
# Tutorial on small RNA-Seq data analysis

13th Annual Meeting of the Bioinformatics Italian Society
June 15-17, 2016, University of Salerno, Italy

# Summary

# 1 Introduction

## 1.1 Motivation

small RNA-Seq technologies have recently come to prominence as technique for characterizing small non-coding RNAs (sncRNAs) and comparative analysis of their abundances. It is used in many area of biology including functional genomics, developmental biology and cancer biology. In small RNA-Seq, mapped reads are typically aggregated to counts at some level of interest, such as miRNAs, tRNAs and other sncRNAs. The count for a given miRNA or other sncRNA its expression. We are very often interested in which miRNAs or other sncRNAs have different levels of expression under different experimental conditions.

## 1.2 Scope of this tutorial

This tutorial introduces use of iMir tool for the analysis of small RNA-Seq data. The goal is to develop familiarity with the several statistical and bioinformatics steps needed to analyze this kind of data

Our tutorial will consist of a hands-on demonstration, walking through an end-to-end analysis of a typical smallRNA-Seq data, going from uploading raw count data to exploring significant variation in miRNAs levels between different groups of samples.

The tutorial will give an overview of the theory behind the iMir tool and it will illustrate its features through examples. We will go through the following step:

1. Design and quality assessment of small RNA-Seq experiments
2. Count data normalization
3. Statistical analysis of differences
4. Downstream interpretative analysis

## 2 Input Data

As a running example we will use the dataset from an experiment performed on MCF-7 cells (human) maintained in two different culture conditions affecting cell cycle progression, e.g. growth-arrest and exponential growth. For each experimental condition, three sequencing replicates will be analyzed to gather a correct estimation of the variability occurring during differential expression analysis. The *.fastq* data are inside the folder *iMir_Samples*:

- Controls:

  SRR554400_Control1

  SRR554401_Control2

  SRR554402_Control3

- Treatment

  SRR554397_Test1

  SRR554398_Test2

  SRR554399_Test3

# 3 iMir Analysis

## 3.1 Run iMir

To run iMir open a terminal on Linux and move in the iMir directory.

Now you can type:

```
>python iMir.py
```

iMir main-window (Figure 1) will appear on screen.

## 3.2 Run complete Analysis

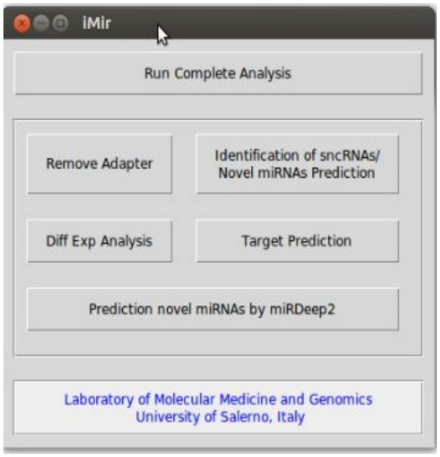


Figure 1

From the window showed in the Figure 1 select: Run Complete Analysis

### 3.3 Output Directory, Adapters Removal and Workflow

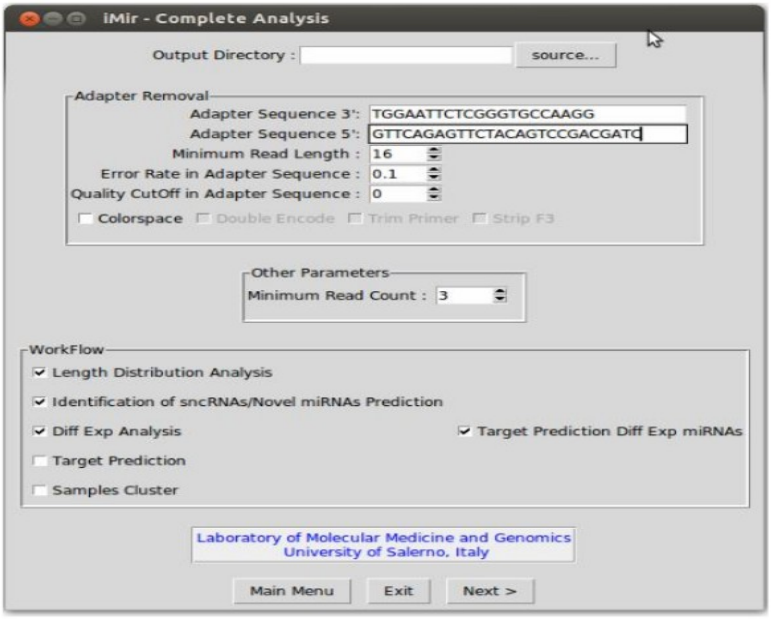


Figure 2

Select the output directory, in the window showed in the Figure 2, and insert these parameters:

**Adapter Sequence 3'**: TGGAATTCTCGGGTGCCAAGG

**Adapter Sequence 5'**: Empty

**Minimum Read Length**: 15

**Error Rate in Adapter Sequence**: 0.1

**Quality Cutoff in Adapter Sequence**: 0

**Colorspace**: No

**Minimum Read Count**: 3

**Select these options: Length Distribution Analysis, Identification of sncRNA/Novel miRNAs Prediction**

### 3.4 Input Folder and selection of samples

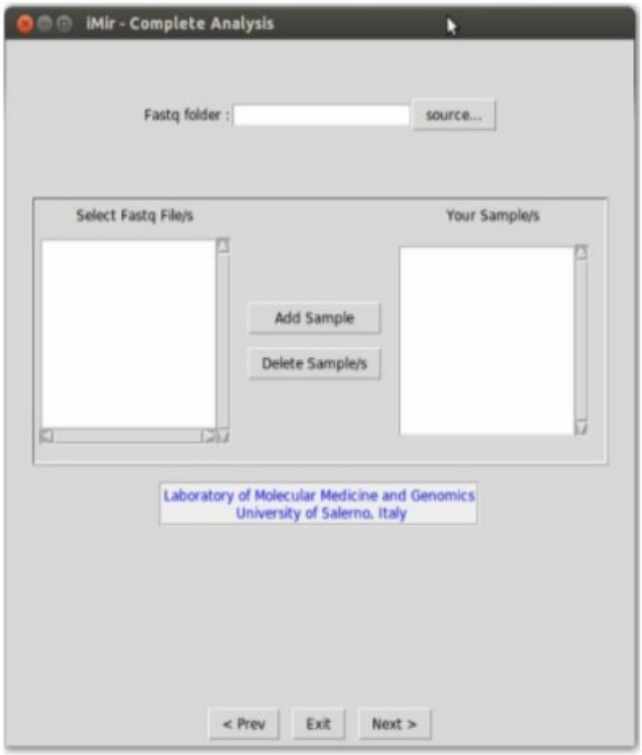


Figure 3

From the window showed in the Figure 3 select the folder of the FastQ files and load the Control sample with the name Controls and Test samples with the name Tests.

### 3.5 Analysis Parameters

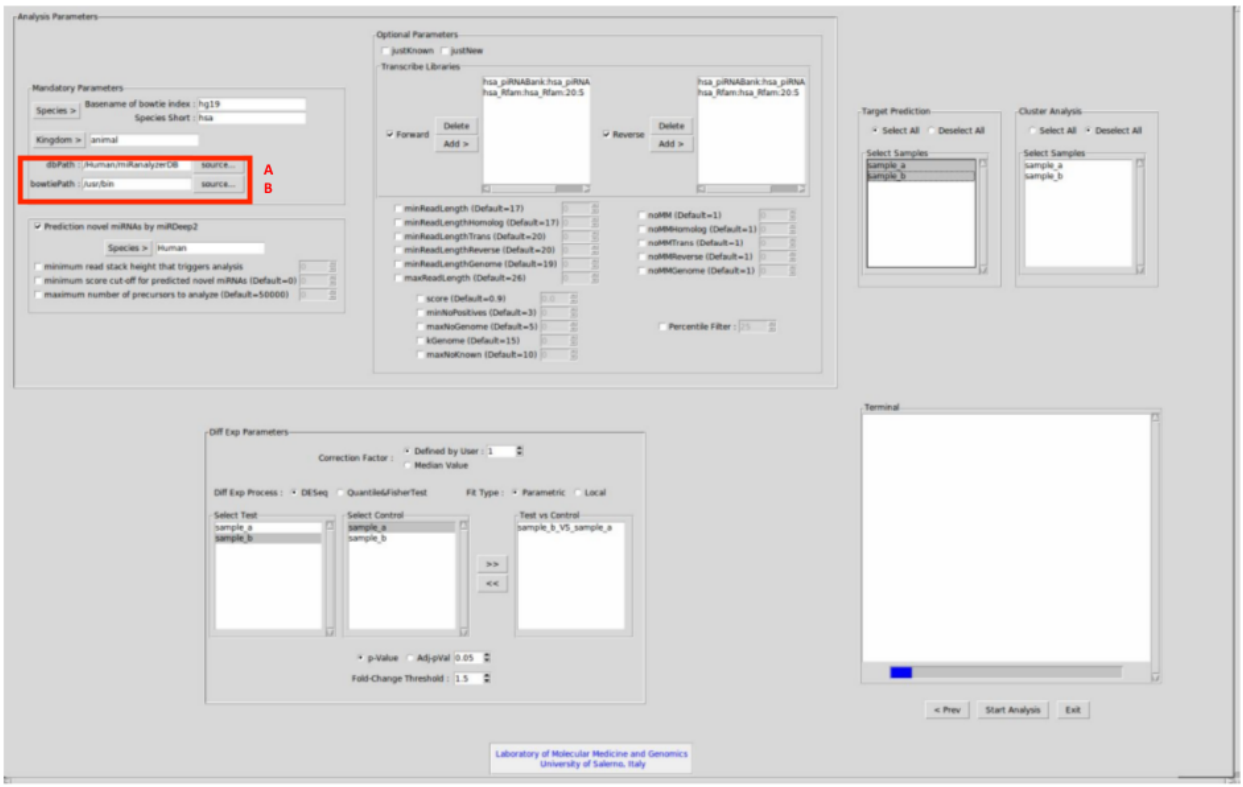


Figure. 4

Finally from the window showed in the Figure 4 select these parameters:

Click on Species and select: Human

Click on Kingdom and select: animal

Click on dbPath source and select the folder of miRanalyzer DB

Click on bowtiePath and select: usr/bin

From the DiffExp Parameters:

Click on Defined by User and type 1

Click on DESeq

Click on Parametric

In the Select Test click on Tests

In the Select Control click on Controls

Click on Adjpval and type: 0.05  Type in Fold Change Threshold: 1.5