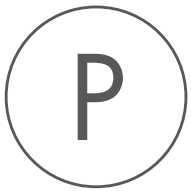

(Beta)

Advanced Peak Shape Tools Plugin

USER MANUAL

User manual for

Advanced Peak Shape Tools Plugin 1.0 beta 7

Windows, Mac OS X and Linux

February 15, 2017

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Running Advanced Peak Shape Tools

Advanced Peak Shape Tools Plugin provides advanced tools to manipulate peak shapes and find regions in sequencing data that exhibit a distinct shape. These tools are used by the ChIP-Seq algorithm (see http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=ChIP_Seq_Analysis.html), but they can also be used interactively to refine the results of ChIP-Seq peak calling or to perform shape-based analyses of sequencing data.

These tools are available under:

Toolbox | **Epigenomics Analysis** () | **Advanced Peak Shape Tools** ()

The **Learn Peak Shape Filter** tool allows you to build a new peak shape filter from sequencing data and a set of positive and negative regions.

The **Apply Peak Shape Filter** tool allows you to apply a peak shape filter to sequencing data to discover regions (peaks), which match a given peak shape.

The **Score Regions** tool allows you to apply a peak shape filter to score genomic regions according how they match a given peak shape.

1.1 Running the Learn Peak Shape Filter tool

The Learn Peak Shape Filter tool allows you to build a new peak shape filter from sequencing data and a set of positive and negative regions. The resulting filter can be used to identify genomic regions whose read coverage profile matches the characteristic shape of the positive examples and does not match the shape of the negative examples. The procedure used to build the peak shape filter is described here http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Learning_peak_shapes.html. An example of such filter is shown in figure 1.1.

To run the Learn Peak Shape Filter plugin:

Toolbox | **Epigenomics Analysis** () | **Advanced Peak Shape Tools** () | **Learn Peak Shape Filter** ()

This will open up the wizard shown in figure 1.2 where you can select the input data (for example the mapped ChIP-Seq reads). Multiple inputs are accepted, provided that they refer to the same

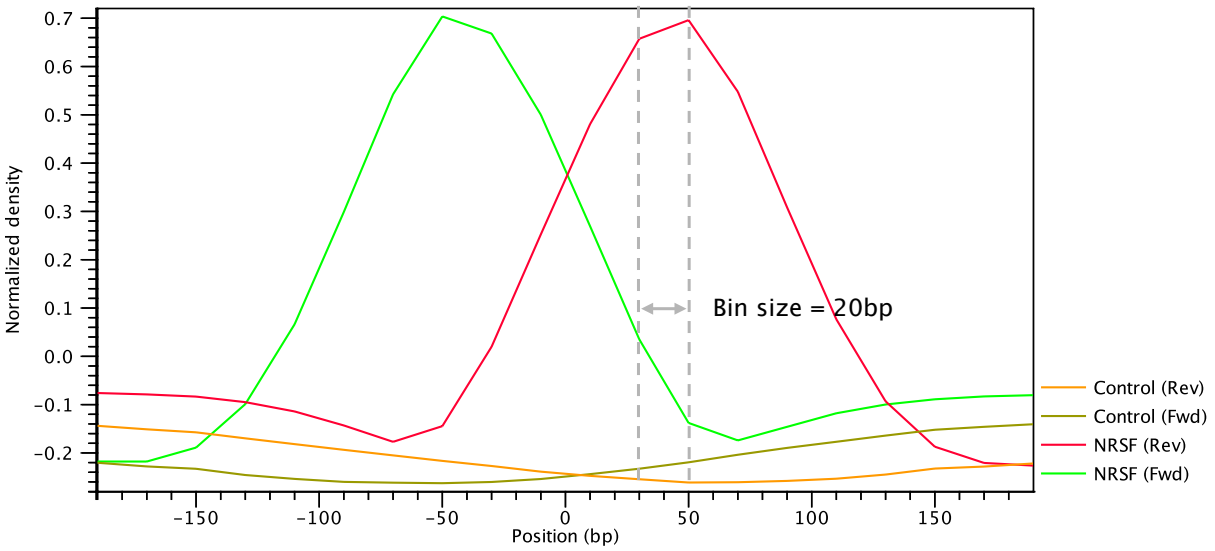

Figure 1.1: *Example of a peak shape filter with a window size of 400bp made up of 20 bins of size 20bp each. The filter was built from ChIP-Seq data of the transcription factor NRFS and a control ChIP-Seq experiment.*

genome. Track based read mappings ( ) and stand-alone read mappings ( ) / ( ) are both accepted.

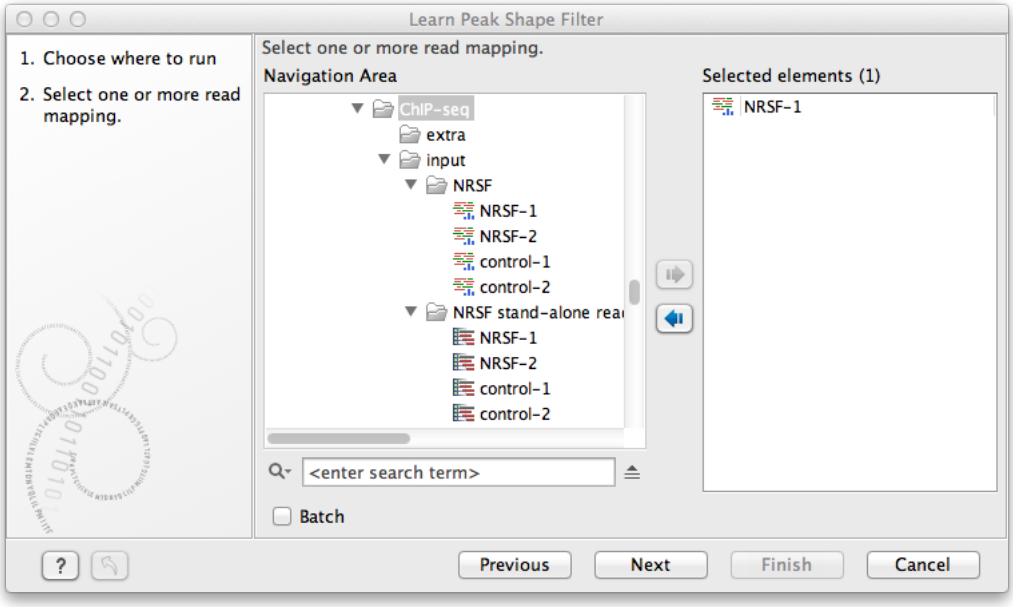

Figure 1.2: *Select the input data for Learn Peak Shape Filter.*

Click on the button labeled **Next** to go to the next wizard step (shown in figure 1.3).

In this wizard step you have the following options:

- **Location of positive regions** An annotation track ( ) containing the location of the positive regions (e.g. ChIP-Seq peaks) that will be used to build the peak shape filter. The set

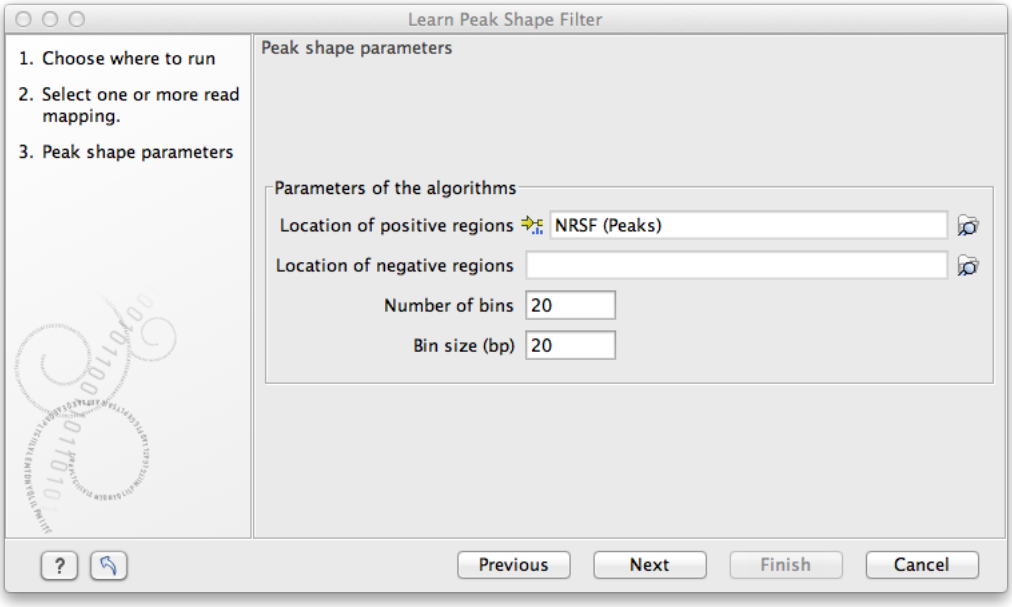

Figure 1.3: *Options for Learn Peak Shape Filter.*

of positive regions should include examples where the shape is clearly exhibited. It is preferable to have fewer peaks with high quality rather than a large amount of ambiguous peaks. Typically, a number of positive peaks greater than 5-10 times the number of bins is sufficient to learn a well-defined shape.

- **Location of negative regions** An annotation track ( ) containing the location of the negative regions that will be used to build the peak shape filter (e.g. background, PCR artifacts or examples of bad peaks from a previous run of the ChIP-Seq analysis tool). If no annotation track is provided, a negative profile will be derived from sequencing noise.

- **Number of bins** The number of bins to use to build the filter. The default value of 20 for the Number of bins parameter should be satisfactory for most uses. Higher values may be useful when the shape to be learned is particularly complex. Note that if the chosen number of bins is very large, the learned peak shape filter may not be smooth and could over-fit the input data. If only few positive regions are available, reducing the number of bins may be helpful.

- **Bin size** The size of each bin in base pairs. The bin size is related to the window size (i.e. the length of the shape to be learned) by the formula Window size $=$ Bin size $\times$ Number of bins (see figure 1.1).

The result of the algorithm will be a **Peak shape filter** ( ), which can then be applied to call peaks or score regions using Apply Peak Shape Filter. After clicking on the button labeled **Next**, you can choose whether you want to open the result directly, or save the results in the **Navigation Area**. If you choose to save the results, you will be asked to specify where you would like to save them.

1.2 Running the Apply Peak Shape Filter tool

The Apply Peak Shape Filter tool allows you to apply a peak shape filter to sequencing data to discover regions (peaks), which match a given peak shape (see http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Applying_peak_shape_filters_call_peaks.html).

To run the Apply Peak Shape Filter plugin:

> **Toolbox** | **Epigenomics Analysis** ( ) | **Advanced Peak Shape Tools** ( ) | **Apply Peak Shape Filter** ( )

This will open up the wizard shown in figure 1.4 where you can select the input data (e.g. mapped ChIP-Seq reads). Multiple inputs are accepted, provided that they refer to the same genome. Track based read mappings ( ) and stand-alone read mappings ( ) / ( ) are both accepted.

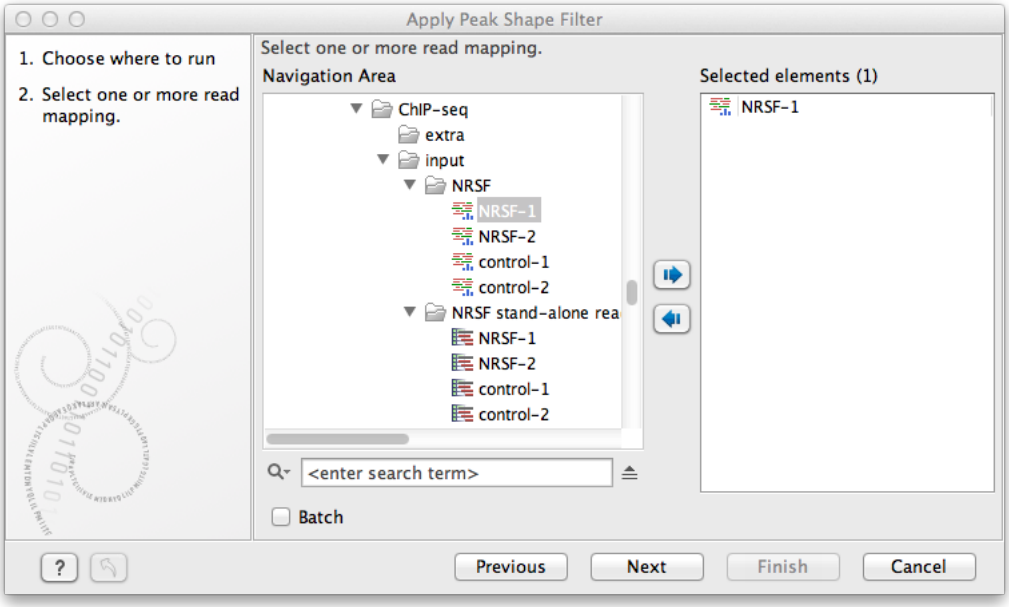

Figure 1.4: *Select the input data for Apply Peak Shape Filter.*

Click on the button labeled **Next** to go to the next wizard step (shown in figure 1.5).

In this wizard step you have the following options:

- **Peak shape filter** The peak shape filter ( ) to apply to the data. Peak shape filters can be obtained as the result of the ChIP-Seq Analysis tool. If no filter is given, a filter is derived from the input data.
- **Maximum P-value for peak calling** The threshold for reporting peaks can be specified by this option.

Click on the button labeled **Next** to go to the wizard step shown in figure 1.6.

In addition to the annotation track with **Peak annotations** ( ) that will always be generated by the algorithm, you can choose to select an additional output type:

- **Peak shape score ( )** A graph track containing the peak shape score. The track shows

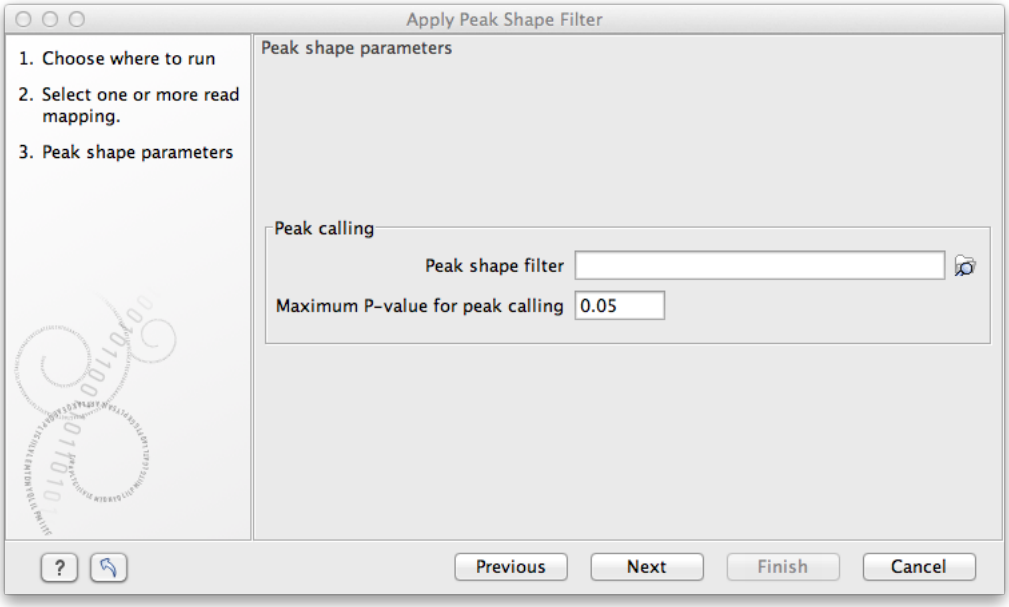

Figure 1.5: *Options for Apply Peak Shape Filter.*

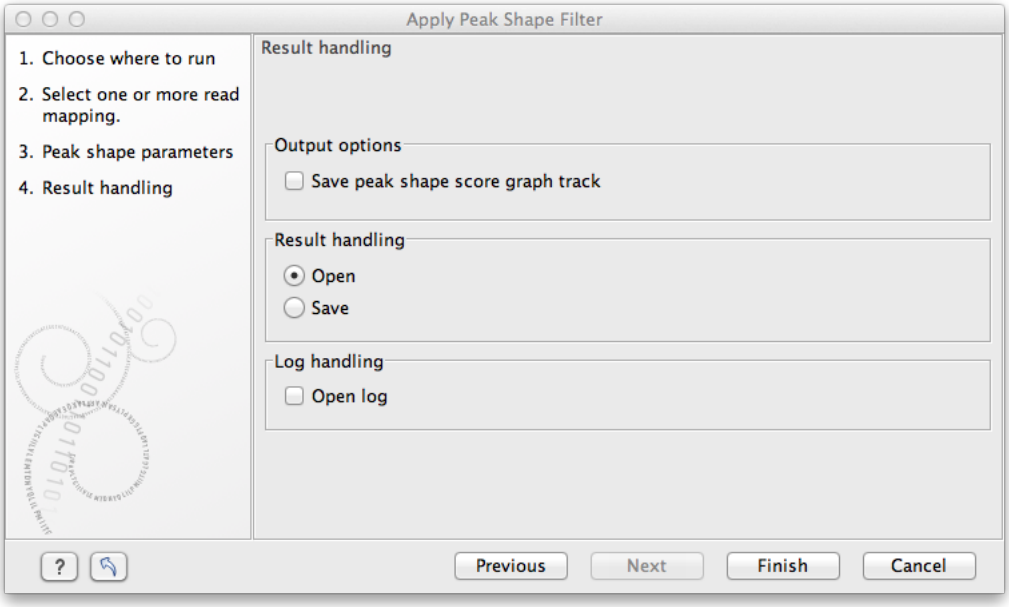

Figure 1.6: *Output options for Apply Peak Shape Filter.*

the peak shape score for each genomic position. To save disk space, only scores greater than zero are reported. For the definition of peak shape score.

Choose whether you want to open the results directly, or save the results in the **Navigation Area**. If you choose to save the results, you will be asked to specify where you would like to save them.

For more information on the **Peak track** () (see the CLC Genomics Workbench manual `http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Peak_track.html`).

1.3 Running the Score Regions tool

The Score Regions tool allows you to apply a new peak shape filter to score genomic regions according how they match a given peak shape. To run the Score Regions plugin:

> **Toolbox** | **Epigenomics Analysis** ( ) | **Advanced Peak Shape Tools** ( ) | **Score Regions** ( )

This will open up the wizard shown in figure 1.7 where you can select the input data (e.g. mapped ChIP-Seq reads). Multiple inputs are accepted, provided that they refer to the same genome. Track based read mappings ( ) and stand-alone read mappings ( ) / ( ) are both accepted.

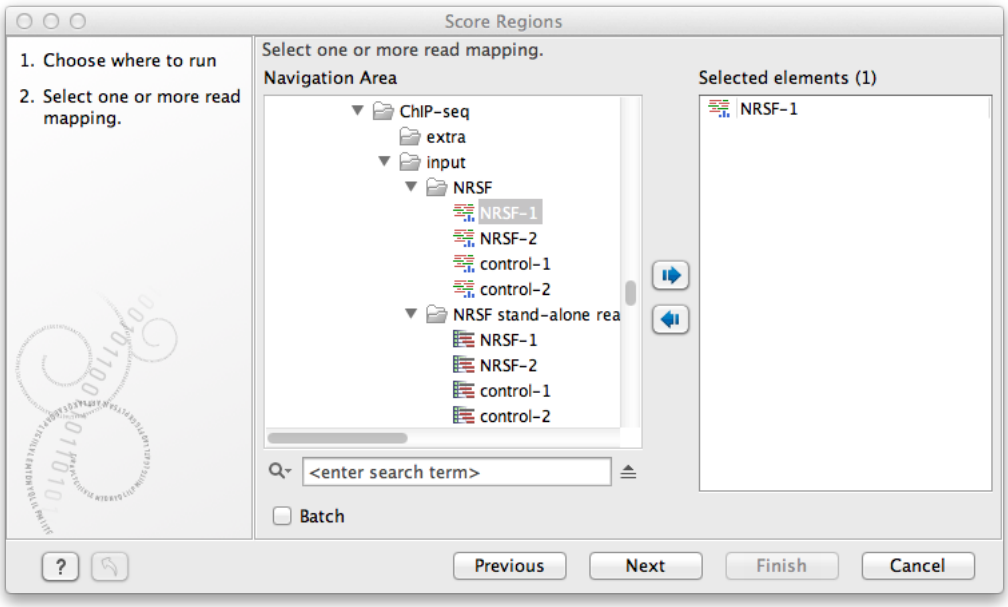

Figure 1.7: *Select the input data for Score Regions.*

Click on the button labeled **Next** to go to the next wizard step (shown in figure 1.8).

In this wizard step you have the following options:

- **Peak shape filter** The peak shape filter ( ) to apply to the data. Peak shape filters can be obtained as the result of the ChIP-Seq Analysis tool.
- **Regions to score** An annotation track ( ) containing the regions where the peak shape will be applied. The peak shape filter will be applied to every genomic position within the interval and the maximum values will be used to score the region.

The result of the algorithm will be an annotation track ( ) of the same type as the regions to score annotation track, where the columns of type **Peak shape score**, **P-value** and **Center of peak** will be added or replaced.

After clicking on the button labeled **Next**, you can choose whether you want to open the result directly, or save the results in the **Navigation Area**. If you choose to save the results, you will be asked to specify where you would like to save them.

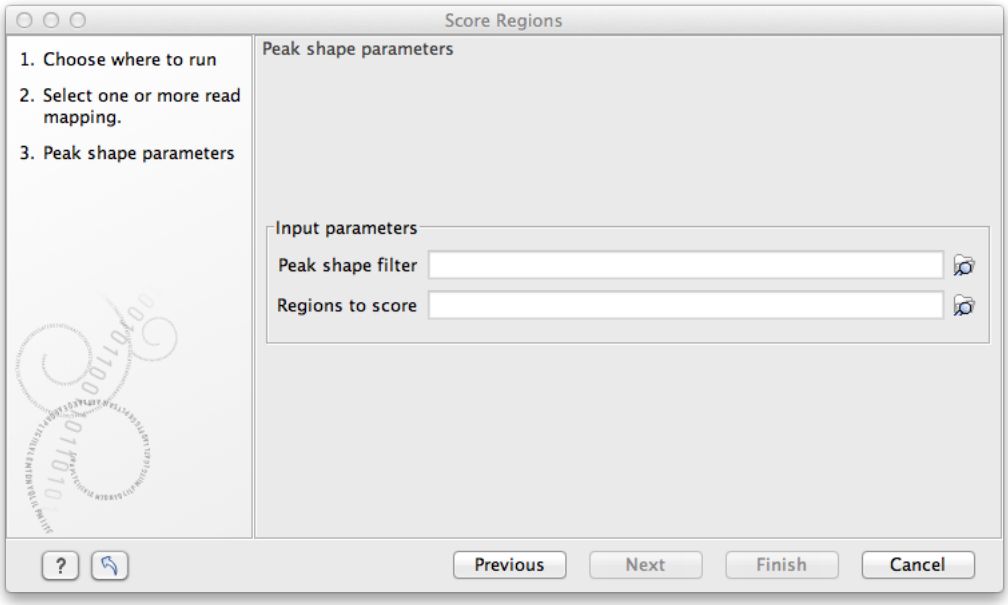

Figure 1.8: *Options for Score Regions.*

Chapter 2

Installation of the Advanced Peak Shape Tools

Advanced Peak Shape Tools is installed as a plugin. Plugins are installed using the plugin manager. In order to install plugins on Windows, the Workbench must be run in administrator mode: Right-click the program shortcut and choose "Run as Administrator". Then follow the procedure described below.

Help in the Menu Bar | Plugins... ( )

or **Plugins ( ) in the Toolbar**

The plugin manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins that are installed.
- **Download Plugins.** This is an overview of available plugins on CLC bio's server.

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 2.1).

Clicking a plugin will display additional information at the right side of the dialog. This will also display a button: **Download and Install**.

Click the Advanced Peak Shape Tools and press **Download and Install**. A dialog displaying progress is now shown, and the plugin is downloaded and installed.

If the Advanced Peak Shape Tools is not shown on the server, and you have it on your computer (for example if you have downloaded it from our website), you can install it by clicking the **Install from File** button at the bottom of the dialog. This will open a dialog where you can browse for the plugin. The plugin file should be a file of the type ".cpa".

When you close the dialog, you will be asked whether you wish to restart the CLC Workbench. The plugin will not be ready for use until you have restarted.

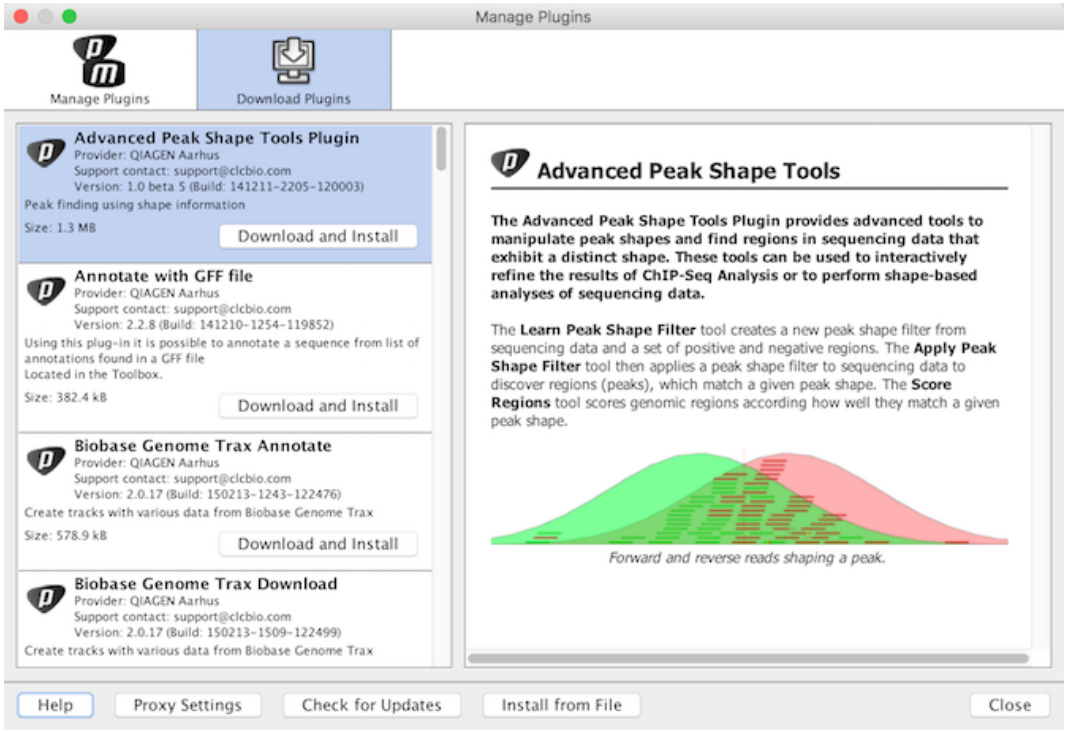

Figure 2.1: *The plugins that are available for download.*

Chapter 3

Uninstall

Plugins are uninstalled using the plugin manager:

Help in the Menu Bar | **Plugins... (** **)**

or **Plugins (** **) in the Toolbar**

This will open the dialog shown in figure 3.1.

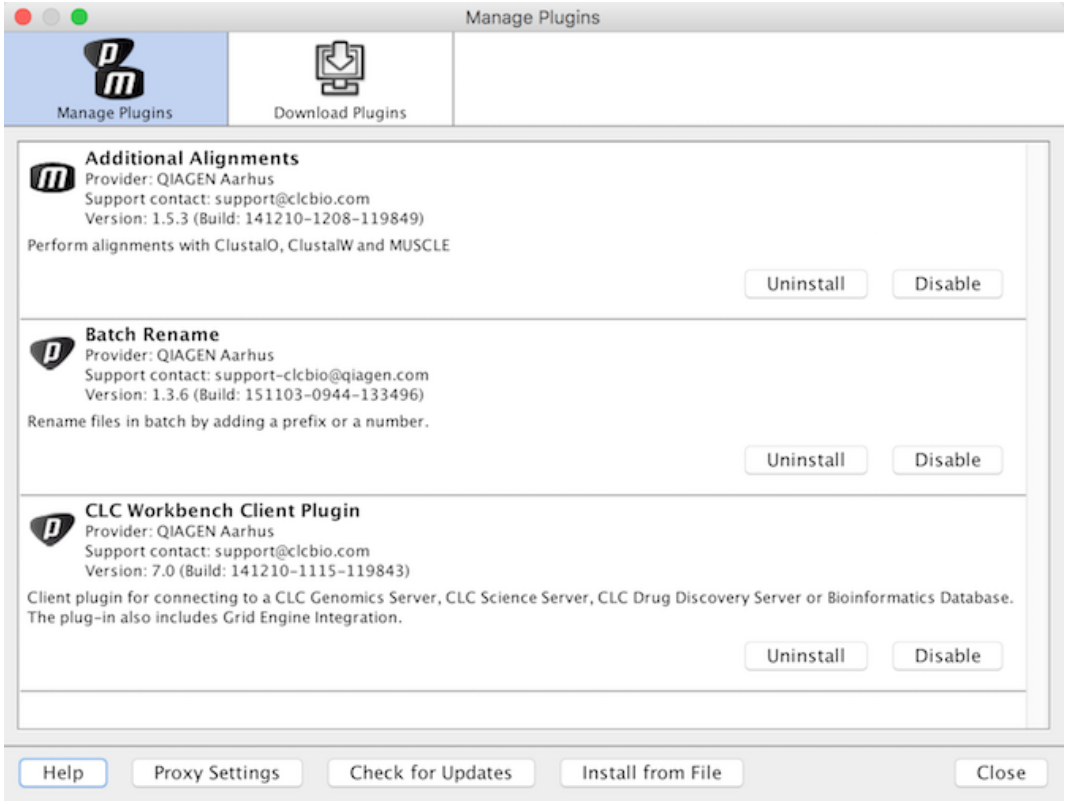

Figure 3.1: *The plugin manager with plugins installed.*

The installed plugins are shown in this dialog. To uninstall:

Click the Advanced Peak Shape Tools | **Uninstall**

If you do not wish to completely uninstall the plugin but you don't want it to be used next time you start the Workbench, click the **Disable** button.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be uninstalled until the workbench is restarted.

Index