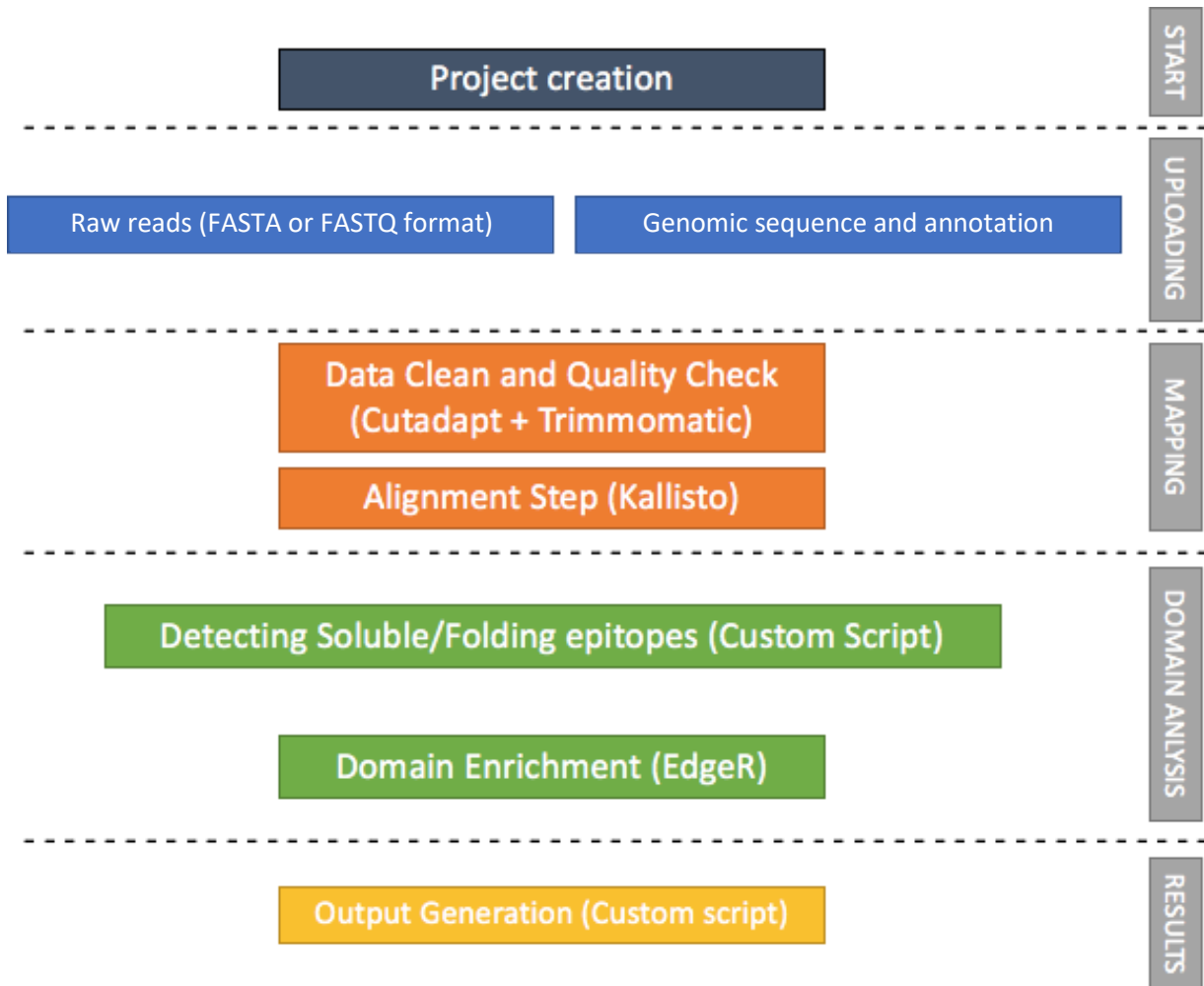


Eukaryote - Tutorial RIDome

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_RIDome.zip

[Download a zip folder containing the raw datasets from this link \(46 MB\).](#)

Mandatory inputs for **InteractomeSeq - Eukaryote** 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).
 - NotSelected.fastq (35 MB)
 - TG2_solid_phase_immobilized.fastq (47 MB)
 - TG2_soluble_biotinylated.fastq (69 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	Delta5Uchl1ssDNA	~ 17m	133 MB
Mapping	Delta5Uchl1tRNA	~ 4m	12 MB
Mapping	invSINEB2ssDNA	~ 4m	7.7 MB
Mapping	invSINEB2tRNA	~ 4m	7.0 MB
Mapping	NotSelected	~ 3m	6.3 MB
Domain Definition	Delta5Uchl1ssDNA	~ 23m	2.4 MB
Domain Definition	Delta5Uchl1tRNA	~ 4m	736 KB
Domain Definition	invSINEB2ssDNA	~ 4m	652 KB
Domain Definition	invSINEB2tRNA	~ 4m	686 KB
Domain Definition	NotSelected	~ 4m	1.5 MB
Domain Enrichment	Delta5Uchl1ssDNA - Delta5Uchl1tRNA	< 5s	193 KB
Domain Enrichment	invSINEB2ssDNA - invSINEB2tRNA	< 5s	226 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 ⓘ

Project Description ⓘ

Email Address ⓘ

Project Type ⓘ

EUKARYOTE

PROKARYOTE

SAVE ↵

CANCEL ✕

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	RIDome
Project ID	jwlf9lhq8cqk5o8t2bdccir488bz %
Project Description	RNA-interacting domainome (RIDome) Project

Project Type	Eukaryote
Project Status	

Creation Date

Tuesday March 10, 2020 - 13:21:41

Expiration Date

Wednesday March 25, 2020 - 13:21:41

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:

- NotSelected.fastq
- TG2_solid_phase_immobilized.fastq
- TG2_soluble_biotinylated.fastq

DataSets

Raw Data Files

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

Dataset loading visualization:

DataSets

Raw Data Files

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

File	Size	Date	Progress	Status	Actions
NotSelected.fastq	35.34 MB	12/20/2019 - 18:05:35	<div style="width: 100%; background-color: green; text-align: center;">100%</div>	✓	REMOVE
TG2_solid_phase_immobilized.fastq	47.26 MB		<div style="width: 68%; background-color: orange; text-align: center;">68%</div>	⚙	CANCEL
TG2_soluble_biotinylated.fastq	69.24 MB		<div style="width: 0%; background-color: orange; text-align: center;">0%</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)

The screenshot shows the 'Mapping Params' section with the following settings:

- Sequencing Type: SINGLE-READ (selected), PAIRED-ENDS
- Read File: NotSelected.fastq
- Mapping Label: NS

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

The screenshot shows the 'Organism' section with the following setting:

- Annotation Strain: Homo Sapiens - GRCh38 99

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

The screenshot shows the 'Adapters' section with the following settings:

- Adapters: Custom Adapters
- Forward Read 5' Adapter: CGCGCATG
- Forward Read 3' Adapter: GCTAG

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**

The screenshot shows the 'Trimming Params' section with the following settings:

- Min Clone Length: 100
- Allowed Mismatches: 3

Buttons: EXECUTE, RESET, CANCEL

This mapping will appear in the list of running mappings.

The screenshot shows the 'Mapping List' section with the following details:

- Buttons: + MAPPING, Refresh, TOTAL: 1, COLUMNS
- Table:

Info	Label	Status	Date	Log	Output	Edit	Delete
✓	NS	Running	11/03/2020 - 11:01:07	👁	📄	✏	✖
- Page Size: 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:

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: GCAGCAAAGCGGCGCGCATGCCACTAGTGGGAT
 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 :: NS

STATUS Mapping Done Completed Processing

Setting Parameters
 Sequencing input: Single-End
 Input dataset: NotSelected.fastq
 Input dataset type: fastq
 Primer 5' read: GCAGCAAAGCGGCGCGCATGCCACTAGTGGGAT
 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.
 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 :: NS

Mapping Output File **DOWNLOAD**

Here are the parameters for the other 3 mappings in this example:

1. TG2_solid_phase_immobilized
SequencingType Single-Read

Forward Read File TG2_solid_phase_immobilized.fastq

Annotation Strain Homo Sapiens - GRCh38 p10

AdapterType Custom

Forward Read 5' Adapter CGCGCATG

Forward Read 3' Adapter GCTAG

Min Clone Length 100

Allowed Mismatches 3

2. TG2_soluble_biotinylated

SequencingType Single-Read

Forward Read File TG2_soluble_biotinylated.fastq.fastq

Annotation Strain Homo Sapiens - GRCh38 p10

AdapterType Custom

Forward Read 5' Adapter CGCGCATG

Forward Read 3' Adapter GCTAG

Min Clone Length 100

Allowed Mismatches 3

This is the list of running mappings:

Mapping List									
+ MAPPING RELOAD TOTAL: 3 COLUMNS									
Info	Label	Status	Date	Log	Output	Edit	Delete		
<input type="checkbox"/>	NS	Done	11/03/2020 - 11:01:07						
<input type="checkbox"/>	SP	Running	11/03/2020 - 11:02:35						
<input type="checkbox"/>	BIO	Running	11/03/2020 - 11:02:52						

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 3 datasets uploaded and mapped by choosing them in the **Mapping** menu, then press **Execute**.

Domain Definition :: Insert

Domain Definition Params

Mapping ⓘ NS

Domain Definition Label ⓘ NS

Organism

Annotation Strain ⓘ Homo Sapiens - GRCh38 99

EXECUTE ⚡
RESET 🗑️
CANCEL ✕

Domain Definition List shows the running domain definitions

Domain Definition List

+ DOMAIN DEFINITION
↻
TOTAL : 3
COLUMNS ▾

Info	Label	Status	Date	Log	Output	Edit	Delete
▼	NS	Running	11/03/2020 - 11:07:06	👁️	📄	✏️	✕
▼	SP	Running	11/03/2020 - 11:07:21	👁️	📄	✏️	✕
▼	BIO	Running	11/03/2020 - 11:07:25	👁️	📄	✏️	✕

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 Definition :: Log :: NS

☰ 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 :: NS

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 :: NS

Domain Definition Output File DOWNLOAD

TOTAL: 7,090 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	185217	185453	236	ENST00000623083.4	185217	195411	WASH9P	-	5	4.0000
▼	1	629373	629433	60	ENST00000416931.1	629062	629433	MTND1P23	+	1	1.0000
▼	1	729826	729958	132	ENST00000506640.2	725885	778626	AL669831.1	-	1	1.0000
▼	1	926845	927238	393	ENST00000620200.4	925741	944581	SAMD11	+	8	3.0000
▼	1	961225	961469	244	ENST00000622660.1	960639	965602	KLHL17	+	1	1.0000
▼	1	1013497	1013961	464	ENST00000649529.1	1013497	1014540	ISG15	+	15	12.0000
▼	1	1217591	1218510	919	ENST00000655486.1	1217512	1228793	SDF4	-	76	7.0000
▼	1	1255782	1255937	155	ENST00000400930.8	1255178	1273833	UBE2J2	-	2	1.0000
▼	1	1255969	1256124	155	ENST00000473215.5	1255268	1273849	UBE2J2	-	2	1.0000
▼	1	1293916	1294561	645	ENST00000353662.4	1293564	1303260	ACAP3	-	10	4.0000

« 1 2 3 ... 709 » 5 10 25 50

Genome Browser JBROWSE

Available Tracks

- Annotation 2
 - Reference sequence
 - GRCh38..99 Annotation
- Domain Definition 3
 - NS
 - BIO
 - SP

Genome Track View Help Share

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

1:99585113..148708464 (49.12 Mb) Go

Reference sequence sequence Zoom in to see sequence

GRCh38..99 Annotation per 500,000 bp

NS

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:

- NotSelected and TG2_solid_phase_immobilized (labeled NS + SP)
- NotSelected and TG2_soluble_biotinylated (labeled NS + BIO)

Domain Enrichment :: Insert

Control Domain Definition NS

Selection Domain Definition SP

Domain Enrichment Label NS + SP

EXECUTE ⚡
RESET 🗑️
CANCEL ✕

Domain Enrichment List

Domain Enrichment List

+ DOMAIN ENRICHMENT
🔄
TOTAL: 2
COLUMNS ▾

Info	Label	Status	Date	Log	Output	Edit	Delete
▼	NS + SP	Done	11/03/2020 - 11:14:46	👁️	📄	✎	✕
▼	NS + BIO	Running	11/03/2020 - 11:15:01	👁️	📄	✎	✕

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 :: NS + SP

☰ STATUS ⬇️
STOP ✕
Domain Enrichment Running ⏸️
Processing

CLOSE ✕

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

Domain Enrichment :: Log :: NS + SP

☰ 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 :: NS + SP

● Common Intervals -

Domain Enrichment Output File [DOWNLOAD](#)

TOTAL: 37
[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	45551070	45551592	522	ENST00000476957.5	45550854	45568624	AKR1A1	+	5	3.6443e-2
<input type="checkbox"/>	1	45551070	45551592	522	ENST00000476957.5	45550854	45568624	AKR1A1	+	5	3.6443e-2
<input type="checkbox"/>	1	45551070	45551592	522	ENST00000476957.5	45550854	45568624	AKR1A1	+	5	3.6443e-2
<input type="checkbox"/>	1	45551072	45551372	300	ENST00000481885.5	45551067	45568025	AKR1A1	+	5	3.6443e-2
<input type="checkbox"/>	1	110401264	110401781	517	ENST00000602318.5	110401253	110407708	LAMTOR5	-	7	4.2838e-2
<input type="checkbox"/>	1	110401264	110401781	517	ENST00000602318.5	110401253	110407708	LAMTOR5	-	7	4.2838e-2
<input type="checkbox"/>	1	209616868	209617266	398	ENST00000455193.1	209616581	209618798	LAMB3	-	15	5.0395e-3
<input type="checkbox"/>	4	39510709	39510896	187	ENST00000503779.5	39510709	39527437	UGDH	-	3	4.0054e-2
<input type="checkbox"/>	5	150401644	150402081	437	ENST00000377795.7	150401637	150412769	CD74	-	18	7.5449e-3
<input type="checkbox"/>	8	143579941	143581015	1074	ENST00000317198.10	143579697	143597388	EEF1D	-	63	3.4227e-2

« 1 2 3 4 »
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			
	NS	01/10/2019 - 14:00:00	Download
	SP	01/10/2019 - 14:00:00	Download
	BIO	01/10/2019 - 14:00:00	Download
Domain Definition			
	NS	01/10/2019 - 14:00:00	Download
	SP	01/10/2019 - 14:00:00	Download
	BIO	01/10/2019 - 14:00:00	Download
Domain Enrichment			
	NS + SP	01/10/2019 - 14:00:00	Download
	NS + BIO	01/10/2019 - 14:00:00	Download

2 5 10 25

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

