Galaxy Project Update

ISMB / ECCB 2015

Dublin, Ireland

Dave Clements and the Galaxy Team Johns Hopkins University





Multi-Sample Analysis

Need to scale to support current and coming experimental designs:

500 specimens or 20 different conditions with 4 replicates each or

 $\bullet \bullet \bullet$

Multi-Sample Analysis: Workflows



Correctly Paired Reads Incorrectly Paired / Unpaired Reads

Multi-Sample: Multiple dataset selection within a tool

FastQC Read Quality reports (Galaxy Tool Version 0.63) Services Versions											
Short read data from your current history											
1: MeOH_REP1_R1.fastq	□ 1: MeOH_REP1_R1.fastq										
Multiple datasets											
🗅 🖄 🗅 No selection		-									
tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA											
Submodule and Limit specifing file											
□ □ No selection		•									
a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter											
✓ Execute											

FastQC Read Quality reports (Galaxy Tool Version 0.63) 🗞 Versions

Options

Short read data from your current history



- 12: R3G_REP3_R2.fastq 11: R3G_REP3_R1.fastq 10: R3G_REP2_R2.fastq 9: R3G_REP2_R1.fastq 8: P3C_PEP1_P2_fastq
- 8: R3G REP1 R2.fasto

This is a batch mode input field. A separate job will be triggered for each dataset.

Contaminant list

 \square

 \square



No selection

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Submodule and Limit specifing file



No selection

a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter



ረግ

-

History	C	\$	
search datasets			8
Cam RNA-Seq Day 2 Test 8 shown			
189.3 MB		۲	•
<u>8: genes_chr12.Q</u> perations data	on n sets	nultij	ple
<u>7: chr12.fa</u>	۲		×
<u>6: R3G_REP3 Mapped & F</u> iltered	۲		×
<u>5: R3G_REP2 Mapped & F</u> iltered	۲		×
<u>4: R3G_REP1 Mapped & F</u> iltered	۲	(×
<u>3: MeOH_REP3 Mapped &</u> <u>Filtered</u>	۲	A	×
<u>2: MeOH_REP2 Mapped &</u> Filtered	۲		×
<u>1: MeOH_REP1 Mapped &</u> Filtered	۲		×

History	C	₽	History	2 🌣 🗆					
search datasets			search datasets	8					
Cam RNA-Seq Day 2 Test 8 shown			Cam RNA-Seq Day 2 Test 8 shown						
189.3 MB		۲	189.3 MB						
8: genes chr1Operationsdata	on m sets	ulti	All None For a	l selected					
<u>7: chr12.fa</u>	۲		8: genes_chr12.gtf						
<u>6: R3G_REP3 Mapped & F</u>	۲		7: chr12.fa						
iltered			6: R3G_REP3 Mapped & Filtered						
<u>5: R3G_REP2_Mapped & F</u> iltered	۲	(5: R3G_REP2 Mapped	<u>& Filtered</u>					
4: R3G REP1 Mapped & F	۲		4: R3G_REP1 Mapped	<u>& Filtered</u>					
3: MeOH REP3 Mapped &			3: MeOH_REP3 Mappe	d & Filtered					
Filtered		0*	2: MeOH_REP2 Mappe	d & Filtered					
2: MeOH_REP2 Mapped & Filtered	۲	ø	I: MeOH_REP1 Mappe	<u>d & Filtered</u>					
<u>1: MeOH_REP1 Mapped &</u> Filtered	۲		×						

History	₽ \$	History	C 🕈	Hist	ory	4	C 🌣 🗆		
search datasets		search datasets		search datasets					
Cam RNA-Seq Day 2 Test 8 shown	:	Cam RNA-Seq Day 2 Test 8 shown			RNA-Sec	q Day 2 Test			
189.3 MB		189.3 MB			189.3 MB				
<u>8: genes_chr1</u> 2_Operations	on multi	All None For all selected			All None For all select				
<u>7: chr12.fa</u>	(*)	<u>8: genes_chr12.</u>	<u>gtf</u>		Hide da Unhide	atasets e datasets			
<u>6: R3G_REP3 Mapped & F</u>	۲	7: chr12.fa			Delete	datasets			
<u>iltered</u>		C 6: R3G REP3 Ma	ipped & Filtere		Undele	te datasets			
<u>5: R3G_REP2 Mapped & F</u> iltered	•	<u>5: R3G_REP2_Ma</u>	ipped & Filtere		Perman Build D	nently delete d Dataset List	atasets		
4: R3G_REP1 Mapped & F	۲	<u>4: R3G REP1 Ma</u>	pped & Filtere		Build D	ataset Pair			
<u>iltered</u>		3: MeOH_REP3	Mapped & Filte		Build L	ist of Dataset I	Pairs		
<u>3: MeOH_REP3 Mapped &</u> <u>Filtered</u>	۲	<u>2: MeOH_REP2 M</u>	Mapped & Filte		2: MeOH I	REP2 Mapped	<u>& Filtered</u>		
2: MeOH_REP2_Mapped & Filtered	۲	✓ <u>1: MeOH_REP1 M</u>	Mapped & Filte		L: MeOH_F	REP1 Mapped	<u>& Filtered</u>		
<u>1: MeOH_REP1 Mapped &</u> Filtered	،	×							

History	C 🕈	History	C 🕈	Hist	ory	S	History	3	¢ (
search datasets		search datasets		search datasets			search datasets		(8
Cam RNA-Seq Day 2 Test 8 shown	:	Cam RNA-Seq Day 2 8 shown	Test	Cam 8 sho	RNA-Seq Day	2 Test	Cam RNA-Seq Day 2 Test			
189.3 MB		189.3 MB	S	189.	3 MB		189.3 MB		•	•
<u>8: genes_chr1</u> 2_Operations data	on multi asets	All None	For all selected	All	None	For all sele	<u>10: R3G</u>			×
<u>7: chr12.fa</u>	۲	8: genes_chr12.gt	<u>tf</u>		Hide dataset	s	a list of datasets			
<u>6: R3G_REP3 Mapped & F</u>	۲	☐ <u>7: chr12.fa</u>			Unhide datas Delete datas	ets ets	<u>9: MeOH</u> a list of datasets			×
<u>iltered</u>		<u> </u>	ped & Filtere		Undelete dat	asets	<u>8: genes_chr12.gtf</u>	۲		×
<u>5: R3G_REP2 Mapped & F</u> iltered	۲	<u>5: R3G REP2 Map</u>	ped & Filtere		Permanently Build Dataset	delete data t List	<u>7: chr12.fa</u>	۲		×
4: R3G_REP1 Mapped & F	۲	<u>4: R3G_REP1 Map</u>	ped & Filtere		Build Dataset	t Pair	6: R3G REP3 Mapped & F			x
		3: MeOH_REP3 Ma	apped & Filte		Build List of	Dataset Pai	<u>iltered</u>		-	
<u>3: MeOH_REP3 Mapped &</u> Filtered	•	2: MeOH_REP2 Ma	apped & Filte		2: MeOH_REP2 M	Mapped &	<u>5: R3G_REP2 Mapped & F</u> iltered	۲		×
<u>2: MeOH_REP2 Mapped &</u> <u>Filtered</u>	۲	✓ <u>1: MeOH_REP1 Ma</u>	apped & Filte		L: MeOH_REP1 M	Mapped &	<u>4: R3G_REP1 Mapped & F</u> iltered	۲		×
1: MeOH_REP1 Mapped & Filtered	۲	×					3: MeOH_REP3 Mapped &	۲		×
							<u>Filtered</u>			
							2: MeOH_REP2 Mapped & Filtered	۲		×
							<u>1: MeOH_REP1 Mapped &</u> Filtered	۲		×

StringTie transcript assembly and quantification (Galaxy 7 Version 1.0.3)	Tool Options
Version 1.0.3) Mapped reads to assemble transcripts from 3: MeOH_REP3 Mapped & Filtered Use Dataset collection assembly Use GFF Reference annotation to use for guiding the assembly 8: genes_chr12.gtf -G Perform abundance estimation only of input transcrip Yes No -e Output additional files for use in Ballgown Yes No -b	StringTie transcript assembly and quantification (Galaxy Tool Version 1.0.3) Mapped reads to assemble transcripts from Image: Construction of the system o
Options Use defaults	Yes No -e
Job Resource Parameters Use default job resource parameters	Output additional files for use in Ballgown Yes No -b Options
	Use defaults 🗸



User Interface: Scratchbook



8

Enable/Disable Scratchbook

search datasets

ESHG 1: Through BWA Mapping &

Merge w/ RG

User 🗸

← ⇒ C fi 🔒	https://test.galaxyp	project.org					Q 5⁄2	3 /	R	≡				
🗧 Galax	Galaxy Analyze			Shared Data -	Visualization -	Cloud -	Admin	Help -	wnk1	4	of 460	^ v] ×]	ΪB
Ŧ		S	gnificant			ŵ	Ŧ	genes	s_chr12.gtf	f 🖻)	C	‡ [П
1	2	3		4		5 🚺	T	•	gene_i				-/	
test_id	gene_id	gene		locus		samı	-		gene_i	d "NINJ2"; ge	abase	⇒ ha1	9	
XLOC_000005	XLOC_000005	WNK1		chr12:8620	88-1020664	MeO	-	•	gene_i	d "NINI2": ge	abast	. IIGI	<u> </u>	
XLOC_000013	XLOC_000013	FKBP4		chr12:2904	107-2914587	MeO		ว	gene_i	d "NINI2": ge	'ОК',			r
XLOC_000021	XLOC_000021	CCND2		chr12:4382	901-4414522	MeO		2	gene_i	d "NINI2": ge	0 val	id line	s	
XLOC_000023	XLOC_000023	RAD51AP1		chr12:4647	949-4669213	MeO		U	gene_i	d "NINI2": ge				
XLOC_000035	XLOC_000035	TAPBPL		chr12:6561	176-6579843	MeO	-		gene_i	d "NINI2": ge		٩	> •	
XLOC_000037	XLOC_000037	SCARNA10		chr12:6603	297-6641132	MeO	-	U	gene_i	d "NINI2": ac		2		
XLOC_000038	XLOC_000038	GAPDH		chr12:6643	584-6647537	MeO	-	•	gene_i	d "NINI2": ac		o dene		
XLOC_000039	XLOC_000039	COPS7A		chr12:6833	149-6841041	MeO	+		gene_i	d "WNK1": a	3001	u 00100	28877	78
XLOC_000042	XLOC_000042	CD4		chr12:6898	637-6929976	MeO	+		gene_i	d "WNK1": a	0003	CCDC77	20011	Ĭ
зпарії різріцу в	<u>//////</u>		1			ŕ	+		gene_i	d " <mark>WNK1</mark> ": a	0005	WNK1		
(†		Enough d	ata to make a	a call			+		gene i	d " <mark>WNK1</mark> "; q	0009	ADIPOR	2	
test_id	gene_id	gene		loc	us		+	0	gene i	d "WNK1"; q	0013	FKBP4		
XLOC_000001	XLOC_000001	LOC1002887	78	chi	12:87983-91263		+	0	gene i	d " <mark>WNK1</mark> "; g				
XLOC_000003	XLOC_000003	CCDC77		chi	12:389222-5518	06	e +	0	gene i	d " <mark>WNK1</mark> "; g	4,	•	∕ ×	
XLOC_000005	XLOC_000005	WNK1		chi	12:862088-1020	664	+	0	gene_i	d " <mark>WNK1</mark> "; g	<u>ran</u>			
XLOC_000009	XLOC_000009	ADIPOR2		chi	12:1800246-189	7845	+	2 • 3	gene_i	d " <mark>WNK1</mark> "; g	1			
XLOC_000013	XLOC_000013	FKBP4		chi	12:2904107-291	4587	+		gene_i	d " <mark>WNK1</mark> "; g	4,		<i>i</i> x	
XLOC_000014	XLOC_000014	ITFG2		chi	12:2921786-293	4237	+		gene_i	d " <mark>WNK1</mark> "; g	<u>ran</u>			
XLOC_000016	XLOC_000016	RHNO1		chi	12:2945981-299	8691	+		gene_i	d " <mark>WNK1</mark> "; g	<u>(pres</u>	sion t	estin	1
XLOC_000021	XLOC_000021	CCND2		chi	12:4382901-441	4522	+	0	gene_i	d " <mark>WNK1</mark> "; g				
XLOC_000023	XLOC_000023	RAD51AP1		chi	12:4647949-466	9213	+	0	gene_i	d " <mark>WNK1</mark> "; g	4,		<i>i</i> x	
XLOC_000025	XLOC_000025	NDUFA9		chi	12:4758263-479	6720		^		-L	100			

User Interface: Scratchbook



User Interface: Histories



0. 11-011

User Interface: Histories

DE. MAAC

Galaxy Analyze Data	Workflow Shared Data - Visualizati	on 👻 Cloud 👻	Admin Help - User -	Using	g 14.9 GB
Done search histories	search all datasets			Cr	reate new
Current History -	Switch to	•	Switch to		Switch to
Cam RNA-Seq Day 2 15 shown, 15 <u>hidden</u>	Cam ChIP-Seq 1 39 shown, 6 <u>hidden</u>		Copy of 'Cam RNA-Seq Day 2' (ac only)	tive items	Cam Ch 32 shown
194.2 MB Search datasets	476.9 MB search datasets		203.0 MB		497.7 M
Drag datasets here to copy them to the current history	39: Subtract on data 36 and data	• • ×	48: Venn Diagram on data 46 and		<u>32: H1h</u> eaks in
26: StringTie on collection 10: Coverage	<u>38: Pou5f1 peaks in both reps</u>	• / ×	<u>data 47</u> <u>47: Cut on data 42</u>	• • ×	<u>31: H1h</u> eaks in
25: StringTie on collection 10: Assemble	37: Concatenate datasets on data 33 and data 32	• / ×	46: Cut on data 44	 Image: A state <	<u>30: H1h</u> eaks in
a list of datasets	36: Nanog peaks in both reps	• # ×	44: Filter on data 42	• / ×	<u>29: H1h</u>
18: StringTie on collection 9: Coverage a list of datasets	35: Concatenate datasets on data 31 and data 30		42: Filter on data 38	• # ×	eaks in 28: MAC
17: StringTie on collection 9: Assembled x transcripts	34: Concatenate on data 31 an d data 30	• / ×	41: Cuffdiff on data 4, data 5, an d others: transcript FPKM trackin g		nd data
a list of datasets	33: H1hesc_Pou5f1_Rep2_chr12 P eaks in BED format	• / ×	40: Cuffdiff on data 4, data 5, an d others: transcript differential e	• / ×	nd data
a list of datasets	32: H1hesc_Pou5f1_Rep1_chr12 P eaks in BED format	• # ×	xpression testing		nd data

User Interface: Data Load

		Please wait.	11 out of 12 remaining.			
Name	Size	Туре	Genome	Settings	Status	
MeOH_REP1_R1.fastq	28.2 MB	Auto-det 🔻 🔾	Additional S 🔻	•	100%	•
MeOH_REP1_R2.fastq	28.2 MB	Auto-det 🔻 🔾	Additional S 🔻	•	79%	ं
□ MeOH_REP2_R1.fastq	28.9 MB	Auto-det 🔻 🔾	Additional S 🔻	\$	0%	ं
MeOH_REP2_R2.fastq	28.9 MB	Auto-det 🔻 🔾	Additional S 🔻	•	0%	ं
MeOH_REP3_R1.fastq	28.9 MB	Auto-det 🔻 🔾	Additional S 🔻	*	0%	ं
MeOH_REP3_R2.fastq	28.9 MB	Auto-det 🔻 🔾	Additional S 🔻	•	0%	ं
R3G_REP1_R1.fastq	23.5 MB	Auto-det 🔻 🔾	Additional S 🔻	•	0%	ं
💷 R3G REP1 R2.fastq	23.5 MB	Auto-det 💌 🛛	Additional S	•		<u> </u>

Tools: Multiple version support

axy	Analy	ze Data	Workflow	Shared Data -	Visualization -	Cloud 🗕 🖌	Admin Help -	User -		ک _ع ل
	1	پر Cuf	fdiff find sig	nificant changes	in transcript	🗞 Versior	ns 🔻 Options	Histo	ry	
mats pulation ures nces nents Analysis ression		expres Trans A tran Omit Yes Discar	cripts Cripts Cript GFF3 of Script GFF3 of Tabular Data No d the tabular	g, and promoter 27: Cuffmerge o or GTF file produ asets r output.	use (G Swi on data 7, data 8, ced by cufflinks, o	tch to 0.0.7 tch to 2.2.1.1 and others: me	erged 👻	47: Cdatane di46: Cidata !group45: Cidata !groupng	Cuffdiff o 5, and ot fferentia uffdiff on 5, and oth os FPKM t uffdiff on 5, and oth os differe	n data 4, hers: ge l expressi data 4, hers: TSS tracking data 4, hers: TSS ntial expr
<u>riation</u> ay Data		Yes Genera	No ate a SQLite data type	44: Cu data ! FPKM	uffdiff on 5, and oth tracking	<u>data 4,</u> 1ers: CDS				
<u>Fools (beta)</u> bly		SAM/ CuffNo not su	BAM orm supports	s either CXB (fror fault: SAM/BAM	n cuffquant) or SA	M/BAM input	• files. Mixing is	<u>43: Cu</u> data <u>!</u> FPKM g	uffdiff on 5, and oth different	<u>data 4,</u> <u>1ers: CDS</u> tial expres
<u>rosatellite Analysis</u> <u>cs</u>		Con	dition Condition					<u>42: Cu</u> data !	uffdiff on 5, and oth	<u>data 4.</u> hers: CDS
re Prediction		Na	ame					overle sting	bading di	ffential ex
<u>DWS</u>		Re	1eOH					<u>41: Cu</u> data !	uffdiff on 5, and oth	<u>data 4,</u> hers: pro

Tool: Options

axy	Analyze Data	Workflow	Shared Data -	Visualization -	Cloud	l √ Admin	Help -	User -		@ 3 U
	1	u ffdiff find sig	nificant changes	in transcript	& 1	/ersions 🔻 C	Options	Histo	ry	
mats pulation ures nces nents Analysis ression	expression Tran A tran Omi Yes Disca	ession, splicing scripts anscript GFF3 of t Tabular Data s No ard the tabular erate SQLite	g, and promoter of 27: Cuffmerge of or GTF file produc asets	use (Galaxy Tool Ve n data 7, data 8, an ced by cufflinks, cu		Question? Search Share Download Requirements See in Tool S	s	47: C data ne di 46: Cu data 5 group 45: Cu data 5 group ng	uffdiff or 5, and ot fferential offdiff on 5, and oth s FPKM t ffdiff on 5, and oth s differe	<u>1 data 4,</u> <u>hers: ge</u> <u>expressi</u> <u>data 4,</u> <u>ers: TSS</u> <u>racking</u> <u>data 4,</u> <u>ers: TSS</u> ntial expr
<u>ay Data</u>	Ye: Gene	Yes No Generate a SQLite database for use with cummeRbund.						<u>44: Cu</u> data 5 FPKM	uffdiff on 5, and oth tracking	<u>data 4,</u> iers: CDS
<u>Fools (beta)</u> bly	SAN Cuff not s	I/BAM Norm supports supported. Def	s either CXB (from ault: SAM/BAM	n cuffquant) or SAM	I/BAM	input files. Mi	▼ xing is	<u>43: Cu</u> data 5 FPKM g	uffdiff on 5, and oth different	<u>data 4,</u> Iers: CDS ial expres
rosatellite Analysis <u>cs</u> re Prediction		ondition Condition Name						42: Cu data 5 overic sting	uffdiff on 5, and oth bading di	data 4, Iers: CDS ffential ex
<u>DWS</u>		MeOH Replicates						<u>41: Cu</u> data 5	uffdiff on 5, and oth	data 4, iers: pro

Tools: New Forms Engine

ху	Analyz	ze Data	Workflow	Shared Data -	Visualization -	Cloud -	Admin	Help -	User -		@ 3	
	1	CuffN not su	orm support	s either CXB (fror fault: SAM/BAM	m cuffquant) or SA	M/BAM inpu	ut files. Mi	xing is	Histor	у		
<u>ats</u> ulation		Con	ndition						47: Cu data 5 ne dif	<u>iffdiff o</u> 5, and ot ferentia	<u>n data 4,</u> <u>hers: ge</u> l express	5
res ces		N N	ame MeOH						<u>46: Cut</u> data 5,	ffdiff on and oth	<u>data 4,</u> 1ers: TSS	
ation Vala		R	eplicates	5: R3G_REP2 Map 1: R3G_REP1 Map 3: MeOH_REP3 Ma 2: MeOH_REP2 Ma 1: MeOH_REP1 Ma	ped & Filtered ped & Filtered apped & Filtered apped & Filtered apped & Filtered				45: Cut data 5, groups ng 44: Cut	ffdiff on and oth differe	<u>data 4,</u> <u>ners: TSS</u> ntial exp data 4,	
<u></u>		2: (Condition						<u>data 5,</u> <u>FPKM t</u>	and oth racking	<u>ners: CDS</u>	2
ools (beta) ly		R R	ame R3G eplicates						<u>43: Cut</u> <u>data 5,</u> <u>FPKM c</u> <u>9</u>	ffdiff on and oth different	<u>data 4,</u> ners: CDS tial expre	
osatellite Analysi s Prediction	<u>s</u>			5: R3G_REP3 Map 5: R3G_REP2 Map 4: R3G_REP1 Map 8: MeOH_REP3 Ma	ped & Filtered ped & Filtered ped & Filtered apped & Filtered				42: Cut data 5, overloa sting	ffdiff on and oth ading di	<u>data 4,</u> 1ers: CDS ffential (
<u>vs</u>		+	Insert Condit	ion	apped & Filtered				<u>41: Cut</u> data 5,	ffdiff on and oth	i data 4, ners: pro	
		Librar	ry normaliza	tion method								

Using Docker for Tool Integration

Tools can now be run in Docker containers.



Tool authors may annotate Docker container ids the tool works with in the Tool XML file

Deployers may specify Docker container ids on a per destination basis.

https://wiki.galaxyproject.org/Admin/Tools/Docker

Interactive environments such as IPython and RStudio can now be invoked from Galaxy. Work contributed by



Björn Grüning University of Freiburg





Eric Rashe Texas A&M University

🗧 Ga	alaxy		Analyze Data	Workflow	Shared	Data 🗸	Visualizatio	on - Admin He	elp v Use	r v	==	Using 93.7 M	мв
Seqid	Source	Туре	Start	End	Score	Strand	Phase	Attributes		÷.	History	00	
##gff-ve	rsion 3												
chr2L	dm3.ensGene.tbl	gene	20385293	20399394		-		Name=FBgn0032	2857;ID=FB	n0032	search datasets	C	2
chr2L	dm3.ensGene.tbl	mRNA	20385293	20399394				ID=FBtr0081370;F	Parent=FBg	n00321	Screencast		
chr2L	dm3.ensGene.tbl	exon	20399206	20399394		-		ID=exon:FBtr0081	1370:1;Pare	nt=FBt	4 shown, 2 deleted		
chr2L	dm3.ensGene.tbl	exon	20385293	20387592		-		Parent=FBtr00813	370;ID=exo	n:FBtrC			
chr2L	dm3.ensGene.tbl	five_prime_UTR	20399206	20399394		-		Parent=FBtr00813	370;ID=five	prime	46.9 MB		<u> </u>
chr2L	dm3.ensGene.tbl	five_prime_UTR	20387532	20387592	а.	-		Parent=FBtr00813	370;ID=five	prime	4: misopy.ipynb	@ # ¥	
chr2L	dm3.ensGene.tbl	start_codon	20387529	20387531		-	0	Parent=FBtr00813	370;ID=star	_codo			
chr2L	dm3.ensGene.tbl	CDS	20386317	20387531		-	0	ID=CDS:FBtr0081	1370:1;Pare	nt=FBt	IPython Notebook		
chr2L	dm3.ensGene.tbl	stop_codon	20386317	20386319		-	0	ID=stop_codon:F8	Btr0081370	1;Pare	format: ipynb, database: <u>7</u>	-	
chr2L	dm3.ensGene.tbl	mRNA	20385293	20399394		-		Parent=FBgn0032	2857;ID=FB	tr0081:	uploaded ipynb file		
chr2L	dm3.ensGene.tbl	exon	20399092	20399394		-		ID=exon:FBtr0081	1369:1;Pare	nt=FBt			
chr2L	dm3.ensGene.tbl	exon	20385293	20387592		-		ID=exon:FBtr0081	1369:2;Pare	nt=FBt	B 0 2 M	۲	
chr2L	dm3.ensGene.tbl	five_prime_UTR	20399092	20399394		-		Parent=FBtr00813	369;ID=five	_prime_	{ ID then		
chr2L	dm3.ensGene.tbl	five_prime_UTR	20387532	20387592	а.	-		ID=five_prime_UT	TR:FBtr008	.369:2;	Python	- <u>(</u>	-
chr2L	dm3.ensGene.tbl	start_codon	20387529	20387531		-	0	ID=start_codon:Fi	Btr0081369	1;Pare	"metadata" RStudio	-	
chr2L	dm3.ensGene.tbl	CDS	20386317	20387531		-	0	Parent=FBtr00813	369;ID=CD	S:FBtr0	"kernelspec": {		
chr2L	dm3.ensGene.tbl	stop_codon	20386317	20386319		-	0	ID=stop_codon:F8	Btr0081369	1;Pare	"disalau assa", "Dubi	2"	
chr2L	dm3.ensGene.tbl	gene	21418852	21419260		+		Name=FBgn0053	802;ID=FB	n0053	display_name : Pyth	ion 2 ,	
chr2L	dm3.ensGene.tbl	transcript	21418852	21419260		+		Parent=FBgn0053	3802;ID=FB	tr0091	"language": "python",		
chr2L	dm3.ensGene.tbl	exon	21418852	21419260		+		ID=exon:FBtr0091	1806:1;Pare	nt=FBt	"name": "python2"		
chr2L	dm3.ensGene.tbl	gene	6036997	6037902		-		ID=FBgn0031766	;Name=FB	n0031	indine i pythone		
chr2L	dm3.ensGene.tbl	mRNA	6036997	6037902		-		ID=FBtr0079222;F	Parent=FBg	n0031	3: dm3.ensGene.gff3	• / ×	t –
chr2L	dm3.ensGene.tbl	five_prime_UTR	6037816	6037902		-		Parent=FBtr00792	222;ID=five	_prime_	~460.000 lines		
chr2L	dm3.ensGene.tbl	start_codon	6037813	6037815		-	0	ID=start_codon:Ff	Btr0079222	1;Pare	format: gff3, database: ?		
chr2L	dm3.ensGene.tbl	CDS	6037129	6037815		-	0	Parent=FBtr00792	222;ID=CD	S:FBtr0	-		
chr2L	dm3.ensGene.tbl	stop_codon	6037129	6037131			0	ID=stop_codon:F8	Btr0079222	1;Pare	uploaded gff3 file		
chr2L	dm3.ensGene.tbl	exon	6036997	6037902		-		ID=exon:FBtr0079	9222:1;Pare	nt=FBt	BACH		
chr2L	dm3.ensGene.tbl	gene	11083459	11085243		-		Name=FBgn0032	2337;ID=FB	n0032		~ ~	·
chr2L	dm3.ensGene.tbl	mRNA	11083459	11085243		-		Parent=FBgn0032	2337;ID=FB	tr0114	display with IGV local		
chr2L	dm3.ensGene.tbl	five_prime_UTR	11085171	11085243	а.	-		Parent=FBtr01144	455;ID=five	prime	1.Seqid 2.Source	3.Type	
chr2L	dm3.ensGene.tbl	start_codon	11085168	11085170		-	0	ID=start_codon:Ff	Btr0114455	1;Pare	##gff-version 3		
chr2L	dm3.ensGene.tbl	exon	11084811	11085243		-		ID=exon:FBtr0114	4455:1;Pare	nt=FBt	chr2L dm3.ensGene.tb)	gene	
chr2L	dm3.ensGene.tbl	exon	11084167	11084347		-		ID=exon:FBtr0114	4455:2;Pare	nt=FBt	chr21 dm3 ensGene th	mDNA	

Galaxy Analyze Data Workflow Shared Data - Visualization - Admin Help - User -	===	Using 93.7 MB
	History	3¢⊡
	search datasets	8
File Edit View Insert Cell Kernel Heip Saving every 120s Saving every 120s	Screencast 4 shown, 2 <u>deleted</u>	
Install Misopy	4: misopy.ipynb	• # X
In [1]: !pip installuserquiet misopy	IPython Notebook format: ipynb, database	:: <u>?</u>
Restart the kernel (Kernel -> Restart), so that IPython finds Misopy	uploaded ipynb file	
Get Samtools		•
<pre>In [1]: from urllib import urlretrieve urlretrieve("http://depot.galaxyproject.org/package/linux/x86_64/samtools !tar -xf samtools.tgz</pre>	<pre>"metadata": { "kernelspec": {</pre>	
<	"display_name": "P	ython 2",
The samtools binary should now be in bin/samtools	"language": "pytho	n",
Generate Indices of the bam Files	"name": "python2"	
 Now the bam files from the Galaxy history can be imported with get(history_id) Then index files of the bam files can be generated via samtools 	3: dm3.ensGene.gff3 ~460,000 lines format: gff3, database:	• • ×
<pre>In [2]: get(5) get(6) !mv 5 461177.bam !mv 6 461178.bam</pre>	uploaded gff3 file	•
!bin/samtools index 461177.bam !bin/samtools index 461178.bam	1.Seqid 2.Source ##gff-version 3 chr2L dm3.ensGene.t chr2L dm3.ensGene.t	3.Type bl gene bl mRNA

🚍 Galaxy	Analyze Data Workflow	Shared Data - Visualization -	Admin Help - Us	er -	===	Using 93.7 MB
TP[v]: Notobook	invthon galaxy n	otobook		Ĥ	istory	2 � □
		OLEDUOK (unsaved changes)	0		search datasets	8
		Cell Toolbar: None		4	creencast shown, 2 <u>deleted</u> 6.9 MB	S
The samtools binary sho	ould now be in bin/samtool	ls		<u>4</u>	misopy.ipynb dm3.ensGene.gff3	• # ×
 Now the barn files fr Then index files of the 	rom the Galaxy history car the bam files can be gener	n be imported with get(history_ic rated via samtools	d)	2: 2.	461178.bam 4 MB	• # ×
<pre>In [2]: get(1) get(2) !mv 1 461177.bam !mv 2 461178.bam</pre>	get(1) get(2) !mv 1 461177.bam !mv 2 461178.bam					•
!bin/samtools inde !bin/samtools inde	ex 461177.bam ex 461178.bam			di di B:	splay with IGV <u>local</u> splay in IGB <u>View</u> inary bam alignments	file
If everything went well th	nere should now be index	files (*.bai) next to the bam files	5.	1	461177.bam	● # ×
In [3]: !ls *.ba* I 461177.bam 461177 Generate Index of an A	.bam.bai 461178.ba	am 461178.bam.bai				
An annotation file is need In [4]: get(10) !mv 10 dm3.ensGene	ded. ≥.gff3					









Bloomington, Indiana

What's coming: *Right here! Right now!*



TT16: RiboGalaxy: a platform for the alignment, analysis and visualization of ribo-seq data

Audrey Michel

The Galaxy Team



Enis Afgan

Dannon Baker

Dan Blankenberg

Dave Bouvier

Marten Cech

John Chilton



Dave Clements

Nate Coraor

Carl Eberhard

Jeremy Goecks

James Taylor





Jen Jackson

Ross Lazarus

Anton Nekrutenko

Nick Stoler



Nitesh Turaga

http://wiki.galaxyproject.org/GalaxyTeam

http://bit.ly/GxyCamFeedback

Dave Clements

Galaxy Project Johns Hopkins University clements@galaxyproject.org



Scalability ...

