

Integration

- Share your data using Dolphin WebAPI (J)
- Compare your experiment results with
 - published data
 - your previous results
- Export meta data
- Export picard metrics and QC results
- Encode submission
- NCBI submission
- GEO import functionality

Processing

Running Pipelines

Common steps:

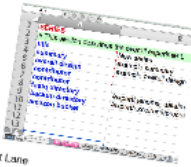
- Barcode separation (novobarcode)
- Adapter removal (small RNA-Seq)
- Fastq quality checks (FastQC)
- Quality filtering (trimomatic)
- 5' 3' Trimming (trimomatic)
- Mapping custom sequence sets and quantification (Bowtie2)
- Mapping common RNAs and quantification (rRNA, miRNA, tRNA, sRNA, piRNA, rRNA)
- Fastq splitter

Pipelines:

- RNA-Seq quantification (RSEM)
- Differential expression analysis (DESeq2)
- support for RSEM, mRNA, tRNA
- TopHat Pipeline (tophat2 and picard tools)
- Chip-Seq pipeline (Peak Calling and aggregation plots)
- Bisulphite Mapping for RRBS and WGBS

Meta-/Data Import

1. Excel Import

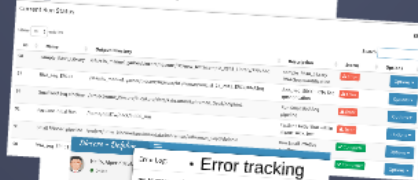


2. Fast Lane

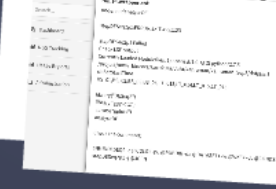


Job Monitoring

• Status screen



• Error tracking



RNA-Seq Data Processing with

Where is dolphin?

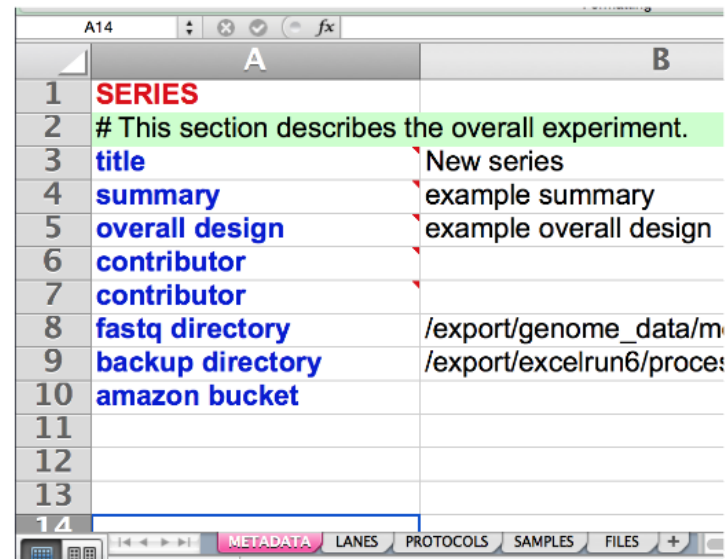
<http://dolphin.umassmed.edu>

Documentation

<http://dolphin.readthedocs.org>

Meta-/Data Import

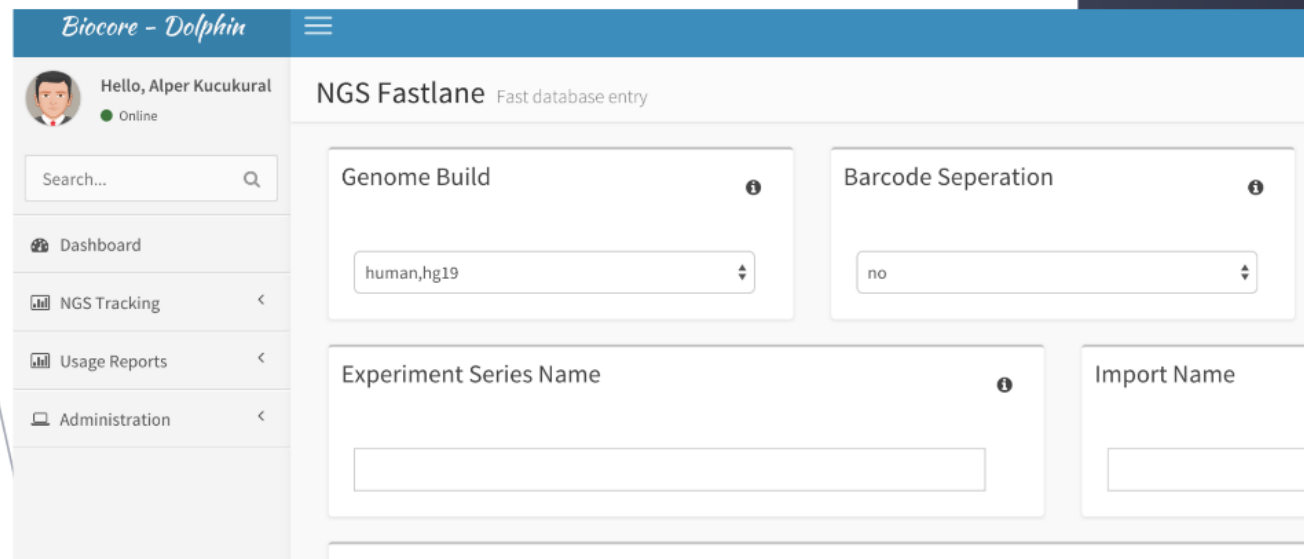
1. Excel Import



The screenshot shows an Excel spreadsheet with a table containing metadata for a series. The table has two columns, A and B. Column A contains labels for various fields, and Column B contains their corresponding values. The first row is a header with the word 'SERIES' in red. The second row is a comment: '# This section describes the overall experiment.' The following rows are labeled 'title', 'summary', 'overall design', 'contributor', 'fastq directory', 'backup directory', and 'amazon bucket'. The values for these fields are: 'New series', 'example summary', 'example overall design', an empty field, '/export/genome_data/m', '/export/excelrun6/proces', and an empty field. The spreadsheet is viewed from the 'METADATA' tab, with other tabs like 'LANES', 'PROTOCOLS', 'SAMPLES', and 'FILES' visible at the bottom.

	A	B
1	SERIES	
2	# This section describes the overall experiment.	
3	title	New series
4	summary	example summary
5	overall design	example overall design
6	contributor	
7	contributor	
8	fastq directory	/export/genome_data/m
9	backup directory	/export/excelrun6/proces
10	amazon bucket	
11		
12		
13		
14		

2. Fast Lane



The screenshot shows the 'Biocore - Dolphin' NGS Fastlane interface. The top bar is blue with the 'Biocore - Dolphin' logo and a hamburger menu. Below the bar, there's a user profile section with a profile picture, the name 'Hello, Alper Kucukural', and a status indicator 'Online'. A search bar is located below the profile. The main content area is titled 'NGS Fastlane Fast database entry'. It contains several input fields: 'Genome Build' with a dropdown menu showing 'human,hg19', 'Barcode Separation' with a dropdown menu showing 'no', 'Experiment Series Name' with a text input field, and 'Import Name' with a text input field. The left sidebar contains a navigation menu with items: 'Dashboard', 'NGS Tracking', 'Usage Reports', and 'Administration'.

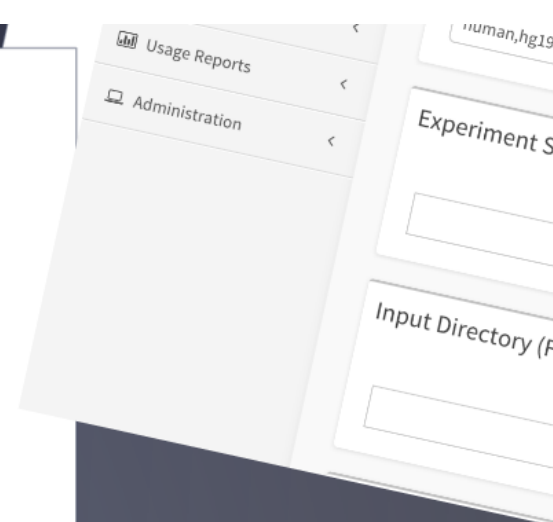
Running Pipelines

Common steps:

- Barcode separation (novobarcoder)
- Adapter removal (small RNA-Seq)
- Fastq quality checks (FastQC)
- Quality filtering (trimomatic)
- 5' 3' Trimming (trimomatic)
- Mapping custom sequence sets and quantification (Bowtie2)
- Mapping common RNAs and quantification (rRNA, miRNA, tRNA, snRNA, piRNA, rsmk)
- Fastq splitter

Pipelines:

- RNA-Seq quantification (RSEM)
- Differential expression analysis (DESeq2: support for RSEM, miRNA, tRNA)
- Tophat Pipeline (tophat2 and picard tools)
- Chip-Seq pipeline (Peak Calling and aggregation plots)
- Bisulphite Mapping for RRBS and WGBS



Job Monitoring

- Status screen

Current Run Status ?

Show entries Search:

ID	Name	Output Directory	Description	Status	Options
98	sample_R848_Library	/s4s/s4s_manuel_garber/biocre/ngsruns/DC/new_tests/sample_R848_Library/RNA-Seq	sample_R848_Library RNA-Seq quantification	⚠ Error	Options
97	RNA_seq_ERCC1	/s4s/s4s_manuel_garber/biocre/ngsruns/DC/human/August_14_2015_ERCC/RNASeq	RNA_seq_ERCC1 RNA-Seq quantification	⚠ Error	Options
94	Small RNASeq pipeline	/project/umw_biocre/kucukura/data/kahraman/kahraman_Sep3/dolphin1	Run Small RNASeq pipeline	⚠ Error	Options
92	Fastlane Initial Run	/home/nm87w/back/initial_run	Fastlane Initial Run within import: nick_test	⚠ Error	Options
91	Small RNASeq pipeline	/project/umw_biocre/kucukura/data/kahraman/kahraman_Sep3/dolphin	Run Small RNASeq pipeline	✓ Complete!	Options
90	RNA_seq_ERCC1			✓ Complete!	Options

Biocre - Dolphin



Hello, Alper Kucuk

● Online

Search...

Dashboard

NGS Tracking

Usage Reports

Administration

Error Log:

- Error tracking

run.94.wrapper.std:

undefinedanalyseDE

stepDESeq2p1:ERROR:EXIT 4551123

stepDESeq2p1 Failed

Check LSF output

Currently Loaded Modulefiles: 1) openssl/1.0.1g 2) python/2.7.5

/project/umw_biocre/kucukura/data/kahraman/kahraman_Sep3/dolphin1

dbSetStartTime

BS_01,BS_02,BS_03,BS_04,LT_013,LT_014,LT_015,LT_016

library("DESeq2")

library("ggplot2")

library("gplots")

analyseDE

Check the command:

@RUNDESEQ2 -c @COLS1 -pu @PUBDIR -da @DATASET1 -w @WKEY -dc @CONDS1 -r

Integration

Share your data using Dolphin WebAPI (JSON)
Compare your experiment results with
published data
Your previous results
Support meta data
Support picard metrics and QC results
Support code submission
Support CBI submission
Support GEO import functionality



Where is dolphin?

<http://dolphin.umassmed.edu>

Documentation

<http://dolphin.readthedocs.org>

Analysis

Reports

- Initial Mapping Results
- FastQC reports
- RNA Quantification tables (for gene and isoforms; raw, and normalized counts)
- DESeq results
- Tophat alignments (bam format, picard metrics)

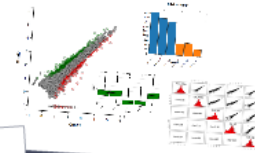


DEBrowser

INTERACTIVE DIFFERENTIAL EXPRESSION ANALYSIS TOOL

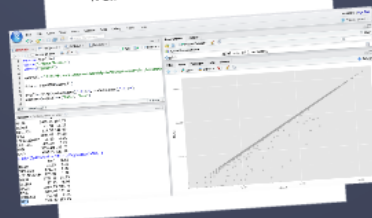
Discover what the data is

- make **complex comparisons** by building custom tables that include results of any supported DE algorithms on any combination of replicates of interest.
- filter **batch effects** and **identify outliers** by All-to-All, IQR, density, and PCA plots
- graphically select from any of the genome wide visualizations of any subset of genes for further in-depth exploration such as **ontology, pathway and disease analysis**.



R-Studio

- Use the output files (tsv format) directly from R-Studio for downstream analysis.
- Example codes for DE-Seq2, hierarchical clustering and heatmaps can be downloadable from reports section.
- Use Dolphin WebAPI to read the data from R-Studio.



Processing with Dolphin



Reports

- Initial Mapping Results
- FastQC reports
- RNA Quantification tables (for gene and isoforms; raw, and normalized counts)
- DESeq results
- Tophat alignments (bam format, picard metrics)

Initial Mapping Results

Show 10 entries

Libname	Total Reads	ercc	rRNA	miRNA	tRNA	snRNA	rmsk	Reads Left
D12_MDDC_Ctrl_0h_t1	12,641,738	0 (0.00 %)	313,493 (2.48 %)	10 (0.00 %)	42 (0.00 %)	580 (0.00 %)	161,544 (1.28 %)	12,166,069 (96.24 %)
D12_MDDC_Lps_12h	19,820,565	2 (0.00 %)	509,640 (2.57 %)	25 (0.00 %)	126 (0.00 %)	1,165 (0.01 %)	246,603 (1.24 %)	19,063,004 (96.18 %)
D12_MDDC_Lps_1h	19,365,298	0 (0.00 %)	539,348 (2.79 %)	60 (0.00 %)	189 (0.00 %)	1,727 (0.01 %)	208,555 (1.08 %)	18,615,419 (96.13 %)
D12_MDDC_Lps_4h	16,456,109	0 (0.00 %)	377,348 (2.29 %)	34 (0.00 %)	120 (0.00 %)	1,071 (0.01 %)	196,941 (1.20 %)	15,880,595 (96.50 %)
D12_MDDC_LpsJaki_12h	16,794,717	0 (0.00 %)	533,782 (3.18 %)	34 (0.00 %)	96 (0.00 %)	919 (0.01 %)	236,679 (1.41 %)	16,023,207 (95.41 %)
D12_MDDC_LpsJaki_1h	16,504,951	0 (0.00 %)	304,800 (1.85 %)	37 (0.00 %)	143 (0.00 %)	1,416 (0.01 %)	203,537 (1.23 %)	15,995,018 (96.91 %)
D12_MDDC_LpsJaki_4h	12,469,253	0 (0.00 %)	272,933 (2.19 %)	35 (0.00 %)	119 (0.00 %)	792 (0.01 %)	155,016 (1.24 %)	12,040,358 (96.56 %)
D12_MDDC_LpsTnfab_1h	16,217,610	0 (0.00 %)	617,120 (3.81 %)	62 (0.00 %)	238 (0.00 %)	2,608 (0.02 %)	268,826 (1.66 %)	15,328,756 (94.52 %)
D12_MDDC_LpsTnfab_4h	16,851,519	0 (0.00 %)	436,401 (2.59 %)	30 (0.00 %)	344 (0.00 %)	1,418 (0.01 %)	238,553 (1.42 %)	16,174,773 (95.98 %)
D13_MDDC_Lps_4h	16,154,426	0 (0.00 %)	832,238 (5.15 %)	39 (0.00 %)	189 (0.00 %)	1,349 (0.01 %)	224,406 (1.39 %)	15,096,205 (93.45 %)

Showing 1 to 10 of 10 entries

--- Select a Result ---

FastQC Summary

Detailed FastQC Results

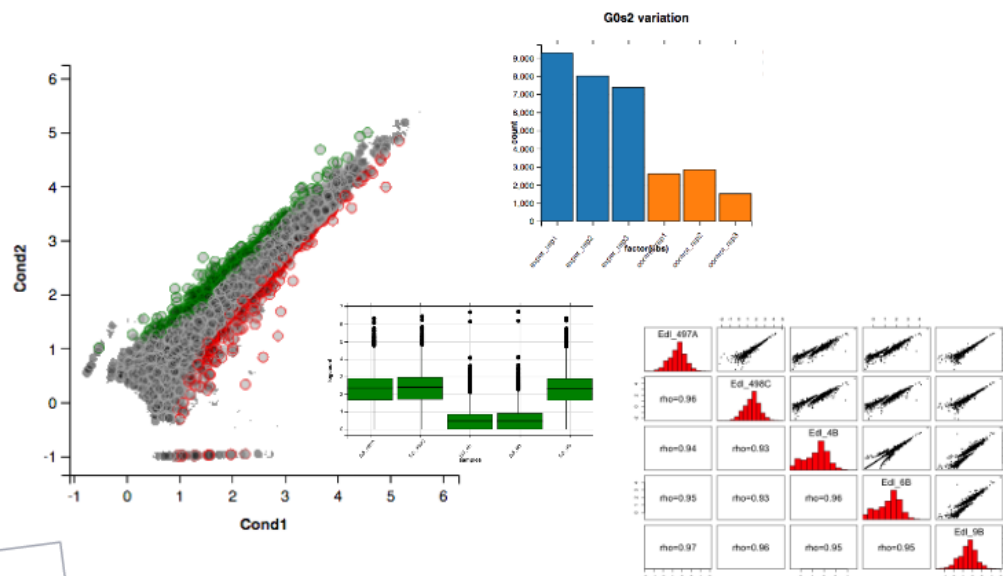
RSEM Results

DEBrowser

INTERACTIVE DIFFERENTIAL EXPRESSION ANALYSIS TOOL

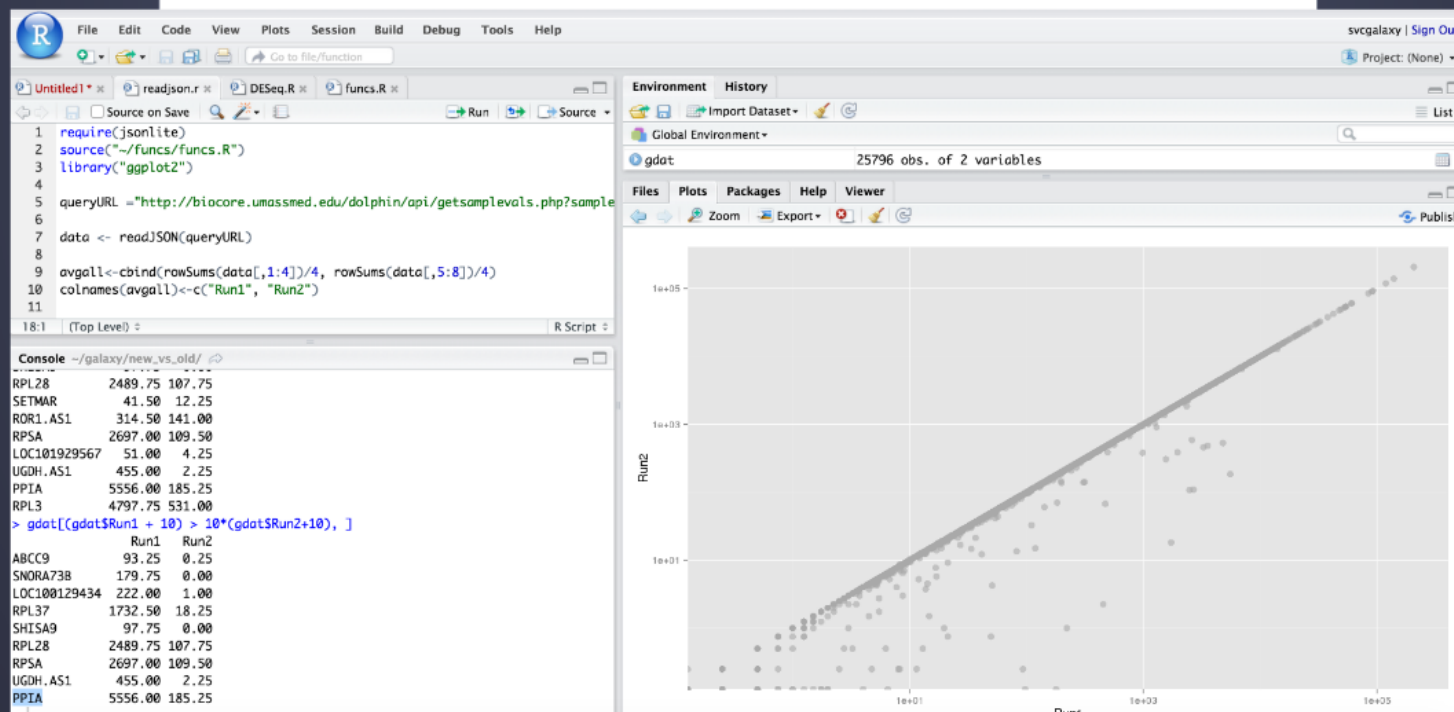
DEBrowser allow the user to:

- make **complex comparisons** by building custom tables that include results of any supported DE algorithms on any combination of replicates of interest.
- detect **batch effects** and **identify outliers** by All-to-All, IQR, density, and PCA plots
- graphically select from any of the genome wide visualizations of any subset of genes for further in-depth exploration such as **ontology, pathway and disease analysis**.



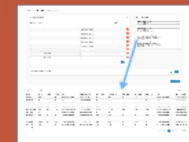
R-Studio

- Use the output files (tsv format) directly from R-Studio for downstream analysis.
- Example codes for DE-Seq2, hierarchical clustering and heatmaps can be downloadable from reports section.
- Use Dolphin WebAPI to read the data from R-Studio.



Integration

- Share your data using Dolphin WebAPI (JSON)
- Compare your experiment results with
 - published data
 - your previous results
- Export meta data
- Export picard metrics and QC results
- Encode submission
- NCBI submission
- GEO import functionality



Where is dolphin?

<http://dolphin.umassmed.edu>

Documentation

NGS Table Creator Custom table creation

Samples Selected

Show 10 entries

Search:

id	Sample Name	Run ID	Delete
31	c_rep3	Run 12: Run 1 RNA-Seq	
33	e_rep2	Run 12: Run 1 RNA-Seq	
34	e_rep3	Run 12: Run 1 RNA-Seq	
35	control_rep1	Run 13: Run 2 RNA-Seq	
36	control_rep2	Run 13: Run 2 RNA-Seq	
37	control_rep3	Run 13: Run 2 RNA-Seq	
38	exper_rep1	Run 13: Run 2 RNA-Seq	
39	exper_rep2	Run 13: Run 2 RNA-Seq	

Showing 1 to 8 of 8 entries

Previous 1 Next

Additional Sample Selection

Generate Table

Report Selection

counts/rRNA.summary.tsv
counts/rRNA.counts.tsv
counts/miRNA.summary.tsv
counts/miRNA.counts.tsv
counts/tRNA.summary.tsv
counts/tRNA.counts.tsv
rsem/genes_expression_tpm.tsv
rsem/genes_expression_expected_count.tsv
rsem/isoforms_expression_tpm.tsv
rsem/isoforms_expression_expected_count.tsv
DESeq2RSEM1/alldetected_genes.tsv
DESeq2RSEM1/selected_log2fc_genes.tsv
DESeq2RSEM1/rscript_genes.R
DESeq2RSEM1/alldetected_isoforms.tsv
DESeq2RSEM1/selected_log2fc_isoforms.tsv
DESeq2RSEM1/rscript_isoforms.R

10 entries per page

Search here...

name	c_rep3	e_rep2	e_rep3	padj	log2FoldChange	control_rep1	control_rep2	control_rep3	exper_rep1	exper_rep2	padj	log2FoldChange
NR_073370	3.04	1	3.04	0.667708137672413	-0.440817461903343	13	6.09	3.04	6.12	1	0.667708137672413	-0.440817461903343
NM_175347	7	7	103	NA	1.7087359505823	17	7	7	5	7	NA	1.7087359505823
NM_029090	1460.53	1028.46	731	0.507280029803894	-0.208258491512925	1797.49	1182.64	1460.53	1197.09	1028.46	0.507280029803894	-0.208258491512925
NR_038166	2444.9	5253.74	2014.74	7.11604506700198e-22	0.804341741957158	3100.19	2210.2	2444.9	4061.55	5253.74	7.11604506700198e-22	0.804341741957158
NM_001252437	181.8	104.8	27.84	4.57778388161203e-08	-1.38940663698532	248.24	141.59	181.8	49.07	104.8	4.57778388161203e-08	-1.38940663698532
NM_008051	0	12	0	0.667708137672413	0.461343515560868	10	0	0	12	12	0.667708137672413	0.461343515560868
NM_145359	248	308	121	0.667708137672413	0.0676492354361472	302	202	248	221	308	0.667708137672413	0.0676492354361472
NM_001033159	115	101	56	0.0288207286554533	-0.531163166815204	173	131	115	75	101	0.0288207286554533	-0.531163166815204
NM_009624	436.43	69.78	69.68	4.46237968308457e-06	-2.08829621664454	970	583	436.43	174.23	69.78	4.46237968308457e-06	-2.08829621664454
NM_025571	199	577	185	0.000253417103214425	0.782439765519644	301	244	199	340	577	0.000253417103214425	0.782439765519644

1 to 10 of 45 entries

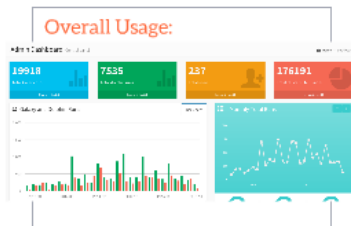
← Previous 1 2 3 4 5 Next →

Where is dolphin?

<http://dolphin.umassmed.edu>

Documentation

<http://dolphin.readthedocs.org>



THANK YOU!!!

Homework:

- Import RNA-Seq samples from GEO.

<https://www.ncbi.nlm.nih.gov/sites/GDSbrowser/>

- Run RNA-Seq pipeline with Quality Checks and filter out rRNA's.
- Make sure you have enough space to process this data (at least 5x space of the total size of the initial files)