

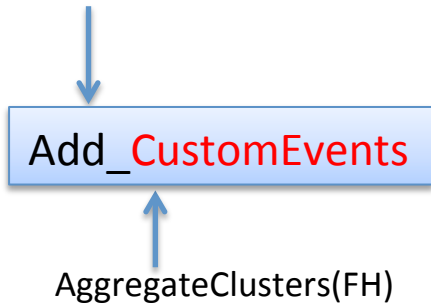
customEvent modules user manual

July 18 2013

BROAD GDAC Team, Juok Cho

1. To import categorical subtype events

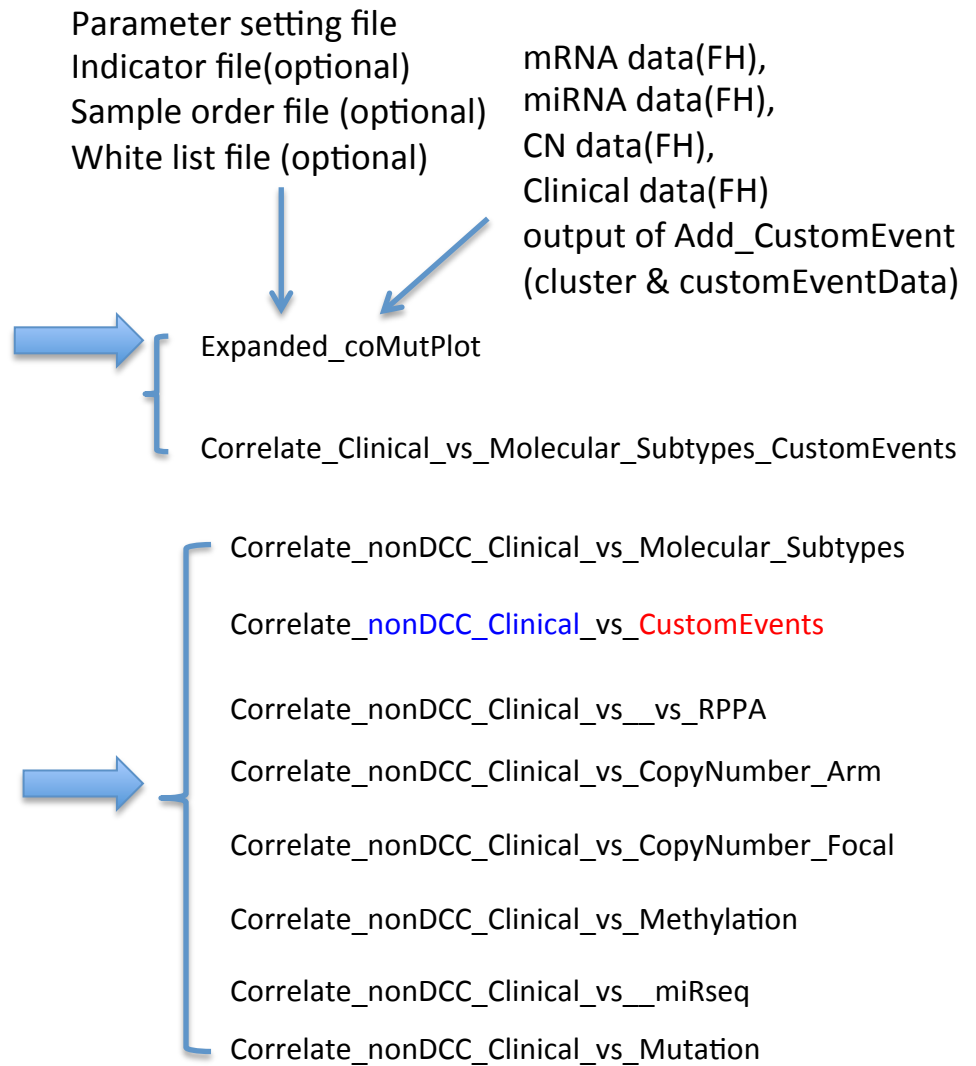
CustomEvents data
(e.g. fusion event data: categorical data)



2. To import nonDCC_Clinical event data

Add_CustomEvents_nonDCC_Clinical

nonDCC_Clinical data
(e.g. risk category data: categorical type data
MACIS score data: continuous type data)



1. How to create custom event data: naming convention

1. Example of events

- fusion, mtCN, ... etc. (CN,mRNA, miRNA will be changed to gene name list)

1. File name

- **eventName.event.categorical.txt**

- **eventName.event.continuous.txt**

(NOTE: eventName should be unique)

2. Data format

1. Column name

- **SampleName, eventName_a, eventName_b, ...**

2. Sample ID in the 'SampleName' column

For categorical event, it should be the same as in 'aggregate_Molecular_Subtype_Clusters'
(e.g. TCGA-BJ-A0YZ-01)

For nonDCC clinical event, it should be the same as in 'Correlate_Clinical_Vs ~'
(e.g. TCGA-BJ-A0YZ)

3. Optional data for expanded coMut plot

- whitelist.txt (genes to include)
- sampleorder.txt (sorting order by important events)
- indicator.txt (e.g. histological type in the panel upper mutation spectrum with legend on the top right coner)

e.g.) Data format

fusion.event.categorical.txt

SampleName	fusion_RET	fusion_PAX8	fusion_NTRK	fusion_NTRK	fusion_BRAF
TCGA-ET-A3BO-01	0	0	0	0	0
TCGA-IM-A3U3-01	NA	NA	NA	NA	NA
TCGA-EL-A3GY-01	0	0	0	0	0
TCGA-DJ-A2PY-01	0	0	0	0	0
TCGA-EL-A3N3-01	0	0	0	0	0
TCGA-EL-A3CN-01	0	0	0	0	0

Risk.event.categorical.txt

1	SampleName	Risk
2	TCGA-BJ-A0YZ	Low
3	TCGA-BJ-A0Z0	Low
4	TCGA-BJ-A0Z2	High
5	TCGA-BJ-A0Z3	Low
6	TCGA-BJ-A0Z5	Intermediate
7	TCGA-BJ-A0Z9	Intermediate
8	TCGA-BJ-A0ZA	Low
9	TCGA-BJ-A0ZB	Intermediate
10	TCGA-BJ-A0ZC	Low
11	TCGA-BJ-A0ZE	Intermediate

MACIS.event.continuous.txt

	A	B
1	SampleName	MACIS
2	TCGA-BJ-A0YZ	6.1
3	TCGA-BJ-A0Z0	5.45
4	TCGA-BJ-A0Z2	8.4
5	TCGA-BJ-A0Z3	3.49
6	TCGA-BJ-A0Z5	7.44
7	TCGA-BJ-A0Z9	7.4
8	TCGA-BJ-A0ZA	6.38
9	TCGA-BJ-A0ZB	8
10	TCGA-BJ-A0ZC	5.45

Whitelist.txt

BRAF
HRAS
NRAS
KRAS
TG
EMG1
EIF1AX

Sampleorder.txt

sample_order
TCGA-ET-A3BO-01
TCGA-IM-A3U3-01
TCGA-EL-A3GY-01
TCGA-DJ-A2PY-01

Indicator.txt

Histological_Type	sample_id
classical	TCGA-BJ-A0ZF-01
classical	TCGA-BJ-A28R-01
classical	TCGA-BJ-A28T-01
classical	TCGA-BJ-A3PR-01

2. Where to put customEvent data

1. Create a directory for <disease type> under reference directory

TCGA AWG reference directory : /xchip/cga/reference/tcga/custom_user_input/

2. Create a directory for each awg run consistent with date in your awg run sample stamp
e.g.)

```
bash:cga1:/xchip/cga/reference/tcga/custom_user_input 1152 $ ls
PANCAN12 THCA
bash:cga1:/xchip/cga/reference/tcga/custom_user_input 1153 $ ls THCA/
2012_03_18 2013_03_18 2013_06_06 2013_06_20 2013_07_01
```

- * Inquire Mike N for permission issue

3. Create folder for each sample set

```
bash:cga1:/xchip/cga/reference/tcga/custom_user_input/THCA/2013_06_20 1129 $ ls
THCA THCA-HistologicalType_Follicular THCA-Mut_BRAF THCA-Mut_RAS THCA-NT THCA-RiskCategory_Intermediate THCA-TM
THCA-HistologicalType_Classical THCA-HistologicalType_Tall_cell THCA-Mut_DARK THCA-NB THCA-RiskCategory_High THCA-RiskCategory_Low THCA-TP
```

1. Put customEvent files under each sample set folder

e.g.) In the custom_user_input_dir for categorical event data

```
bash:cga1:/xchip/cga/reference/tcga/custom_user_input/THCA/2013_06_20/THCA-TP 1133 $ ls
BCclusters.event.categorical.txt custom_user_inputB_dir indicator.txt miRNA.event.continuous.txt whitelist.txt
CN.event.continuous.txt fusion.event.categorical.txt mRNA.event.continuous.txt sampleorder.txt
```

e.g.) In the custom_user_inputB_dir for nonDCC clinical data

```
bash:cga1:/xchip/cga/reference/tcga/custom_user_input/THCA/2013_06_20/THCA-TP/custom_user_inputB_dir 1140 $ ls
M.event.categorical.txt MACIS.event.continuous.txt N.event.categorical.txt N_neg_Nlb_outcome.event.categorical.txt Risk.event.categorical.txt T.event.categorical.txt
```

3. How to let FH know the location of customEvent data

: You need to set a directory annotation for your TCGA awg reference directory

1. For categorical customEvent data

'fiss annot_set <workspaceName> <sset=samplesetName> **custom_user_input_dir** <directory path>'

e.g. fiss annot_set awg_thca__2013_06_20 sset=THCA-classical custom_user_input_dir /xchip/cga/reference/tcga/custom_user_input/THCA/2013_03_18/THCA-classical

2. For nonDCC clinical data

'fiss annot_set <workspaceName> <sset=samplesetName> **custom_user_inputB_dir** <directory path>'

e.g. fiss annot_set awg_thca__2013_06_20 sset=THCA custom_user_inputB_dir /xchip/cga2/jcho/THCA/data_freeze/reference.dir_06_06/THCA/customEvent_B

NOTE: For each sample set, it requires setting these annotations for each sample set

TIP. How to rerun?

:To re-run 'Add_CustomEvent' with updated files under reference directory, you can create symlink to your reference directory and use different symlink name for updated <directory path> value when you re-do 'fiss annot_set'. Otherwise, FH doesn't realize the file under the reference directory is changed.

e.g. reference.dir.0620 -> /xchip/cga/reference/tcga/custom_user_input/THCA/2013_06_20