

The CGDS-R library

Anders Jacobsen

December 6, 2017

Contents

1	Introduction	1
2	The CGDS R interface	2
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	10
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	12
3	Examples	12
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.cbiportal.org/>, as an argument.

```
> library(cgdsr)
> # Create CGDS object
> mycgds = CGDS("http://www.cbiportal.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	ampca_bcm_2016
11	blca_mskcc_solit_2014
12	blca_mskcc_solit_2012
13	blca_plasmacytoid_mskcc_2016
14	blca_bgi
15	blca_dfarber_mskcc_2014
16	blca_tcga_pub
17	blca_tcga
18	lgg_tcga
19	brca_metabric
20	brca_bccrc
21	brca_broad
22	brca_sanger
23	brca_tcga_pub2015
24	brca_tcga_pub
25	brca_tcga
26	brca_bccrc_xenograft_2014
27	cellline_ccle_broad
28	cesc_tcga
29	chol_nccs_2013
30	chol_nus_2012
31	chol_tcga
32	lcll_broad_2013
33	c1l_iuopa_2015
34	ccrcc_utokyo_2013
35	coadread_dfci_2016
36	coadread_genentech
37	coadread_tcga_pub
38	coadread_tcga
39	coadread_mskcc
40	ctcl_columbia_2015
41	cscd_dfarber_2015

42	pact_jhu_2011
43	desm_broad_2015
44	dlbc_broad_2012
45	esca_broad
46	esca_tcga
47	escc_icgc
48	escc_ucla_2014
49	es_iocurie_2014
50	gbc_shanghai_2014
51	egc_tmucih_2015
52	nsclc_unito_2016
53	prad_cpcg_2017
54	gct_msk_2016
55	gbm_tcga_pub2013
56	gbm_tcga_pub
57	gbm_tcga
58	hnsb_broad
59	hnsb_jhu
60	hnsb_tcga_pub
61	hnsb_tcga
62	liad_inserm_fr_2014
63	hcc_inserm_fr_2015
64	all_stjude_2013
65	all_stjude_2015
66	panet_shanghai_2013
67	chol_jhu_2013
68	kich_tcga_pub
69	kich_tcga
70	kirc_bgi
71	kirc_tcga_pub
72	kirc_tcga
73	kirp_tcga
74	lihc_amc_prv
75	lihc_riken
76	lihc_tcga
77	lgg_ucsf_2014
78	luad_broad
79	luad_mskcc_2015
80	luad_tcga_pub
81	luad_tcga
82	luad_tsp
83	lusc_tcga_pub
84	lusc_tcga
85	dlbc_tcga
86	msk_impact_2017
87	lung_msk_2017
88	prad_mskcc_2017
89	mpnst_mskcc
90	plmeso_nyu_2015
91	mcl_idibips_2013

92	mbl_broad_2012
93	mbl_icgc
94	mbl_pcgp
95	mbl_sickkids_2016
96	skcm_broad_dfarber
97	lgggbm_tcga_pub
98	meso_tcga
99	prad_su2c_2015
100	mm_broad
101	ccrcc_irc_2014
102	brca_igr_2015
103	mds_tokyo_2011
104	cellline_nci60
105	odg_msk_2017
106	npc_nusingapore
107	nbl_amc_2012
108	nbl_ucologne_2015
109	nepc_wcm_2016
110	skcm_vanderbilt_mskcc_2015
111	hnsc_mdanderson_2013
112	ov_tcga_pub
113	ov_tcga
114	mel_tsam_liang_2017
115	nsclc_tcga_broad_2016
116	paad_icgc
117	paad_qcmg_uq_2016
118	paad_tcga
119	paad_utsu_2015
120	panet_jhu_2011
121	thca_tcga_pub
122	es_dfarber_broad_2014
123	pcpg_tcga
124	thyroid_mskcc_2016
125	pcnsl_mayo_2015
126	prad_broad_2013
127	prad_broad
128	prad_fhcr
129	prad_mskcc
130	prad_tcga_pub
131	prad_tcga
132	prad_mskcc_2014
133	prad_mskcc_chenyl_organoids_2014
134	prad_mich
135	hnc_mskcc_2016
136	nccrcc_genentech_2014
137	rms_nih_2014
138	sarc_mskcc
139	sarc_tcga
140	skcm_broad
141	skcm_tcga

142 skcm_yale
 143 scco_mskcc
 144 sclc_clcgp
 145 sclc_jhu
 146 sclc_ucologne_2015
 147 stad_pfizer_uhongkong
 148 stad_tcga_pub
 149 stad_tcga
 150 stad_utokyo
 151 stad_uhongkong
 152 stes_tcga_pub
 153 egc_msk_2017
 154 urcc_mskcc_2016
 155 crc_msk_2017
 156 tgct_tcga
 157 brca_mbcproject_wagle_2017
 158 tet_nci_2014
 159 thym_tcga
 160 thca_tcga
 161 ucs_jhu_2014
 162 ucs_tcga
 163 ucec_tcga_pub
 164 ucec_tcga
 165 uvm_tcga
 166 panet_arcnet_2017
 167 skcm_ucla_2016
 168 past_dkfz_heidelberg_2013

1 Acinar Cell Carcinoma of the Pancreas (Johns Hopkins)
 2 Acute Myeloid Leukemia
 3 Acute Myeloid Leukemia
 4 Adenoid Cystic Carcinoma (Flinn)
 5 Adenoid Cystic Carcinoma (Flinn)
 6 Adenoid Cystic Carcinoma (Flinn)
 7 Adenoid Cystic Carcinoma (Flinn)
 8 Adenoid Cystic Carcinoma of the Breast
 9 Adrenocortical Carcinoma
 10 Ampullary Carcinoma (Baylor College of Medicine)
 11 Bladder Cancer
 12 Bladder Cancer
 13 Bladder Cancer, Plasmacytoid Variant
 14 Bladder Urothelial Carcinoma
 15 Bladder Urothelial Carcinoma (Dana Farber & MIT)
 16 Bladder Urothelial Carcinoma
 17 Bladder Urothelial Carcinoma
 18 Brain Lower Grade Glioma
 19 Breast Cancer (METABRIC, Nature)
 20 Breast Invasive Carcinoma (British Columbia)
 21 Breast Invasive Carcinoma
 22 Breast Invasive Carcinoma

23 Breast Invasive C
 24 Breast Invasive Car
 25 Breast Invasive Car
 26 Breast cancer patient xenografts (Brit
 27 Cancer Cell Line Encyclopedia (No
 28 Cervical Squamous Cell Carcinoma and Endocervical Adenocar
 29 Cholangiocarcinoma (National Cancer Centre of
 30 Cholangiocarcinoma (National University of
 31 Cholangiocar
 32 Chronic Lymphocytic L
 33 Chronic Lymphocytic Leu
 34 Clear Cell Renal Cell Carcinoma
 35 Colorectal Adenocarcinoma
 36 Colorectal Adenocarcinom
 37 Colorectal Adenocar
 38 Colorectal Adenocar
 39 Colorectal Adenocarcinoma Triplets
 40 Cutaneous T Cell Lymphoma (C
 41 Cutaneous squamous cell carcinoma (D
 42 Cystic Tumor of the Pancreas
 43 Desmoplastic Melanoma (Broad
 44 Diffuse Large B-Cell L
 45 Esophageal Adenocarcinon
 46 Esophageal Car
 47 Esophageal Squamous Cell Car
 48 Esophageal Squamous Cell Carcin
 49 Ewing Sarcoma (Institut
 50 Gallbladder Carcinoma
 51 Gastric Adenocar
 52 Genetic Characterization of NSCLC young adult patients (University o
 53 Genomic Hallmarks of Prostate Adenocarcinon
 54 Genomic Profile of Patients with Advanced Germ Ce
 55 Glio
 56 Gliobl
 57 Glioblastoma Mult
 58 Head and Neck Squamous Cell Carci
 59 Head and Neck Squamous Cell Carcinoma (Jo
 60 Head and Neck Squamous Cell Car
 61 Head and Neck Squamous Cell Car
 62 Hepatocellular Adenoma
 63 Hepatocellular Carcinoma
 64 Hypodiploid Acute Lymphoid Leukemia
 65 Infant MLL-Rearranged Acute Lymphoblastic Leukemia
 66 Insulinoma (I
 67 Intrahepatic Cholangiocarcinoma (Johns Hopkins U
 68 Kidney Chromophob
 69 Kidney Chrom
 70 Kidney Renal Clear Cell Carci
 71 Kidney Renal Clear Cell Car
 72 Kidney Renal Clear Cell Car

73	Kidney Renal Papillary Cell Car
74	Liver Hepatocellular Carcin
75	Liver Hepatocellular Carcin
76	Liver Hepatocellular Car
77	Low-Grade Gli
78	Lung Adenoca
79	Lung A
80	Lung Adenocar
81	Lung Adenocar
82	Lung Adenoca
83	Lung Squamous Cell Car
84	Lung Squamous Cell Car
85	Lymphoid Neoplasm Diffuse Large B-cell Ly
86	MSK-IMPACT Clinical Sequencing Co
87	MSK-IMPACT Clinical Sequencing Cohort for Non-Small Cell Cancer (M
88	MSK-IMPACT Clinical Sequencing Cohort in Prostate Cancer (MSK, JC
89	Malignant Peripheral Nerve Sheath Tum
90	Malignant Pleural Mesotheli
91	Mantle Cell Lym
92	Medullobla
93	Medullobl
94	Medullobl
95	Medulloblasto
96	Melanoma (Broad
97	Merged Cohort of LGG
98	Mesoth
99	Metastatic Prostate Cancer, SU2C/PCF Dream Team (R
100	Multiple Myeloma
101	Multiregion Sequencing of Clear Cell Renal Cell Carci
102	Mutational profiles of metastatic br
103	Myelodysp
104	NCI-60 Cell Lin
105	NGS in Anaplastic Oligodendroglioma and Anaplastic Oligoastrocytomas tumo
106	Nasopharyngeal Carcinoma (
107	Neuroblastoma (A
108	Neuroblasto
109	Neuroendocrine Prostate Cancer (
110	Next generation sequencing (NGS) of pre-treatment metastatic melanoma samples (MSK, JC
111	Oral Squamous Cell Carcinoma (MD And
112	Ovarian Serous Cystadenocar
113	Ovarian Serous Cystadenocar
114	Paired-exome sequencing of acral melano
115	Pan-Lung Can
116	Pancreatic Adenocar
117	Pancreatic Adenocar
118	Pancreatic Adenocar
119	Pancreatic Canc
120	Pancreatic Neuroendocrine Tumors (Johns Hopkins
121	Papillary Thyroid C
122	Pediatric Ewing Sarcoma

123	Pheochromocytoma and Paragan
124	Poorly-Differentiated and Anaplastic Thyroid
125	Primary Central Nervous System Lymphoma (Mayo Cli
126	Prostate Adenocarcinoma
127	Prostate Adenocarcinoma (Broa
128	Prostate Adenocarcinoma (Fred Hut
129	Prostate Adenocarcinoma
130	Prostate Adenoc
131	Prostate Adenocar
132	Prostate Adenocarcinoma CN
133	Prostate Adenocarcinoma Or
134	Prostate Adenocarcinoma, Metastat
135	Recurrent and Metastatic Head & Neck Cance
136	Renal Non-Clear Cell Carcinoma (
137	Rhabdomyosarcoma
138	Sarcoma (MS
139	S
140	Skin Cutaneous M
141	Skin Cutaneous Me
142	Skin Cutaneous Melan
143	Small Cell Carcinoma of the Ova
144	Small Cell Lung Canc
145	Small Cell Lung Cancer (John
146	Small Cell Lung Cance
147	Stomach Adenocarcinoma (Pfizer
148	Stomach Adenocar
149	Stomach Adenocar
150	Stomach Adenocarcinoma
151	Stomach Adenocarci
152	TCGA data for Esophagus-Stomach C
153	Targeted Sequencing of 341 samples from metastatic esophagogastric cancer patients (
154	Targeted gene sequencing in 62 high-grade primary Unclassified Renal Cell Ca
155	Targeted sequencing of 1134 samples from metastatic colorectal cancer sampl
156	Testicular Germ Cell
157	The Metastatic Breast Cancer Project (
158	Thymic Epithelial Tu
159	T
160	Thyroid Car
161	Uterine Carcinosarcoma (Johns Hopkins Un
162	Uterine Carcinos
163	Uterine Corpus Endometrial Car
164	Uterine Corpus Endometrial Car
165	Uveal Me
166	Whole-Genome Sequencing of Pancreatic Neuroendoc
167	Whole-exome sequences (WES) of pretreatment melanom
168	Whole-genome sequencing of pilocytic astrocytoma

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)]

      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)]
```

	case_list_id	case_list_name
1	gbm_tcga_3way_complete	All Complete Tumors
2	gbm_tcga_all	All Tumors
3	gbm_tcga_sequenced	Sequenced Tumors
4	gbm_tcga_cna	Tumor Samples with CNA data
5	gbm_tcga_methylation_hm27	Tumor Samples with methylation data (HM27)
6	gbm_tcga_methylation_hm450	Tumor Samples with methylation data (HM450)
7	gbm_tcga_mrna	Tumor Samples with mRNA data (Agilent microarray)
8	gbm_tcga_rna_seq_v2_mrna	Tumor Samples with mRNA data (RNA Seq V2)
9	gbm_tcga_mrna_U133	Tumor Samples with mRNA data (U133 microarray only)
10	gbm_tcga_rppa	Tumor Samples with RPPA data
11	gbm_tcga_cnaseq	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(1:2)]
```

	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

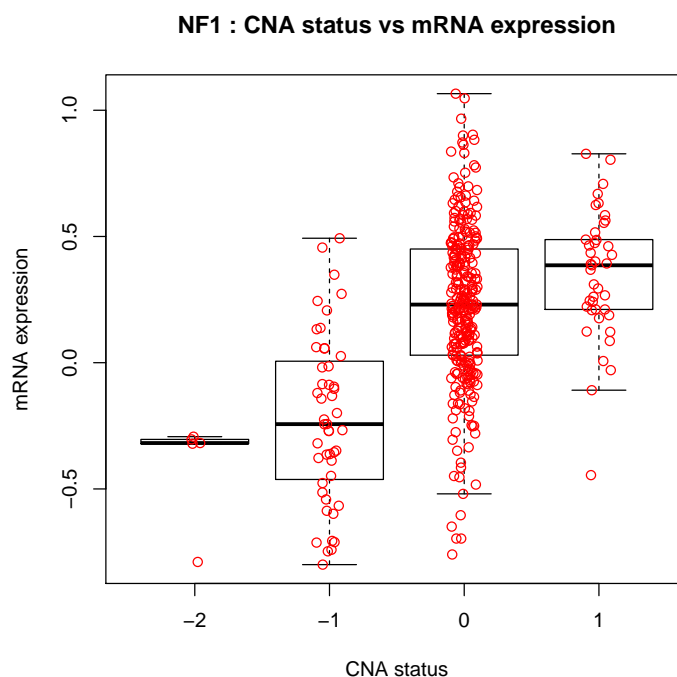
As a simple example, we will generate a plot of the association between copy number alteration (CNA) status and mRNA expression change for the NF1

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)
```

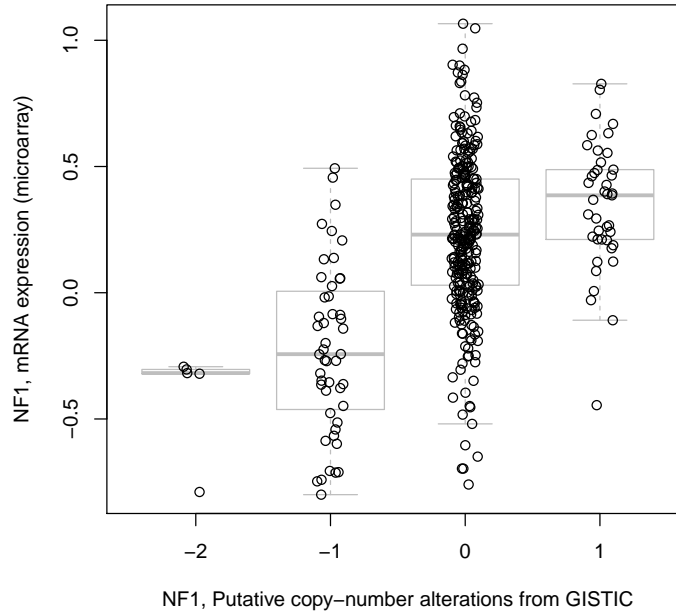
	gbm_tcga_gistic	gbm_tcga_mrna
TCGA.02.0001.01	-1	NaN
TCGA.02.0003.01	0	NaN
TCGA.02.0006.01	0	NaN
TCGA.02.0007.01	0	NaN
TCGA.02.0009.01	0	NaN
TCGA.02.0010.01	0	NaN

```
> 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')
```



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
```



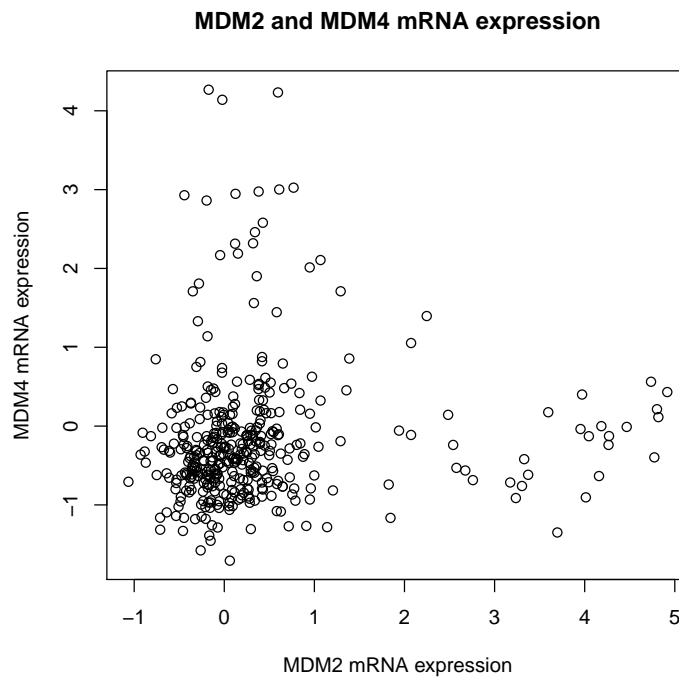
3.2 Example 2: MDM2 and MDM4 mRNA expression levels in glioblastoma

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)
```

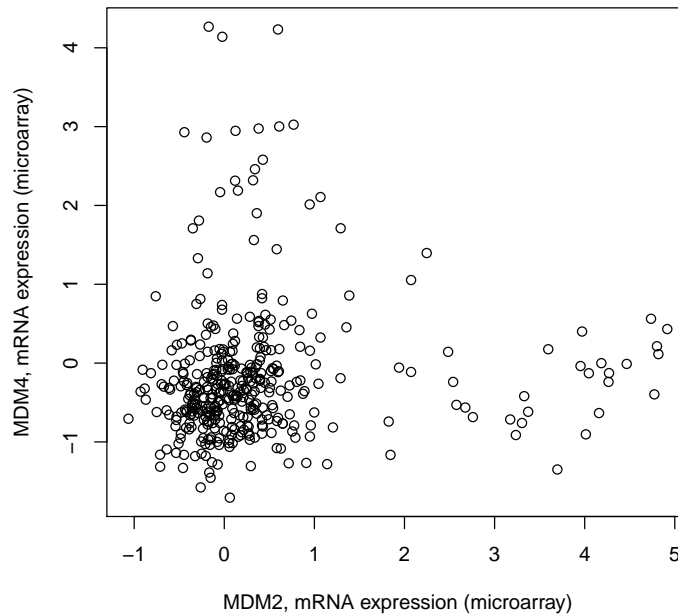
	MDM2	MDM4
TCGA.02.0001.01	NaN	NaN
TCGA.02.0003.01	NaN	NaN
TCGA.02.0006.01	NaN	NaN
TCGA.02.0007.01	NaN	NaN
TCGA.02.0009.01	NaN	NaN
TCGA.02.0010.01	NaN	NaN

```
> 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_median_Zscores", "prad_mskcc_pri")
> 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_met")
> 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')
```

