

MAPPING RIBO-SEQ AND RNA-SEQ SAMPLES to GENOME in RIBOGALAXY

This is the third part of the tutorial. Here the user can learn how to map Ribo-Seq and matching RNA-Seq reads to a genome and prepare files that can be transferred to GWIPs-viz and visualised there using the RiboGalaxy instance available at <https://ribogalaxy.genomicsdatascience.ie/>. At the end of the tutorial the user will be able to see the alignments in the genome context.

As input, you will need clean FASTQ files from part 1 of the tutorial (no adapters, UMIs, untemplated additions, barcodes and ncRNA contaminants). We assume here that you have the files in the RiboGalaxy instance and we are going to continue working with it.

Mapping RIBO-seq to a genome

Let's briefly go through the main steps.

1. First we will map clean reads (ribosome protected fragments) to the genome using **Bowtie Genome Alignment**.
2. Then we sort alignments based on coordinates using **Samtools sort**.
3. Next we will deduplicate a sample based on UMIs using **UMI-tools deduplicate**.
4. Then we again sort alignments based on coordinates using Samtools sort.
5. We need to obtain chromosome sizes by using the **Get Chromosome Sizes** tool.
6. Using the deduplicated bam file from step 4, now we can create ribosome profile in bed format using **Create Ribosome Profiles** tool.
7. Next we will convert the bed file to BigWig so that it can be uploaded and visualised in GWIPS-viz by using **Convert a BED File to a BigWig** tool.
8. In order to upload the resulting BigWig track file to GWIPS-viz, we will need to create a link using the Generate **Custom Track** tool.

1. Map reads (RPFs) that were cleaned from rRNA/tRNA contamination (after the last step of STAGE I) on the genome using **Bowtie Genome Alignment** in **GWIPS-Viz (genomic alignment) branch**. You can choose the built-in index *Homo sapiens (hg38) Genome*. Click **Execute**.

Bowtie Genome Alignment Align to the Genome with Bowtie (Galaxy Version 1.6.0) ☆ ▾

Will you select a reference genome from your history or use a built-in index?

Use a built-in index ▾

Built-ins were indexed using default options

Select a reference genome

Homo sapiens (hg38) Genome ▾

if your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?

Single-end ▾

FASTQ file

12: Bowtie ncRNA Removal on data 10: unmapped reads (L) ▾

2. Sort the resulting bam file by coordinate using **Samtools sort** from **GWIPS-Viz (genomic alignment) branch**. Click **Execute**.

Samtools Sort Sort Alignment Files (SAM/BAM) (Galaxy Version 2.0.5) ☆ ▾

BAM File

58: Bowtie Genome Alignment on data 12: mapped reads ▾

Primary sort key

coordinate ▾

3. This step can be performed if you want to deduplicate your sample (remove PCR duplicates) using UMIs. It can be done with **UMI-tools deduplicate** from **UMI and barcodes**. Input file is bam sorted by coordinate. You can also choose to output log file. Click **Execute**.

UMI-tools deduplicate Extract UMI from fastq files (Galaxy Version 1.1.2+galaxy2) ☆ ▾

Reads to deduplicate in SAM or BAM format

59: Samtools Sort on data 58 ▾

Output log?

☒ Yes

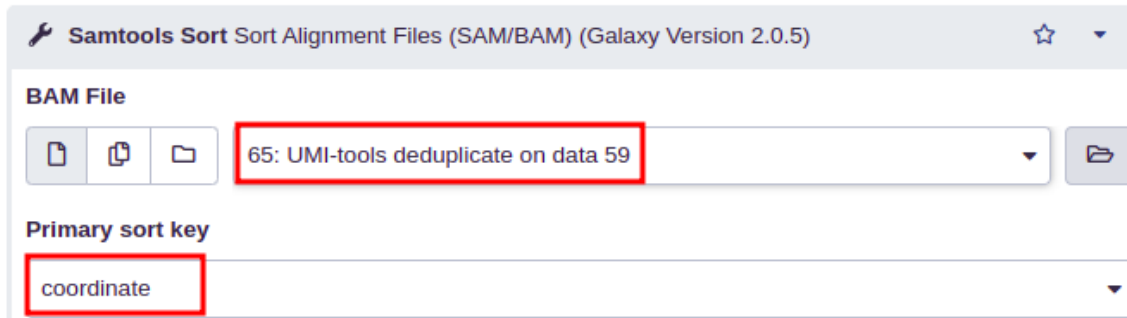
Choose if you want to generate a text file containing logging information (--log)

You can see in the log file that only 3639 reads were in input and 3403 are left after deduplication. The original sample had about 100k reads. This significant downsampling happened because most of the reads were multimappers (remember

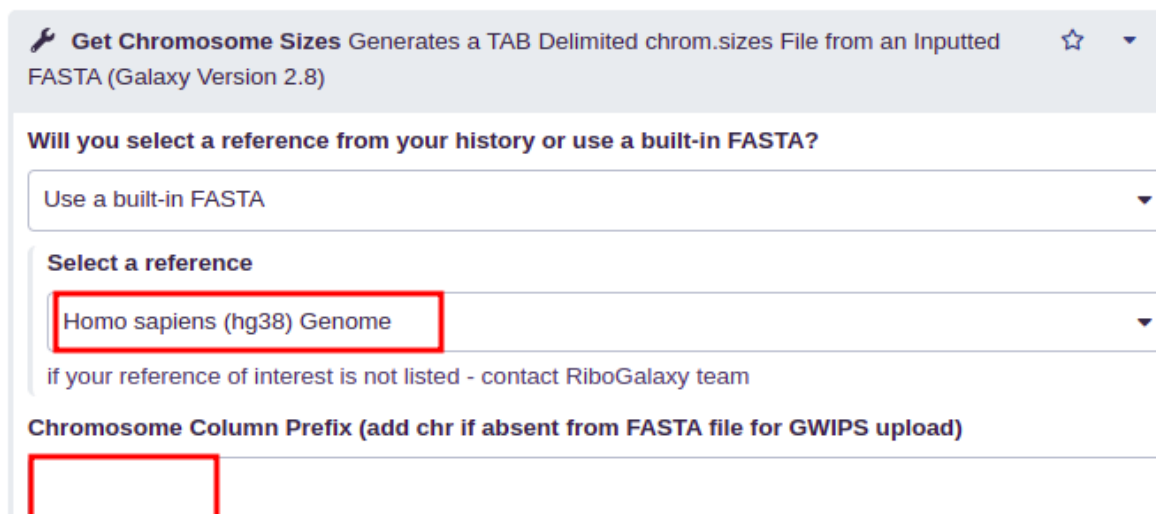
that we had to tick 'Allow ambiguously mapped reads' in part 2 of the tutorial?) and when we prepare samples for GWIPs-viz, we remove multimappers.

```
2023-02-04 15:24:03,797 INFO Reads: Input Reads: 3639
2023-02-04 15:24:03,797 INFO Number of reads out: 3403
```

4. Sort the resulting deduplicated bam file by coordinate using **Samtools sort** from **GWIPS-Viz (genomic alignment) branch**. Click **Execute**.



5. Now we will get chromosome sizes for subsequent transformation of alignment to the ribosome profile. We will use the **Get Chromosome Sizes** tool from **GWIPS-Viz (genomic alignment) branch**. Select built-in fasta, Homo sapiens (hg38) Genome. Click **Execute**.



6. Create ribosome profile in bed format from sorted by coordinate deduplicated bam file from step 4 using **Create Ribosome Profiles** tool in **GWIPS-Viz (genomic alignment) branch**. Choose built-in fasta, *Homo sapiens (hg38) Genome*. Click **Execute**.

Create Ribosome Profiles Produce a Sorted BED File of A-Site profiles of RPFs. (Galaxy ☆ Version 1.1)

BAM file to process

73: Samtools Sort on data 65

Offset to use

15

Use 15 for elongating ribosomes, 12 for initiating and 0 form RNA-seq reads

Will you select a reference from your history or use a built-in FASTA?

Use a built-in FASTA

Select a reference

Homo sapiens (hg38) Genome

7. Next we need to convert the bed file from step 6 to BigWig file. BigWig can be uploaded to genome browsers, e.g. GWIPS-viz. We will use the **Convert a BED File to a BigWig** tool from **GWIPS-Viz (genomic alignment) branch**. Another input is chromosome sizes file from step 3. Click 'Execute'.

Convert a BED File to a BigWig Convert Ribosome Profile BED File to a BigWig (Galaxy ☆ Version 1.2)

Bed File

75: Create Ribosome Profiles on data 73

Chromosome Sizes

74: Get Chromosome Sizes

8. Finally, we can generate the file for easy upload of the BigWig file generated on step 7 to the GWIPS-viz browser. We will use **Generate Custom Track** from **GWIPS-Viz (genomic alignment) branch**. You will need to copy a link to the BigWig file and use it as input. Add name and description of the sample, as well as any chromosome position of interest, e.g. chr7:5,526,409-5,530,601 (it can be changed in genome browser). Click 'Execute'. This tool will output a file containing a link. You need to download this file and then upload to the GWIPS-viz browser.

Generate Custom Track Generates a Custom Track File for GWIPS-viz. (Galaxy Version 0.2)

File Type
bigWig

URL of File (copy link from history)
https://ribogalaxy.genomicsdatascience.ie/datasets/94b5131a2d9d1199/display?to_ext=bigwig

Name of this sample
RIBO

Description of this sample
RIBO

chromosome position
chr7:5,526,409-5,530,601

Get this from Gwips.ucc.ie. Hg38 example = chr9:136,848,259-136,851,600

History

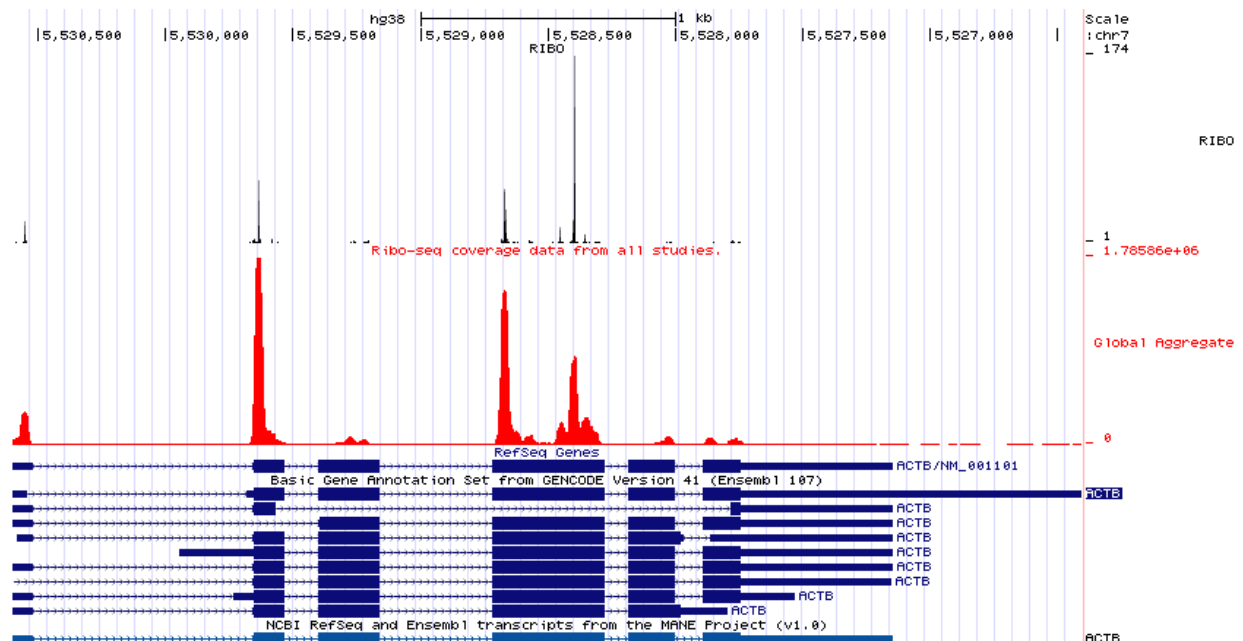
search datasets

Unnamed history
60 shown, 8 deleted, 8 hidden
45.76 GB

77: Generate Custom Track
2 lines
format: txt, database: ?
Copy link text into the dialog box at # or download the file and upload to GWIPS browser position chr15:44,711,487-44,716 track type=bigWig name="RIBO" descriptive

In order to upload this file, go to <https://gwips.ucc.ie/cgi-bin/hgGateway>, select **My Data, Custom Tracks** and **add custom track**. Choose **file** (or copy and paste its content) and **submit**.

You then can take a look at the data. Depending on your browser configurations, you will see different tracks. One of them should be the track 'RIBO' (you can right click on that and choose 'full' to see its expanded version). Also we enabled 'Global Aggregate' of Elongating Ribosomes (A-sites), track coloured in red. You can see here that there are similarities between these tracks.



Mapping RNA-seq to a genome

The main steps are similar to the Ribo-Seq sample except for some last steps.

1. First we will map clean reads (ribosome protected fragments) to the genome using **Bowtie Genome Alignment**.
2. Then we sort alignments based on coordinates using **Samtools sort**.
3. We need to obtain chromosome sizes by using the Get **Chromosome Sizes tool**.
4. Using the sorted BAM file from step 2, we will create a coverage file in BED format using **BedTools Genome Coverage**.
5. Next we will convert the BED file to BigWig so that it can be uploaded and visualised in GWIPS-viz by using **Convert a BED File to a BigWig** tool.
6. In order to upload the resulting BigWig track file to GWIPS-viz, we will need to create a link using the Generate **Custom Track** tool.

1. Map clean RNA-seq reads to the genome using **Bowtie Genome Alignment** in **GWIPS-Viz (genomic alignment) branch**. You can choose the built-in index *Homo sapiens (hg38) Genome*. Click 'Execute'.

2. Sort the resulting bam file by coordinate using **Samtools sort** from **GWIPS-Viz (genomic alignment) branch**. Click 'Execute'.

3. Now we will get chromosome sizes for subsequent transformation of alignment to the ribosome profile. We will use the **Get Chromosome Sizes** tool from **GWIPS-Viz (genomic alignment) branch**. Select built-in fasta, Homo sapiens (hg38) Genome. Click 'Execute'.

Get Chromosome Sizes Generates a TAB Delimited chrom.sizes File from an Inputted FASTA (Galaxy Version 2.8)

Will you select a reference from your history or use a built-in FASTA?

Use a built-in FASTA

Select a reference

Homo sapiens (hg38) Genome

if your reference of interest is not listed - contact RiboGalaxy team

Chromosome Column Prefix (add chr if absent from FASTA file for GWIPS upload)

4. Create coverage file in BED format from sorted by coordinate BAM file from step 2 using **BedTools Genome Coverage** tool in **GWIPS-Viz (genomic alignment) branch**. Choose built-in fasta, *Homo sapiens (hg38) Genome*. Click 'Execute'.

BedTools Genome Coverage Compute Read Coverage Over An Entire Genome (Galaxy Version 2.30.2)

Input type

BAM

BAM file

81: Samtools Sort on data 79

5. Next we need to convert the bed file from step 4 to BigWig file. BigWig can be uploaded to genome browsers, e.g. GWIPS-viz. We will use the **Convert a BED File to a BigWig** tool from **GWIPS-Viz (genomic alignment) branch**. Another input is chromosome sizes file from step 3. Click 'Execute'.

Convert a BED File to a BigWig Convert Ribosome Profile BED File to a BigWig (Galaxy Version 1.2)

Bed File

82: BedTools Genome Coverage on data 81

Chromosome Sizes

80: Get Chromosome Sizes

6. Finally, we can transform the BigWig file from the previous step for easy upload to the GWIPS-viz browser. We will use **Generate Custom Track** from **GWIPS-Viz (genomic alignment) branch**. You will need to copy a link to the BigWig file and use it as input. Add name and description of the sample, as well as any chromosome position of interest, e.g. chr7:5,526,409-5,530,601 (it can be changed in genome browser). Click **Execute**. This tool will output a file containing a link. You need to download this file and then upload to the GWIPS-viz browser.

Generate Custom Track Generates a Custom Track File for GWIPS-viz. (Galaxy Version 0.2)

File Type
bigWig

URL of File (copy link from history)
https://ribogalaxy.genomicsdata.science.ie/datasets/f6a51c66e5717573/display?to_ext=bigwig

Name of this sample
RNA

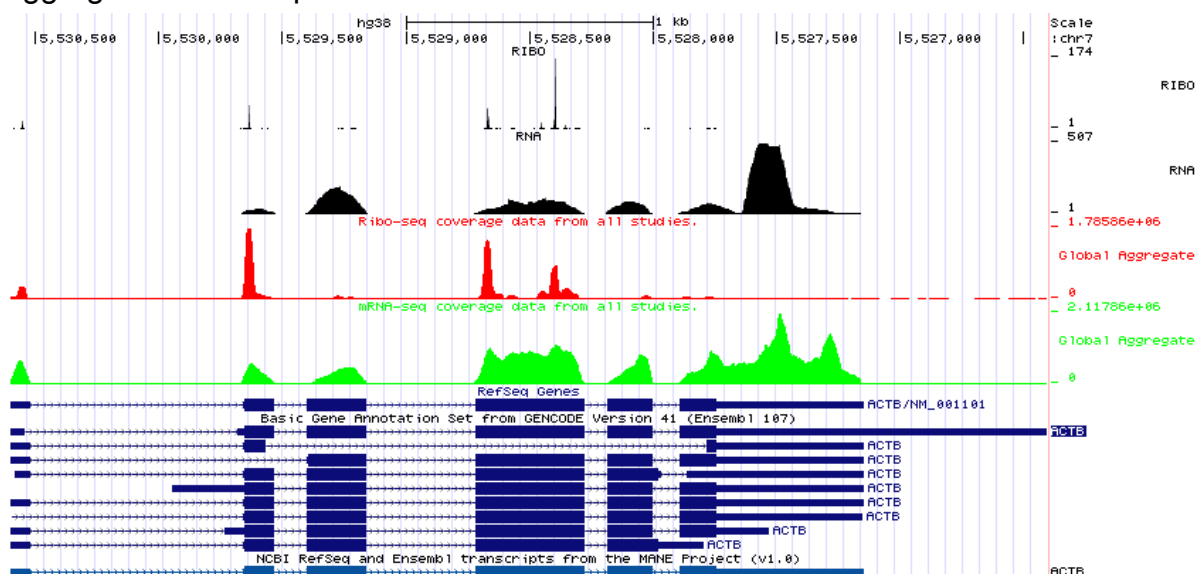
Description of this sample
RNA

chromosome position
chr7:5,526,409-5,530,601

History
search datasets
Unnamed history
66 shown, 15 deleted, 11 hidden
45.78 GB

84: Convert a BED File to a BigWig on data 80 and data 82
46.4 KB
format: bigwig, database: hg38
Copy link CSC main test display in IGB View display with IGV web current local

In order to upload this file, go to <https://gwips.ucc.ie/cgi-bin/hgGateway>, select **My Data, Custom Tracks** and **add custom track**. **Choose the file** (or paste its content) and **submit**. You then can take a look at the data. Depending on your browser configurations, you will see different tracks. One of them should be the track 'RIBO', another one is 'RNA'. Also we enabled 'Global Aggregate' of Elongating Ribosomes (A-sites), track coloured in red and 'Global Aggregate' of RNA-seq reads, track is green. You can see here that there are similarities between RNA-seq aggregate and our uploaded 'RNA' track.



Workflow for Ribo-Seq sample

For the Ribo-seq sample, we also created a Workflow which was tailored (all the parameters are set) to process the particular read structure from raw data to BED file. Generation of file compatible with GWIPs-viz requires an extra step. First, you need to choose 'All workflows' on the left panel and then select 'Run workflow' for GWIPs_viz_pipeline.

The screenshot shows the Galaxy RiboGalaxy interface. On the left, the 'Tools' panel is visible with a search bar and a list of categories. The 'WORKFLOWS' category is highlighted, and 'All workflows' is selected. In the main panel, a table lists workflows. The first workflow, 'imported: GWIPs_viz_pipeline', is highlighted. A red box highlights the 'Run workflow' button in the 'Bookmarked' column for this workflow.

Name	Tags	Updated	Sharing	Bookmarked
▼ imported: GWIPs_viz_pipeline raw fastq/fastq.gz files to BigWig files and links acceptable for GWIPs-viz		2 months ago		<input type="checkbox"/>
▼ imported: Trips_viz_pipeline Process data for upload to Trips-Viz		2 months ago		<input type="checkbox"/>

Next you need to make sure that the input file is correct and hit 'Run Workflow'.

Workflow: imported: GWIPs_viz_pipeline

✓ Run Workflow

2: Input dataset

202: RIBO_human.fq

Now can can just monitor the progress of the Workflow:

✓ Successfully invoked workflow **imported: GWIPs_viz_pipeline**.
You can check the status of queued jobs and view the resulting data by refreshing the History pane, if this has not already happened automatically.

Invocation 1...

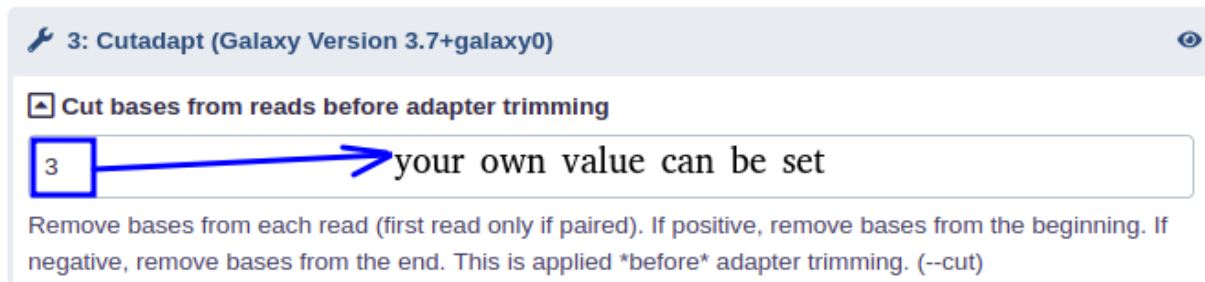
14 of 14 steps successfully scheduled.

5 of 13 jobs complete....

- Inputs
- Outputs
- Steps

Once it's finished, you will have a BED file and you can use **Generate Custom Track** from **GWIPS-Viz (genomic alignment) branch**.

The workflow parameters are customisable:



3: Cutadapt (Galaxy Version 3.7+galaxy0)

▲ Cut bases from reads before adapter trimming

3 → your own value can be set

Remove bases from each read (first read only if paired). If positive, remove bases from the beginning. If negative, remove bases from the end. This is applied *before* adapter trimming. (--cut)

You can also create your own workflows.