NYU, Cancer Res 2015)  
 83                                    Mantle Cell Lymphoma (IDIBIPS, PNAS 2013)  
 84                                    Medulloblastoma (Broad, Nature 2012)  
 85                                    Medulloblastoma (ICGC, Nature 2012)  
 86                                    Medulloblastoma (PCGP, Nature 2012)  
 87                                    Melanoma (Broad/Dana Farber, Nature 2012)  
 88                                    Merged Cohort of LGG and GBM (TCGA, Cell 2016)  
 89                                    Mesothelioma (TCGA, Provisional)  
 90                                    Metastatic Prostate Cancer, SU2C/PCF Dream Team (Robinson et al., Cell 2015)

91 Multiple Myeloma (Broad, Cancer Cell 2014)  
 92 Multiregion Sequencing of Clear Cell Renal Cell Carcinoma (IRC, Nat Genet 2014).  
 93 Mutational profiles of metastatic breast cancer (France, 2016)  
 94 Myelodysplasia (Tokyo, Nature 2011)  
 95 NCI-60 Cell Lines (NCI, Cancer Res. 2012)  
 96 Nasopharyngeal Carcinoma (Singapore, Nat Genet 2014)  
 97 Neuroblastoma (AMC Amsterdam, Nature 2012)  
 98 Neuroblastoma (Broad, Nat Genet 2013)  
 99 Neuroendocrine Prostate Cancer (Trento/Cornell/Broad 2016)  
 100 Oral Squamous Cell Carcinoma (MD Anderson, Cancer Discov 2013)  
 101 Ovarian Serous Cystadenocarcinoma (TCGA, Nature 2011)  
 102 Ovarian Serous Cystadenocarcinoma (TCGA, Provisional)  
 103 Pan-Lung Cancer (TCGA, Nat Genet 2016)  
 104 Pancreatic Adenocarcinoma (ICGC, Nature 2012)  
 105 Pancreatic Adenocarcinoma (QCMG, Nature 2016)  
 106 Pancreatic Adenocarcinoma (TCGA, Provisional)  
 107 Pancreatic Cancer (UTSW, Nat Commun 2015)  
 108 Pancreatic Neuroendocrine Tumors (Johns Hopkins University, Science 2011)  
 109 Papillary Thyroid Carcinoma (TCGA, Cell 2014)  
 110 Pediatric Ewing Sarcoma (DFCI, Cancer Discov 2014)  
 111 Pheochromocytoma and Paraganglioma (TCGA, Provisional)  
 112 Poorly-Differentiated and Anaplastic Thyroid Cancers (MSKCC, JCI 2016)  
 113 Primary Central Nervous System Lymphoma (Mayo Clinic, Clin Cancer Res 2015)  
 114 Prostate Adenocarcinoma (Broad/Cornell, Cell 2013)  
 115 Prostate Adenocarcinoma (Broad/Cornell, Nat Genet 2012)  
 116 Prostate Adenocarcinoma (Fred Hutchinson CRC, Nat Med 2016)  
 117 Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)  
 118 Prostate Adenocarcinoma (TCGA, Cell 2015)  
 119 Prostate Adenocarcinoma (TCGA, Provisional)  
 120 Prostate Adenocarcinoma CNA study (MSKCC, PNAS 2014)  
 121 Prostate Adenocarcinoma Organoids (MSKCC, Cell 2014)  
 122 Prostate Adenocarcinoma, Metastatic (Michigan, Nature 2012)  
 123 Recurrent and Metastatic Head & Neck Cancer (JAMA Oncology, 2016)  
 124 Renal Non-Clear Cell Carcinoma (Genentech, Nat Genet 2014)  
 125 Rhabdomyosarcoma (NIH, Cancer Discov 2014)  
 126 Sarcoma (MSKCC/Broad, Nat Genet 2010)  
 127 Sarcoma (TCGA, Provisional)  
 128 Skin Cutaneous Melanoma (Broad, Cell 2012)  
 129 Skin Cutaneous Melanoma (TCGA, Provisional)  
 130 Skin Cutaneous Melanoma (Yale, Nat Genet 2012)  
 131 Small Cell Carcinoma of the Ovary (MSKCC, Nat Genet 2014)  
 132 Small Cell Lung Cancer (CLCGP, Nat Genet 2012)  
 133 Small Cell Lung Cancer (Johns Hopkins, Nat Genet 2012)  
 134 Small Cell Lung Cancer (U Cologne, Nature 2015)  
 135 Stomach Adenocarcinoma (Pfizer and UHK, Nat Genet 2014)  
 136 Stomach Adenocarcinoma (TCGA, Nature 2014)  
 137 Stomach Adenocarcinoma (TCGA, Provisional)  
 138 Stomach Adenocarcinoma (U Tokyo, Nat Genet 2014)  
 139 Stomach Adenocarcinoma (UHK, Nat Genet 2011)  
 140 TCGA data for Esophagus-Stomach Cancers (TCGA, Nature 2017)



```

141                                     Testicular Germ Cell Cancer (TCGA, Provisional)
142                                     Thymic Epithelial Tumors (NCI, Nat Genet 2014)
143                                     Thymoma (TCGA, Provisional)
144                                     Thyroid Carcinoma (TCGA, Provisional)
145                                     Unclassified Renal Cell Carcinoma (MSKCC 2016)
146         Uterine Carcinosarcoma (Johns Hopkins University, Nat Commun 2014)
147                                     Uterine Carcinosarcoma (TCGA, Provisional)
148         Uterine Corpus Endometrial Carcinoma (TCGA, Nature 2013)
149         Uterine Corpus Endometrial Carcinoma (TCGA, Provisional)
150                                     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_tcg')[,c(1:2)]

      genetic_profile_id
1          gbm_tcg_rppa
2    gbm_tcg_rppa_Zscores
3          gbm_tcg_gistic
4          gbm_tcg_mrna_U133
5    gbm_tcg_mrna_U133_Zscores
6    gbm_tcg_mrna_median_Zscores
7    gbm_tcg_rna_seq_v2_mrna
8    gbm_tcg_rna_seq_v2_mrna_median_Zscores
9          gbm_tcg_linear_CNA
10         gbm_tcg_methylation_hm27
11         gbm_tcg_methylation_hm450
12         gbm_tcg_mutations
13         gbm_tcg_mrna
      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 z-Scores (microarray)
7    mRNA expression (RNA Seq V2 RSEM)
8    mRNA Expression z-Scores (RNA Seq V2 RSEM)
9    Relative linear copy-number values

```

```

10             Methylation (HM27)
11             Methylation (HM450)
12             Mutations
13             mRNA expression (microarray)

```

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)]

      case_list_id
1   gbm_tcga_3way_complete
2   gbm_tcga_all
3   gbm_tcga_sequenced
4   gbm_tcga_cna
5   gbm_tcga_methylation_hm27
6   gbm_tcga_methylation_hm450
7   gbm_tcga_rna_seq_v2_mrna
8   gbm_tcga_mrna_U133
9   gbm_tcga_rppa
10  gbm_tcga_cnaseq

      case_list_name
1   All Complete Tumors
2   All Tumors
3   Sequenced Tumors
4   Tumor Samples with CNA data
5   Tumor Samples with methylation data (HM27)
6   Tumor Samples with methylation data (HM450)
7   Tumor Samples with mRNA data (RNA Seq V2)
8   Tumor Samples with mRNA data (U133 microarray only)
9   Tumor Samples with RPPA data
10  Tumor Samples with sequencing and CNA data

```

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(1:5,)]
```

	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. 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(1:5),]
```

	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

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),]
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

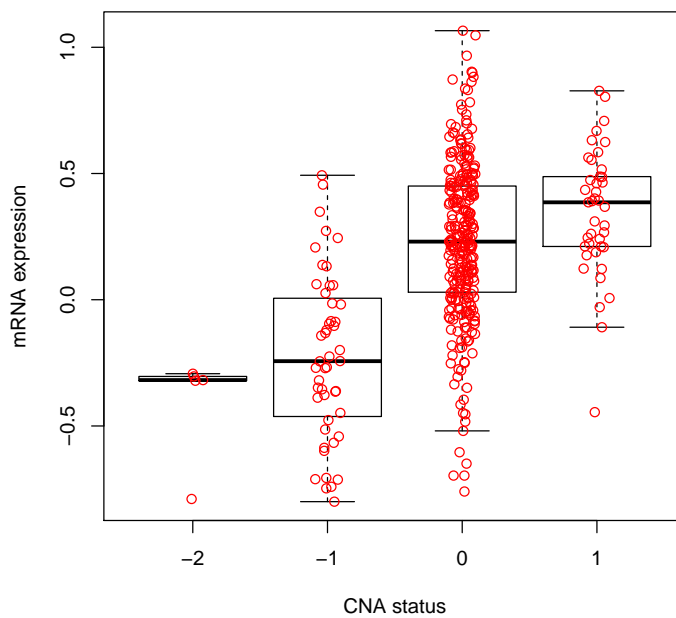
As a simple example, we will generate a plot of the association between copy number alteration (CNA) status and mRNA expression change for the NF1 tumor suppressor gene in glioblastoma. This plot is very similar to Figure 2b in the TCGA research network paper on glioblastoma (McLendon et al. 2008). The 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")
> 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

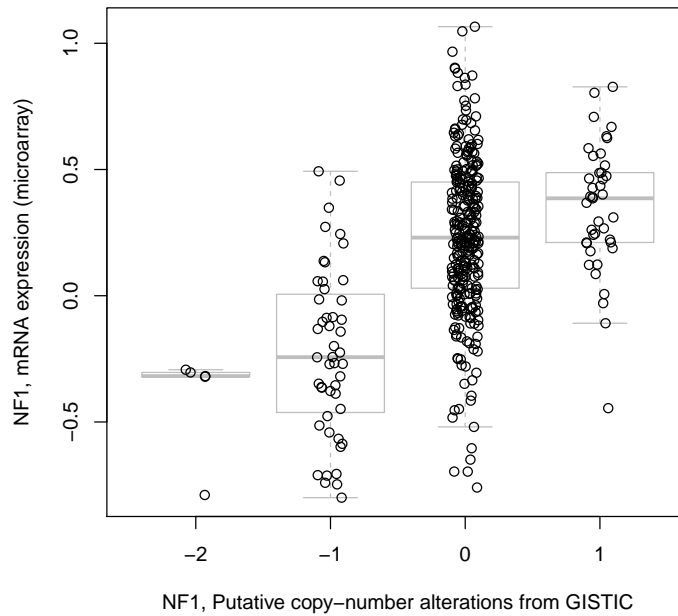
```
> boxplot(df[,2] ~ df[,1], main="NF1 : CNA status vs mRNA expression", xlab="CNA status",
> stripchart(df[,2] ~ df[,1], vertical=T, add=T, method="jitter",pch=1,col='red')
```

**NF1 : CNA status vs mRNA expression**



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  
[1] TRUE
```



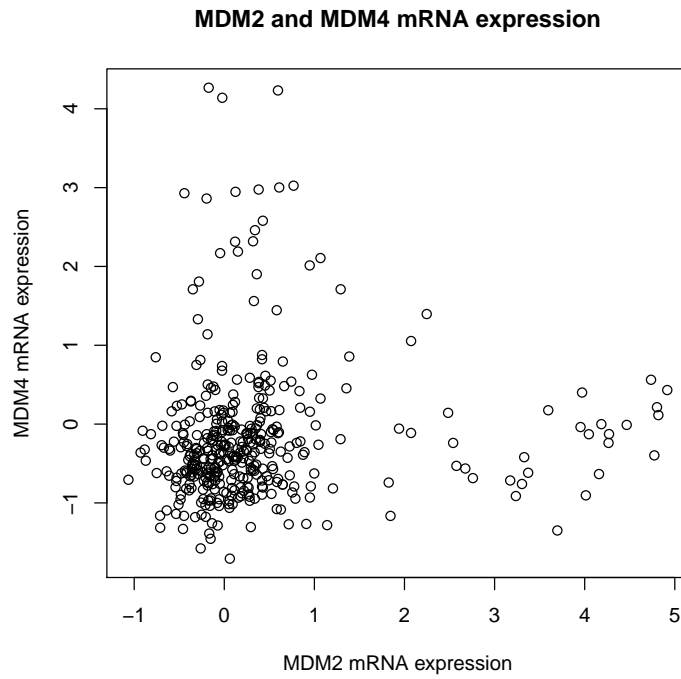
### 3.2 Example 2: MDM2 and MDM4 mRNA expression levels in glioblastoma

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).

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

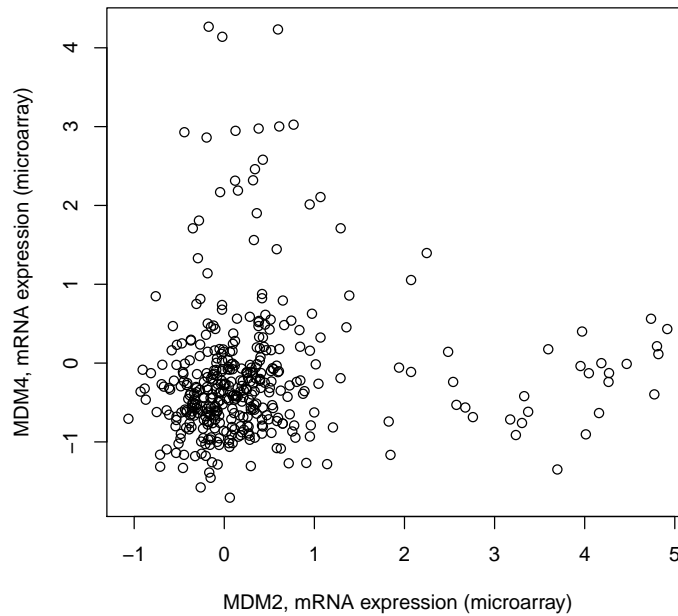
	MDM2	MDM4
TCGA.02.0001.01	NaN	NaN
TCGA.02.0003.01	NaN	NaN
TCGA.02.0006.01	NaN	NaN
TCGA.02.0007.01	NaN	NaN
TCGA.02.0009.01	NaN	NaN
TCGA.02.0010.01	NaN	NaN

```
> 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")  
[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", "prad_mskcc_primary")
> 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", "prad_mskcc_mets")
> 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')
```

