

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: calIntegrator Usage

Date entered: 08/23/2011

Release: Up to calIntegrator 1.3

Solution

calIntegrator 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, calIntegrator 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 calIntegrator version 1.3 or later, involves visualizing the data in the form of color-coded maps via a newly added feature known as the Integrative Genomics Viewer (IGV).

Querying the Data

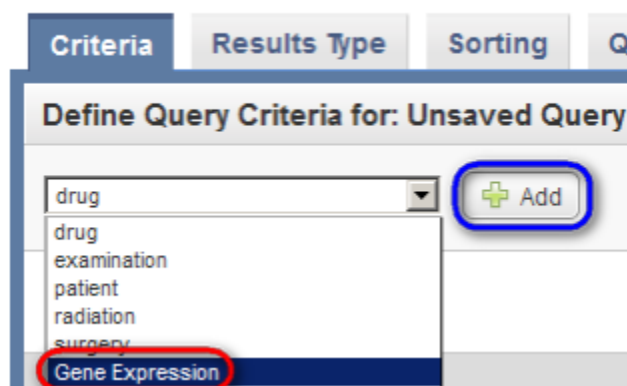
First, let's do a sample query of a calIntegrator study to illustrate the limitations in displaying query results. In our example, the study we will query is entitled 'TCGA Ovarian', which is selected in the 'Public Studies' drop-down list at the top of the calIntegrator home page. To start querying,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



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 Name + Add

Gene Expression Gene Name Gene Symbol(s) (comma separated list) or blank brca1
for all genes
Platform Name 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: Unsaved Query

Gene Expression 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 Gene Symbol(s) (comma separated list) or blank brca1
for all genes
Platform Name AgilentG4502A_07_singleADF

patient gender equals FEMALE

☐ or ☒ 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 None	Select Subject All None	Subject ID	
	<input checked="" type="checkbox"/>	TCGA-29-1688	
	<input checked="" type="checkbox"/>	TCGA-30-1868	
	<input checked="" type="checkbox"/>	TCGA-13-1491	
	<input checked="" type="checkbox"/>	TCGA-13-0751	
	<input checked="" type="checkbox"/>	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 Results Type:

☒ Annotation
☐ Gene Expression
☐ Copy Number
☐ Integrative Genomics Viewer
☐ Heat Map Viewer

Annotation result type - will display tabular data, including column selection.
Gene Expression result type - will display a gene expression data matrix.
Copy Number result type - will display segmentation data with Chromosome position.
Integrative Genomics Viewer - will display a link to the IGV Viewer.
HeatMap Viewer - will display a link to the Heat Map Viewer.

Select Columns for Results

drug

☐ bcr_drug_barcode
☐ days_to_drug_therapy_end
☐ days_to_drug_therapy_start
☐ drug_category
☐ drug_name

Select All
Unselect All

examination

☐ eastern_cancer_oncology_group
☐ karnofsky_performance_score
☐ performance_status_scale_timing
☐ progression_determined_by

Select All
Unselect All

patient

☐ days_to_tumor_progression
☐ days_to_tumor_recurrence
☒ gender
☐ histological_type
☐ hormonal_therapy

Select All
Unselect All

radiation

☐ Units
☐ anatomic_treatment_site
☐ bcr_radiation_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 All
Unselect All

Select Reporter Type:
☒ Reporter Id
☐ Gene

Reporter Id - use reporter in the gene expression criteria
Gene - use gene in the gene expression criterion.

Run 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 Results for: Unsaved Query

Results per Page: 20
Apply

509 items found, displaying 1 to 20. [First/Prev] 1, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Select Image All None	Select Subject All None	Subject ID	gender
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-13-0903	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-04-1652	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-61-2096	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-25-1632	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-29-1768	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-59-2354	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-29-1707	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-10-0931	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-13-0888	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-24-2023	FEMALE
<input type="checkbox"/>	<input checked="" type="checkbox"/>	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.

CriteriaResults TypeSortingQuery ResultsSave query as...

Select Results Type:

☐ Annotation

☒ Gene Expression

☐ Integrative Genomics Viewer

Annotation result type - will display tabular data, including column selection.

Gene Expression result type - will display a gene expression data matrix.

Select Reporter Type:

☐ Reporter Id

☒ Gene

Reporter Id - will display all reporter ids.

Gene - will display a median value for each gene.

Select Results Orientation:

☒ Genes in Columns / Subjects in Rows

☐ Genes in Rows / Subjects in Columns

Genes in Rows / Subjects in Columns - will display the subjects and samples along the top and genes and reporters along the s

Genes in Columns / Subjects in Rows - will display the genes and reporters along the top and subjects and samples along the s

few reporters.

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.

CriteriaResults TypeSortingQuery ResultsSave query as...

Query Results for: test

Subject ID	Sample ID	Gene	BRCA1
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

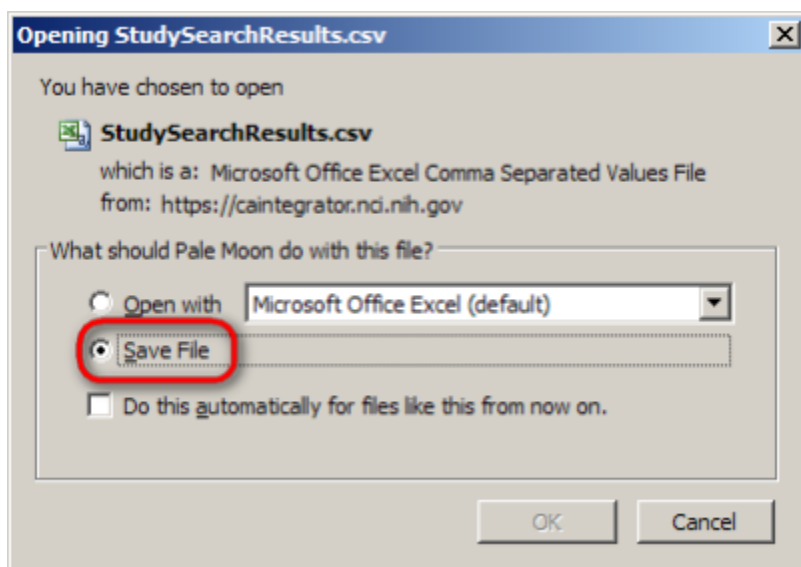
Workaround 1: Manually Exporting Spreadsheets

So far, we've seen that calIntegrator 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.

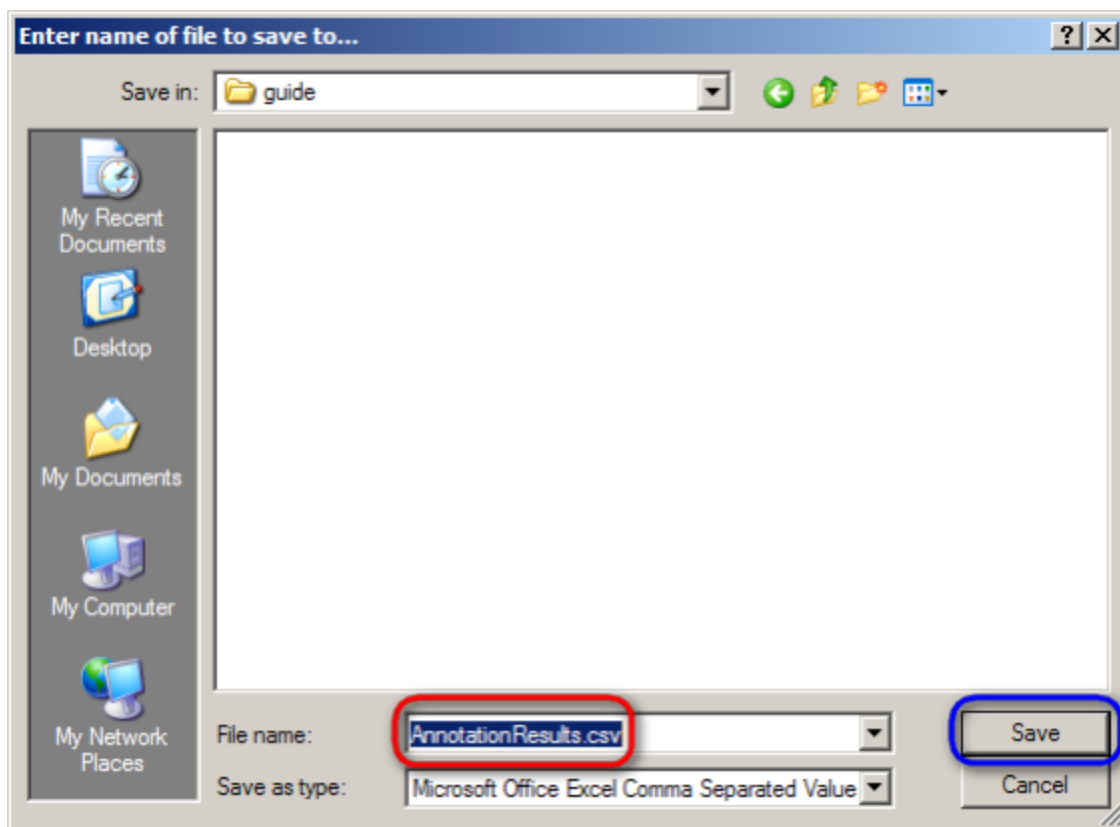
Criteria	Results Type	Sorting	Query Results	Save query as...
Query Results for: test				Results per Page: 20 Apply
525 items found, displaying 1 to 20. (First Prev 1, 2, 3, 4, 5, 6, 7, 8 Next Last)				
Select Image All None	Select Subject All None	Subject ID	gender	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-29-2414	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-10-0933	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-24-0979	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-24-2290	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-13-0794	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-13-0791	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-30-1882	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-24-1467	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-81-2102	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-24-0970	FEMALE	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	TCGA-25-1628	FEMALE	

Export To CSV Save Subject List View in Integrative Genomics Viewer View in Heat 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.



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 drop-down list at the top, then enter 'AnnotationResults.csv' for the file name and click on the 'Save' button.



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'.

	A	B	C
1	Subject ID	gender	
2	TCGA-29-2414	FEMALE	
3	TCGA-10-0933	FEMALE	
4	TCGA-24-0979	FEMALE	
5	TCGA-24-2290	FEMALE	
6	TCGA-13-0794	FEMALE	

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.

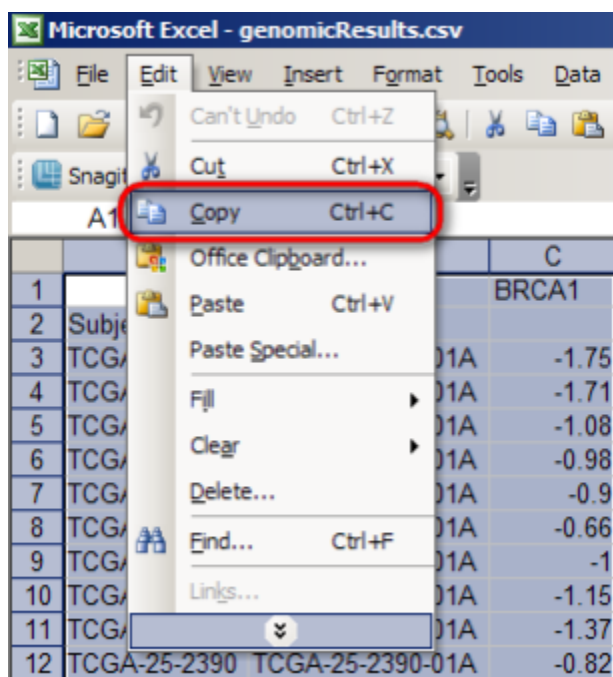
Criteria	Results Type	Sorting	Query Results	Save query as...
Query Results for: test				
		Gene	BRCA1	
Subject ID	Sample ID			
TCGA-61-2094	TCGA-61-2094-01A		-1.75	
TCGA-25-1328	TCGA-25-1328-01A		-1.71	
TCGA-61-2088	TCGA-61-2088-01A		-1.08	
TCGA-61-2092	TCGA-61-2092-01A		-0.98	
TCGA-25-2393	TCGA-25-2393-01A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

Export To CSV View in Integrative Genomics Viewer

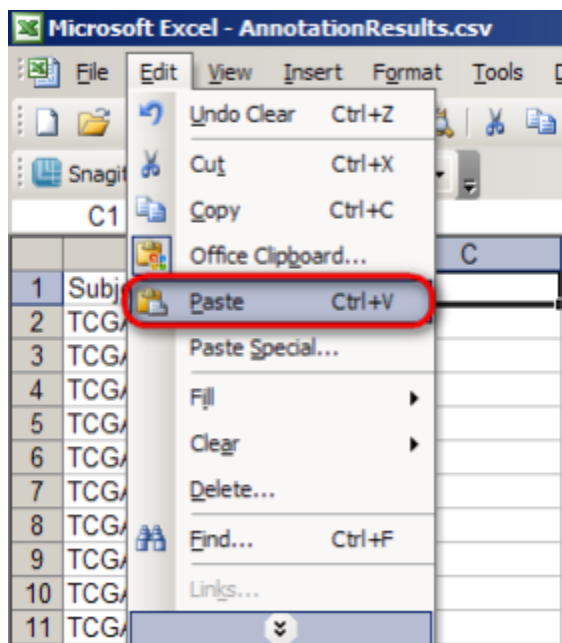
Then, repeat the steps we followed previously to save the resulting CSV file as 'genomicResults.CSV' and open it in Excel. A partial screenshot of the Excel window is shown below. The spreadsheet consists of four columns: 'Subject ID', 'Sample ID', 'Gene Name', and 'BRCA1'. Since the 'Gene Name' column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

Microsoft Excel - genomicResults.csv				
File Edit View Insert Format Tools Data Window Help				
Snagit Window				
A1	fx			
	A	B	C	D
1			Gene Name	BRCA1
2	Subject ID	Sample ID		
3	TCGA-61-2094	TCGA-61-2094-01A		-1.75
4	TCGA-25-1328	TCGA-25-1328-01A		-1.71
5	TCGA-61-2088	TCGA-61-2088-01A		-1.08
6	TCGA-61-2092	TCGA-61-2092-01A		-0.98
7	TCGA-25-2393	TCGA-25-2393-01A		-0.9

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



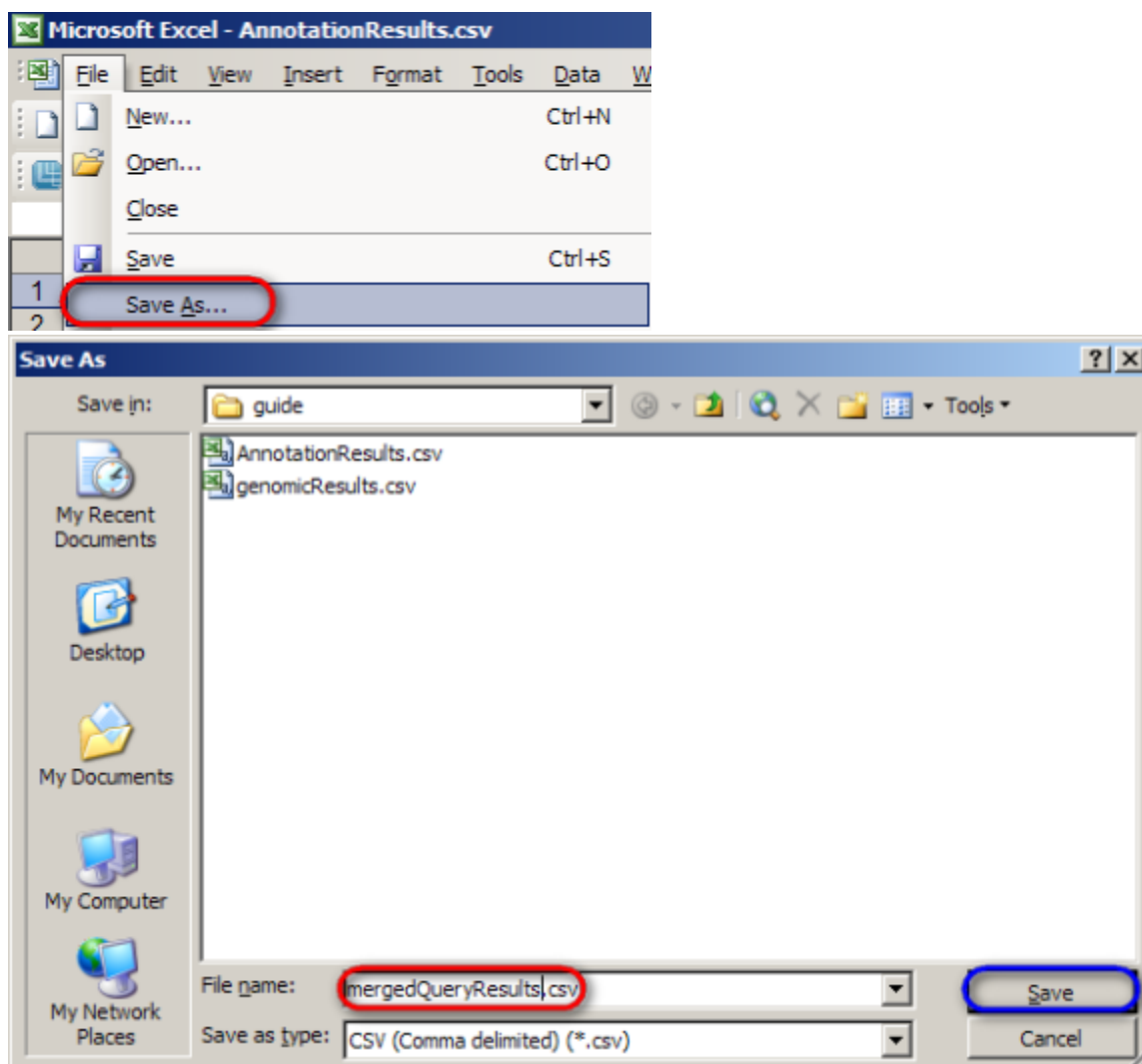
Finally, switch back to the 'AnnotationResults.csv' spreadsheet, click on the empty cell to the right of the 'Gender' column heading, and select 'Paste' from the 'Edit' menu.



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

	A	B	C	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 calIntegrator's limited options for displaying query results is a newly added feature 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 Results Type:

☒ Annotation ☐ Gene Expression ☐ Copy Number ☒ Integrative Genomics Viewer ☐ 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:

☐ Annotation ☐ Gene Expression ☐ Copy Number ☒ Integrative Genomics Viewer ☐ Heat Map Viewer

Annotation result type - will display tabular data, including column selection.
Gene Expression result type - will display a gene expression data matrix.
Copy Number result type - will display segmentation data with Chromosome position.
Integrative Genomics Viewer - will display a link to the IGV Viewer.
Heatmap Viewer - will display a link to the Heat Map Viewer.

Select Columns for Results

drug	examination
<input type="checkbox"/> bcr_drug_barcode	<input type="checkbox"/> eastern_cancer_oncology_group
<input type="checkbox"/> days_to_drug_therapy_end	<input type="checkbox"/> karnofsky_performance_score
<input type="checkbox"/> days_to_drug_therapy_start	<input type="checkbox"/> performance_status_scale_timing
<input type="checkbox"/> drug_category	<input type="checkbox"/> progression_determined_by
<input type="checkbox"/> drug_name	
<input type="checkbox"/>	
<input type="button" value="Select All"/> <input type="button" value="Unselect All"/>	<input type="button" value="Select All"/> <input type="button" value="Unselect All"/>

patient	radiation
<input type="checkbox"/> days_to_last_followup	<input type="checkbox"/> Units
<input type="checkbox"/> days_to_tumor_progression	<input type="checkbox"/> anatomic_treatment_site
<input type="checkbox"/> days_to_tumor_recurrence	<input type="checkbox"/> bcr_radiation_barcode
<input checked="" type="checkbox"/> gender	<input type="checkbox"/> days_to_radiation_therapy_end
<input type="checkbox"/> histological_type	<input type="checkbox"/> days_to_radiation_therapy_start
<input type="checkbox"/> hormonal_therapy	
<input type="button" value="Select All"/> <input type="button" value="Unselect All"/>	<input type="button" value="Select All"/> <input type="button" value="Unselect All"/>

surgery
<input type="checkbox"/> bcr_surgery_barcode
<input type="checkbox"/> days_to_procedure
<input type="checkbox"/> procedure_type
<input type="button" value="Select All"/> <input type="button" value="Unselect All"/>

Select Reporter Type: ☒ Reporter Id ☐ Gene
Reporter Id - use reporter in the gene expression criterion.
Gene - use gene in the gene expression criterion.

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'.



National Cancer Institute

caINTEGRATOR

build: CAINTEGRATOR_R1_3_0_GA | date: '06/16/2011 10:58 AM'

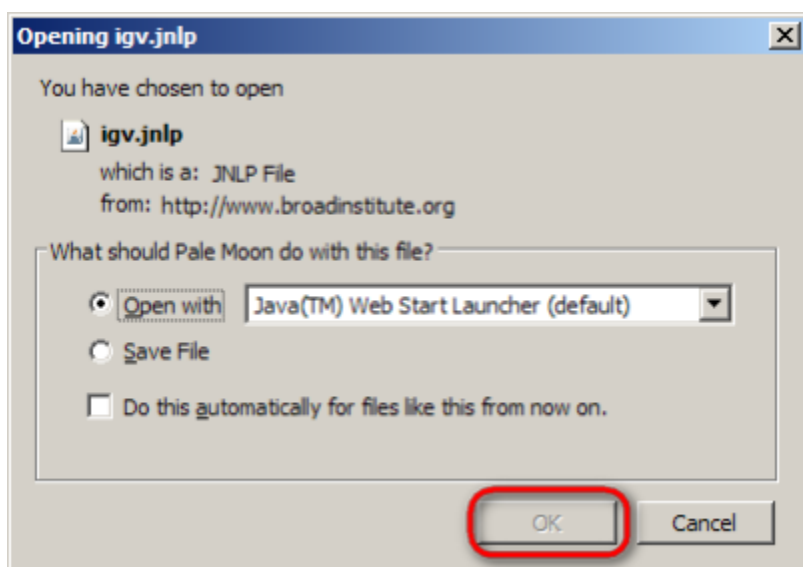
Integrative Genomics Viewer

Current Status:

Job Finished

[Launch Integrative Genomics Viewer](#)

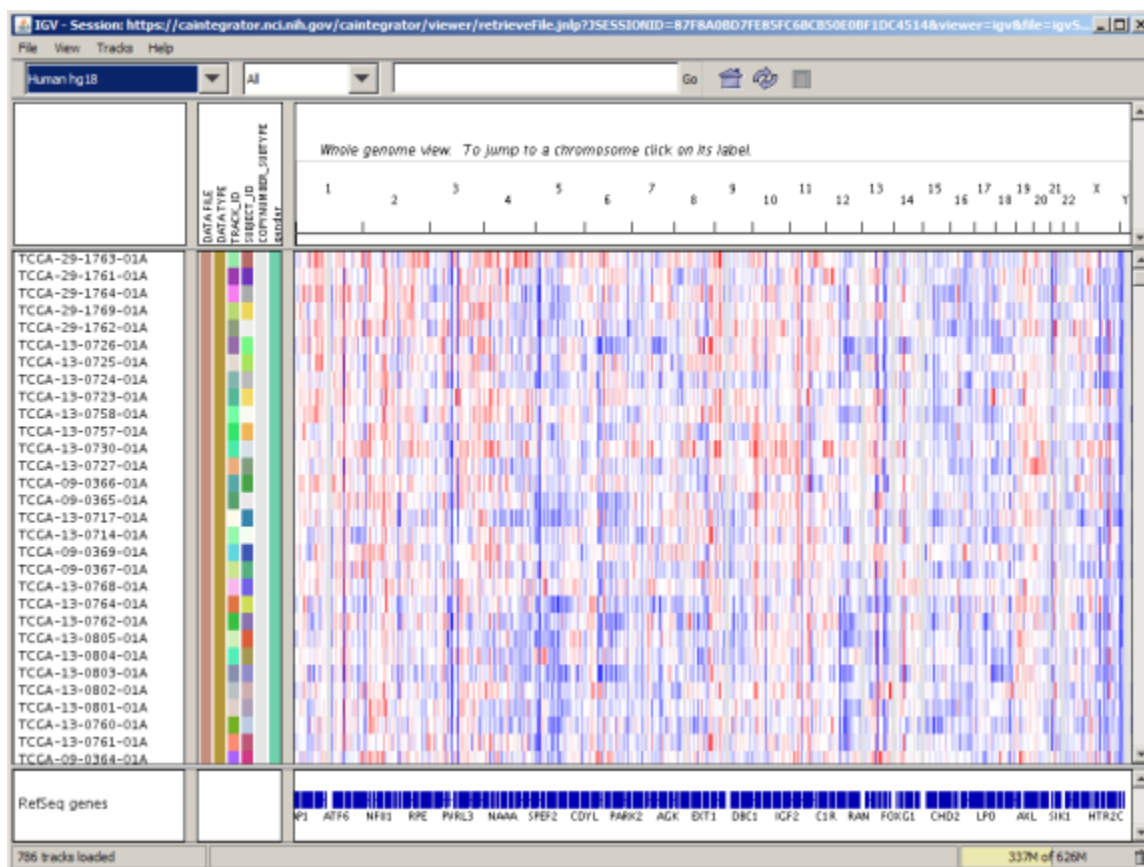
Assuming that you have the JRE properly installed, you should now see a dialog entitled 'Opening igv.jnlp'. Make sure that the 'Open with' button is selected, then click on the 'OK' button at the bottom to launch the IGV.



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 [caIntegrator End User Forum](#).