Outline

Sequencing (NGS) in general

Sequencing data analysis in general

Kathi Zarnack and Julian König data

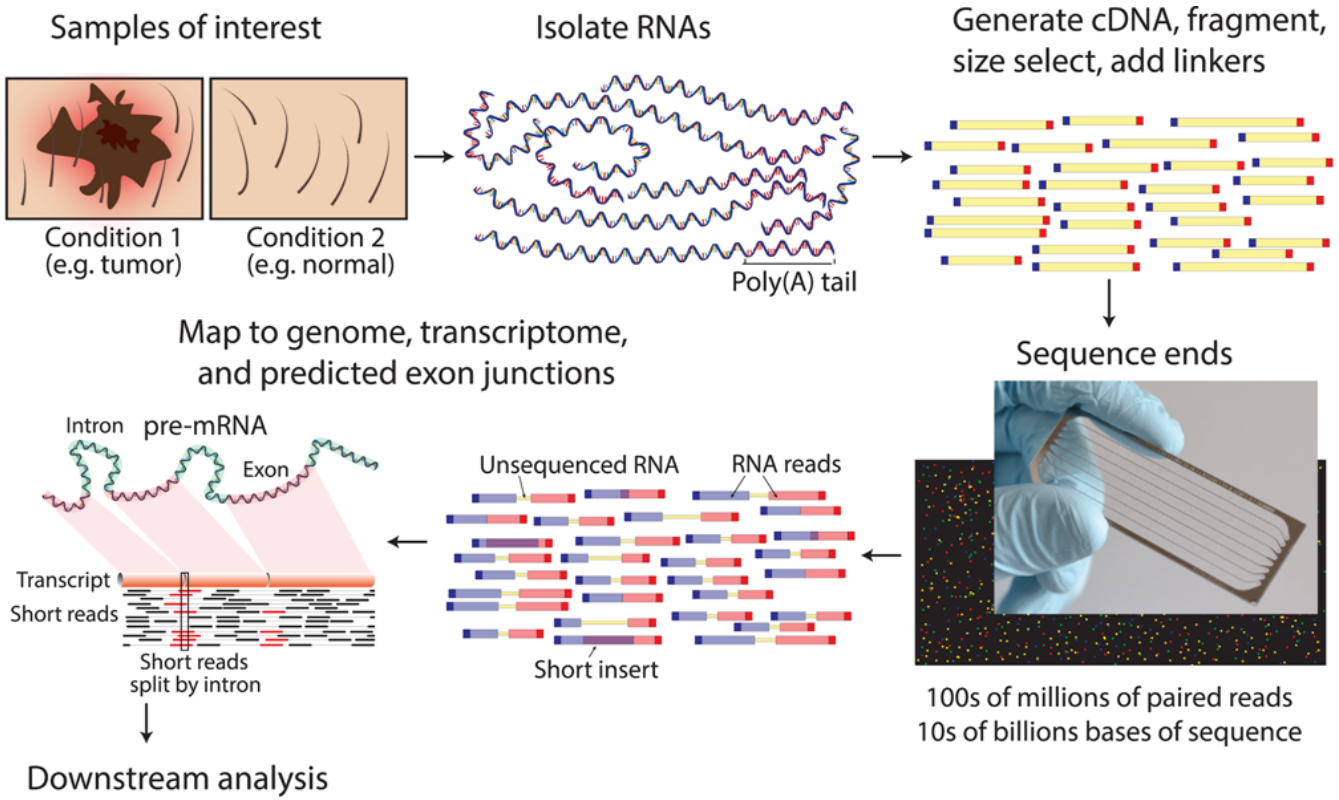
Results (selected)

Galaxy

RNA-Seq data analysis

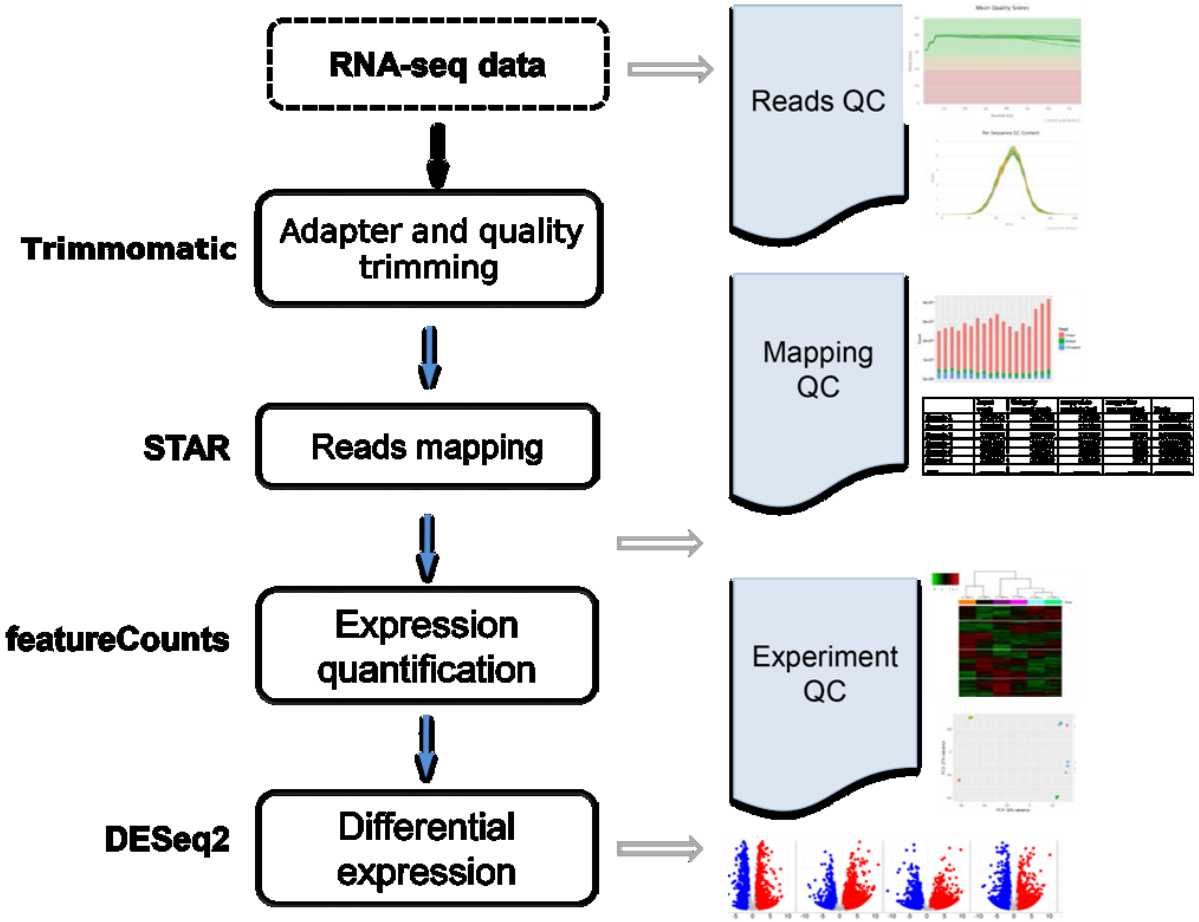
PLoS Comput Biol. 2015
Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud.
Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL.

Sequencing (NGS) in general



PLoS Comput Biol. 2015
Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud.
Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL.

Sequencing data analysis in general



Cell. 2013
Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements.
Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J.

Kathi Zarnack data

Cell

Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of *Alu* Elements

Kathi Zarnack,[1,8] Julian König,[2,8] Mojca Tajnik,[2,3] Iñigo Martincorena,[1] Sebastian Eustermann,[2] Isabelle Stévant,[1] Alejandro Reyes,[4] Simon Anders,[4] Nicholas M. Luscombe,[1,5,6,7,*] and Jernej Ule[2,*]
[1]European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
[3]Faculty of Medicine, University of Ljubljana, Vrazov trg 2, SI-1104 Ljubljana, Slovenia
[4]EMBL, Genome Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany
[5]UCL Genetics Institute, Department of Genetics, Environment and Evolution, University College London, Gower Street, London WC1E 6BT, UK
[6]Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK
[7]Okinawa Institute for Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa 904-0495, Japan
[8]These authors contributed equally to this work
*Correspondence: nicholas.luscombe@ucl.ac.uk (N.M.L.), jule@mrc-lmb.cam.ac.uk (J.U.)
http://dx.doi.org/10.1016/j.cell.2012.12.023

Cell. 2013
Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements.
Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J.

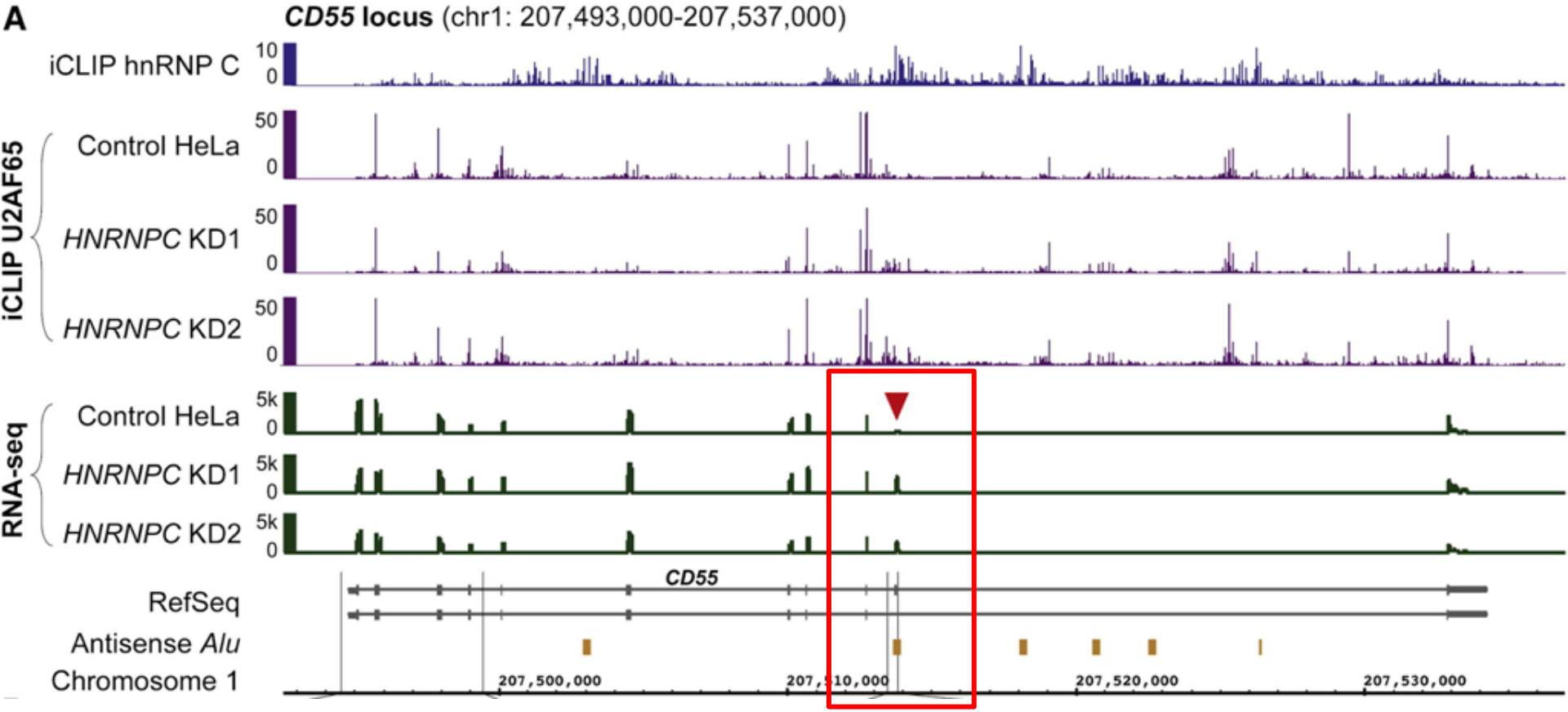
Summary of the results

There are ~650,000 ***Alu* elements** in transcribed regions of the human genome. These elements contain **cryptic splice sites**, so they are in constant **danger** of aberrant **incorporation into mature transcripts**. Despite posing a major threat to transcriptome integrity, **little is known** about the molecular **mechanisms preventing their inclusion**. Here, we present a mechanism for protecting the human transcriptome from the aberrant exonization of transposable elements. Quantitative **iCLIP data** show that the **RNA-binding protein hnRNP C competes** with the **splicing factor U2AF65** at many genuine and **cryptic splice sites**. **Loss of hnRNP C** leads to **formation** of previously **suppressed *Alu* exons**, which severely **disrupt transcript function**. Minigene experiments explain disease-associated mutations in *Alu* elements that hamper hnRNP C binding. Thus, by preventing U2AF65 binding to *Alu* elements, **hnRNP C** plays a **critical role** as a **genome-wide sentinel protecting the transcriptome**. The findings have important implications for human evolution and disease.

Cell. 2013

Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements.

Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J.



Goal of the practical

Get from the **raw sequencing data** to the **gene expression** (**RNA-Seq**)

Analyze **RNA-Seq** data and get **differential gene expression** and **expression** of individual **exons** (example at gene CD55 gene)

Show **coverage cryptic exon(s)** (example at gene CD55)

Do **everything** in **less that half a day**

Galaxy practical

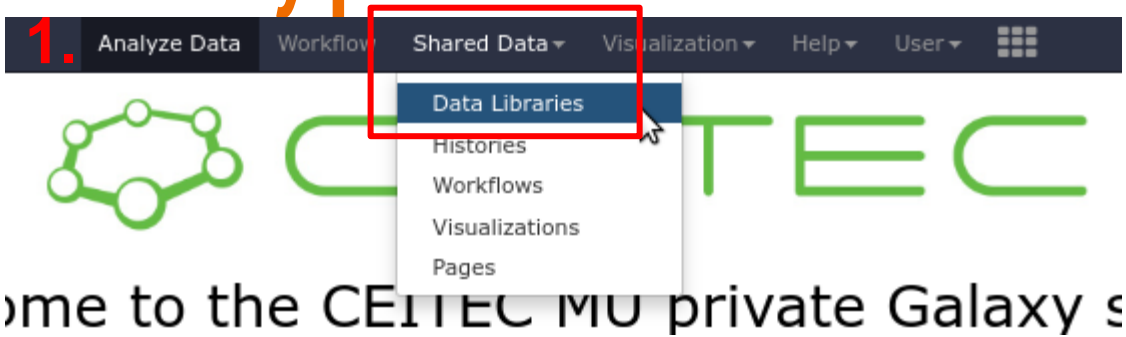
Get the data

Or you just load the **preloaded data**

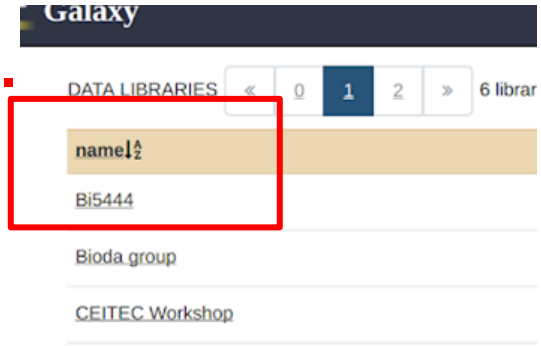
```
Shared Data -> Data Libraries -> Bi5444 -> RNA-Seq
```

Galaxy practical

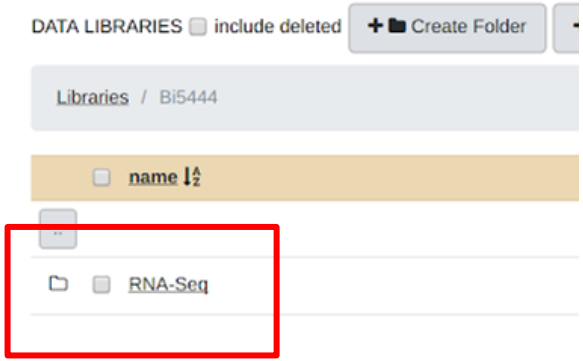
1.



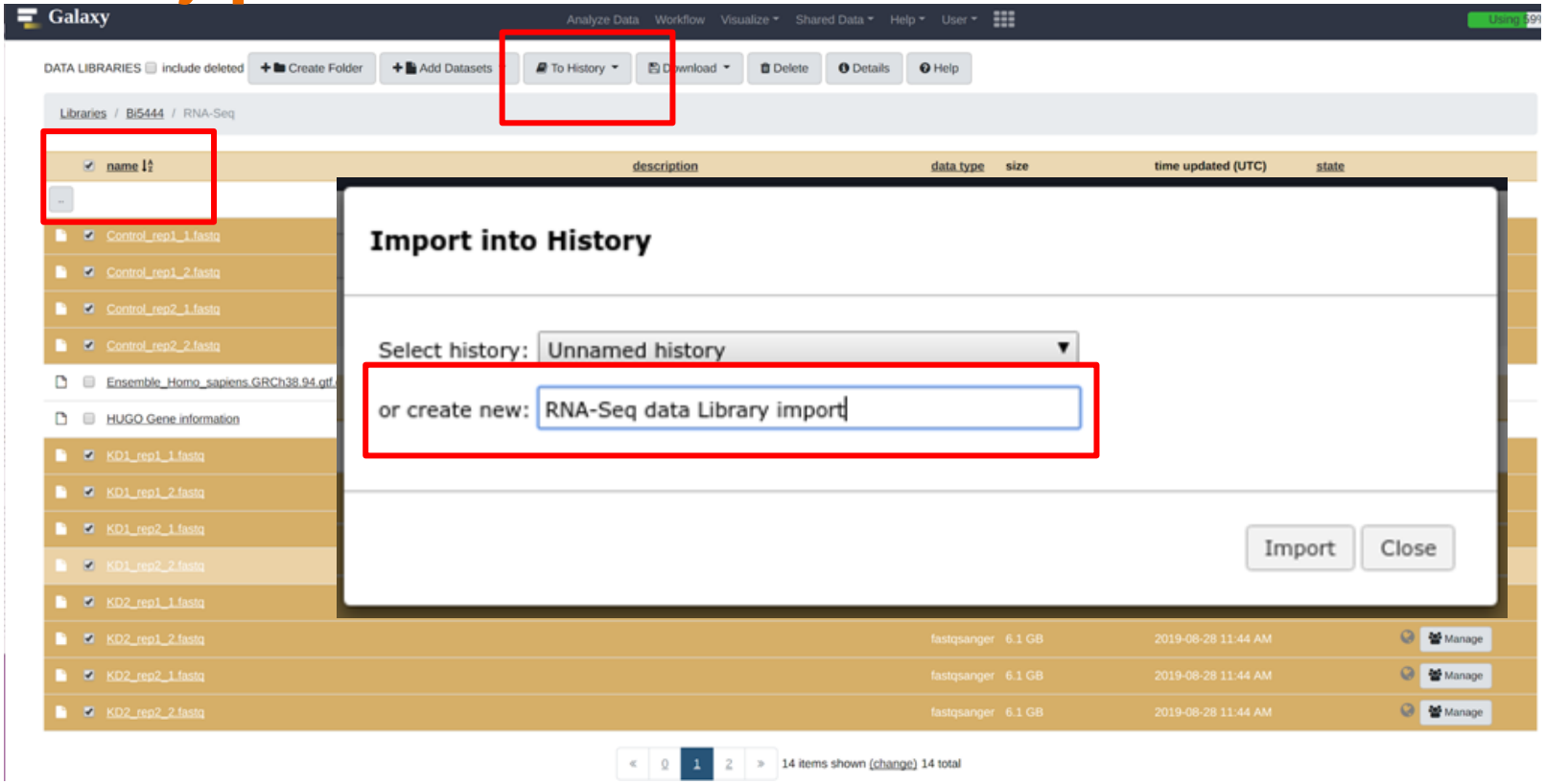
2.



3.



Galaxy practical



Galaxy practical

Get the data

Galaxy practical

Initial quality check

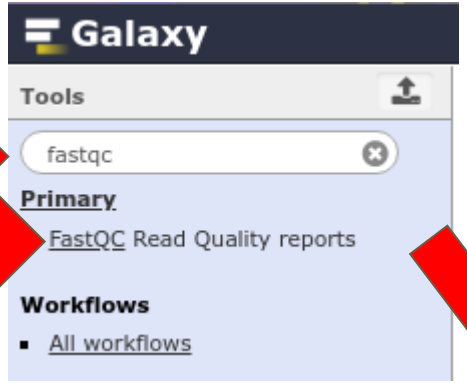
Check the **raw** reads **quality**

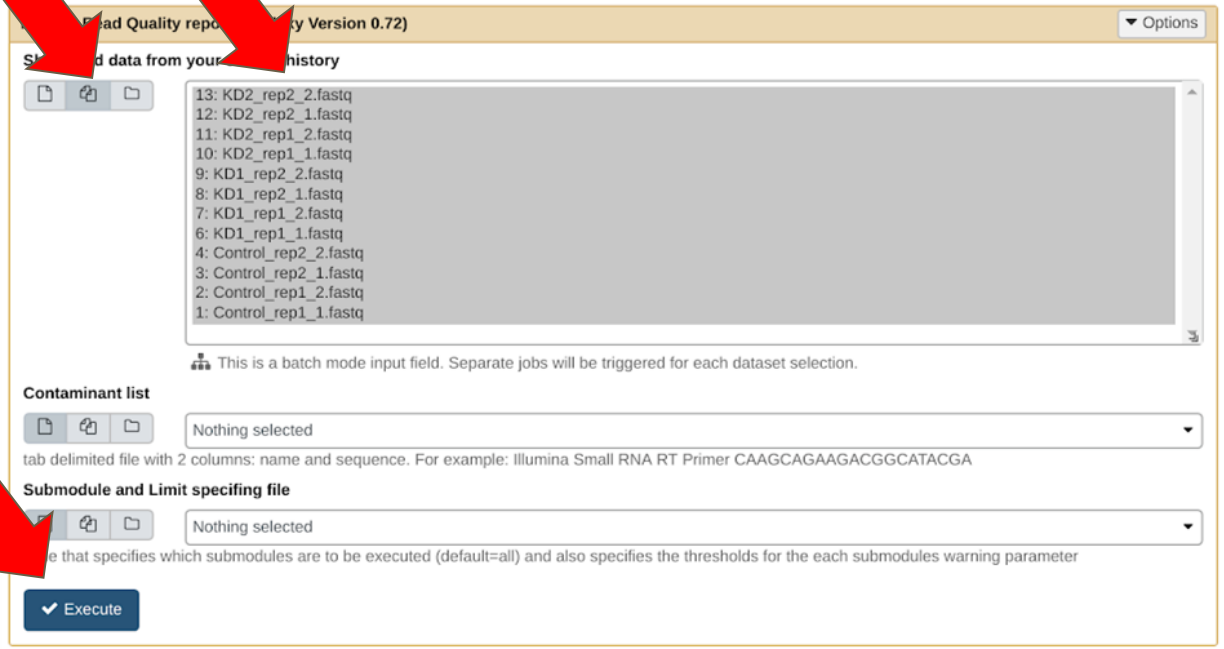
Using `FastQC` tool

Input **FASTQ**, output **HTML**

Galaxy practical

Initial quality check





Galaxy practical

Initial quality check

It is still running, right?

But without that, we cannot proceed

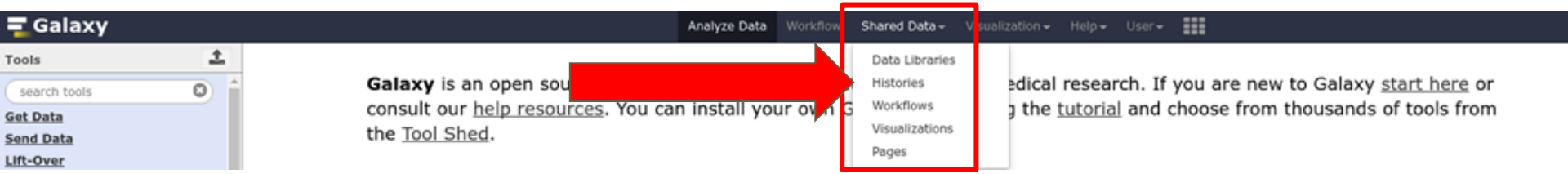
We have a solution! :)

Import **Galaxy history**

Galaxy practical

Initial quality check

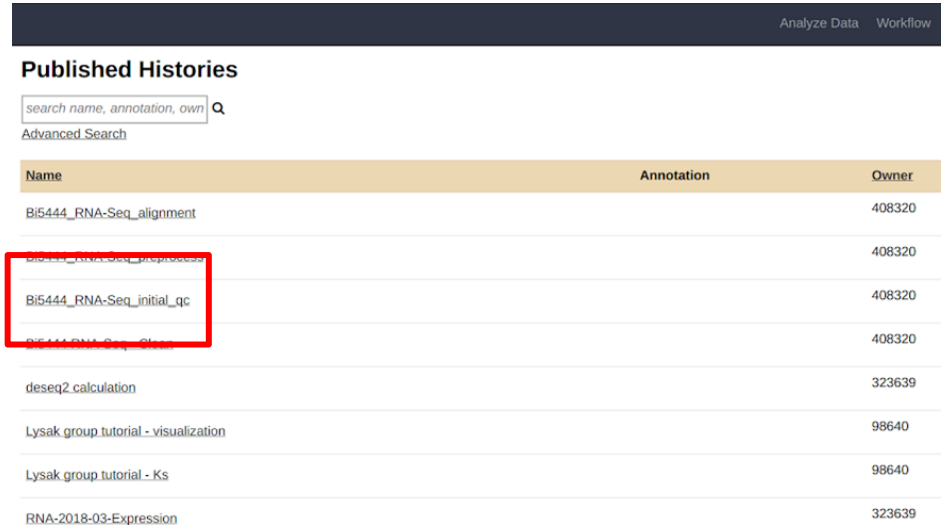
Import Galaxy history



Galaxy practical

Initial quality check

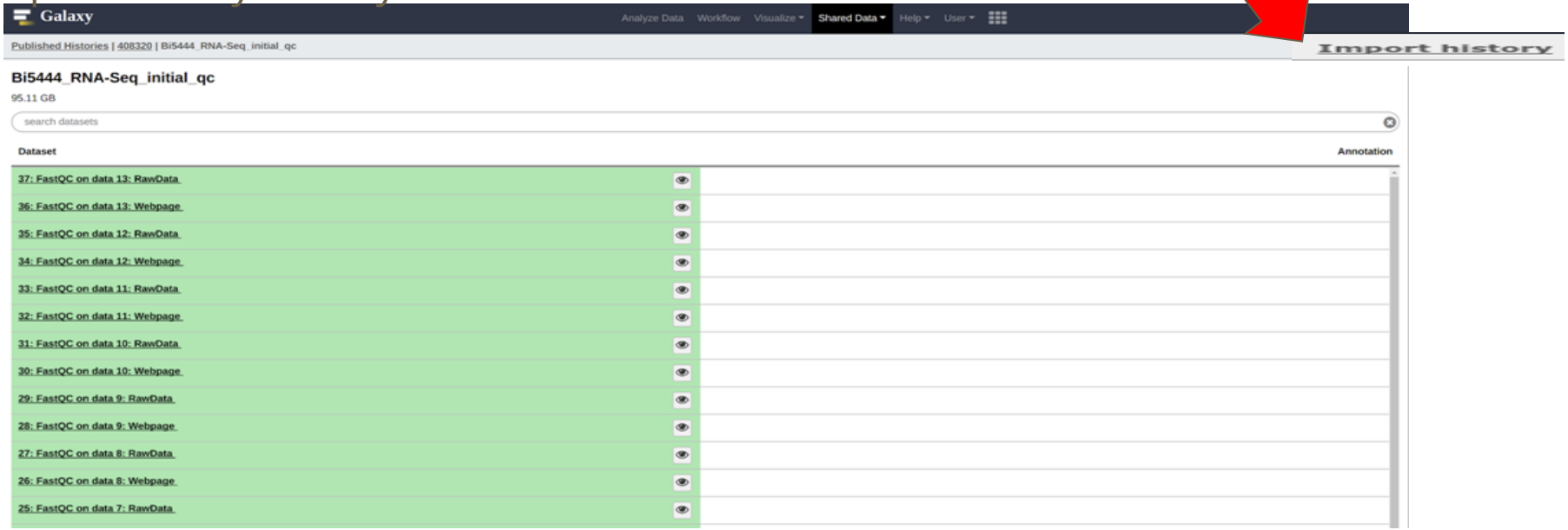
Import Galaxy **history**



Galaxy practical

Initial quality check

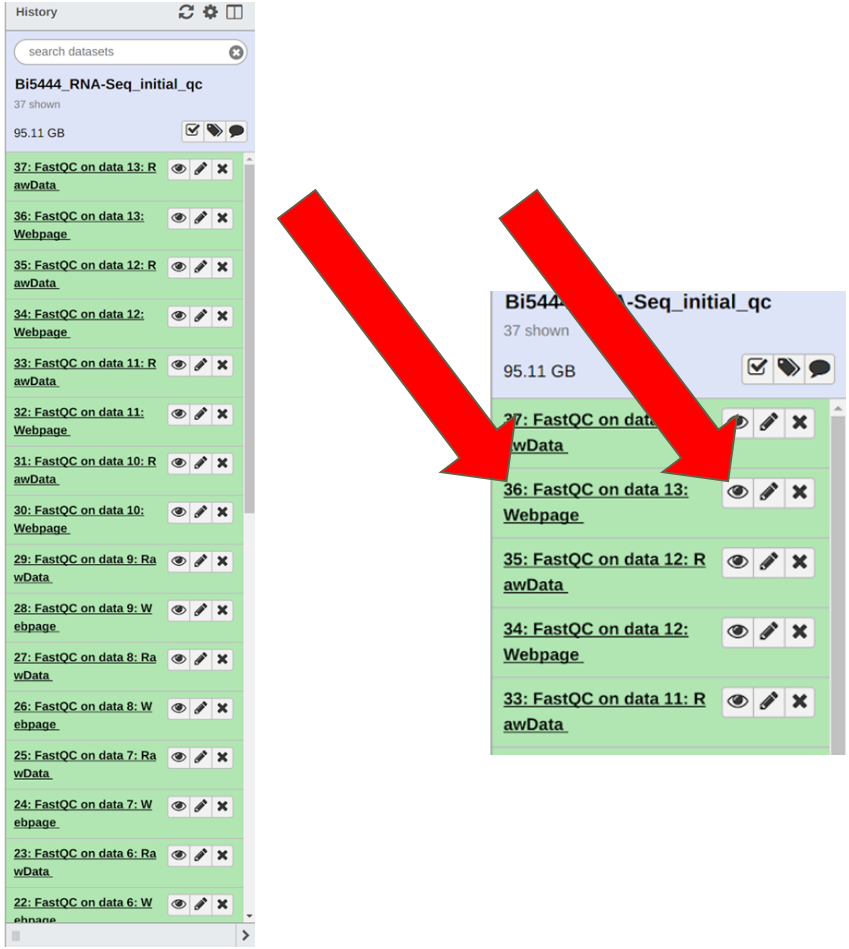
Import Galaxy history



Galaxy practical

Initial quality check

HTML report(s)



Galaxy practical

Initial quality check

HTML report(s)

But there is **too many of them**

> **`MultiQC`** - makes you life simpler
>
> This time, you can **try it on your own**!

Galaxy practical

Summary of the logs

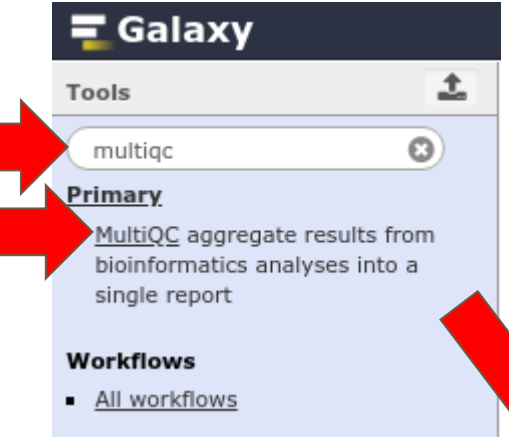
Summarize the output **logs**

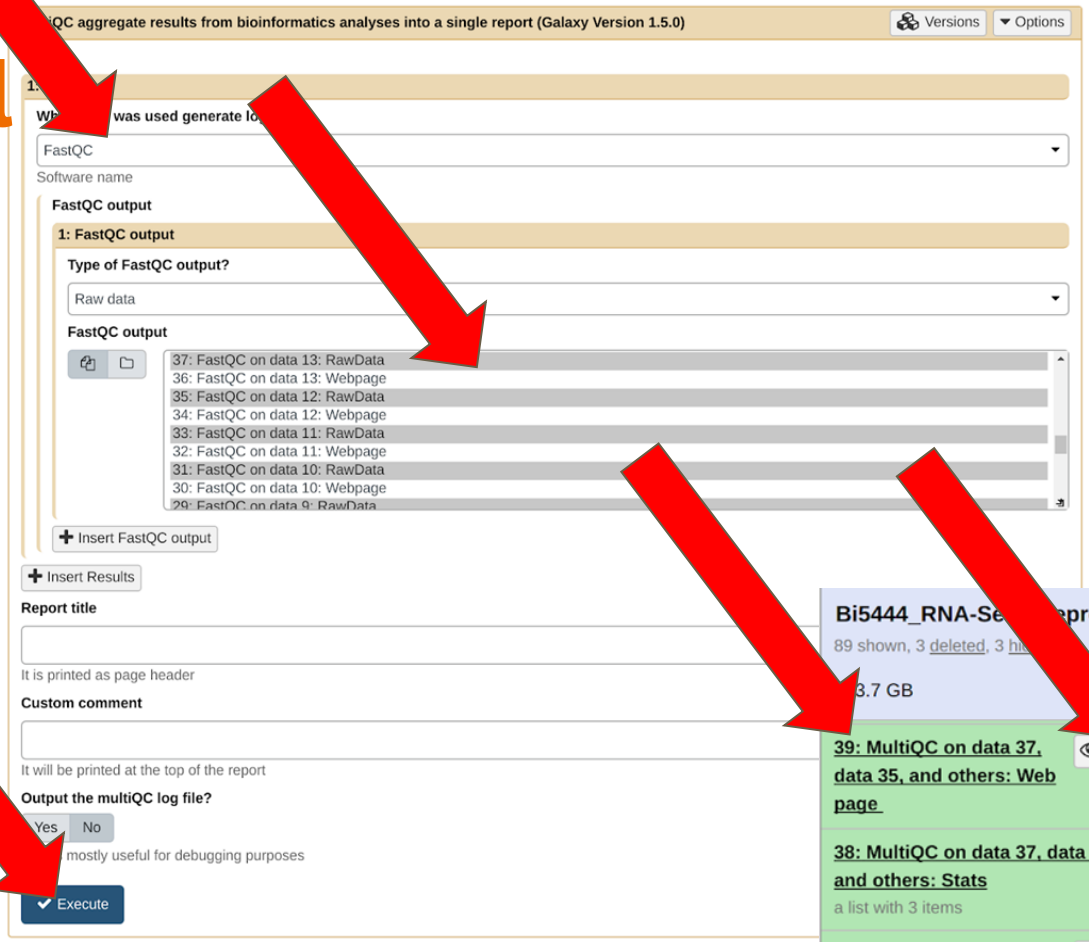
Using **`MultiQC`** tool

Input **LOG(s)**, output **HTML**

Galaxy practical

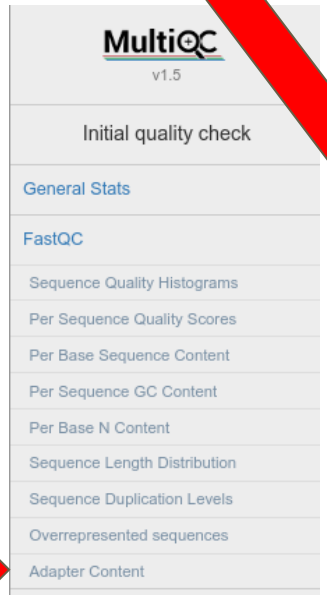
Summary of the logs

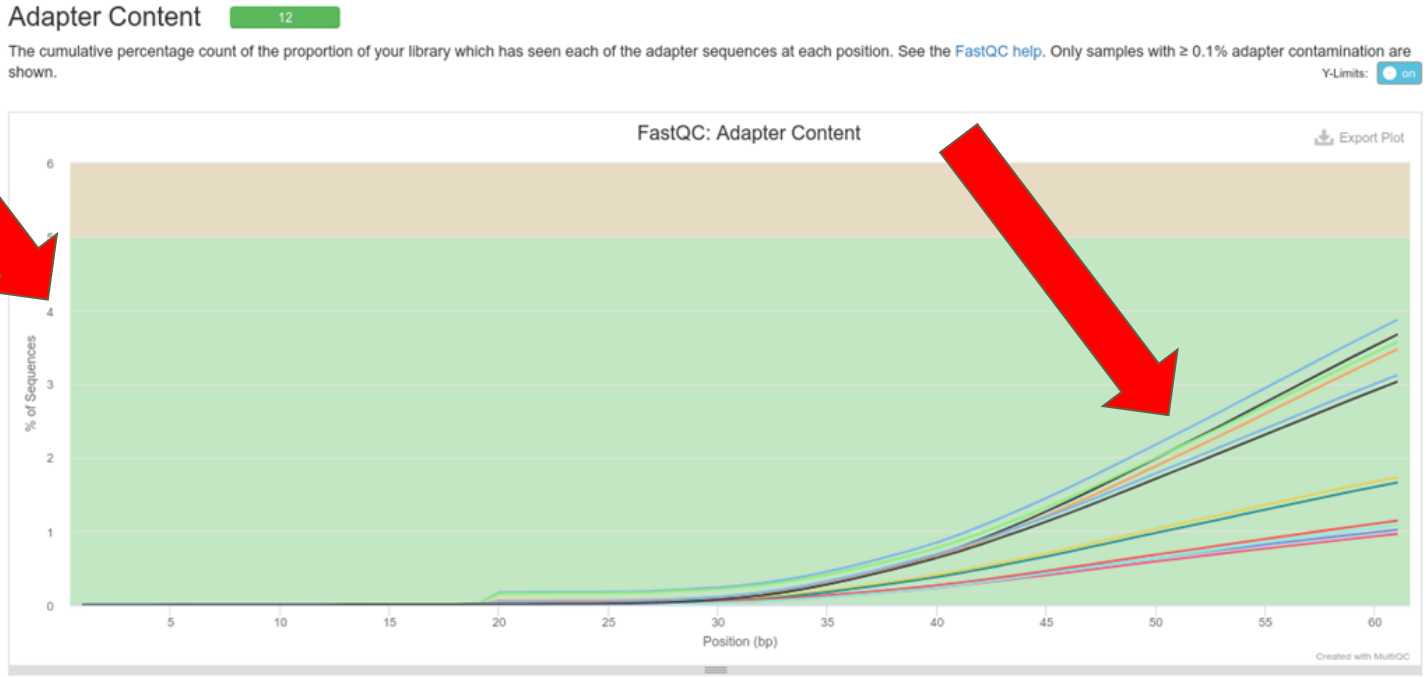




Galaxy practical

Initial quality check - Adapter content





Galaxy practical

Read preprocessing

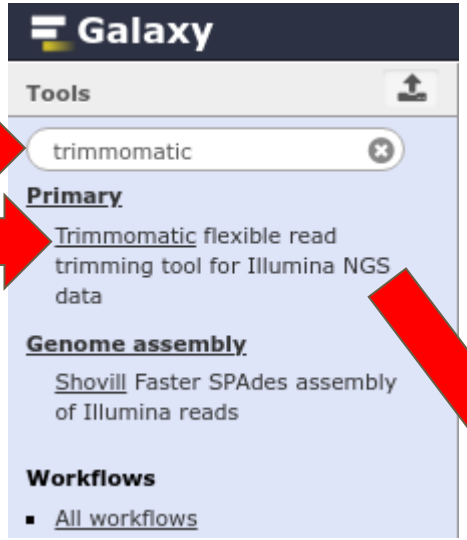
Remove **adapters** and/or trim **low-quality** ends

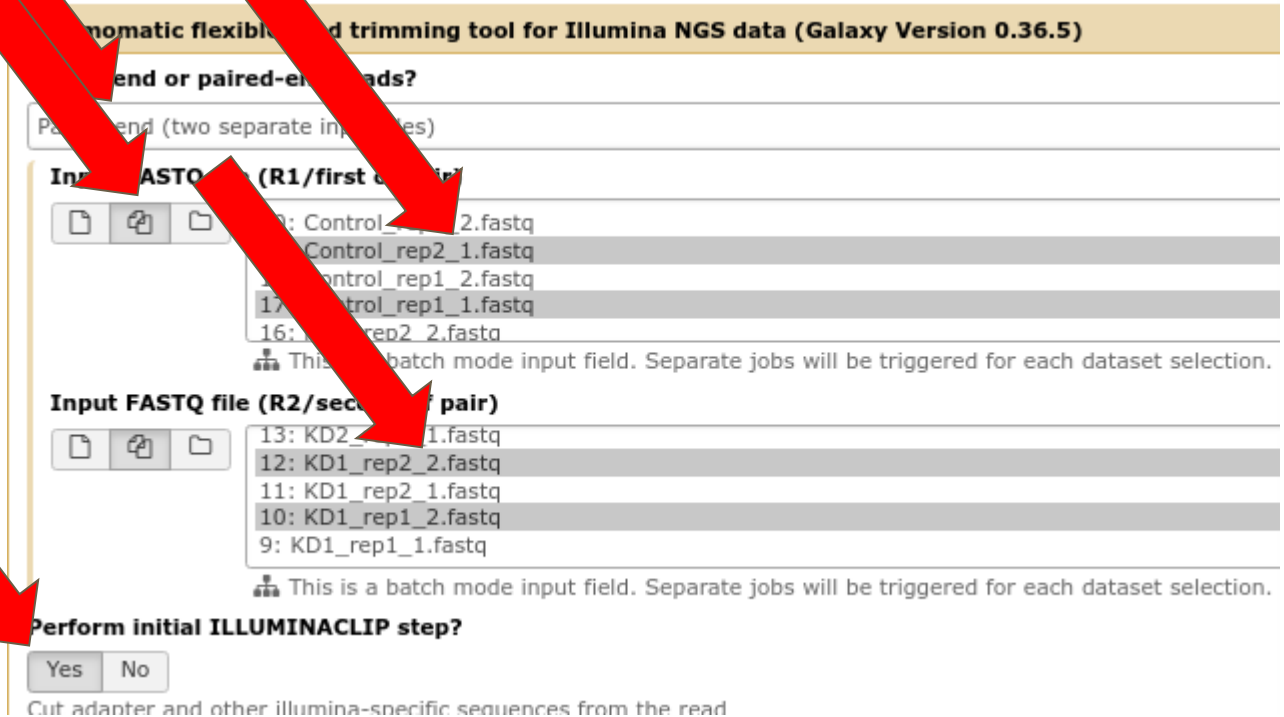
Using `Trimmomatic` trimmer

Input **FASTQ**, output **FASTQ**

Galaxy practical

Read preprocessing





Galaxy practical

Read preprocessing

Perform initial ILLUMINACLIP step?

Yes | No

adapter and other illumina-specific sequences from the read

Select standard adapter sequences or provide custom?

Custom

Custom adapter sequences in fasta format

```
>adapter
AGATCGGAAGAGC
```

Write sequences in the fasta format.

Adapter sequence (partial):

```
>adapter

AGATCGGAAGAGC
```

Trimmomatic Operation

1: Trimmomatic Operation

Select Trimmomatic operation to perform

Sliding window trimming (SLIDINGWINDOW)

Number of bases to average across

4

Average quality required

5

+ Insert Trimmomatic Operation

✔ Execute

Galaxy practical

Read preprocessing

Galaxy practical

Preprocessing quality check

Check the **preprocessed** reads **quality and summarize**

Using **`FastQC`** & **`MultiQC`** tools

Input **FASTQ/LOG**, output **HTML**

Galaxy practical

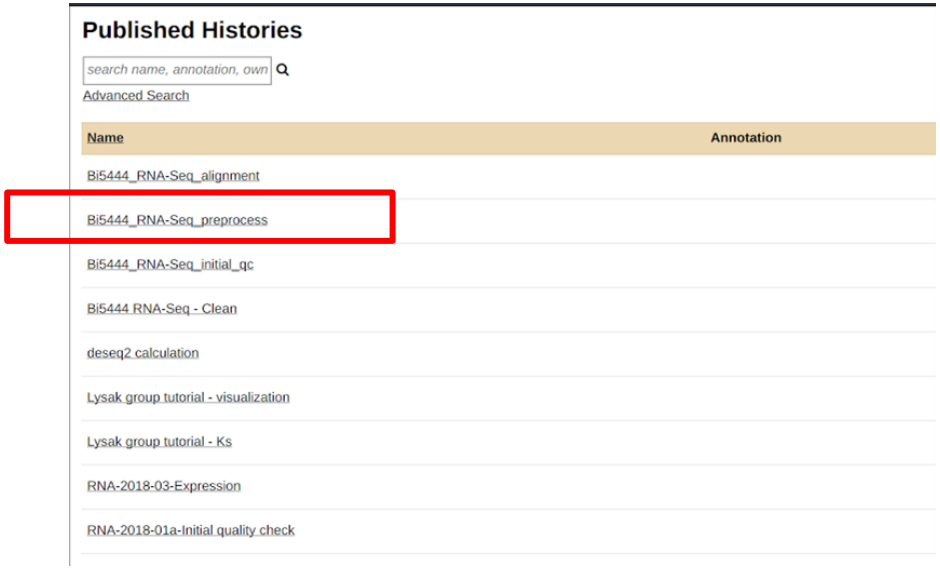
Preprocessing quality check

Please, run the **`FastQC`** and **`MultiQC`** on the preprocessed files and check the adapter content

Galaxy practical

Preprocessing quality check

`Share Data -> Histories -> Bi5444_RNA-Seq_preprocess`

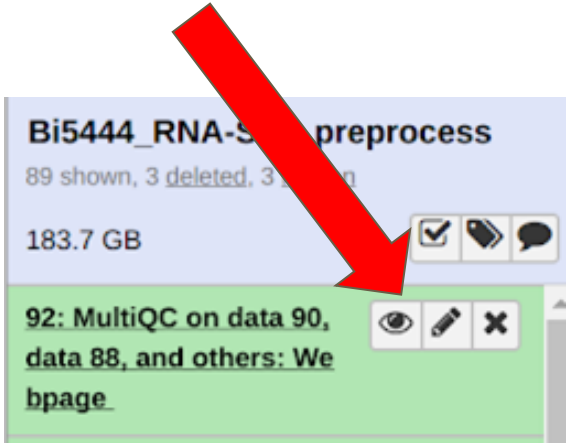


Galaxy practical

Preprocessing quality check

Check the **preprocessed** reads **quality & summarize**

Are all the **bad** things **gone**?



Galaxy practical

Preprocessing quality check

Adapter Content 12

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position. See the FastQC help. Only samples with ≥ 0.1% adapter contamination are shown.

No samples found with any adapter contamination > 0.1%

Galaxy practical

Preprocessing quality check

Check the **preprocessed** reads **quality & summarize**

Are all the **bad** things **gone**?

Actually, for **modern aligners** such as `STAR`, it **doesn't** matter that **much**

They can perform **soft-clipping**

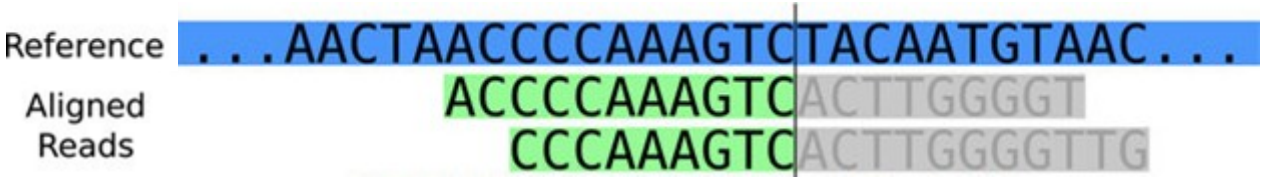
Galaxy practical

Soft-clipping in alignment

Hiding of **non-matching** parts of the **reads**

Can overcome **remaining adapter** or **low-quality** sequences

Only to **specified limits** (in `STAR` the default is max. **33%** of the read length)



Soft-clipped part

BMC genomics. 2014
SHEAR: sample heterogeneity estimation and assembly by reference.
Landman SR, Hwang TH, Silverstein KA, Li Y, Dehm SM, Steinbach M, Kumar V.

Galaxy practical

Alignment to genome

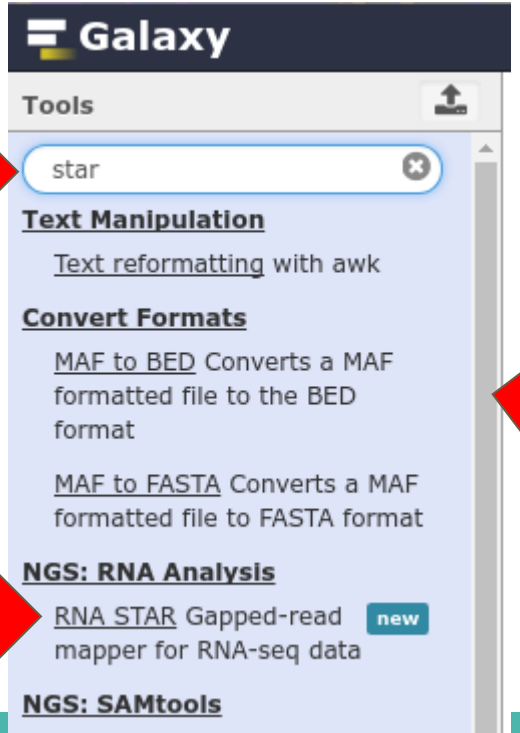
Align **RNA-Seq** data to a genome

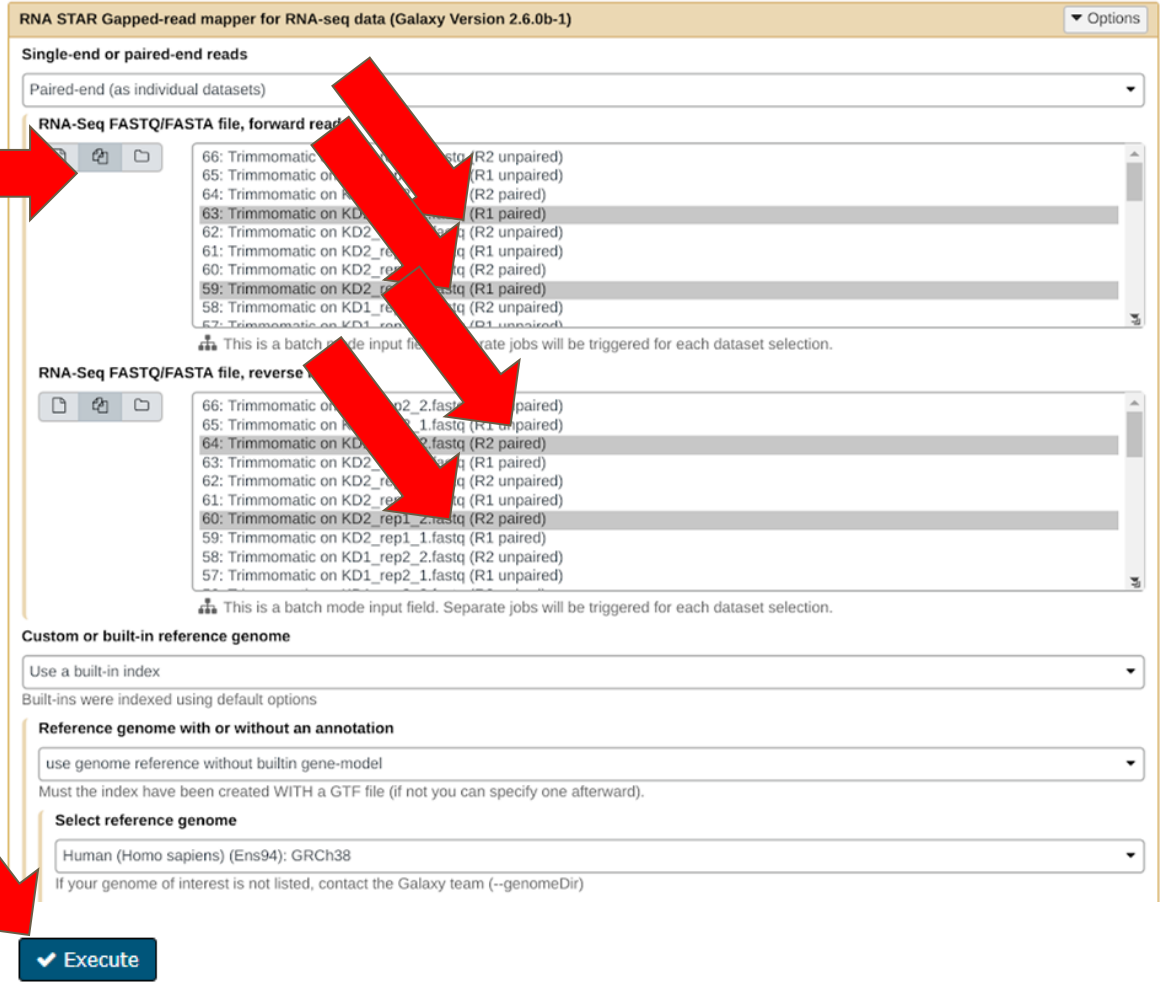
Using **STAR** aligner

Input **FASTQ**, output **BAM**

Galaxy practical

Alignment to genome





Sit back, wait and relax

You don’t need to know what’s happening or how long it’s going to take.

Galaxy practical

Alignment to genome

`Share Data -> Histories -> Bi5444_RNA-Seq_alignment`

Galaxy practical

Alignment to genome

`STAR` performs well even **with defaults**

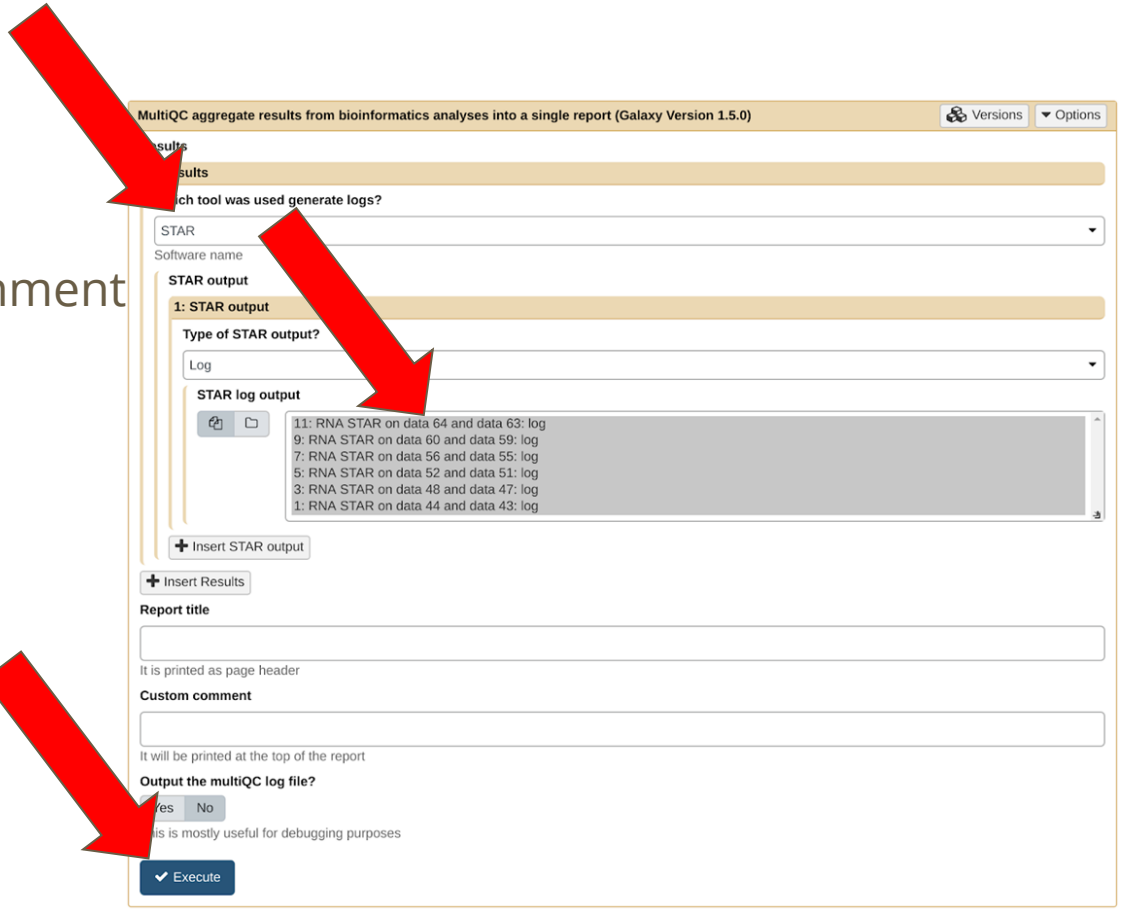
Main output is the **BAM** file

This is one of the few files worth to **keep** and **save**

Galaxy practical

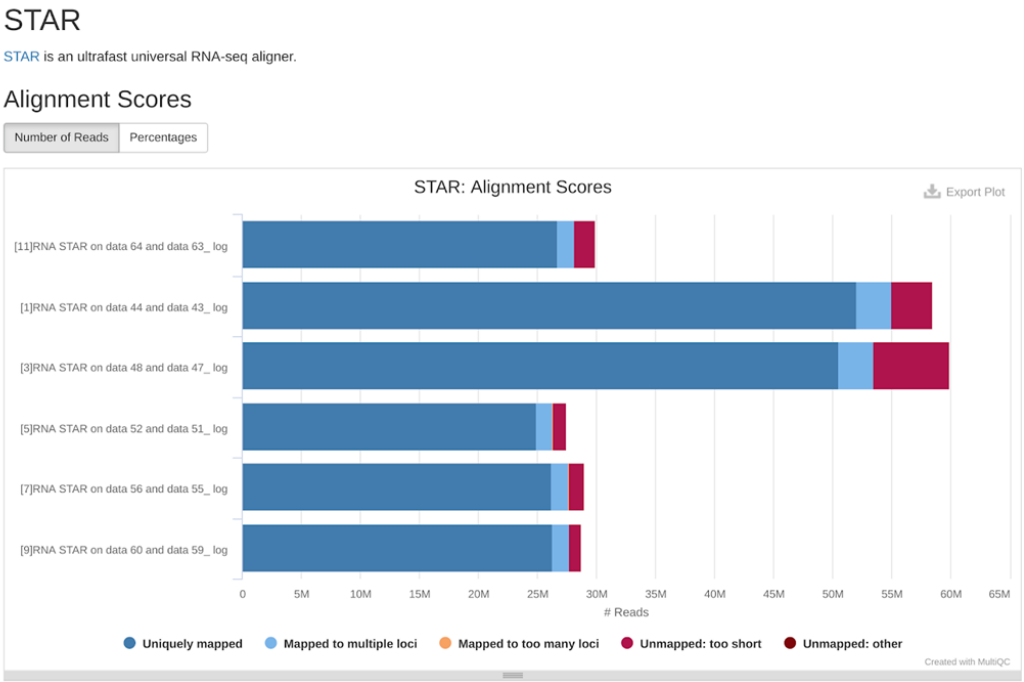
Quality control of alignment

Run **MultiQC** to assess the alignment



Galaxy practical

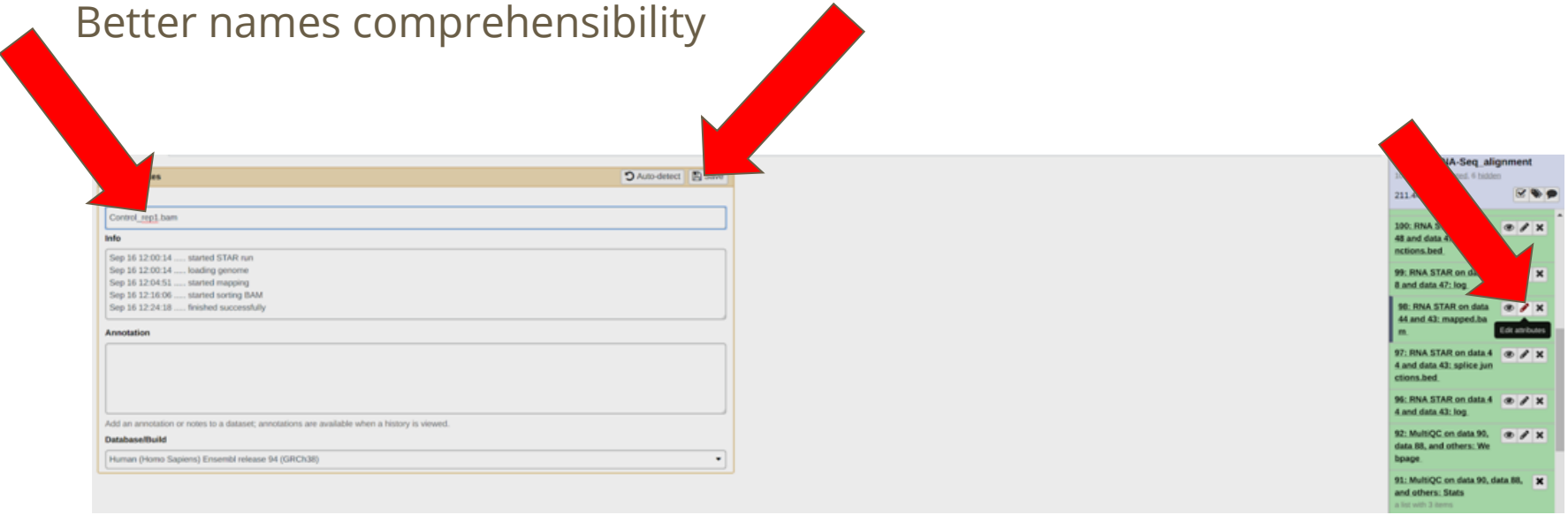
Quality control of alignment



Galaxy practical

Rename and tags

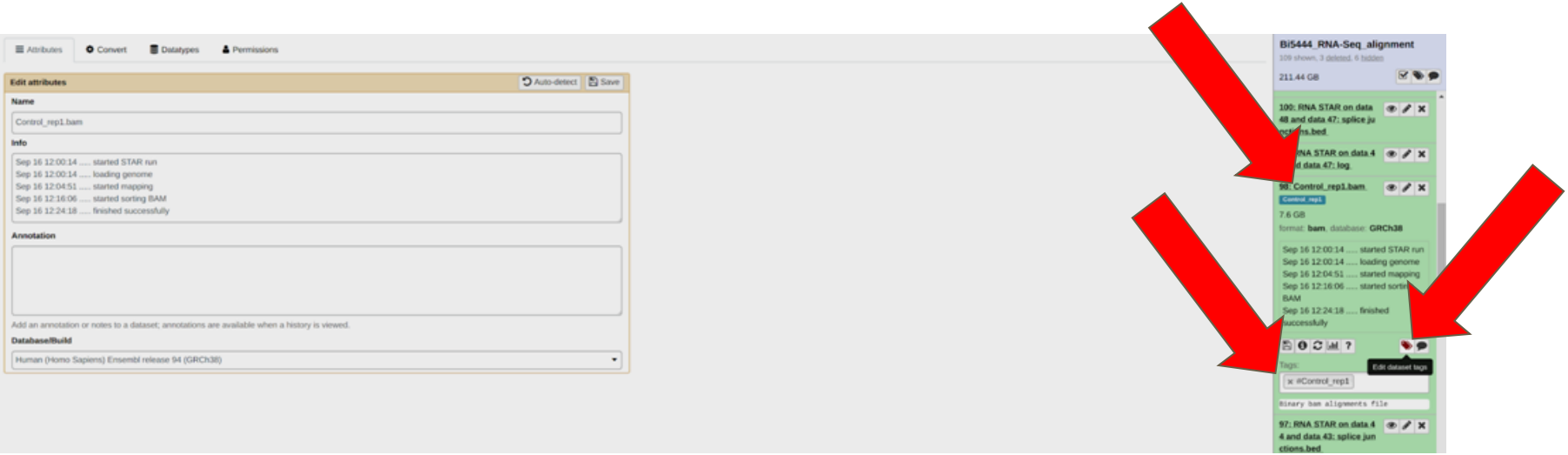
Better names comprehensibility



Galaxy practical

Rename and tags

Better names comprehensibility



Galaxy practical

Gene counts

For the **raw gene counts** (expression) you need to have a list of genes and their positions in the genome - gene annotation

Using **`UCSC Main table browser`**

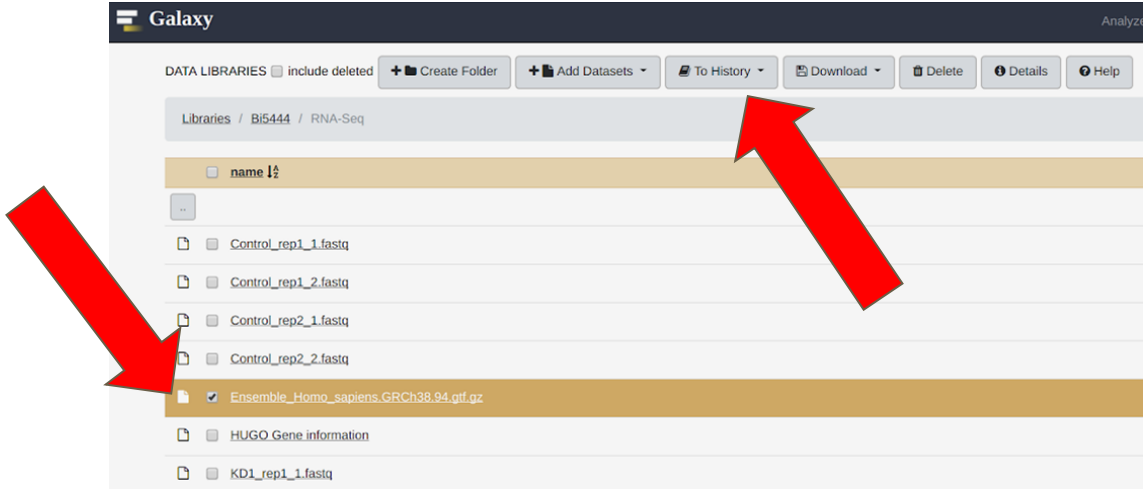
Input nothing, output **GTF**

Galaxy practical

Gene counts

Share Data -> Data Libraries -> Bi5444

-> RNA-Seq -> Ensemble_Homo_sapiens.GRCh38.94.gtf.gz



Galaxy practical

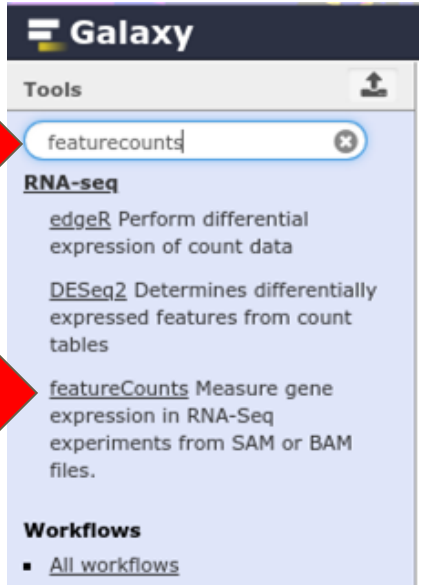
Gene counts

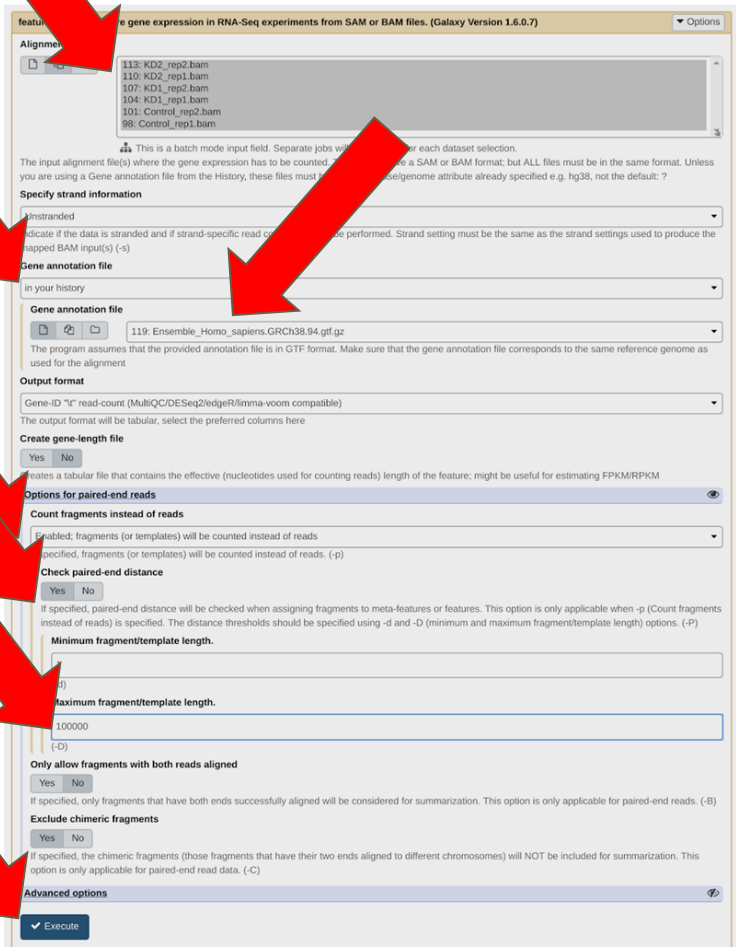
Get the **raw gene counts**

- Using `featureCounts` tool
- Input **BAM** and annotation **GTF**, output **TXT** (raw gene counts)

Galaxy practical

Gene counts



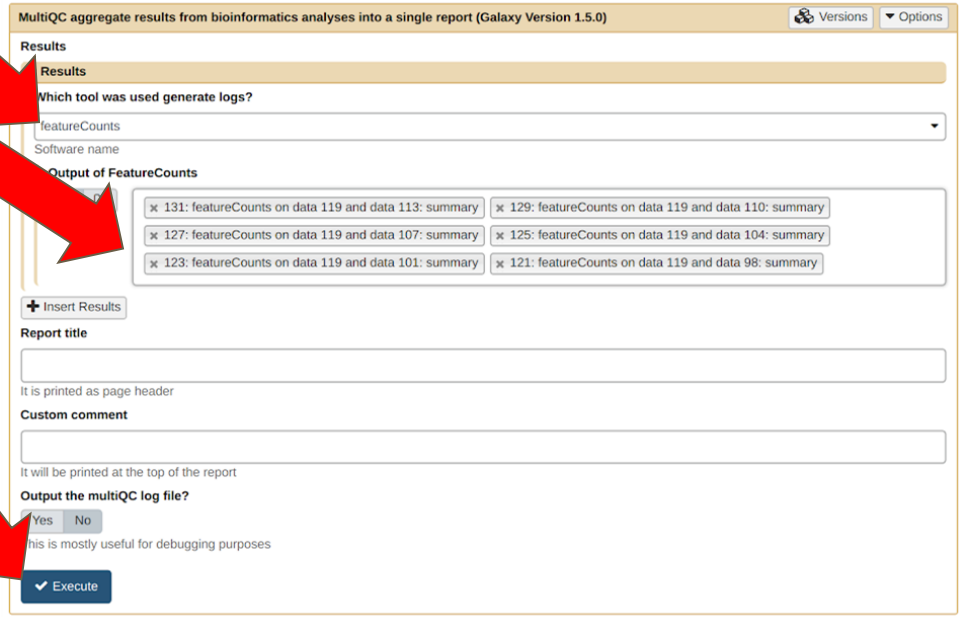


Galaxy practical

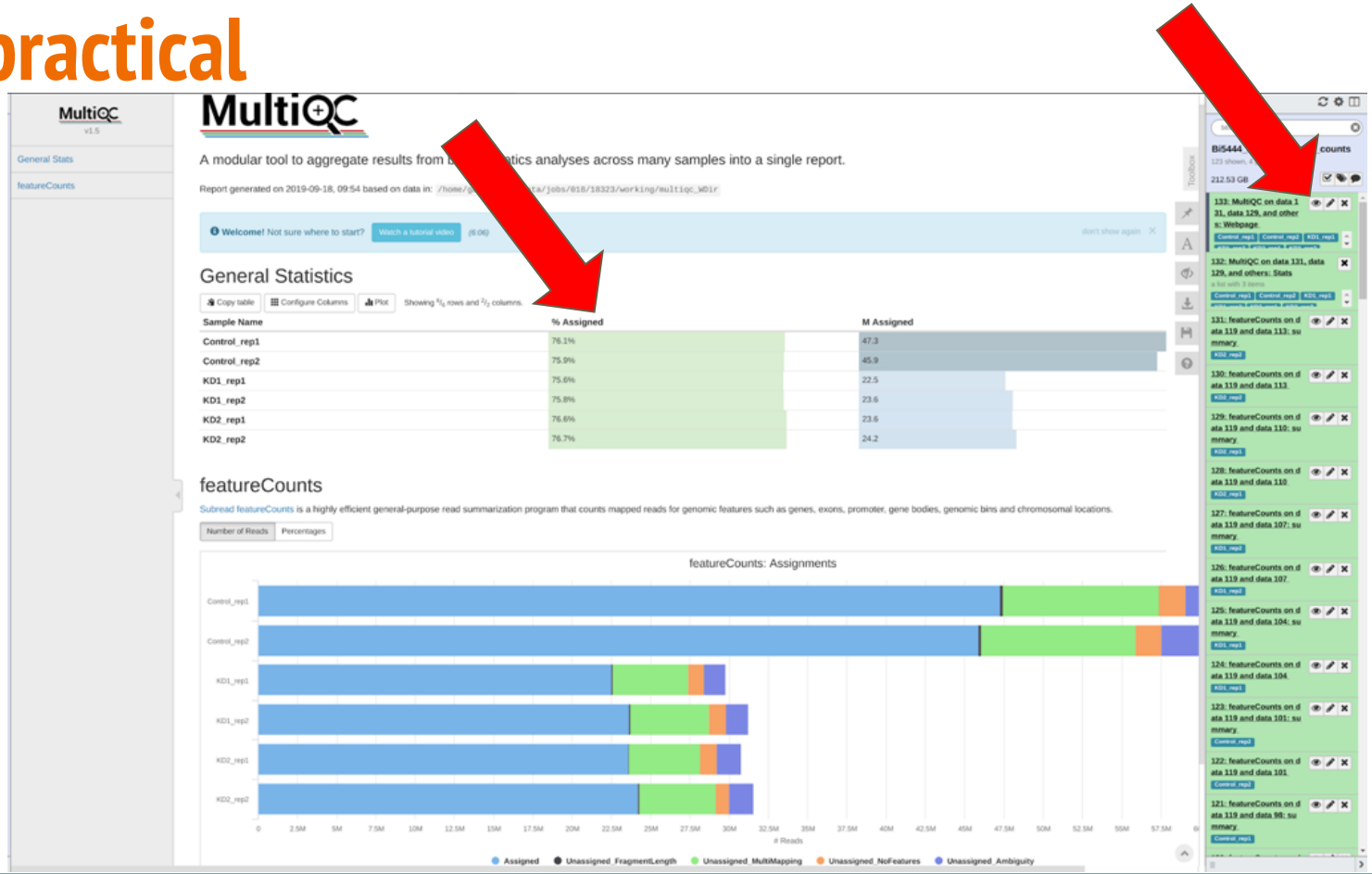
Gene counts

Quality control of gene counts

Again **MultiQC**



Galaxy practical



Galaxy practical

Differential gene expression

Get **differential gene expression** from the raw counts

Using `edgeR` and `DESeq2` tools

Galaxy practical

Differential gene expression - note

Optimally, the experiment **should** be **designed** with at least **three biological replicates**

However, if the **data** are only "**supportive**" **two replicates** is **enough**

Bioinformatics. 2013
Liu Y, Zhou J, White KP. *RNA-seq differential expression studies: more sequence or more replication?*
Rna. 2016
Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M. *How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?.*

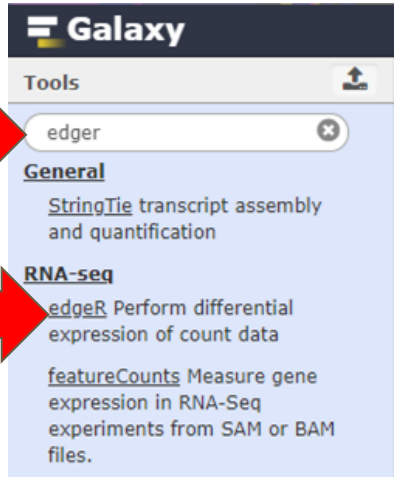
Galaxy practical

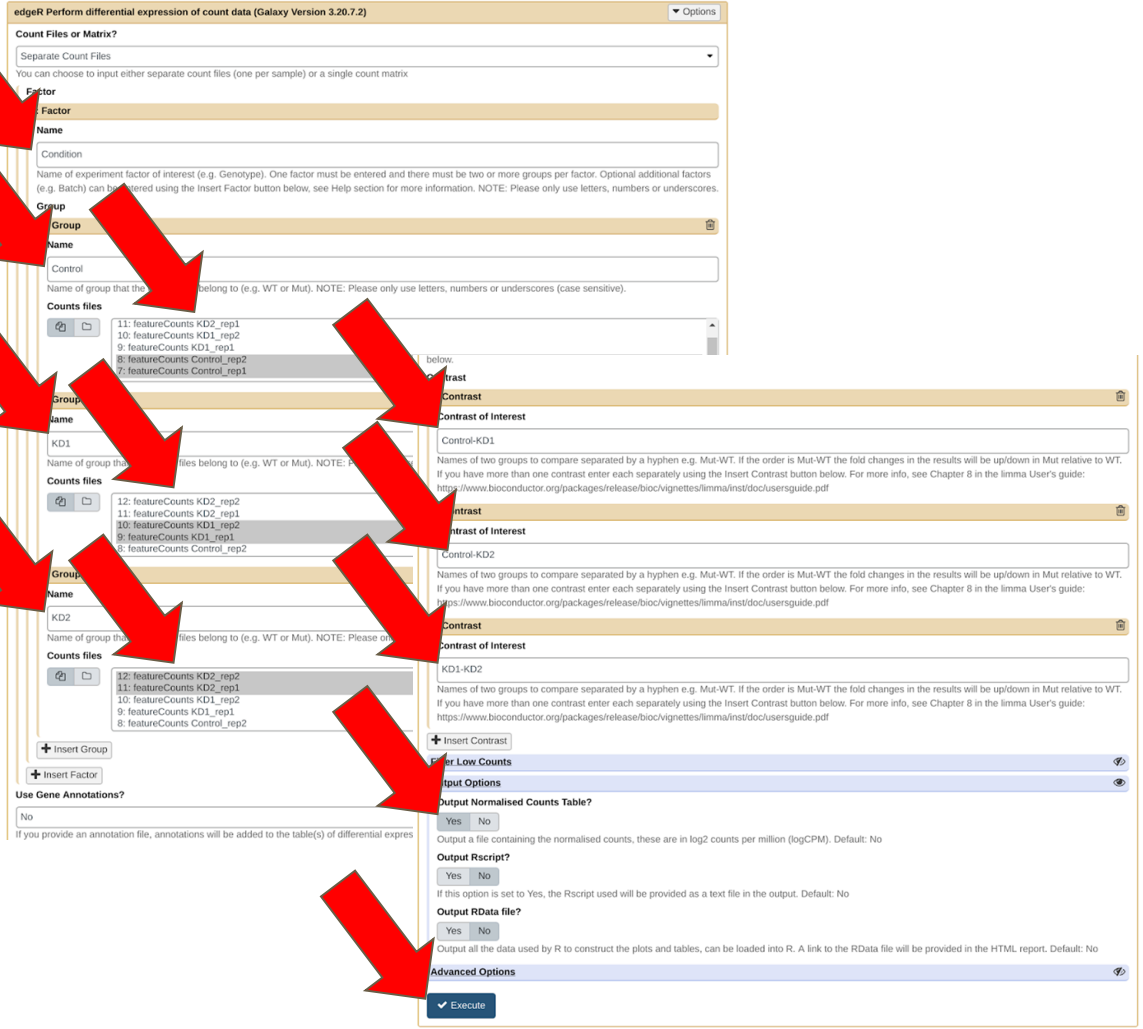
Differential gene expression

`Shared Data -> Histories -> Bi5444_RNA-Seq_DE_start`

Galaxy practical

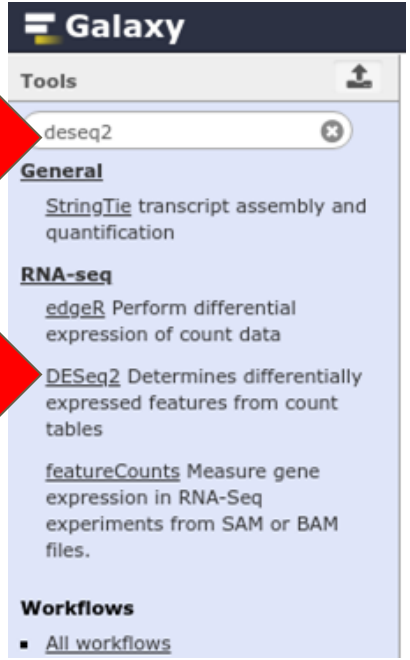
Differential gene expression

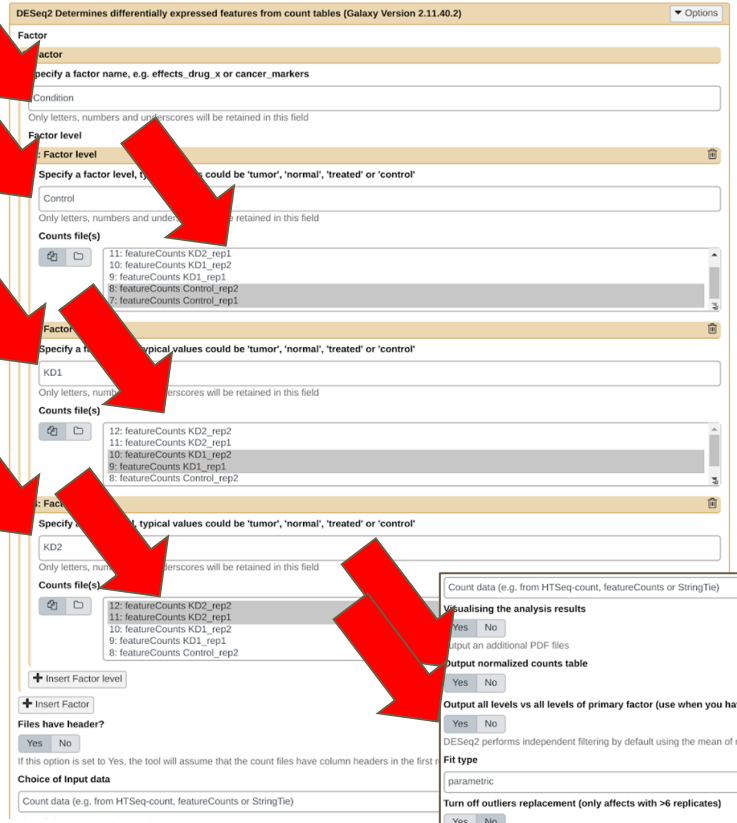


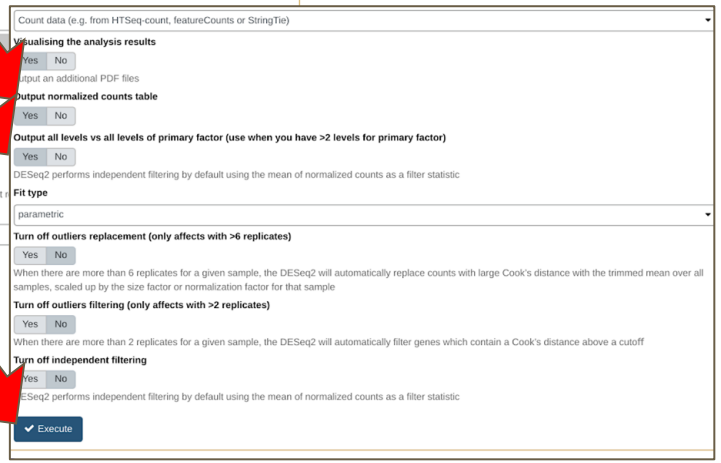


Galaxy practical

Differential gene expressio





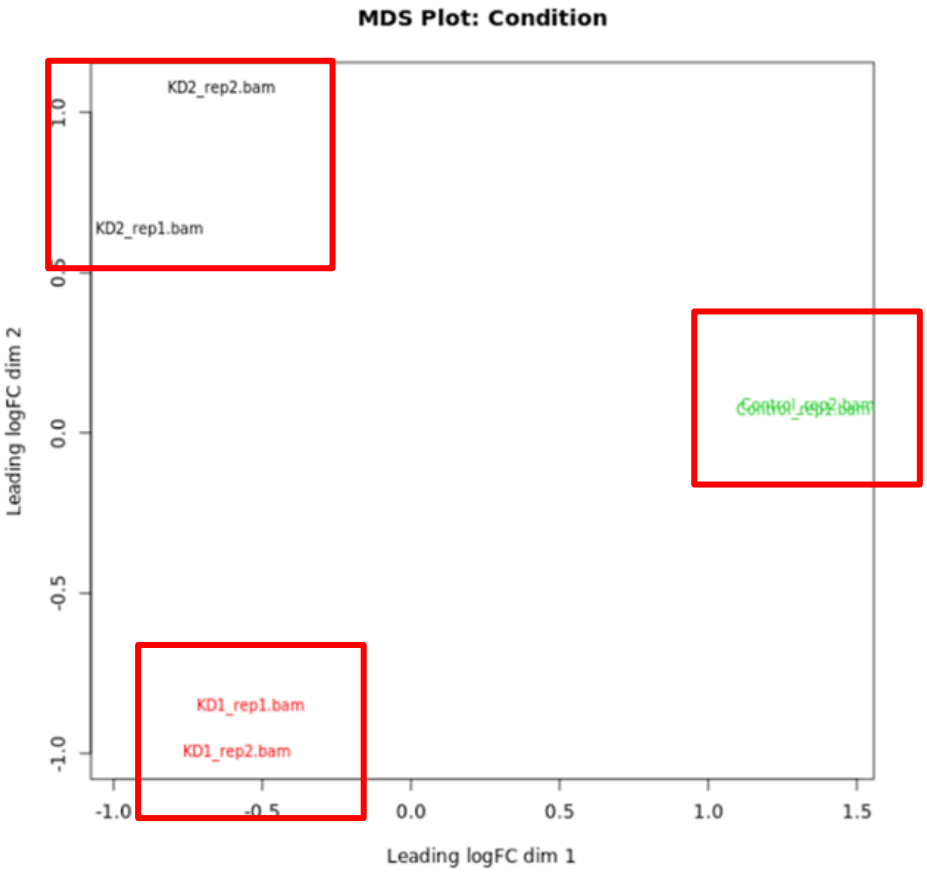


Galaxy practical

Differential gene expression

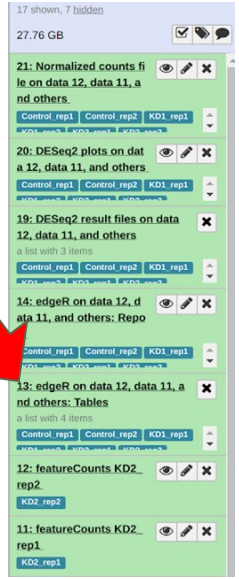
edgeR Analysis Output:

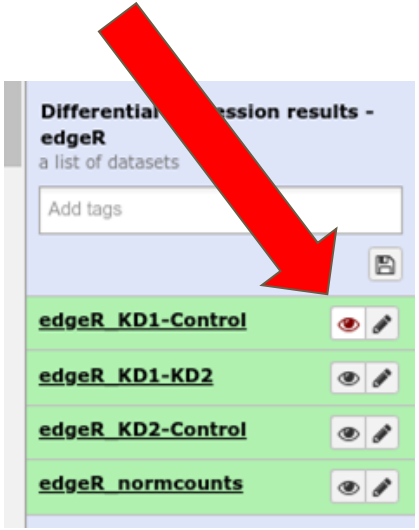
Links to PDF copies of plots are in 'Plots' section below.



Galaxy practical

Differential gene expression





1	2	3	4	5	6
GeneID	logFC	logCPM	F	PValue	FDR
ENSG00000092199	5.52142742878572	5.96346716277756	1462.34679277333	1.22114253157414e-318	7.17238065920072e-314
ENSG00000204253	6.22215133875183	4.69131980192735	1235.82902592066	8.6488589246041e-270	2.53995364468311e-265
ENSG00000107159	3.22278923144336	4.09433985683369	491.850213415037	7.9546272105367e-109	1.55738343070291e-104
ENSG00000179954	2.85904668532772	4.41691242725208	456.438217554854	3.85225443835843e-101	5.65655411092455e-97
ENSG00000148204	2.70384863075395	4.17441363845286	376.393049584693	1.77411170612e-83	2.08404902117916e-79
ENSG00000185633	4.22087072157582	2.41739313442468	358.912648383713	5.83527380494571e-80	5.71224678222477e-76
ENSG00000204560	2.10676298079	5.93980087019778	334.313475363922	1.29573528700066e-74	1.08721445831406e-70
ENSG00000099998	3.08600815349408	3.0528996991664	331.974819220342	4.17737211080344e-74	2.78770162619006e-70
ENSG00000197375	2.12482893907467	5.68742165552465	331.930250691321	4.2716122645289e-74	2.78770162619006e-70
ENSG00000169231	2.13377519397903	5.38735924665232	322.649201605679	4.44958188736524e-72	2.61346192154398e-68
ENSG00000157224	2.09164554071042	5.5519631536178	312.92589063867	5.78761591740723e-70	3.09032382644467e-66
ENSG00000157193	-1.90372177023031	7.73296494003751	302.848534430836	8.99503831407289e-68	4.40269646147559e-64
ENSG00000172346	2.62579315698578	3.626882383666	300.257776550655	3.29206236846498e-67	1.48737910162916e-63
ENSG00000139625	2.05291597695154	5.38217952232118	300.04966484212	3.65368754800453e-67	1.53285241522889e-63
ENSG00000005448	2.22003359290287	4.52871246394248	292.482024287755	1.61720906558702e-65	6.33245163115024e-62
ENSG00000128510	1.8486894818772	8.12307851756572	283.525213683999	1.43626871828282e-63	5.27245269802135e-60
ENSG00000090615	1.88529858198942	6.1597319328219	278.352935895823	1.91671384929033e-62	6.62224634929807e-59
ENSG00000067182	1.82548908657389	7.5115858556092	277.402224202048	3.08612249175523e-62	1.00701891418469e-58
ENSG00000158050	-1.84668605179623	6.23928993329961	276.101339804859	7.13793321382711e-62	2.20656056481124e-58
ENSG00000064666	-1.85118591689233	6.57299128227469	275.326792870349	8.72967729082304e-62	2.56368797838246e-58
ENSG00000123999	2.19745940516381	4.15118492163223	275.179670155556	9.39746483123374e-62	2.62838141363204e-58
ENSG00000262406	-1.90770662742228	5.0062325295315	267.427649306659	4.56986484650027e-60	1.22005005345088e-56

Galaxy practical

Gene symbol annotation

But we do **not see any gene symbol/names** which **we** all **like**

Merge with **HUGO information** (https://www.genenames.org/)

```
Share Data -> Data Libraries -> Bi5444

-> RNA-Seq -> HUGO Gene information
```

Galaxy practical

Gene symbol annotation

Merge with **HUGO information**

Galaxy practical

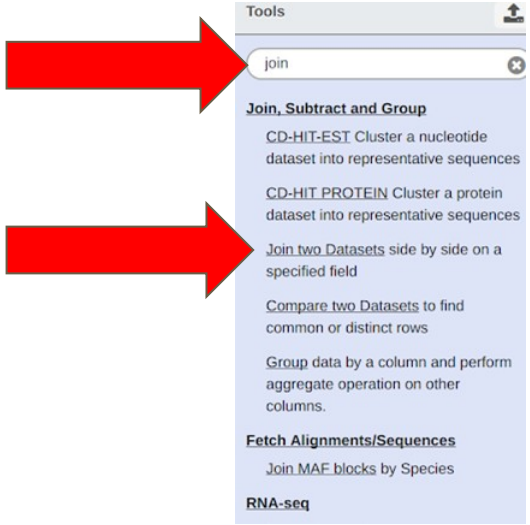
Gene symbol annotation

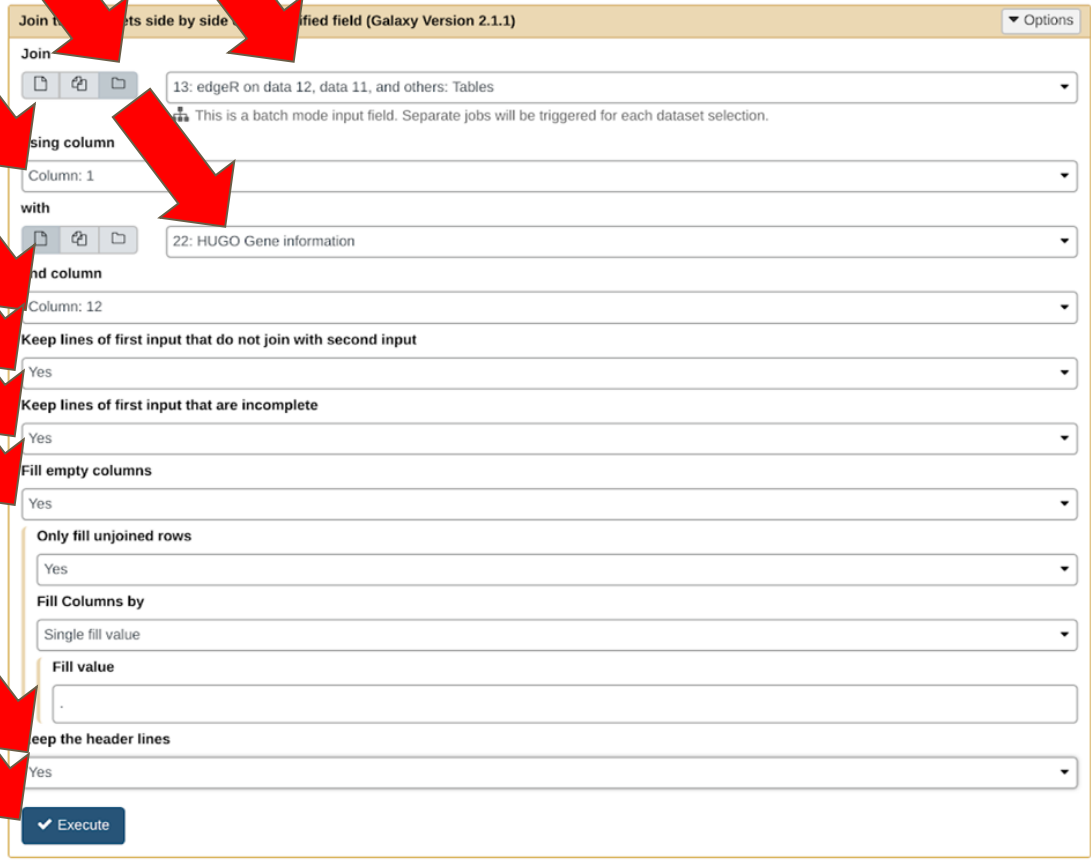
1	2	3
HGNC ID	Approved Symbol	Approved Name
HGNC:5	A1BG	alpha-1-B glycoprotein
HGNC:37133	A1BG-AS1	A1BG antisense RNA 1
HGNC:24086	A1CF	APOBEC1 complementation factor
HGNC:6	A1S9T~withdrawn	symbol withdrawn, see UBA1
HGNC:7	A2M	alpha-2-macroglobulin
HGNC:27057	A2M-AS1	A2M antisense RNA 1
HGNC:23336	A2ML1	alpha-2-macroglobulin like 1
HGNC:41022	A2ML1-AS1	A2ML1 antisense RNA 1
HGNC:41523	A2ML1-AS2	A2ML1 antisense RNA 2
HGNC:8	A2MP1	alpha-2-macroglobulin pseudogene 1
HGNC:9	A2MR~withdrawn	symbol withdrawn, see LRP1
HGNC:10	A2MRAP~withdrawn	symbol withdrawn, see LRPAP1
HGNC:30005	A3GALT2	alpha 1,3-galactosyltransferase 2
HGNC:18149	A4GALT	alpha 1,4-galactosyltransferase (P blood group)
HGNC:17968	A4GNT	alpha-1,4-N-acetylglucosaminyltransferase

9	10	11	12	13
RefSeq IDs	Entrez Gene ID(supplied by NCBI)	RefSeq(supplied by NCBI)	Ensembl ID(supplied by Ensembl)	UCSC ID(supplied by UCSC)
NM_130786	1	NM_130786	ENSG00000121410	uc002qsd.5
NR_015380	503538	NR_015380	ENSG00000268895	uc002qse.3
NM_014576	29974	NM_001198818	ENSG00000148584	uc057tgv.1
NM_000014	2	NM_000014	ENSG00000175899	uc001qvk.2
NR_026971	144571	NR_026971	ENSG00000245105	uc009zgj.2
NM_144670	144568	NM_001282424	ENSG00000166535	uc001quz.6
	100874108		ENSG00000256661	uc058kxy.1
	106478979		ENSG00000256904	uc058kyb.1
NG_001067	3	NR_040112	ENSG00000256069	
NM_001080438	127550	NM_001080438	ENSG00000184389	uc031plq.1
NM_017436	53947	NM_001318038	ENSG00000128274	uc062ewl.1
NM_016161	51146	NM_016161	ENSG00000118017	uc003ers.2

Galaxy practical

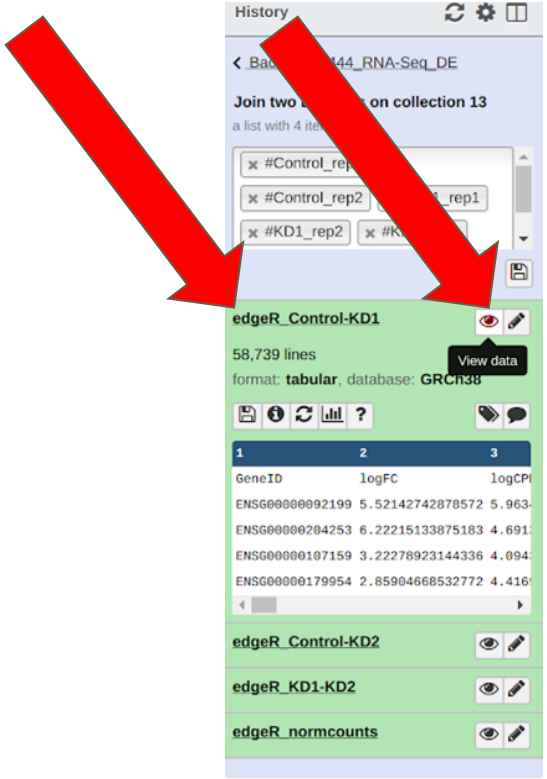
Gene symbol annotation





Galaxy practical

Differential gene expression



Galaxy practical

Differential gene expression

1	2	3	4	5	6	7	8	9	10	11
GeneID	logFC	logCPM	F	PValue	FDR	HGNC ID	Approved Symbol	Approved Name	Status	Previous Symbols
ENSG00000092199	5.52142742878572	5.96346716277756	1462.34679277333	1.22114253157414e-318	7.17238065920072e-314	HGNC:5035	HNRNPC	heterogeneous nuclear ribonucleoprotein C (C1/C2)	Approved	HNRPC
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
ENSG00000107159	3.22278923144336	4.09433985683369	491.850213415037	7.9546272105367e-109	1.55738343070291e-104	HGNC:1383	CA9	carbonic anhydrase 9	Approved	
ENSG00000179954	2.85904668532772	4.41691242725208	456.438217554854	3.85225443835843e-101	5.65655411092455e-97	HGNC:26641	SSC5D	scavenger receptor cysteine rich family member with 5 domains	Approved	
ENSG00000148204	2.70384863075395	4.17441363845286	376.393049584693	1.77411170612e-83	2.08404902117916e-79	HGNC:18688	CRB2	crumbs 2, cell polarity complex component	Approved	
ENSG00000185633	4.22087072157582	2.41739313442468	358.912648383713	5.83527380494571e-80	5.71224678222477e-76	HGNC:29836	NDUFA4L2	NDUFA4, mitochondrial complex associated like 2	Approved	
ENSG00000204560	2.10676298079	5.93980087019778	334.313475363922	1.29573528700066e-74	1.08721445831406e-70	HGNC:2739	DHX16	DEAH-box helicase 16	Approved	DDX16
ENSG00000099998	3.08600815349408	3.0528996991664	331.974819220342	4.17737211080344e-74	2.78770162619006e-70	HGNC:4260	GGT5	gamma-glutamyltransferase 5	Approved	GGTLA1
ENSG00000197375	2.12482893907467	5.68742165552465	331.930250691321	4.2716122645289e-74	2.78770162619006e-70	HGNC:10969	SLC22A5	solute carrier family 22 member 5	Approved	CDSP
ENSG00000169231	2.13377519397903	5.38735924665232	322.649201605679	4.44958188736524e-72	2.61346192154398e-68	HGNC:11787	THBS3	thrombospondin 3	Approved	
ENSG00000157224	2.09164554071042	5.5519631536178	312.92589063867	5.78761591740723e-70	3.09032382644467e-66	HGNC:2034	CLDN12	claudin 12	Approved	
ENSG00000157193	-1.90372177023031	7.73296494003751	302.848534430836	8.99503831407289e-68	4.40269646147559e-64	HGNC:6700	LRP8	LDL receptor related protein 8	Approved	
ENSG00000172346	2.62579315698578	3.626882383666	300.257776550655	3.29206236846498e-67	1.48737910162916e-63	HGNC:30359	CSDC2	cold shock domain containing C2	Approved	
ENSG00000139625	2.05291597695154	5.38217952232118	300.04966484212	3.65368754800453e-67	1.5328524152289e-63	HGNC:6851	MAP3K12	mitogen-activated protein kinase kinase kinase 12	Approved	ZPK
ENSG00000005448	2.22003359290287	4.52871246394248	292.482024287755	1.61720906558702e-65	6.33245163115024e-62	HGNC:25770	WDR54	WD repeat domain 54	Approved	
ENSG00000128510	1.8486894818772	8.12307851756572	283.525213683999	1.43626871828282e-63	5.27245269802135e-60	HGNC:15740	CPA4	carboxypeptidase A4	Approved	
ENSG00000090615	1.88529858198942	6.1597319328219	278.352935895823	1.91671384929033e-62	6.62224634929807e-59	HGNC:4426	GOLGA3	golgin A3	Approved	
ENSG00000067182	1.82548908657389	7.5115858556092	277.402224202048	3.08612249175523e-62	1.00701891418469e-58	HGNC:11916	TNFRSF1A	TNF receptor superfamily member 1A	Approved	TNFR1
ENSG00000158050	-1.84668605179623	6.23928993329961	276.101339804859	7.13793321382711e-62	2.20656056481124e-58	HGNC:3068	DUSP2	dual specificity phosphatase 2	Approved	
ENSG00000064666	-1.85118591689233	6.57299128227469	275.326792870349	8.72967729082304e-62	2.56368797838246e-58	HGNC:2156	CNN2	calponin 2	Approved	
ENSG00000123999	2.19745940516381	4.15118492163223	275.179670155556	9.3974648312374e-62	2.62838141363204e-58	HGNC:6065	INHA	inhibin subunit alpha	Approved	
ENSG00000262406	-1.90770662742228	5.0062325295315	267.427649306659	4.56986484650027e-60	1.22005005345088e-56	HGNC:7158	MMP12	matrix metallopeptidase 12	Approved	
ENSG00000159792	1.87315272517581	5.81432694173605	265.276959847098	1.34261834082599e-59	3.42863861949629e-56	HGNC:9529	PSKH1	protein serine kinase H1	Approved	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

Galaxy practical

Alignment coverage

Visualization of **coverage** of **aligned** data (and expressed exons)

Using **`deepTools`** -> **`bamCoverage`**

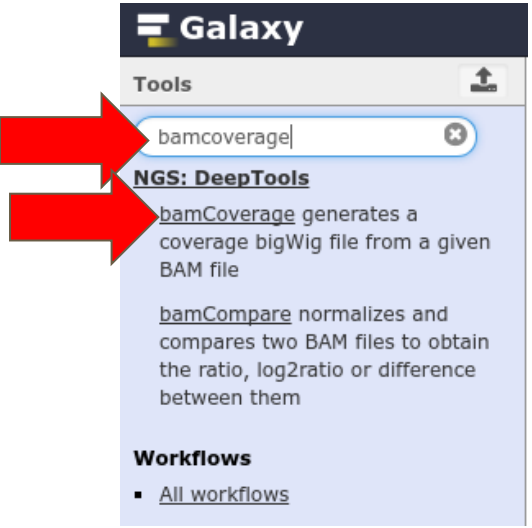
Input **BAM**, output **BIGWIG**

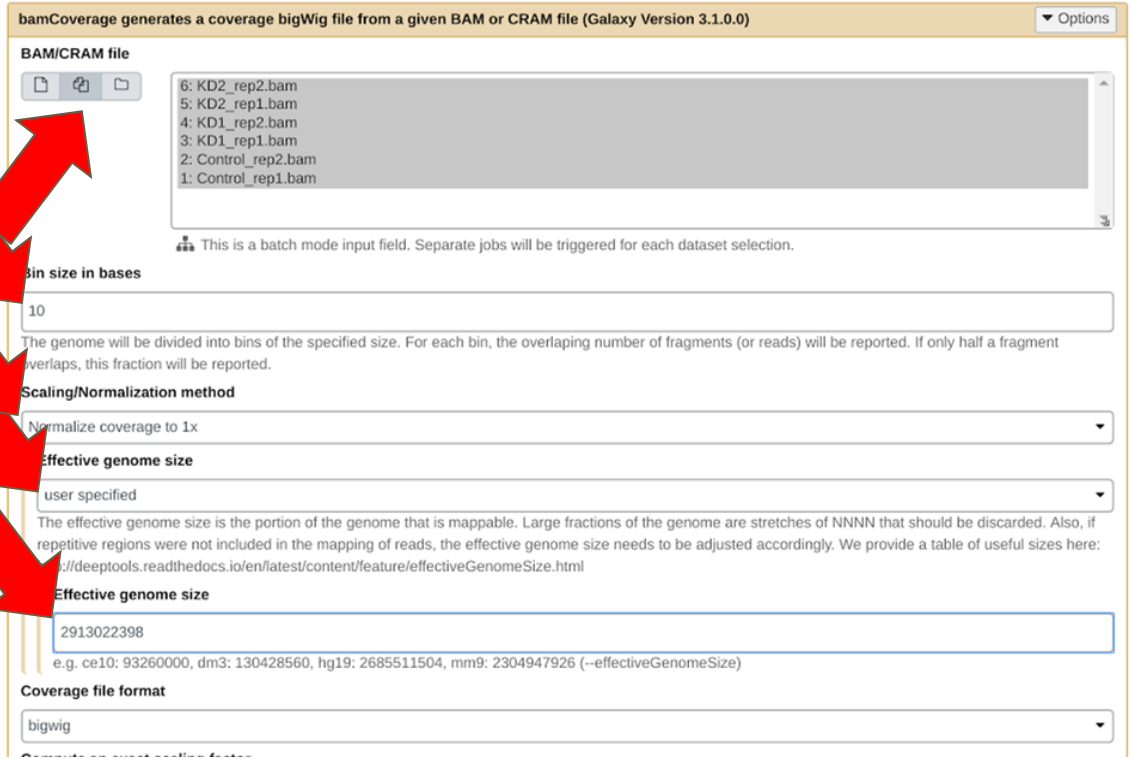
Effective genome size (hg38): **2913022398**

Galaxy practical

https://deeptools.readthedocs.io/en/develop/content/feature/effectiveGenomeSize.html

Alignment coverage





Galaxy practical

Alignment coverage

BIGWIG coverage

Galaxy practical

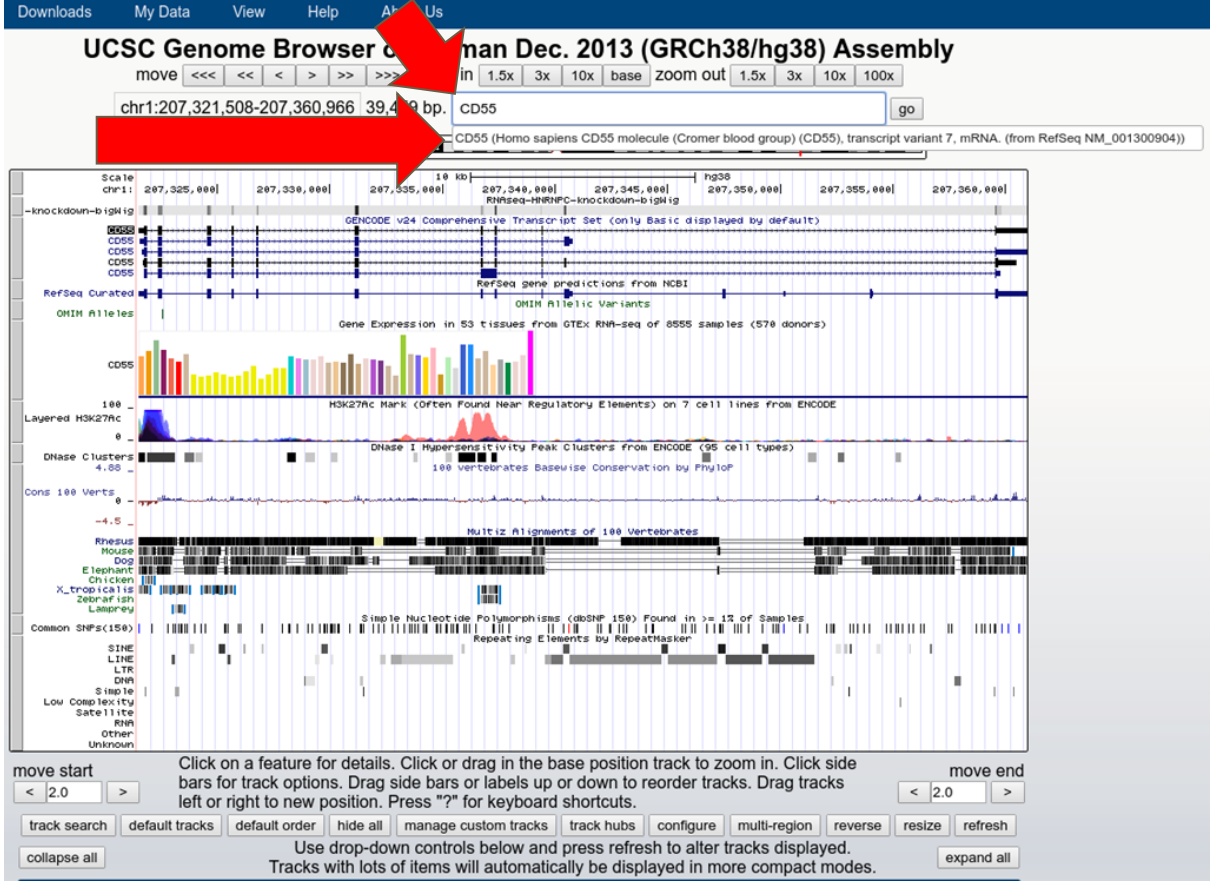
Alignment coverage

BIGWIG visualization

Galaxy practical

Alignment coverage

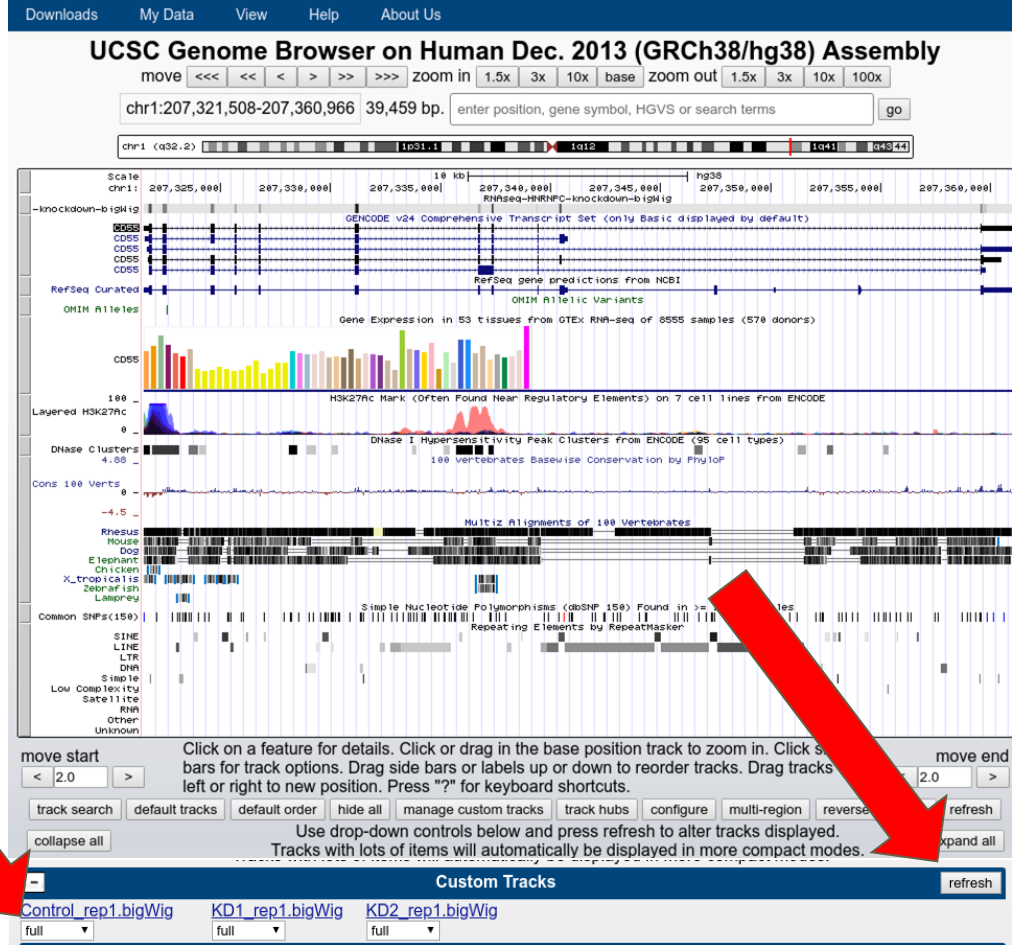
CD55 region



Galaxy practical

Alignment coverage

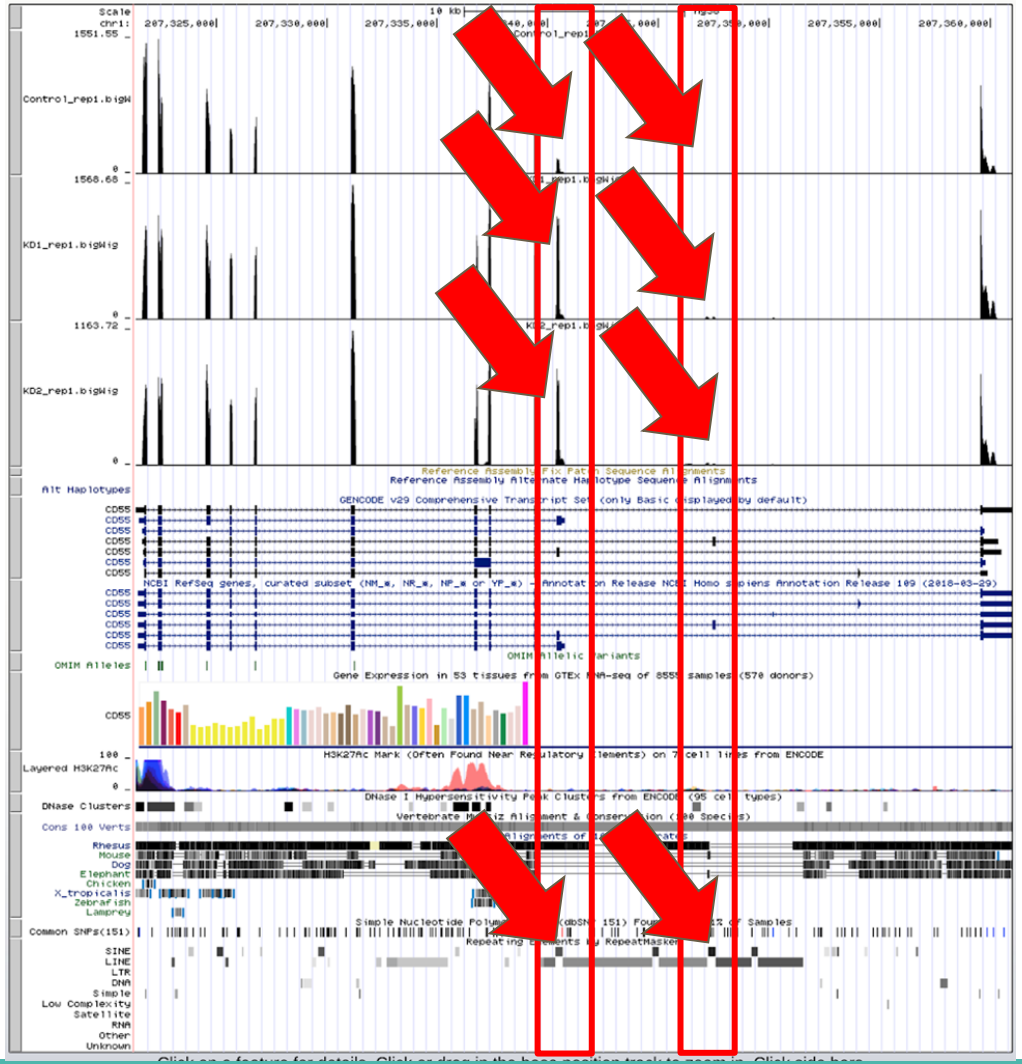
BIGWIG size



Galaxy practical

Alignment coverage

Now you do it for the **other two BIGWIG files**


Scale
chr1:
207,325,000
207,330,000
207,335,000
207,355,000
207,360,000
10 kb
1551.55
Control_rep1.bigW
0
1568.68
KD1_rep1.bigWig
0
1163.72
KD2_rep1.bigWig
0
Alt Haplotypes
CD55
OMIM Alleles
Gene Expression in 53 tissues from GTEx RNA-seq of 8555 samples (570 donors)
100
Layered H3K27Ac
0
DNase Clusters
Cons 100 Verts
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Lamprey
Common SNPs(151)
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

Galaxy practical

Alignment coverage

If something went wrong, history of DE and coverage visualization

`Shared Data -> Histories -> Bi5444_RNA-Seq_DE_full_history`

RNA-Seq data analysis - pipeline in Galaxy

1. **Initial quality check** - `FastQC`
    - Check for overall quality of the data, number of reads, read length distribution, ...
2. **Preprocessing** - `Trimmomatic`
    - Remove adapters, low quality ends, unwanted sequences, ...
3. **Alignment** - `STAR`
    - Map reads to the reference genome
4. **Alignment quality check** - `STAR log, featureCounts`
    - Check overall alignment statistics
5. **Genome coverage (peaks)** - `bamCoverage`
    - Get overview of mapped positions in the genome
6. **Gene annotation** - `UCSC Main table browser`
    - Get gene annotations for reference genome
7. **Quantification** - `featureCounts`
    - Get gene read counts
8. **Differential gene expression** - `edgeR`,`DESeq2` (genes)
    - Differences between conditions

RNA-Seq data analysis - other possibilities

1. **Initial quality check** - `FastQC`
   - Check for overall quality of the data, number of reads, read length distribution, ...
2. **Preprocessing** - `Cutadapt, BBTools, seqtk`
   - Remove adapters, low quality ends, unwanted sequences, ...
3. **Alignment** - `GSNAP, Bowtie2`
   - Map reads to the reference genome
4. **Alignment quality check** - `Picard tools, RSeQC, Qualimap`
   - Check overall alignment statistics
5. **Genome coverage (peaks)** - `STAR + bedGraphToBigWig, Bedtools`
   - Get overview of mapped positions in the genome
6. **Gene annotation** - `UCSC, Ensembl, NCBI`
   - Get gene annotations for reference genome
7. **Quantification** - `RSEM, HTSeq, Salmon, Kallisto`
   - Get gene read counts
8. **Differential gene expression** - `DEXSeq` (exons), `baySeq` (genes)
   - Differences between conditions
9. **Gene ontology and pathways -** `g:Profiler, KEGG`
   - Check ontologies and pathways for selected genes