

calnt 0007 - Format of caArray supplementary files that can be loaded into calIntegrator

Question: What is the format of caArray supplementary files that can be loaded into calIntegrator?

Topic: calIntegrator Usage

Release: calIntegrator 1.2 and above

Date entered: 08/31/2010

Answer

This knowledge base entry describes the format for creating supplemental files for use by calIntegrator. The supplemental files described here are to be added to an experiment in caArray prior to configuring a study in calIntegrator.

The file itself is a tab-delimited text file. The file extension can be anything, though .txt is typically used. The name of each supplemental file has to be unique within a caArray experiment.

Inside the file, each row in the file contains the data from one reporter. Each column in the file must have a unique header name. (That is, you can not give two different columns the same column name.)

There are 2 supported formats:

1. SINGLE SAMPLE FORMAT
2. MULTIPLE SAMPLE FORMAT

SINGLE SAMPLE FORMAT

The format requirements for the single sample format file are following:

1. Minimum of two required columns.
2. One column must contain the reporter/probe name.
3. One column must contain the value be reported by the reporter.
4. The file can have additional columns, though other than reporter/probe name and value mentioned above, the rest will be ignored.
5. One SINGLE SAMPLE FORMAT file for each sample in the experiment.

Refer to the illustration below for an example of SINGLE SAMPLE FORMAT.

Probe	log2signalvalue
AGI HUM1 OLIGO A 23 P142277	-0.699
A 24 P255523	0.259
AGI HUM1 OLIGO A 23 P350454	0.34
VIRS 10001995	0.149
A 24 P381602	1.151
VIRS 20002073	0.115
NM 002046 GAPD 3END 1001	-0.853
AGI HUM1 OLIGO A 24 P659906	-0.009
AGI HUM1 OLIGO A 23 P32285	-0.384
AGI HUM1 OLIGO A 32 P516804	1.15
A 24 P307858	0.262
AGI HUM1 OLIGO A 32 P2333	-0.384
AGI HUM1 OLIGO A 24 P352608	1.475

MULTIPLE SAMPLE FORMAT

The format requirements for the multiple sample format file are following:

1. one column must contain the reporter/probe name.
2. each additional column contains the reporter values such that there is one column per sample.
3. one MULTIPLE SAMPLE FORMAT file for the whole experiment.
4. CAVEAT: Currently the MULTIPLE SAMPLE FORMAT is slower to load than the SINGLE SAMPLE FORMAT for platforms other than Agilent Copy Number. This performance issue is expected to be improved in a future release.

Refer to the illustration below for an example of MULTIPLE SAMPLE FORMAT.

Probe	TCGA-02-0001-01C-01R	TCGA-02-0002-01A-01R	TCGA-02-0003-01A-01R
AGI HUM1 OLIGO A 23 P142277	-0.699	-1.16	-0.223
A 24 P255523	0.259	0.7	2.19
AGI HUM1 OLIGO A 23 P350454	0.34	0.378	0.385
VIRS 10001995	0.149	-0.013	-0.215
A 24 P381602	1.151	2.075	2.278
VIRS 20002073	0.115	-0.382	-0.108
NM 002046 GAPD 3END 1001	-0.853	-1.385	-2.335
AGI HUM1 OLIGO A 24 P659906	-0.009	-0.37	0.152
AGI HUM1 OLIGO A 23 P32285	-0.384	-0.47	-0.589
AGI HUM1 OLIGO A 23 P516804	1.15	0.636	0.825

Have a comment?

Please leave your comment in the [calIntegrator End User Forum](#).