

次世代シーケンサーデータの解析手法  
第15回RNA-seq解析(その3)  
ウェブ資料

寺田朋子、清水謙多郎、門田幸二\*

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W1: スタート地点

The screenshot shows the Galaxy web interface. The main content area displays a workflow table with two columns: '1' and '2'. The table lists various tools and their outputs, including #KEY, EBG00001128470, LGG\_00001, and LGG\_00014. The right-hand side features a 'History' panel with a search bar and a list of datasets, including '54: Column Join on data 52, data 50, and data 48', '53: htseq-count on data 4 4 and data 47 (no feature)', '52: htseq-count on data 4 4 and data 47', '51: htseq-count on data 4 4 and data 46 (no feature)', and '50: htseq-count on data 4'. The left-hand side contains a 'Tools' panel with a search bar and various tool categories like 'Get Data', 'Send Data', 'Collection Operations', etc.

1	2
#KEY	htseq-count on data 44 and data 45_2
EBG00001128470	130
EBG00001128476	15
EBG00001128500	99
EBG00001128509	166
EBG00001128529	0
LGG_00001	195
LGG_00002	513
LGG_00003	21
LGG_00004	138
LGG_00005	239
LGG_00006	680
LGG_00007	6
LGG_00008	54
LGG_00009	11
LGG_00010	13
LGG_00011	1412
LGG_00012	1373
LGG_00013	703
LGG_00014	14

# W2: 新規ヒストリー

前回(第14回)のW9-7と同じ画面です。①View all histories。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and a 'Using 1%' indicator. A red arrow points to the 'View all histories' button in the History panel.

**Tools**

- Get Data
- Send Data
- Collection Operations
- Expression Tools
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Datamash
- GENOMIC FILE MANIPULATION
- FASTA/FASTQ
- FASTQ Quality Control
- SAM/RAM

1	2
#KEY	htseq-count on data 44 and data 45_2
EBG00001128470	130
EBG00001128476	15
EBG00001128500	99
EBG00001128509	166
EBG00001128529	0
LGG_00001	195
LGG_00002	513
LGG_00003	21
LGG_00004	138
LGG_00005	239
LGG_00006	680
LGG_00007	6
LGG_00008	54
LGG_00009	11
LGG_00010	13
LGG_00011	1412
LGG_00012	1373
LGG_00013	703
LGG_00014	14

**History**

search datasets

**GSE107337\_3samples**

54 shown

4.52 GB

- 54: Column Join on data 52, data 50, and data 48
- 53: htseq-count on data 4 4 and data 47 (no feature)
- 52: htseq-count on data 4 4 and data 47
- 51: htseq-count on data 4 4 and data 46 (no feature)
- 50: htseq-count on data 4

https://usegalaxy.org/history/view\_multiple

# W2: 新規ヒストリー

前回(第14回)のW9-7と同じ画面です。①View all histories。私の場合はこんな感じになります。

The screenshot shows the Galaxy web interface. The browser address bar is `usegalaxy.org/history/view_multiple`. The Galaxy logo and navigation menu are visible at the top. Below the navigation bar, there are search boxes for 'search histories' and 'search all datasets'. The main content area is divided into two panels. The left panel, titled '現在のヒストリー' (Current History), shows a list of datasets under the history 'GSE107337\_3samples'. The right panel, titled 'Unnamed history', is currently empty and shows a message 'ヒストリーは空です' (History is empty).

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

search histories search all datasets Create new

現在のヒストリー

Switch to

**GSE107337\_3samples**  
54 shown  
4.52 GB

search datasets

Drag datasets here to copy them to the current history

- 54: Column Join on data 52, data 50, and data 48
- 53: htseq-count on data 44 and data 47 (no feature)
- 52: htseq-count on data 44 and data 47
- 51: htseq-count on data 44 and data 46 (no feature)
- 50: htseq-count on data 44

**Unnamed history**  
(empty)

search datasets

ヒストリーは空です

# W2: 新規ヒストリー

前回(第14回)のW9-7と同じ画面です。①View all histories。私の場合はこんな感じになります。赤枠が新規ヒストリーを作成済みの状態と同じなので、この場合は②Switch toを押す。

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with the Galaxy logo and various menu items like 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. Below this is a search bar with 'search histories' and 'search all datasets' options. The main content area is divided into two panels. The left panel, titled '現在のヒストリー', shows a list of histories with details like 'GSE107337\_3samples' and '54 shown'. The right panel, titled 'Unnamed history', is currently empty and contains a message 'ヒストリーは空です'. A red arrow points to the 'Switch to' button between the two panels, which has a circled '2' next to it.

# W2: 新規ヒストリー

前回(第14回)のW9-7と同じ画面です。①View all histories。私の場合はこんな感じになります。赤枠が新規ヒストリーを作成済みの状態と同じなので、この場合は②Switch toを押す。こんな感じで③空のヒストリーが、④現在のヒストリーになっていればOK。

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

search histories search all datasets Create new

現在のヒストリー Switch to

Unnamed history

(empty)

search datasets

Drag datasets here to copy them to the current history

i ヒストリーは空です

GSE107337\_3samples

54 shown

4.52 GB

search datasets

54: Column Join on data 52, data 50, and data 48

53: htseq-count on data 44 and data 47 (no feature)

52: htseq-count on data 44 and data 47

51: htseq-count on data 44 and data 46 (no feature)

50: htseq-count on data 44 and data 46

49: htseq-count on data 44 and data 45 (no feature)

48: htseq-count on data 44 and data 45

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (RAM)

参考

もし①View all historiesをやった時にこんな感じになっていた場合は、②Create newでよい。

# W2: 新規ヒストリー

The screenshot shows the Galaxy web interface. The browser address bar displays 'usegalaxy.org/history/view\_multiple'. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets', and a 'Create new' button. A red arrow with the number '2' points to the 'Create new' button. The main content area shows a list of histories under the heading '現在のヒストリー'. The first history is 'GSE107337\_3samples' with 54 shown items and 4.52 GB of data. Below it, there are four more history entries, each with a description and icons for viewing, editing, and deleting.



# W2: 新規ヒストリー

①のところをtrans\_mapとして、ヒストリー名を変更。

The screenshot shows the Galaxy web interface. The browser address bar is `usegalaxy.org/history/view_multiple`. The Galaxy navigation bar includes options like 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets', and a 'Create new' button. The main content area is divided into two panels. The left panel, titled '現在のヒストリー' (Current History), shows an 'Unnamed history' with a red arrow pointing to it and a callout box that says 'ヒストリーの名前を変更するにはクリック' (Click to change the history name). Below this is a 'search datasets' bar and a message 'ヒストリーは空です' (History is empty). The right panel, titled 'Switch to GSE107337\_3samples', shows a list of datasets and jobs. The jobs are listed as follows:

Job ID	Description	View	Edit	Delete
54	Column Join on data 52, data 50, and data 48	👁	✎	✖
53	htseq-count on data 44 and data 47 (no feature)	👁	✎	✖
52	htseq-count on data 44 and data 47	👁	✎	✖
51	htseq-count on data 44 and data 46 (no feature)	👁	✎	✖
50	htseq-count on data 44 and data 46	👁	✎	✖
49	htseq-count on data 44 and data 45 (no feature)	👁	✎	✖
48	htseq-count on data 44 and data 45	👁	✎	✖
47	Bowtie2 on data 43, data 28, and data 27: aligned reads (RAM)	👁	✎	✖

# W2: 新規ヒストリー

①のところをtrans\_mapとして、ヒストリー名を変更。変更後。

The screenshot shows the Galaxy web interface. The browser address bar displays 'usegalaxy.org/history/view\_multiple'. The Galaxy navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets', along with a 'Create new' button. The main content area is divided into two panels. The left panel, titled '現在のヒストリー', shows a new history named 'trans\_map' which is currently empty. A message at the bottom of this panel states 'ヒストリーは空です'. The right panel, titled 'Switch to', shows a history named 'GSE107337\_3samples' containing 54 datasets. A list of datasets is displayed, including '54: Column Join on data 52, data 50, and data 48', '53: htseq-count on data 44 and data 47 (no feature)', '52: htseq-count on data 44 and data 47', '51: htseq-count on data 44 and data 46 (no feature)', '50: htseq-count on data 44 and data 46', '49: htseq-count on data 44 and data 45 (no feature)', and '48: htseq-count on data 44 and data 45'. Each dataset entry has icons for viewing, editing, and deleting.

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy | Histories', 'usegalaxy.org/history/view\_multiple', and various menu items like 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. Below the navigation bar, there are search bars for 'search histories' and 'search all datasets', and a 'Create new' button. The main content area is divided into two panels. The left panel, titled '現在のヒストリー' (Current History), shows a history named 'trans\_map' which is currently empty. A red arrow with the number '1' points to the 'trans\_map' header. The right panel, titled 'Switch to GSE107337\_3samples', shows a list of 54 datasets. A red arrow with the number '2' points to the '54: Column Join on data 52, data 50, and data 48' dataset. The datasets listed include '54: Column Join on data 52, data 50, and data 48', '53: htseq-count on data 44 and data 47 (no feature)', '52: htseq-count on data 44 and data 47', '51: htseq-count on data 44 and data 46 (no feature)', '50: htseq-count on data 44 and data 46', '49: htseq-count on data 44 and data 45 (no feature)', '48: htseq-count on data 44 and data 45', and '47: Bowtie2 on data 43, data 28, and data 27: aligned reads (RAM)'. Each dataset entry has icons for viewing, editing, and deleting.

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy | Histories', 'usegalaxy.org/history/view\_multiple', and 'Galaxy' with sub-menus for 'データ解析', 'ワークフロー', '可視化する', '共有データ', and 'ヘルプ'. Below the navigation bar are search boxes for 'search histories' and 'search all datasets', and a 'Create new' button. The main content area is divided into two panels. The left panel, titled '現在のヒストリー' (Current History), shows a 'trans\_map' history which is empty. A blue message box at the bottom of this panel says 'ヒストリーは空です' (History is empty). The right panel, titled 'GSE107337\_3samples', shows a list of datasets. The list includes: 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM); 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM); 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM); 44: ASM2650v1.gff3; 43: ASM2650v1.fa; 42: FastQC on data 28: RawData; 41: FastQC on data 28: Webpage; 40: FastQC on data 27: RawData. A red arrow labeled '2' points to the 'x' icon of dataset 46. A red arrow labeled '3' points to a red bracket grouping datasets 43 and 44.

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。例えば④.faファイルを赤矢印の先にドラッグアンドドロップすると...

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy | Histories', 'usegalaxy.org/history/view\_multiple', and 'Galaxy' with sub-menus for 'データ解析', 'ワークフロー', '可視化する', and '共有データ'. Below the navigation bar are search bars for 'search histories' and 'search all datasets', and a 'Create new' button. The main content area is divided into two panels. The left panel, titled '現在のヒストリー', shows the 'trans\_map' history, which is currently empty. A blue message box at the bottom of this panel says 'ヒストリーは空です'. The right panel, titled 'GSE107337\_3samples', shows a list of data items. Item 43, '43: ASM2650v1.fa', is highlighted with a red box and a red arrow pointing to it from the 'trans\_map' history area. Other items in the list include '47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)', '46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)', '45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)', '44: ASM2650v1.gff3', '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', and '40: FastQC on data 27: RawData'.

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。例えば④.faファイルを赤矢印の先にドラッグアンドドロップすると...こんな感じになります。

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy

データ解析 ワークフロー 可視化する 共有データ

search histories search all datasets

現在のヒストリー

trans\_map

(empty)

search datasets

Drag datasets here to copy them to the current history

43: ASM2650v1.fa

ヒストリーは空です

Switch to

GSE107337\_3samples

54 shown

4.52 GB

search datasets

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。例えば④.faファイルを赤矢印の先にドラッグアンドドロップすると...こんな感じになります。⑤コピー完了後の状態。

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy

データ解析 ワークフロー 可視化する 共有データ

search histories search all datasets

現在のヒストリー

Switch to

**trans\_map**  
1 shown  
2.92 MB

search datasets

1: ASM2650v1.fa

**GSE107337\_3samples**  
54 shown  
4.52 GB

search datasets

- 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)
- 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)
- 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)
- 44: ASM2650v1.gff3
- 43: ASM2650v1.fa
- 42: FastQC on data 28: RawData
- 41: FastQC on data 28: Webpage
- 40: FastQC on data 27: RawData



# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。例えば④.faファイルを赤矢印の先にドラッグアンドドロップすると...こんな感じになります。⑤コピー完了後の状態。⑥.gff3ファイルについても同様にコピーする。

The screenshot shows the Galaxy web interface with two history panes. The left pane, titled 'trans\_map', contains one dataset: '1: ASM2650v1.fa'. The right pane, titled 'GSE107337\_3samples', contains a list of datasets from 40 to 47. A red box highlights dataset '44: ASM2650v1.gff3' in the right pane, with a red arrow pointing to it from a red circle containing the number '6'. Another red arrow points from the '44: ASM2650v1.gff3' dataset in the right pane to a search bar in the left pane, which contains the text '44: ASM2650v1.gff3'. The search bar in the left pane is highlighted with a black box.

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy

データ解析 ワークフロー 可視化する 共有データ

search histories search all datasets

現在のヒストリー

Switch to

**trans\_map**  
1 shown  
2.92 MB

search datasets

Drag datasets here to copy them to the current history

44: ASM2650v1.gff3

1: ASM2650v1.fa

**GSE107337\_3samples**  
54 shown  
4.52 GB

search datasets

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

# W3: データのコピー

①trans\_mapで解析を進めていくので、ここに必要なデータをコピーする。まずは、トランスクリプトーム配列の作成に必要なゲノム配列ファイル(ASM2650v1.fa)とアノテーションファイル(ASM26501v1.gff3)から。②下部に移動して探す。③発見。例えば④.faファイルを赤矢印の先にドラッグアンドドロップすると...こんな感じになります。⑤コピー完了後の状態。⑥.gff3ファイルについても同様にコピーする。コピー完了後の状態。

Galaxy | Histories

usegalaxy.org/history/view\_multiple

Galaxy

データ解析 ワークフロー 可視化する 共有データ

search histories search all datasets

現在のヒストリー

Switch to

**trans\_map**  
2 shown  
4.57 MB

search datasets

2: ASM2650v1.gff3

1: ASM2650v1.fa

**GSE107337\_3samples**  
54 shown  
4.52 GB

search datasets

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

# W3: データのコピー

次は、マップする側のデータをコピー。①のあたりまで下部に移動して、Trimmomatic実行後のリードデータを利用する。まずは、この画面上で見えている②と③をドラッグアンドドロップでコピー。

The screenshot shows the Galaxy web interface with two history panels. The left panel, titled "trans\_map", shows two datasets: "2: ASM2650v1.gff3" and "1: ASM2650v1.fa". The right panel, titled "GSE107337\_3samples", shows a list of Trimmomatic processing jobs. Red arrows and numbers 1, 2, and 3 point to specific elements: arrow 1 points to the bottom of the job list, arrow 2 points to a bracket grouping jobs 19 and 20, and arrow 3 points to a bracket grouping jobs 23 and 24.

Job ID	Description	View	Edit	Delete
25	Trimmomatic on SRR6322567_1 (R1 unpaired)	👁	✎	✖
24	Trimmomatic on SRR6322567_2 (R2 paired)	👁	✎	✖
23	Trimmomatic on SRR6322567_1 (R1 paired)	👁	✎	✖
22	Trimmomatic on SRR6322564_2 (R2 unpaired)	👁	✎	✖
21	Trimmomatic on SRR6322564_1 (R1 unpaired)	👁	✎	✖
20	Trimmomatic on SRR6322564_2 (R2 paired)	👁	✎	✖
19	Trimmomatic on SRR6322564_1 (R1 paired)	👁	✎	✖

# W3: データのコピー

次は、マップする側のデータをコピー。①のあたりまで下部に移動して、Trimmomatic実行後のリードデータを利用する。まずは、この画面上で見えている②と③をドラッグアンドドロップでコピー。コピー完了後の状態。

The screenshot shows the Galaxy web interface with two history panels. The left panel, titled 'trans\_map', contains six items. The right panel, titled 'GSE107337\_3samples', contains seven items. Red boxes and arrows highlight specific items in both panels, with red lightning bolts and circled numbers 2 and 3 indicating the items to be copied.

History Panel	Item ID	Description	Copy Status
trans_map	6	Trimmomatic on SRR632 2567_2 (R2 paired)	Highlighted (Red box)
	5	Trimmomatic on SRR632 2567_1 (R1 paired)	Highlighted (Red box)
	4	Trimmomatic on SRR632 2564_2 (R2 paired)	Highlighted (Red box)
	3	Trimmomatic on SRR632 2564_1 (R1 paired)	Highlighted (Red box)
	2	ASM2650v1.gff3	Highlighted (Red box)
	1	ASM2650v1.fa	Highlighted (Red box)
GSE107337_3samples	25	Trimmomatic on SRR6322567_1 (R1 unpaired)	Not highlighted
	24	Trimmomatic on SRR6322567_2 (R2 paired)	Highlighted (Red box)
	23	Trimmomatic on SRR6322567_1 (R1 paired)	Highlighted (Red box)
	22	Trimmomatic on SRR6322564_2 (R2 unpaired)	Not highlighted
	21	Trimmomatic on SRR6322564_1 (R1 unpaired)	Not highlighted
	20	Trimmomatic on SRR6322564_2 (R2 paired)	Highlighted (Red box)
	19	Trimmomatic on SRR6322564_1 (R1 paired)	Highlighted (Red box)

# W3: データのコピー

次は、マップする側のデータをコピー。①のあたりまで下部に移動して、Trimmomatic実行後のリードデータを利用する。まずは、この画面上で見えている②と③をドラッグアンドドロップでコピー。コピー完了後の状態。④少し上部に移動して、⑤最後のサンプルデータのコピー。

The screenshot shows the Galaxy web interface with two history panels. The left panel, titled "trans\_map", contains 6 datasets. The right panel, titled "GSE107337\_3samples", contains 10 datasets. A red arrow labeled "4" points to the right panel, and a red arrow labeled "5" points to a group of datasets in the right panel.

History Panel	Dataset ID	Description	Size	
trans_map	6	Trimmomatic on SRR632 2567_2 (R2 paired)	760.36 MB	
	5	Trimmomatic on SRR632 2567_1 (R1 paired)		
	4	Trimmomatic on SRR632 2564_2 (R2 paired)		
	3	Trimmomatic on SRR632 2564_1 (R1 paired)		
	2	ASM2650v1.gff3		
	1	ASM2650v1.fa		
GSE107337_3samples	31	FastQC on data 19: Webpage	4.52 GB	
	30	Trimmomatic on SRR6322569_2 (R2 unpaired)		
	29	Trimmomatic on SRR6322569_1 (R1 unpaired)		
	28	Trimmomatic on SRR6322569_2 (R2 paired)		
	27	Trimmomatic on SRR6322569_1 (R1 paired)		
	26	Trimmomatic on SRR6322567_2 (R2 unpaired)		
	25	Trimmomatic on SRR6322567_1 (R1 unpaired)		

# W3: データのコピー

次は、マップする側のデータをコピー。①のあたりまで下部に移動して、Trimmomatic実行後のリードデータを利用する。まずは、この画面上で見えている②と③をドラッグアンドドロップでコピー。コピー完了後の状態。④少し上部に移動して、⑤最後のサンプルデータのコピー。コピー完了後の状態。

The screenshot shows the Galaxy web interface with two history panels. The left panel, titled 'trans\_map', contains a list of Trimmomatic jobs. Jobs 8 and 7 are highlighted with a red box. The right panel, titled 'GSE107337\_3samples', contains a list of Trimmomatic jobs. Jobs 27 and 28 are grouped with a red bracket and a red arrow labeled '5' pointing to them.

Job ID	Description	View	Edit	Delete
8	Trimmomatic on SRR632 2569_2 (R2 paired)	👁	✎	✖
7	Trimmomatic on SRR632 2569_1 (R1 paired)	👁	✎	✖
6	Trimmomatic on SRR632 2567_2 (R2 paired)	👁	✎	✖
5	Trimmomatic on SRR632 2567_1 (R1 paired)	👁	✎	✖
4	Trimmomatic on SRR632 2564_2 (R2 paired)	👁	✎	✖
31	FastQC on data 19: Webpage	👁	✎	✖
30	Trimmomatic on SRR6322569_2 (R2 unpaired)	👁	✎	✖
29	Trimmomatic on SRR6322569_1 (R1 unpaired)	👁	✎	✖
28	Trimmomatic on SRR6322569_2 (R2 paired)	👁	✎	✖
27	Trimmomatic on SRR6322569_1 (R1 paired)	👁	✎	✖
26	Trimmomatic on SRR6322567_2 (R2 unpaired)	👁	✎	✖
25	Trimmomatic on SRR6322567_1 (R1 unpaired)	👁	✎	✖

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W4: 解析準備完了

①データ解析のところを押して、いつもの3画面の状態に戻す。

The screenshot shows the Galaxy web interface. The browser address bar contains 'usegalaxy.org/history/view\_multiple', with a red circle and the number '1' highlighting the 'view\_multiple' part. The interface includes a search bar, a 'Switch to' dropdown, and two panels of history items. The left panel, titled 'trans\_map', shows 8 items (1.29 GB). The right panel, titled 'GSE107337\_3samples', shows 54 items (4.52 GB). Both panels have a 'search datasets' input field and icons for viewing, editing, and deleting items.

History Item	View	Edit	Delete
8: Trimmomatic on SRR632 2569_2 (R2 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7: Trimmomatic on SRR632 2569_1 (R1 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6: Trimmomatic on SRR632 2567_2 (R2 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5: Trimmomatic on SRR632 2567_1 (R1 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4: Trimmomatic on SRR632 2564_2 (R2 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31: FastQC on data 19: Webpage	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30: Trimmomatic on SRR6322569_2 (R2 unpaired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29: Trimmomatic on SRR6322569_1 (R1 unpaired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28: Trimmomatic on SRR6322569_2 (R2 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27: Trimmomatic on SRR6322569_1 (R1 paired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26: Trimmomatic on SRR6322567_2 (R2 unpaired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25: Trimmomatic on SRR6322567_1 (R1 unpaired)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



# W4: 解析準備完了

①データ解析のところを押して、いつもの3画面の状態に戻す。解析準備完了状態。②ヒストリーパネルで下部に移動。

The screenshot shows the Galaxy web interface. The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a "Try Galaxy on the Cloud" banner with the text "Now you can have a personal Galaxy within the infinite Universe".

The left sidebar contains a "Tools" section with a search bar and various tool categories: "Get Data", "Send Data", "Collection Operations", "Expression Tools", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", and "SAM/RAM".

The right sidebar shows the "History" panel with a search bar and a list of datasets. The datasets are:

- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)

A red arrow with the number "2" points to the "6: Trimmomatic on SRR 6322567\_2 (R2 paired)" entry in the history panel.

# W4: 解析準備完了

①データ解析のところを押して、いつもの3画面の状態に戻す。解析準備完了状態。②ヒストリーパネルで下部に移動後。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and a 'Using 1%' indicator. The left sidebar lists various tool categories such as 'Tools', 'Get Data', 'Send Data', 'Collection Operations', 'Expression Tools', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', and 'SAM/RAM'. The main content area displays a welcome message for Galaxy and a 'Galaxy Help' banner with the text 'Got Questions? Get Answers.' and the URL 'help.galaxyproject.org'. The right sidebar shows the 'History' panel with a search bar and a list of datasets. The list includes items 1 through 6, with item 3, 'Trimmomatic on SRR 6322564\_1 (R1 paired)', highlighted in green and marked with a red arrow and the number '2'. The bottom of the page shows a 'Tweets by @galaxyproject' section.

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar lists various tool categories: 'Tools', 'Get Data', 'Send Data', 'Collection Operations', 'Expression Tools', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', and 'SAM/RAM'. The main content area displays a workflow history titled 'trans\_map' with 8 shown items. The items are listed as follows:

- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)
- 3: Trimmomatic on SRR 6322564\_1 (R1 paired)
- 2: ASM2650v1.gff3
- 1: ASM2650v1.fa

Red arrows with numbers 1 and 2 point to the workflow items. Arrow 1 points to item 2, and arrow 2 points to item 3. A tooltip 'データを表示' is visible near item 1. In the center of the main panel, there is a black banner with the text: 'Galaxy Help Got Questions? Get Answers. help.galaxyproject.org'. Below the banner, there are 'Tweets by @galaxyproject' and a URL: 'https://usegalaxy.org/datasets/bbd44e69cb8906b565183755561ae85a/display/?preview=True'.

# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示後。

The screenshot shows the Galaxy web interface. The central panel displays a table of genomic features with the following columns: Seqid, Source, Type, and Size. The table contains several rows of data, including gene, mRNA, exon, and CDS features for sequence FM179322. A red arrow labeled '1' points to the 'CDS' row for FM179322. The right panel shows the job history, with a job titled '2: ASM2650v1.gff3' highlighted in green. A red arrow labeled '2' points to this job. The job history also shows other Trimmomatic jobs on SRR data.

Seqid	Source	Type	Size
##gff-version 3			
##sequence-region FM179322 1 3010111			
#!genome-build European Nucleotide Archive ASM2650v1			
#!genome-version ASM2650v1			
#!genome-date 2015-02			
#!genome-build-accession GCA_000026505.1			
#!genebuild-last-updated 2015-02			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	European Nucleotide Archive	supercontig	
###			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	ena	gene	

# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示後。③アノテーションファイル中には様々なfeaturesが含まれているので、gene領域のみの塩基配列情報を抽出するための前処理を行う。

The screenshot shows the Galaxy web interface with a GFF file being processed. The central panel displays a table of features with columns for Seqid, Source, Type, and Size. The right sidebar shows a history panel with a list of jobs. Red arrows and boxes highlight specific elements:

- Arrow 1 points to the 'CDS' feature in the table.
- Arrow 2 points to the '2: ASM2650v1.gff3' entry in the history panel.
- Arrow 3 points to the 'gene' feature in the table.

Seqid	Source	Type	Size
##gff-version 3			
##sequence-region FM179322 1 3010111			
#!genome-build European Nucleotide Archive ASM2650v1			
#!genome-version ASM2650v1			
#!genome-date 2015-02			
#!genome-build-accession GCA_000026505.1			
#!genebuild-last-updated 2015-02			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	European Nucleotide Archive	supercontig	
###			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	ena	gene	

History panel:

- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)
- 3: Trimmomatic on SRR 6322564\_1 (R1 paired)
- 2: ASM2650v1.gff3
- 1: ASM2650v1.fa

# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示後。③アノテーションファイル中には様々なfeaturesが含まれているので、gene領域のみの塩基配列情報を抽出するための前処理を行う。④Filter and Sort。

The screenshot shows the Galaxy web interface. The left sidebar contains a list of tool categories, with 'Filter and Sort' highlighted by a red arrow and the number '4'. The main panel displays a table of genomic features for FM179322. The right sidebar shows the 'History' panel with a list of jobs, including Trimmomatic and ASM2650v1.gff3.

Seqid	Source	Type	Score
##gff-version 3			
##sequence-region FM179322 1 3010111			
#!genome-build European Nucleotide Archive ASM2650v1			
#!genome-version ASM2650v1			
#!genome-date 2015-02			
#!genome-build-accession GCA_000026505.1			
#!genebuild-last-updated 2015-02			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	European Nucleotide Archive	supercontig	
###			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	ena	gene	

# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示後。③アノテーションファイル中には様々なfeaturesが含まれているので、gene領域のみの塩基配列情報を抽出するための前処理を行う。④Filter and Sort。こんな感じになります。④が一番上に来るくらいまで、⑤ツール選択パネルを下部に移動。

The screenshot shows the Galaxy web interface with a table of GFF data and a list of tools. A red arrow labeled '4' points to the 'Filter and Sort' tool, and another red arrow labeled '5' points to the 'Extract features from GFF data' tool.

Seqid	Source	Type	S
##gff-version 3			
##sequence-region FM179322 1 3010111			
#!genome-build European Nucleotide Archive ASM2650v1			
#!genome-version ASM2650v1			
#!genome-date 2015-02			
#!genome-build-accession GCA_000026505.1			
#!genebuild-last-updated 2015-02			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	European Nucleotide Archive	supercontig	
###			
FM179322	ena	gene	
FM179322	ena	mRNA	
FM179322	ena	exon	
FM179322	ena	CDS	
###			
FM179322	ena	gene	

Tools list:

- Get Data
- Send Data
- Collection Operations
- Expression Tools
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort (4)
- GFF
- Extract features from GFF data (5)

History:

- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)
- 3: Trimmomatic on SRR 6322564\_1 (R1 paired)
- 2: ASM2650v1.gff3
- 1: ASM2650v1.fa



# W5: GFFの前処理

①アノテーションファイルの中身を、②中央パネル上に表示後。③アノテーションファイル中には様々なfeaturesが含まれているので、gene領域のみの塩基配列情報を抽出するための前処理を行う。④Filter and Sort。こんな感じになります。④が一番上に来るくらいまで、⑤ツール選択パネルを下部に移動。こんな感じ。

The screenshot shows the Galaxy web interface. The 'Tools' panel on the left is expanded to show 'Filter and Sort' (marked with a red arrow and the number 4) and 'GFF' (marked with a red arrow and the number 5). The 'History' panel on the right shows a list of jobs, with the top job highlighted in green.

Seqid	Source	Type
##gff-version 3		
##sequence-region FM179322 1 3010111		
#!genome-build European Nucleotide Archive ASM2650v1		
#!genome-version ASM2650v1		
#!genome-date 2015-02		
#!genome-build-accession GCA_000026505.1		
#!genebuild-last-updated 2015-02		
FM179322	ena	gene
FM179322	ena	mRNA
FM179322	ena	exon
FM179322	ena	CDS
###		
FM179322	European Nucleotide Archive	supercontig
###		
FM179322	ena	gene
FM179322	ena	mRNA
FM179322	ena	exon
FM179322	ena	CDS
###		
FM179322	ena	gene

History:

- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)
- 3: Trimmomatic on SRR 6322564\_1 (R1 paired)
- 2: ASM2650v1.gff3
- 1: ASM2650v1.fa

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- **W6: Extract features**
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W6-1 : Extract feature

①アノテーションファイルの中身を、②中央パネル上に表示後。③アノテーションファイル中には様々なfeaturesが含まれているので、gene領域のみの塩基配列情報を抽出するための前処理を行う。④Filter and Sort。こんな感じになります。④が一番上に来るくらいまで、⑤ツール選択パネルを下部に移動。こんな感じ。⑥ **Extract features**が、⑦GFFファイル中の、目的のfeature(この場合はgeneという)情報のみ抽出するときに利用するプログラム。⑥をクリック。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy' and 'usegalaxy.org'. The main content area is divided into several panels:

- Filter and Sort:** This panel is active and shows options for filtering and sorting data. A red arrow points to the 'Filter and Sort' panel.
- Tools:** This panel is visible below the 'Filter and Sort' panel. It contains a search bar and a list of tools. The 'Extract features' tool is highlighted with a red arrow and a circled '6'.
- Table:** A table of genomic features is displayed in the center. The table has columns for 'Seqid', 'Source', and 'Type'. The data rows are as follows:

Seqid	Source	Type
##gff-version 3		
##sequence-region FM179322 1 3010111		
#!genome-build European Nucleotide Archive A		
#!genome-version ASM2650v1		
#!genome-date 2015-02		
#!genome-build-accession GCA_000026505.1		
#!genebuild-last-updated 2015-02		
FM179322	ena	gene
FM179322	ena	mRNA
FM179322	ena	exon
FM179322	ena	CDS
###		
FM179322	European Nucleotide Archive	supercontig
###		
FM179322	ena	gene
FM179322	ena	mRNA
FM179322	ena	exon
FM179322	ena	CDS
###		
FM179322	ena	gene
- Tools List:** A list of tools is shown on the right, including 'Trimmomatic on SRR 6322567\_2 (R2 paired)', 'Trimmomatic on SRR 6322567\_1 (R1 paired)', 'Trimmomatic on SRR 6322564\_2 (R2 paired)', 'Trimmomatic on SRR 6322564\_1 (R1 paired)', 'ASM2650v1.gff3', and 'ASM2650v1.fa'. A red arrow points to the 'Extract features' tool in the 'Tools' panel.

# W6-2: Extract features

The screenshot displays the Galaxy web interface for the 'Extract features' tool. The browser address bar shows 'usegalaxy.org'. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and a 'Using 1%' indicator.

**Tools Panel:** A search bar for tools is present. Under 'Filter and Sort', options include filtering data on any column using simple expressions, sorting data in ascending or descending order, and selecting lines that match an expression. Under 'GFF', the 'Extract features from GFF data' tool is highlighted. Other options include filtering GFF data by attribute, feature count, or attribute values\_list. Under 'Join, Subtract and Group', 'Datamash' is listed.

**Tool Configuration:** The 'Extract features from GFF data (Galaxy Version 1.0.0)' tool is selected. It has a 'Favorite' button and an 'Options' dropdown. The 'Select GFF data' section shows a file selection icon, a folder icon, and a dropdown menu with '2: ASM2650v1.gff3'. The 'From' section has a dropdown menu set to 'Column 1 / Sequence name'. The 'Extract features' section has a 'Select/Unselect all' checkbox and a text input field. Below this is a multi-select list instruction: 'Multi-select list - hold the appropriate key while clicking to select multiple columns'. An 'Execute' button is at the bottom.

**History Panel:** A search bar for datasets is present. The history shows a dataset named 'trans\_map' with a size of 1.29 GB. The history list includes:

- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)
- 4: Trimmomatic on SRR 6322564\_2 (R2 paired)
- 3: Trimmomatic on SRR 6322564\_1 (R1 paired)
- 2: ASM2650v1.gff3
- 1: ASM2650v1.fa

The bottom of the browser shows the URL: [https://usegalaxy.org/tool\\_runner?tool\\_id=Extract\\_features1](https://usegalaxy.org/tool_runner?tool_id=Extract_features1)

# W6-3: Extract feature

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。

The screenshot shows the Galaxy web interface for the 'Extract features from GFF data' tool. The tool title is 'Extract features from GFF data (Galaxy Version 1.0.0)'. The 'Select GFF data' field contains '2: ASM2650v1.gff3'. The 'From' dropdown is set to 'Column 1 / Sequence name'. The 'Extract features' section has a 'Select/Unselect all' checkbox. The 'Execute' button is visible. The right panel shows a history list with items 1 through 6, including 'ASM2650v1.gff3' and 'Trimmomatic on SRR' entries.

# W6-4: Extract feature

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。③Fromのところで、どの列でフィルタリングを行うかを指定します。

Galaxy

usegalaxy.org

Galaxy

データ解析 ワークフロー 可視化する 共有データ

Tools

search tools

Filter and Sort

Filter data on any column using simple expressions

Sort data in ascending or descending order

Select lines that match an expression

GFF

Extract features from GFF data

Filter GFF data by attribute using simple expressions

Filter GFF data by feature count using simple expressions

Filter GTF data by attribute values\_list

Join, Subtract and Group

Datamash

Extract features from GFF data (Galaxy Version 1.0.0)

Select GFF data

2: ASM2650v1.gff3

From

Column 1 / Sequence name

Extract features

Select/Unselect all

Multi-select list - hold the appropriate key while clicking to select multiple columns

Execute

What it does

is tool extracts selected features from GFF data

search datasets

trans\_map

8 shown

1.29 GB

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

5: Trimmomatic on SRR 6322567\_1 (R1 paired)

4: Trimmomatic on SRR 6322564\_2 (R2 paired)

3: Trimmomatic on SRR 6322564\_1 (R1 paired)

2: ASM2650v1.gff3

1: ASM2650v1.fa

https://usegalaxy.org/tool\_runner?tool\_id=Extract\_features1

# W6-5: Extract feature

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。③Fromのところ、どの列でフィルタリングを行うかを指定します。④今は3列目 (Column 3) でフィルタリングを行いたいのので、⑤Column 3 / Featureを選択する。

Galaxy

usegalaxy.org

Galaxy

データ解析 ワークフロー 可視化する 共有データ

Tools

search tools

Filter and Sort

Filter data on any column using simple expressions

Sort data in ascending or descending order

Select lines that match an expression

GFF

Extract features from GFF data

Filter GFF data by attribute using simple expressions

Filter GFF data by feature count using simple expressions

Filter GTF data by attribute values\_list

Join, Subtract and Group

Datamash

Extract features from GFF data (Galaxy Version 1.0.0)

Select GFF data

2: ASM2650v1.gff3

From

Column 1 / Sequence name

Column 2 / Source

Column 3 / Feature

Column 7 / Strand

Column 8 / Frame

Execute

What it does

This tool extracts selected features from GFF data

8 shown

1.29 GB

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

5: Trimmomatic on SRR 6322567\_1 (R1 paired)

4: Trimmomatic on SRR 6322564\_2 (R2 paired)

3: Trimmomatic on SRR 6322564\_1 (R1 paired)

2: ASM2650v1.gff3

1: ASM2650v1.fa

# W6-6: Extract feature

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。③Fromのところ、どの列でフィルタリングを行うかを指定します。④今は3列目 (Column 3) でフィルタリングを行いたいのので、⑤Column 3 / Featureを選択する。⑥変更完了。次は、どのfeatureを抽出するかを指定します。⑦をクリック。

The screenshot shows the Galaxy web interface for the 'Extract features' tool. The 'From' dropdown is set to 'Column 3 / Feature' (marked with a red arrow and '6'). The 'Extract features' section has a multi-select list (marked with a red arrow and '7') for selecting columns. The right panel shows a list of GFF files, with '6: Trimmomatic on SRR 6322567\_2 (R2 paired)' selected.

File Name	View	Edit	Delete
6: Trimmomatic on SRR 6322567_2 (R2 paired)	<input checked="" type="checkbox"/>		
5: Trimmomatic on SRR 6322567_1 (R1 paired)	<input type="checkbox"/>		
4: Trimmomatic on SRR 6322564_2 (R2 paired)	<input type="checkbox"/>		
3: Trimmomatic on SRR 6322564_1 (R1 paired)	<input type="checkbox"/>		
2: ASM2650v1.gff3	<input type="checkbox"/>		
1: ASM2650v1.fa	<input type="checkbox"/>		



# W6-7: Extract feature

The screenshot shows the Galaxy web interface. The main tool panel is titled 'Extract features from GFF data (Galaxy Version 1.0.0)'. A dropdown menu for 'Select' is open, showing a list of features: gene, mRNA, exon, CDS, supercontig, rRNA\_gene, and rRNA. The 'gene' option is highlighted in blue, and a red arrow with a circled '8' points to it. Below the dropdown is a text input field and a 'Multi-select list - hold the appropriate key while clicking to select multiple columns' instruction. At the bottom of the tool panel is a blue 'Execute' button. To the right of the tool panel, a list of data items is visible, including '5: Trimmomatic on SRR 6322567\_1 (R1 paired)', '4: Trimmomatic on SRR 6322564\_2 (R2 paired)', '3: Trimmomatic on SRR 6322564\_1 (R1 paired)', '2: ASM2650v1.gff3', and '1: ASM2650v1.fa'. Each item has icons for visibility, edit, and delete.

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。③Fromのところ、どの列でフィルタリングを行うかを指定します。④今は3列目 (Column 3) でフィルタリングを行いたいので、⑤Column 3 / Featureを選択する。⑥変更完了。次は、どのfeatureを抽出するかを指定します。⑦をクリック。こんな感じになるので、目的の⑧geneを指定。第14回のW8-4と同じようなことをやっていると同理解すればよい。

# W6-8: Extract feature

こんな感じになります。①Select GFF dataのところでは、②が見えています。これは、右側のヒストリーパネル上にある利用可能なGFFファイルがこれしかないからです。もし複数の候補GFFファイルがある場合は、②のところが対象とするファイルになっているかを確認しておく必要があります。③Fromのところ、どの列でフィルタリングを行うかを指定します。④今は3列目 (Column 3) でフィルタリングを行いたいのので、⑤Column 3 / Featureを選択する。⑥変更完了。次は、どのfeatureを抽出するかを指定します。⑦をクリック。こんな感じになるので、目的の⑧geneを指定。第14回のW8-4と同じようなことをやっているかと理解すればよい。⑨実行。

The screenshot shows the Galaxy web interface for the 'Extract features' tool. The 'From' dropdown is set to 'Column 3 / Feature'. The 'Extract features' section has a search box containing 'gene'. A red arrow with the number 9 points to the 'Execute' button. The right panel shows a list of files with their corresponding features.

File	Feature	View	Edit	Delete
6322567_1 (R1 paired)				
4: Trimmomatic on SRR				
6322564_2 (R2 paired)				
3: Trimmomatic on SRR				
6322564_1 (R1 paired)				
2: ASM2650v1.gff3				
1: ASM2650v1.fa				

# W6-9: Extract features

The screenshot displays the Galaxy web interface. At the top, a yellow banner indicates '実行中...' (Running...). The main content area shows a green notification box with a checkmark, stating: 'Executed **Extract features** and successfully added 1 job to the queue. The tool uses this input: 

- 2: ASM2650v1.gff3

 It produces this output: 

- 9: **Extract features on data 2**

 You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

The History panel on the right shows a list of jobs:

- 9: **Extract features on data 2**
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)

The left sidebar contains the 'Tools' section with a search bar and various tool categories: Filter and Sort, GFF, and Join, Subtract and Group. The 'Extract features' tool is listed under the GFF category.

# W6-10: Extract features

The screenshot displays the Galaxy web interface. At the top, a yellow banner indicates '実行完了。' (Execution completed). The main content area shows a green notification box with a checkmark, stating: 'Executed **Extract features** and successfully added 1 job to the queue. The tool uses this input: 

- 2: ASM2650v1.gff3

 It produces this output: 

- 9: **Extract features on data 2**

 You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

 The right-hand 'History' panel shows a list of jobs under the heading 'trans\_map' (9 shown, 1.29 GB). The jobs listed are: 

- 9: **Extract features on data 2**
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)

 The left-hand 'Tools' panel is visible, showing search and filter options.

# W6-11: Extract features

実行完了。①出力結果の中身を中央パネル上で表示。

The screenshot shows the Galaxy web interface. The central panel displays the output of the 'Extract features' tool, which is a GFF file. The output is a table with columns: Seqid, Source, Type, Start, End, and Score. The data rows show gene features for the sequence region FM179322.1. A red arrow with the number 1 points to the '9: Extract features on data 2' entry in the History panel on the right.

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

# W6-12: Extract feature

実行完了。①出力結果の中身を中央パネル上で表示。確かに、②3列目のfeatureがgeneばかりになっていて妥当ですね。

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools search tools

Filter and Sort

Filter data on any column using simple expressions

Sort data in ascending or descending order

Select lines that match an expression

GFF

Extract features from GFF data

Filter GFF data by attribute using simple expressions

Filter GFF data by feature count using simple expressions

Filter GTF data by attribute values\_list

Join, Subtract and Group

Datamash

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

History search datasets

trans\_map

9 shown

1.29 GB

9: Extract features on data 2

8: Trimmomatic on SRR 6322569\_2 (R2 paired)

7: Trimmomatic on SRR 6322569\_1 (R1 paired)

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

5: Trimmomatic on SRR 6322567\_1 (R1 paired)

https://usegalaxy.org/datasets/bbd44e69cb8906b58eb078...

# W6-13: Extract feature

実行完了。①出力結果の中身を中央パネル上で表示。確かに、②3列目のfeatureがgeneばかりになっていて妥当ですね。③を押して、Filter and Sortカテゴリを閉じる。

The screenshot shows the Galaxy web interface. The 'Filter and Sort' panel is highlighted with a red arrow and the number 3. The central panel displays a table of GFF data with columns: Seqid, Source, Type, Start, End, Score. The right panel shows a history of jobs, including 'Extract features on data 2'.

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

# W6-14: Extract feature

実行完了。①出力結果の中身を中央パネル上で表示。確かに、②3列目のfeatureがgeneばかりになっていて妥当ですね。③を押して、Filter and Sortカテゴリを閉じる。閉じた後。

The screenshot shows the Galaxy web interface. The central panel displays a table of genomic features with the following columns: Seqid, Source, Type, Start, End, and Score. The table contains several header lines (###) and data rows for gene features from the European Nucleotide Archive (ENA).

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

The right-hand History panel shows a list of recent jobs, with the most recent one highlighted in green: "9: Extract features on data 2". Other jobs include "8: Trimmomatic on SRR 6322569\_2 (R2 paired)", "7: Trimmomatic on SRR 6322569\_1 (R1 paired)", "6: Trimmomatic on SRR 6322567\_2 (R2 paired)", and "5: Trimmomatic on SRR 6322567\_1 (R1 paired)".



# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W7-1 : bedtools GetFastaBed

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains tool categories: 'Tools', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', 'Nanopore', 'Convert Formats', 'Lift-Over', and 'COMMON GENOMICS TOOLS'. A red arrow points to the 'BED' category, which has a '1' in a red circle next to it.

The main content area displays a table of genomic features:

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

The right sidebar shows the 'History' section with a search bar and a list of workflow steps:

- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)

# W7-2: bedtools GetFasta

①次はBED。②bedtools GetFastaBedプログラムを利用して、トランスクリプトーム配列取得を行います。②をクリック。

The screenshot shows the Galaxy web interface. The 'Tools' panel on the left lists various bedtools. A red arrow with the number '1' points to 'bedtools ExpandBed', and another red arrow with the number '2' points to 'bedtools GetFastaBed'. The central panel displays a table of genomic data with columns: Seqid, Source, Type, Start, End, and Score. The table contains several header lines (starting with ##) and data rows for gene intervals. The right panel shows the 'History' section with a search bar and a list of recent jobs, including '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', '6: Trimmomatic on SRR 6322567\_2 (R2 paired)', and '5: Trimmomatic on SRR 6322567\_1 (R1 paired)'.

Seqid	Source	Type	Start	End	Score
##gff-version 3					
##sequence-region FM179322 1 3010111					
#!genome-build European Nucleotide Archive ASM2650v1					
#!genome-version ASM2650v1					
#!genome-date 2015-02					
#!genome-build-accession GCA_000026505.1					
#!genebuild-last-updated 2015-02					
FM179322	ena	gene	1	1350	.
###					
###					
FM179322	ena	gene	1524	2663	.
###					
FM179322	ena	gene	3157	3369	.
###					
FM179322	ena	gene	3366	4484	.
###					
FM179322	ena	gene	4516	6477	.
###					
FM179322	ena	gene	6540	9152	.
###					

# W7-3: bedtools GetFasta

①次はBED。②bedtools GetFastaBedプログラムを利用して、トランスクリプトーム配列取得を行います。②をクリック。中央パネル上に操作画面が表示されます。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and a list of tools including 'bedtools ExpandBed', 'bedtools GroupByBed', 'bedtools ClosestBed', 'bedtools GetFastaBed', 'bedtools MultiCovBed', 'bedtools MergeBED', and 'bedtools SubtractBed'. The main panel displays the configuration for 'bedtools GetFastaBed'. It includes a 'BED/bedGraph/GFF/VCF file' section with a file selection button and a dropdown menu showing '9: Extract features ...'. Below this is a 'Choose the source for the FASTA file' section with a dropdown menu showing 'History'. The 'FASTA file' section has a file selection button and a dropdown menu showing '1: ASM2650v1.fa'. There is also a 'Use the 'name' column in the BED file for the FASTA headers in the output FASTA file' section with 'Yes' and 'No' buttons. The right sidebar shows a 'History' section with a search bar and a list of datasets including '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', '6: Trimmomatic on SRR 6322567\_2 (R2 paired)', and '5: Trimmomatic on SRR 6322567\_1 (R1 paired)'. Each dataset entry has icons for viewing, editing, and deleting.

# W7-4: bedtools GetFasta

①次はBED。②bedtools GetFastaBedプログラムを利用して、トランスクリプトーム配列取得を行います。②をクリック。中央パネル上に操作画面が表示されます。③これが目に入った瞬間に、BED形式だけでなく他の座標情報を取り扱う形式も入力として受け入れているのだと判断できる。

The screenshot shows the Galaxy web interface with the **bedtools GetFastaBed** tool selected. The tool description states: "use intervals to extract sequences from a FASTA file (Galaxy Version 2.29.0)".

The configuration panel includes the following fields:

- BED/bedGraph/GFF/VCF file**: A file selection field with a dropdown menu showing "9: Extract features ...". A red arrow with the number "3" points to this field.
- Choose the source for the FASTA file**: A dropdown menu set to "History".
- FASTA file**: A file selection field with a dropdown menu showing "1: ASM2650v1.fa".
- Use the 'name' column in the BED file for the FASTA headers in the output FASTA file**: Radio buttons for "Yes" and "No".

The **History** panel on the right shows a list of datasets:

- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)

# W7-5: bedtools GetFasta

①次はBED。②bedtools GetFastaBedプログラムを利用して、トランスクリプトーム配列取得を行います。②をクリック。中央パネル上に操作画面が表示されます。③これが目に入った瞬間に、BED形式だけでなく他の座標情報を取り扱う形式も入力として受け入れているのだと判断できる。③の部分で選択可能なものは、④デフォルトのヒストリー9と、⑤gene以外のfeatureも含む前処理前のヒストリー2。当然デフォルトのままで行う。

The screenshot shows the Galaxy web interface for the **bedtools GetFastaBed** tool. The tool description states: "Use intervals to extract sequences from a FASTA file (Galaxy Version 2.29.0)".

**Tools List (Left Panel):**

- bedtools ExpandBed** replicate lines based on lists of values in columns
- bedtools GroupByBed** group by common cols and summarize other cols
- bedtools ClosestBed** find the closest, potentially non-overlapping interval
- bedtools GetFastaBed** use intervals to extract sequences from a FASTA file
- bedtools MultiCovBed** counts coverage from multiple BAMs at specific intervals
- bedtools MergeBED** combine overlapping/nearby intervals into a single interval
- bedtools SubtractBed** remove

**Configuration Panel (Center):**

- Input:** BED/bedGraph/GFF/VCF file
- Choose the source:** A dropdown menu is open, showing options: "9: Extract features on data 2" (highlighted), "2: ASM2650v1.gff3", and "1: ASM2650v1.fa".
- FASTA file:** "1: ASM2650v1.fa"
- Options:** "Use the 'name' column in the BED file for the FASTA headers in the output FASTA file" (Yes/No)

**History Panel (Right):**

- 9 shown
- 1.29 GB
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
- 5: Trimmomatic on SRR 6322567\_1 (R1 paired)

Red arrows point to specific elements: ③ points to the tool description, ④ points to the selected dropdown option, and ⑤ points to the second option in the dropdown.

# W7-6: bedtools GetFasta

①中央パネルを下部に移動していく。一番下までいくと、このプログラムの文献情報が見られますので、適切に引用しましょう。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and a list of tools including 'bedtools ExpandBed', 'bedtools GroupByBed', 'bedtools ClosestBed', 'bedtools GetFastaBed', 'bedtools MultiCovBed', 'bedtools MergeBED', and 'bedtools SubtractBed'. The main panel displays the configuration for the 'bedtools GetFastaBed' tool. It includes a search bar, a 'Favorite' button, 'Versions', and 'Options' buttons. The description states: 'bedtools GetFastaBed use intervals to extract sequences from a FASTA file (Galaxy Version 2.29.0)'. The configuration section includes a 'BED/bedGraph/GFF/VCF file' input with a dropdown menu showing '9: Extract features ...'. Below this is a 'Choose the source for the FASTA file' dropdown menu set to 'History'. The 'FASTA file' input shows '1: ASM2650v1.fa'. There is a checkbox for 'Use the 'name' column in the BED file for the FASTA headers in the output FASTA file' which is currently unchecked. The right sidebar shows the 'History' section with a search bar and a list of datasets. A red arrow with the number '1' points to the '9: Extract features on data 2' dataset in the history list.

# W7-7: bedtools GetFasta

①中央パネルを下部に移動していく。一番下までいくと、このプログラムの文献情報が見られますので、適切に引用しましょう。オプションって読み飛ばしがちですが、②鎖の方向性が重要となる局面もあるので、一応気にはとめておいた方がよいと思います。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools Panel (Left):** A search bar for tools. Under the 'BED' category, 'bedtools GetFastaBed' is highlighted. Its description reads: 'use intervals to extract sequences from a FASTA file'.
- Configuration Panel (Center):** Shows options for the 'bedtools GetFastaBed' tool:
  - 'Use the 'name' column in the BED file for the FASTA headers in the output FASTA file': Yes (selected), No.
  - 'Report extract sequences in a tab-delimited format instead of in FASTA format': Yes (selected), No.
  - 'Force strandedness': Yes (selected), No. A red arrow with the number '2' points to this option.
  - 'Treat split/spliced BAM or BED12 entries as distinct BED intervals when computing coverage.': Yes (selected), No.
- History Panel (Right):** Shows a list of datasets. A red arrow with the number '1' points to the top of this panel. The list includes:
  - 9: Extract features on data 2
  - 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
  - 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
  - 6: Trimmomatic on SRR 6322567\_2 (R2 paired)
  - 5: Trimmomatic on SRR 6322567\_1 (R1 paired)



# W7-8: bedtools GetFasta

①中央パネルを下部に移動していく。一番下までいくと、このプログラムの文献情報が見られますので、適切に引用しましょう。オプションって読み飛ばしがちですが、②鎖の方向性が重要となる局面もあるので、一応気にはとめておいた方がよいと思います。③実行。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools Panel (Left):** Lists various tools under the 'BED' category. The 'bedtools GetFastaBed' tool is highlighted, with a description: 'use intervals to extract sequences from a FASTA file'.
- Configuration Panel (Center):** Shows the tool's options. The option 'Treat split/spliced BAM or BED12 entries as distinct BED intervals when computing coverage.' is set to 'No'. Below this is an 'Execute' button, which is highlighted with a red arrow and the number '3'.
- History Panel (Right):** Shows a list of previous jobs. The job '9: Extract features on data 2' is highlighted in green, with a red arrow and the number '1' pointing to it. Other jobs listed include '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', '6: Trimmomatic on SRR 6322567\_2 (R2 paired)', and '5: Trimmomatic on SRR 6322567\_1 (R1 paired)'.

# W7-9: bedtools GetFastaBed

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

BED

- bedtools ExpandBed** replicate lines based on lists of values in columns
- bedtools GroupByBed** group by common cols and summarize other cols
- bedtools ClosestBed** find the closest, potentially non-overlapping interval
- bedtools GetFastaBed** use intervals to extract sequences from a FASTA file
- bedtools MultiCovBed** counts coverage from multiple BAMs at specific intervals
- bedtools MergeBED** combine overlapping/nearby intervals into a single interval
- bedtools SubtractBed** remove

Executed **bedtools GetFastaBed** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- 9: Extract features on data 2
- 1: ASM2650v1.fa

It produces this output:

- 10: bedtools GetFastaBed on data 1 and data 9

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

**trans\_map**

10 shown

1.29 GB

- 10: bedtools GetFastaBed on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

実行完了。①ヒストリー10の中身を中央パネル  
上に表示。

# W7-10: bedtools GetFastaBed

Galaxy

usegalaxy.org

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

BED

- bedtools ExpandBed** replicate lines based on lists of values in columns
- bedtools GroupByBed** group by common cols and summarize other cols
- bedtools ClosestBed** find the closest, potentially non-overlapping interval
- bedtools GetFastaBed** use intervals to extract sequences from a FASTA file
- bedtools MultiCovBed** counts coverage from multiple BAMs at specific intervals
- bedtools MergeBED** combine overlapping/nearby intervals into a single interval
- bedtools SubtractBed** remove

Executed **bedtools GetFastaBed** and successfully added 1 job to the queue.

The tool uses 2 inputs:

- 9: Extract features on data 2
- 1: ASM2650v1.fa

It produces this output:

- 10: bedtools GetFastaBed on data 1 and data 9

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

trans\_map

10 shown

1.29 GB

- 10: bedtools GetFastaBed on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W7-11: bedtools GetF

実行完了。①ヒストリー10の中身を中央パネル  
上に表示。②確かにmulti-FASTA形式ファイル  
が作成されていて妥当ですね。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar lists various tools, including 'bedtools ExpandBed', 'bedtools GroupByBed', 'bedtools ClosestBed', 'bedtools GetFastaBed', 'bedtools MultiCovBed', 'bedtools MergeBED', and 'bedtools SubtractBed'. The main panel displays the output of the 'bedtools GetFastaBed' tool, which is a multi-FASTA file. A red box highlights the output, and a red arrow points to the tool entry in the history panel. The history panel shows a list of jobs, with the top job being '10: bedtools GetFastaBed'.

Galaxy  
usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools  
search tools

BED

- bedtools ExpandBed** replicate lines based on lists of values in columns
- bedtools GroupByBed** group by common cols and summarize other cols
- bedtools ClosestBed** find the closest, potentially non-overlapping interval
- bedtools GetFastaBed** use intervals to extract sequences from a FASTA file
- bedtools MultiCovBed** counts coverage from multiple BAMs at specific intervals
- bedtools MergeBED** combine overlapping/nearby intervals into a single interval
- bedtools SubtractBed** remove

This dataset is large and only the first megabyte is shown below.  
[Show all](#) | [Save](#)

```
>FM179322:0-1350
ATGCCCAATTTAGAGGAGCTTTGGGCTTACCTGAATGATAAATTCGGTGAAGAGTTGA
>FM179322:1523-2663
ATGAAATTTACGATTACCCGATCCACATTTCTGAAAACCTTGAATGATGTTTCCCGGG
>FM179322:3156-3369
ATGACAACCATCAAGATCACCACCGCATTTTTGACATTGGGCCAGTTTCTGAAAGAGG
>FM179322:3365-4484
ATGAAGCTGGATCATTTAACGTTAAAAAATATCGCAATTACGCAACGGTCGATACGG
>FM179322:4515-6477
GTGACGGACAAGAAAGAATCGGCCGAAGAAAAGAAAGAAGAACTGGCAGCAGAATATG
>FM179322:6539-9152
ATGGATGATCGCCAAGAAAGCCGGATTACGAATGTGAATCTCGGCGAAACAATGCGTA
>FM179322:9369-9471
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:9844-10075
ATGATGACGACAACCTTAGTTTGTACGATGCCTTAGGTGAGGATCAACTCCAAGGTA
>FM179322:10478-10682
ATGAATATGGACGAGAGTTTACGGAGACGACAGCTGATGACTATAAACAAATTCGG
>FM179322:11005-11437
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:11563-11860
ATGGCTGAAACCAAATATGAAGTCACTTACATCATTCGTCGGATCTGGATGAAGCTG
>FM179322:11889-12480
```

History  
search datasets

trans\_map  
10 shown  
1.29 GB

- 10: bedtools GetFastaBed on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W7-12: bedtools GetF

実行完了。①ヒストリー10の中身を中央パネル上に表示。②確かにmulti-FASTA形式ファイルが作成されていて妥当ですね。③の部分をクリックすれば、④配列数が2,949個だとわかる。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar lists various tools under the 'BED' category, including 'bedtools ExpandBed', 'bedtools GroupByBed', 'bedtools ClosestBed', 'bedtools GetFastaBed', 'bedtools MultiCovBed', 'bedtools MergeBED', and 'bedtools SubtractBed'. The main panel displays a dataset with a warning: 'This dataset is large and only the first megabyte is shown below.' Below the warning is a list of genomic coordinates and sequence fragments. The right sidebar shows the 'History' panel with a search bar and a list of datasets. The top entry is '10: bedtools GetFastaB ed on data 1 and data 9', which is highlighted in green. A red arrow labeled '3' points to this entry. Below the entry, it shows '2,949 sequences' and 'フォーマット: fasta, データベース: ?'. A red arrow labeled '4' points to the '2,949 sequences' text. The bottom of the history panel shows a preview of the FASTA output: '>FM179322:0-1350 ATGCCCAATTTAGAGGAGCTTTGGGCTTACCTGAATGATAAAATCCGTGAAGAGTTGA'.

# W7-13: bedtools GetF

実行完了。①履歴10の中身を中央パネル上に表示。②確かにmulti-FASTA形式ファイルが作成されていて妥当ですね。③の部分をクリックすれば、④配列数が2,949個だとわかる。⑤を押して、BEDカテゴリを閉じる。

The screenshot shows the Galaxy web interface. The 'Tools' panel on the left has a search bar and a list of tool categories. 'BED' is highlighted with a red arrow and a circled '5'. The central panel displays a large dataset with a warning message: 'This dataset is large and only the first megabyte is shown below.' Below the warning is a list of genomic coordinates and sequence fragments. The 'History' panel on the right shows a list of recent jobs, with '10: bedtools GetFastaB ed on data 1 and data 9' at the top.

Galaxy

usegalaxy.org

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

BED

VCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Fetch Sequences/Alignments

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

Variant Calling

This dataset is large and only the first megabyte is shown below.  
Show all | Save

```
>FM179322:0-1350
ATGCCCAATTTAGAGGAGCTTTGGGCTTACCTGAATGATAAATTCGCGAAGAGTTGA
>FM179322:1523-2663
ATGAAATTTACGATTACCCGATCCACATTTCTTGAAAACCTTGAATGATGTTTCCCGGG
>FM179322:3156-3369
ATGACAACCATCAAGATCACCACCGCATTTTTGACATTGGGCCAGTTTCTGAAAGAGG
>FM179322:3365-4484
ATGAAGCTGGATCATTTAACGTTAAAAAATATCGCAATTACGCAACGGTCGATACGG
>FM179322:4515-6477
GTGACGGACAAGAAAGAAATCGGCCGAAGAAAAGAAAGAAGAACTGGCAGCAGAATATG
>FM179322:6539-9152
ATGGATGATCGCCAAGAAAGCCGGATTACGAATGTGAATCTCGGCAGAAACAATGCGTA
>FM179322:9369-9471
ATGTTAAACAATGTGAGTTTAAACCGCCGCTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:9844-10075
ATGATGACGACAACCTTAGTTTGTACGATGCCTTAGGTGAGGATCAACTCCAAGGTA
>FM179322:10478-10682
ATGAATATGGACGAGAGTTTACGGGAGACGACAGCTGATGACTATAAACAAATTCGG
>FM179322:11005-11437
ATGTTAAACAATGTGAGTTTAAACCGCCGCTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:11563-11860
ATGGCTGAAACCAATATGAAGTCACTTACATCATTCGTCGGATCTGGATGAAGCTG
>FM179322:11889-12480
```

History

search datasets

trans\_map

10 shown

1.29 GB

10: bedtools GetFastaB ed on data 1 and data 9

9: Extract features on data 2

8: Trimmomatic on SRR 6322569\_2 (R2 paired)

7: Trimmomatic on SRR 6322569\_1 (R1 paired)

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- **W8: Kallisto quant**
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W8-1 : Kallisto quant

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' panel with a search bar and categories: BED, VCF/BCF, Nanopore, Convert Formats, Lift-Over, COMMON GENOMICS TOOLS, Operate on Genomic Intervals, Fetch Sequences/Alignments, GENOMICS ANALYSIS, Assembly, Annotation, Mapping, and Variant Calling. A red arrow with the number 1 points to the 'Mapping' category. The main content area displays a warning: 'This dataset is large and only the first megabyte is shown below. Show all | Save'. Below the warning is a list of genomic coordinates and sequence reads. The right sidebar shows a 'History' panel with a search bar and a list of jobs: '10: bedtools GetFastaB ed on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'.



# W8-1 : Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq

Galaxy

usegalaxy.org/datasets/edit

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

ChIP-seq

RNA-seq

Multiple Alignments

Phenotype Association

Regional Variation

STR-FM: Microsatellite Analysis

Chromosome Conformation

METAGENOMICS

Metagenomic Analysis

Mothur

GENOMICS TOOLKITS

Picard

This dataset is large and only the first megabyte is shown below.  
Show all | Save

```
>FM179322:0-1350
ATGCCCAATTTAGAGGAGCTTTGGGCTTACCTGAATGATAAATTCGGTGAAGAGTTGA
>FM179322:1523-2663
ATGAAATTTACGATTACCCGATCCACATTCTTGAAAACCTTGAATGATGTTTCCCGGG
>FM179322:3156-3369
ATGACAACCATCAAGATCACCACCGCATTTTTGACATTGGGCCAGTTTCTGAAAGAGG
>FM179322:3365-4484
ATGAAGCTGGATCATTTAACGTTAAAAAATATCGCAATTACGCAACGGTCGATACGG
>FM179322:4515-6477
GTGACGGACAAGAAAGAATCGGCCGAAGAAAAGAAAGAAGAAGTGGCAGCAGAATATG
>FM179322:6539-9152
ATGGATGATCGCCAAGAAAGCCGGATTACGAATGTGAATCTCGGCGAAACAATGCGTA
>FM179322:9369-9471
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:9844-10075
ATGATGACGACAACCTTAGTTTGTACGATGCCTTAGGTGAGGATCAACTCCAAGGTA
>FM179322:10478-10682
ATGAATATGGACGAGAGTTTACGGGAGACGACAGCTGATGACTATAAACAATTCCTGG
>FM179322:11005-11437
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:11563-11860
ATGGCTGAAACCAATATGAAGTCACTTACATCATTCGTCGGATCTGGATGAAGCTG
>FM179322:11889-12480
```

History

search datasets

trans\_map

10 shown

1.29 GB

10: bedtools GetFastaB ed on data 1 and data 9

9: Extract features on data 2

8: Trimmomatic on SRR 6322569\_2 (R2 paired)

7: Trimmomatic on SRR 6322569\_1 (R1 paired)

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W8-1 : Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq  
。今回利用する③Kallisto quantはこの段階で  
見えてはいますが、④一応少し下部に移動。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'RNA-seq': 'Cuffmerge', 'Filter Combined Transcripts', 'GffCompare', 'DESeq2', 'goseq', 'Kallisto pseudo', and 'Kallisto quant'. A red arrow labeled '3' points to 'Kallisto pseudo'. Another red arrow labeled '4' points to 'Filter Combined Transcripts'. The main content area displays a dataset view with a warning message: 'This dataset is large and only the first megabyte is shown below. Show all | Save'. Below the warning is a list of sequence reads. The right sidebar shows a 'History' section with a search bar and a list of datasets: 'trans\_map', '10: bedtools GetFastaB ed on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'.

# W8-1 : Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq。今回利用する③Kallisto quantはこの段階で見えてはいますが、④一応少し下部に移動。こんな感じ。⑤Salmonも有名です。

Galaxy

usegalaxy.org/datasets/edit

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

DESeq2 Determines differentially expressed features from count tables

goseq tests for overrepresented gene categories

**Kallisto pseudo - run**  
pseudoalignment on RNA-Seq transcripts

**Kallisto quant - quantify abundances of RNA-Seq transcripts**

Salmon quant Perform dual-phase, reads or mapping-based estimation of transcript abundance from RNA-seq reads

StringTie merge transcripts

StringTie transcript assembly and quantification

DEXSeq Determines differential exon

This dataset is large and only the first megabyte is shown below.  
Show all | Save

```
>FM179322:0-1350
ATGCCCAATTTAGAGGAGCTTTGGGCTTACCTGAATGATAAATTCGGTGAAGAGTTGA
>FM179322:1523-2663
ATGAAATTTACGATTACCCGATCCACATTTCTTGAAACCTTGAATGATGTTTCCCGGG
>FM179322:3156-3369
ATGACAACCATCAAGATCACCACCGCATTTTTGACATTGGGCCAGTTTCTGAAAGAGG
>FM179322:3365-4484
ATGAAGCTGGATCATTTAACGTTAAAAAATATCGCAATTACGCAACGGTCGATACGG
>FM179322:4515-6477
GTGACGGACAAGAAAGAATCGGCCGAAGAAAAGAAAGAAGAACTGGCAGCAGAATATG
>FM179322:6539-9152
ATGGATGATCGCCAAGAAAGCCGGATTACGAATGTGAATCTCGGCGAAACAATGCGTA
>FM179322:9369-9471
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:9844-10075
ATGATGACGACAACCTTAGTTTGTACGATGCCTTAGGTGAGGATCAACTCCAAGGTA
>FM179322:10478-10682
ATGAATATGGACGAGAGTTTACCGGAGACGACAGCTGATGACTATAAACAAATTCGG
>FM179322:11005-11437
ATGTTAAACAATGTGAGTTTAAACCGCCGTCTGACGAAGGAACCAGAAGTGTAAAAA
>FM179322:11563-11860
ATGGCTGAAACCAAATATGAAGTCACTTACATCATTCGTCGGATCTGGATGAAGCTG
>FM179322:11889-12480
```

History

search datasets

trans\_map

10 shown

1.29 GB

10: bedtools GetFastaB ed on data 1 and data 9

9: Extract features on data 2

8: Trimmomatic on SRR 6322569\_2 (R2 paired)

7: Trimmomatic on SRR 6322569\_1 (R1 paired)

6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W8-1 : Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq。今回利用する③Kallisto quantはこの段階で見えてはいますが、④一応少し下部に移動。こんな感じ。⑤Salmonも有名です。③をクリック。

The screenshot shows the Galaxy web interface at [usegalaxy.org/datasets/edit](https://usegalaxy.org/datasets/edit). The top navigation bar includes "Galaxy", "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". A "Using 1%" indicator is visible on the right.

The "Tools" panel on the left contains a search bar and a list of tools. The tool "Kallisto pseudo - run" is highlighted with a red circle containing the number 3. Other tools listed include DESeq2, goseq, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq.

The central area displays a warning: "This dataset is large and only the first megabyte is shown below." Below the warning, the first megabyte of sequence data is shown, including coordinates and sequence reads.

The "History" panel on the right shows a search bar and a list of datasets. The dataset "trans\_map" is selected, showing a size of 1.29 GB. A list of operations is shown, including "10: bedtools GetFastaB ed on data 1 and data 9", "9: Extract features on data 2", "8: Trimmomatic on SRR 6322569\_2 (R2 paired)", "7: Trimmomatic on SRR 6322569\_1 (R1 paired)", and "6: Trimmomatic on SRR 6322567\_2 (R2 paired)".

# W8-1 : Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq。今回利用する③Kallisto quantはこの段階で見えてはいますが、④一応少し下部に移動。こんな感じ。⑤Salmonも有名です。③をクリック。こんな感じになり、中央パネルにKallisto quantの操作画面が表示されます。

The screenshot displays the Galaxy web interface for configuring the Kallisto quant tool. The top navigation bar shows the Galaxy logo and the URL 'usegalaxy.org'. The left sidebar contains a 'Tools' panel with a search bar and a list of tools, including DESeq2, goseq, Kallisto pseudo, Kallisto quant (selected), Salmon quant, StringTie merge, StringTie, and DEXSeq. The central panel shows the configuration for 'Kallisto quant', including options for 'Reference transcriptome for quantification' (set to 'Use a built-in transcriptome'), 'Select a reference transcriptome' (set to 'No options available'), 'Single-end or paired reads' (set to 'Single-end'), and 'Reads in FASTQ format' (set to '8: Trimmomatic o...'). The 'Average fragment length' is set to 200. The right sidebar shows a 'History' panel with a search bar and a list of recent datasets and operations, including '10: bedtools GetFastaBe d on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'. The interface is clean and modern, with a dark header and a light background.

# W8-2: リファレンス

まずはリファレンス配列の指定。デフォルトの①  
Use a built-in transcriptomeから...

The screenshot shows the Galaxy web interface with the 'Kallisto quant' tool selected. The 'Reference transcriptome for quantification' dropdown menu is highlighted with a red arrow and a circled '1', indicating the step to select a built-in transcriptome. The 'History' panel on the right shows a list of datasets, including '10: bedtools GetFastaBed on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'.

# W8-2: リファレンス

まずはリファレンス配列の指定。デフォルトの① Use a built-in transcriptomeから、② Use a transcriptome from historyに変更。

The screenshot displays the Galaxy web interface. The main content area shows the configuration for the 'Kallisto quant' tool. Under the 'Reference transcriptome for quantification' section, the dropdown menu is open, and the option 'Use a transcriptome from history' is highlighted with a red arrow and a circled '2'. The 'History' panel on the right shows a list of datasets, including '10: bedtools GetFastaBed on data 1 and data 9' and '9: Extract features on data 2'.

# W8-2: リファレンス

まずはリファレンス配列の指定。デフォルトの① Use a built-in transcriptomeから、② Use a transcriptome from historyに変更。こんな感じになります。

The screenshot displays the Galaxy web interface for the Kallisto quant tool. The main configuration area includes:

- Kal list:** Favorite, Versions, Options
- o quant:** quantify abundances of RNA-Seq transcripts (Galaxy Version 0.46.0.4)
- Reference transcriptome for quantification:** Use a transcriptome from history
- FASTA reference transcriptome:** 10: bedtools GetF...
- Single-end or paired reads:** Single-end
- Reads in FASTQ format:** 8: Trimmomatic o...
- Average fragment length:** 200

The History panel on the right shows a list of datasets, with '10: bedtools GetFastaBe d on data 1 and data 9' selected. Other datasets include '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'.



# W8-2: リファレンス

まずはリファレンス配列の指定。デフォルトの① Use a built-in transcriptomeから、② Use a transcriptome from historyに変更。こんな感じになります。③リファレンス配列の候補として、④デフォルトのヒストリー10と、⑤ヒストリー1が選択可能ですが、④デフォルトが正解。

The screenshot displays the Galaxy web interface for the Kallisto quant tool. The 'Reference transcriptome for quantification' dropdown is set to 'Use a transcriptome from history'. In the 'FASTA reference transcriptome' section, a search for '10: bedtools GetFastaBed on data 1 and data 9' is shown, with a dropdown menu displaying this option as selected (marked with 4) and another option '1: ASM2650v1.fa' (marked with 5). The 'History' panel on the right shows a list of datasets, with the top entry '10: bedtools GetFastaBed on data 1 and data 9' highlighted in green.

# W8-2: リファレンス

まずはリファレンス配列の指定。デフォルトの① Use a built-in transcriptomeから、② Use a transcriptome from historyに変更。こんな感じになります。③リファレンス配列の候補として、④デフォルトのヒストリー10と、⑤ヒストリー1が選択可能ですが、④デフォルトが正解。つまりこの状態。

The screenshot displays the Galaxy web interface for the Kallisto quant tool. The main configuration area includes:

- Kal list:** Favorite, Versions, Options
- o quant:** quantify abundances of RNA-Seq transcripts (Galaxy Version 0.46.0.4)
- Reference transcriptome for quantification:** Use a transcriptome from history
- FASTA reference transcriptome:** 10: bedtools GetF...
- Single-end or paired reads:** Single-end
- Reads in FASTQ format:** 8: Trimmomatic o...
- Average fragment length:** 200

The History panel on the right shows a list of datasets, with '10: bedtools GetFastaBe d on data 1 and data 9' selected. Other visible datasets include '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'.

# W8-3: Pairedに変更

The screenshot shows the Galaxy web interface. The main panel is titled "FASTA reference transcriptome". Under the "Single-end or paired reads" section, a dropdown menu is currently set to "Single-end". A red arrow with the number "1" points to this dropdown, indicating the instruction to change it to "Paired". Below this, the "Reads in FASTQ format" section shows a dropdown set to "8: Trimmomatic o...". The "Average fragment length" is set to 200, and the "Estimated standard deviation of fragment length" is set to 20. The "Perform sequence based bias correction" section has "Yes" selected. The right-hand "History" panel shows a list of jobs, including "10: bedtools GetFastaBe d on data 1 and data 9", "9: Extract features on data 2", "8: Trimmomatic on SRR 6322569\_2 (R2 paired)", "7: Trimmomatic on SRR 6322569\_1 (R1 paired)", and "6: Trimmomatic on SRR 6322567\_2 (R2 paired)".

# W8-3: Pairedに変更

①少しページ下部に移動。②リードファイルはペアエンドデータなので、③Pairedに変更。

The screenshot shows the Galaxy web interface. The main panel is titled "FASTA reference transcriptome" and contains a dropdown menu for "Single-end or paired reads". The dropdown is open, showing "Single-end" and "Paired" options. The "Paired" option is selected and highlighted in blue. Two red arrows with numbers 2 and 3 point to the dropdown and the selected option, respectively. The "History" panel on the right shows a list of datasets, including "10: bedtools GetFastaBe d on data 1 and data 9".

# W8-3: Pairedに変更

①少しページ下部に移動。②リードファイルはペアエンドデータなので、③Pairedに変更後。

The screenshot shows the Galaxy web interface. The main panel is titled "FASTA reference transcriptome" and has several configuration options:

- FASTA reference transcriptome:** 10: bedtools GetF...
- Single-end or paired reads:** Paired
- Collection or individual datasets:** Individual files
- Forward reads:** 8: Trimmomatic...
- Reverse reads:** 8: Trimmomatic...
- Perform sequence based bias correction:** Yes
- Number of bootstrap samples:** (--bias)

The right-hand panel is titled "History" and shows a list of datasets. The top dataset is highlighted in green:

- 10: bedtools GetFastaBe d on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。

The screenshot shows the Galaxy web interface with the following components:

- Header:** Galaxy logo, navigation menu (データ解析, ワークフロー, 可視化する, 共有データ, ヘルプ, ユーザー), and a progress indicator (Using 1%).
- Tools Panel (Left):** Search tools, DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq.
- Main Tool Configuration (Center):**
  - FASTA reference transcriptome:** Input field: 10: bedtools GetF...
  - Single-end or paired reads:** Paired
  - Collection or individual datasets:** Individual files
  - Forward reads:** 8: Trimmomatic... (highlighted with a red arrow and '1')
  - Reverse reads:** 8: Trimmomatic...
  - Perform sequence based bias correction:** Yes/No buttons, (--bias)
  - Number of bootstrap samples:** (partially visible)
- History Panel (Right):** Search datasets, trans\_map, 10 shown, 1.29 GB. List of jobs:
  - 10: bedtools GetFastaBe d on data 1 and data 9
  - 9: Extract features on data 2
  - 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
  - 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
  - 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、② Multiple datasetsに変更。

The screenshot shows the Galaxy web interface. The main panel is titled "FASTA reference transcriptome". It has a search bar for tools and a dropdown menu for "Collection or individual datasets" set to "Individual files". A red arrow points to the "Multiple datasets" option in this dropdown. Below this, there are two input fields for "8: Trimmomatic..." with file selection icons. The "History" panel on the right shows a list of jobs, including "10: bedtools GetFastaBe d on data 1 and data 9".

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、②Multiple datasetsに変更。こんな感じになります。

The screenshot shows the Galaxy web interface with a workflow configuration for 'FASTA reference transcriptome'. The interface includes a top navigation bar with 'Galaxy' and 'usegalaxy.org', and a main workspace with several panels:

- Tools:** A search bar and a list of tools including DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq.
- FASTA reference transcriptome:** The main workflow configuration panel. It shows a search for '10: bedtools GetF...' and a dropdown menu for 'Single-end or paired reads' set to 'Paired'. Below this, there is a section for 'Collection or individual datasets' set to 'Individual files'. The 'Forward reads' section contains a list of 10 items, each representing a Trimmomatic job on a specific dataset (e.g., '8: Trimmomatic on SI', '7: Trimmomatic on SI', etc.). A note below this list states: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.'
- History:** A panel on the right showing a list of datasets. The top entry is 'trans\_map' (10 shown, 1.29 GB). Below it, a list of 10 datasets is shown, each with a green background and icons for viewing, editing, and deleting. The datasets are: '10: bedtools GetFastaBe d on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'. The bottom of the history panel shows 'Reverse reads'.



# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、② Multiple datasetsに変更。こんな感じになります。③を押して、Forward側に相当するヒストリーを指定する。

The screenshot shows the Galaxy web interface. The main panel is titled "FASTA reference transcriptome". It has several sections: "Single-end or paired reads" (set to Paired), "Collection or individual datasets" (set to Individual files), and "Forward reads" (expanded to show a list of datasets: 8: Trimmomatic on SI, 7: Trimmomatic on SI, 6: Trimmomatic on SI, 5: Trimmomatic on SI, 4: Trimmomatic on SI, 3: Trimmomatic on SI). A red arrow with the number 3 points to the folder icon next to the list. Below the list is a note: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection." The "Reverse reads" section is partially visible at the bottom. The "History" panel on the right shows a list of jobs, with the top job highlighted in green: "10: bedtools GetFastaBed on data 1 and data 9".

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、② Multiple datasetsに変更。こんな感じになります。③を押して、Forward側に相当するヒストリーを指定する。こんな感じになります。

Type to Search

Label	Details	Time
10: bedtools GetFastaBed on data 1 and data 9	fasta	2019-11-25 08:22
9: Extract features on data 2	gff3	2019-11-25 08:20
8: Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:53
7: Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:53
6: Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:45
5: Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:45
4: Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:44
3: Trimmomatic on SRR6322564_1 (R1 paired)	fastasanger.gz	2019-11-23 06:44

Cancel Ok

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

Reverse reads

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、②Multiple datasetsに変更。こんな感じになります。③を押して、Forward側に相当するヒストリーを指定する。こんな感じになります。④下のほうに移動して、リードファイル群に相当するヒストリーの全貌が見られるようにしておいて、⑤Forward側に相当する3つのR1 pairedを選択して、⑥Ok。

Job ID	Job Name	File Type	Timestamp
9	Extract features on data 2	gff3	2019-11-25 08:20
8	Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:53
7	Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:53
6	Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:45
5	Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:45
4	Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:44
3	Trimmomatic on SRR6322564_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:44
2	ASM2650v1.gff3	gff3	2019-11-25 02:11
1	ASM2650v1.fasta	fasta	2019-11-25 08:20

# W8-4: Forward側

次が、リードファイル指定の本番。まずは① Forward側。ここでは計3サンプル分の指定を行うので、②Multiple datasetsに変更。こんな感じになります。③を押して、Forward側に相当するヒストリーを指定する。こんな感じになります。④下のほうに移動して、リードファイル群に相当するヒストリーの全貌が見られるようにしておいて、⑤Forward側に相当する3つのR1 pairedを選択して、⑥Ok。こんな感じになります。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', and '共有データ'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main workspace is titled 'FASTA reference transcriptome' and shows a file list with '10: bedtools GetF...' selected. Below this, the 'Single-end or paired reads' dropdown is set to 'Paired'. The 'Collection or individual datasets' dropdown is set to 'Individual files'. Under 'Forward reads', a list of files is shown, with '8: Trimmomatic on SRR6322569\_2 (R2 paired)' selected. A note below the list states: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' The right panel shows a list of jobs, with '10: bedtools GetFastaBed on data 1 and data 9' highlighted in green. Other jobs in the list include '9: Extract features on data 2', '8: Trimmomatic on SRR6322569\_2 (R2 paired)', '7: Trimmomatic on SRR6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR6322567\_2 (R2 paired)'.

# W8-5: Reverse側

次は①Reverse側を指定すべく、②少しページ下部に移動したところ。

The screenshot shows the Galaxy web interface. The 'Tools' panel on the left lists various bioinformatics tools. The main workspace is divided into 'Forward reads' and 'Reverse reads' sections. The 'Reverse reads' section has a dropdown menu with '8: Trimmomatic...' selected. Below it are options for 'Perform sequence based bias correction' (Yes/No) and 'Number of bootstrap samples' (0). The 'History' panel on the right shows a list of jobs, with job 10 highlighted in green. A red arrow labeled '1' points to the 'Reverse reads' section, and a red arrow labeled '2' points to the 'History' panel.

# W8-5: Reverse側

次は①Reverse側を指定すべく、②少しページ下部に移動したところ。③Multiple datasets。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main workspace is divided into 'Forward reads' and 'Reverse reads' sections. The 'Reverse reads' section has a dropdown menu set to '8: Trimmomatic...' and a 'Multiple datasets' button highlighted with a red arrow and the number '3'. Below this is a 'Perform sequence based bias correction' section with 'Yes' and 'No' buttons, and a 'Number of bootstrap samples' input field set to '0'. The right sidebar shows a 'History' section with a search bar and a list of datasets, including '10: bedtools GetFastaBed on data 1 and data 9', '9: Extract features on data 2', '8: Trimmomatic on SRR 6322569\_2 (R2 paired)', '7: Trimmomatic on SRR 6322569\_1 (R1 paired)', and '6: Trimmomatic on SRR 6322567\_2 (R2 paired)'. The browser address bar shows 'usegalaxy.org'.

# W8-5: Reverse側

次は①Reverse側を指定すべく、②少しページ下部に移動したところ。③Multiple datasets。④Browse Datasets。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main content area is divided into 'Forward reads' and 'Reverse reads' sections. Both sections show a list of jobs (8, 7, 6, 5, 4, 3) and a 'Browse Datasets' button. A red arrow with the number '4' points to the 'Browse Datasets' button in the 'Reverse reads' section. The right sidebar contains a 'History' section with a search bar and a list of datasets, including 'trans\_map' and '10: bedtools GetFastaBed on data 1 and data 9'.

# W8-5: Reverse側

次は①Reverse側を指定すべく、②少しページ下部に移動したところ。③Multiple datasets。④Browse Datasets。⑤Reserve側に相当する3つのR2 pairedを選択して、⑥Ok。

The screenshot shows the Galaxy web interface with a dataset selection dialog open. The dialog has a search bar and a list of datasets. Three datasets are selected, indicated by green highlights and red arrows labeled '5'. The 'Ok' button is highlighted with a red arrow labeled '6'.

Dataset ID	Description	Format	Created
9	Extract features on data 2	gff3	2019-11-25 08:20
8	Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:53
7	Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:53
6	Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:45
5	Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:45
4	Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-11-23 06:44
3	Trimmomatic on SRR6322564_1 (R1 paired)	fastqsanger.gz	2019-11-23 06:44
2	ASM2650v1.gff3	gff3	2019-11-25 02:11
1	ASM2650v1.fasta	fasta	2019-11-25 08:20



# W8-5: Reverse側

次は①Reverse側を指定すべく、②少しページ下部に移動したところ。③Multiple datasets。④Browse Datasets。⑤Reverse側に相当する3つのR2 pairedを選択して、⑥Ok。

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main workspace is divided into 'Forward reads' and 'Reverse reads' sections. Both sections show a list of jobs (3-8) for 'Trimmomatic on SRR 6322567\_2 (R2 paired)'. Below each list is a note: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' The right sidebar shows a 'History' section with a search bar and a list of jobs. Job 6, 'Trimmomatic on SRR 6322567\_2 (R2 paired)', is highlighted in green. Other jobs in the history include 'bedtools GetFastaBed on data 1 and data 9', 'Extract features on data 2', 'Trimmomatic on SRR 6322569\_2 (R2 paired)', and 'Trimmomatic on SRR 6322569\_1 (R1 paired)'.

# W8-6: 実行

様々なオプションを眺めながら、①下部に移動。②ブートストラップのサンプリング回数のデフォルトは0、つまりブートストラップは行わないということのようです(本文中で言及あり)。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the tool "Perform sequence based bias correction". The "Number of bootstrap samples" field is set to 0, with a red arrow labeled "2" pointing to it. The "Seed for the bootstrap sampling" field is set to 42. The "History" panel on the right shows a list of jobs, with a red arrow labeled "1" pointing to the bottom of the list. The jobs listed are:

- 10: bedtools GetFastaBe d on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

# W8-6: 実行

様々なオプションを眺めながら、①下部に移動。②ブートストラップのサンプリング回数のデフォルトは0、つまりブートストラップは行わないということのようです(本文中で言及あり)。③さらに下部に移動。④Execute。

The screenshot shows the Galaxy web interface for the Kallisto quant tool. The main configuration area includes:

- Seed for the bootstrap sampling:** Input field with value 42. Default: 42 (--seed).
- Search for fusions:** Radio buttons for Yes and No.
- Library strandness information:** Dropdown menu set to Unstranded.
- Output pseudoalignments in BAM format:** Radio buttons for Yes and No.
- Execute button:** A blue button with a checkmark and the text "Execute". A red arrow with a circled number 4 points to it. A tooltip below it reads "Execute: Kallisto quant (0.46.0.4)".

The right-hand side of the interface shows the **History** panel, which lists previous jobs. A red arrow with a circled number 3 points to the job list. The history items include:

- 10: bedtools GetFastaBed on data 1 and data 9
- 9: Extract features on data 2
- 8: Trimmomatic on SRR 6322569\_2 (R2 paired)
- 7: Trimmomatic on SRR 6322569\_1 (R1 paired)
- 6: Trimmomatic on SRR 6322567\_2 (R2 paired)

## W8-7: 実行中...

The screenshot shows the Galaxy web interface at usegalaxy.org. The main content area displays a successful job execution for **Kallisto quant**. A green checkmark icon indicates success, with the message: "Executed **Kallisto quant** and successfully added 3 jobs to the queue."

The tool uses 7 inputs:

- 10: **bedtools GetFastaBed** on data 1 and data 9
- 3: **Trimmomatic** on SRR6322564\_1 (R1 paired)
- 5: **Trimmomatic** on SRR6322567\_1 (R1 paired)
- 7: **Trimmomatic** on SRR6322569\_1 (R1 paired)
- 4: **Trimmomatic** on SRR6322564\_2 (R2 paired)
- 6: **Trimmomatic** on SRR6322567\_2 (R2 paired)
- 8: **Trimmomatic** on SRR6322569\_2 (R2 paired)

It produces 6 outputs:

- 11: **Kallisto quant** on data 4, data 3, and data 10: Abundances (HDF5)
- 12: **Kallisto quant** on data 4, data 3,

The right sidebar shows the **History** panel with a search for datasets. The **trans\_map** dataset is selected, showing 16 shown items and 1.29 GB. The history list includes several Kallisto quant jobs on various data sets, with the most recent job (16) being the one just executed.

# W8-7: 実行中...

実行中...。数分後の状態。①入力は全部で7つ。②リファレンス配列、③forward側リード、④reverse側リード。

The screenshot shows the Galaxy web interface with a job titled "Executed Kallisto quant and successfully added 3 jobs to the queue" in a green box. The job details are as follows:

The tool uses 7 inputs:

- 10: bedtools GetFastaBed on data 1 and data 9
- 3: Trimmomatic on SRR6322564\_1 (R1 paired)
- 5: Trimmomatic on SRR6322567\_1 (R1 paired)
- 7: Trimmomatic on SRR6322569\_1 (R1 paired)
- 4: Trimmomatic on SRR6322564\_2 (R2 paired)
- 6: Trimmomatic on SRR6322567\_2 (R2 paired)
- 8: Trimmomatic on SRR6322569\_2 (R2 paired)

It produces 6 outputs:

- 11: Kallisto quant on data 4, data 3, and data 10: Abundances (HDF5)
- 12: Kallisto quant on data 4, data 3,

The History panel on the right shows a list of jobs, with the most recent ones being Kallisto quant jobs on various data sets, including "trans\_map" and "Abundances (tabular)" and "Abundances (HDF5)".

Red arrows in the image point to:

- ①: The job title "Executed Kallisto quant and successfully added 3 jobs to the queue".
- ②: The input list, specifically the first item "10: bedtools GetFastaBed on data 1 and data 9".
- ③: The input list, specifically the items "3: Trimmomatic on SRR6322564\_1 (R1 paired)" and "5: Trimmomatic on SRR6322567\_1 (R1 paired)".
- ④: The input list, specifically the items "7: Trimmomatic on SRR6322569\_1 (R1 paired)" and "4: Trimmomatic on SRR6322564\_2 (R2 paired)".

# W8-8: 実行完了

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

DESeq2 Determines differentially expressed features from count tables

goseq tests for overrepresented gene categories

Kallisto pseudo - run pseudoalignment on RNA-Seq transcripts

Kallisto quant - quantify abundances of RNA-Seq transcripts

Salmon quant Perform dual-phase, reads or mapping-based estimation of transcript abundance from RNA-seq reads

StringTie merge transcripts

StringTie transcript assembly and quantification

DEXSeq Determines differential exon

Executed **Kallisto quant** and successfully added 3 jobs to the queue.

The tool uses 7 inputs:

- 10: bedtools GetFastaBed on data 1 and data 9
- 3: Trimmomatic on SRR6322564\_1 (R1 paired)
- 5: Trimmomatic on SRR6322567\_1 (R1 paired)
- 7: Trimmomatic on SRR6322569\_1 (R1 paired)
- 4: Trimmomatic on SRR6322564\_2 (R2 paired)
- 6: Trimmomatic on SRR6322567\_2 (R2 paired)
- 8: Trimmomatic on SRR6322569\_2 (R2 paired)

It produces 6 outputs:

- 11: Kallisto quant on data 4, data 3, and data 10: Abundances (HDF5)
- 12: Kallisto quant on data 4, data 3,

History

search datasets

trans\_map

16 shown

1.29 GB

16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)

14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)

# W8-8: 実行完了

①実行完了。このときは約5分。②中央パネル下部に移動。

The screenshot shows the Galaxy web interface at usegalaxy.org. The main content area is a green box with the following text:

(R2 paired)

It produces 6 outputs:

- 11: Kallisto quant on data 4, data 3, and data 10: Abundances (HDF5)
- 12: Kallisto quant on data 4, data 3, and data 10: Abundances (tabular)
- 13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)
- 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)
- 15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)
- 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

The History panel on the right shows a list of datasets. A red arrow with the number 2 points to the entry for job 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular), which is highlighted in green.

# W8-8: 実行完了

①実行完了。このときは約5分。②中央パネル下部に移動。③出力は全部で6つ。計3サンプルのデータをジョブに投げているので、④1つのサンプルにつき2種類の出力(HDF5とtabular)となっているのだと解釈できます。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' panel with a search bar and a list of tools including DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The central panel displays a job titled '(R2 paired)' with the text 'It produces 6 outputs:' followed by a list of 6 items, each with a red arrow pointing to the number 4. The right panel shows the 'History' view with a search bar and a list of 6 datasets, each with a red arrow pointing to the number 4.

Galaxy

usegalaxy.org

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

DESeq2 Determines differentially expressed features from count tables

goseq tests for overrepresented gene categories

Kallisto pseudo - run pseudoalignment on RNA-Seq transcripts

Kallisto quant - quantify abundances of RNA-Seq transcripts

Salmon quant Perform dual-phase, reads or mapping-based estimation of transcript abundance from RNA-seq reads

StringTie merge transcripts

StringTie transcript assembly and quantification

DEXSeq Determines differential exon

(R2 paired)

It produces 6 outputs:

- 11: Kallisto quant on data 4, data 3, and data 10: Abundances (HDF5)
- 12: Kallisto quant on data 4, data 3, and data 10: Abundances (tabular)
- 13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)
- 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)
- 15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)
- 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

trans\_map

16 shown

1.29 GB

16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)

14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)



# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。

The screenshot shows the Galaxy web interface. The main panel displays a table with 4 columns: target\_id, length, eff\_length, and est\_cou. The table contains 20 rows of data. The History panel on the right shows a list of datasets, with two entries highlighted in green and annotated with red arrows and circled numbers. Arrow 1 points to entry 15, and arrow 2 points to entry 16.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	9
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	3
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	5
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	17
FM179322:14018-15497	1479	1258.25	17
FM179322:15489-16509	1020	799.253	17
FM179322:16548-16743	195	36.7899	17
FM179322:16821-17079	258	77.7218	17

History panel entries:

- 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)
- 15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)
- 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)
- 13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

DESeq2 Determines differentially expressed features from count tables

goseq tests for overrepresented gene categories

Kallisto pseudo - run pseudoalignment on RNA-Seq transcripts

Kallisto quant - quantify abundances of RNA-Seq transcripts

Salmon quant Perform dual-phase, reads or mapping-based estimation of transcript abundance from RNA-seq reads

StringTie merge transcripts

StringTie transcript assembly and quantification

DEXSeq Determines differential exon

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	
FM179322:1523-2663	1140	919.253	1
FM179322:3156-3369	213	47.1821	
FM179322:3365-4484	1119	898.253	
FM179322:4515-6477	1962	1741.25	1
FM179322:6539-9152	2613	2392.25	3
FM179322:9369-9471	102	29.4316	2.28
FM179322:9844-10075	231	59.1546	
FM179322:10478-10682	204	41.5832	
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	1
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.02
FM179322:13201-13708	507	291.66	
FM179322:14018-15497	1479	1258.25	
FM179322:15489-16509	1020	799.253	
FM179322:16548-16743	195	36.7899	
FM179322:16821-17079	258	77.7218	

History

search datasets

trans\_map

16 shown

1.29 GB

16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)

14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo - run, Kallisto quant, Salmon quant, StringTie merge transcripts, StringTie transcript assembly and quantification, and DEXSeq. The main panel displays a table with 4 columns: target\_id, length, eff\_length, and est\_cou. The table contains 20 rows of data. The right sidebar shows a 'History' panel with a search bar and a list of datasets. A red arrow labeled '4' points to a download icon (a square with a downward arrow) in the history panel, which is highlighted in green. The history panel also shows the size of the datasets (1.29 GB) and icons for viewing, editing, and deleting.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	9
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	3
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	5
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	3
FM179322:14018-15497	1479	1258.25	3
FM179322:15489-16509	1020	799.253	3
FM179322:16548-16743	195	36.7899	3
FM179322:16821-17079	258	77.7218	3

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。

The screenshot shows the Galaxy web interface. The main panel displays a table with columns labeled 1, 2, 3, and 4. The table contains data for various target IDs, including target\_id, length, eff\_length, and est\_cou. A red arrow labeled ⑤ points to the table. On the right, the History panel shows a dataset named '16: Kallisto quant on data 8, data 7, and data 1 0: Abundances (tabular)' with 2,950 lines. A red arrow labeled ⑥ points to the 'tabular' format information in the History panel. The left sidebar shows various tools like DESeq2, goseq, Kallisto pseudo-run, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	9
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	3
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	1
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	1
FM179322:14018-15497	1479	1258.25	1
FM179322:15489-16509	1020	799.253	1
FM179322:16548-16743	195	36.7899	1
FM179322:16821-17079	258	77.7218	1

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。

The screenshot shows the Galaxy web interface. On the left, there is a 'Tools' panel with a search bar and several tool options like DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main panel displays a table with 4 columns: target\_id, length, eff\_length, and est\_cou. The table contains 20 rows of data. On the right, there is a 'History' panel with a search bar and a list of datasets. A specific history entry is highlighted in green, showing the command 'kallisto quant on data 8, data 7, and data 10:'. Below the command, there is a small preview of the tabular data. Red arrows point to the download icon (7) and the history entry (6).

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	544
FM179322:1523-2663	1140	919.253	1703
FM179322:3156-3369	213	47.1821	79
FM179322:3365-4484	1119	898.253	388
FM179322:4515-6477	1962	1741.25	1703
FM179322:6539-9152	2613	2392.25	388
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	544
FM179322:10478-10682	204	41.5832	544
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	1703
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	544
FM179322:14018-15497	1479	1258.25	544
FM179322:15489-16509	1020	799.253	544
FM179322:16548-16743	195	36.7899	544
FM179322:16821-17079	258	77.7218	544

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。⑧上に戻って、⑨をもう一度クリックすると元に戻ります。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', and '共有データ'. The 'Tools' panel on the left lists various analysis tools like DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main panel displays a table with columns '1', '2', and '3'. The table contains data for various target IDs, including 'target\_id', 'length', 'eff\_length', and 'est\_cou'. A 'trans\_map' panel on the right shows a search for datasets, with a result for '16: Kallisto quant on data 8, data 7, and data 1 0: Abundances (tabular)'. This result is highlighted in green, and a red arrow labeled '8' points to the 'tabular' format. Another red arrow labeled '9' points to the '16: Kallisto quant on data 8, data 7, and data 1 0: Abundances (tabular)' entry.

1	2	3
target_id	length	eff_length
FM179322:0-1350	1350	1129.25
FM179322:1523-2663	1140	919.253
FM179322:3156-3369	213	47.1821
FM179322:3365-4484	1119	898.253
FM179322:4515-6477	1962	1741.25
FM179322:6539-9152	2613	2392.25
FM179322:9369-9471	102	29.4316
FM179322:9844-10075	231	59.1546
FM179322:10478-10682	204	41.5832
FM179322:11005-11437	432	221.392
FM179322:11563-11860	297	106.883
FM179322:11889-12480	591	372.299
FM179322:12571-12808	237	63.0443
FM179322:12947-13139	192	35.2526
FM179322:13201-13708	507	291.66
FM179322:14018-15497	1479	1258.25
FM179322:15489-16509	1020	799.253
FM179322:16548-16743	195	36.7899
FM179322:16821-17079	258	77.7218

trans\_map  
16 shown  
1.29 GB

16: Kallisto quant on data 8, data 7, and data 1 0: Abundances (tabular)  
2,950 lines  
フォーマット: tabular,  
データベース: ?

```
{
  "n_targets": 2949,
  "n_bootstraps": 0,
  "n_processed": 2464045,
  "n_pseudoaligned": 1758959,
  "n_unique": 1733113,
  "p_pseudoaligned": 71.4,
```

# W8-9: tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。⑧上に戻って、⑨をもう一度クリックすると元に戻ります。こんな感じ。

The screenshot shows the Galaxy web interface. On the left, there is a 'Tools' sidebar with a search bar and several tool descriptions including DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq. The main panel displays a table with columns labeled 1, 2, and 3. The table contains data for various target IDs, lengths, and efficiency values. On the right, there is a 'trans\_map' dataset list showing 16 datasets, with a size of 1.29 GB. The list includes entries like '16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)', '15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)', '14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)', and '13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)'. Each entry has icons for viewing, editing, and deleting.

1	2	3
target_id	length	eff_length
FM179322:0-1350	1350	1129.25
FM179322:1523-2663	1140	919.253
FM179322:3156-3369	213	47.1821
FM179322:3365-4484	1119	898.253
FM179322:4515-6477	1962	1741.25
FM179322:6539-9152	2613	2392.25
FM179322:9369-9471	102	29.4316
FM179322:9844-10075	231	59.1546
FM179322:10478-10682	204	41.5832
FM179322:11005-11437	432	221.392
FM179322:11563-11860	297	106.883
FM179322:11889-12480	591	372.299
FM179322:12571-12808	237	63.0443
FM179322:12947-13139	192	35.2526
FM179322:13201-13708	507	291.66
FM179322:14018-15497	1479	1258.25
FM179322:15489-16509	1020	799.253
FM179322:16548-16743	195	36.7899
FM179322:16821-17079	258	77.7218

# W8-10:HDF5

The screenshot shows the Galaxy web interface. The main panel displays a table with the following data:

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	
FM179322:14018-15497	1479	1258.25	
FM179322:15489-16509	1020	799.253	
FM179322:16548-16743	195	36.7899	
FM179322:16821-17079	258	77.7218	

The History panel on the right shows several Kallisto quantification jobs. Two entries are highlighted with red arrows and circled numbers:

- Entry 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular) (marked with ①)
- Entry 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular) (marked with ②)



参考

①HDF5はバイナリデータなので、②中央パネル上に表示させようとしても...③ダウンロードになってしまいます。

# W8-10:HDF5

The screenshot shows the Galaxy web interface. The main panel displays a table with the following data:

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	5
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	
FM179322:14018-15497	1479	1258.25	
FM179322:15489-16509	1020	799.253	

The History panel on the right shows several Kallisto quantification jobs. The most recent job is highlighted in green:

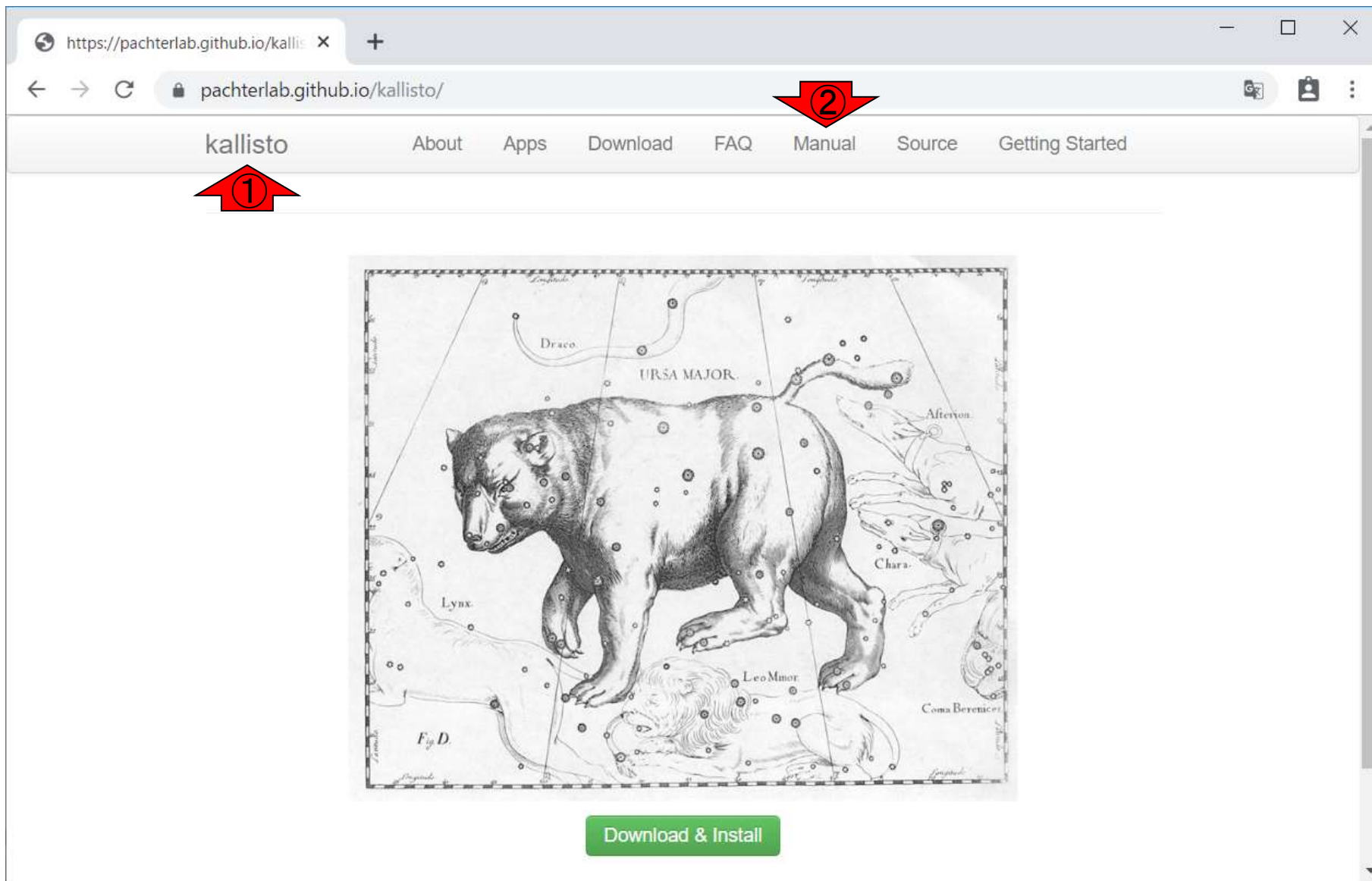
- 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)
- 15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)
- 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

A red arrow points to a file named 'Galaxy15-[Kallisto\_....h5]' in the browser's download bar, which is circled with a red '3'.

# Contents

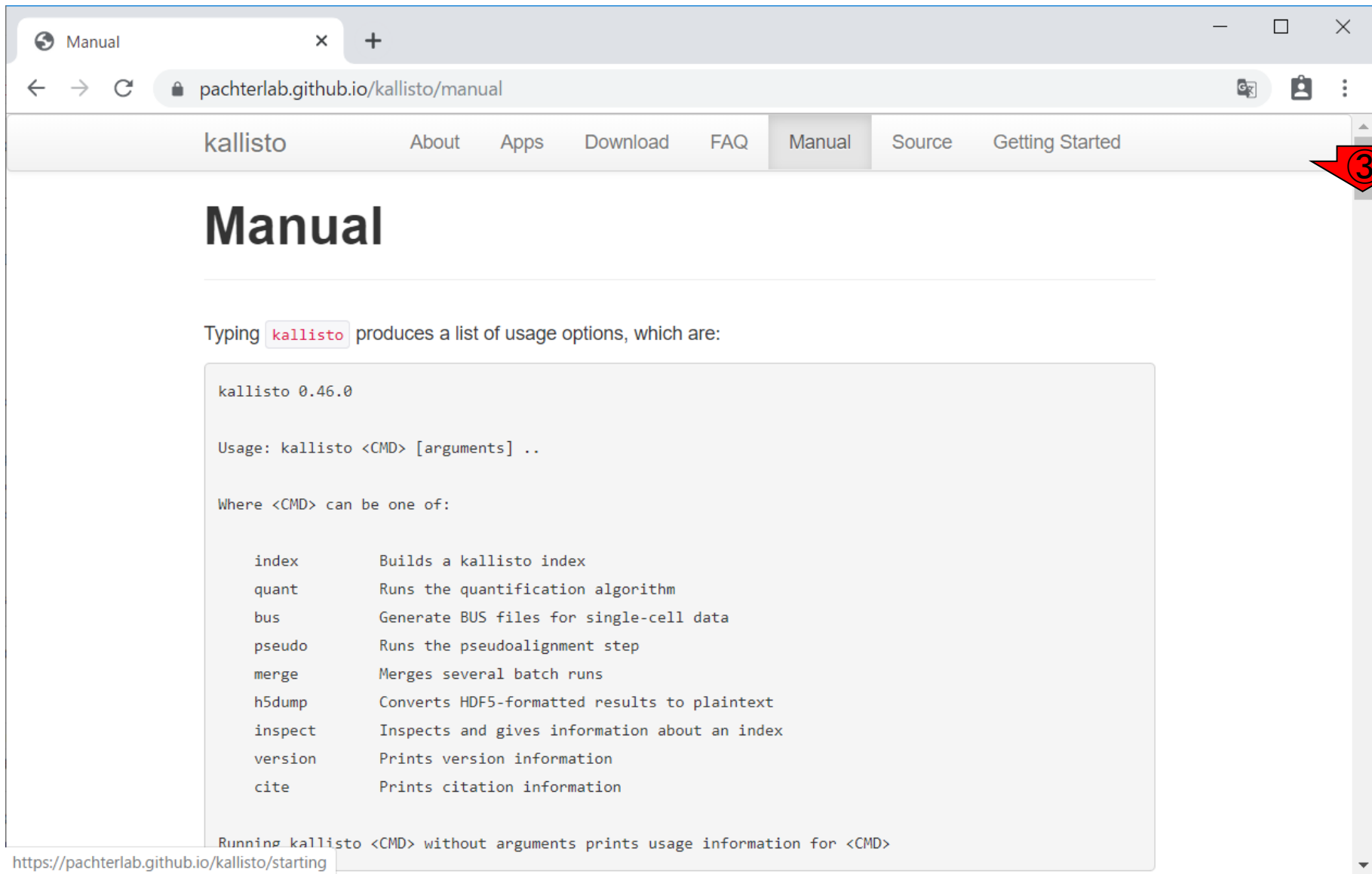
- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W9-1: マニュアル



# W9-1: マニュアル

①Kallistoのページ。②Manual。③ページ下部に移動。



Manual

kallisto About Apps Download FAQ Manual Source Getting Started

## Manual

Typing `kallisto` produces a list of usage options, which are:

```
kallisto 0.46.0

Usage: kallisto <CMD> [arguments] ..

Where <CMD> can be one of:

    index      Builds a kallisto index
    quant      Runs the quantification algorithm
    bus        Generate BUS files for single-cell data
    pseudo     Runs the pseudoalignment step
    merge      Merges several batch runs
    h5dump     Converts HDF5-formatted results to plaintext
    inspect    Inspects and gives information about an index
    version    Prints version information
    cite       Prints citation information
```

Running `kallisto <CMD>` without arguments prints usage information for `<CMD>`

<https://pachterlab.github.io/kallisto/starting>

# W9-1: マニュアル

①Kallistoのページ。②Manual。③ページ下部に移動。このあたり。④KallistoのHDF5形式ファイルには、ブートストラップの結果が格納されているようだ。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the average fragment length. Typical Illumina libraries produce fragment lengths ranging from 180–200 bp but it's best to determine this from a library quantification with an instrument such as an Agilent Bioanalyzer. For paired-end reads, the average fragment length can be directly estimated from the reads and the program will do so if `-l` is not used (this is the preferred run mode). For reads that are produced by 3'-end sequencing, the `--single-overhang` option does not discard reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because of the large amount of data that may be produced when the number of bootstrap samples is high, kallisto outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** files by default:

- **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth**
- **abundances.tsv** is a plaintext file of the abundance estimates. It does not contains bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).
- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.



# W9-1: マニュアル

①Kallistoのページ。②Manual。③ページ下部に移動。このあたり。④KallistoのHDF5形式ファイルには、ブートストラップの結果が格納されているようだ。HDF5の中身を見るためには専用のViewerが必要だが、⑤ブートストラップ結果の有効な利用法は、sleuthという発現変動解析プログラム上で用いること。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the Illumina libraries produce fragment lengths ranging from 180–200 bp but it's best to determine this from a library quantification with an instrument such as an Agilent Bioanalyzer. For paired-end reads, the average fragment length can be directly estimated from the reads and the program will do so if `-l` is not used (this is the preferred run mode). For reads that are produced by 3'-end sequencing, the `--single-overhang` option does not discard reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because of the large amount of data that may be produced when the number of bootstrap samples is high, **kallisto** outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** files by default:

- **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth**
- **abundances.tsv** is a plaintext file of the abundance estimates. It does not contains bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).
- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.

⑤

# W9-2: HDF5 binary

①を眺めることで、HDF5形式というものがよくわからなくても、バイナリファイルだということはこれを見た段階でわかる。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the average fragment length. Typical Illumina libraries produce fragment lengths ranging from 180–200 bp but it's best to determine this from a library quantification with an instrument such as an Agilent Bioanalyzer. For paired-end reads, the average fragment length can be directly estimated from the reads and the program will do so if `-l` is not used (this is the preferred run mode). For reads that are produced by 3'-end sequencing, the `--single-overhang` option does not discard reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because of the large amount of data that may be produced when the number of bootstrap samples is high, **kallisto** outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** <sup>①</sup> by default:

- **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth**
- **abundances.tsv** is a plaintext file of the abundance estimates. It does not contains bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).
- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.

# W9-2: HDF5 binary

①を眺めることで、HDF5形式というものがよくわからなくても、バイナリファイルだということはこれを見た段階でわかる。②を眺めることで、③.h5ファイルのみがbootstrap estimates情報を含んでいるのだと認識する。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the average fragment length. Typical Illumina libraries produce fragment lengths ranging from 180–200 bp but it's best to determine this from a library quantification with an instrument such as an Agilent Bioanalyzer. For paired-end reads, the average fragment length can be directly estimated from the reads and the program will do so if `-l` is not used (this is the preferred run mode). For reads that are produced by 3'-end sequencing, the `--single-overhang` option does not discard reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because of the large amount of data that may be produced when the number of bootstrap samples is high, **kallisto** outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** files by default:

- ③ **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth**
- **abundances.tsv** is a plaintext file of the abundance estimates. It does not contains bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).
- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.



# W9-2: HDF5 binary

①を眺めることで、HDF5形式というものがよくわからなくても、バイナリファイルだということはこれを見た段階でわかる。②を眺めることで、③.h5ファイルのみがbootstrap estimates情報を含んでいるのだと認識する。また、④を眺めることで、気合をいれてHDF5ファイルの中身を見なくてもよいと判断する。ましてやW8-6でNumber of bootstrap samplesをデフォルトの0で実行しているので、実質的に⑤.tsvのタブ区切りテキストファイルと同じ中身なのだろうと判断する。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the Illumina libraries produce fragment lengths ranging from 180–200 bp but quantification with an instrument such as an Agilent Bioanalyzer. For pair length can be directly estimated from the reads and the program will do so (run mode). For reads that are produced by 3'-end sequencing, the `--single` reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because bootstrap samples are produced when the number of bootstrap samples is high, **kallisto** outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** files by default:

③ **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth** ④  
⑤ **abundances.tsv** is a plaintext file of the abundance estimates. It does not contain bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` ② can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).

- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- **W10: 定量結果の解説**
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W10-1: 結果の解説

W8-9と同じく、①タブ区切りテキストファイル(.tabular)の中身を、②中央パネル上に表示させた状態。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains a 'Tools' section with a search bar and several tool descriptions: DESeq2, goseq, Kallisto pseudo - run, Kallisto quant, Salmon quant, StringTie merge transcripts, StringTie transcript assembly and quantification, and DEXSeq. The main panel displays a table with 4 columns: 1, 2, 3, and 4. The table contains 18 rows of data. A red box highlights the table. The right sidebar contains a 'History' section with a search bar and a list of datasets. A red arrow labeled '1' points to the first dataset in the history: '16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)'. Another red arrow labeled '2' points to the table in the main panel.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	
FM179322:1523-2663	1140	919.253	1
FM179322:3156-3369	213	47.1821	
FM179322:3365-4484	1119	898.253	
FM179322:4515-6477	1962	1741.25	1
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.28
FM179322:9844-10075	231	59.1546	
FM179322:10478-10682	204	41.5832	
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	1
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.02
FM179322:13201-13708	507	291.66	
FM179322:14018-15497	1479	1258.25	
FM179322:15489-16509	1020	799.253	
FM179322:16548-16743	195	36.7899	
FM179322:16821-17079	258	77.7218	

①1列目 (target\_id) は配列名。

# W10-2: 1列目

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with 'Galaxy' and 'usegalaxy.org'. Below this is a menu with options like 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. A red arrow with the number '1' points to the 'データ解析' menu item.

The main content area is divided into three sections:

- Tools:** A search bar and a list of tools including DESeq2, goseq, Kallisto pseudo, Kallisto quant, Salmon quant, StringTie merge, StringTie, and DEXSeq.
- Table:** A table with 4 columns. The first column is labeled '1' and contains 'target\_id'. The other columns are labeled '2', '3', and '4'. The table contains 16 rows of data.
- History:** A search bar and a list of history items. The first item is highlighted in green and reads: '16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)'. Other items include '15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)', '14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)', and '13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)'.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	5
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	3
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	1
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	1
FM179322:14018-15497	1479	1258.25	1
FM179322:15489-16509	1020	799.253	1
FM179322:16548-16743	195	36.7899	1
FM179322:16821-17079	258	77.7218	1

# W10-3: 2-3列目

①1列目 (target\_id) は配列名。②2列目 (length) は配列長、③3列目 (eff\_length) は有効配列長 (effective length)。この有効配列長については、Kallistoのマニュアル中に説明があります。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into 'Tools' on the left and a table of results in the center. The table has four columns: 1 (target\_id), 2 (length), 3 (eff\_length), and 4 (est\_count). Red arrows labeled ② and ③ point to the length and eff\_length columns respectively. The right sidebar shows a 'History' panel with a search bar and a list of jobs, including 'Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)' and 'Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)'.

1	2	3	4
target_id	length	eff_length	est_cou
FM179322:0-1350	1350	1129.25	5
FM179322:1523-2663	1140	919.253	17
FM179322:3156-3369	213	47.1821	3
FM179322:3365-4484	1119	898.253	3
FM179322:4515-6477	1962	1741.25	17
FM179322:6539-9152	2613	2392.25	33
FM179322:9369-9471	102	29.4316	2.284
FM179322:9844-10075	231	59.1546	5
FM179322:10478-10682	204	41.5832	1
FM179322:11005-11437	432	221.392	49.7
FM179322:11563-11860	297	106.883	17
FM179322:11889-12480	591	372.299	24
FM179322:12571-12808	237	63.0443	14
FM179322:12947-13139	192	35.2526	2.024
FM179322:13201-13708	507	291.66	1
FM179322:14018-15497	1479	1258.25	1
FM179322:15489-16509	1020	799.253	1
FM179322:16548-16743	195	36.7899	1
FM179322:16821-17079	258	77.7218	1

# W10-4: eff\_length

①1列目 (target\_id) は配列名。②2列目 (length) は配列長、③3列目 (eff\_length) は有効配列長 (effective length)。この有効配列長については、Kallistoのマニュアル中に説明があります。W9-2で示したKallistoのマニュアル中の、④のリンク先です。

have samples that span multiple FASTQ files.

In the case of single-end reads, the `-l` option must be used to specify the average fragment length. Typical Illumina libraries produce fragment lengths ranging from 180–200 bp but it's best to determine this from a library quantification with an instrument such as an Agilent Bioanalyzer. For paired-end reads, the average fragment length can be directly estimated from the reads and the program will do so if `-l` is not used (this is the preferred run mode). For reads that are produced by 3'-end sequencing, the `--single-overhang` option does not discard reads where the expected fragment size goes beyond the transcript start.

The number of bootstrap samples is specified using `-b`. Note that because of the large amount of data that may be produced when the number of bootstrap samples is high, **kallisto** outputs bootstrap results in HDF5 format. The `h5dump` command can be used afterwards to convert this output to plaintext, however most convenient is to analyze bootstrap results with **sleuth**.

`kallisto quant` produces three **output** files by default:

- **abundances.h5** is a HDF5 binary file containing run info, abundance estimates, bootstrap estimates, and transcript length information length. This file can be read in by **sleuth**
- **abundances.tsv** is a plaintext file of the abundance estimates. It does not contains bootstrap estimates. Please use the `--plaintext` mode to output plaintext abundance estimates. Alternatively, `kallisto h5dump` can be used to output an HDF5 file to plaintext. The first line contains a header for each column, including [estimated counts](#), [TPM](#), [effective length](#).
- **run\_info.json** is a json file containing information about the run

## Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordingly.
- `-t, --threads` specifies the number of threads to be used both for pseudoalignment and running bootstrap. The default value is 1 thread, specifying more than the number of bootstraps or the number of cores on your machine has no additional effect.

④

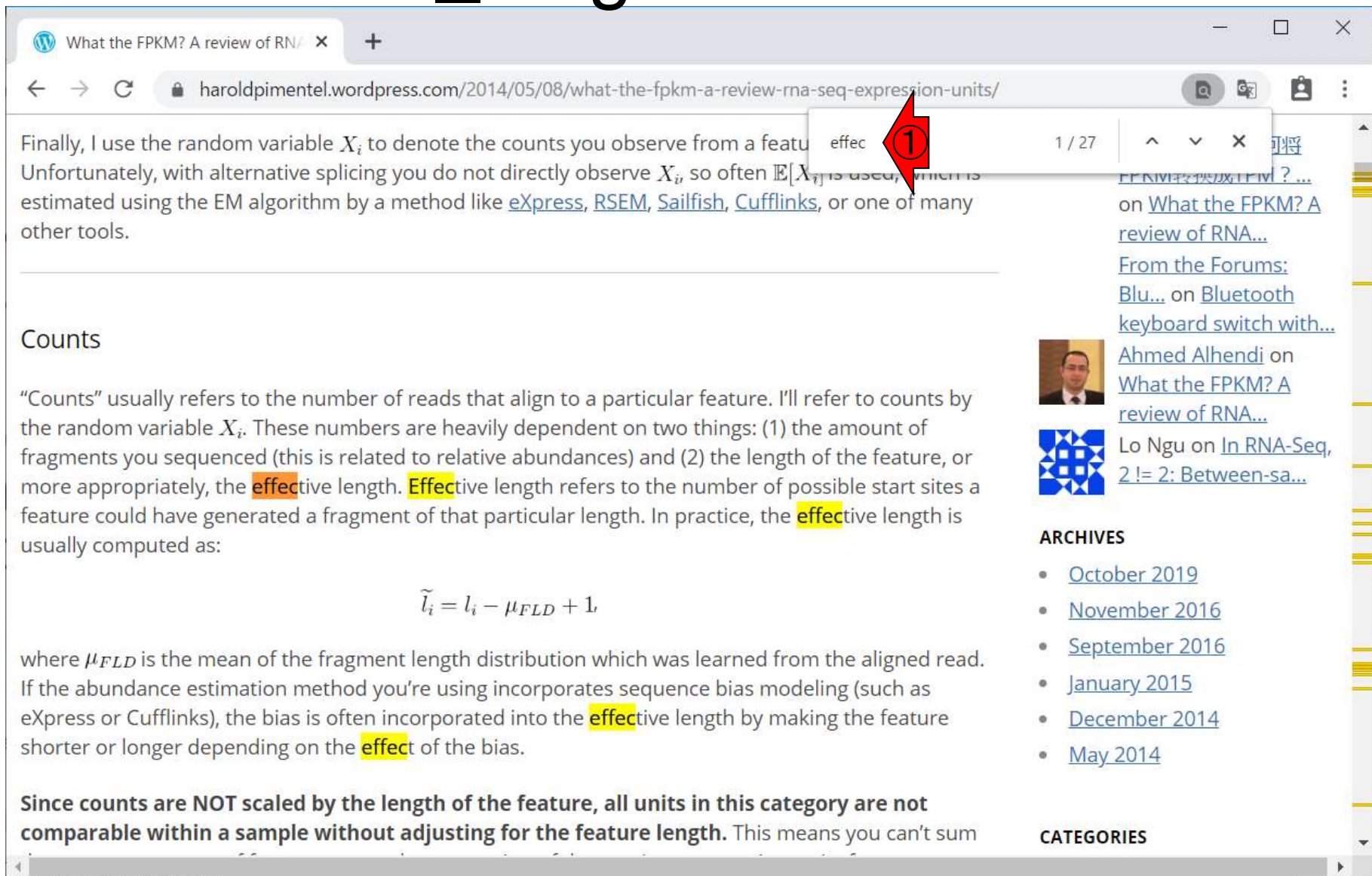
# W10-4: eff\_length

The screenshot shows a web browser window displaying a WordPress blog post. The browser's address bar shows the URL: `haroldpimentel.wordpress.com/2014/05/08/what-the-fpkm-a-review-rna-seq-expression-units/`. The page title is "The farrago" with a subtitle "A place for my thoughts on comp bio, statistics, cycling([a-zA-Z]+)". Below the title is a search bar and a "RECENT POSTS" section. The main content of the post is titled "What the FPKM? A review of RNA-Seq expression units" and includes a paragraph: "This post covers the units used in RNA-Seq that are, unfortunately, often misused and misunderstood. I'll try to clear up a bit of the confusion here." and another paragraph: "The first thing one should remember is that without between sample normalization (a topic for a later post), **NONE of these units are comparable across experiments. This is a result of RNA-Seq being a relative measurement, not an absolute one.**"

# W10-4: eff\_length

(リンクをクリックすると)こんな感じになります。

①「**efec**」でページ内検索したところ。



What the FPKM? A review of RNA-seq expression units

Finally, I use the random variable  $X_i$  to denote the counts you observe from a feature  $i$ . Unfortunately, with alternative splicing you do not directly observe  $X_i$ , so often  $\tilde{X}_i$  is used, which is estimated using the EM algorithm by a method like [eXpress](#), [RSEM](#), [Sailfish](#), [Cufflinks](#), or one of many other tools.

## Counts

“Counts” usually refers to the number of reads that align to a particular feature. I’ll refer to counts by the random variable  $X_i$ . These numbers are heavily dependent on two things: (1) the amount of fragments you sequenced (this is related to relative abundances) and (2) the length of the feature, or more appropriately, the **effective length**. **Effective length** refers to the number of possible start sites a feature could have generated a fragment of that particular length. In practice, the **effective length** is usually computed as:

$$\tilde{l}_i = l_i - \mu_{FLD} + 1,$$

where  $\mu_{FLD}$  is the mean of the fragment length distribution which was learned from the aligned read. If the abundance estimation method you’re using incorporates sequence bias modeling (such as [eXpress](#) or [Cufflinks](#)), the bias is often incorporated into the **effective length** by making the feature shorter or longer depending on the **effect** of the bias.

**Since counts are NOT scaled by the length of the feature, all units in this category are not comparable within a sample without adjusting for the feature length.** This means you can’t sum

Search results for "efec": 1 / 27

ARCHIVES

- [October 2019](#)
- [November 2016](#)
- [September 2016](#)
- [January 2015](#)
- [December 2014](#)
- [May 2014](#)

CATEGORIES



# W10-4: eff\_length

(リンクをクリックすると)こんな感じになります。  
①「efec」でページ内検索したところ。effective lengthに関する記述が②など赤下線部分に確かにあります。

What the FPKM? A review of RNA-seq expression units

Finally, I use the random variable  $X_i$  to denote the counts you observe from a feature. Unfortunately, with alternative splicing you do not directly observe  $X_i$ , so often  $\mathbb{E}[X_i]$  is used, which is estimated using the EM algorithm by a method like [eXpress](#), [RSEM](#), [Sailfish](#), [Cufflinks](#), or one of many other tools.

## Counts

“Counts” usually refers to the number of reads that align to a particular feature. I’ll refer to counts by the random variable  $X_i$ . These numbers are heavily dependent on two things: (1) the amount of fragments you sequenced (this is related to relative abundances) and (2) the length of the feature, or more appropriately, the effective length. Effective length refers to the number of possible start sites a feature could have generated a fragment of that particular length. In practice, the effective length is usually computed as:

$$\tilde{l}_i = l_i - \mu_{FLD} + 1,$$

where  $\mu_{FLD}$  is the mean of the fragment length distribution which was learned from the aligned read. If the abundance estimation method you’re using incorporates sequence bias modeling (such as [eXpress](#) or [Cufflinks](#)), the bias is often incorporated into the effective length by making the feature shorter or longer depending on the effect of the bias.

**Since counts are NOT scaled by the length of the feature, all units in this category are not comparable within a sample without adjusting for the feature length.** This means you can’t sum

ARCHIVES

- [October 2019](#)
- [November 2016](#)
- [September 2016](#)
- [January 2015](#)
- [December 2014](#)
- [May 2014](#)

CATEGORIES

# W10-4: eff\_length

(リンクをクリックすると)こんな感じになります。  
①「efec」でページ内検索したところ。effective lengthに関する記述が②など赤下線部分に確かにあります。この②effective lengthは、③実際の配列長と、④マップされたリードの平均フラグメント長から算出できるようです。

What the FPKM? A review of RNA-Seq  
haroldpimentel.wordpress.com/2014/05/08/what-the-fpkm-a-review-rna-

Finally, I use the random variable  $X_i$  to denote the counts you observe from a feature. Unfortunately, with alternative splicing you do not directly observe  $X_i$ , so often  $\mathbb{E}[X_i]$  is used, which is estimated using the EM algorithm by a method like eXpress, RSEM, Sailfish, Cufflinks, or one of many other tools.

## Counts

“Counts” usually refers to the number of reads that align to a particular feature. I’ll refer to counts by the random variable  $X_i$ . These numbers are heavily dependent on two things: (1) the amount of fragments you sequenced (this is related to relative abundances) and (2) the length of the feature, or more appropriately, the effective length. Effective length refers to the number of possible start sites a feature could have generated a fragment of that particular length. In practice, the effective length is usually computed as:

$$\tilde{l}_i = l_i - \mu_{FLD} + 1$$

where  $\mu_{FLD}$  is the mean of the fragment length distribution which was learned from the aligned read. If the abundance estimation method you’re using incorporates sequence bias modeling (such as eXpress or Cufflinks), the bias is often incorporated into the effective length by making the feature shorter or longer depending on the effect of the bias.

**Since counts are NOT scaled by the length of the feature, all units in this category are not comparable within a sample without adjusting for the feature length.** This means you can’t sum

What the FPKM? A review of RNA-Seq  
From the Forums:  
Ahmed Alhendi on What the FPKM? A review of RNA-Seq  
Lo Ngu on In RNA-Seq, 2 != 2: Between-sa...

## ARCHIVES

- October 2019
- November 2016
- September 2016
- January 2015
- December 2014
- May 2014

## CATEGORIES

# W10-5: 4列目

①4列目 (est\_counts) は、カウント数の推定値 (estimated counts)。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. A red arrow points to the '共有データ' dropdown menu, which is labeled with a circled '1'. Below the navigation bar, there is a table with columns labeled 2, 3, 4, and 5. The table headers are length, eff\_length, est\_counts, and tpm. The table contains 20 rows of data. The 'History' panel on the right shows a list of jobs, including 'Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)' and 'Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)'. The 'Tools' panel on the left shows a search bar and a list of tools including DESeq2, goseq, Kallisto pseudo-run, Kallisto quant, Salmon quant, StringTie merge transcripts, StringTie transcript assembly and quantification, and DEXSeq.

	2	3	4	5
	length	eff_length	est_counts	tpm
350	1350	1129.25	544	79.711
13-2663	1140	919.253	1703	306.542
16-3369	213	47.1821	79	277.051
15-4484	1119	898.253	388	71.4732
15-6477	1962	1741.25	1760	167.248
19-9152	2613	2392.25	3313	229.153
19-9471	102	29.4316	2.28423	12.8421
14-10075	231	59.1546	571	1597.19
178-10682	204	41.5832	41	163.146
105-11437	432	221.392	49.7158	37.1572
163-11860	297	106.883	1738	2690.61
189-12480	591	372.299	2430	1080
171-12808	237	63.0443	1420	3726.95
147-13139	192	35.2526	2.02408	9.50051
101-13708	507	291.66	37	20.9911
118-15497	1479	1258.25	26	3.41913
189-16509	1020	799.253	31	6.41783
148-16743	195	36.7899	2	8.99521
121-17079	258	77.7218	83	176.704

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W11-1:5列目

①5列目 (tpm) は、いわゆるTPM (Transcripts per million) 値と呼ばれるもの。

The screenshot shows the Galaxy web interface with a table of transcript data. The table has columns for transcript ID, length, eff\_length, est\_counts, and tpm. A red arrow points to the 'tpm' column header. The 'History' panel on the right shows a list of jobs, with the most recent ones highlighted in green.

	2	3	4	5
	length	eff_length	est_counts	tpm
350	1350	1129.25	544	79.711
13-2663	1140	919.253	1703	306.542
16-3369	213	47.1821	79	277.051
15-4484	1119	898.253	388	71.4732
15-6477	1962	1741.25	1760	167.248
19-9152	2613	2392.25	3313	229.153
19-9471	102	29.4316	2.28423	12.8421
14-10075	231	59.1546	571	1597.19
178-10682	204	41.5832	41	163.146
105-11437	432	221.392	49.7158	37.1572
163-11860	297	106.883	1738	2690.61
189-12480	591	372.299	2430	1080
171-12808	237	63.0443	1420	3726.95
147-13139	192	35.2526	2.02408	9.50051
101-13708	507	291.66	37	20.9911
118-15497	1479	1258.25	26	3.41913
189-16509	1020	799.253	31	6.41783
148-16743	195	36.7899	2	8.99521
121-17079	258	77.7218	83	176.704

History:

- 16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)
- 15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)
- 14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)
- 13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)

# W11-2: マニュアル

W10-4のKallistoマニュアルの少し下のほうに移動して、①TPMの説明が見える状態。このマニュアル中では、②のように説明されている。

What the FPKM? A review of RNA-seq expression units

haroldpimentel.wordpress.com/2014/05/08/what-the-fpkm-a-review-rna-seq-expression-units/

Again, the methods in this section allow for comparison of features with different length WITHIN a sample but not BETWEEN samples.

TPM

①

Transcripts per million (TPM) is a measurement of the proportion of transcripts in your pool of RNA.

②

Since we are interested in taking the length into consideration, a natural measurement is the rate, counts per base ( $X_i/\tilde{l}_i$ ). As you might immediately notice, this number is also dependent on the total number of fragments sequenced. To adjust for this, simply divide by the sum of all rates and this gives the proportion of transcripts  $i$  in your sample. After you compute that, you simply scale by one million because the proportion is often very small and a pain to deal with. In math:

$$\text{TPM}_i = \frac{X_i}{\tilde{l}_i} \cdot \left( \frac{1}{\sum_j \frac{X_j}{\tilde{l}_j}} \right) \cdot 10^6.$$

TPM has a very nice interpretation when you're looking at transcript abundances. As the name suggests, the interpretation is that if you were to sequence one million full length transcripts, TPM is the number of transcripts you would have seen of type  $i$ , given the abundances of the other transcripts in your sample. The last "given" part is important. The denominator is going to be different between experiments, and thus is also sample dependent which is why you cannot directly compare TPM between samples. While this is true, TPM is probably the most stable unit across experiments, though you still shouldn't compare it across experiments.

# W11-3: 2行目の転写物

①FM179322:1523-2663の、②カウント数は79で、③TPM値は306.541。

	2	3	4	5
	length	eff_length	est_counts	tpm
350	1350	1129.25	544	79.711
13-2663	1140	919.253	79	306.542
16-3369	213	47.1821	79	277.051
15-4484	1119	898.253	388	71.4732
15-6477	1962	1741.25	1760	167.248
19-9152	2613	2392.25	3313	229.153
19-9471	102	29.4316	2.28423	12.8421
14-10075	231	59.1546	571	1597.19
178-10682	204	41.5832	41	163.146
105-11437	432	221.392	49.7158	37.1572
163-11860	297	106.883	1738	2690.61
189-12480	591	372.299	2430	1080
171-12808	237	63.0443	1420	3726.95
147-13139	192	35.2526	2.02408	9.50051
101-13708	507	291.66	37	20.9911
118-15497	1479	1258.25	26	3.41913
189-16509	1020	799.253	31	6.41783
148-16743	195	36.7899	2	8.99521
121-17079	258	77.7218	83	176.704

# W11-4: 4列目

①4列目 (est\_counts) の数値を全部足すと、1,758,959となる。100万よりも1.759倍ほど大きい数字であるということがポイント。

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

	2	3	4	5
	length	eff_length	est_counts	tpm
350	1350	1129.25	544	79.711
13-2663	1140	919.253	1703	306.542
16-3369	213	47.1821	79	277.051
15-4484	1119	898.253	388	71.4732
15-6477	1962	1741.25	1760	167.248
19-9152	2613	2392.25	3313	229.153
19-9471	102	29.4316	2.28423	12.8421
14-10075	231	59.1546	571	1597.19
178-10682	204	41.5832	41	163.146
105-11437	432	221.392	49.7158	37.1572
163-11860	297	106.883	1738	2690.61
189-12480	591	372.299	2430	1080
171-12808	237	63.0443	1420	3726.95
147-13139	192	35.2526	2.02408	9.50051
101-13708	507	291.66	37	20.9911
118-15497	1479	1258.25	26	3.41913
189-16509	1020	799.253	31	6.41783
148-16743	195	36.7899	2	8.99521
121-17079	258	77.7218	83	176.704

History

search datasets

trans\_map

16 shown

1.29 GB

16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)

14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)



# W11-5:3列目

①4列目 (est\_counts) の数値を全部足すと、1,758,959となる。100万よりも1.76倍ほど大きい数字であるということがポイント。

Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化ツール 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

	2	3	4	5
	length	eff_length	est_counts	tpm
3-2663	1350	1129.25	544	79.711
6-3369	213	47.1821	79	277.051
5-4484	1119	898.253	388	71.4732
5-6477	1962	1741.25	1760	167.248
9-9152	2613	2392.25	3313	229.153
9-9471	102	29.4316	2.28423	12.8421
4-10075	231	59.1546	571	1597.19
78-10682	204	41.5832	41	163.146
05-11437	432	221.392	49.7158	37.1572
63-11860	297	106.883	1738	2690.61
89-12480	591	372.299	2430	1080
71-12808	237	63.0443	1420	3726.95
47-13139	192	35.2526	2.02408	9.50051
01-13708	507	291.66	37	20.9911
18-15497	1479	1258.25	26	3.41913
89-16509	1020	799.253	31	6.41783
48-16743	195	36.7899	2	8.99521
21-17079	258	77.7218	83	176.704

History

search datasets

trans\_map

16 shown

1.29 GB

16: Kallisto quant on data 8, data 7, and data 10: Abundances (tabular)

15: Kallisto quant on data 8, data 7, and data 10: Abundances (HDF5)

14: Kallisto quant on data 6, data 5, and data 10: Abundances (tabular)

13: Kallisto quant on data 6, data 5, and data 10: Abundances (HDF5)

# W11-6: ヒストリー16

①ヒストリー16の、②FPKM値の、③総和は3,435,847.75。④CPK値の総和は6,043,515.33。

	A	B	C	D	E	F	G	H	I
1	target_id	length	eff_length	est_counts	tpm	cpm	cpk	fpkm	tpm
2932	FM179322:2992403-2993243	840	619.32	166.00	44.35	94.37	268.04	152.38	44.35
2933	FM179322:2993351-2993537	186	32.20	66.00	339.16	37.52	2049.71	1165.30	339.16
2934	FM179322:2993964-2995050	1086	865.25	297.00	56.80	168.85	343.25	195.15	56.80
2935	FM179322:2995287-2995473	186	32.20	9.00	46.25	5.12	279.51	158.90	46.25
2936	FM179322:2995426-2995603	177	28.06	11.00	64.86	6.25	392.01	222.86	64.86
2937	FM179322:2995992-2997192	1200	979.25	334.00	56.44	189.89	341.08	193.91	56.44
2938	FM179322:2997188-2997677	489	274.64	59.00	35.55	33.54	214.83	122.14	35.55
2939	FM179322:2997678-2998443	765	544.44	125.00	37.99	71.06	229.60	130.53	37.99
2940	FM179322:2998648-2998855	207	43.49	192.00	730.43	109.16	4414.35	2509.64	730.43
2941	FM179322:2998994-2999216	222	52.85	5.00	15.65	2.84	94.61	53.79	15.65
2942	FM179322:2999512-3001132	1620	1399.25	105.00	12.42	59.69	75.04	42.66	12.42
2943	FM179322:3001258-3003160	1902	1681.25	630.00	62.00	358.17	374.72	213.04	62.00
2944	FM179322:3003200-3004589	1389	1168.25	172.00	24.36	97.79	147.23	83.70	24.36
2945	FM179322:3004646-3004901	255	75.57	11.75	25.73	6.68	155.52	88.42	25.73
2946	FM179322:3004939-3005707	768	547.44	678.00	204.93	385.46	1238.50	704.11	204.93
2947	FM179322:3005724-3006561	837	616.32	500.00	134.24	284.26	811.27	461.22	134.24
2948	FM179322:3006590-3006947	357	155.58	167.00	177.61	94.94	1073.38	610.23	177.61
2949	FM179322:3007541-3007682	141	17.48	1698.00	16071.30	965.34	97126.81	55218.35	16071.24
2950	FM179322:3007920-3009357	1437	1216.25	31179.00	4241.79	17725.83	25635.35	14574.16	4241.80
2951									
2952	sum	871.71	662.98	1758959.00	999999.89	1000000.00	6043515.33	3435847.75	1000000.00

# W11-7: ヒストリー14

①ヒストリー14の、②FPKM値の、③総和は4,005,980.92。④CPK値の総和は4,714,290.42。

	A	B	C	D	E	F	G	H	I
1	target_id	length	eff_length	est_counts	tpm	cpm	cpk	fpkm	tpm
2932	FM179322:2992403-2993243	840	634.00	29.00	9.70	24.64	45.74	38.87	9.70
2933	FM179322:2993351-2993537	186	34.20	14.00	86.85	11.90	409.41	347.90	86.85
2934	FM179322:2993964-2995050	1086	879.93	147.00	35.44	124.91	167.06	141.96	35.44
2935	FM179322:2995287-2995473	186	34.20	15.00	93.05	12.75	438.66	372.75	93.05
2936	FM179322:2995426-2995603	177	29.35	14.00	101.17	11.90	476.95	405.29	101.17
2937	FM179322:2995992-2997192	1200	993.93	91.00	19.42	77.33	91.56	77.80	19.42
2938	FM179322:2997188-2997677	489	286.03	33.00	24.47	28.04	115.37	98.04	24.47
2939	FM179322:2997678-2998443	765	559.00	80.00	30.36	67.98	143.11	121.61	30.36
2940	FM179322:2998648-2998855	207	46.24	62.00	284.41	52.68	1340.77	1139.32	284.40
2941	FM179322:2998994-2999216	222	55.62	23.00	87.72	19.54	413.51	351.38	87.72
2942	FM179322:2999512-3001132	1620	1413.93	113.00	16.95	96.02	79.92	67.91	16.95
2943	FM179322:3001258-3003160	1902	1695.93	491.00	61.41	417.23	289.52	246.02	61.41
2944	FM179322:3003200-3004589	1389	1182.93	274.00	49.13	232.83	231.63	196.83	49.13
2945	FM179322:3004646-3004901	255	78.80	46.00	123.83	39.09	583.75	496.04	123.82
2946	FM179322:3004939-3005707	768	562.00	383.00	144.56	325.46	681.50	579.11	144.56
2947	FM179322:3005724-3006561	837	631.00	301.00	101.19	255.78	477.02	405.35	101.19
2948	FM179322:3006590-3006947	357	162.53	147.00	191.85	124.91	904.45	768.56	191.85
2949	FM179322:3007541-3007682	141	15.78	980.00	13177.70	832.76	62123.22	52789.37	13177.64
2950	FM179322:3007920-3009357	1437	1230.93	5610.00	966.75	4767.11	4557.53	3872.77	966.75
2951									
2952	sum	871.71	674.54	1176813.00	#####	1000000	4714290.42	4005980.92	#####

# W11-8: ヒストリー12

①ヒストリー12の、②FPKM値の、③総和は5,275,181.61。④CPK値の総和は5,659,410.02。

	A	B	C	D	E	F	G	H	I
1	target_id	length	eff_length	est_counts	tpm	cpm	cpk	fpkm	tpm
2932	FM179322:2992403-2993243	840	625.24	15.00	4.24	13.98	23.99	22.36	4.24
2933	FM179322:2993351-2993537	186	34.52	9.00	46.07	8.39	260.70	243.00	46.07
2934	FM179322:2993964-2995050	1086	871.18	110.00	22.31	102.53	126.27	117.69	22.31
2935	FM179322:2995287-2995473	186	34.52	13.00	66.54	12.12	376.57	351.00	66.54
2936	FM179322:2995426-2995603	177	29.91	13.00	76.80	12.12	434.64	405.13	76.80
2937	FM179322:2995992-2997192	1200	985.18	87.00	15.60	81.09	88.31	82.31	15.60
2938	FM179322:2997188-2997677	489	277.54	31.00	19.74	28.90	111.70	104.11	19.74
2939	FM179322:2997678-2998443	765	550.48	78.00	25.04	72.70	141.70	132.08	25.04
2940	FM179322:2998648-2998855	207	46.26	217.00	828.91	202.27	4691.15	4372.66	828.91
2941	FM179322:2998994-2999216	222	55.27	14.00	44.76	13.05	253.29	236.09	44.76
2942	FM179322:2999512-3001132	1620	1405.18	85.00	10.69	79.23	60.49	56.38	10.69
2943	FM179322:3001258-3003160	1902	1687.18	283.00	29.64	263.79	167.74	156.35	29.64
2944	FM179322:3003200-3004589	1389	1174.18	131.00	19.71	122.11	111.57	103.99	19.71
2945	FM179322:3004646-3004901	255	77.17	26.00	59.53	24.23	336.90	314.03	59.53
2946	FM179322:3004939-3005707	768	553.48	720.00	229.86	671.12	1300.87	1212.55	229.86
2947	FM179322:3005724-3006561	837	622.24	403.00	114.44	375.64	647.66	603.69	114.44
2948	FM179322:3006590-3006947	357	156.86	261.00	294.01	243.28	1663.95	1550.98	294.01
2949	FM179322:3007541-3007682	141	17.50	4387.00	44295.40	4089.16	250685.71	233666.17	44295.38
2950	FM179322:3007920-3009357	1437	1222.18	8464.00	1223.69	7889.36	6925.33	6455.16	1223.68
2951									
2952	sum	871.71	667.43	1072837.00	999999.89	1000000.00	5659410.02	5275181.61	1000000.00

# Contents

- W1: スタート地点
- W2: 新規ヒストリー
- W3: データのコピー
- W4: 解析準備完了
- W5: GFFの前処理
- W6: Extract features
- W7: bedtools GetFastaBed
- W8: Kallisto quant
- W9: Kallistoのマニュアル
- W10: 定量結果の解説
- W11: CPM, CPK, FPKM, and TPM
- W12: 全サンプルでKallisto quantを実行

# W12-1:9サンプル

①Kallisto\_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene\_IDと対応づけている。

自動保存  Kallisto\_9samples.xlsx - Excel 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有 コメント

A1   fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1					pH4.5_1h			pH4.5_24h			pH7_CCG			Sum	
2	Serial	gene ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0	0	0	0	0	0	0	0	0	0
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92

Sheet1 100%

# W12-1:9サンプル

①Kallisto\_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene\_IDと対応づけている。④の列は、②の領域で座標順に並べようとしてもうまくいかないの、それを実現したいときに用いるための情報。

自動保存 ● オフ Kallisto\_9samples.xlsx

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検

④

②

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1					pH4.5_1h			pH4.5_24h			pH7_CCG			Sum		
2	Serial	gene ID	target id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG	
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659	
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210	
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259	
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847	
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0	0	0	0	0	0	0	0	0	0	
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505	
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130	
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191	
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081	
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154	
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180	
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14	
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410	
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116	
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151	
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406	
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997	
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435	
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11	
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92	

Sheet1

100%

# W12-1:9サンプル

①Kallisto\_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene\_IDと対応づけている。④の列は、②の領域で座標順に並べようとしてもうまくいかないので、それを実現したいときに用いるための情報。⑤の部分のみ座標が順番に並んでいないことがわかる。これはカウントデータをマージした後に、②と③を対応づけていく過程で、「あれ、なんかおかしいぞ」と気づき、アノテーションファイル中で、⑤の遺伝子群に対して、②と③の情報の対応関係を一つ一つ丁寧にチェックして、④のSerial番号を手作業で割り当てるなどしてこのファイルを完成させている。このあたりの作業はわりと泥臭いです。

自動保存 ● オフ Kallisto\_9samples.xlsx

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1					pH4.5_1h																					
2	Serial	gene ID	target id	rep1	rep2	rep3	rep4	rep5	rep6	rep7	rep8	rep9	rep10	rep11	rep12	rep13	rep14	rep15	rep16	rep17	rep18	rep19	rep20	rep21	rep22	rep23
3	499	EBG00001128470	FM179322:506641-506842	21	262	275																				
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124																				
5	1487	EBG00001128500	FM179322:1489175-1489541	79	20948	28807																				
6	500	EBG00001128509	FM179322:507096-507297	16	333	337																				
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0																				
8	1	LGG_00001	FM179322:0-1350	105	328	447																				
9	2	LGG_00002	FM179322:1523-2663	230	880	1115																				
10	3	LGG_00003	FM179322:3156-3369	4	61	82																				
11	4	LGG_00004	FM179322:3365-4484	50	234	316																				
12	5	LGG_00005	FM179322:4515-6477	150	370	514																				
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365																				
14	7	LGG_00007	FM179322:9369-9471	1	8	13																				
15	8	LGG_00008	FM179322:9844-10075	11	72	114																				
16	9	LGG_00009	FM179322:10478-10682	4	23	21																				
17	10	LGG_00010	FM179322:11005-11437	4	21	28																				
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011																				
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686																				
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610																				
21	14	LGG_00014	FM179322:12947-13139	1	8	8																				
22	15	LGG_00015	FM179322:13201-13708	8	100	108																				

Sheet1

100%



# W12-1:9サンプル

赤枠内が9サンプル分のカウント情報をマージした結果。黄色部分が第15回のGalaxy上で実際に行ったものに対応する。⑥ヒストリー16、⑦ヒストリー14、⑧ヒストリー12。

自動保存  Kallisto\_9samples.xlsx - Excel 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有 コメント

A1

	A	B	C	D			E			F			G			H			I			J			K			L			M	N	O
				pH4.5_1h			pH4.5_24h			pH7_CCG			Sum																				
	Serial	gene ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG															
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659																		
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210																		
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259																		
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847																		
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0	0	0	0	0	0	0	0	0	0																		
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505																		
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130																		
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191																		
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081																		
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154																		
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180																		
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14																		
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410																		
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116																		
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151																		
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406																		
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997																		
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435																		
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11																		
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92																		

Sheet1 100%

# W12-1:9サンプル

赤枠内が9サンプル分のカウント情報をマージした結果。黄色部分が第15回のGalaxy上で実際に行ったものに対応する。⑥ヒストリー16、⑦ヒストリー14、⑧ヒストリー12。例えば⑥のヒストリー16のカウント情報はW10-5でも見られる。但し、⑨アノテーションファイル中の並びはここからなので、⑩の部分から見ていくと、確かに一致していると納得できる。

自動保存 ● オフ Kallisto\_9samples.xlsx

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1					pH4.5_1h			pH4.5_24h			pH7_CCG			Sum		
2	Serial	gene ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG	
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659	
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210	
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259	
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847	
7	1394	EBG00001128529	FM179322:1393560-1393560	0	0	0	0	0	0	0	0	0	0	0	0	
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505	
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130	
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191	
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081	
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154	
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180	
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14	
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410	
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116	
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151	
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406	
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997	
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435	
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11	
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92	

Sheet1

100%

①が反復データをマージした結果。

# W12-2: Table1の前説

The screenshot shows an Excel spreadsheet with the following data table:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1					pH4.5_1h			pH4.5_24h			pH7_CCG		Sum		
2	Serial	gene ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0	0	0	0	0	0	0	0	0	0
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92

# W12-2: Table1の前説

①が反復データをマージした結果。例えば②558という数値は、③21 + 262 + 275の和。原著論文では平均と書いてありましたが、和をとったほうがその値に近いこともあり、気分的にそうしています。また、一般には和をとるものだと思います。①の情報が表1で示されています。

自動保存 ● オ Kallisto\_9samples.x

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1					pH4.5_1h			pH4.5_24h			pH7_CCG			Sum		
2	Serial	gene ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG	
3	499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659	
4	2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210	
5	1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259	
6	500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847	
7	1394	EBG00001128529	FM179322:1393560-1393708	0	0	0	0	0	0	0	0	0	0	0	0	
8	1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505	
9	2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130	
10	3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191	
11	4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081	
12	5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154	
13	6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180	
14	7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14	
15	8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410	
16	9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116	
17	10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151	
18	11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406	
19	12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997	
20	13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435	
21	14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11	
22	15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92	

Sheet1

100%