



SINGLE-CELL COPY-NUMBER ANALYSIS

## **VIZBI 2015 WORKSHOP**

Robert Aboukhalil  
raboukha@cshl.edu

# What is Ginkgo?



# Outline

Experimental  
approaches

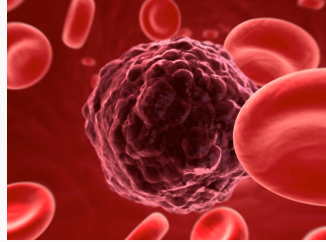
Computational  
approaches for  
CNV analysis

Demo

Hands-on



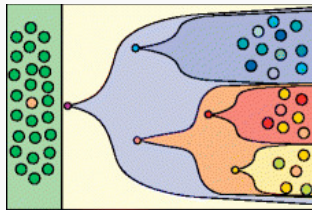
# Single-cell sequencing



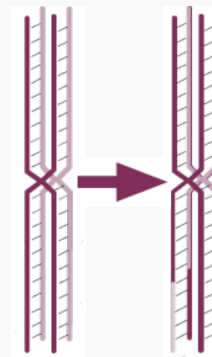
Circulating tumor cells



Neuronal mosaicism



Clonal evolution in tumors



Recombination/  
crossover in germ cells

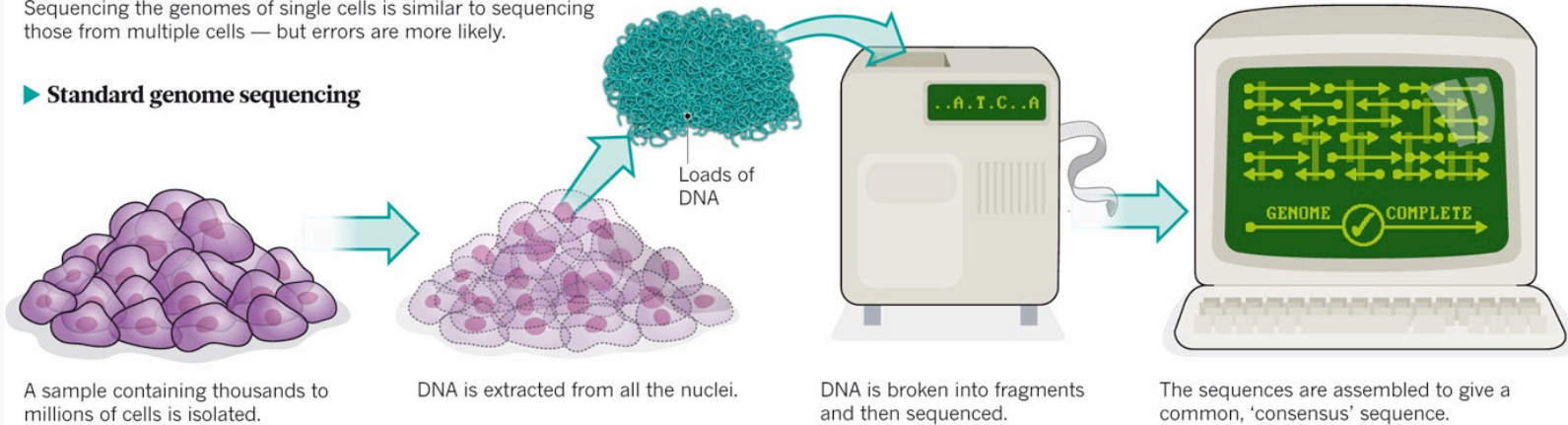


# Single-cell vs. bulk sequencing

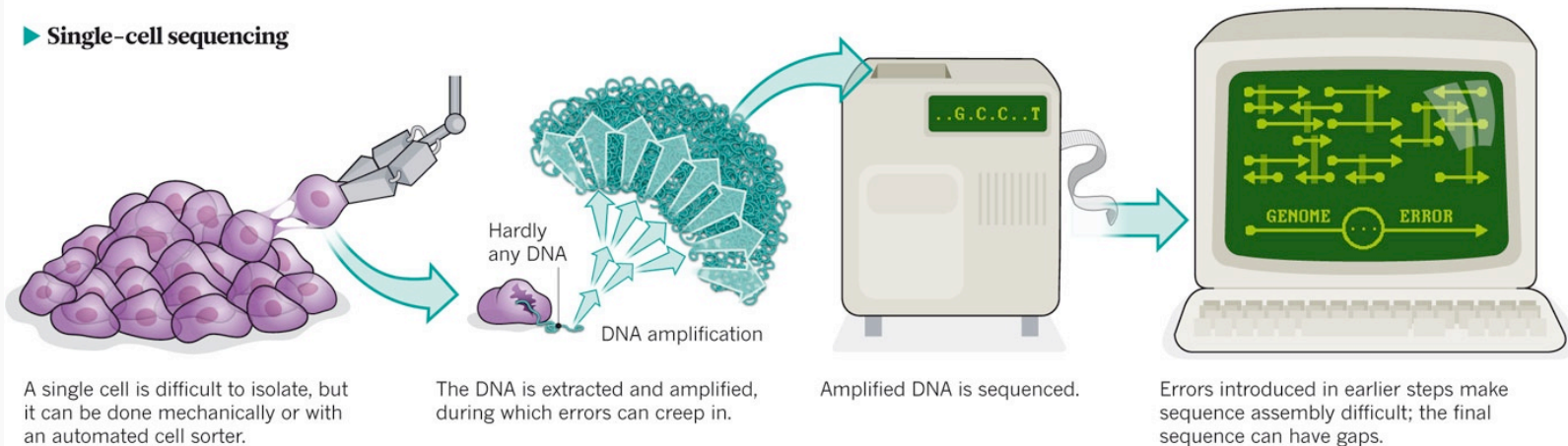
## ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

### ► Standard genome sequencing



### ► Single-cell sequencing

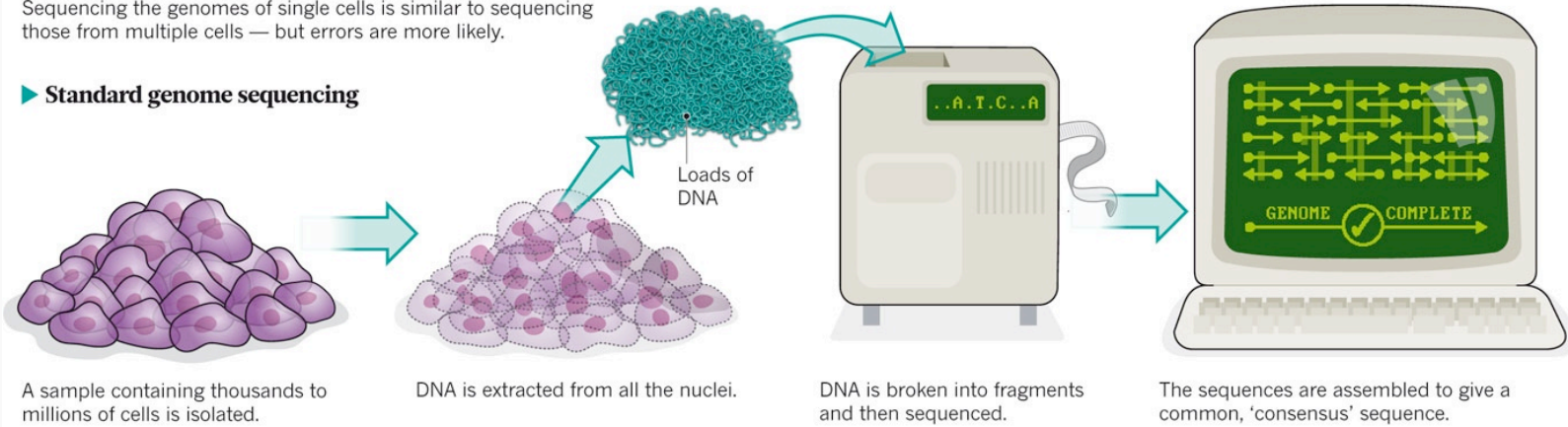


# Single-cell vs. bulk sequencing

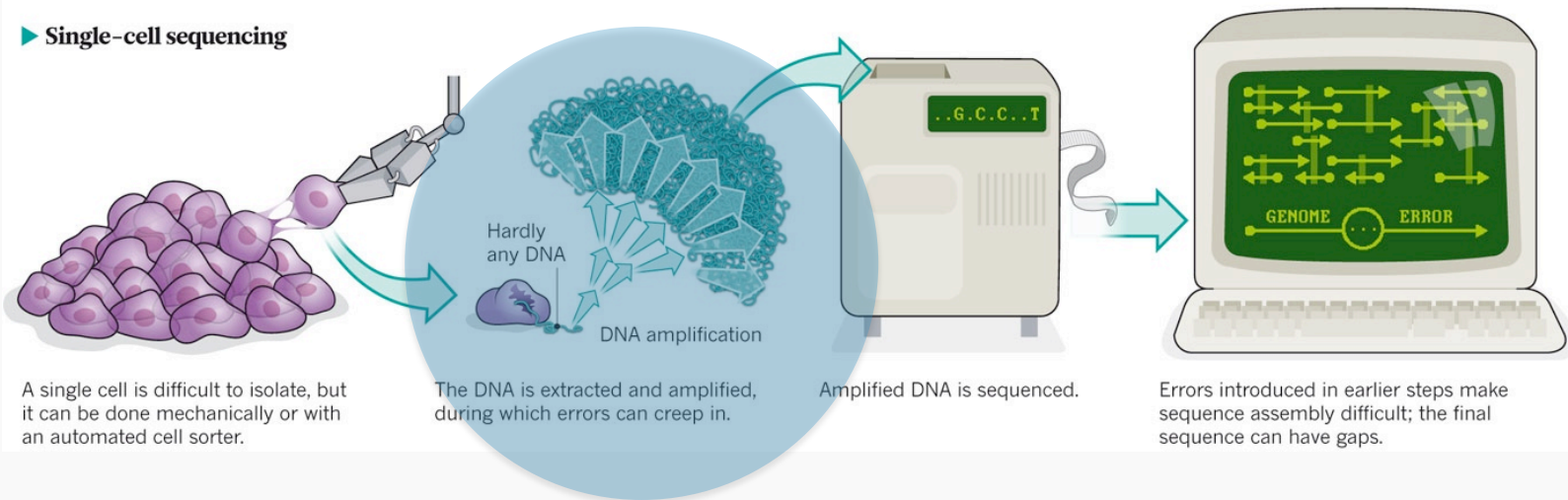
## ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

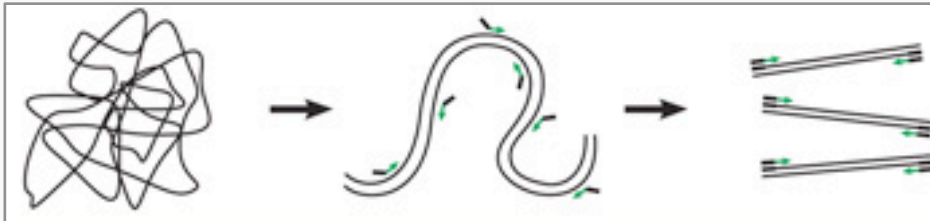
### ► Standard genome sequencing



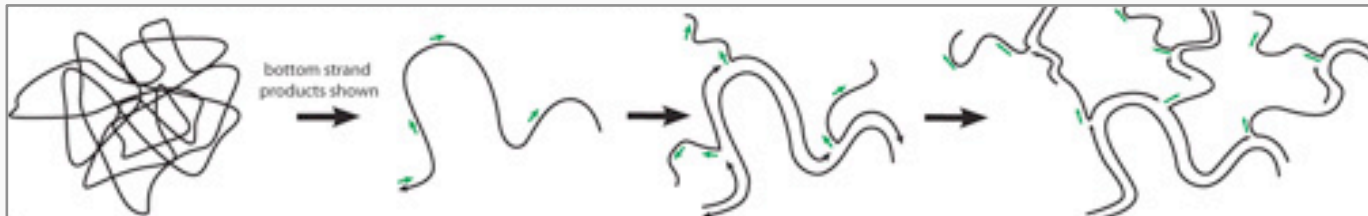
### ► Single-cell sequencing



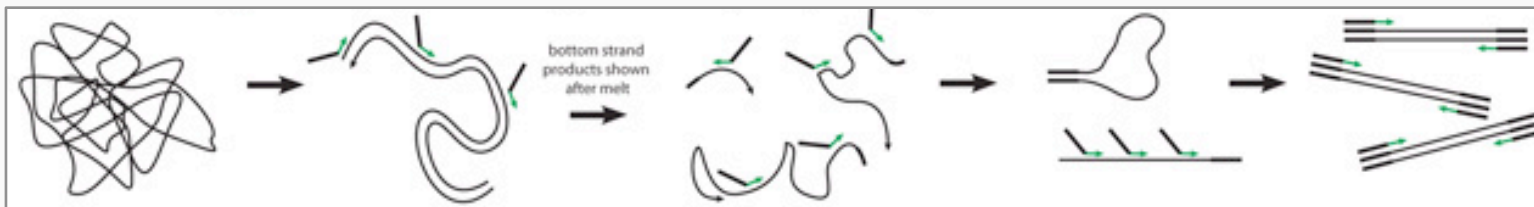
# Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)



MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

# Whole Genome Amplification (WGA) methods

**CSH Cold Spring Harbor Laboratory**

**bioRxiv** beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results [bit.ly/1EkCCOO](https://bit.ly/1EkCCOO)

**Interactive analysis and quality assessment of single-cell copy-number variations**

Tyler Garvin , Robert Aboukhalil , Jude Kendall , Timour Baslan , Gurinder S Atwal , James Hicks , Michael Wigler , Michael Schatz  
doi: <http://dx.doi.org/10.1101/011346>

DOP-PCR (Degenerate Oligonucleotide-Primed PCR)

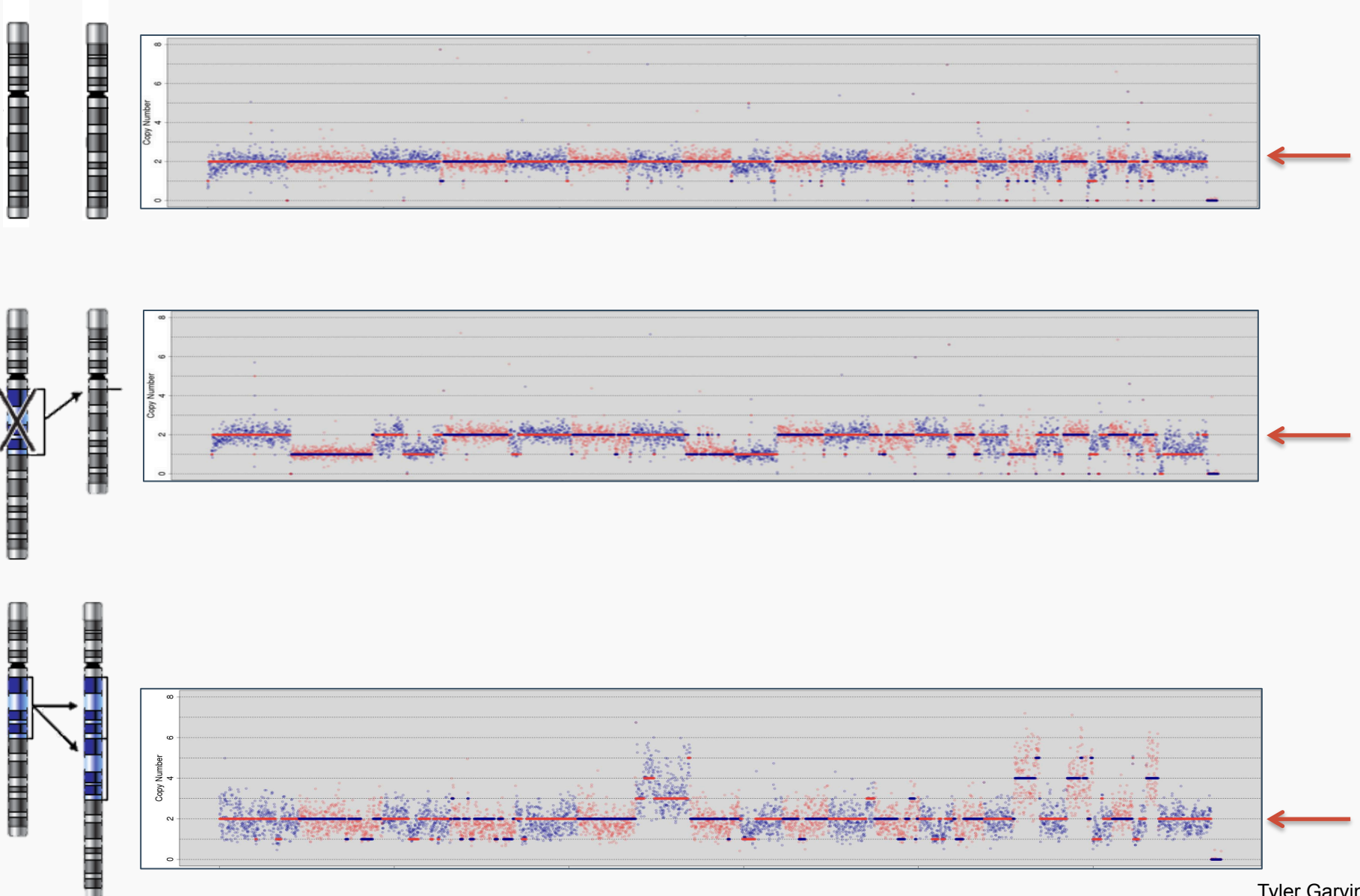
MDA (Multiple Displacement Amplification)

MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

products shown after melt



# Copy-number variant analysis



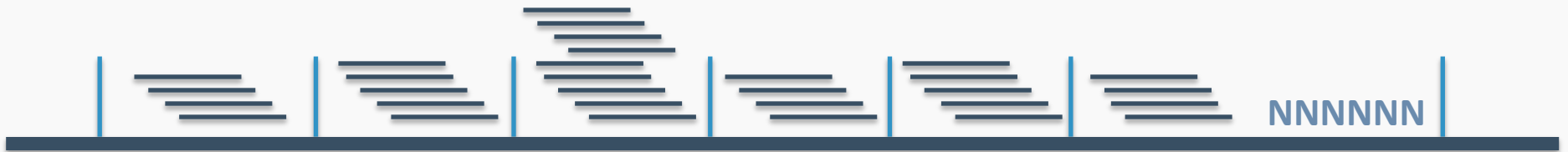
# Copy-number variant analysis



---

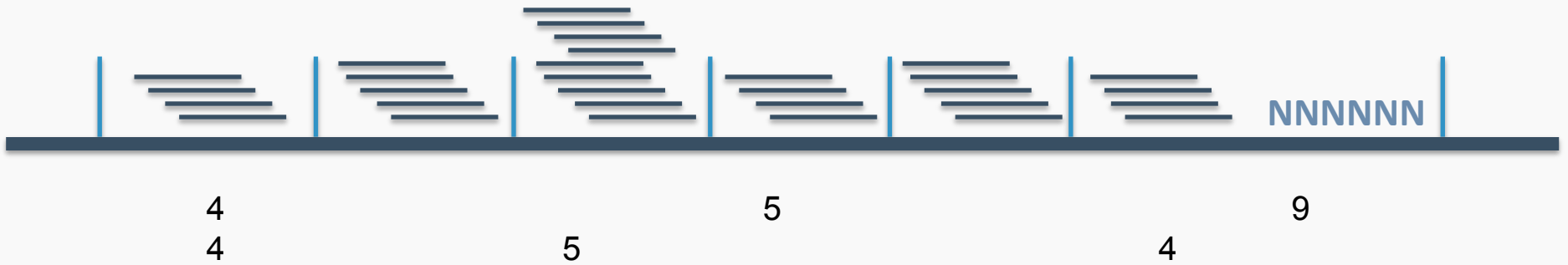
Low coverage allows us to study copy-number variants

# Copy-number variant analysis



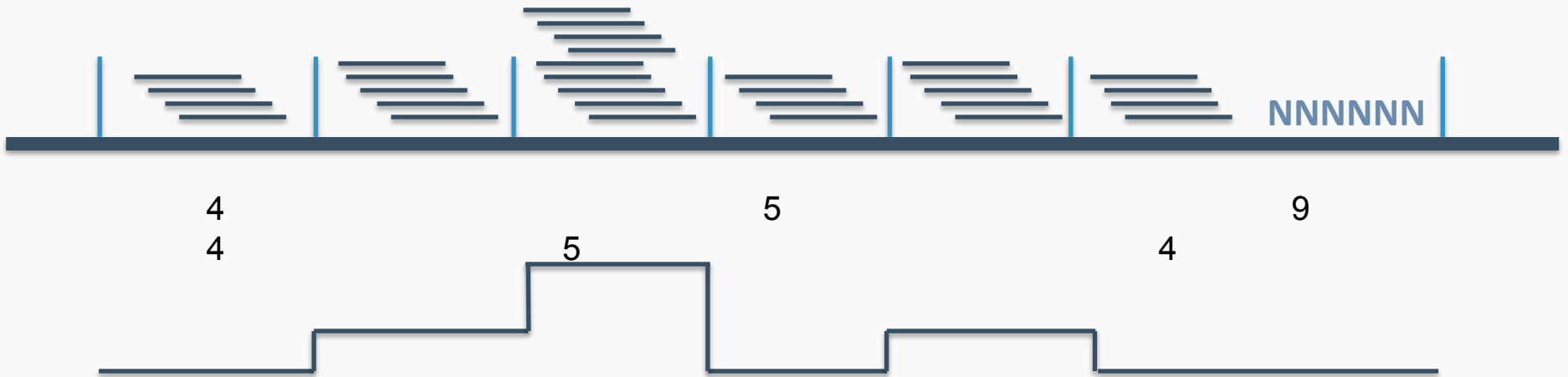
Divide genome into “bins” with ~50 – 100 reads / bin

# Copy-number variant analysis

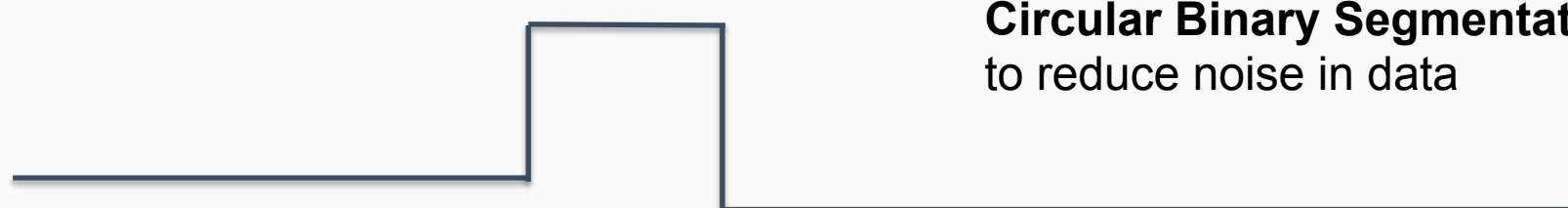




# Copy-number variant analysis



# Copy-number variant analysis



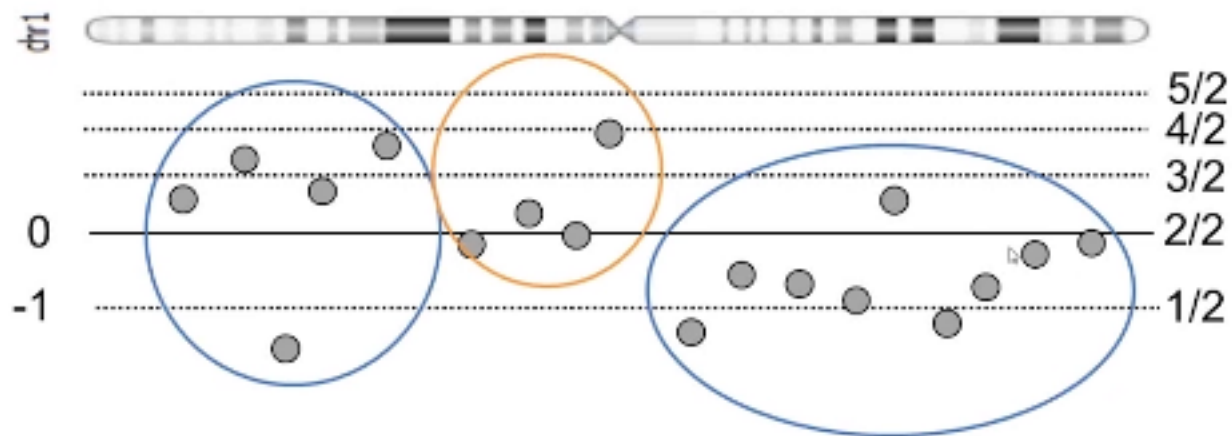
**Circular Binary Segmentation (CBS)**  
to reduce noise in data

4.5

9

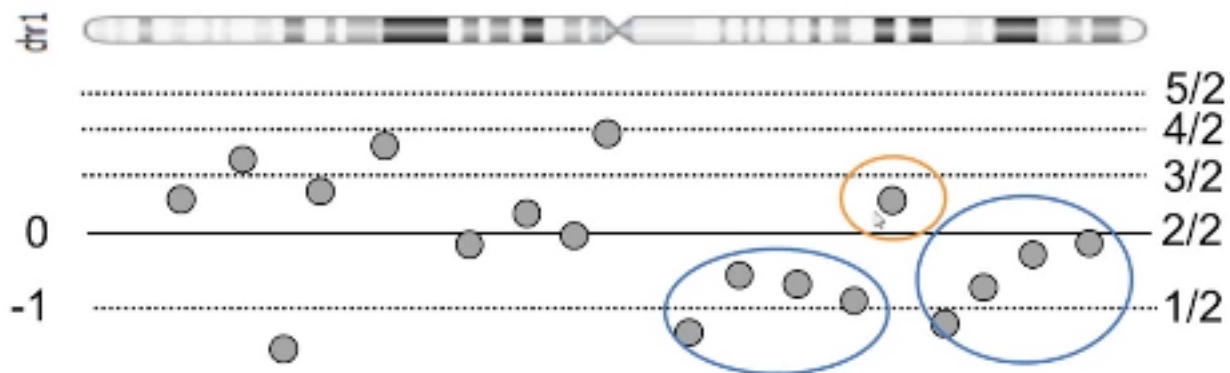
# Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors



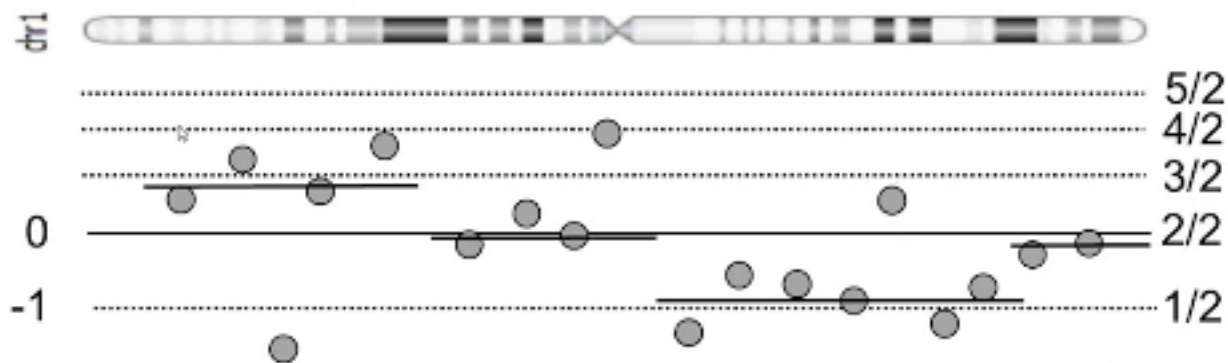
# Circular Binary Segmentation

- Recursively divide up the genome until identify segments that have probe distribution different than neighbors

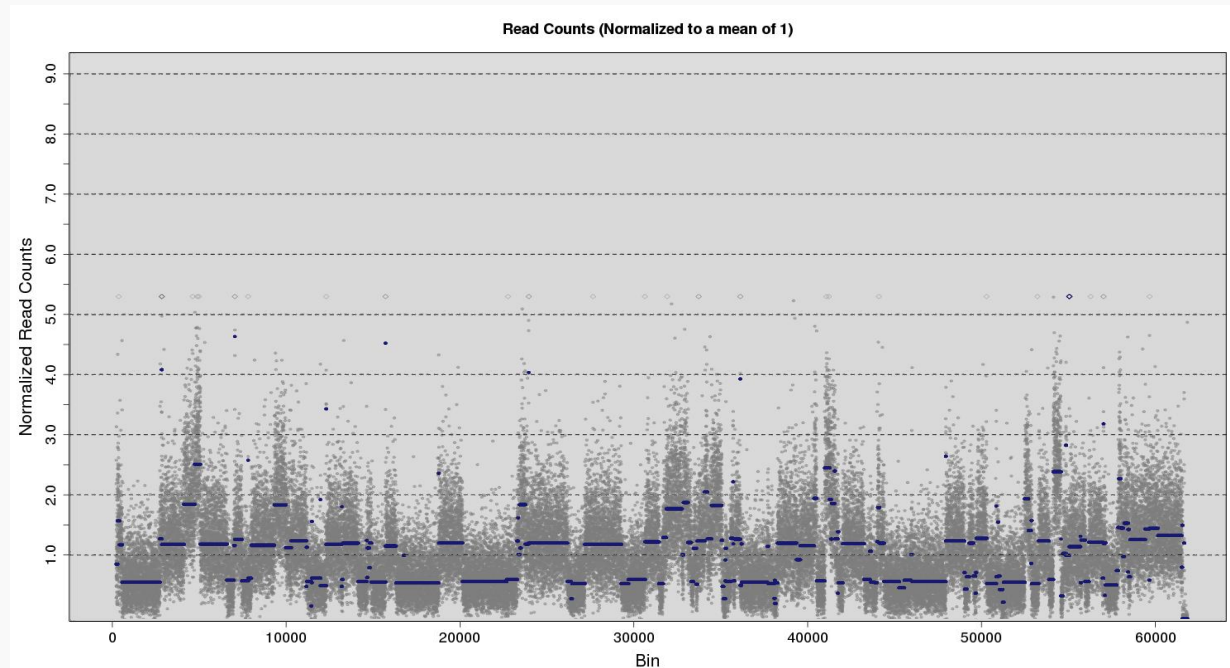


# Circular Binary Segmentation

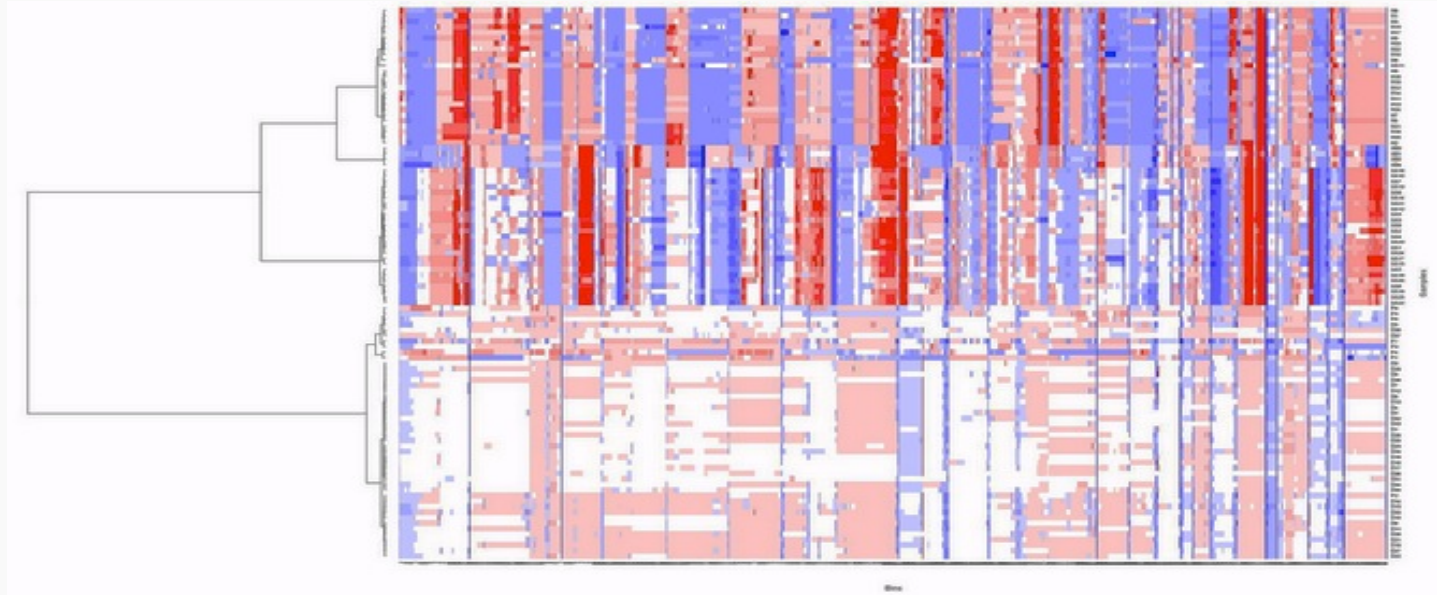
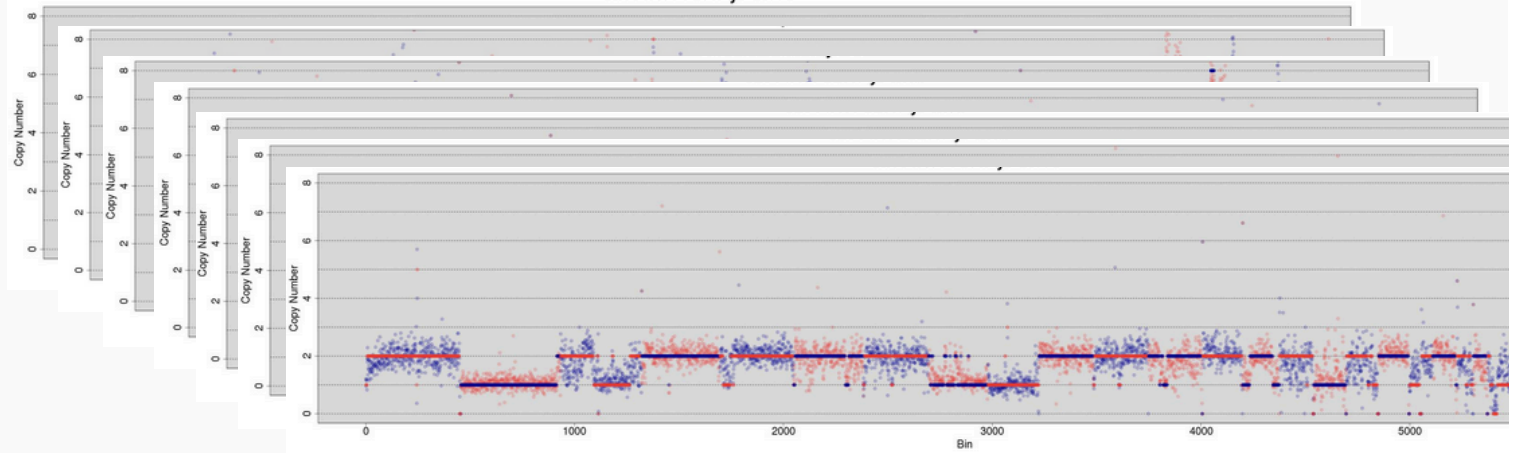
- Recursively divide up the genome until identify segments that have probe distribution different than neighbors



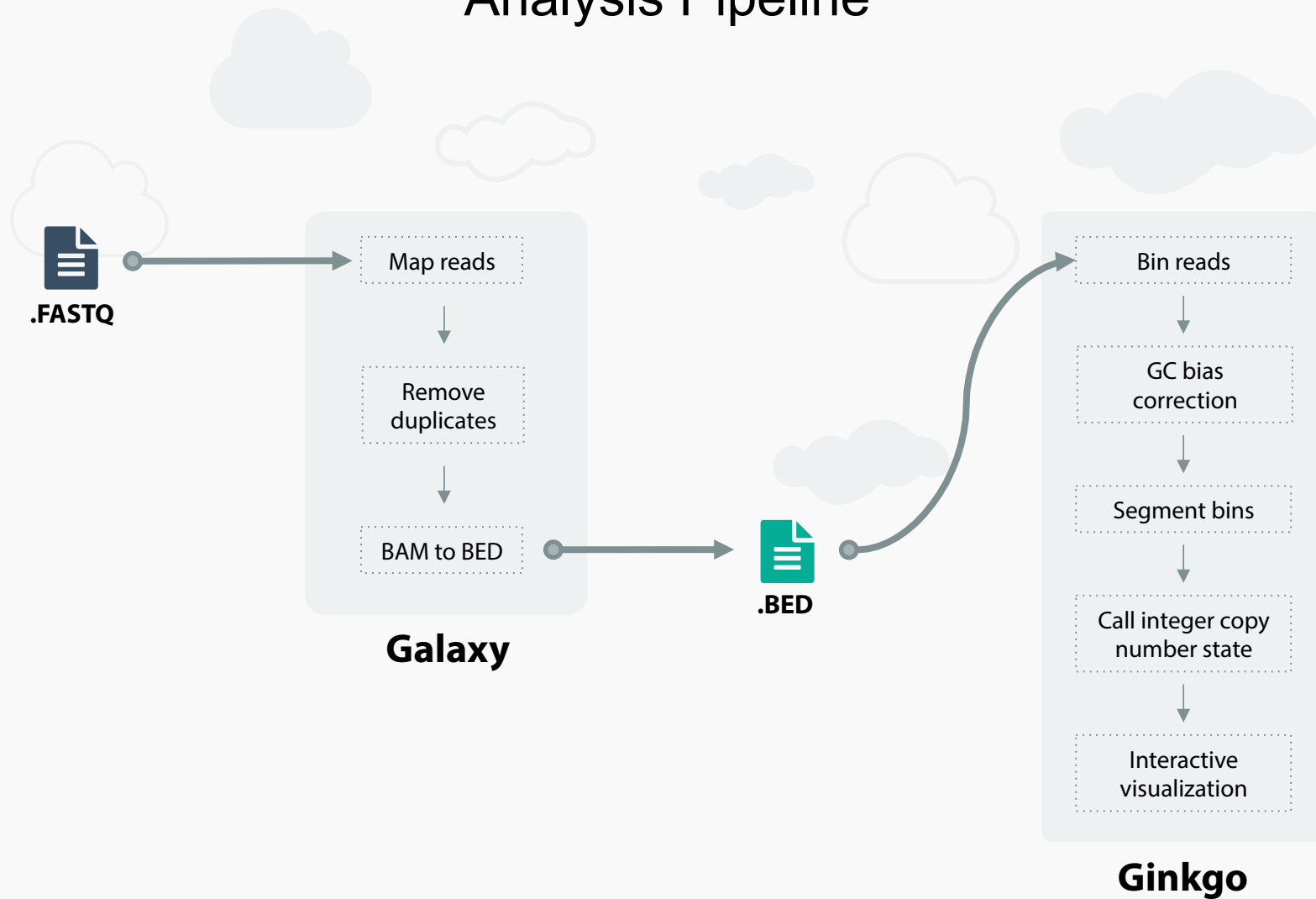
# Copy-number variant analysis



# Copy-number variant analysis



# Analysis Pipeline





# Galaxy intro



# Demo – Galaxy Pipeline

Galaxy <https://usegalaxy.org> Using 0 bytes

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑

- [Get Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snpEff](#)
- [BEDTools](#)
- [Genome Diversity](#)
- [EMBOSS](#)
- [Regional Variation](#)
- [FASTA manipulation](#)
- [Evolution](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [Motif Tools](#)
- [NGS TOOLBOX BETA](#)
- [NGS: SAM Tools](#)

**Galaxy** is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

## Galaxy 101

### Start small

The very first tutorial you need

**PENN STATE**

**JOHNS HOPKINS UNIVERSITY**

**TACC**

**iPlant Collaborative**

[The Galaxy Team](#) is a part of [the Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

History ↻ ⚙️

Unnamed history ✓

0 bytes

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

**Tweets** Follow

**PyPI Recent Updates** @pypi\_updates 6h  
 planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/100CxY1>)  
Retweeted by Galaxy Project

**Galaxy Project** @galaxyproject 9h  
 #usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5iQ](https://pic.twitter.com/URIINYs5iQ)  
Show Photo

**Yannick Wurm** @yannick\_ 12 Mar  
 Fun genome #bioinformatics #iob in

Tweet to @galaxyproject

Galaxy

https://usegalaxy.org

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0 bytes

Tools

search tools

Get Data

Lift-Over

Text Manipulation

Convert Formats

Filter and Sort

Join, Subtract and Group

NGS: QC and manipulation

NGS: Mapping

NGS: BAM Tools

NGS: Picard

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Graph/Display Data

Phenotype Association

snpEff

BEDTools

Genome Diversity

EMBOSS

Regional Variation

FASTA manipulation

Evolution

Multiple Alignments

Metagenomic analyses

Motif Tools

NGS TOOLBOX BETA

NGS: SAM Tools

Galaxy is an open source, web-based platform for data intensive bioinformatics research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

GXYcast 1  
2015

Tweets

PyPI Recent Updates @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxyproject.org> [bit.ly/1O0CxY1](http://bit.ly/1O0CxY1))  
Retweeted by Galaxy Project  
Expand

Galaxy Project @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5IQ](https://pic.twitter.com/URIINYs5IQ)  
Show Photo

Yannick Wurm @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

Tweet to @galaxyproject

History

search datasets

Unnamed history

0 bytes

This history is empty. You can [load your own data](#) or [get data from an external source](#)

PENNSSTATE

JOHNS HOPKINS UNIVERSITY

TACC

iPlant Collaborative

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

https://usegalaxy.org/user/create

Galaxy <https://usegalaxy.org>

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0 bytes

Tools

- Get Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA
- NGS: SAM Tools

### Create account

**Email address:**

**Password:**

**Confirm password:**

**Public name:**

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least four characters in length and contain only lower-case letters, numbers, and the '-' character.

**Subscribe to mailing list:**  
  
See [all Galaxy project mailing lists](#).

Please register only one account. The usegalaxy.org service is provided free of charge and has limited computational and data storage resources. Registration and usage of multiple accounts is tracked and such accounts are subject to termination and data deletion.

[Terms and Conditions for use of this service](#)

History

Unnamed history  
0 bytes

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

<https://usegalaxy.org>

Galaxy <https://usegalaxy.org> Using 0 bytes

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA
- NGS: SAM Tools

Login

Email address:

Password:

[Forgot password? Reset here](#)

Login

OpenID Login

OpenID URL:

Or, authenticate with your  account.

Login

[Terms and Conditions for use of this service](#)

History

Unnamed history 0 bytes

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

Galaxy

https://usegalaxy.org

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0 bytes

Tools

search tools

- Get Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA
- NGS: SAM Tools

History

search datasets

Unnamed history

0 bytes

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Waiting for usegalaxy.org...

You are now logged in as raboukha@cshl.edu.  
You can [go back to the page you were visiting](#) or [go to the home page](#).

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

Want help?  
Get answers.

**BioStars**  
GALAXY EXPLAINED

**Tweets** [Follow](#)

**PyPI Recent Updates** @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxyproject.org> [bit.ly/100CxY1](http://bit.ly/100CxY1))  
Retweeted by Galaxy Project  
Expand

**Galaxy Project** @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINyS5IQ](https://pic.twitter.com/URIINyS5IQ)  
[Show Photo](#)

**Yannick Wurm** @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

Tweet to @galaxyproject

History

Unnamed history  
0 bytes

Annotation:  
[Click here to edit annotation](#)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).



Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User


Tools





**Get Data**  
[EBI SRA](#) [ENA SRA](#)

**NGS: Mapping**  
[Lastz](#) map short reads against reference sequence

**Workflows**  
[All workflows](#)

**Galaxy** is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).



[The Galaxy Team](#) is a part of [the Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).


**History**


**Unnamed history**  
0 bytes


Annotation:  
[Click here to edit annotation](#)

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

**Tweets** [Follow](#)

 **PyPI Recent Updates** @pypi\_updates 6h  
 planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/100CxY1>)  
Retweeted by Galaxy Project

 **Galaxy Project** @galaxyproject 9h  
 #usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINy5iQ](https://pic.twitter.com/URIINy5iQ)

 **Yannick Wurm** @yannick\_ 12 Mar  
 Fun genome #bioinformatics #iob in

Tweet to @galaxyproject

Galaxy <https://usegalaxy.org> Using 0%

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools SRA

Get Data  
[EBI SRA](#) [ENA SRA](#)  
 NGS: Mapping  
[Lastz](#) map short reads against reference sequence  
 Workflows  
 All workflows

EMBL-EBI **ENA** European Nucleotide Archive

Home Search & Browse Submit & Update About ENA Support

Search [input] Search  
 Examples: [BN000065](#), [histone](#)  
 Advanced Sequence

Please subscribe to [ena-announce](mailto:ena-announce@ebi.ac.uk) mailing list here: [listserver.ebi.ac.uk/mailman/listin...](http://listserver.ebi.ac.uk/mailman/listin...) to receive alerts about ENA services.

EMBL-EBI **Services** Research Training About us

EMBL-EBI	Services	Research	Training	Industry	About us
<a href="#">News</a> <a href="#">Brochures</a> <a href="#">Contact us</a> <a href="#">Intranet</a>	<a href="#">By topic</a> <a href="#">By name (A-Z)</a> <a href="#">Help &amp; Support</a>	<a href="#">Overview</a> <a href="#">Publications</a> <a href="#">Research groups</a> <a href="#">Postdocs &amp; PhDs</a>	<a href="#">Overview</a> <a href="#">Train at EBI</a> <a href="#">Train outside EBI</a> <a href="#">Train online</a> <a href="#">Contact organisers</a>	<a href="#">Overview</a> <a href="#">Members Area</a> <a href="#">Workshops</a> <a href="#">SME Forum</a> <a href="#">Contact Industry programme</a>	<a href="#">Overview</a> <a href="#">Leadership</a> <a href="#">Funding</a> <a href="#">Background</a> <a href="#">Collaboration</a> <a href="#">Jobs</a> <a href="#">People &amp; groups</a> <a href="#">News</a> <a href="#">Events</a> <a href="#">Visit us</a> <a href="#">Contact us</a>

EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
 Copyright © EMBL-EBI 2015 | EBI is an outstation of the European Molecular Biology Laboratory | [Privacy](#) | [Cookies](#) | [Terms of use](#)

History search datasets Unnamed history 0 bytes Annotation: [Click here to edit annotation](#)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

[https://usegalaxy.org/tool\\_runner/data\\_source\\_redirect?tool\\_id=ebi\\_sra\\_main](https://usegalaxy.org/tool_runner/data_source_redirect?tool_id=ebi_sra_main)

# SRR054622

The screenshot shows a web browser window with the URL <https://usegalaxy.org>. The main content area displays the ENA (European Nucleotide Archive) search interface. The search bar contains the accession number **SRR054622**. Below the search bar, there are links for "Advanced" and "Sequence". The page header includes navigation tabs for "Services", "Research", "Training", and "About us". The left sidebar contains a "Tools" section with a search bar containing "SRA" and a "Workflows" section with a link to "All workflows". The right sidebar shows a "History" section with a search bar and a message: "This history is empty. You can load your own data or get data from an external source". The footer contains contact information for EMBL-EBI, including the address "Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK" and the phone number "+44 (0)1223 49 44 44".

Galaxy  
https://usegalaxy.org

Galaxy  
Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools  
SRA  
Get Data  
EBI SRA ENA SRA  
NGS: Mapping  
Lastz map short reads against reference sequence  
Workflows  
All workflows

EMBL-EBI  
Services Research Training About us  
ENA  
European Nucleotide Archive  
SRR054622  
Examples: BN000065, histone  
Search  
Advanced  
Sequence  
Home Search & Browse Submit & Update About ENA Support  
Please subscribe to ena-announce mailing list here: [listserv.ebi.ac.uk/mailman/listin...](mailto:listserv.ebi.ac.uk/mailman/listin...) to receive alerts about ENA services.

History  
search datasets  
Unnamed history  
0 bytes  
Annotation:  
Click here to edit annotation  
This history is empty. You can load your own data or get data from an external source

EMBL-EBI  
News  
Brochures  
Contact us  
Intranet  
Services  
By topic  
By name (A-Z)  
Help & Support  
Research  
Overview  
Publications  
Research groups  
Postdocs & PhDs  
Training  
Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers  
Industry  
Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme  
About us  
Overview  
Leadership  
Funding  
Background  
Collaboration  
Jobs  
People & groups  
News  
Events  
Visit us  
Contact us  
EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK +44 (0)1223 49 44 44  
Copyright © EMBL-EBI 2015 | EBI is an outstation of the European Molecular Biology Laboratory | Privacy | Cookies | Terms of use

Galaxy <https://usegalaxy.org> Plug-in blocked Using 0%

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools EMBL-EBI Services Research Training About us History

SRA +

**Get Data**  
[EBI SRA](#) [ENA SRA](#)

**NGS: Mapping**  
[Lastz](#) map short reads against reference sequence

**Workflows**  
 All workflows

**ENA**  
 European Nucleotide Archive

Search  Advanced Sequence

Examples: [BN000065](#), [histone](#)

Home **Search & Browse** [Submit & Update](#) [About ENA](#) [Support](#)

Please subscribe to ena-announce mailing list here: [listserv.ebi.ac.uk/mailman/listin...](mailto:listserv.ebi.ac.uk/mailman/listin...) to receive alerts about ENA services.

**Read: SRR054622**

Illumina Genome Analyzer sequencing; Single Cells from Breast Tumor T10

[Send Feedback](#) [Download: XML](#)

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
Cold Spring Harbor Laboratory		ILLUMINA	Illumina Genome Analyzer	2,789,858	133,913,184

Library Layout	Library Strategy	Library Source	Library Selection	Library Name
SINGLE	WGS	GENOMIC	RANDOM	SC.WGA.LIB

**Broker Name**  
NCBI

Navigation Read Files

This table contains the files for run SRR054622  
[Download files](#)

**View:** [TEXT](#) **Download:** [TEXT](#)

[Select columns](#)

search datasets +

**Unnamed history**  
 0 bytes ☑ 🗑 💬

Annotation:  
[Click here to edit annotation](#)

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

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑

SRA ✕

**Get Data**

[EBI SRA](#) [ENA SRA](#)

**NGS: Mapping**

[Lastz](#) map short reads against reference sequence

**Workflows**

- [All workflows](#)

## Read: SRR054622

Illumina Genome Analyzer sequencing; Single Cells from Breast Tumor T10

**View:** [XML](#) [Send Feedback](#) [Download:](#) [XML](#)

Submitting Centre	Run Date	Platform	Model	Read Count	Base Count
Cold Spring Harbor Laboratory		ILLUMINA	Illumina Genome Analyzer	2,789,858	133,913,184

Library Layout	Library Strategy	Library Source	Library Selection	Library Name
SINGLE	WGS	GENOMIC	RANDOM	SC.WGA.LIB

**Broker Name**  
NCBI

Navigation Read Files

---

This table contains the files for run SRR054622

[Download files](#)

**View:** [TEXT](#) **Download:** [TEXT](#)

[Select columns](#)

**Showing results 1 - 1 of 1 results**

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (gala)
<a href="#">SRP002535</a>	<a href="#">SRP002535</a>	<a href="#">SAMN00014736</a>	<a href="#">SRS074087</a>	<a href="#">SRX021401</a>	<a href="#">SRR054622</a>	<a href="#">9606</a>	<a href="#">Homo sapiens</a>	Illumina Genome Analyzer	SINGLE	<a href="#">File 1</a>	<a href="#">File 1</a>

**History** ↻ ⚙ 📄

search datasets ✕

**Unnamed history**

0 bytes ☑ 🗑 💬

Annotation:

[Click here to edit annotation](#)

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)



Galaxy

https://usegalaxy.org

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

SRA

Get Data

- EBI SRA ENA SRA

NGS: Mapping

- Lastz map short reads against reference sequence

Workflows

- All workflows

History

search datasets

Unnamed history

1 shown

0 bytes

Annotation:

[Click here to edit annotation](#)

1: EBI SRA: SRR05462

2 File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

A job has been successfully added to the queue – resulting in the following dataset:

1: EBI SRA: SRR054622 File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

You are now being redirected back to [Galaxy](#)

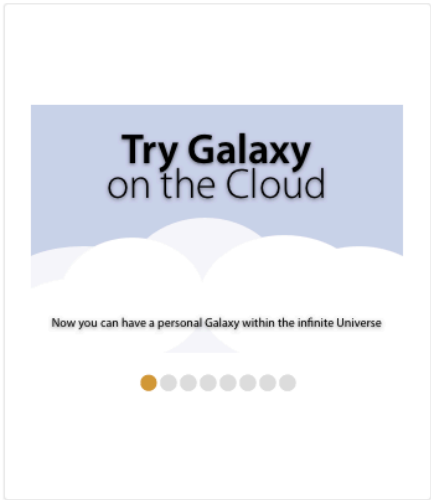
Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).



Now you can have a personal Galaxy within the infinite Universe

**Tweets** [Follow](#)

**PyPI Recent Updates** @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/100CxY1>)  
Retweeted by Galaxy Project  
Expand

**Galaxy Project** @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5IQ](https://pic.twitter.com/URIINYs5IQ)  
[Show Photo](#)

**Yannick Wurm** @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

**History**  0 bytes

Annotation:  
[Click here to edit annotation](#)

1: [EBI SRA: SRR05462](#)

2 File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

PENNSYLVANIA STATE UNIVERSITY  
JOHNS HOPKINS UNIVERSITY  
TACC  
iPlant Collaborative

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).



Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

**Galaxy 101**  
Start small  
The very first tutorial you need

Tweets

**PyPI Recent Updates** @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/100CxY1>)  
Retweeted by Galaxy Project  
Expand

**Galaxy Project** @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5iQ](https://pic.twitter.com/URIINYs5iQ)  
Show Photo

**Yannick Wurm** @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

Tweet to @galaxyproject

History

search datasets

Unnamed history  
1 shown  
0 bytes

Annotation:  
[Click here to edit annotation](#)

1: [EBI SRA: SRR05462](#)  
2 File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

PENNSSTATE      JOHNS HOPKINS UNIVERSITY      TACC      iPlant Collaborative

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).




Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).







**Tweets** [Follow](#)

**PyPI Recent Updates** @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/1O0CxY1>)  
Retweeted by Galaxy Project  
Expand

**Galaxy Project** @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5IQ](https://pic.twitter.com/URIINYs5IQ)  
[Show Photo](#)

**Yannick Wurm** @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

History  394.8 MB     
Annotation: [Click here to edit annotation](#)  
**1: EBI SRA: SRR054622 FI**     
[le: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz)

PENNSSTATE  JOHN HOPKINS UNIVERSITY  TACC  iPlant Collaborative 

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑


- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snpEff](#)
- [BEDTools](#)
- [Genome Diversity](#)
- [EMBOSS](#)
- [Regional Variation](#)
- [FASTA manipulation](#)
- [Evolution](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [Motif Tools](#)
- [NGS TOOLBOX BETA](#)


**Galaxy** is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

## Running Your Own


### Understanding how Galaxy works

An in-depth tutorial







**PENN STATE**



**JOHNS HOPKINS UNIVERSITY**



**TACC**



**iPlant Collaborative**

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

History ↻ ⚙️ 📄

**Unnamed history**  
1 shown


394.8 MB 📄 🗑️ 🗨️


Annotation:  
[Click here to edit annotation](#)


**1: EBI SRA: SRR054622 F** 👁️ ✎️ ✕

**le: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622.fastq.gz** Edit attributes

**Tweets** 🐦 Follow

 **PyPI Recent Updates** @pypi\_updates 6h  
planemo 0.6.0: Command-line utilities to assist in building tools for the Galaxy project (<http://galaxypr... bit.ly/100CxY1>)  
🔄 Retweeted by Galaxy Project  
Expand

 **Galaxy Project** @galaxyproject 9h  
#usegalaxy Histories now support advanced searching and offer multiple other improvements.  
[wiki.galaxyproject.org/Histories](http://wiki.galaxyproject.org/Histories)  
[pic.twitter.com/URIINYs5IQ](https://pic.twitter.com/URIINYs5IQ)  
📷 Show Photo

 **Yannick Wurm** @yannick\_ 12 Mar  
Fun genome #bioinformatics #job in

Tweet to @galaxyproject

<https://usegalaxy.org/datasets/bbd44e69cb8906b570be9a6fd5dd080/edit>

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Attributes Convert Format Datatype Permissions

### Edit Attributes

**Name:**  
EBI SRA: SRR054622 File: ftp://ftp.sra

**Info:**

**Annotation / Notes:**  
  
Add an annotation or notes to a dataset; annotations are available when a history is viewed.

**Database/Build:**  
----- Additional Species Are Below -----

Save

Auto-detect  
This will inspect the dataset and attempt to correct the above column values if they are not accurate.

History

Unnamed history  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622     
File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

<https://usegalaxy.org/datasets/bbd44e69cb8906b570be9a6fd5dd080/edit>

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Attributes Convert Format Datatype Permissions

Change data type

New Type:

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

History

Unnamed history  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622     
File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

<https://usegalaxy.org/datasets/bbd44e69cb8906b570be9a6fd5dd080/edit#datatype>

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Attributes Convert Format Datatype Permissions

Change data type

New Type:

- fastq
- fasta
- fastq
- fastqcssanger
- fastqillumina
- fastqsanger**
- fastqsolexa
- featable
- fitch
- ...

dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

History

Unnamed history  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622  
File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Attributes Convert Format **Datatype** Permissions

Change data type

New Type:  
fastqsanger

This will change the datatype of the existing dataset but *not* modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

Save

History

Unnamed history  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622  
File: <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz>

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Changed the type of dataset 'EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz' to fastqsanger

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

Info:

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

History

Unnamed history  
1 shown

394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Galaxy <https://usegalaxy.org> Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Changed the type of dataset 'EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz' to fastqsanger

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:

Info:

Annotation / Notes:

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

History

Unnamed history  
1 shown

394.8 MB

Annotation:  
[Click here to edit annotation](#)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy

https://usegalaxy.org

Analyze Data Workflow **Shared Data** Visualization Cloud Help User

Using 0%

Tools

search tools

Get Data  
Send Data  
Lift-Over  
Text Manipulation  
Convert Formats  
Filter and Sort  
Join, Subtract and Group  
NGS: QC and manipulation  
NGS: Mapping  
NGS: BAM Tools  
NGS: Picard  
Extract Features  
Fetch Sequences  
Fetch Alignments  
Get Genomic Scores  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
Phenotype Association  
snpEff  
BEDTools  
Genome Diversity  
EMBOSS  
Regional Variation  
FASTA manipulation  
Evolution  
Multiple Alignments  
Metagenomic analyses  
Motif Tools  
NGS TOOLBOX BETA

Galaxy is an open source, web-based, data intensive biomedical research. If you are new to Galaxy, visit our [help resources](#).

**kmer-SVM**  
Galaxy server for identifying predictive regulatory sequence features in NGS data

**Galaxy / Beer Lab**

Data Libraries  
Data Libraries Beta  
Published Histories  
**Published Workflows**  
Published Visualizations  
Published Pages

Yvan Le Bras @Yvan2935 28m  
With @cmonjeaud we love the new @galaxyproject release! Using Klast from @Korilog on @ebiogenouest Galaxy instance pic.twitter.com/N9MNhiuj8f  
Retweeted by Galaxy Project  
Show Photo

Nate Coraor @natefoo 1h  
It's official, #usegalaxy finally has versions! This is 15.03.1: RT @galaxyproject The March 2015 Release is here! bit.ly/galaxy201503  
Retweeted by Galaxy Project  
Expand

Tweet to @galaxyproject

History

search datasets

Unnamed history  
2 shown, 3 hidden  
721.5 MB

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

PENNSSTATE  
JOHNS HOPKINS UNIVERSITY  
TACC  
iPlant Collaborative

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at [Johns Hopkins University](#).

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

https://usegalaxy.org/workflow/list\_published

Galaxy | Published Workflows x

https://usegalaxy.org/workflow/list\_published

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

### Published Workflows

search name, annotation, owner, and tag Q

Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
Ginkgo Preprocessing		robertaboukhalil	★★★★★		~3 minutes ago
Galaxy Variant 101		galaxyproject	★★★★★		Mar 06, 2015
Trim Reads Based on Quality (imported from uploaded file)		jdlim	★★★★★		Feb 28, 2015
RNAseqTRAPLINE	RNA sequencing data analysis in a Transparent Reproducible and Automated Pipeline - TRAPLINE.	mwolfien	★★★★★	de fastq rnaseq tophat2 cummerbund	Feb 25, 2015
TF_ayalawol_302504840		ayalalol	★★★★★		Feb 19, 2015
Bioinfo525_lab4		jing-07	★★★★★	rnaseq	Feb 09, 2015
Lab_3_Bioinformatica		mususu	★★★★★		Feb 07, 2015
Picard Quality stat on BAM		umar	★★★★★		Jan 26, 2015
Quality stat on BAM		umar	★★★★★		Jan 26, 2015
Quality Statistics work flow		umar	★★★★★		Jan 26, 2015
GATK b37 unified genotyper (old)		umar	★★★★★		Jan 26, 2015
mpileup Hg19 MQ 40 BQ 30		umar	★★★★★		Jan 26, 2015
SAM filter Picard duplicate removal GATK recalibration mpileup Hg 19 successful run		umar	★★★★★		Jan 26, 2015
Mpileup BED test		umar	★★★★★		Jan 26, 2015

https://usegalaxy.org/library/index

Galaxy | Published Workflow x

https://usegalaxy.org/u/robertaboukhalil/w/ginkgo-prep

Galaxy Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Published Workflows | robertaboukhalil | Ginkgo Preprocessing

**Galaxy Workflow ' Ginkgo Preprocessing'**

Annotation:

Step	Annotation
<b>Step 1: Input dataset</b>	
<b>Input Dataset</b> select at runtime	
<b>Step 2: Bowtie2</b>	
<b>Is this library mate-paired?</b> Single-end	
<b>FASTQ file</b> Output dataset 'output' from step 1	
<b>Write unaligned reads to separate file(s)</b> False	
<b>Will you select a reference genome from your history or use a built-in index?</b> Use a built-in index	
<b>Select a reference genome</b> hg19	
<b>Specify the read group for this file?</b> No	
<b>Parameter Settings</b> Use defaults	
<b>Step 3: Filter SAM or BAM, output SAM or BAM</b>	
<b>SAM or BAM File to Filter</b> Output dataset 'output' from step 2	
<b>Header in output</b> Include Header	

https://usegalaxy.org/workflow/imp?id=c332d1bbd62e139d

About this Workflow

**Author**  
robertaboukhalil

**Related Workflows**  
[All published workflows](#)  
[Published workflows by robertaboukhalil](#)

**Rating**  
Community (0 ratings, 0.0 average) ★★★★★  
Yours ★★★★★

**Tags**  
Community: none  
Yours:

Import workflow

Workflow home x

← → ↻ <https://usegalaxy.org/workflow> ☆ 🗨️ ☰

**Galaxy** Analyze Data **Workflow** Shared Data ▾ Visualization ▾ Cloud ▾ Help ▾ User ▾ Using 0%

### Your workflows

[Create new workflow](#) [Upload or import workflow](#)

Name	# of Steps
Ginkgo Preprocessing ▾	5

### Workflows shared with you by others

No workflows have been shared with you.

### Other options

[Configure your workflow menu](#)

Workflow home x

https://usegalaxy.org/workflow

Galaxy Analyze Data **Workflow** Shared Data Visualization Cloud Help User Using 0%

### Your workflows

Create new workflow Upload or import workflow

Name	# of Steps
Ginkgo Preprocessing	5

Workflow shared with you by others

No workflows

Other workflows

Configure workflow

- Edit
- Run
- Share or Publish
- Download or Export
- Copy
- Rename
- View
- Delete

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Galaxy [https://usegalaxy.org/root?workflow\\_id=c332d1bbd62e139d](https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d) Using 0%

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

### Running workflow "Ginkgo Preprocessing"

Expand All Collapse

**Step 1: Input dataset**

**Input Dataset**

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

type to filter

**Step 2: Bowtie2** (version 0.2)

**Step 3: Filter SAM or BAM, output SAM or BAM** (version 1.1.1)

**Step 4: rmdup** (version 1.0.0)

**Step 5: Convert from BAM to BED** (version 0.1.0)

Send results to a new history

**Run workflow**

**History**

**Unnamed history**  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

**1: EBI SRA: SRR054622 Fi**     
**le: ftp://ftp.sra.ebi.ac.uk**  
**/vol1/fastq/SRR054/SRR054622/SR**  
**R054622.fastq.gz**

Galaxy [https://usegalaxy.org/root?workflow\\_id=c332d1bbd62e139d](https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d) Using 0%

**Galaxy** Analyze Data Workflow Shared Data Visualization Cloud Help User

**Tools**

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

### Running workflow "Ginkgo Preprocessing" Expand All Collapse

**Step 1: Input dataset**

**Input Dataset**

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

type to filter

**Step 2: Bowtie2** (version 0.2)

**Step 3: Filter SAM or BAM, output SAM or BAM** (version 1.1.1)

**Step 4: rmdup** (version 1.0.0)

**Step 5: Convert from BAM to BED** (version 0.1.0)

Send results to a new history

**Run workflow**

**History**  Using 0%

**Unnamed history**  
1 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

**1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz**

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
5 shown  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

- 5: rmdup on data 3 (as BED)
- 4: rmdup on data 3
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 2: Bowtie2 on data 1: aligned reads
- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

Get Data  
Send Data  
Lift-Over  
Text Manipulation  
Convert Formats  
Filter and Sort  
Join, Subtract and Group  
NGS: QC and manipulation  
NGS: Mapping  
NGS: BAM Tools  
NGS: Picard  
Extract Features  
Fetch Sequences  
Fetch Alignments  
Get Genomic Scores  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
Phenotype Association  
snpEff  
BEDTools  
Genome Diversity  
EMBOSS  
Regional Variation  
FASTA manipulation  
Evolution  
Multiple Alignments  
Metagenomic analyses  
Motif Tools  
NGS TOOLBOX BETA

History

search datasets

Unnamed history  
5 shown  
394.8 MB  
Annotation:  
Click here to edit annotation

5: rmdup on data 3 (as BED)  
4: rmdup on data 3  
3: Filter SAM or BAM, output SAM or BAM on data 2: bam  
2: Bowtie2 on data 1: aligned reads  
1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

Get Data  
Send Data  
Lift-Over  
Text Manipulation  
Convert Formats  
Filter and Sort  
Join, Subtract and Group  
NGS: QC and manipulation  
NGS: Mapping  
NGS: BAM Tools  
NGS: Picard  
Extract Features  
Fetch Sequences  
Fetch Alignments  
Get Genomic Scores  
Operate on Genomic Intervals  
Statistics  
Graph/Display Data  
Phenotype Association  
snpEff  
BEDTools  
Genome Diversity  
EMBOSS  
Regional Variation  
FASTA manipulation  
Evolution  
Multiple Alignments  
Metagenomic analyses  
Motif Tools  
NGS TOOLBOX BETA

History

search datasets

Unnamed history  
4 shown, 1 hidden  
394.8 MB  
Annotation:  
Click here to edit annotation

5: rmdup on data 3 (as BED)  
4: rmdup on data 3  
3: Filter SAM or BAM, output SAM or BAM on data 2: bam  
1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
4 shown, 1 hidden  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

- 5: rmdup on data 3 (as BED)
- 4: rmdup on data 3
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools

NGS TOOLBOX BETA

History

search datasets

Unnamed history  
3 shown, 2 hidden  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

- 5: rmdup on data 3 (as BED)
- 4: rmdup on data 3
- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
3 shown, 2 hidden  
394.8 MB

Annotation:  
[Click here to edit annotation](#)

- 5: rmdup on data 3 (as BED)
- 4: rmdup on data 3
- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
2 shown, 3 hidden

394.8 MB

Annotation:  
[Click here to edit annotation](#)

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

4: rmdup on data 3

5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User

Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
2 shown, 3 hidden

394.8 MB

Annotation:  
[Click here to edit annotation](#)

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

4: rmdup on data 3

5: rmdup on data 3 (as BED)

Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

History

search datasets

Unnamed history  
2 shown, 3 hidden

721.5 MB

Annotation:  
[Click here to edit annotation](#)

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

2: Bowtie2 on data 1: aligned reads

3: Filter SAM or BAM, output SAM or BAM on data 2: bam

4: rmdup on data 3

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz



Galaxy

https://usegalaxy.org/root?workflow\_id=c332d1bbd62e139d

Analyze Data Workflow Shared Data Visualization Cloud Help User Using 0%

Tools

search tools

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- NGS: QC and manipulation
- NGS: Mapping
- NGS: BAM Tools
- NGS: Picard
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- snpEff
- BEDTools
- Genome Diversity
- EMBOSS
- Regional Variation
- FASTA manipulation
- Evolution
- Multiple Alignments
- Metagenomic analyses
- Motif Tools
- NGS TOOLBOX BETA

Successfully ran workflow "Ginkgo Preprocessing". The following datasets have been added to the queue:

- 1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz
- 2: Bowtie2 on data 1: aligned reads
- 3: Filter SAM or BAM, output SAM or BAM on data 2: bam
- 4: rmdup on data 3
- 5: rmdup on data 3 (as BED)

History

search datasets

Unnamed history  
2 shown, 3 hidden

721.5 MB

Annotation:  
[Click here to edit annotation](#)

5: rmdup on data 3 (as BED)

1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz

View data

https://usegalaxy.org/datasets/bbd44e69cb8906b514ab3fface8a1d1f/display?preview=True

Galaxy [https://usegalaxy.org/root?workflow\\_id=c332d1bbd62e139d](https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d) Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑

search tools ✕

- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snpEff](#)
- [BEDTools](#)
- [Genome Diversity](#)
- [EMBOSS](#)
- [Regional Variation](#)
- [FASTA manipulation](#)
- [Evolution](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [Motif Tools](#)
- [NGS TOOLBOX BETA](#)

1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+

History ↺ ⚙️ 📄

search datasets ✕

**Unnamed history**  
2 shown, 3 hidden

721.5 MB 👁️ 🗑️ 🗨️

Annotation:  
[Click here to edit annotation](#)

**5: rmdup on data 3 (as BED)** 👁️ 🗑️ ✕

**1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz** 👁️ 🗑️ ✕

<https://usegalaxy.org/datasets/bbd44e69cb8906b514ab3fface8a1d1f/display/?preview=True>

Galaxy [https://usegalaxy.org/root?workflow\\_id=c332d1bbd62e139d](https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d) Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑

search tools ✕

- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snpEff](#)
- [BEDTools](#)
- [Genome Diversity](#)
- [EMBOSS](#)
- [Regional Variation](#)
- [FASTA manipulation](#)
- [Evolution](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [Motif Tools](#)
- NGS TOOLBOX BETA

1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+
chr10	182965	183013	SRR054622.1414433	42	-

History ↺ ⚙️ 📄

search datasets ✕

**Unnamed history**  
2 shown, 3 hidden

721.5 MB ✔️ 🗑️ 🗨️

Annotation:  
[Click here to edit annotation](#)

---

**5: rmdup on data 3 (as BED)** 👁️ ✎️ ✕

~1,000,000 regions  
format: **bed**, database: **hg19**

🗨️ 🗑️ 🗨️

Annotation:  
[Click here to edit annotation](#)

display in IGB [View](#)  
display at Ensembl [Current](#)  
display at RViewer [main](#)  
display at UCSC [main](#)

1. Chrom	2. Start	3. End	4. Name
chr10	61988	62036	SRR054622.2038687
chr10	69093	69141	SRR054622.2024583
chr10	69098	69146	SRR054622.481587
chr10	69099	69147	SRR054622.739359
chr10	72859	72907	SRR054622.1184885
chr10	92035	92083	SRR054622.2364537

---

**1: EBI SRA: SRR054622 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz** 👁️ ✎️ ✕

Galaxy [https://usegalaxy.org/root?workflow\\_id=c332d1bbd62e139d](https://usegalaxy.org/root?workflow_id=c332d1bbd62e139d) Using 0%

Analyze Data Workflow Shared Data Visualization Cloud Help User

Tools ↑

search tools ✕

- [Get Data](#)
- [Send Data](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Convert Formats](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [NGS: QC and manipulation](#)
- [NGS: Mapping](#)
- [NGS: BAM Tools](#)
- [NGS: Picard](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Phenotype Association](#)
- [snpEff](#)
- [BEDTools](#)
- [Genome Diversity](#)
- [EMBOSS](#)
- [Regional Variation](#)
- [FASTA manipulation](#)
- [Evolution](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [Motif Tools](#)
- [NGS TOOLBOX BETA](#)

1	2	3	4	5	6
chr10	61988	62036	SRR054622.2038687	30	+
chr10	69093	69141	SRR054622.2024583	30	+
chr10	69098	69146	SRR054622.481587	30	+
chr10	69099	69147	SRR054622.739359	30	+
chr10	72859	72907	SRR054622.1184885	30	-
chr10	92035	92083	SRR054622.2364537	30	-
chr10	101602	101650	SRR054622.2476079	30	-
chr10	108416	108464	SRR054622.501574	35	-
chr10	108416	108464	SRR054622.2321097	25	+
chr10	110108	110156	SRR054622.1050882	34	-
chr10	112519	112567	SRR054622.1459050	30	+
chr10	112525	112573	SRR054622.651721	26	+
chr10	113723	113771	SRR054622.2235401	30	-
chr10	114074	114122	SRR054622.2190816	30	+
chr10	114078	114126	SRR054622.1006253	30	+
chr10	124019	124067	SRR054622.2081693	42	+
chr10	130959	131007	SRR054622.213507	40	+
chr10	132896	132944	SRR054622.1964746	34	+
chr10	132905	132953	SRR054622.161640	35	+
chr10	132910	132958	SRR054622.1169143	34	-
chr10	135289	135337	SRR054622.1757379	42	-
chr10	137966	138014	SRR054622.1142440	42	+
chr10	141733	141781	SRR054622.1035785	42	+
chr10	143964	144012	SRR054622.1722443	42	-
chr10	148479	148527	SRR054622.879161	42	+
chr10	148669	148717	SRR054622.1663231	26	+
chr10	168517	168565	SRR054622.1615104	42	-
chr10	168526	168574	SRR054622.1661154	42	-
chr10	168651	168699	SRR054622.2720453	30	+
chr10	173788	173836	SRR054622.2499973	42	-
chr10	176020	176068	SRR054622.1912131	42	+
chr10	179265	179313	SRR054622.1985607	40	-
chr10	182939	182987	SRR054622.1731700	42	-
chr10	182940	182988	SRR054622.1479478	42	-
chr10	182960	183008	SRR054622.531468	40	+

History ↺ ⚙

search datasets ✕

**Unnamed history**  
2 shown, 3 hidden

721.5 MB ✔ 🗑

Annotation:  
[Click here to edit annotation](#)

**5: rmdup on data 3 (as BED)** 👁 ✎ ✕

~1,000,000 regions  
format: **bed**, database: **hg19**

🗑 🗑 🗑 🗑

**Download** 🗑

[Click here to edit annotation](#)

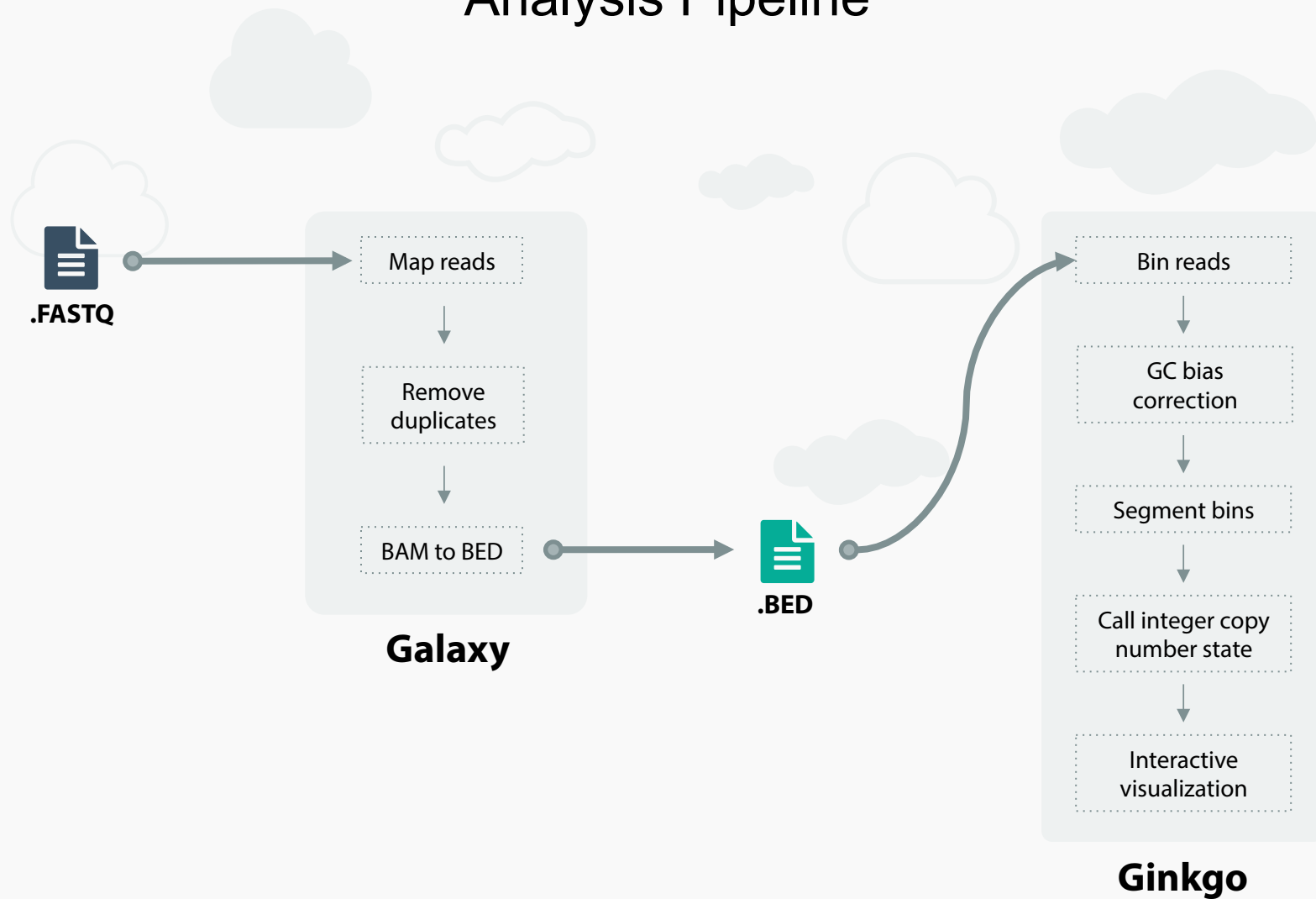
display in IGB [View](#)  
display at Ensembl [Current](#)  
display at RViewer [main](#)  
display at UCSC [main](#)

1. Chrom	2. Start	3. End	4. Name
chr10	61988	62036	SRR054622.2038687
chr10	69093	69141	SRR054622.2024583
chr10	69098	69146	SRR054622.481587
chr10	69099	69147	SRR054622.739359
chr10	72859	72907	SRR054622.1184885
chr10	92035	92083	SRR054622.2364537

**1: EBI SRA: SRR054622 File:** <ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR054/SRR054622/SRR054622.fastq.gz> 👁 ✎ ✕

[https://usegalaxy.org/datasets/bbd44e69cb8906b514ab3fface8a1d1f/display?to\\_ex...](https://usegalaxy.org/datasets/bbd44e69cb8906b514ab3fface8a1d1f/display?to_ex...)

# Analysis Pipeline



## Hands-on with Ginkgo



**Ginkgo:** [qb.cshl.edu/ginkgo](http://qb.cshl.edu/ginkgo)

**Sample Data:** [qb.cshl.edu/ginkgo/vizbi2015.tar](http://qb.cshl.edu/ginkgo/vizbi2015.tar)