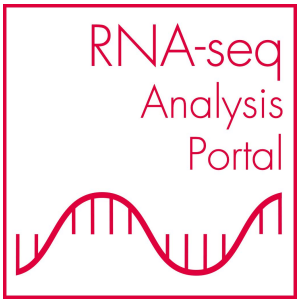


RNA-seq Analysis Portal

USER MANUAL

Manual for

QIAGEN RNA-seq Analysis Portal 5.1

August 20, 2024

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Introduction

This manual is for *QIAGEN RNA-seq Analysis Portal 5.1*.

General information

Technical assistance

If you have questions about or difficulties using QIAGEN *RNA-seq Analysis Portal*, please contact us by filling in the web-based contact form at `https://www.qiagen.com/service-and-support/technical-support/technical-support-form/`.

We also appreciate your feedback to help us improve our products and documentation.

Policy statement

It is QIAGEN's policy to improve products as new techniques and components become available. QIAGEN reserves the right to change the specifications of products at any time.

How to cite RNA-seq Analysis Portal

Please use the following when citing this product:

QIAGEN RNA-seq Analysis Portal 5.1 (QIAGEN, Aarhus, Denmark), `https://rnaportal.qiagen.com`

Latest improvements

Latest improvements are available at `https://digitalinsights.qiagen.com/qiagen-rna-seq-analysis-portal-latest-improvements/`

Intended use statement

QIAGEN RNA-seq Analysis Portal is for research use only.

General Description

RNA-seq Analysis Portal is a cloud-based platform for analyzing your RNA-seq expression data, accessible from a web browser. In just a few steps, the predefined analysis pipelines let you calculate expression levels, identify differential expression, as well as identify significant pathways, regulators, diseases and functions.

The graphical views and interactive plots enable seamless exploration and refinement of your RNA-seq analysis results. *RNA-seq Analysis Portal* is available to user with a My QIAGEN account and to users with a QIAGEN Ingenuity Pathway Analysis (IPA) account. My QIAGEN users can pass results to QIAGEN GeneGlobe Design and Analysis Hub for exploration and planning of next steps, for example qPCR or dPCR-based follow up experiments for your filtered target genes. QIAGEN IPA users can upload results to their QIAGEN IPA account for detailed downstream analysis.

RNA-seq Analysis Portal is powered by QIAGEN CLC Genomics and QIAGEN Ingenuity Pathway Analysis.

Deployment

RNA-seq Analysis Portal is deployed on Amazon Web Services, US East (Northern Virginia) Region.

GeneGlobe Design and Analysis Hub is deployed on Microsoft Azure, West Europe (Netherlands).

Ingenuity Pathway Analysis is deployed on a QIAGEN Secure Private Cloud, US (Virginia).

All external access to *RNA-seq Analysis Portal* is SSL encrypted using the TLSv1.2_2019 policy specified here: `https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/secure-connections-supported-viewer-protocols-ciphers.html#secure-connections-supported-ciphers`

All communication between *RNA-seq Analysis Portal* and *GeneGlobe Design and Analysis Hub* or *Ingenuity Pathway Analysis* is SSL encrypted.

Supported sample kits and sequencing platforms

RNA-seq Analysis Portal supports analysis of FASTQ data generated on Illumina® instruments using the following sample kits:

QIAGEN

- QIAseq® Stranded mRNA Select Kit
- QIAseq® Stranded Total RNA Lib Kit
- QIAseq® miRNA Library Kit
- QIAseq® UPX 3' Transcriptome Kit
- QIAseq® FastSelect RNA Lib Kit
- QIAseq® UPXome RNA Lib Kit

Illumina

- TruSeq® Stranded Total RNA Library Prep (Human/Rat, Gold, Globin)
- Illumina® Stranded Total RNA Prep with Ribo-Zero Plus

New England Biolabs

- NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina

Roche Sequencing solutions

- KAPA® RNA HyperPrep Kit

Takara Bio

- SMARTer® Stranded Total RNA Sample Prep Kit - HI Mammalian
- SMARTer® Stranded Total RNA Sample Prep Kit - Low Input Mammalian

Thermo Fisher Scientific

- Collibri™ Stranded RNA Library Prep Kit for Illumina Systems

RNA-seq Analysis Portal supports analysis of FASTQ and BAM data generated on Thermo Fisher® Ion Torrent® instruments using the following kit:

QIAGEN

- QIAseq® miRNA Library Kit

Data is analyzed using dedicated analysis workflows. Details specific to a given analysis workflow are provided in later sections. The main steps in an expression analysis are:

- Reads are aligned and counted to produce expression values for the individual samples (Align and count).
- Expression levels between groups of samples are compared in an experiment, producing one or more differential expression outputs (Create experiment).

Accessing RNA-seq Analysis Portal

RNA-seq Analysis Portal is accessed using a web browser and is available for users with either a My QIAGEN user account or a QIAGEN Ingenuity Pathway Analysis (IPA) user account as detailed in the following sections.

Supported browsers

RNA-seq Analysis Portal is supported on recent versions of Chrome®, Firefox®, and Safari®. The *RNA-seq Analysis Portal* interface is optimized for a minimum screen width of 1024 pixels, but a larger resolution should be preferred.

Login with My QIAGEN account

With a My QIAGEN user account you can access *RNA-seq Analysis Portal* either via the "QIAGEN GeneGlobe Design and Analysis Hub" or directly via bookmarked URL or link:

- **QIAGEN GeneGlobe Data Analysis Center:** `https://geneglobe.qiagen.com/analyze`. How to log in is described below.
- **Direct link:** `https://rnaportal.qiagen.com`. Use your My QIAGEN account credentials to log in.

Logging in from the QIAGEN GeneGlobe Data Analysis Center

To log into *RNA-seq Analysis Portal* from GeneGlobe Data Analysis Center (figure 1):

1. Select one of the supported sample kits, or click the button **Non-QIAGEN Kits** and select the third-party kit provider.
2. Click **Start your analysis**. This opens *RNA-seq Analysis Portal* in a new browser tab.

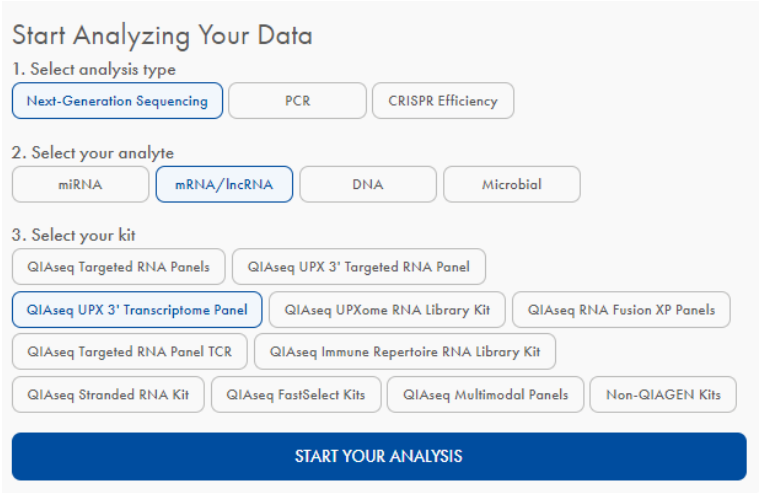


Figure 1: Accessing RNA-seq Analysis Portal from the QIAGEN GeneGlobe Data Analysis Center.

Login with QIAGEN IPA account

With a QIAGEN IPA account you can access *RNA-seq Analysis Portal* either via QIAGEN IPA or directly via bookmarked URL or link:

- **QIAGEN IPA:** In the QIAGEN IPA application, click on the **Go to secondary analysis** button.
- **Direct link:** `https://rnaportal.qiagen.com/rnaportalui/#/uap/`. Use your QIAGEN IPA account credentials to log in.

Analysis credits for RNA-seq Analysis Portal

To run the initial Align and count analysis in *RNA-seq Analysis Portal* you will need analysis credits. Instructions on how to add analysis credits to your account are described in the following sections. The process differs depending on whether you are logged in with a My QIAGEN or a QIAGEN IPA account.

Credits for My QIAGEN accounts

When logged in with a My QIAGEN account, you can add analysis credits to your account using either the lot number of your QIAseq RNA sample kit or an analysis credit code following the

steps below. If you will be using a QIAseq RNA sample kit lot number, you can skip the step *Purchase analysis credits*.

1. **Purchase analysis credits.** You can purchase analysis credits on `https://geneglobe.qiagen.com/analyze/analysis-credits`. Choose between *GeneGlobe Analyze RNA-seq (12)*, or *GeneGlobe Analyze RNA-seq (24)*. Once you have placed your order, you will receive an email with the analysis credit code.

2. **Redeem analysis credits.** You can redeem analysis credits using a lot number from one of the supported QIAseq RNA sample kits or an analysis credit code at `https://geneglobe.qiagen.com/analyze/analysis-credits/redeem-credits`. For instructions on how to locate the lot number, see figure 2.

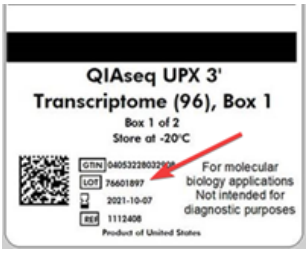


Figure 2: The lot number is on the label of your RNA sample kit box.

Your analysis credit balance and credit history is provided on the QIAGEN GeneGlobe Data Analysis Center dashboard.

Your analysis credit balance is also shown at the top, right side of the Analysis Portal page, described in the chapter Getting Started.

Credits for QIAGEN IPA accounts

When logged into *RNA-seq Analysis Portal* with a QIAGEN IPA account, you are using credits associated with your QIAGEN IPA license. Should you run out of credits, please contact bioinformaticssales@qiagen.com.

Your analysis credit balance is shown at the top, right side of the Analysis Portal page, described in the chapter Getting Started.

Getting Started

RNA-seq Analysis Portal guides you through your RNA expression analysis in 5 steps:

1. Upload sequencing data
2. Align and count
3. Create experiment
4. Filter
5. What's next

When you log into *RNA-seq Analysis Portal*, you will find steps 1-3 at the top of the Analysis Portal page (figure 3).

Your analysis credits, including the number reserved for running analyses, are displayed in the top right corner, just under your name.

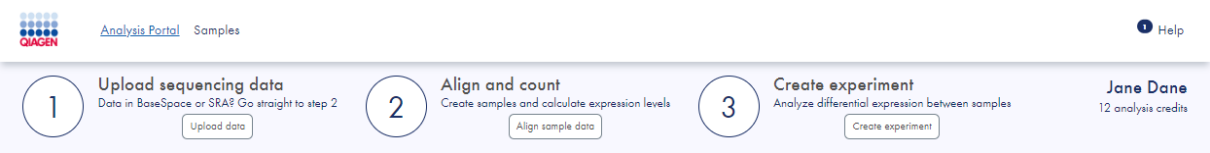


Figure 3: Analysis steps 1 to 3 are accessed from the top of the Analysis Portal page. This is also where you find your analysis credit balance.

For first-time users, onboarding pop-ups will appear to describe functionality and guide new users through the flow of uploading, aligning, creating experiments and sending the results back to GeneGlobe. You can click the "Don't show introductions tips" to stop the dispaly of these helper tips, but just close it to see more tips on later stages. If you want to re-enable these pop-ups this option is available from the 'Help' menu.

Upload sequencing data 1

Before you can start an analysis, the raw data files (FASTQ or BAM) must be available to *RNA-seq*

Analysis Portal. This section describes how to upload data from your local system. If your sequencing data is stored in Illumina BaseSpace or SRA, go to the step Align and count.

To upload data from your local system:

1. Click on the **Upload data** button at the top of the Analysis Portal page to open the *Upload sequencing data* dialog.
2. Select up to 500 files at a time in the box on the left side, either by drag-and-dropping or by clicking on **browse** to select them. Each FASTQ or BAM file can be up to 70 GB in size.
3. In the **File to sample** section on the right, check that, if using FASTQ files, they are grouped into samples as expected (figure 4). Grouping is based on FASTQ file naming. If the sample grouping is not as expected, please check and adjust the file names as necessary. See Appendix A for information about the expected naming pattern.
4. Click on **Upload**.

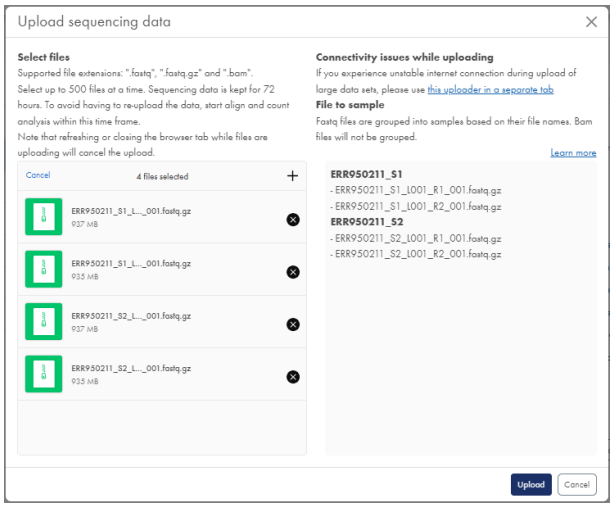


Figure 4: Selection of locally stored data in the Upload sequencing data dialog.

You can monitor upload progress by clicking the **Uploads** icon next to Help at the right hand side of the navigation bar. When an upload completes, a **Align sample data** button will appear next to it in the **Uploads** dropdown (figure 5).

You are now ready to start analyzing your data. You can proceed in two ways:

- Click on the **Align sample data** button in the **Uploads** dropdown (figure 5). This will open the *Align and count* dialog with the sample data pre-selected and lets you to skip a couple of steps. Follow the Align and count instructions from bullet 3.
- Follow the instructions under Align and count from the beginning.

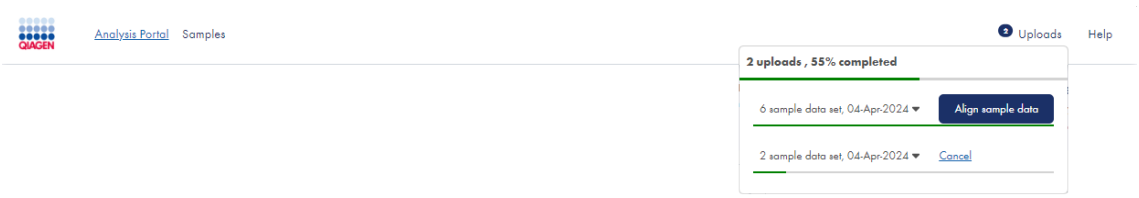


Figure 5: Upload progress is visible from the Uploads icon.

Note: Do not refresh your browser while files are being uploaded as this will cancel the upload.

Uploaded data is kept for 72 hours. To avoid having to upload the data once again, start the align and count analysis within this time frame.

If you upload files for the same sample in different upload actions, those files will still be grouped together as expected and be represented as one sample in the *Align and count* dialog.

Should you experience connectivity issues like time-outs or unstable internet connection, you can use the alternative uploader. Click the link in the upper right corner in the Upload dialog in the "Connectivity issues" section. This opens in a separate tab in your browser. This tab must stay open during the upload and not be refreshed. Select files as described above and click 'Upload'. The alternative uploader lets the application do multiple retries and reconnections and keeps the session active for +20 hours.

Align and count (2)

The Align and count analysis step analyzes the data, producing expression levels (count data) for the samples.

There is an analysis credit cost to this step, which is why you need to have the required amount of analysis credits in your account to start analysis. The analysis credit cost equals the number of samples that will be created from this step. For most sample kits, this is the same as the number of input sample data sets. For a kit like the QIAseq UPX 3' Transcriptome Kit, it equals the number of wells selected. Your analysis credit balance is visible from the top right of the Analysis Portal page. The analysis credit cost will be calculated once you have made your selections from the *Align and count* dialog drop-downs as described below. If you do not have sufficient analysis credits, you may follow the link to purchase more.

1. Click the **Align sample data** button at the top of the Analysis Portal page. This opens the *Align and count* dialog (figure 6).
2. Select the data to analyze. This can be data from the Sequence Read Archive (SRA), BaseSpace or data uploaded to the *RNA-seq Analysis Portal* earlier:

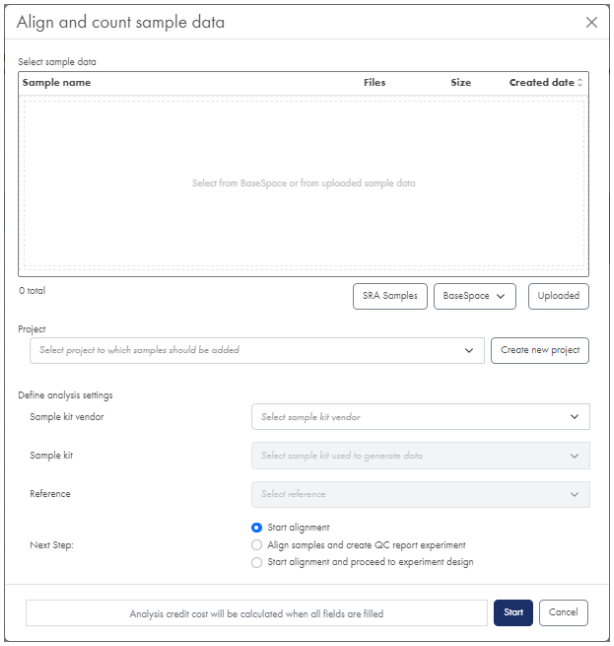


Figure 6: The Align and count dialog.

(a) **Selecting SRA data**

i. Click on the **SRA samples** button to open the *Sequence Read Archive - data retrieval* dialog (figure 7).

ii. In the *SRA accession IDs* field, enter the identifiers separated by comma, semi-colon, tab or line break if needed.

To search SRA for relevant identifiers, click on the *Sequence Read Archive's advanced search* link to open SRA in a new tab. Look for identifiers beginning with SRR, SRX, SRS, or SRP to point to one or more samples.

iii. Click on **Search** to populate project summary info, a link to associated metadata, and a list of samples for each identifier.

iv. Click on the *Metadata* link next to each project to download the metadata needed later for setting up experiments. This opens the relevant page in SRA.

v. Select the samples to analyze and click on **Add**.

(b) **Selecting BaseSpace data**

i. Click on the **BaseSpace** dropdown and select the region where you data is stored. This redirects you to Illumina's Sign In page (figure 8).

ii. Enter your BaseSpace credentials and click on **Sign In**. This takes you back to *RNA-seq Analysis Portal*.

iii. In the *Select sample data from BaseSpace* dialog, select the relevant project and click on **Open**.

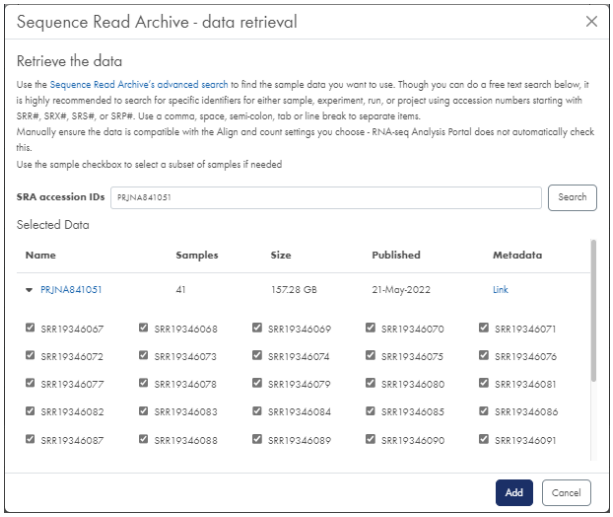


Figure 7: The Sequence Read Archive data retrieval dialog allows you to search and select data from SRA.

iv. Select the data to use and click on **Select**.

Your BaseSpace session will expire after three hours of being inactive.

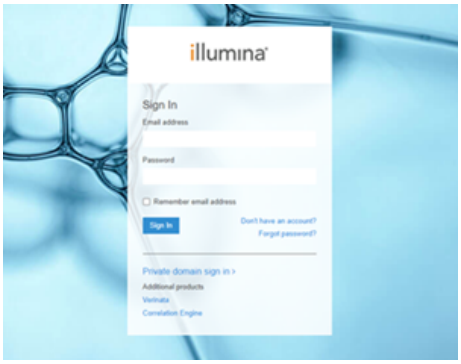


Figure 8: The Illumina Sign In page

(c) **Selecting previously uploaded data**

i. Click on the **Uploaded** button.

ii. In the *Select from uploaded sample data* dialog, select the data to use. Use the filter and sort options to limit the list of samples to select from.

iii. Click on **Select**

Your sample data will now be listed in the *Align and count* dialog. If you wish to refine the selection by adding or removing items on the list this can be done by either clicking the "x" next to the sample name to delete it from selection, or reopening one of the input selection types mentioned above.

3. Define the project that your samples should be added to. Select an existing project from the drop-down list, or create a new project by clicking on **Create new project**.

The choice of project will determine the analysis workflow *version* with which your samples will be processed, see the section About analysis workflow versions and sample comparability.

4. From the *Sample kit vendor* drop-down list, select the vendor of your sample kit.
5. From the *Sample kit* drop-down list, select the sample kit relevant for your data.

   For the QIAseq UPXome RNA Lib Kit or QIAseq FastSelect RNA Lib Kit, select the dropdown item that corresponds to the applied cDNA synthesis protocol: N6-T RT primer, ODT-T RT primer, or combined N6-T RT and ODT-T RT primers.
6. If you used spike-ins, check the *Spike-ins* box.
7. From the *Reference* drop-down list, select the relevant reference data option (figure 9).

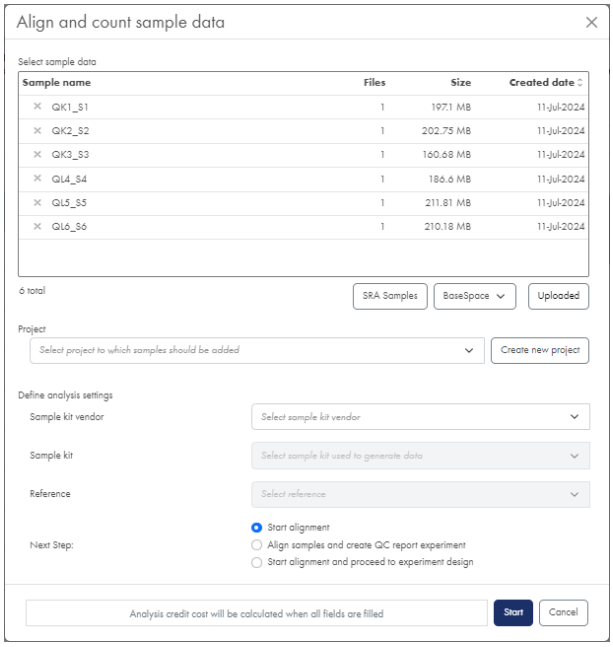


Figure 9: In the Align and count dialog, select sample data, project, vendor, sample kit and reference.

If you selected the QIAseq UPX 3' Transcriptiome Kit, you must specify if you used the custom R2 primer:

8. Custom R2 primer (QIAseq D Read 2 Primer I): Check the box if you used the Custom R2 primer.

   See Align and count dialog for additional guidance.

Both the QIAseq UPX 3' Transcriptiome Kit and QIAseq UPXome RNA Lib Kit rely on demultiplexed sample data for the analysis. A well-selection option is presented in the demultiplexing step

if you select only one input sample. You can specify number and location of samples to be demultiplexed from the input file. "Use already demultiplexed sample data" option will be automatically enabled if multiple input are selected, as we assume that the data has been demultiplexed elsewhere.

9. Well selection: Click on **Open well-picker** and select the wells used when generating your data (figure 10). Then click on **OK**.
   See Align and count dialog for additional guidance.

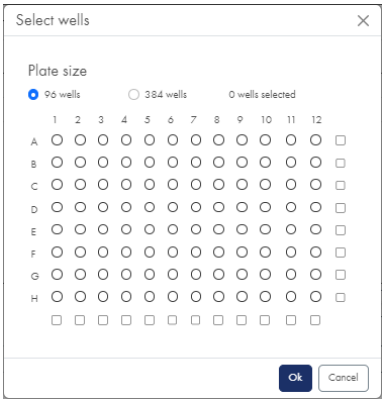


Figure 10: Well selection is required for the QIAseq UPX 3' Transcriptiome Kit and QIAseq UPXome RNA Lib kits.

Once all mandatory selections have been made, the analysis credit cost will be displayed at the bottom of the dialog (figure 11). If you have insufficient analysis credits in your account, you will need to purchase more before you will be able to proceed.

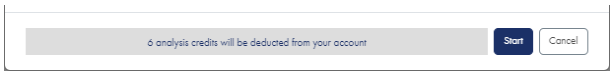


Figure 11: The analysis credit cost will be displayed at the bottom of the *Align and count* dialog once all mandatory selections have been made.

10. Under *Next Step*, select how to proceed when you click **Start**:

    - *Start alignment*. Align and count analysis will start and the dialog will close. When you later wish to set up an experiment with your samples, follow the instructions under Create experiment from the beginning.
    - *Align samples and create QC report experiment.* Align and count analysis will start and a simple *QC Experiment* report will be created in the selected project. As no sample metadata has been selected for grouping of samples, the QC experiment report will not include PCA plot under the Samples tab. Apart from this, content is the same as for the Experiment summary and QC report of a full experiment.

- *Start alignment and proceed to experiment design*. Align and count analysis will start and you will be taken to page 2 of the *Create experiment* dialog with your newly created samples pre-selected. With this, you skip a number of steps and can follow the Create experiment instructions from bullet 6.

11. Click on **Start** to submit the data for analysis.

Depending on your preferences, you will receive an email notification when your samples are done, see *Email notification setup* in the User Interface chapter.

Create experiment 3

The Create experiment step is for setting up differential expression analysis for groups of samples based on sample metadata. A brief description is below. Please refer to the Create experiment dialog section later in this manual for details.

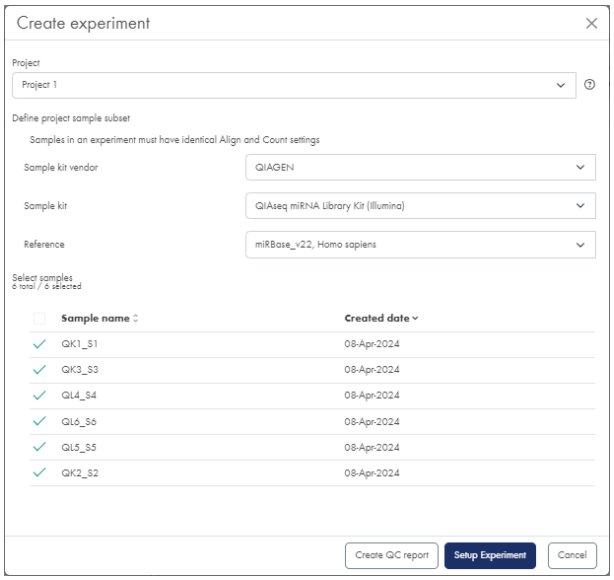


Figure 12: On page 1 of the Create experiment dialog you select project and samples.

1. Click on the **Create experiment** button at the top of the Analysis Portal page. Alternatively, go into the project you selected in the *Align and count* dialog and click on **Create experiment** there. This opens the *Create experiment* dialog (figure 12).
2. Select the project with the samples to be analyzed. If you clicked **Create experiment** from within a project, that project will be preselected.
3. Define a subset of the project samples by selecting the relevant Vendor, Sample kit and Reference.

It is only possible to run differential expression analyses using samples analyzed with the same sample kit and reference settings.

4. From the subset now listed in the bottom half of the dialog, select the samples to analyze.
5. Select how to proceed:
    - Click on **Create QC report** to create an experiment containing only the Experiment summary and QC. This allows you to see sample statistics without running a full experiment. The dialog will close.
    - Click on **Setup experiment** to go to page 2 of the dialog and proceed as described in the following.

At the top of page 2 of the Create Experiment dialog, your settings and sample selection from page 1 are summarized (figure 13).

6. Fill in the Experiment name and Experiment description fields.
7. Define the sample grouping, either by adding a sample metadata file, or by adding attributes and values manually. Samples can be divided into groups based on the values of each attribute.

In the Experimental design section, specify the details of the analysis.

8. In the *Test differential expression due to* drop-down menu, select the attribute to test for expression effects. In figure 13, we selected *Tissue*.
9. In the *While controlling for* drop-down menu, you can select an attribute to correct for. In figure 13, we did not select an attribute to correct for.
10. In the *Experimental setup (comparisons)* drop-down menu, select the type of comparison to run:
    - Across groups (ANOVA-like)
    - All group pairs
    - Against control group

In the example shown in figure 13, the choice of attribute *Tissue* results in the six selected samples being divided in two groups based on the values of that attribute, *Normal* and *Tumor*.

11. If you selected *Against control group*, you need to specify a control group attribute value, i.e. the group to compare the remaining groups to.

The number of differential expressions that will created is stated at the bottom of the dialog.

12. Click on **Start**.

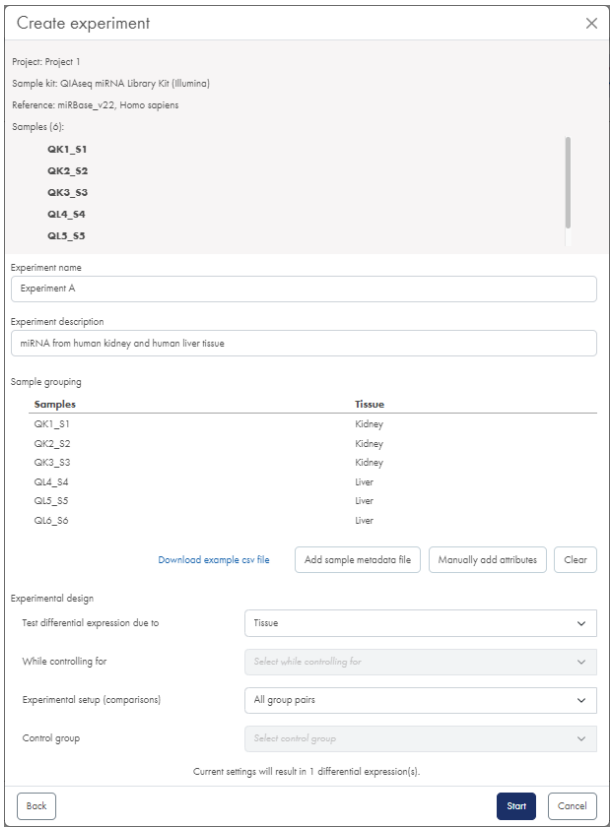


Figure 13: In Create experiment dialog page 2 you define how samples should be compared.

Your experiment will now be created and will be listed under the specified project.

When the analysis is complete, the status of the Experiment overview changes from *In progress…* to *Done* and an additional tile, *Experiment summary and QC*, is added.

Depending on your preferences, you will receive an email notification when your experiment is done, see *Email notification setup* in the User Interface chapter.

Results of the differential expression analyses can now be inspected and refined.

View and filter results

1. Click on one of the differential expression results on the Experiment analyses overview to open up the differential expression result view (figure 14). This view is explained under

Differential expression view for more info on this step.

2. Filter the feature table at the top left by adjusting the p-value and fold change thresholds. You can do this using the fields in the **Filter features** section, in the center at the bottom, or by dragging the corresponding threshold bars in the volcano plot.

3. Click on **Advanced filters** to access additional filter criteria like expression values and biotypes.

4. Apply a biological insights filter by selecting an item from the Biological insights table, at the bottom on the left.

Figure 14: Adjust the p-value and fold change thresholds, and apply advanced and biological insights filters to adjust the feature table contents. The figure shows the 'What's next'-feature available to My QIAGEN users.

When you are satisfied with the filtering and the set of features in the feature table, you can save your filter.

5. Click on **Save**.

6. In the *Save filter* dialog, either go with the auto-generated filter name, or choose a name of your own.

7. Click **Save**.

What's next - send results downstream

Depending on whether you are logged in with a My QIAGEN account or a QIAGEN IPA account, you can now either send results to GeneGlobe *What Next - Follow-up Experiment Planner* or upload them to your IPA account, as described in the following.

Next step for My QIAGEN users

When logged in with a My QIAGEN account, you can send results to GeneGlobe *What Next - Follow-up Experiment Planner*:

1. Select a saved filter (figure 15).

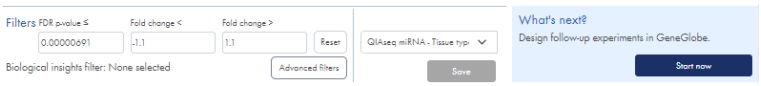


Figure 15: The Filter and What's next sections from where you can save filters and send results to GeneGlobe.

2. In the *What's next* box on the bottom right, click on the **Start now** button to send the filtered list of features and the full lists of Canonical Pathways, and - when applicable - Upstream regulator (miRNA) to GeneGlobe *What Next - Follow-up Experiment Planner*.
3. GeneGlobe *What Next - Follow-up Experiment Planner* opens in a new tab. Depending on your browser settings, you may need to first allow pop-ups. See Design follow-up experiments in GeneGlobe for instructions on how to proceed in GeneGlobe.

Next step for QIAGEN IPA users

When logged in with a QIAGEN IPA account, you can send results to your IPA account:

1. In the *Send to QIAGEN IPA* box on the bottom right, click on the **Send** button (figure 16).

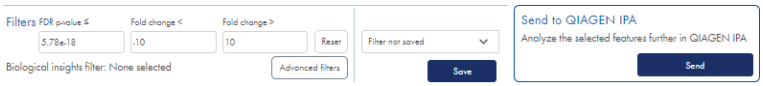


Figure 16: The Send to QIAGEN IPA box is located at the bottom right of the differential expression view.

2. In the dialog, enter *Name of IPA project folder*. Enter the name of an existing project to have your data added to this, or enter a new name to create a new project in your IPA account.
3. *Name of dataset* is the name of *RNA-seq Analysis Portal* filter.
4. Select *Upload only* or *Upload and start Core analysis in IPA*.
5. Add a optional Data set description.

6. Click on **Send**.
7. Go to QIAGEN IPA to proceed with the analysis. IPA knowledgebase documentation is available here: `https://qiagen.my.salesforce-sites.com/KnowledgeBase/KnowledgeNavigatorPage?id=kA41i000000L6rMCAS&categoryName=IPA`

Design follow-up experiments in GeneGlobe

In GeneGlobe, based on the list of Genes or miRNAs, miRNA upstream regulators (selected analyses), and canonical pathways sent from *RNA-seq Analysis Portal*, you can identify products for conducting follow-up experiments like biomarker verification.

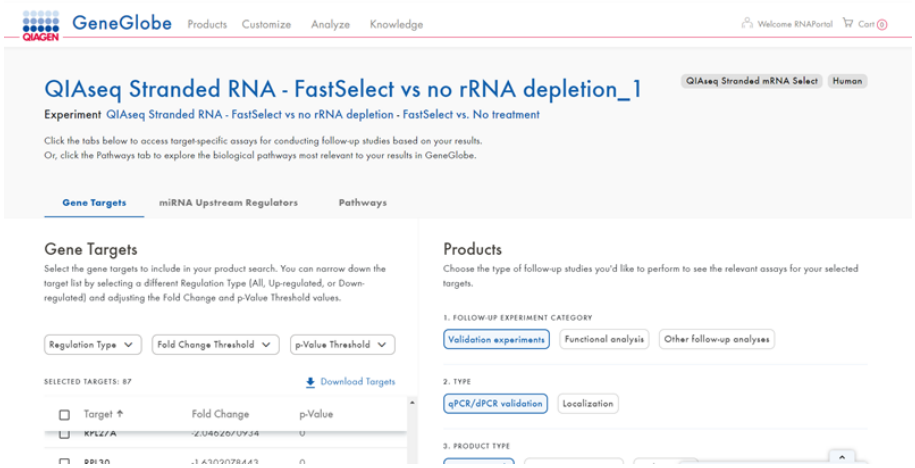


Figure 17: In GeneGlobe, filter your features futher by adjusting thresholds or selecting features individually.

You can toggle between the *Genes/miRNA*, *miRNA upstream regulators*, and *Pathways* tabs (figure 17).

By selecting Genes/miRNAs or miRNA upstream regulators of interest and choosing your type of follow-up experiment, you will be guided through the QIAGEN product portfolio and presented with a list of relevant products:

1. On the *Genes/miRNA* and *miRNA upstream regulators* tabs, select one or more features (figure 18).
2. Under *Products* on the right, select *Follow-up experiment category* and *Type*.
3. Choose *Product type* to see a list of suggested products.
    - *Custom panels* and *Customized products*. These will be designed specifically for your list of targets. To order a custom product, click on **CREATE & ORDER** to launch the product design tool.

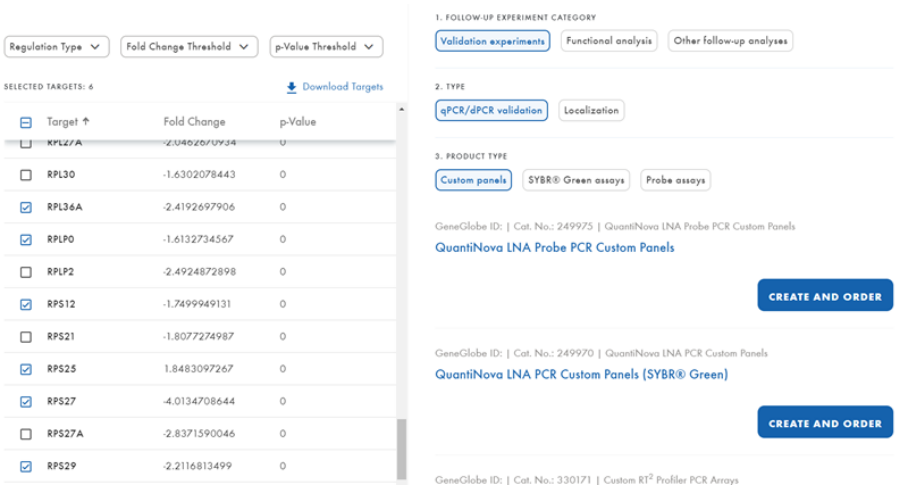


Figure 18: The list of suggested products depends on your selection

- Other product categories list ready-to-order assays that match your selection of genes/miRNAs or miRNA upstream regulators. Click on *Product specification* to see additional details. To purchase, click on **CONFIGURE** and then **ADD TO CART** (figure 19).

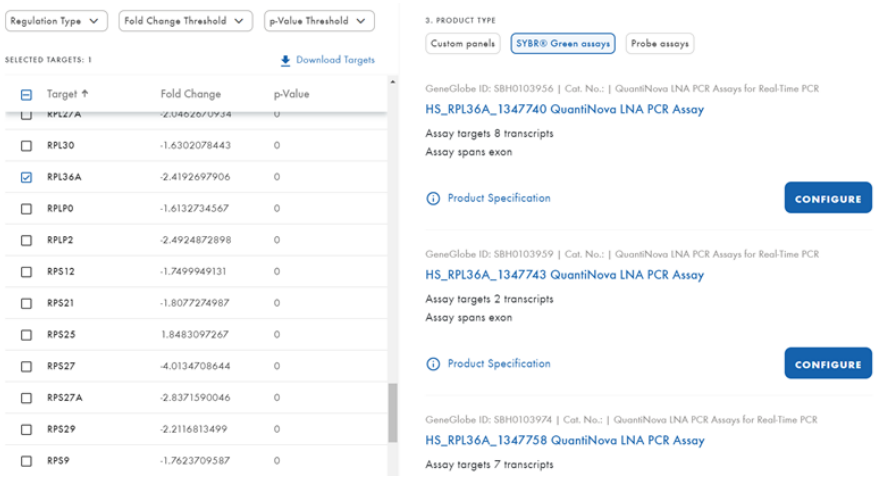


Figure 19: Browse the list of Ready-to-order products relevant for your targets.

The *Pathway* tab allows you to explore pathway maps for the canonical pathways identified in *RNA-seq Analysis Portal*. Here, you can browse through the pathway maps, explore the biology behind your findings and save genes in the pathways to the gene lists in your My GeneGlobe account.

You can access your *RNA-seq Analysis Portal* projects, gene lists and custom designs from the My GeneGlobe menu at the top of the page (figure 20).

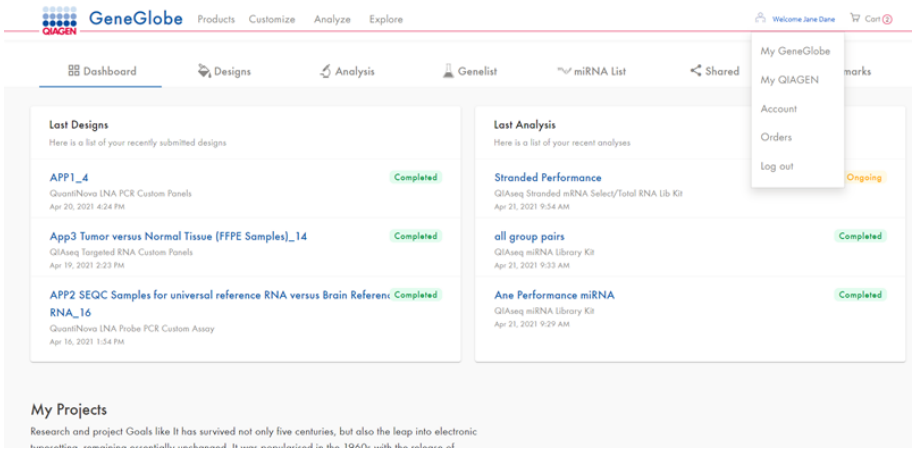


Figure 20: Access My GeneGlobe from the top menu.

User Interface

The *RNA-seq Analysis Portal* interface has two main pages, **Analysis Portal**, and **Samples**. You navigate between these via the navigation bar at the top. The pages are described in detail in later sections.

Figure 21: The RNA-seq Analysis Portal user interface includes a top navigation bar.

The **Help** menu, at the right hand side of the navigation bar, contains links to product resources like the user manual and the *Contact support* form (figure 21). In addition, the menu holds the following options and sections:

- **Share experiment with Support**. When on Project view or Differential expression view, the **Help** menu will feature the option *Share experiment with Support*. This will allow you to share the active experiment with QIAGEN Technical Support for investigation or troubleshooting. To link the shared experiment to an existing support case, enter the support case number in the designated field. The support case number is provided in the subject line of the emails you have received from QIAGEN Technical Services.
  Your experiment will be copied to a dedicated Support account.
- **Email notification setup**. By default, you will receive an email when a batch of samples or an experiment is done, or when a colleague sends you a copy of a project. The **Help** menu option *Email notification setup* allows you to adjust your preferences for when to receive *RNA-seq Analysis Portal* notification emails (figure 22).
- **Pending project copies**. A list of the copies of projects that were shared by you, or shared with you by someone else (see Send copy of project).

To **log out** of *RNA-seq Analysis Portal* and your My QIAGEN account, click on the icon in the footer.

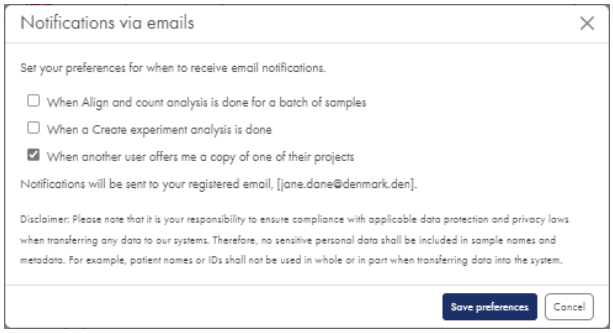


Figure 22: Via the Email notification setup you can opt out of receiving emails when samples or experiments are done, and when colleagues send copies of projects to you.

Click on the ⓘ icon (figure 23) in the footer to find links to the end user and data processing agreements and third party software notices.

Figure 23: Log out by clicking on the icon at the far right of the RNA-seq Analysis Portal footer.

RNA-seq Analysis Portal terminology and data organization

Terminology

The following terminology is used for different types of data and output:

- **Sample data**. Raw sequencing files grouped based on file naming. Input to the *Align and count* analysis step.
- **Samples**. Expression level samples, i.e. count data. Output from the *Align and count* analysis step.
- **Differential expression**. A comparison of expression level differences between two or more groups of samples. The *Create experiment* analysis step generates one or more differential expressions.
- **Experiment summary and QC report**. An overview of the samples within a particular experiment, the metadata assigned to them, and selected analysis quality control metrics.
- **Comparison**. A comparison of two or three differential expression analysis results. Output from *Compare analyses*.
- **Analysis credits**. The cost of one Align and count analysis is one analysis credit. Note that for the QIAseq UPX 3' Transcriptome Kit, the number of Align and count analyses and

hence the total credit cost will depend on the number of wells selected. If a sample is deleted while the Align and count analysis is still in progress, or if the analysis should fail, the analysis credit is returned to your account.

Data organization

Data is organized into projects and experiments:

- Every sample belongs to a project, specified in the *Align and count* dialog.
- A project is a collection of samples and experiments.
- An experiment belongs to a project, specified in the *Create experiment* dialog. It is created from samples within the specific project, often just a subset. An experiment will contain one *Experiment summary and QC* report and one or more differential expression analysis results, all based on the same metadata attribute(s). It can also contain one or more comparisons.

Analysis Portal page

The *Analysis Portal* page includes multiple layers, described in further details below and in the Viewing Results chapter:

- Front page with 1-2-3 button bar and Project overview
- Project view
- Analysis result view

The breadcrumb at the top of each view can be used for orientation and navigation through the layers (figure 24).

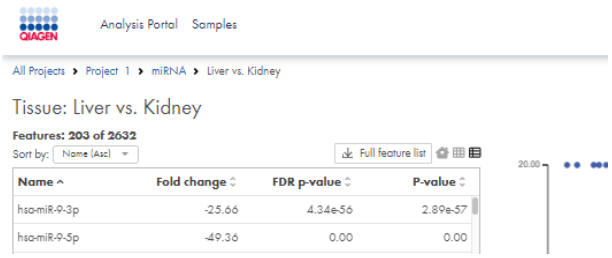


Figure 24: Use the breadcrumb underneath the QIAGEN logo to navigate through projects, experiments and analyses.

Front page and Projects overview

The *Analysis Portal* front page includes two main sections: The top bar, from where you initiate analysis steps 1, 2 and 3, and below it the *Projects* overview.

The 1-2-3 buttons are described in the Getting Started section of this manual.

Projects overview

The *Projects* table lists the name, creation data and number of samples in each project within your *RNA-seq Analysis Portal* account (figure 25).

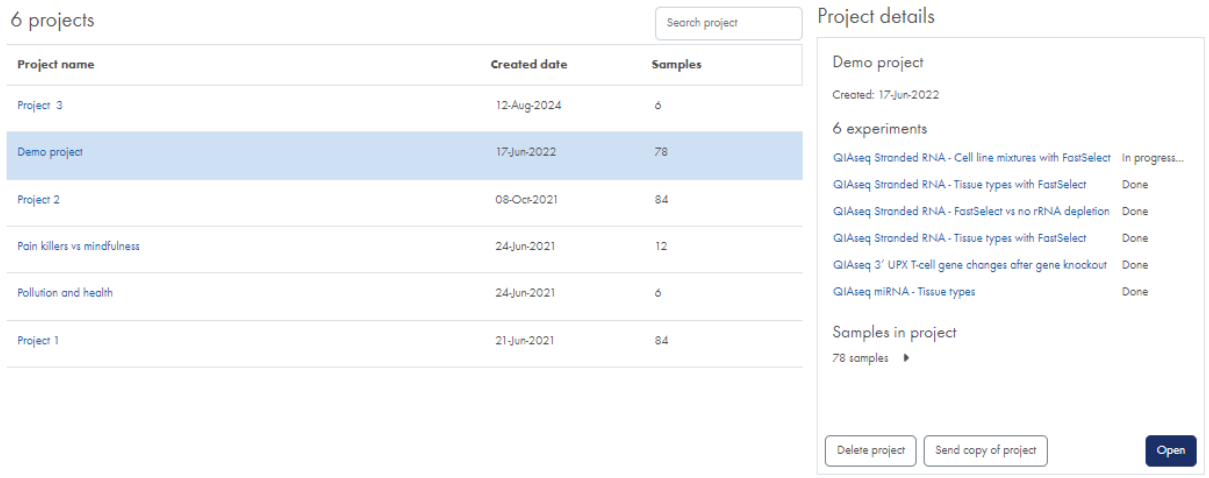


Figure 25: The Projects overview with Project details panel on the right.

The *Project details* panel on the right provides additional information about the selected project, including a list of experiments and the number of samples within the project. The number of samples label can be expanded to see a list of the individual samples in the project, including date and with which panels they have been analyzed. Experiment status is listed here, indicating for each experiment, if it is *In progress..*, or *Done* and ready for review. At the bottom of the panel, you find the following buttons:

- **Delete project**. Deletes the project, including contained experiments, analyses and filters as well as the associated samples.
- **Send copy of project**. Allows you to send a copy of the project to a specified user, see Send copy of project.
- **Open**. Opens the selected project. Alternatively, you can navigate to a specific project by clicking on the project name in the projects list.

To navigate to a specific project, either click on the name of the project or click on the **Open**

button in the *Project details* panel. Alternatively, click on the name of a specific experiment to enter the project with that experiment selected.

Project view

The *Project view* on the left hand side presents the same information that is found in the *Project details* panel on *Projects overview* page. Specifically: project name, description, a list of experiments and the number of samples in the project in the same expandable list. For each experiment the status is provided, indicating whether the Align and count analysis step is still running (*In progress..*, *Done*) (figure 26).

Clicking any of the experiment titles in the left panel will bring up the details of that experiment including the individual analyses.

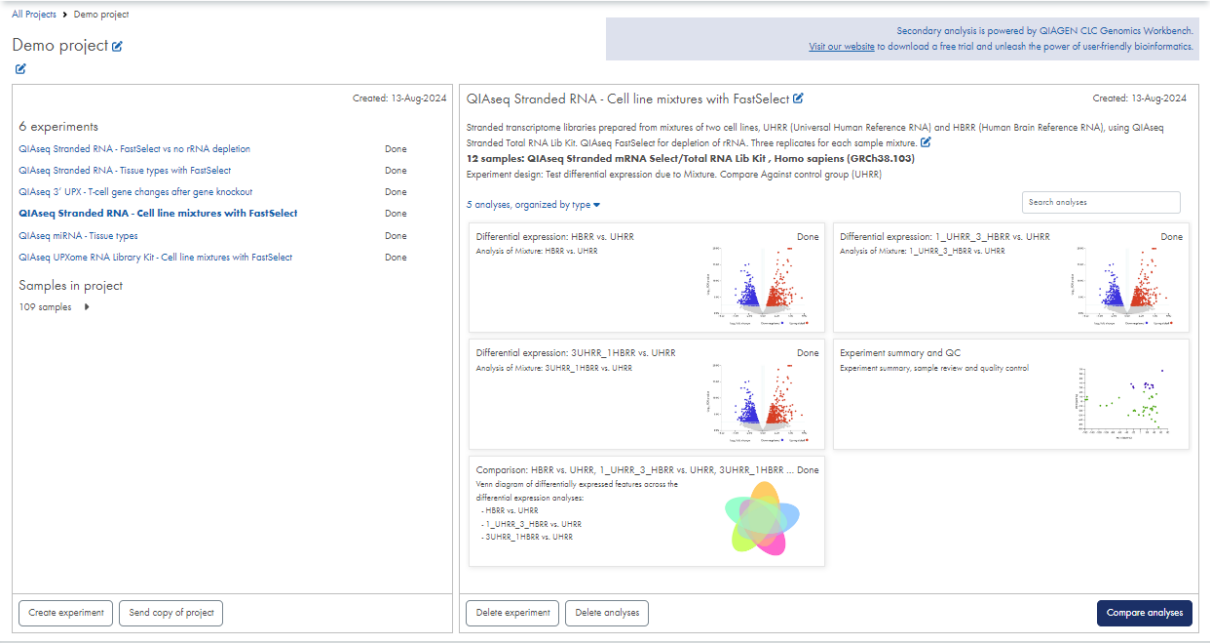


Figure 26: The project view with experiment details is on the right. The different result types are indicated by their images.

To create a new experiment within the project, click on **Create experiment** at the top of the experiment list. This opens the *Create experiemnt dialog* with the project preselected.

To send a copy of the project to someone else, click on **Send copy of project**, see Send copy of project.

The experiment view on the right provides details about the selected experiment, including a summary of the experimental design. The tiles underneath represent the different outputs. The three result types are indicated using images: *Experiment summary and QC* tiles are indicated by a PCA plot, *Differential expression* tiles are identified by a volcano plot, and *Comparison* tiles are

identified by a Venn diagram.

While the *Create experiment* analysis step is still in progress, an *In progress..* a tile will be displayed for each planned differential expression.

To view results, click on the tile of interest.

The various analysis result views are described in detail in the Viewing Results chapter.

You can delete both an entire experiment and selected analyses via the *Delete* buttons below the tile area. The **Delete experiment** button will delete the selected experiment, including contained analyses and filters. The **Delete analyses** button will let you select which analyses to delete. If you delete a differential expression analysis, the Comparisons in which this differential expression is contained will be deleted as well.

Click on the **Compare analyses** button below the tile area to create a comparison. See the Compare analyses section for information about this functionality.

Send copy of project

From both the *Projects overview* and the *Project view* you can send a copy of a project including the experiments, analyses and associated samples to someone else. Once accepted by the recipient, the copy will exist as an independent copy in their *RNA-seq Analysis Portal* account.

To send a copy of a project to someone else:

1. Select the relevant project.
2. Click the **Send copy of project** button to open the *Send copy of project* dialog (figure 27).
3. In the dialog, enter the email address of the recipient. The person does not have to be an existing *RNA-seq Analysis Portal* user. If you send a copy of project to an existing *RNA-seq Analysis Portal* user, make sure to use the email address associated with their *RNA-seq Analysis Portal* account.
4. Enter an optional message. This will appear in the notification email sent to the recipient.
5. Click **Send copy**

Depending on their preferences (see *Email notification setup* in the User Interface chapter), an email is sent to the recipient to let them know that a copy of an *RNA-seq Analysis Portal* project was sent to them and that they must log into *RNA-seq Analysis Portal* to accept the project copy.

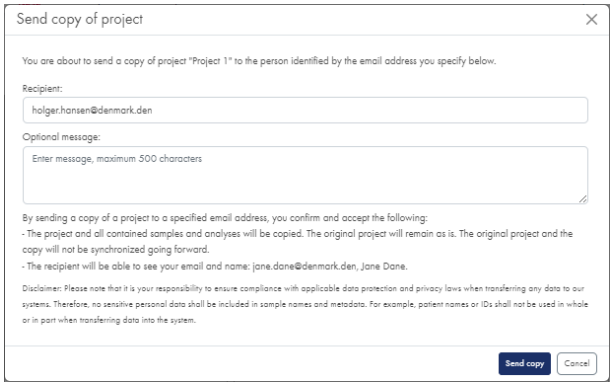


Figure 27: In the Send copy of project dialog, specify the email address of the person to whom you wish to send a copy of your project. You also have the option to enter a comment. comment.

A badge next to the **Help** icon will advertise the number of pending project copies, both those sent by you and those sent to you by others. You find a list of your pending project copies in the **Help** menu bottom half. Click on **View full list** to open the *Pending project copies* dialog with a table view representation (figure 28).

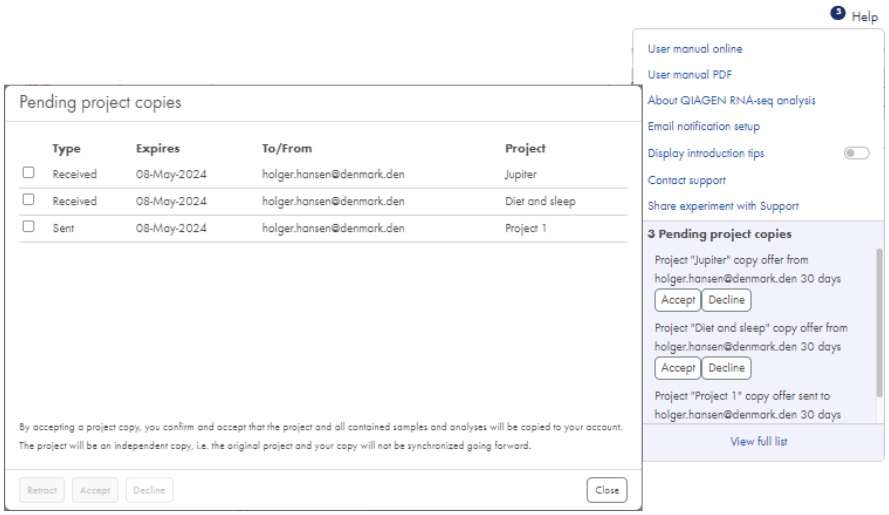


Figure 28: The Help menu with Pending project copies listing (right) and the Pending project copies dialog (left).

Retract copy of project.

You can withdraw a project copy offer extended by you until it has been accepted by the recipient. To do so, click on **Retract** next to the relevant item in the **Help** menu, or select the item in the *Pending project copies* dialog and click **Retract**.

Accept or decline copy of project

When a copy of an *RNA-seq Analysis Portal* project has been sent to you, you must log into the application to accept or decline. You must do so within 30 days after which the project copy offer

expires.

As you log in, a message dialog will show for a few seconds, and again later as you navigate between pages and views (figure 29). In this dialog, click on **View full list** to open the *Pending project copies* dialog (figure 28). Select the relevant items and click on **Accept** or **Decline**. Alternatively, go to the **Help** menu to accept or decline from there.

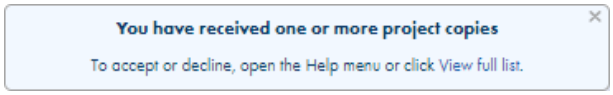


Figure 29: A message dialog will notify you if you have pending project copies.

Once you accept a received project copy, the project will appear in your *Projects overview* alongside your existing projects, and the associated samples will show up on the *Samples* page.

Samples page

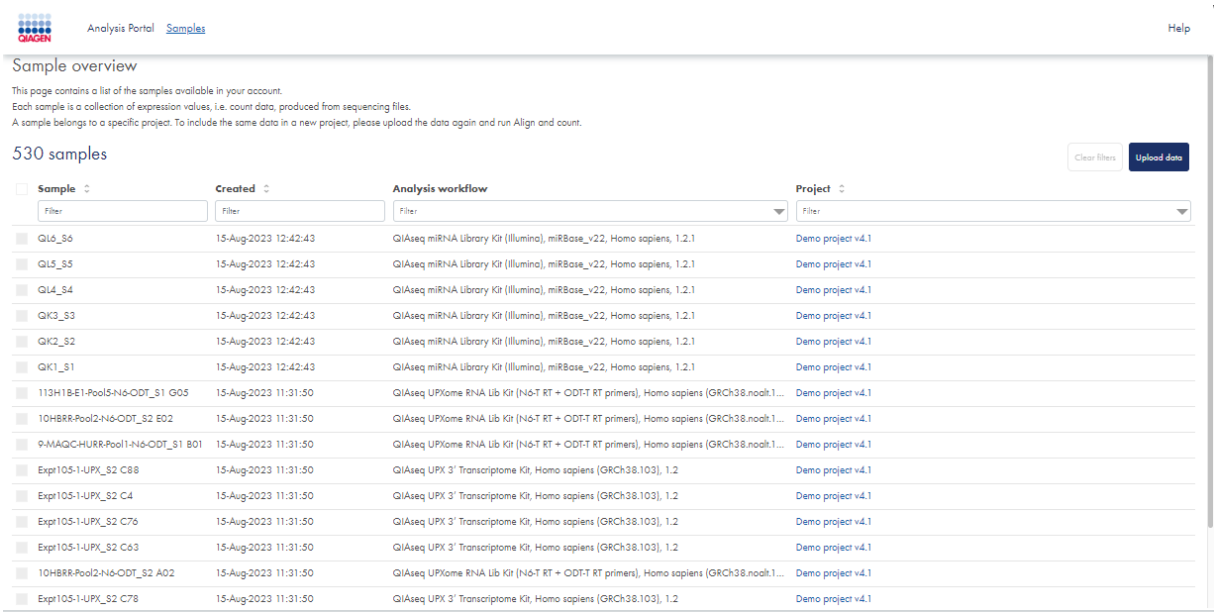


Figure 30: The RNA-seq Analysis Portal Samples page

The *Samples* page provides an overview of the samples in your *RNA-seq Analysis Portal* account (figure 30). The following information is provided:

- Sample name
- Created date. Contains the date that the count data was created for that sample, or the text: *In progress..*, which indicates the *Align and count* analysis is still running.

- Sample kit analysis workflow. Combines Sample kit and reference as selected in the *Align and count* dialog and lists the analysis workflow version in parenthesis.
- Project. The project to which the sample belongs. Click on the project name to go to the project.

Filter table columns to list entries that match your criteria, e.g. all samples in a specific project. Click on **Clear filters** at the top right to remove all filters.

Delete and move samples

You can delete and move samples as long as they are not part of an experiment. Samples that cannot be deleted or moved will appear with disabled checkboxes.

Delete samples

To delete samples, select the checkboxes on the left and click the **Delete samples** button below the table.

If you delete a sample while the analysis is still in progress, the corresponding analysis credit will be released back into your My QIAGEN account.

Move samples

To move samples, select the checkboxes on the left and click the **Move samples** button below the table to bring up the *Move selected samples* dialog (figure 31). Specify the project to which you wish to move the samples by either selecting an existing project or creating a new one.

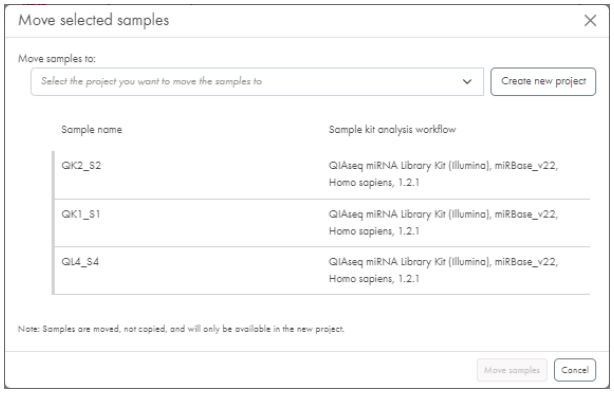


Figure 31: The Move selected samples dialog where you will need to first select the destination project.

You cannot move samples into a project that already contains samples with identical sample kit-reference combination, but different analysis workflow version. This restriction ensures comparability of samples within a project.

As an example, say you have a destination project with samples of the following type and version: *QIAseq miRNA Library Kit miRBase_v22, Homo sapiens (GRCh38.103), 1.1*. What samples could you move into this project?

- QIAseq miRNA Library Kit miRBase_v22, Homo sapiens (GRCh38.103), 1.1? Yes - same sample kit-reference combination and same version.
- QIAseq miRNA Library Kit miRBase_v22, Homo sapiens (GRCh38.103), 1.0? No - same sample kit-reference combination, but different version.
- QIAseq miRNA Library Kit miRBase_v22, Mus musculus (GRCm38.101), 1.0? Yes - different reference, so no problem.
- QIAseq UPX 3' Transcriptome Kit, Homo sapiens (GRCh38.103), 1.0? Yes - different sample kit, so no problem.

If all selected samples are compatible with the destination project, this is indicated by vertical green bars next to all the samples (figure 32).

If only a subset of selected samples can be moved into the destination project, compatible samples will be indicated with vertical green bars, and incompatible samples will be striken out (figure 32). As you click **Move samples**, only compatible samples will be moved.

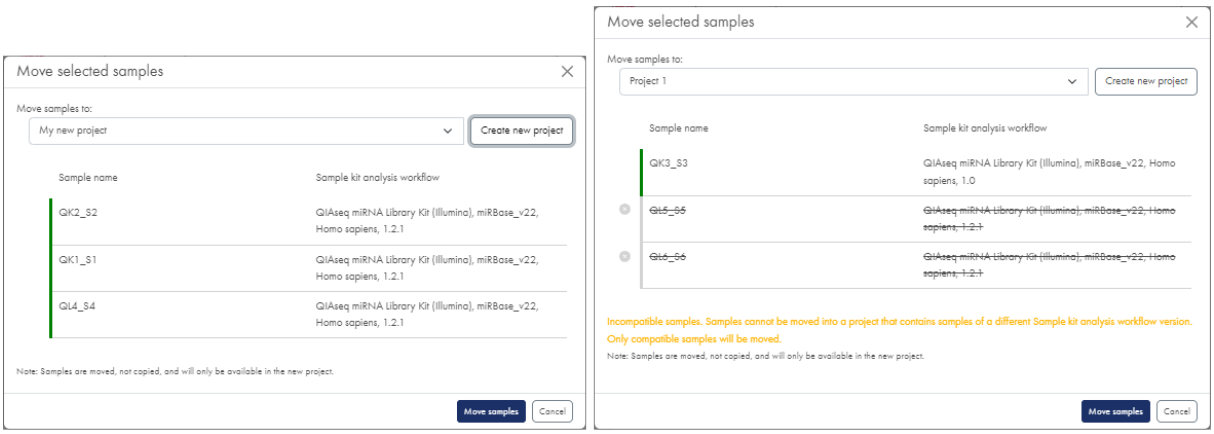


Figure 32: Left: All samples are compatible with the destination project. Right: a subset of the samples will be ignored as they are not compatible with the destination projecct.

If you select samples that are mutually incompatible, i.e. they have sample kit-reference combination, but different analysis workflow version, you will need to remove some before you can click **Move samples** (figure 33).

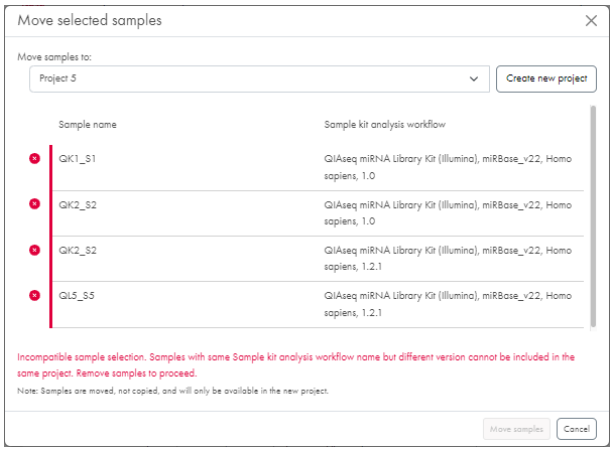


Figure 33: If samples are not inter-compatible you need to remove some before you can move the rest.

Starting Analysis

The Getting Started chapter of this manual provides a quick tutorial for starting *RNA-seq Analysis Portal* analyses, including a run through of the *Align and count* and *Create experiment* dialogs. This chapter provides more detail about some of the options and parameters in those dialogs, as well as describing the *Compare analyses* dialog.

Align and count dialog

QIAseq UPX 3' Transcriptome Kit and QIAseq UPXome RNA Lib Kit specific settings

The data generated with QIAseq UPX 3' Transcriptome and QIAseq UPXome RNA Lib Kits can be either multiplexed or demultiplexed. If your data has been demultiplexed elsewhere, you will select multiple input files in the top of the dialog. This will disable the demultiplexing well-picker functionality and pre-select the "Use already demultiplexed sample data".

When your data is multiplexed, a single file contains several samples, why an initial demultiplex analysis step is required. The below settings relate to this analysis step and are specific to the QIAseq UPX 3' Transcriptome kits.

Custom R2 primer (QIAseq D Read 2 Primer I)

This setting applies only to the QIAseq UPX 3' Transcriptome Kit.

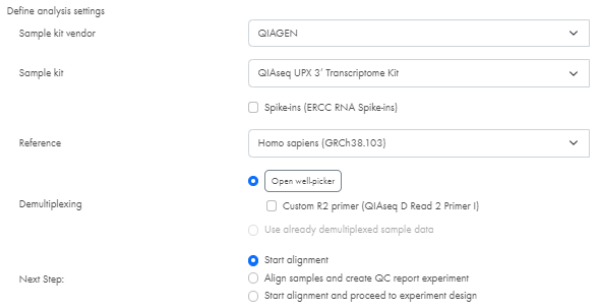


Figure 34: The custom R2 primer setting is available for the QIAseq UPX 3' Transcriptome Kit only.

The QIAseq D Read 2 Primer I option determines which adapter list is used for demultiplexing and is essential for proper processing of the sequencing reads.

For information about the custom R2 primer, please refer to the QIAseq UPX 3' Transcriptome Handbook, available from the *Product Resources* tab on the QIAseq UPX 3' Transcriptome Kit product page: https://www.qiagen.com/products/discovery-and-translational-research/next-generation-sequencing/rna-sequencing/three-rnaseq/qiaseq-upx-3-transcriptor kits/

Well selection

This setting is sample data specific. This is why you can only analyze one QIAseq UPX 3' Transcriptome or QIAseq UPXome RNA Lib Kit sample data item at a time.

1. If this option is available, specify whether you used a 96-well or a 384-well plate.
2. Select the wells used on the well plate in the laboratory. Make sure to select the correct wells, including considering the plate orientation. Use the checkboxes to the right of and below the wells to select an entire row or column (figure 35).

   Correct well selection is important because reads from a specific well all contain the same barcode. During demultiplexing, the first step in UPX data analysis, reads are grouped into samples based on the barcodes they carry. One sample is created per indicated barcode, i.e. per selected well. If well selection here is not identical to the actual wells used in the laboratory, the demultiplexing step will look for the wrong barcodes, leading to samples with very few or no reads.

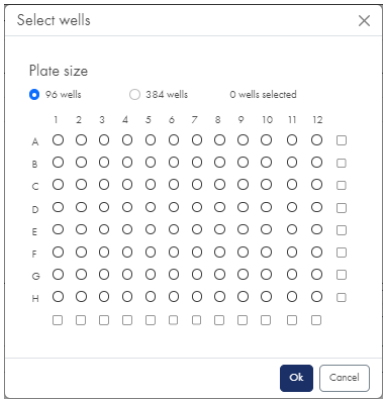


Figure 35: Select the wells that were used on the actual well plate in the laboratory.

Create experiment dialog

After samples have been selected in the first step of the *Create experiment* dialog, you define how the experiment should be set up in the next step, i.e. how samples should be compared. At this point you can either click **Create QC report** to create an experiment containing only the

Experiment summary and QC analysis report or you can click **Setup experiment** to proceed. The main sections in this next step are *Sample grouping* and *Experimental design*.

Sample grouping

In this section, you provide one or more sample metadata attributes and associated values based on which samples are grouped.
Initially, the selected samples will be listed with no attributes (figure 36). From here, you can add attributes from a sample metadata file or you can add them manually.

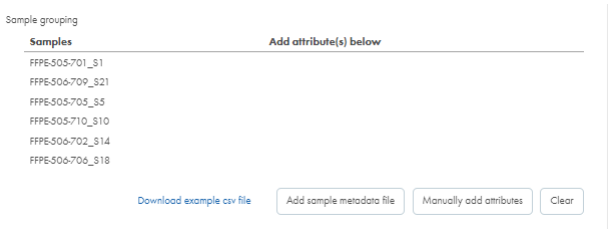


Figure 36: The Sample grouping section with no attributes specified.

Add attributes from sample metadata file

1. Click the **Add sample metadata file** button.
2. Select the relevant metadata file in .csv, .txt, or .xlsx format from your file system.

   The file must be comma or semi-colon separated, and the first column must have the header *Sample* or *Run* and contain the sample names. The header of the remaining columns specifies attributes. To be available for selection in the *Experimental design* section, attributes must have values defined for all samples.
   When using Excel with files containing special characters, choose the UTF-8 option when saving the file to .csv format.
   Click on the *Download example of csv file* link to see an example of a correctly formatted file.

   If the names of the selected samples match the sample name in the metadata file, you should now see the attributes and sample-specific values listed in the *Sample grouping* section (figure 37).
3. If needed, you can click on **Clear** to remove all attributes and values to start over.

Add attributes manually

1. Click on **Manually add attributes** to open the dialog *Add sample metadata attribute manually* (figure 38).

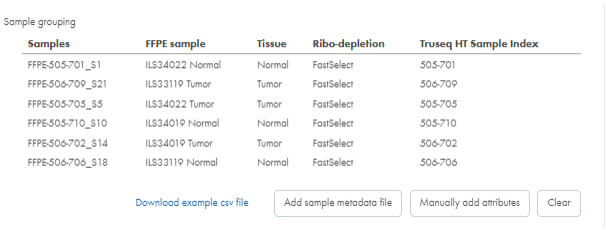


Figure 37: The Sample grouping section enriched with metadata attributes added from a csv file.

2. Specify the name of the attribute, e.g. "Diet".
3. Enter the attribute values, separated by commas, e.g. "Low calorie intake, High fat, Low fat, High protein".
4. Under *Add attribute options to samples*, for each sample select the appropriate attribute value from the drop-down list.
5. Click on **Add to samples**.
6. Repeat the above steps to add additional attributes.
   Attributes and values will be listed in the *Sample grouping* section (figure 39).

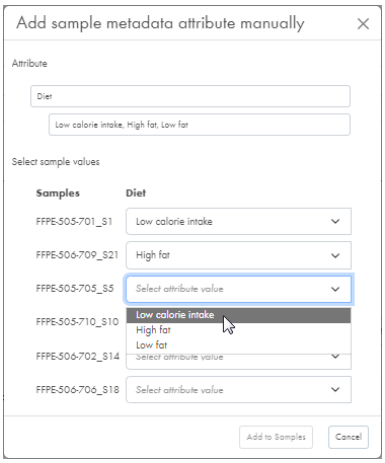


Figure 38: When adding metadata manually, you must specify the name and values for each attribute and then select the relevant value for each sample.

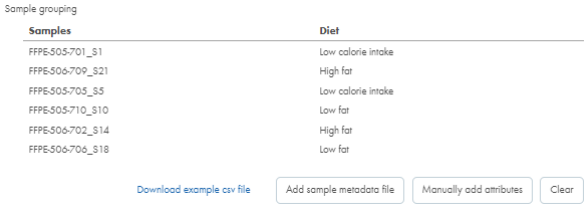


Figure 39: The Sample grouping section enriched with manually added metadata attributes.

7. If desired, you can click on **Clear** to remove all attributes and values to start over.

Experimental design

In this section you specify the attributes to use for a differential expression analysis, and how samples should be compared.

Test differential expression due to

Select the attribute to test for expression effects. In figure 40, we show an example where the effects of different diets on gene expression will be tested.

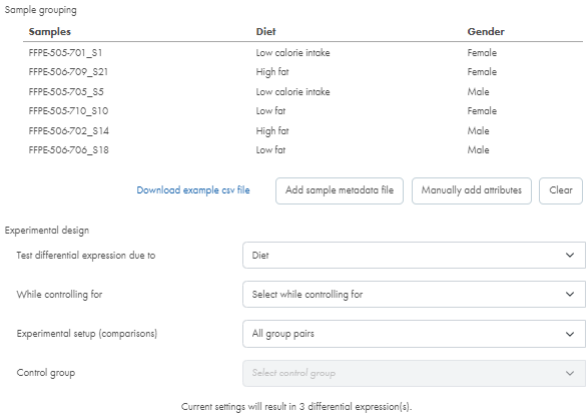


Figure 40: Designing an experiment to test if diet affects gene expression.

You will only be able to select attributes that have values defined for all samples, and for which at least two sample groups exist.

While controlling for

This selection is optional. Use this to specify confounding factors, i.e. factors that are not of primary interest, but may affect gene expression. In figure 40 we did not make a selection, but we could have selected "Gender" if we wanted to remove differences in gene expression that could be ascribed to gender.

Experimental setup (comparisons)

This selection determines the type of comparison done. Different numbers of differential expression outputs result, depending on the option selected.

- **Across groups (ANOVA-like).** Tests differential expression across all sample groups. In figure 40, we are testing on the attribute "Diet", which has three values (Low calorie intake, High fat, and Low fat). Here one differential expression output would result, "Due to Diet". **Note:** The statistics behind this analysis requires that replicates are present for each sample group. If this requirement is not met, the analysis will fail.

- **All group pairs.** Tests differential expression between all pairs of groups. In the Diet experiment example, choosing this option would produce a total of 3 differential expression outputs:
  - Low calorie intake vs High fat, and Low fat
  - Low calorie intake vs Low fat
  - High fat vs Low fat
- **Against control group.** Tests differential expression between sample groups and a specified control group. In the Diet example, if we specified "Low calorie intake" as the control group, we would get the following two differential expression outputs:
  - High fat vs Low calorie intake
  - Low fat vs Low calorie intake

Control group

If you opt for the *Against control group* option, you must select an attribute value to specify the control group that all remaining sample groups will be compared to.

Summarizing number of analyses

Depending on how you set up your experiment, a number of differential expression analyses will be created. The number is reported when all settings have been fulfilled.

Compare analyses dialog

Use this functionality too compare results of two or three differential expressions, e.g. to identify genes or miRNAs that were differentially expressed across the differential expression results.

1. On the Project view, click the **Compare analyses** button below the analyses tiles. This opens the *Compare differential expressions* dialog (figure 41).
   This button is disabled if the experiment contains only one differential expression result.
2. Select the differential expressions (two or three) to be compared.

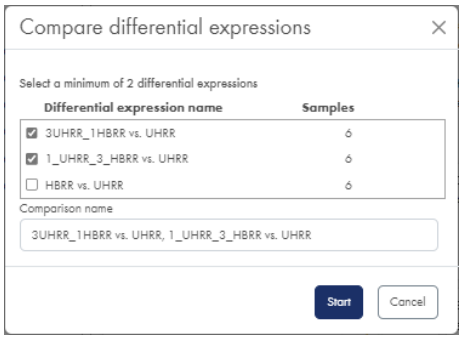


Figure 41: In the Compare differential expression dialog, select the differential expressions to compare.

Viewing Results

This chapter provides details of the various result views.

Please refer to the User Interface chapter for descriptions of the layout of other aspects of the *RNA-seq Analysis Portal* user interface, including the project and experiment overview.

Experiment summary and QC report

On the experiment analyses overview, the *Experiment summary and QC* tile is indicated by the PCA plot image (figure 42).

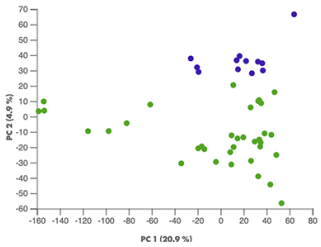


Figure 42: The Experiment summary and QC analysis tile includes a picture of a PCA plot.

Click on the tile to open it.

The *Experiment summary and QC* analysis result is a report divided into three tabs, *Experiment summary*, *Samples*, and *Quality control*.

Print or download report. To print or download the report, click on **Print view** above the three tabs (figure 43). Using your browser's print functionality, you can select a printer, or choose to save to PDF.

Experiment summary

The *Experiment summary* tab lists key information about the experimental design, including the

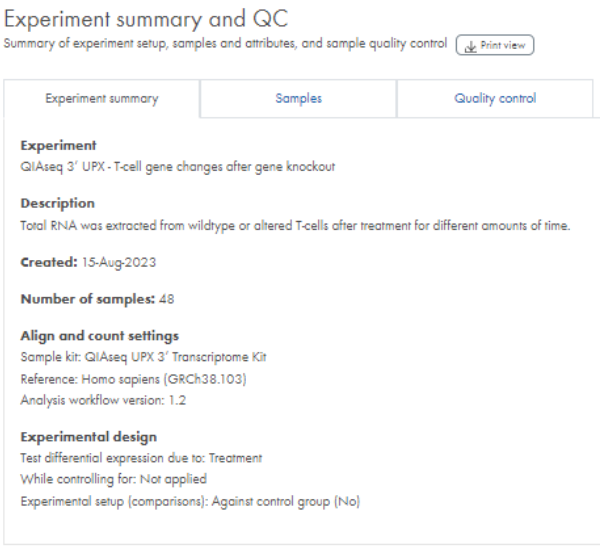


Figure 43: The first tab on the Experiment summary and QC report contains the Experiment summary.

attribute on which the experiment is based, and underlying samples (figure 43).

Samples

The *Samples* tab contains a list of the samples included in the experiment and the specified sample attributes. Alongside this is a Principal Component Analysis (PCA) plot, colored by the attribute on which the experiment is based, indicated above the list (figure 44).

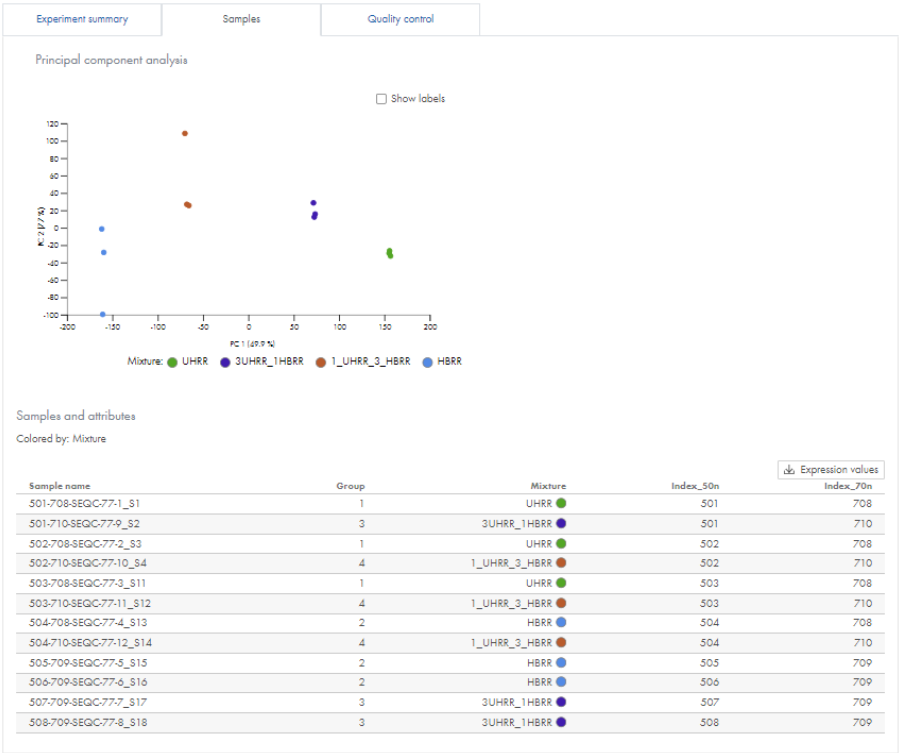


Figure 44: The Samples tab lists samples and their attributes, and allows you to download expression values.

The PCA plot can help with identifying outlying samples for quality control purposes, and can provide you with a feel for the principal causes of variation in the dataset. If you have outlier samples, you may want to run the experiment again without these. If the *Quality control* tab metrics for outlier samples are skewed, this is additional evidence that you should consider removing them from your experiment.

Download PCA plot. To download the PCA plot, hover the plot and click on **PCA plot**. In the dialog, select image resolution, and click on **Download** (figure 45).

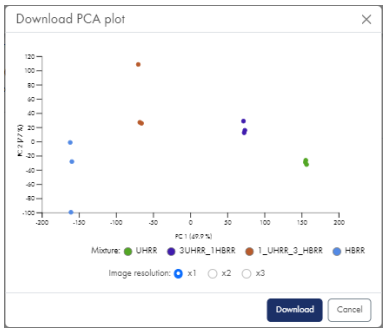


Figure 45: The Download PCA plot dialog with options for image resolution.

Download expression values. To download sample expression values, click on **Expression values** above the table and select the expression values you wish to download. The available expression values differ between sample kits.

miRNA sample kit expression values:

- Mature miRNA values: Name, count, CPM (Counts per million, TMM normalized)
- Mature miRNA and piRNA counts. This option is only available for human, mouse and rat samples.
- IsomiR counts: Name, sequence, count, ambiguity tag (true/false)

RNA sample kit expression values:

- Gene values - Total exon reads, RPKM (Reads Per Kilobase of exon model per Million mapped reads), TPM (Transcripts per million), CPM (Counts per million, TMM normalized)
- Transcript values - Total transcript reads, RPKM (Reads Per Kilobase of exon model per Million mapped reads), TPM (Transcripts per million), CPM (Counts per million, TMM normalized). Transcript values are not supported for 3' protocols such as QIAseq UPX 3' Transcriptome Kit.

RNA sample kit expression value downloads will contain all genes/transcripts of the reference genome, i.e. also biotypes not included in the differential expression analysis. A biotype column specifies the biotype of each feature.

CPM, RPKM, and TPM measures and the tool used to generate IsomiR counts are described in more detail in the QIAGEN CLC Genomics Workbench and Biomedical Genomics Analysis plugin manuals:

- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=_expression_browser.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Extract_IsomiR_Counts.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Expression_settings.html

Quality control

The *Quality control* tab displays several tables with sample quality control metrics from the individual *Align and count* analysis steps. The available metrics differ between sample kits.

The top table, *Quality control summary*, provides an overview of selected metrics like the number of reads, average quality scores, and mapped reads.

QIAseq miRNA Library Kit

The *Quality control* tab contains the two related tables *Unique search sequences* and *Reads*. Unique search sequences are groups of identical reads. The ratio between corresponding *Unique search sequences* values and *Reads* values indicates the redundancy of the particular biotype. In general, a higher level of redundancy is expected for miRNAs than for other biotypes. High redundancy for other biotypes may be indicative of issues with sequencing samples. For additional information on QC metrics, see Appendix E.

RNA sample kits

The *Quality control summary* metrics *Mapped to total rRNA* represents the sum of the corresponding cytoplasmic rRNA and mitochondrial rRNA (Mt_rRNA) values and indicates the level of rRNA contamination. It can be used as a measure of whether the rRNA depletion protocol worked as expected. For additional information on QC metrics, see Appendix E.

The *Biotype distribution* chart and table give insight into where reads mapped (figure 46). For a poly-A enrichment experiment, you expect that the majority of reads correspond to "protein_coding" regions. For other experiment types, a variety of non-coding RNA regions may also be observed. For an rRNA depletion protocol, the percentage of reads mapping to rRNA and Mt_rRNA combined should usually be <15%.

To download the biotype chart, hover the chart and click on **Biotype chart**.

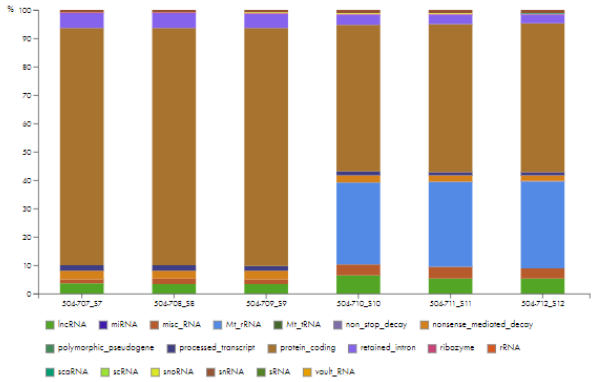


Figure 46: The biotype chart shows the origin of the mapped reads. The large proportion of Mt_rRNA reads for three of the samples is an indication that no rRNA depletion protocol was used, or that if one was used, it did not work as expected.

The sections *Taxonomic profile of unmapped reads* and *Taxonomic profiling summary* will hold taxonomic profiling results for samples with more than 100,000 reads and where less than 75% of the reads could be mapped. A high level of unmapped reads could indicate contamination. The taxonomic profiling of the unmapped reads detects if those unmapped reads may originate from bacteria or archaea, and if so from which phyla. The plot (figure 47)and table under *Taxonomic profile of unmapped reads* show the relative abundance at phylum level.

To download the taxonomic profile chart, hover the chart and click on **Taxonomic profile chart chart**.

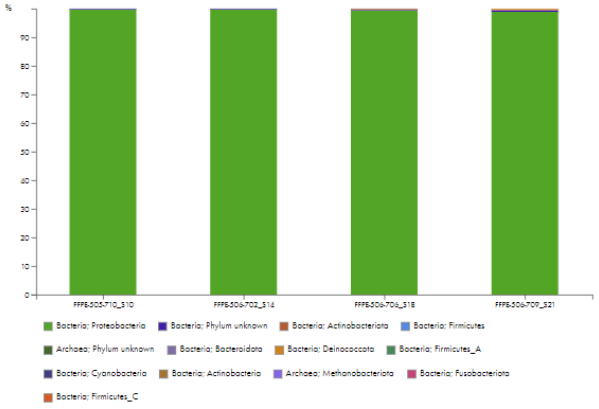


Figure 47: The taxonomic profile chart shows the relative abundance of phyla in unmapped reads of samples where less than 75% of the reads could be mapped.

Differential expression view

The differential expression view is the main result view of *RNA-seq Analysis Portal* (figure 48). It consists of a number of different elements, all linked and interactive.

A differential expression analysis can be for comparisons between two groups of samples or comparisons across all sample groups. For two-group analyses, the name of the differential expression will be [group 1 attribute value] vs. [group 2 attribute value], e.g. "Tumor vs. Normal". For across-groups analyses, the name will reflect the attribute for which the analysis was performed: (attribute): (group 1) vs. (group 2), e.g. "Mixture: UHRR vs HBRR".

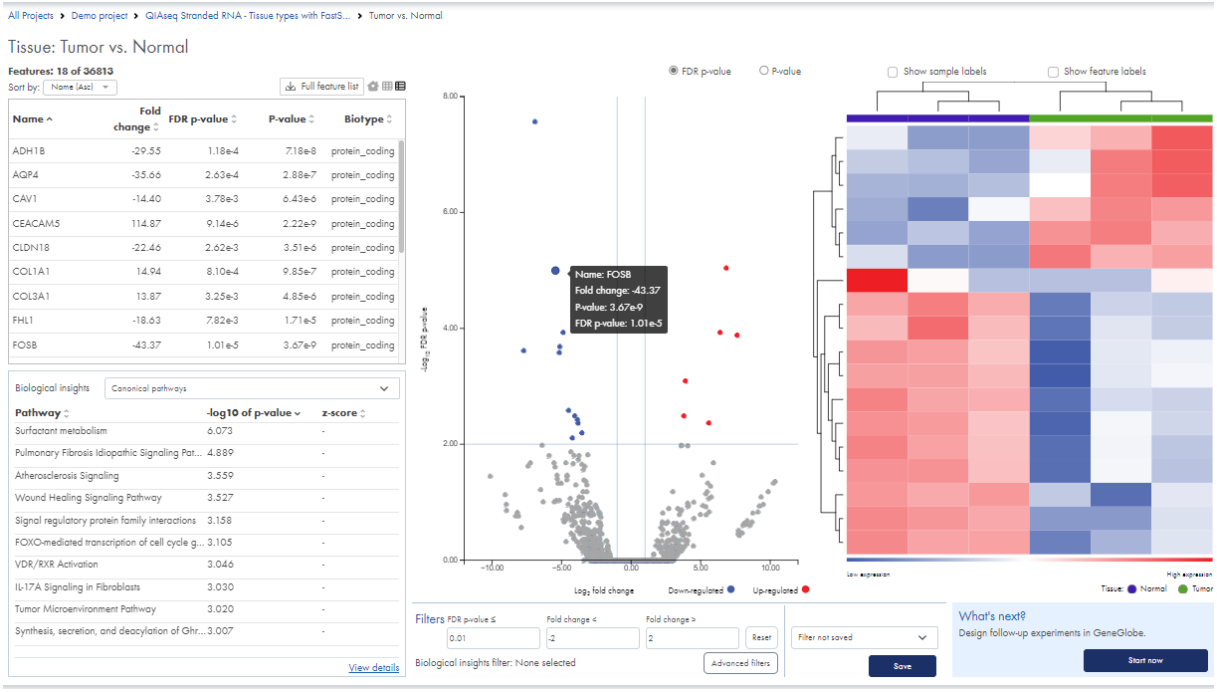


Figure 48: The differential expression view comprises a feature table, a biological insights table, a volcano plot, and a heat map.

Feature table

Using the icons at the top right of the table, you can switch between three modes:

- Features in table view
- Features in grid view
- qPCR normalization genes

For miRNA analyses, the table contains mature miRNAs. For RNA analyses, the table contains protein coding genes and long non-coding RNA (lncRNA, incl. subtypes).

For differential expression analyses between two groups of samples, the fold change indicates the expression levels in group 1 relative to group 2. For example:

- If expression values in group 1 are twice as large as in group 2, the fold change will be +2.
- If expression values in group 2 are twice as large as in group 1, the fold change will be -2.

For differential expression analyses across groups, the fold change represents the maximum pairwise fold change between any two of the groups.

Features in table view and Features in grid view

These views list differentially expressed genes or miRNAs features that meet the applied filter settings described in the Filtering section.

Select one or more features to have these highlighted in the volcano plot and heat map.

Click on **Download full list** above the table to get the full table in .xlsx format.

qPCR normalization genes

This view contains 20 genes or miRNAs that are stably expressed across the conditions analyzed in the differential expression. These genes are candidates for inclusion in follow-up qPCR assays, where they can be used to normalize expression levels for qPCR experiments on the same conditions.

Fold change and minimum CPM (Counts per Million, TMM-adjusted) values for each group are provided.

How this list is generated is described in the section Create experiment analysis workflow.

Select one or more features to have these highlighted in the volcano plot.

When on this view, filtering will be disabled.

For a list of qPCR normalization genes to be available, the underlying samples must be analyzed with sample kit analysis workflows listed as *most recent* for *RNA-seq Analysis Portal* 2.5 or later, see Table 2 in Appendix C.

Volcano plot

The volcano plot shows the relationship between gene/miRNA p-values and fold changes among the samples. The log2 fold changes are plotted on the x-axis, and the -log10 p-values are plotted on the y-axis. You toggle between regular p-values and FDR p-values via the selection on the top

right. Features of interest are typically those in the upper left and upper right hand corners of the volcano plot, as these have large fold changes (lie far from x = 0) and are statistically significant (have large y-values).

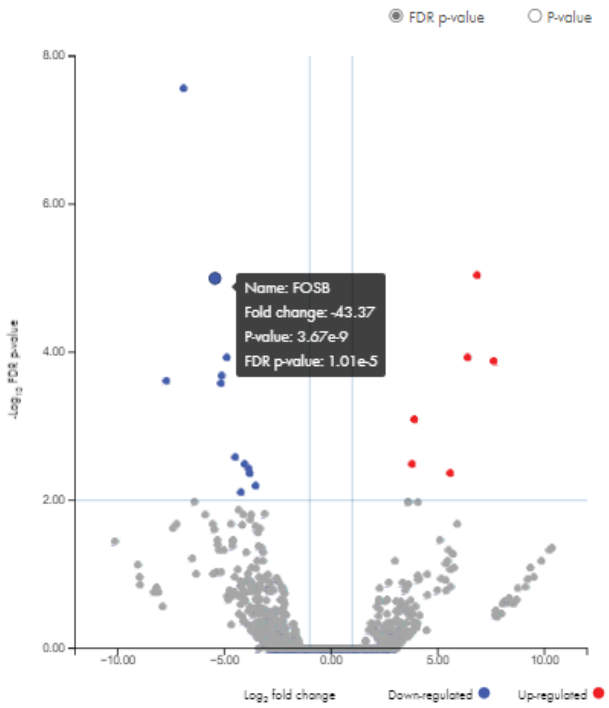


Figure 49: Hover over features in the volcano plot to see names and values.

Features that meet the applied filter settings will be colored blue if down-regulated or red if up-regulated in group 1 relative to group 2.

Hover over the individual features to see name, fold change and p-value (figure 49).

The vertical and horizontal bars represent the filtering settings. These can be dragged to adjust the thresholds. The values in the filtering fields below will change accordingly.

Download volcano plot. To download the volcano plot, hover the plot and click on **Volcano plot**.

Heat map

The two dimensional heat map shows expression levels. Each column corresponds to one sample, and each row corresponds to a feature (gene or miRNA). The samples and features are both hierarchically clustered. The attribute values on which the analysis was performed are added as an overlay (figure 50).

The coloring of the heat map fields indicate expression values of the individual features. Blue colors signify low expression and red colors signify high expression. Hover over an field to see

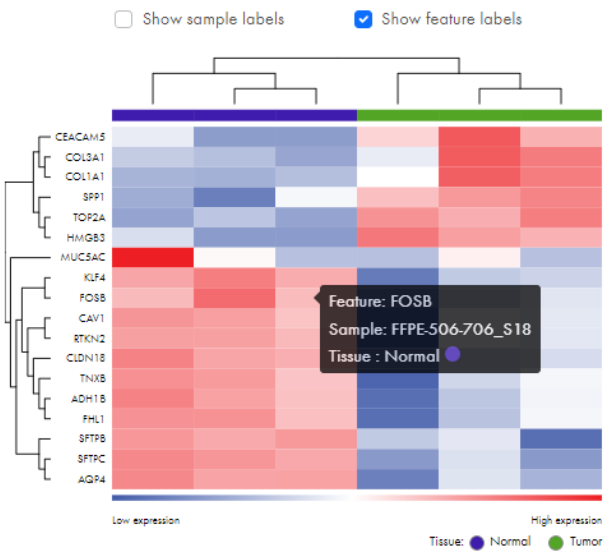


Figure 50: Hover over fields in the heat map to see feature and sample names.

the feature and sample name.

Download heatmap. To download the heatmap, hover the plot and click on **Heatmap**.

Biological insights table

The biological insights table at the bottom left is built using QIAGEN Ingenuity Pathway (IPA) results for your list of differentially expressed genes or miRNAs. The drop-down contains two to four categories, depending on the sample kit (figure 51):

- *Canonical pathways*
- *Upstream regulators* (RNA sample kits only)
- *Upstream regulators (miRNA)* (RNA sample kits only)
- *Diseases and functions*

Each tab holds up to ten entries.

The information in this table helps with identifying the most significant pathways, whether activated or inhibited, as well as potential upstream regulators and causal relationships associated with your data.

For each table entry, the -log10 p-value and z-score are listed.

Click on **View details** to get an expanded biological insights table with additional information and values (figure 52). Click on **Learn about these values**, at the top right side, to go to QIAGEN

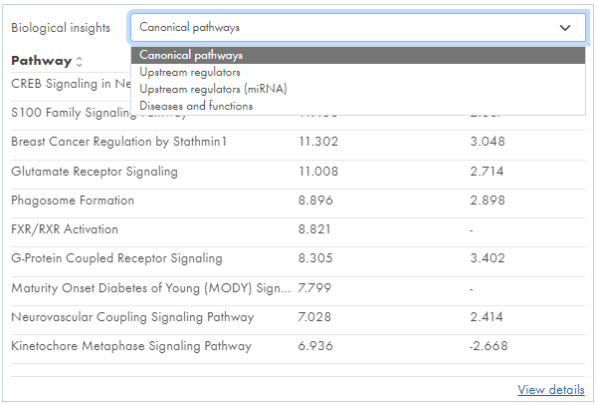


Figure 51: The biological insights table reports QIAGEN Ingenuity Pathway Analysis results in three or four categories.

Ingenuity Pathway Analysis documentation, where information on the statistical calculations behind the values is provided.

Note that for any particular biological insights result, a p-value will always be present, but a z-score may not be. This is expected and is due to how the pathways and biological relationships are constructed.

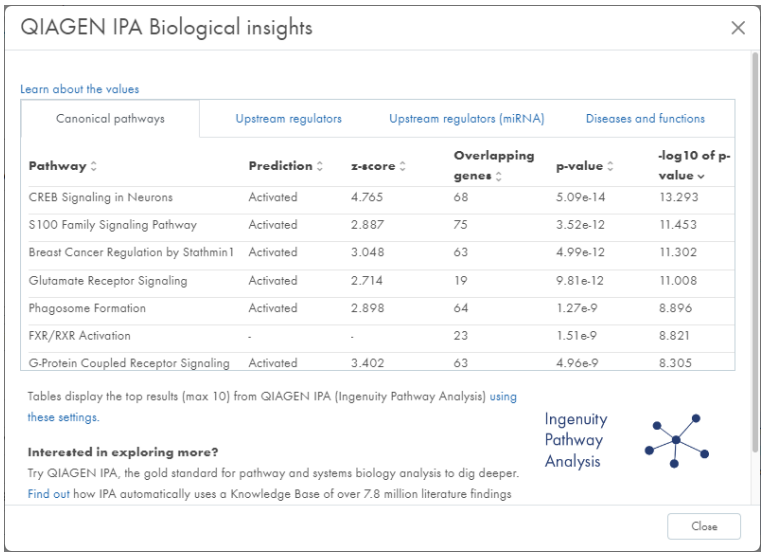


Figure 52: The expanded biological insights table provides additional information.

On the smaller biological insights table, you can click an item to apply this as a biological insights filter. This limits your feature table and heat map contents and volcano plot coloring to only genes or miRNAs associated with the particular item, e.g. a Canonical Pathway. Other filter settings still apply (figure 53). Click on the same item again to remove the biological insights filter.

Figure 53: Click on a biological insight table item to limit entries in the feature table and heat map, and coloring on the volcano plot to genes or miRNAs associated with the particular item.

Filter

Underneath the volcano plot you find the **Filter** section (figure 54). You can filter feature table, volcano plot and heat map based on minimum and maximum fold changes, and on regular p-values or FDR p-values depending on the volcano plot p-value selection. You can enter or paste values in the field to adjust thresholds, or you can use the horizontal and vertical bars on the volcano plot to do this.

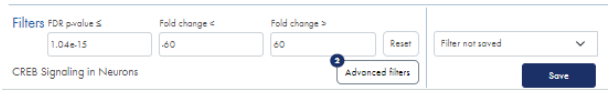


Figure 54: Adjust fold change and p-value, apply advanced filtering and select a biological insights filter to narrow down the list of genes or miRNAs.

Additional filtering options are available from the **Advanced filters** button. In the resulting dialog, you can filter on biotypes (when available), minimum, mean and maximum CPM ((Counts per Million, TMM-adjusted) values for each sample group, and minimum, mean and maximum CPM values across groups (figure 55). You can apply multiple filters, one at a time:

1. Select criterion
2. Select operator and enter value, or select one or more biotypes
3. Click **Add filter**

Figure 55: With Advanced filters, you can filter on biotypes and CPM expression values.

Each advanced filter is added as a filter tag below the dialog table and can be removed again by clicking the 'x'.

Once you have added the desired filters, click **Apply filter criteria to feature list** to close the dialog.

An badge next to the **Advanced filters** button indicates the number of applied advanced filters (figure 54).

Text below the regular filter fields signals whether a biological insights filter has been applied.

Click on **Save** to save the filter including biological insights and advanced filtering.

Click on **Reset** to reset filtering, or choose one of your previously saved filters from the drop-down list on the right.

P-value and FDR p-value

- P-value. Standard, uncorrected p-value. Genes/transcripts that are not observed in any sample have undefined p-values and are reported as '-'.
- FDR p-value. The false discovery rate corrected p-value. The FDR-corrected p-value will always be larger than the uncorrected p-value.
  The differential expression analysis includes the step *Filter on average expression for FDR correction*. This filters away some genes prior to the FDR correction. As a result, those genes will have undefined FDR p-values, reported as '-'.

The p-value controls the chance of getting a false positive result. When you apply a statistical test and use a p-value cut-off of 0.05, you should expect 5% of your significant results to be false positives.

When you carry out many tests, e.g. test differential expression for many genes in one experiment, you run into the multiple testing problem: If in a multiple-test scenario 10000 tests turn out

significant and for each of these you use a 0.05 p-value cut-off, in total you should expect 500 of those significant results to be false positives. This approach may not be useful in practice.

The False Discovery Rate (FDR) p-value addresses this problem and allows you to control the overall false positive rate in the multiple testing scenario. In this approach it is not the chance of making a false positive call in each individual test that is being controlled, but rather the proportion of false positive tests among all significant test. When you use a FDR-corrected p-value cut-off of 0.05, you should expect 5% of your significant tests to be false positives. Mathematically, the FDR-corrected p-values are obtained from examining the distribution of the traditional p-values across all tests performed, and identifying cut-offs between significant and non-significant values.

Typically, the standard p-value is used in cases where only a few tests are being performed in parallel, whereas the FDR corrected p-value is used when many tests are being performed. Analyzing differential expression across many genes or miRNAs is a example of the latter, i.e. for this use case, the FDR p-value is normally recommended. Using the FDR p-value will result in fewer false positive calls, without much loss of sensitivity.

Send results to GeneGlobe

When logged in with a My QIAGEN account, you can upload the filter results to GeneGlobe *My Projects* by clicking on **Start now** (figure 56). This sends the filtered list of features and the full lists of Canonical Pathways, and - when applicable - Upstream regulator (miRNA) to GeneGlobe *My Projects*.

GeneGlobe *My Projects* opens up in a new tab. Depending on your browser settings, you may need to first allow pop-ups.

Within the same differential expression, you can send multiple saved filter results to GeneGlobe *My Projects*. This can be useful if, for instance, you wish to explore several lists of features associated with different Canonical Pathways.

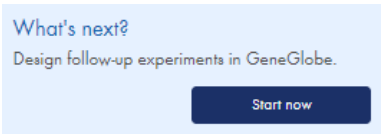


Figure 56: Click on What's next to send results to GeneGlobe My Projects.

Send results to QIAGEN IPA

When logged in with a QIAGEN IPA account, *RNA-seq Analysis Portal* integrates with the projects

in your QIAGEN IPA account. When you click on **Send** (figure 57), the filtered list of features is sent to your QIAGEN IPA account.

Within the same differential expression, you can send multiple filter results to your QIAGEN IPA account.

Figure 57: Click on Send to QIAGEN IPA to send results to your QIAGEN IPA account.

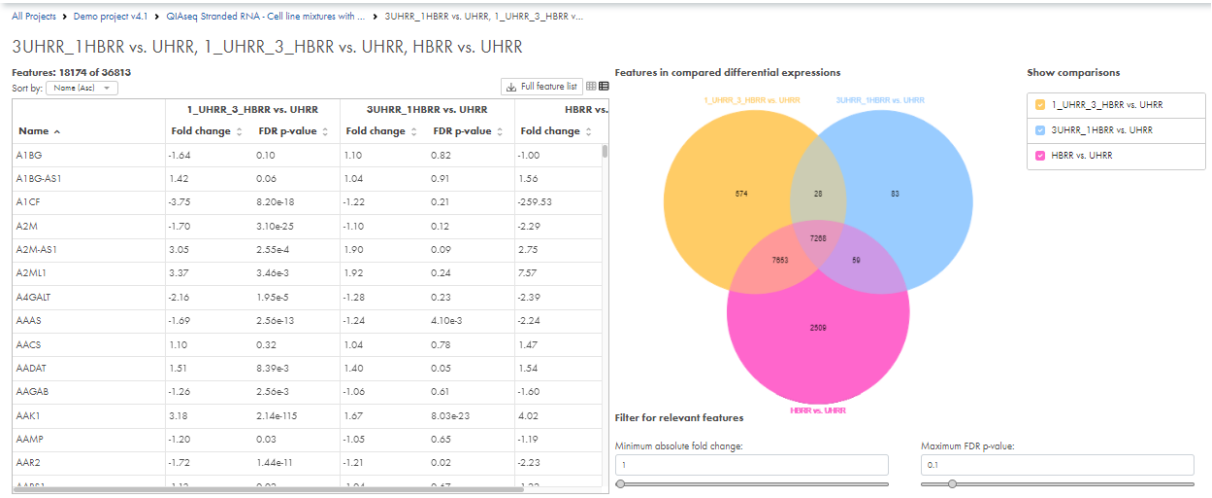
Comparison view

The *Comparison view* contains a table and a Venn diagram (figure 58). These contain the same features and are linked. Using the fields on the right, you can adjust the FDR p-value and fold change thresholds and thereby include more or fewer features.

The Venn diagram demonstrates the overlap of differentially expressed features in two or three differential expression analyses. The number of unique and shared features are given by the values on the plot. Click on the plot to select certain areas and have the corresponding features highlighted in the table.

Download Venn diagram. To download the Venn diagram, hover the diagram and click on **Venn diagram**.

The table holds features with FDR p-values and fold change values from the differential expression analyses that were compared. Click on **Download full list**, above the table, to get the full table in .xlsx format.



Name	1_UHRR_3_HBRR vs. UHRR		3UHRR_1HBRR vs. UHRR		HBRR vs.
	Fold change	FDR p-value	Fold change	FDR p-value	Fold change
A1BG	-1.64	0.10	1.10	0.82	-1.00
A1BG-AS1	1.42	0.06	1.04	0.91	1.56
A1CF	-3.75	8.20e-18	-1.22	0.21	-259.53
A2M	-1.70	3.10e-25	-1.10	0.12	-2.29
A2M-AS1	3.05	2.55e-4	1.90	0.09	2.75
A2ML1	3.37	3.46e-3	1.92	0.24	7.57
A4GALT	-2.16	1.95e-5	-1.28	0.23	-2.39
AAAS	-1.69	2.56e-13	-1.24	4.10e-3	-2.24
AACS	1.10	0.32	1.04	0.78	1.47
AADAT	1.51	8.39e-3	1.40	0.05	1.54
AAGAB	-1.26	2.56e-3	-1.06	0.61	-1.60
AAK1	3.18	2.14e-115	1.67	8.03e-23	4.02
AAMP	-1.20	0.03	-1.05	0.65	-1.19
AAR2	-1.72	1.44e-11	-1.21	0.02	-2.23

Figure 58: The Venn diagram shows the overlap of differentially expressed genes or miRNAs in different differential expression analyses.

RNA-seq Analysis Portal analysis explained

RNA-seq Analysis Portal facilitates analysis of FASTQ or BAM data generated on Illumina® or Themo Fisher® instruments for the appropriate sample kits listed under Supported sample kits.

This chapter describes the secondary analysis, which for all sample kit types consists of two steps handled by separate analysis workflows:

1. Align and count - alignment and counting of sequencing reads. Outputs are samples containing expression counts.
2. Create experiment - calculation of differential expression between groups of samples. Outputs are one or more differential expression analyses.

The details of the two analysis workflows differ for different sample kits. Details of these differences are described in the sections below.

In addition, this chapter covers Taxonomic profiling of unmapped reads and Comparison of differential expression analyses.

Align and count analysis workflows

The supported sample kits fall into four groups based on the *Align and count* analysis workflows:

- QIAseq miRNA Library Kit
- QIAseq UPX 3' Transcriptome Kit
- QIAseq UPXome RNA Lib Kit
- Demultiplexed RNA sample kits (remaining sample kits)

Align and count - QIAseq miRNA Library Kit

The analysis workflow is based on the Biomedical Genomics Analysis *QIAseq miRNA Quantification*

analysis workflow described at

https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=QIAseq_miRNA_Quantification.html

It includes the steps below. Specific parameters are listed in Appendix C.

- **Create UMI Reads for miRNA.** Sequencing reads are merged to UMI reads based on the unique molecular indexes (UMIs).
- **Quantify miRNA.** miRNAs are counted and annotated using miRBase and custom databases. Expression levels are generated for the mature miRNAs.

The tools are described in more detail in the QIAGEN CLC Genomics Workbench and Biomedical Genomics Analysis plugin manuals:

- https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=Create_UMI_Reads_miRNA.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Quantify_miRNA.html

Align and count - QIAseq UPX 3' Transcriptome Kit

The analysis workflow is based on the Biomedical Genomics Analysis QIAseq *Quantify QIAseq UPX 3'* analysis workflow described at

https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=_Quantify_QIAseq_UPX_3_ready_to_use_workflow.html

It includes the steps below. Specific parameters are listed in Appendix C.

- **Demultiplex Reads.** Reads are grouped into samples based on the sample-specific adapters.
- **Remove and Annotate with Unique Molecular Index.** The UMI and common sequence prefixes are removed from the reads. Instead, a UMI annotation is added to each read to retain fragment identity.
- **Trim Reads.** Short, internal polyA sequences are removed from R1 reads, and the short R2 reads are discarded.
- **Trim Reads.** PolyA and polyG sequences are removed from R1.

- **Create UMI Reads from Reads.** Sequencing reads are merged to UMI reads based on the unique molecular indexes (UMIs).
- **Trim Reads.** Low quality nucleotides, ambiguous nucleotides, and adapter sequences are removed.
- **RNA-Seq Analysis.** UMI reads are mapped to transcripts, and expression values for each gene are obtained by summing the transcript counts belonging to the gene.

The tools are described in more detail in the QIAGEN CLC Genomics Workbench and Biomedical Genomics Analysis plugin manuals:

- https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=Demultiplex_QIAseq_UPX_3_reads.html
- https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=Remove_Annotate_with_Unique_Molecular_CLCIndex.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Trim_Reads.html
- https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2101/index.php?manual=Create_UMI_Reads_from_Reads.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=RNA_Seq_Analysis.html

Align and count - QIAseq UPXome RNA Lib Kit

RNA-seq Analysis Portal supports three cDNA synthesis protocols: N6-T RT primer, ODT-T RT primer, or combined N6-T RT and ODT-T RT primers. The analysis workflows are based on the corresponding Biomedical Genomics Analysis QIAseq template workflows described under https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2220/index.php?manual=UPXome_RNA.html

N6-T RT primer and ODT-T RT primer protocols

For the N6-T RT primer and ODT-T RT primer protocols, the analysis workflows include the following steps. Specific parameters are listed in Appendix C.

- **Demultiplex Reads.** Reads are grouped into samples based on the sample-specific adapters.

- **Trim Reads.** Low quality, ambiguous nucleotides and adapter sequences are removed.
- **RNA-Seq Analysis.** Reads are mapped to transcripts, and expression values for each gene are obtained by summing the transcript counts belonging to the gene.

N6-T RT + ODT-T RT primer protocol

For the combined N6-T RT and ODT-T RT primer protocol, the analysis workflow includes two sections, A and B. The steps involved in each section are listed below. Specific parameters are listed in Appendix C.

Analysis workflow section A:

- **Demultiplex Reads.** Reads are grouped into samples based on the sample-specific adapters.
- **Trim Reads.** Adapter sequences are removed.

Analysis workflow section B:

Section B is split in two subsections, B1 and B2. B1 creates gene level expression results using all reads as input (N6-T and ODT-T reads combined). B2 generates transcript level expression values from N6-T reads only. Both subsections contain the following steps:

- **Trim Reads.** Low quality, ambiguous nucleotides are removed.
- **RNA-Seq Analysis.** Reads are mapped to transcripts. For the gene expression, expression values for each gene are obtained by summing the transcript counts belonging to the gene.

Results from the B1 subsection are used for subsequent differential gene expression analysis and as input for the Experiment summary and QC report Quality control tab. The Quality control tab is split into two sections and shows statistics for all reads in one section and the N6 reads in a separate section.

The tools are described in more detail in the QIAGEN CLC Genomics Workbench and Biomedical Genomics Analysis plugin manuals:

- `https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/2202/index.php?manual=Detect_Wells.html`
- `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Trim_Reads.html`

- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=RNA_Seq_Analysis.html

Align and count - Demultiplexed RNA sample kits

This *Align and count* analysis workflow is used for the remaining sample kits. These do not leverage unique molecular indexes. Sample data must be demultiplexed prior to being uploaded to *RNA-seq Analysis Portal*.

Sample kits

- QIAseq Stranded mRNA Select/Stranded Total RNA Lib Kit (QIAGEN)
- QIAseq FastSelect RNA Lib Kit (QIAGEN)
- QIAseq UPX 3' Transcriptome Kit (with already-demultiplexed input data selected)
- QIAseq UPXome RNA Lib Kit (with already-demultiplexed input data selected)
- TruSeq Stranded Total RNA Library Prep (Human/Rat, Gold, Globin)(Illumina)
- Illumina Stranded Total RNA Prep with Ribo-Zero Plus (Illumina)
- NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs)
- KAPA RNA HyperPrep Kit (Roche Sequencing solutions)
- SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian/Low Input Mammalian (Takara Bio)
- Collibri Stranded RNA Library Prep Kit for Illumina Systems (Thermo Fisher Scientific)

The analysis workflow includes the steps below. Parameters for the individual sample kits are listed in Appendix C.

- **Trim Reads.** Low quality and ambiguous nucleotides and adapter sequences are removed.
- **RNA-Seq Analysis.** Reads are mapped to transcripts, and expression values for each gene are obtained by summing the transcript counts belonging to the gene.

The tools are described in more detail in the QIAGEN CLC Genomics Workbench manual:

- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Trim_Reads.html

- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=RNA_Seq_Analysis.html

Create experiment analysis workflow

The differential expression analysis report summarizes the analysis results from the following tools. Specific parameters for the individual sample kits are listed in Appendix C.

- **Differential Expression for RNA-Seq.** Performs a differential expression test using multi-factorial statistics based on a negative binomial GLM. The tool supports paired designs and can control for batch effects.
- **Create Heat Map for RNA-Seq.** Samples and features are hierarchically clustered to create a two dimensional heat map of expression values. Each column contains data for a particular sample, and each row contains data for a particular feature.
  The following filtering and normalization is carried out when producing the heat map:
  - Log CPM (Counts per Million) values are calculated for each feature. The CPM calculation uses the effective library sizes as calculated by the TMM normalization.
  - A Z-score normalization is performed across samples for each gene: the counts for each gene are mean centered, and scaled to unit variance.
  - Genes or miRNAs with zero expression across all samples or invalid values (NaN or +/-Infinity) are removed.

The tools are described in more detail in the QIAGEN CLC Genomics Workbench manual:

- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Differential_Expression_RNA_Seq.html
- https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Create_Heat_Map_RNA_Seq.html

IPA submission and analysis. Prior to being passed on to IPA, the list of features is marked to indicate which to consider for the IPA analysis. The approach is as follows:

1. Select features with an FDR p-value <0.05. If this leaves more than 1000 features, then:
2. Select the 1000 features with the largest absolute fold change, keeping the ratio of up- and down-regulated features.

The uploaded features are used as input for an IPA Core Analysis.

Identification of qPCR normalization genes. The following filtering and selection is applied to identify stably expressed genes that can be used for normalization is e.g. qPCR assays.

1. Filter features:
    - Absolute fold change <1.03
    - FDR p-value >0.05
    - Minimum CPM (Counts per Million, TMM-adjusted) for each group >25
2. Select the 20 features with smallest absolute fold change

Taxonomic profiling of unmapped reads

For RNA samples with more than 100,000 reads and where less than 75% of the reads map, an additional taxonomic profiling analysis is performed with the aim of detecting potential contamination from bacteria and archaea.

The analysis uses the tool *Taxonomic Profiling*. This takes as input the reads that go unmapped in the Align and count analysis and maps these to a reference database of complete archaea and bacteria genomes. If a read is found to map to multiple genomes in the reference database, it will be assigned to the lowest common ancestor.

Parameter settings and reference database for the Taxonomic Profiling analysis are independent of sample kit and reference:

Taxonomic Profililng	
Reference index	*QMI-PTDB Genus (v2.0)*
Filter host reads	No
Auto-detect paired distances	Yes
Minumum seed length	30
Adjust read count abundances	Yes

The Taxonomic Profiling tool is described in more detail in the CLC Microbial Genomics Module manual: `https://resources.qiagenbioinformatics.com/manuals/clcmgm/2200/index.php?manual=Taxonomic_Profiling.html`.

Taxonomic profiling reference index. The reference index *QMI-PTDB Genus (v2.0)* database was obtained using the tool *Download Curated Microbial Reference Database*. In the previous version the database (QMI-PTDB - Approx. 22GB (Jan2022)) was enriched with six additional species

Taxonomic level	Kingdom	Phylum	Class	Order	Family	Genus	Species
Classifications	2	64	140	374	872	4381	7794

Table 1: Taxonomic summary of the *QMI-PTDB Genus (v2.0)* **taxonomic profiling reference index.**

(Mycoplasma hominis, Mycoplasma fermentans, Mycoplasma salivarium, Mycoplasma arginini, Mycoplasma orale, Escherichia coli O157:H7). These are now integrated in the database off the shelf. The index covers a total of 8626 species. See Table 1 for a taxonomic summary.

The tools used for downloading curated and custom databases and for creating taxonomic indexes are described in the CLC Microbial Genomics Module manual at `https://resources.qiagenbioinformatics.com/manuals/clcmgm/2200/index.php?manual=Databases_Taxonomic_Analysis.html`.

Compare analyses

The *Compare analyses* functionality of *RNA-seq Analysis Portal* uses the tool *Create Venn Diagram for RNA-Seq*. This is described in the QIAGEN CLC Genomics Workbench manual: `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/2103/index.php?manual=Create_Venn_Diagram_RNA_Seq.html`.

Appendix A - Sequence file naming pattern

RNA-seq Analysis Portal supports analysis of FASTQ files. To be appropriately grouped into samples during upload, the sequence files names must adhere to the standard Illumina file naming. If you have FASTQ files from e.g. a Thermo Fisher Ion Torrent sequencer, you can manually rename them to fit your grouping needs:

```
<sample-name>_S<sample_number>_L<lane (0-padded to 3 digits)>_R<read number>_001.fastq(.gz)
```

Example of single lane, paired end data set consisting of two FASTQ files that will be grouped into one sample, TC-35-A_S1:

```
TC-35-A_S1_L001_R1_001.fastq.gz
TC-35-A_S1_L001_R2_001.fastq.gz
```

Example of a four lane, paired end data set consisting of eight FASTQ file that will be grouped into one sample, 501-708-SEQC-77-1_S1:

```
501-708-SEQC-77-1_S1_L001_R1_001.fastq.gz
501-708-SEQC-77-1_S1_L001_R2_001.fastq.gz
501-708-SEQC-77-1_S1_L002_R1_001.fastq.gz
501-708-SEQC-77-1_S1_L002_R2_001.fastq.gz
501-708-SEQC-77-1_S1_L003_R1_001.fastq.gz
501-708-SEQC-77-1_S1_L003_R2_001.fastq.gz
501-708-SEQC-77-1_S1_L004_R1_001.fastq.gz
501-708-SEQC-77-1_S1_L004_R2_001.fastq.gz
```

For single lane, single read data, as what may you see from the QIAseq miRNA Library Kit, each sample is made up from one FASTQ file. The following two FASTQ files are converted into the two samples, QL4_S4 and QL6_S6.

```
QL4_S4_L001_R1_001.fastq
QL6_S6_L001_R1_001.fastq
```

On the *RNA-seq Analysis Portal sequencing data uploader* tab, once you have selected your FASTQ files, the **Files to samples** section on the right provides a sample grouping preview (figure 59). If

grouping is not as expected this is likely due to the FASTQ file names not following the expected format.

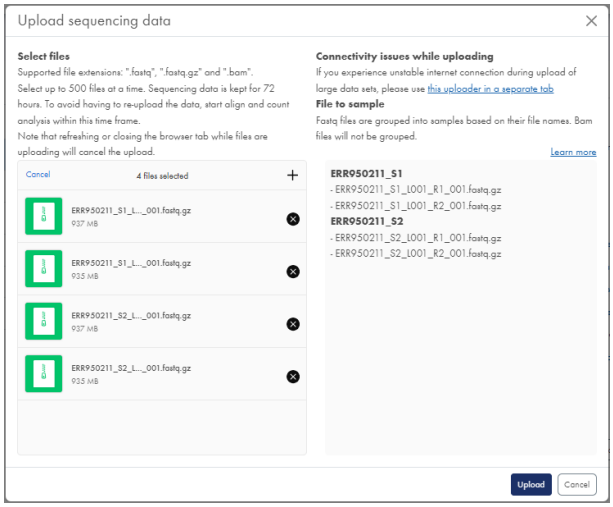


Figure 59: The Files to samples preview provides a check of sample grouping.

Appendix B - Reference and annotation information

This appendix holds a list of the supported species incl. source of reference sequence and gene annotations, and a section on reference resources relevant to specific sample kit analysis workflows. When a reference is updated, the older version will be marked with (Legacy) in the selection menu.

General references and annotations

- Arabidopsis thaliana
- Bos taurus
- Caenorhabditis elegans
- Canis lupus familiaris
- Danio rerio
- Drosophila melanogaster
- Equus caballus
- Gallus gallus (GRCg6a.105) (legacy)
- Gallus gallus (GRCg7b.110)
- Gasterosteus aculeatus
- Homo sapiens (GRCh38.103)
- Homo sapiens (GRCh38.noalt.106)
- Macaca fascicularis
- Macaca mulatta
- Mus musculus

- Oncorhynchus mykiss
- Oryctolagus cuniculus
- Rattus norvegicus (Rnor6.0.103) (legacy)
- Rattus norvegicus (mRatBN7.2.110)
- Saccharomyces cerevisiae
- Schizosaccharomyces pombe
- Sus scrofa
- Zea mays

Arabidopsis thaliana (TAIR10.52)

Common name	Thale cress
Assembly	TAIR10
Ensembl release	52
Reference sequence	`http://ftp.ensemblgenomes.org/pub/plants/release-52/fasta/arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.dna.toplevel.fa.gz`
Ensembl annotations	`http://ftp.ensemblgenomes.org/pub/plants/release-52/gff3/arabidopsis_thaliana/Arabidopsis_thaliana.TAIR10.52.gff3.gz`

Bos taurus (ARS-UCD1.2.105)

Common name	Cow
Assembly	ARS-UCD1.2
Ensembl release	105
Reference sequence	`http://ftp.ensembl.org/pub/release-105/fasta/bos_taurus/dna/Bos_taurus.ARS-UCD1.2.dna.primary_assembly.*.fa.gz` (*: 1-29, X, MT)
Ensembl annotations	`http://ftp.ensembl.org/pub/release-105/gff3/bos_taurus/Bos_taurus.ARS-UCD1.2.105.chr.gff3.gz`

Caenorhabditis elegans (WBcel235.105)

Common name	Roundworm
Assembly	WBcel235
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/caenorhabditis_elegans/dna/Caenorhabditis_elegans.WBcel235.dna_sm.toplevel.fa.gz
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/caenorhabditis_elegans/Caenorhabditis_elegans.WBcel235.105.gff3.gz

Canis lupus familiaris (ROS_Cfam_1.0.105)

Common name	Dog (labrador retriever breed)
Assembly	ROS_Cfam_1.0
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/canis_lupus_familiaris/dna/Canis_lupus_familiaris.ROS_Cfam_1.0.dna.primary_assembly.*.fa.gz (*: 1-38, X, Y)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/canis_lupus_familiaris/Canis_lupus_familiaris.ROS_Cfam_1.0.105.chr.gff3.gz

Danio rerio (GRCz11.10.105)

Common name	Zebra fish
Assembly	GRCz11.10
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/danio_rerio/dna/Danio_rerio.GRCz11.dna.chromosome.*.fa.gz (*: 1-25, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/danio_rerio/Danio_rerio.GRCz11.105.chr.gff3.gz

Drosophila melanogaster (BDGP6.32.105)

Common name	Fruit fly
Assembly	BDGP6.32
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/drosophila_melanogaster/dna/Drosophila_melanogaster.BDGP6.32.dna.primary_assembly.*.fa.gz (*: 2L, 2R, 3L, 3R, 4, X, Y, mitochondrion_genome)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/drosophila_melanogaster/Drosophila_melanogaster.BDGP6.32.105.chr.gff3.gz

Equus caballus (EquCab3.0.105)

Common name	Horse
Assembly	EquCab3.0
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/equus_caballus/dna/Equus_caballus.EquCab3.0.dna.primary_assembly.*.fa.gz (*: 1-31, X)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/equus_caballus/Equus_caballus.EquCab3.0.105.chr.gff3.gz

Gallus gallus (GRCg6a.105) (legacy)

Common name	Chicken
Assembly	GRCg6a
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/gallus_gallus/dna/Gallus_gallus.GRCg6a.dna.chromosome.*.fa.gz (*: 1-33, W, Z, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/gallus_gallus/Gallus_gallus.GRCg6a.105.chr.gff3.gz

Gallus gallus (GRCg7b.110)

Common name	Chicken
Assembly	GRCg7b
Ensembl release	110
Reference sequence	http://ftp.ensembl.org/pub/release-110/fasta/gallus_gallus/dna/Gallus_gallus.bGalGal1.mat.broiler.GRCg7b.dna.primary_assembly.*.fa.gz (*: 1-33, W, Z, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-110/gff3/gallus_gallus/Gallus_gallus.bGalGal1.mat.broiler.GRCg7b.110.gff3.gz

Gasterosteus aculeatus (GAculeatus_UGA_version5.112)

Common name	Three-spined stickleback
Assembly	GAculeatus_UGA_version5
Ensembl release	112
Reference sequence	https://ftp.ensembl.org/pub/release-112/fasta/gasterosteus_aculeatus/dna/Gasterosteus_aculeatus.GAculeatus_UGA_version5.dna.primary_assembly.*.fa.gz (*: I, II, III, IV, IX, V, VI, VII, VIII, X, XI, XII, XIII, XIV, XIX, XV, XVI, XVI
Ensembl annotations	https://ftp.ensembl.org/pub/release-112/gff3/gasterosteus_aculeatus/Gasterosteus_aculeatus.GAculeatus_UGA_version5.112.gff3.gz

Gasterosteus aculeatus (GAculeatus_UGA_version5.112)

Homo sapiens (GRCh38.103)

Common name	Human
Assembly	GRCh38.p13
Ensembl release	103
Reference sequence	http://ftp.ensembl.org/pub/release-103/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.chromosome.*.fa.gz (*: 1-22, MT, X, Y)
Ensembl annotations	http://ftp.ensembl.org/pub/release-103/gff3/homo_sapiens/Homo_sapiens.GRCh38.103.chr.gff3.gz

Homo sapiens (GRCh38.noalt.106)

Reference sequence includes unlocalized scaffolds and unplaced scaffolds.

Common name	Human
Assembly	GRCh38
Ensembl release	106
Reference sequence	`https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz`
Ensembl annotations	`http://ftp.ensembl.org/pub/release-106/gff3/homo_sapiens/Homo_sapiens.GRCh38.106.chr.gff3.gz`

Macaca fascicularis (Macaca_fascicularis_6.0.105)

Common name	Crab-eating macaque
Assembly	Macaca_fascicularis_6.0
Ensembl release	105
Reference sequence	`http://ftp.ensembl.org/pub/release-105/fasta/macaca_fascicularis/dna/Macaca_fascicularis.Macaca_fascicularis_6.0.dna.primary_assembly.*.fa.gz` (*: 1-20, X)
Ensembl annotations	`http://ftp.ensembl.org/pub/release-105/gff3/macaca_fascicularis/Macaca_fascicularis.Macaca_fascicularis_6.0.105.chr.gff3.gz`

Macaca mulatta (Mmul_10.105)

Common name	Rhesus macaque
Assembly	Macaca_mulatta.Mmul_10
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/macaca_mulatta/dna/Macaca_mulatta.Mmul_10.dna.primary_assembly.*.fa.gz (*: 1-20, X, Y, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/macaca_mulatta/Macaca_mulatta.Mmul_10.105.chr.gff3.gz

Mus musculus (GRCm38.101)

Common name	Mouse
Assembly	GRCm38.p6
Ensembl release	101
Reference sequence	http://ftp.ensembl.org/pub/release-101/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.chromosome.*.fa.gz (*: 1-19, MT, X, Y)
Ensembl annotations	http://ftp.ensembl.org/pub/release-101/gff3/mus_musculus/Mus_musculus.GRCm38.101.chr.gff3.gz

Oncorhynchus mykiss (Omyk_1.0.105)

Common name	Rainbow trout
Assembly	Omyk_1.0
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/oncorhynchus_mykiss/dna/Oncorhynchus_mykiss.Omyk_1.0.dna.primary_assembly.*.fa.gz (*: 1-29)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/oncorhynchus_mykiss/Oncorhynchus_mykiss.Omyk_1.0.105.chr.gff3.gz

Oryctolagus cuniculus (OryCun2.0.105)

Common name	Rabbit
Assembly	OryCun2.0
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/oncorhynchus_mykiss/dna/Oncorhynchus_mykiss.Omyk_1.0.dna.primary_assembly.*.fa.gz (*: 1-21, X, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/oryctolagus_cuniculus/Oryctolagus_cuniculus.OryCun2.0.105.chr.gff3.gz

Rattus norvegicus (Rnor6.0.103) (legacy)

Common name	Rat
Assembly	Rnor_6.0
Ensembl release	103
Reference sequence	http://ftp.ensembl.org/pub/release-103/fasta/rattus_norvegicus/dna/Rattus_norvegicus.Rnor_6.0.dna.chromosome.*.fa.gz (*: 1-20, MT, X, Y)
Ensembl annotations	http://ftp.ensembl.org/pub/release-103/gff3/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.103.chr.gff3.gz

Rattus norvegicus (mRatBN7.2.110)

Common name	Rat
Assembly	mRatBN_7.2
Ensembl release	110
Reference sequence	https://ftp.ensembl.org/pub/release-110/fasta/rattus_norvegicus/dna/Rattus_norvegicus.mRatBN7.2.dna.primary_assembly.*.fa.gz (*: 1-20, MT, X, Y)
Ensembl annotations	https://ftp.ensembl.org/pub/release-110/gff3/rattus_norvegicus/Rattus_norvegicus.mRatBN7.2.110.chr.gff3.gz

Saccharomyces cerevisiae (R64-1-1.105)

Common name	Baker's yeast, brewer's yeast, budding yeast
Assembly	R64-1-1
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/saccharomyces_cerevisiae/dna/Saccharomyces_cerevisiae.R64-1-1.dna.chromosome.*.fa.gz (*: I, II, III, IV, IX, V, VI, VII, VIII, X, XI, XII, XIII, XIV, XV, XVI, Mito)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/saccharomyces_cerevisiae/Saccharomyces_cerevisiae.R64-1-1.105.gff3.gz

Schizosaccharomyces pombe (ASM294v2.52)

Common name	Fission yeast
Assembly	ASM294v2
Ensembl release	52
Reference sequence	http://ftp.ensemblgenomes.org/pub/fungi/release-52/fasta/schizosaccharomyces_pombe/dna/Schizosaccharomyces_pombe.ASM294v2.dna.chromosome.*.fa.gz (*: I, II, III, MT)
Ensembl annotations	http://ftp.ensemblgenomes.org/pub/fungi/release-52/gff3/schizosaccharomyces_pombe/Schizosaccharomyces_pombe.ASM294v2.52.chr.gff3.gz

Sus scrofa (Sscrofa11.1.105)

Common name	Pig
Assembly	Sscrofa11.1
Ensembl release	105
Reference sequence	http://ftp.ensembl.org/pub/release-105/fasta/sus_scrofa/dna/Sus_scrofa.Sscrofa11.1.dna.chromosome.*.fa.gz (*: 1-18, X, Y, MT)
Ensembl annotations	http://ftp.ensembl.org/pub/release-105/gff3/sus_scrofa/Sus_scrofa.Sscrofa11.1.105.chr.gff3.gz

Zea mays (Zm-B73-REFERENCE-NAM-5.0.57)

Common name	Maize
Assembly	Zm-B73-REFERENCE-NAM-5.0
Ensembl release	57
Reference sequence	`https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-57/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.chromosome.*.fa.gz` `(*: 1-10, nonchromosomal)`
Ensembl annotations	`https://ftp.ebi.ac.uk/ensemblgenomes/pub/release-57/plants/gff3/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.57.chr.gff3.gz`

Sample kit specific resources

QIAseq miRNA Library Kit

- **miRBase**: mirbase_v22, `http://mirbase.org/`
- **spike-ins**: QIAseq miRNA Library QC Spike-ins
- **piRNA**: `https://www.pirnadb.org/`
  - Caenorhabditis elegans (WBcel235.105): piRNAdb.cel.v1_7_6
  - Drosophila melanogaster (BDGP6.32.105): piRNAdb.cel.v1_7_6
  - Homo sapiens (GRCh38.103): piRNAdb.hsa.v1_7_6
  - Mus musculus (GRCm38.101): piRNAdb.mmu.v1_7_6
  - Rattus norvegicus (Rnor6.0.103): piRNAdb.rno.v1_7_6
  - Rattus norvegicus (mRatBN7.2.110): piRNAdb.rno.v1_7_6
- **GtRNAdb**: `http://gtrnadb.ucsc.edu/`
  Applied to analysis workflow versions v1.0 and v1.1 only.
  - Homo sapiens (GRCh38.103): hg38-tRNAs.fa, release 18
  - Mus musculus (GRCm38.101): mm10-tRNAs.fa,release 18
  - Rattus norvegicus (Rnor6.0.103): rn6-tRNAs.fa, release 18

QIAseq Stranded mRNA Select Kit, QIAseq Stranded Total RNA Lib Kit, and QIAseq UPX 3' Transcriptome Kit

- **Spike-ins**: ERCC RNA spike-ins, `https://shop.nist.gov/ccrz__ProductDetails?sku=2374&cclcl`

Appendix C - Analysis workflow versions, species support and parameters

This appendix provides information on analysis workflow and sample comparability, gives an overview of analysis workflow versions and species support, and lists secondary analysis workflow parameters for the individual sample kits and analyses.

About analysis workflow versions and sample comparability

To ensure comparability of samples within a project, the analysis workflow version for a specific sample kit-reference combination is restricted to a single version. As an example, if *QIAseq Stranded mRNA Select/Total RNA Lib Kit, Homo sapiens (GRCh38.103)* samples were included in Project A in *RNA-seq Analysis Portal* 1.0, Project A will be locked to the analysis workflow version of *RNA-seq Analysis Portal* 1.0, i.e. v1.0. Consequently, *QIAseq Stranded mRNA Select/Total RNA Lib Kit, Homo sapiens (GRCh38.103)* samples added to this project at a later point in time will be processed with analysis workflow 1.0 regardless the version of *RNA-seq Analysis Portal*.

If you create a new project in the *Align and count* dialog, samples will be processed with the most recent analysis workflow version.

To have samples processed with an older analysis workflow version, e.g. to ensure that they can be included in experiments with older samples, in the *Align and count* dialog select an existing project with samples processed with the desired analysis workflow version.

The version of the analysis workflow with which samples were processed can be found from the *Experiment summary* tab of the *Sample and Quality control*, as well as from the *Samples* page.

Analysis workflow versions in RNA-seq Analysis Portal

Table 2 provides an overview of the most recent sample kit analysis workflow versions in each version of *RNA-seq Analysis Portal*.

RNA-seq Analysis Portal version	1.0	1.1	2.0	2.5, 3.0	3.0.1	4.0, 4.1	5.0, 5.1
QIAseq miRNA Library Kit	v1.0	v1.0	v1.1	v1.2	v1.2	v1.2	v1.2
QIAseq UPX 3' Transcriptome Kit	v1.0	v1.0	v1.1	v1.2	v1.2	v1.2	v1.2
QIAseq UPXome RNA Lib Kit				v1.0	-	-	-
QIAseq UPXome RNA Lib Kit (N6-T RT + ODT-T RT primers)					v1.0	v1.0	v1.0
QIAseq UPXome RNA Lib Kit (ODT-T RT primer)					v1.0	v1.0	v1.0
QIAseq UPXome RNA Lib Kit (N6-T RT primer)					v1.0	v1.0	v1.0
QIAseq FastSelect RNA Lib Kit (N6-T RT + ODT-T RT primers)						v1.0	v1.0
QIAseq FastSelect RNA Lib Kit (ODT-T RT primers)						v1.0	v1.0
QIAseq FastSelect RNA Lib Kit (N6-T RT)						v1.0	v1.0
QIAseq Stranded mRNA Select/Stranded Total RNA Lib Kit	v1.0	v1.1	v1.2	v1.2	v1.2	v1.2	v1.2
TruSeq Stranded Total RNA Library Prep		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1
Illumina Stranded Total RNA Prep		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1
NEBNext Ultra II Directional RNA Library Prep Kit		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1
KAPA RNA HyperPrep Kit		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1
SMARTer Stranded Total RNA Sample Prep Kit		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1
Collibri Stranded RNA Library Prep Kit		v1.0	v1.1	v1.1	v1.1	v1.1	v1.1

Table 2: Overview of most recent sample kit analysis workflow versions in each version of RNA-seq Analysis Portal.

Sample kit analysis workflows and species support

Species support depends on the sample kit analysis workflow. Table 3 gives an overview of species support for the two categories of analysis workflows, miRNA and RNA:

- miRNA
  - QIAseq miRNA Library Kit
- RNA
  - QIAseq UPX 3' Transcriptome Kit
  - QIAseq UPXome RNA Lib Kit
  - QIAseq FastSelect RNA Lib Kit (N6-T RT + ODT-T RT primers)
  - QIAseq FastSelect RNA Lib Kit (N6-T RT primer)
  - QIAseq FastSelect RNA Lib Kit (ODT-T RT primer)
  - QIAseq UPXome RNA Lib Kit (N6-T RT + ODT-T RT primers)
  - QIAseq UPXome RNA Lib Kit (N6-T RT primer)
  - QIAseq UPXome RNA Lib Kit (ODT-T RT primer)
  - QIAseq Stranded mRNA Select/Stranded Total RNA Lib Kit
  - TruSeq Stranded Total RNA Library Prep (Human/Rat, Gold, Globin)
  - Illumina Stranded Total RNA Prep with Ribo-Zero Plus
  - NEBNext Ultra II Directional RNA Library Prep Kit for Illumina
  - KAPA RNA HyperPrep Kit
  - SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian/Low Input Mammalian
  - Collibri Stranded RNA Library Prep Kit for Illumina Systems

Biological insights support. Table 3 also indicates whether biological insights are supported. For species other than human, mouse and rat, Ingenuity Pathway Analysis (IPA) will map the gene identifiers according to HomoloGene to the corresponding human, mouse, and rat ortholog information in the QIAGEN Knowledge Base. Therefore, biological insights results will reflect content for human, mouse, and rat.

	miRNA	Biological insights for miRNA	RNA	Biological insights for RNA
Arabidopsis thaliana	x		x	
Bos taurus	x		x	x
Caenorhabditis elegans	x		x	x
Canis lupus familiaris	x		x	x
Danio rerio	x		x	x
Drosophila melanogaster	x		x	x
Equus caballus	x		x	x
Gallus gallus	x		x	x
Gasterosteus aculeatus			x	
Homo sapiens	x	x	x	x
Macaca fascicularis			x	x
Macaca mulatta	x		x	x
Mus musculus	x	x	x	x
Oncorhynchus mykiss			x	x
Oryctolagus cuniculus	x		x	x
Rattus norvegicus	x	x	x	x
Saccharomyces cerevisiae			x	x
Schizosaccharomyces pombe			x	
Sus scrofa	x		x	x
Zea mays			x	

Table 3: Species support for miRNA and RNA analysis workflows.

QIAseq miRNA Library Kit (Illumina)

The QIAseq miRNA Library Kit (Illumina) analysis workflow is based on miRBase, why only species present in miRBase are supported.

Ingenuity Pathway Analysis for miRNA data is restricted to human, mouse and rat. For this reason, biological insights results will be generated for these species only.

For an overview of what species are supported for the QIAseq miRNA Library Kit (Illumina), see Sample kit analysis workflows and species support.

Analysis workflow versions

- v1.0 - RNA Analysis Portal 1.0, RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0
- v1.2 - RNA-seq Analysis Portal 2.5, 3.0, 3.0.1, 4.0, 4.1, 5.0, 5.1

Align and count analysis workflow parameters

Only applied options and parameter are listed.

Create UMI Reads for miRNA	
Minimum length of miRNA	15
Maximum length of miRNA	55
UMI length	12
Maximum size of small UMI groups	1
Common sequence	AACTGTAGGCACCATCAAT
Maximum differences in common sequence	1

Quantify miRNA	
Minimum supporting count	1
Prioritized species	Selected species
Allow length-based isomiRs	Yes
- Additional upstream bases	2
- Additional downstream bases	2
- Missing upstream bases	2
- Missing downstream bases	2
Maximum mismatches	2
Strand specific	Yes
Minimum sequence length	18
Maximum sequence length	25
Annotation records	v1.0, v1.1: miRBase, piRNA, tRNA, rRNA, mRNA, snoRNA, and lncRNA/scRNA/ncRNA/snRNA v1.2: miRBase, piRNA (when available)

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Small RNA
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq miRNA Library Kit (Thermo Fisher)

The QIAseq miRNA Library Kit (Thermo Fisher) analysis workflow is based on miRBase, why only species present in miRBase are supported.

Ingenuity Pathway Analysis for miRNA data is restricted to human, mouse and rat. For this reason, biological insights results will be generated for these species only.

For an overview of what species are supported for the QIAseq miRNA Library Kit (Thermo Fisher), see Sample kit analysis workflows and species support.

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 5.0, 5.1

Align and count analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Adapter trimming by Trim adapter list	5prime_Ion_Torrent_miRNA_trim_adapter
Create UMI Reads for miRNA	
Minimum length of miRNA	15
Maximum length of miRNA	55
UMI length	12
Maximum size of small UMI groups	1
Common sequence	AACTGTAGGCACCATCAAT
Maximum differences in common sequence	1

Quantify miRNA	
Minimum supporting count	1
Prioritized species	Selected species
Allow length-based isomiRs	Yes
- Additional upstream bases	2
- Additional downstream bases	2
- Missing upstream bases	2
- Missing downstream bases	2
Maximum mismatches	2
Strand specific	Yes
Minimum sequence length	18
Maximum sequence length	25
Annotation records	v1.0: miRBase, piRNA (when available)

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Small RNA
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq UPX 3’ Transcriptome Kit

Analysis workflow versions

- v1.0 - RNA Analysis Portal 1.0, RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0
- v1.2 - RNA-seq Analysis Portal 2.5, 3.0, , 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed. Version numbers such as "v1.0" refer to the analysis workflow version.

Demultiplex Reads	
Allow mismatches	v1.0, v1.1: No v1.2: Yes
Naming	v1.0, v1.1: [row][column] (e.g. Sample1 A1, Sample1 A2) v1.2: [cellID] (e.g. Sample1 C1, Sample1 C2)

Remove and Annotate with Unique Molecular Index	
Read structure	Paired end reads (Index on Read 2)
Number of bases to remove	12
Start position of unique molecular index	0
Length of unique molecular index	12
Trim read-through common sequence and UMI	Yes
- Part of insert to search for	27
- Part of common sequence and UMI to search for	3
- Number of errors allowed in match	3

Trim Reads - Trim short polyA from middle of R1, discard R2	
Adapter trimming by Trim adapter list	UPX_3_prime_adapter_list
Discard short reads	Yes
- Minimum length	15

Trim Reads - Trim polyA and polyG from R1	
Trim homopolymers from 3’	polyA, polyG
Discard short reads	Yes
- Minimum length	15

Create UMI Reads from Reads	
Read structure	Single end reads
Minimum UMI read length	20
Minimum average quality score	20
Minimum UMI group size	1
Set ambiguous nucleotides to N	Yes
Coarse grouping	
- Hasher type	Simple k-mer hasher
- k-mer length	16
- Number of hashes	16
- Similarity factor	2
Fine grouping	
- Hasher type	Simple k-mer hasher
- k-mer length	5
- Number of hashes	16
- Segment length	40
- Minimum similarity (same UMI)	10
- Minimum similarity (similar UMI)	10

Trim Reads - Trim low quality and ambiguous ends	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Discard short reads	Yes
- Minimum length	15

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Forward
Library type	3’ sequencing
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq UPXome RNA Lib Kit

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 2.5, 3.0 (replaced by three workflows - one for each set of primer combinations.)

Analysis workflow parameters

Only applied options and parameter are listed.

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq UPXome RNA Lib Kit (N6-T RT primer)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes
Remove 16 bases from Read2 5’ end (adapter)	Yes

RNA-Seq Analysis	
Mismatch cost	3
Insertion cost	3
Deletion cost	3
Length fraction	0.4
Similarity fraction	0.9
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq UPXome RNA Lib Kit (ODT-T RT primer)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes
Remove 16 bases from Read2 5’ end (adapter)	Yes

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	3’ sequencing
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq UPXome RNA Lib Kit (N6-T + ODT-T RT primers)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Analysis workflow section A

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads; adapter trim	
Automatic read-through adapter trimming	Yes
Remove 16 bases from Read 2 5'end (adapter)	Yes

Analysis workflow section B1 (Gene expression)

Trim Reads; quality and ambiguity trim	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3'	PolyA, PolyG, PolyT
Discard short reads	Yes

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Analysis workflow section B2 (Transcript expression)

Trim Reads; extract N6-T reads	
Discard reads with ODT-T primer	Yes

Trim Reads; quality and ambiguity trim	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes

RNA-Seq Analysis	
Mismatch cost	3
Insertion cost	3
Deletion cost	3
Length fraction	0.4
Similarity fraction	0.9
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq FastSelect RNA Lib Kit (N6-T RT primer)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes
Remove 16 bases from Read2 5’ end (adapter)	Yes

RNA-Seq Analysis	
Mismatch cost	3
Insertion cost	3
Deletion cost	3
Length fraction	0.4
Similarity fraction	0.9
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq FastSelect RNA Lib Kit (ODT-T RT primer)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes
Remove 16 bases from Read2 5’ end (adapter)	Yes

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	3’ sequencing
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq FastSelect RNA Lib Kit (N6-T + ODT-T RT primers)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Analysis workflow section A

Demultiplex Reads	
Allow mismatches	Yes
Naming	[row][column, padded to two digits] (e.g. Sample1 A01, Sample1 A10, Sample1 C07)

Trim Reads; adapter trim	
Automatic read-through adapter trimming	Yes
Remove 16 bases from Read 2 5'end (adapter)	Yes

Analysis workflow section B1 (Gene expression)

Trim Reads; quality and ambiguity trim	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3'	PolyA, PolyG, PolyT
Discard short reads	Yes

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Analysis workflow section B2 (Transcript expression)

Trim Reads; extract N6-T reads	
Discard reads with ODT-T primer	Yes

Trim Reads; quality and ambiguity trim	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Trim homopolymers from 3’	PolyA, PolyG, PolyT
Discard short reads	Yes

RNA-Seq Analysis	
Mismatch cost	3
Insertion cost	3
Deletion cost	3
Length fraction	0.4
Similarity fraction	0.9
Auto-detect paired distances	No
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

QIAseq Stranded mRNA Select/Stranded Total RNA Lib Kit

Analysis workflow versions

- v1.0 - RNA Analysis Portal 1.0
- v1.1 - RNA-seq Analysis Portal 1.1
- v1.2 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed. Version numbers such as "v1.0" refer to the analysis workflow version.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	v1.0: Not applied v1.1: Yes, TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Forward
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

TruSeq Stranded Total RNA Library Prep (Human/Rat, Gold, Globin)

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, , 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

Illumina Stranded Total RNA Prep with Ribo-Zero Plus

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, , 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	Nextera adapter
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

NEBNext Ultra II Directional RNA Library Prep Kit for Illumina

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, , 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

KAPA RNA HyperPrep Kit

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, , 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Reverse
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian/Low Input Mammalian

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Forward
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

Collibri Stranded RNA Library Prep Kit for Illumina Systems

Analysis workflow versions

- v1.0 - RNA-seq Analysis Portal 1.1
- v1.1 - RNA-seq Analysis Portal 2.0, 2.5, 3.0, 3.0.1, 4.0, 4.1, 5.0, 5.1

Analysis workflow parameters

Only applied options and parameter are listed.

Trim Reads	
Trim using quality scores	Yes
- Quality limit	0.05
Trim ambiguous nucleotides	Yes
- Maximum number of ambiguities	2
Automatic read-through adapter trimming	Yes
Adapter trimming by Trim adapter list	TruSeq adapters
Trim homopolymers from 3’	PolyG

RNA-Seq Analysis	
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.8
Similarity fraction	0.8
Auto-detect paired distances	Yes
Maximum number of hits for a read	10
Strand specific	Forward
Library type	Bulk
Ignore broken pairs	Yes
Expression value	Total counts

Create experiment analysis workflow parameters

Only applied options and parameter are listed.

Differential Expression for RNA-Seq	
Technology	Whole transcriptome RNA-Seq
Normalization method	TMM
Filter on average expression for FDR correction	Yes

Create Heat Map for RNA-Seq	
Distance measure	Euclidean distance
Linkage criteria	Complete linkage

Appendix D - RNA-seq Analysis Portal for QIAGEN IPA users

RNA-seq Analysis Portal is available to both users with a My QIAGEN account and users with a QIAGEN Ingenuity Pathway Analysis (IPA) account, see Accessing RNA-seq Analysis Portal. With a QIAGEN IPA account, you can upload the results from your *RNA-seq Analysis Portal* differential expression analysis to your QIAGEN IPA account for tertiary in-depth analysis.

The general functionality of *RNA-seq Analysis Portal* remains the same as when logging in with a My QIAGEN account. There are however a few differences:

1. Analysis credit are associated with your QIAGEN IPA account. A certain number of analysis credits may be part of your QIAGEN IPA account, see Credits for QIAGEN IPA accounts.
2. In addition to the initial IPA results available to all users in the *Biological insights table*, you can upload the results from your *RNA-seq Analysis Portal* differential expression analysis to your QIAGEN IPA account for in-depth tertiary analysis, see Next step for QIAGEN IPA users.

Appendix E - Quality control details

This appendix is a detailed explanation of the QC elements in the Experiment summary and QC report Quality control tab. The analyses in *RNA-seq Analysis Portal* leverage tools and workflows from *QIAGEN CLC Genomics Workbench*. For further detail, please refer to the product manuals:

- *QIAGEN CLC Genomics Workbench*: `https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Introduction_CLC_Genomics_Workbench.html`
- *Biomedical Genomics Analysis plugin*: `https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php?manual=Introduction.html`

The quality control report contains metrics helpful for assessing data quality and the outcome of the individual analysis steps. The metrics are derived from the sample creation process (Align and count).

QIAseq miRNA Library Kits

Quality control summary

The summary table presents key data points from the Quality control tab.

- **Sample name**: The name of the sample.
- **Reads**: The number of reads in the sample data.
  - Low numbers could indicate library prep issues.
- **UMI Reads**: Unique Molecular Indexes are used to join reads with the same amplification origin into UMI reads. This allows for better quantification of the miRNAs by eliminating any library amplification and sequencing bias.
- **Avg Q Score, UMI Reads**: The average quality score of the UMI reads.
  - Numbers less than 30 indicate poor quality library prep or instrument runs.

- **UMI reads annotated**: The number of UMI reads annotated with any of the databases involved in the analysis.
  - Low percentage numbers could indicate sample prep issues or contamination.
- **UMI reads annotated with miRBase**: The number of UMI reads annotated with record from the miRBase database.

Creation of UMI reads

Detailed QC from the process of creating the UMI reads as single consensus reads, from reads that have the same Unique Molecular Index.

- **Input reads**: The number of reads in the sample data. This is the same as the 'Reads' column in the Summary.
  - Low numbers in any samples could indicate library prep issues.
- **Avg Q Score, input reads**: The average quality score of the sample data.
- **Discarded reads**: Reads where the common sequence from the UMI is not found, or where the lengths of the miRNA or UMI do not fulfill the predefined criteria.
- **UMI groups**: The number of UMI groups. The UMI process identifies similar reads via the index and joins them in a group of reads.
- **Merged UMI groups**: Although indexed as different reads, some UMI reads may originate from the same biological read or fragment and have the same genetic code. In the process of creating UMIs, reads with the same index are grouped. Identical groups are subsequently merged into a Merged UMI group. This final consensus is the 'UMI read' used in downstream analysis. This column is similar to 'UMI Reads' in the summary table. The terms 'UMI reads' and 'UMI' are used interchangeably throughout the report.
- **Avg Q score: UMI reads**: The average quality score of the UMI reads.
  - Numbers less than 30 indicate poor quality library prep or instrument runs.
- **Avg reads per UMI**: The average number of raw reads in each UMI read.
  - Should be greater than one.
- **UMIs with less than 9 reads**: This column together with 'Max reads per UMI' highlights the extreme end of the UMI grouping distribution and should be seen as indications of potential problems in sequencing or library prep. For most applications, the ideal merged UMI group

size will be 2-4 reads. Larger UMI groups consume sequencing capacity without providing additional benefits.

  - A high percentage is preferable. Numbers less than 90 indicate sample prep issues.

- **Max reads per UMI**: This indicates the extreme end of the UMI grouping to highlight distribution and potential problems in sequencing.

Annotation Records Found

- **miRBase (species of interest)**: 'Records in source' indicates the number total number of Precursor miRNAs (pre-miRNAs) in the database used for annotating the features. Note that miRBase records are pre-miRNAs, whereas sample findings indicate the numbers of corresponding mature miRNAs.
  For each sample, the number and percentage indicates how many of the database records were seen in the sample.

- **piRNA (piRNAdb_species)**: 'Records in source' indicates the total number of records available in piRNAdb for the given species.
  For each sample, the number and percentage indicates how many of the database records were seen in the sample. See reference appendix for more information on piRNA version.

Unique search sequences

For annotating the reads with database information, the analysis collapses the reads into unique search sequences. Collapsing identical reads into unique search sequences significantly reduces the number of miRNA reads in the subsequent annotation step and thereby saves computational time. The annotations are subsequently transferred to the UMI reads used in the expression analysis.

- **Annotated (of total)**: Unique search sequences annotated with database records.

- **Annotated with miRBase (species) (of total)**: Unique search sequences annotated with miRBase database records.

- **Annotated with piRNA (piRNAdb_species) (of total)**: Unique search sequences annotated with piRNA database records.

- Additional databases can be listed here depending on how the workflow has been set up.

- **Unannotated (of total)**: Unique search sequences that did not match database records.

- **Total (of total)**: The number of unique search sequences in the sample.

Reads

Distribution of UMI reads annotated with the records from the applied databases. Annotations are transferred from the unique search sequences.

- **Annotated (of total)**: Reads annotated with database records.
- **Annotated with miRBase (species) (of total)**: Reads annotated with miRBase database records. Note that miRBase annotations include reads annotated outside mature regions, which are ignored in the counts and differential expression analyses. This can result in a higher number shown here compared to the number of features in your differential expression analysis or in expression value downloads.
- **Annotated with piRNA (piRNAdb_species) (of total)**: Reads annotated with piRNA database records.
- Additional databases can be listed here depending on how the workflow has been set up.
- **Unannotated (of total)**: Reads that did not match database records.
- **Total (of total)**: The number of reads in the sample.

Spike-in quality control

This section is available if the *Spike-ins* option was selected in the *Align and Count* dialog.

- **Spike-ins detected**: The number of spike-ins detected relative to the spike-ins used.
- **UMI reads mapped to spike-ins**: The number of UMI reads that mapped to the detected spike-ins.
- **% of total UMI reads**: Percentage of UMI reads that mapped to spike-ins.

QIAseq UPX 3' Transcriptome Kits

Quality control summary

The summary table presents key data points from the Quality control tab.

- **Sample name**: The name of the sample.
- **Reads**: The number of reads in the sample data.

  - Low numbers indicate library prep issues.
  - Spillover into unused wells causes those wells to return small numbers of reads.

- **Trimmed reads**: The number of trimmed input reads.
- **UMI Reads**: Unique Molecular Indexes join similar reads into UMI reads. This allows for better quantification of the RNAs by eliminating any library amplification and sequencing bias.
- **Avg Q score, UMI reads**: The average quality score of the UMI reads.
  - Numbers less than 30 indicate poor quality library prep or instrument runs.
- **Trimmed UMI reads**: A second quality trimming is performed on the created UMI reads before mapping.
- **Mapped**: Percentage of UMI reads that mapped to the selected reference.
  - If the percentage is low, check that you selected the correct reference species.
- **Mapped to total rRNA**: Percentage of reads mapped to ribosomal RNA.
  - If rRNA was not depleted, percentages can vary from 10%-30%.
  - If rRNA was depleted, percentages should be low, about or less than 1%.
  - Samples not depleted of rRNA or samples with high rRNA percentage can still be used for differential expression, but expression values such as TPM and RPKM may not be comparable to those of other samples. To troubleshoot the issues in future experiments, check for rRNA depletion prior to library preparation. Also, if an rRNA depletion kit was used, check that the kit matches the species being studied.

Trimming, raw reads

Metrics for trimming of raw input reads. If the number of reads and average read length decrease substantially after trimming, it indicates poor reads quality.

- **Reads before trim**: The number of raw reads prior to trimming. This value is also shown in the 'Reads' column in the summary table.
- **Avg length before trim**: The average read length before trimming.
- **Reads after trim**: The number of raw reads after trimming. This value is also shown in the 'Trimmed reads' column in the summary table.
- **Avg length after trim**: The average read length after trimming.

Creation of UMI reads

Metrics from the process of creating UMI reads from reads that have the same Unique Molecular Index.

- **Read pairs and single reads annotated with UMIs**: The percentage of reads that were annotated with UMIs.
- **Input Reads**: The number of raw reads used as input for the creation of UMI reads.
- **Avg Q score, input reads**: The average quality score of the input raw reads.
  - Numbers less than 30 indicate poor quality library preps or instrument runs.
- **Detected barcode length**: The sequence length of the UMI barcode is defined in the UMI protocol. This value can be used as a check to see if the analysis detected the expected length.
- **UMIs**: The number of created UMI reads.
- **Avg reads per UMI**: The average number of raw reads in each UMI read.
  - Should be greater than one. For most applications, the ideal UMI group size will be around 2-4.
- **UMIs with more than 10 reads (Pct of UMIs) (Pct of reads)**: This column highlights the extreme end of the UMI grouping distribution and should be seen as indications of potential problems in sequencing or library prep. For most applications, the ideal merged UMI group size will be around 2-4 reads. Larger UMI groups consume sequencing capacity without providing additional benefits.
  - A low percentage is preferable. If the percentage is higher than 5, or there are large variation between samples, a re-evaluation of the sample prep may be needed.
- **Max reads per UMI**: This indicates the extreme end of the UMI grouping to highlight distribution and potential problems in sequencing.
- **Avg Q Score, UMI reads**: The average quality score of the UMI reads. This should be higher than the average quality score for the input reads.
  - Numbers less than 30 indicate poor quality library preps or instrument runs.

Trimming, UMI reads

Metrics for trimming of merged UMI reads. If the number of UMI reads and average read length decreases substantially after trimming, it indicates poor quality UMI reads. Trimming is performed on the UMI reads even though they are merged from reads that have already been quality trimmed. Very few reads should be trimmed at this step, and if a considerable amount are trimmed it could indicate of a problem in the UMI process or further upstream in the pipeline.

- **UMI reads before trim**: The number of UMI reads before trimming.
- **Avg length before trim**: The average UMI read length before trimming.
- **UMI reads after trim**: The number UMI reads after trimming.
- **Avg length after trim**: The average UMI read length after trimming.

Spike-in quality control

This section is available if the *Spike-ins* option was selected in the *Align and Count* dialog.

- **Number of spike-ins detected**: The number of spike-ins detected relative to the spike-ins used.
- **R2**: Correlation of expected and sequenced spike-ins using the Pearson Correlation coefficient
  - When samples have a poor correlation (R2 < 0.8) between known and measured spike-in concentrations, it indicates problems with the spike-in protocol or a more serious problem with the sample.
- **Reads mapped to spike-ins**: The number of reads that mapped to the detected spike-ins.
  - If fewer than 10,000 reads mapped to spike-ins, consider using more spike-in mix in future experiments.
- **Lower limit of detection (attomoles/ul)**: Spike-ins concentration measurement. The lower limit of detection is the lowest concentration spike-in to which at least 3 reads map. This provides a rough estimate of the minimal concentration of mRNA that can be detected in this sample.

Mapping statistics

Metrics for the step of mapping UMI reads to the reference genome.

- **UMI reads**: The number of UMI reads used as input for the mapping step.
- **Paired (yes/no)**: Indication of whether the input reads are paired reads. This should fit with the applied protocol.
- **Reads mapped**: The percentage of the UMI reads that were mapped to the reference. This excludes both reads that were ignored due to wrong strand and reads that could not be mapped to the reference.
- **Strand-specific setting**: Read direction of UMI reads. This should fit the applied protocol.
- **Forward % of reads mapped**: Percent of UMI reads mapped in the forward direction.
- **Reverse % of reads mapped**: Percent of UMI reads mapped in the reverse direction.
- **Ignored reads % (wrong strand**): The percentage of UMI reads that were ignored because the did not map to the strand defined by the strand-specific setting.
  - A percentage greater than 20-25% indicates that the wrong strand protocol may have been used in library prep.

Mapped by type

This section describes the relative mapping of the UMI reads in terms of the type of target.

- **Mapped to gene**: Percentage of UMI reads that map to genes.
- **Mapped to gene, intron**: Percentage of UMI reads that mapped partly or entirely within an intron.
- **Mapped to gene, exon**: Percentage of UMI reads that mapped entirely within an exon or to an exon-exon junction.
- **Mapped to intergenic region**: Percentage of UMI reads that mapped partly or entirely between genes.

Biotype Distribution

Metrics covering the biotype distribution. The plot and the table show which biotypes are found in the samples and at which relative abundance. The names and classification of the biotypes is based on the Ensembl definitions found here: `http://www.ensembl.org/info/genome/genebuild/biotypes.html`.

Taxonomic profile of unmapped reads

Taxonomic profiling is performed for samples with a high level of unmapped reads as this can indicate sample contamination. If all samples have low levels of unmapped reads, this section will be empty. Plot and table show the relative abundance at phylum level. Reads that map equally well to two or more phyla are assigned to the common ancestor (kingdom level).

Taxonomic profiling summary

Information about which taxonomic levels were found in the data sample and how many different taxa were found on each level.

- **Kingdom**
- **Phylum**
- **Total reads**: The number of reads used as input for this step, i.e. reads that did not map the reference genome.
- **Classified reads**: The number of reads that mapped to the taxonomic profiling database.
- **Unclassified reads**: The number of reads that mapped to neither the reference nor the taxonomic profiling database. These are reads of unknown origin. If this number constitutes a significant portion of the input reads, it may be due to the selection of the wrong reference.

QIAseq Stranded RNA Library Kits (FastSelect RNA Library Kit, Stranded mRNA/Stranded Total RNA Library Kits), QIAseq UPXome Kit, and all third-party kits

Quality control summary

The summary table presents key data points from the Quality Control Report.

- **Sample name**: The name of the sample.
- **Reads**: The number of reads in the sample data.
  - Low numbers in any samples indicate failed library prep.
- **Trimmed reads**: Quality trimming is performed on the input reads.
- **Mapped/Mapped in pairs**: The number of reads that mapped to the reference. If the input was paired reads, this value will be the number of reads that mapped in pairs, i.e reads that mapped as broken pairs are left out.

  - A low percentage can be an indication that the wrong reference was selected, that data is of poor quality, or a sign of sample contamination.

- **Mapped to total rRNA**: Percentage of reads mapped to ribosomal RNA (rRNA) or mitochondrial ribosomal RNA (MtrRNA).
  - If rRNA was not depleted, percentages can vary from 10%-30%.
  - If rRNA was depleted, percentages should be low, about or less than 1%.
  - Samples not depleted of rRNA or samples with higher percentages can still be used for differential expression, but expression values such as TPM and RPKM may not be comparable to those of other samples. To troubleshoot the issues in future experiments, check for rRNA depletion prior to library preparation. Also, if an rRNA depletion kit was used, check that the kit matches the species being studied.

Trimming

Metrics for trimming of raw input reads. If the number of reads and average read length decrease substantially after trimming, it indicates poor read quality.

- **Reads before trim**: The number of raw reads prior to trimming. This value is also shown in the 'Reads' column in the summary table.
- **Avg length before trim**: The average read length before trimming.
- **Reads after trim**: The number of raw reads after trimming. This value is also shown in the 'Trimmed reads' column in the summary table.
- **Avg length after trim**: The average read length after trimming.

Spike-in quality control

This section is available if the *Spike-ins* option was selected in the *Align and Count* dialog.

- **Number of spike-ins detected**: The number of spike-ins detected relative to the spike-ins used.
- **R2**: Correlation of expected and sequenced spike-ins using the Pearson Correlation coefficient
  - When samples have a poor correlation (R2 < 0.8) between known and measured spike-in concentrations, it indicates problems with the spike-in protocol or a more serious problem with the sample.

- **Reads mapped to spike-ins**: The number of reads that mapped to the detected spike-ins.
  - If fewer than 10,000 reads mapped to spike-ins, consider using more spike-in mix in future experiments.
- **Lower limit of detection (attomoles/ul)**: Spike-ins concentration measurement. The lower limit of detection is the lowest concentration spike-in to which at least 3 reads map. This provides a rough estimate of the minimal concentration of mRNA that can be detected in this sample.

Mapping statistics

Metrics for the step of mapping reads to the reference genome.

- **Reads**: The number of reads used as input for the mapping step.
- **Paired (yes/no)**: Indication of whether the input reads are paired reads. This should fit with the applied protocol.
- **Reads mapped (in pairs)**: The percentage of the reads that were mapped to the selected reference. Excludes both reads that were ignored due to wrong strand and reads that could not be mapped to the reference.
- **Strand-specific setting**: Read direction of reads. This should fit the applied protocol.
- **Forward % of reads mapped**: Percent of reads mapped in the forward direction.
- **Reverse % of reads mapped**: Percent of reads mapped in the reverse direction.
- **Ignored reads % (wrong strand**): The percentage of reads that were ignored because the did not map to the strand defined by the strand-specific setting.
  - If percentages are greater than 20-25% that the wrong strand protocol may have been used in library prep.

Mapped by type

This section describes the relative mapping of the UMI reads in terms of the type of target.

- **Mapped to gene**: Percentage of reads that map to genes.
- **Mapped to gene, intron**: Percentage of reads that mapped partly or entirely within an intron.
- **Mapped to gene, exon**: Percentage of reads that mapped entirely within an exon or to an exon-exon junction.

- **Mapped to intergenic region**: Percentage of reads that mapped partly or entirely between genes.

Biotype Distribution

Metrics covering the biotype distribution. The plot and the table show which biotypes are found in the samples and at which relative abundance. The names and classification of the biotypes is based on the Ensembl definitions found here: `http://www.ensembl.org/info/genome/genebuild/biotypes.html`.

Taxonomic profile of unmapped reads

Taxonomic profiling is performed for samples with a high level of unmapped reads as this can indicate sample contamination. If all samples have low levels of unmapped reads, this section will be empty. Plot and table show the relative abundance at phylum level. Reads that map equally well to two or more phyla are assigned to the common ancestor (kingdom level).

Taxonomic profiling summary

Information about which taxonomic levels were found in the data sample and how many different taxa were found on each level.

- **Kingdom**
- **Phylum**
- **Total reads**: The number of reads used as input for this step, i.e. reads that did not map the reference genome.
- **Classified reads**: The number of reads that mapped to the taxonomic profiling database.
- **Unclassified reads**: The number of reads that mapped to neither the reference nor the taxonomic profiling database. These are reads of unknown origin. If this number constitutes a significant portion of the input reads, it may be due to the selection of the wrong reference.