Using Galaxy for NGS Analysis in a collaborative environment

2nd Swiss Galaxy Workshop
October 1, 2014, Bern

Hans-Rudolf Hotz (hrh@fmi.ch)

Friedrich Miescher Institute for Biomedical Research
Basel, Switzerland

Friedrich Miescher Institute

- funded by the Novartis Research Foundation
- affiliated institute of Basel University

325 employees
(incl. 97 PhD students, 103 Post Docs)

Epigenetics
(7 research groups)

Cancer
(8 research groups)

Neurobiology
(8 research groups)

Technology Platforms

Computational Biology – Cell Sorting – Imaging and Microscopy – *C. elegans*
Functional Genomics – Histology – Mass Spectrometry – Protein Structure

FMI
Friedrich Miescher Institute
for Biomedical Research

The Computational Biology platform is providing support for....

the “average” lab scientist, using computers to:

- draw plasmids
- do BLAST searches
- use Excel

the “modern” lab scientist, using computers to:

- analyze NGS data with R/Bioconductor scripts

FMI
Friedrich Miescher Institute
for Biomedical Research

The Computational Biology platform is providing support for....

the “average” lab scientist, using computers to:

draw plasmids

do BLAST searches

use Excel

?

the “modern” lab scientist, using computers to:

analyze NGS data with R/Bioconductor scripts

FMI

Friedrich Miescher Institute
for Biomedical Research

The Computational Biology platform is providing support for....

the “average” lab scientist, using computers to:

?

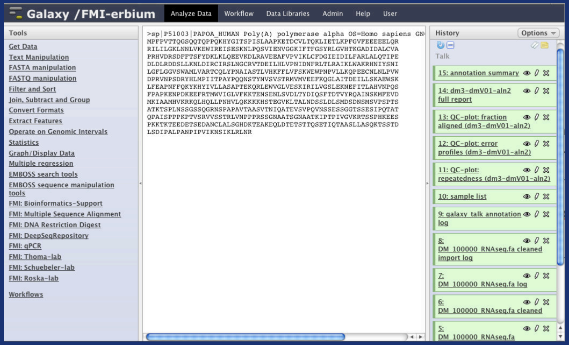
→

the “modern” lab scientist, using computers to:

draw plasmids

do BLAST searches

use Excel



Galaxy /FMI-erbium
Analyze Data
Workflow
Data Libraries
Admin
Help
User
Tools
Get Data
Text Manipulation
FASTA manipulation
FASTQ manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Operate on Genomic Intervals
Statistics
Graph/Display Data
Multiple regression
EMBOSS search tools
EMBOSS sequence manipulation tools
FMI: Bioinformatics-Support
FMI: Multiple Sequence Alignment
FMI: DNA Restriction Digest
FMI: DeepSeqRepository
FMI: qPCR
FMI: Thoma-lab
FMI: Schuebeler-lab
FMI: Roska-lab
Workflows
History
Options
Talk
15: annotation summary
14: dm3-dmV01-aln2 full report
13: QC-plot: fraction aligned (dm3-dmV01-aln2)
12: QC-plot: error profiles (dm3-dmV01-aln2)
11: QC-plot: repeatedness (dm3-dmV01-aln2)
10: sample list
9: galaxy talk annotation log
8: DM 100000 RNAseq.fa cleaned import log
7: DM 100000 RNAseq.fa log
6: DM 100000 RNAseq.fa cleaned
5: DM 100000 RNAseq.fa

```
>sp|P51003|PAPOA_HUMAN Poly(A) polymerase alpha OS=Homo sapiens GN
MPFPVTTQGSQQTQPPQKHYGITSPISLAAPKETDCVLTQKLIETLKPFGVFEEEEELQR
RILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGADIDALCVA
PRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFARLALQTIPE
DLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAKRHNIYSNI
LGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEECNLNLPVW
DPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAITDEILLSKAEWSK
LFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEKNEFITLAHVNPQS
FPAPKENPDKEEFRTMWVIGLVFKKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFEVD
MKIAAMHVKRKQLHQLLPNHVLQKKKKHSTEGVKLTALNDSSLDLSMDSDNSMSVPSPTS
ATKTSPLNSSGSSQGRNSPAPAVTAASVTNIQATEVSVPQVNSSESSGGTSSESIPQTAT
QPAISPPPKPTVSRVVSSTRLVNPPPRSSGNAATSGNAATKIPTPIVGVKRTSSPHKEES
PKKTKTEEDETSEDANCLALSGHDKTEAKEQLDTETSTTQSETIQTAASLLASQKTSSTD
LSDIPALPANPIPVIKNSIKLRLNR
```

analyze NGS data with R/Bioconductor scripts

http://galaxyproject.org/

FMI

Friedrich Miescher Institute for Biomedical Research

why are we using Galaxy

Galaxy

- open source software / no license fee
- it provides a standard set of Bioinformatics tools
- we can add our own scripts and tools
- the Galaxy community is huge
- a local installation is simple to set up
- it is flexible (we can adjust it to our needs)

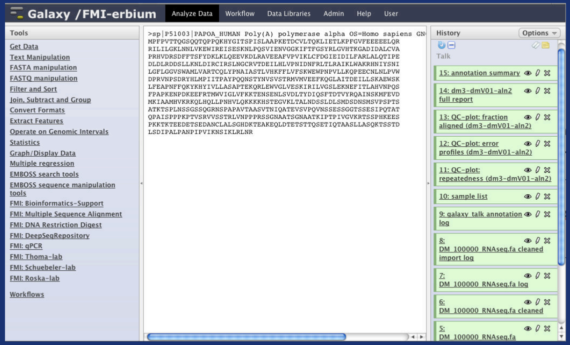
in use at the FMI since 2007

FMI
Friedrich Miescher Institute
for Biomedical Research

Galaxy as a stepping stone

the “average” lab scientist, using computers to:

- draw plasmids
- do BLAST searches
- use Excel


Galaxy /FMI-erbium
Analyze Data
Workflow
Data Libraries
Admin
Help
User
Tools
Get Data
Text Manipulation
FASTA manipulation
FASTQ manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Operate on Genomic Intervals
Statistics
Graph/Display Data
Multiple regression
EMBOSS search tools
EMBOSS sequence manipulation tools
FMI: Bioinformatics-Support
FMI: Multiple Sequence Alignment
FMI: DNA Restriction Digest
FMI: DeepSeqRepository
FMI: qPCR
FMI: Thoma-lab
FMI: Schuebeler-lab
FMI: Roska-lab
Workflows
History
Options
Talk
15: annotation summary
14: dm3-dmV01-aln2 full report
13: QC-plot: fraction aligned (dm3-dmV01-aln2)
12: QC-plot: error profiles (dm3-dmV01-aln2)
11: QC-plot: repeatedness (dm3-dmV01-aln2)
10: sample list
9: galaxy talk annotation log
8: DM 100000 RNAseq.fa cleaned import log
7: DM 100000 RNAseq.fa log
6: DM 100000 RNAseq.fa cleaned
5: DM 100000 RNAseq.fa

```
>sp|P51003|PAPOA_HUMAN Poly(A) polymerase alpha OS=Homo sapiens GN
MPFPVTTQGSQQTQPPQKHYGITSPISLAAPKETDCVLTQKLIETLKPFGVFEEEEELQR
RILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGADIDALCVA
PRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFARLALQTIPE
DLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAKRHNIYSNI
LGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEECNLNLPVW
DPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAITDEILLSKAEWSK
LFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEKNEFITLAHVNPQS
FPAPKENPDKEEFRTMWVIGLVFKKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFEVD
MKIAAMHVKRKQLHQLLPNHVLQKKKHSTEGVKLTALNDSSLDLSMDSDNSMSVPSPTS
ATKTSPLNSSGSSQGRNSPAPAVTAASVTNIQATEVSVPQVNSSESSGGTSSESIPQTAT
QPAISPPPKPTVSRVVSSTRLVNPPPRSSGNAATSGNAATKIPTPIVGVKRTSSPHKEES
PKKTKTEEDETSEDANCLALSGHDKTEAKEQLDTETSTTQSETIQTAASLLASQKTSSTD
LSDIPALPANPIPVIKNSIKLRLNR
```

the “modern” lab scientist, using computers to:

analyze NGS data with R/Bioconductor scripts

http://galaxyproject.org/

FMI
Friedrich Miescher Institute
for Biomedical Research

Galaxy as a stepping stone

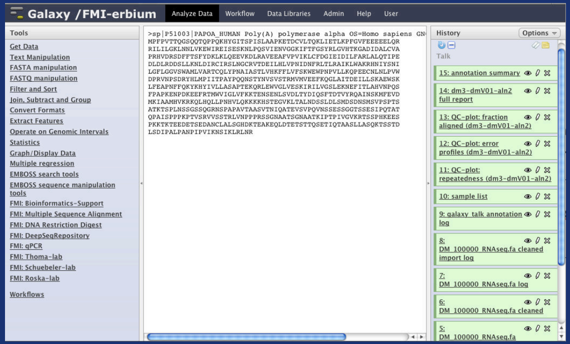
for learning Bioinformatics....

the "average" lab scientist, using computers to:

- draw plasmids
- do BLAST searches
- use Excel

the "modern" lab scientist, using computers to:

- analyze NGS data with R/Bioconductor scripts



```
>sp|P51003|PAPOA_HUMAN Poly(A) polymerase alpha OS=Homo sapiens GN
MPFPVTTQGSQQTQPPQKHYGITSPISLAAPKETDCVLTQKLIETLKPFGVFEEEEELQR
RILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGADIDALCVA
PRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFARLALQTIPE
DLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAKRHNIYSNI
LGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEECNLNLPVW
DPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAITDEILLSKAEWSK
LFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEKNEFITLAHVNPQS
FPAPKENPDKEEFRTMWVIGLVFKKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFEVD
MKIAAMHVKRKQLHQLLPNHVLQKKKKHSTEGVKLTALNDSSLDLSMDSDNSMSVPSPTS
ATKTSPLNSSGSSQGRNSPAPAVTAASVTNIQATEVSVPQVNSSESSGGTSSESIPQTAT
QPAISPPPKPTVSRVVSSTRLVNPPPRSSGNAATSGNAATKIPTPIVGVKRTSSPHKEES
PKKTKTEEDETSEDANCLALSGHDKTEAKEQLDTETSTTQSETIQTAASLLASQKTSSTD
LSDIPALPANPIPVIKNSIKLRLNR
```

15: annotation summary

14: dm3-dmV01-aln2 full report

13: QC-plot: fraction aligned (dm3-dmV01-aln2)

12: QC-plot: error profiles (dm3-dmV01-aln2)

11: QC-plot: repeatedness (dm3-dmV01-aln2)

10: sample list

9: galaxy_talk annotation log

8: DM_100000_RNAseq.fa cleaned import log

7: DM_100000_RNAseq.fa log

6: DM_100000_RNAseq.fa cleaned

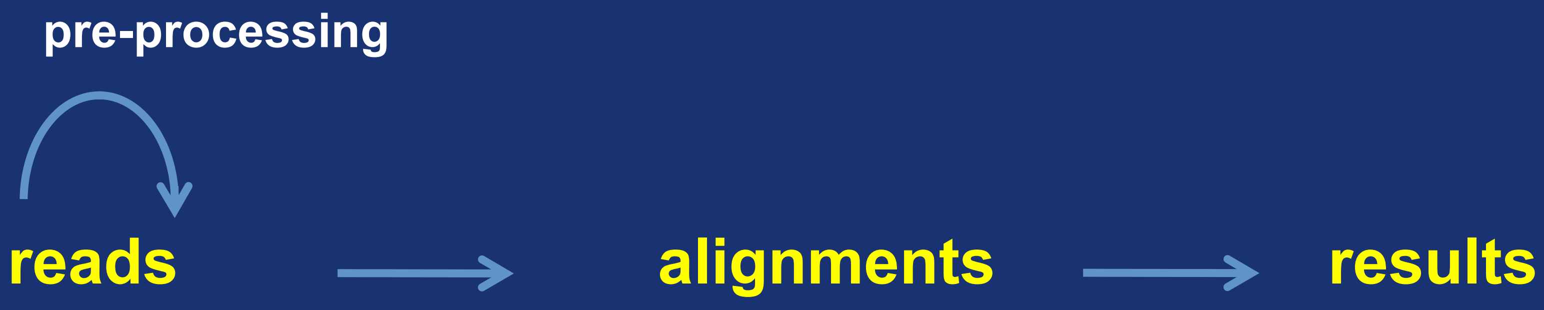
5: DM_100000_RNAseq.fa

http://galaxyproject.org/

... which is more than pressing a red button

FMI
Friedrich Miescher Institute
for Biomedical Research

quantification and analysis of NGS reads

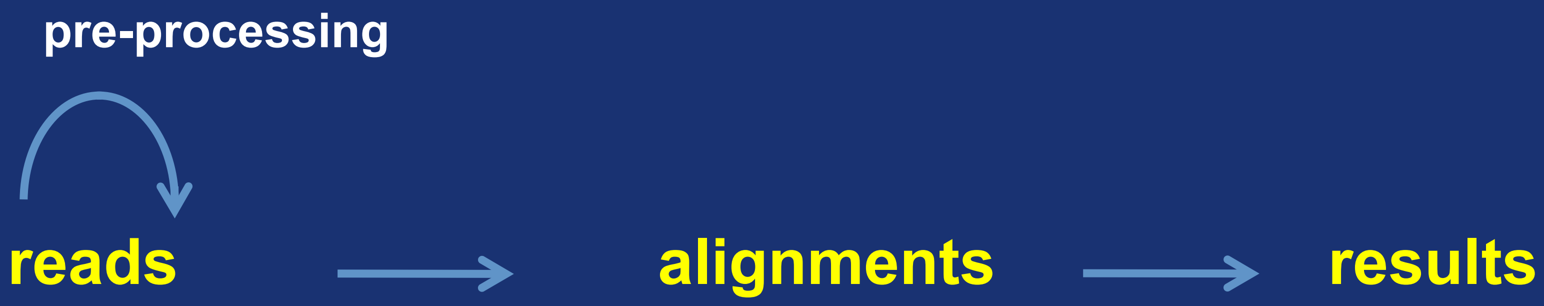


pre-processing

reads → alignments → results

counts
wig files
QC reports
methylation

quantification and analysis of NGS reads *with Galaxy*



pre-processing

reads

alignments

results

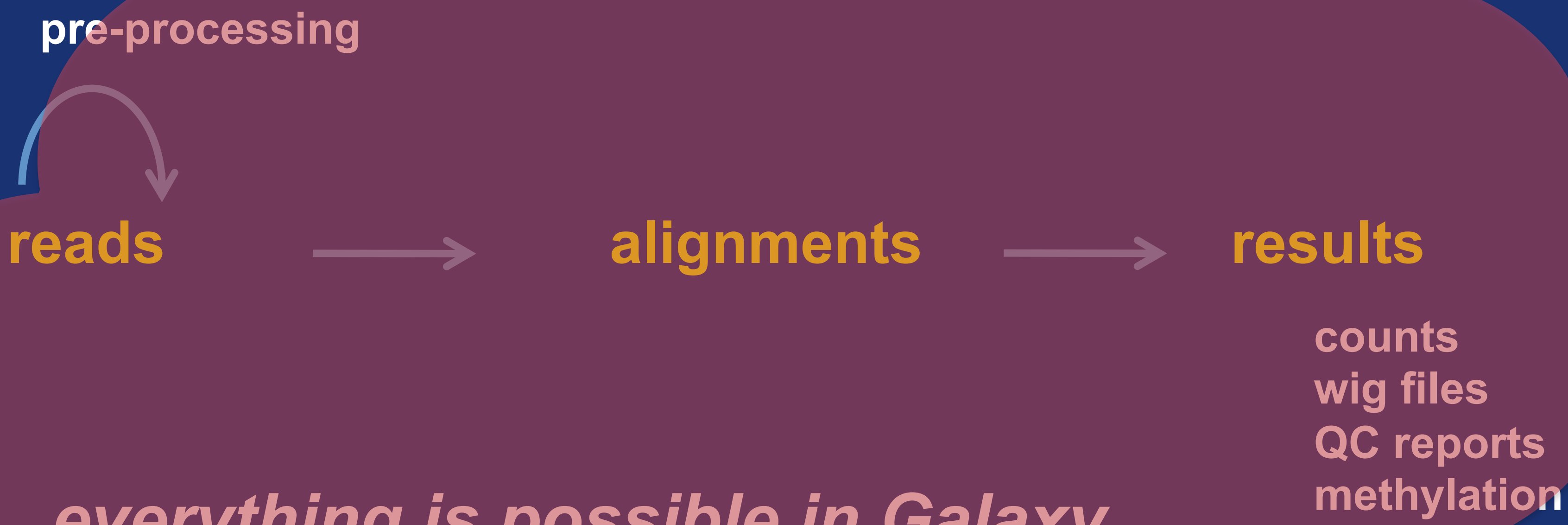
counts
wig files
QC reports
methylation

everything is possible in Galaxy

as long as you can run the tool on the command line, you can incorporate it into Galaxy.

FMI
Friedrich Miescher Institute
for Biomedical Research

quantification and analysis of NGS reads *with Galaxy*



pre-processing

reads → alignments → results

counts
wig files
QC reports
methylation

everything is possible in Galaxy

as long as you can run the tool on the command line, you can incorporate it into Galaxy.

- data is hidden in Galaxy
- data gets duplicated

FMI
Friedrich Miescher Institute
for Biomedical Research

hidden in Galaxy

Galaxy

reads

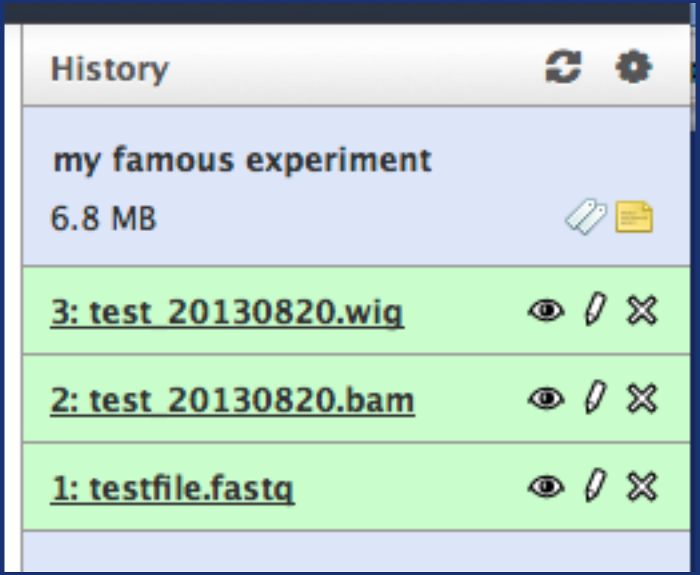
→ **fastq file**

align reads

→ **BAM file**

results

→ **wig file**



History

my famous experiment

6.8 MB

3: test_20130820.wig

2: test_20130820.bam

1: testfile.fastq

FMI
Friedrich Miescher Institute
for Biomedical Research

hidden in Galaxy, not really.....

- export Galaxy results

- sharing Galaxy histories

- Galaxy pages

data duplication, not really.....

- Galaxy data libraries

hidden in Galaxy, not really.....

- export Galaxy results

- sharing Galaxy histories

- Galaxy pages

data duplication, not really.....

- Galaxy data libraries

.... once you have started to work on the command line, you don't want to go back, no matter how brilliant Galaxy is

FMI

Friedrich Miescher Institute
for Biomedical Research

a two-way stepping stone?

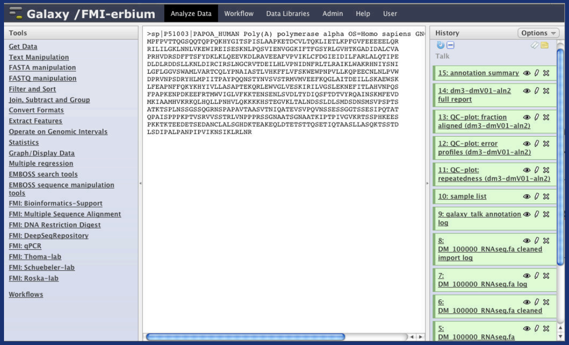
the “average” lab scientist, using computers to:

- draw plasmids
- do BLAST searches
- use Excel

?

the “modern” lab scientist, using computers to:

- analyze NGS data with R/Bioconductor scripts


Galaxy /FMI-erbium
Analyze Data
Workflow
Data Libraries
Admin
Help
User
Tools
Get Data
Text Manipulation
FASTA manipulation
FASTQ manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Operate on Genomic Intervals
Statistics
Graph/Display Data
Multiple regression
EMBOSS search tools
EMBOSS sequence manipulation tools
FMI: Bioinformatics-Support
FMI: Multiple Sequence Alignment
FMI: DNA Restriction Digest
FMI: DeepSeqRepository
FMI: qPCR
FMI: Thoma-lab
FMI: Schuebeler-lab
FMI: Roska-lab
Workflows
History
Options
Talk
15: annotation summary
14: dm3-dmV01-aln2 full report
13: QC-plot: fraction aligned (dm3-dmV01-aln2)
12: QC-plot: error profiles (dm3-dmV01-aln2)
11: QC-plot: repeatedness (dm3-dmV01-aln2)
10: sample list
9: galaxy_talk annotation log
8: DM_100000_RNAseq.fa cleaned import log
7: DM_100000_RNAseq.fa log
6: DM_100000_RNAseq.fa cleaned
5: DM_100000_RNAseq.fa

```
>sp|P51003|PAPOA_HUMAN Poly(A) polymerase alpha OS=Homo sapiens GN
MPFPVTTQGSQQTQPPQKHYGITSPISLAAPKETDCVLTQKLIETLKPFGVFEEEEELQR
RILILGKLNNLVKEWIREISESKNLPQSVIENVGGKIFTFGSYRLGVHTKGADIDALCVA
PRHVDRSDFFTSFYDKLKLQEEVKDLRAVEEAFVPVIKLCFDGIEIDILFARLALQTIPE
DLDLRDDSLLKNLDIRCIRSLNGCRVTDEILHLVPNIDNFRLTLRAIKLWAKRHNIYSNI
LGFLGGVSWAMLVARTCQLYPNAIASTLVHKFFLVFSKWEWPNPVLLKQPEECNLNLPVW
DPRVNPSDRYHLMPIITPAYPQQNSTYNVSVSTRMVMVEEFKQGLAITDEILLSKAEWSK
LFEAPNFFQKYKHYIVLLASAPTEKQRLEWVGLVESKIRILVGSLEKNEFITLAHVNPQS
FPAPKENPDKEEFRTMWVIGLVFKKTENSENLSVDLTYDIQSFTDTVYRQAINSKMFEVD
MKIAAMHVKRKQLHQLLPNHVLQKKKHSTEGVKLTALNDSSLDLSMDSDNSMSVPSPTS
ATKTSPLNSSGSSQGRNSPAPAVTAASVTNIQATEVSVPQVNSSESSGGTSSESIPQTAT
QPAISPPPKPTVSRVVSSTRLVNPPPRSSGNAATSGNAATKIPTPIVGVKRTSSPHKEES
PKKTKTEEDETSEDANCLALSGHDKTEAKEQLDTETSTTQSETIQTAASLLASQKTSSTD
LSDIPALPANPIPVIKNSIKLRLNR
```

http://galaxyproject.org/

?

collaboration?

FMI
Friedrich Miescher Institute
for Biomedical Research

storing data outside of Galaxy

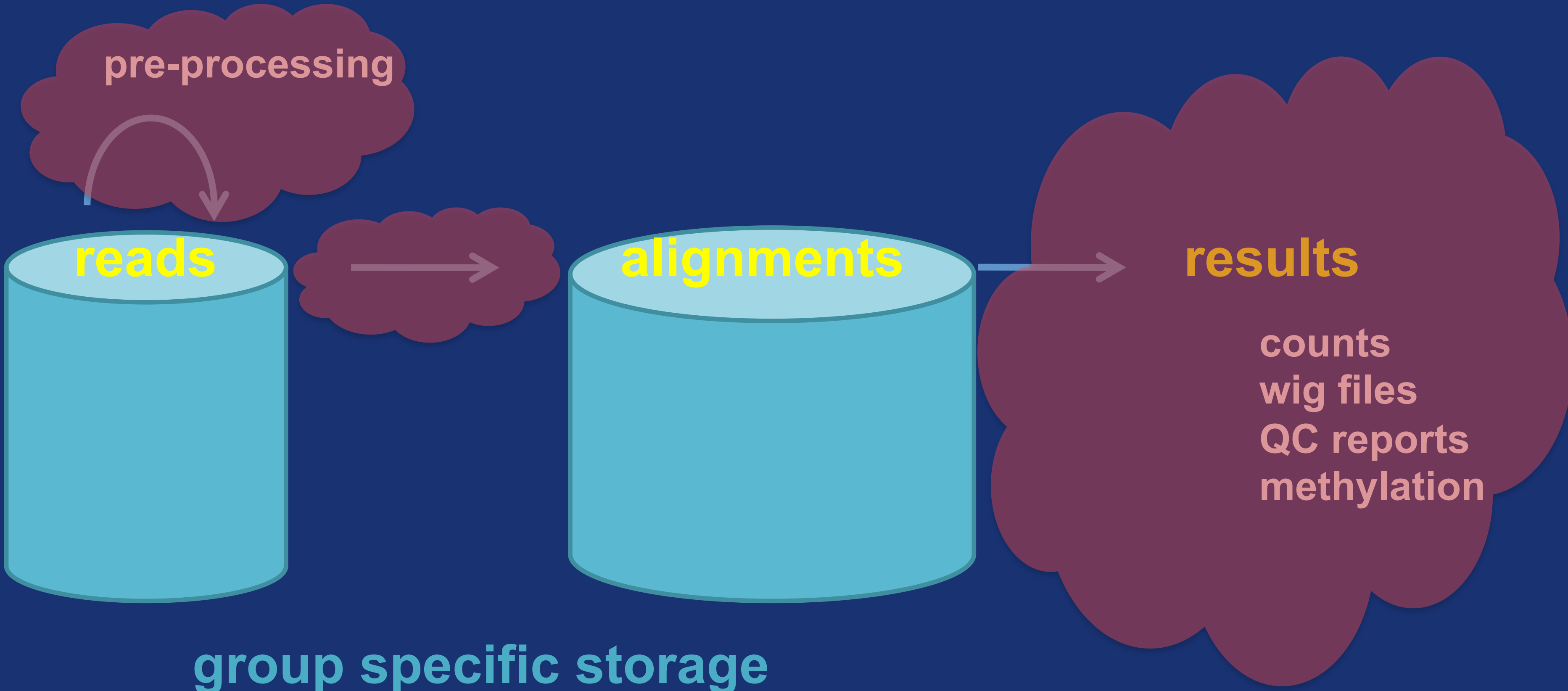
- raw data (fastq) files are in central/group specific repositories

- the Galaxy ‘aligner’ stores the BAM file in a group specific repository and creates just a 'log file' as history item

- the Galaxy ‘count’ tool uses the 'log file' as input

this is not really best (Galaxy) practice, but it allows to collaborate with non-Galaxy users ....and reproducibility is still guaranteed

quantification and analysis of NGS reads with Galaxy



group specific storage
(accessible on the command line)

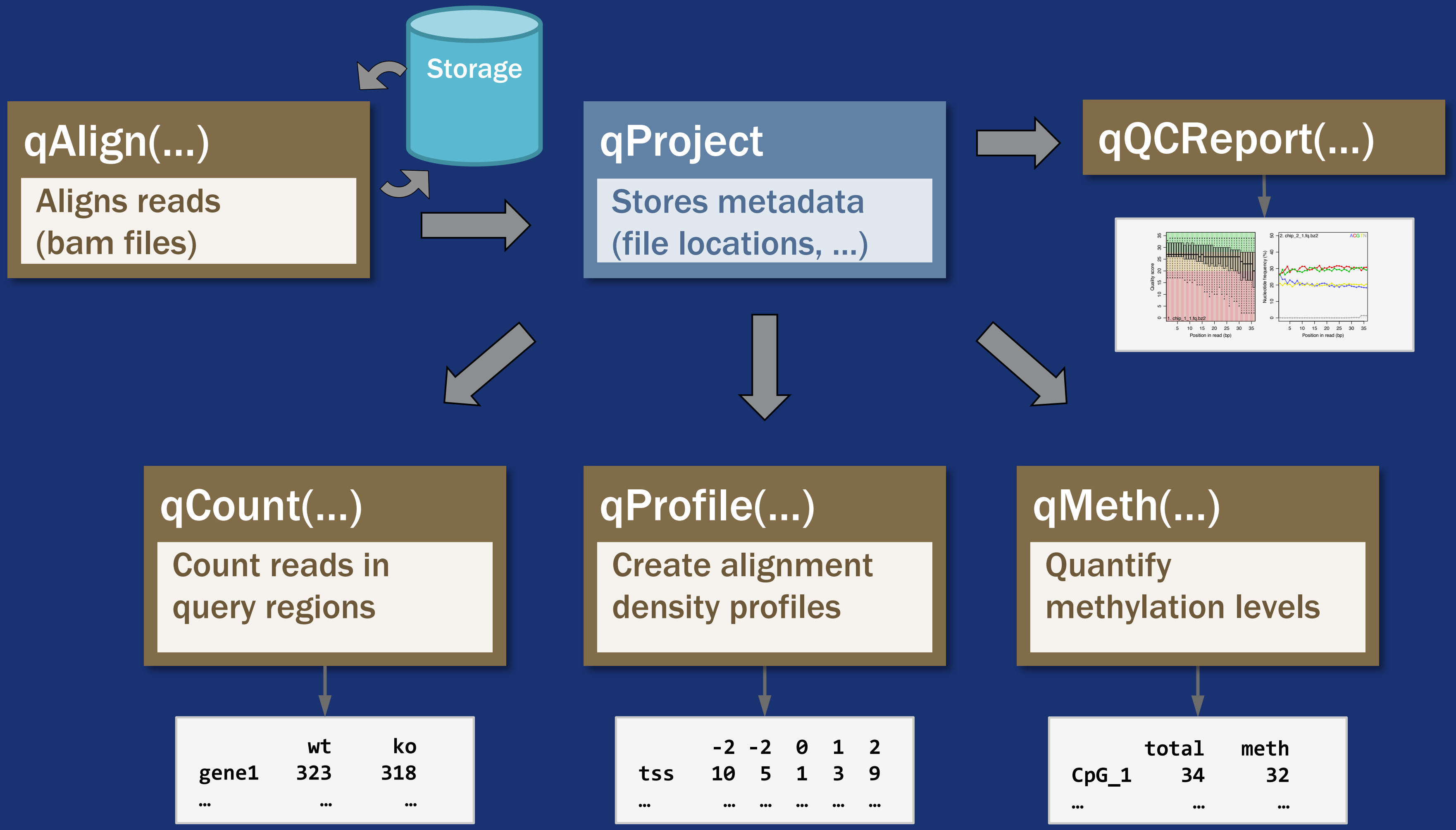
FMI
Friedrich Miescher Institute
for Biomedical Research

The (new) FMI NGS pipeline

Bioconductor package: QuasR
(Quantification and Analysis of Short Reads)

- package that provides an end-to-end analysis solution for tag counting applications
- ships with the aligners Bowtie and SpliceMap
- creates alignments from within R
- provides an additional layer of abstraction on top of pre-existing tools in Bioconductor
- makes use of Bioconductor genome and annotation packages

QuasR parts



```
             wt       ko
gene1       323      318
…             …        …
```

```
        -2  -2   0   1   2
tss     10   5   1   3   9
…        …   …   …   …   …
```

```
         total    meth
CpG_1       34      32
…            …       …
```

slide from Michael Stadler

simple RNAseq workflow with QuasR

raw files: `QuasR_rna_1_1.fastq.gz`
`QuasR_rna_1_2.fastq.gz`

align with Bowtie to: **hg19**

Bioc package: ***BSgenome.Hsapiens.UCSC.hg19***

```
qAlign("samples.txt","BSgenome.Hsapiens.UCSC.hg19")
```

raw counts for: **UCSC known genes**

Bioc package: ***TxDb.Hsapiens.UCSC.hg19.knownGene***

```
qCount(project,"TxDb.Hsapiens.UCSC.hg19.knownGene")
```

simple RNAseq workflow with QuasR

define samples: `samples.txt`

```
FileName                SampleName
/group_data/example/raw/QuasR_rna_1_1.fastq.gz  sampleA
/group_data/example/raw/QuasR_rna_1_2.fastq.gz  sampleB
```

R command to create alignments:

```
> library(QuasR)

> project <- qAlign("samples.txt",
                     "BSgenome.Hsapiens.UCSC.hg19",
                     alignmentsDir="/group_data/example/bam/")
```

FMI
Friedrich Miescher Institute
for Biomedical Research

simple RNAseq workflow with QuasR

```
> project <- qAlign("samples.txt",
              "BSgenome.Hsapiens.UCSC.hg19",
              alignmentsDir="/group_data/example/bam/")
Loading required package: Biostrings
Loading required package: XVector
alignment files missing - need to:
    create 2 genomic alignment(s)
will start in ..9s..8s..7s..6s..5s..4s..3s..2s..1s
Testing the compute nodes...OK
Loading QuasR on the compute nodes...OK
Available cores:
nodeNames
xenon1.fmi.ch
            1
Performing genomic alignments for 2 samples. See progress in
the log file:
/tmp/freiburg_example/QuasR_log_4ba87ed8d616.txt
Genomic alignments have been created successfully

>
```

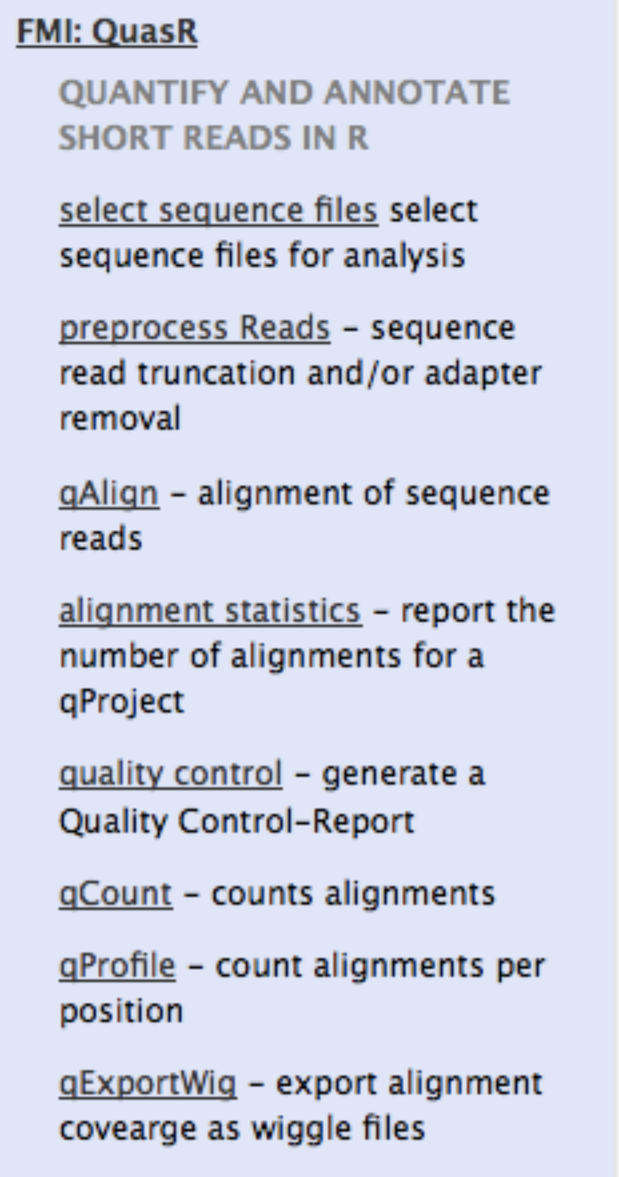
simple RNAseq workflow with QuasR

```
> project
Project: qProject
 Options    : maxHits           : 1
              paired            : no
              splicedAlignment: FALSE
              bisulfite         : no
              snpFile           : none
 Aligner    : Rbowtie v1.4.5 (parameters: -m 1 --best --strata)
 Genome     : BSgenome.Hsapiens.UCSC.hg19 (BSgenome)

 Reads      : 2 files, 2 samples (fastq format):
   1. QuasR_rna_1_1.fastq.gz  sampleA (phred33)
   2. QuasR_rna_1_2.fastq.gz  sampleB (phred33)

 Genome alignments: directory: /group_data/example/bam
   1. QuasR_rna_1_1_4ba8447d4806.bam
   2. QuasR_rna_1_2_4ba819e654f5.bam

 Aux. alignments: none
>
```

simple RNAseq workflow with QuasR

```
> library(TxDb.Hsapiens.UCSC.hg19.knownGene)
> counts <- qCount(project,TxDb.Hsapiens.UCSC.hg19.knownGene)
extracting gene regions from TranscriptDb...done
counting alignments...done
collapsing counts by query name...done
>
> dim(counts)
[1] 23459     3
>
> counts[counts[,"sampleA"]>0,]
       width sampleA sampleB
126792  2792      31      31
51150   5082     324     324
55845   1176     820     794
6201    3286      30      41
7293    1751      17      16
7428    4560     135     140
8784    1388       7       9
>
```

and now with Galaxy



FMI: QuasR

QUANTIFY AND ANNOTATE SHORT READS IN R

select sequence files select sequence files for analysis

preprocess Reads – sequence read truncation and/or adapter removal

qAlign – alignment of sequence reads

alignment statistics – report the number of alignments for a qProject

quality control – generate a Quality Control-Report

qCount – counts alignments

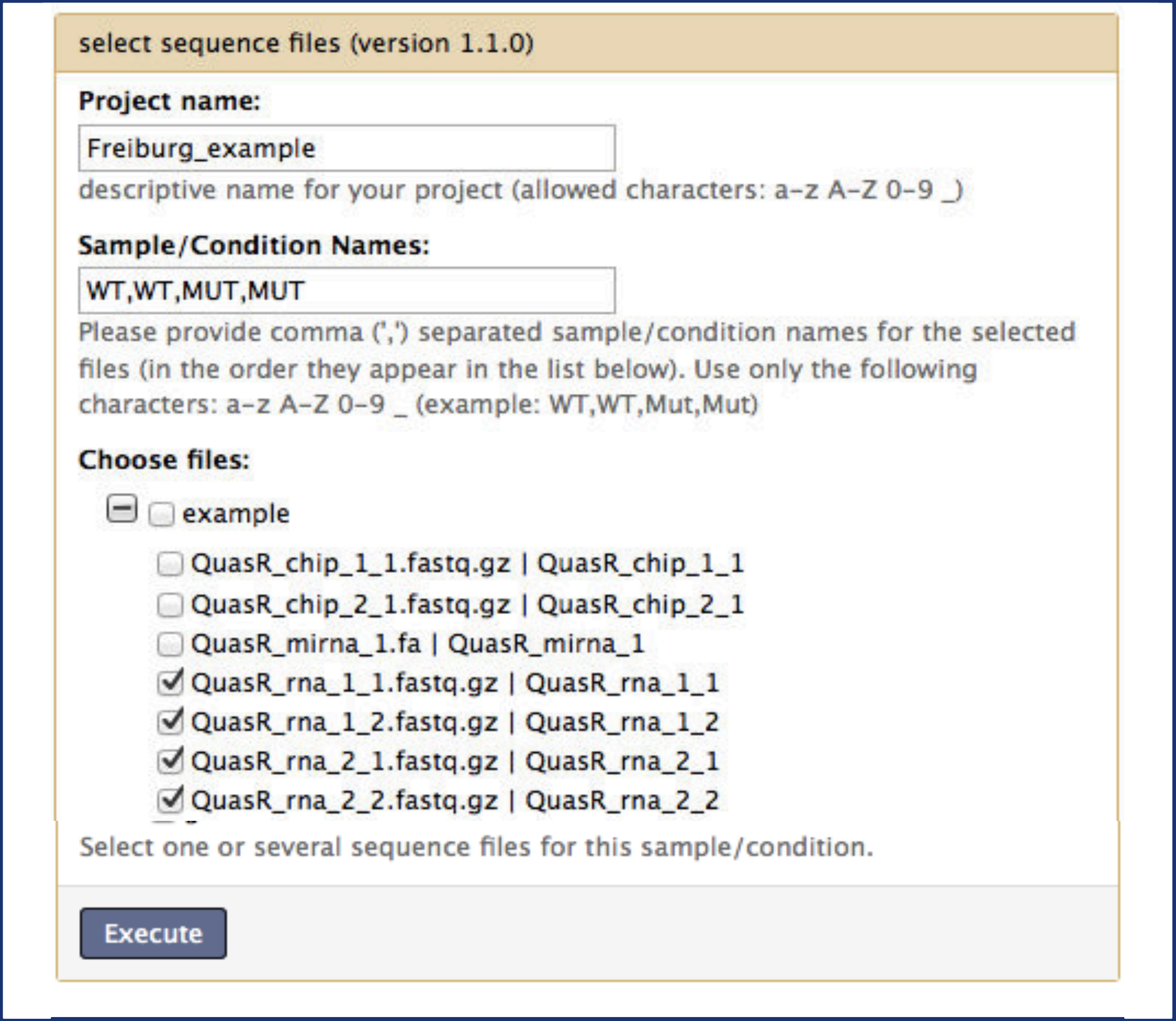
qProfile – count alignments per position

qExportWig – export alignment covearge as wiggle files

rscripts executing the QuasR steps

Friedrich Miescher Institute
for Biomedical Research

select files and assign sample names



select sequence files (version 1.1.0)

Project name:

Freiburg_example

descriptive name for your project (allowed characters: a-z A-Z 0-9 _)

Sample/Condition Names:

WT,WT,MUT,MUT

Please provide comma (',') separated sample/condition names for the selected files (in the order they appear in the list below). Use only the following characters: a-z A-Z 0-9 _ (example: WT,WT,Mut,Mut)

Choose files:

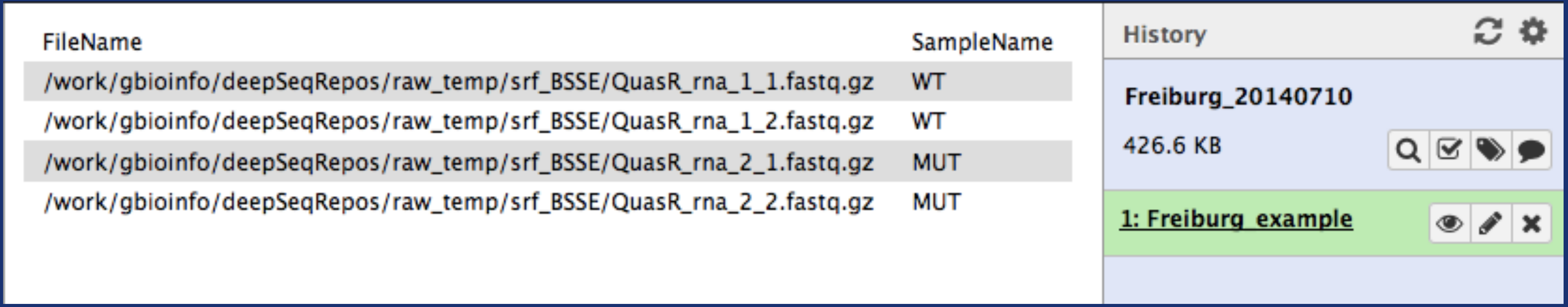
⊟ □ example

- □ QuasR_chip_1_1.fastq.gz | QuasR_chip_1_1
- □ QuasR_chip_2_1.fastq.gz | QuasR_chip_2_1
- □ QuasR_mirna_1.fa | QuasR_mirna_1
- ☑ QuasR_rna_1_1.fastq.gz | QuasR_rna_1_1
- ☑ QuasR_rna_1_2.fastq.gz | QuasR_rna_1_2
- ☑ QuasR_rna_2_1.fastq.gz | QuasR_rna_2_1
- ☑ QuasR_rna_2_2.fastq.gz | QuasR_rna_2_2

Select one or several sequence files for this sample/condition.

FMI
Friedrich Miescher Institute
for Biomedical Research

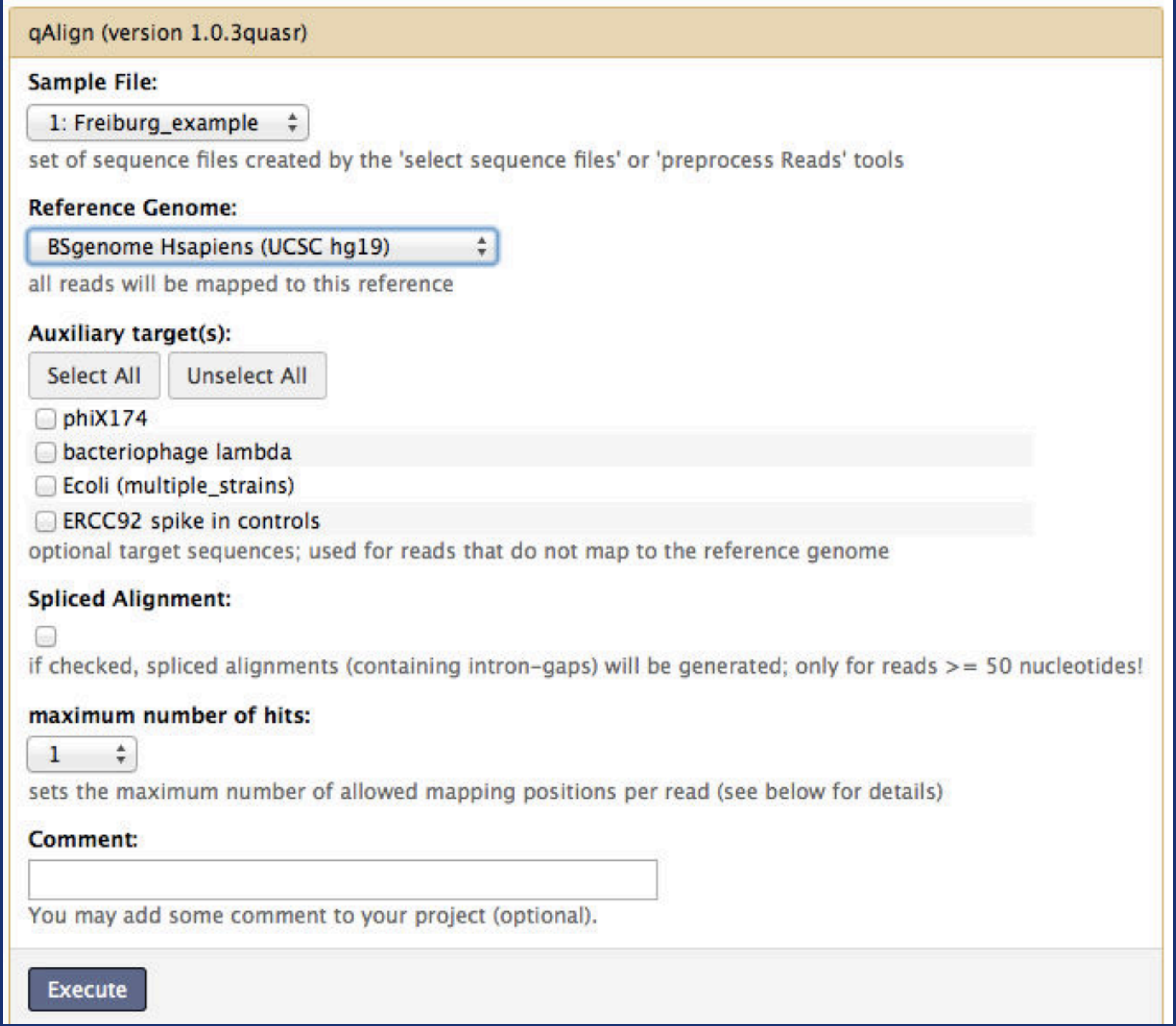
select files and assign sample names



FileName	SampleName
/work/gbioinfo/deepSeqRepos/raw_temp/srf_BSSE/QuasR_rna_1_1.fastq.gz	WT
/work/gbioinfo/deepSeqRepos/raw_temp/srf_BSSE/QuasR_rna_1_2.fastq.gz	WT
/work/gbioinfo/deepSeqRepos/raw_temp/srf_BSSE/QuasR_rna_2_1.fastq.gz	MUT
/work/gbioinfo/deepSeqRepos/raw_temp/srf_BSSE/QuasR_rna_2_2.fastq.gz	MUT

FMI
Friedrich Miescher Institute
for Biomedical Research

start alignments



qAlign (version 1.0.3quasr)

Sample File:

1: Freiburg_example

set of sequence files created by the 'select sequence files' or 'preprocess Reads' tools

Reference Genome:

BSgenome Hsapiens (UCSC hg19)

all reads will be mapped to this reference

Auxiliary target(s):

- □ phiX174
- □ bacteriophage lambda
- □ Ecoli (multiple_strains)
- □ ERCC92 spike in controls

optional target sequences; used for reads that do not map to the reference genome

Spliced Alignment:

□

if checked, spliced alignments (containing intron-gaps) will be generated; only for reads >= 50 nucleotides!

maximum number of hits:

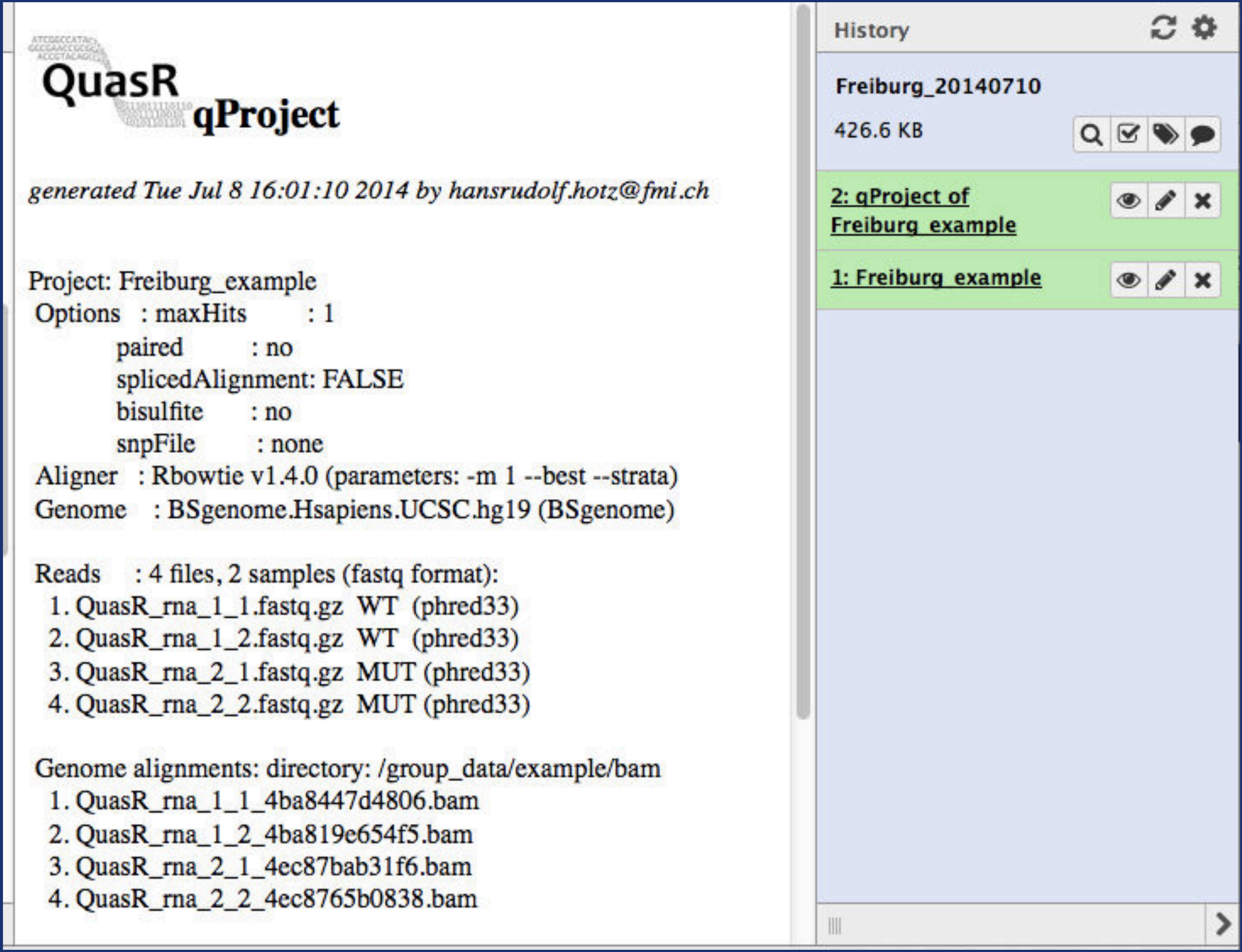
1

sets the maximum number of allowed mapping positions per read (see below for details)

Comment:

You may add some comment to your project (optional).

qProject

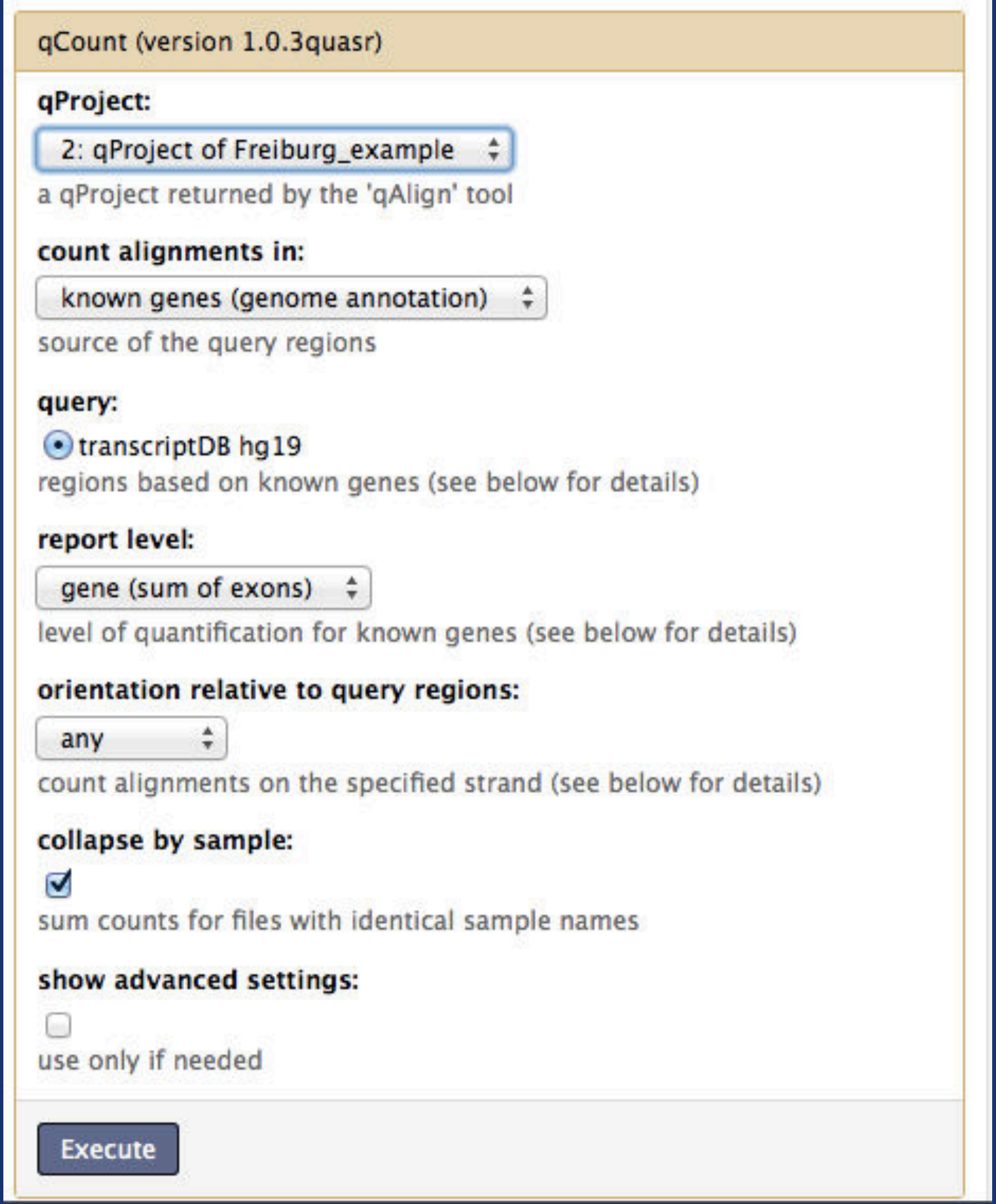


QuasR qProject

generated Tue Jul 8 16:01:10 2014 by hansrudolf.hotz@fmi.ch

```
Project: Freiburg_example
 Options   : maxHits        : 1
             paired          : no
             splicedAlignment: FALSE
             bisulfite       : no
             snpFile         : none
 Aligner   : Rbowtie v1.4.0 (parameters: -m 1 --best --strata)
 Genome    : BSgenome.Hsapiens.UCSC.hg19 (BSgenome)

 Reads     : 4 files, 2 samples (fastq format):
   1. QuasR_rna_1_1.fastq.gz  WT  (phred33)
   2. QuasR_rna_1_2.fastq.gz  WT  (phred33)
   3. QuasR_rna_2_1.fastq.gz  MUT (phred33)
   4. QuasR_rna_2_2.fastq.gz  MUT (phred33)

 Genome alignments: directory: /group_data/example/bam
   1. QuasR_rna_1_1_4ba8447d4806.bam
   2. QuasR_rna_1_2_4ba819e654f5.bam
   3. QuasR_rna_2_1_4ec87bab31f6.bam
   4. QuasR_rna_2_2_4ec8765b0838.bam
```

qCount



qCount (version 1.0.3quasr)

qProject:

2: qProject of Freiburg_example

a qProject returned by the 'qAlign' tool

count alignments in:

known genes (genome annotation)

source of the query regions

query:

● transcriptDB hg19

regions based on known genes (see below for details)

report level:

gene (sum of exons)

level of quantification for known genes (see below for details)

orientation relative to query regions:

any

count alignments on the specified strand (see below for details)

collapse by sample:

☑

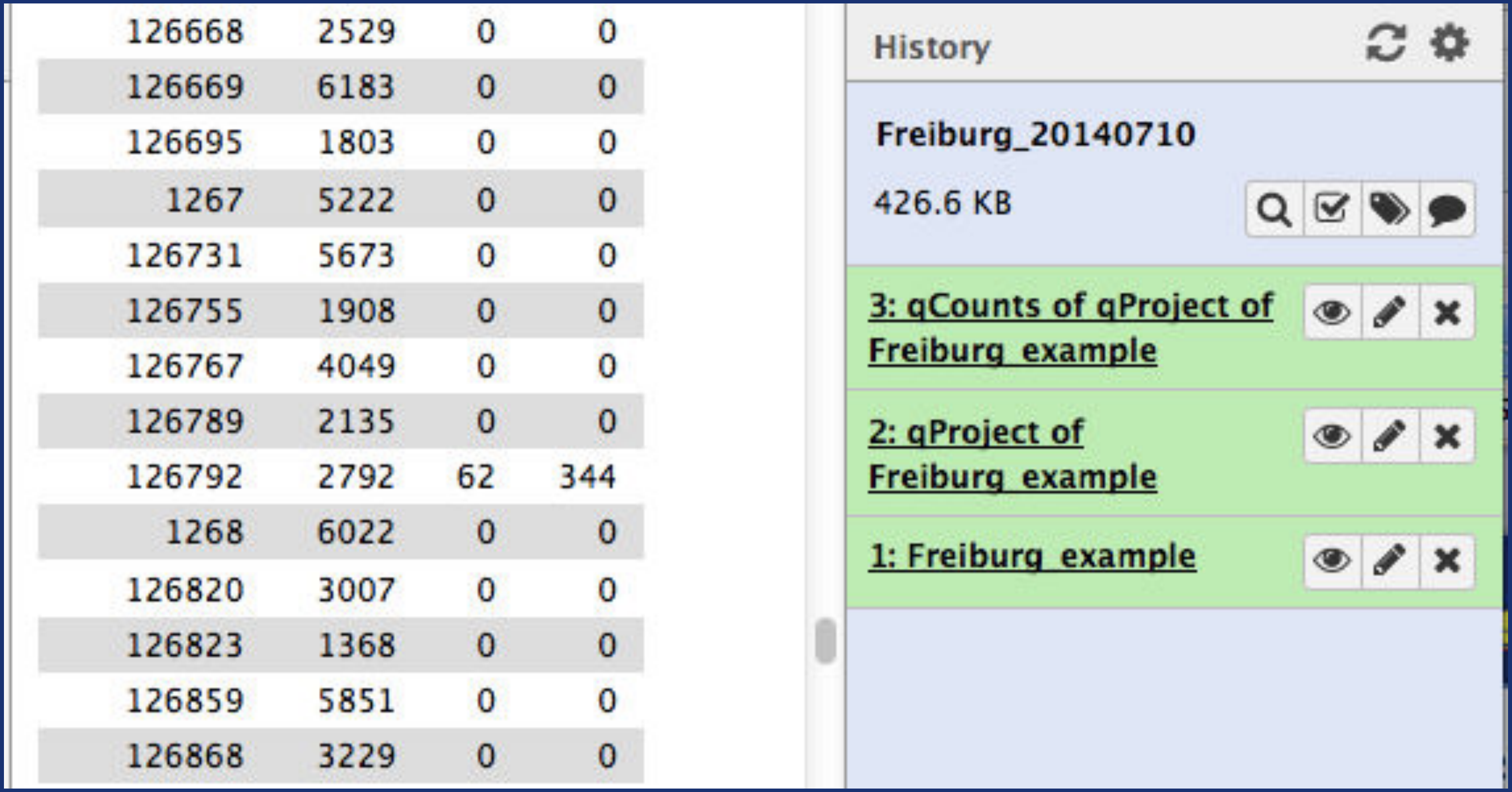
sum counts for files with identical sample names

show advanced settings:

□

use only if needed

qCount



126668	2529	0	0
126669	6183	0	0
126695	1803	0	0
1267	5222	0	0
126731	5673	0	0
126755	1908	0	0
126767	4049	0	0
126789	2135	0	0
126792	2792	62	344
1268	6022	0	0
126820	3007	0	0
126823	1368	0	0
126859	5851	0	0
126868	3229	0	0

History

Freiburg_20140710

426.6 KB

3: qCounts of qProject of Freiburg_example

2: qProject of Freiburg_example

1: Freiburg_example

redo on command line

samples.2.txt

```
FileName                SampleName
/group_data/example/raw/QuasR_rna_1_1.fastq.gz  sampleA
/group_data/example/raw/QuasR_rna_1_2.fastq.gz  sampleB
/group_data/example/raw/QuasR_rna_2_1.fastq.gz  sampleC
/group_data/example/raw/QuasR_rna_2_2.fastq.gz  sampleD
```

redo on command line

```
> project2 <- qAlign("samples.2.txt",
+               "BSgenome.Hsapiens.UCSC.hg19",
+               alignmentsDir="/group_data/example/bam/")
all necessary alignment files found
>
> alignments(project2)
$genome
                            FileName  SampleName
1 .../QuasR_rna_1_1_4ba8447d4806.bam  sampleA
2 .../QuasR_rna_1_2_4ba819e654f5.bam  sampleB
3 .../QuasR_rna_2_1_4ec87bab31f6.bam  sampleC
4 .../QuasR_rna_2_2_4ec8765b0838.bam  sampleD

$aux
data frame with 0 columns and 0 rows

>
```

Using Galaxy for NGS Analysis in a collaborative environment

the “average” lab scientist, using Galaxy

the “modern” lab scientist, using R/Bioconductor

The same raw data is used in both procedures and the alignments will not be generated twice.

FMI

Friedrich Miescher Institute
for Biomedical Research

Acknowledgment

Galaxy

Michael Stadler Christian Hundsrucker

Anita Lerch Tim Roloff Maria Florescu

Dimos Gaidatzis Stefan Grzybek Lukas Burger

and all the people from the "Galaxy Communtiy"

http://www.bioconductor.org/packages/release/bioc/html/QuasR.html