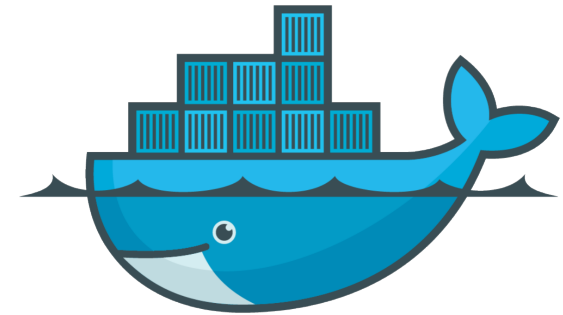


RRBS Analysis

Konstantin Riege
FLI Jena

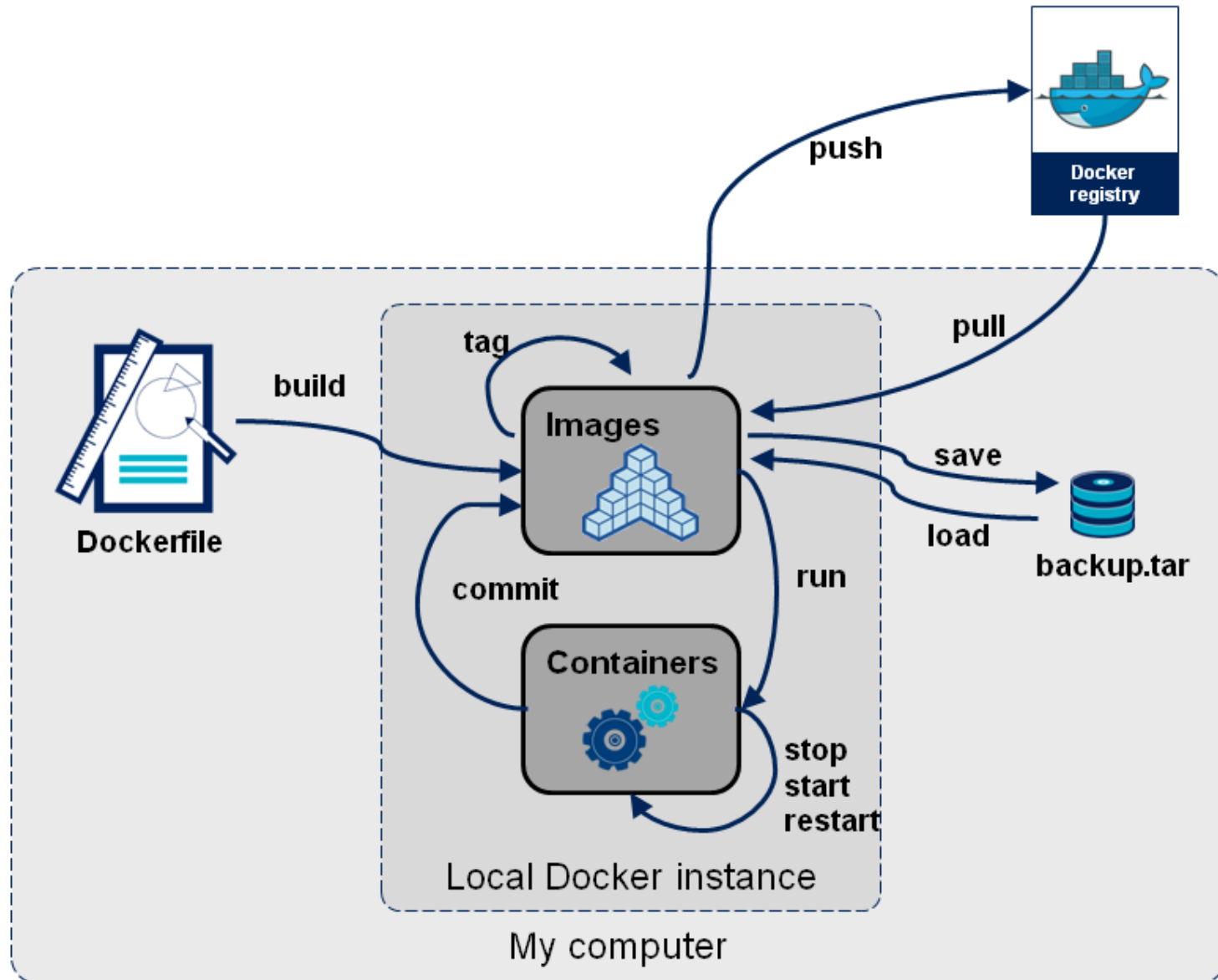
Rstudio Installation

- 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



docker

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.gol.fa	1.7 MB	Auto-detect	Additional Speci...		100%
hg38.gol.gtf	1.6 MB	Auto-detect	Additional Speci...		100%
hg38.promoter.bed	816 b	Auto-detect	Additional Speci...		100%
rbs_1T.N1.R1.fastq	14.4 MB	Auto-detect	Additional Speci...		Adding to history...
rbs_1T.N1.R2.fastq	14.4 MB	Auto-detect	Additional Speci...		0%
rbs_1T.N2.R1.fastq	20.9 MB	Auto-detect	Additional Speci...		0%
rbs_1T.N2.R2.fastq	20.9 MB	Auto-detect	Additional Speci...		0%

Type (set all): Auto-detect Genome (set all): Additional Species Are B...

2 **3**

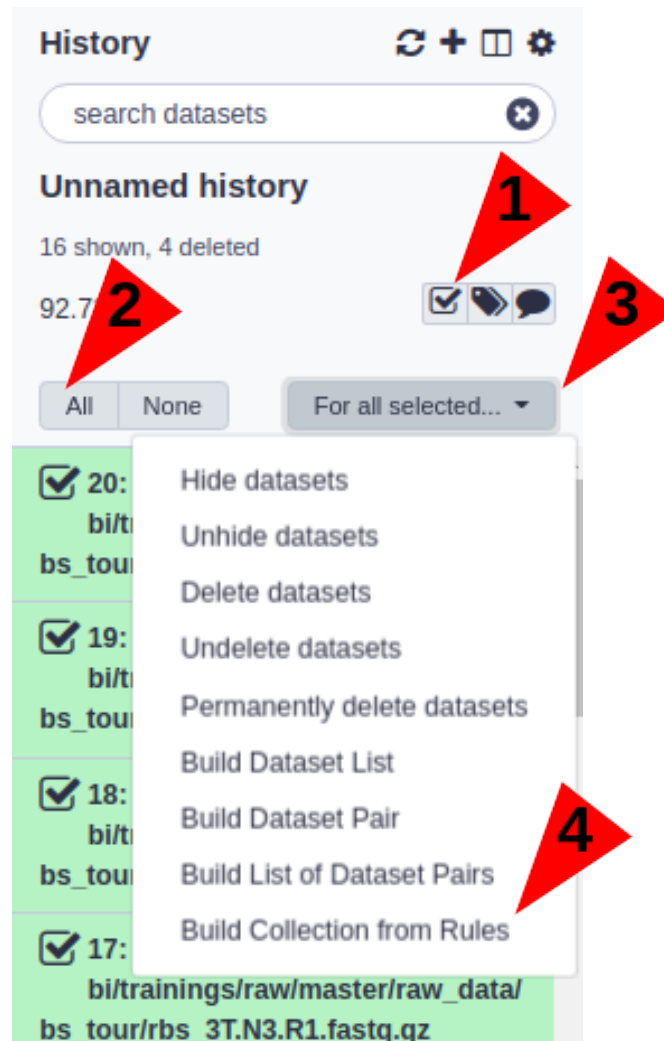
Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

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

Build Rules for Creating Collection(s)

Use this form to describe rules for building collection(s) from the spec name" to group datasets into multiple collections.

Rules ✎

- Add column for hid. ✎ ✕
- Add column for name. ✎ ✕
- Add new column using .+rbs_(\dT).Nld.(R\d) applied to column B ✎ ✕
- Set column B as List Identifier(s) ✎ ✕

✕

Sort

Remove Column(s)

Split Column(s)

Swap Column(s)

Add / Modify Column Definitions

+ Rules ▲ + Filter ▲

+ Column ▲

Hide original elements: ☐ Add nametag for name: ☐

A	
20	https://g
19	https://g
18	https://g
17	...
16	
15	
14	
13	
12	
11	
10	
9	
8	
7	
6	
5	https://g
4	

List Identifier(s) ✕

- B ✕
- ... Assign Another Column

+ Add Definition ▼

Paired-end Indicator

Collection Name

General Purpose Tag(s)

Group Tag(s)

History ↺ + 📄 ⚙️

search datasets ✕

Unnamed history

20 shown, 8 deleted, 16 hidden

92.72 MB

☑️ 🏷️ 💬

60: 1T_R1	✕
a list with 4 items	
55: 1T_R2	✕
a list with 4 items	
50: 3T_R1	✕
a list with 4 items	
45: 3T_R2	✕
a list with 4 items	

Quality Control

Tools 1

fastqc

FASTA/FASTQ

- [Combine FASTA and QUAL into FASTQ](#)
- [Manipulate FASTQ reads on various attributes](#)
- [FastQC:Read QC reports using FastQC](#)
- [fastp - fast all-in-one preprocessing for FASTQ files](#)

FASTQ Quality Control 2

- [FastQC Read Quality reports](#)

Mapping

- [Map with PerM for SOLiD and Illumina](#)

FastQC Read Quality reports (Galaxy Version 0.72)

Short read data from your current history 3

33: rna_3T.N4.fastq
32: rna_3T.N3.R2.fastq
31: rna_3T.N3.R1.fastq
30: rna_3T.N2.R2.fastq
29: rna_3T.N2.R1.fastq 4

This is a batch mode input

Contaminant list

Nothing selected

tab delimited file with 2 columns: name and sequence.

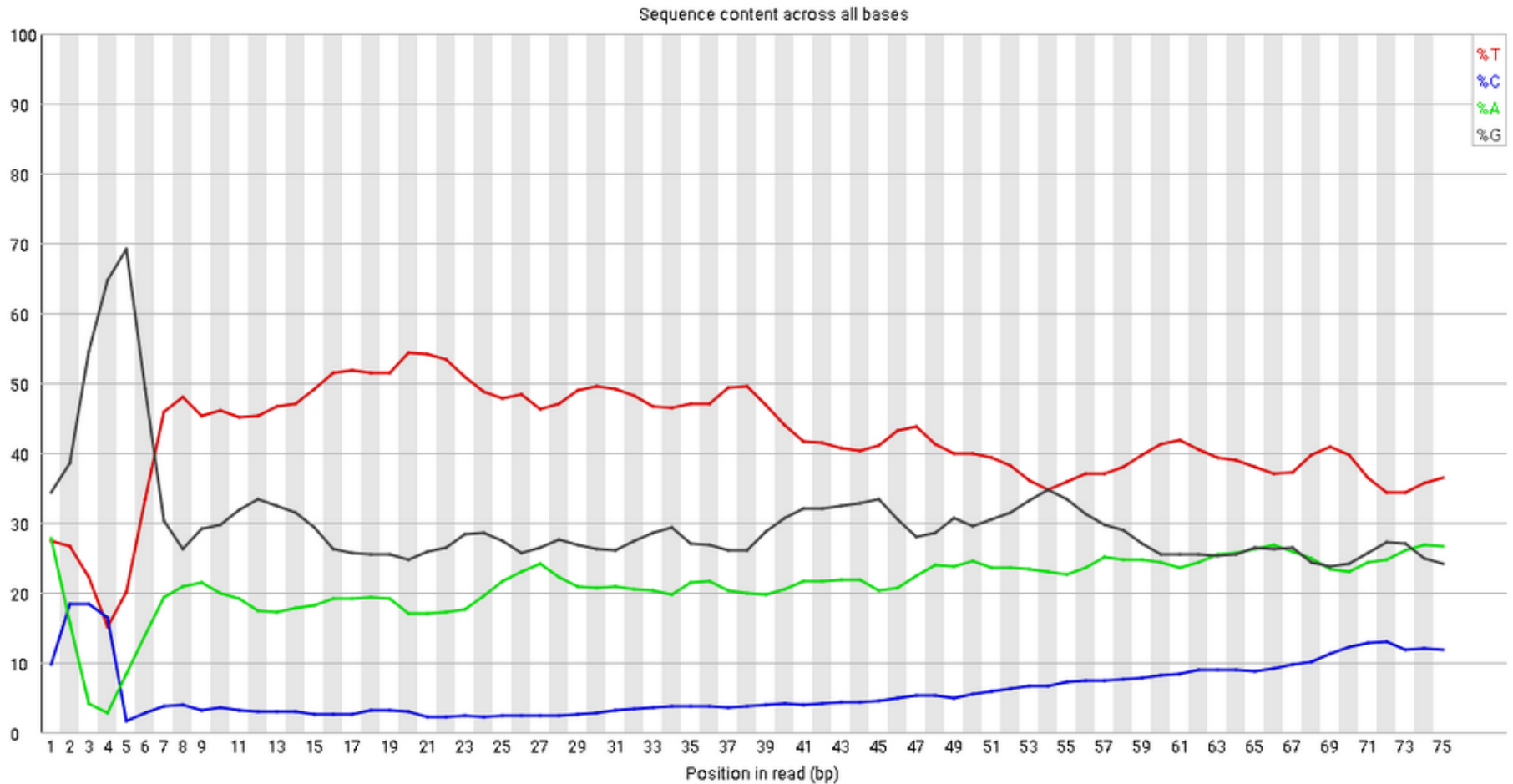
Submodule and Limit specifying file

Nothing selected

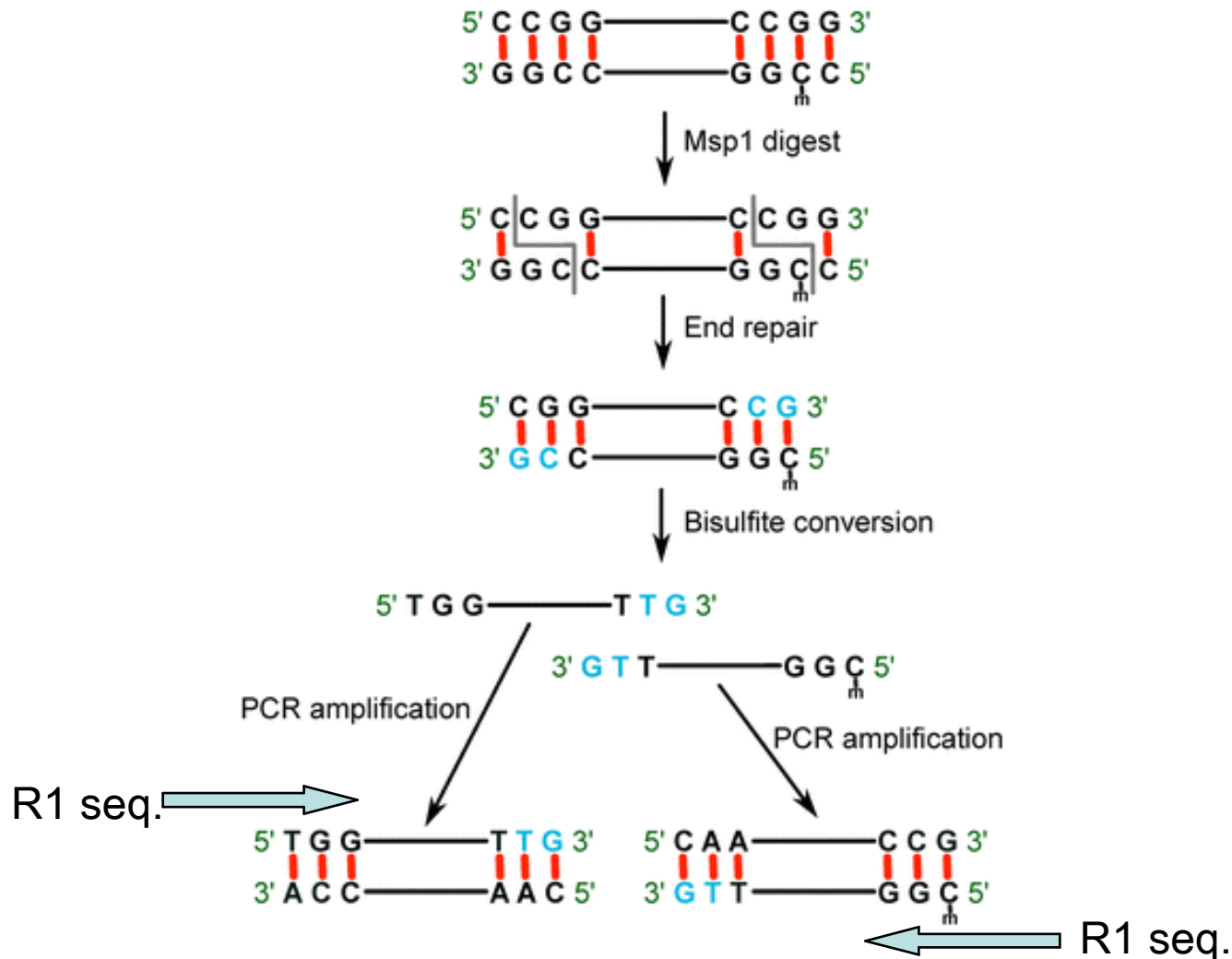
a file that specifies which submodules are to be executed

☒ Execute

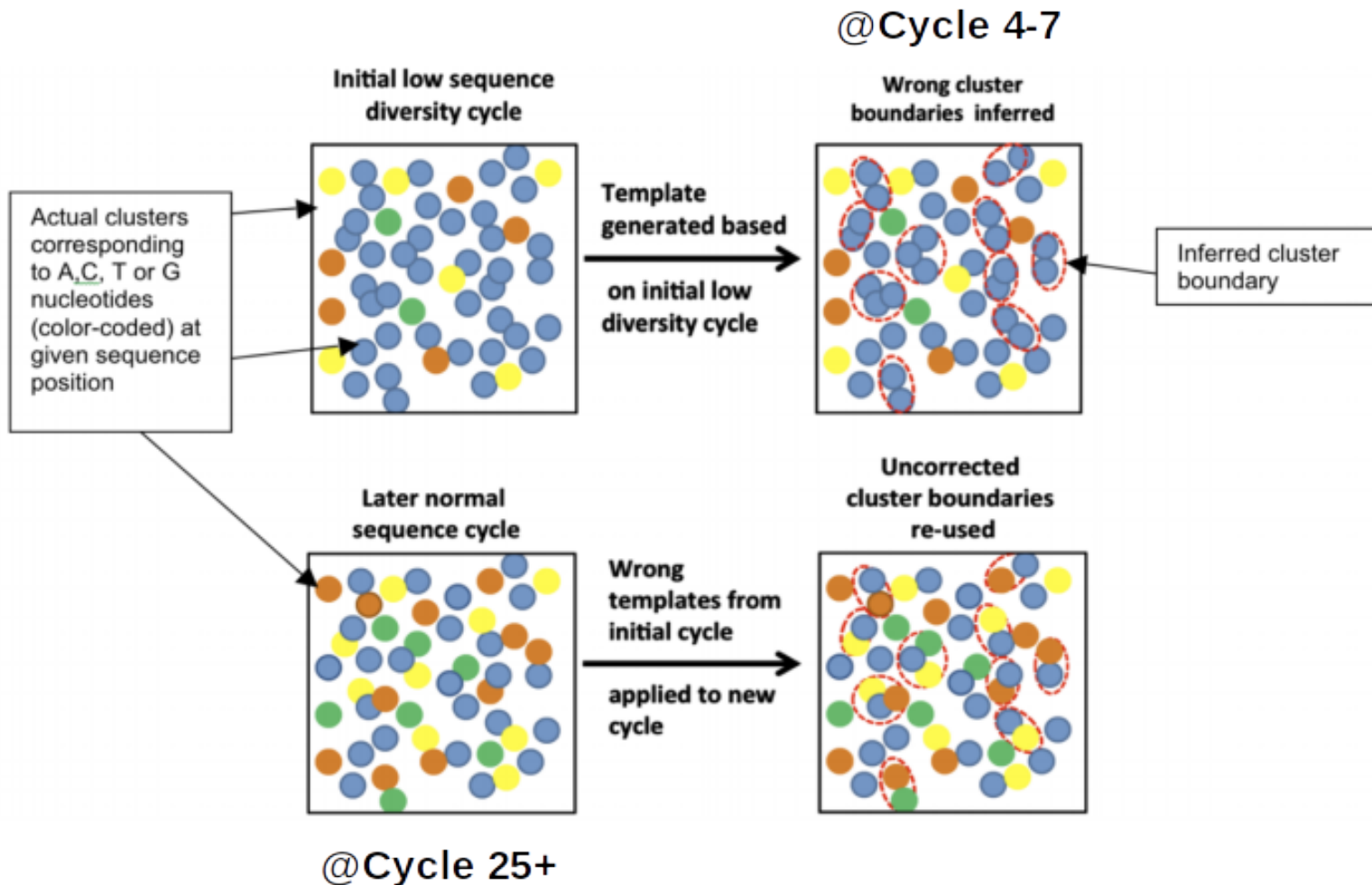
Quality Control



CpG enrichment by MspI

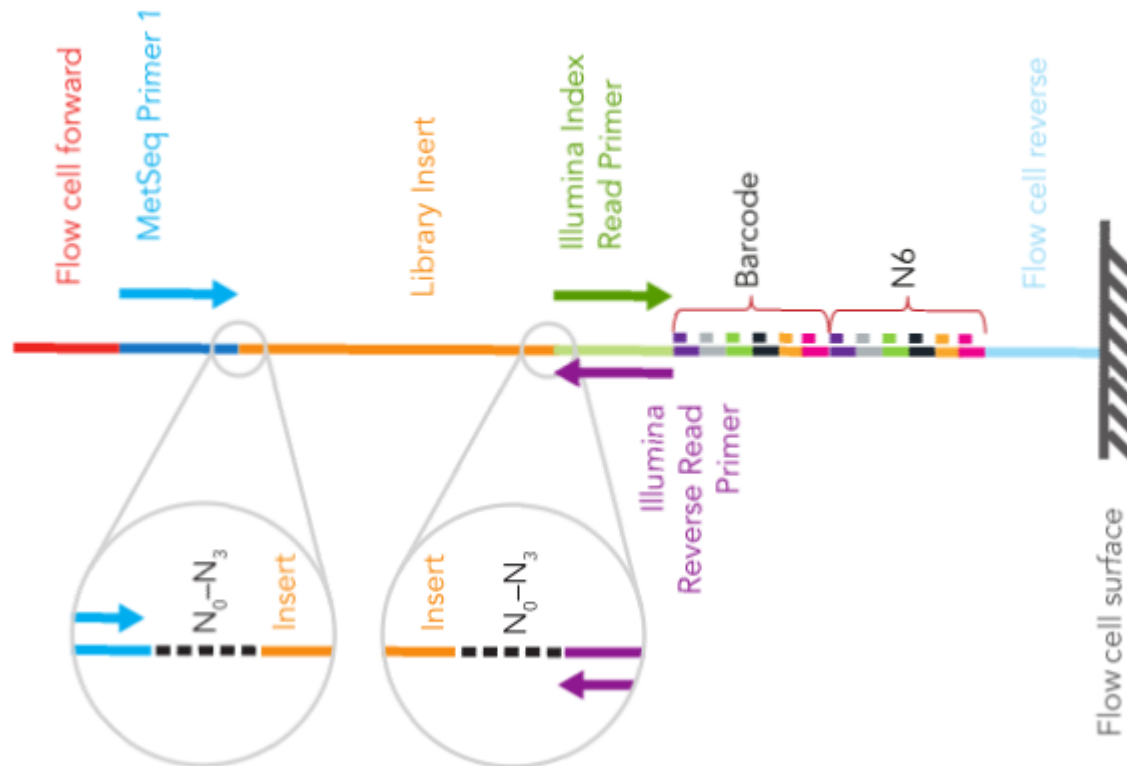


Low complexity bias

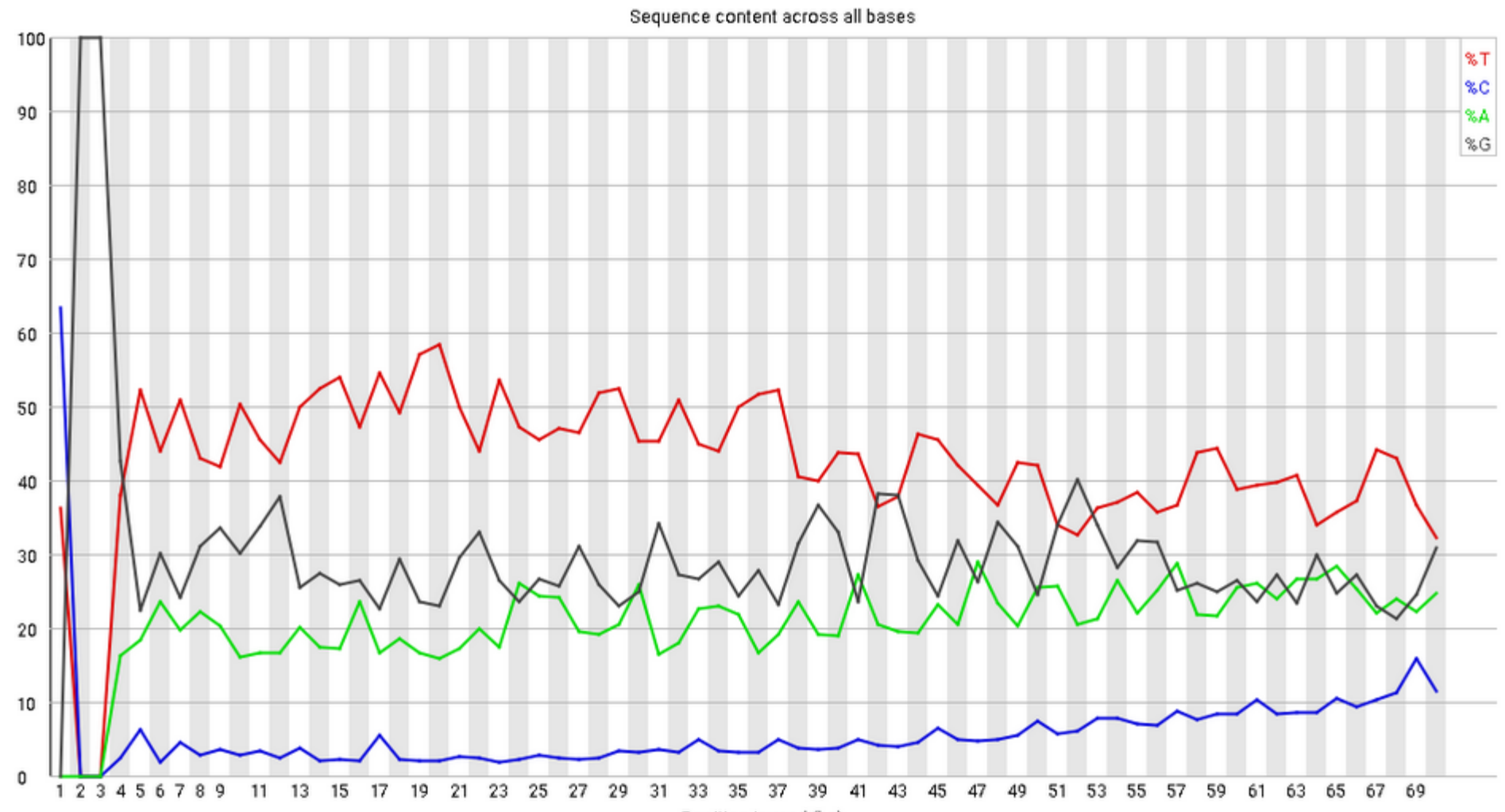


Diversity Adapter

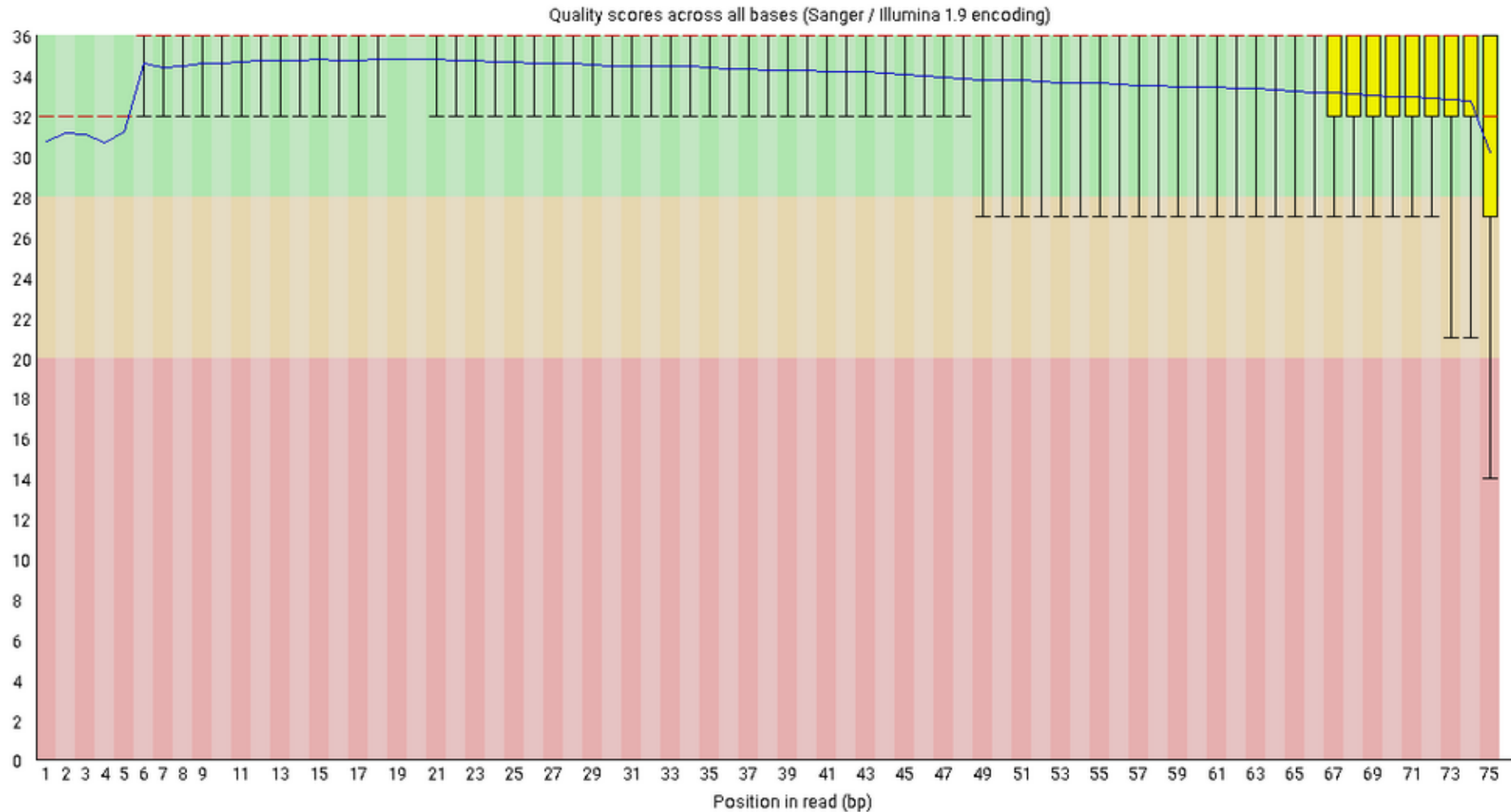
0 diversity bases added	5' <u>YGG</u> (BCO)	BCO = Bisulfite-Converted Original genomic strand
1 diversity base added	5' <u>DYGG</u> (BCO)	Y = C or T, depending on methylation status
2 diversity bases added	5' <u>DDYGG</u> (BCO)	D = A, G, or T
3 diversity bases added	5' <u>RDDYGG</u> (BCO)	R = A or G



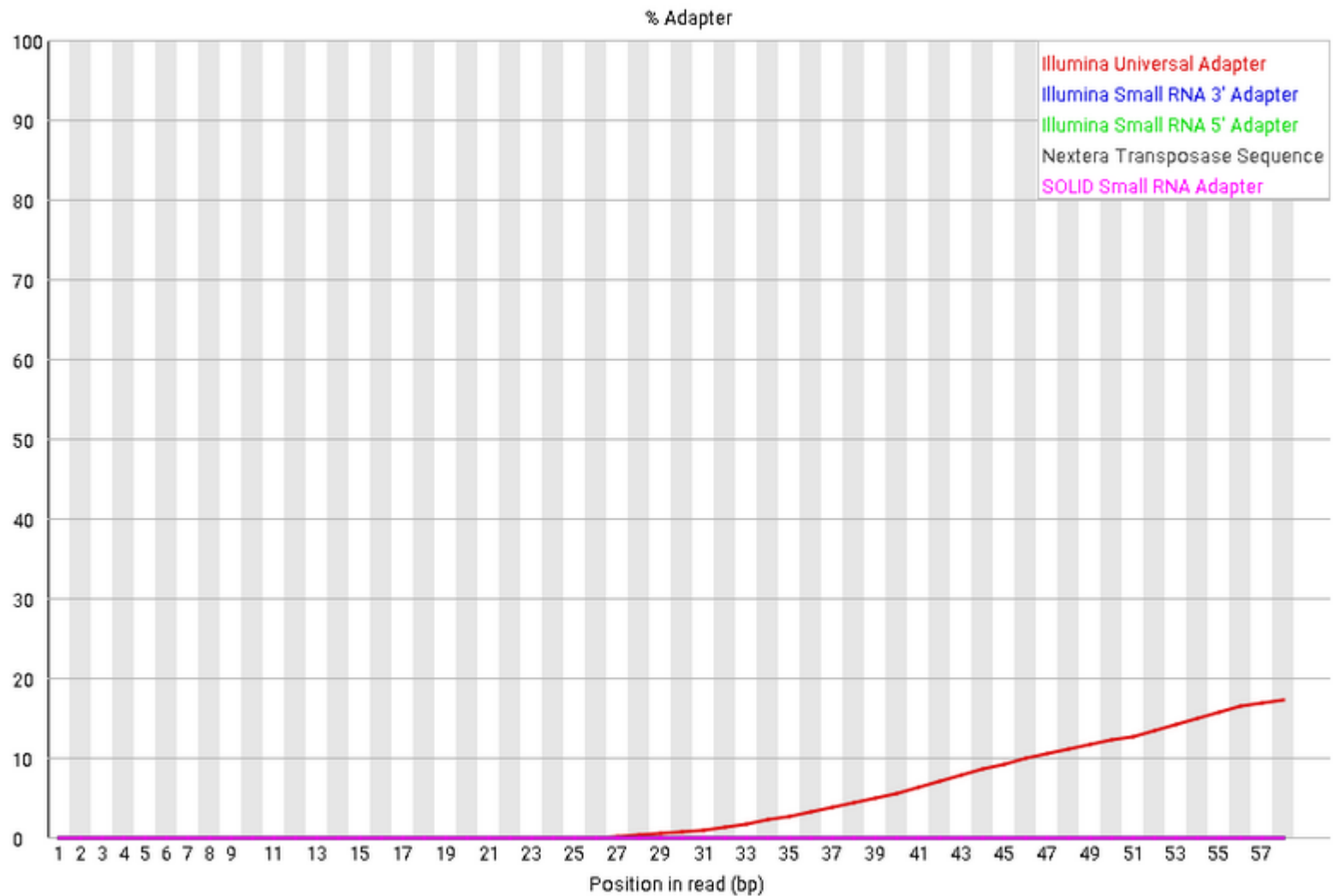
Quality Control



Quality Control



Adapter Clipping



Adapter Clipping

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

Trimming & Clipping

Tools

cutadapt

FASTA/FASTQ

Cutadapt Remove adapter sequences from Fastq/Fasta

Genome editing

MAGeCK count - collect sgRNA read counts from read mapping files

Workflows

All workflows

anno2bed

Cutadapt Remove adapter sequences from Fastq/Fasta (Galaxy Version 1.16.5)

☆ Favorite 🔄 Versions ▼ Options

Single-end or Paired-end reads?

Paired-end

FASTQ/A file #1

60: 1T_R1

👤 This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Should be of datatype "fastq.gz" or "fasta"

FASTQ/A file #2

55: 1T_R2

👤 This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Should be of datatype "fastq.gz" or "fasta"

Read 1 Options

3' (End) Adapters

Trimming & Clipping

Read 1 Options

3' (End) Adapters

1: 3' (End) Adapter

Source

Enter custom 3' adapter name (Optional)

Enter custom 3' adapter sequence

(-a)

2: 3' (End) Adapter

Source

Enter custom 3' adapter name (Optional)

Enter custom 3' adapter sequence

Read 2 Options

3' (End) Adapters

1: 3' (End) Adapters

Source

Enter custom 3' adapter name (Optional)

Enter custom 3' adapter sequence

(-A)

2: 3' (End) Adapters

Source

Enter custom 3' adapter name (Optional)

Enter custom 3' adapter sequence

Minimum length

Discard trimmed reads that are shorter than an initial primer is not counted. Value c

Maximum length

Discard trimmed reads that are longer than an initial primer is not counted. Value c

Max N

Discard reads with more than this num

Pair filter

Which of the reads in a paired-end rea

Read Modification Options

Quality cutoff

Upload GOI 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	Auto-detect	— Additional Speci...	⚙	100% ✓
hg38.goi.gtf	1.6 MB	Auto-detect	— Additional Speci...	⚙	100% ✓
hg38.promoter.bed	816 b	Auto-detect	— Additional Speci...	⚙	100% ✓
rbs_1T.N1.R1.fastq	14.4 MB	Auto-detect	— Additional Speci...	⚙	Adding to history... ⌛
rbs_1T.N1.R2.fastq	14.4 MB	Auto-detect	— Additional Speci...	⚙	0% ⌛
rbs_1T.N2.R1.fastq	20.9 MB	Auto-detect	— Additional Speci...	⚙	0% ⌛
rbs_1T.N2.R2.fastq	20.9 MB	Auto-detect	— Additional Speci...	⚙	0% ⌛

2

Type (set all): Auto-detect

Genome (set all): — Additional Species Are B...

3

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

Learn how the NASA uses Galaxy!

Mapping

The image shows a screenshot of the Galaxy web interface for the Bismark Mapper tool. The left sidebar contains a search bar with 'bismark' entered (1). Below it, the 'FASTA/FASTQ' section lists 'Trim Galore!' (2) and 'Bismark Mapper Bisulfite reads mapper' (3). The 'Mapping' section lists 'Bismark Deduplicate' (4), 'Bismark Pretty Report' (5), 'Bismark Meth Extractor' (6), and 'Bismark bisulfite mapper (bowtie)' (7). A red 'X' is placed over the 'Bismark bisulfite mapper (bowtie)' option. The 'Workflows' section lists 'All workflows' (8). The main panel shows the 'Bismark Mapper Bisulfite reads mapper' tool (9) with version 0.20.0.1. It asks 'Will you select a reference genome in your history or use a built-in index?' and provides options to 'Generate Bismark indexes from Genome (fastq)' or 'Built-ins were indexed using default options'. The 'Select the reference genome' section shows a file 'hg38.goi.fa' (10). The 'Is this library paired?' section has a 'Paired-end' option (11). The 'Paired End Pairs' section shows 'Mate pair 1' (12) and 'Mate pair 2' (13). The 'FASTQ or FASTA files' section shows a list of files (14) and a note 'This is a batch mode input field. Separate jobs will be created for each file.' (15). The 'Minimum insert size for valid paired-end alignments' is set to 0 (16) and the 'Maximum insert size for valid paired-end alignments' is set to 500 (17).

Tools

bismark

FASTA/FASTQ

- Trim Galore! (2) Trimmer of paired-end adapter
- Bismark Mapper Bisulfite reads mapper (3)

Mapping

- Bismark Deduplicate (4) Deduplicates reads mapped by Bismark
- Bismark Pretty Report (5) Generates a graphical HTML report page from report outputs of Bismark
- Bismark Meth Extractor (6) Reports on methylation status of reads mapped by Bismark
- Bismark bisulfite mapper (bowtie) (7)

Workflows

- All workflows (8)

Bismark Mapper Bisulfite reads mapper (9) Version 0.20.0.1

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

Generate Bismark indexes from Genome (fastq) or Galaxy history

Built-ins were indexed using default options

Select the reference genome

hg38.goi.fa (10)

Is this library paired?

Paired-end (11)

Paired End Pairs (12)

1: Paired End Pairs

Mate pair 1 (13)

Mate pair 2 (14)

FASTQ or FASTA files (15)

This is a batch mode input field. Separate jobs will be created for each file.

Minimum insert size for valid paired-end alignments

0 (16)

Maximum insert size for valid paired-end alignments

500 (17)

Alignment postprocessing

1 Tools

2 samtools view

3 Filter SAM or BAM

4 Filter SAM or BAM output SAM or BAM files on FLAG MAPQ RG LN or by region (Galaxy Version 1.8)

5 Include header

6 Minimum MAPQ quality score

7 Skip alignments with any of these flag bits set

8 The alignment or this read is not primary

X VCF/BCF

! Only output alignments with all of these flag bits set

SAM/BAM

- [SAM-to-BAM](#) convert SAM to BAM
- [Slice BAM](#) by genomic regions
- [BAM-to-SAM](#) convert BAM to SAM
- [Filter SAM or BAM](#) output SAM or BAM files on FLAG MAPQ RG LN or by region
- [Filter SAM or BAM](#) files on FLAG MAPQ RG LN or by region

VCF/BCF

- [bcftools call](#) SNP/indel variant calling from VCF/BCF

Variant Calling

- [bcftools call](#) SNP/indel variant calling from VCF/BCF

Workflows

- [All workflows](#)

SAM or BAM file to filter

176: Bismark Mapper on data 123, data 124, and data 1: mapped reads (as bam)
174: Bismark Mapper on data 121, data 120, and data 1: mapped reads (as bam)
172: Bismark Mapper on data 117, data 116, and data 1: mapped reads (as bam)
170: Bismark Mapper on data 113, data 112, and data 1: mapped reads (as bam)
168: Bismark Mapper on data 109, data 108, and data 1: mapped reads (as bam)

This is a batch mode input field. Separate jobs will be triggered for each dataset

Header in output

Include header

Minimum MAPQ quality score

1

(-q)

Filter on bitwise flag

yes

Only output alignments with all of these flag bits set

☐ Select/Unselect all

- ☐ Read is paired
- ☐ Read is mapped in a proper pair
- ☐ The read is unmapped
- ☐ The mate is unmapped
- ☐ Read strand
- ☐ Mate strand
- ☐ Read is the first in a pair
- ☐ Read is the second in a pair
- ☐ The alignment or this read is not primary
- ☐ The read fails platform/vendor quality checks
- ☐ The read is a PCR or optical duplicate
- ☐ Supplementary alignment

(-f)

Skip alignments with any of these flag bits set

☐ Select/Unselect all


- ☐ Read is paired
- ☐ Read is mapped in a proper pair
- ☒ The read is unmapped
- ☐ The mate is unmapped
- ☐ Read strand
- ☐ Mate strand
- ☐ Read is the first in a pair
- ☐ Read is the second in a pair
- ☒ The alignment or this read is not primary
- ☐ The read fails platform/vendor quality checks
- ☐ The read is a PCR or optical duplicate
- ☐ Supplementary alignment


(-F)

Alignment postprocessing

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

Alignment postprocessing

Tools 



FASTA/FASTQ

[UMI-tools group](#) Extract UMI from fastq files

SAM/BAM

[Filter pileup](#) on coverage and SNPs




[Generate pileup](#) from BAM dataset




[SAM-to-BAM](#) convert SAM to BAM

[Samtools flagstat](#) tabulate descriptive stats for BAM dataset


[samtools mpileup](#) multi-way pileup of variants

Samtools sort order of storing aligned sequences (Galaxy Version 2.0.2)


BAM File   

203: Filter SAM or BAM, output SAM or BAM on data 176: bam
202: Filter SAM or BAM, output SAM or BAM on data 174: bam
201: Filter SAM or BAM, output SAM or BAM on data 172: bam
200: Filter SAM or BAM, output SAM or BAM on data 170: bam
199: Filter SAM or BAM, output SAM or BAM on data 168: bam
198: Filter SAM or BAM, output SAM or BAM on data 166: bam
197: Filter SAM or BAM, output SAM or BAM on data 164: bam
196: Filter SAM or BAM, output SAM or BAM on data 162: bam
176: Bismark Mapper on data 125, data 124, and data 1: mapped

 This is a batch mode input field. Separate jobs will be triggered

Primary sort key

 **Execute**

Infer 5mC methylation level

Tools

1

2

methyldackel

Epigenetics

MethylDackel A tool for processing bisulfite sequencing alignments

Workflows

- All workflows

MethylDackel tool for processing bisulfite sequencing alignments (Galaxy Version 0.3.0.1)

3

4

5

6

7

8

Load reference genome from

History

Use the following dataset as reference sequence

1: hg38.goi.fa

REFERENCE SEQUENCE; You can upload a FASTA sequence to the history and use it as reference

sorted_alignments.bam

211: Samtools sort
210: Samtools sort
209: Samtools sort
208: Samtools sort
207: Samtools sort

This is a batch mode

What do you want to do?

Extract methylation metrics from an alignment

Merge per-Cytosine metrics from CpG

Yes No

(--mergeContext)

By default, any alignment marked as a duplicate is ignored.

Yes No

By default, if only one read in a pair aligns (a singleton) then

Yes No

By default, paired-end alignments with the properly-paired concordant and discordant is based on your aligner setting:

Yes No

Minimum MAPQ threshold to include an alignment (default 10)

0

Infer median methylation level

1	2	3	4	5	6			
track 1								
chr10	1	2	3	4	5	6		
track 2								
chr10	1	2	3	4	5	6		
chr10	chr10	track type="bedGraph" description="output CpG merged methylation levels"						
chr10	chr10	1	2	3	4	5	6	
chr10	chr10	chr10	track type="bedGraph" description="output CpG merged methylation levels"					
chr10	chr10	chr10	chr10	2271	2273	50	2	2
chr10	chr10	chr10	chr10	2349	2351	25	1	3
chr10	chr10	chr10	chr10	9788	9790	3	4	102
chr10	chr10	chr10	chr10	9794	9796	29	31	74
chr10	chr10	chr10	chr10	9804	9806	34	37	70
chr10	chr10	chr10	chr10	9862	9864	50	48	47
chr10	chr10	chr10	chr10	9922	9924	75	15	5
chr10	chr10	chr10	chr10	9936	9938	100	5	0
chr10	chr10	chr10	chr10	9958	9960	20	1	4
chr10	chr10	chr10	chr10	12433	12435	85	6	1
chr10	chr10	chr10	chr10	12460	12462	100	7	0
chr10	chr10	chr10	chr10	12514	12516	14	1	6
chr10	chr10	chr10	chr10	35213	35215	13	4	25
chr10	chr10	chr10	chr10	35229	35231	48	14	15
chr10	chr10	chr10	chr10	35257	35259	41	12	17
chr10	chr10	chr10	chr10	35266	35268	13	4	25
chr10	chr10	chr10	chr10	35273	35275	37	11	18
chr10	chr10	chr10	chr10	35287	35289	9	3	29
chr10	chr10	chr10	chr10	37004	37006	0	0	5
chr10	chr10	chr10	chr10	37009	37011	50	2	2

chrgoi	9788	9790	3	8	232
chrgoi	9788	9790	5	20	325
chrgoi	9788	9790	5	19	347
chrgoi	9788	9790	4	12	229
chrgoi	9794	9796	26	63	176
chrgoi	9794	9796	37	126	206
chrgoi	9794	9796	48	175	183
chrgoi	9794	9796	41	99	139
chrgoi	9804	9806	31	77	165
chrgoi	9804	9806			
chrgoi	9804	9806			
chrgoi	9804	9806			
chrgoi	9862	9864			
chrgoi	9862	9864			
chrgoi	9862	9864			
chrgoi	9862	9864			
chrgoi	9862	9864			
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	35257	35259			47

Intersect with promoter regions

Chrom	Start	End	Name
chr1	37553	38553	ENSG00000136859
chr1	85184	86184	ENSG00000110934
chr1	88182	89182	ENSG00000137948
chr1	157238	158238	ENSG00000121797
chr1	167073	168073	ENSG00000073067
chr1	177515	178515	ENSG00000166801
chr1	229319	230319	ENSG00000269190
chr1	239797	240797	ENSG00000073067
chr1	242795	243795	ENSG00000073067
chr1	250261	251261	ENSG00000073067
chr1	414709	415709	ENSG00000073067
chr1	420561	421561	ENSG00000073067
chr1	427763	428763	ENSG00000073067
chr1	619900	620900	ENSG00000073067
chr1	622898	623898	ENSG00000073067
chr1	805693	806693	ENSG00000073067
chr1	812937	813937	ENSG00000088533
chr1	849720	850720	ENSG00000137872
chr1	1443843	1444843	ENSG00000139343
chr1	1711347	1712347	ENSG00000166444
chr1	1748149	1749149	ENSG00000137868
chr1	1751147	1752147	ENSG00000171992

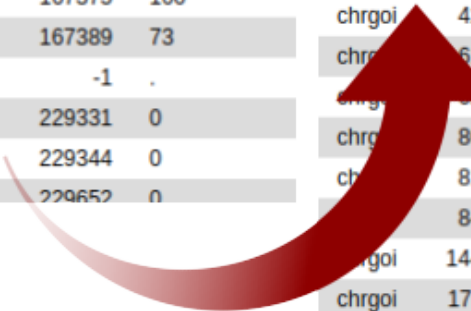
chr1	2271	2273	65
chr1	2349	2351	47.5
chr1	9788	9790	4.25
chr1	9794	9796	38
chr1	9804	9806	49
chr1	9862	9864	46
chr1	9922	9924	78.5
chr1	9936	9938	81.33333333
chr1	9958	9960	65.66666667
chr1	12433	12435	24.33333333
chr1	12460	12462	100
chr1	12514	12516	71.66666667
chr1	35213	35215	11.5
chr1	35229	35231	56.5
chr1	35257	35259	47

chr1	85184	86184	ENSG00000110934	chr1	85189	85193	41.66666667
chr1	85184	86184	ENSG00000110934	chr1	85201	85203	100
chr1	85184	86184	ENSG00000110934	chr1	85229	85231	50
chr1	85184	86184	ENSG00000110934	chr1	85234	85236	50
chr1	85184	86184	ENSG00000110934	chr1	85243	85245	0
chr1	88182	89183	ENSG00000137948	.	-1	-1	.
chr1	157238	158239	ENSG00000121797	.	-1	-1	.
chr1	167073	168074	ENSG00000073067	chr1	167297	167299	30.33333333
chr1	167073	168074	ENSG00000073067	chr1	167301	167303	66
chr1	167073	168074	ENSG00000073067	chr1	167364	167366	87.5
chr1	167073	168074	ENSG00000073067	chr1	167373	167375	100
chr1	167073	168074	ENSG00000073067	chr1	167387	167389	73
chr1	177515	178516	ENSG00000166801	.	-1	-1	.
chr1	229319	230319	ENSG00000269190	chr1	229327	229331	0
chr1	229319	230319	ENSG00000269190	chr1	229342	229344	0
chr1	229319	230319	ENSG00000269190	chr1	229650	229652	0

Infer median promoter me level

chr1	85184	86184	ENSG00000110934	chr1	85189	85193	41.66666667
chr1	85184	86184	ENSG00000110934	chr1	85201	85203	100
chr1	85184	86184	ENSG00000110934	chr1	85229	85231	50
chr1	85184	86184	ENSG00000110934	chr1	85234	85236	50
chr1	85184	86184	ENSG00000110934	chr1	85243	85245	0
chr1	88182	89183	ENSG00000137948	.	-1	-1	.
chr1	157238	158239	ENSG00000121797	.	-1	-1	.
chr1	167073	168074	ENSG00000073067	chr1	167297	167299	30.33333333
chr1	167073	168074	ENSG00000073067	chr1	167301	167303	66
chr1	167073	168074	ENSG00000073067	chr1	167364	167366	87.5
chr1	167073	168074	ENSG00000073067	chr1	167373	167375	100
chr1	167073	168074	ENSG00000073067	chr1	167387	167389	73
chr1	177515	178516	ENSG00000166801	.	-1	-1	.
chr1	229319	230319	ENSG00000269190	chr1	229327	229331	0
chr1	229319	230319	ENSG00000269190	chr1	229342	229344	0
chr1	229319	230319	ENSG00000269190	chr1	229650	229652	0

chr1	37553	38553	ENSG00000136859	.
chr1	85184	86184	ENSG00000110934	40.27777778
chr1	88182	89183	ENSG00000137948	.
chr1	157238	158239	ENSG00000121797	.
chr1	167073	168074	ENSG00000073067	71.36666667
chr1	177515	178516	ENSG00000166801	.
chr1	229319	230319	ENSG00000269190	16.4
chr1	239797	240797	ENSG00000127951	.
chr1	242795	243796	ENSG00000189280	45
chr1	250261	251262	ENSG00000196208	46
chr1	414709	415710	ENSG00000170276	.
chr1	420561	421562	ENSG00000129009	.
chr1	427763	428764	ENSG00000172005	.
chr1	619900	620900	ENSG00000143850	2.135416667
chr1	622898	623899	ENSG00000132334	8.333333333
chr1	805693	806694	ENSG00000188060	33
chr1	812937	813938	ENSG00000008853	.
chr1	849720	850721	ENSG00000137872	0
chr1	1443843	1444844	ENSG00000139343	25.375
chr1	1711347	1712347	ENSG00000166444	16.29166667
chr1	1748149	1749149	ENSG00000137868	80
chr1	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

Tools

join

Collection Operations

- [Column Join](#) on Collections

Text Manipulation

- [Histogram](#) of a numeric column
- [Join two files](#) on column allowing a small difference
- [Subtract](#) the intervals of two datasets
- [Sort](#) data in ascending or descending order
- [Join two files](#)
- [Multi-Join](#) (combine multiple files)
- [Join](#) the intervals of two datasets side-by-side

Filter and Sort

- [Query Tabular](#) using sqlite sql

Join, Subtract and Group

- [Datamash](#) (operations on tabular data)
- [Reverse](#) columns in a tabular file
- [Transpose rows/columns](#) in a

Join two files Galaxy Version 1.1.1

1st file

253: DGE.tsv

Column to use from 1st file

Column: 1

2nd file

DME.tsv

Column to use from 2nd file

Column: 1

Output lines appearing in

Both 1st & 2nd file.

First line is a header line

☐ Yes ☒ No

Use if first line contains column headers. It will not be sorted.

Ignore case

☐ Yes ☒ No

Sort and Join key column values regardless of upper/lower case letters.

Value to put in unpaired (empty) fields

0

☒ Execute

Scatterplot

Download "Galaxy254-[Join_on_data_248_and_data_253].tabular?"

File Name: Galaxy254-[Join_on_data_248_and_data_253].tabular
File Size: 930 bytes
Source: usegalaxy.eu

☐ Always Save to Default Download Location

Save Save As... Open Cancel

1	2	3
ENSG00000008853	0.5429356	0.0
ENSG00000073067	1.880033	-4.761904762
ENSG00000110934	2.232084	-80.0
ENSG00000121797	0.1168624	0.0
ENSG00000127951	0.3433029	0.0
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

History

search datasets

Rostock
246 shown, 8 deleted, 1 hidden
868.02 MB

254: Join on data 248 and data 253
22 lines
format: tabular, database: ?

1	2	3
ENSG00000008853	0.542935631009544	19
ENSG00000073067	1.88003372546044	-4
ENSG00000110934	2.23208401512721	-5
ENSG00000121797	0.116862409130137	0.
ENSG00000127951	0.343302983265895	0.

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',
  repeat = TRUE) +
  gradient_color(c('blue', 'white', 'red')) +
  geom_density_2d();
pdf('scatter.pdf');
ggMarginal(p, type = 'boxplot');
graphics.off();
```

Scatterplot

The image shows the RStudio interface with four red arrows pointing to specific elements:

- Arrow 1** points to the **Environment** pane on the right, which shows the data frame `df` with 22 observations and 3 variables.
- Arrow 2** points to the **Run** button in the top toolbar.
- Arrow 3** points to the **Source** pane on the left, which contains the R code for creating the scatterplot.
- Arrow 4** points to the **Files** pane on the right, which shows the file `scatter.pdf` in the `project` directory.

The R code in the Source pane is as follows:

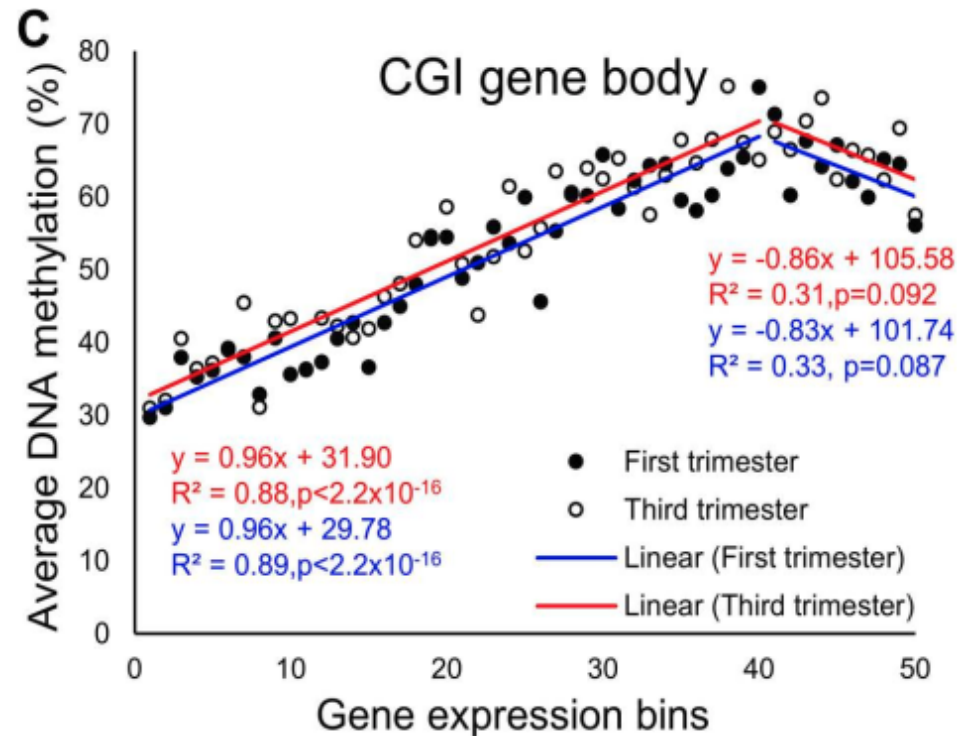
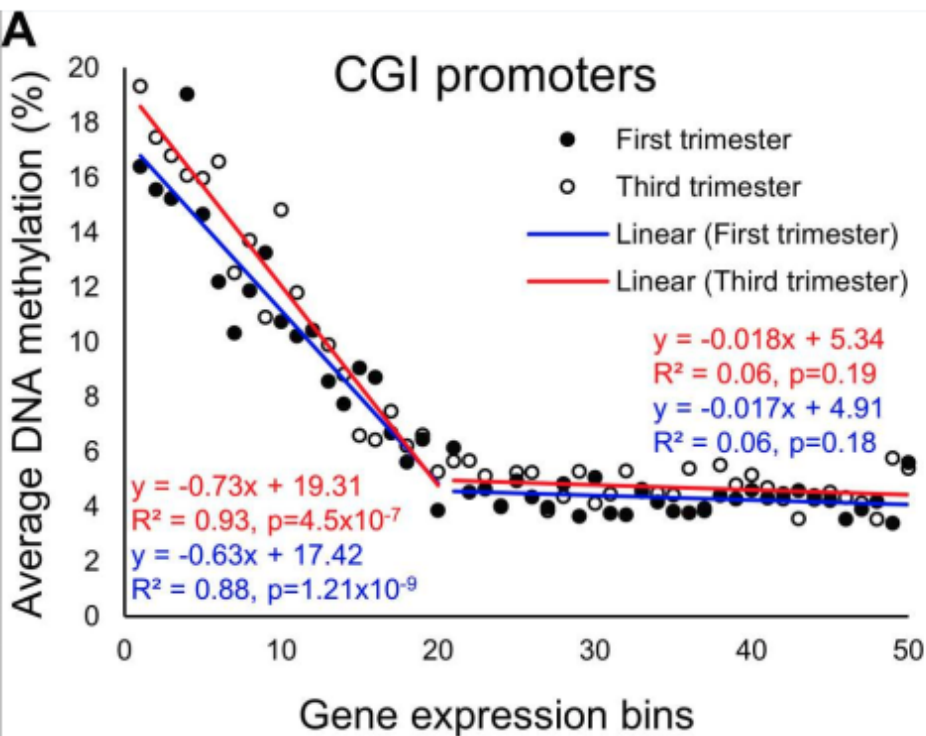
```
1 library('ggpubr');
2 library('ggExtra');
3 df=read.table('scatter.tsv', header = FALSE, sep = '\t');
4 colnames(df) = c('gene', 'DGE', 'DME');
5 p = ggscatter(df, x = 'DME', y = 'DGE', color = 'DGE', label = 'gene', repel = TRUE) +
6   gradient_color(c('blue', 'white', 'red')) +
7   geom_density_2d();
8 pdf('scatter.pdf');
9 ggMarginal(p, type = 'boxplot');
10 graphics.off();
11
```

The Console pane at the bottom shows the execution of the same code:

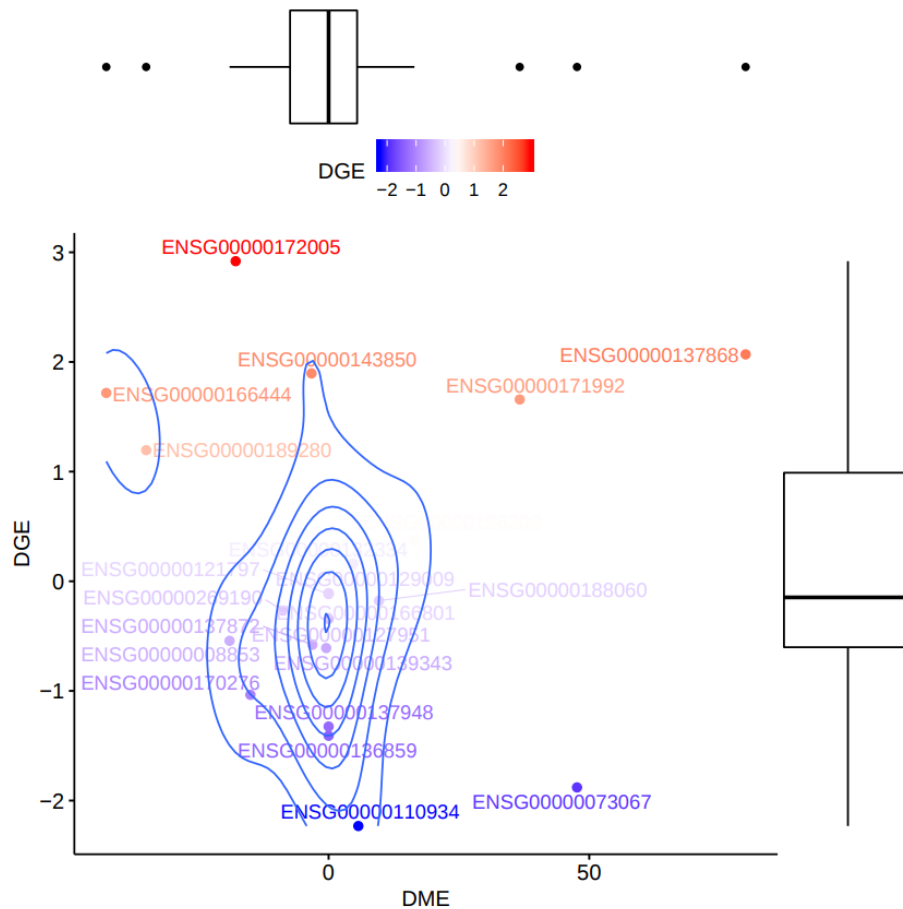
```
> library('ggpubr');
> library('ggExtra');
> df=read.table('scatter.tsv', header = FALSE, sep = '\t');
> colnames(df) = c('gene', 'DGE', 'DME');
> p = ggscatter(df, x = 'DME', y = 'DGE', color = 'DGE', label = 'gene', repel = TRUE) +
+   gradient_color(c('blue', 'white', 'red')) +
+   geom_density_2d();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

An **Upload Files** dialog box is open in the center, showing the target directory `/cloud/project` and a **Choose File** button. A tip at the bottom of the dialog states: "TIP: To upload multiple files or a directory, create a zip file. The zip file will be automatically expanded after upload."

Recap



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

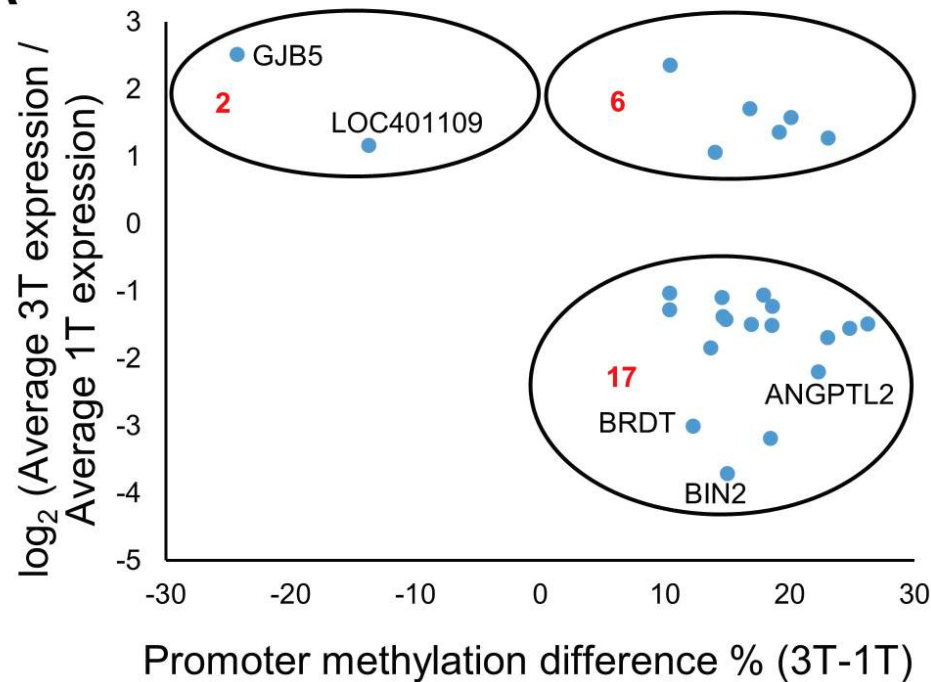
RAB42: Onkogen

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

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

Comparison

A



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