



Tutorial

RNA-Seq with Ingenuity Pathway Analysis

June 20, 2019

— Sample to Insight —

RNA-Seq with Ingenuity Pathway Analysis

This tutorial will demonstrate the collaborative power of the CLC Genomics Workbench RNA-Seq tools and the Ingenuity Pathway Analysis plugin to analyze and interpret RNA-Seq expression data.

Prerequisites For this tutorial, you must be working with the CLC Genomics Workbench and have installed the Ingenuity Pathway Analysis plugin. How to install plugins is described here: <http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html>.

In addition, you must have secured access to Ingenuity Pathway Analysis services. A free trial account for Ingenuity Pathway Analysis can be created using this link: <http://www.ingenuity.com/products/ipa/try-ipa-for-free>. Please note that setting up an account can take 1-4 business days.





Download and import the data

This tutorial will use RNA-Seq data for 4 tumor/normal tissue pairs. To allow the completion of the tutorial in a reasonable amount of time, we have reduced the reads to only 5% of the original dataset, selecting them so they map only to chromosome 20.

The data set includes:

- A subset of the reads from 4 paired samples.
- An Excel spreadsheet that contains the metadata for each tumor/normal pair.
- Three reference sequence for the human hg38 genome, as well as genes and mRNA reference tracks for chromosome 20.

Go through the following steps to download and import the data into the Workbench.

1. Download the sample data from our website: http://resources.qiagenbioinformatics.com/testdata/RNA_Seq_IPA20.zip and unzip the file on your local machine.
2. In *CLC Genomics Workbench*, import the reads via the toolbar: **File | Import**  | **Standard Import** . Choose the file called "reads" and the three reference data files containing data for chromosome 20. Leave the import type set to Automatic and save the imported data in a new folder you can create in your Navigation Area (called for example RNA-Seq tutorial).
3. Import the metadata table via the toolbar: **File | Import**  | **Import Metadata** .
4. A wizard opens. In the first window, select the Excel spreadsheet with the sample information. The contents of the Excel spreadsheet populates the table situated at the bottom of the dialog (figure 1). Click **Next**.
5. Click on the Navigation button next to "Location of data", and find the folder containing the imported reads. Right-click the name of the folder and select "Add folder contents recursively" (figure 2). Click **OK**. You can see that the data association is successful

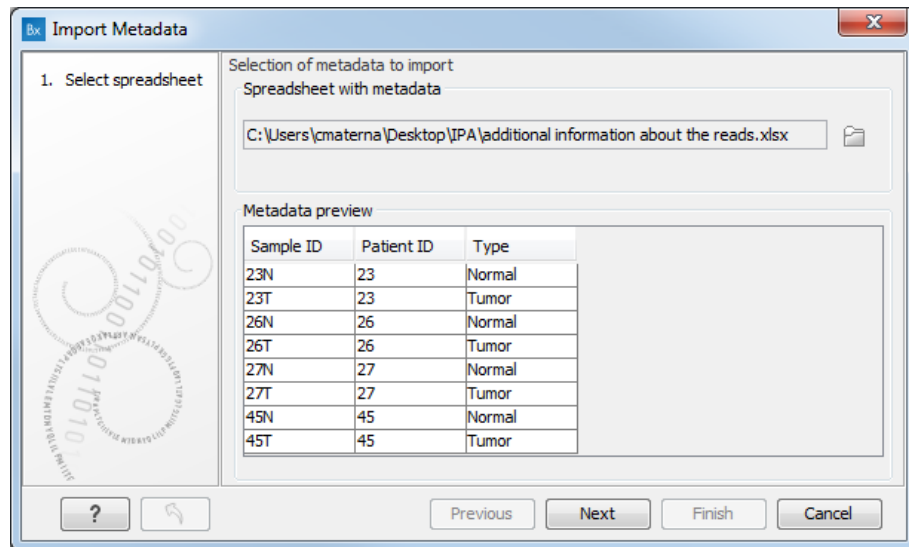


Figure 1: In the *Import Metadata* wizard, select the spreadsheet with the sample information.

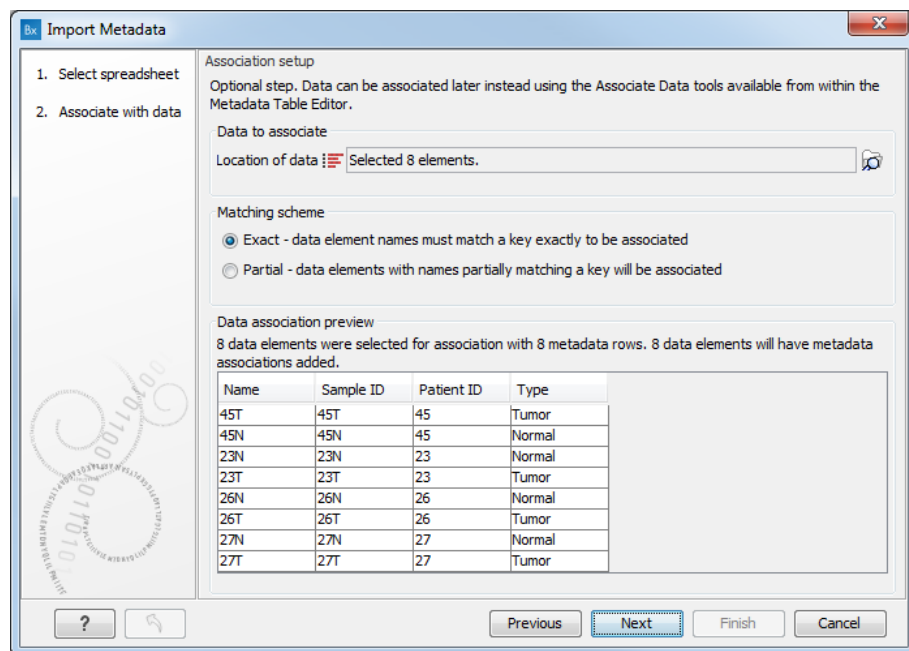
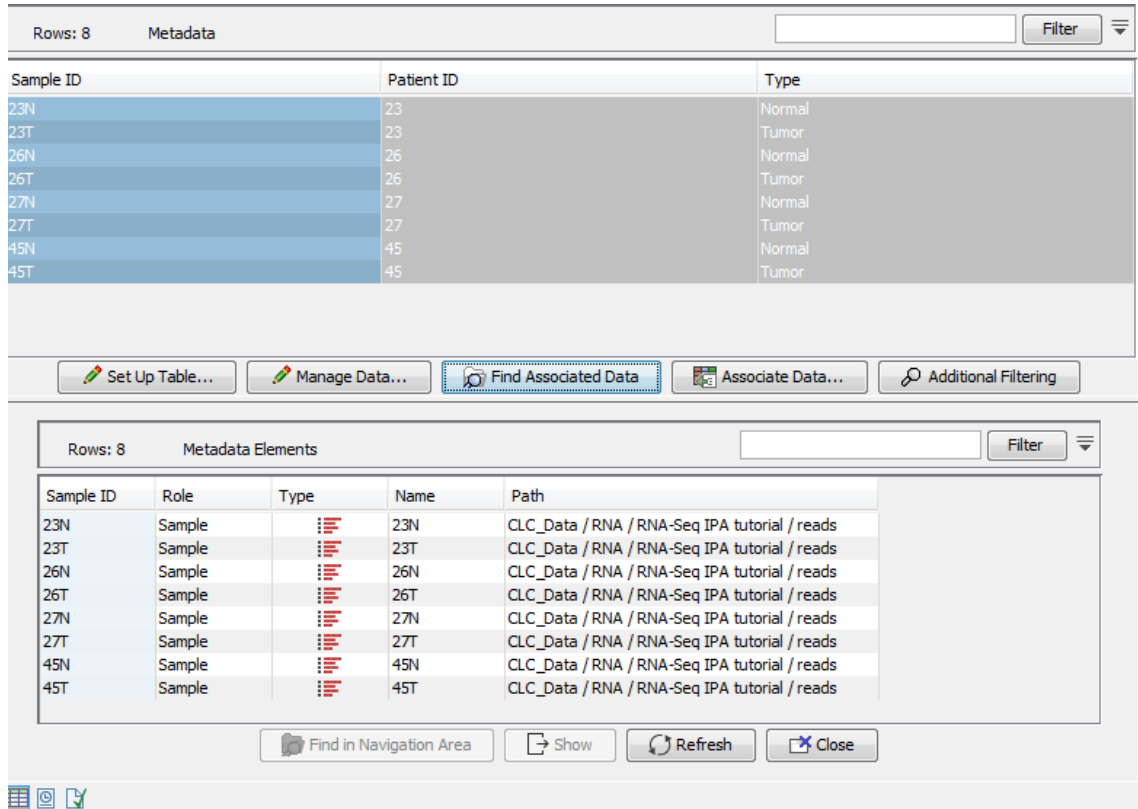


Figure 2: In the *Import Metadata* wizard, select the reads to associate with the metadata.

because the metadata has filled in the columns of the Data association table at the bottom of the dialog. Click **Next**.

6. Select the "RNA-Seq tutorial" folder to save the resulting metadata table.
7. Now open the metadata table you just saved (called "Samples" as can be seen in the upper part of figure 3). Select all rows in the table, and click "Find Associated Data". This will show that all rows are associated correctly with the corresponding reads.

Once the data has been downloaded and imported, you should see the folder and files in the Navigation Area as shown in figure 4.



The screenshot shows two tables in the Ingenuity Pathway Analysis software. The top table is titled 'Metadata' and contains 8 rows of sample information. The bottom table is titled 'Metadata Elements' and shows the hierarchical path for each sample's data.

Sample ID	Patient ID	Type
23N	23	Normal
23T	23	Tumor
26N	26	Normal
26T	26	Tumor
27N	27	Normal
27T	27	Tumor
45N	45	Normal
45T	45	Tumor

Sample ID	Role	Type	Name	Path
23N	Sample		23N	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
23T	Sample		23T	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
26N	Sample		26N	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
26T	Sample		26T	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
27N	Sample		27N	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
27T	Sample		27T	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
45N	Sample		45N	CLC_Data / RNA / RNA-Seq IPA tutorial / reads
45T	Sample		45T	CLC_Data / RNA / RNA-Seq IPA tutorial / reads

Figure 3: In the Metadata table, check that associations between reads and the metadata have happened successfully.

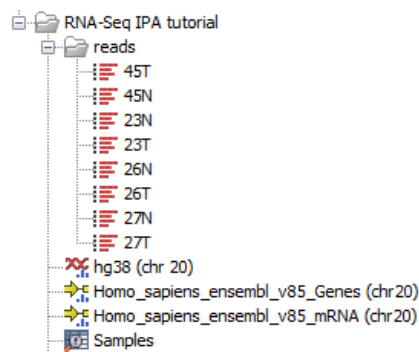


Figure 4: The data set has been imported.

RNA-Seq quantification

To map the reads to the reference, we will use the RNA-Seq tool. Go to:

Toolbox | RNA-Seq Analysis | RNA-Seq Analysis 

1. In the window called "Select sequencing reads", check the "Batch" box before selecting the folder containing the imported reads files. Right-click on the folder to choose the option "Add folder contents" (figure 5).

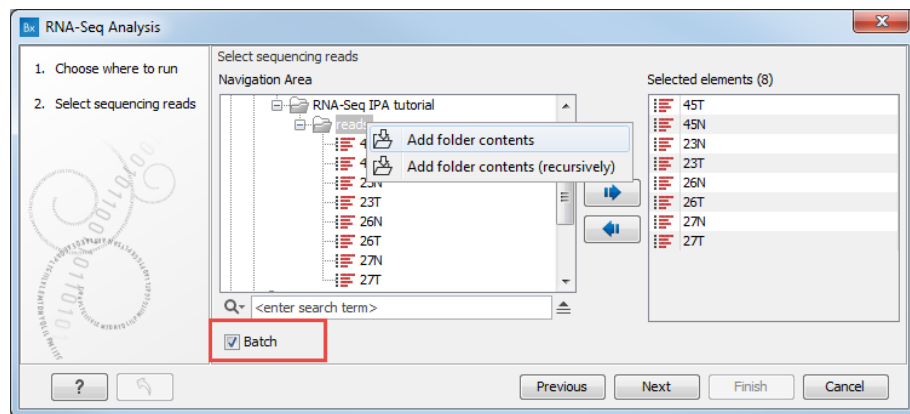


Figure 5: Right-click on the folder to be able to choose the option "Add folder contents"

2. In the next wizard window review the batch unit settings and click **Next**.
3. Set up the wizard step so that it looks like figure 6. Click **Next**.

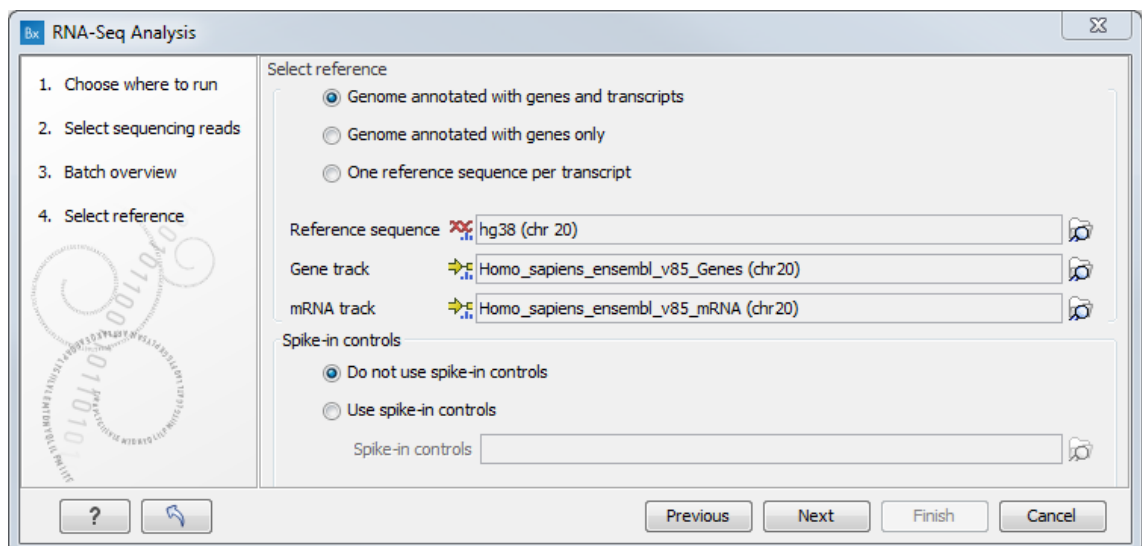


Figure 6: Keep both "Genome annotated with genes and transcripts" and "Do not use spike-in controls" checked.

4. Use default parameters for the "Mapping options" dialog. You can use the **Reset** button if you are not sure whether you have previously changed the parameters for the tool. Click **Next**.

5. In the "Expression level options" dialog, leave all parameters as default.
6. Save the results in a subfolder called RNA-Seq analysis.

After starting the tool, you can use the progress bar to see how it is progressing (figure 7).

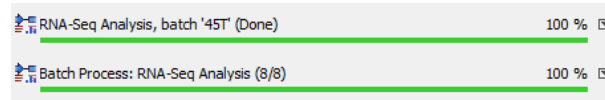


Figure 7: A process called "Batch Process" indicates how many batches have been completed, while the one called "RNA-Seq Analysis" shows the analysis progress of a particular batch unit.

From expression tracks to Ingenuity Pathway Analysis

When the mapping is completed, we will use the "Analyze Expression Data and Upload Comparisons to IPA" workflow (figure 8) to analyze the results.

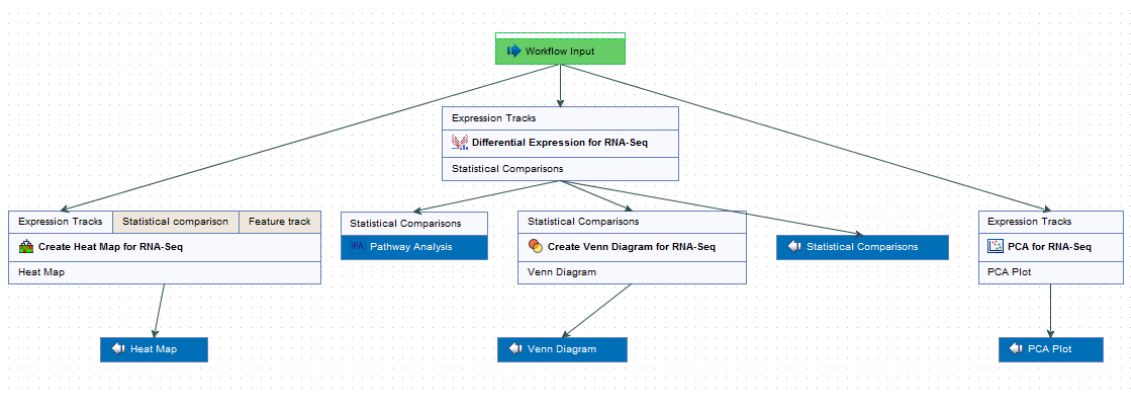


Figure 8: The workflow results in two folders containing the results for each family.

1. We will use the metadata table to quickly find the tracks to analyze using this ready-to-use workflow (figure 9).
 - Open (or re-open if you had left the tab "Samples" opened in the View Area of the workbench) the metadata table, and select all rows.
 - Click **Find Associated Data**.
 - Click the "Role" column title twice to sort the elements by title, and select the 8 rows of "Transcript expression". Alternatively, you can enter "Transcript expression" in the Filter box, which will keep only the 8 rows containing transcript-level expression tracks.
 - Right-click the selected rows. Choose **Toolbox | Ready-to-use-workflows | Whole Transcriptome Sequencing | Analyze Expression Data and Upload Comparisons to IPA** to start the workflow (figure 9).
2. A wizard window opens for the workflow. Click **Next** again to accept the selected tracks. Note that even if you start the workflow from the Toolbox, the tracks selected in the metadata table will also be automatically used as input.

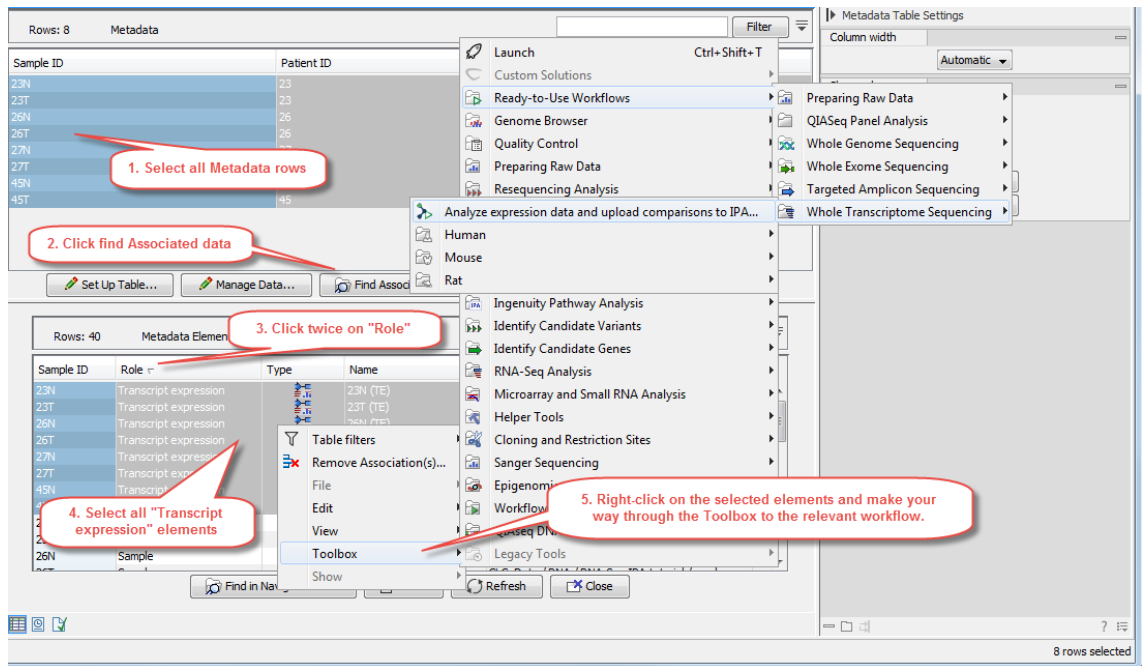


Figure 9: Starting a workflow from the metadata table.

- Set the parameters as shown in figure 10. This will result in an analysis of Tumor vs. Normal samples, while controlling for Patient ID. This means that a "paired" statistical test will be carried out. When finished, click **Next**.

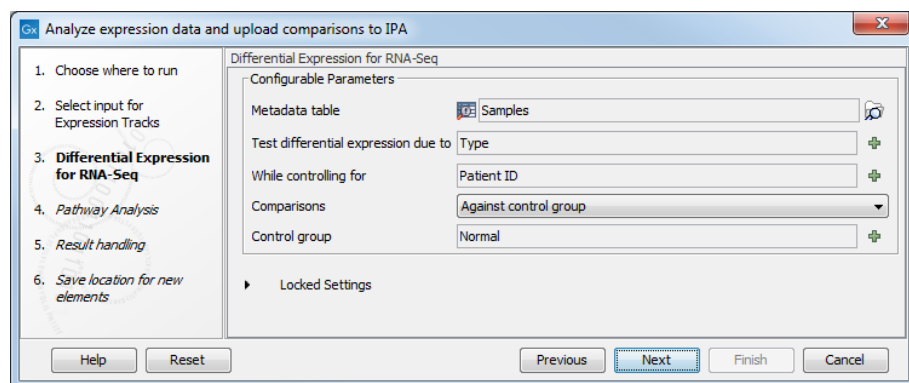


Figure 10: Set the parameters for the "Differential Expression for RNA-Seq" window.

- In the "Pathway analysis" step, enter your Ingenuity username and password, and use default parameters for the rest of the options as shown in figure 11.
- In the "Result handling" window, choose the location to save the results (for example a new subfolder you can call "IPA"), and click **Finish**. The workflow will now execute.

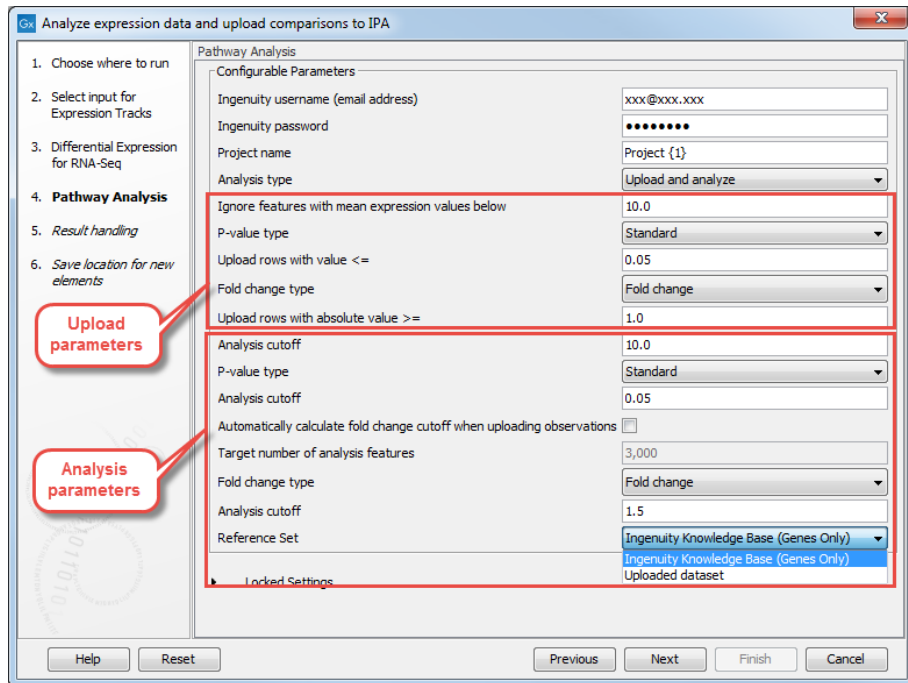


Figure 11: Enter your Ingenuity username and password and set the parameters for the "Pathway analysis" window.

Results interpretation

When the workflow is completed, several new objects are created (figure 12) and the results are also uploaded to IPA (you will get an email).

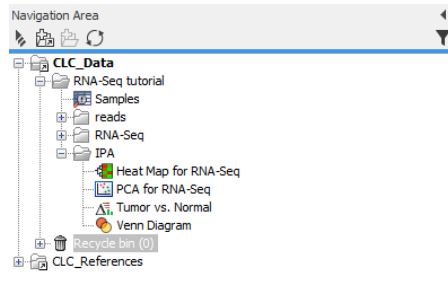


Figure 12: The results as seen in your Navigation Area.

For the moment we will focus on the results present in the Navigation Area.

PCA plot Open the PCA plot by double-clicking on "PCA for RNA-Seq" in the Navigation Area. The plot is first colored by Patient ID. We can see that the points are rather scattered without any clear clustering. Change the coloring scheme to "Type" in the side panel palette **Metadata | Symbol color | Type**. Now we can see that the samples cluster along the first principal component by Type (figure 13). This is consistent with a hypothesis that there is a difference in gene expression between tumor and normal samples.

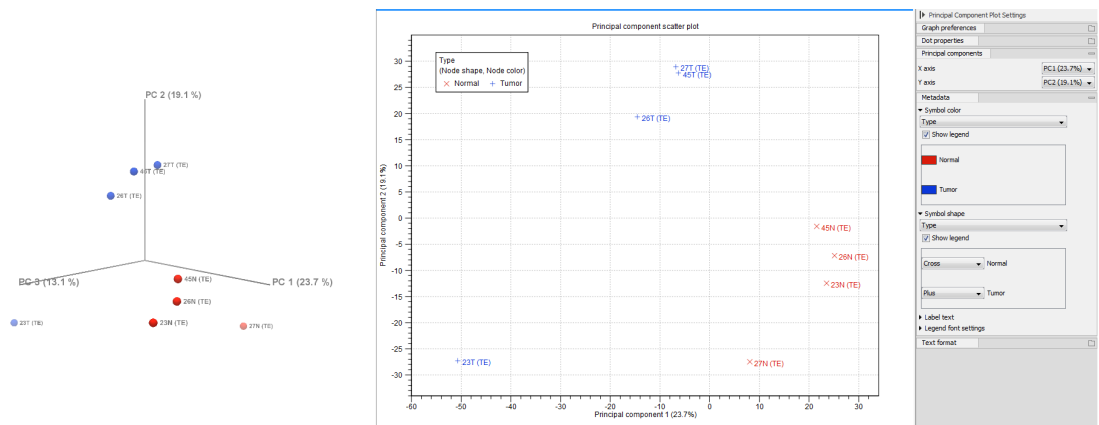


Figure 13: The PCA plot when samples are colored by Type. Click on the 3D icon in the lower left corner to see the plot in 3D as on the left hand side of this figure.

Heat map Next, open the Heat Map for RNA-Seq from the Navigation Area. The hierarchical clustering algorithm is pre-configured in the Ready-to-Use workflow to select the 25 "most interesting" transcripts in the heat map, based on the coefficient of variation (relative standard deviation). The samples are also clustered on the horizontal axis, in unsupervised clustering.

We expect the samples to cluster by Type. To see if this is the case, add "Type" as a metadata layer in the Metadata Layer side-panel (figure 14). Now we can see that the samples cluster by tumor vs. normal, as we expected. Note that you can click-and-drag to move the heat map legend for optimal visibility.

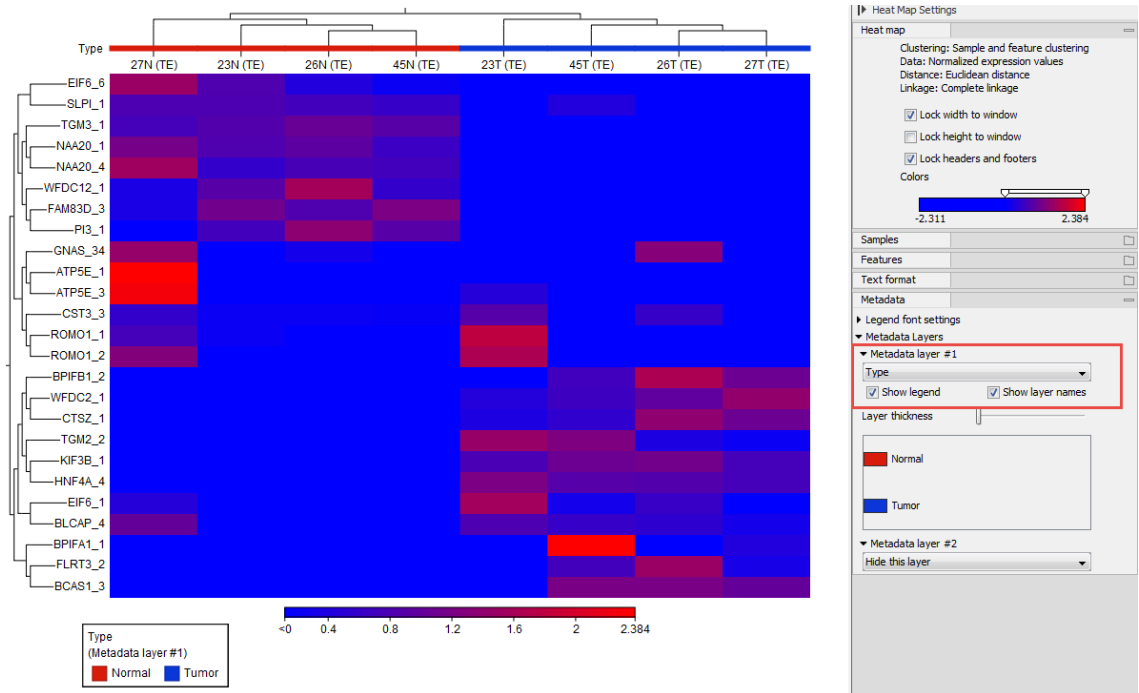


Figure 14: The Metadata side panel options for Heat Maps.

Statistical comparison The statistical comparison track "Tumor vs. Normal" is the result of the statistical analysis that was carried out by the workflow. Open the "Tumor vs. Normal" item: it will automatically display the table view of the track.

We will first investigate the volcano plot of the results: click the volcano plot icon (📊) at the bottom of the table (highlighted in red in figure 15). This demonstration dataset has very few reads and very few significant transcripts under FDR-correction, therefore the volcano plot is not very dense. To improve visibility in this particular case, choose "P-value" in the "P-value type" drop down menu in the Side Panel. You can select any "interesting" transcripts by simply clicking the point representing them. This will turn those points red and their names will be displayed next to the point.

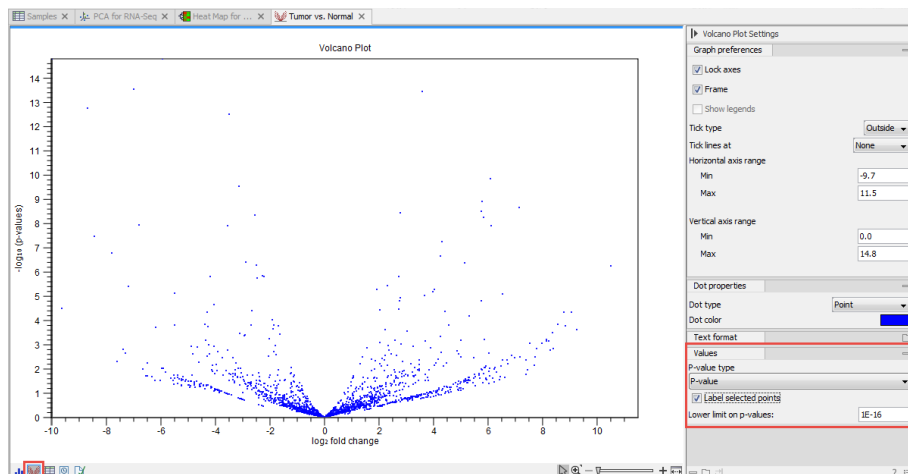


Figure 15: Volcano plot and the corresponding side panel options.

Note that any transcripts you have selected in the volcano plot are also selected in the table. It also works the other way round, so a selection you make in the table will also be selected in the volcano plot. The fact that table and plot are linked allows us to highlight the transcripts in the plots by filtering them in the table.

- Go back to the table view and click on the arrow next to the "Filter" button at the top of the table to open up the advanced filter box.
- Select **FDR p-value** in the first field, **abs value <** in the second, and type **0.05** in the third.
- Add a new filtering criterion by clicking on the green + button.
- As you did for the first filter, select **Fold change, abs value >**, and type **1.5**.
- Click on the button "Filter".
- Select all remaining rows in the table and look at the volcano plot to see the transcripts you have highlighted in red.

Note that some transcripts are too close to the borders of the plot to have their name correctly displayed. To fix this, you can increase the plot size by editing the axes ranges in the side panel section called "Graph preferences".

Advanced batching by patient

Sometimes it can be useful to compare Tumor vs. Normal results on a patient-specific basis. For this, we can execute in batch mode a modified version of the Ready-to-Use workflow to do the analysis.

1. First reorganize the RNA-seq analysis folder by moving all (TE) results file in a subfolder called "Transcript Expression".
2. To edit the "Analyze Expression Data and Upload Comparisons to IPA" workflow, first select and then right-click the workflow name in the Toolbox **Ready-to-use-workflows | Whole Transcriptome Sequencing | Analyze Expression Data and Upload Comparisons to IPA**. Choose "Open copy of workflow".
3. Select the "Create Heat Map for RNA-Seq" tool together with its output, and remove it by pressing Delete on your keyboard. Similarly, remove the "PCA for RNA-Seq" tool and its output, as well as the "Venn diagram" tool and its output.
4. The workflow only has one input. As we would like to compare Tumor samples versus Normal samples, add a second input, by right-clicking "Expression tracks" and selecting "Connect to workflow input". Rename the input elements to "Tumor" and "Normal", respectively.
5. The finished workflow should look like in (figure 16).

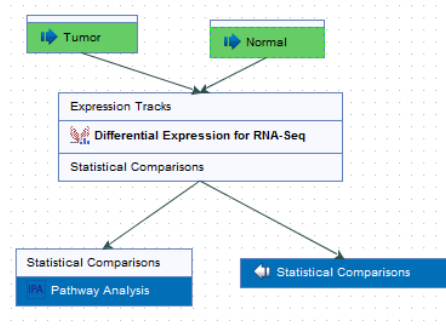


Figure 16: Workflow after editing, ready to batch.

6. Save the workflow in the RNA-Seq tutorial folder in the Navigation Area.
7. To run in batch, the workflow needs to be installed in the Toolbox of your workbench. After saving the workflow in the Navigation Area, The **Installation** button becomes active, and by clicking on it you will activate a wizard. Enter the requested information and in the next window, choose to "Install the workflow on your local computer". It will now appear in the folder "Workflows" of the Toolbox.
8. Select and right-click the name of the newly installed workflow and choose to "Run in batch mode" (figure 17).
9. In the "Import Metadata" wizard,
 - You are first asked to specify a metadata table as an Excel spreadsheet. Select the same Excel spreadsheet that you selected when importing the metadata table.

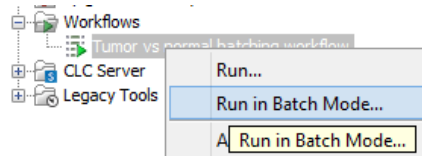


Figure 17: Right-click on the name of the workflow to choose the "Run in Batch Mode" option.

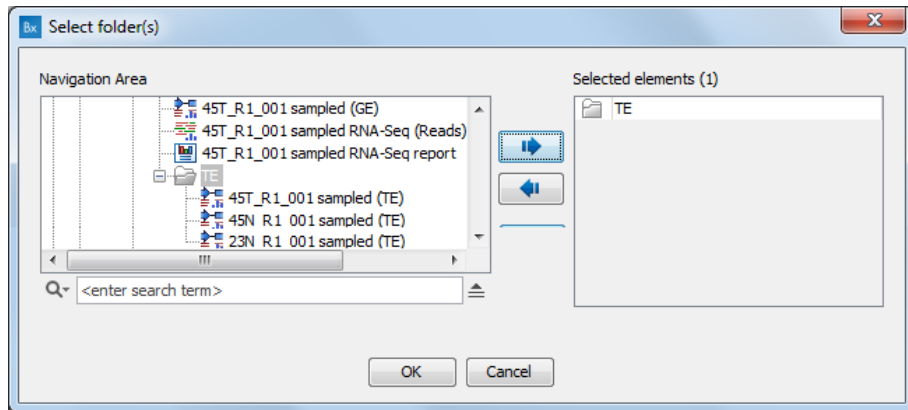


Figure 18: Select the subfolder in which you rearrange the transcript expression files.

- Next, select the subfolder that contains all transcript tracks to use in the analysis (figure 18).
- Finally, select "Partial" matching.
- The table should indicate that a correct match could be found with green marks next to the Sample IDs. Click **Next**.

10. In the "Select grouping parameters and analysis inputs" window, set up the parameters as in figure 19.

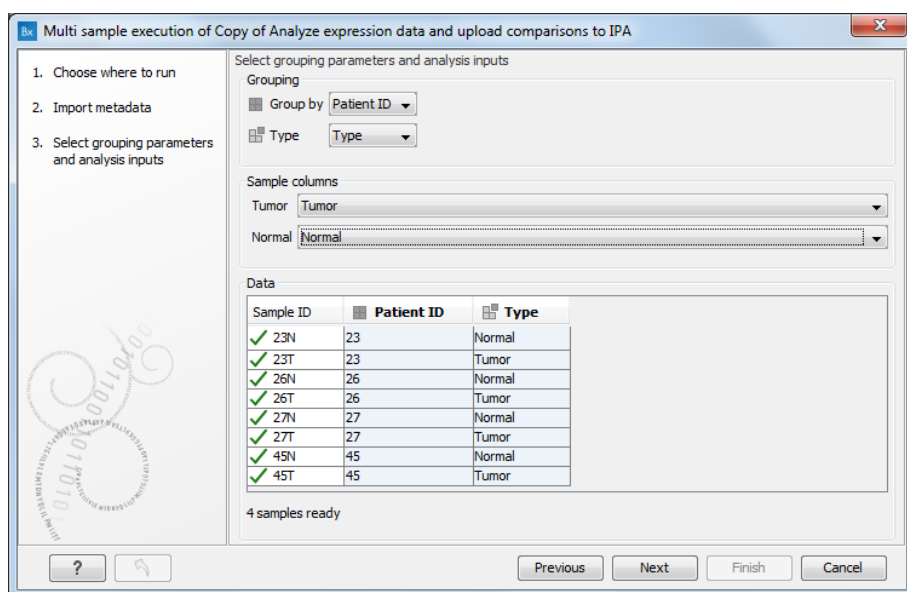


Figure 19: Parameters for the "Select grouping parameters and analysis inputs" window.

- In the "Differential Expression for RNA-Seq" window, set up the parameters for the algorithms that will be run as shown in figure 20.

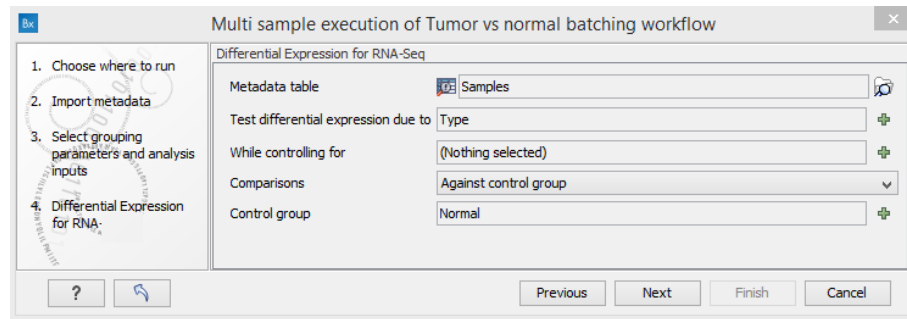


Figure 20: Parameters for the statistical tool.

- In the next window, use the default parameters for the Pathway Analysis tool and enter your Ingenuity username and password.
- For clarity, choose to save the results into a new folder called "IPA batch".

After you click Finish, the algorithms will start to run. They will run 4 times in total, as four executions will be carried out for the 4 patients.

Results of the advanced batching by patient

We will now analyze the results of the advanced batching of the workflow. Each batch unit is saved in a different subfolder within the "IPA batch folder".

Venn diagram

1. First, rename the output objects by appending the patient ID to the names of the resulting statistical comparison tracks (figure 21).

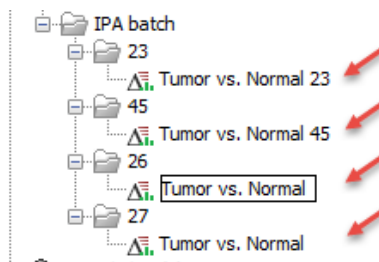


Figure 21: Rename the "Tumor vs. Normal" object in each folder by appending the folder name to it.

2. Next, go to: **Toolbox | RNA-Seq Analysis | Create Venn Diagram for RNA-Seq**. You can also start this tool by using the Launch button and then beginning to type "venn".
3. In the first wizard window, select the four output tracks you just renamed.
4. Choose to open the result.

After a few seconds, the Venn diagram is displayed in the View Area of your workbench (see figure 22).

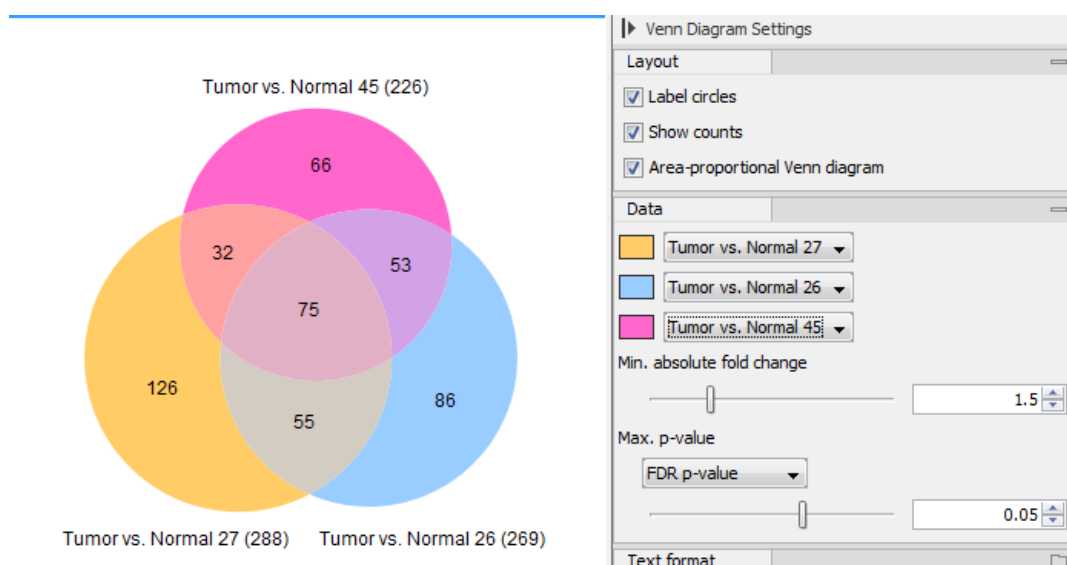


Figure 22: An example of Venn Diagram. Results might differ slightly from the ones you see in the figure.

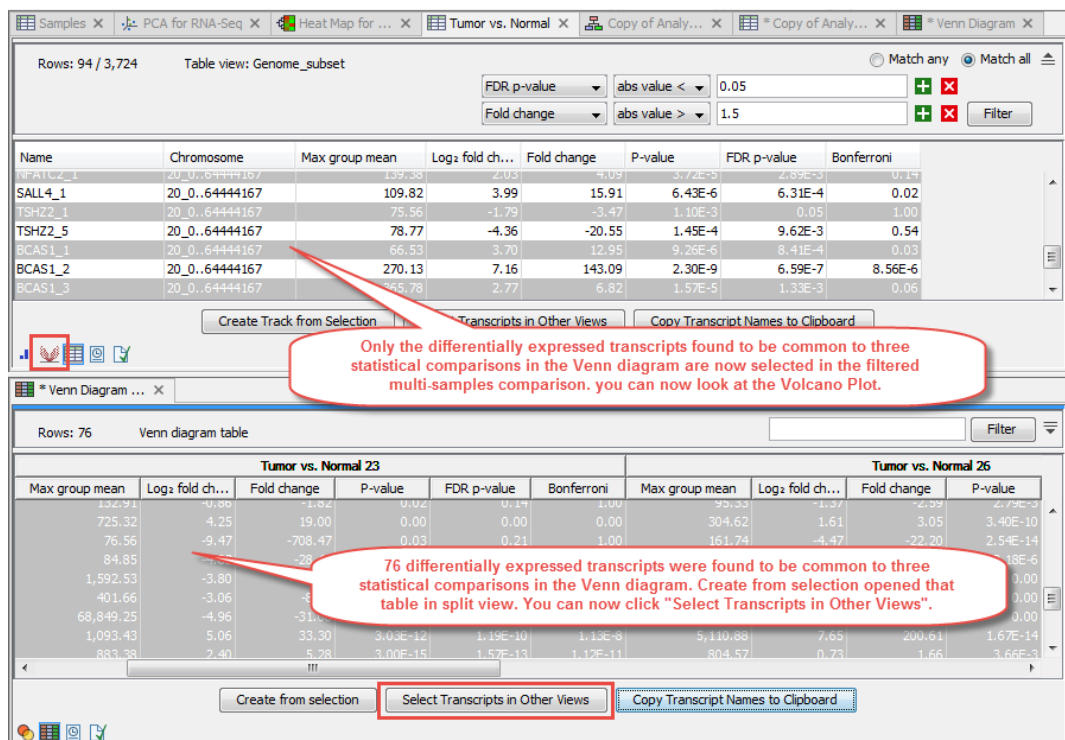
Tutorial

The Venn diagram shows how many transcripts were detected to be differentially expressed in each comparison. Note: only 3 comparisons can be visualized at a time. You can select which comparisons should be visualized in the side panel. In this example, 76 transcripts were detected to be differentially expressed in all three comparisons.

Select the middle intersection with the transcripts that were differentially expressed in all comparisons. With the selection maintained, go to the table view of the Venn diagram. The same transcripts will still be selected in the table. It can be hard to find them, however you can create a new selection only containing selected transcripts by clicking the **Create from selection** button. This will open a new table in split view. This table contains only the 76 transcripts that were differentially expressed in all three comparisons for the example above.

If you want to see how these transcripts look in the overall comparison of Tumor vs. Normal generated before the batch analysis (see figure 23):

- Select all transcripts from the table you "created from selection", and click "Select Transcript in Other Views".
- Reopen the tab holding original Tumor vs. Normal comparison (with the samples for all patients) in split table view. The filters set up earlier are still active.
- Click on "Select Transcripts in Other Views" of the "created from selection" table.



The screenshot shows the Ingenuity Pathway Analysis interface. The top window displays a table of transcripts with columns: Name, Chromosome, Max group mean, Log₂ fold ch..., Fold change, P-value, FDR p-value, and Bonferroni. A red callout points to the 'Transcripts in Other Views' button, stating: "Only the differentially expressed transcripts found to be common to three statistical comparisons in the Venn diagram are now selected in the filtered multi-samples comparison. you can now look at the Volcano Plot." Below this, the 'Venn Diagram ... X' window is open, showing a table with columns for 'Tumor vs. Normal 23' and 'Tumor vs. Normal 26'. A red callout points to the 'Create from selection' button, stating: "76 differentially expressed transcripts were found to be common to three statistical comparisons in the Venn diagram. Create from selection opened that table in split view. You can now click 'Select Transcripts in Other Views'." The 'Create from selection' button is highlighted with a red box.

Figure 23: Select a subset of transcripts and analyze them in the context of the multi samples comparison.

You can see that only 34 of the 76 differentially expressed transcripts are predicted to be significant in the multi-sample analysis. Switching to the Volcano plot view, you can see where they are located compared to other transcripts.

Tutorial

For the sake of completeness, we mention that the "Select transcripts in other views" button will allow you to do synchronized selections between expression tracks, statistical comparison tracks (including volcano plots), and the table view of Venn diagrams.

Finally, you can also Launch IPA to look for genes and associated pathways that are differentially expressed between tumor and normal samples. Remember that we are only analyzing genes present on chromosome 20 in this tutorial, and thus IPA might not output a comprehensive picture. For more information on IPA, see <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>.
