

Small RNA Analysis using Illumina Data June 27, 2019

Sample to Insight -

-

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark Telephone: +45 70 22 32 44 · www.qiagenbioinformatics.com · ts-bioinformatics@qiagen.com



Small RNA Analysis using Illumina Data

This tutorial shows how to go through the initial steps of analyzing a small RNA data set. It shows how to analyze one sample including how to trim off adapter sequences and extract the small RNAs, how to count the small RNAs, inspect and check the results and finally how to annotate the small RNAs to identify known miRNAs and other non-coding RNAs.

The tutorial is based on the study published by Stark et al., 2010. In this study, 12 samples from melanoma cell cultures were sequenced using an Illumina Genome Analyzer II. In this tutorial, we will analyze one of these samples, the MELB sample, which represents a primary melanocyte cell.

Downloading and importing the raw data

- 1. Download the fastq file and the associated metadata using the **Download | Search for Reads in SRA** button.
 - In the SRA search window that opens in the View Area (figure 1), enter SRR038853 in the search field and click on **Start search**.

Figure 1: Searching the SRA database to download the raw reads.

- Clicking on the Run Accession number will display the title and abstract of the publication associated with the selected run in the "SRA Preview" panel on the right hand side.
- Click on the button "Download Reads and Metadata".
- Make sure the **Discard read names** and **Discard quality scores** checkboxes are checked. Information about read names and quality scores are not used in this analysis so it would just take up disk space if imported with the data. Click **Next**.
- Choose to **Save** the reads in a new folder you can call **Small RNA tutorial** and click **Finish**.

After a short while, the reads have been imported, together with a metadata table. Open the sequence list file and place your mouse on the tab. After one second, you will see a small tool tip with information about the number of reads in the file as shown in figure 2.

I SRR038853 ×
20 Sequence List: SRR038853 No nam Size: 2,070,061 sequences 20 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
No name TGAGGTAGTAGGTTGTATAGTTTCGTATGGCGTGTT
20 I Noname TGAGGTAGTAGATTGTATAGTTTCGGATGCCGTCTT
Figure 2: This data set contains about two million reads.

2. Download the list of homo sapiens non-coding RNAs from ENSEMBL

ftp://ftp.ensembl.org/pub/release-84/fasta/homo_sapiens/ncrna/Homo_ sapiens.GRCh38.ncrna.fa.gz.

3. Import the list with the standard import option:

File | Import | Standard import

Select the compressed Homo_sapiens.GRCh37.75.ncrna file. Click Next, then Finish.

Trimming adapters and counting the reads

The next step in the analysis is to trim off the partial adapter sequences and subsequently to count how many copies there are of each of the resulting small RNAs. For this you will have to create an adapter list using the oligonucleotide sequences disclosed in the Illumina Customer Sequence Letter found at this address: https://support.illumina.com/downloads/illumina-adapter-sequences-document-100000002694.html.¹

To trim an adapter, you have to search for the reverse complement of the relevant adapter sequence. Here we want to trim the 3'RNA adapter called "v1.5 Small RNA 3' Adapter". Copy the sequence between the 5rApp and 3ddC additional.

- 1. Go to File | New | Trim Adapter List (=)
- 2. Click on the button **Add Row** () found at the bottom of the View Area in the New Trim Adapter List.
- 3. Give the adapter a name.
- 4. Paste the 3'RNA adapter called "v1.5 Small RNA 3' Adapter" (figure 3).
- 5. Choose to Remove the adapter and following sequence.
- 6. For reads without adapters, chose to "Discard the read". Click Next.
- 7. Keep alignments scores as they are set by default.
- 8. In the "Match thresholds section", choose to **Allow end matches**, and increase the **Minimum score at end** to 6 (figure 4).
- 9. Click on the button labeled **Finish** to create the trim adapter list.

¹Oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved. No sponsorship or affiliation. Link provided for convenience. QIAGEN not responsible for content at link.



Trim Adapter	×
1. Create or edit adapter	Create or edit adapter
2. Set scoring	
	Trim adapter
	Name 3' RNA
	Reverse Complement
	When an adapter is found
	AGCT
	\bigcirc Remove the adapter and preceding sequence (5' trim)
	Remove the adapter and following sequence (3' trim) Discard the read
	For reads without adapters
	© Keep the read
	Discard the read
Help Reset	Previous Next Finish Cancel

Figure 3: Creating the adapter list. Note that the sequence written in the figure is a mock one to respect Illumina copyrights.

Gx Trim Adapter		×
1. Create or edit adapter	Set scoring	
2. Set scoring	Alignment scores costs Mismatch cost 2 + Gap cost 3 +	
	Match thresholds Allow internal matches Allow end matches Allow both internal and end matches Minimum internal score Image: Allow both internal score Minimum end score 6 +	
Help	t	Previous Next Finish Cancel

Figure 4: Setting scores for the adapter.

- 10. Save the generated trim adapter list as **New Trim Adapter List** in the **Navigation Area**. You can do this by clicking on the tab and dragging and dropping the trim adapter list to the desired destination, or you can go to **File** in the menu bar and the choose **Save as**.
- 11. Now we are ready to trim the adapters from our reads. Launch () the **Extract and Count** () tool.

This opens a dialog where you select the SRR038853 sample as shown in figure 5. Click **Next**.

- 12. In the "Set trim options" dialog, make sure the checkbox for "Perform custom adapter trimming before counting" is checked (it is by default) and click **Next**.
- 13. In the next window (figure 6), select the trim adapter list.



	Select sequencing reads	
. Select sequencing read	Navigation Area	Selected elements (1)
. Set trim options	Small RNA tutorial	▲ IF SRR038853
. Adapter trimming	Homo_sapiens.GRCh38.ncrna	-
. Sequence filtering	Q- <enter search="" term=""></enter>	≜
Bandt bandling	🔲 Batch	

Figure 5: Selecting the sample for extracting and counting the small RNAs.

Gx Extract and Count						×
 Select sequencing reads Set trim options Adapter trimming Sequence filtering 	Adapter trimming	rough adapter trimmi Trim adapter list 🗮 🗌 Use color space	ing New trim adapter list		ଲ୍ଲ	
5. Result handling	Preview Number of reads 1,	,000	Number of nucleotides	36,000	Avg. lengt	n 36
178	Name	Matches found	Reads discarde	d Nucleotides	removed A	vg. length
rear of the second s	3' RNA	570	430	25,642	18	.2
Help Reset			Pre	evious Next	Finis	h Cancel

Figure 6: Trimming the raw data for adapter sequence.

In the preview panel below you can see the number of matches found among the first 1000 reads for this adapter. We will see more statistics on this for the full data set later on - this preview is just intended to support the user when defining the adapter trim setting. Click **Next**.

14. You will now see the dialog shown in figure 7. The most important settings here are the minimum and maximum lengths for the tags you wish to include when counting and how many copies there have to be for a particular tag to be included in the output.

Leave these options set to the default values and click Next.

Gx Extract and Count	x
1. Select sequencing reads	Set parameters
	Trim bases
Set trim options	Remove 5' terminal nucleotides 1
3. Adapter trimming	Remove 3' terminal nucleotides 1
4. Sequence filtering	Filter on length
5. Result handling	Discard reads below length 15
10	Discard reads above length 55
10	Sampling threshold
10 American	Minimum sampling count 1
Help Reset	Previous Next Finish Cancel

Figure 7: Adjusting options for counting the small RNAs.

15. You are now asked to specify the output options, as shown in figure 8. The default is to output a **Sample**, which is the table of all the small RNAs and their counts, and to create a report showing summary statistics. Leave the default settings, choose to **Open** the results rather than Save them, and click **Finish**.



Gx Extract and Count		
Select sequencing reads Set trim options Adapter trimming Sequence filtering Sequence filtering	Result handling Output options	
000	Result handling Open Save	
To an a statement	Log handling	
Help Reset	Previous Next Finish Cancel	

Figure 8: Selecting the results to output.

Interpreting the adapter trim report

Once the analysis is complete, two tabs will be opened. First, we take a look at the report (figure 9).



Figure 9: The small RNA counting report.

The report is meant to be used as a quality check, mainly to see that the adapter trimming worked as expected.

In this example, it shows that more than 83% of the reads were trimmed. The trim settings set earlier implied that if no adapter sequence was found, the read would have been discarded. A graph showing a distribution of the read lengths before and after trimming. In this example, there is a very nice distribution with a peak around 22 bp which is expected for miRNAs.



Save and close the report.

Investigating the small RNA sample

You should now see the small RNA sample. There are more than 50,000 unique small RNAs in the sample. You can filter and sort the sample, and you can extract subsets using the buttons at the bottom of the view.

1. Sort the table on **Length** by clicking on that column header once. This should sort the table so that the entries with the shortest lengths are at the top (figure 10).

			Filter	ן =	Small RNA sam	ple Settings
Rows: 88,470			Fliter	•	Column width	_
Small RNA	Expression values	Length	Count			Automatic 👻
CCCGGGCGTCGGCACCC	1.00	17		1 🔺	Show column	
AGCGACATTGTCTCTGGGTTTC	1.00	22		1 =		
TAAAAATTTCGGTTGGTA	1.00	18		1		Small RNA
TATCACCATCTGAAATCGTTTA	1.00	22		1		Expression values
AGCAGCAGTGGACAGGGCGGCGA	1.00	23		1		
CTACAGTCCCACGACC	1.00	16		1		V Lengun
TACCACAGGGTTGAACCACTGACA	1.00	24		1		Count
ACAGCAAGACGAGAAGACCC	1.00	20		1		Select All
AGGGGTATGAGTCTCGGT	1.00	18		1		
ATAGTTAGTACGCTTCAT	1.00	18		1		Deselect All
ACCTGGCATACAATGGAGATTTCTGT	1.00	26		1		
CAACGGAACCCCAAAAGCAGCTGT	1.00	24		1		
ATAGCTAAGGTCCAAA	1.00	16		1		
				1		
CGGAGAGAAAGGCAGTTCCTTA	1.00	22		1		
TCACTGGGAATTCAGCCTCTGAA	1.00	23		1		
AAGCCCTTACCCCAAAAAGAAT	1.00	22		1		
ACTGGACTTGGAGGCACAATGC	1.00	22		1 *		
Extract Reads	ads 🛛 📑 Extract Sm	all RNAs 🛛 🗰 Crea	ate Sample from Selec.)		
🖾 🖚 🛽 🖪					- 다 태	? ⊨

Figure 10: The small RNA sample table with all possible columns selected for view.

 Now click the row at the top of the table so it is selected, then click on the Extract Reads button. Keep the option DNA and choose to Open the result. Click on the button labeled Finish.

You should now be able to see the original read sequence(s) and a trim annotation as shown in figure 11.

	Trim
No name	GAAGCACATGAGACTCCCTCGCATGCCGTCTTCTGC

Figure 11: A sequencing read with the trim annotation.

3. Clicking the **Double stranded** checkbox in the Side Panel to the right under **Sequence Layout**, you can see the minus strand as well, and you can see that the adapter sequence has a perfect match in the reverse orientation here (figure 12).



Figure 12: A sequencing read shown as double stranded with the trim annotation.

4. Save and close the small RNA sample.



Downloading miRBase and annotating the sample

The next step in the analysis is to annotate the small RNA sample so known small RNAs can be identified. We use two sources for the annotation here:

- miRBase, which will be used to identify known miRNAs
- a set of other known non-coding RNAs, which were downloaded at the beginning of this tutorial.
- 1. You can download the latest version of miRBase directly via the Workbench by using the **Download miRBase (**) tool.
 - Search for the tool using the Launch (ot p
 ot) button.
 - Choose to Save the files in the folder created for this tutorial and click on Finish.
- 2. Launch the Annotate and Merge Counts (main tool to start annotating.
- 3. This opens a dialog where you select the **sra-data Small RNA sample** as shown in figure **13**. Click **Next**.

Select Small RNA sample or experiments of such Specify annotation resources Specify match parameters	Select Small RNA samples or experiments of such Navigation Area Small RNA tutorial Q	Selected elements (1)
Heln Reset	Previo	us Next Einish Cancel

Figure 13: Selecting the sample for annotating the small RNAs.

4. You are now presented with the dialog shown in figure 14. At the top, check to use miRBase and select () the miRBase - Release 21 file that you have just downloaded. Below, check the Use other resource and select () the Homo_sapiens.GRCh38.ncrna file that you imported in the beginning of this tutorial. Leave the other option set such that miRBase has the highest priority.

Gx Annotate and Merge Count	2	X
1. Select Small RNA samples or experiments of such	Specify annotation resources miRBase	
2. Specify annotation resources	✓ Use miRBase ImiRBase - Release 21	ର
 miRBase species Specify match parameters 	Other source	
5. Result handling	Immo_sapiens.GRCh38.ncma	ଇ
1017670 M	Priority miRBase has highest priority miRBase has lowest priority 	
Help Reset	Previous Next Finish Canc	el

Figure 14: Setting miRBase and the other non-coding RNAs as annotation sources.

The miRBase file contains a list of precursor sequences with specification of the mature 5' and in some cases the mature 3' regions. This information is used to categorize the annotated small RNAs. The "Other resource" does not include this kind of information and is used here in order to identify known small RNAs that are not miRNAs. Note that you could include several sequence lists here if you had other sources of non-coding small RNAs.

Click Next.

5. In this window (figure 15), select **Homo sapiens** and then select **Mus musculus**. Note that the order in which these are listed is important as it is the order in which the resources will be used for the purpose of annotation. The sample is human, so human annotations should have the first priority and thus should be the first source listed. Since there may be miRNAs that have not yet been identified in human but have an ortholog in mouse, including the mouse miRNAs may provide useful information. Click **Next**.

Gx Annotate and Merge Count	ts				×
 Select Small RNA samples or experiments of such 	Specify miRBase species Prioritization of miRBase species				
and	Species			Selected species	
2. Specify annotation	Acacia auriculiformis			Homo sapiens	
resources	Acacia mangium			Mus musculus	
3. miRBase species	Acyrthosiphon pisum		10		
A State of the second s	Aedes aegypti				
4. Specify match parameters	Aegilops tauschii				
The second second	Amborella trichopoda		=		
5. Result handling	Amphimedon queenslandica				
2 O Manutalina	Anolis carolinensis	-			
	<u> </u>				
Help Reset			Previous	Next Finish	Cancel

Figure 15: Prioritizing species for annotation.

6. In the "match parameters" dialog (figure 16), leave all settings at their default values and click **Next**.

Gx Annotate and Merge Counts	×
1. Select Small RNA samples or experiments of such	Specify match parameters
2. Specify annotation	Mature length variants (IsomiRs)
resources	Additional downstream bases 2 -
3. miRBase species	Missing upstream bases 2
4. Specify match parameters	
S. Result in J. Re	Alignment settings Maximum mismatches 2
1001 11	Strand specific alignment
	Use color space
Help Reset	Previous Next Finish Cancel

Figure 16: Thresholds for annotating.

7. In the dialog shown in figure 17, make sure all options except the "Create unannotated samples/experiments" are checked, choose to **Save** the results and click **Next**. You can now select the folder in which to save them. Click on the button labeled **Finish**.



GX Annotate and Merge Coun	ts X
1. Select Small RNA samples or experiments of such	Result handling Output options
 Specify annotation resources 	Create unannotated samples/experiments Create annotated samples/experiments
3. miRBase species	Create grouped samples/experiments, grouping by Precursor/Reference
 Specify match parameters Result handling 	 Create grouped samples/experiments, grouping by Mature Create report
6. Save location for new elements	Result handling Open Save
TTBUTOT	Log handling
Help Reset	Previous Next Finish Cancel

Figure 17: Select the sample grouped on mature 5'.

Analyzing the annotated samples

For a detailed description of the output, please refer to the user manual (press the **F1** key to display it).

In this example we focus on a few specific miRNAs to illustrate how the annotation and grouping of samples work and show some possibilities for interacting with the data.

Looking at the grouped sample

Open the **SRR038853 small RNA Sample annotated** (\Rightarrow). We want to look at *mir-29a*, so type this into the filter at the top of the table and click on the button labeled **Filter**. This will list all the tags that have been mapped to the mir-29a precursor sequence from miRBase (figure 18). If you sort the table by **Count** (clicking the count column header twice to show the highest count on top of the table), you can see that most of these are exact matches of the mature 3' miRNA (Match type is set to Mature 3'). The rest are variants and length variants.

Small RNA Length Count r Name Resource Match type Mismatches TAGCACCATCTGAAATCGGTTA 22 30689 mir-29a Homo sapiens Mature 3' TAGCACCATCTGAAATCGGTT 21 1224 mir-29a Homo sapiens Mature 3' TAGCACCATCTGAAATCGGTTA 22 685 mir-29a Homo sapiens Mature 3' super variant TAGCACCATCTGAAATCGGTA 22 645 mir-29a Homo sapiens Mature 3' super variant TAGCACCATCTGAAATCGGTTA 22 241 mir-29a Homo sapiens Mature 3' super variant TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' super variant TAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAAATCGGTTA 23 126 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo] ₹
TAGCACCATCTGAAATCGGTTA 22 30689 mar-29a Homo sapiens Mature 3' TAGCACCATCTGAATCGGGTT 21 1224 mir-29a Homo sapiens Mature 3' TAGCACCATCTGAATCGGGTA 22 685 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTA 23 377 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 241 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 22 165 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3'	
TAGCACCATCTGAAATCGGTT 21 1224 mr-29a Homo sapiens Mature 3' sub TAGCACCATCTGAAATCGGTA 22 685 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 377 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 241 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 241 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 23 240 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 186 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTT 23 150 mr-29a Homo sapiens Mature 3' sub TAGCACCATCTGAAATCGGTTA 23 150 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 148 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 147 mr-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 148 mr-29a Homo sapiens Mature 3' varian	• 0
TAGCACCATCTGAAATCGGGTA 22 665 mr-23a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 377 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 241 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 165 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens	0 =
TAGCACCATCTGAAATCGGTTA 23 377 mr-29a Homo sapiens Mature 3' super variant TAGCACCATCTGAAATCGGTTT 22 241 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 166 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 132 mir-29a Homo sapiens <td>1</td>	1
TACCACCATCTGAAATCGGTT 22 241 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant CTACCACCATCTGAAATCGGTA 23 240 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 186 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 186 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' sub TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 147 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens	1
TAGCACCATCTGAAATCGGTTA 22 240 mir-29a Homo sapiens Mature 3' variant CTAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' super TAGCACCATCGGAAATCGGTTA 22 186 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGGAAATCGGTTA 20 165 mir-29a Homo sapiens Mature 3' super TAGCACCATCGGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' super TAGCACCATCGGAAATCGGTTA 22 149 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGGAAATCGGCTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGGAATCGGCTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAATCGGCTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAATCGGTTA 22 132 mir-29a Homo sapiens <td< td=""><td>1</td></td<>	1
CTAGCACCATCTGAAATCGGTTA 23 240 mir-29a Homo sapiens Mature 3' super TAGCACCATCGGAAATCGGTTA 22 186 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAAATCGGT 20 155 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAAATCGGTTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTTA 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens <t< td=""><td>1</td></t<>	1
TAGCACCATCGGAATCGGTTA 22 186 mir-29a Homo sapiens Mature 3' variant TAGCACCATCGAATCGGTT 20 165 mir-29a Homo sapiens Mature 3' sub TAGCACCATCTGAATCGGTT 23 150 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTT 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant TCGACCATCTGAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant	0
TAGCACCATCTGAAATCGGT 20 165 mir-29a Homo sapiens Mature 3' sub TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' sub TAGCACCATCTGAAATCGGTTA 23 150 mir-29a Homo sapiens Mature 3' subra TAGCACCATCTGAAATCGGTTA 22 140 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGGTA 22 137 mir-29a Homo sapiens Mature 3' variant CTAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant	1
TAGCACCATCTGAAATCGGTTAT 23 150 mir-29a Homo sapiens Mature 3' super TAGCACCATCTGAAATCGGTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTA 22 137 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGTT 22 137 mir-29a Homo sapiens Mature 3' variant TCAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant TCAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' variant	0
TAGCACCATCTGAAATCTGTTA 22 148 mir-29a Homo sapiens Mature 3' variant TAGCACCATCTGAAATCGGCTA 22 137 mir-29a Homo sapiens Mature 3' variant CTAGCACCATCTGAAATCGGCTA 22 132 mir-29a Homo sapiens Mature 3' variant CTAGCACCATCTGAAATCGGCTA 22 122 mir-29a Homo sapiens Mature 3' sub/super VECACCATCTGAAATCGGTT 22 127 mir-29a Homo sapiens Mature 3' sub/super	0
TAGCACCATCTGAAATCGGGTA 22 137 mir-29a Homo sapiens Mature 3' variant CTAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' sub/super VCACCATCTGAAATCCGGTT 22 132 mir-29a Homo sapiens Mature 3' sub/super	1
CTAGCACCATCTGAAATCGGTT 22 132 mir-29a Homo sapiens Mature 3' sub/super	1
TACCACCATCTCACATCCCTTA 22 117 mir 205 Home explored Mature 2' variant	0
TAGCACCATCTOACATCGOTTA 22 117 mm-298 Homo sapiens Mature 5 Variant	1
TAGCACCATCTGCAATCGGTTA 22 115 mir-29a Homo sapiens Mature 3' variant	1
TAGCACCATCTGAAATCGG 19 111 mir-29a Homo sapiens Precursor	0
TAGCACCATCTTAAATCGGTTA 22 93 mir-29a Homo sapiens Mature 3' variant	1 -
III Extract Reads III Extract Trimmed Reads III Extract Small RNAs Create Sample from Selection	

Figure 18: Showing all tags annotated with mir29a.

For expression analysis, it can make sense to look at all the variants of the same miRNA as one entity rather than 196 items as it is the case here. Open the sra-data Small RNA sample grouped (P) and type in mir-29a in the filter. You now have two lines representing all the tags that have been annotated with mir-29a. The upper row shows mir-29a from human and the lower row shows mir-29a from mouse. The total number of counts for the human mir-29a is 33,109. The number of reads in different categories are shown, e.g. 2,584 for the exact mature 3'. You also see a few tags annotated with the mouse ortholog, but this could be noise due to sequencing errors.

Double-click the human mir-29a row to open the mapping of all the tags to the precursor sequence (see figure 19).



Figure 19: Showing the alignment against the mir29a precursor sequence.

The tags are colored to reflect the counts, which are also shown in numbers next to the name to the left. As the exact mature 3' is very dominant in terms of count, it is the only one standing out in a different color. Note that it may be necessary to adjust the user settings in the side panel to obtain the same view as shown in figure 19.

Tracking back from the sample grouped on mature

Now, open the **SRR038853 Small RNA** sample grouped on mature (P) and look at the first row with let-7f-1//let-7f-2 in the Name column. The let-7f miRNA is annotated in miRBase with two different precursor sequences. This means that when the tags are annotated, they are assigned either to let-7f-1 or let-7f-2. The sample grouped on mature merges the tags from precursors sharing the same mature 5' sequence (the sequence itself is shown in the **Feature ID** column).

Open the **SRR038853 Small RNA sample grouped** and enter let-7f in the filter. The two precursor variants are now displayed, and you can see that the numbers sum to the numbers given in the sample grouped on mature: 25,349 + 25,344 = 50,693 (see the **Mature 5'** column in figure 20).



Rows: 2 / 6,539												e	t-7f			Filter
Expression values	Name	Resource	Exact ma	at Mature	5' - U	Jnique ex L	Unique ma	Exact mat	Matur	e 3'	Unique ex	Unique mat	Exact othe	er Othe	er	Total
25,349.00	let-7f-2	Homo sapiens		1557	25349	0	3	3	0	0	C	0	24	9514	255873	2812
25,344.00	let-7f-1	Homo sapiens		1556	25344	0	1	.4	0	10	C	10	24	9361	255697	2810
			- Open Read Mappin		Treate Sam	inle from Selecti	on P	voression val	e Matu	o 5'	_1					
			- observes and the		areare being	the more served.					•					
s 🔛 🖸 🕼																
S 🗄 🛛 🖌																
🖼 🔠 🔯 🚺	RNA sample gr	. ×														
🖼 🔛 💿 🚺	RNA sample gr	. x														Filter
E E O L	RNA sample gr	. x														Filter
SRR038853 Small Rows: 559 Feature ID	RNA sample gr.	. ×	Name /		Res	source		Matd	ı type	Exa	t mature	Mature	Unique	exact m	Unique r	Filter
Rows: 559	RNA sample gr.	Expression values	Name /		Res	source no sapiens		Matu	i type	Exa	t mature	Mature	Unique	exact m	Unique r	Filter nature 47
Rows: 559 GAGGIAGIAGATIGIA CALLERATION	RNA sample gr	Expression values	Name , let-7f-1//et-7f-2 let-7f-2		Res	source no sapiens no sapiens		Matu Matur Matur	type 5' 2 3'	Exa	:t mature 3113 0	Mature 5	Unique 0693 0	exact m	Unique r	Filter nature 47 0
Rows: 559 Redure ID Redure ID RAGGTAGTAGATIGTA GAGGTAGTAGTTGTA	RNA sample gr ATAGTT CTTTCC ACAGTT	Expression values 50,693,07 0.00 1,287,00	Name ,)let-7f-1//let-7f-2)let-7g		Res Hom Hom	source 10 sapiens 10 sapiens 10 sapiens		Matu Matur Matur Matur	type 5' 2 3' 2 5'	Exa	:t mature 3113 0 73	Mature	Unique 0693 0 1287	e exact m	Unique r 0 3	Filter nature 47 0 1251
Rows: 559 Feature ID TCACGGTACTACTGT CAGGGTAGTAGTTGTT CGAGGTAGTAGTTGTT CGGCGCAAGCTACTGC	RNA sample gr ATAGTT CTTTCC ACAGTT CCTTGCT	Expression values 50,653,07 0,00 1,287,00 15,50,00	Name > let-7f-1//et-7f-2 let-7g let-7g let-7		Res Hom Hom Hom	source no sapiens no sapiens no sapiens no sapiens		Matu Matur Matur Matur Matur Matur	i type 5' 3' 5' 3'	Exa	t mature 3113 0 73 0	Mature	Unique 1693 0 1287 15	e exact m	Unique r 0 3 0	Filter nature 47 0 1251 15

Figure 20: The sample grouped on mature joins the counts of precursors sharing the same mature sequence.

RNA folding

One of the advantages of *CLC Genomics Workbench* is the integration between various tools. We are now going to explore the RNA secondary structure prediction tool using this miRNA.

1. Right-click on let-7f-2 (Homo sapiens) in the filtered **SRR038853 Small RNA sample** grouped and choose to "Open read mapping" (see figure 21).

													Filter
Expression values Name	Resource	Exact mat	Mature 5' 🗸	Unique ex	Unique ma	Exact mat	Mature 3'	Unique ex	Unique ma	Exact other	Other	Tr	otal
314,839.00 let-7f-2 300,322.00 let-7f-1	Open Read Mapping Create Sample from Selection	254121 254122	314839 300322	0	15616 1108	6 () 8 ()	0 1	0 (0) ()) 11	789 . 784	-	980 862	31581 30119
		and Mapping	Create S	ample from Sele	ection E	xpression value	e: Mature 5'	•					
S 🛅 🛛 🖓													
⊑ [let-7f-2 (H] ×													
		20 1			40 I			60 I		80			
	miRNA						miF	RNA					
let-7f-2 (Homo saniens		CTACATIO	TATACTIT	TACCOTC		TOTTOGA	CATAACT	ATACAGTO	TACTOTO	TTCCCAC	c		
let H 2 (Hollie Supjetio	Rename Sequence			1400010	11400007		GATAACI	ATAOAGTO	-IACIGIC	TTOCCAG	0		
Consensu	Select Sequence		ATAGTTA	AAGGT									
Mature 5' (ambiguous) (254.12	Extract New Consensus	Sequence	TACTT										
mature 5 (ambiguous) (254, 12	🚾 Open Sequence	,											
ature 5' super variant (ambiguous) (18,31'	BR Delete Open Sequence) • Canada Danaidi an	ATAGTTA										
Mature 5' super (14,24)	Sort Reads by Alignmen	t Start Position	ATAGTTT										
	Sort Reads by Length												
Mature 5' sub (ambiduous) (8,91t	BLAST Selection against	NCBI	ATAGT										
			TATT										
Mature 5' variant (ambiguous) (2,647	BLAST Selection against	Local Data	010111										
Mature 5' variant (ambiguous) (2,64 햚 [let-7f-2 (H] ×	BLAST Selection against	Local Data											
Mature 5' variant (ambiguous) (2,64 & [let-7f-2 (H] ×	BLAST Selection against	Local Data	40			60 I			80 I				

Figure 21: Showing the folding of the mir-29a precursor.

- 2. In the read mapping, right-click the let-7f-2 (Homo sapiens) label at the top of the mapping and select **Open Sequence**.
- 3. This will open this sequence in a new view. However, it is still part of the mapping and the grouped sample (this is denoted by the square brackets around its name in the tab of this new view).
- 4. We can now predict the secondary structure of this sequence by launching the tool **Predict Secondary Structure** () (choose the tool that belongs to the RNA Structure folder).



- 5. The sequence let-7f-2 is preselected. Click **Next** and **Next** using the default settings. Uncheck the output option to add annotations and click on the **Finish** button.
- 6. Switch to the **Secondary Structure 2D View** () to see the predicted structure.
- 7. If your views are not already split, drag the tabs of the views to create a set-up as shown in figure 22 and select using the mouse either in the secondary structure or the reference in the mapping view and you will be able to follow the selections across the views.



Figure 22: A split view showing the secondary structure of the RNA together with the length variants.

8. Close the views. You are prompted to save the changes, which in this case is the addition of the secondary structure to the precursor sequence.



Bibliography

[Stark et al., 2010] Stark, M. S., Tyagi, S., Nancarrow, D. J., Boyle, G. M., Cook, A. L., Whiteman, D. C., Parsons, P. G., Schmidt, C., Sturm, R. A., and Hayward, N. K. (2010). Characterization of the melanoma mirnaome by deep sequencing. *PLoS One*, 5(3):e9685.