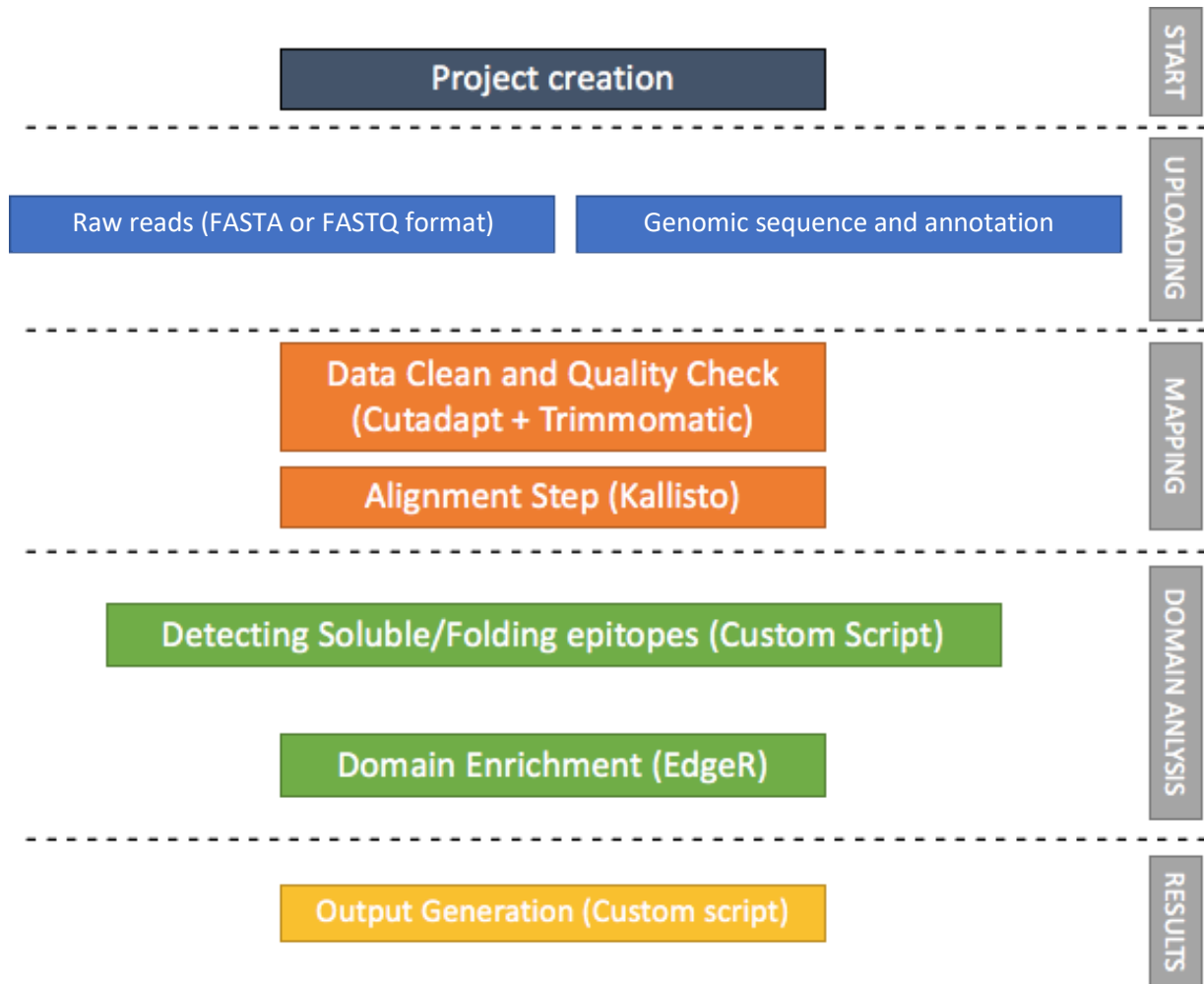


Eukaryote - Tutorial RnaBindProt

This introductory section provides an overview of **Eukaryote** pipeline drafting and design. The vertical gray rectangles correspond to the website sections.



This document shows how to perform all the steps of an example analysis with the data provided for Eukaryote pipeline. For further information about the forms and the interpretation of the results, refer to the User Guide.

Input Files: dataset Eukaryote_RnaBindProt.zip

[Download a zip folder containing the raw datasets from this link \(105 MB\).](#)

Mandatory inputs for **InteractomeSeq - Eukaryote RnaBindProt** execution are:

- Raw Data files, FASTA or FASTQ format for query reads are allowed in the input, therefore the web interface additionally allows the submission of compressed files (gz format) to reduce the time of data upload (DataSets).
 - Delta5Uchl1ssDNA.fastq (86 MB)
 - Delta5Uchl1tRNA.fastq (7 MB)
 - invSINEB2ssDNA.fastq (4 MB)
 - invSINEB2tRNA.fastq (4 MB)
 - NotSelected.fastq (4 MB)

InteractomeSeq requires at least two datasets. The input datasets must be generated with the same sequencing platform.

Time and memory usage for this tutorial

Analysis Step	File	Excution time	Output size
Mapping	NS	~ 3m	11 MB
Mapping	SP	~ 3m	12 MB
Mapping	BIO	~ 3m	15 MB
Domain Definition	NS	~ 3m	829 KB
Domain Definition	SP	~ 3m	945 KB
Domain Definition	BIO	~ 4m	999 KB
Domain Enrichment	NS + SP	< 5s	308 KB
Domain Enrichment	NS + BIO	< 5s	309 KB

Create a project


Click the button **START** to create a new project. Give the project a name (mandatory), a description (optional) and an email address (to receive messages during the execution. Select **EUKARYOTE** as project type and **SAVE**.

Create a Project

Project Name ⓘ	Rna Binding Protein
Project Description ⓘ	Rna Binding Protein Project
Email Address ⓘ	my@email.com
Project Type ⓘ	<input checked="" type="radio"/> EUKARYOTE <input type="radio"/> PROKARYOTE

The project now has an ID that is a link, active for 15 days. Save this link if you didn't provide an email address (the link will be sent via email).

Information Summary



Project Name	Rna Binding Protein
Project ID	z1ggax97ctbr4fd5ldubmzhuat5 %
Project Description	Rna Binding Protein Project
Project Type	Eukaryote
Project Status	
Creation Date	Monday March 9, 2020 - 16:34:08
Expiration Date	Tuesday March 24, 2020 - 16:34:08

Uploading

Click on **UPLOADING** to upload the FASTQ datasets.

Select the annotation to use for the dataset mapping. In this case, select **HOMO SAPIENS**. Pre-loaded nucleotide sequences of all protein-coding transcripts and annotation file are available for Homo Sapiens and Mus Musculus genome. The genome assembly version for Homo Sapiens is the GRCh38 downloaded from NCBI and the annotation was downloaded from GenCode consortium <https://www.genecodegenes.org/human/>. The genome assembly version for Mus Musculus is the GRCm38 downloaded from NCBI and the annotation was downloaded from GenCode consortium <https://www.genecodegenes.org/mouse/>.

Annotation

Organism

Annotation Strain **HOMO SAPIENS - GRCH38 P10**

PREVIEW

Chromosome	Start	End	Strand	Gene Name	Description
1	65419	71585	+	ENST00000641515.1	ID=ENST00000641515.1;gene_id=ENSG00000186092.5; gene_name=OR4F5;protein_id= ENSP00000493376.1;olfactory receptor family 4 subfamily F member 5 [Source:HGNC Symbol;Acc:HGNC:14825]
1	69055	70108	+	ENST00000335137.4	ID=ENST00000335137.4;gene_id=ENSG00000186092.5; gene_name=OR4F5;protein_id= ENSP00000334393.3;olfactory receptor family 4 subfamily F member 5 [Source:HGNC Symbol;Acc:HGNC:14825]
1	450703	451697	-	ENST00000426406.3	ID=ENST00000426406.3;gene_id=ENSG00000284733.1; gene_name=OR4F29;protein_id= ENSP00000409316.1;olfactory receptor family 4 subfamily F member 29 [Source: HGNC Symbol;Acc:HGNC:31275]
1	685679	686673	-	ENST00000332831.4	ID=ENST00000332831.4;gene_id=ENSG00000284662.1; gene_name=OR4F16;protein_id= ENSP00000329982.2;olfactory receptor family 4 subfamily F member 16 [Source: HGNC Symbol;Acc:HGNC:15079]
1	923928	939291	+	ENST00000420190.6	ID=ENST00000420190.6;gene_id=ENSG00000187634.11; gene_name=SAMD11;protein_id= ENSP00000411579.2;sterile alpha motif domain containing 11 [Source:HGNC Symbol; Acc:HGNC:28706]

Navigation: 1 2 3 ... 100 5 10 25

In the **DataSets** section, click on **SELECT FILE** or drag&drop files to upload FASTQ datasets:

- Delta5Uchl1ssDNA.fastq.gz
- Delta5Uchl1tRNA.fastq.gz
- invSINEB2ssDNA.fastq.gz
- invSINEB2tRNA.fastq.gz
- NotSelected.fastq.gz


DataSets



Raw Data Files

FASTA/FASTQ Format **SELECT FILE** Drop File

Dataset loading visualization:

Raw Data Files

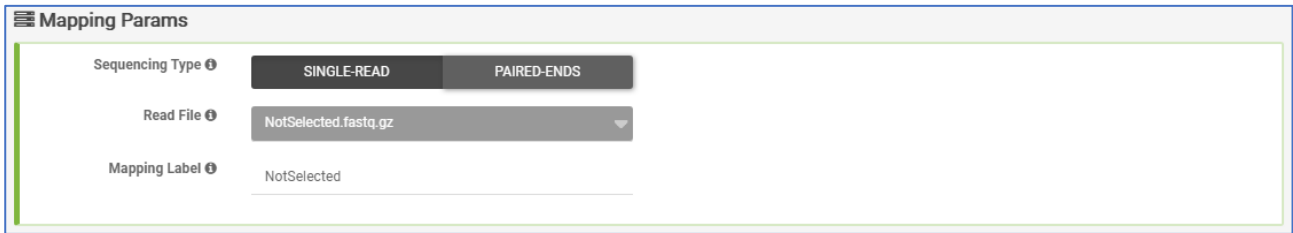
FASTA/FASTQ Format  **SELECT FILE** Drop File

File	Size	Date	Progress	Status	Actions
Delta5Uchl1ssDNA.fastq.gz	86.10 MB		<div style="width: 79%;"><div style="width: 79%; background-color: #ffc107;">79%</div></div>		 CANCEL
Delta5Uchl1tRNA.fastq.gz	7.27 MB		<div style="width: 0%;"><div style="width: 0%; background-color: #ffc107;">0%</div></div>		
invSINEB2ssDNA.fastq.gz	3.71 MB		<div style="width: 0%;"><div style="width: 0%; background-color: #ffc107;">0%</div></div>		
invSINEB2tRNA.fastq.gz	4.30 MB		<div style="width: 0%;"><div style="width: 0%; background-color: #ffc107;">0%</div></div>		
NotSelected.fastq.gz	3.52 MB		<div style="width: 0%;"><div style="width: 0%; background-color: #ffc107;">0%</div></div>		

Mapping

In the **MAPPING** section, by clicking on the button **+ MAPPING**, 4 sub-sections will appear on the screen:

1. **Mapping Params.** Select single-end reads (in this example: NotSelected)



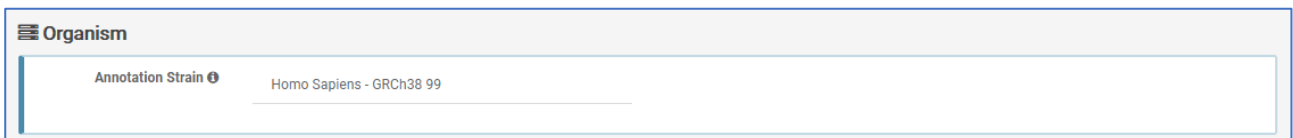
Mapping Params

Sequencing Type SINGLE-READ PAIRED-ENDS

Read File

Mapping Label

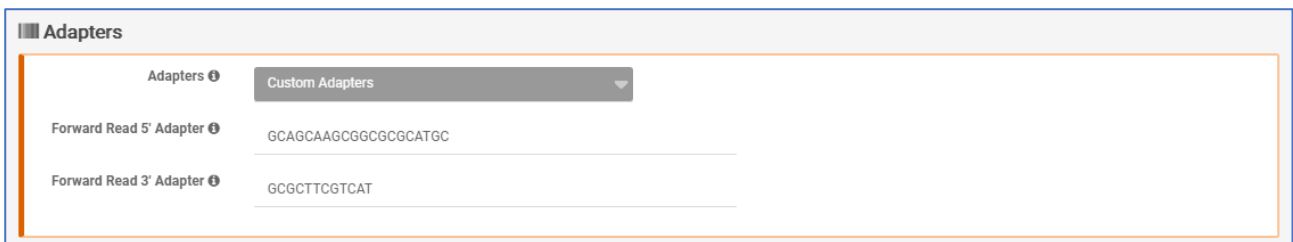
2. **Organism.** Preloaded FASTA file that will be used as reference to align the sequences (Homo Sapiens).



Organism

Annotation Strain

3. **Adapters.** Select **Custom Adapters** and insert:
 - a. Forward Read 5' Adapter GCAGCAAGCGGCGCATGC
 - b. Forward Read 3' Adapter GCGCTTCGTCAT



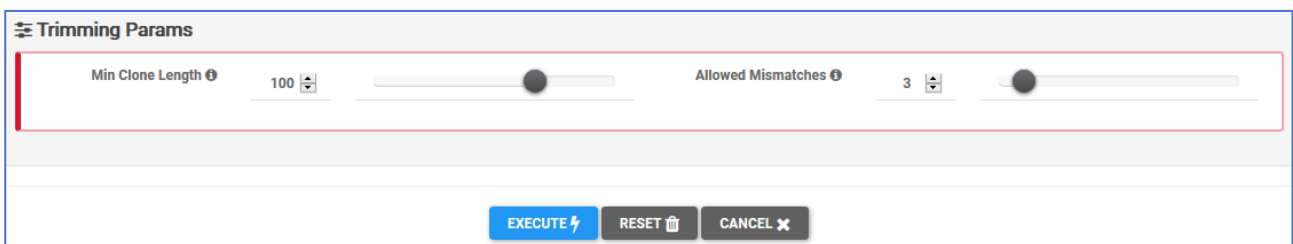
Adapters

Adapters

Forward Read 5' Adapter

Forward Read 3' Adapter

4. **Trimming Params.** Selection of minimum length of sequence and number of allowed mismatches, reads below this thresholds will be discarded. Use the default, and click **EXECUTE**

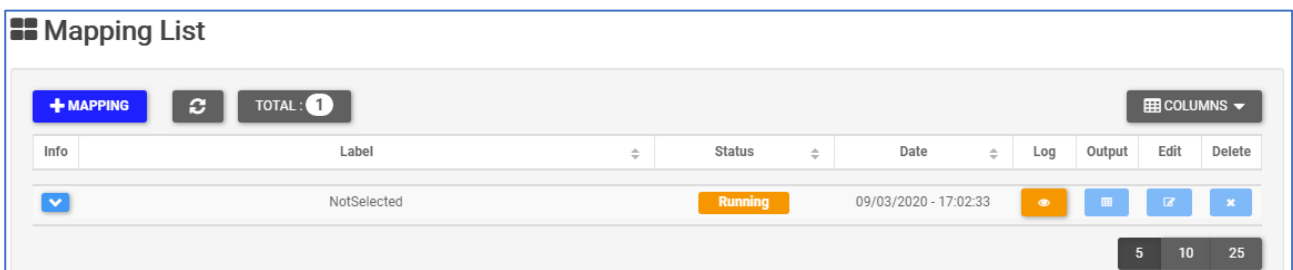


Trimming Params

Min Clone Length

Allowed Mismatches

This mapping will appear in the list of running mappings.



Mapping List

TOTAL: 1

Info	Label	Status	Date	Log	Output	Edit	Delete
<input type="button" value="v"/>	NotSelected	Running	09/03/2020 - 17:02:33	<input type="button" value="👁"/>	<input type="button" value="📄"/>	<input type="button" value="✎"/>	<input type="button" value="✕"/>

Info – Drop-down menu with information of Mapping input file.

Label – Sample label.

Status – When the execution ends successfully, the button turns green, otherwise, it turns red.

Date – Day and time of analysis execution

Log – Button that hides/opens a box with execution log file. When the execution is running, the log shows the possibility to stop it:

☰ Mapping :: Log :: NotSelected

☰ Mapping :: Log :: NS .

☰ STATUS ⌵ STOP ✕ Mapping Running ⌚ Processing

```

Setting Parameters
Sequencing input: Single-End
Input dataset: NotSelected.fastq
Input dataset type: fastq
Primer 5' read: GCAGCAAGCGGCGCATGCCACTAGTGGGAT
Primer 3' read: ATCTCCGCTAGCGGCAAACCAATCCC
Number of reads: 67587
Sample Name: NS___
Sequence file: NS___Homo_sapiens.GRCh38.99.fasta
Fasta name file: NS___
Cutadapt version: 1.12
-----
Single-end analysis start.
Trimming of 5 primers is complete.
  
```

When the process has finished, the log shows summary information:

☰ Mapping :: Log :: NotSelected

☰ STATUS ⌵ Mapping Done ✓ Completed Processing

```

Setting Parameters
Sequencing input: Single-End
Input dataset: NotSelected_forward.fastq
Input dataset type: fastq
Primer 5' read: GCAGCAAGCGGCGCATGC
Primer 3' read: GCGCTTCGTCAT
Number of reads: 155708
Sample Name: NotSelected
Sequence file: NotSelected.Homo_sapiens.GRCh38.99.fasta
Fasta name file: NotSelected
Cutadapt version: 1.12
-----
Single-end analysis start.
Trimming of 5 primers is complete.
Trimming of 3 primers is complete.
Trimming complete. Starting mapping.
Estimating average and standard deviation of fragment length complete.
Kallisto mapping complete.
Kallisto mapping complete.
Sorting of BAM file complete.
Conversion SAM to BAM is complete.
Conversion SAM to BAM is complete.
Conversion SAM to BAM is complete.
Conversion SAM to BAM is complete.
Conversion SAM to BAM is complete.
  
```

Output – Hides/opens panel with download button

☰ Mapping :: Output :: NotSelected



Mapping Output File [↓ DOWNLOAD](#)


























Here are the parameters for the other 4 mappings in this example:

1. Delta5Uchl1ssDNA
 - SequencingType** Single-Read
 - Forward Read File** Delta5Uchl1ssDNA.fastq.gz
 - Annotation Strain** Homo Sapiens - GRCh38 p10
 - AdapterType** Custom
 - Forward Read 5' Adapter** GCAGCAAGCGGCGCGCATGC
 - Forward Read 3' Adapter** GCGCTTCGTCAT
 - Min Clone Length** 100
 - Allowed Mismatches** 3
2. Delta5Uchl1tRNA
 - SequencingType** Single-Read
 - Forward Read File** Delta5Uchl1tRNA.fastq.gz
 - Annotation Strain** Homo Sapiens - GRCh38 p10
 - AdapterType** Custom
 - Forward Read 5' Adapter** GCAGCAAGCGGCGCGCATGC
 - Forward Read 3' Adapter** GCGCTTCGTCAT
 - Min Clone Length** 100
 - Allowed Mismatches** 3
3. invSINEB2ssDNA
 - SequencingType** Single-Read
 - Forward Read File** invSINEB2ssDNA.fastq.gz
 - Annotation Strain** Homo Sapiens - GRCh38 p10
 - AdapterType** Custom
 - Forward Read 5' Adapter** GCAGCAAGCGGCGCGCATGC
 - Forward Read 3' Adapter** GCGCTTCGTCAT
 - Min Clone Length** 100
 - Allowed Mismatches** 3
4. invSINEB2tRNA
 - SequencingType** Single-Read
 - Forward Read File** invSINEB2tRNA.fastq.gz
 - Annotation Strain** Homo Sapiens - GRCh38 p10
 - AdapterType** Custom
 - Forward Read 5' Adapter** GCAGCAAGCGGCGCGCATGC
 - Forward Read 3' Adapter** GCGCTTCGTCAT
 - Min Clone Length** 100
 - Allowed Mismatches** 3

This is the list of running mappings:

Mapping List

+ MAPPING  **TOTAL: 5**  COLUMNS ▾

Info	Label	Status	Date	Log	Output	Edit	Delete
	Delta5Uchl1ssDNA	Done	09/03/2020 - 12:26:11				
	Delta5Uchl1tRNA	Done	09/03/2020 - 12:26:36				
	invSINEB2ssDNA	Done	09/03/2020 - 12:27:02				
	invSINEB2tRNA	Done	09/03/2020 - 12:27:29				
	NotSelected	Done	09/03/2020 - 12:29:01				

5 10 25

Domain Analysis

Domain analysis is composed by two sheets:

1. **Domain Definition**
2. **Domain Enrichment**

Domain Definition takes as input the mapping file previously generated. Define the domains for each of the 5 datasets uploaded and mapped by choosing them in the **Mapping** menu, then press **Execute**.

Domain Definition :: Insert

Domain Definition Params

Mapping ⓘ NotSelected

Domain Definition Label ⓘ NotSelected

Organism

Annotation Strain ⓘ Homo Sapiens - GRCh38 p10

EXECUTE ⚡
RESET 🗑️
CANCEL ✕

Domain Definition List shows the running domain definitions

Domain Definition List

+ DOMAIN DEFINITION
↻
TOTAL : 1
COLUMNS ▾

Info	Label	Status	Date	Log	Output	Edit	Delete
ⓘ	NotSelected	Running	09/03/2020 - 17:07:29	👁️	📄	🔗	✕

5
10
25

Info – Drop-down menu with information of Mapping input file.

Label – Sample label.

Status – When the execution ends successfully, the button turns green, otherwise, it turns red.

Date – Day and time of analysis execution

Log – Button that hides/opens a box with execution log file. When the execution is running, the log shows the possibility to stop it:

Domain Definition :: Log :: NotSelected

STATUS ⓘ
STOP ✕
Domain Definition Running ⓘ
Processing

Computing the depth-of-coverage complete.

CLOSE ✕

When the process has finished, the log shows summary information:

Domain Definition :: Log :: NotSelected

STATUS Domain Definition Done Completed Processing

Computing the depth-of-coverage complete.
 Computing the breadth-of-coverage complete.
 Bam2bed complete.
 Read count complete.
 Max depth coverage computing complete.
 Percentile depth filtering complete.
 Raw definition of domains complete.
 Computing domain start and end complete.
 Parsing output complete

Output – Hides/opens panel with output preview

Domain Definition :: Output :: NotSelected

Domain Definition Output File DOWNLOAD

TOTAL: 11,771 RESET FILTERING RESET SORTING COLUMNS

Info	Chromosome	Clone Start	Clone End	Clone Length	Transcript	Transcript Start	Transcript End	Gene	Strand	Read Count	Average Depth
▼	1	14528	15267	739	ENST00000488147.1	14404	29570	WASH7P	-	7	1.0000
▼	1	629358	629433	75	ENST00000416931.1	629062	629433	MTND1P23	+	1	1.0000
▼	1	631227	632375	1148	ENST00000414273.1	631074	632616	MTCO1P12	+	15	3.0000
▼	1	633894	634221	327	ENST00000514057.1	633696	634376	MTATP6P1	+	25	9.0000
▼	1	944877	945663	786	ENST00000483767.5	944204	947060	NOC2L	-	34	2.0000
▼	1	954527	954973	446	ENST00000487214.1	954426	959309	NOC2L	-	7	2.0000
▼	1	1013507	1014061	554	ENST00000649529.1	1013497	1014540	ISG15	+	24	5.0000
▼	1	1047333	1047615	282	ENST00000466223.1	1047201	1047865	AGRN	+	6	2.0000
▼	1	1087925	1088044	119	ENST00000442117.5	1087575	1092813	C1orf159	-	5	5.0000
▼	1	1217522	1218482	960	ENST00000655486.1	1217512	1228793	SDF4	-	120	6.0000

1 2 3 ... 1178 5 10 25 50

Genome Browser JBROWSE

Available Tracks filter tracks

- Annotation 2
 - Reference sequence
 - GRCh38.99 Annotation
- Domain Definition 1
 - NotSelected

Genome Track View Help Share

0 50,000,000 100,000,000 150,000,000 200,000,000

1,99591111..149382394 (49.79 Mb) Go

Reference sequence Zoom in to see sequence

GRCh38.99 Annotation per 500,000 bp Zoom in to see sequence

NotSelected Zoom in to see sequence

Domain Enrichment takes as input the Control and Selection output of Domain Definition step. Enrich domains for each defined domain by selecting the Genomic and the target domains. In the current example, select:

- Delta5Uchl1ssDNA and Delta5Uchl1tRNANotSelected
- invSINEB2ssDNA and invSINEB2tRNA

Domain Enrichment :: Insert

Control Domain Definition i Delta5Uchl1ssDNA

Selection Domain Definition i Delta5Uchl1tRNA

Domain Enrichment Label i Delta5Uchl1ssDNA - Delta5Uchl1tRNA

EXECUTE ⚡
RESET 🗑️
CANCEL ✕

Domain Enrichment List

Domain Enrichment List

+ DOMAIN ENRICHMENT
TOTAL: 2
COLUMNS ▾

Info	Label	Status	Date	Log	Output	Edit	Delete
▾	Delta5Uchl1ssDNA - Delta5Uchl1tRNA	Done	09/03/2020 - 13:23:20	👁️	📄	✎	✕
▾	invSINEB2ssDNA - invSINEB2tRNA	Done	09/03/2020 - 13:23:26	👁️	📄	✎	✕

5
10
25

Info – Drop-down menu with information of Domain Definition input file.

Label – Sample label.

Status – When the execution ends successfully, the button turns green, otherwise, it turns red.

Date – Day and time of analysis execution

Log – Button that hides/opens a box with execution log file. When the execution is running, the log shows the possibility to stop it:

Domain Enrichment :: Log :: Delta5Uchl1ssDNA - Delta5Uchl1tRNA

STATUS ▾
STOP ✕
Domain Enrichment Running ⏸
Processing

CLOSE ✕

When the process has finished, the log shows summary information:

Domain Enrichment :: Log :: Delta5Uchl1ssDNA - Delta5Uchl1tRNA

STATUS ▾
Domain Enrichment Done ✓
Completed Processing

Parsing input file complete.
 Parsing input file complete.
 Bedtools intersect of common domains complete.
 Bedtools intersect of unique domains complete.
 Parsing files for edgeR analysis complete.
 Differential expression analysis complete.
 Parsing output file with common domains complete.
 Parsing output file with unique domains complete.

Output – Hides/opens panel with output preview

Domain Enrichment :: Output :: Delta5Uchl1ssDNA - Delta5Uchl1tRNA

● Common Intervals -

Domain Enrichment Output File [DOWNLOAD](#)

TOTAL: 44
[RESET FILTERING](#)
[RESET SORTING](#)
[COLUMNS](#)

Info	Chromosome	Clone Start	Clone End	Clone Length	Transcript	Transcript Start	Transcript End	Gene	Strand	Read Count	Adjust PValue
<input type="checkbox"/>	1	39565139	39565956	817	ENST00000470443.5	39565106	39575781	PABPC4	-	6	2.9094e-3
<input type="checkbox"/>	1	154207104	154207740	636	ENST00000640799.1	154207098	154220636	C1orf43	-	3	1.0201e-2
<input type="checkbox"/>	1	154207104	154207740	636	ENST00000640799.1	154207098	154220636	C1orf43	-	3	8.8027e-3
<input type="checkbox"/>	2	85595756	85595875	119	ENST00000414390.5	85595748	85596820	RNF181	+	1	1.5924e-2
<input type="checkbox"/>	3	49674486	49674814	328	ENST00000463616.1	49674486	49676232	APEH	+	2	1.0201e-2
<input type="checkbox"/>	3	134351853	134351972	119	ENST00000273411.2	134351852	134356561	RPL39P5	-	2	3.4401e-2
<input type="checkbox"/>	5	140564828	140565122	294	ENST00000623481.2	140564828	140567117	SLC35A4	+	5	2.8317e-2
<input type="checkbox"/>	6	33416462	33417219	757	ENST00000462802.5	33416442	33418043	CUTA	-	7	1.6797e-2
<input type="checkbox"/>	6	33416462	33417219	757	ENST00000462802.5	33416442	33418043	CUTA	-	7	1.6797e-2
<input type="checkbox"/>	6	33416462	33417219	757	ENST00000462802.5	33416442	33418043	CUTA	-	7	1.6797e-2

« 1 2 3 4 5 » 5 10 25 50

[Unique Intervals](#) +

Genome Browser [JBROWSE](#)

Results

The summary of all the analysis performed is showed in the results page

Results

TOTAL : 3 COLUMNS ▾

Analysis ▾ ↗ ↘

	Label	Date	Download
Mapping			
	Delta5Uchl1ssDNA	01/10/2019 - 14:00:00	↓
	Delta5Uchl1tRNA	01/10/2019 - 14:00:00	↓
	invSINEB2ssDNA	01/10/2019 - 14:00:00	↓
	invSINEB2tRNA	01/10/2019 - 14:00:00	↓
	NotSelected	01/10/2019 - 14:00:00	↓
Domain Definition			
	Delta5Uchl1ssDNA	06/03/2020 - 10:51:39	↓
	Delta5Uchl1tRNA	06/03/2020 - 10:53:58	↓
	invSINEB2ssDNA	06/03/2020 - 10:54:01	↓
	invSINEB2tRNA	06/03/2020 - 10:54:04	↓
	NotSelected	06/03/2020 - 10:57:42	↓
Domain Enrichment			
	Delta5Uchl1ssDNA - Delta5Uchl1tRNA	06/03/2020 - 11:05:16	↓
	invSINEB2ssDNA - invSINEB2tRNA	06/03/2020 - 11:05:20	↓

2 5 10 25

And a genome browser with the results aligned to the reference genome

