calnt 0013 - How To Combine Annotation and Gene Expression Data When Displaying Query Results

Problem: How to combine annotation and gene expression data when displaying query results

Topic: calntegrator Usage

Date entered: 08/23/2011

Release: Up to calntegrator 1.3

Solution

calntegrator allows users to query studies by both clinical annotation data (i.e., subject age and gender) and by gene expression data (i.e., numerical reporter IDs representing gene expression levels). When displaying the results for a compound query, calntegrator can display either the annotation data or the gene expression data individually, but it lacks the ability to display the two together side-by-side.

This how-to guide presents two different workarounds for this limitation. The first, and less sophisticated workaround, involves manually exporting the annotation and expression data separately as spreadsheet files, then merging the two spreadsheets afterwards. The second workaround, only available with calntegrator version 1.3 or later, involves visualizing the data in the form of color-coded maps via a newly added featured known as the Integrative Genomics Viewer (IGV).

Querying the Data

First, let's do a sample query of a calntegrator study to illustrate the limitations in displaying query results. In our example, the study we will query is entitled 'TCGA Ovarian', which is selected in the 'Public Studies' drop-down list at the top of the calntegrator home page. To start querying, click on the 'Search TCGA Ovarian' link in the navigation menu at the left.

National Cancer Ins	titute	U.S. N
CaINTEGRATOR	2011 10 58 AMF	My Studies: (TCGA Ovariar)
TCGA 💮	Welcome to TCGA Ovarian	
	14	arre: TCGA Ovarian
TCGA OVARIAN Nome Search TCGA Ovarian	Descrip	tion: The Cancer Genome Atlas (TCGA) program was launched by the The National Ca Genome Research Institute (NHGR) to create a comprehensive atlas of the geno common types of cancer. TCGA ovarian (TCGA-OV) datasets stored in caIntegral more

To begin a sample query of the 'TCGA Ovarian' study (highlighted in red), click on the 'Search TCGA Ovarian' link (highlighted in blue).

Next, on the 'Search' page, under the 'Criteria' tab, select 'Gene Expression' from the drop-down list, then click on the 'Add' button to the right.

Search TCGA Ovarian



In the 'Gene Expression' drop-down list that appears below, select 'Gene Name', then in the 'Gene Symbol' field to its right, type in 'BRCA1', and in the 'Platform Name' drop-down list, select the entry that begins with the phrase 'Agilent'.

Define Query Criteria for: Unsaved Query		
Gene Expression		
Gene Expression Gene Name	Gene Symbol(s) (comma separated list) or blar for all genes	brcat
	Platform Name	Please Select
0.0.		AgilentG4502A_07_singleADF

Next, go back to the top drop-down list, and this time select 'patient', then click on the 'Add' button again.

Define Query Criteria for: Un	saved Que
Gene Expression drug examination patient	Add

In the 'patient' drop-down list that appears below, select 'Gender', then in the drop-down list to its right, select 'Female', and finally, click on the 'Run Query' button at the bottom.

Define Query Criteria for: Unsaved Query				
patient 💽 🔶 Add				
Gene Expression Gene Name	Gen for o Plat	e Symbol(s) (comma separated list) or bla Il genes orm Name	AgilentG4502A_07_singleADF	() () () () () () () () () () () () () (
patient	• eq	als	FEMALE	×
C or C and				
		Run Query		

The results are displayed under the 'Query Results' tab as a list of subject IDs who meet the specified criteria (i.e., female subjects with available expression data for the BRCA1 gene).

Query Results for: Unsaved Query		Results p
525 items found, displaying 1 to 20. [First/Prev] 1, 2, 3, 4, 5, 6, 7,	8 [Next/Last]	
Select Image All L None	Select Subject	Subject ID
	N	TCGA-29-1688
	v	TCGA-30-1866
	v	TCGA-13-1491
	<u>र</u>	TCGA-13-0751
	v	TCGA-04-1331

You can customize these results by first clicking on the 'Results Type' tab at the top. To show the annotation data in the results, first make sure that the 'Annotation' button is selected under the 'Select Results Type' heading, then check off 'gender' in the scrollable list box under the 'patient' heading, and finally, click on the 'Run Query' button at the bottom.

Criteria	Results Type	Sorting	Query Results	Save query	as	
Select Res	ults Type:				_	
Select Col	Anno Gene Copy Heath	C Gen lation result type - Expression result Number result type ative Genomics Vi dap Viewer - will d	e Expression C Copy will display tabular data, in type - will display a gene e - will display segmentatio ever - will display a link to the Heat M	Number C Integ including column sel hypression data mat in data with Chromot the ISV Viewer. ap Viewer.	grative G lection. rix. some posi	ienomics Viewer C Heat Map Viewer
drug					examin	ation
ber_drug_t days_to_dr days_to_dr drug_categ drug_name	barcode ug_therapy_end ug_therapy_start pory				east karr perf prog	tern_cancer_oncology_group hofsky_performance_score formance_status_scale_timing gression_determined_by
- died	onaciourni				Gener	Chibeletchi
patient					radiatio	on
days_to_tu days_to_tu gender histologica Select All	mor_progression mor_recurrence I_type therapy Unselect All				Unit anat bcr_ day: day: Selec	s tomic_treatment_site _radiation_barcode s_to_radiation_therapy_end s_to_radiation_therapy_start t All Unselect All
surgery					Select R	teporter Type: [•] Reporter Id [•] Gene
bcr_surger days_to_pr procedure	y_barcode ocedure _type				Reporter i Gene - us	d - use reporter in the gene expression criterio. e gene in the gene expression criterion.
Select All	Unselect All			Ru	in Query	

Now, the 'Query Results' tab re-appears with updated results - namely, a new column headed 'Gender' to the right of the 'Subject ID' column.

Criteria	Results Type	Sorting	Query Results	Save query as		
Query Res	ults for: Unsave	d Query			Results per Page:	20 💌 Apply
509 items found,	displaying 1 to 20. [First	(Prev] 1, 2, 3, 4, 9	5, 6, 7, 8 [Next/Last]			
Select Image All None		Se All	lect Subject		Subject ID	gender
		R	2		TCGA-13-0903	FEMALE
		F	7		TCGA-04-1652	FEMALE
		R	z		TCGA-61-2096	FEMALE
		R	7		TCGA-25-1632	FEMALE
		R	z		TCGA-29-1768	FEMALE
		F	7		TCGA-59-2354	FEMALE
		R	z		TCGA-29-1707	FEMALE
		R	7		TCGA-10-0931	FEMALE
		R	2		TCGA-13-0888	FEMALE
		R	7		TCGA-24-2023	FEMALE
		R.	z		TCGA-31-1955	FEMALE

To show the gene expression data in the results, click back on the 'Results Type' tab at the top, then select the 'Gene Expression' button under 'Select Results Type', the 'Gene' button next to 'Select Reporter Type', and the 'Genes in Columns' button next to 'Select Results Orientation'. Finally, click on the 'Run Query' button at the bottom.

Criteria	Results Type	Sorting	Query Results	Save query as	
Select Re	sults Type:	_			
	C Ann Anno Gene	otation Gene station result type - e Expression result	a Expression C Integ will display tabular data, i type - will display a gene	rative Genomics Viewer Including column selection. expression data matrix.	
	Select F Repo	Reporter Type: offer Id - will display a - will display a me	C Reporter Id Gen y all reporter ids. adian value for each gene.		
	Select F Gen Gen few /	Results Orientat to in Rows / Subjec to in Columns / Sub reporters.	ion: Genes in Columns - will display bjects in Rows - will display	nns / Subjects in Rows y the subjects and samples al y the genes and reporters alo	C Genes in Rows / Subjects in Columns ong the top and genes and reporters along the a ng the top and subjects and samples along the a
				Run Query	

The results are displayed under the 'Query Results' tab as a list of subject IDs with available BRCA1 expression data, with the corresponding sample IDs and expression levels for each subject.

Criteria	Results Type	Sorting	Query Results	Save query as		
Query Res	ults for: test					
					Gene	BRCA1
Subject ID			Sample ID			
TCGA-61-2094			TCGA-61-2094-01A			-1.75
TCGA-25-1328			TCGA-25-1328-01A			-1.71
TCGA-61-2088			TCGA-61-2088-01A			-1.08
TCGA-61-2092			TCGA-61-2092-01A			-0.98
TCGA-25-2393			TCGA-25-2393-01A			-0.9

Workaround 1: Manually Exporting Spreadsheets

So far, we've seen that calntegrator can display annotation and gene expression query results individually, but not side-by-side. One workaround for this limitation is to manually export the annotation and expression data separately as spreadsheet files, then merge the two spreadsheets afterwards. To begin, let's go back to the 'Query Results' tab for the annotation results type and click on the 'Export to CSV' button at the bottom left.

Criteria	Results Type	Sorting	Query Results	Save query as			
Query Res	sults for: test				F	lesults per Page: 20	 Apply
525 items found,	displaying 1 to 20. (Firs	(Prev] 1, 2, 3, 4,	5, 6, 7, 8 [Next/Last]				
Select Image		S	elect Subject		Subject ID		gender
		1	z		TCGA-29-2414		FEMALE
		1	7		TCGA-10-0933		FEMALE
		1	2		TCGA-24-0979		FEMALE
		1	7		TCGA-24-2290		FEMALE
		1	7		TCGA-13-0794		FEMALE
		1	7		TCGA-13-0791		FEMALE
		1	z		TCGA-30-1892		FEMALE
		1	7		TCGA-24-1467		FEMALE
		1	2		TCGA-61-2102		FEMALE
		1	7		TCGA-24-0970		FEMALE
		1	7		TCGA-25-1628		FEMALE
T							
	Export To C:	sv)	🔚 Save Subject List	🔡 View in Integ	grative Genomics Viewer	📑 View in He	at Map Viewer

A dialog will then appear prompting you whether you want to open or save the CSV file. Click on the 'Save File' button, then click on 'OK' below it.

Opening StudySearchResults.csv	×
You have chosen to open	
StudySearchResults.csv which is a: Microsoft Office Excel Comma Separated Values File from: https://caintegrator.nci.nih.gov	
What should Pale Moon do with this file?	
Save File	
Do this <u>a</u> utomatically for files like this from now on.	
OK Cancel	

Another dialog will then appear prompting you for the name and location of the saved file. Choose a convenient location to save the file to from the dropdown list at the top, then enter 'AnnotationResults.csv' for the file name and click on the 'Save' button.

Enter name of fil	e to save to		? ×
Save in:	🗀 guide	💽 🧿 🕬 🔛 •	
My Recent Documents			
My Documents			
My Computer			
My Network	File name:	Annotation Results.csv	Save
Places	Save as type:	Microsoft Office Excel Comma Separated Value	Cancel

Next, let's open the CSV file we just saved in a spreadsheet application. In this guide, we'll use Microsoft Excel 2003. A partial screenshot of the Excel window is shown below. The spreadsheet consists of two columns: 'Subject ID' and 'Gender'.

M	licroso	ft Exc	el - Ai	nnotal	tionR	esul	ts.cs	v
:2	<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	Inse	rt F	= <u>o</u> rma	t 1	<u>[</u> 0
1	2	3		3	<u>a</u> , ['	Sec. 1	ä,	9
10	Snagit	💇	Windo	w			• 📮	
	A1		•	ţ.	λ Sι	ubjec	t ID	
		А		E	3		С	
1	Subje	ct ID		gend	er			
2	TCGA	-29-2	414	FEM/	ALE			
3	TCGA	-10-0	933	FEM/	ALE			
4	TCGA	-24-0	979	FEM/	ALE			
5	TCGA	-24-2	290	FEM/	ALE			
6	TCGA	-13-0	794	FEM	ALE			

Now, go back to the 'Query Results' tab for the gene expression results type and click on the 'Export to CSV' button at the bottom left.

Criteria	Results Type	Sorting	Query Results	Save query as			
Query Res	ults for: test						
						Gene	BRCA1
Subject ID			Sample ID				
TCGA-61-2094			TCGA-61-2094-01A				-1.75
TCGA-25-1328			TCGA-25-1328-01A				-1.71
TCGA-61-2088			TCGA-61-2088-01A				-1.08
TCGA-61-2092			TCGA-61-2092-01A				-0.98
TCGA-25-2393			TCGA-25-2393-01A				-0.9
TCGA-25-1871			TCGA-25-1871-01A				-0.66
TCGA-25-1329			TCGA-25-1329-01A				-1.0
TCGA-25-1325			TCGA-25-1325-01A				-1.15
TCGA-25-2399			TCGA-25-2399-01A				-1.37
TCGA-25-2390			TCGA-25-2390-01A				-0.82
TCGA-25-2391			TCGA-25-2391-01A				-1.08
TCGA-25-2392			TCGA-25-2392-01A				-1.31
TCGA-25-1635			TCGA-25-1635-01A				-0.72
X							
-			🗣 Export To CSV	View in li	ntegrative Genom	ics Viewer	

Then, repeat the steps we followed previously to save the resulting CSV file as 'genomicResults.CSV' and open it in Excel. A partial screenshot of the Excel window is shown below. The spreadsheet consists of four columns: 'Subject ID', 'Sample ID', 'Gene Name', and 'BRCA1'. Since the 'Gene Name' column is empty, we can delete it, leaving three columns.

N	licrosoft Exc	cel - g	enomicRe	esults.csv	,			
:2	<u>F</u> ile <u>E</u> dit	<u>V</u> iew	Insert	F <u>o</u> rmat	<u>T</u> ools	<u>D</u> ata	<u>W</u> indow	<u>H</u> el
1	💕 🛃 🕻		🖪 🖪	I 🍄 🛍	8	la 🚨	• 🛷 🖂 🔊	-
: @	Snagit 🛃	Windo	w	-	Ŧ			
	A1	•	fx					
	Α			В		С	D	
1					Ge	ne Nam	e BRCA1	1
2	Subject ID		Sample I	D				
3	TCGA-61-2	2094	TCGA-61	-2094-01	A		-1	.75
4	TCGA-25-1	328	TCGA-25	-1328-01	A		-1	.71
5	TCGA-61-2	2088	TCGA-61	-2088-01	A		-1	.08
6	TCGA-61-2	2092	TCGA-61	-2092-01	A		-0	.98
7	TCGA-25-2	393	TCGA-25	-2393-01	A		J .	0.9

Now, highlight the entire contents of the spreadsheet (527 rows total), then select 'Copy' from the 'Edit' menu.

Microsoft Excel - genomicResults.csv							
:2	Eile	Edit	View	Insert	F <u>o</u> rma	at T	ools <u>D</u> ata
1	2	5	Can't <u>U</u> n	do C	trl+Z		አ 🗈 🛍
10	Snagit	¥	Cu <u>t</u>	0	trl+X		
	A1	-	Copy	C	trl+C) *	
			Office C	lip <u>b</u> oard			С
1		2	Paste	C	trl+V		BRCA1
2	Subje	_	-				
3	TCG/		Paste S	ecal)1A	-1.75
4	TCG/		Fill		•)1A	-1.71
5	TCG/)1A	-1.08
6	TCG		Cle <u>a</u> r		•)1A	-0.98
7	TCG		Delete)1A	-0.9
8	TCG	AA	Eind		white.)1A	-0.66
9	TCG	ara	<u>Cing</u>		.0171)1A	-1
10	TCG		Links)1A	-1.15
11	TCG		(¥)1A	-1.37
12	TCG/	4-25-	2390 T	CGA-2	5-2390-	01A	-0.82

Finally, switch back to the 'AnnotationResults.csv' spreadsheet, click on the empty cell to the right of the 'Gender' column heading, and select 'Paste' from the 'Edit' menu.

M	licros	oft Ex	ccel - Anr	notation	Results	s.csv	
:2	Eile	Edit	View	Insert	Format	Tools	C
1	2	5	<u>U</u> ndo Cle	ar Ctr	l+Z	1 X 🛛	
: @	Snagit	⋇	Cut	Ctr	I+X		
	C1		Copy	Ctr	I+C		
			Office Cli	p <u>b</u> oard		С	
1	Subj	2	Paste	Ctr	l+V	_	-
L 2							
3	TCG	_	Paste Sp	ecial	_		-
3	TCG/ TCG/	Τ	Paste Sp Fill	ecial	•		
3 4 5	TCG/ TCG/ TCG/	Τ	Paste Sp Fill	ecial	•		
3 4 5 6	TCG/ TCG/ TCG/ TCG/	Ī	Paste Sp Fill Cle <u>a</u> r	ecial	•		
3 4 5 6 7	TCG/ TCG/ TCG/ TCG/ TCG/		Paste ≦p Fill Cle <u>a</u> r Delete	ecial	•		
3 4 5 6 7 8	TCG/ TCG/ TCG/ TCG/ TCG/ TCG/	AA	Paste Sp Fill Clear Delete	ecial	•		
3 4 5 6 7 8 9	TCG/ TCG/ TCG/ TCG/ TCG/ TCG/	<i>8</i> 4	Paste Sp Fill Clear Delete Find	ecial Ctr	• •		
3 4 5 6 7 8 9 10	TCG/ TCG/ TCG/ TCG/ TCG/ TCG/ TCG/	<i>8</i> 8	Paste Sp Fill Clear Delete Find Links	ecial Ctr)) [+F		

The resulting spreadsheet should appear as in the screenshot below of the first ten rows:

	A	В	С	D	E
1	Subject ID	gender	Subject ID	Sample ID	BRCA1
2	TCGA-29-2414	FEMALE	TCGA-61-2094	TCGA-61-2094-01A	-1.75
3	TCGA-10-0933	FEMALE	TCGA-25-1328	TCGA-25-1328-01A	-1.71
4	TCGA-24-0979	FEMALE	TCGA-61-2088	TCGA-61-2088-01A	-1.08
5	TCGA-24-2290	FEMALE	TCGA-61-2092	TCGA-61-2092-01A	-0.98
6	TCGA-13-0794	FEMALE	TCGA-25-2393	TCGA-25-2393-01A	-0.9
7	TCGA-13-0791	FEMALE	TCGA-25-1871	TCGA-25-1871-01A	-0.66
8	TCGA-30-1892	FEMALE	TCGA-25-1329	TCGA-25-1329-01A	-1
9	TCGA-24-1467	FEMALE	TCGA-25-1325	TCGA-25-1325-01A	-1.15
10	TCGA-61-2102	FEMALE	TCGA-25-2399	TCGA-25-2399-01A	-1.37

Finally, save the merged spreadsheet by selecting 'Save As' from the 'File' menu, then entering 'mergedQueryResults.csv' as the file name in the 'Save As' dialog and clicking on the 'Save' button.



Workaround 2: The Integrated Genomics Viewer (IGV)

The other workaround for calntegrator's limited options for displaying query results is a newly added featured known as the Integrative Genomics Viewer (IGV), which allows the annotation and expression data to be visualized side-by-side in the form of color-coded maps. To begin using the IGV, switch back to the 'Results Type' tab on the 'Search TCGA Ovarian' page and click on the 'Integrative Genomics Viewer' button under the 'Select Results Type' heading.

Search TCGA Ovarian

Criteria	Results Type	Sorting	Query Results	Save query as	
Select Re	sults Type:				
	Anno	tation C Gen	e Expression Copy	Number C Integrative 0	Genomics Viewer C Heat Map Viewer

Next, check off 'gender' in the scrollable list box under the 'patient' heading, and finally, click on the 'View in Integrative Genomics Viewer' button at the bottom.

Select Results Type:	
C Annotation C Gene Expression C Copy Number	Integrative Genomics Viewer C Heat Map Viewer
Annotation result type - will display tabular data, including co Gene Expression result type - will display a gene expression (Copy Number result type - will display segmentation data with Integrative Genomics Viewer - will display a link to the IGV Vie HeatMap Viewer - will display a link to the Heat Map Viewer.	lumn selection. data matrix. Chromosome position. awer.
Select Columns for Results	
drug	examination
bcr_drug_barcode days_to_drug_therapy_end days_to_drug_therapy_start drug_category drug_name Select All Unselect All	easterm_cancer_oncology_group karnofsky_performance_score performance_status_scale_timing progression_determined_by Select All Unselect All
patient ays_to_last_tollowup days_to_tumor_progression days_to_tumor_recurrence gender histological_type hormonal_therapy Select All Unselect All	radiation Tradiation Tradiation_treatment_site Cradiation_barcode days_to_radiation_therapy_end days_to_radiation_therapy_start Select All Unselect All
surgery bcr_surgery_barcode days_to_procedure procedure_type	Select Reporter Type: Reporter Id Gene Reporter Id use reporter in the gene expression criterion. Gene - use gene in the gene expression criterion.
Select All Unselect All	ntegrative Genomics Viewer

NOTE: Before you launch the IGV, make sure you have version 6 or later of the Java Runtime Environment (JRE) installed on your machine. The JRE can be downloaded from the Oracle Web site at http://www.oracle.com/technetwork/java/javase/downloads/jre-6u27-download-440425.html

Now, a new window will open in your Web browser entitled 'Integrative Genomics Viewer'. Click on the blue link entitled 'Launch Integrative Genomics Viewer'.



Integrative Genomics Viewer

Current Status:

Job Finished

Launch Integrative Genomics Viewer

Assuming that you have the JRE properly installed, you should now see a dialog entitled 'Opening igv.jnlp'. Make sure that the 'Open with' button is selected, then click on the 'OK' button at the bottom to launch the IGV.

Opening igv.jnlp	×
You have chosen to open	
🔊 igv.jnlp	
which is a: JNLP File	
from: http://www.broadinstitute.org	
What should Pale Moon do with this file?	
Open with Java(TM) Web Start Launcher (default)	
C Save File	
Do this automatically for files like this from now on.	
OK Cancel	

Depending on how the JRE is configured on your machine, you may see a dialog entitled 'Warning – Security' stating, 'The application's digital signature cannot be verified. Do you want to run the application?' If this is the case, simply click on the 'Run' button at the bottom to continue.



The main IGV window is shown below. The application is complex and thus its documentation is beyond the scope of this tutorial. For an in-depth tutorial, visit the official IGV Web site, hosted by the Broad Institute of Harvard and MIT, at http://www.broadinstitute.org/igv/QuickStart.



Have a comment?

Please leave your comment in the calntegrator End User Forum.