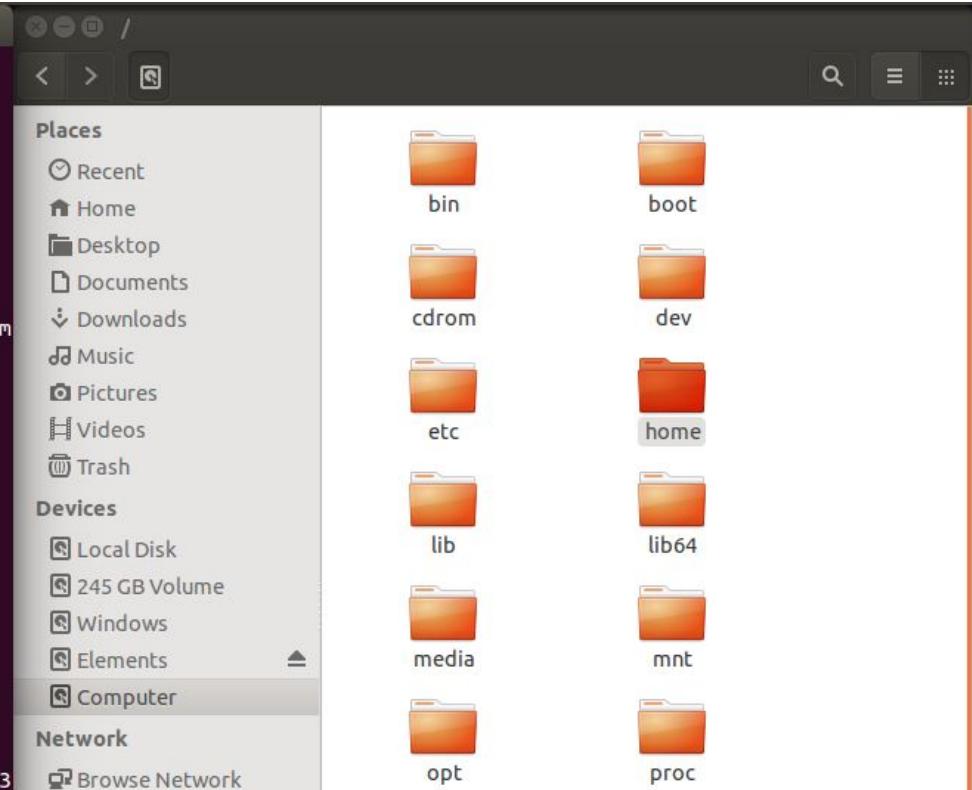


Wprowadzenie do środowiska Galaxy

Warszawa 11.04.2017
intelliseq OpenExome

CLI vs GUI

```
navanshu@pavilion:~$ ls -a -l /
total 104
drwxr-xr-x 23 root root 4096 Aug  9  2015 .
drwxr-xr-x 23 root root 4096 Aug  9  2015 ..
drwxr-xr-x  2 root root 4096 Jun  2 17:58 bin
drwxr-xr-x  4 root root 4096 Jun  2 18:11 boot
drwxr-xr-x  2 root root 4096 Aug  9  2015 cdrom
drwxr-xr-x 16 root root 4320 Jun  4 20:47 dev
drwxr-xr-x 144 root root 12288 Jun  4 20:47 etc
drwxr-xr-x  3 root root 4096 Aug  9  2015 home
lrwxrwxrwx  1 root root    33 Aug  9  2015 initrd.img -> boot/initrd.im
-24-generic
drwxr-xr-x 23 root root 4096 Jun  2 17:51 lib
drwxr-xr-x  2 root root 4096 Jun  2 17:43 lib64
drwx----- 2 root root 16384 Aug  9  2015 lost+found
drwxr-xr-x  3 root root 4096 Sep  3  2015 media
drwxr-xr-x  2 root root 4096 Apr 11  2014 mnt
drwxr-xr-x  4 root root 4096 Jun  3 14:26 opt
dr-xr-xr-x 235 root root   0 Jun  5  2016 proc
drwx-----  4 root root 4096 May 27 19:47 root
drwxr-xr-x 25 root root  800 Jun  4 21:14 run
drwxr-xr-x  2 root root 12288 Jun  2 17:57 sbin
drwxr-xr-x  2 root root 4096 Apr 17  2014 srv
dr-xr-xr-x 13 root root   0 Jun  5  2016 sys
drwxrwxrwt  6 root root 4096 Jun  5 00:04 tmp
drwxr-xr-x 10 root root 4096 Apr 17  2014 usr
drwxr-xr-x 14 root root 4096 May 30 22:22 var
lrwxrwxrwx  1 root root    30 Aug  9  2015 vmlinuz -> boot/vmlinuz-3.13
```



CLI - interfejs wiersza poleceń

```
bowtie2 -x hg19 -U test.fq -S test.sam
```

GUI - graficzny interfejs użytkownika

Bowtie2 - map reads against reference genome (Galaxy Version 2.2.6.2)

Versions Options

Is this single or paired library

Single-end

FASTQ file

No fastqsanger dataset available.
Must be of datatype "fastqsanger"

Write unaligned reads (in fastq format) to separate file(s)

Yes No
--un/--un-conc; This triggers --un parameter for single reads and --un-conc for paired reads

Write aligned reads (in fastq format) to separate file(s)

Yes No
--al/--al-conc; This triggers --al parameter for single reads and --al-conc for paired reads

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

Use a built-in genome index
Built-ins were indexed using default options. See 'Indexes' section of help below

Select reference genome

Human (Homo sapiens) (b37): hg19
If your genome of interest is not listed, contact the Galaxy team

What is Galaxy?

~~Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research~~

Galaxy is a tool to convert all those ugly linux command line tools into web accessible buttons

```
marpiech@edison:~/html/smm
"; exon_id "ENSMUST00000129094_1"; gene_name "ENSMUSG00000024006";
chr17  stdin  exon  28991298          28991418          .
gene_id "ENSMUSG00000024006"; transcript_id "ENSMUST00000129094"; exon_number "2"
"; exon_id "ENSMUST00000129094_2"; gene_name "ENSMUSG00000024006";
chr17  stdin  exon  28989977          28990148          .
gene_id "ENSMUSG00000084632"; transcript_id "ENSMUST00000122683"; exon_number "1"
"; exon_id "ENSMUST00000122683_1"; gene_name "ENSMUSG00000084632";
chr17  stdin  exon  28991405          28991420          .
gene_id "ENSMUSG00000024006"; transcript_id "ENSMUST00000142366"; exon_number "1"
"; exon_id "ENSMUST00000142366_1"; gene_name "ENSMUSG00000024006";
chr17  stdin  exon  28992417          28992468          .
gene_id "ENSMUSG00000024006"; transcript_id "ENSMUST00000142366"; exon_number "2"
"; exon_id "ENSMUST00000142366_2"; gene_name "ENSMUSG00000024006";
chr17  stdin  exon  28998792          28998894          .
gene_id "ENSMUSG00000024006"; transcript_id "ENSMUST00000142366"; exon_number "3"
"; exon_id "ENSMUST00000142366_3"; gene_name "ENSMUSG00000024006";
chr17  stdin  exon  28997564          29000144          .
gene_id "ENSMUSG00000024006"; transcript_id "ENSMUST00000145298"; exon_number "1"
"; exon_id "ENSMUST00000145298_1"; gene_name "ENSMUSG00000024006";
marpiech@edison:~/html/smm$ cat ensGene.gtf | grep chr17 | grep -P "\t2[7-8][0-9]*[0-9]*[0-9]*\t" > short.gtf
marpiech@edison:~/html/smm$
```



The Galaxy interface is a web-based platform for data intensive biomedical research. It features a sidebar with various tools and a main panel displaying a 'Try G on the Web' banner and a list of recent workflows.

Screenshots:

- Terminal View:** Shows a command-line session on a Linux system (edison) running a script to filter a GTF file (ensGene.gtf) to extract entries for chromosome 17 and save it as short.gtf.
- Galaxy Interface:** Shows the Galaxy web application. The sidebar contains links to various tools: Get Data, Send Data, LiftOver, Text Miner, Data Store, Compose Formulas, Filter and Sort, Join, Subtract and Group, NGS: QC and manipulation, NGS: Mapping, NGS: RNA-seq, NGS: SMMtools, NGS: BAM Tools, NGS: BED Tools, NGS: VCF Manipulation, Extract Features, Fetch Sequences, Fetch Alignments, Get Genome Scores, Operate on Genomic Intervals, Statistics, Output Display Data, Phenotype Association, seqkit, BEDTools, Genome Diversity.
- Workflow View:** Shows a list of recent workflows, each with a thumbnail, name, and status (e.g., 'Cuffdiff on data 52_d' is completed).

Logos at the bottom:

- PENNSTATE
- JOHNS HOPKINS
- TACC
- iPlant Collaborative

Czym jest Galaxy?

Galaxy to platforma pozwalająca na tworzenie graficznego interfejsu użytkownika (GUI) dla narzędzi CLI

Inne właściwości: tworzenie scenariuszy przetwarzania danych, dzielenie się danymi, powtarzalność analiz

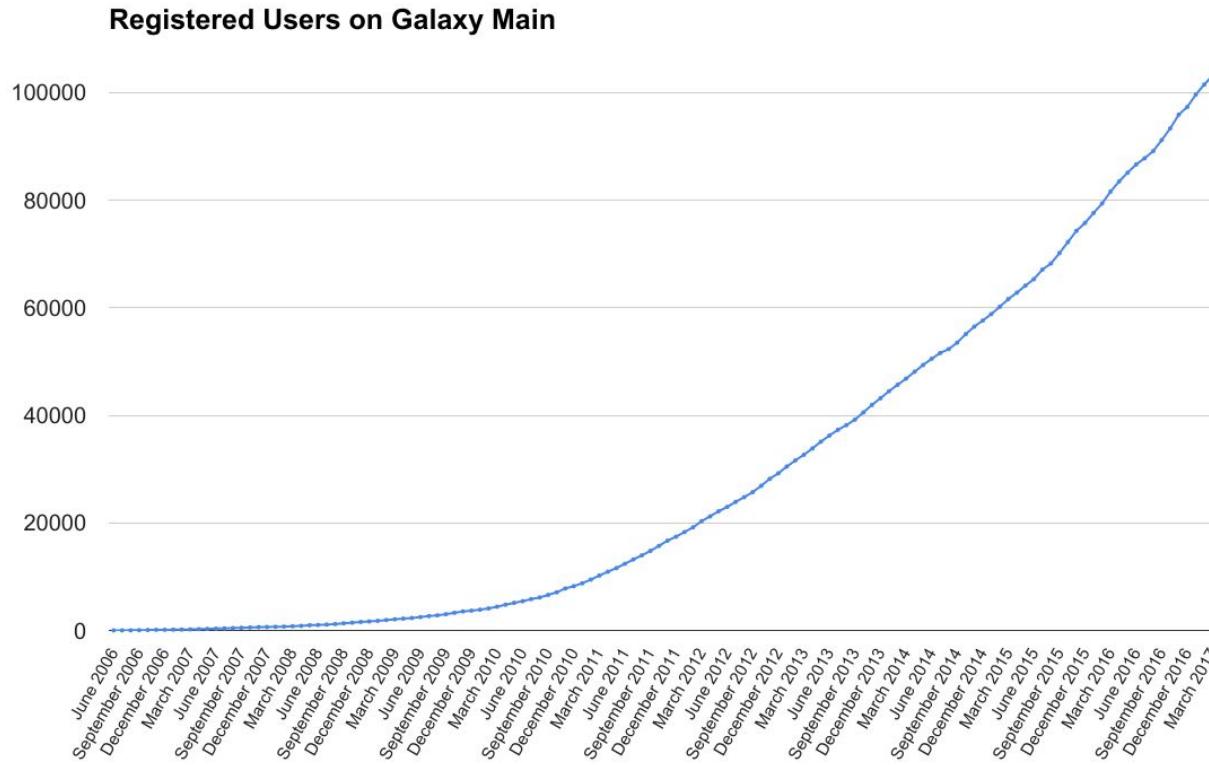
Początkowo rozwijane przez Emory University oraz Penn State University.
Obecnie przez ponad 10 grup i pojedynczych ludzi.

Jak uruchomić Galaxy?

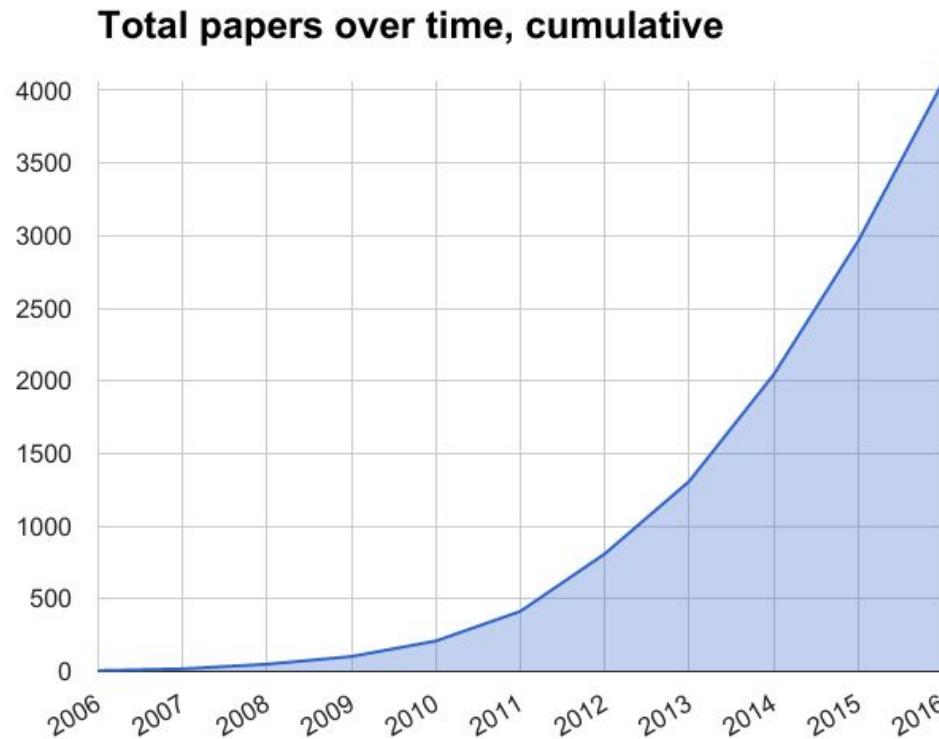
Wpisać adres w przeglądarce

np. usegalaxy.org

Ille osób używa galaxy



Ile osób używa Galaxy?



Publiczne instancje Galaxy

Date	# Servers
2011/07	15
2012/01	21
2012/07	20
2013/01	25
2013/02	27
2013/07	35
2014/01	54
2014/07	60
2015/01	70
2015/07	73
2016/01	84
2016/08	85
2017/01	90
2017/04	95

Galaxy Tool Shed - repozytorium narzędzi

Tools

Galaxy Project Tool Shed

The Galaxy Project [Tool Shed](#) contains contributed software and workflows that can be installed at any Galaxy site.

Date	Repos	Unique Owners	Valid Tools
2014/07	1,140	254	2,538
2014/10	1,236	273	2,684
2015/01	1,355	293	2,914
2015/07	1,726	335	3,374
2016/02	2,093	380	3,676
2016/08	2,582	438	4,167
2017/01	2,689	466	4,300

Interfejs Galaxy

The screenshot shows the Galaxy web interface with several red diagonal annotations:

- A large red diamond-shaped annotation covers the left sidebar, containing the text "Analysis modules".
- A smaller red diamond-shaped annotation covers the central workspace area, containing the text "Data browser".
- A red diamond-shaped annotation covers the right sidebar, containing the text "History".

Tools Panel (Left):

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract, group
- Extract Features
- Fetch Sequences
- File Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- Human Genome Variation
- EMBOSS

Central Workspace:

This dataset is large and only the first megabyte is shown below.
Show all | Save

```
#1:2:4:885:Y  
GTTTTCATCTTAAATTGTACCGCAAAANNNCCACC  
+1:2:4:885:Y  
B#BB7#B#B#%;AA7#AB7B<ABAAA#?A#####  
#1:2:4:1211:Y  
GATAAAAGATTCTCATGAGTTTCCAAGATCATTNNITCAA  
+1:2:4:1211:Y  
B>B#AACAA#<#;ABC7AA#7BBB#7BBB-##B#-#B#  
#1:2:4:1280:Y  
ACTTGTGTAAATNATCTTTGTCCTCCAGGCT#NNTCC  
+1:2:4:1280:Y  
BA>BB#B7AA#%JABBBB->;AA#B#B#B#B#B#B#  
#1:2:4:341:Y  
ACATTTGGTTTNCACCAACCTGAAACTGTAAGGCGCT  
+1:2:4:341:Y  
#A7#BBC<CB#-#4#BBCA#<#BB#B#B#B#B#B#>#B#A  
#1:2:4:396:Y  
GCTTTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT  
+1:2:4:396:Y  
BBCB#BBCB#BBCB#BBCB#BBCB#BBCB#BBCB#BBCB#<%>B#<  
#1:2:4:1118:Y  
GGCTGATATTTCAGGAATCAGCTCTCCAGAAGNTGTT  
+1:2:4:1118:Y  
B#A#B#A1>A#%;?>B#SAB#;1ABABBBB#A%;<(B?  
#1:2:4:1172:Y  
AAATCCGATTCNTCGGATTCTTCCCCAACTANNGCAGA  
+1:2:4:1172:Y  
BCACCB#CCB#%;B>BC3ACC#CABC#9ABBB#<#####  
#1:2:4:1229:Y  
TGAAACAGTTGNCATCCCTGCCCTGATTCTNACTGT  
+1:2:4:1229:Y  
BCACCCBC#C#>CC#6#CCBBCACA#8#CBC7#%>BCB#  
#1:2:4:1189:Y  
CATTTCTTANTCTCTTCTCAATAAACTCTNNTGGTC  
+1:2:4:1189:Y  
9ACCCCCCCC#>BCCCCCCC#B#CCCC#%>CCCC  
#1:2:4:862:Y  
ACCTGAGCCAGNATAATGCTTCACTGTTCCNNGCCG  
+1:2:4:862:Y  
BBB#B#B#B#B#A#A#B#A#B#B#B#B#B#B#B#B#B#B#B#  
#1:2:4:473:Y  
GATTTCCTTATTGATTCACTCCGAAAGTTGAAANNNCTCT
```

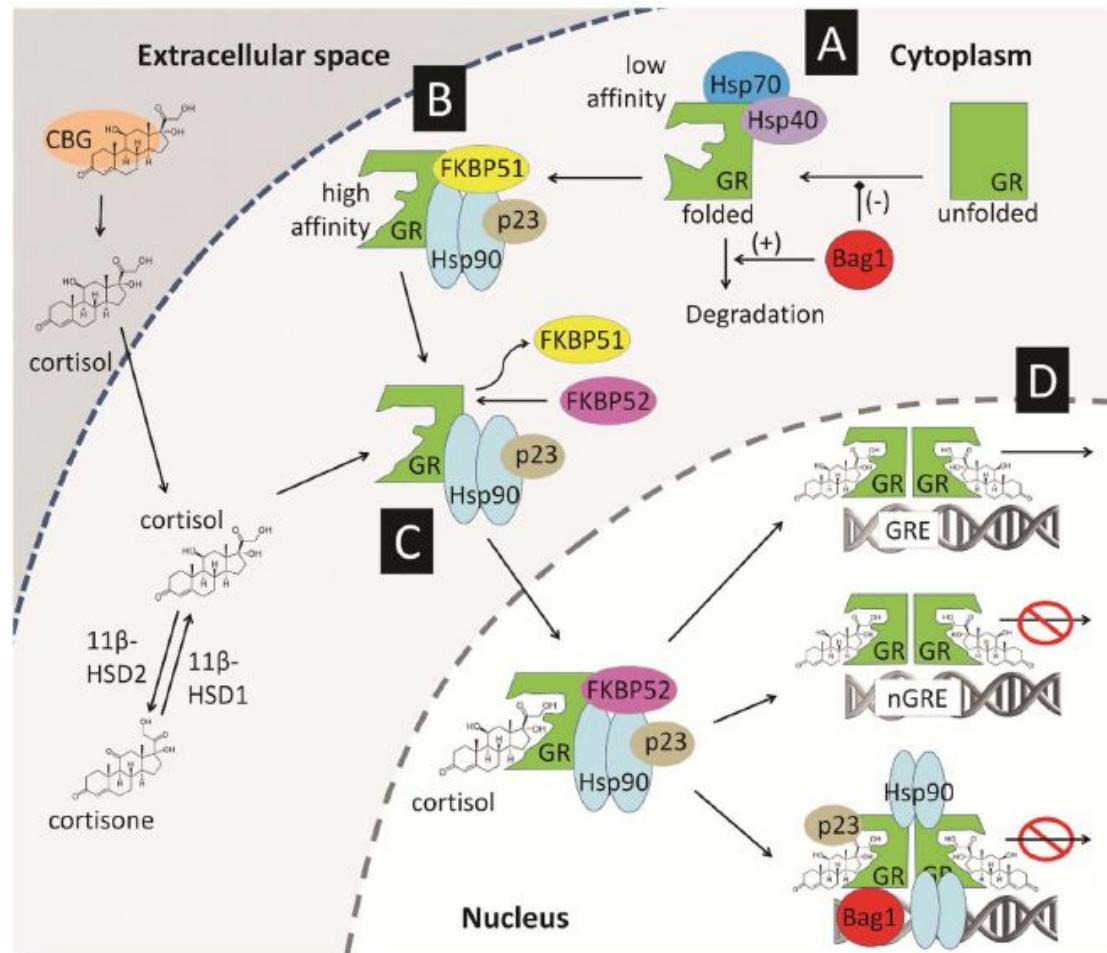
History Panel (Right):

- NA18603.chrom1.ILLUMINA
- Map with Bowtie
- FASTQ Masker
- FASTQ Groomer
- FASTQ Groomer
- FASTQ Groomer
- http://biocluster.ucr.edu/~tbackman/query.fasta
- fpt://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data/NA18969/sequence_read/SRR004506.filt.fastq

Analiza danych RNA-seq

Dane

Astrocyty
wyizolowane z
prążkowia myszy 4
godziny po iniekcji
deksametazonu



Experiment



Animal
Model
C57BL/6J
inbred mice

Dexamethasone
treatment
4 mg/kg
2h



Brain tissue
dissection
cell separation
total RNA
extraction

Library
preparation



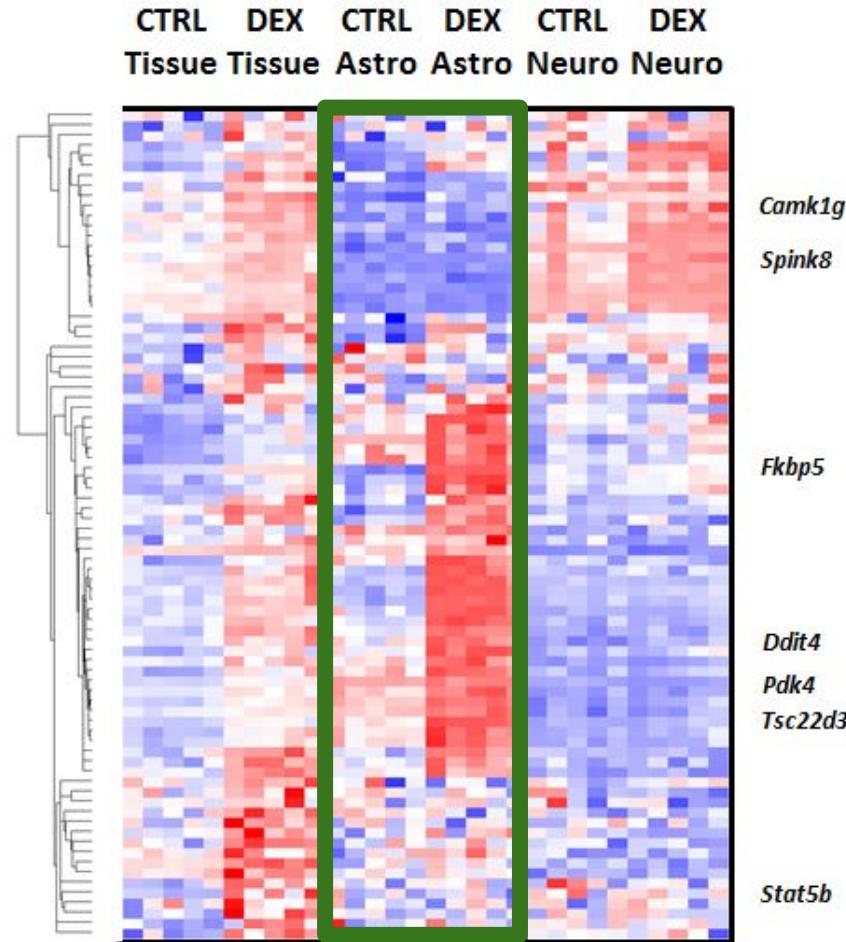
NGS sequencing
Ion Proton



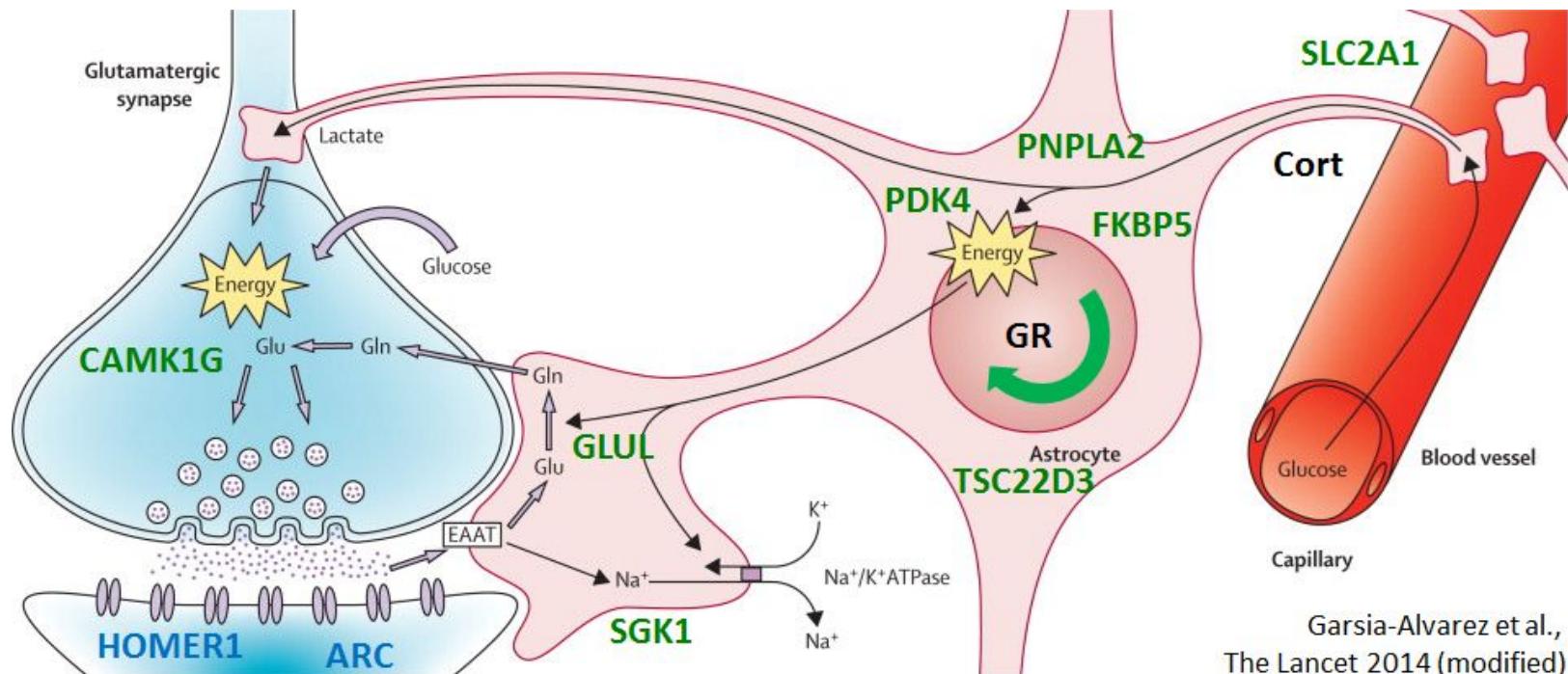
Data
analysis

Results

Brain (striatum)



Summary



Garsia-Alvarez et al.,
The Lancet 2014 (modified)

Utworzenie nowego użytkownika

Screenshot of the Galaxy Docker Build interface showing the registration process.

The top navigation bar includes: Analyze Data, Workflow, Shared Data, Visualization, Help, User (with a dropdown menu), and a grid icon. The User dropdown menu has two options: Login and Register, with Register highlighted and circled in red.

The main content area displays a message: "Hello, your Galaxy Docker container is running!" followed by instructions to customize the page by creating a welcome.html file. It also features a cartoon illustration of a whale carrying shipping containers, labeled "Docker In Galaxy".

The left sidebar lists various tools and resources:

- Tools: search tools, Get Data, Collection Operations, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features, Fetch Sequences, Fetch Alignments, Statistics, Graph/Display Data, NGS: RNA Analysis, NGS: QC and manipulation, BAM-to-SAM convert BAM to SAM, Plot Data.
- Workflows: All workflows.

The right sidebar shows a History section with an empty history named "Unnamed history". A note indicates that the history is empty and suggests loading data from an external source.

Page footer: 178.216.200.146:8080/user/create

Utworzenie nowego użytkownika

The screenshot shows the Galaxy Docker Build interface. On the left, there's a sidebar titled 'Tools' containing various bioinformatics tools like 'Get Data', 'Text Manipulation', 'Statistics', and 'NGS: QC and manipulation'. The main area is titled 'Create User' and contains fields for 'Email address' (new-user@example.com), 'Password' (a masked password), 'Confirm password' (also a masked password), and 'Public name' (new-user). A note below the public name field specifies that it must be at least three characters long and contain only lowercase letters, numbers, dots, underscores, and dashes. A red circle highlights the 'Email address' and 'Password' fields. At the bottom of the 'Create User' form is a 'Submit' button. To the right of the main form is a 'History' panel which is currently empty, displaying a message encouraging users to load their own data or get data from an external source.

Zmiana nazwy historii

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Help User

Using 0 bytes

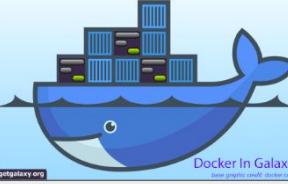
Tools

- search tools
- [Get Data](#)
- [Send Data](#)
- [Collection Operations](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [NGS: RNA Analysis](#)
- [NGS: QC and manipulation](#)
- [BAM-to-SAM convert BAM to SAM](#)
- [Plot Data](#)
- [Workflows](#)
 - All workflows

Hello, your Galaxy Docker container is running!

To customize this page you can create a welcome.html page in your directory mounted to /export.

Configuring Galaxy » Installing Tools » Guided Tour »



Galaxy is an open platform for supporting data intensive research. Galaxy is developed by [The Galaxy Team](#) with the support of many contributors. The Galaxy Docker project is supported by the University of Freiburg, part of de.NBI.

The Galaxy Project is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience at Penn State](#), and [Johns Hopkins University](#).

History

search datasets

rna-seq-tutorial

(empty) Click to rename history

This history is empty. You can load your own data or get data from an external source

A red circle highlights the "Click to rename history" button in the History panel.

Pobranie danych

Screenshot of the Galaxy web interface showing the "Get Data" tool being used to download files from the web or upload from disk.

The main interface shows a list of datasets on the right, including:

- RNA-seq-tutorial (36 shown, 10 deleted)
- 1: A1_DEX_trimm ed.bam
- 15: A1_CTRL_trimm ed.bam
- 14: A5_DEX_trimm ed.fq
- 13: A5_CTRL_trimm ed.fq
- 12: A4_DEX_trimm ed.fq
- 11: A4_CTRL_trimm ed.fq
- 10: A3_DEX_trimm ed.fq
- 9: A3_CTRL_trimm ed.fq
- 8: A2_DEX_trimm ed.fq
- 7: A2_CTRL_trimm ed.fq
- 6: A1_DEX_trimm ed.fq
- 5: 23847391_3339 6705.qtf
- 4: mm10-ch17.fasta
- 3: FastQC on data 1: RawData
- 2: FastQC on data 1: Webpage
- 1: A1_CTRL_trimm ed.fq

The "Get Data" tool dialog is open in the center, with two red circles highlighting specific areas:

- A red circle surrounds the URL input field, which contains the URL: `http://www.sanger.ac.uk/Software/galaxy/distro/test-data/FASTQ/1/A1_CTRL_trimm/ed.fq`.
- A red circle surrounds the "Paste/Fetch data" button, which is highlighted in blue.

The dialog also includes fields for Name, Size, Type, Genome, Settings, and Status, and buttons for Choose local file, Choose FTP file, Paste/Fetch data, Pause, Reset, Start, and Close.

Dane do pobrania

http://178.216.203.113/training-data/rna-seq/fastq/A1_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A2_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A3_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A4_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A5_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A1_DEX_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A2_DEX_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A3_DEX_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A4_DEX_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/fastq/A5_DEX_trimmed.fq.gz

Dane do pobrania

http://178.216.203.113/training-data/rna-seq/fastq/A1_CTRL_trimmed.fq.gz

http://178.216.203.113/training-data/rna-seq/resources/23847391_33396705.gtf

Format fastq



The screenshot shows a Mozilla Firefox browser window displaying a FASTQ sequence. The sequence consists of four lines per read, starting with an '@' symbol. The first two lines represent the sequence and quality scores, while the third and fourth lines provide sequencing platform-specific information.

```
@read1
AGCTTATCCTCTGCTACCCCCGGGTTAGCGCACTTGATGTATTACAGC
+
BA1@CC7CBCCC9C8;B2@>C?B@B@B3=9?@B1:AB7B?B8B?B6B.7.
@read2
TTGGGCCGGGATCTCCAGAACCATATGGATGTGATCCACACAGCATTCTGC
+
?>?B@)<?@,AA7A@C<C?=@@B;+)?B5*@2=@+=BB,=B6C>AB@B24
@read3
TATGCTCAAGAAGGGGCTGATGAGTTGGTGTTCACGATATCACTGCCTC
+
A3AB:B1:B;9/0BBCB<BB@AA0?BB9:BB<A@BB@7@6@<A@@@<3
```


Sprawdzenie jakości

The screenshot shows the Galaxy web interface with the title "Galaxy / Galaxy Docker Build". The main panel displays the "FastQC Read quality reports (Galaxy Version 0.67)" tool. A red circle highlights the "Short read data from your current history" dropdown menu, which contains the option "1: A1_CTRL_trimmed.fq". Below this, the "Contaminants" dropdown is set to "Nothing selected". The "Submodule and Limits" dropdown also shows "Nothing selected". At the bottom of the tool panel is a "Execute" button. To the right of the tool panel is a "History" sidebar titled "rna-seq-tutorial" showing 36 items. Most items are green, indicating successful execution, such as "24: A5 DEX trimm ed.bam" and "23: A5 CTRL trimm ed.bam". Some items are greyed out, like "1: A1 CTRL trimm ed.bam". The bottom of the history sidebar has navigation arrows.

FastQC Read quality reports (Galaxy Version 0.67)

Short read data from your current history

1: A1_CTRL_trimmed.fq

Contaminants

Nothing selected

Submodule and Limits

Nothing selected

Execute

Execute: FastQC (0.67)

FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis.

The main functions of FastQC are:

- Import of data from BAM, SAM or FastQ/FastQ.gz files (any variant),
- Providing a quick overview to tell you in which areas there may be problems
- Summary graphs and tables to quickly assess your data
- Export of results to an HTML based permanent report
- Offline operation to allow automated generation of reports without running the interactive application

FastQC

This is a Galaxy wrapper. It merely exposes the external package [FastQC](#) which is documented at [FastQC](#). Kindly acknowledge it as well as this tool if you use it. FastQC incorporates the [Picard-tools](#) libraries for sam/bam processing.

The contaminants file parameter was borrowed from the independently developed fastqcwrapper contributed to the Galaxy Community Tool Shed by J. Johnson. Adaption to version 0.11.2 by T. McGowan.

Inputs and outputs

FastQC is the best place to look for documentation - it's very good. A summary follows below for those in a tearing hurry.

This wrapper will accept a Galaxy fastq, fastq.gz, sam or bam as the input read file to check. It will also take an optional file containing a list of contaminants information, in the form of a tab-delimited file with 2 columns, name and sequence. As another option the tool takes a custom limits.txt file that allows setting the warning thresholds for the different modules and also specifies which modules to include in the output.

The tool produces a basic text and a HTML output file that contain all of the results, including the following:

- Basic statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences

Mapowanie do genomu referencyjnego

Galaxy / Galaxy Docker Build Analyze Data Workflow Shared Data Visualization Admin Help User Using 1.1 GB

Tools
Get Data Send Data Collection Operations Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Fetch Alignments Statistics Graph/Display Data NGS: RNA Analysis NGS: QC and manipulation BAM-to-SAM Plot Data Workflows All workflows

HISAT2 A fast and sensitive alignment program (Galaxy Version 2.0.5)

Input data format: FASTQ
Single end or paired reads?
Reads:
14: A5_DEX_trimmed.fq
13: A5_CTRL_trimmed.fq
12: A4_DEX_trimmed.fq
11: A4_CTRL_trimmed.fq
10: A3_DEX_trimmed.fq
This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Write unaligned reads (in fastq format) to separate file(s)
Yes No
(-un)

Write aligned reads (in fastq format) to separate file(s)
Yes No
(-al)

Source for the reference genome to align against
Use a genome from history
Built-in references were created using default options
Select the reference genome:
mm10-ch17.fasta

Primary alignments
Search for at most K distinct, primary alignments for each read. Primary alignments mean alignments whose alignment score is equal or higher than any other alignments. The search terminates when it can't find more distinct valid alignments, or when it finds K, whichever happens first. The alignment score for a paired-end alignment equals the sum of the alignment scores of the individual mates. Each reported read or pair alignment beyond the first has the SAM 'secondary' bit (which equals 256) set in its FLAG field. For reads that have more than K distinct, valid alignments, hisat2 does not guarantee that the K alignments reported are the best possible in terms of alignment score. HISAT2 is not designed with large values for -k in mind, so when aligning reads to long repetitive genomic features (HGFMs) (-k)

Maximum number of seeds that will be extended
1000

HISAT2, like other aligners, uses seed-and-extend approaches. HISAT2 tries many seeds and skips the rest of the seeds. Large values for --max-seeds in large values for --max-seeds in mind, and when aligning reads to long repetitive genomic features (HGFMs) (-k)

Report secondary alignments
Yes No
(-secondary)

Alignment options

History
ma-seq-tutorial
36 shown, 10 deleted
1.1 GB
All None For all selected...
A3_DEX_trimmed.bam
A3_CTRL_trimmed.bam
A2_DEX_trimmed.bam
A2_CTRL_trimmed.bam
A1_DEX_trimmed.bam
A1_CTRL_trimmed.bam
A5_DEX_trimmed.fq
A5_CTRL_trimmed.fq
A4_DEX_trimmed.fq
A4_CTRL_trimmed.fq
A3_DEX_trimmed.fq
A3_CTRL_trimmed.fq
A2_DEX_trimmed.fq
A2_CTRL_trimmed.fq
A1_DEX_trimmed.fq
23847391_33396705.gtf
mm10-ch17.fasta
FastQC on data 1: RawData
FastQC on data 1: Webpage
A1_CTRL_trimmed.fq

Spliced alignment parameters ->
Specify alignment parameters ->
GTF file with known splice sites ->
23847391_33396705.gtf

Format SAM/BAM

Zmiana nazwy pliku

Screenshot of the Galaxy web interface showing the "Galaxy / Galaxy Docker Build" tool. The main area displays the "Edit Attributes" form for a dataset named "A1_DEX_trimmed.bam". The "Name" field is set to "A1_DEX_trimmed.bam". The "Info" section shows "Building DifferenceCoversample" and "Building sPrime". The "Annotation / Notes" section is empty. The "Database/Build" section shows "unspecified (?)". A "Save" button is present, along with an "Auto-detect" link. A note below the "Auto-detect" link states: "This will inspect the dataset and attempt to correct the above column values if they are not accurate." To the right, the "History" panel shows a list of datasets, with the entry "16: A1_DEX trimmed ed.bam" highlighted and circled in red.

Galaxy / Galaxy Docker Build

Attributes Convert Format Datatype Permissions

Get Data

Upload File from your computer

UCSC Main table browser

UCSC Test table browser

UCSC Archaea table browser

EBI SRA ENA SRA

Get Microbial Data

BioMart Ensembl server

CBI Rice Mart rice mart

GrameneMart Central server

modENCODE fly server

Flymine server

Flymine test server

modENCODE modMine server

MouseMine server

Ratmine server

YeastMine server

metabolicMine server

modENCODE worm server

WormBase server

Wormbase test server

ZebrafishMine server

EuPathDB server

HbVar Human Hemoglobin Variants and Thalassemias

GenomeSpace import from file browser

Send Data

Collection Operations

Text Manipulation

Filter and Sort

Join, Subtract and Group

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 1.1 GB

History

search datasets

ma-seq-tutorial

36 shown, 10 deleted

1.1 GB

16: A1_DEX trimmed ed.bam

15: A1 CTRL trimmed ed.bam

14: A5 DEX trimmed ed.fq

13: A5 CTRL trimmed ed.fq

12: A4 DEX trimmed ed.fq

11: A4 CTRL trimmed ed.fq

10: A3 DEX trimmed ed.fq

9: A3 CTRL trimmed d.fq

8: A2 DEX trimmed d.fq

7: A2 CTRL trimmed d.fq

6: A1 DEX trimmed d.fq

5: 23847391_3339 6705.of

4: mm10-ch17.fast

3: FastQC on data 1: RawData

2: FastQC on data 1: Webpage

1: A1 CTRL trimmed d.fq

178.216.200.146:8080/datasets/1e8ab44153008be8/edit

Obliczenie statystyki (Cuffdiff)

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 1.1 GB

Tools

Get Data Send Data Collection Operations Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Fetch Alignments Statistics Graph/Display Data NGS: RNA Analysis NGS: QC and manipulation BAM-to-SAM convert BAM to SAM Plot Data Workflows All workflows

Cuffdiff find significant changes in transcript expression, splicing, and promoter use (Galaxy Version 2.2.1.5)

Transcripts

5: 23847391_33396705.gtf

A transcript GFF3 or GTF file produced by cufflinks, cuffcompare, or other source.

Omit Tabular Datasets

Yes No Discard the tabular output.

Generate SQLite

Yes No Generate a SQLite database for use with cummeRbund.

Input data type

SAM/BAM

CuffNorm supports either CXB (from cuffquant) or SAM/BAM input files. Mixing is not supported. Default: SAM/BAM

Condition

1: Condition

Condition name: CTRL

Replicates:

- 19: A3_CTRL_trimmed.bam
- 18: A2_DEX_trimmed.bam
- 17: A2_CTRL_trimmed.bam
- 16: A1_DEX_trimmed.bam
- 15: A1_CTRL_trimmed.bam

2: Condition

Condition name: DEX

Replicates:

- 20: A3_DEX_trimmed.bam
- 19: A2_CTRL_trimmed.bam
- 18: A2_DEX_trimmed.bam
- 17: A2_CTRL_trimmed.bam
- 16: A1_DEX_trimmed.bam

+ Insert condition

Library normalization method

geometric

Dispersion estimation method

pooled

If using only one sample per condition, you must use 'blind.'

False Discovery Rate

0.05

History

search datasets

rna-seq-tutorial

36 shown, 10 deleted

1.1 GB

All None For all selected...

20: A3_DEX_trimmed.bam
19: A3_CTRL_trimmed.bam
18: A2_DEX_trimmed.bam
17: A2_CTRL_trimmed.bam
16: A1_DEX_trimmed.bam
15: A1_CTRL_trimmed.bam
14: A5_DEX_trimmed.fa
13: A5_CTRL_trimmed.fa
12: A4_DEX_trimmed.fa
11: A4_CTRL_trimmed.fa
10: A3_DEX_trimmed.fa
9: A3_CTRL_trimmed.fa
8: A2_DEX_trimmed.fa
7: A2_CTRL_trimmed.fa
6: A1_DEX_trimmed.fa
5: 23847391_33396705.gtf
4: mm10-ch17.fasta
3: FastQC on data 1: RawData
2: FastQC on data 1: Webpage
1: A1_CTRL_trimmed.fa

Spliced alignment parameters ->
Specify alignment parameters ->
GTF file with known splice sites ->
23847391_33396705.gtf

Obliczenie statystyki (Cuffdiff)

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 1.1 GB

Tools search tools

Get Data Send Data Collection Operations Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Fetch Alignments Statistics Graph/Display Data NGS: RNA Analysis NGS: QC and manipulation BAM-to-SAM convert BAM to SAM Plot Data Workflows All workflows

Cuffdiff find significant changes in transcript expression, splicing, and promoter use (Galaxy Version 2.2.1.5)

Transcripts

5: 23847391_33396705.gtf

A transcript GFF3 or GTF file produced by cufflinks, cuffcompare, or other source.

Omit Tabular Datasets

Yes No Discard the tabular output.

Generate SQLite

Yes No Generate a SQLite database for use with cummeRbund.

Input data type

SAM/BAM CuffNorm supports either CXB (from cuffquant) or SAM/BAM input files. Mixing is not supported. Default: SAM/BAM

Condition

1: Condition

Condition name CTRL

Replicates

19: A3_CTRL_trimmed.bam
18: A2_DEX_trimmed.bam
17: A2_CTRL_trimmed.bam
16: A1_DEX_trimmed.bam
15: A1_CTRL_trimmed.bam

2: Condition

Condition name DEX

Replicates

20: A3_DEX_trimmed.bam
19: A3_CTRL_trimmed.bam
18: A2_DEX_trimmed.bam
17: A2_CTRL_trimmed.bam
16: A1_DEX_trimmed.bam

+ Insert condition

Library normalization method

geometric

Dispersion estimation method

pooled

If using only one sample per condition, you must use 'blind.'

False Discovery Rate

0.05

History

search datasets

rna-seq-tutorial

36 shown, 10 deleted

1.1 GB

All None For all selected...

20: A3_DEX_trimmed.bam
19: A3_CTRL_trimmed.bam
18: A2_DEX_trimmed.bam
17: A2_CTRL_trimmed.bam
16: A1_DEX_trimmed.bam
15: A1_CTRL_trimmed.bam
14: A5_DEX_trimmed.fa
13: A5_CTRL_trimmed.fa
12: A4_DEX_trimmed.fa
11: A4_CTRL_trimmed.fa
10: A3_DEX_trimmed.fa
9: A3_CTRL_trimmed.fa
8: A2_DEX_trimmed.fa
7: A2_CTRL_trimmed.fa
6: A1_DEX_trimmed.fa
5: 23847391_33396705.gtf
4: mm10-ch17.fasta
3: FastQC on data 1: RawData
2: FastQC on data 1: Webpage
1: A1_CTRL_trimmed.fa

Filtrowanie

Filter data on any column using simple expressions (Galaxy Version 1.1.0) ▾ Options

Filter

   78: Cuffdiff on data 69, data 67, and others: gene differential expression testing ▾

Dataset missing? See TIP below.

With following condition

c14=='yes'

Double equal signs, ==, must be used as shown above. To filter for an arbitrary string, use the Select tool.

Number of header lines to skip

1

Execute

Wizualizacja IGV

70: HISAT2 on data 12, d
ata 11, and data 10

42.2 MB
format: **bam**, database: **mm10_v2**

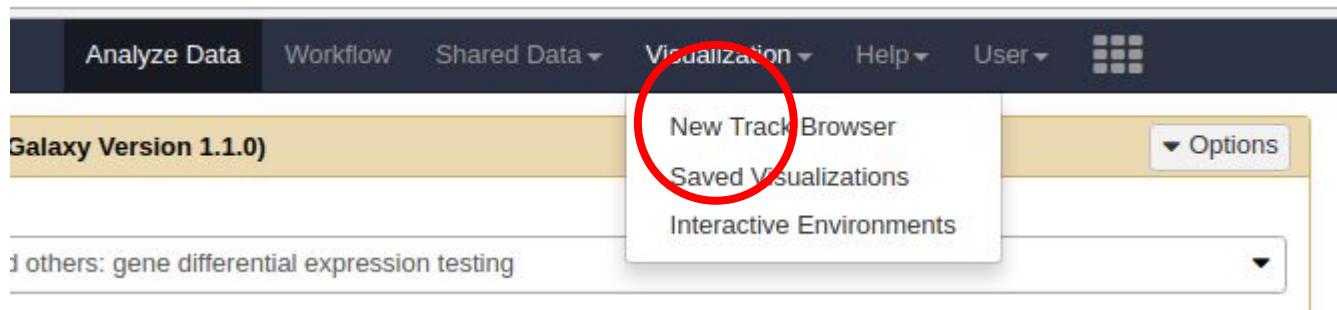
Building DifferenceCoverSample
Building sPrime
Building sPrimeOrder
V-Sorting samples
V-Sorting samples time: 00:00:02
Allocating rank array
Ranking v-sort output
Ranking v-sort output time: 00:00:01
Invoking Larsson-Sadakane on ra



display with [IGV local](#)
display in [IGV View](#)
[display at bam://bam.iobio.io](#)

Binary bam alignments file

Wizualizacja



New Visualization

Browser name:

Reference genome build (dbkey):

 ▾

Is the build not listed here? [Add a Custom Build](#)

Cancel

Create

Wizualizacja

<http://hgdownload.cse.ucsc.edu/goldenPath/mm10/bigZips/mm10.chrom.sizes>

lub Google: mm10 chrom sizes

New Build

Name (eg: Hamster):
mm10

Key (eg: hamster_v1):
mm10_v2

Definition:

[FASTA](#) [Len File](#) [Len Entry](#)

chr1	195471971
chr2	182113224
chrX	171031299
chr3	160039680
chr4	156508116
chr5	151834684
chr6	149736546
chr7	145441459

Submit

Wizualizacja Track browser

Zmień właściwości plików do wizualizacji (genom referencyjny)

Dodaj pliki do wizualizacji

Analiza Whole Exome Sequencing

Ashkenazi Trio (chr22)

Znaleźć rzadkie mutacje homozygotyczne mające wpływ na białko

Dane

<http://178.216.203.113/training-data/exome/patient.bam>

<http://178.216.203.113/training-data/exome/mother.bam>

<http://178.216.203.113/training-data/exome/father.bam>

<http://178.216.203.113/training-data/exome/exac22.vcf>

Wczytanie genomu

Screenshot of the Galaxy Docker Build interface.

The top navigation bar includes: Analyze Data, Workflow, Shared Data (highlighted), Visualization, Admin, Help, User, and a grid icon.

A red circle highlights the "Shared Data" dropdown menu, which is open to show options: Data Libraries, Histories, Workflows, Visualizations, and Stories.

The main content area displays a message: "Hello, your Galaxy Docker container is up and running!" followed by instructions to customize a welcome.html page.

Below the message are three buttons: Configuring Galaxy, Installing Tools, and Guided Tour.

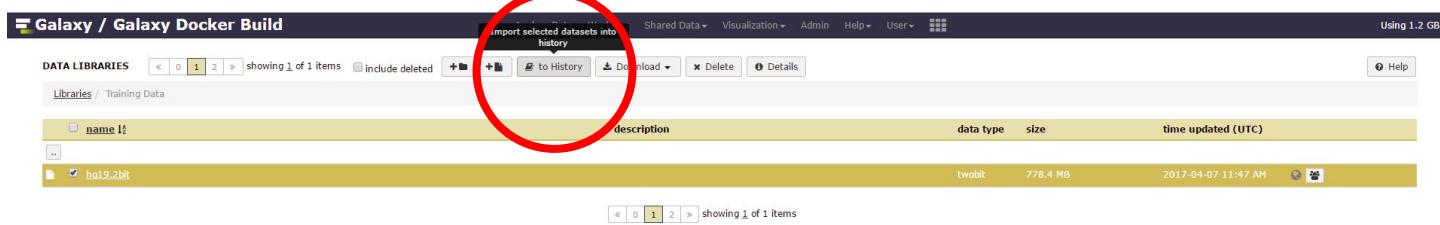
A central graphic features a blue whale carrying several shipping containers, with the text "Docker In Galaxy" and "getgalaxy.org".

The right sidebar shows a "History" section with four datasets:

Dataset	Size	Actions
4: hg19.2bit	95.17 MB	
3: father.bam		
2: patient.bam		
1: mother.bam		

The bottom left corner shows the URL: 178.216.200.146:8080/library/list

Wczytanie genomu



The screenshot shows the Galaxy web interface with the title "Galaxy / Galaxy Docker Build". The main area displays a table of datasets in the "Training Data" library. One dataset, "hg19.2bit", is highlighted with a red circle. The table columns include name, description, data type, size, and time updated (UTC). The "hg19.2bit" entry has a checked checkbox next to it. The interface includes navigation buttons (back, forward, search) and a toolbar with various icons.

name	description	data type	size	time updated (UTC)
hg19.2bit		twobit	778.4 MB	2017-04-07 11:47 AM

Variant calling (freebayes)

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User Using 4.2 GB

Tools

- search tools
- Get Data
- Send Data
- Collection Operations
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Statistics
- Graph/Display Data
- NGS: RNA Analysis
- NGS: OC and manipulation
- BAM-to-SAM convert BAM to SAM
- Plot Data
- Visualization
- NGS: Variant Analysis
- Workflows
- All workflows

FreeBayes bayesian genetic variant detector (Galaxy Version 1.0.2.29-2)

Choose the source for the reference genome

BAM file

- 3: father.bam
- 2: patient.bam
- 1: mother.bam

Use the following dataset as the reference sequence

hg19.fasta

Limit variant calling to a set of regions?

Do not limit

Sets -targets or --region options

Choose parameter selection level

1. Simple diploid calling

Select how much control over the freebayes run you need

Execute

Execute: FreeBayes (1.0.2.29-2)

FreeBayes is a Bayesian genetic variant detector designed to find small polymorphisms, specifically SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment.

See <https://github.com/ekg/freebayes> for details on FreeBayes.

Description

Provided BAM file(s) and a reference. FreeBayes will provide VCF output on standard out describing SNPs, indels, and complex variants in samples in the input alignments.

By default, FreeBayes will consider variants supported by at least 2 observations in a single sample (-C) and also by at least 20% of the reads from a single sample (-F). These settings are suitable to low to high depth sequencing in haploid and diploid samples, but users working with polyploid or pooled samples may wish to adjust them depending on the characteristics of their sequencing data.

FreeBayes is capable of calling variant haplotypes shorter than a read length where multiple polymorphisms segregate on the same read. The maximum distance between polymorphisms phased in this way is determined by the --max-complex-gap, which defaults to 3bp. In practice, this can comfortably be set to half the read length.

Ploidy may be set to any level (-p), but by default all samples are assumed to be diploid. FreeBayes can model per-sample and per-region variation in copy-number (-A) using a copy-number variation map.

FreeBayes can act as a frequency-based pooled caller and describe variants and haplotypes in terms of observation frequency rather than called genotypes. To do so, use --pooled-continuous and set input filters to a suitable level. Allele observation counts will be described by AO and RO fields in the VCF output.

Galaxy-specific options

Galaxy allows five levels of control over FreeBayes options provided by Choose parameter selection level menu option. These are:

1. Simple diploid calling: The simplest possible FreeBayes application. Equivalent of using FreeBayes with only a BAM input and no other parameter options.
2. Simple diploid calling with filtering and coverage: Same as #1 plus two additional options: -O (standard filters: --min-mapping-quality 30 --min-base-quality 20 --min-supporting-allele-qsum 0 --genotype-varinat-

History

exome-tutorial

3.07 GB

5: hg19.fasta

4: hg19.bit

3: father.bam

2: patient.bam

1: mother.bam

SnpSift filtrowanie na jakość

The screenshot shows the Galaxy Docker Build interface with the SnpSift Filter tool open. The filter criteria are set to `(QUAL >= 30) & (DP >= 10)`. The History panel on the right shows several datasets: father-primitiv.vcf, father.vcf, hg19.fasta, hg19.2bit, father.bam, patient.bam, and mother.bam.

SnpSift Filter Variants using arbitrary expressions (Galaxy Version 4.1.1)

Variant input file in VCF format: 8: father-primitiv.vcf

Filter criteria: `(QUAL >= 30) & (DP >= 10)`

Inverse filter: No

Show how that variant matches filter expression

Filter mode: Retain entries that pass filter, remove other entries

Execute: Execute

Execute: SnpSift Filter (4.1.1)

You can filter a VCF file using arbitrary expressions, for instance "(QUAL > 30) | (exists INDEL) | (countHet() > 2)". The actual expressions can be quite complex, so it allows for a lot of flexibility.

Some examples:

- I want just the variants from the second million bases of chr1:
`(CHROM = 'chr1') & (POS > 1000000) & (POS < 2000000)`
- Filter value is either 'PASS' or it is missing:
`(FILTER = 'PASS') | (na FILTER)`
- I want to filter lines with an ANN annotation EFFECT of 'frameshift_variant' (for vcf files using Sequence Ontology terms):
`(ANN(*).EFFECT has 'frameshift_variant')`

Important According to the specification, there can be more than one EFFECT separated by & (e.g. 'missense_variant&splice_region_variant', thus using has operator is better than using equality operator (=)). For instance 'missense_variant&splice_region_variant' = 'missense_variant' is false, whereas 'missense_variant&splice_region_variant' has 'missense_variant' is true.

- I want to filter lines with an EFF of 'FRAME_SHIFT' (for vcf files using Classic Effect names):
`(EFF[*].EFFECT = 'FRAME_SHIFT')`
- I want to filter out samples with quality less than 30:
`(QUAL > 30)`
- ...but we also want InDels that have quality 20 or more:
`((exists INDEL) & (QUAL >= 20)) | (QUAL >= 30)`
- ...or any homozygous variant present in more than 3 samples:
`(countHom() > 3) | ((exists INDEL) & (QUAL >= 20)) | (QUAL >= 30)`
- ...or any heterozygous sample with coverage 25 or more:

Snpeff download annotacja wariantów

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User Using 4.2 GB

Tools

Snpeff Download a new database (Galaxy Version 4.1.0)

Select the genome version you want to download (e.g. GRCh37.74)

hg19
https://snpeff-data.github.io/databases/v4_1/

Execute: Snpeff Download (4.1.0) tool, please go to:
http://snpeff.sourceforge.net/Snpeff_manual.html

Citations Show BibTeX

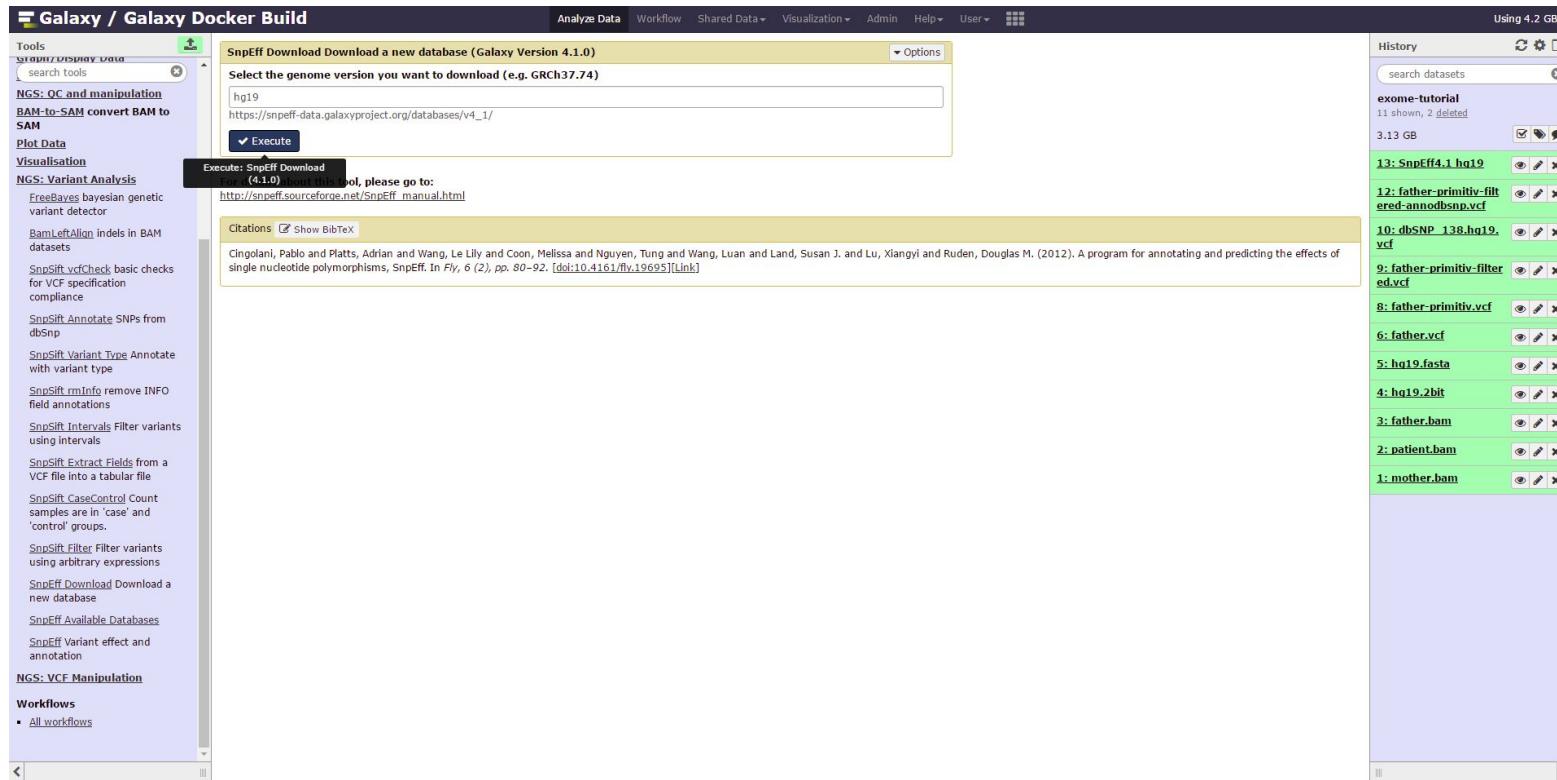
Cingolani, Pablo and Platts, Adrian and Wang, Le Lily and Coon, Melissa and Nguyen, Tung and Wang, Luan and Land, Susan J. and Lu, Xiangyi and Ruden, Douglas M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, Snpeff. In *Fly*, 6 (2), pp. 80–92. [doi:10.4161/fly.19695] [Link]

History

search datasets

exome-tutorial
11 shown, 2 deleted
3.13 GB

13: Snpeff4.1 hg19
12: father-primitiv-filter
ered-annodbsnp.vcf
10: dbSNP_138.hg19.vcf
9: father-primitiv-filter.ed.vcf
8: father-primitiv.vcf
6: father.vcf
5: hg19.fasta
4: hg19.2bit
3: father.bam
2: patient.bam
1: mother.bam



Snpeff annotacja wariantów

Galaxy / Galaxy Docker Build Analyze Data Workflow Shared Data Visualization Admin Help User Using 4.2 GB

Tools
NGS: QC and manipulation
BAM-to-SAM convert BAM to SAM
Plot Data
Visualisation
NGS: Variant Analysis
FreeBayes bayesian genetic variant detector
BamEffAllan indels in BAM datasets
SnpSift vcfCheck basic checks for VCF specification compliance
SnpSift Annotate SNPs from dbSNP
SnpSift Variant_Type Annotate with variant type
SnpSift rmInfo remove INFO field annotations
SnpSift Intervals Filter variants using intervals
SnpSift Extract Fields from a VCF file into a tabular file
SnpSift CaseControl Count samples are in 'case' and 'control' groups.
SnpSift Filter variants using arbitrary expressions
SnpSift Download Download a new database
SnpSift Available Databases
SnpEff Variant effect and annotation
NGS: VCF Manipulation
Workflows
• All workflows

SnpEff Variant effect and annotation (Galaxy Version 4.1.0)

Sequence changes (SNPs, MNPs, InDels)
Input format: VCF
Output format: VCF (only if input is VCF)
Genome source: Reference genome from your history
SnpEff4.1 Genome Data
Upstream / Downstream length: 5000 bases
Set size for splice sites (donor and acceptor) in bases: 2 bases
Default: 2
spliceRegion Settings
Annotation options:
 Select/Unselect all
 Perform 'cancer' comparisons (somatic vs. germline)
 Only use canonical transcripts
 Use gene ID instead of gene name (VCF output)
 Add loss of function (LOF) and nonsense mediated decay (NMD) tags
 Add OICR tag in VCF file
 Only use regulation tracks
 Use Classic Effect names and amino acid variant annotations (NON_SYNONYMOUS_CODING vs missense_variant and G180R vs p.Gly180Arg/c.538G>C)
 Override classic and use HGVS annotations for amino acid annotations (p.Gly180Arg/c.538G>C vs G180R)
 Override classic and use Sequence Ontology terms for effects (missense_variant vs NON_SYNONYMOUS_CODING)
 Use 'EFF' field compatible with older versions (instead of 'ANN')
 Do not add HGVS annotations.

History
search datasets
exome-tutorial
11 shown, 2 deleted
3.13 GB
13: SnpEff4.1 hg19
12: father-primitiv-filtered-annodbsnp.vcf
10: dbSNP_138.hg19.vcf
9: father-primitiv-filter.ed.vcf
8: father-primitiv.vcf
6: father.vcf
5: hg19.fasta
4: hg19.2bit
3: father.bam
2: patient.bam
1: mother.bam

SnpSift filtrowanie na genotyp

The screenshot shows the Galaxy Docker Build interface. On the left, a sidebar lists various NGS tools, including SnpSift variants for VCF files. The main panel displays the "SnpSift Filter" tool configuration. A red box highlights the filter criteria expression:

```
(isHet(GEN[0]) & isHet(GEN[1]) & isHom(GEN[2]) & isVariant(GEN[2]))
```

The "Filter criteria" section contains the expression `QUAL >= 30 & DP >= 10`. Below it, the "Filter mode" dropdown is set to "Retain entries that pass filter, remove other entries". At the bottom, there is an "Execute" button and a link to the tool's documentation.

The right side of the interface shows the "History" panel, which lists several datasets: father-primitiv.vcf, father.vcf, hq19.fasta, hq19.2bit, father.bam, patient.bam, and mother.bam.

SnpSift filtrowanie na wpływ na białko

The screenshot shows the Galaxy Docker Build interface with the SnpSift Filter tool open. The filter criteria section contains the expression `(ANN[*].IMPACT = 'HIGH') | (ANN[*].IMPACT = 'MODERATE')`, which is highlighted with a red box.

Filter criteria

```
QUAL >= 30 & DP >= 10
```

Filter mode

Execute

Important According to the specification, there can be more than one EFFECT separated by & (e.g. 'missense_variant&splice_region_variant', thus using has operator is better than using equality operator (=)). For instance 'missense_variant&splice_region_variant' = 'missense_variant' is false, whereas 'missense_variant&splice_region_variant' has 'missense_variant' is true.

- I want to filter lines with an ANN annotation EFFECT of 'frameshift_variant' (for vcf files using Sequence Ontology terms):
(ANN(*).EFFECT has 'frameshift_variant')
- I want to filter lines with an EFF of 'FRAME_SHIFT' (for vcf files using Classic Effect names):
(EFF[*].EFFECT = 'FRAME_SHIFT')
- I want to filter out samples with quality less than 30:
(QUAL > 30)
- ...but we also want InDels that have quality 20 or more:
(exists INDEL) & (QUAL >= 20) | (QUAL >= 30)
- ...or any homozygous variant present in more than 3 samples:
(countHom() > 3) | ((exists INDEL) & (QUAL >= 20)) | (QUAL >= 30)
- ...or any heterozygous sample with coverage 25 or more:

SnpSift annotowanie wariantów

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

NGS: QC and manipulation

BAM-to-SAM convert BAM to SAM

Plot Data

Visualisation

NGS: Variant Analysis

FreeBayes bayesian genetic variant detector

BamLeftAlign indels in BAM datasets

SnpSift vcfCheck basic checks for VCF specification compliance

SnpSift Annotate SNPs from dbSNP

SnpSift Variant Type Annotate with variant type

SnpSift rmInfo remove INFO field annotations

SnpSift Intervals Filter variants using intervals

SnpSift Extract Fields from a VCF file into a tabular file

SnpSift CaseControl Count samples in 'case' and 'control' groups.

SnpSift Filter variants using arbitrary expressions

SnpSift Download Download a new database

SnpSift Available Databases

SnpSift Variant effect and annotation

NGS: VCF Manipulation

Workflows

All workflows

SnpSift Annotate SNPs from dbSNP (Galaxy Version 4.1.1)

Variant input file in VCF format
1: father-primitiv-filtered.vcf

VCF File with ID field annotated (e.g. dnsNP.vcf)
2: dbSNP_138.hg19.vcf

The ID field for a variant in input will be assigned from a matching variant in this file.

Only annotate ID field (do not add INFO field)
 Yes No

Allow unsorted VCF files
 Yes No

This option will load the entire 'database' VCF file into memory (which may not be practical for large 'database' VCF files). Otherwise, both the database and the input VCF files should be sorted by position (Chromosome sort order can differ between files).

Execute

Execute: SnpSift Annotate (4.1.1)

Annotating only the ID field from dbSnp137.vcf

Input VCF:
#CHROM POS ID REF ALT QUAL FILTER INFO
22 16157571 . T G 0.0 FAIL NS=53
22 16346045 . T C 0.0 FAIL NS=244
22 16350245 . C A 0.0 FAIL NS=192

Annotated Output VCF:
#CHROM POS ID REF ALT QUAL FILTER INFO
22 16157571 . T G 0.0 FAIL NS=53
22 16346045 rs56234788 T C 0.0 FAIL NS=244
22 16350245 rs2905295 C A 0.0 FAIL NS=192

Annotating both the ID and INFO fields from dbSnp137.vcf

Input VCF:
#CHROM POS ID REF ALT QUAL FILTER INFO
22 16157571 . T G 0.0 FAIL NS=53
22 16346045 . T C 0.0 FAIL NS=244
22 16350245 . C A 0.0 FAIL NS=192

Annotated Output VCF:
#CHROM POS ID REF ALT QUAL FILTER INFO
22 16157571 . T G 0.0 FAIL NS=53
22 16346045 rs56234788 T C 0.0 FAIL NS=244;RSPOS=16346045;GMAF=0.162248628884826;dbSNPBuildID=129;SSR=0;SAO=0;VP=0501000000000000100000100;WGT=0;VC=SNV;SLO;GNO
22 16350245 rs2905295 C A 0.0 FAIL NS=192;RSPOS=16350245;GMAF=0.230804387568556;dbSNPBuildID=101;SSR=1;SAO=0;VP=0500000000000000100000140;WGT=0;VC=SNV;GNO

For details about this tool, please go to:
http://snpeff.sourceforge.net/SnpEff_manual.html
<http://snpeff.sourceforge.net/SnpSift.html#annotate>

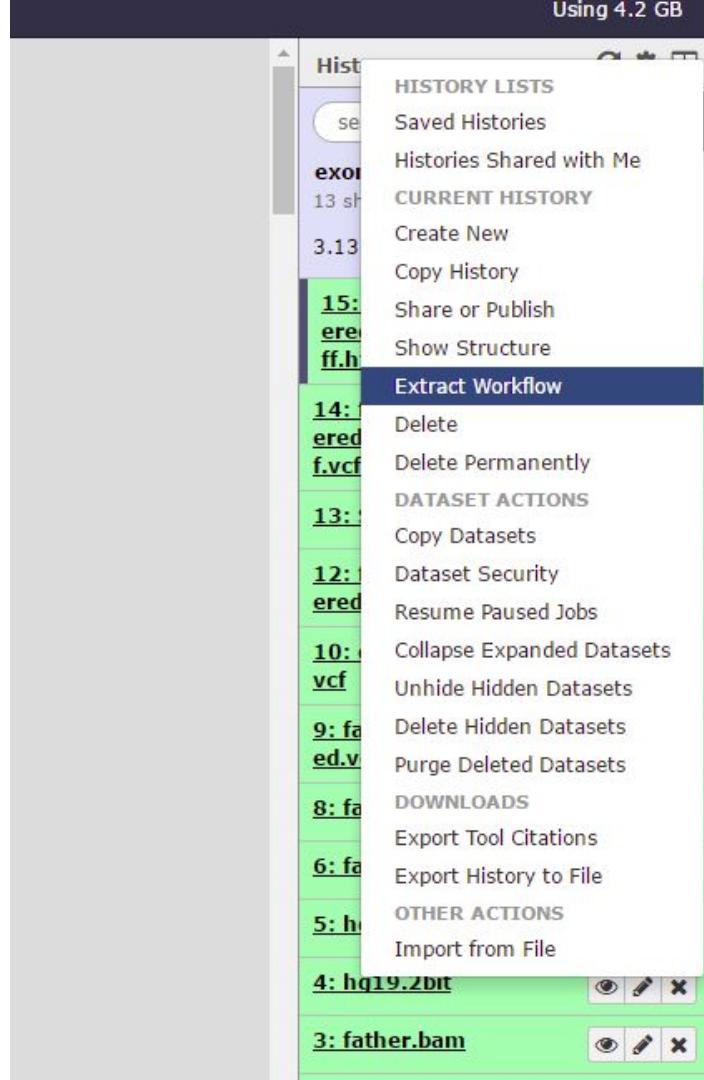
History

search datasets

exome-tutorial
9 shown, 2 deleted
3.08 GB

10: dbSNP_138.hg19.vcf
9: father-primitiv-filter.ed.vcf
8: father-primitiv.vcf
6: father.vcf
5: hg19.fasta
4: hg19.2bit
3: father.bam
2: patient.bam
1: mother.bam

Workflows



Workflows

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User Using 4.2 GB

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
vcf-calling	6

Shared with you by others

vcf-calling
shared with you.

menu

- vcf-calling ▾
- Edit**
- Run
- Share or Download
- Copy
- Rename
- View
- Delete

Workflows

Galaxy / Galaxy Docker Build

Analyze Data Workflow Shared Data Visualization Admin Help User Using 4.2 GB

Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
vcf-calling	6

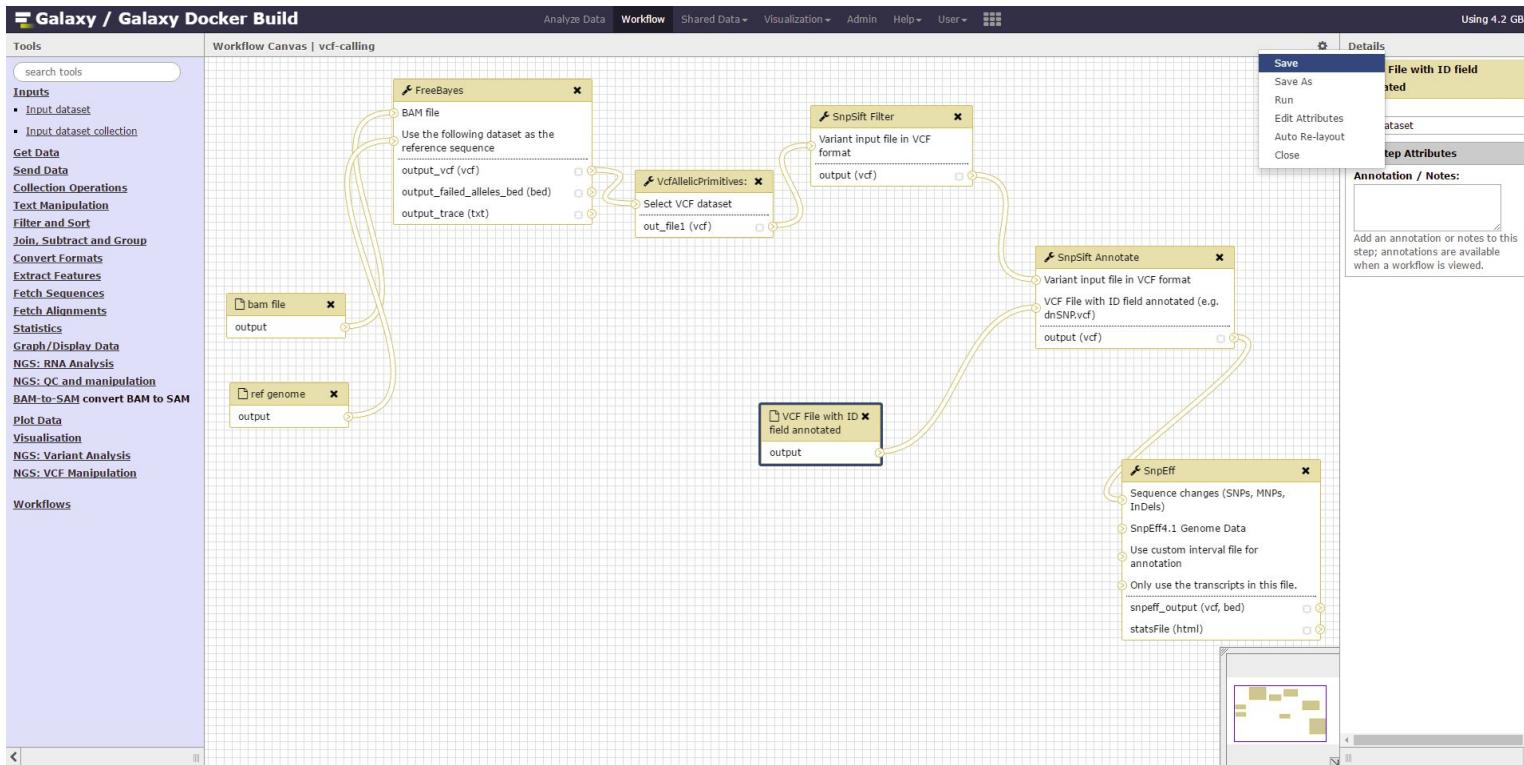
Shared with you by others

vcf-calling
shared with you.

menu

- vcf-calling ▾
- Edit**
- Run
- Share or Download
- Copy
- Rename
- View
- Delete

Workflows



Workflows

Galaxy / Galaxy Docker Build

Tools

- Get Data
 - Upload File from your computer
 - UCSC Main table browser
 - UCSC Test table browser
 - UCSC Archaea table browser
 - EBI SRA ENA SRA
 - Get Microbial Data
 - BioMart Ensembl server
 - CBL Rice Mart rice mart
 - GrameneMart Central server
 - modENCODE fly server
 - Flymine server
 - Flymine test server
 - modENCODE modMine server
 - MouseMine server
 - Ratmine server
 - YeastMine server
 - metabolomine server
 - modENCODE worm server
 - WormBase server
 - Wormbase test server
 - ZebrafishMine server
 - EuPathDB server
 - HbVar Human Hemoglobin Variants and Thalassemias
 - GenomeSpace import from file browser
- Send Data
- Collection Operations
- Text Manipulation

Tool

Analyze Data Workflow Shared Data Visualization Admin Help User

History items created

- 1 mother.bam
 - Treat as input dataset: mother.bam
- 2 patient.bam
 - Treat as input dataset: patient.bam
- 3 father.bam
 - Treat as input dataset: father.bam
- 4 hg19.2bit
 - Treat as input dataset: hg19.2bit
- 5 hg19.fasta
- 6 father.vcf
- 8 father-primitiv.vcf
- 9 father-primitiv-filtered.vcf
- 10 dbSNP_138.hg19.vcf
 - Treat as input dataset: dbSNP_138.hg19.vcf
- 12 father-primitiv-filtered-annodbsnp.vcf
- 13 SnpEff4.1 hg19
- 14 father-primitiv-filtered-annodbsnp-snpeff.vcf

History

- exome-tutorial
- 13 shown, 2 deleted
- 3.13 GB
- 15: father-primitiv-filtered-annodbsnp-snpeff.html
- 14: father-primitiv-filtered-annodbsnp-snpeff.vcf
- 13: SnpEff4.1 hg19
- 12: father-primitiv-filtered-annodbsnp.vcf
- 10: dbSNP_138.hg19.vcf
- 9: father-primitiv-filtered.vcf
- 8: father-primitiv.vcf
- 6: father.vcf
- 5: hg19.fasta
- 4: hg19.2bit
- 3: father.bam
- 2: patient.bam
- 1: mother.bam

Data Sharing

