LEXOGEN
Enabling complete transcriptome sequencing

BLUEBEE
High Performance Genomics

SLAM DUNK

SLAMseq Data Analysis Pipeline
SLAMdunk on Bluebee Platform

User Guide

Catalog Numbers:
061, 062 (SLAMseq Kinetics Kits)
015 (QuantSeq 3′ mRNA-Seq Library Prep Kits)

063UG147V0100

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 SLAMseq and 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). SLAMseq is a trademark of the Institute of Molecular Biotechnology GmbH (IMBA).

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 SLAMseq and QuantSeq kits as well as SLAMdunk data analysis pipeline 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 does not convey the rights to resell, distribute, further sublicense, 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

SLAMdunk was developed in a collaboration between the Zuber Group at the Research Institute of Molecular Pathology (IMP), the von Haesler Group at the Max F. Perutz Laboratories (MFPL) and the Ameres Group at the Institute of Molecular Biotechnology (IMBA) at the Vienna Biocenter. An opens source version of SLAMdunk was made available on Github https://github.com/t-neumann/slamdunk/.

When describing a procedure for publication using this Pipeline product, please refer to it as Lexogen's SLAMdunk integrated data analysis pipeline on Bluebee platform.

For citing a peer reviewed scientific publication of the whole workflow involving the SLAMseq method, QuantSeq kit and SLAMdunk data analysis please refer to: Herzog V. et. al., (2017) Thiol-linked alkylation to assess expression dynamics. *Nature Methods*, doi:10.1038/nmeth. 4435.

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

1. Overview 4
2. Registration 5
3. Get connected 6
4. Upload your data 7
5. Pipeline execution 8
   - 5.1 Data selection 8
   - 5.2 Pipeline configuration 9
   - 5.3 Workflow 10
   - 5.4 Output 11
6. Revision History 12

LEXOGEN

Enabling complete transcriptome sequencing

1. Overview

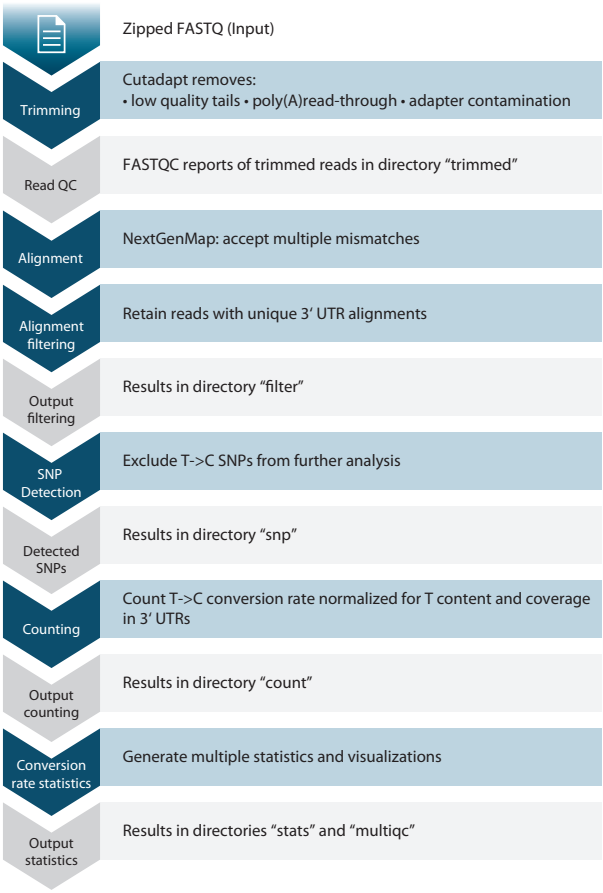
Uploading Your Data

The input to start SLAMdunk data analysis pipeline is zipped fastq file. Uploading and downloading data is handled by the Bluebee Service Connector.

Processing Your Data

A run describes the data analysis run for a certain number of samples. The allocated size of the purchased coupon in GB reflects the size per run with a maximum data input. E.g. if you have a coupon of 12 GB, you can run as many samples as you need but not exceeding total of 12 GB per run. This can be 24 samples of 500 MB each or 48 samples of 250 MB each etc.

IMPORTANT! **Please note that one coupon allows one run not exceeding allocated size. Unused GB will be lost.**



Viewing and Retrieving Your Data

The Bluebee Service Connector also caters for the download of the results. This can be fully automated by defining download rules. Alternatively, the users can login to the platform via the web GUI to view the results online.

Defining Access to Data (for e.g., Core Facilities)

Access to data is controlled through the concept of projects whereby fine grained access to data can be defined. Data is always owned by a specific account within the system, the administrators of this account can setup specific sub-projects and link the appropriate data to become available to designated users. Once a sub-project has been setup, individual users of different accounts can be invited to join the project as a viewer (view only) or contributor (allowing the person to upload and process additional data).

2. Registration

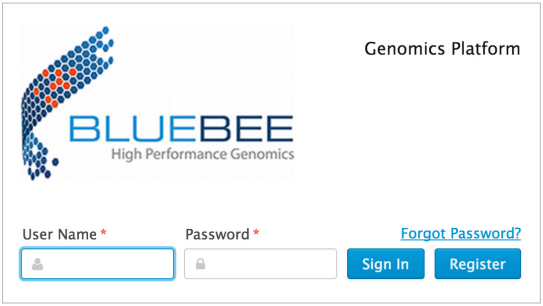
For your registration, please keep your kit activation code handy, you will need it to verify your access.

To register yourself as a first time user of the Bluebee analysis platform, go to:
www.bluebee.com/Lexogen

Click the login button:

LOGIN TO THE PLATFORM

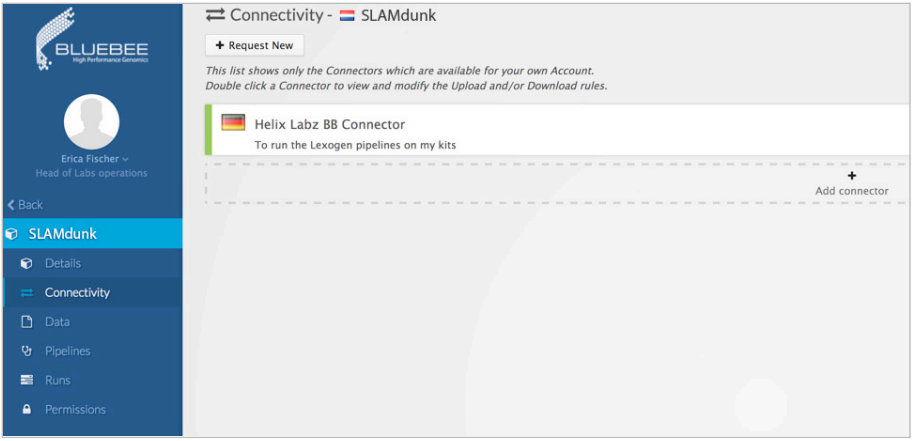
And start the Registration process, by clicking the REGISTER button on the right:



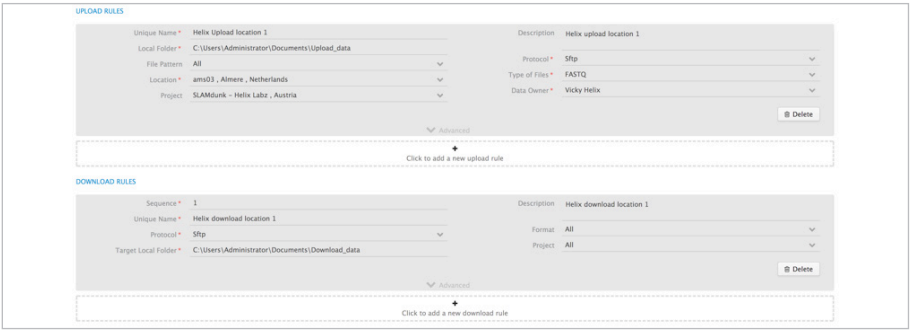
Fill out your kit code in the first part, click CONTINUE, and fill out all other data as explained in the video tutorial you will find on www.bluebee.com/Lexogen

3. Get connected

Once you are registered, please install and configure a Bluebee Connector. You will find the link in your SLAMdunk Project under the Connectivity option. 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 sequencer and initiates the transfer of raw data in an automated way (no manual intervention required). Additionally, the software encrypts, compresses the raw sequence data and drastically accelerates the data transfer to the appropriate Bluebee compute centre.



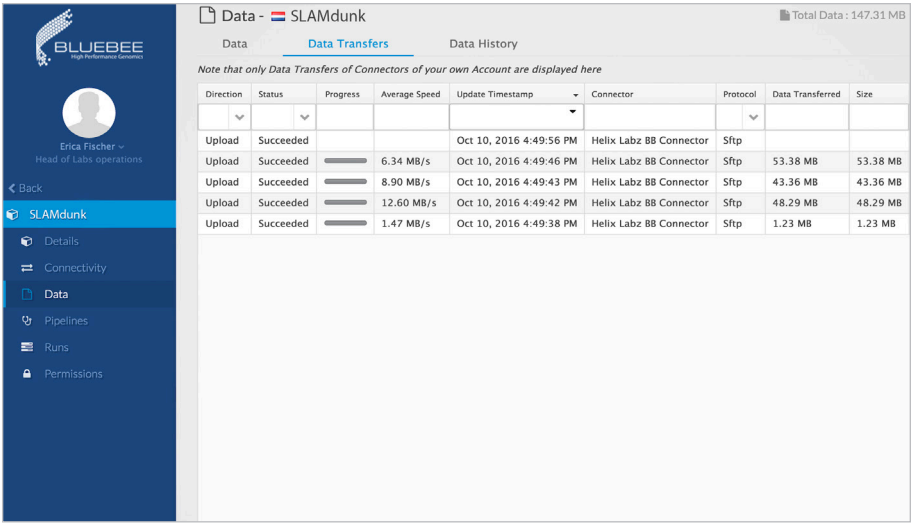
Here, you can configure the location where the Connector needs to get your fastq files (Upload rule), and the location where the connector needs to put files you scheduled for download (Download rule).



After you filled out all those details, press the Install button at the bottom of the screen. The connector software for the operating system you choose will be downloaded automatically. After you install this software, and start up the connector, you're all set to go.

4. Upload your data

Once your connector is installed and active, it will automatically upload your data to Bluebee. Please check your transfers under the DATA option in your SLAMdunk project.



While configuring your upload directory, please note that the Bluebee Connector only picks up from **subfolders** in your Local Folder to transfer sequencer data to the platform.

5. Pipeline execution

5.1. Data selection

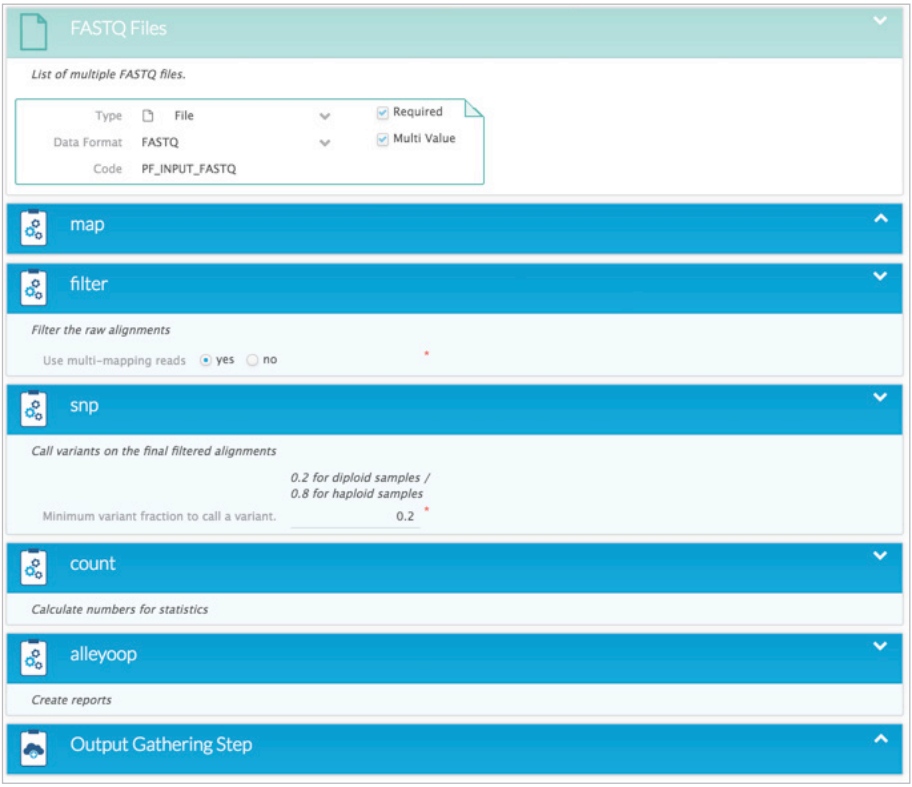
Prior to executing SLAMdunk, the user must select the files which are to be processed. Ideally, the user will select all samples belonging to an experiment/timecourse. This guarantees that summaries and comparative overviews are derived for all conditions of interest. To make the resulting reports and graphics more readable we recommend renaming fastq files to the names of their respective samples.

5.2. Pipeline configuration

The workflow of the SLAMdunk pipeline can be influenced in two ways.

1. Discard or retain multi-mapping reads. If the latter option is chosen a multi-mapping read will be retained only if it maps uniquely to an annotated 3' UTR region.
2. The sensitivity of the SNP variant caller can be adjusted for haploid or diploid RNA samples.

Both options are specified during pipeline selection.



5.3. Workflow

Once a new run has been started the SLAMdunk pipeline executes 5 major steps.

1. The *map* step is used to map reads to a given genome using NextGenMap's SLAMSeq alignment settings. Output files from this step will be individual bam files for each of the samples in the input.

2. The *filter* step is used to filter the raw alignments from the *map* step using multiple quality criteria to obtain the final alignments for all subsequent analyses. Results of this step can be found in directory "filter".

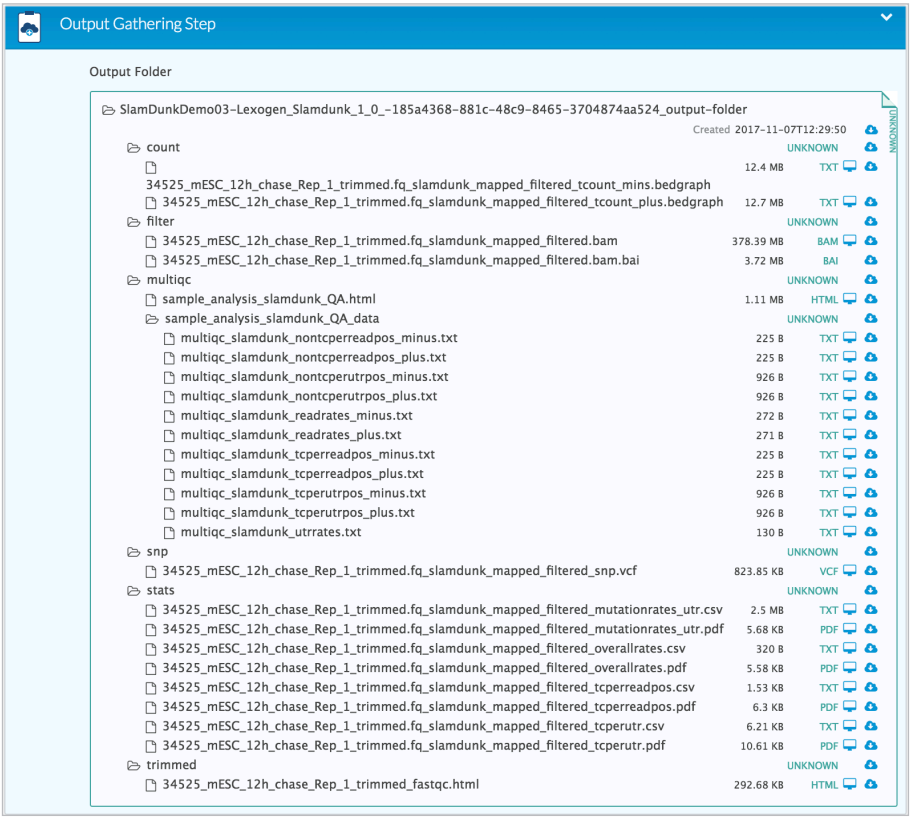
3. The *snp* step is used to call variants on the final filtered alignments of the *filter* step. Any called T->C SNPs from this step will be excluded in the subsequent analyses to reduce the false-positive number. Output files from this step will be vcf files for each of the samples in the input. Results of this step can be found in directory "snp".

4. The *count* step calculates all relevant numbers on statistics of SLAMSeq reads for each given 3'UTR. Central output will be *tcount* files for the samples in the input. Results of this step can be found in directory "count".

5. The alleyoop step will produce summary statistics, a PCA for all samples, and several conversion rate plots (read based, UTR based, read position based). Results of this step can be found in file summary_PCA.pdf and directories "multiqc" and "stats".

5.4. Output

The results of the SLAMdunk pipeline are stored in the directories mentioned in the previous section. These directories are in the output folder of the executed run. Results can either be downloaded in batch or by selecting the relevant files.



The *tcount* file is the central output file of *slamdunk*. It contains all results, conversion rates and other statistics for each UTR which is the main starting point for any subsequent analysis that will follow e.g. transcript half-life estimates or DE analysis.

Tcount files are essentially tab-separated text files containing one line entry per 3'UTR supplied by the user. Each entry contains the following columns:

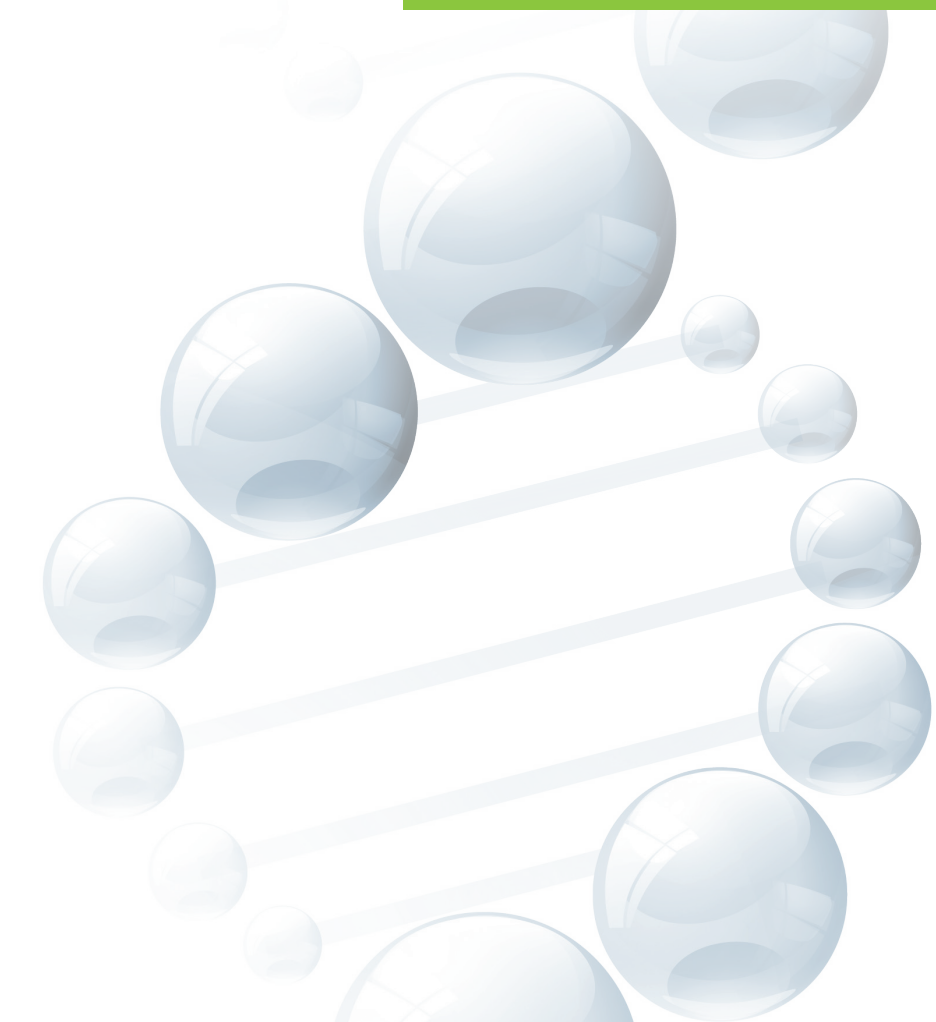
Column	Datatype	Description
chromosome	String	Chromosome on which the 3′ UTR resides
start	Integer	Start position of the 3′ UTR
end	Integer	End position of the 3′ UTR
name	String	Name or ID of the 3′ UTR supplied by the user
length	Integer	Length of the 3′ UTR
strand	String	Strand of the 3′ UTR
conversionRate	Float	Average conversion rate (see Conversion rates)
readsCPM	Float	Normalized number of reads as counts per million
Tcontent	Integer	Number of Ts in the 3′ UTR (As for - strand UTRs)
coverageOnTs	Integer	Cumulative coverage on all Ts in the 3′ UTR
conversionsOnTs	Integer	Cumulative number of T>C conversions in the 3′ UTR
readCount	Integer	Number of reads aligning to the 3′ UTR
tcReadCount	Integer	Number of reads with T>C conversions aligning to the 3′ UTR
multimapCount	Integer	Number of reads considered as multimappers aligning to the 3′ UTR

6. Revision History

Publication No. / Revision Date	Change	Page
063UG147V0100 Nov. 9, 2017	Initial Release.	

LEXOGEN

Enabling complete transcriptome sequencing



SLAMseq Data Analysis Pipeline SLAMdunk on Bluebee Platform

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: info@lexogen.com
© Lexogen, 2017