# NGS Analysis Using Galaxy

- Sequences and Alignment Format
- Galaxy overview and Interface
- Getting Data in Galaxy
- Analyzing Data in Galaxy
  - Quality Control
  - Mapping Data
- History and workflow
- Galaxy Exercises

# UCR Galaxy homepage (https://galaxy.bioinfo.ucr.edu)

### 🗧 Galaxy

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
<u>Statistics</u>
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: Mapping
NGS: Indel Analysis
NGS: RNA Analysis
<

Institute for Integrative
Genome Biology

### Welcome to IIGB's Galaxy Server!

### Overview

Galaxy is an open, highly customizable, web-based platform for the analysis of next generation sequence data and many other biological data types. It enables users to run computationally demanding next generation sequencing analysis tasks on powerful server hardware from a graphical web browser-based user interface rather than the Linux command-line. A subset of of application supported by Galaxy is given in the left pane. Much more detailed descriptions of Galaxy's basic functionalities including user tutorials are available <u>here</u>.

### Why Local Galaxy Service?

There are many advantages of using a local Galaxy server here at UCR rather than public test instances of Galaxy available on the internet. The most important are: (1) shorter waiting queues for analysis tasks; (2) elimination of time consuming uploads of large data sets; (3) support for analyzing much larger data sets than this is possible on public services; (4) the ability to customize software tools and database collections.

### How to Gain Access?

This instance of Galaxy runs on IIGB's high performance compute (HPC) infrastructure, called Biocluster. As such its usage is covered by the annual registration fee for this infrastructure (see <u>here</u> for details). Users with an active Biocluster account can access this Galaxy service using their existing user name and password without any extra cost. New account requests for this service can be sent to <u>support@biocluster.ucr.edu</u>.

### Additional Databases and Sofware Tools

Support requests for including additional reference genomes and software tools on IIGB's Galaxy server can be sent to <a href="mailto:support@biocluster.ucr.edu">support@biocluster.ucr.edu</a>

### Workshops on Galaxy

Past and future UCR workshop events on using Galaxy are listed <u>here</u>. The user manual from previous workshops can be accessed <u>here</u>.

Enter IIGB's Galaxy Service To enter this service, click here.

# SNP-seq Analysis dataset

- Data source: SRR038850 sample from experiment published by Kaufman *et al* (2012, GSE20176)
  - Understand the target molecular mechanisms underlying AP1 function
- Fastq File:

http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/SRR038850.fastq

• TAIR10 Genome:

http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/tair10chr.fasta

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20176

# SNP-seq pipeline



# Upload data

Go to "Get Data", click open "Upload File from your computer". Then specify the following list of URLs in URL/Text box

http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/SRR038850.fastq http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/tair10chr.fasta

💳 Galaxy	Analyze Data Workflow Shared Data - Visualization - Help - User -			Using 0%
Tools	Upload File (version 1.1.3)	н	listory	C 0
search tools <u>Get Data</u> <u>Upload File</u> from your computer <u>UCSC Main</u> table browser <u>UCSC Test</u> table browser <u>UCSC Archaea</u> table browser <u>BX</u> table browser <u>EBI SRA ENA SRA</u> <u>Get Microbial Data</u>	Opload File (Version 1.1.3)         File Format:         Auto-detect         Which format? See help below         File:         Browse         No file selected.         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 (iby the site administrator).         URL/Text:         http://biocluster.ucr.edu/~rkaundal         /Galaxy_workshop/Snpseq/SRR038850.fastq         http://biocluster.ucr.edu/~rkaundal         /Galaxy_workshop/Snpseq/tair10chr.fasta	(if enabled	Innamed history bytes Your history is em Data' on the left p	Phys. Click 'Get bane to start
<u>BioMart</u> Test server <u>CBI Rice Mart</u> rice mart <u>GrameneMart</u> Central server	Files uploaded via FTP:       File     Size			
modENCODE fly server Flymine server Flymine test server modENCODE modMine server MouseMine server Ratmine server YeastMine server metabolicMine server modENCODE worm server WormBase server	Your FTP upload directory contains no files.         To upload files greater than 2GB in size, it is recommended that you upload your files on any webserver as biocluster and then paste the URL to your file in the URL/Text box. You may use filezilla and transfer the file to your .html directory in your biocluster account. The url of your file will be (http://biocluster.ucr.edu/~username/filename)         Convert spaces to tabs:			
Wormbase test server	Auto-detect  The system will attempt to detect Axt, Fasta, Fastosolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) formats.	s. If your file		>

# Convert Fastq file to sanger format

- Select NGS: QC and manipulation and Fastq groomer
- The FASTQ Groomer tool is used to verify and convert between the known FASTQ variants.
- After grooming, the user is presented with a valid FASTQ format that is accepted by all downstream analysis tools.

🗧 Galaxy	Analyze Data Workflow Shared Data - Visualization - Help - User -		Using 0%
Tools	FASTQ Groomer (version 1.0.4)	History	C ¢
Graph /Display Data	File to groop	Unnamed history	
Regional Variation	1: http://bioluster.ucr.edu/~rkaundal/Galaxy_workshop/Snpseq/SRR038850.fastq	182.4 MB	0 🖻
Multiple regression	Input FASTO quality scores type:		
Multivariate Analysis	Saner 0	<u>2:</u>	
Evolution	Advanced Options	/Galaxy workshop/	/Snpseq/tair10ch
Motif Tools	Hild Avanced Options	<u>r.fasta</u>	
Multiple Alignments		1.	@ // %
Metagenomic analyses	Execute	http://biocluster.uc	r.edu/~rkaundal
FASTA manipulation		/Galaxy workshop/	/Snpseq/SRR0388
NGS: QC and manipulation	What it does	50.fastq	
FastQC: Comprehensive QC	This to do a second conversion enters solution to the CACTO format		
reporting for short read sequence	This tool offers several conversions options relating to the FASTQ format.		
Barcode Splitter (ngs-tools)	When using <i>Basic</i> options, the output will be <i>sanger</i> formatted or <i>cssanger</i> formatted (when the input is Color Space Sanger).		
Clip adapter sequences	When converting, if a quality score fails outside of the target score range, it will be coerced to the closest available value (i.e. the minimum or maximum).		
ILLUMINA FASTQ	When converting between Solexa and the other formats, quality scores are mapped between Solexa and PHRED scales using the equations found in <u>Cock PJ</u> , Eiclds CI Corto N Henry ML Discore DM The Sanora FASTO file format for sequences with quality scores and the Solex/Illumina FASTO variants. Nucleic		
FASTQ Groomer convert between	Acids Res 2009 Dec 16.		
various FASTQ quality formats	When converting between color space (csSanger) and base/sequence space (Sanger, Illumina, Solexa) formats, adapter bases are lost or gained; if gained,		
FASTQ splitter on joined paired	the base 'G' is used as the adapter. You cannot convert a color space read to base space if there is no adapter present in the color space sequence. Any		
end reads	masked or ambiguous nucleotides in base space will be converted to 'N's when determining color space encoding.		
FASTQ joiner on paired end reads			
FASTQ Summary Statistics by	Quality Score Comparison		
column	555555555555555555555555555555555555555		
ROCHE-454 DATA			
Build base quality distribution			
Select high quality segments			
Combine FASTA and QUAL into	33 59 64 73 104 126		
FASTQ	S - Sanger Phred+33, 93 values (0, 93) (0 to 60 expected in raw reads)		
AB-SOLID DATA	I - Illumina 1.3 Phred+64, 62 values (0, 62) (0 to 40 expected in raw reads)		
Convert SOLiD output to faste	X - Solexa Solexa+64, 67 values (-5, 62) (-5 to 40 expected in raw reads)		
Compute quality statistics for	Diagram adapted from http://en.wikipedia.org/wiki/FASTQ_format		
SOLID data	G Output from Illumina 1.8+ pipelines are Sanger encoded.		

# Features available in history panel

- View, Edit, Delete file
- Size of the file
- Save the file
- Repeat the analysis

• / × 1: http://biocluster.ucr.edu/~rkaundal /Galaxy\_workshop/Snpseg/SRR0388 50.fastq 66.8 MB format: fastq, database: ? uploaded fastq file 🔚 🛈 🕑 (1) 📑 @SRR038850.12 HWI-EAS038:3:1:2:1948 ld CAAGCATCTTTTTTGAATTTCCCATTTATCCGTTTA +SRR038850.12 HWI-EAS038:3:1:2:1948 1€ @?@<@?@BBAAB@>>===?@AB?7=<:6>@A@:?:6 @SRR038850.16 HWI-EAS038:3:1:2:1261 1c AGTAGGAGCTTAGGCTCCCAAAGGCACGTGTCGCTG

# **FASTQ Summary Statistics**

• To understand the quality properties of the reads, one can run the FASTQ Summary Statistics tool from NGS: QC and manipulation.

💳 Galaxy	Analyze Data Workflow Shared Data - Visualization - Help - User -	Using 0%
Tools	FASTQ Summary Statistics (version 1.0.0)	History C O
wavelet Analysis		Unnamed history
Graph/Display Data	FASTQ File:	
Regional Variation	3: rASTQ Groomer on data 1	249.2 MB
Multiple regression	Europe -	3: FASTQ Groomer on data 💿 🖉 💥
Multivariate Analysis	Execute	1
Evolution		
Motif Tools	This tool creates summary statistics on a FASTQ file.	2: O / X
Multiple Alignments	1 TIP: This statistics report can be used as input for the <b>Boxplot</b> tools.	/Galaxy workshop/Snpseg/tair10ch
Metagenomic analyses		r.fasta
FASTA manipulation	C	- 0.00
NGS: QC and manipulation	The output file will contain the following fields:	1: O U X
FastQC: Comprehensive QC	column = column number (1 to 36 for a 36-cycles read Solexa file)	/Galaxy workshop/Snpseg/SRR0388
reporting for short read sequence	count = number of bases found in this column.	50.fastq
Barcode Splitter (ngs-tools)	min = Lowest quality score value found in this column. max = Highest quality score value found in this column.	
Clin adapter sequences	sum = Sum of quality score values for this column.	
<u>enp</u> dapter sequences	mean = Mean quality score value for this column.	
ILLUMINA FASTQ	Q1 = 1st quartile quality score.	
FASTQ Groomer convert between	med = Median quality score.	
various FASTQ quality formats	Q3 = 510 quality score.	
FASTQ splitter on joined paired	IW = 'Left-Whisker' value (for boxplotting).	
end reads	rW = 'Right-Whisker' value (for boxplotting).	
FASTO joiner on paired end reads	outliers = Scores falling beyond the left and right whiskers (comma separated list).	
	A_Count = Count of A nucleotides found in this column.	
Column	G Count = Count of G' nucleotides found in this column.	
country	T_Count = Count of 'T' nucleotides found in this column.	
ROCHE-454 DATA	N_Count = Count of 'N' nucleotides found in this column.	
Build base quality distribution	Other_Nucs = Comma separated list of other nucleotides found in this column.	
Select high quality segments	For example:	
Combine FASTA and OUAL inte		
EASTO	#count count min max sum mean 01 med 03 IQR 1W rW outliers A Count C Count C Count N Count other Dases ( 1 1433355 2 33 45606057 31 4306291875 32 0 33 0 31 0 13 33 2 4 5 6 7 9 10 11 21 31 44 15 16 17	
	2 14336356 2 34 441135033 30.7703737965 30.0 33.0 33.0 3.0 26 34 2,4,5,6,7.8,9,10,11,12,13,14,15,16,17.	
AB-SOLID DATA	3 14336356 2 34 433659182 30.2489127642 29.0 32.0 33.0 4.0 23 34 2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,	
Convert SOLiD output to fastq	4 14336356 2 34 433635331 30.2472490917 29.0 32.0 33.0 4.0 23 34 2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,	
	5 14336356 2 34 432498583 30.167957813 29.0 32.0 33.0 4.0 23 34 2,4,5,6,7,8,9,10,11,12,13,14,15,16,17,	

# FASTQ Quality control

• To understand the quality properties of the reads, one can also run the FASTQC: Read QC reports from NGS: QC and manipulation.

- Galaxy	Analyze Data Workflow Shared Data + Visualization + Help + User +	2 2 2 8 2 2 8 9 9	Using 0%
Tools	FastQC: Comprehensive QC version 0.53 0	History	C 0
search tools	Short read data from your current history: 3: FASTQ Groomer on data 1	Workshop_test 249.3 MB	47 🖻
Get Data Send Data ENCODE Tools	Title for the output file - to remind you what the job was for: FastQC Letters and numbers only please - other characters will be removed	4: FASTQ Summary Statistics on data 3	• 0 %
Text Manipulation	Contaminant list:	3: FASTQ Groomer o	<u>n data</u> 👁 🖉 💥
Join, Subtract and Group	tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA	2: http://biocluster.ucr	● Ø ⊠ .edu/~rkaundal
Extract Features	Execute	/Galaxy workshop/! r.fasta	Snpseq/tair10ch
Fetch Sequences Fetch Alignments Get Genomic Scores Operate on Genomic Intervals	i) Purpose Quote from FastQC FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a modular set of analyses which you can use to give a quick impression of whether your data has any problems of which you should be aware before doing any further analysis.	1: http://biocluster.ucr /Galaxy_workshop/: 50.fastq	
Statistics Wavelet Analysis Graph/Display Data Regional Variation Multiple regression Multivariate Analysis	The main functions of FastQC are: Import of data from BAM, SAM or FastQ files (any variant) Providing a quick overview to tell you in which areas there may be problems Summary graphs and tables to quickly assess your data Export of results to an HTML based permanent report Offline operation to allow automated generation of reports without running the interactive application <u>FastQC</u> is the best place to look for documentation – it's very good. Some features of the Galaxy wrapper you are using are described below.		
Motif Tools Multiple Alignments Metagenomic analyses FASTA manipulation	1 This Galaxy Tool You are using <u>FastOC</u> in Galaxy. This is easy because it has been packaged into a Galaxy tool by the Intergalactic Utilities Commission. It exposes the external package <u>FastQC</u> which is documented at <u>FastQC</u> Kindly acknowledge it as well as this tool if you use it. FastQC incorporates the <u>Picard-tools</u> libraries for sam/bam processing. The contaminants file parameter was borrowed from the independently developed fastqcwrapper contributed to the Galaxy Community Tool Shed by Jim Johnson.		
NGS: QC and manipulation FastQC: Comprehensive QC reporting for short read sequence Barcode Splitter (ngs-tools)	1 Inputs and outputs This wrapper will accept a Galaxy fastq, sam or bam as the input read file to check. It will also take an optional file containing a list of contaminants information, in the form of a tab- delimited file with 2 columns, name and sequence.		
Clip adapter sequences	EastQC produces a single HTML output file which is slightly adjusted so it looks good in Galaxy that contains all of the results, including the following:		
ILLUMINA FASTQ FASTQ Groomer convert between various FASTQ quality formats	Per base sequence quality Per sequence quality scores Per base sequence content		
FASTQ splitter on joined paired end reads	Per pase GL content Per sequence GC content Per base N content		
FASTQ joiner on paired end reads FASTQ Summary Statistics by column	sequence Length Distribution Sequence Duplication Levels Overrepresented sequences Kmer Content		
ROCHE-454 DATA	All except Basic Statistics and Overrepresented sequences are plots.		

# Quality control output

- Galaxy		Analyze Data	Workflow	Shared Data +	Visualization <del>-</del>	Help <del>~</del> Use			Using 0%
Tools	FASTQ_Groomer_on_data_1 FastQC Repo	rt						History	0 0
( sourch tools	R Factor Banat							Workshop test	
search tools	Sat 6 Dec 2014							249 6 MB	02 📑
<u>Get Data</u>	FASTQ_Groomer_on_data_1							21510 110	~
Send Data								5: FastQC FASTQ Gro	oomer 👁 🛛 🛛
ENCODE Tools	Summary							on data 1.html	
<u>Lift-Over</u>								4: FASTQ Summary	• / %
Text Manipulation								Statistics on data 3	
Filter and Sort	• Basic Statistics							3. FASTO Groomer o	ndata @ // %
Join, Subtract and Group	• V Per base sequence quality							<u>1</u>	
Convert Formats									- 0.00
Extract Features	Per sequence quality scores							<u>Z:</u> http://biocluster.ucr	.edu/∼rkaundal
Fetch Alignments	• U Per base sequence content							/Galaxy workshop/S	Snpseg/tair10ch
Get Genomic Scores								<u>r.fasta</u>	
Operate on Genomic Intervals	• <u>Per base GC content</u>							1:	• / ×
Statistics	Per seguence GC content							http://biocluster.ucr	.edu/~rkaundal
Wavelet Analysis								/Galaxy workshop/S	Snpseq/SRR0388
Graph/Display Data	<ul> <li>Per base N content</li> </ul>							<u>SU.Tastq</u>	
Regional Variation									
Multiple regression	• Sequence Length Distribution								
Multivariate Analysis	<ul> <li>Sequence Duplication Levels</li> </ul>								
Evolution									
Motif Tools	• <u>Overrepresented sequences</u>								
Multiple Alignments	Kmer Content								
Metagenomic analyses									
FASTA manipulation	Pagia Statistics								
NGS: QC and manipulation	Dasic Statistics								
FastQC: Comprehensive QC	Maasura Valer								
reporting for short read sequence	Filenama EASTO Groomer on	oto 1							
Barcode Splitter (ngs-tools)	File type Conventional base calls	ala_1							
Clip adapter sequences	Encoding Sanger / Illumina 1 9								
ILLUMINA FASTQ	Total Sequences 386588								
FASTQ Groomer convert between	Filtered Sequences 0								
various FASTQ quality formats	Sequence length 36								
FASTQ splitter on joined paired end reads	%GC 42								

# Quality control reports







Per sequence GC content

# Quality filter

 This tool filters reads based on quality scores. NGS: QC and manipulation -> Generic FASTQ manipulation->Filter FASTQ reads by quality score and length



# Alignment with BWA

- BWA is a fast and accurate short read aligner that allows mismatches and indels (<u>Link</u>)
- Go to "NGS: Mapping" and click on "Map with BWA". (Settings)



# SAM to BAM format conversion

- Produce an indexed BAM file based on a sorted input SAM file.
- Go to "NGS: SAM Tools", then click open "SAM-to-BAM".

💳 Galaxy	Analyze Data Workflow Shared Data → Visualization → Help → User →		Using 0%
Tools	SAM-to-BAM (version 1.1.3)	History	C 🕈
<u>Graph/Display Data</u> <u>Regional Variation</u> Multiple regression	Choose the source for the reference list: History	Workshop_test 286.1 MB	47 🖻
Multivariate Analysis Evolution Motif Tools	SAM file to convert: 7: Map with BWA for Illumina on data 6 and data 2: mapped reads \$ Using reference file:	<u>7: Map with BWA for</u> Illumina on data 6 and mapped reads	● Ø X data 2:
Multiple Alignments	2: http://biocluster.ucr.edu/~rkaundal/Galaxy_workshop/Snpseq/ta 💲	6: Filter FASTQ on data	<u>3</u> • / ×
Metagenomic analyses FASTA manipulation NGS: OC and manipulation	Execute	5: FastQC_FASTQ Groo on data 1.html	<u>mer</u> ● Ø ⊗
NGS: Mapping NGS: Indel Analysis	What it does This tool uses the SAMTools toolkit to produce an indexed BAM file based on a sorted input	4: FASTQ Summary Statistics on data 3	• 0 %
NGS: RNA Analysis NGS: SAM Tools	SAM file.	3: FASTQ Groomer on a 1	<u>data</u> ●ℓX
<u>BCF Tools Cat</u> This tool allows the user to concatenate BCF files. <u>bcftools view</u> Converts BCF format to VCF format	<b>Citation</b> For the underlying tool, please cite <u>Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer</u> N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–	<u>2:</u> http://biocluster.ucr.ec /Galaxy_workshop/Snp r.fasta	
<u>BCF Tools Index</u> This tool allows the user to index sorted BCF for random access. MPileup SNP and indel caller	<u>9.</u>	1: http://biocluster.ucr.ec /Galaxy_workshop/Snj 50.fastg	
<u>SAM-to-BAM</u> converts SAM format to BAM format			
bcftools view Converts BCF format			

# Variants calling with Mpileup

- SNP and INDEL caller to Generate BCF (Binary Variant Format) for one or multiple BAM files
- Go to "NGS: SAM Tools", then click open "MPileup SNP and indel caller".

- Galaxy		Analyze Data	Workflow	Shared Data <del>-</del>	Visualization <del>+</del>	Help <del>+</del>	User <del>-</del>		Using 0%
Tools Convert Formats	MPileup (version 0.0.2							History	C 🗘
Extract Features	Choose the source fo	r the reference l	ist:					Workshop_test	
Fetch Sequences	History 🗘							289.5 MB	42 🖻
Fetch Alignments	BAM files							0. CAN	
Get Genomic Scores	BAM file 1							and data 7: converter	ata z 👁 🖉 🐹 d BAM
Operate on Genomic Intervals	BAN MC I								
Statistics	BAM TILE:	data 2 and data	7: convorto	DAM	•			7: Map with BWA for	
Wavelet Analysis	C. SAM to BAN OF	Gata 2 and Gata	7. converte					mapped reads	iu uata z.
Graph/Display Data	Add new BAM file								0.00
Regional Variation								<u>6: Filter FASTQ on da</u>	
Multiple regression	Using reference file:	an ada ta da sa da			/			5: FastQC_FASTQ_Gro	oomer 👁 🖉 🕱
Multivariate Analysis	2: http://biocluster.u	cr.edu/~rkaunda	I/Galaxy_wo	orkshop/Snpseq	tair10chr.fasta 🗧	J		on data 1.html	
Evolution	Genotype Likelihood	Computation:						4: FASTO Summary	@ / S?
Motif Tools	Perform genotype lik	elihood computat	ion	¢				Statistics on data 3	~ ~ ~
Multiple Alignments	Phred-scaled gap ext	ension sequenci	ng error pr	robability:					
Metagenomic analyses	20							3: FASTQ Groomer or	ndata 👁 🖉 💥
FASTA manipulation	Coefficient for model	ing homopolyme	er errors.:					-	
NGS: QC and manipulation	100	<b>3</b> , ., .,						<u>2:</u>	• 0 🛛
NGS: Mapping								http://biocluster.ucr.	<u>.edu/~rkaundal</u> .nnseg/tair10ch
NGS: RNA Analysis	Perform INDEL calling							r.fasta	inpocq/tan roth
NGS: SAM Tools	Perform INDEL Calling	<b>↓</b>							- 0.00
BCE Tools Cat This tool allows the	Skip INDEL calling if	he average per-	sample de	pth is above:				1: http://biocluster.ucr.	edu/∼rkaundal
user to concatenate BCF files.	250							/Galaxy_workshop/S	npseq/SRR0388
bcftools view Converts BCF format	Phred-scaled gap op	en sequencing e	rror probab	bility:				50.fastq	
to VCF format	40								
BCF Tools Index This tool allows	Platform for INDEL ca	ndidates							
the user to index sorted BCF for		indiduces	_						
random access.	Add new Platform fo	INDEL candidate	:S						
MPileup SNP and indel caller	Set advanced options	:							
<u>SAM-to-BAM</u> converts SAM format to BAM format	Basic 💠								
<u>bcftools view</u> Converts BCF format to VCF format	Execute								
NGS: GATK Tools (beta)	1111 12 - 1								
NGS: Peak Calling	what it does								
NGS: Simulation	Generate BCF or pi	eup for one or m	ultiple BAM	files. Alignment	records are group	ed by sam	ple identifiers in @RG header		
SNP/WGA: Data; Filters	lines. If sample ider	tifiers are absent	, each input	t file is regarded	as one sample.				
Phenotype Association									
VCF Tools	Settings								
	secongs								

# Bcftools view

- Converts BCF format to VCF format.
- Go to "NGS: SAM Tools", then click open "bcftools view"

- Galaxy	Analyze Data Workflow Shared Data → Visualization → Help → User →		Using 0%
Tools	bcftools view (version 0.0.1)	History	C 🕈
Fetch Sequences	Choose a bcf file to view:	Warkshan tast	
Fetch Alignments	9: MPileup on data 2 and data 8 💠	workshop_test	2
Get Genomic Scores	Retain all possible alternate alleles at variant sites:	294.5 MB	~/ <b>E</b>
Operate on Genomic Intervals	No ¢	10: MPileup on data	<u>2 and</u> • 0 %
<u>Statistics</u>	Output in the BCF format. The default is VCF.:	data 8 (log)	
Wavelet Analysis	No ¢	9: MPileun on data 2	and @ 0.82
Graph/Display Data	Sequence dictionary (list of chromosome names) for VCE->BCE conversion -	data 8	
Regional Variation	No +		<u>.</u>
Multiple regression		8: SAM-to-BAM on d	
Multivariate Analysis	Indicate PL is generated by 1921 or before (ordering is different).:	and data 7. converte	
Evolution		7: Map with BWA for	• 0 ×
Motif Tools	Suppress all individual genotype information:	Illumina on data 6 au	nd data 2:
Multiple Alignments	No ÷	mapped reads	
Metagenomic analyses	Skip sites where the REF field is not A/C/G/T:	6: Filter FASTQ on da	ata 3 👁 🖉 🕱
FASTA manipulation	No ÷		- 0.00
NGS: QC and manipulation	The input is VCF instead of BCF.:	5: FastQC_FASTQ_Gr	oomer @ 0 XX
NGS: Mapping	No ¢	on outer antim	
NGS: Indel Analysis	Uncompressed BCF output.:	4: FASTQ Summary	• 0 ×
NGS: RNA Analysis	No ¢	Statistics on data 3	
NGS: SAM Tools	Call variants using Bayesian inference. Automatically performs max-likelihood inference only:	3: FASTQ Groomer o	n data 👁 🖉 🕱
BCF Tools Cat This tool allows the	Yes +	1	
user to concatenate BCF files.	Parform may likelihood informer only including actimating the site allele frequency testing Hardy Weinberg	2.	@ / %
bcftools view Converts BCF format	equilibrium and testing associations with LRT:	http://biocluster.ucr	.edu/~rkaundal
to VCF format	No \$	/Galaxy_workshop/S	inpseq/tair10ch
BCF Tools Index This tool allows	Call par-scample genetymes at variant sites:	<u>r.fasta</u>	
the user to index sorted BCF for random access.	Can per sample genotypes at variant sites.	1:	• 0 ×
		http://biocluster.ucr	.edu/~rkaundal
MPileup SNP and indel caller	Use alternate INDEL-to-SNP mutation rate, default 0.15.:	/Galaxy_workshop/S	inpseq/SRR0388
<u>SAM-to-BAM</u> converts SAM format		<u>50.18514</u>	
to BAM format	variant_filter:		
bcftools view Converts BCF format			
to ver tormat	Specify scaled mutation rate for variant calling, default is 0.001.:		
NGS: GATK Tools (beta)	No ¢		
NGS: Peak Calling	Output variant sites only.:		
NGS: Simulation	Yes \$		
SNP/WGA: Data; Filters			
Phenotype Association	Execute		
VCF Tools			
Workflows	What it does:		
<ul> <li>All workflows</li> </ul>			

This tool converts BCF files into VCF files using BCFtools view from the SAMtools set of utilities:

# Rename history

- Galaxy	Analyze Data Workflow Shared Data - Visualization - Admin Help - User -	Using	g 0%
Tools		History	0 0
search tools	UCR Institute for Integrative Genome Biology	Unnamed history 23.0 Mt Click to rename history	0
Send Data		2: Sout on data 1	n ∞
ENCODE Tools		2: Sort on data 1	000
Lift-Over	Welcome to IIGB's Galaxy Server!	1: UCSC Main on Human:	• 0 ×
Text Manipulation	Overview	knownGene (genome)	
Filter and Sort	Galaxy is an open, highly customizable, web-based platform for the analysis		
Join, Subtract and Group	of next generation sequence data and many other biological data types. It		
Convert Formats	enables users to run computationally demanding next generation sequencing analysis tasks on powerful server hardware from a graphical web		
Extract Features	browser-based user interface rather than the Linux command-line. A subset		
Fetch Sequences	of of application supported by Galaxy is given in the left pane. Much more		
Fetch Alignments	are available here.		
Get Genomic Scores			
Operate on Genomic Intervals	Why Local Galaxy Service?		
Statistics	There are many advantages of using a local Galaxy server here at UCR rather		
Wavelet Analysis	important are: (1) shorter waiting queues for analysis tasks: (2) elimination of		
Graph/Display Data	time consuming uploads of large data sets; (3) support for analyzing much		
Regional Variation	larger data sets than this is possible on public services; (4) the ability to		
Multiple regression	customize software tools and uatabase conections.		
	How to Gain Access?		
Evolution Motif Tools	This instance of Galaxy runs on IIGB's high performance compute (HPC)		
Multiple Alignments	infrastructure, called Biocluster. As such its usage is covered by the annual		
Metagenomic analyses	active Biocluster account can access this Galaxy service using their existing		
inclugerorine unuryses	user name and password without any extra cost. New account requests for		
S	this service can be sent to <u>support@biocluster.ucr.edu</u> .		>

# Extract workflow from history

- Galaxy	Analyze Data Workflow Shared Data - Visualization - Admin Help - User -		Using 0%	%
Tools       search tools       Get Data       Send Data	UCR Institute for Integrative Genome Biology	Histor Unna 23.0 2: So	HISTORY LISTS Saved Histories Histories Shared with Me CURRENT HISTORY	
ENCODE Tools Lift-Over Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats Extract Features Fetch Sequences Fetch Alignments Get Genomic Scores	Welcome to IIGB's Galaxy Server! Overview Galaxy is an open, highly customizable, web-based platform for the analysis of next generation sequence data and many other biological data types. It enables users to run computationally demanding next generation sequencing analysis tasks on powerful server hardware from a graphical web browser-based user interface rather than the Linux command-line. A subset of of application supported by Galaxy is given in the left pane. Much more detailed descriptions of Galaxy's basic functionalities including user tutorials are available <u>here</u> .		Create New Copy History Copy Datasets Share or Publish Extract Workflow Dataset Security Resume Paused Jobs Collapse Expanded Datasets Include Deleted Datasets Include Hidden Datasets	ets
Statistics Wavelet Analysis Graph/Display Data Regional Variation Multiple regression Multivariate Analysis Evolution Motif Tools Multiple Alignments Metagenomic analyses	<ul> <li>Why Local Galaxy Service?</li> <li>There are many advantages of using a local Galaxy server here at UCR rather than public test instances of Galaxy available on the internet. The most important are: (1) shorter waiting queues for analysis tasks; (2) elimination of time consuming uploads of large data sets; (3) support for analyzing much larger data sets than this is possible on public services; (4) the ability to customize software tools and database collections.</li> <li>How to Gain Access?</li> <li>This instance of Galaxy runs on IIGB's high performance compute (HPC) infrastructure, called Biocluster. As such its usage is covered by the annual registration fee for this infrastructure (see <u>here</u> for details). Users with an active Biocluster account can access this Galaxy service using their existing user name and password without any extra cost. New account requests for</li> </ul>		Unhide Hidden Datasets Delete Hidden Datasets Purge Deleted Datasets Show Structure Export to File Delete Delete Permanently OTHER ACTIONS Import from File	
javascript:void(0);	this service can be sent to <u>support@biocluster.ucr.edu</u> .	1	P	

# Exercise 2: RNA-seq Analysis

- Data source: RNA-seq experiment SRA023501
- Four Samples:

Samples	Factors	Fastq
AP3_f14	AP3	http://biocluster.ucr.edu/~rkaundal/Galaxy_workshop/Rnaseq/SRR064154.fastq
AP3_f14	AP3	http://biocluster.ucr.edu/~rkaundal/Galaxy_workshop/Rnaseq/SRR064155.fastq
T1_f14	TRL	http://biocluster.ucr.edu/~rkaundal/Galaxy_workshop/Rnaseq/SRR064166.fastq
T1_f14	TRL	http://biocluster.ucr.edu/~rkaundal/Galaxy_workshop/Rnaseq/SRR064167.fastq

# RNA-seq Analysis workflow



# Upload Data

- Upload four fastq files with URL
- Upload tair10chr.fa with URL (http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/tair10chr.fasta)
- Upload TAIR10.GTF with URL, specify the format "gtf" (<u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/TAIR10.GTF</u>)

	💳 Galaxy	Analyze Data Workflow Shared Data - Visualization - Help - User -			Using 0%
	Tools	Upload File (version 1.1.3)		History	C \$
	search tools	File Format:		Unnamed history	
	Get Data	Auto-detect		0 bytes	0
$\land$	Upload File from your computer	Which format? See help below		1 Your history is em	pty. Click 'Get
V	UCSC Main table browser	File:		Data' on the left pa	ane to start
	UCSC Test table browser	Choose File No file chosen TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or F	TP (if enabled by		
	UCSC Archaea table browser	the site administrator).			
	<u>BX</u> table browser	URL/Text:			
	EBI SRA ENA SRA	http://biocluster.ucr.edu/~rkaundal/Galaxy_ workshop/Rnaseg/SRR064154 fasto			
	Get Microbial Data	http://biocluster.ucr.edu/~rkaundal/Galaxy_			
	BioMart Central server	workshop/Rnaseq/SRR064155.tastq http://biocluster.ucr.edu/~rkaundal/Galaxy			
	BioMart Test server	Here you may specify a list of URLs (one per line) or paste the contents of a file.			
	CBI Rice Mart rice mart	Files uploaded via FTP:			
	GrameneMart Central server	File Size Date			
	modENCODE fly server	Your FTP upload directory contains no files.			
	Elymine server	To upload files greater than 2GB in size, it is recommended that you upload your files on any webserver as biocluster and then paste the UR to your file in the URI /Text box. You may use filezilla and transfer the file to your, html directory in your biocluster account. The url of you	L.		
	Flymine test server	file will be (http://biocluster.ucr.edu/~username/filename)			
	modENCODE modMine server	Convert spaces to tabs:			
	MouseMine server	Yes Use this option if you are entering intervals by hand			
	<u>Ratmine</u> server	Genome:			
	YeastMine server	unspecified (?)			
	metabolicMine server				
	modENCODE worm server	Execute			
	WormBase server				
	Wormbase test server	Auto-detect			
	EuPathDB server	The system will attempt to detect Axt, Fasta, Fastqsolexa, Gff, Gff3, Html, Lav, Maf, Tabular, Wiggle, Bed and Interval (Bed with headers) form	ats. If your file is		
	EncodeDB at NHGRI	rows). You can still coerce the system to set your data to the format you think it should be. You can also upload compressed files, which will	automatically be		
	EpiGRAPH server	decompressed.	101		

# Fastq Groomer

- Select NGS: QC and manipulation and Fastq groomer
- Run Fastq Groomer for all the 4 fastq sequences.

- Galaxy	Analyze Data Workflow Shared Data - Visualization - Help - User -	382	Using 0%
Tools	FASTO Groomer (version 1.0.4)	History	C 0
Graph/Display Data		Unnamed history	
Regional Variation	File to groom:		2. 🗪
Multiple regression	1. http://biocluster.ucr.edu/~rkaindal/Galaxy_workshop/khased/skk064134.lastq 👻	003.0 MB	Ø 🖻
Multivariate Analysis	Input FASTQ quality scores type:	<u>7:</u>	• 1 ×
Evolution	Sanger ¢	http://biocluster.ucr	.edu/~rkaundal
Motif Tools	Advanced Options:	/Galaxy workshop/F	Reased/SRR0641
Multiple Alignments	Hide Advanced Options	UTHASIQ	
Metagenomic analyses		<u>6:</u>	• / ×
FASTA manipulation	Execute	http://biocluster.ucr	<u>.edu/~rkaundal</u>
NGS: QC and manipulation		66.fastg	(naseq/skk0041
FastQC: Comprehensive QC reporting for short read sequence	What it does	5:	@ / X
Barcode Splitter (ngs-tools)	This tool offers several conversions options relating to the FASTQ format.	http://biocluster.ucr	.edu/~rkaundal
	When using Basic options, the output will be sanger formatted or cssanger formatted (when the input is Color Space Sanger).	/Galaxy workshop/F	<u>Inaseq/TAIR10.</u>
Clip adapter sequences	When converting, if a quality score falls outside of the target score range, it will be coerced to the closest available value (i.e. the minimum or maximum).		
ILLUMINA FASTQ	When converting between Solexa and the other formats, quality scores are manned between Solexa and PHRED scales using the equations found in Cock PL	<u>4:</u>	• 0 ×
FASTQ Groomer convert between various FASTQ quality formats	Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2009 Dec 16.	http://biocluster.ucr /Galaxy workshop/F	<u>.edu/~rkaundal</u> <u>Rnaseq/tair10ch</u>
FASTQ splitter on joined paired	When converting between color space (csSanger) and base/sequence space (Sanger, Illumina, Solexa) formats, adapter bases are lost or gained; if gained, the		- 0.00
enu reaus	base o is used as the adapter. Tou cannot convert a color space read to base space in there is no adapter present in the color space sequence. Any masked or ambiauous nucleotides in base space will be converted to 'N's when determining color space normality.	<u>2:</u> http://biocluster.ucr	edu/~rkaundal
FASTQ joiner on paired end reads		/Galaxy workshop/F	Rnaseq/SRR0641
FASTQ Summary Statistics by column	Quality Score Comparison	55.fastq	
ROCHE-454 DATA	\$	<u>1:</u> http://biocluster.ucr	● Ø X redu/~rkaundal
Build base quality distribution		/Galaxy workshop/F	Rnaseq/SRR0641
Select high quality segments	!"#\$%&'()*+,/0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopgrstuvwxyz{ }~	<u>34.18510</u>	
Combine FASTA and QUAL into FASTQ	33 59 64 73 104 126		
AB-SOLID DATA	S - Sanger Phred+33, 93 values (0, 93) (0 to 60 expected in raw reads)		
Convert SOLiD output to fastq	X - Solexa Solexa+64, 67 values (-5, 62) (-5 to 40 expected in raw reads)		
Compute quality statistics for	Diagram adapted from http://en.wikipedia.org/wiki/FASTQ format		
SOLD Uata	Output from Illumina 1.8+ pipelines are Sanger encoded.		
<u>Draw quality score boxplot</u> for SOLiD data			



# Alignment with TopHat

- TopHat is a fast splice junction mapper for RNA-Seq reads, it can identify splice junctions between exons.
- Go to "NGS RNA analysis", click open "Tophat for illumina"
- Similarly repeat this process for all the 4 fastq groomed sequences



# Find Significant Changes

- Cuffdiff find significant changes in transcript expression.
- Go to "NGS RNA analysis", click open "Cuffdiff"

Toris     C. dfdff (version 0.0.7)     History     Q ⊂ Q ⊂       Cet Genomic Kores     Statistic     Statistic     Statistic       Statistic     Statistic     Statistic     Statistic       Statistic     Conditions     Conditions     Statistic       Conditions     Conditions     Statistic       Conditions     Conditions     Statistics       Conditions     Statistics     Statistics       Conditions     Conditions     Statistics       Conditions     Statistics     Statistics       Conditions     Conditions     Statistics       Conditions     Replicate Conditions     Statistics       Multiple Conservation Conditions     Statistics     Statistics       Multiple Conservation Conditions     Statistics     Statistics       Conditions     Replicate Conditions     Statistics     Statistics       Multiple Conservation Conditions     Statistics     Statistics     Statistics       Multiple Conservation Conditions     Statistics </th <th>- Galaxy</th> <th>Analyze Data Workflow 9</th> <th>Shared Data 🗸 Visualization 🛨</th> <th>Help + User +</th> <th>Using 1%</th>	- Galaxy	Analyze Data Workflow 9	Shared Data 🗸 Visualization 🛨	Help + User +	Using 1%
Get Genomic Scores       55: Tophat for Illumina com # 0 2%         Questa to Gettamonic Internals       55: Tophat for Illumina com # 0 2%         Waviet Analysis       Gettamonic Internals         Grant And Display Data trains       57: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       53: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       54: Tophat for Illumina com # 0 2%         Multipation       56: Tophat for Illumina com # 0 2%         Multipation       56: Tophat for Illumina com # 0 2%         Multipation       56: Tophat for Illumina com # 0 2%         Multipation       56: Tophat for Illumina com # 0 2%         Multipation       56: Tophat for Illumina com # 0 2%	Tools Cuffdiff (version 0.0.	)			History 2 🗢
Operate on Genomic Intervals         74: http://bioluter.aur.edu/~fsauad1/Galax_work/http/fmaseqTARED.OTF_1         data 11 and data 4: accented.htts           Statistics         Conditions         94: http://bioluter.aur.edu/~fsauad1/Galax_work/http://fmaseqTARED.OTF_1         data 11 and data 4: accented.htts           Wavelet Analysis         Conditions         94: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 11 and data 4: accented.htts         data 11 and data 4: accented.htts           Multiple Constraintion         04: atta 12 and data 4: accented.htts         d	Get Genomic Scores				56: Tophat for Illumina on ⊕ Ø 🛛
Statistics       A transcript CFF 30 CFT file produced by cuffinis, cuffcompare, or other source.       \$5: Tophs for filmmins on # 0 2%         Maxiet Anabalis       Granul/Display Data       Barling and an 4: splice       January 2000         Regional Variation       Name:       Statistice       January 2000         Multiple representation       Regional Variation       Statistice       Statistice         NSS. Model Anabylis       Regional Variation       Statistice       Statistice </td <td>Operate on Genomic Intervals 74: http://biocluste</td> <th>.ucr.edu/~rkaundal/Galaxy_wo</th> <td>rkshop/Rnaseg/TAIR10.GTF 🛟</td> <td></td> <td>data 11 and data 4: accepted_hits</td>	Operate on Genomic Intervals 74: http://biocluste	.ucr.edu/~rkaundal/Galaxy_wo	rkshop/Rnaseg/TAIR10.GTF 🛟		data 11 and data 4: accepted_hits
Wavelt Ambais     Conditions     dest 1 and data 4: splice       Graph/UsiaNDA     Name:     mattions       Multiple registion     Name:     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Multiple registion     Repicates     Shift Tophat for Illumina on # 0 %       Shift Tophat for Illumina on data 9 and data 4: accepted hits # <t< td=""><td>Statistics A transcript GFF3 or</td><th>TF file produced by cufflinks, c</th><td>uffcompare, or other source.</td><td></td><td>55: Tophat for Illumina on @ 0 🕱</td></t<>	Statistics A transcript GFF3 or	TF file produced by cufflinks, c	uffcompare, or other source.		55: Tophat for Illumina on @ 0 🕱
Graphilos     Condition 1     Junctions       Regional Variation     Sat: Tophat for Illumina on # 0 %       Multiple regression     Sat: Tophat	Wavelet Analysis Conditions				data 11 and data 4: splice
Residencial Variation       Name:       Trated       S2-Tophat for illumina on $\ll 0 \times 2$ Multiple registion       S2-Tophat for illumina on $\ll 0 \times 2$ S2-Tophat for illumina on $\ll 0 \times 2$ Multiple registion       S2-Tophat for illumina on $\ll 0 \times 2$ S2-Tophat for illumina on $\ll 0 \times 2$ Multiple registion       S2-Tophat for illumina on $\ll 0 \times 2$ S2-Tophat for illumina on $\ll 0 \times 2$ Multiple Alignments       Method registion       S2-Tophat for illumina on $\ll 0 \times 2$ Multiple Alignments       Repicate 1       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 1       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 1       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 1       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 1       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 2       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 2       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 2       S2-Tophat for illumina on $\ll 0 \times 2$ Motion       Repicate 2       S2-Tophat for illumina on $\ll 0 \times 2$ Matter       Condition 2       S2-Tophat for illumina on $\ll 0 \times 2$ Mater       S2-Tophat for illumina on $\ll 0 \times 2$ <td>Graph/Display Data Condition 1</td> <th></th> <td></td> <td></td> <td>junctions</td>	Graph/Display Data Condition 1				junctions
Multiple reference     fata 11 and data 4: deteriors       Multiple reference     fata 11 and data 4: deteriors       Replicates     fata 11 and data 4: deteriors       Replicates     fata 11 and data 4: deteriors       Multiple Alignments     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2       Multiple Reference     fata 10 and data 4: accepted_hits 2	Regional Variation				54: Tophat for Illumina on ⊕ Ø 🖄
Multivadate Analysis       S3: Tophit for illumina on 40 %         Multivadate Analysis       S3: Tophit for illumina on 40 %         Multivadate Analysis       Replicate 1         Multivadate Analysis       S2: Tophit for illumina on 40 %         Multivadate Analysis       S2: Tophit for illumina on 40 %         Multivadate Analysis       Replicate 1         Multivadate Analysis       S2: Tophit for illumina on 40 %         Multivadate Analysis       Replicate 1         Multivadate Analysis       Replicate 1         Multivadate Analysis       Replicate 1         Multivadate Analysis       Replicate 2         NGS: Indel Analysis       Replicate 2         Multivadate Analysis       Replicate 2         Multivadate Analysis       Replicate 2         DESign Otermines differential exercised formexering form of multimina on adata 9 and data 4: accepted_hits 6       Replicate 2         DESign Otermines differential exercised formexering form of multimina on adata 9 and data 4: accepted_hits 6       Replicate 2         DESign Otermines differential exercised formexering form of multimina on adata 4: accepted_hits 6       Replicate 2         DESign Otermines differential exercised formexering form of multimina on adata 4: accepted_hits 6       Replicate 2         Norticates       Replicate 1       Replicate 2         DESign Oter	Multiple regression Trastad				data 11 and data 4: deletions
Louistion Moit Loois Mutiple Alignments Mutiple Alignments Mutiple Alignments Mutiple Alignments Mutiple Alignments Mate ageomethy Respicate 1 Add replicate 1 Add replicate 1 Add replicate 1 Remove Replicate 1 Add replicate 2 Add replicate 3 BESS Determines differentially expressed transcripts from read alignments DESSing Determines differentially expressed due to and data 4: accepted , hits 0 Remove Replicate 2 Condition 2 NAA-SEQ Tophat for Illumina on od ata 10 and data 4: accepted , hits 0 Remove Replicate 1 Add replicate Remove Replicate 1 Remove Replicate 2 Condition 2 NAA-SEQ Tophat for Illumina on od ata 10 and data 4: accepted , hits 0 Remove Replicate 1 Remove Replicate 1 Remove Replicate 1 Replicate 2 Add replicate So Tophat for Illumina on 0 / % data 8 and data 4: accepted , hits 0 Replicate 2 Add replicate 1 Remove Replicate 1 Remove Replicate 1 Remove Replicate 1 Replicate 2 Add replicate 1 Remove Replicate 1 Remove Replicate 1 Add replicate 2 Add replicate 3 Add replicate 2 Ad	Multivariate Analysis				53: Tophat for Illumina on 👁 🖉 💥
Multiple       Replicate 1       S2: Tophat for: Illumina on $0 / 22$ Multiple       Add replicate:       S2: Tophat for: Illumina on $0 / 22$ MS: SC: Cand manipulation       Remove Replicate 1       S2: Tophat for: Illumina on $0 / 22$ MS: Standal Analysis       Remove Replicate 1       S2: Tophat for: Illumina on $0 / 22$ MS: Standal Analysis       Remove Replicate 2       S2: Tophat for: Illumina on $0 / 22$ MS: Standal Analysis       Remove Replicate 2       S2: Tophat for: Illumina on $0 / 22$ MS: Standal Analysis       Remove Replicate 2       S2: Tophat for: Illumina on $0 / 22$ MS: Standal Analysis       Remove Replicate 2       S2: Tophat for: Illumina on $0 / 22$ DESseg Deternines differentially expressed instructures from each at sceneted hits       data 10 and data 4: accepted hits         data Band data 4: splice junctions       S3: Tophat for: Illumina on $0 / 22$ DESseg Inference of differentially expressed in standata 4: splice junctions       S3: Tophat for: Illumina on $0 / 22$ Replicate 1       Add new Replicate       S3: Tophat for: Illumina on $0 / 22$ Replicate 5       Replicate 6       S3: Tophat for: Illumina on $0 / 22$ Replicate 6       Nme:       S3: Tophat for: Illumina on $0 / 22$ DESseg Deternines differentially expressed for Illumina on $0 / 22$ S3: Tophat	Evolution Replicates				data 11 and data 4: insertions
Matagenomic analyzes     Add replicate:     Sz. Tophat for illumina on data 8 and data 4; accepted_hits       Matagenomic analyzes     Add replicate:     Sz. Tophat for illumina on data 8 and data 4; accepted_hits       MGS: Mapping     Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Remove Replicate 2     Sg. Tophat for illumina on data 9 and data 4; accepted_hits       MGS: Mapping     Sg. Tophat for illumina on data 9 and data 4; accepted_hits     Sg. Tophat for illumina on data 4; accepted_hits       MGS: Mapping     Sg. Tophat for illumina on data 9 and data 4; accepted_hits     Sg. Tophat for illumina on data 4; accepted_hits       MGS: Mapping     Sg. Tophat for illumina on data 9 and data 4; accepted_hits     Sg. Tophat for illumina on data 4; accepted_hits       MGS: Mapping     MGS: Mapping     Sg. Tophat for illumina on data 4; accepted_hits     Sg. Tophat for illumina on data 4; accepted_hits       MGS: Mapping     MGS: Mapping     MGS: Mapping     Sg. Tophat for illumina on data 4; accepted_hits       MGS: MGA Mappi	Multiple Alignments Replicate 1				57: Tenhat for Illuming on @ 0 %
PASTA manipulation       48: Tophat for Illumina on data 8 and data 4: accepted_hits 9       51: Tophat for Illumina on 0 2         MGS: Mapping       Remove Replicate 1       Sc. Oraphat for Illumina on 0 2         NGS: Indel Analysis       Remove Replicate 2       Sc. Oraphat for Illumina on 0 2         Solitions       Add replicate 2       Sc. Oraphat for Illumina on 0 2         Solitions       Remove Replicate 2       Sc. Oraphat for Illumina on 0 2         DEScap Determines differential evon usage in RNA-Seq       Add replicate 2       Sc. Tophat for Illumina on 0 2         DEScap Differential evon usage in RNA-Seq       Condition 2       Name:       Sc. Tophat for Illumina on 0 2         DEScap Differential evon usage in RNA-Seq       Untreated       Sc. Tophat for Illumina on 0 4       Sc. Tophat for Illumina on 0 4         NRM:       Remove Replicate 1       Name:       Sc. Tophat for Illumina on 0 4       Sc. Tophat for Illumina on 0 4         DEScap Differential gene evon usage in RNA-Seq       Untreated       Sc. Tophat for Illumina on 0 4       Sc. Tophat for Illumina on 0 4       Sc. Tophat for Illumina on 0 4         Replicate 1       Replicate 1       Replicate 1       Sc. Tophat for Illumina on 0 4       Sc. Tophat for	Metagenomic analyses Add replicate:				data 10 and data 4: accepted hits
NG3: OC and manipulation       Remove Replicate 1       Si: Tophat for Illumina on % 0 %         NG3: Mapping       NG3: Mapping       Replicate 1         NG3: MAPPING       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         Gutfdift find significant changes in transcript expression, splicing, and promoter use       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         DESeg Determines differentially expression analysis based on the negative binomial distribution       Add new Replicate 2       Si: Tophat for Illumina on % 0 %         DESeg Differential gene expression analysis based on the negative binomial distribution       Condition 2       Si: Tophat for Illumina on % 0 %         NMA: Seq data Using Trinity RNA-Seq data       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         Replicate 1       Name:       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         Tiphat Seg Differential gene expression analysis based on the negative binomial distribution       Condition 2       Si: Tophat for Illumina on % 0 %         Tiphat Seg Differential gene expression using RNA-seq data       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         Replicate 1       Name:       Si: Tophat for Illumina on % 0 %       Si: Tophat for Illumina on % 0 %         Tiphat Seg Differential gene expression using RNA-seq data Si deletions       Si: Tophat for Illumina on	FASTA manipulation 48: Tophat for	Illumina on data 8 and data 4: a	ccepted_hits 💠		
NGS: Mapping     Immunose replicate 1       NGS: Indel Analysis     Feplicate 2       Cutfield find significant changes in transmission replicate 2     Sci. Tophat for Illumina on the 9 pince 9: Tophat for Illu	NGS: OC and manipulation	ate 1			51: Tophat for Illumina on (1) (2) (2)
NGS: Indel Analysis       Replicate 2       So: Tophat for Illumina on data 9 and data 4: accepted_hits \$       So: Tophat for Illumina on $@ 0 @ 0$ Cutfiding find significant changes in transcript expression, and promoter use       DESseg Determines differentially expression (analysis based on the negative binomial data 4: insertions)       So: Tophat for Illumina on $@ 0 @ 0$ DESseg Inference of differential econ uses in RNA-Seq       Condition 2       So: Tophat for Illumina on $@ 0 @ 0$ DESseg Inference of differential econ uses in RNA-Seq       Condition 2       So: Tophat for Illumina on $@ 0 @ 0$ DESseg Differencial gene expression analysis based on the negative binomial distribution       Condition 2       So: Tophat for Illumina on $@ 0 @ 0$ Timity De novo assembly of RNA-Seq       Replicate 1       Add replicate:       So: Tophat for Illumina on $@ 0 @ 0$ RunA-seq       Tophat for Illumina on data 10 and data 4: accepted_hits \$       Replicate 1       So: Tophat for Illumina on $@ 0 @ 0$ RunA-seq       Tophat for Illumina on data 10 and data 4: accepted_hits \$       Replicate 1       So: Tophat for Illumina on $@ 0 @ 0$ RunA-seq data       Tophat for Illumina on data 10 and data 4: accepted_hits \$       Replicate 1       So: Tophat for Illumina on $@ 0 @ 0$ RunA-seq data       Tophat for Illumina on data 10 and data 4: accepted_hits \$       Replicate 1       So: Tophat for Illumina on $@ 0 @ 0$ RunA-seq data	NGS: Mapping	ite I			junctions
NGS: RNA Analysis       So: Iophat for Illumina on % 0 %         Lufdiff find significant changes in transcript expression, splicing, and promoter use       Add replicate:       44: Tophat for Illumina on % 0 %         DESea Determines differentially expressed transcripts from read alignments       Add new Replicate 2       49: Tophat for Illumina on % 0 %         DESea Differential gene expression analysis based on the negative binomial distribution       Condition 2       47: Tophat for Illumina on % 0 %         NRM-Seq Untrasted binomial distribution       Condition 2       Name:       98: Tophat for Illumina on % 0 %         VRM-Seq data Using Trinty RNA-Seq data       Condition 2       86: Tophat for Illumina on % 0 %         Tophat for Illumina find splice junctions using RNA-seq data       Condition 2       86: Tophat for Illumina on % 0 %         Replicate 1       Add replicate:       43: Tophat for Illumina on % 0 %         Tophat for Illumina find splice junctions using RNA-seq data       For Illumina on data 10 and data 4: accepted_hits 4       44: Tophat for Illumina on % 0 %         Replicate 2       Add replicate:       52: Tophat for Illumina on data 10 and data 4: accepted_hits 4       44: Tophat for Illumina on % 0 %         Replicate 1       Add replicate:       52: Tophat for Illumina on data 10 and data 4: accepted_hits 4       44: Tophat for Illumina on % 0 %       44: Tophat for Il	NGS: Indel Analysis Replicate 2				
Cutfailf find significant changes in transform expression analysis based on the negative binomial distribution       44: Tophat for Illumina on data 9 and data 4: accepted_hits ‡       44: Tophat for Illumina on $@ 0$ % data 10 and data 4: insertions         DESeg Determines differential expressed transcripts from read alignments       Add new Replicate 2       43: Tophat for Illumina on $@ 0$ % data 8 and data 4: insertions       48: Tophat for Illumina on $@ 0$ % data 8 and data 4: insertions         DESeg Determines differential expression analysis based on the negative binomial distribution       Condition 2       46: Tophat for Illumina on $@ 0$ % data 8 and data 4: jolice junctions         Trinity De novo assembly of RNA-Seq data Using Trinity RNA-Seq data Using Trinity RNA-Seq data Using Trinity RNA-Seq data       Replicate 1         Tophat for Illumina on data 10 and data 4: accepted_hits ‡       49: Tophat for Illumina on $@ 0$ % data 8 and data 4: accepted_hits ‡         Tophat for Illumina on data 10 and data 4: accepted_hits ‡       49: Tophat for Illumina on $@ 0$ % data 8 and data 4: accepted_hits ‡         Tophat for Illumina on data 10 and data 4: accepted_hits ‡       49: Tophat for Illumina on $@ 0$ % data 8 and data 4: accepted_hits ‡         Tophat for Illumina on data 10 and data 4: accepted_hits ‡       49: Tophat for Illumina on $@ 0$ % data 8 and data 4: accepted_hits ‡         Tophat for Illumina on data 10 and data 4: accepted_hits ‡       49: Tophat for Illumina on $@ 0$ % data 8 and data 4: deletions         Tophat for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on	NGS: RNA Analysis				50: Tophat for Illumina on (19) 7 X
in transcript expression,       splicing, and promoter use       d3: Tophat for Illumia on ⊕ 0 %         DESeq Determines differentially       expressed transcripts from read alignments       d3: Tophat for Illumina on ⊕ 0 %         DEXSeg Inference of differential exon usage in RNA-Seq       Condition 2       data 8 and data 4: accepted hits         DEXSeg Inferencial gene expression analysis based on the negative binomial distribution       Condition 2       data 8 and data 4: splice junctions         Tinity De novo assembly of RNA-Seq       Replicate 1       dd replicate:       dd replicate         Tophat for Illumina find splice junctions using RNA-seq data       Tophat for Illumina on data 10 and data 4: accepted_hits ‡       dd: Tophat for Illumina on ⊕ 0 %         Tophat for SUID Find splice junctions using RNA-seq data       Replicate 2       dd: Tophat for Illumina on ⊕ 0 %         Tophat for SUID Find splice junctions using RNA-seq data       Replicate 2       dd: and data 4: accepted_hits ‡         Tophat for SUID Find splice junctions using RNA-seq data       Replicate 2       dd: and data 4: accepted_hits ‡         Tophat for SUID Find splice junctions using RNA-seq data       S5: Tophat for Illumina on data 11 and data 4: accepted_hits ‡       data 9 and data 4: insertions         S6: Tophat for Illumina on data 11 and data 4: accepted_hits ‡       data 9 and data 4: insertions       data 9 and data 4: insertions         S6: Tophat for Illumina on data 11 and data 4: accepted_hits ‡	Cuffdiff find significant changes 44: Tophat for	Illumina on data 9 and data 4: a	ccepted_hits 💠		
DSSeq Determines differentially expressed transcripts from read alignments       Add new Replicate 2       48: Tophat for Illumina on $\oplus$ 0 % data 8 and data 4: accepted hits         DEXSeq Inferentially exon usage in RNA-Seq       Condition 2       49: Tophat for Illumina on $\oplus$ 0 % data 8 and data 4: splice junctions         DEXSeq Differential exon usage in RNA-Seq       Condition 2       44: Tophat for Illumina on $\oplus$ 0 % data 8 and data 4: splice junctions         DEXSeq Differential exon usage in RNA-Seq       Condition 2       45: Tophat for Illumina on $\oplus$ 0 % data 8 and data 4: splice junctions         DEXSeq Solution and the regative binomial distribution       Intinity De novo assembly of Replicates       Name: Untreated Replicate 1       45: Tophat for Illumina on $\oplus$ 0 % data 8 and data 4: insertions         NAM-Seq junctions using RNA-seq data express Quantify the abundances of a set of target sequences       Replicate 2       43: Tophat for Illumina on $\oplus$ 0 % data 9 and data 4: deletions         Add replicate 2       Add new Replicate 2       43: Tophat for Illumina on $\oplus$ 0 % data 9 and data 4: insertions         Stophat for Subp RNA-seq data express Quantify the abundances of a set of target sequences       56: Tophat for Illumina on data 1: and data 4: accepted hits $\oplus$ Replicate 2       43: Tophat for Illumina on $\oplus$ 0 % data 9 and data 4: insertions         Add new Replicate 2       Add new Replicate 2       43: Tophat for Illumina on $\oplus$ 0 % data 9 and data 4: insertions	in transcript expression,		· · · · ·		49: Tophat for Illumina on ● Ø ※
DEsc       Determines differentially expressed transcripts from read alignments       Add new Replicate       48: Tophat for Illumina on @ / % data 8 and data 4: accepted tota 8 and data 4: splice junctions         DEScs       DEScs       Differential gene expression analysis based on the negative binomial distribution       Condition 2       48: Tophat for Illumina on @ / % data 8 and data 4: splice junctions         DEScs       DEScs       Differential gene expression analysis based on the negative binomial distribution       Condition 2       48: Tophat for Illumina on @ / % data 8 and data 4: splice junctions         Tinity De novo assembly of RNA-Seq data Using Trinity RNA-Seq data Using Trinity       Replicate 1       43: Tophat for Illumina on @ / % data 9 and data 4: accepted hits ‡         Tophat for Illumina Find splice junctions using RNA-seq data       Tophat for Splice for RNA-seq data       Replicate 2         Add replicate:       56: Tophat for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on @ / % data 9 and data 4: insertions         Tophat for Sulp Find splice junctions using RNA-seq data       56: Tophat for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on @ / % data 9 and data 4: insertions         Add new Replicate 2       Add new Replicate 2       4dd new Replicate 2       4dd new Replicate 2         Aud new Replicate       11: FASTO Groomer on @ / % data 7       4/ % data 7 </td <td>Remove Replic</td> <th>ate 2</th> <td></td> <td></td> <td>data 10 and data 4. Insertions</td>	Remove Replic	ate 2			data 10 and data 4. Insertions
alignments       Add new Replicate       data 8 and data 4; accepted hits         DEXseq Inference of differential exon usage in RNA-Seq       Condition 2       data 8 and data 4; accepted hits         DEscace Differential exon usage in RNA-Seq       Condition 2       data 8 and data 4; accepted hits         DEscace Differential exon usage in RNA-Seq       Mame:       Untreated       data 8 and data 4; collations         Descace Differential expression analysis based on the negative binomial distribution       Trinity De novo assembly of RNA-Seq data Using Trinity       Replicate 1       data 8 and data 4; collations         NAM-Seq       Add rew Replicate       Replicate 1       data 8 and data 4; collations       data 8 and data 4; collations         Trinity De novo assembly of RNA-Seq       Replicate 1       Add reglicate:       data 9 and data 4; collations       data 9 and data 4; collations         Tophat for Illumina on data 10 and data 4; accepted_hits ‡       Replicate 1       data 9 and data 4; splice junctions       data 9 and data 4; collations         Tophat for SOLID Find splice       Add replicate 2       data 4; accepted_hits ‡       data 9 and data 4; collations       data 9 and data 4; collations         Add new Replicate 2       Add replicate 2       data 9 and data 4; collations       data 9 and data 4; collations       data 9 and data 4; collations         Tophat for Illumina on data 11 and data 4; accepted_hits ‡       fa	DESeq Determines differentially expressed transcripts from read				48: Tophat for Illumina on 👁 🖉 💥
DEXSeq Inference of differential exon usage in RNA-Seq       47: Tophat for Illumina on @ 0 % data 8 and data 4: splice junctions         DESeq2 Differential gene expression analysis based on the negative binomial distribution       10: Tophat for Illumina on @ 0 % data 8 and data 4: splice junctions         Trinity De novo assembly of RNA-seq data Using Trinity       Replicates       8 epicate 1         NAME:       52: Tophat for Illumina on data 10 and data 4: accepted_hits ‡       44: Tophat for Illumina on @ 0 % data 8 and data 4: insertions         Tophat 2 Gapped-read mapper for RNA-seq data       Tophat for SOLID Find splice junctions using RNA-seq data       43: Tophat for Illumina on @ 0 % data 9 and data 4: accepted_hits ‡         Tophat 2 Gapped-read mapper for RNA-seq data       Add replicate:       4dit replicate:       4dit replicate:         Tophat 2 Gapped-read mapper for RNA-seq data       Sci Tophat for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on @ 0 % data 9 and data 4: deletions         Sci Tophat 2 Gapped-read mapper for RNA-seq data       Sci Tophat for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on @ 0 % data 9 and data 4: deletions         Sci Tophat 4 for Illumina on data 11 and data 4: accepted_hits ‡       42: Tophat for Illumina on @ 0 % data 9 and data 4: insertions         Add replicate       Add replicate       41: Tophat for Illumina on @ 0 % data 9 and data 4: insertions         Sci Tophat 5 or SoliD Find splice junctions       Sci Tophat for Illumina on d	alignments Add new Replica	2			data 8 and data 4: accepted_hits
exon usage in RNA-Seq     Name:     data 8 and data 4: splice junctions       DESeq2 Differential gene expression analysis based on the negative binomial distribution     Name:     data 8 and data 4: splice junctions       Trinity De novo assembly of RNA-Seq data Using Trinity     Replicate 1     Add replicate:       Tophat for Illumina Find splice junctions using RNA-seq data     S2: Tophat for Illumina on $\emptyset 0$ % data 9 and data 4: accepted_hits ‡       Tophat for Illumina Find splice junctions using RNA-seq data     Replicate 1       Replicate 2     Replicate 2       Add replicate:     42: Tophat for Illumina on $\emptyset 0$ % data 9 and data 4: accepted_hits ‡       Tophat for SOLID Find splice junctions using RNA-seq data     Replicate 2       Add replicate 2     Add replicate 2       Add replicate 2     Add replicate 2       Add new Replicate 2     Add new Replicate 3	DEXSeg Inference of differential				47: Tophat for Illumina on 👁 🖉 💥
DESeq2 Differential gene expression analysis based on the negative binomial distribution     Marie: Untreated     46: Tophat for Illumina on @ 0 %       Trinity De novo assembly of RNA-Seq data Using Trinity RNA-Seq data Using Trinity RNA-Seq data Using Trinity RNA-Seq data Using RNA-seq data     Add replicate: S2: Tophat for Illumina on data 10 and data 4: accepted_hits \$     43: Tophat for Illumina on @ 0 %       Tophat for Illumina Find splice junctions using RNA-seq data     Replicate 1     Add replicate 1       Tophat for SULD Find splice junctions using RNA-seq data     Replicate 2     43: Tophat for Illumina on @ 0 %       Add replicate:     S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$     43: Tophat for Illumina on @ 0 %       Keplicate 2     Add replicate:     43: Tophat for Illumina on @ 0 %       Add replicate:     S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$     42: Tophat for Illumina on @ 0 %       Add replicate 2     Add replicate 1     Add replicate 2     42: Tophat for Illumina on @ 0 %       Add replicate:     S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$     42: Tophat for Illumina on @ 0 %       S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$     41: Tophat for Illumina on @ 0 %       Add new Replicate 2     Add new Replicate 2     42: Tophat for Illumina on @ 0 %       Add new Replicate 2     Add new Replicate 2     42: Tophat for Illumina on @ 0 %	exon usage in RNA-Seq				data 8 and data 4: splice junctions
expression analysis based on the negative binomial distribution       Untreated       data 8 and data 4: deletions         Trinity De novo assembly of RNA-Seq data Using Trinity       Replicate 1       data 8 and data 4: insertions         RNA-SEQ       Add replicate:       52: Tophat for Illumina on data 10 and data 4: accepted_hits \$       data 8 and data 4: deletions         Tophat for Illumina Find splice junctions using RNA-seq data       Replicate 2       data 8 and data 4: accepted_hits \$       data 9 and data 4: accepted_hits         Tophat for SOLiD Find splice junctions using RNA-seq data       S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$       data 9 and data 4: deletions         Keplicate 2       Add replicate:       S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$       data 9 and data 4: deletions         express Quantify the abundances of a set of target sequences from sampled subsequences       ast of target sequences from sampled subsequences       data 0 and data 4: insertions       data 9 and data 4: insertions         Add new Replicate 2       Add new Replicate 2       Add new Replicate 2       data 9 and data 4: insertions         Add new Replicate       Add new Replicate 4       data 9 and data 4: insertions       data 7	DESeq2 Differential gene				46: Tophat for Illumina on 👁 🖉 💥
the negative binomial distribution       Replicates       45: Tophat for Illumina on (*) (*) (*) (*) (*) (*) (*) (*) (*) (*)	expression analysis based on Untreated				data 8 and data 4: deletions
Trinity De novo assembly of RNA-Seq data Using Trinity     Replicate 1     43: 10phat for Illumina on @ 0 & data 4: insertions       RNA-SEQ     Add replicate:     44: Tophat for Illumina on @ 0 & data 4: accepted_hits \$       Tophat for Illumina Find splice junctions using RNA-seq data     Replicate 1     43: Tophat for Illumina on @ 0 & data 4: accepted_hits \$       Tophat for SOLiD Find splice junctions using RNA-seq data     Replicate 2     43: Tophat for Illumina on @ 0 & data 4: accepted_hits \$       Tophat for SOLiD Find splice junctions using RNA-seq data     S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$     42: Tophat for Illumina on @ 0 & data 4: deletions       Replicate 2     Add new Replicate 2     42: Tophat for Illumina on @ 0 & data 4: deletions     42: Tophat for Illumina on @ 0 & data 4: data 4: deletions       eXpress Quantify the abundances of a set of target sequences from sampled subsequences     Add new Replicate 2     4dd new Replicate 2       Add new Replicate     Add new Replicate 0     11: FASTO Groomer on @ 0 & data 7	the negative binomial Replicates				45: Tankat far Illuming on @ 0.
RNA-Seq data Using Trinity       Add replicate:       52: Tophat for Illumina on data 10 and data 4: accepted_hits \$         Tophat for Illumina Find splice       Remove Replicate 1       44: Tophat for Illumina on @ 0 & data 9 and data 4: accepted hits         Tophat for Illumina Find splice       Replicate 2       43: Tophat for Illumina on @ 0 & data 9 and data 4: accepted_hits \$         Tophat for SOLID Find splice       S6: Tophat for Illumina on data 11 and data 4: accepted_hits \$       42: Tophat for Illumina on @ 0 & data 9 and data 4: deletions         express Quantify the abundances of a set of target sequences from sampled subsequences       Add new Replicate 2       42: Tophat for Illumina on @ 0 & data 2 and data 4: insertions         Add new Replicate 1       Add new Replicate 2       6       42: Tophat for Illumina on @ 0 & data 4: insertions	Tripity Do novo assembly of Replicate 1				data 8 and data 4: insertions
RNA-SEQ       52: Tophat for Illumina on data 10 and data 4: accepted_hits ‡       44: Tophat for Illumina on @ 0 & data 9 and data 4: accepted_hits         Tophat for Illumina Find splice junctions using RNA-seq data       Remove Replicate 1       43: Tophat for Illumina on @ 0 & data 9 and data 4: accepted_hits         Tophat for SOLID Find splice junctions using RNA-seq data       Replicate 2       42: Tophat for Illumina on @ 0 & data 9 and data 4: accepted_hits ‡         Tophat for SOLID Find splice junctions using RNA-seq data       Express Quantify the abundances of a set of target sequences from sampled subsequences       56: Tophat for Illumina on @ 0 & data 4: insertions         Add new Replicate 2       Add new Replicate 2       4dit new Replicate 2       4dit new Replicate 2         Add new Replicate 2       Add new Replicate 2       4dit new Replicate 2       4dit new Replicate 2	RNA-Seq data Using Trinity Add replicate:				
Tophat for Illumina Find splice junctions using RNA-seq data       Remove Replicate 1       43: Tophat for Illumina on @ 0 & & data 9 and data 4: splice junctions         Tophat for RNA-seq data       Replicate 2       43: Tophat for Illumina on @ 0 & & data 9 and data 4: splice junctions         Tophat for SOLiD Find splice junctions using RNA-seq data       Add replicate:       56: Tophat for Illumina on data 11 and data 4: accepted_hits \$         express Quantify the abundances of a set of target sequences from sampled subsequences       Remove Replicate 2       41: Tophat for Illumina on @ 0 & & data 9 and data 4: insertions         Add new Replicate 2       Add new Replicate 2       60 & & 0	RNA-SEO 52: Tophat for	Illumina on data 10 and data 4:	accepted_hits 🗘		44: Tophat for Illumina on <ul> <li>44: Tophat for Illumina on</li> <li>44: A second data 4: accented hits</li> </ul>
Implation	Tophat for Illumina Find collins	ate 1			and y und dute in accepted into
Tophat2 Gapped-read mapper       Replicate 2       Add replicate       42: Tophat for Illumina on @ 0 & data 9 and data 4: spice junctions         Tophat for SOLiD Find splice       Junctions using RNA-seq data       56: Tophat for Illumina on data 11 and data 4: accepted_hits \$       42: Tophat for Illumina on @ 0 & data 9 and data 4: deletions         eXpress Quantify the abundances of a set of target sequences from sampled subsequences       Remove Replicate 2       Add new Replicate       41: Tophat for Illumina on @ 0 & data 9 and data 4: insertions	junctions using RNA-seq data				43: Tophat for Illumina on
for RNA-seq data       42: Tophat for Illumina on @ 0 & data 9 and data 4: deletions         Tophat for SOLiD Find splice       56: Tophat for Illumina on data 11 and data 4: accepted_hits \$         junctions using RNA-seq data       6: Tophat for Illumina on data 11 and data 4: accepted_hits \$         eXpress Quantify the abundances of a set of target sequences from sampled subsequences       Add new Replicate         Add new Replicate       Add new Replicate	Tophat2 Gapped-read mapper Replicate 2				uala 9 and data 4: splice junctions
Tophat for SOLID Find splice       56: Tophat for Illumina on data 11 and data 4: accepted_hits \$       data 9 and data 4: deletions         gunctions using RNA-seq data       56: Tophat for Illumina on data 11 and data 4: accepted_hits \$       data 9 and data 4: deletions         eXpress Quantify the abundances of a set of target sequences from sampled subsequences       Add new Replicate 2       data 9 and data 4: insertions	for RNA-seq data				42: Tophat for Illumina on ④ Ø 💥
junctions using RNA-seq data <u>eXpress</u> Quantify the abundances of a set of target sequences from sampled subsequences	Tophat for SOLiD Find splice 56: Tophat for	Illumina on data 11 and data 4:	accepted_hits \$		data 9 and data 4: deletions
eXpress Quantify the abundances of a set of target sequences from sampled subsequences     data 9 and data 4: insertions       Add new Replicate     Add new Replicate	junctions using RNA-seq data	-			41: Tophat for Illumina on 👁 🖉 💥
abundances of a set of target sequences from sampled subsequences Add new Replicate	eXpress Quantify the Remove Replic	ate 2			data 9 and data 4: insertions
Add new Replicate	abundances of a set of target				11: FASTO Groomer on @ / 🕱
	subsequences Add new Replica	e			data 7
	FILTERING	1			10: EASTO Groomer on O
Add new Condition	Add new Condition				

# Cuffdiff Output

- TSS... files report on Transcription Start Sites
- splicing... report on splicing
- CDS... track coding region expression
- transcript... track transcripts
- gene... rolls up the transcripts into their genes
  - gene/transcript FPKM tracking: gives information about the gene/ transcript (length, nearest ref id, TSS, etc) and the confidence intervals for FPKM for each condition.
  - gene/transcript differential expression testing: gives the expression change between groups, a status of whether there was enough data for that value to be accurate (OK is good, FAIL and NOTEST are bad. LOWDATA is somewhere in between). Finally, it gives a p-value.
  - see more details... Link

# Find Significant Changes

• Cuffdiff find significant changes in transcript expression

=_ Galaxy		Analyz	<b>Data</b> Workflow	Shared Data <del>-</del> V	Visualization 😽 🛛 H	lelp → User →			Using 1%
Tools	test_id	gene_id	gene	locus	sample_1	sample_2	status	value	History 2 ¢
Motif Tools	AT1G01010.1	AT1G01010	ANAC001	Chr1:3630-5899	Treated	Untreated	OK	71.3	
Multiple Alignments	AT1G01020.1	AT1G01020	ARV1	Chr1:5927-8737	Treated	Untreated	NOTEST	23.08	73: Cuffdiff on data 48,
Metagenomic analyses	AT1G01020.2	AT1G01020	ARV1	Chr1:5927-8737	Treated	Untreated	OK	295.5	Gata 44, and others: transcript
FASTA manipulation	AT1G01030.1	AT1G01030	NGA3	Chr1:11648-13714	Treated	Untreated	NOTEST	4.011	
NGS: OC and manipulation	AT1G01040.1	AT1G01040	DCL1	Chr1:23145-33153	Treated	Untreated	OK	41.64	72: Cuffdiff on data 48, 💿 🖉 💥
NGS: Mapping	AT1G01040.2	AT1G01040	DCL1	Chr1:23145-33153	Treated	Untreated	ОК	167.0	data 44, and others: transcript
NGS: Indel Analysis	AT1G01046.1	AT1G01046	MIR838A	Chr1:23145-33153	Treated	Untreated	OK	643.3	41.622 lines
NGS: RNA Analysis	AT1G01050.1	AT1G01050	AtPPa1	Chr1:23145-33153	Treated	Untreated	OK	1028	format: tabular. database: ?
Cuffdiff find singlificant shapped	AT1G01060.1	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	OK	65.49	Log: tool progress cuffdiff v2.1.1
in transcript expression	AT1G01060.2	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	NOTEST		(4046M) Log: tool progress
splicing, and promoter use	AT1G01060.3	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	OK	53.52	[18:33:49] Loading reference
	AT1G01060.4	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	OK	91.48	annotation. [18:33:54] Inspecting
DESeq Determines differentially	AT1G01060.5	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	NOTEST		maps and determining fragment
alignments	AT1G01070.1	AT1G01070	AT1G01070	Chr1:38751-40944	Treated	Untreated	OK	12.9	Modeling fragment count
angiments	AT1G01070.2	AT1G01070	AT1G01070	Chr1:38751-40944	Treated	Untreated	NOTEST	0.002298	overdispersion. [18:36:19] Modeling
DEXSeq Inference of differential	AT1G01073.1	AT1G01073	AT1G01073	Chr1:44676-44787	Treated	Untreated	NOTEST		fragme
exon usage in RNA-Seq	AT1G01080.1	AT1G01080	AT1G01080	Chr1:45295-47019	Treated	Untreated	OK	644.5	🔲 🛈 🚵 📥 🛛 🖉 🖻
DESeq2 Differential gene	AT1G01080.2	AT1G01080	AT1G01080	Chr1:45295-47019	Treated	Untreated	NOTEST	3.854	
expression analysis based on	AT1G01090.1	AT1G01090	PDH-E1 ALPHA	Chr1:47484-49286	Treated	Untreated	OK	3101	1 2 3 4
the negative binomial	AT1G01100.1	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK	1277	test_id gene_id gene locus
distribution	AT1G01100.2	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK	75.39	AT1G01010.1 AT1G01010 ANAC001 Chr1:3
Trinity De novo assembly of	AT1G01100.3	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	ок		AT1G01020.1 AT1G01020 ARV1 Chr1:5
RNA-Seq data Using Trinity	AT1G01100.4	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	ок	1547	AT1G01020.2 AT1G01020 ARV1 Chr1:5
RNA-SEQ	AT1G01110.1	AT1G01110	IOD18	Chr1:52238-54692	Treated	Untreated	OK	212.6	AT1G01030.1 AT1G01030 NGA3 Chr1:1
Tankat for Illumina Find calles	AT1G01110.2	AT1G01110	IOD18	Chr1:52238-54692	Treated	Untreated	OK	577.0	AT1G01040.1 AT1G01040 DCL1 Chr1:2
junctions using RNA-seq data	AT1G01115.1	AT1G01115	AT1G01115	Chr1:56623-56740	Treated	Untreated	NOTEST		
Tankat2 Canad and many	AT1G01120.1	AT1G01120	KCS1	Chr1:57268-59167	Treated	Untreated	OK	821.7	71 Cuffdiff on data 48 A 7 5
<u>Topnatz</u> Gapped-read mapper	AT1G01130.1	AT1G01130	AT1G01130	Chr1:61904-63811	Treated	Untreated	OK	1695	data 44 and others: gene EPKM
Tor KinA-seq uata	AT1G01140.1	AT1G01140	CIPK9	Chr1:64165-67625	Treated	Untreated	OK	2756	tracking
Tophat for SOLID Find splice	AT1G01140.2	AT1G01140	СІРК9	Chr1:64165-67625	Treated	Untreated	NOTEST	1.754	
junctions using RNA-seq data	AT1G01140.3	AT1G01140	СІРК9	Chr1:64165-67625	Treated	Untreated	NOTEST		70: Cuffdiff on data 48, 💿 🖉 💥
eXpress Quantify the	AT1G01150.1	AT1C01150	AT1G01150	Chr1:70114-72138	Treated	Untreated	ОК	49.53	data 44, and others: gene
abundances of a set of target	AT1G01160.1	AT1C01160	GIF2	Chr1:72338-74737	Treated	Untreated	ОК	3057	differential expression testing
sequences from sampled	AT1G01160.2	AT1C01160	GIF2	Chr1:72338-74737	Treated	Untreated	OK		69: Cuffdiff on data 48 @ 0 %
subsequences	AT1G01170.1	AT1G01170	AT1G01170	Chr1:72338-74737	Treated	Untreated	OK		data 44, and others: TSS groups
FILTERING	AT1G01170.2	AT1G01170	AT1G01170	Chr1:72338-74737	Treated	Untreated	OK	5C	FPKM tracking
Filter Combined Transcripts	AT1G01180.1	AT1G01180	AT1G01180	Chr1:75582-76758	Treated	Untreated	OK	91.99	
using tracking file	AT1G01183.1	AT1G01183	MIR165A	Chr1:78931-79032	Treated	Untreated	NOTEST	1.5.5.000	b8: Cuttdiff on data 48,
NGS: SAM Tools	AT1G01190.1	AT1G01190	CYP78A8	Chr1:83044-84864	Treated	Untreated	ок	39.14	differential expression testing

# Output Files from Galaxy

- SNP-seq
  - Save Bam file BWA generated
  - Save .bai file (index of BAM) BWA generated
  - Save vcf file Samtools mpileup generated
  - Already saved them at:
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/</u>
- RNA-seq
  - Save four Bam files and four .bai files
  - Already saved them at:
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/</u>

# Downloading bam index file (.bai)

💳 Galaxy		Analyze	Data Workflow	Shared Data 😽 👌	Visualization <del>-</del>	Help <del>+</del> User <del>+</del>			Using	1%
Tools	test_id	gene_id	gene	locus	sample_	1 sample_2	status	value	History	•
Graph/Display Data	AT1G01010.1	AT1G01010	ANAC001	Chr1:3630-5899	Treated	Untreated	ОК	71.3	group unstang	
Regional Variation	AT1G01020.1	AT1G01020	ARV1	Chr1:5927-8737	Treated	Untreated	NOTEST	23.08	60: Cuffdiff on data 48, 💿 🖟	2 🐹
Multiple regression	AT1G01020.2	AT1G01020	ARV1	Chr1:5927-8737	Treated	Untreated	OK	295.5	data 44, and others: genes read	
Multivariate Analysis	AT1G01030.1	AT1G01030	NGA3	Chr1:11648-13714	4 Treated	Untreated	NOTEST	4.011	group tracking	
Evolution	AT1G01040.1	AT1G01040	DCL1	Chr1:23145-33153	3 Treated	Untreated	OK	41.64	59: Cuffdiff on data 48.	8
Evolution Matif Table	AT1G01040.2	AT1G01040	DCL1	Chr1:23145-33153	3 Treated	Untreated	OK	167.0	data 44, and others: isoforms r	ead
MOTIFIOOIS	AT1G01046.1	AT1G01046	MIR838A	Chr1:23145-33153	3 Treated	Untreated	OK	643.3	group tracking	
Multiple Alignments	AT1G01050.1	AT1G01050	AtPPa1	Chr1:23145-33153	3 Treated	Untreated	ОК	1028		
Metagenomic analyses	AT1G01060.1	AT1G01060	LHY	Chr1:33378-37871	1 Treated	Untreated	ОК	65.49	56: Tophat for Illumina on @ 6	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
FASTA manipulation	AT1G01060.2	AT1G01060	LHY	Chr1:33378-37871	1 Treated	Untreated	NOTEST		15 6 MD	<u>IIS</u>
NGS: QC and manipulation	AT1G01060.3	AT1G01060	LHY	Chr1:33378-37871	1 Treated	Untreated	OK	53.52	format: ham database: ?	
NGS: Mapping	AT1G01060.4	AT1G01060	LHY	Chr1:33378-37871	Treated	Untreated	OK	91.48	TopHat v2 0.9 Settings: Output fi	les:
NGS: Indel Analysis	AT1G01060.5	AT1G01060	LHY	Chr1:33378-37871	1 Treated	Untreated	NOTEST		"/tmp/tmpUBe3bw/dataset_2455	.*.e
NGS: RNA Analysis	AT1G01070.1	AT1G01070	AT1G01070	Chr1:38751-40944	4 Treated	Untreated	OK	12.9	bwt" Line rate: 6 (line is 64 bytes)	
Cuffdiff find significant changes	AT1G01070.2	AT1G01070	AT1G01070	Chr1:38751-40944	4 Treated	Untreated	NOTEST	0.002298	Lines per side: 1 (side is 64 bytes	)
in transcript expression,	AT1G01073.1	AT1G01073	AT1G01073	Chr1:44676-44787	7 Treated	Untreated	NOTEST		Offset rate: 5 (one in 32) FTable	
splicing, and promoter use	AT1G01080.1	AT1G01080	AT1G01080	Chr1:45295-47019	Treated	Untreated	ОК	644.5	chars: 10 Strings: unpacked Max	
DESeg Determines differentially	AT1G01080.2	AT1G01080	AT1G01080	Chr1:45295-47019	Treated	Untreated	NOTEST	3.854		
expressed transcripts from read	AT1G01090.1	AT1G01090	PDH-E1 ALPHA	Chr1:47484-49286	5 Treated	Untreated	OK	3101		
alignments	AT1G01100.1	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK	1277	Download Dataset	
DEVSec Inference of differential	AT1G01100.2	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK	75.39	ADDITIONAL FILES	
exon usage in RNA-Seg	AT1G01100.3	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK		Download bam_index	
	AT1G01100.4	AT1G01100	AT1G01100	Chr1:50074-51199	Treated	Untreated	OK	1547	55: Tophat for Illumina on 👁 🕼	1 22
DESeq2 Differential gene	AT1G01110.1	AT1G01110	IQD18	Chr1:52238-54692	2 Treated	Untreated	ок	212.6	data 11 and data 4: splice	
the negative binomial	AT1G01110.2	AT1G01110	IQD18	Chr1:52238-54692	2 Treated	Untreated	ок	577.C	junctions	
distribution	AT1G01115.1	AT1G01115	AT1G01115	Chr1:56623-56740	D Treated	Untreated	NOTEST			
Trinity Do nove accombly of	AT1G01120.1	AT1G01120	KCS1	Chr1:57268-59167	7 Treated	Untreated	OK	821.7	54: Tophat for Illumina on @ 6	~ 23
RNA-Seg data Using Trinity	AT1G01130.1	AT1G01130	AT1G01130	Chr1:61904-63811	1 Treated	Untreated	OK	1695	data 11 and data 4. deletions	
nov see oute osnig minty	AT1G01140.1	AT1G01140	CIPK9	Chr1:64165-67625	5 Treated	Untreated	OK	2756	53: Tophat for Illumina on @ 6	2
RNA-SEQ	AT1G01140.2	AT1G01140	CIPK9	Chr1:64165-67625	5 Treated	Untreated	NOTEST	1.754	data 11 and data 4: insertions	
Tophat for Illumina Find splice	AT1G01140.3	AT1G01140	CIPK9	Chr1:64165-67625	5 Treated	Untreated	NOTEST			
junctions using RNA-seq data	AT1G01150.1	AT1G01150	AT1G01150	Chr1:70114-72138	3 Treated	Untreated	ОК	49.53	52: Tophat for Illumina on @ (	×
Tophat2 Gapped-read mapper	AT1G01160.1	AT1G01160	GIF2	Chr1:72338-74737	7 Treated	Untreated	ОК	3057	uata 10 anu uata 4: accepted_n	15

# Outline

- What is Galaxy
- Galaxy for Bioinformaticians
- Galaxy for Experimental Biologists
- Using Galaxy for NGS Analysis
- NGS Data Visualization and Exploration Using IGV

# Why IGV

- IGV is an integrated visualization tool of large data types
- View large dataset easily
- Faster navigation on browsing
- Run it locally on your computer
- Easy to use interface



http://www.broadinstitute.org/igv/

# IGV download



		Spe	cify genor	nic re	gion				Zo	om in	
A. thaliana (TAIR 10	) 🛟 Chr1	¢ Cł	nr1:42,220-52,077	G	• 着 🖣	• 🏟	• 🖪 🗙 🗖				
	43,000	bp 44 I	4,000 bp 45,000 bp 	46,00 	0 bp 47	9,839 bp ,000 bp I	48,000 bp	49,000 bp	50,000 bp	51,000 bp	Ruler
RR064154.bam Coverage	[0 - 835]									¥ .	0
SRR064154.bam											× ×
RR064155.bam Coverage	[0 - 712]			-				terration and attacks a	<u></u>	1.	0
RR064155.bam						I					×.
RR064166.bam Coverage	[0 - 843]										٥
SRR064166.bam											Track
RR064167.bam Coverage	[0 - 1251]										0
3RR064167.bam											A
Gene			AT1G01073.1	AT10	G01080.1	<	AT1G0	< <mark> &lt;</mark> 1090.1	AT1G011	00.2	
xpression_diff.gff3				AT1	IG01080		PDH-E1	ALPHA	AT1G01	100	

# Load data

Select genome: Click the genome drop-down list in the toolbar and select the genome

### Select chromosome

000			IGV		
A. thaliana (TAIR 10)	\$ All \$		Go 👚 🔺 🕨	🧼 🔳 🛪 🏳	-
A. thaliana (TAIR 10)	All     Chr1	Selected geno the genome d Filter: Human hg19 Human (1kg, t Human (1kg, t Human (1kg, t Human hg18 Human (1kg, t Human hg17 Human hg16 Mouse mm10 Mouse mm10 Mouse mm7 Mouse (12951 Chimp (panTro Rhesus (rheMa Macaca fascicu Rat (rn4) Rat (rn5) Dog (canFam2 Dog (canFam2 Dog (canFam2 Cow (bosTau7	Go Go Go Go Construction of the second secon		Chr4 Chr5
Gene		i hans and the History of Strain Strain Strain Strain Strain Strain	a Ball de la fata de la Constancia de la c	and the second second second	er besen linde held. Here the book of social and a second second second second second second second second second
					758M of 864M

# Load data files

• Load from URL, file, server

File Genomes V	iew	Tra	cks Reg	ions	Tools	Genon	neSpace	Help									
Load from File Load from URL Load from Server Load from DAS	•		Chr1	•	Chr1			C	io 👚	•	ġ.	7 🗖	X				
New Session Open Session Save Session		- <b>-</b>		1			10 mb 		30 mb –		4	20 mb 			Ĕ		₽- 30 mi 
Save Image																	
Exit																	
Gene																	
		1010.1	AT1G0874	0 AT1G1	5560 A	T1G22580	AT1G30030	AT1G35663	AT1G4207	0 AT1G	49440	AT1G56	080.1	AT1G6418	0.1 AT	G70820.1	AT1G78095
	Chr1:	29,694,0	54													418	M of 761M

# Toolbar

- Genome drop-down box: loads a genome
- Chromosome drop-down box: zooms to a chromosome
- Search box: Displays the chromosome location being shown. To scroll to a different location, enter the gene name, locus or track name and click Go.
- Whole genome view: Zooms to whole genome view.
- Define a region: Defines a region of interest on the chromosome.
- Zoom slider: Zooms in and out on a chromosome.

# Change Display Options

- IGV offers several display options for tracks
- Zoom in and Zoom out
- Modify Track Height
- Sort the Tracks
- Filter the Tracks
- Group the Tracks
- Sort Tracks based on Region of Interest

# Variants Visualization in IGV

- Load TAIR10 genome to IGV
- Load BAM file "SRR038850.bam" to IGV with "Load from URL"
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/SAM-to-BAM\_BAM.bam</u>
- Load VCF file "var.raw.vcf" to IGV with "Load from URL"
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/var.raw.vcf</u>

# Zoom In Screen

## Zoom in to : Chr5:57,073-57,142

A. thaliana (TAIR 10)	Chr5       Chr5:57,073-57,142       Go
	57,080 bp         57,100 bp         57,110 bp         57,120 bp         57,130 bp         57,140 bp
var.raw.vcf /space/g2mainm_input_0.bam	
SAM-to-BAM_BAM.bam Coverage	[0 - 10.00]
SAM-to-BAM_BAM.bam	
Sequence 🔿 Gene	A A A G C C C A C A T T A A T G G C A T T G A T T G C G T T G A G G T C A A T T G A T G G A A A A A T A A T C A C G T C A C G T A A A A A G F T C
5 tracks loaded Chr5:	:57,088 543M of 681M

# Variants Visualization in IGV

- Load TAIR10 genome to IGV
- Load BAM file "SRR064154.bam" to IGV with "Load from URL"
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/SRR064154.fastq.bam</u>
- Load VCF file "var.raw.vcf" to IGV with "Load from URL"
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Snpseq/gatk.vcf</u>

# Zoom in position (chr5:6,435-6,475)



# Zoom in position (chr5:6,435-6,475)



# Right click on track for more options



# Show all bases in IGV

00				IGV			
A. thaliana (TAIR 10	) ‡	Chr5	¢ Chr5:6,435-6,475	Go 😭	i 🔹 🕨 🗖	] 🗙 🏳	+
gatk.vcf SM	NAME DATA TYDE		6,440 bp 	6,450 bp I	41 bp	6,460 bp 	€,470 bp
SRR064154.fastq.bam Covera	ge	[0-104]         T       G         T	A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       A       T       G       C       A       G         A       A       T       G       C       A       G       G         A       A       T       G       C       A       G         A	C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A       G       C         C       T       C       C       A <t< td=""><td>T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G</td><td>A       G       T       T       G       C         A       G       T       T       G       C         T       G       C       T       G       C         T       G       C       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G</td><td>A C A A G G C A A G G G A T C A C A A G G G A T C A G A T C A</td></t<>	T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G       T         T       G       C       A       G	A       G       T       T       G       C         A       G       T       T       G       C         T       G       C       T       G       C         T       G       C       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G       C         A       G       T       T       G	A C A A G G C A A G G G A T C A C A A G G G A T C A G A T C A
Sequence • Gene	•	T G C	AAATGCAG L H L	CTCCAGC EL	T G T A G T Q L AT5G01020.1	AGTTGC/	A C A A G G G A T C A
5 tracks C	hr5:6,46	4					506M of 1,033M

# **RNA-seq Results Visualization**

- Load four BAM files to IGV
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/SRR064154.bam</u>
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/SRR064155.bam</u>
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/SRR064166.bam</u>
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseg/SRR064167.bam</u>
- Load gene differential expression GFF3 file "expression\_diff.gff3" to IGV
  - <u>http://biocluster.ucr.edu/~rkaundal/Galaxy\_workshop/Rnaseq/expression\_diff.gff3</u>

# Exercise 2: RNA-seq Result Visualization

## Zoom in to Chr1:41,351-51,208

000				IGV			
A. thaliana (TAIR 10)	*	Chr1	¢ Chr1:41,946-51,803	Go 런 🖣	► 🛷	🔲 💥 🖵	+
		-					
		-					
		-4			837 bn —		
	NAME DATA TYPI DATA FILE	00 bp 4	3,000 bp 44,000 bp	45,000 bp 46,000 bp	47,000 bp	48,000 bp 49,000 bp	50,000 bp 51,000 bp
SRR064154.bam Coverage		[0 - 835]				at an	
SRR064154.bam							
SRR064155.bam Coverage		[0 - 712]				an tar Anna tar a tar	
SRR064155.bam							
SRR064166.bam Coverage		[0 - 843]				a to do define and the address of the same star	
SRR064166.bam							
SRR064167.bam Coverage		[0 - 1251]					
SRR064167.bam							
Gene			AT1G01	073.1 AT1G01080.1	<	AT1G01090.1	AT1G01100.2
expression_diff.gff3				AT1G01080		PDH-E1 ALPHA	AT1G01100
11 tracks loaded Chr1	1:51,577						619M of 987M

# Exercise 2: RNA-seq Result Visualization

## Zoom in Chr1:49,457-51,457



# UCR Galaxy homepage (https://galaxy.bioinfo.ucr.edu)

### 🗧 Galaxy

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
<u>Statistics</u>
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: Mapping
NGS: Indel Analysis
NGS: RNA Analysis
<

Institute for Integrative
Genome Biology

### Welcome to IIGB's Galaxy Server!

### Overview

Galaxy is an open, highly customizable, web-based platform for the analysis of next generation sequence data and many other biological data types. It enables users to run computationally demanding next generation sequencing analysis tasks on powerful server hardware from a graphical web browser-based user interface rather than the Linux command-line. A subset of of application supported by Galaxy is given in the left pane. Much more detailed descriptions of Galaxy's basic functionalities including user tutorials are available <u>here</u>.

### Why Local Galaxy Service?

There are many advantages of using a local Galaxy server here at UCR rather than public test instances of Galaxy available on the internet. The most important are: (1) shorter waiting queues for analysis tasks; (2) elimination of time consuming uploads of large data sets; (3) support for analyzing much larger data sets than this is possible on public services; (4) the ability to customize software tools and database collections.

### How to Gain Access?

This instance of Galaxy runs on IIGB's high performance compute (HPC) infrastructure, called Biocluster. As such its usage is covered by the annual registration fee for this infrastructure (see <u>here</u> for details). Users with an active Biocluster account can access this Galaxy service using their existing user name and password without any extra cost. New account requests for this service can be sent to <u>support@biocluster.ucr.edu</u>.

### Additional Databases and Sofware Tools

Support requests for including additional reference genomes and software tools on IIGB's Galaxy server can be sent to <a href="mailto:support@biocluster.ucr.edu">support@biocluster.ucr.edu</a>

### Workshops on Galaxy

Past and future UCR workshop events on using Galaxy are listed <u>here</u>. The user manual from previous workshops can be accessed <u>here</u>.

Enter IIGB's Galaxy Service To enter this service, click here.

# **Thank You**