

# ERCC Assessment

The ERCC (External RNA Control Consortium) developed a set of RNA standards for quality control in microarray, qPCR, and sequencing applications. These RNA standards are spiked-in RNA with known concentrations and composition (i.e. sequence length and GC content). They can be used to evaluate sensitivity and accuracy of RNA-seq data.

The ERCC analysis is performed on unaligned data, if the ERCC RNA standards have been added to the samples. There are 92 ERCC spiked-in sequences with different concentrations and different compositions. The idea is that the raw data will be aligned (with Bowtie) to the known ERCC-RNA sequences to get the count of each ERCC sequence. This information is available within Partek Flow and will be used to plot the correlation between the observed counts and the expected concentration. If there is a high correlation between the observed counts versus the expected concentration, you can be confident that the quantified RNA-seq data are reliable. Partek Flow supports Mix1 and Mix 2 ERCC formulations. Both formulations use the same ERCC sequences, but each sequence is present at different expected concentrations. If both Mix 1 and Mix 2 formulations have been used, ExFold comparison can be performed to compare the observed and expected Mix1:Mix2 ratio for each spike-in.

To start ERCC assessment, select an unaligned reads node and choose **ERCC** in the context sensitive menu. If all samples in the project have used the Mix 1 or Mix 2 formulation, choose the appropriate radio button at the top (Figure 1).

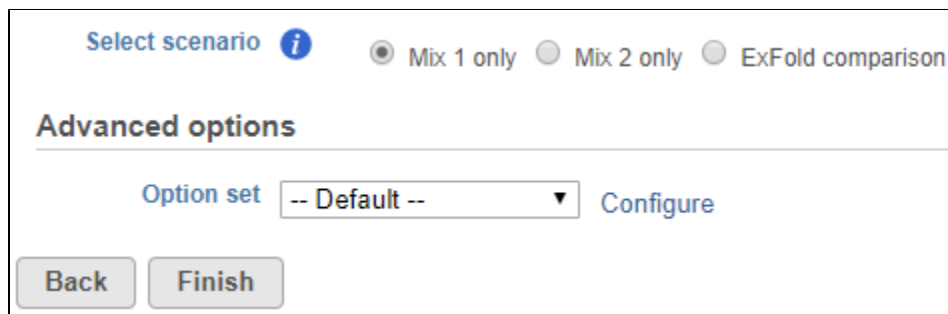
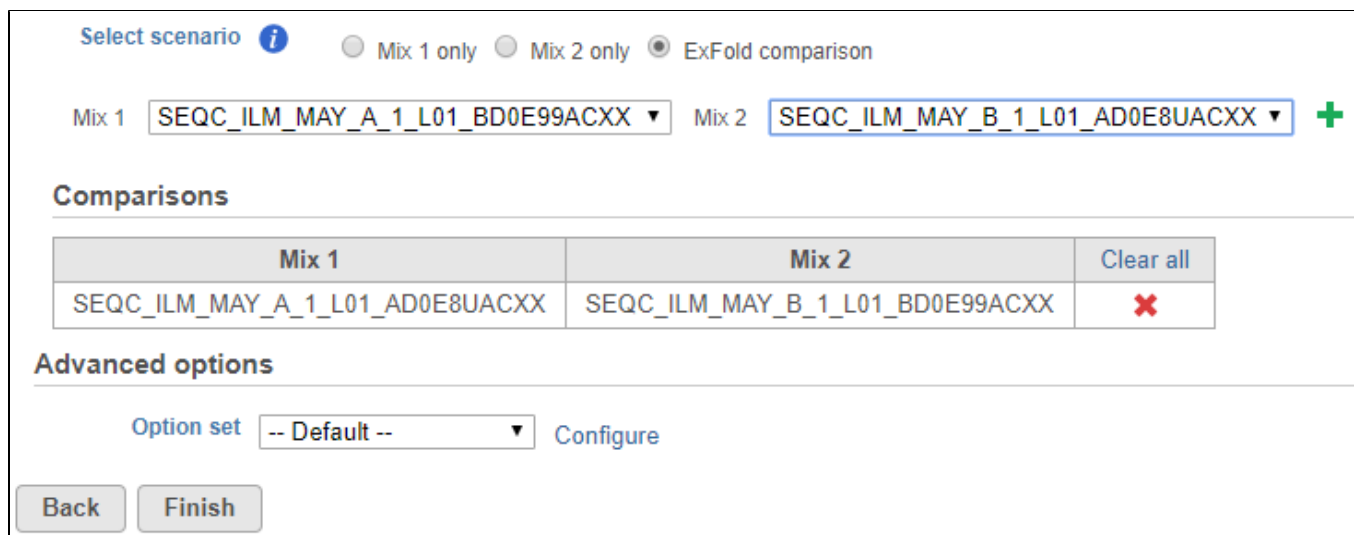


Figure 1. Setting up advanced options for the alignment of ERCC controls using Bowtie

If some samples have been treated with the Mix 1 formulation and others have been treated with the Mix 2 formulation, choose the **ExFold comparison** radio button (Figure 2). Set up the pairwise comparisons by choosing the Mix 1 and Mix 2 samples that you wish to compare from the drop-down lists, followed by the green plus (+) icon. The selected pair of samples will be added to the table below.



Mix 1	Mix 2	Clear all
SEQC_ILM_MAY_A_1_L01_AD0E8UACXX	SEQC_ILM_MAY_B_1_L01_BD0E99ACXX	✗

Figure 2. ExFold comparison can be performed between specified Mix 1 and Mix 2 pairs of samples

You can change the Bowtie parameters by clicking **Configure** before the alignment (Figure 1), although the default parameters work fine for most data. Once the task has been set up correctly, select **Finish**.

ERCC task report starts with a table (Figure 3), which summarizes the result on the project level. The table shows which samples use the Mix 1 or Mix 2 formulation. The total number of alignments to the ERCC controls are also shown, which is further divided into the total number of alignments to the forward strand and the reverse strand. The summary table also gives the percentage of ERCC controls that contain alignment counts (i.e. are present). Generally, the fraction of present controls should be as high as possible, however, there are certain ERCC controls that may not contain alignment counts due to their low concentration; that information is useful for evaluation of the sensitivity of the RNA-seq experiment. The coefficient of determination (R squared) of the present ERCC controls is listed in the next column. As a rule of a thumb, you should expect a good correlation between the observed alignment counts and the actual concentration, or else the RNA-seq quantification results may not be accurate. Finally, the last two columns give estimates of bias with regards to sequence length and GC content, by giving the correlation of the alignment counts with the sequence length and the GC content, respectively.

Optional columns								
Sample name ↕	Mix ↕	Alignments to controls ↕	Forward strand ↕	Reverse strand ↕	% controls present ↕	R squared ↕	Sequence length correlation ↕	% GC correlation ↕
SEQC_ILM_MAY_A_1_L01_AD0E8UACXX	Mix 1	102,902	51,451	51,451	66.30% (61 out of 92)	0.94	0.05	-0.07
SEQC_ILM_MAY_A_1_L01_BD0E99ACXX	Mix 1	107,746	53,873	53,873	66.30% (61 out of 92)	0.95	0.08	-0.06
SEQC_ILM_MAY_B_1_L01_AD0E8UACXX	Mix 2	276,020	138,010	138,010	75.00% (69 out of 92)	0.96	0.08	-0.11
SEQC_ILM_MAY_B_1_L01_BD0E99ACXX	Mix 2	260,464	130,232	130,232	73.91% (68 out of 92)	0.93	0.06	-0.06

Rows per page 25 |< << (1 of 1) >> >>|

Figure 3. Summary of ERCC assessment. Each row is a sample (an example is shown)

If ExFold comparison was enabled, an extra table will be produced in the ERCC task report (Figure 4). Each row in the table is a pairwise comparison. This table lists the percentage of ERCC controls present in the Mix 1 and Mix 2 samples and the R squared for the observed vs expected Mix1:Mix2 ratios.

Comparison ↕	% Mix 1 controls present ↕	% Mix 2 controls present ↕	R squared ↕
SEQC_ILM_MAY_A_1_L01_AD0E8UACXX - SEQC_ILM_MAY_B_1_L01_BD0E99ACXX	66.30% (61 out of 92)	73.91% (68 out of 92)	0.89
SEQC_ILM_MAY_A_1_L01_BD0E99ACXX - SEQC_ILM_MAY_B_1_L01_AD0E8UACXX	66.30% (61 out of 92)	75.00% (69 out of 92)	0.91

Rows per page 25 |< << (1 of 1) >> >>|

Figure 4. Summary of ExFold comparison. Each row is a different pairwise comparison

The ERCC spike-ins plot (Figure 5) shows the regression lines between the actual spike-in concentration (x-axis, given in log<sub>2</sub> space) and the observed alignment counts (y-axis, given in log<sub>2</sub> space), for all the samples in the project. The samples are depicted as lines, and the probes with the highest and lowest concentration are highlighted as dots. The regression line for a particular sample can be turned off by simply clicking on the **sample name** in the legend beneath the plot.

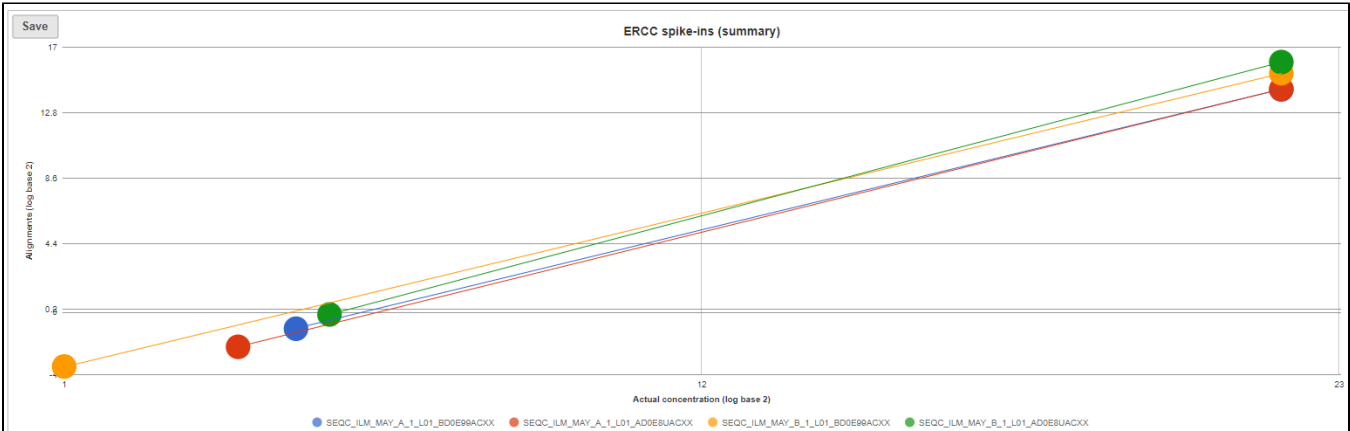


Figure 5. ERCC spike-ins plot. Lines (one per sample) correspond to regression lines between actual spike-in concentrations and observed number of alignment

Optionally, you can invoke a principal components analysis plot (**View PCA**), which is based on RPKM-normalised counts, using the ERCC sequences as the annotation file (not shown).

For more details, go to the sample-level report (Figure 6) by selecting a **sample name** on the summary table. First, you will get a comprehensive scatter plot of observed alignment counts (y-axis, in log<sub>2</sub> space) vs. the actual spike-in concentration (x-axis, in log<sub>2</sub> space). Each dot on the plot represents an ERCC sequence, coloured based on GC content and sized by sequence length (plot controls are on the right).

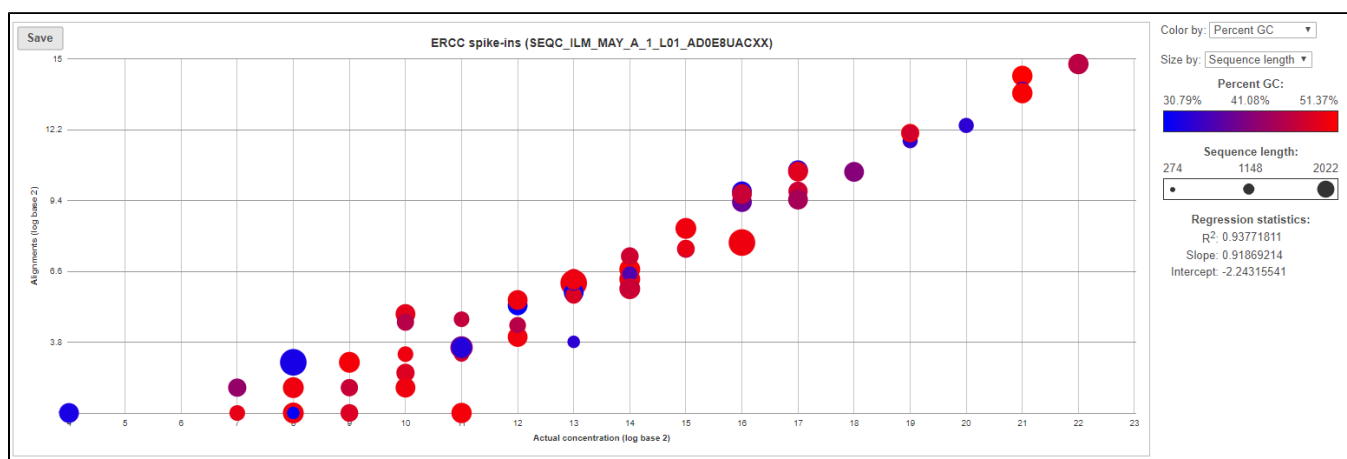


Figure 6. Scatter plot of actual observed alignment counts vs. probe concentration for each ERCC control within a sample. Each dot is an ERCC control, color-coded by percent GC content and sized by sequence length.

The table (Figure 7) lists individual controls, with their actual concentration, alignment counts, sequence length, and % GC content. The table can be downloaded to the local computer by selecting the **Download** link.

Control	Actual (Concentration)	Alignments	Sequence length	Percent GC
ERCC-00075	2	0	1,023	36.27%
ERCC-00057	2	0	1,021	49.76%
ERCC-00048	2	0	992	47.68%
ERCC-00083	4	0	1,022	34.54%
ERCC-00117	8	2.00	1,136	51.14%
ERCC-00098	8	0	1,143	51.44%
ERCC-00061	8	0	1,136	50.09%
ERCC-00017	16	0	1,136	51.41%
ERCC-00138	16	0	1,024	32.91%
ERCC-00012	16	0	994	50.91%
ERCC-00086	16	0	1,020	32.25%
ERCC-00123	32	0	1,022	35.62%
ERCC-00104	32	2.00	2,022	32.74%
ERCC-00142	32	0	493	49.90%
ERCC-00024	32	0	536	46.27%
ERCC-00016	32	0	844	48.46%
ERCC-00041	32	0	1,122	45.37%
ERCC-00081	32	2.00	534	48.69%
ERCC-00097	64	0	523	36.14%
ERCC-00156	64	0	494	48.99%
ERCC-00158	64	0	1,027	34.27%
ERCC-00164	64	0	1,022	37.08%
ERCC-00168	64	0	1,024	34.18%
ERCC-00147	128	6.00	1,023	36.46%
ERCC-00109	128	0	536	45.52%

Rows per page: 25
(1 of 4)
Download

Figure 7. Table report for ERCC controls within a sample. The default sort order is by column Control; the example in the figure is sorted by Actual (Concentration) to highlight the relationship between the control concentration and number of alignments

For more details on ExFold comparisons, select a **comparison name** in the ExFold summary table (Figure 8). First, you will get a comprehensive scatter plot of observed Mix1:Mix2 ratios (y-axis, in log2 space) vs. the expected Mix1:Mix2 ratio (x-axis, in log2 space). Each dot on the plot represents an ERCC sequence, coloured based on GC content and sized by sequence length (plot controls are on the right).

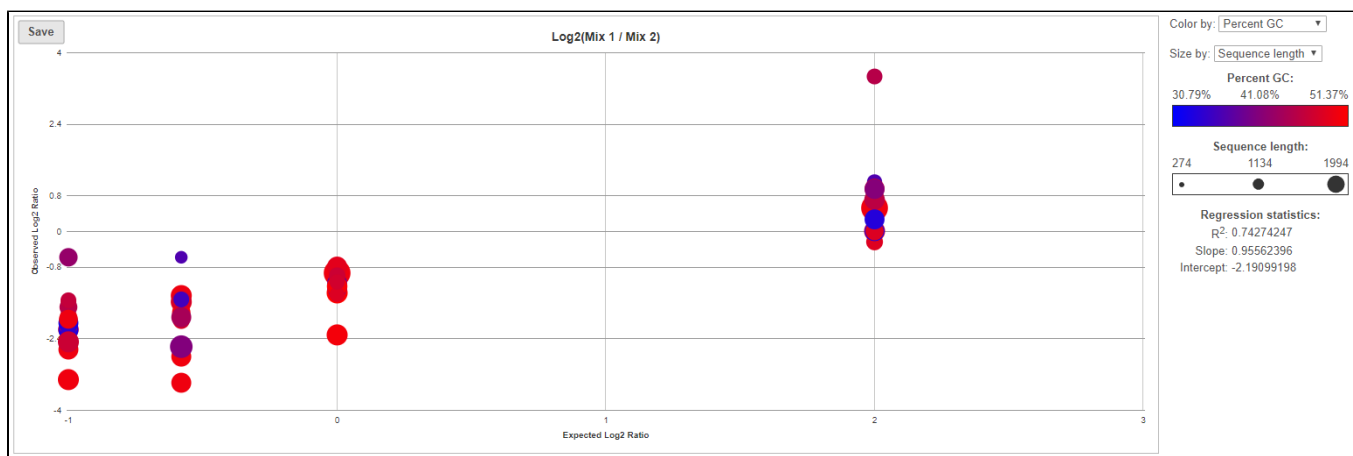


Figure 8. Scatter plot of actual observed Mix1:Mix2 ratios vs. expected Mix1: Mix2 ratios for each ERCC control within a sample. Each dot is an ERCC control, coloured by GC content and sized by concentration

The table (Figure 9) lists individual controls, with each samples' alignment counts, together with the observed and expected Mix1:Mix2 ratios. The table can be downloaded to the local computer by selecting the **Download** link.

ERCC ID	Alignments Mix 1 (SEQC_ILM_MAY_A_1_L01_AD0E8UACXX)	Alignments Mix 2 (SEQC_ILM_MAY_B_1_L01_BD0E99ACXX)	Observed Mix 1 : Mix 2 ratio	Expected Mix 1 : Mix 2 ratio
ERCC-00002	20,262	93,072	0.22	0.50
ERCC-00003	1,532	7,096	0.22	0.50
ERCC-00004	5,222	3,308	1.58	4.00
ERCC-00009	862	2,194	0.39	1.00
ERCC-00012	0	4	0	0.67
ERCC-00013	4	6	0.67	0.50
ERCC-00014	0	4	0	0.50
ERCC-00016	0	0	?	0.67
ERCC-00017	0	0	?	4.00
ERCC-00019	22	26	0.85	4.00
ERCC-00022	178	784	0.23	0.50
ERCC-00024	0	2	0	0.67
ERCC-00025	70	134	0.52	1.00
ERCC-00028	8	8	1.00	4.00
ERCC-00031	0	10	0	1.00
ERCC-00033	8	0	?	4.00
ERCC-00034	30	52	0.58	1.00
ERCC-00035	102	266	0.38	1.00
ERCC-00039	0	22	0	0.67
ERCC-00040	0	0	?	0.67
ERCC-00041	0	0	?	0.50
ERCC-00042	642	1,438	0.45	1.00
ERCC-00043	860	3,572	0.24	0.50
ERCC-00044	78	212	0.37	0.67
ERCC-00046	3,442	16,176	0.21	0.50

Figure 9. Table report for ERCC ExFold comparison within a sample

## Additional Assistance

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.



Your Rating:      Results:      43 rates