



Galaxy Introduction

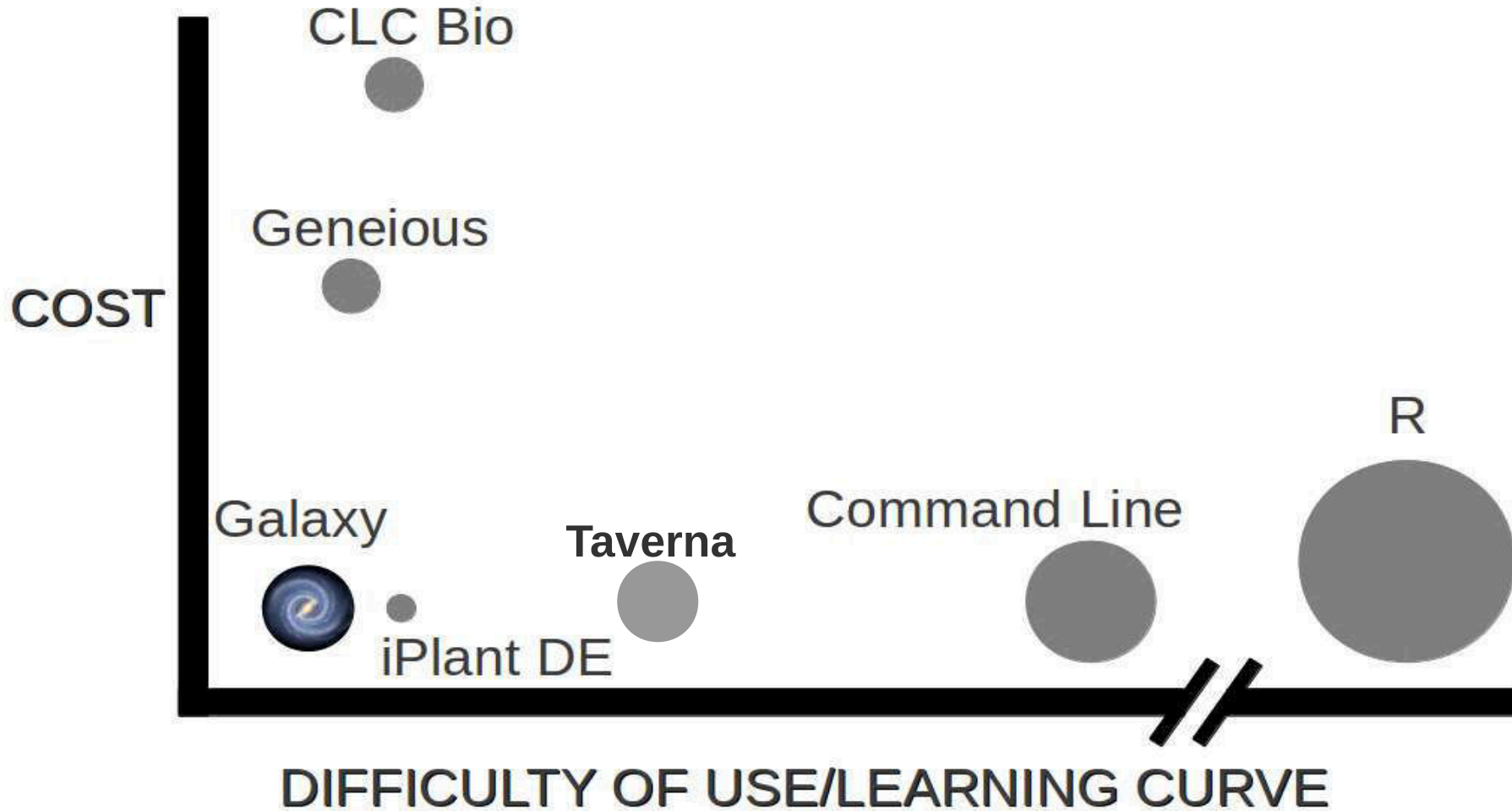
Leon Mei



Acknowledgement



Galaxy



Size of dot indicates flexibility/power

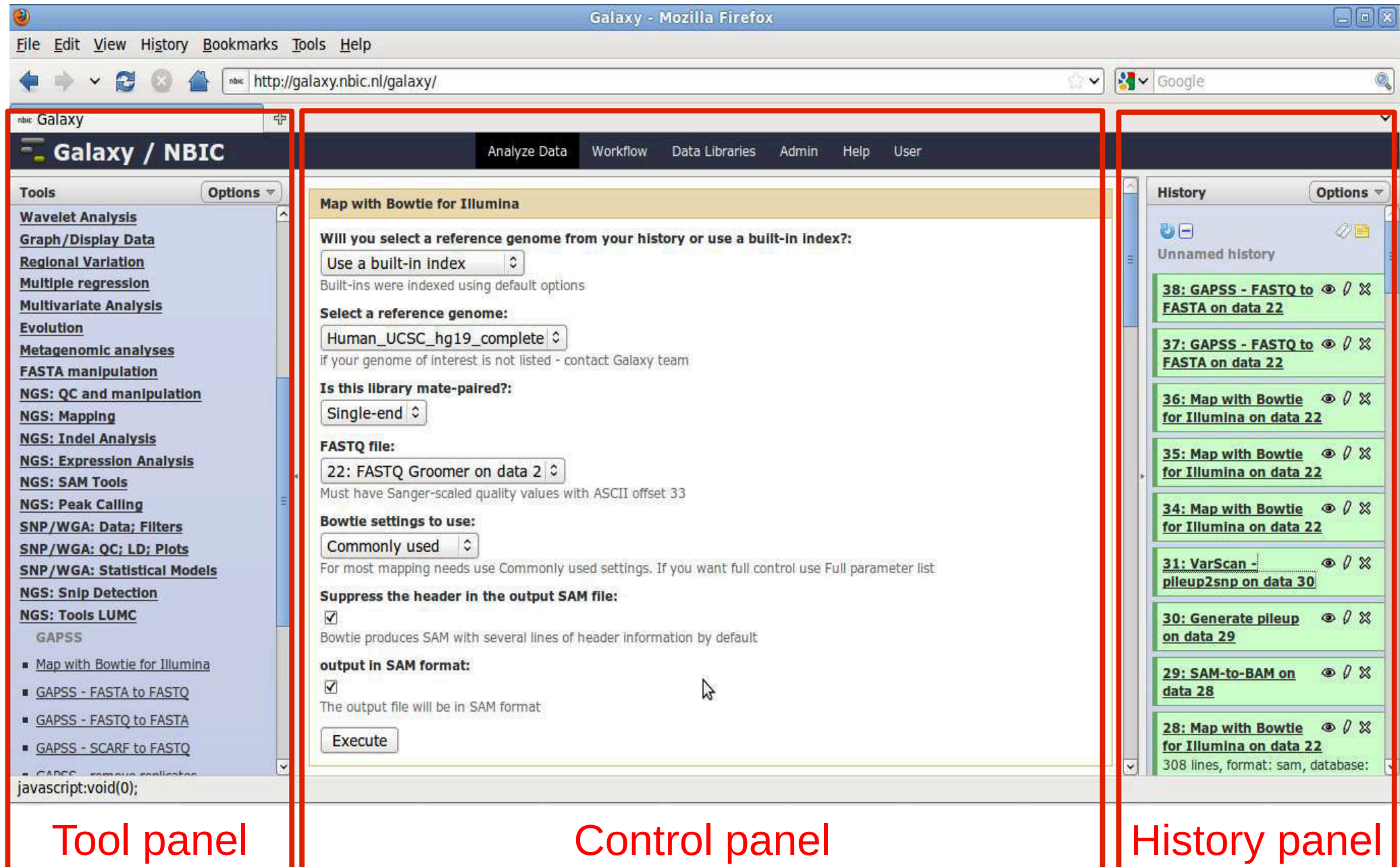
Advantage

- Open source and active community
 - [Galaxy community conference starting from 2010](#)
- Web-based client
- History tracking for reproducibility
- Most of NGS tools available and easy tool/data installation
- Use of external resources (UCSC, biomart)
- Support heterogeneous resource
- Support versioning

(Current) Limitations

- No user specific tool panel
- Can not check a job's progress
 - `stdout`, `stderr`
- Accounting of user CPU hours
- Tools are installed centrally by admins
- Only support linear and relatively simple workflow

User interface



The screenshot shows the Galaxy web interface in Mozilla Firefox. The browser address bar displays `http://galaxy.nbic.nl/galaxy/`. The interface is divided into three main panels:

- Tool panel (left):** Contains a list of tools under the heading "Tools / NBIC". The "GAPSS" category is expanded, showing tools like "Map with Bowtie for Illumina", "GAPSS - FASTA to FASTQ", "GAPSS - FASTQ to FASTA", and "GAPSS - SCARF to FASTQ".
- Control panel (center):** Displays the configuration for the "Map with Bowtie for Illumina" tool. It includes fields for selecting a reference genome (set to "Human_UCSC_hg19_complete"), a FASTQ file (set to "22: FASTQ Groomer on data 2"), and various settings like "Bowtie settings to use" (set to "Commonly used") and "Suppress the header in the output SAM file" (checked). An "Execute" button is at the bottom.
- History panel (right):** Shows a list of recent jobs. The top job is "38: GAPSS - FASTQ to FASTA on data 22". Other jobs include "37: GAPSS - FASTQ to FASTA on data 22", "36: Map with Bowtie for Illumina on data 22", "35: Map with Bowtie for Illumina on data 22", "34: Map with Bowtie for Illumina on data 22", "31: VarScan - pileup2snp on data 30", "30: Generate pileup on data 29", "29: SAM-to-BAM on data 28", and "28: Map with Bowtie for Illumina on data 22".

Tool panel

Control panel

History panel

Upload input files

Galaxy / NBIC

galaxy.nbic.nl

Galaxy / NBIC

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 68%

Tools

search tools

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- BX table browser
- EBI SRA ENA SRA
- Get Microbial Data
- BioMart Central server
- BioMart Test 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

Upload File (version 1.1.3)

File Format: Auto-detect

Which format? See help below

File: 031_withoutControl.recode.LOH.homo.vcf

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

Local disk/
External URL

Files uploaded via FTP:

File	Size	Date
<input type="checkbox"/> neo4j_workshopdata.tillWPimport.tgz	3.3 GB	08/19/2013 08:05:13 AM
<input type="checkbox"/> L3230WG_1.bam	24.1 GB	08/16/2013 11:12:43 AM
<input type="checkbox"/> allbio.txt	576 bytes	08/19/2013 07:52:24 AM

FTP

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at galaxy.nbic.nl using your Galaxy credentials (email address and password).

Convert spaces to tabs: Yes

Use this option if you are entering intervals by hand.

Genome: Human Feb. 2009 (GRCh37/hg19) (hg19)

History

imported: Ele_Test
6.7 GB

35: 031_withoutControl.recode.vcf

34: Galaxy40-[Tophat2 on ctrl accepted hits].bam (Genome Coverage BedGraph)

33: Galaxy45-[Tophat2 on treat accepted hits].bam (Genome Coverage BedGraph)

30: Galaxy40-[Tophat2 on ctrl accepted hits].bam

29: Galaxy45-[Tophat2 on treat accepted hits].bam

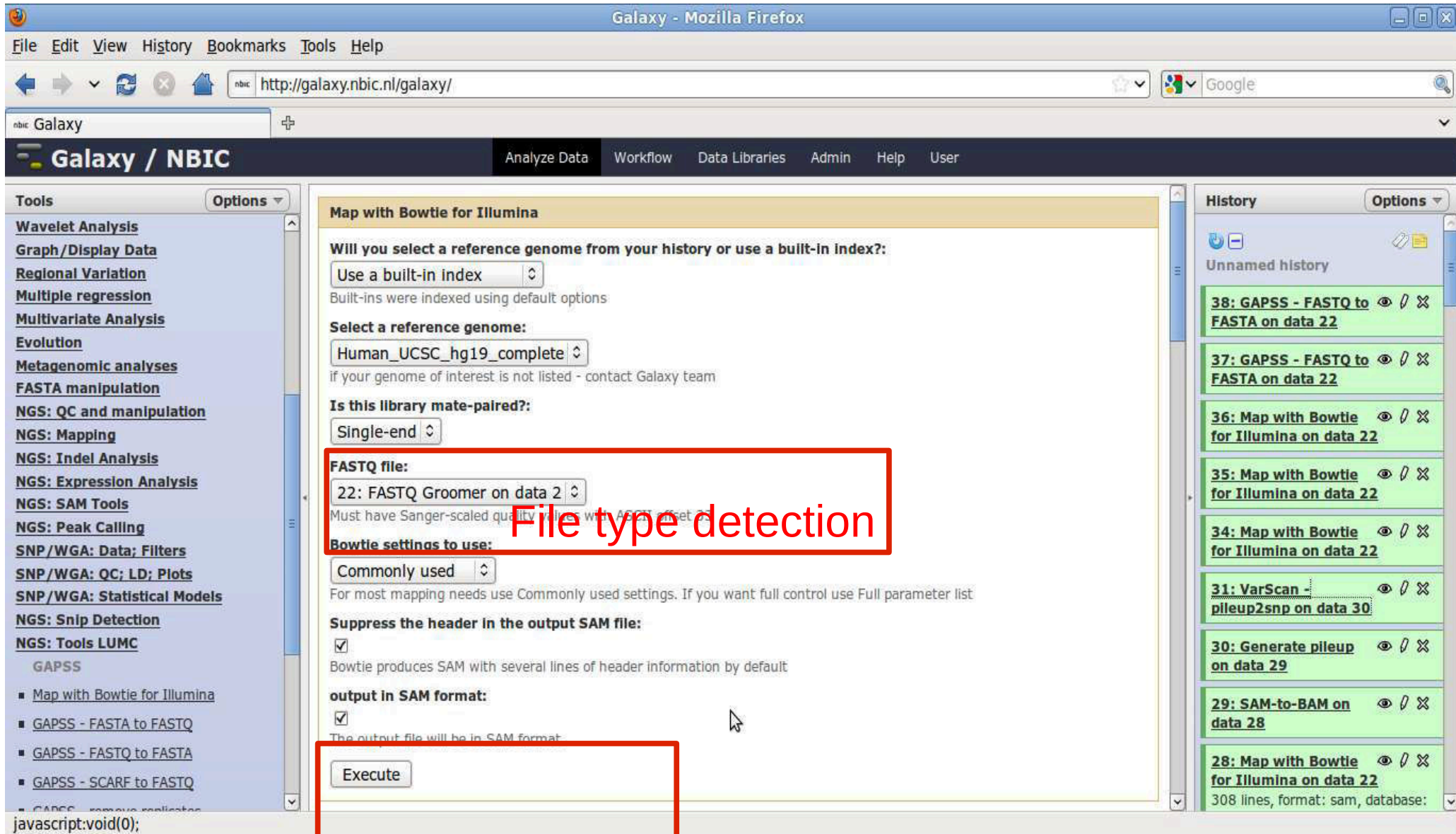
28: Map with BWA for Illumina on data 27: mapped reads

27: FASTQ Groomer on data 4

26: Wiggle file from SAM-to-BAM on data 11: converted BAM

23: Generate pileup on data 21: converted pileup

Execute a task



Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://galaxy.nbic.nl/galaxy/

Galaxy / NBIC

Analyze Data Workflow Data Libraries Admin Help User

Tools Options

- Wavelet Analysis
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- FASTA manipulation
- NGS: QC and manipulation
- NGS: Mapping
- NGS: Indel Analysis
- NGS: Expression Analysis
- NGS: SAM Tools
- NGS: Peak Calling
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- NGS: Snip Detection
- NGS: Tools LUMC
 - GAPSS
 - Map with Bowtie for Illumina
 - GAPSS - FASTA to FASTQ
 - GAPSS - FASTQ to FASTA
 - GAPSS - SCARF to FASTQ
 - GAPSS - remove replicates

Map with Bowtie for Illumina

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

Built-ins were indexed using default options

Select a reference genome:
 Human_UCSC_hg19_complete

If your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?:
 Single-end

FASTQ file:
 22: FASTQ Groomer on data 2

Must have Sanger-scaled quality values with ASCII offset 27

Bowtie settings to use:
 Commonly used

For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:

Bowtie produces SAM with several lines of header information by default

output in SAM format:

The output file will be in SAM format

Execute

History Options

Unnamed history

- 38: GAPSS - FASTQ to FASTA on data 22
- 37: GAPSS - FASTQ to FASTA on data 22
- 36: Map with Bowtie for Illumina on data 22
- 35: Map with Bowtie for Illumina on data 22
- 34: Map with Bowtie for Illumina on data 22
- 31: VarScan pileup2snp on data 30
- 30: Generate pileup on data 29
- 29: SAM-to-BAM on data 28
- 28: Map with Bowtie for Illumina on data 22

308 lines, format: sam, database:

Execute button

Run and View Results

Galaxy / NBIC

galaxy.nbic.nl

Galaxy / NBIC Analyze Data Workflow Shared Data Visualization Admin Help User Using 68%

Tools

search tools

Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Wavelet Analysis
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Motif Tools
Multiple Alignments
Metagenomic analyses
FASTA manipulation
NGS: QC and manipulation
NGS: Assembly

tracking_id	class_code	nearest_ref_id	gene_id	gene_short_name	tss_id
LOC729737	-	-	LOC729737	LOC729737	TSS18150
FAM138A	-	-	FAM138A	FAM138A	TSS8229
OR4F5	-	-	OR4F5	OR4F5	TSS14112
DDX11L1	-	-	DDX11L1	DDX11L1	TSS14523
WASH7P	-	-	WASH7P	WASH7P	TSS7359
OR4F16	-	-	OR4F16	OR4F16	TSS12406
LOC100133331	-	-	LOC100133331	LOC100133331	TSS12776
OR4F3	-	-	OR4F3	OR4F3	TSS4836
LOC100132287	-	-	LOC100132287	LOC100132287	TSS12041
LOC100133331	-	-	LOC100133331	LOC100133331	TSS12041
LOC100288069	-	-	LOC100288069	LOC100288069	TSS8527
LOC643837	-	-	LOC643837	LOC643837	TSS10020
LINC00115	-	-	LINC00115	LINC00115	TSS17932
FAM41C	-	-	FAM41C	FAM41C	TSS20399
LOC100130417	-	-	LOC100130417	LOC100130417	TSS19778
KLHL17	-	-	KLHL17	KLHL17	TSS17211
PLEKHN1	-	-	PLEKHN1	PLEKHN1	TSS11813
SAMD11	-	-	SAMD11	SAMD11	TSS27721
NOC2L	-	-	NOC2L	NOC2L	TSS17475
C1orf170	-	-	C1orf170	C1orf170	TSS8429
HES4	-	-	HES4	HES4	TSS26535
ISG15	-	-	ISG15	ISG15	TSS16010
AGRN	-	-	AGRN	AGRN	TSS19857
RNF223	-	-	RNF223	RNF223	TSS15689
C1orf159	-	-	C1orf159	C1orf159	TSS13942
LOC254099	-	-	LOC254099	LOC254099	TSS28723
MIR200B	-	-	MIR200B	MIR200B	TSS383
MIR200A	-	-	MIR200A	MIR200A	TSS9879
MIR429	-	-	MIR429	MIR429	TSS19367
TLL10	-	-	TLL10	TLL10	TSS23566,TSS5031

History

21: SAM-to-BAM on data 11: converted BAM

15: Cufflinks on data 12 and data 2: assembled transcripts

14: Cufflinks on data 12 and data 2: transcript expression

13: Cufflinks on data 12 and data 2: gene expression **View data**

format: tabular, database: hg19
cufflinks v2.1.1 cufflinks -q
--no-update-check -l 300000 -F
0.100000 -j 0.150000 -p 4 -G
/mnt/galaxyData/database/files
/028/dataset_28940.dat

1	2	3
tracking_id	class_code	nearest_ref_id
LOC729737	-	-
FAM138A	-	-
OR4F5	-	-
DDX11L1	-	-
WASH7P	-	-

12: SAM-to-BAM on data 11: converted BAM

View result

Galaxy workflows

- 3 ways to create Galaxy workflow
 - Extract from your own history
 - Use workflow design panel
 - Shared by other users in Galaxy (local or remote)
- Scalability
 - Parallelism within one workflow
 - API available for automation

Extract a workflow from history

The screenshot shows the Galaxy web interface for CTMM-Trait Galaxy. The main content area displays a list of tools used in the current history, with a 'Create Workflow' button highlighted in red and the word 'create' written in red next to it. The 'History' panel on the right shows a list of history items, with the 'Extract Workflow' option highlighted in red and the word 'extract' written in red next to it.

The interface includes a search bar for tools, a list of tool categories (e.g., TRAIT TOOLSHED TOOLS, TRAIT TEST TOOLS), and a central area for creating a workflow. The workflow name is 'Workflow constructed from history 'RNAseq alignment''. The 'Create Workflow' button is highlighted in red, and the word 'create' is written in red next to it.

The 'History' panel on the right shows a list of history items, with the 'Extract Workflow' option highlighted in red, and the word 'extract' is written in red next to it.

Drag & drop design panel

Galaxy / CTMM-Trait... x

galaxy-sandbox.trait-ctmm.cloudlet.sara.nl/workflow/editor?id=ff5476bcf6c921fa

Galaxy / CTMM-Trait Sandbox Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 0%

Tools

search tools

Get Data

CGTAG

CGAtools

Annotation

Visualisation

File Manipulation

CONFERO

Confero Platform

OTHER

RNA-Seq

Filter and Sort

NGS: RNA Analysis

VCF Tools

saskia_testtools

BioAssist Demo Tools

bioassistedemo

bas

File Concatenation concatenates 2 files

rwwh

Tools by Freek

ruslan's cat tool

java_try

More Tools by Freek

thang_example

iPER test

RSeQC

stef_test

VCFlib tools

Workflow control

Inputs

Workflow Canvas | ECCB2014

Input dataset x

output

FastQC:Read QC x

Short read data from your current history

Contaminant list

html_file (html)

Sickle x

Paired-End Forward Strand FastQ Reads

Paired-End Reverse Strand FastQ Reads

output_single

output_paired1 (fastq, fastqsanger, fastqillumina, fastqsolexa)

output_paired2 (fastq, fastqsanger, fastqillumina, fastqsolexa)

output_paired_single (fastq, fastqsanger, fastqillumina, fastqsolexa)

FASTQ Groomer x

File to groom

output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)

FASTQ Groomer x

File to groom

output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)

Input dataset x

output

FastQC:Read QC x

Short read data from your current history

Contaminant list

html_file (html)

Sickle x

Paired-End Forward Strand FastQ Reads

Paired-End Reverse Strand FastQ Reads

output_single

output_paired1 (fastq, fastqsanger, fastqillumina, fastqsolexa)

output_paired2 (fastq, fastqsanger, fastqillumina, fastqsolexa)

FASTQ Groomer x

File to groom

output_file (fastqsanger, fastqcssanger, fastqsolexa, fastqillumina)

Details

Edit Workflow Attributes

Name:
ECCB2014

Tags:

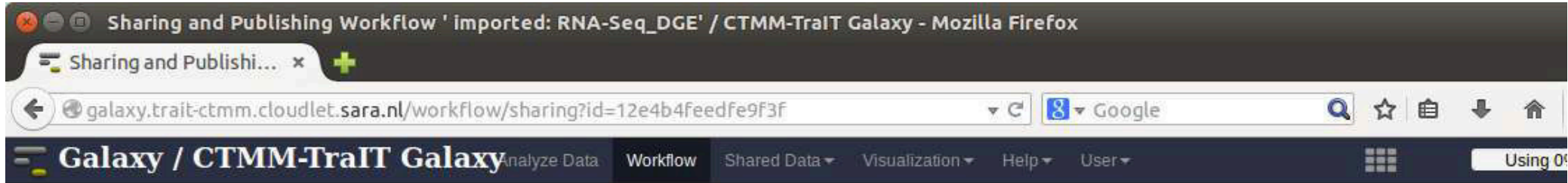
Apply tags to make it easy to search for and find items with the same tag.

Annotation / Notes:

Describe or add notes to workflow

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

Share & Publish



Share or Publish Workflow 'imported: RNA-Seq_DGE'

Make Workflow Accessible via Link and Publish It

This workflow is currently restricted so that only you and the users listed below can access it. You can:

[Make Workflow Accessible via Link](#)

Generates a web link that you can share with other people so that they can view and import the workflow.

[Make Workflow Accessible and Publish](#)

Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's [Published Workflows](#) section, where it is publicly listed and searchable.

Share Workflow with Individual Users

You have not shared this workflow with any users.

[Share with a user](#)

Multiple levels of sharing

[Back to Workflows List](#)

Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement

SERGEI KOSAKOVSKY POND^{1,2,*}, SAMIR WADHAWAN^{3,6*}, FRANCESCA CHIAROMONTE⁴, GURUPRASAD ANANDA^{1,3}, WEN-YU CHUNG^{1,3,7}, JAMES TAYLOR^{1,5}, ANTON NEKRUTENKO^{1,3} and THE GALAXY TEAM^{1*}

Correspondence should be addressed to [SKP](#), [JT](#), or [AN](#).

How to use this document

This document is a live copy of supplementary materials for [the manuscript](#). It provides access to the **exact** analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them; click on the green plus to import an item. To import workflows you must [create a Galaxy account](#) (unless you already have one) – a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN:

[+](#) **Galaxy History | Galaxy vs MEGAN** [+](#) [↗](#)
Comparison of Galaxy vs. MEGAN pipeline.

This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3A**):

[+](#) **Galaxy History | metagenomic analysis** [+](#) [↗](#)

This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3B**):

[+](#) **Galaxy Workflow | metagenomic analysis** [+](#) [↗](#)

About this Page

Author

aun1



Related Pages

[All published pages](#)

[Published pages by aun1](#)

Rating

Community
(6 ratings, 5.0 average)



Yours



Tags

Community:

[paper](#)

[galaxy](#)

[megan](#)

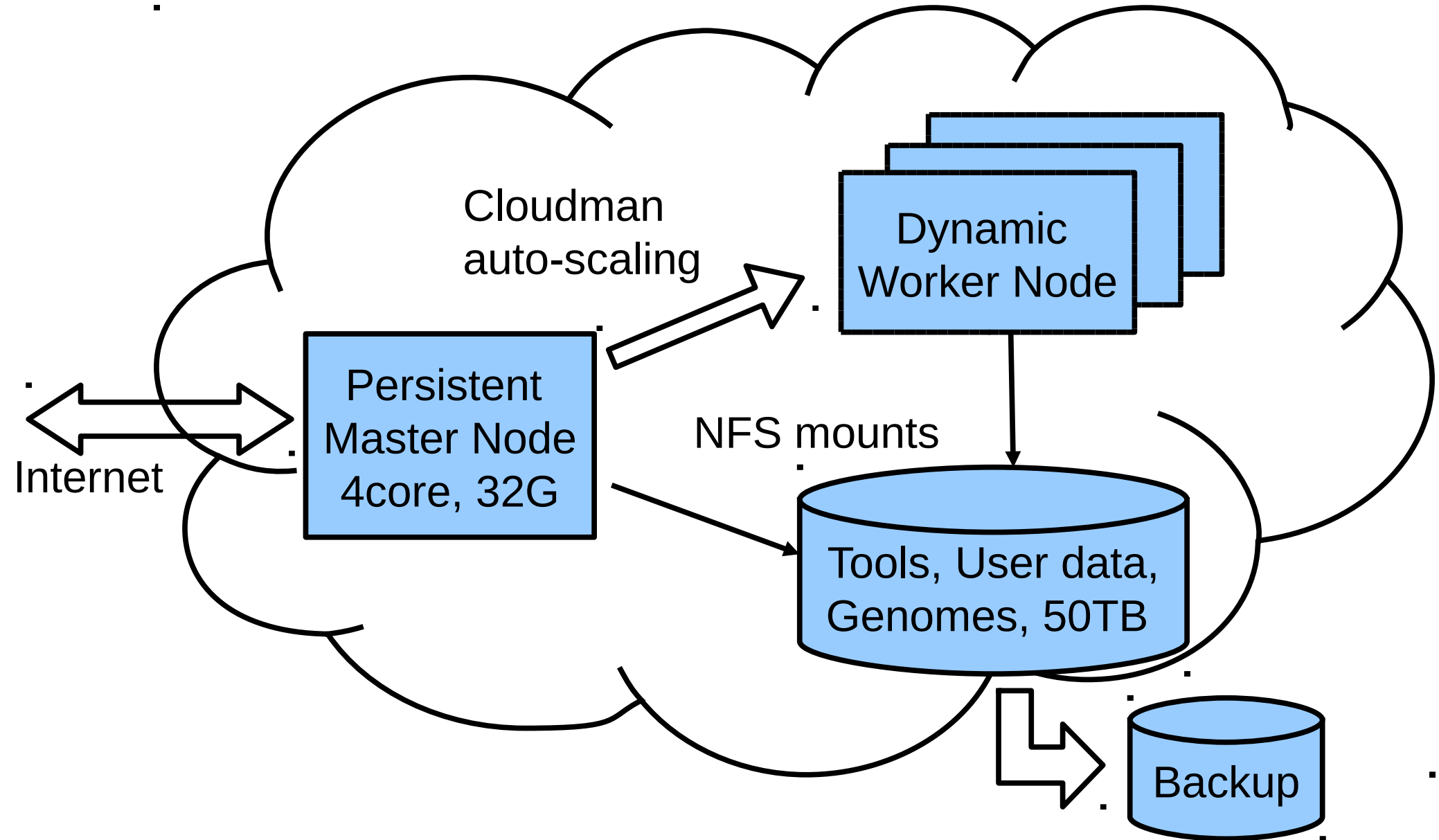
Yours:



More about computing

- Galaxy can run on a local server, cluster, cloud
- Full DRMAA support
- Support of heterogeneous resource
 - Define run requirements in `job_conf.xml`

Galaxy Cloudman



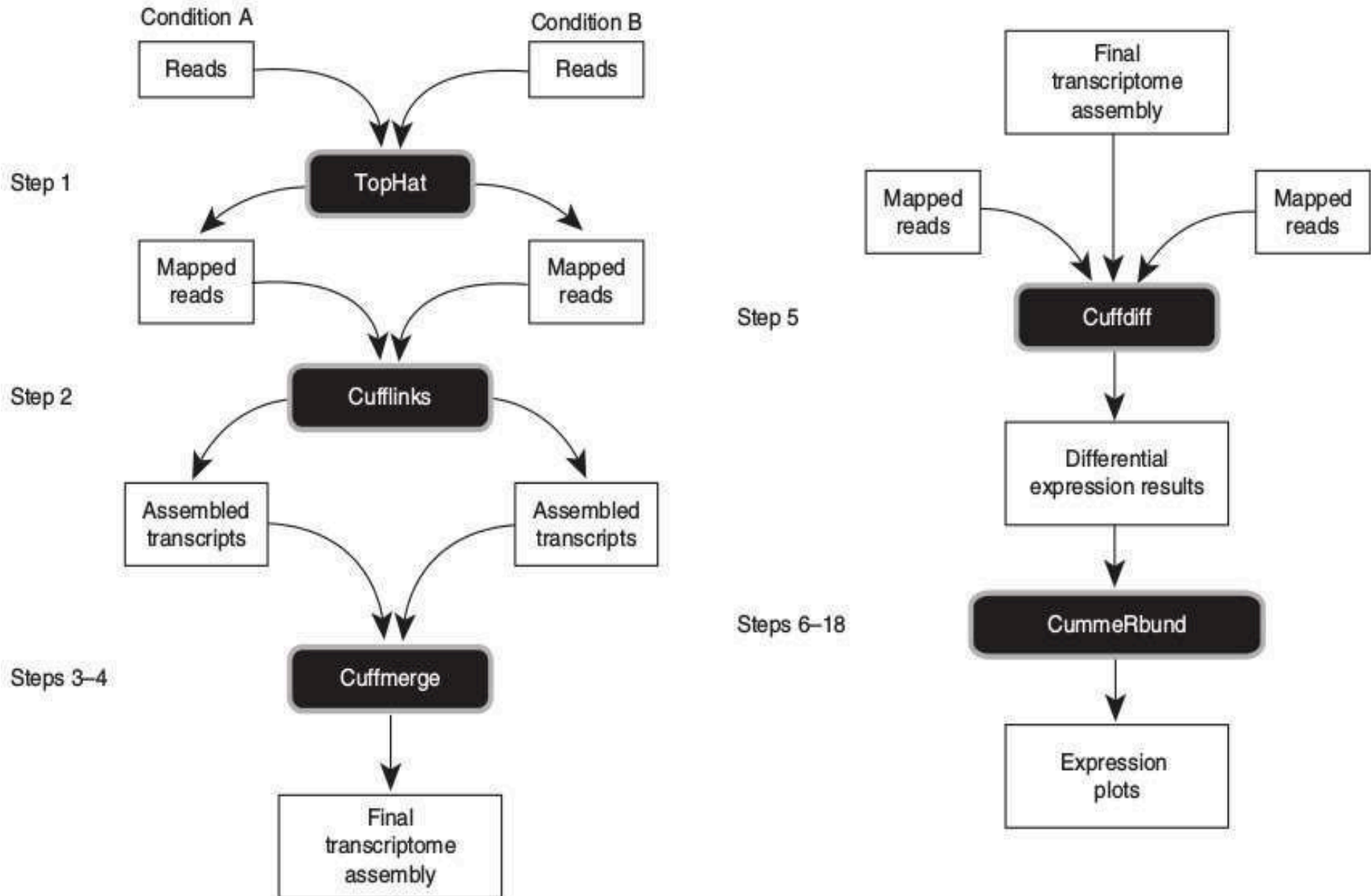
RNAseq Workflow

Leon Mei

General RNAseq pipeline

- Input data
 - Sequencing files (fastq), gene annotation (gtf)
- Pre-alignment
 - QC and Data cleaning
- Alignment
 - Use a specialized (RNA) aligner
- Expression (gene, transcripts) analysis
 - Quantification, up/down regulated genes, p-values
- Transcript assembly
 - New transcripts, alternative splicing
- Pathway analysis
 - Clustering algorithms, Taverna, etc.

Tuxedo pipeline



Scaling up

- Computational requirement of RNAseq analysis
 - Aligning 2-10M pairs per hour, ~1 day per sample
- Needs from large sequencing projects
 - <http://www.geuvadis.org>, 465 RNAseq samples
 - BBMRI-NL BIOS, 4000 RNAseq samples
- Running all samples in parallel (several times)
 - Galaxy is not best place to scale up
 - Dutch life science grid is used for BIOS

Hands-on

- <http://tinyurl.com/eccbgalaxy>