ERCC_Analysis Plugin
For use with: ERCC RNA Spike-In Control Mixes
Publication Part Number: 4479068 Rev. C

SUBJECT: ERCC Analysis

The ERCC_Analysis plugin is intended to help with ERCC RNA Spike-in Controls. It enables you to quickly determine whether or not the ERCC results indicate a problem with library preparation or the PGM run. This bulletin provides the following information:

■ What are the ERCC Spike-in Controls? ........................................................................... 1
■ Requirements for the ERCC_Analysis plugin ................................................................. 2
■ Enable the ERCC_Analysis Plugin .................................................................................. 2
■ Manually run the ERCC_Analysis plugin ......................................................................... 3
■ View the analysis results .................................................................................................. 6
■ Interpret the data .............................................................................................................. 7

Appendix A Definitions ........................................................................................................ 8
Appendix B Parameters used for tmap .................................................................................. 8

What are the ERCC Spike-in Controls?

The External RNA Controls Consortium (ERCC) is an ad-hoc group of academic, private, and public organizations. The National Institute of Standards and Technology (NIST)-hosted ERCC has been working to develop a common set of RNA controls that can be used in multiple gene expression platforms such as quantitative RT-PCR, microarrays, and next-generation sequencing (NGS) technologies.

The outcome of the ERCC effort is a reference library of NIST-certified DNA plasmids that are designed to produce a set of transcripts 250–2000 nt in length that mimic natural eukaryotic mRNAs.

The ERCC RNA Spike-In Control Mixes are pre-formulated sets of 92 polyadenylated transcripts from the ERCC plasmid reference library. The transcripts are traceable through the manufacturing process to the NIST plasmid reference material.

For more information on ERCC RNA Spike-In Control Mixes, please refer to the ERCC RNA Spike-In Control Mixes User Guide (Pub. no. 4455352).
Requirements for the ERCC_Analysis plugin

<table>
<thead>
<tr>
<th>Tool</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td>Torrent Suite™ v3.4.2 or higher</td>
</tr>
<tr>
<td></td>
<td>IMPORTANT! To ensure proper trimming of the sequences prior to alignment with a reference, for all libraries prepared with the Ion Total RNA-seq Kit v2, you must select an option from the barcode pull-down menu.</td>
</tr>
<tr>
<td>Web browser</td>
<td>• Chrome™ • Opera® • Firefox® • Safari® • Internet Explorer® [Version 7+]</td>
</tr>
</tbody>
</table>

Enable the ERCC_Analysis Plugin

To enable the ERCC_Analysis plugin:

1. Select Plugins in the pull-down menu after clicking the gear icon in the upper right corner of the Torrent Browser. Scroll to the ERCC_Analysis plugin.
2. Select Enabled to manually run the plugin. If you want every run to automatically call the ERCC_Analysis plugin, select Auto-Run. You can select both Auto-Run and Enabled.
Manually run the ERCC_Analysis plugin

To manually run the ERCC_Analysis plugin:

1. Click the DATA tab in the Torrent Browser.

2. In the Rep Report Name column, scroll to the run results you want to view and select the run. In the Report column, select a run report.

   **Note:** To open the report in a new browser window, click the reports icon .

3. Scroll to the bottom of the report to **Plugin Summary**, then click **Select Plugins To Run** to open the Plugin List.

   If you do not see the ERCC_Analysis plugin listed, refer to the previous section “Enable the ERCC_Analysis Plugin” on page 2.

4. In the Plugin List window, select **ERCC_Analysis**.
ERCC Analysis

Manually run the ERCC_Analysis plugin

A popup window appears showing the values for minimum acceptable R-squared, minimum acceptable counts per transcript, ERRC pool used, and the barcode used (if any). For a description of what each of these means, see the instructions in “Configure ERCC_Analysis (Optional)

Minimum acceptable R-squared value: .85
The current default minimum R-squared is .85.

Minimum acceptable counts per ERCC transcript: 3
The current default minimum counts per transcript is 3.

ERCC Pool used (1 or 2): 1
Barcode of interest: 1p003_006

Enter the barcode EXACTLY as it appears at the top of this report, under “Barcode name” (you may want to copy and paste). Enter one barcode only, to see multiple barcodes you will need to run plugin more than once. Leave blank if not using barcodes.

To change the default values, go to the Config tab of the Torrent Suite.

5. Click Submit to start the run. You can check the status of the run by clicking Report or Log.

   Note: Enter the barcode exactly as it appears in the Barcode Name column at the top left of the report, or copy and paste it in. If multiple barcodes were used in the run, the plugin will need to be run for each barcode.

For analysis runs with total reads under 1,000,000, the plugin will normally take 2-3 minutes to run on the hardware described on the Torrent Server product page. For larger runs, the plugin will approximately take an additional 1-2 minutes per million total reads (for example, a run with 5 million reads may take 10-15 minutes). These run times are offered only as a guideline. If your Torrent Server is busy with other processing, run times will be longer.

After the ERCC analysis is completed, you can view the analysis results.

Configure ERCC_Analysis (Optional)

You can optionally change the R-squared value to set a default value for the summary report screen:

1. If the window showing the minimum acceptable R-squared value is open, close it. Then scroll to the top of the Torrent Browser and click the gear icon at the upper right corner, and select Plugins from the pull-down menu.

2. In the Plugins section and ERCC_Analysis row, click the corresponding gear icon under the Manage heading, and select Configure from the pull-down menu.

3. In the ERCC Plugin Configuration window, enter a value between 0 and 1 as your minimum acceptable R-squared value (a lower value is indicated by a red light in the summary report). You can override this value on a per-run basis at the time of submission.

4. Enter a value for minimum acceptable counts per transcript. This value determines which transcripts are used in calculating R-squared. For example, if this is set to 3, then any transcript for which there are only 1 or 2 counts, will not be considered in the R-squared calculation.
5. Enter the ERCC Pool used (either 1 or 2). Most users will want to set the default to 1, but may use Pool 2 for some runs (and can indicate this at run time rather than changing the plugin configuration).

6. Enter a barcode value. For most users the default barcode value will probably be blank (no barcode), but a barcode may be entered at run time.

7. Click **Submit**.

8. Scroll to the top of the Torrent Browser and select the **DATA** tab.

9. In the Rep Report Name column, scroll to the run results you want to analyze, and select the link.

10. Scroll to the bottom of the report to Plugin Summary, then click the button **Select Plugins To Run** to open the Plugin List.
    
    If you do not see the ERCC_Analysis plugin listed, refer to “Enable the ERCC_Analysis Plugin” on page 2.

11. In the Plugin List window, select **ERCC_Analysis**.

12. In the popup window, click **Submit** to start the run.
View the analysis results

After the status of the ERCC plugin has changed to “Completed,” click ERCC_Analysis.html in the Plugin Summary section to open the ERCC Report and view analysis results:

![ERCC Report](image)

View error information

If the ERCC_Analysis plugin status changes to “Error” after you click ERCC_Analysis.html, then something went wrong during the running of the plugin. In this case, look at the error log:

1. Return to the Plugin Summary, then click the log file icon to see the error log:

![Log for detailedReport](image)

2. Scroll to the bottom of the system log file to view information about what went wrong with the analysis.
An Error status for the ERCC_Analysis plugin should be a rare event. An error does not indicate that the ERCC_Analysis results are bad, but rather that the ERCC_Analysis plugin itself failed to run.

**Interpret the data**

The ERCC Report screen (shown on page 6) displays the ERCC Dose Response plot. The points are color-coded, based on mapping quality. There is also a trendline, based on the parameters shown in tabular form to the right of the graph.

The y-axis of the plot is the log (base 2) of the raw counts found for the transcript in question. The x-axis is also logarithmic, but represents the known relative concentration of the ERCC transcripts. Ideally, the points all fall on a straight line.

More realistically, in the good case, the raw counts and relative concentration should at least correlate with a high R-squared (for example, 0.9 or higher). The table to the right of the plot (shown on page 6) shows the R-squared value found for this plot, as well as the Slope, Y-intercept, and N (number of transcripts found) values. Although there are 92 transcripts in the ERCC mix, it is not expected that all 92 will be detected. The number of transcripts detected depends on the sequencing depth. In addition, any transcripts for which the number of counts is less than the “minimum acceptable counts per ERCC transcript”, will not be used in the R-squared calculation. The graph will show a dotted line at this minimum counts value, to highlight which points on the graph (i.e. which transcripts) have been excluded from the R-squared calculation. For more information on ERCC analysis, refer to the following resources:

- Figure 2, Analysis of ERCC read counts, in *Sensitivity of RNA-Seq using Ion semiconductor sequencing: a comparison to microarrays and qPCR*.
- The Ion Torrent white paper, *Methods, tools, and pipelines for analysis of Ion RNA-Seq data*.
- The information on the ERCC ExFold RNA Spike-In Mix product page.

**View transcript details**

If you want to look at the details regarding a particular transcript, there are two methods you can use:

- Hover your mouse-cursor over a point on the ERCC Dose Response plot to display a popup window that shows details about that transcript (the name, reads, and coverage plots).

If several points are very close together on the plot and it is difficult to hover over the point you are interested in, you can zoom in on the plot to more easily distinguish points:

  a. Use your mouse to draw a box around the point of interest and magnify it.
  b. To zoom out to the full view of the ERCC Dose Response plot, either double-click the plot, or click the button labeled **Reset Zoom**.

or

- Scroll to the particular transcript, and click the [+] next to the transcript name. This method shows the same information, plus a few additional pieces. See Appendix A, “Definitions” if the meaning of any of these pieces is unclear.
Appendix A  Definitions

<table>
<thead>
<tr>
<th>Coverage Depth</th>
<th>The minimum and maximum number of reads covering bases in the transcript. If coverage is 100%, the minimum value will be &gt; 0.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>The number of base positions covered by at least one read.</td>
</tr>
<tr>
<td>Start Sites</td>
<td>The number of base positions that are the start site for a read.</td>
</tr>
<tr>
<td>Unique Start Sites</td>
<td>The number of start sites that have only one read starting at the site.</td>
</tr>
<tr>
<td>Coverage CV</td>
<td>Coefficient of Variation for coverage = average coverage / stddev coverage for the entire transcript.</td>
</tr>
</tbody>
</table>

Appendix B  Parameters used for tmap

For information about how the ERCC_Analysis plugin is working, use the tmap call:

```
tmap mapall -f $REFERENCE -r ${TSP_FILEPATH_FASTQ} -a 1 -g 0 -n 8 stage1 map1 --seed-length 18 stage2 map2 map3 --seed-length 18
```

$REFERENCE is a constant that refers to the reference file location.

${TSP_FILEPATH_FASTQ} similarly refers to the *.fastq file location.

At the time of publication of this bulletin, it is not possible to change these parameters when you are using the ERCC_Analysis plugin.