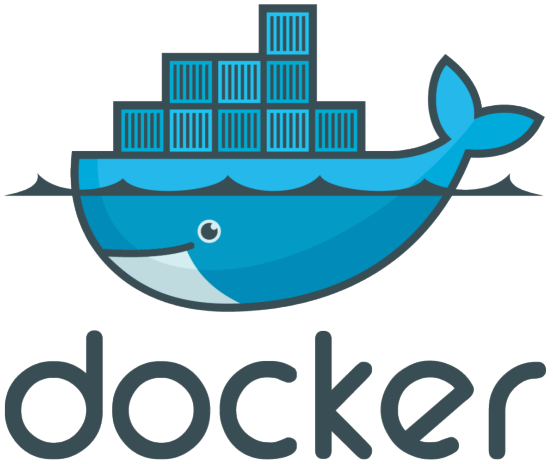
RRBS Analysis

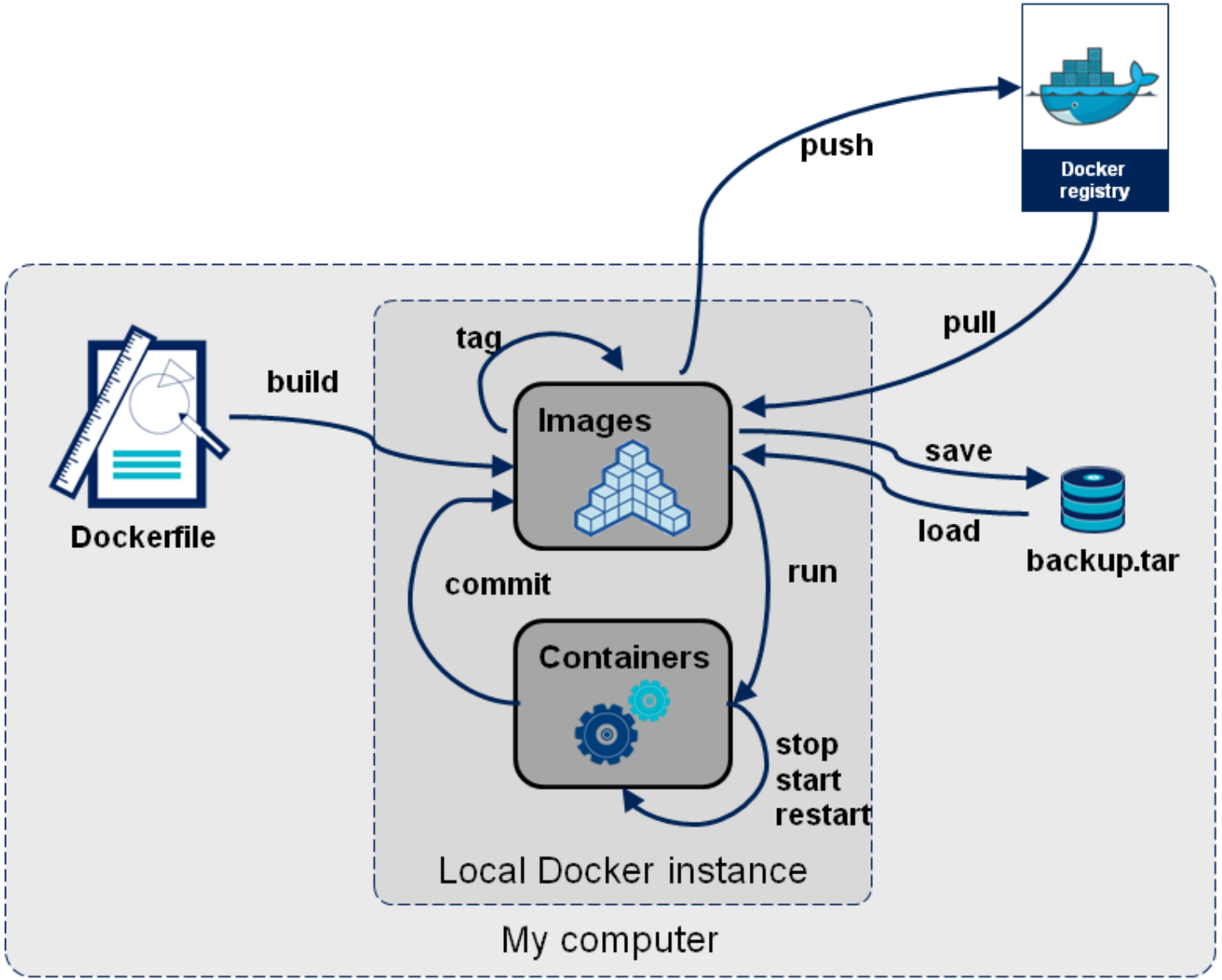
Konstantin Riege
FLI Jena

Rstudio Installation

- Docker is a virtualization technique to setup an uniform Linux environment in which necessary software can be installed
- The whole environment can be shipped as an “image”
- Docker “containers” can be executed on any operating system



R-Studio Installation



R-Studio Installation

- Open up a terminal and type 2 commands

```
docker pull rocker/rstudio

docker run -d -name rstudio --rm -p
8080:8787 -e PASSWORD=rstudio
rocker/rstudio
```

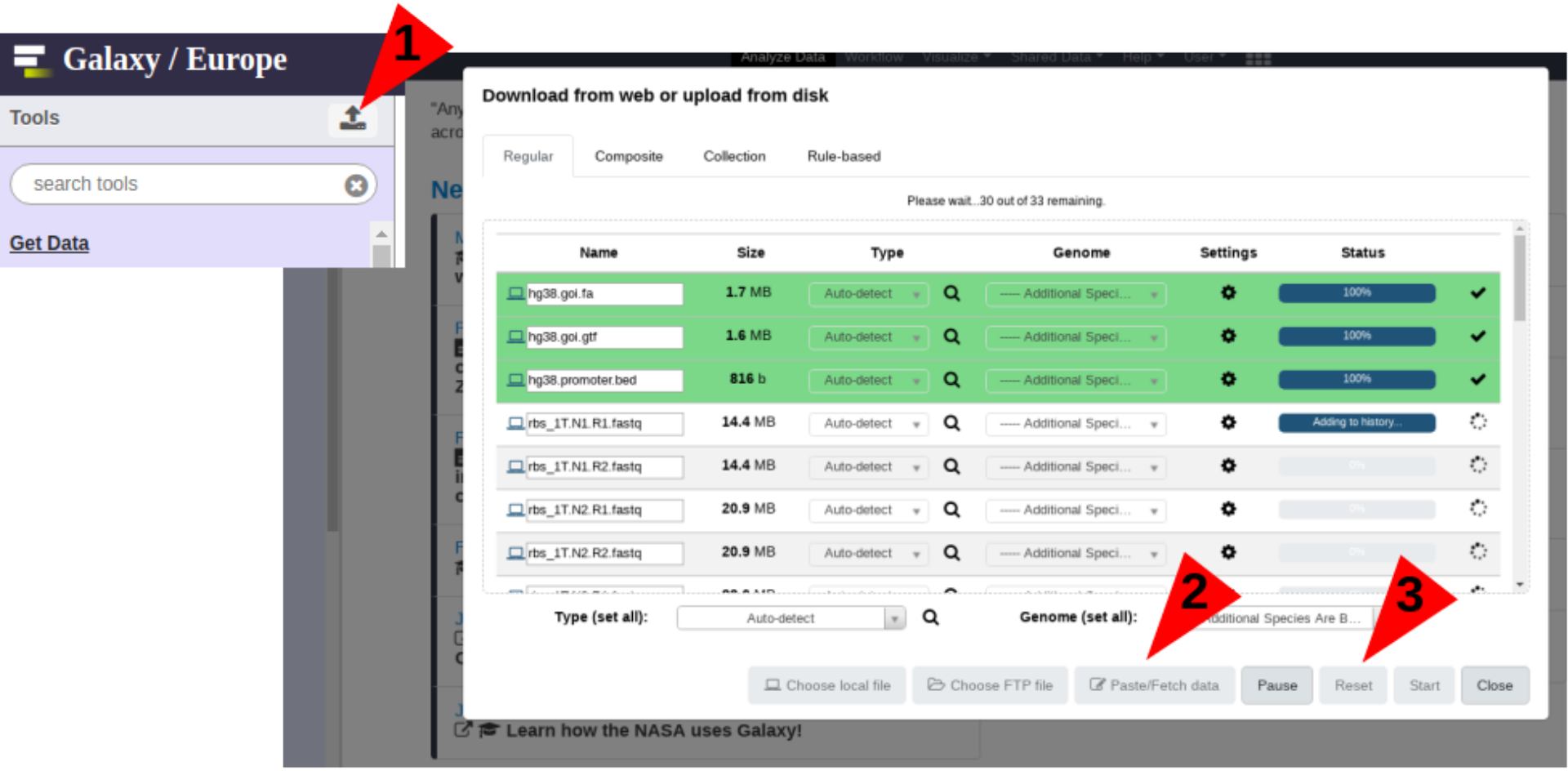
- Open webbrowser on localhost:8080

```
docker stop rstudio
```

Upload Data

https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N1.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N1.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N2.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N2.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N3.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N3.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N4.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_1T.N4.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N1.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N1.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N2.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N2.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N3.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N3.R2.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N4.R1.fastq.gz
https://github.com/destairdenbi/trainings/raw/master/raw_data/bs_tour/rbs_3T.N4.R2.fastq.gz

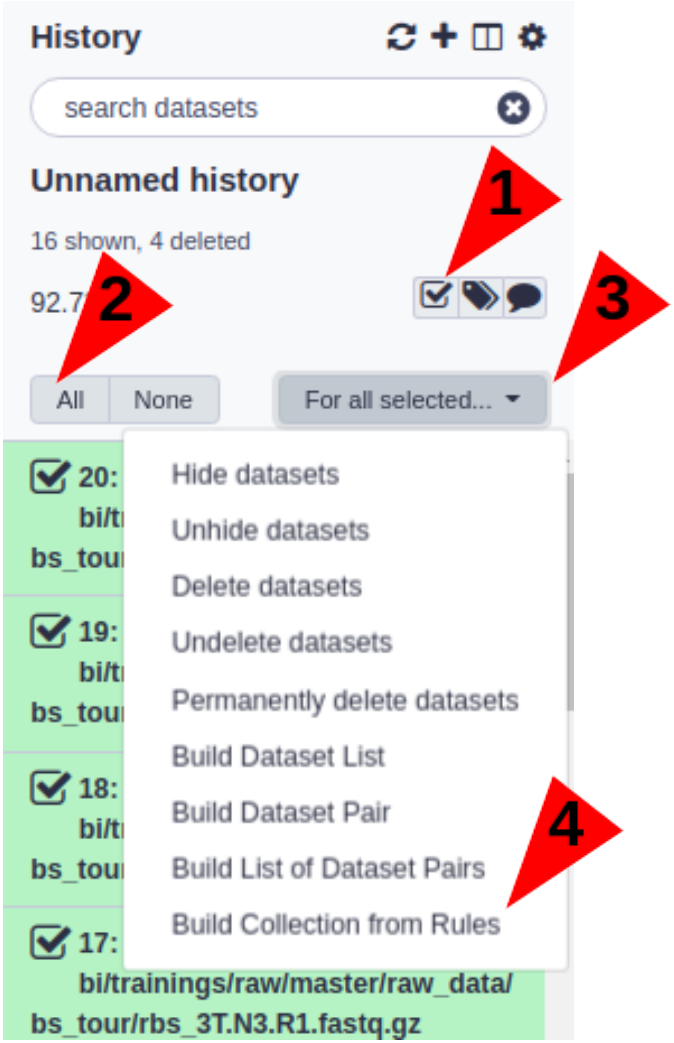
Upload Data


Galaxy / Europe
Tools
search tools
Get Data
1
Download from web or upload from disk
Regular
Composite
Collection
Rule-based
Please wait...30 out of 33 remaining.
Name
Size
Type
Genome
Settings
Status
hg38.goi.fa
1.7 MB
hg38.goi.gtf
1.6 MB
hg38.promoter.bed
816 b
rbs_1T.N1.R1.fastq
14.4 MB
rbs_1T.N1.R2.fastq
14.4 MB
rbs_1T.N2.R1.fastq
20.9 MB
rbs_1T.N2.R2.fastq
20.9 MB
Auto-detect
----- Additional Speci...
100%
Adding to history...
Type (set all):
Genome (set all):
2
3
Choose local file
Choose FTP file
Paste/Fetch data
Pause
Reset
Start
Close
Learn how the NASA uses Galaxy!

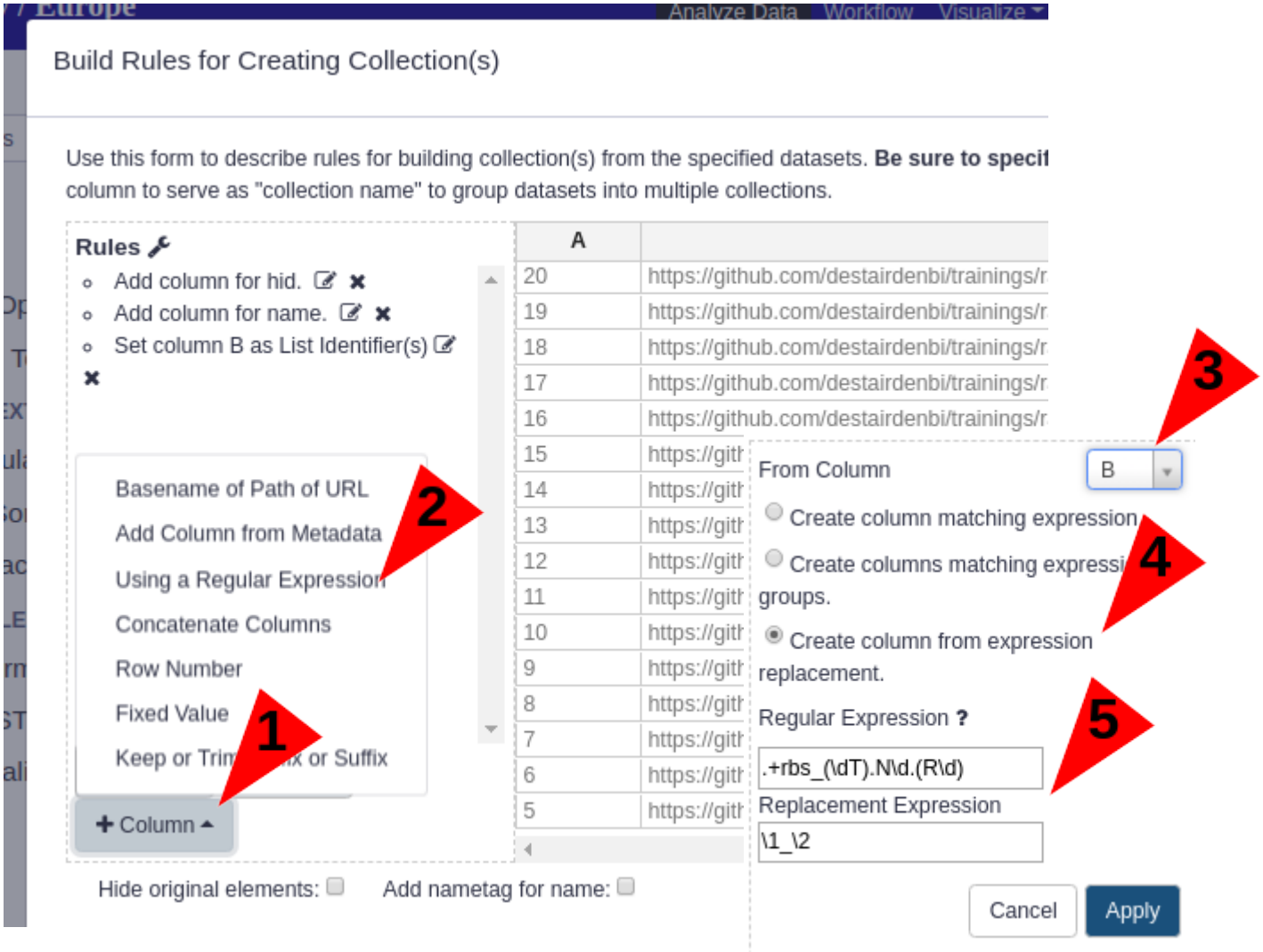
Dataset Collections

- Enable one-select-batch operations
- Keep history small
- Allows to track input file names

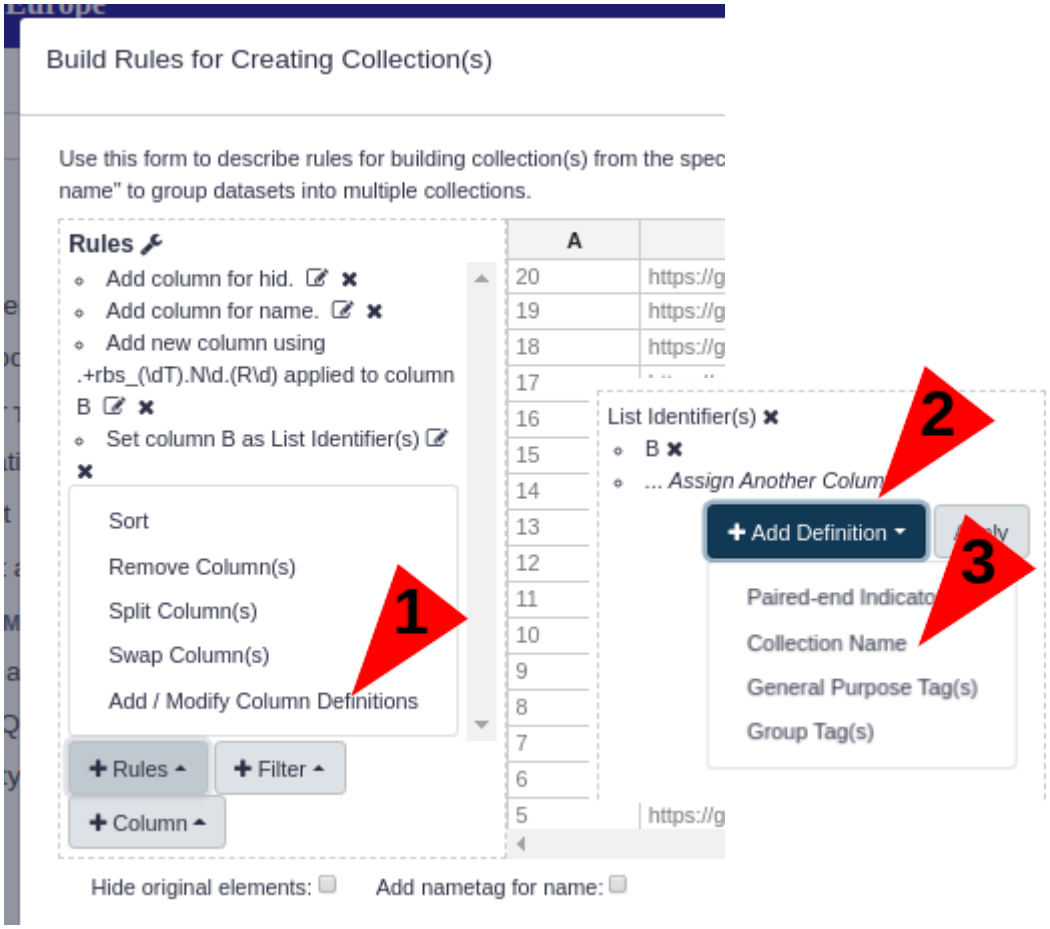
Create Collections

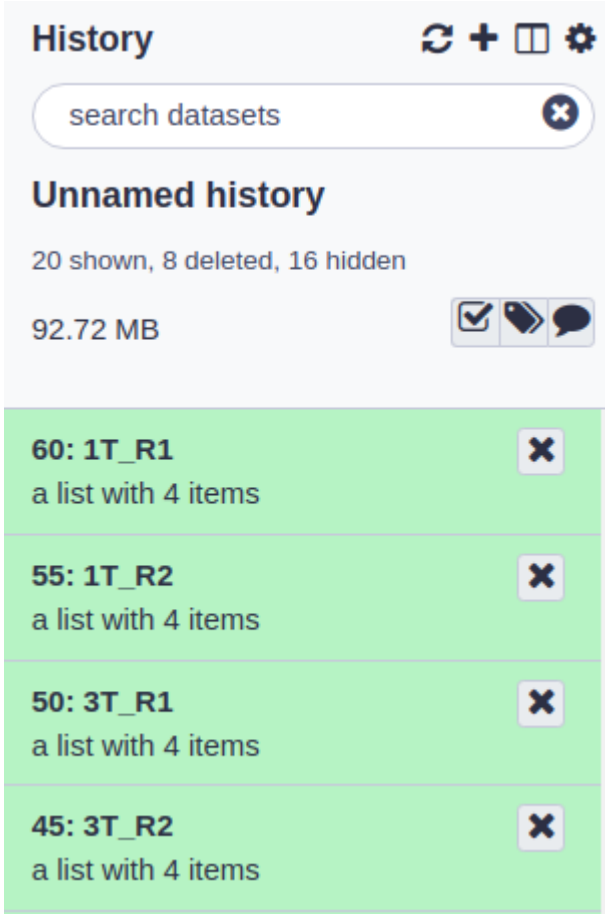


Create Collections

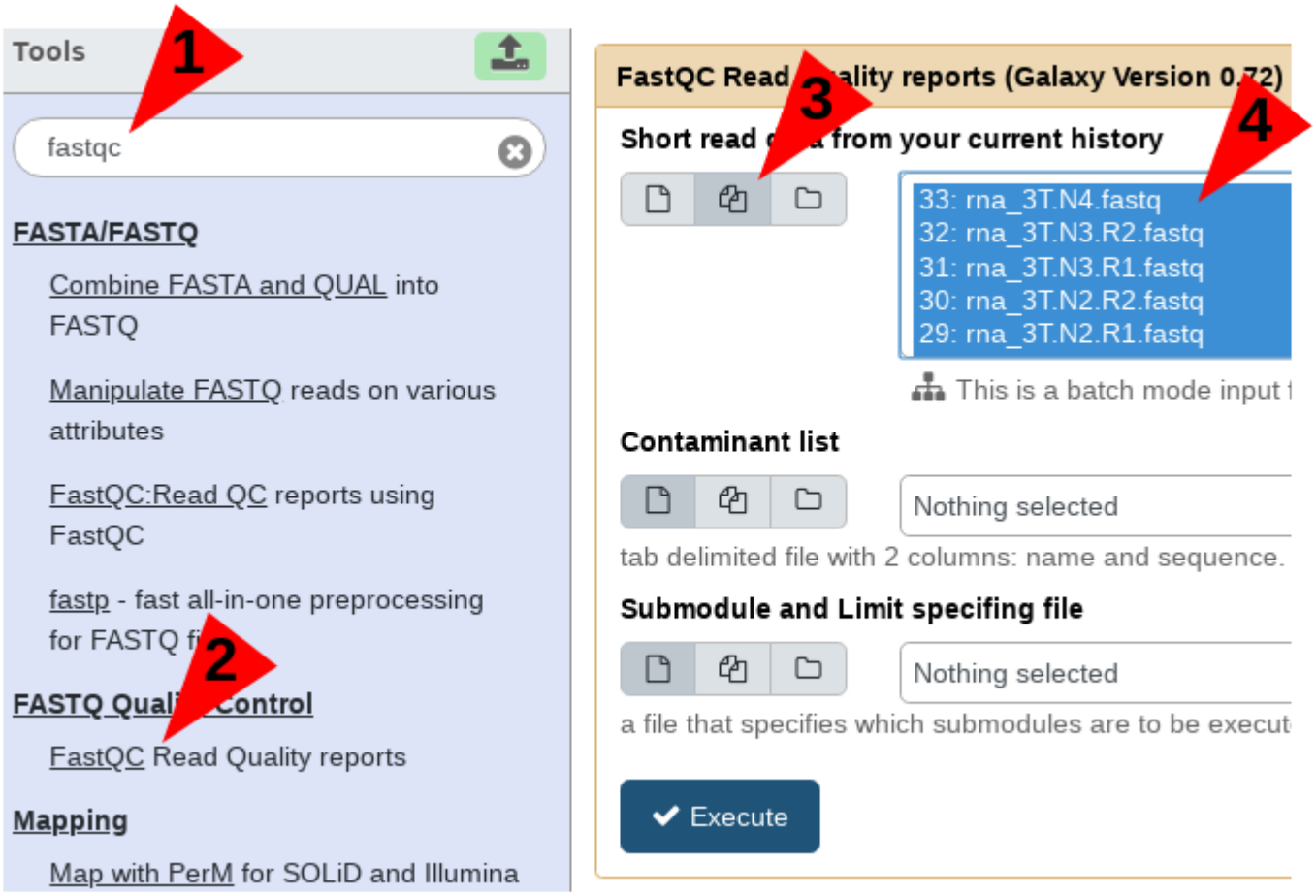


Create Collections

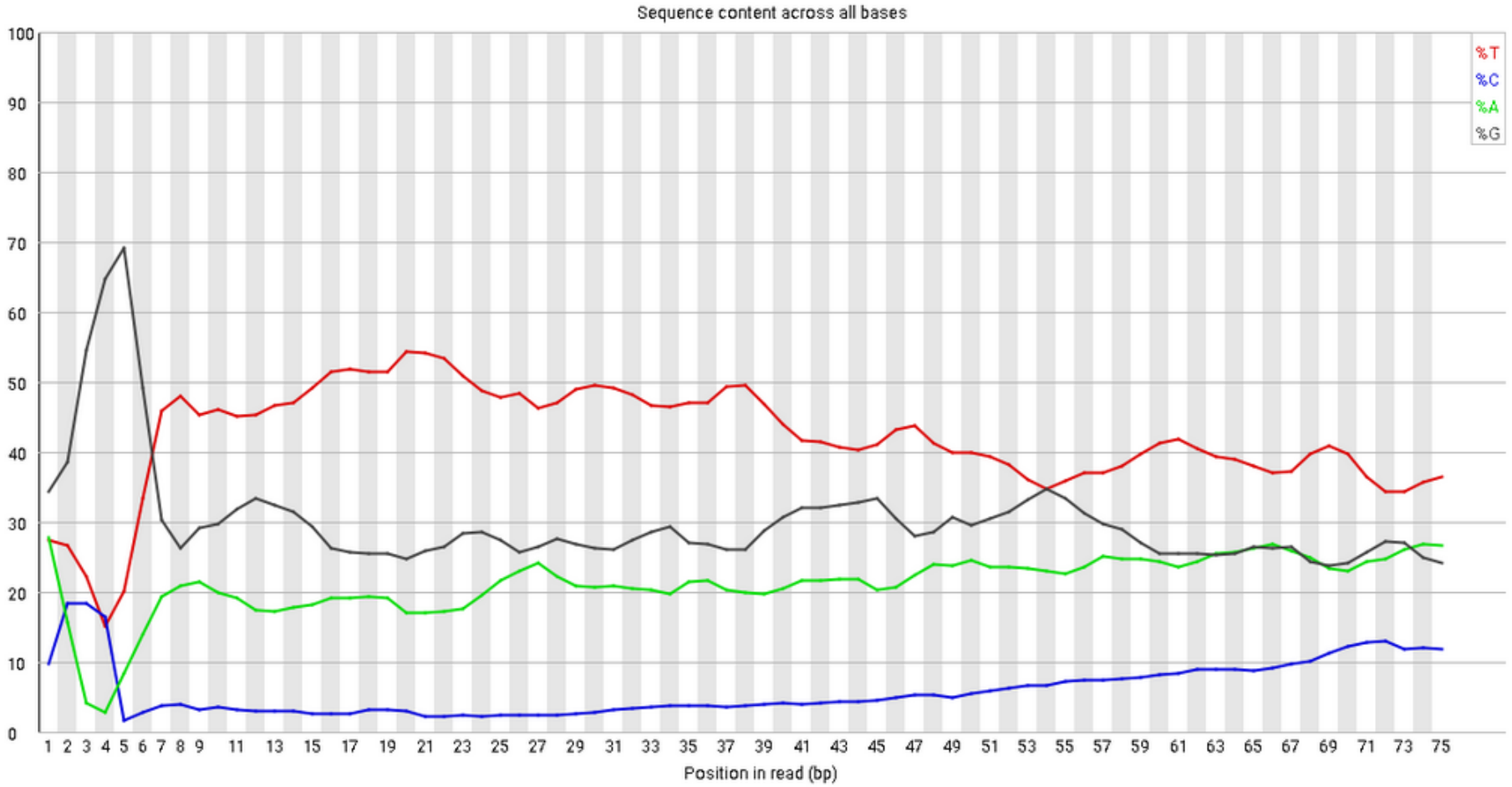




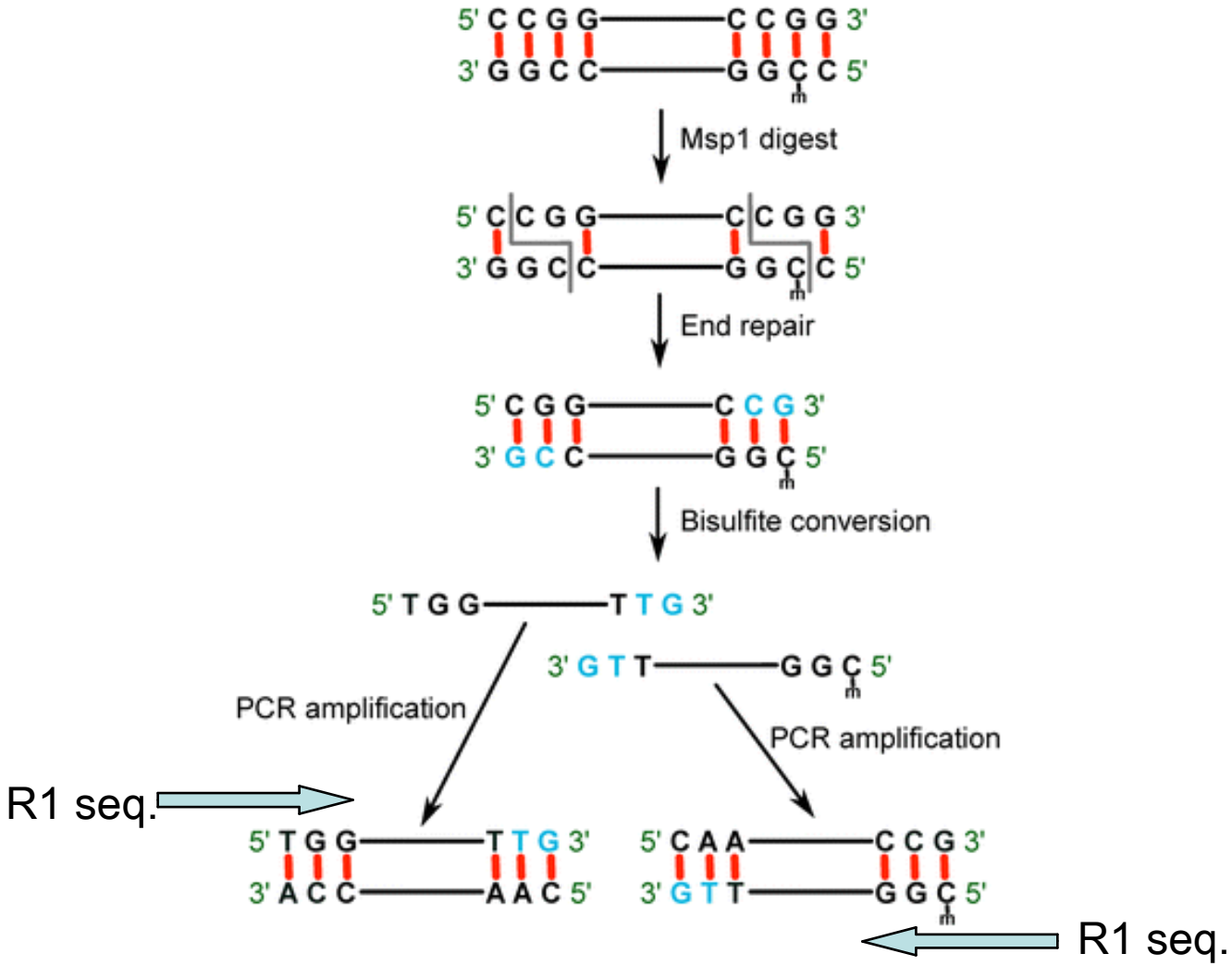
Quality Control



Quality Control

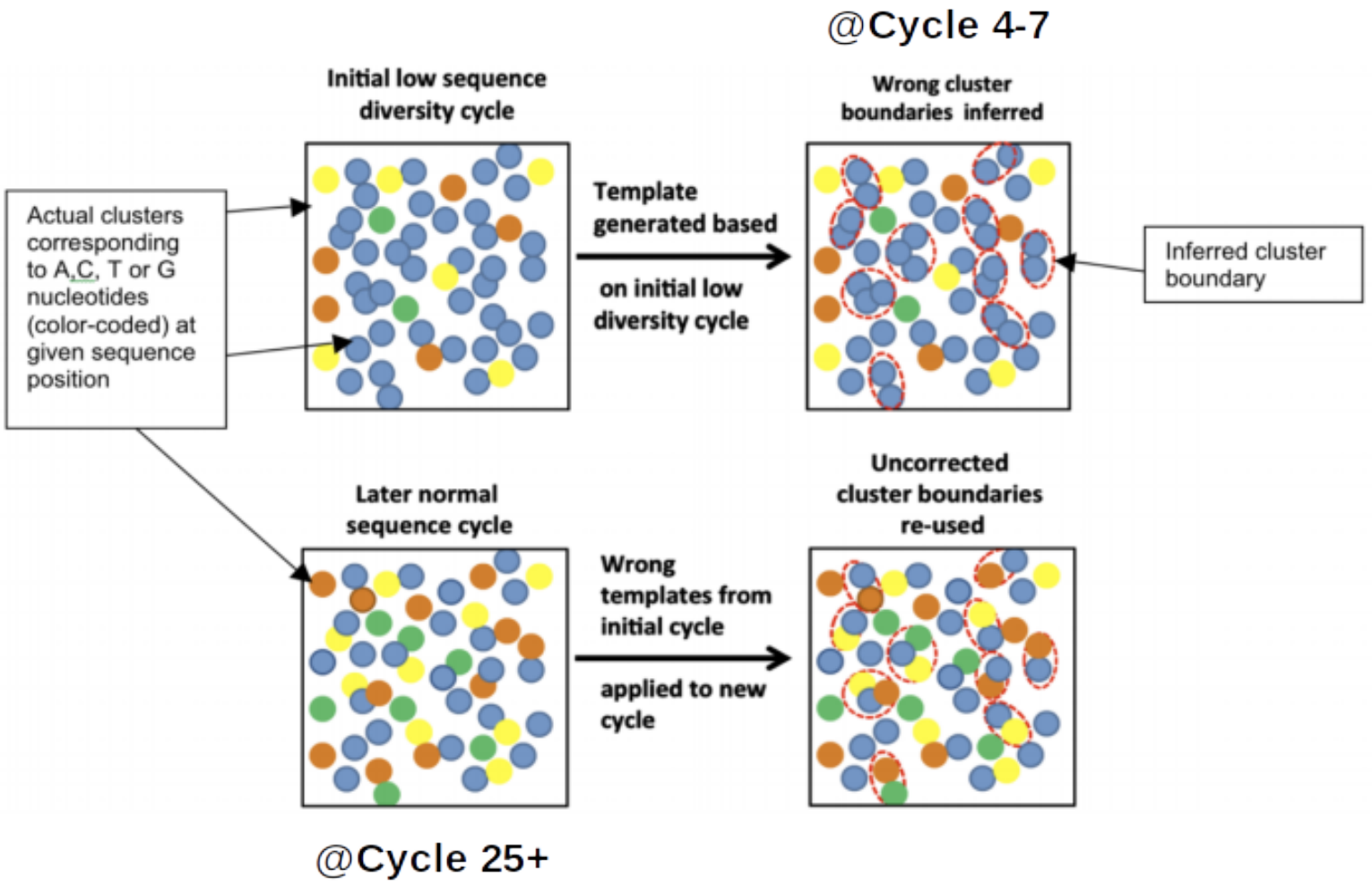


CpG enrichment by MspI



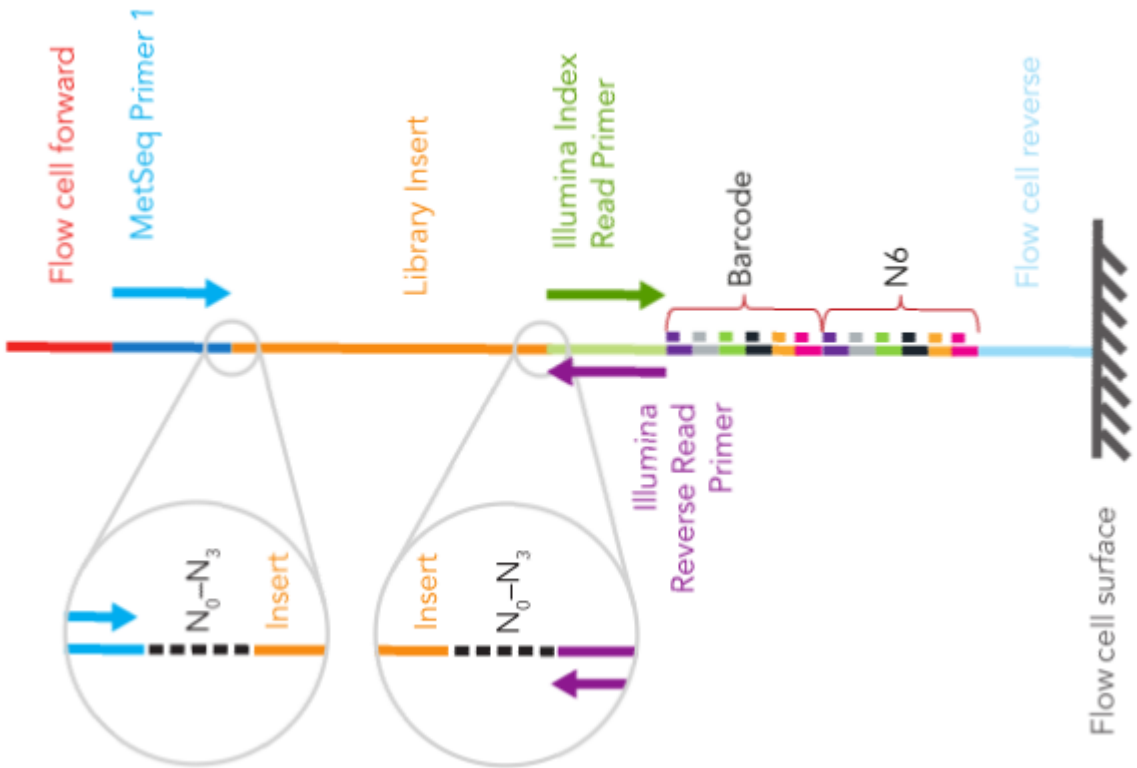
Baheti, S., Kanwar, R., Goelzenleuchter, M. et al. BMC Genomics (2016)

Low complexity bias

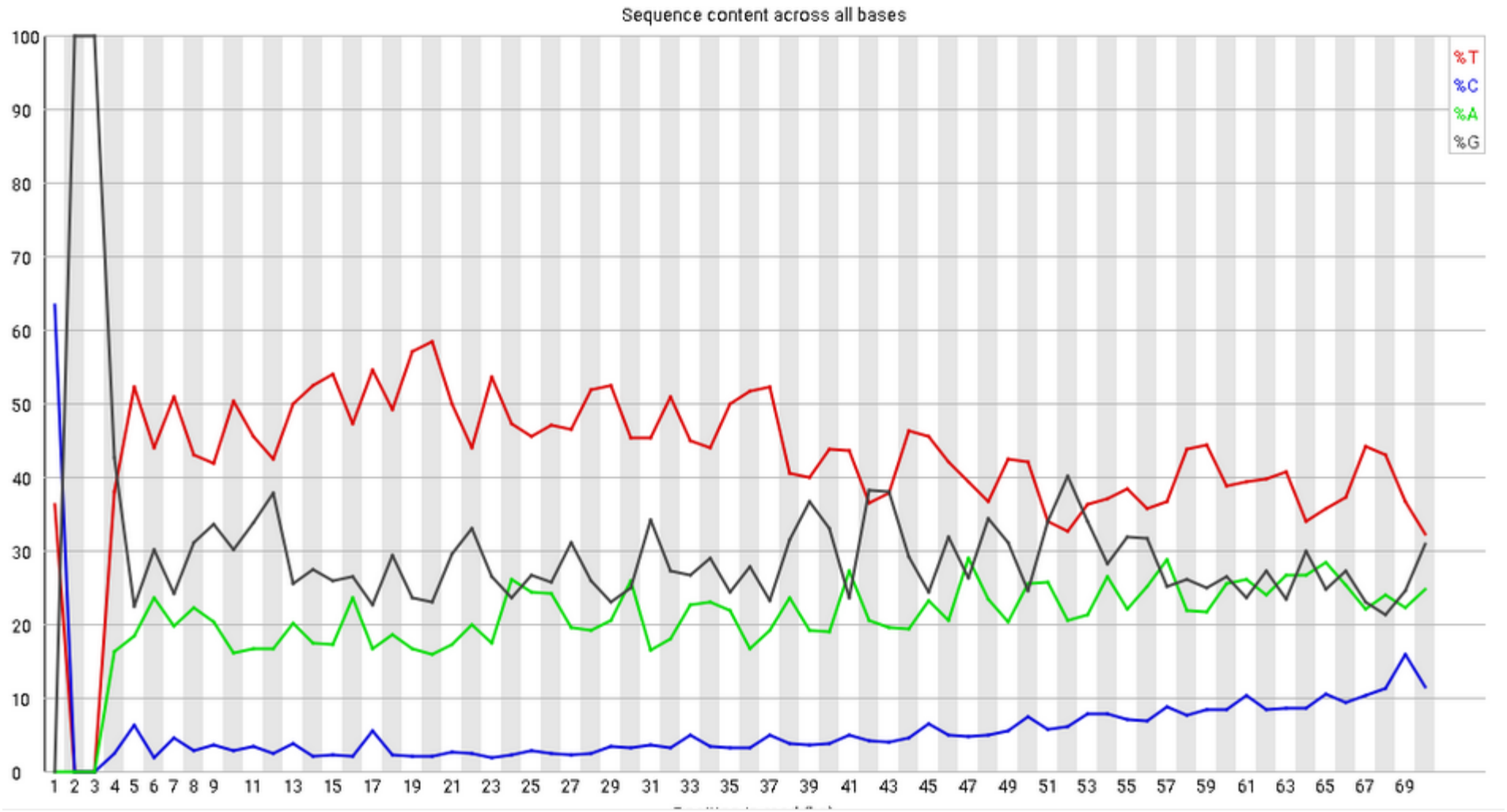


Diversity Adapter

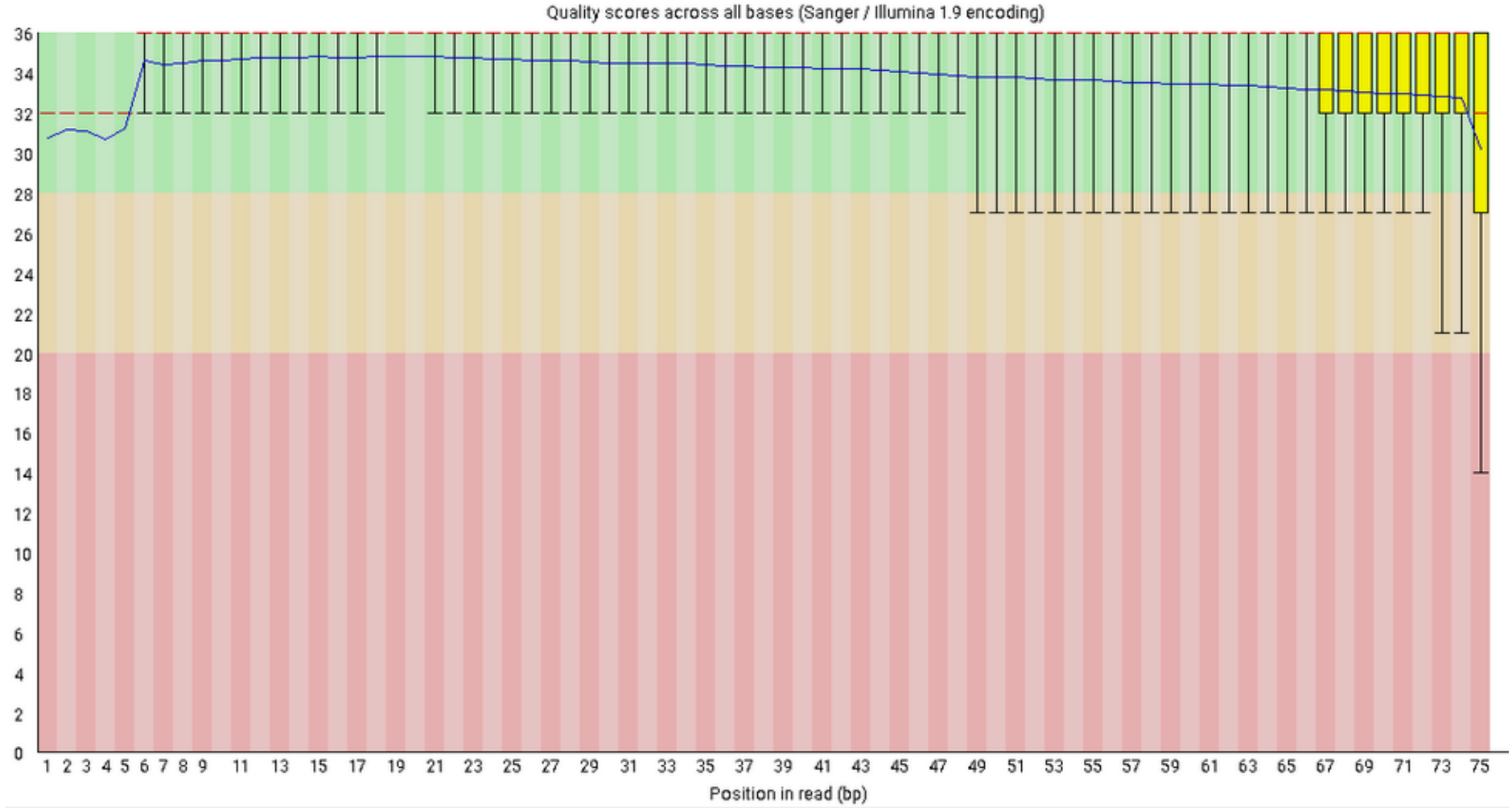
0 diversity bases added	5′	**YGG(BCO)**	**BCO** = Bisulfite-Converted Original genomic strand
1 diversity base added	5′	**DYGG(BCO)**	**Y** = C or T, depending on methylation status
2 diversity bases added	5′	**DDYGG(BCO)**	**D** = A, G, or T
3 diversity bases added	5′	**RDDYGG(BCO)**	**R** = A or G



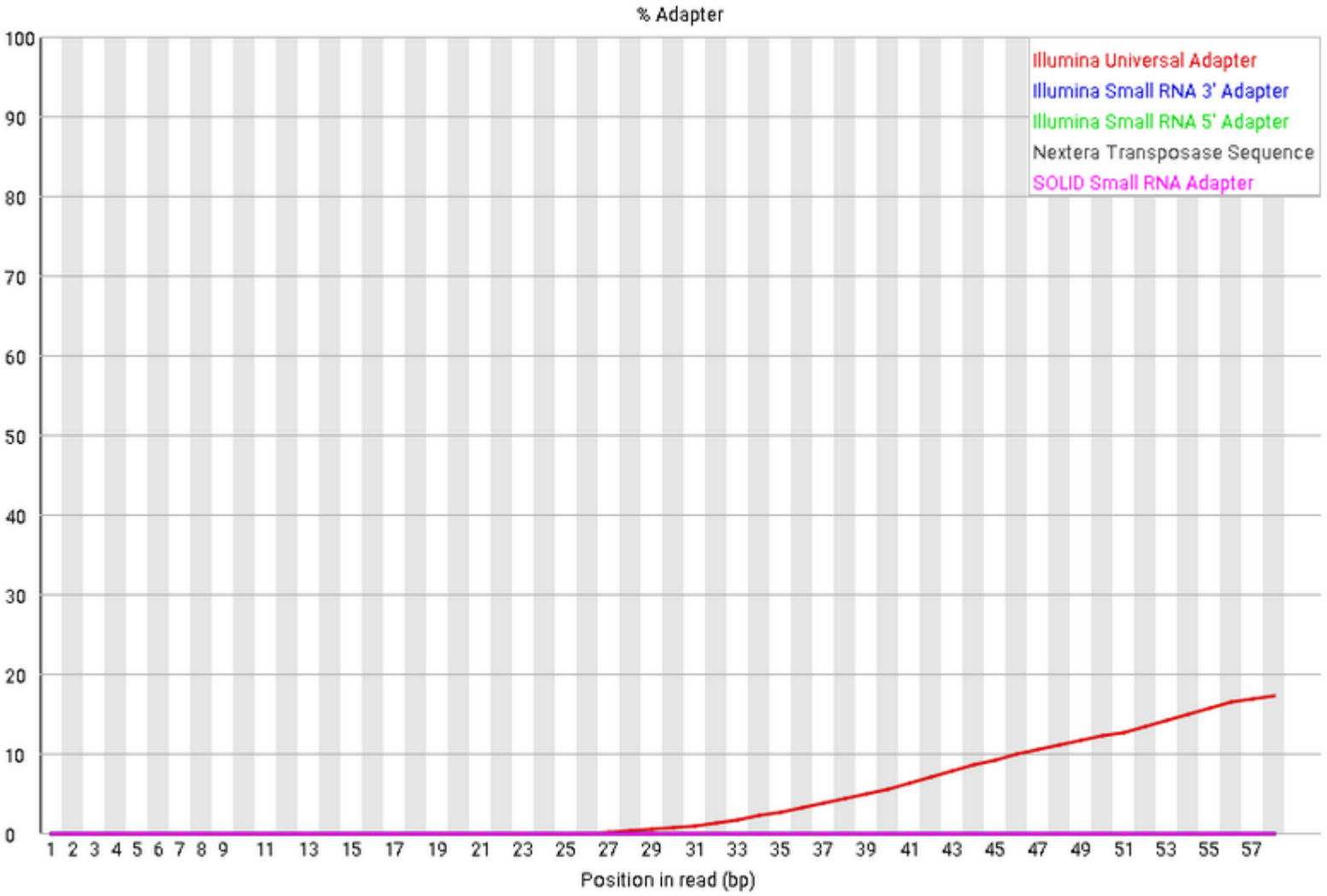
Quality Control



Quality Control



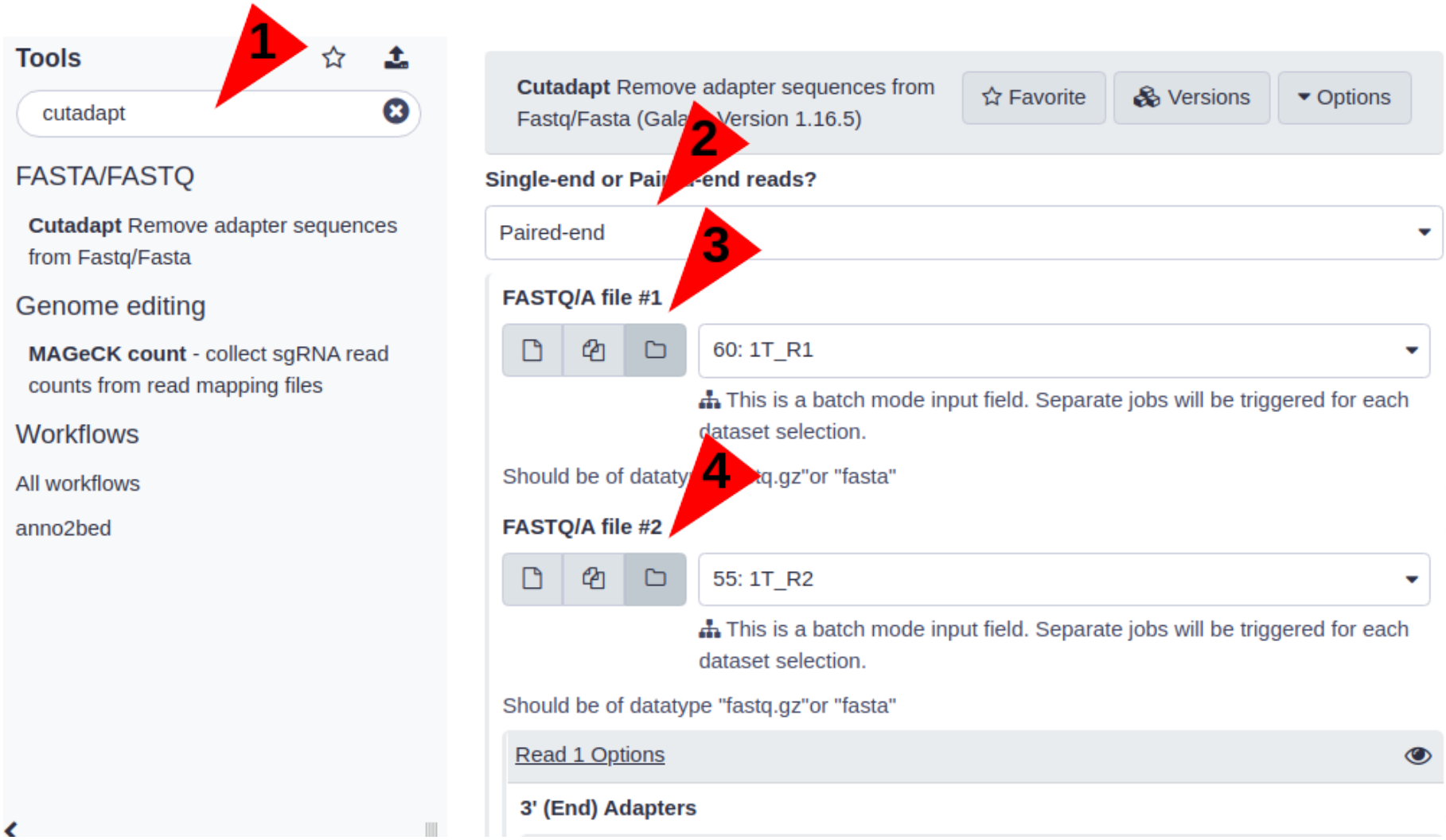
Adapter Clipping



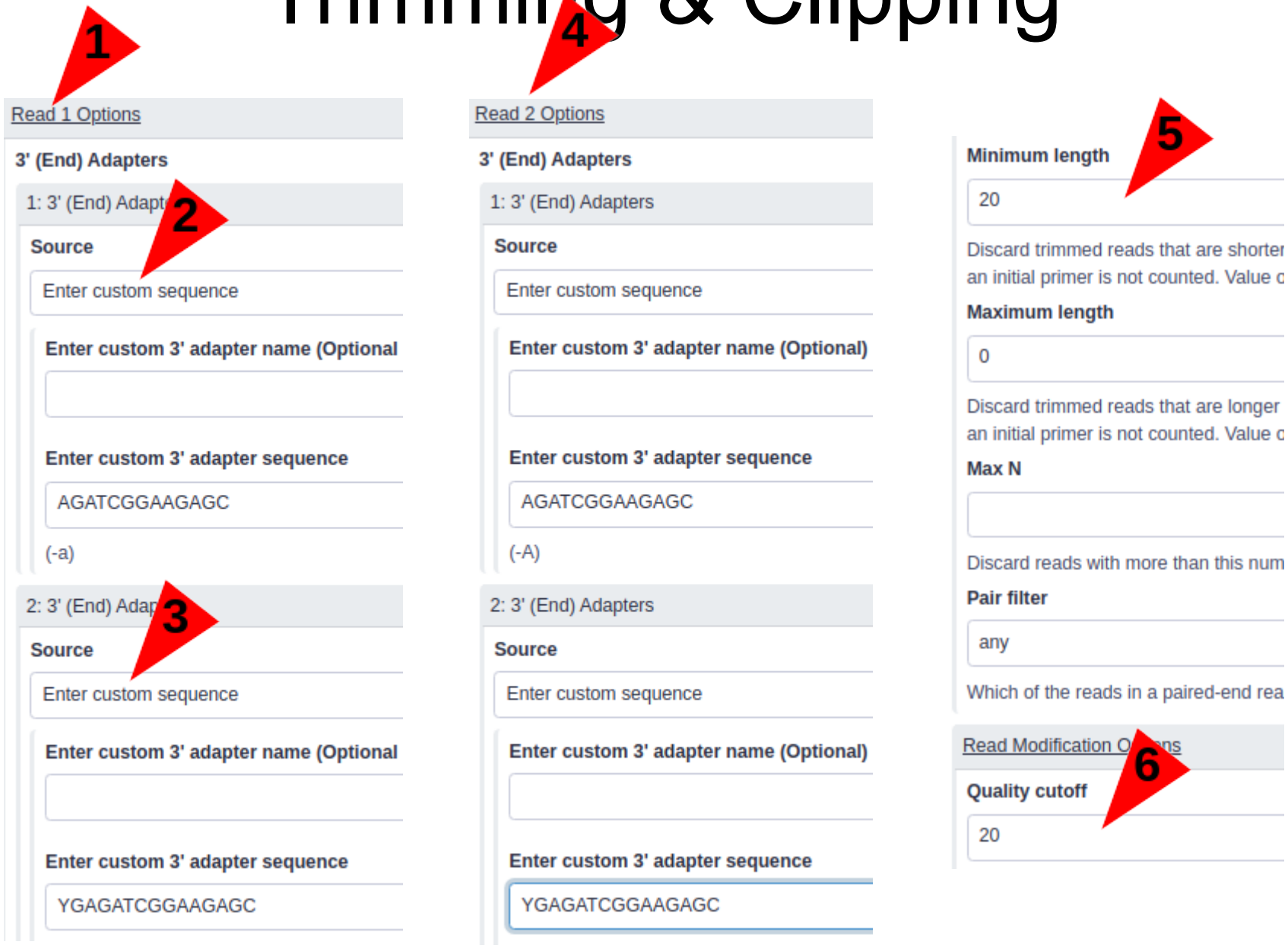
Adapter Clipping

- Universal Illumina adapter: AGATCGGAAGAGC
- Tackle end-repair bias in methylation with second adapter sequence YGAGATCGGAAGAGC

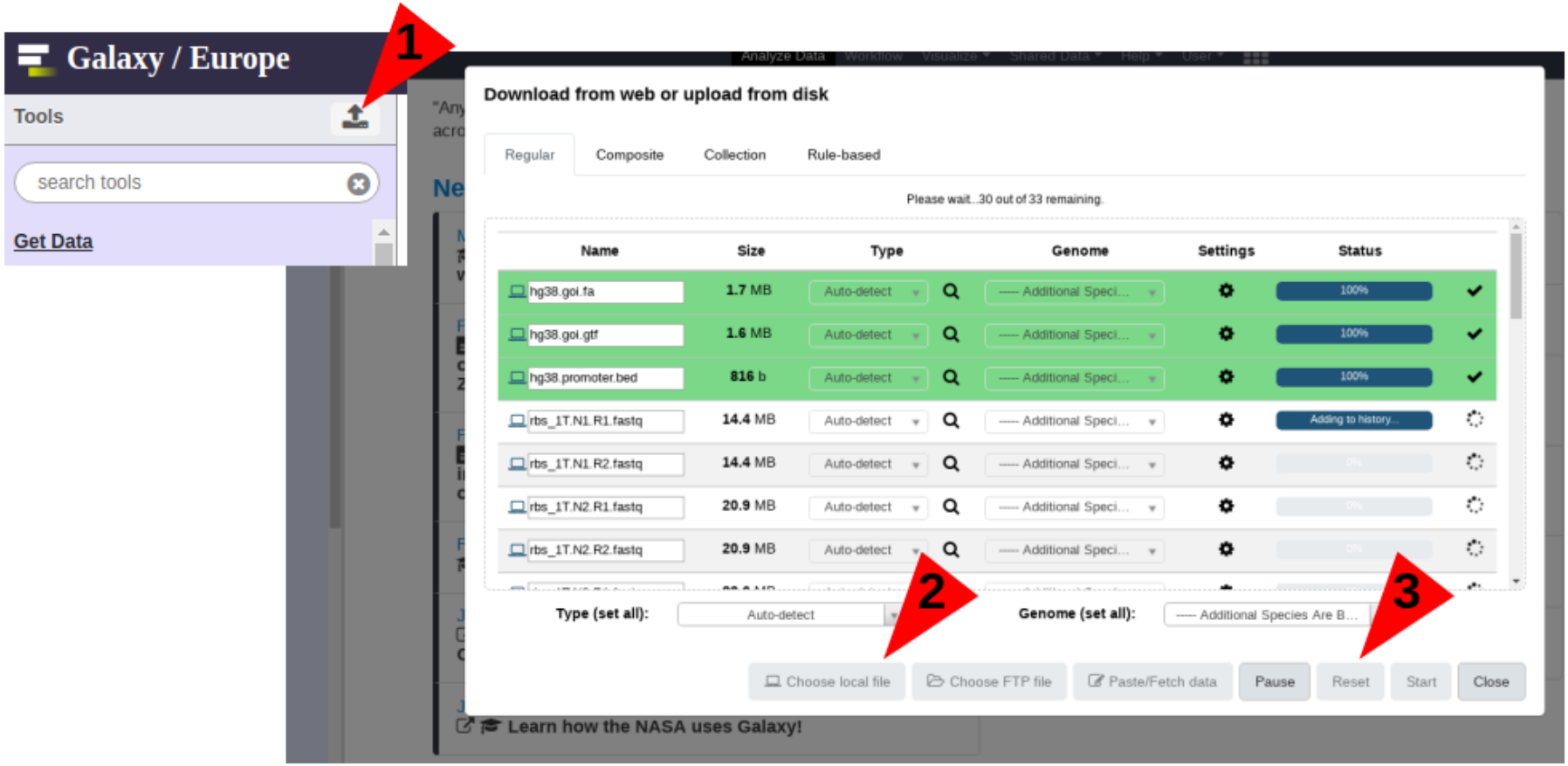
Trimming & Clipping



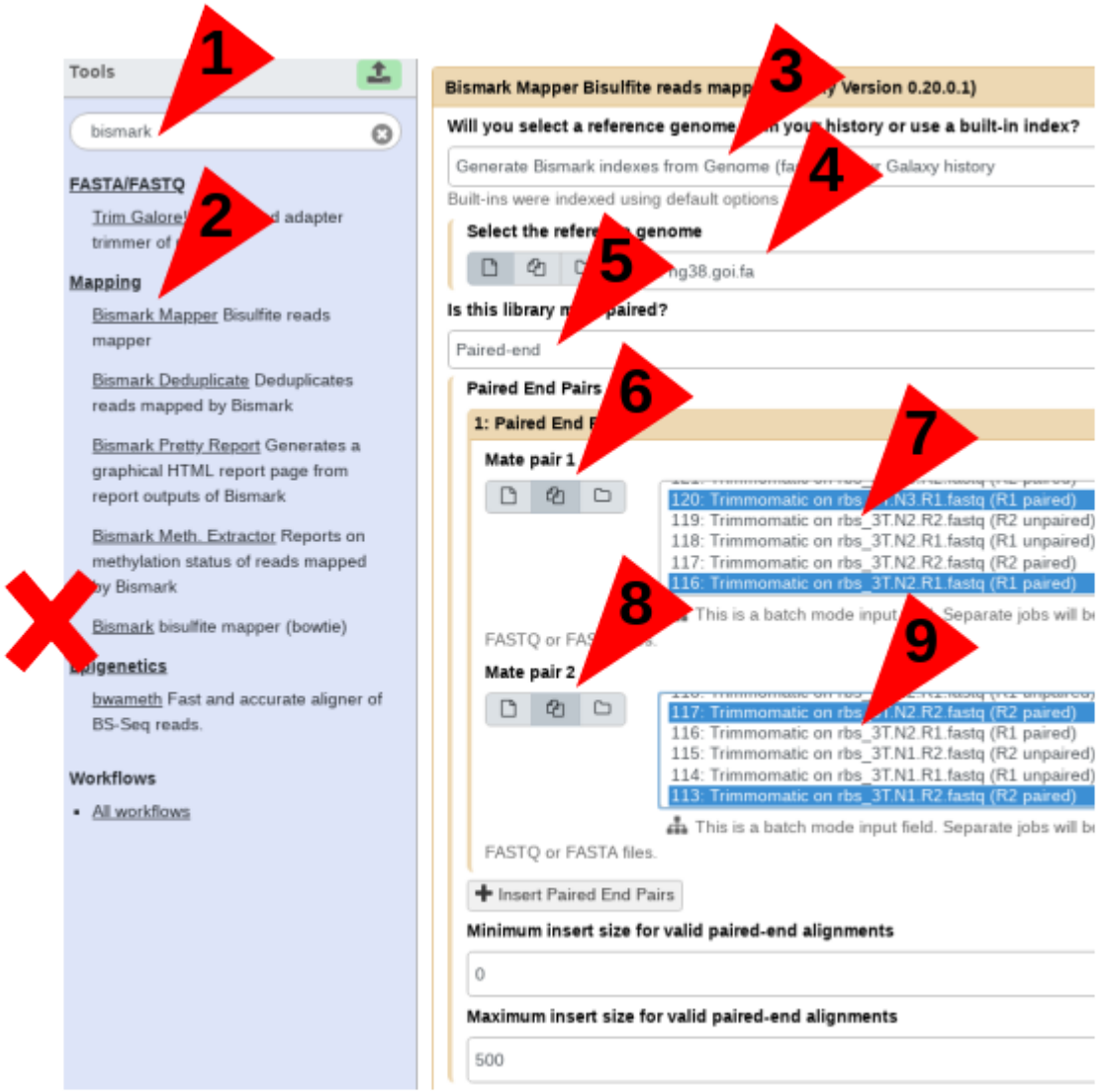
Trimming & Clipping



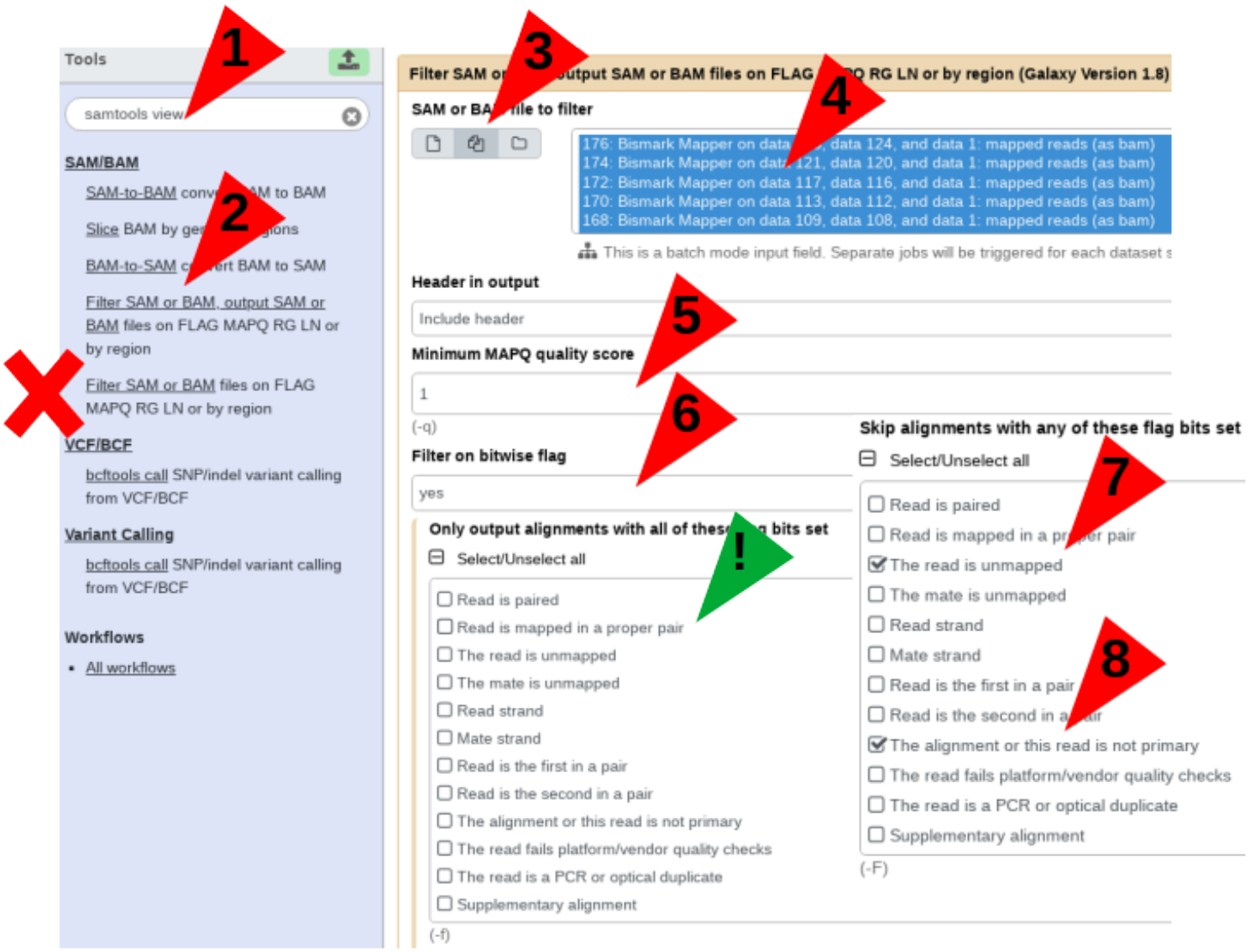
Upload GOI Data



Mapping


1
2
3
4
5
6
7
8
9
Tools
bismark
FASTA/FASTQ
Trim Galore!
adapter trimmer of
Mapping
Bismark Mapper Bisulfite reads mapper
Bismark Deduplicate Deduplicates reads mapped by Bismark
Bismark Pretty Report Generates a graphical HTML report page from report outputs of Bismark
Bismark Meth. Extractor Reports on methylation status of reads mapped by Bismark
Bismark bisulfite mapper (bowtie)
Epigenetics
bwameth Fast and accurate aligner of BS-Seq reads.
Workflows
All workflows
Bismark Mapper Bisulfite reads mapp
Version 0.20.0.1)
Will you select a reference genome
history or use a built-in index?
Generate Bismark indexes from Genome (fa
Galaxy history
Built-ins were indexed using default options
Select the refere
genome
ng38.goi.fa
Is this library
paired?
Paired-end
Paired End Pairs
1: Paired End
Mate pair 1
120: Trimmomatic on rbs
.N3.R1.fastq (R1 paired)
119: Trimmomatic on rbs_3T.N2.R2.fastq (R2 unpaired)
118: Trimmomatic on rbs_3T.N2.R1.fastq (R1 unpaired)
117: Trimmomatic on rbs_3T.N2.R2.fastq (R2 paired)
116: Trimmomatic on rbs_3T.N2.R1.fastq (R1 paired)
This is a batch mode input
Separate jobs will b
FASTQ or FA
Mate pair 2
117: Trimmomatic on rbs
.N2.R2.fastq (R2 paired)
116: Trimmomatic on rbs_3T.N2.R1.fastq (R1 paired)
115: Trimmomatic on rbs_3T.N1.R2.fastq (R2 unpaired)
114: Trimmomatic on rbs_3T.N1.R1.fastq (R1 unpaired)
113: Trimmomatic on rbs_3T.N1.R2.fastq (R2 paired)
This is a batch mode input field. Separate jobs will b
FASTQ or FASTA files.
Insert Paired End Pairs
Minimum insert size for valid paired-end alignments
0
Maximum insert size for valid paired-end alignments
500

Alignment postprocessing

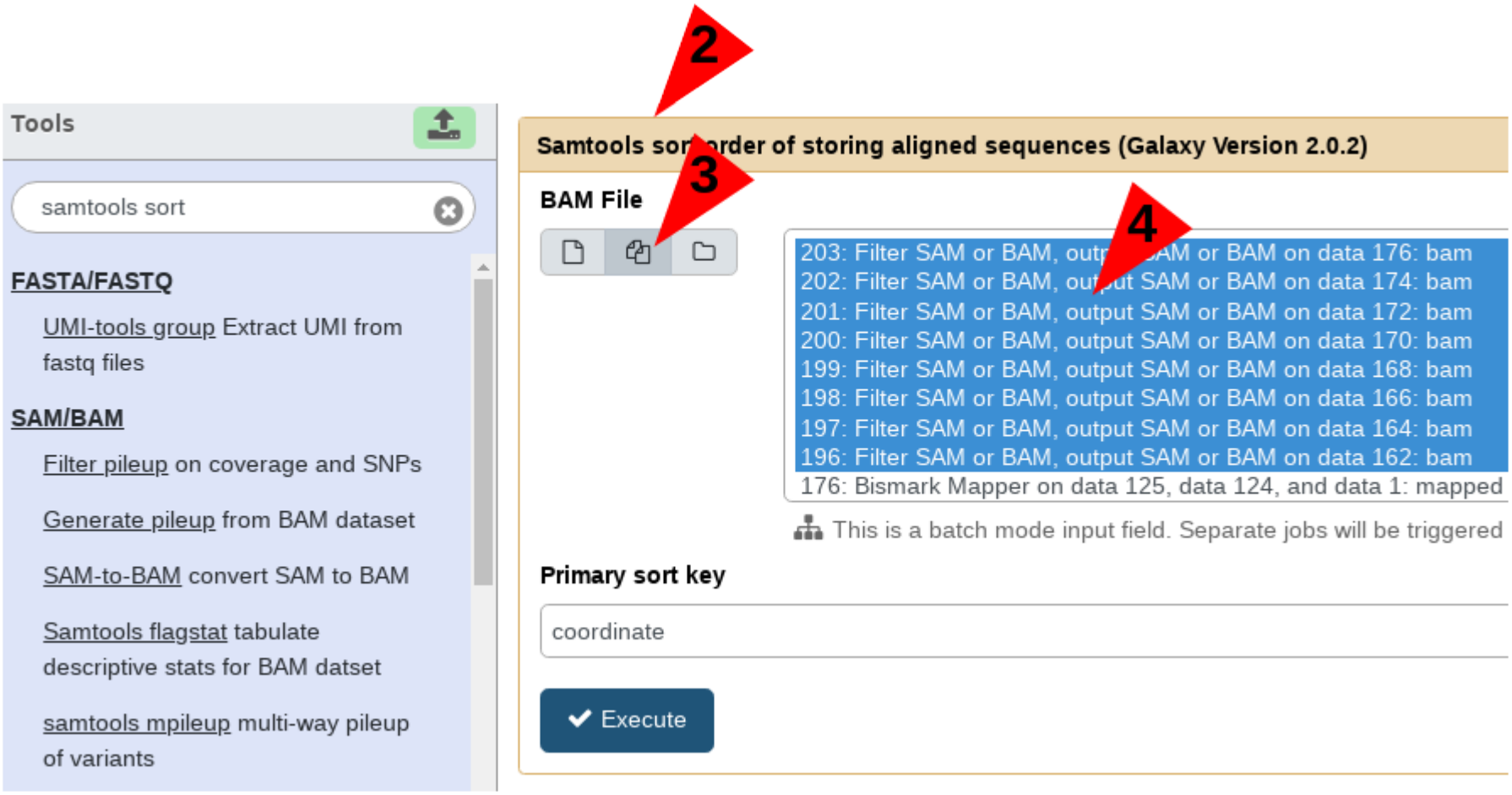


Alignment postprocessing

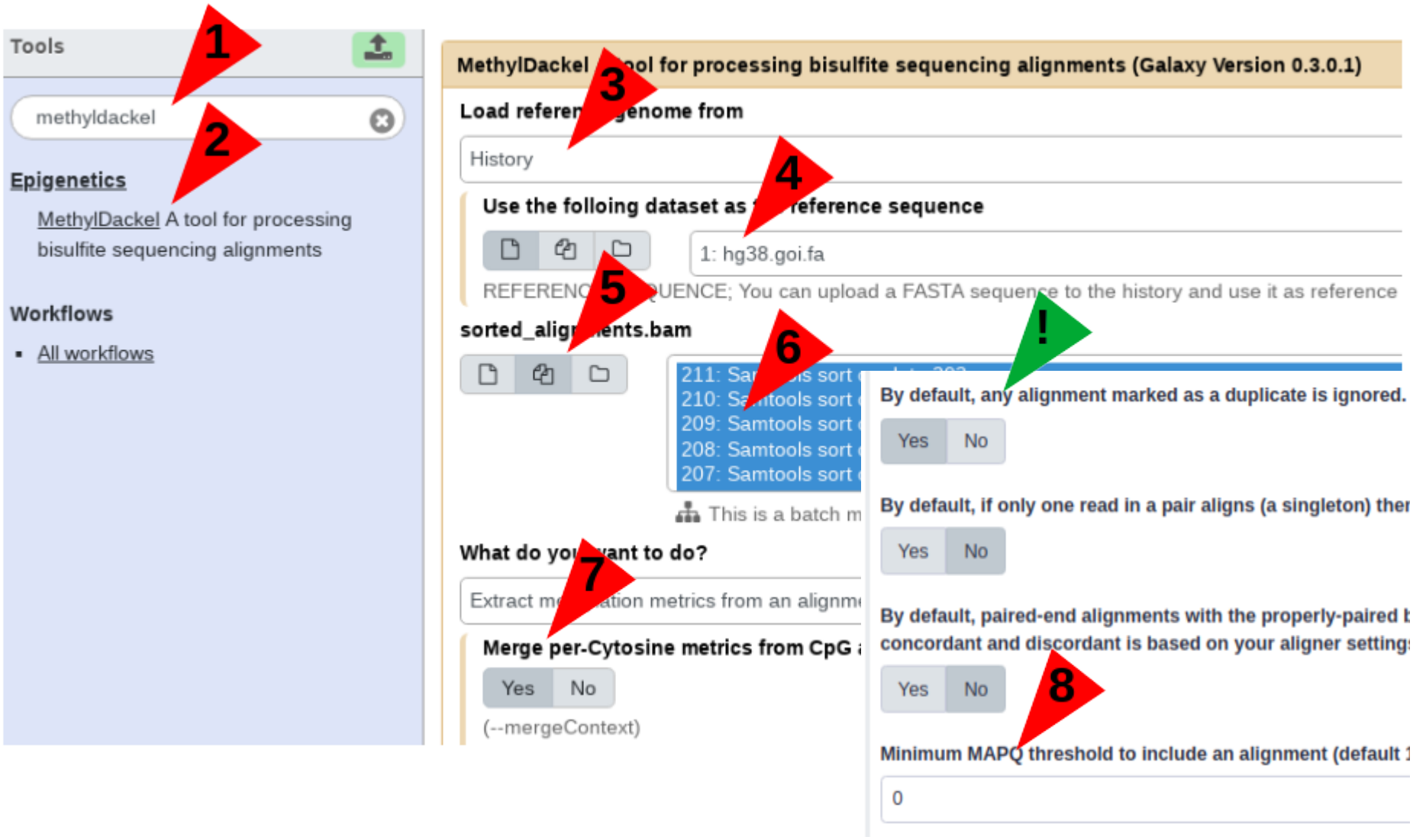
- WGBS: Remove duplicates (second uniquely mapped read with identical sequence and mapping position)
- RRBS: NO!

- Tool box: samtools rmdup

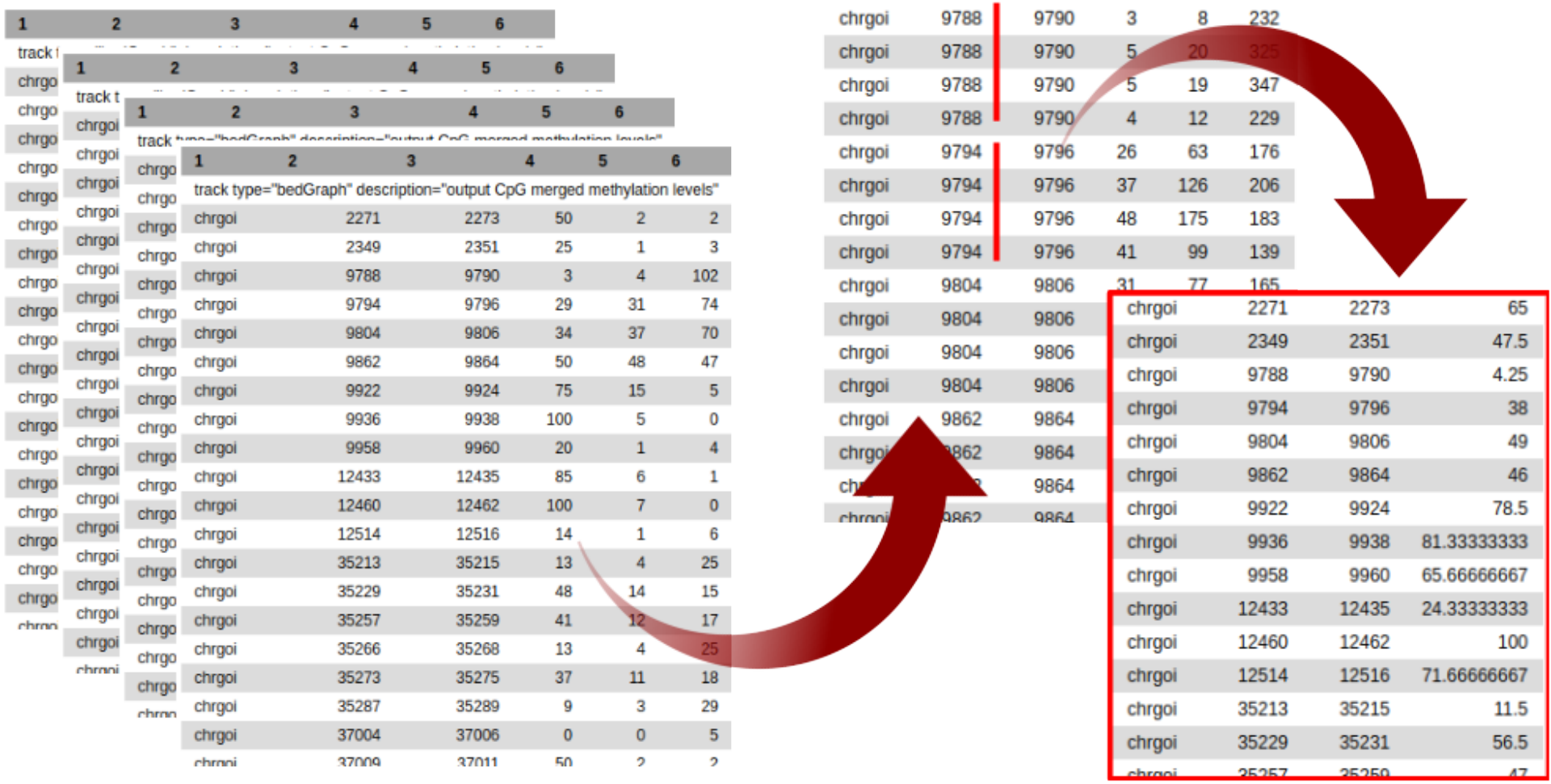
Alignment postprocessing



Infer 5mC methylation level



Infer median methylation level



Intersect with promoter regions

Chrom	Start	End	Name
chrgoi	37553	38553	ENSG00000136859
chrgoi	85184	86184	ENSG00000110934
chrgoi	88182	89183	ENSG00000137948
chrgoi	157238	158239	ENSG00000121797
chrgoi	167073	168074	ENSG00000073067
chrgoi	177515	178516	ENSG00000166801
chrgoi	229319	230319	ENSG00000269190
chrgoi	239797	240797	ENSG00000269190
chrgoi	242795	243795	ENSG00000269190
chrgoi	250261	251261	ENSG00000269190
chrgoi	414709	415709	ENSG00000269190
chrgoi	420561	421561	ENSG00000269190
chrgoi	427763	428763	ENSG00000269190
chrgoi	619900	620900	ENSG00000269190
chrgoi	622898	623898	ENSG00000269190
chrgoi	805693	806693	ENSG00000269190
chrgoi	812937	813938	ENSG00000008853
chrgoi	849720	850721	ENSG00000137872
chrgoi	1443843	1444844	ENSG00000139343
chrgoi	1711347	1712347	ENSG00000166444
chrgoi	1748149	1749149	ENSG00000137868
chrgoi	1751147	1752148	ENSG00000171992

chrgoi	2271	2273	65
chrgoi	2349	2351	47.5
chrgoi	9788	9790	4.25
chrgoi	9794	9796	38
chrgoi	9804	9806	49
chrgoi	9862	9864	46
chrgoi	9922	9924	78.5
chrgoi	9936	9938	81.33333333
chrgoi	9958	9960	65.66666667
chrgoi	12433	12435	24.33333333
chrgoi	12460	12462	100
chrgoi	12514	12516	71.66666667
chrgoi	35213	35215	11.5
chrgoi	35229	35231	56.5

chrgoi	85184	86184	ENSG00000110934	chrgoi	85189	85193	41.66666667
chrgoi	85184	86184	ENSG00000110934	chrgoi	85201	85203	100
chrgoi	85184	86184	ENSG00000110934	chrgoi	85229	85231	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85234	85236	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85243	85245	0
chrgoi	88182	89183	ENSG00000137948	.	-1	-1	.
chrgoi	157238	158239	ENSG00000121797	.	-1	-1	.
chrgoi	167073	168074	ENSG00000073067	chrgoi	167297	167299	30.33333333
chrgoi	167073	168074	ENSG00000073067	chrgoi	167301	167303	66
chrgoi	167073	168074	ENSG00000073067	chrgoi	167364	167366	87.5
chrgoi	167073	168074	ENSG00000073067	chrgoi	167373	167375	100
chrgoi	167073	168074	ENSG00000073067	chrgoi	167387	167389	73
chrgoi	177515	178516	ENSG00000166801	.	-1	-1	.
chrgoi	229319	230319	ENSG00000269190	chrgoi	229327	229331	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229342	229344	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229650	229652	0

Infer median promoter me level

chrgoi	85184	86184	ENSG00000110934	chrgoi	85189	85193	41.66666667
chrgoi	85184	86184	ENSG00000110934	chrgoi	85201	85203	100
chrgoi	85184	86184	ENSG00000110934	chrgoi	85229	85231	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85234	85236	50
chrgoi	85184	86184	ENSG00000110934	chrgoi	85243	85245	0
chrgoi	88182	89183	ENSG00000137948	.	-1	-1	.
chrgoi	157238	158239	ENSG00000121797	.	-1	-1	.
chrgoi	167073	168074	ENSG00000073067	chrgoi	167297	167299	30.33333333
chrgoi	167073	168074	ENSG00000073067	chrgoi	167301	167303	66
chrgoi	167073	168074	ENSG00000073067	chrgoi	167364	167366	87.5
chrgoi	167073	168074	ENSG00000073067	chrgoi	167373	167375	100
chrgoi	167073	168074	ENSG00000073067	chrgoi	167387	167389	73
chrgoi	177515	178516	ENSG00000166801	.	-1	-1	.
chrgoi	229319	230319	ENSG00000269190	chrgoi	229327	229331	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229342	229344	0
chrgoi	229319	230319	ENSG00000269190	chrgoi	229650	229652	0

chrgoi	37553	38553	ENSG00000136859	.
chrgoi	85184	86184	ENSG00000110934	40.27777778
chrgoi	88182	89183	ENSG00000137948	.
chrgoi	157238	158239	ENSG00000121797	.
chrgoi	167073	168074	ENSG00000073067	71.36666667
chrgoi	177515	178516	ENSG00000166801	.
chrgoi	229319	230319	ENSG00000269190	16.4
chrgoi	239797	240797	ENSG00000127951	.
chrgoi	242795	243796	ENSG00000189280	45
chrgoi	250261	251262	ENSG00000196208	46
chrgoi	414709	415710	ENSG00000170276	.
chrgoi	420561	421562	ENSG00000129009	.
chrgoi	427763	428764	ENSG00000172005	.
chr	619900	620900	ENSG00000143850	2.135416667
	898	623899	ENSG00000132334	8.333333333
chrg	805693	806694	ENSG00000188060	33
ch	812937	813938	ENSG00000008853	.
	849720	850721	ENSG00000137872	0
	goi 1443843	1444844	ENSG00000139343	25.375
chrgoi	1711347	1712347	ENSG00000166444	16.29166667
chrgoi	1748149	1749149	ENSG00000137868	80
chrgoi	1751147	1752148	ENSG00000171992	36.66666667

Calculate log2 fold-change

chrgoi	37553	38553	ENSG00000136859	.	chrgoi	37553	38553	ENSG00000136859	.
chrgoi	85184	86184	ENSG00000110934	34.57407407	chrgoi	85184	86184	ENSG00000110934	40.27777778
chrgoi	88182	89183	ENSG00000137948	.	chrgoi	88182	89183	ENSG00000137948	.
chrgoi	157238	158239	ENSG00000121797	.	chrgoi	157238	158239	ENSG00000121797	.
chrgoi	167073	168074	ENSG00000073067	23.71428571	chrgoi	167073	168074	ENSG00000073067	71.36666667
chrgoi	177515	178516	ENSG00000166801	.	chrgoi	177515	178516	ENSG00000166801	.
chrgoi	229319	230319	ENSG00000269190	25.16666667	chrgoi	229319	230319	ENSG00000269190	16.4
chrgoi	239797	240797	ENSG00000127951	.	chrgoi	239797	240797	ENSG00000127951	.
chrgoi	242795	243796	ENSG00000189280	80	chrgoi	242795	243796	ENSG00000189280	45
chrgoi	250261	251262	ENSG00000196208	29.53703704	chrgoi	250261	251262	ENSG00000196208	46
chrgoi	414709	415710	ENSG00000170276	15	chrgoi	414709	415710	ENSG00000170276	.
chrgoi	420561	421562	ENSG00000129009	.	chrgoi	420561	421562	ENSG00000129009	.
chrgoi	427763	428764	ENSG00000172005	17.83333333	chrgoi	427763	428764	ENSG00000172005	.
chrgoi	619900	620900	ENSG00000143850	5.416666667	chrgoi	619900	620900	ENSG00000143850	2.135416667
chrgoi	622898	623899	ENSG00000132334	3.571428571	chrgoi	622898	623899	ENSG00000132334	8.333333333
chrgoi	805693	806694	ENSG00000188060	23.33333333	chrgoi	805693	806694	ENSG00000188060	33
chrgoi	812937	813938	ENSG00000008853	19	chrgoi	812937	813938	ENSG00000008853	.
chrgoi	849720	850721	ENSG00000137872	3.111111111	chrgoi	849720	850721	ENSG00000137872	0
chrgoi	1443843	1444844	ENSG00000139343	25.83333333	chrgoi	1443843	1444844	ENSG00000139343	25.375
chrgoi	1711347	1712347	ENSG00000166444	58.9375	chrgoi	1711347	1712347	ENSG00000166444	16.29166667
chrgoi	1748149	1749149	ENSG00000137868	.	chrgoi	1748149	1749149	ENSG00000137868	80
chrgoi	1751147	1752148	ENSG00000171992	.	chrgoi	1751147	1752148	ENSG00000171992	36.66666667

Tools

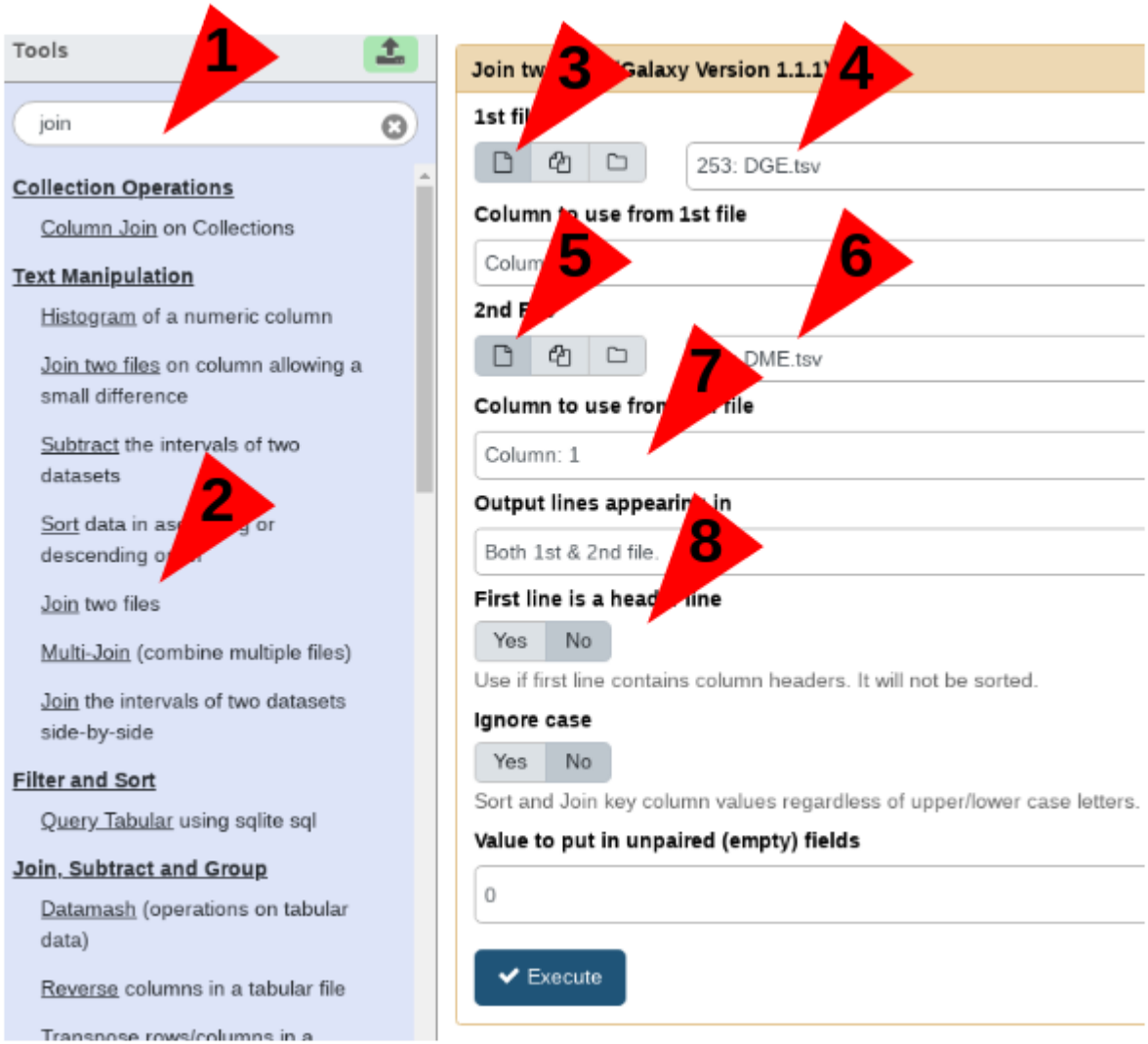
- tail
- cat
- sort
- bedtools merge
- bedtools intersect
- join
- awk

```
{print $4"\t"($5 < $9 ? (log($9+1)-log($5+1))/log(2)
: -(log($5+1)-log($9+1))/log(2))}
```

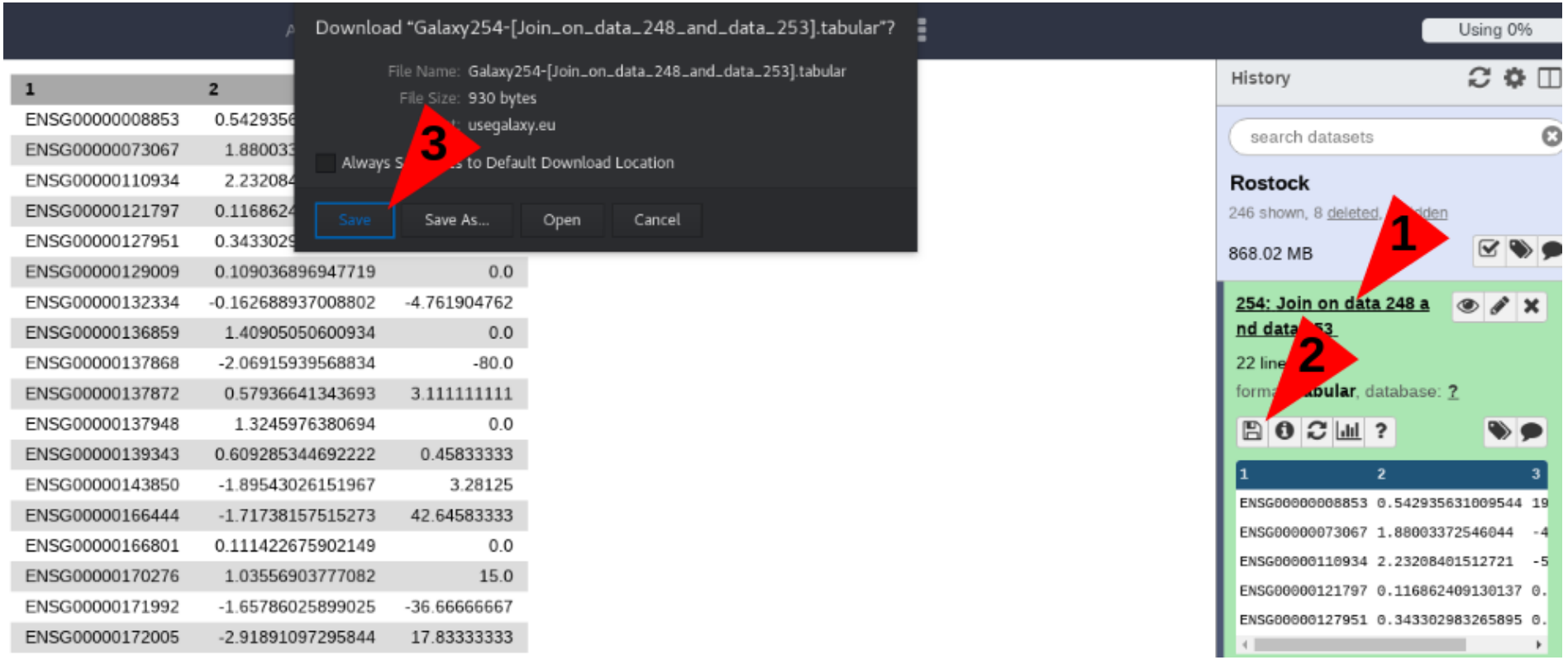
Scatterplot

- Join DGE table and DME table
- Download files
- Switch to R

Scatterplot



Scatterplot

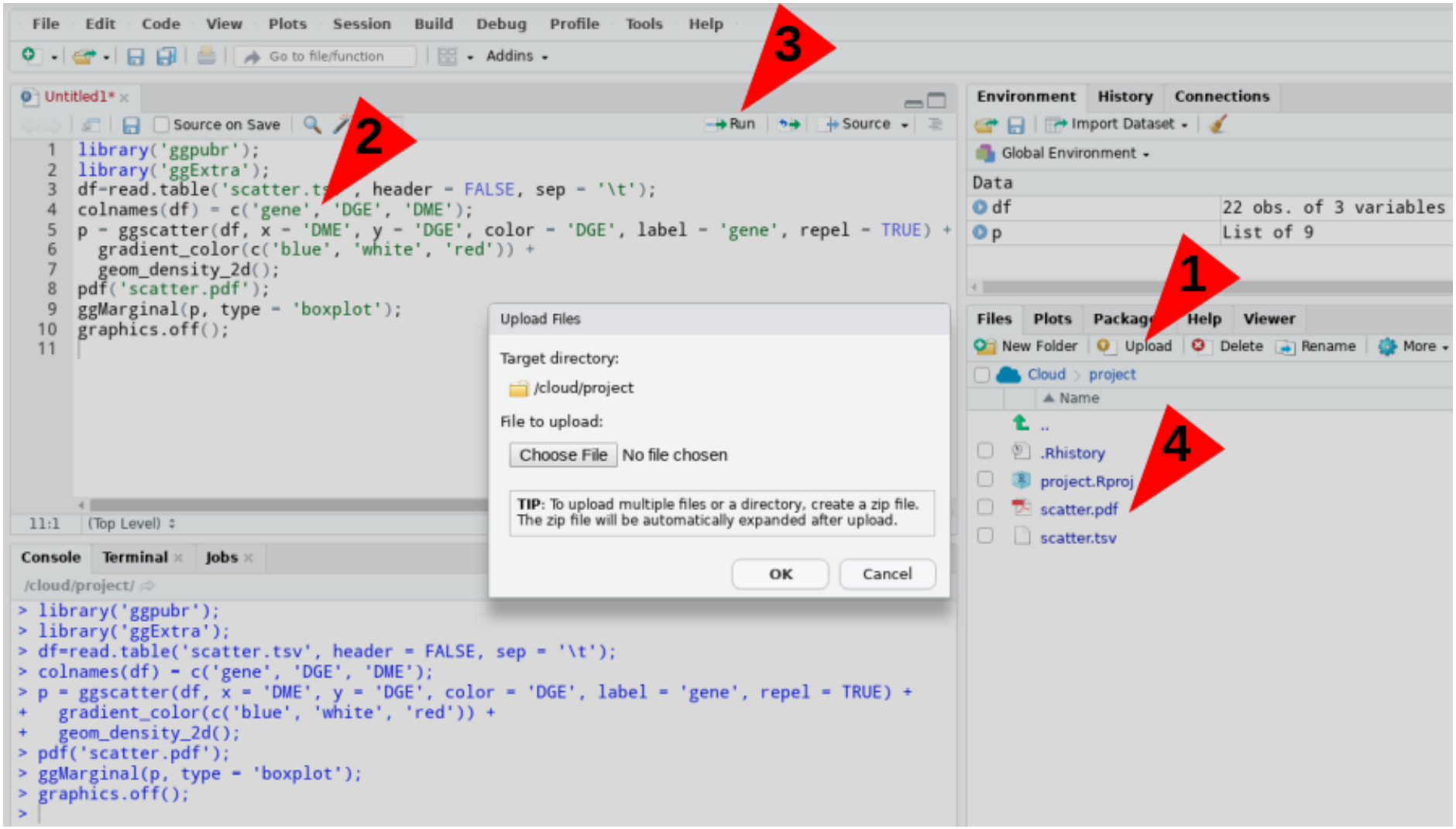


1	2	
ENSG00000008853	0.5429356	
ENSG00000073067	1.880033	
ENSG00000110934	2.232084	
ENSG00000121797	0.1168624	
ENSG00000127951	0.3433029	
ENSG00000129009	0.109036896947719	0.0
ENSG00000132334	-0.162688937008802	-4.761904762
ENSG00000136859	1.40905050600934	0.0
ENSG00000137868	-2.06915939568834	-80.0
ENSG00000137872	0.57936641343693	3.111111111
ENSG00000137948	1.3245976380694	0.0
ENSG00000139343	0.609285344692222	0.45833333
ENSG00000143850	-1.89543026151967	3.28125
ENSG00000166444	-1.71738157515273	42.64583333
ENSG00000166801	0.111422675902149	0.0
ENSG00000170276	1.03556903777082	15.0
ENSG00000171992	-1.65786025899025	-36.66666667
ENSG00000172005	-2.91891097295844	17.83333333

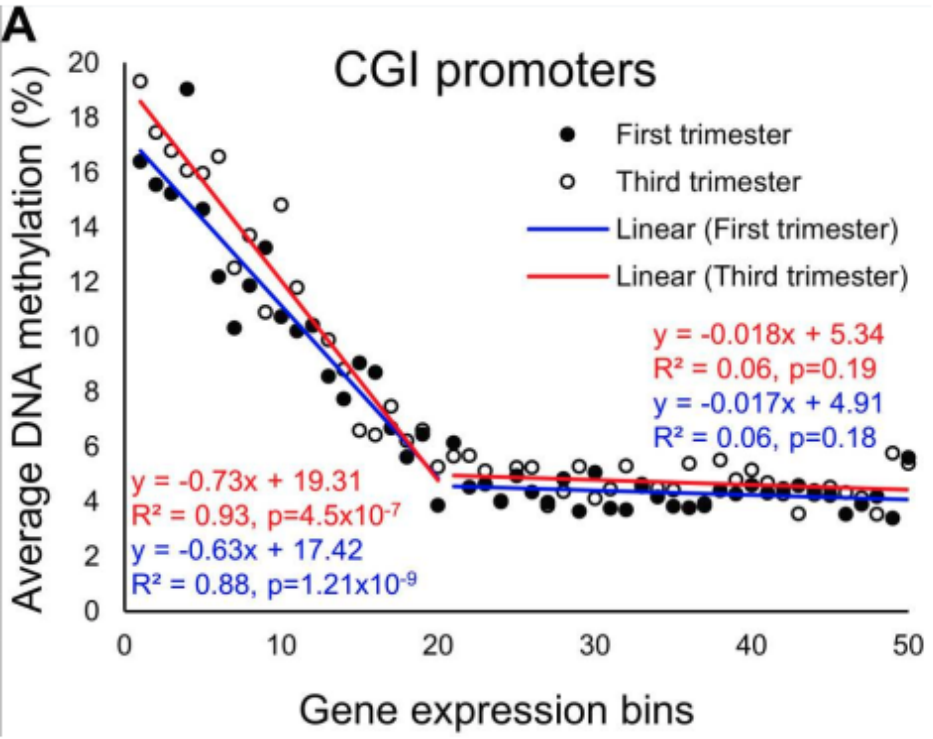
Scatterplot

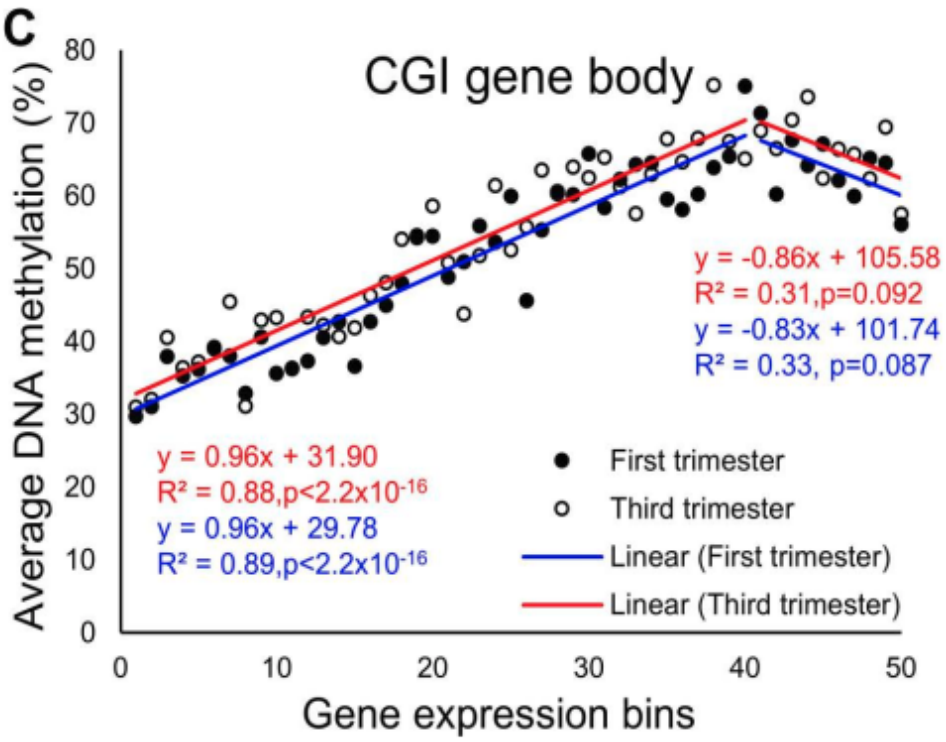
```
library('ggpubr');
library('ggExtra');
df=read.table('scatter.tsv', header = FALSE, sep = '\t');
colnames(df) = c('gene', 'DGE', 'DME');
p = ggscatter(df, x = 'DME', y= 'DGE', color = 'DGE', label = 'gene',
repel = TRUE) +
  gradient_color(c('blue', 'white', 'red')) +
  geom_density_2d();
pdf('scatter.pdf');
ggMarginal(p, type= 'boxplot');
graphics.off();
```

Scatterplot

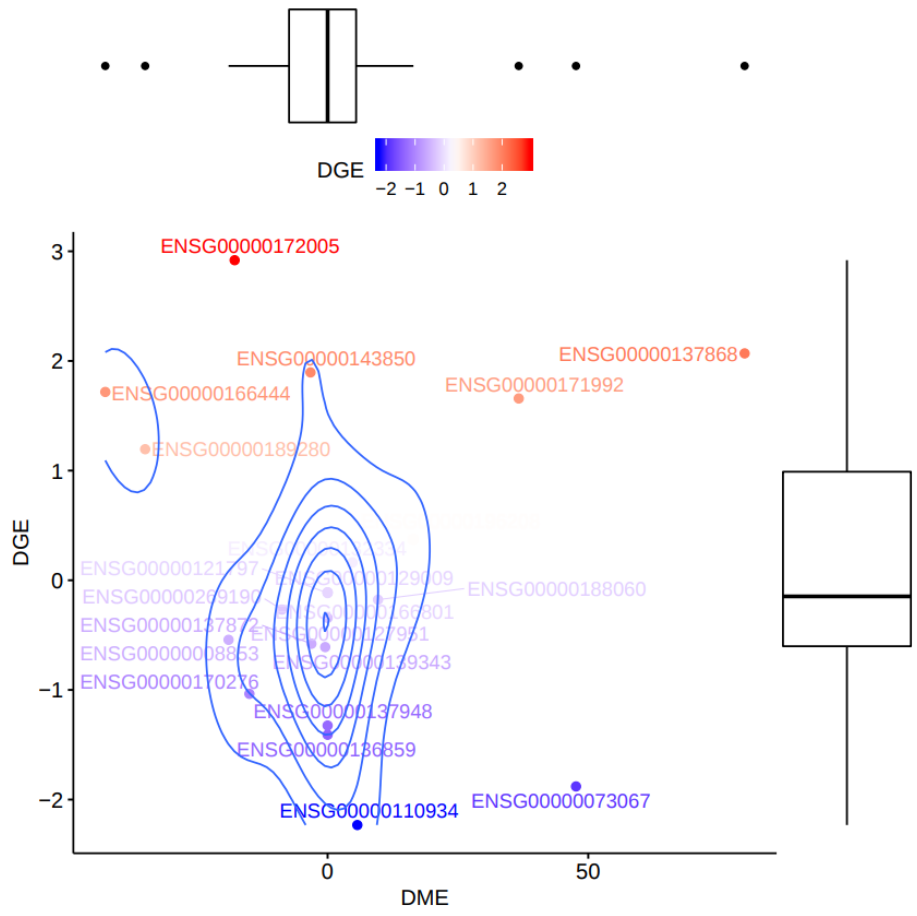


Recap


A
CGI promoters
Average DNA methylation (%)
Gene expression bins
First trimester
Third trimester
Linear (First trimester)
Linear (Third trimester)
y = -0.018x + 5.34
R² = 0.06, p=0.19
y = -0.017x + 4.91
R² = 0.06, p=0.18
y = -0.73x + 19.31
R² = 0.93, p=4.5x10-7
y = -0.63x + 17.42
R² = 0.88, p=1.21x10-9


C
CGI gene body
Average DNA methylation (%)
Gene expression bins
y = -0.86x + 105.58
R² = 0.31,p=0.092
y = -0.83x + 101.74
R² = 0.33, p=0.087
y = 0.96x + 31.90
R² = 0.88,p<2.2x10-16
y = 0.96x + 29.78
R² = 0.89,p<2.2x10-16
First trimester
Third trimester
Linear (First trimester)
Linear (Third trimester)

Scatterplot



Possible impact of epigenetic modification on expression level of top DEGs:

UP (Q1)
- STRA6: Vitamin A metabolism
- SYNPO: Cell motility

UP (Q2)
- MAL: T cell differentiation protein
- ST5: Tumor suppressor
- GJB5: Intercellular signalling

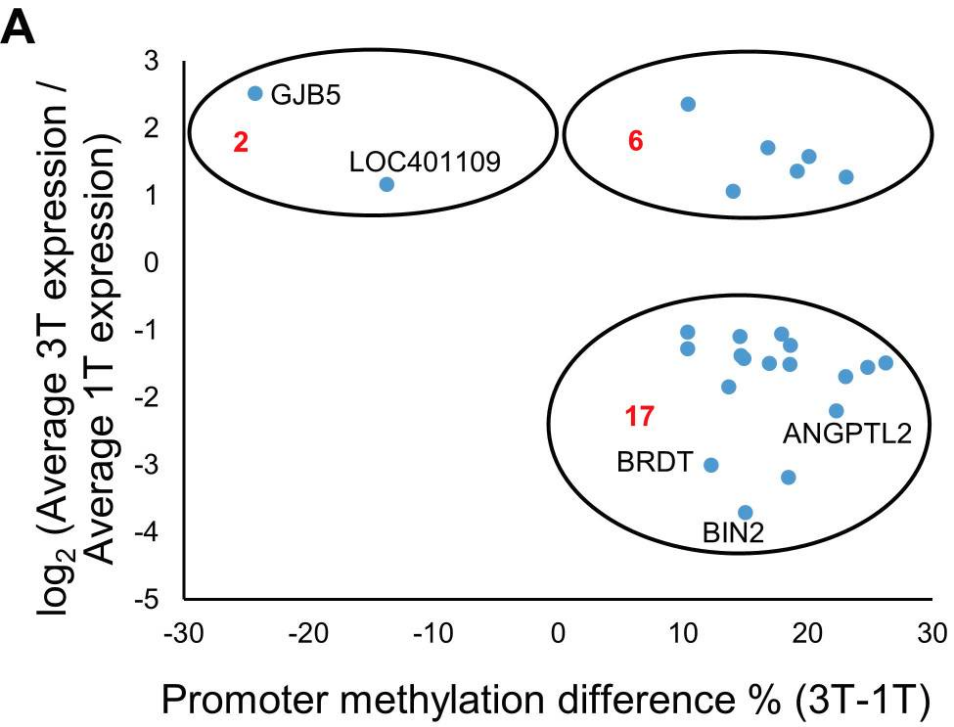
DOWN (Q3)
- HSPB2: Heat shock protein
- RHOBTB2: Tumor suppressor
- SEMA6D: Secretome protein

DOWN (Q4)
- CYP2W1: hormone synthesis, vit. D metabolism
- BIN2: Onkogen
- RAB42: Onkogen

→ Enriched immune system processes due to fetus protective function of placenta

→ Vitamins A and D known modulate pro-inflammatory immune response

Comparison



UP (Q1)
- STRA6: Vitamin A metabolism
- SYNPO: Cell motility

UP (Q2)
- MAL: T cell differentiation protein
- ST5: Tumor suppressor
- GJB5: Intercellular signalling

DOWN (Q3)
- HSPB2: Heat shock protein
- RHOBTB2: Tumor suppressor
- SEMA6D: Secretome protein

DOWN (Q4)
- CYP2W1: hormone synthesis, vit. D metabolism
- BIN2: Onkogen
- RAB42: Onkogen

→ Similar pattern
→ Similar extrema