LEXOGEN
Enabling complete transcriptome sequencing

BlueBee
an Illumina® company

QUANT™ SEQ
Sequencing that counts

QuantSeq 3' mRNA-Seq Integrated Data Analysis Pipelines on BlueBee® Genomics Platform

User Guide

Catalog Numbers:
090 - 094 (QuantSeq Data Analysis on BlueBee® Genomics Platform)

015UG108V0310

FOR RESEARCH USE ONLY. NOT INTENDED FOR DIAGNOSTIC OR THERAPEUTIC USE.
INFORMATION IN THIS DOCUMENT IS SUBJECT TO CHANGE WITHOUT NOTICE.
Lexogen does not assume any responsibility for errors that may appear in this document.

PATENTS AND TRADEMARKS
The QuantSeq 3′ mRNA-Seq Library Prep Kits are covered by issued and/or pending patents. QuantSeq™ is a trademark of Lexogen. Lexogen is a registered trademark (EU, CH, USA). Lexogen UDI 12 nt Unique Dual Index design and UDI sequences are covered by issued and/or pending patents. BlueBee® is a registered trademark of BlueBee Holding BV (an Illumina® Inc. company).

All other brands and names contained in this user guide are the property of their respective owners.

Lexogen does not assume responsibility for patent infringements or violations that may occur with the use of its products.

LIABILITY AND LIMITED USE LABEL LICENSE: FOR RESEARCH USE ONLY
This document is proprietary to Lexogen. The QuantSeq kits and Data Analysis Pipelines are intended for use in research and development only. They need to be handled by qualified and experienced personnel to ensure safety and proper use. Lexogen does not assume liability for any damage caused by the improper use or the failure to read and explicitly follow this user guide. Furthermore, Lexogen does not assume warranty for merchantability or suitability of the product for a particular purpose.

The purchase of the product is subject to Lexogen general terms and conditions (www.lexogen.com/terms-and-conditions/) and does not convey the rights to resell, distribute, further sub-license, repackage, or modify the product or any of its components. This document and its content shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way without the prior written consent of Lexogen.

For information on purchasing additional rights or a license for use other than research, please contact Lexogen.

WARRANTY
Lexogen is committed to providing excellent products. Lexogen warrants that the product performs to the standards described in this user guide up to the expiration date. Should this product fail to meet these standards due to any reason other than misuse, improper handling, or storage, Lexogen will replace the product free of charge or issue a credit for the purchase price. Lexogen does not provide any warranty if product components are replaced with substitutes.

Under no circumstances shall the liability of this warranty exceed the purchase price of this product.

We reserve the right to change, alter, or modify any product without notice to enhance its performance.

LITERATURE CITATION
When describing a procedure for publication using this product, please refer to it as Lexogen's QuantSeq 3′ mRNA-Seq Kit and integrated Data Analysis Pipeline on BlueBee® Genomics platform.

CONTACT INFORMATION

Lexogen GmbH
Campus Vienna Biocenter 5
1030 Vienna, Austria
www.lexogen.com
E-mail: info@lexogen.com

Support
E-mail: support@lexogen.com
Tel. +43 (0) 1 3451212-41
Fax. +43 (0) 1 3451212-99

Table of Contents

LEXOGEN

Enabling complete transcriptome sequencing

1. Overview

BlueBee® Genomics provides an easily accessible and secure platform for performing automated analysis of QuantSeq (Illumina®) sequencing data. The QuantSeq Data Analysis Pipelines are available for a range of species and are adapted specifically for the analysis of QuantSeq 3′ mRNA-Seq data. Access to the Data Analysis Pipelines is possible with the activation codes that are provided with the following QuantSeq kits / modules:

- All QuantSeq 3′ mRNA-Seq Library Prep Kits for Illumina FWD and REV
- UMI Second Strand Synthesis Module for QuantSeq FWD (Illumina, Read 1)

Getting started with the BlueBee® Genomics Platform for QuantSeq Data Analysis is easy!

Register and Activate Your Code

Activation codes are provided with the respective QuantSeq kit or module. Enter the code when registering on the BlueBee® Genomics Platform (see **2. Registration**). Activation codes provided with QuantSeq FWD kits grant access to both FWD and FWD-UMI analysis pipelines. Please ensure you select the correct pipeline according to the library preparation method.

Connect and Upload Your Data

Zipped individual FASTQ files are the required input for QuantSeq Data Analysis Pipelines. Lexogen provides complimentary BlueBee® activation codes for QuantSeq kit customers of up to 1.5 GB. For larger files, activation codes can be purchased at Lexogen's online store. Uploading and downloading data can be handled by the BlueBee® Service Connector, a Cloud Connector (BaseSpace, Google, Amazon), and/or the Drag & Drop browser option (see **3. Get Connected**).

Select and Run Your Pipeline

Based on the activation code that comes with your QuantSeq kit, the correct data analysis pipeline (FWD / FWD-UMI or REV) will be linked to your account. You only need to select the target genome (i.e., species) for read alignment (see **4. Run Your Pipeline**).

View and Retrieve Your Data

Results of the analysis pipeline can be viewed and downloaded via the web interface as well as through the BlueBee® Service and Cloud Connectors (BaseSpace, Google, Amazon) (see **3. Get Connected**). Your data and results are not stored indefinitely and should be downloaded promptly after the pipeline run is completed (see **5. Explore Your Results**).

2. Registration

For registration, please keep your kit activation code handy as you will need it to verify your access. The activation code can be found on the side of the inner reagent box of the QuantSeq kits (FWD and REV, Cat. No. 015 and 016) or on the Information Card inside of the reagent boxes of the QuantSeq with UDI kits (Cat. No. 113 - 115 and 129 - 131). The code includes a combination of letters and numbers in the following format:

7YTM-D2IP-QAMX-HAPT

To register yourself as a first time user of the BlueBee® Analysis Platform, go to: https://lexogen.bluebee.com/quantseq.

First, click **Register** to begin the registration process.

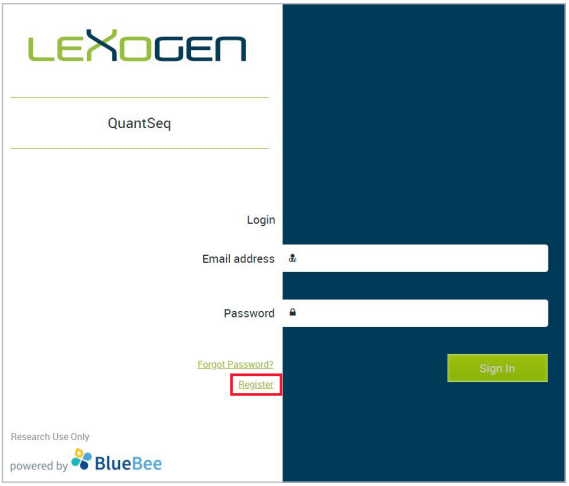

Figure 1. Login page.

Enter your email address and activation code in the required field and click **NEXT**. Fill out all other data to complete the registration.

Activation codes registered after September 20, 2018 are valid for two years. The input file size is limited to 1.5 GB per FASTQ(.gz) file. If you have larger input files or for further inquiries, please contact support@lexogen.com. Codes for additional runs (up to 1.5 GB, Cat. No. 090, 091) or larger input files (Cat. No. 093, 094) can also be purchased from Lexogen.

NOTE: The activation codes are provided with a fixed number of data analysis pipeline runs equal to number of library preps included in the purchased kit or module. This enables one data analysis pipeline run for each library i.e., FASTQ file. The Differential Expression Pipeline can be run up to 500 times using each activation code. Please note, it is not possible to reupload the aligned bam files and re-run the pipeline.

3. Get Connected

There are three options to upload your FASTQ files: BlueBee® Service Connector, Cloud Connector, and Drag & Drop.

3.1 BlueBee® Service Connector

Once registered, you can install and configure the BlueBee® Connector. Go to **Settings** in the top right corner and scroll down to **Local Connector Settings**.

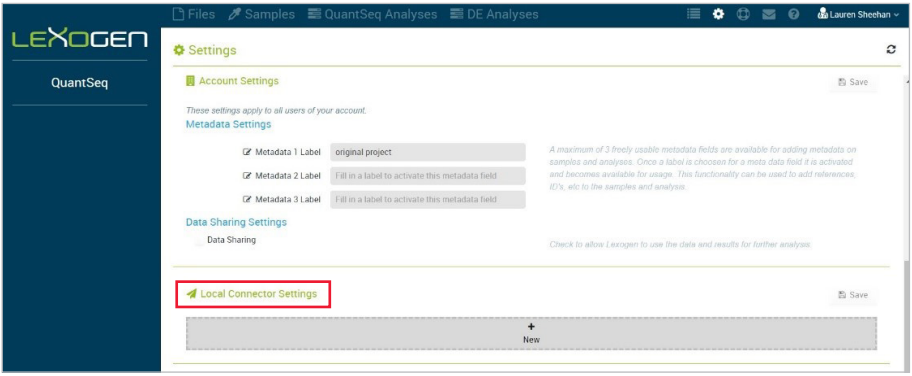

Figure 2. Local Connector settings.

Click **New** to add information on the BlueBee® Service Connector. The BlueBee® Service Connector is a lightweight piece of software enabling several key functionalities and security features for use of the BlueBee® service. The connector uniquely identifies the client and facilitates end-to-end process auditing. It collects data coming straight from the specified server, computer, or sequencer and initiates the transfer of raw data in an automated way (no manual intervention required). The software not only encrypts but also compresses the data, thereby drastically accelerating the data transfer to the appropriate computing centre.

After clicking New, configure what folder the connector pulls FASTQ files from (**upload location**) and into what folder the connector drops output files you scheduled for download (**download location**). Please note, if you would like the connector to only upload or download data, leave the respective location empty.

After filling in the upload and download details, press **Save** at the top right of the screen. The connector software for the operating system you choose will be downloaded automatically. You will see the progress of your uploads on the left of your homepage under **Local Connector Uploads**. Once files are finished uploading, they will be located under the **Files** tab.

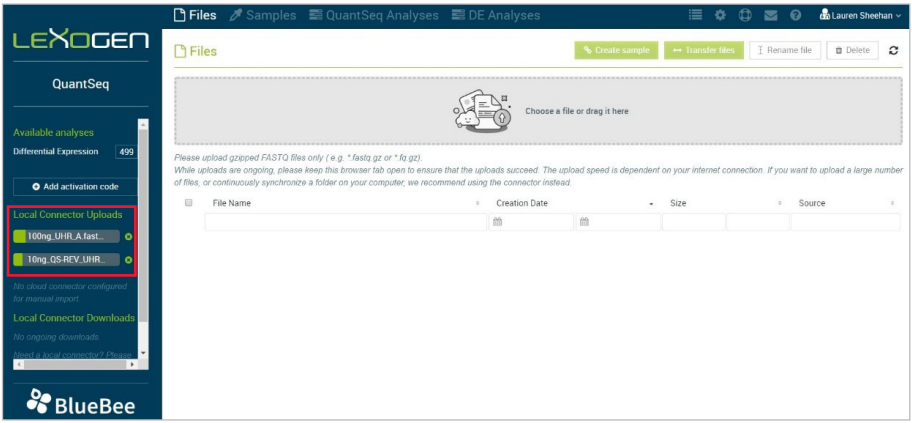

Figure 3. Uploads in progress with the BlueBee® Service Connector.

3.2 Cloud Connectors

You can import data to BlueBee® from the following cloud providers: BaseSpace, Google, and Amazon. To begin, click on the **+** button on the respective connector you wish to import data from.

PLEASE NOTE: It is possible to have multiple Cloud Connectors. For each Cloud Connector added, provide the connector with a unique name.

To make a connection between the bucket or project located on the cloud provider, specific information is required. Instructions on how to find this information are listed below under each respective provider. After entering this specific information, continue with creating an import rule.

3.2.1 BaseSpace

In the instructions below, we are assuming the user already has a project containing data. If you still need to create a project and upload data, please consult the BaseSpace help pages.
To connect the BlueBee® Cloud Connector with BaseSpace you will need:

- **URL:** The user can keep the predefined URL.
- **Access Token:** BaseSpace provides online information on how to request your token (scroll down to 'Steps for developing your first App') or you can follow the next steps:

1. Go to the *My Apps* page within the BaseSpace developers environment or follow this link.
2. If you haven't created an application yet, follow these additional steps first:
    a. Click on 'Steps for developing your first App'
    b. Click on 'Register'
    c. Accept the Terms and Conditions

3. Click the 'Create a new Application' button:
    a. Name: has to be unique, e.g., company name + 'BlueBee connector'
    b. Provide your company or organization name
    c. As App Type select 'Web/Desktop = Other'
    d. Short Description: e.g., 'BlueBee connector'
    e. Click Create Application
    f. Fill in the Release Notes and Long Description as desired
    g. Select a Category as desired
    h. Select 'read' for the Application Permission Settings
    i. 'Save'
4. Scroll up to find the *Credentials* tab, copy the Access Token, and paste it into the corresponding field in the Cloud Connector on the BlueBee® platform.

3.2.2 Google

To connect to Google, you will need a JSON file with your credentials:

1. On the Google Cloud Platform, navigate to the 'Service accounts' page under *AIM & Admin* section.
2. Follow the steps for creating a new service account.
    - The service account permissions should at least contain the role 'Storage Object Viewer' for Storage.
3. Create a key for the service account of type JSON.
4. Download the JSON file and upload this in the Cloud Connector for Google on the BlueBee® platform.

3.2.3 Amazon

Amazon provides information on how to get your access token in their documentation online. Follow the instructions for creating an IAM user (Console) and use the characteristics below while setting up the user:

- Type of access = Programmatic access
- Minimal permission for user group = ListBucket & GetObject

In the last step of the user creation, you can either copy the Access Key ID and the secret Access key or download the CSV file containing this information necessary for configuring the cloud connector on the platform.

Import Rule for Cloud Connectors

Depending on which cloud provider you want to import from, other fields might be shown in the import rule.

- **Automatic import:** If ticked, files uploaded to this cloud project/bucket will immediately be imported to BlueBee®. To import files manually, untick this box and use the 'List' button.

- **Unique Name:** Name you want to give to this import rule.
- **Bucket (Amazon, Google):** Name of the bucket containing data you want to import.
- **BaseSpace project (BaseSpace):** Name of the project containing data you want to import from BaseSpace.

Save the import rule after configuring it. It can be edited afterwards.

Importing Data from a Cloud Connector

There are two way of importing data: Automatically or Manually

Automatic Import: This can be enabled with a tickbox in the import rule. When you upload files into your cloud project/bucket, these files will immediately be imported to the BlueBee® platform as soon as the upload is complete.

Manual Import: Manual import can be used for several reasons such as:

- You do not want all of the files in your cloud bucket/project to be imported to BlueBee®. In this case, make sure automatic import is disabled.
- You want to import data that has been imported before.

For both actions you can use the 'List' button present in the import rule of the cloud connector. By default, all files will be listed that match your import rule and that have not been imported yet. To view files that have been imported before, matching the import rule, enable the tickboxes on top.

4. Run Your Pipeline

Once data is uploaded, check off the small box to the left of the file you wish to analyze and click **Create Sample**. Enter the information related to your file (i.e., species, Lexogen kit, description of file, etc.). Once finished, click the **Create** button on the bottom left. Your samples can now be found under the **Samples** tab.

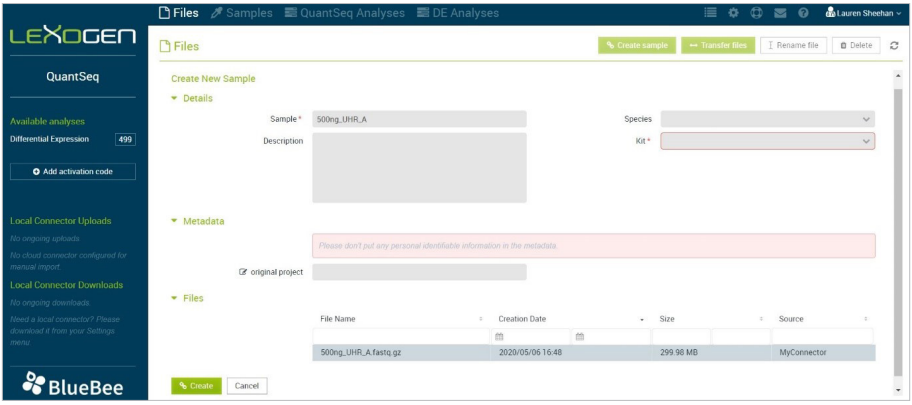

Figure 4. Creating a new sample.

To start a pipeline run, under the Samples tab, select the samples you would like to analyze and click the **QuantSeq Analyses** button. You will then need to select the appropriate pipeline and reference genome from the drop-down menus. Click **Analyze**.

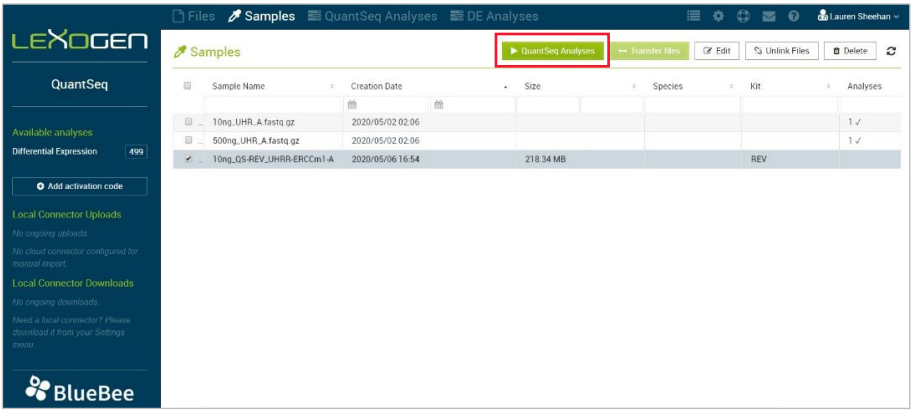

Figure 5. Running the QuantSeq Analyses Pipeline.

ATTENTION: The current pipelines are compatible with single-read data only. If you have paired-end sequencing data, please upload only the Read 1 FASTQ files. Please note that paired-end sequencing is not recommended for QuantSeq FWD libraries.
Under the **QuantSeq Analyses** tab, all runs will appear here and progress can be monitored under the **Status** column. Once complete, **Succeeded** will appear under the Status column and you can click the sample name under the **Analysis ID** column for details of that run.

NOTE: Please refresh your browser to obtain real-time status on the progress of uploads under the QuantSeq Analyses tab.

The REV Data Analysis Pipeline is only accessible using activation codes from QuantSeq REV kits. The FWD and FWD-UMI pipelines are accessible using activation codes provided with QuantSeq FWD kits. When linking pipelines to your project, please ensure you select the correct pipeline for your species, according to the library preparation method. The FWD-UMI pipeline is specifically designed for libraries prepared with the UMI Second Strand Synthesis Module for QuantSeq FWD (Illumina, Read 1, Cat. No. 081). Running the standard FWD pipeline for UMI libraries, and vice versa, will result in run failure.

ATTENTION: Users no longer have to merge input FASTQ files. Simply upload your files and when creating a sample, select all files that should merged (Multi-file sample option). The QuantSeq pipeline will then merge these files when the analysis begins.

The Differential Expression Analysis Pipeline performs pair-wise comparisons of multiple read count files per condition to generate differential expression predictions. Up to 500 pair-wise comparisons can be performed with each activation code.

4.1 QuantSeq FWD / REV

The QuantSeq FWD and REV Data Analysis Pipelines are available for multiple species. The pipelines process the FASTQ files through Trimming, Read Alignment, and Quality Control steps. The output includes read QC reports, alignment statistics and .bam files, and read count files. The read count files are used as input for the Differential Expression Pipeline, where specific read count files can be selected for the desired pair-wise comparisons. The Differential Expression Pipeline outputs the lists of differentially expressed genes, as well as a variety of graphical analyses of the results.

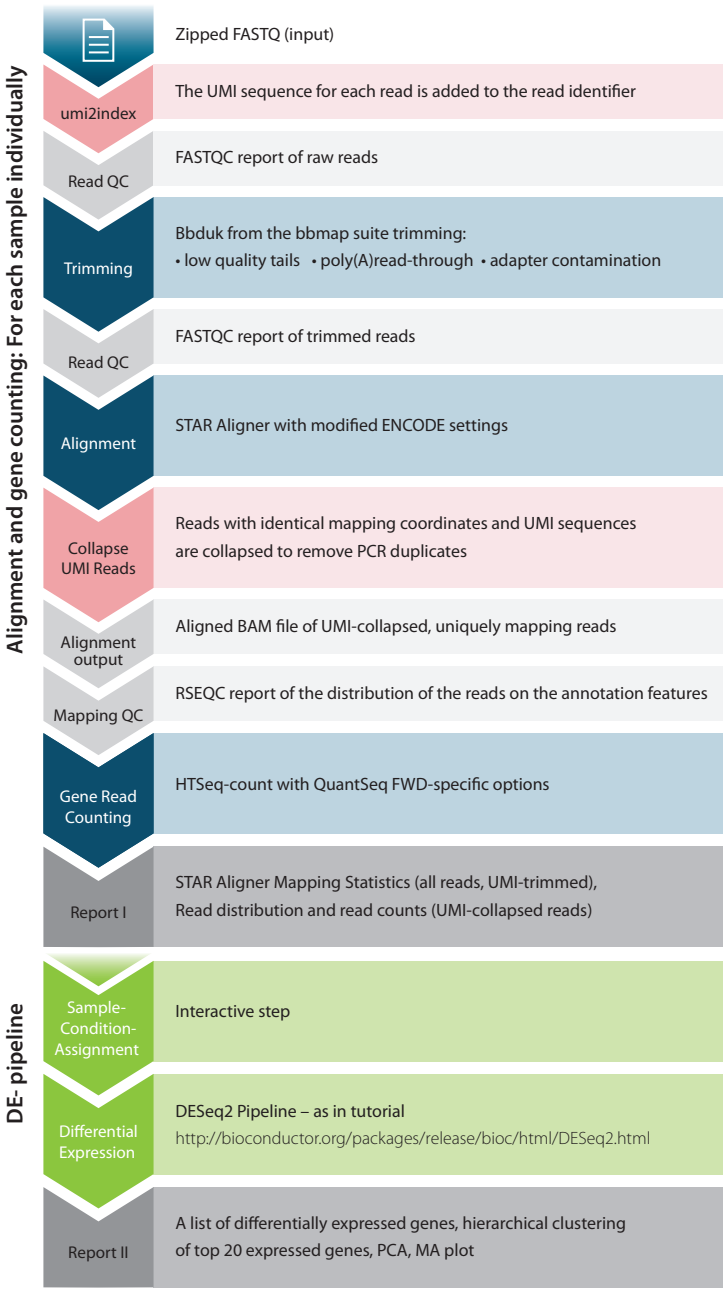

Figure 6. The QuantSeq FWD / REV Data Analysis Pipeline Workflow. Steps in blue represent the data analysis processes, grey bars indicate the results and output files generated. Steps in green indicate the data analysis.

4.2 QuantSeq FWD-UMI

The QuantSeq FWD-UMI Data Analysis Pipeline is specific for QuantSeq FWD libraries that contain Unique Molecular Identifiers (UMIs) in Read 1. The QuantSeq FWD-UMI Data Analysis Pipeline is accessible using the activation code provided with QuantSeq FWD Library Prep Kits. Please ensure you select the correct FWD-UMI pipeline for your species when setting up the run. Running the standard FWD pipeline for QuantSeq FWD-UMI data, and vice versa, will result in run failure.

The FWD-UMI Data Analysis Pipeline includes two additional steps compared to the standard QuantSeq FWD Data Analysis Pipeline:

- First, the umi2index process adds the 6 nucleotide UMI sequence to the identifier of each read and trims the UMI from the start of each read. This generates a new FASTQ file, which is then processed through trimming and alignment.
- Second, after alignment, the mapped reads are collapsed according to the UMI sequence of each read. Reads are collapsed if they have the same mapping coordinates (CIGAR string) and identical UMI sequences. Collapsing reads in this manner removes PCR duplicates.

The output of the QuantSeq FWD-UMI Data Analysis Pipeline includes read QC reports, alignment statistics, and read count files as for the standard QuantSeq FWD Data Analysis Pipeline. In addition, the RSEQC report and read count files are generated from the UMI-collapsed read counts only. The UMI-collapsed read count files can be used as input for the Differential Expression Pipeline.

5. Explore Your Results

Under the **QuantSeq Analyses** tab, you can click a sample to view the results. Navigate this section by clicking one of the options on the left side: Analysis Detail, FastQC Untrimmed, FastQC Trimmed, IGV Browser, and Output Files.

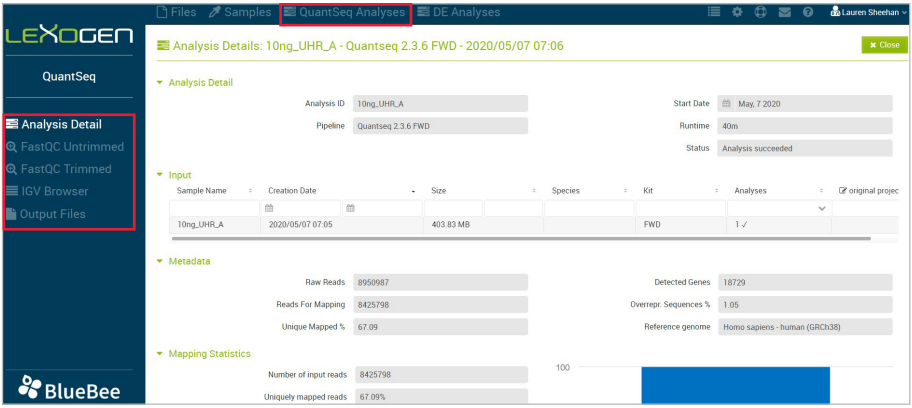

Figure 7. QuantSeq Analyses data output.

You can directly download your files using the **Output Files** option. You can either download files with your browser or with your connector.

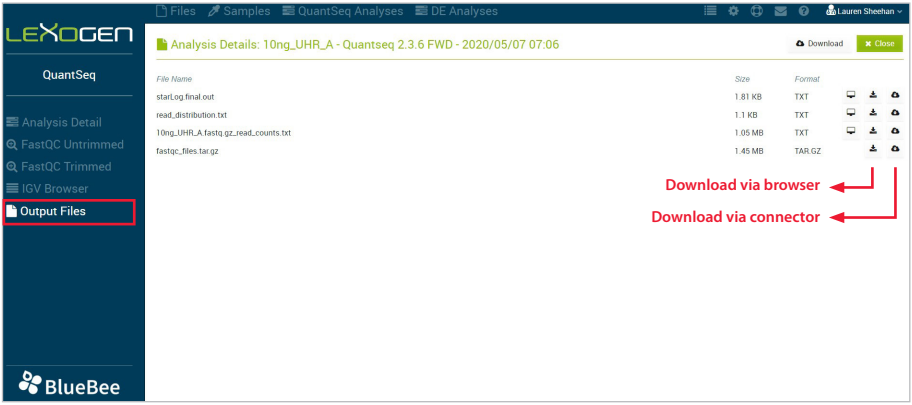

Figure 8. Downloading your output files.

Alternatively, you can select multiple files and click the **Download** button to perform a batch download.

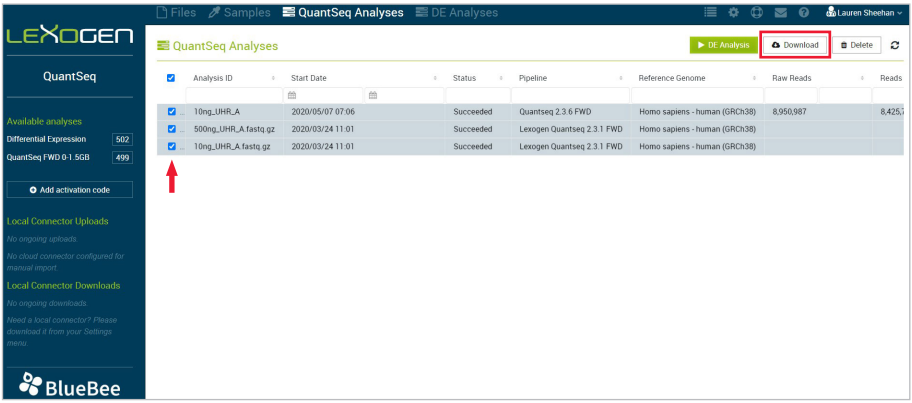

Figure 9. Output data batch downloads.

Data Storage

BlueBee® provides secure cloud-based storage of the results files generated by the pipeline runs. However, results are not stored permanently. We recommend downloading all results files as soon as the pipeline run has finished.

File type	Duration kept in BlueBee®
Uploaded FASTQ files	If not used in a pipeline, 6 days after upload
Output .bam files	1 month (and then permanently deleted)
Output files (except .bam files)	1 year

6. Technical Parameters

The following technical parameters are used in the QuantSeq Data Analysis FWD and REV Pipelines. The pipeline is defined by the type of QuantSeq library (FWD or REV) and the species (e.g., Human, Mouse, etc.). For questions about the Differential Expression (DE) QuantSeq pipeline, please contact support@lexogen.com.

Trimming: bbduk v35.92

```
bbduk.sh in=$fq out=${fq_clean} ref=$resource_dir/polyA.
fa.gz,$resource_dir/truseq-rna.fa.gz k=13 ktrim=r useshortkmers=t
mink=5 qtrim=r trimq=10 minlength=20
```

Quality trimming is performed to remove adapter sequences and poly(A) tails. The referenced truseq_rna.fa.gz file contains the adapter sequences for trimming and is included in the resource files from the BBTools suite. The polyA.fa.gz file can be simply generated to include two sequences: a poly(A) sequence of 18 As, and a poly(T) sequence of 18 Ts. For detailed information on the trimming parameters please consult the BBTools user guide, available here: https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/

For a detailed explanation of bbduk parameters, see all: https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh.

Read QC: FastQC v0.11.5

```
fastqc -o ${out_dir} -t 8 --nogroup $fq $fq_clean
```

FastQC generates a report of sequencing read quality. The --nogroup argument disables grouping of bases for reads longer than 50 bp.

For information and to download the FastQC tool, please see: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Alignment: STAR v2.5.2a

```
STAR --runThreadN 8 --genomeDir $resource_dir/index
--readFilesIn ${fq_clean} --outFilterType BySJout
--outFilterMultimapNmax 20 --alignSJoverhangMin 8
--alignSJDBoverhangMin 1 --outFilterMismatchNmax 999
--outFilterMismatchNoverLmax 0.6 --alignIntronMin 20
--alignIntronMax 1000000 --alignMatesGapMax 1000000
--outSAMattributes NH HI NM MD --outSAMtype BAM SortedByCoordinate
--outFileNamePrefix $out_prefix
```

STAR aligns the trimmed reads to the reference genome. The $resource_dir/index parameter specifies the reference genome to use and is predetermined according to the selected pipeline (e.g., QuantSeq FWD, Human).

For a detailed description of the arguments, please consult the STAR manual available here: https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf

Read Indexing: samtools index v1.3

```
samtools index $bam
```

The Samtools index command creates indexed versions of the aligned read .bam files, which allows for faster processing and visualization of aligned read data.

For more information on Samtools, please consult the comprehensive user guide here: http://www.htslib.org/doc/samtools-1.3.html.

Gene Read Counting: HTSeq-count v0.6.0

```
htseq-count -m intersection-nonempty -s yes -f bam -r pos $bam
$resource_dir/annotation.gtf > $bam_dir/read_counts.txt
```

QuantSeq is a stranded protocol. For the QuantSeq FWD pipeline the argument `-s yes` indicates stranded in the sense orientation. For the QuantSeq REV pipeline `-s reverse` is used. The annotation.gtf file is a predefined .gtf file for the specified pipeline.

For detailed user information and to download HTSeq-count, please see: http://htseq.readthedocs.io.

Mapping QC: RSeQC v2.6.4

```
read_distribution.py -i $bam -r $resource_dir/annotation.bed >
$bam_dir/read_distribution.txt
```

The annotation.bed file is a predefined .bed file for the specified pipeline. The read_distribution.py command is part of the RSeQC tools package and provides a breakdown of reads mapping to annotated features (e.g., CDS_Exons, 5' UTR_Exons, 3' UTR_Exons, Introns, and Intergenic Regions).

For information on the read_distribution.py command and the RSeQC tools package, please see: http://rseqc.sourceforge.net/#read-distribution-py.

UMI-Specific Analyses

The umi2index and UMI read collapsing tool have been specifically developed by Lexogen for the analysis of QuantSeq FWD Read 1 UMI data. Therefore, technical parameters for these processes are not provided. These tools can be provided in the form of Linux/Ubuntu-compatible binaries upon request. Please contact support@lexogen.com.

7. Support

If you have any questions, please fill out a support ticket under the **Support** tab or contact us at support@lexogen.com.

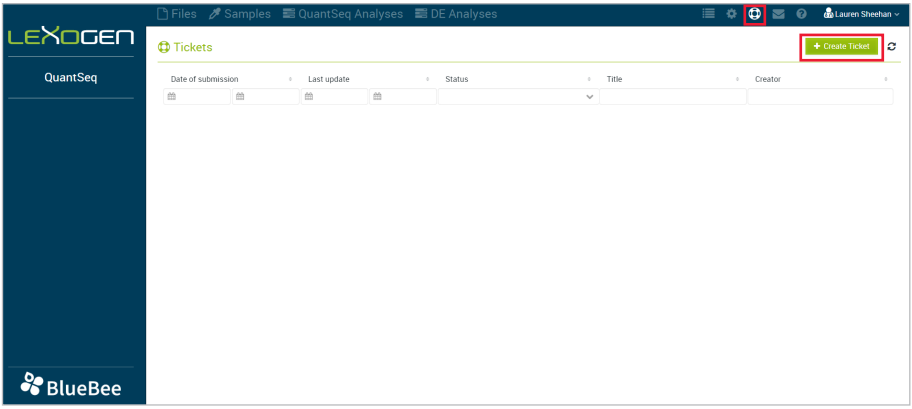

Figure 10. Create a ticket.

Additionally, you can find our frequently asked questions by clicking on the **Help** tab.

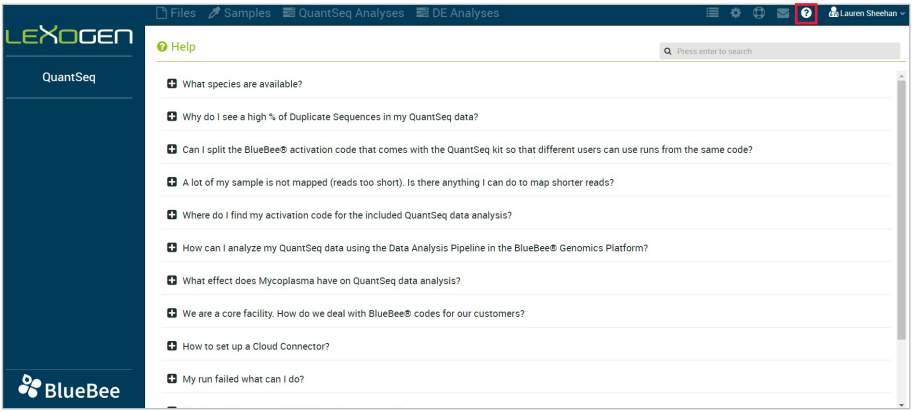

Figure 11. Frequently asked questions.

8. Data Security

BlueBee® can ensure compliance with data protection and security requirements, also for clinical sequencing data. For more information, the BlueBee® Genomics Platform Data Security and Compliance Whitepaper can be found in the downloads section on www.bluebee.com. For further information on data security and data protection, please contact BlueBee® directly via www.bluebee.com.

9. Revision History

Publication No.	Change	Page
015UG108V0310 Jan. 20, 2021	Updated Figure 6.	12
	Contact information updated to support@lexogen.com.	19
015UG108V0300 Jul. 20, 2020	Updated User Guide to fit the new BlueBee Web Solutions interface; adjustments of all text and images.	1 - 22
015UG108V0201 Oct. 19, 2018	Updated BlueBee Logo.	1
	Updated Figure 1.	5
015UG108V0200 Oct. 9, 2018	Updated overview and chapter text explanations with additional information on activation code use.	4 - 16
	Added QuantSeq FWD-UMI Data Analysis pipeline workflow figure and information.	8 - 11
	Updated technical parameter commands for trimming and read alignment.	13 - 14
	Added information about data storage duration and data security.	12, 16
015UG108V0141 Aug. 9, 2017	Text and hyperlinks updated.	4, 7, 10
015UG108V0140 Jul. 3, 2017	Added new section "Technical parameters".	8
015UG108V0130 Mar. 28, 2017	Added new "Overview" section.	4
	Added further text information about connector functionality.	4
015UG108V0120 Jan. 23, 2017	Replaced image of initial connector dialogue.	5
	Added image of advanced connector dialogue.	5
	Added text for advanced connector dialogue.	5
015UG108V0100 Oct. 19, 2016	Initial Release.	

LEXOGEN

Enabling complete transcriptome sequencing

Associated Products:

015 (QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina (FWD))
016 (QuantSeq 3′ mRNA-Seq Library Prep Kit for Illumina (REV) with Custom Sequencing Primer)
081 (UMI Second Strand Synthesis Module for QuantSeq FWD (Illumina, Read 1))
113, 129 - 131, or 115 (QuantSeq 3′ mRNA-Seq Library Prep Kit FWD with UDI 12 nt Set A1, A2, A3, A4, or A1-A4)
114 (QuantSeq 3′ mRNA-Seq Library Prep Kit FWD with UDI 12 nt Set B1)

QuantSeq 3′ mRNA-Seq Integrated Data Analysis Pipelines on BlueBee® Genomics Platform User Guide

Lexogen GmbH
Campus Vienna Biocenter 5
1030 Vienna, Austria
Telephone: +43 (0) 1 345 1212-41
Fax: +43 (0) 1 345 1212-99
E-mail: support@lexogen.com