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

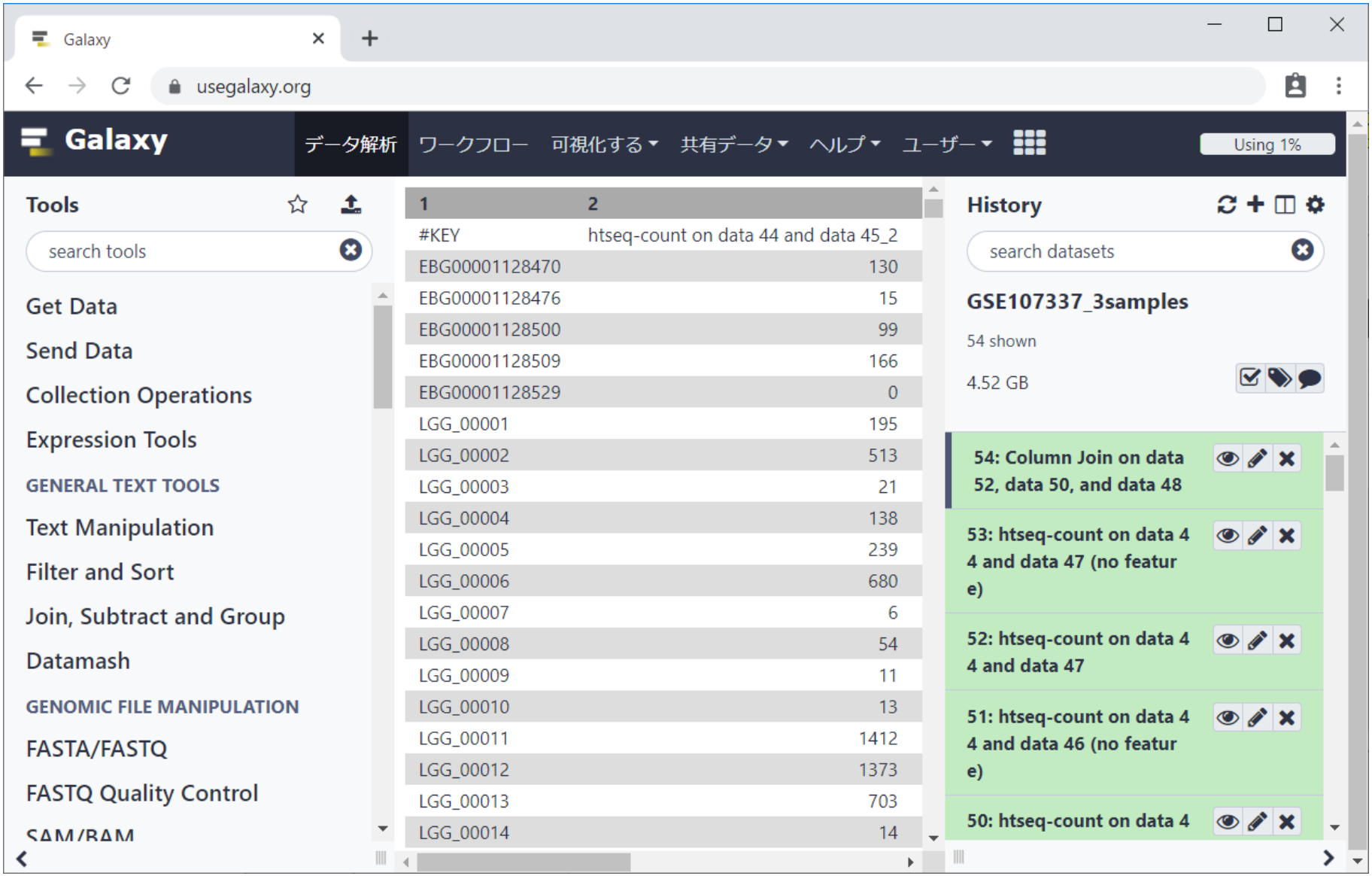
寺田朋子、清水謙多郎、門田幸二*

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

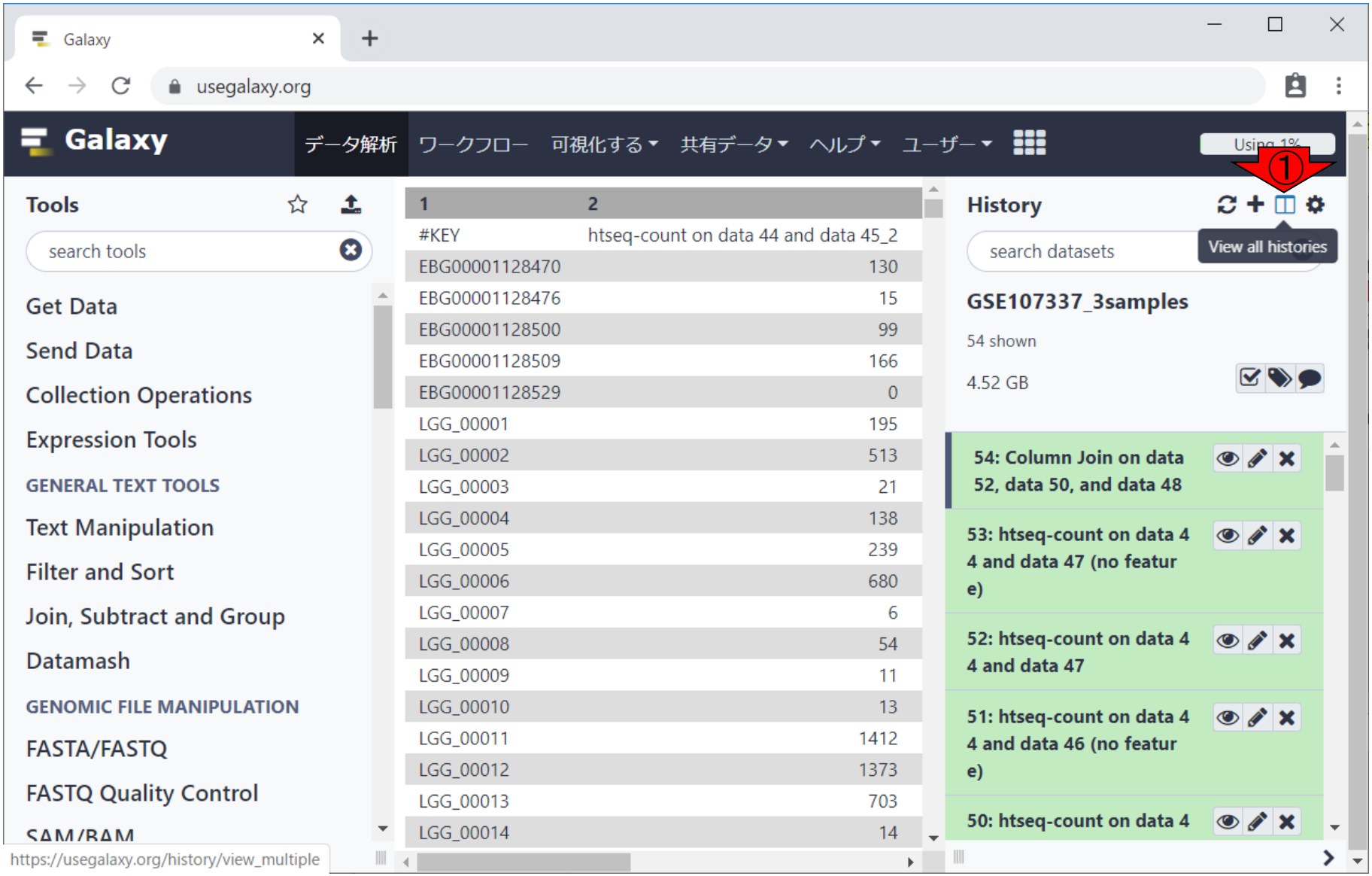
前回（第14回）のW9-7と同じ画面です。

W1：スタート地点



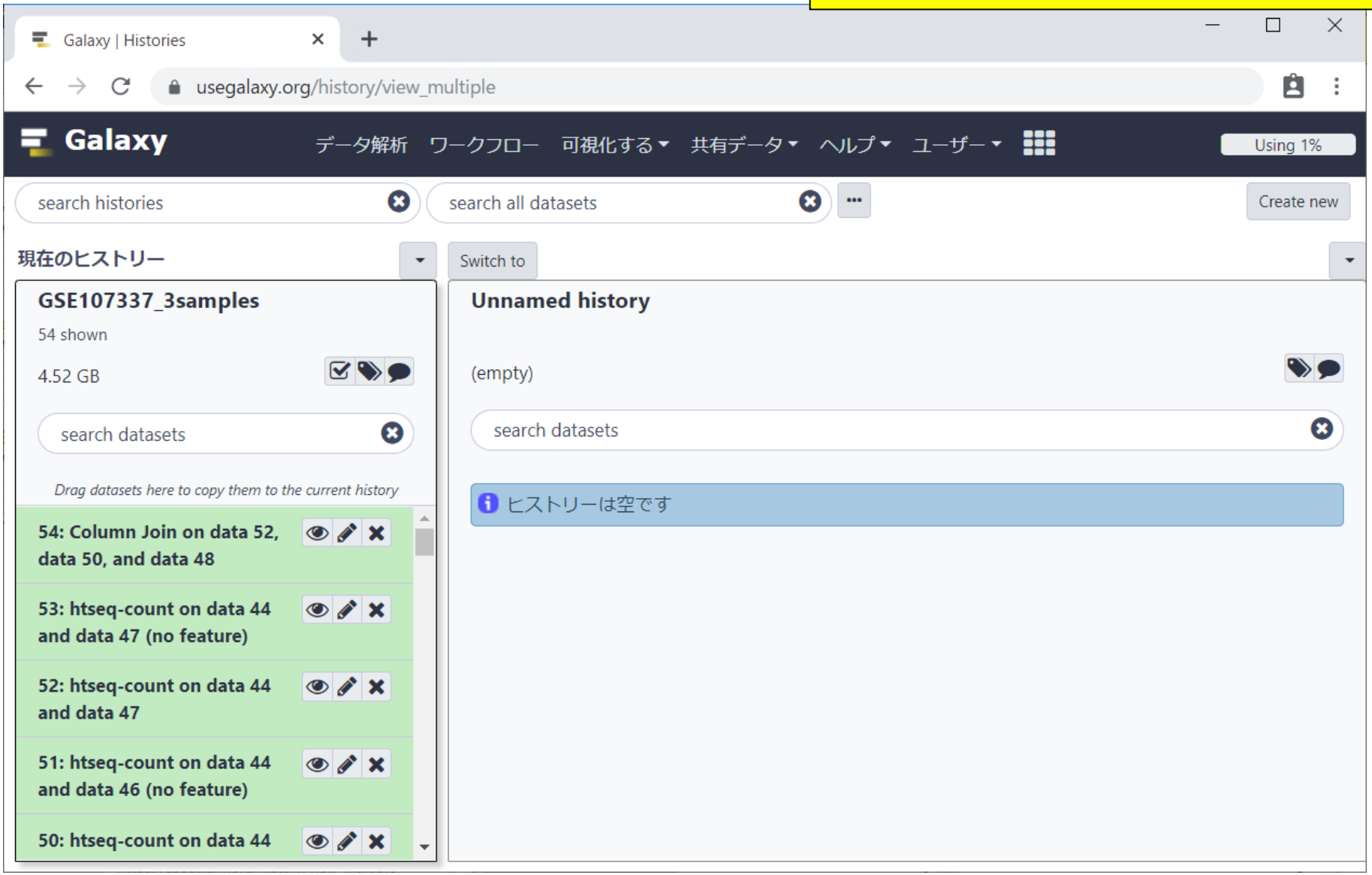
W2：新規ヒストリー

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



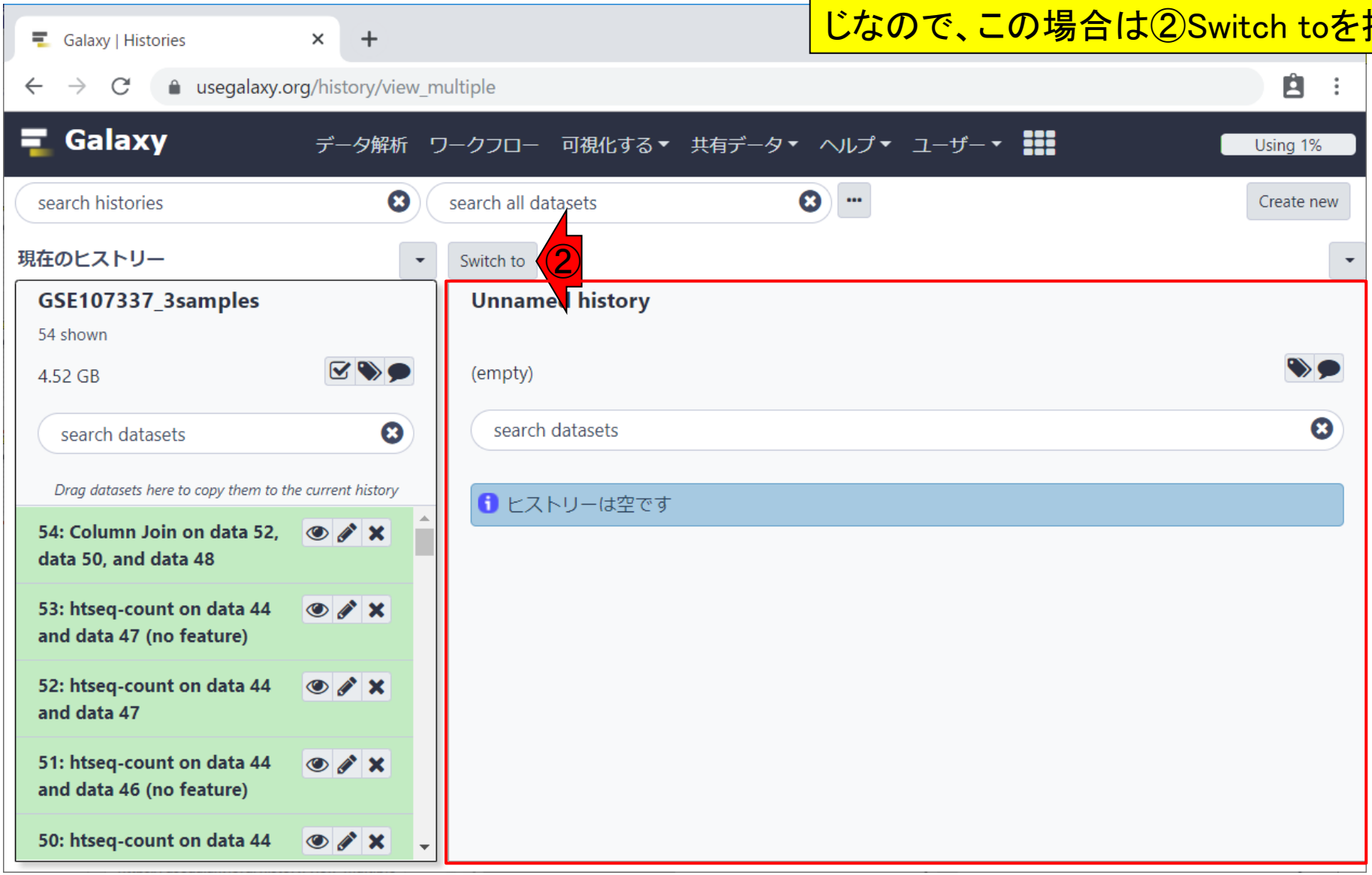
W2：新規ヒストリー

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。



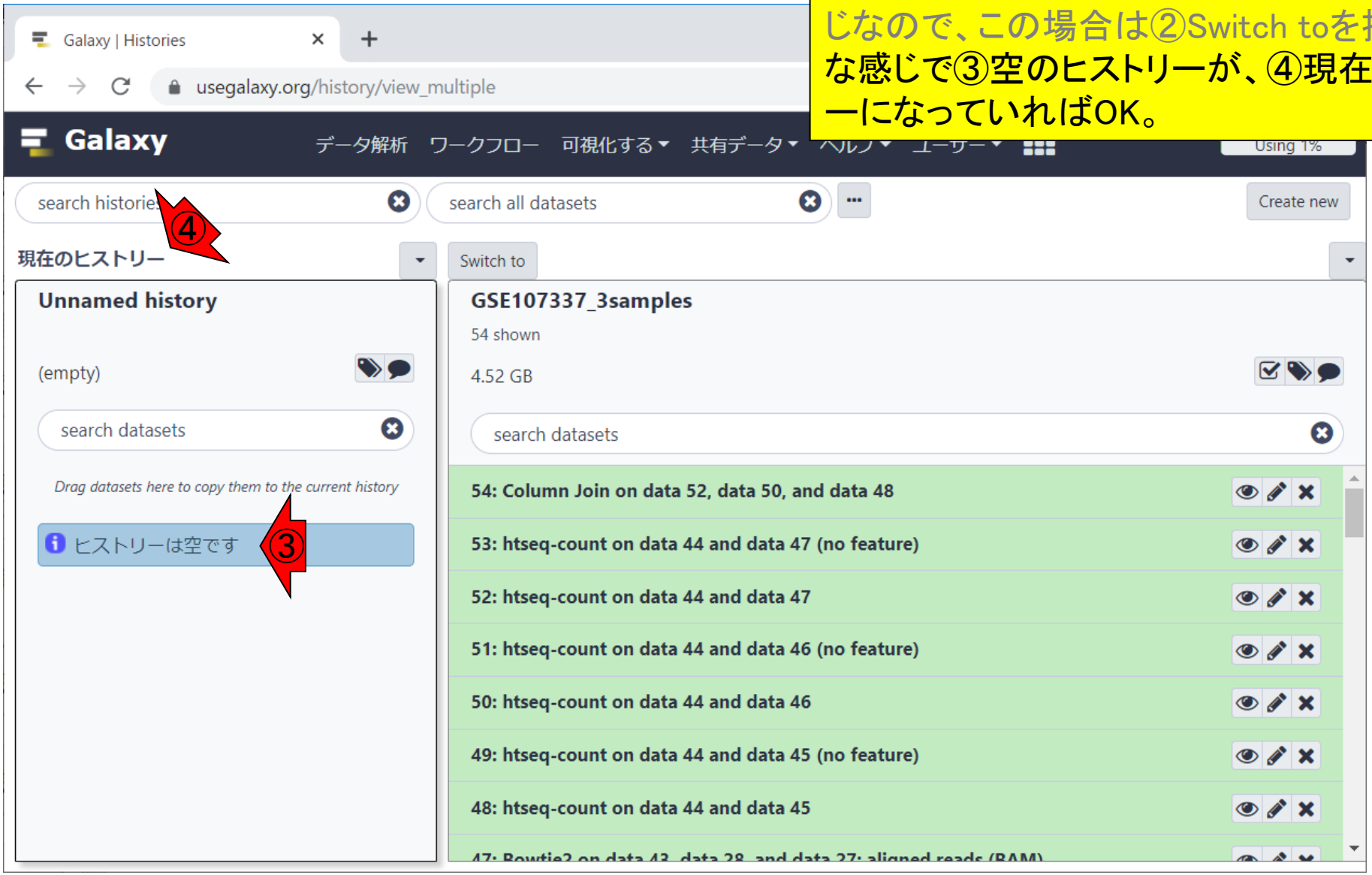
W2：新規ヒストリー

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。



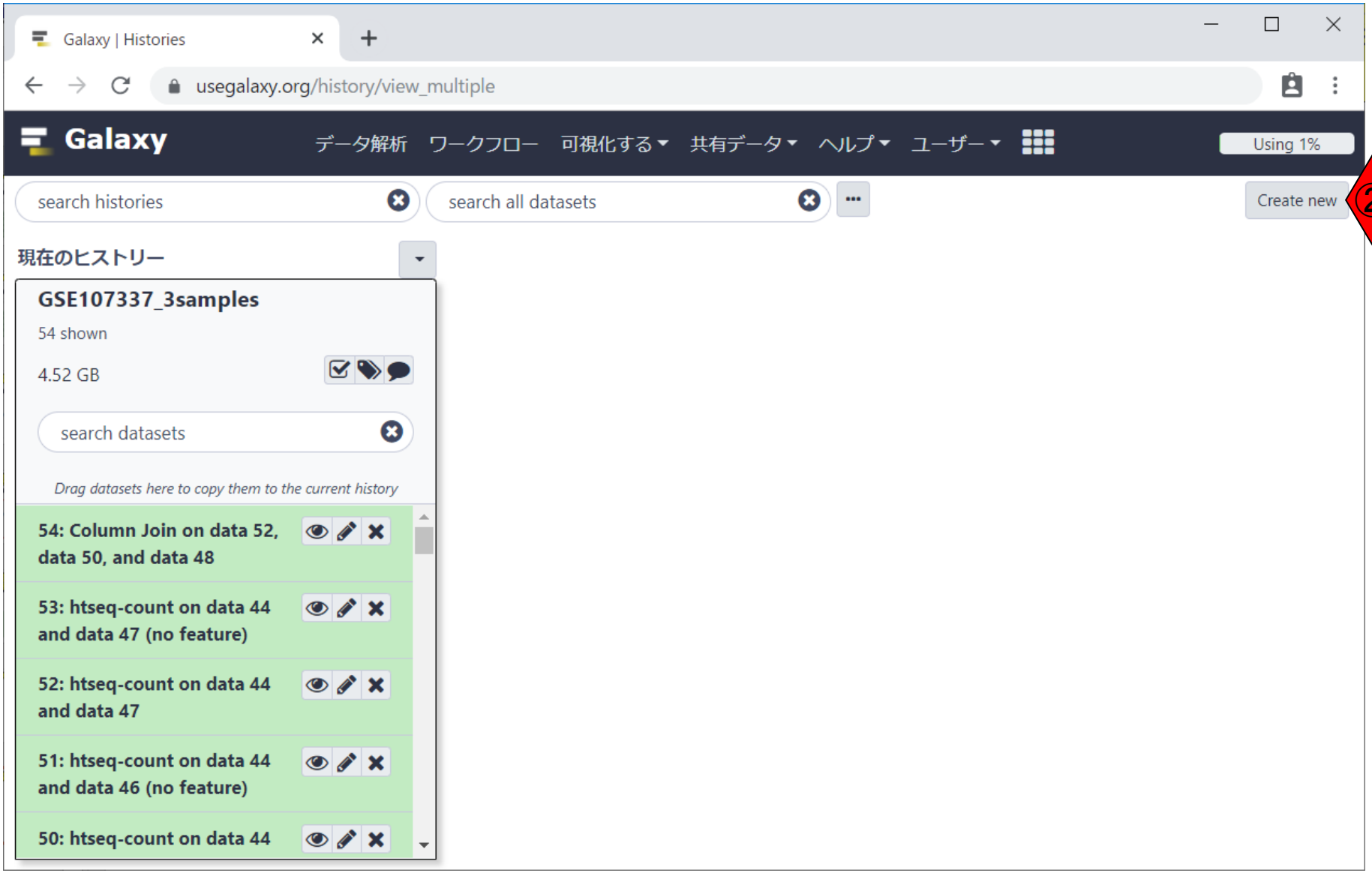
W2：新規ヒストリー

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。



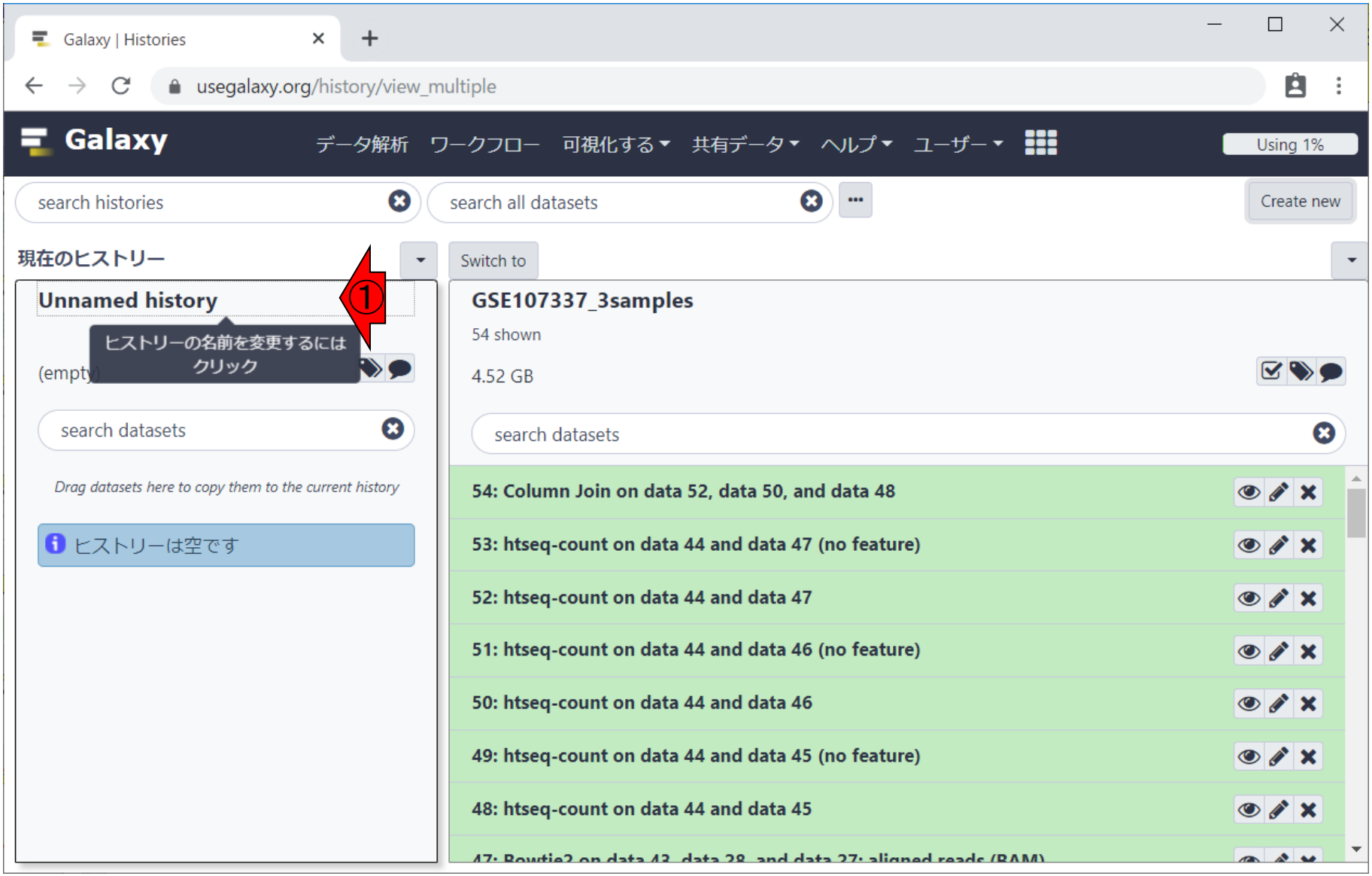
W2：新規ヒストリー

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。



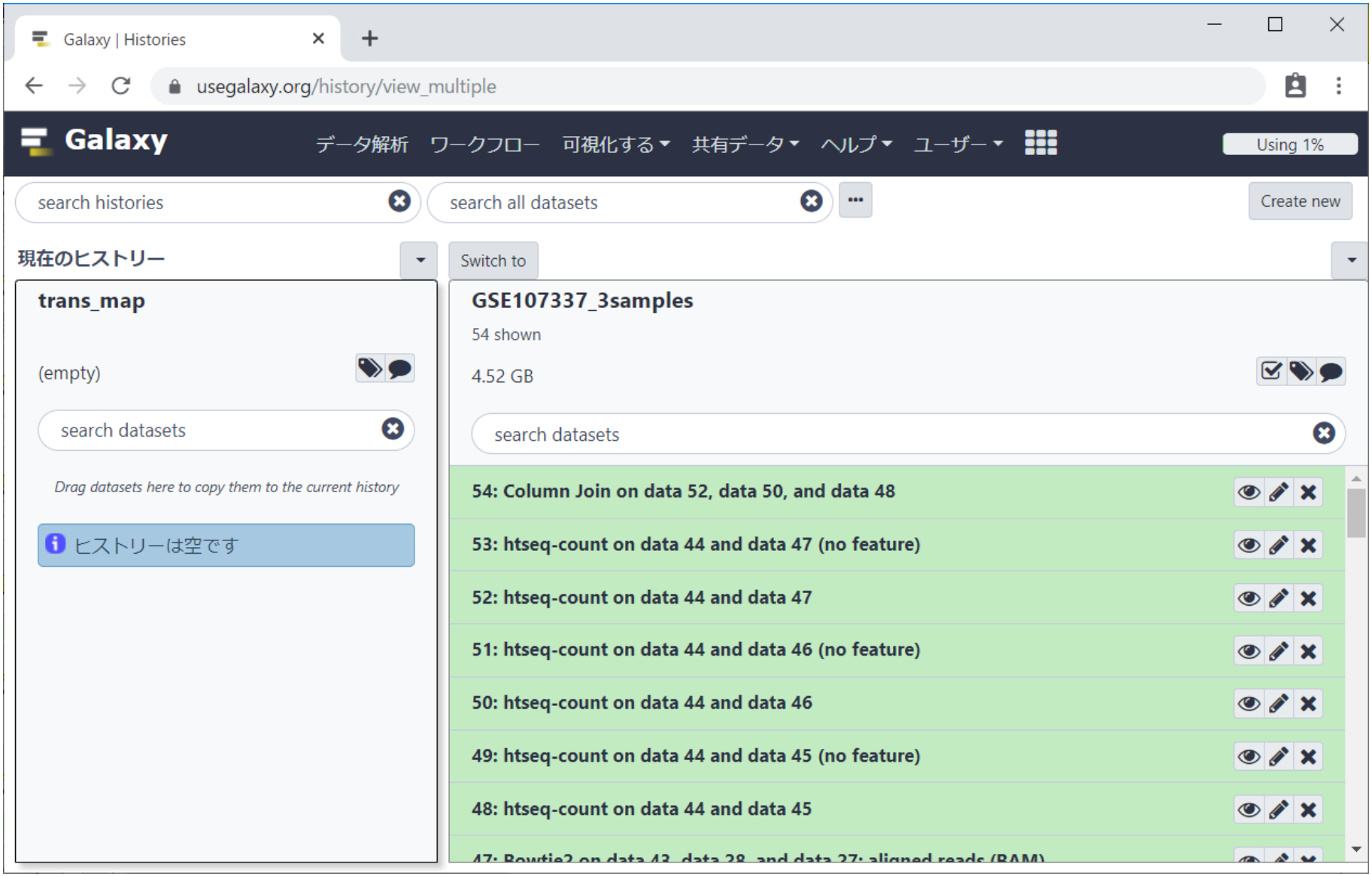
W2：新規ヒストリー

①のところをtrans_mapとして、ヒストリー名を変更。



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。

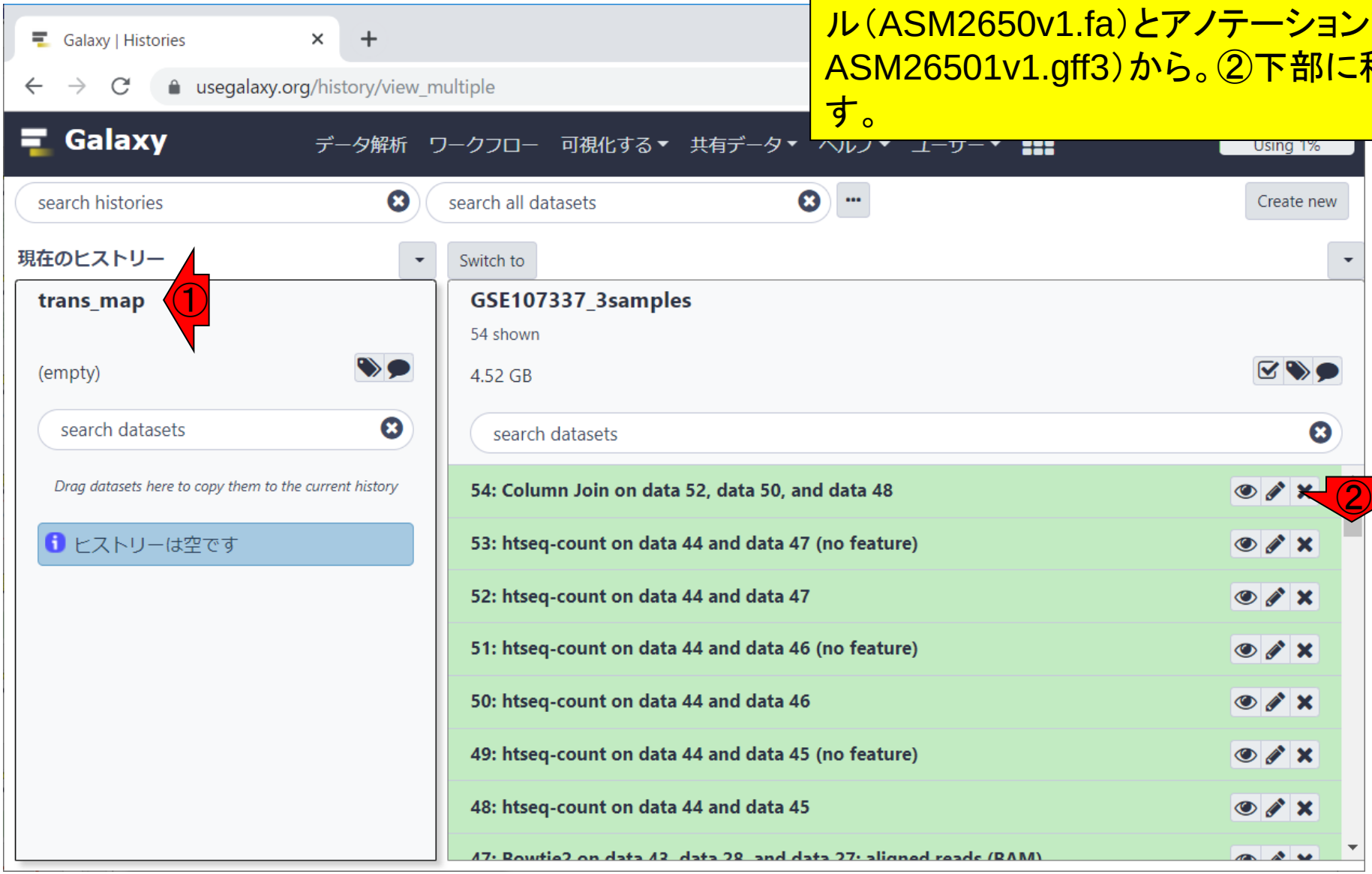


Contents

- 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

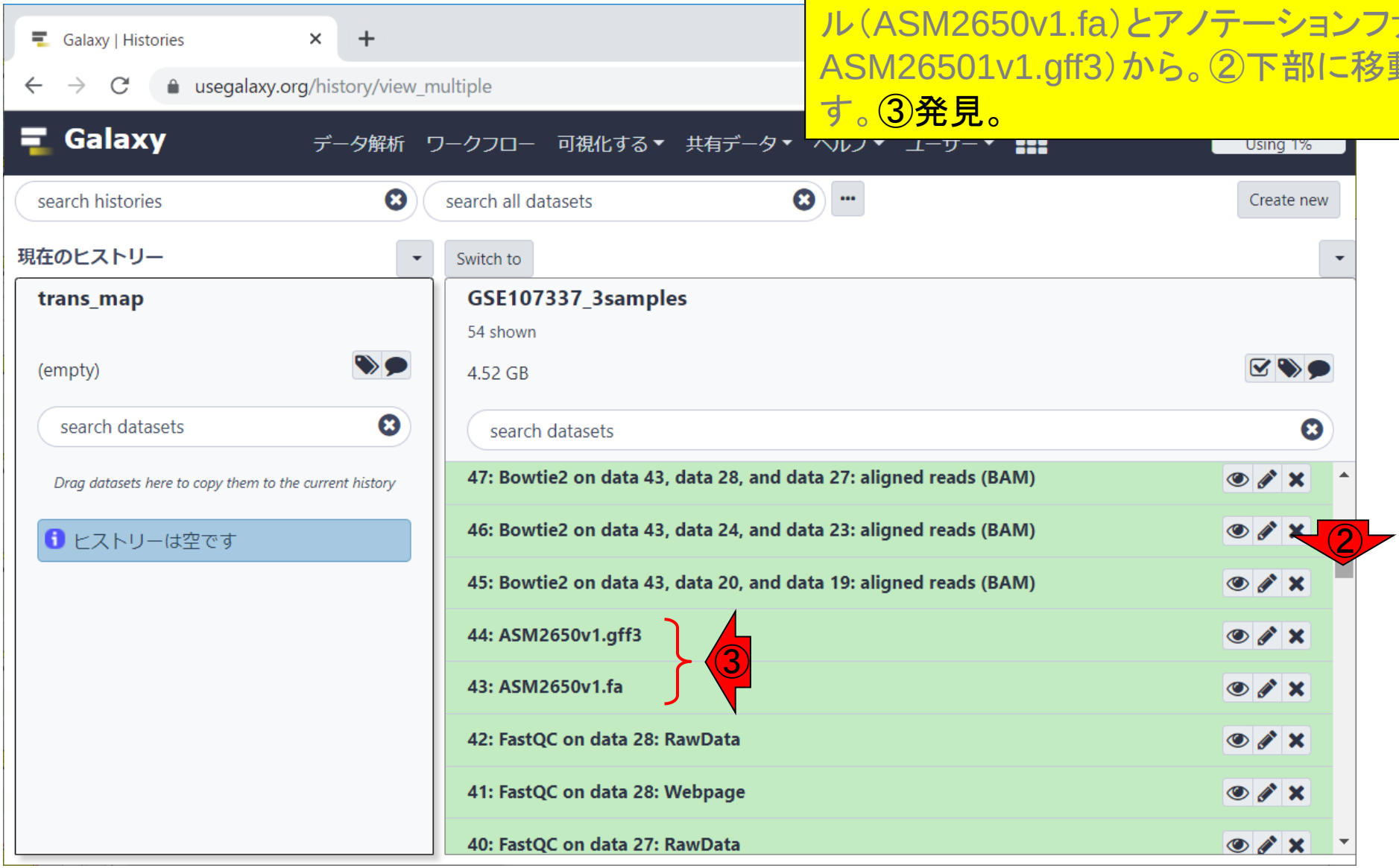
W3：データのコピー

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。



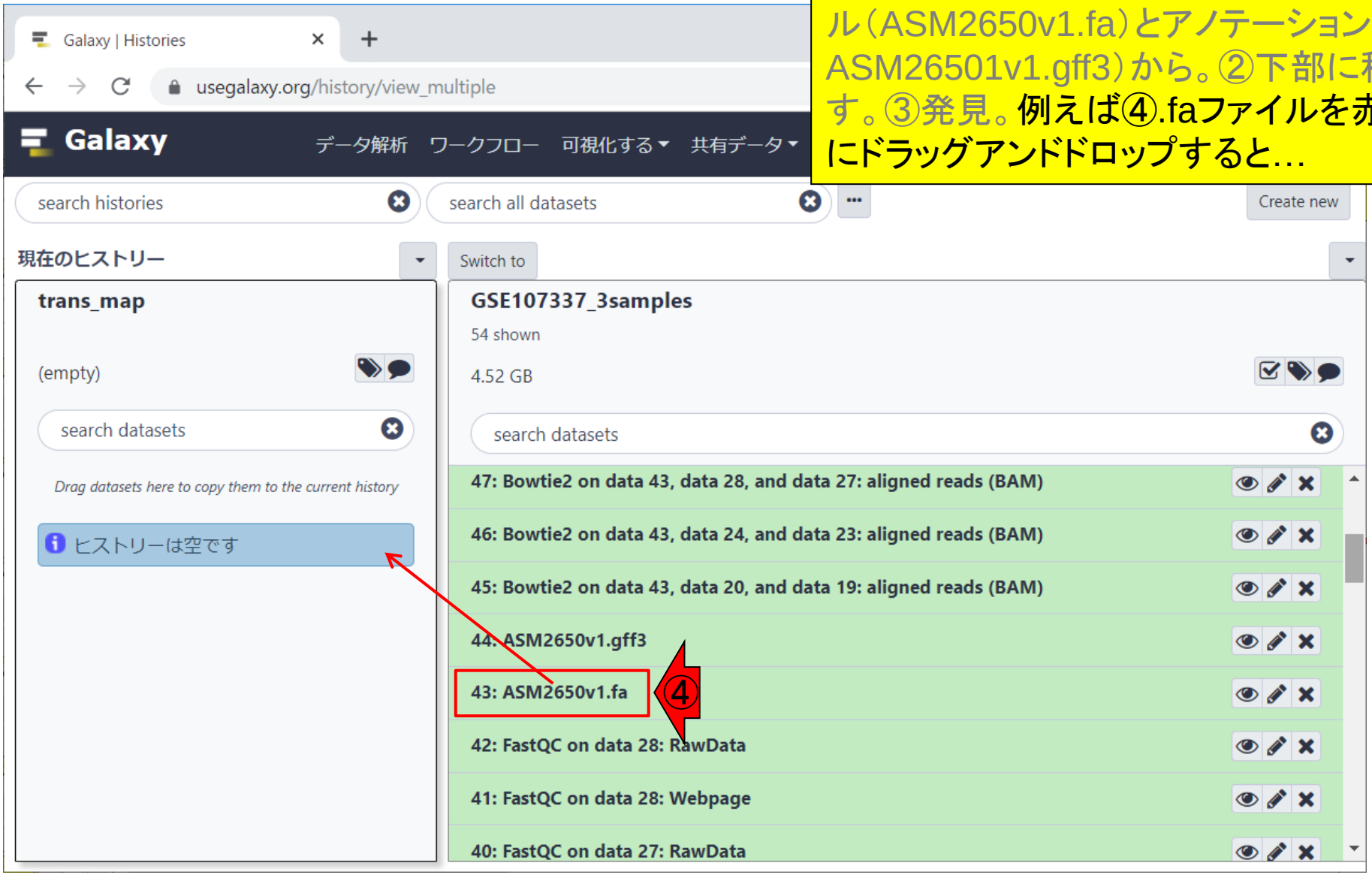
W3：データのコピー

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。



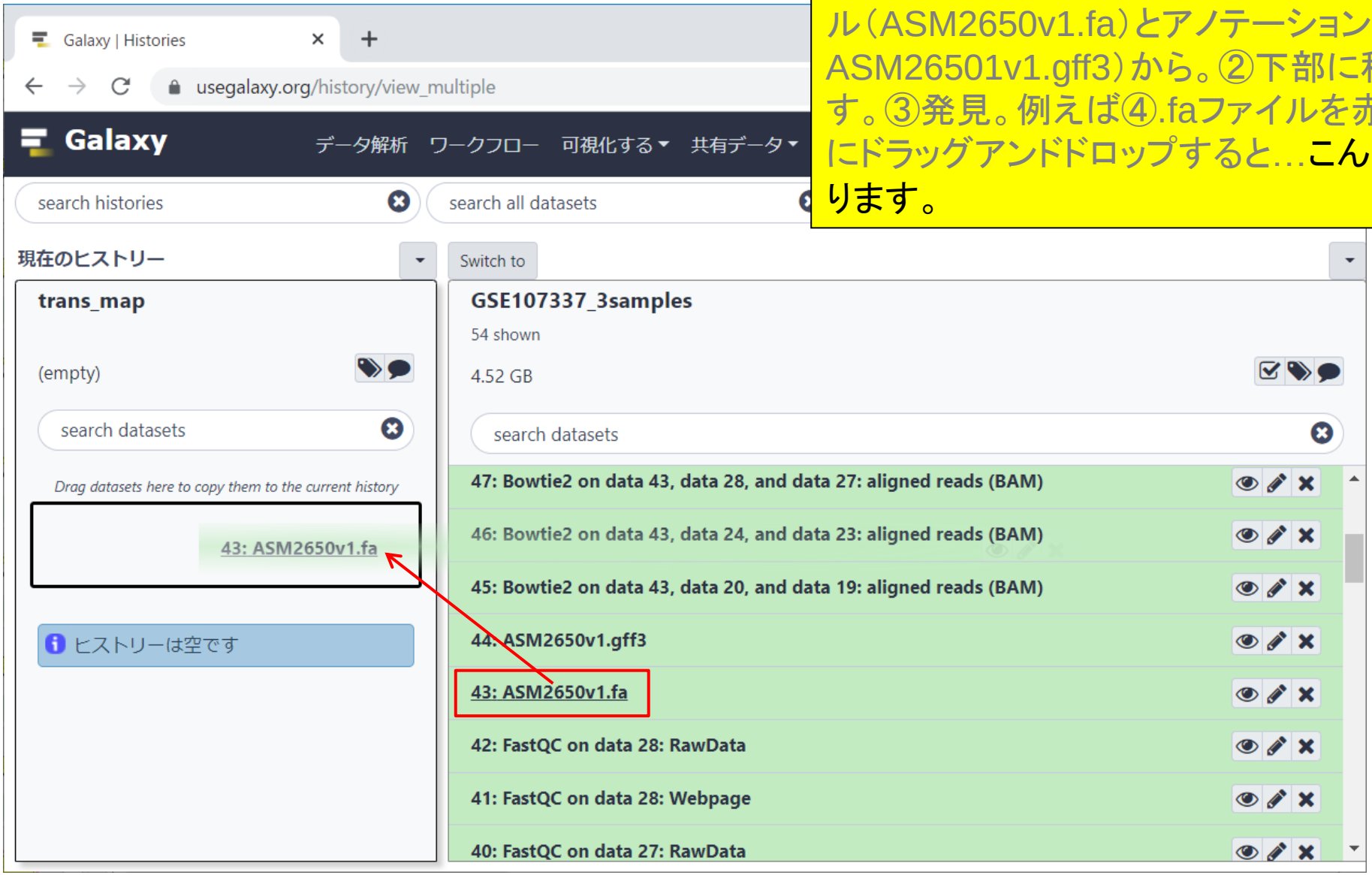
W3：データのコピー

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…



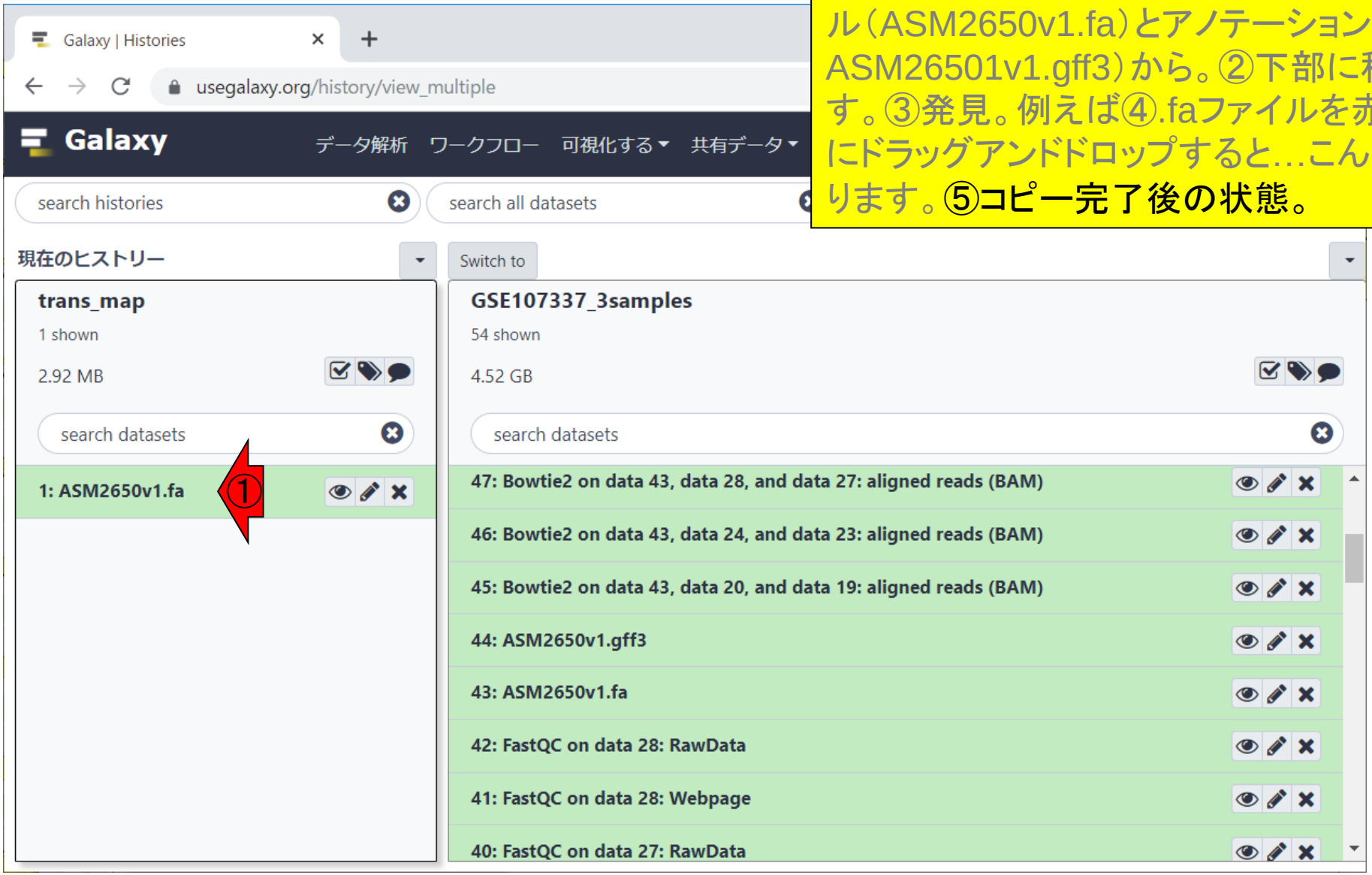
W3：データのコピー

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。



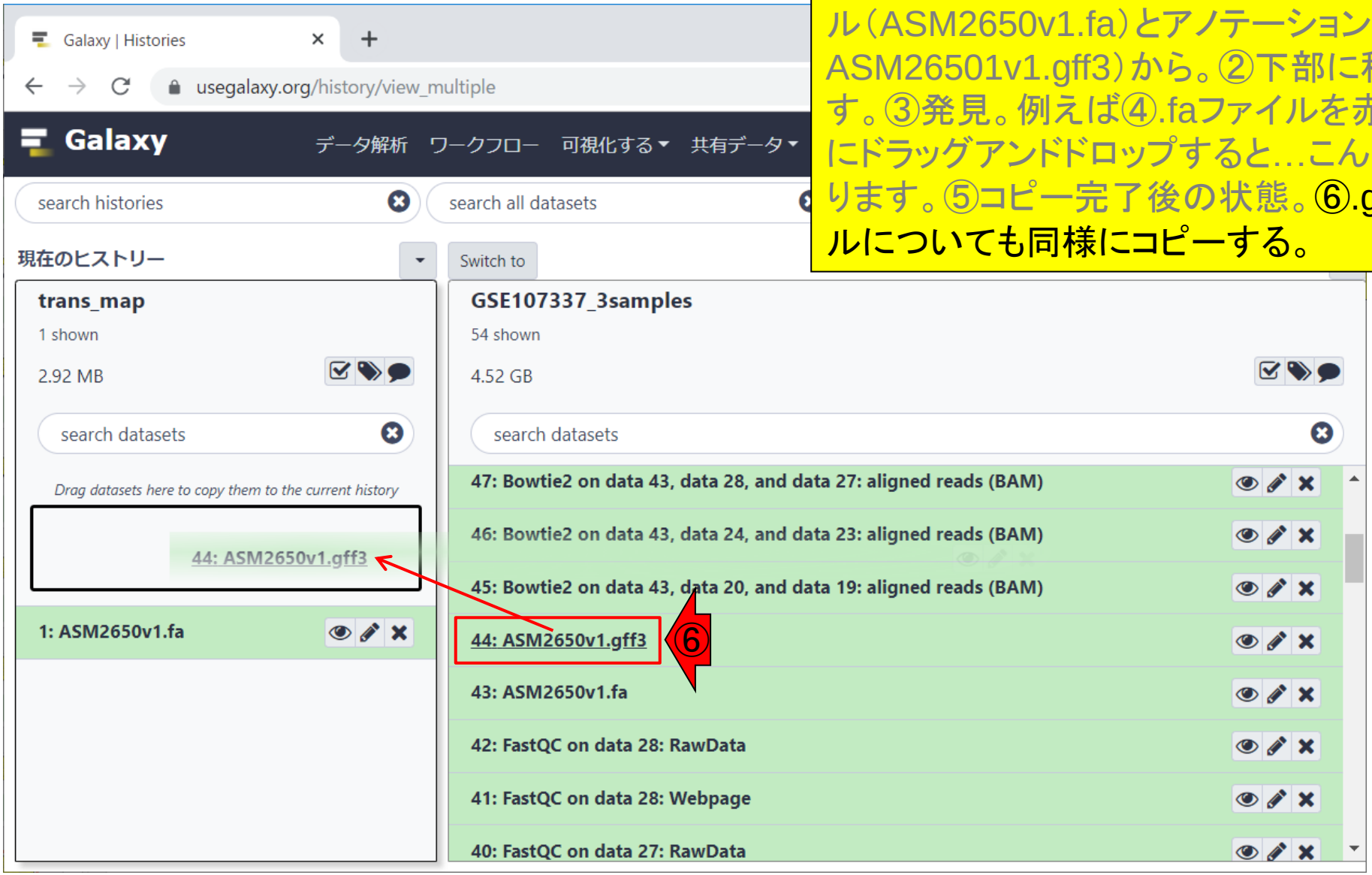
W3：データのコピー

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。



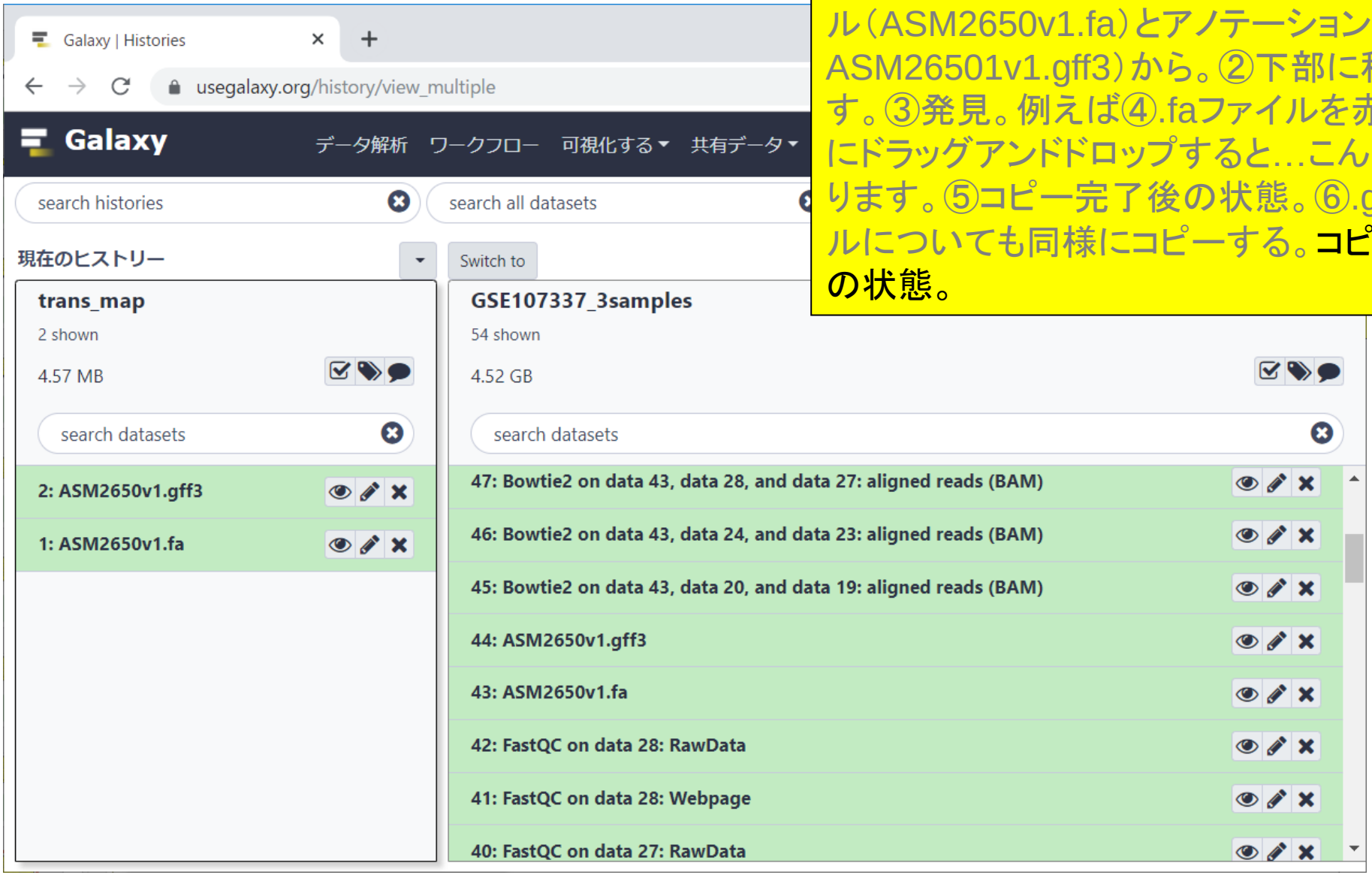
W3：データのコピー

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。



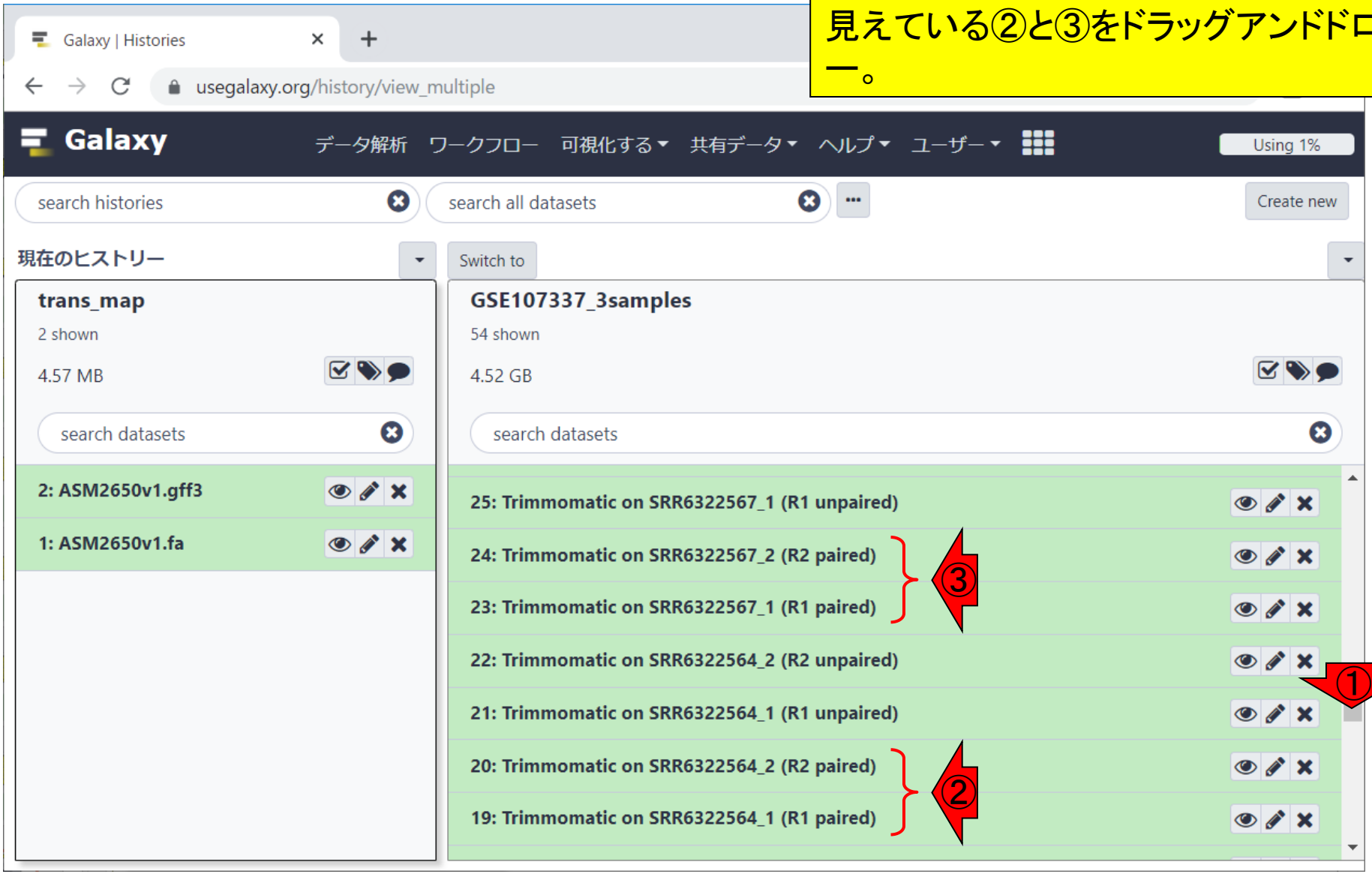
W3：データのコピー

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。**



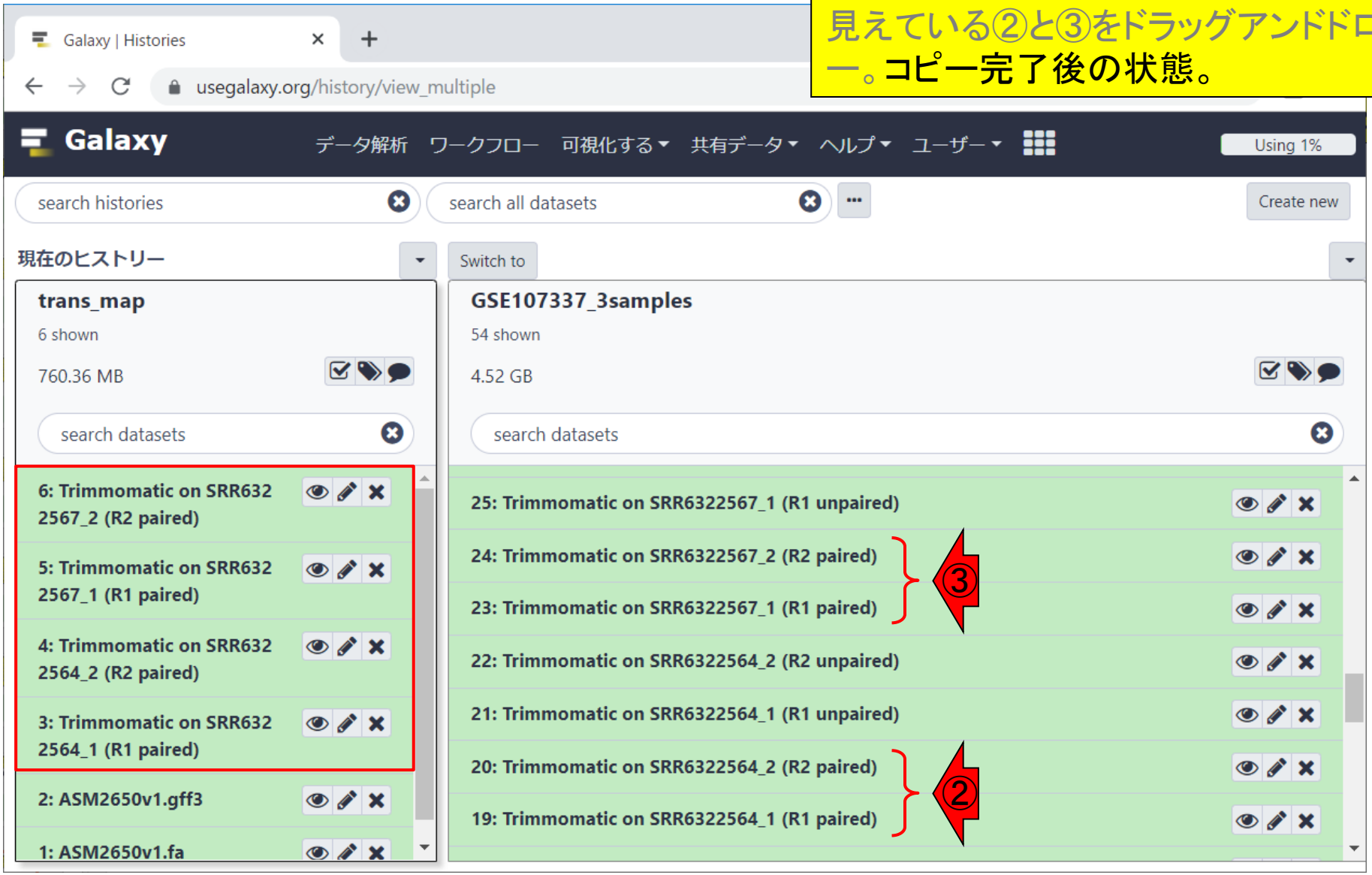
W3：データのコピー

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。



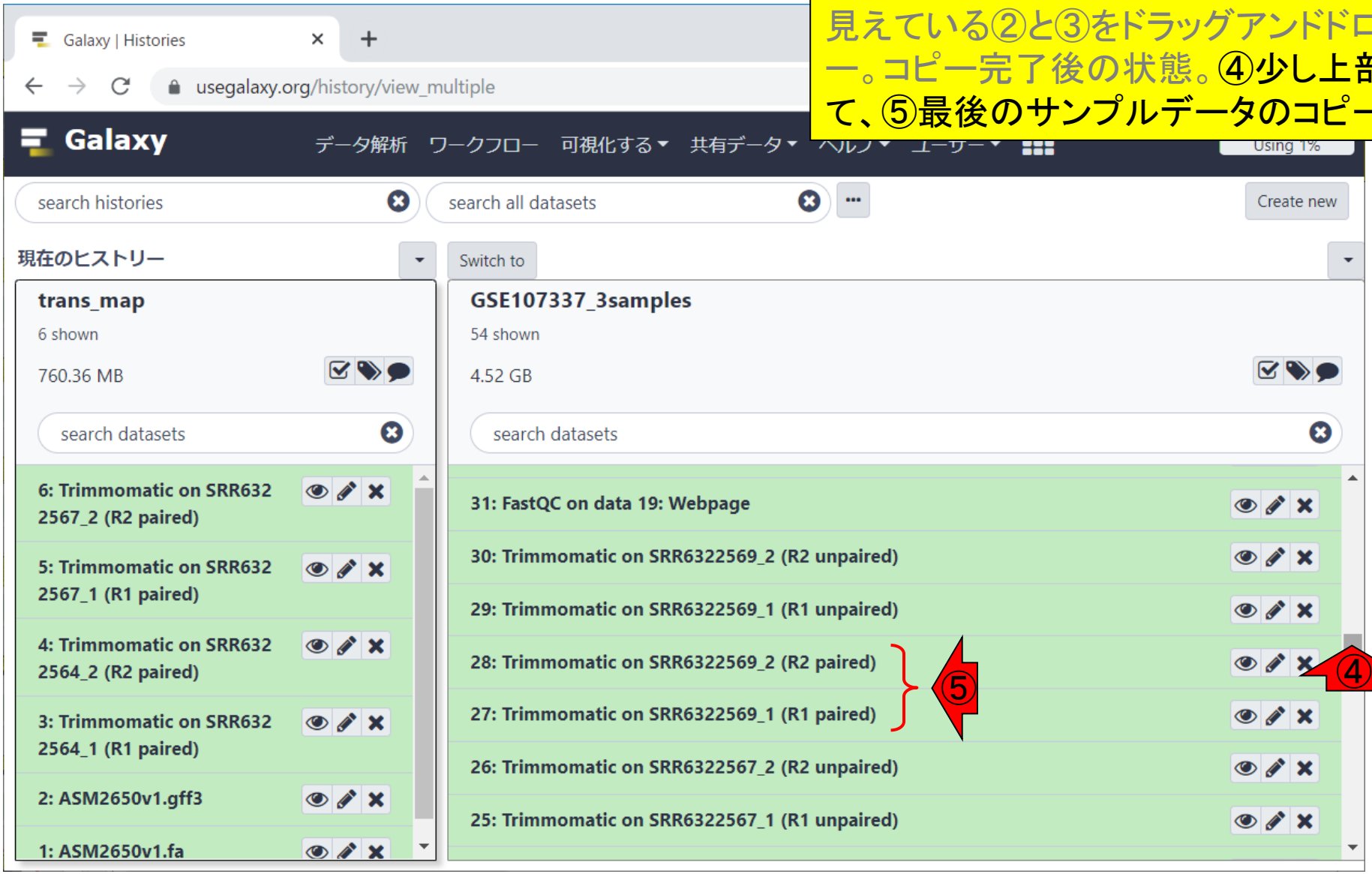
W3：データのコピー

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。



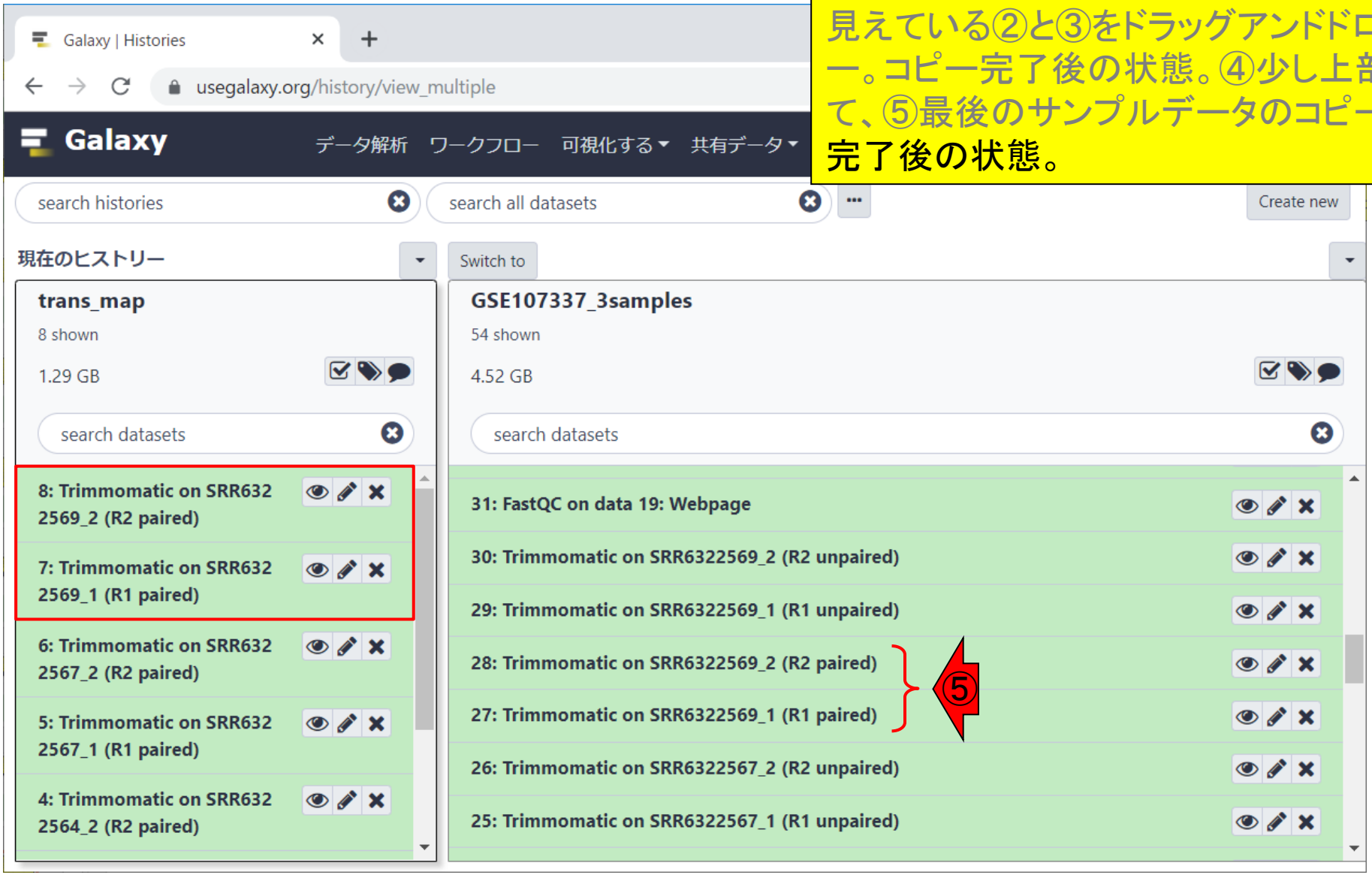
W3：データのコピー

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。



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。

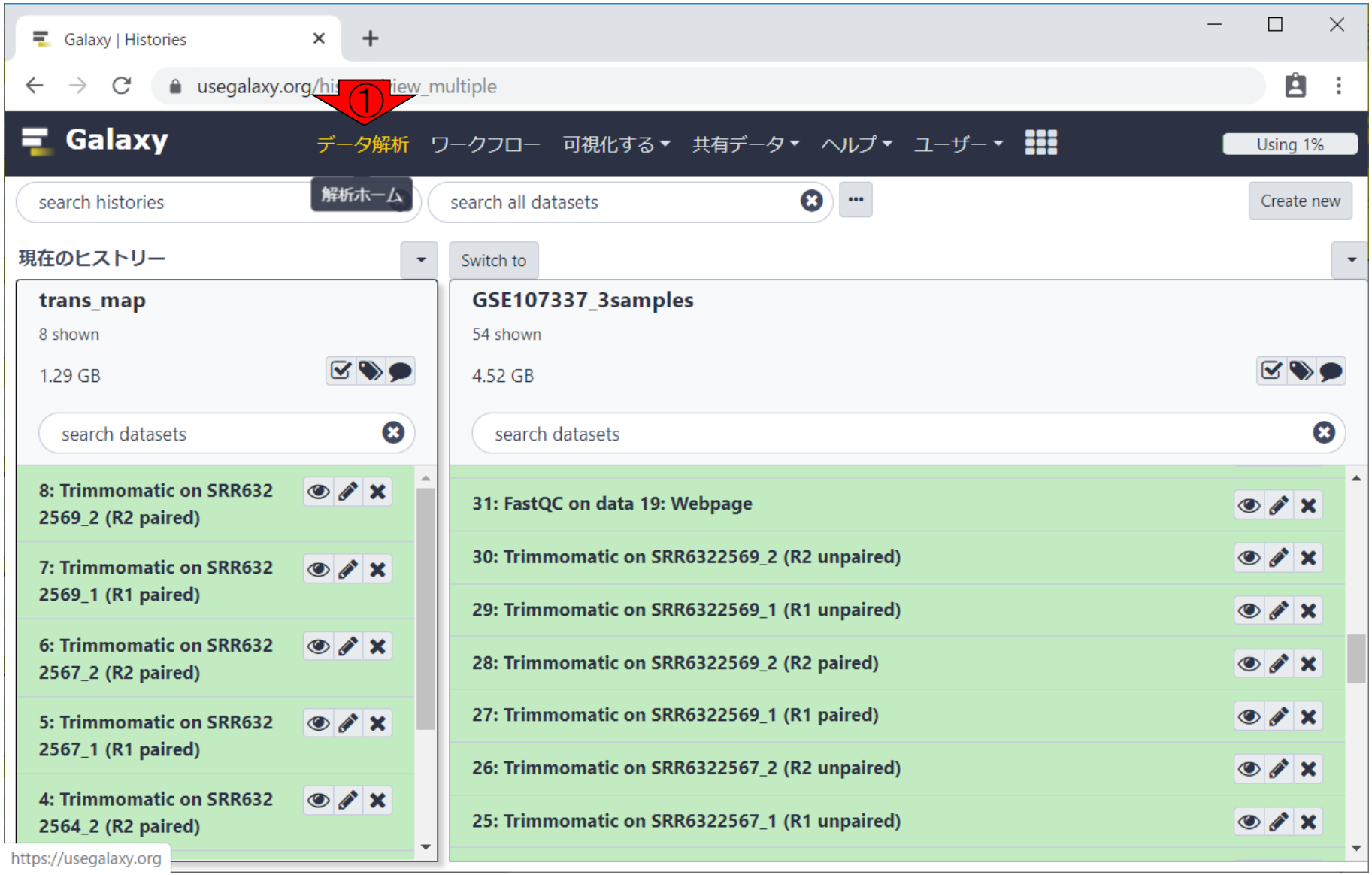


Contents

- 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

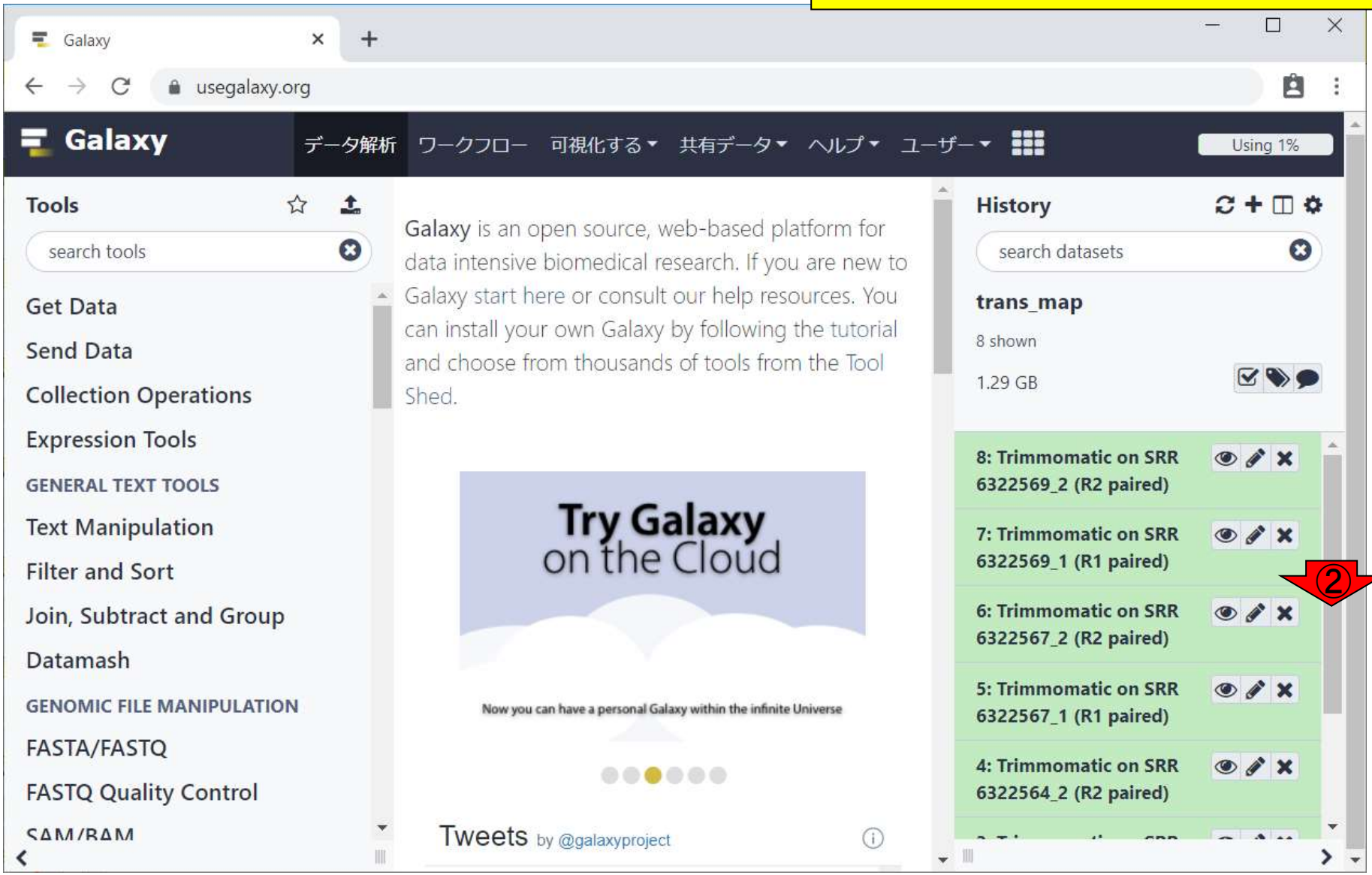
W4:解析準備完了

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。



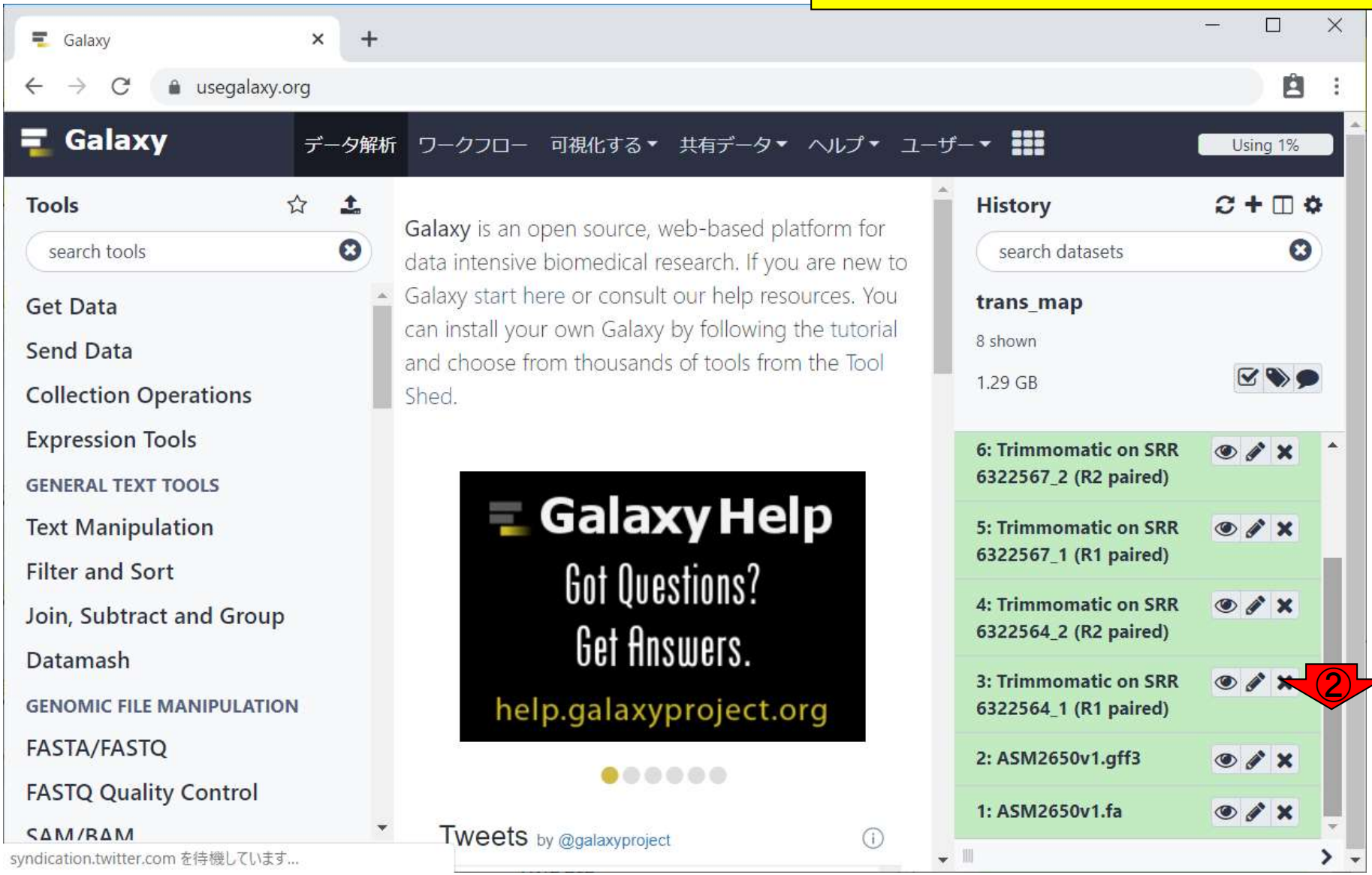
W4：解析準備完了

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。



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。

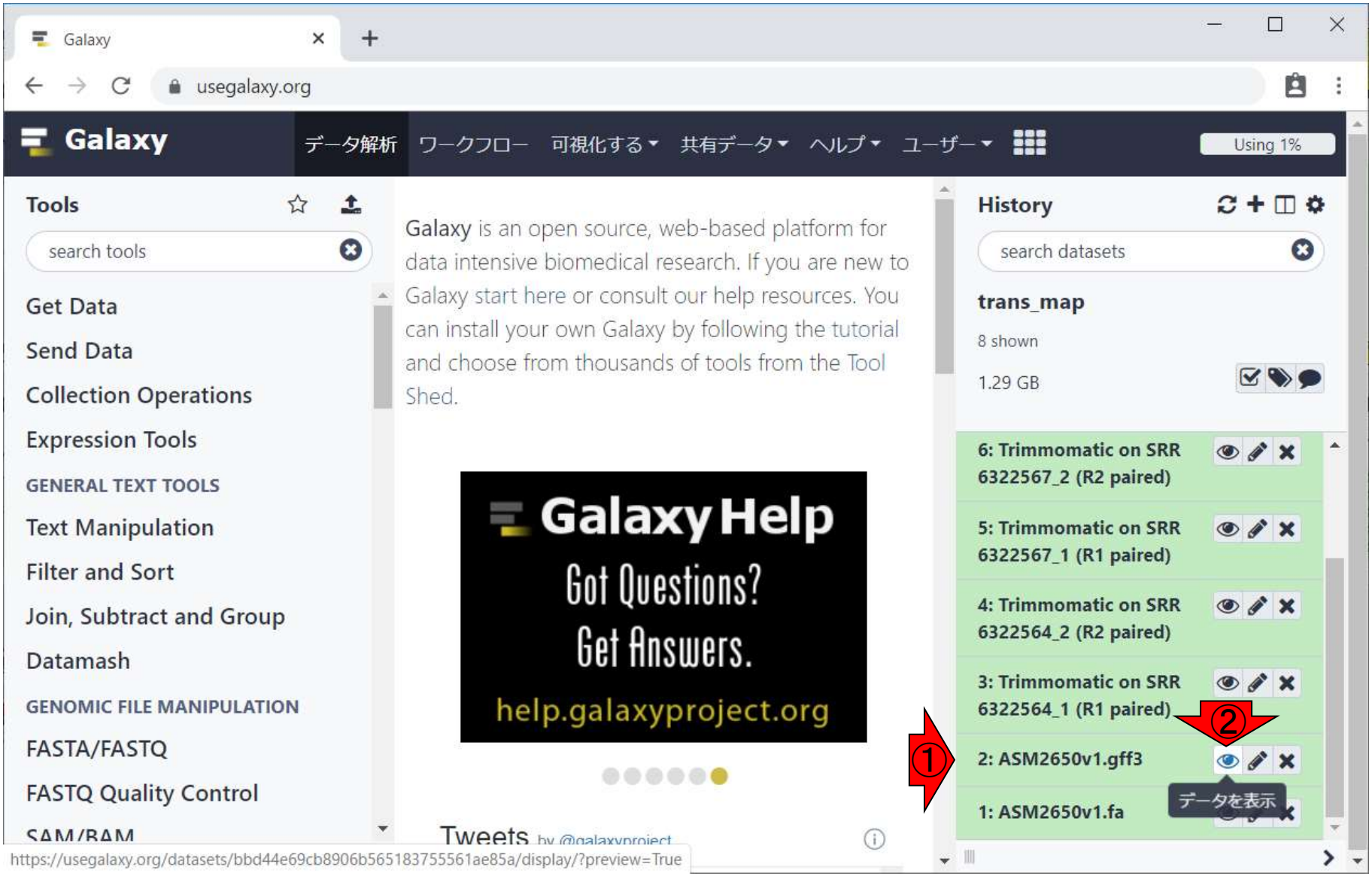


Contents

- 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

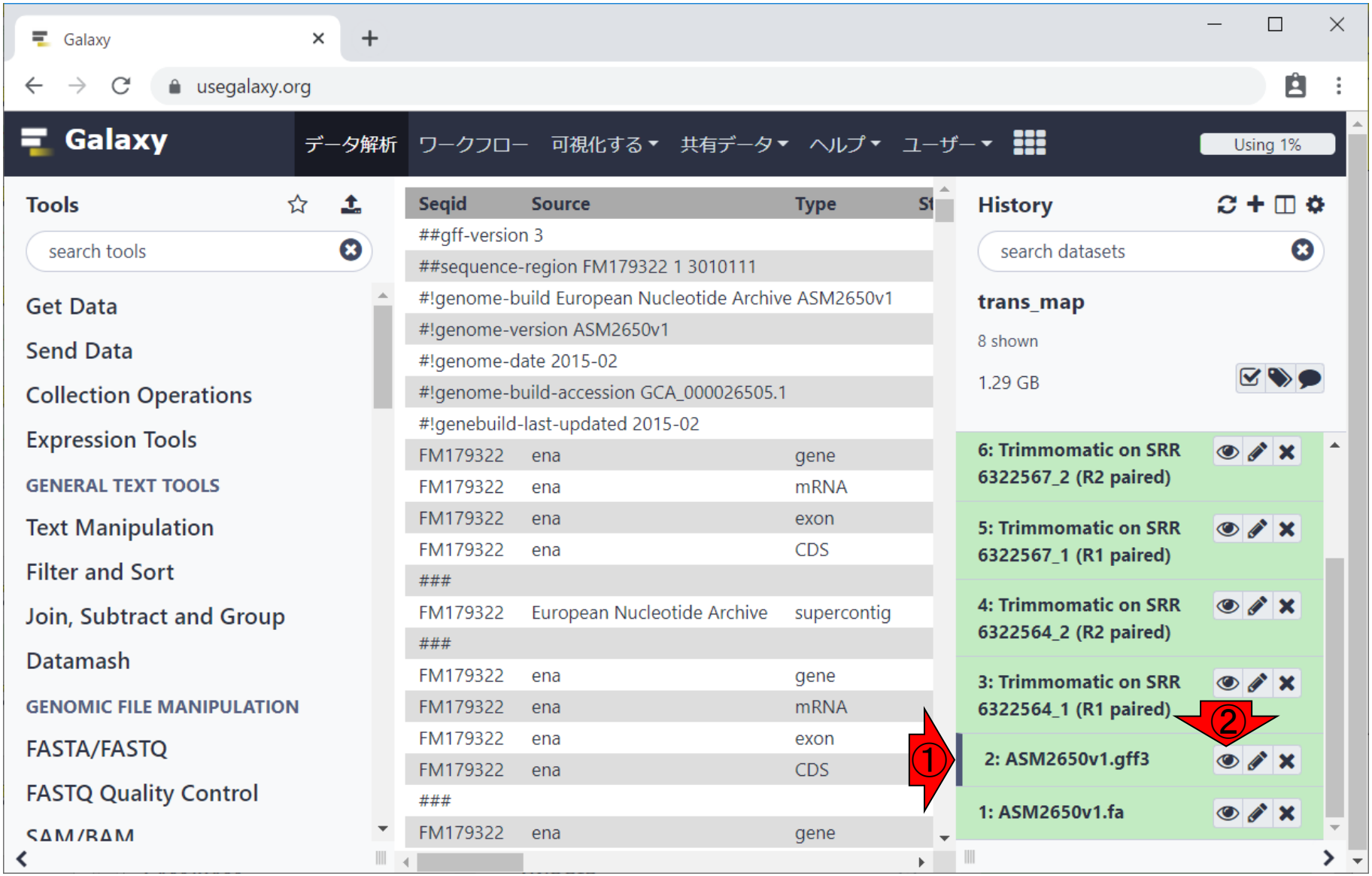
W5：GFFの前処理

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。



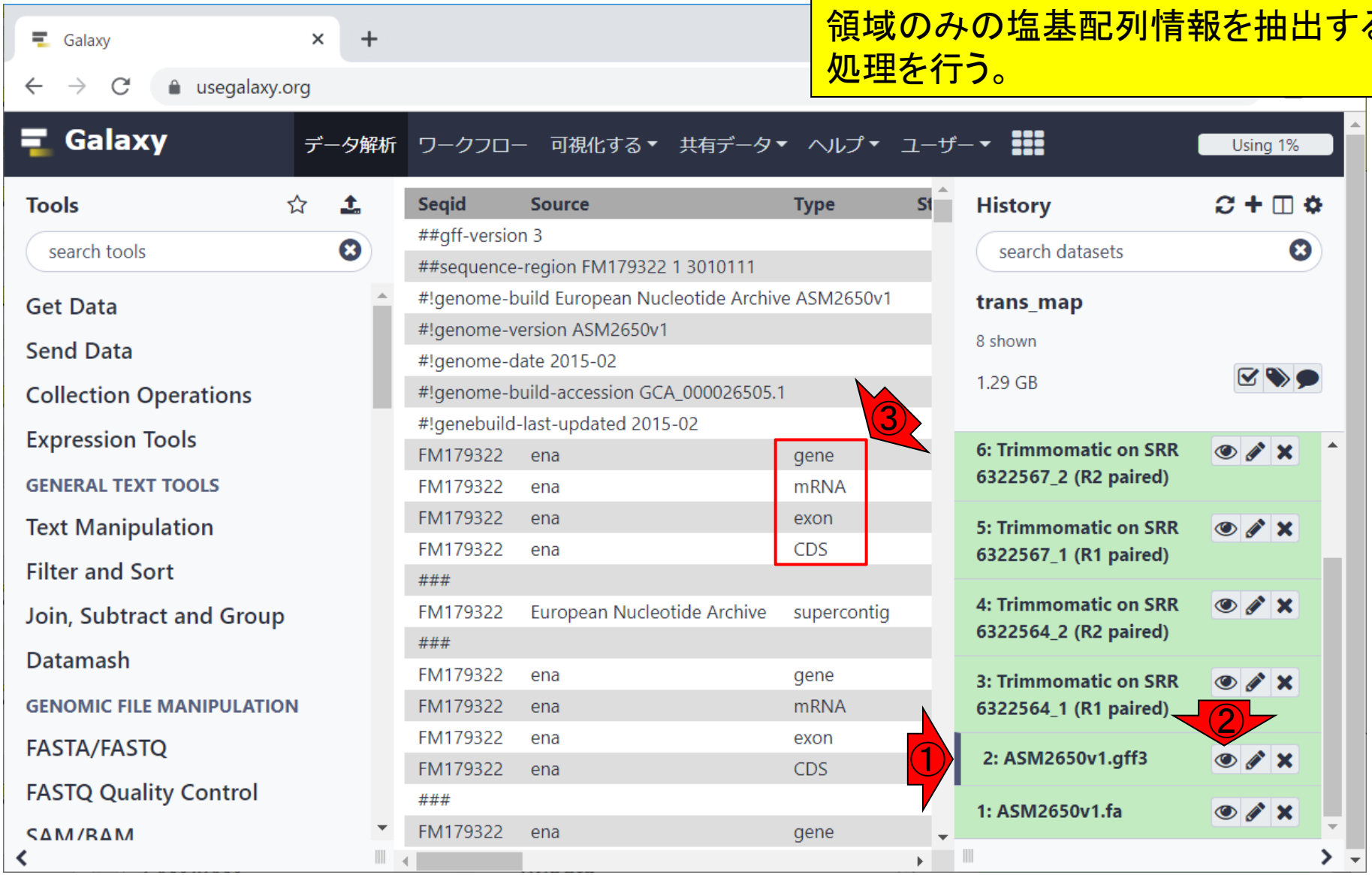
W5：GFFの前処理

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。



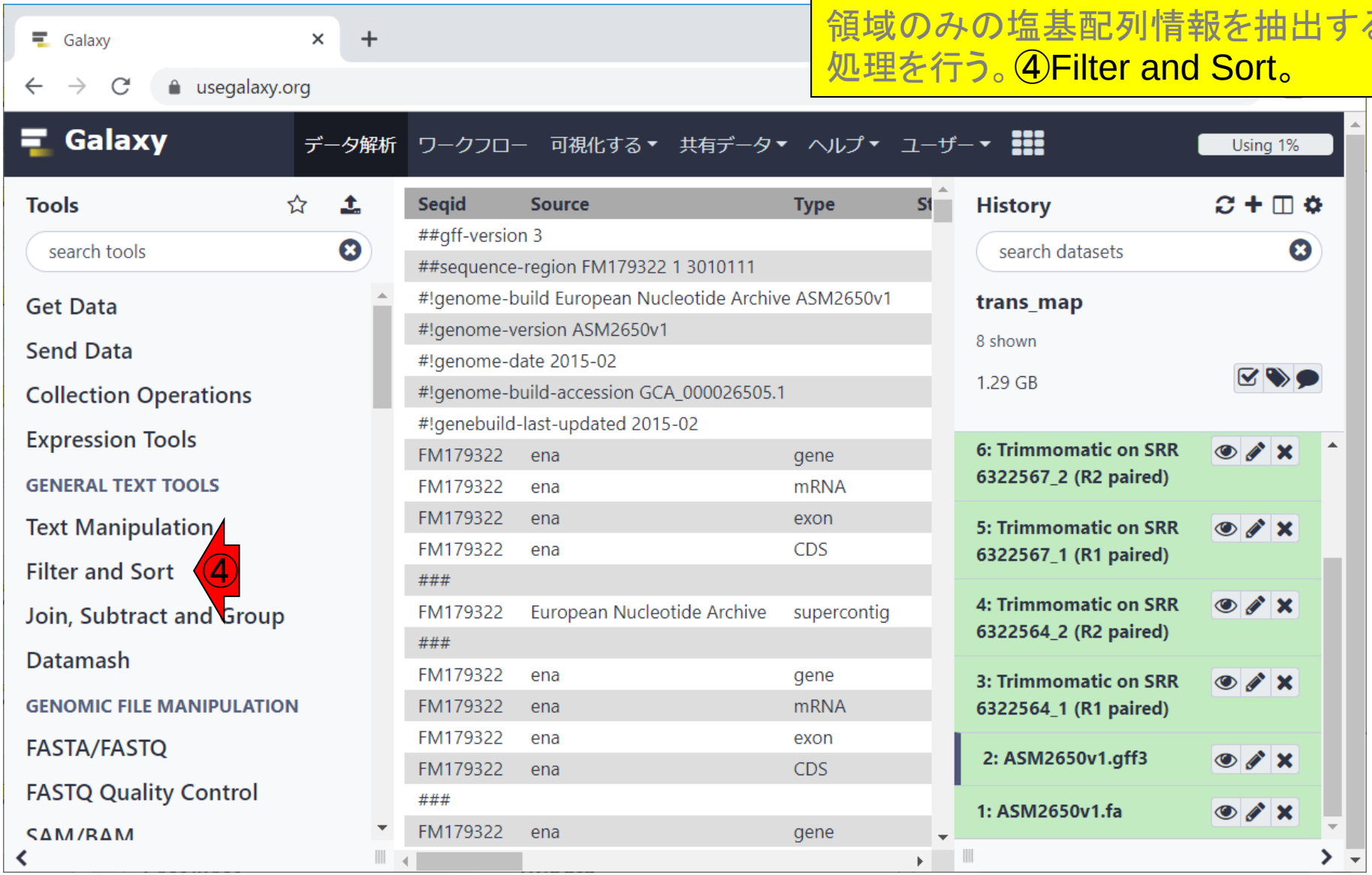
W5：GFFの前処理

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。



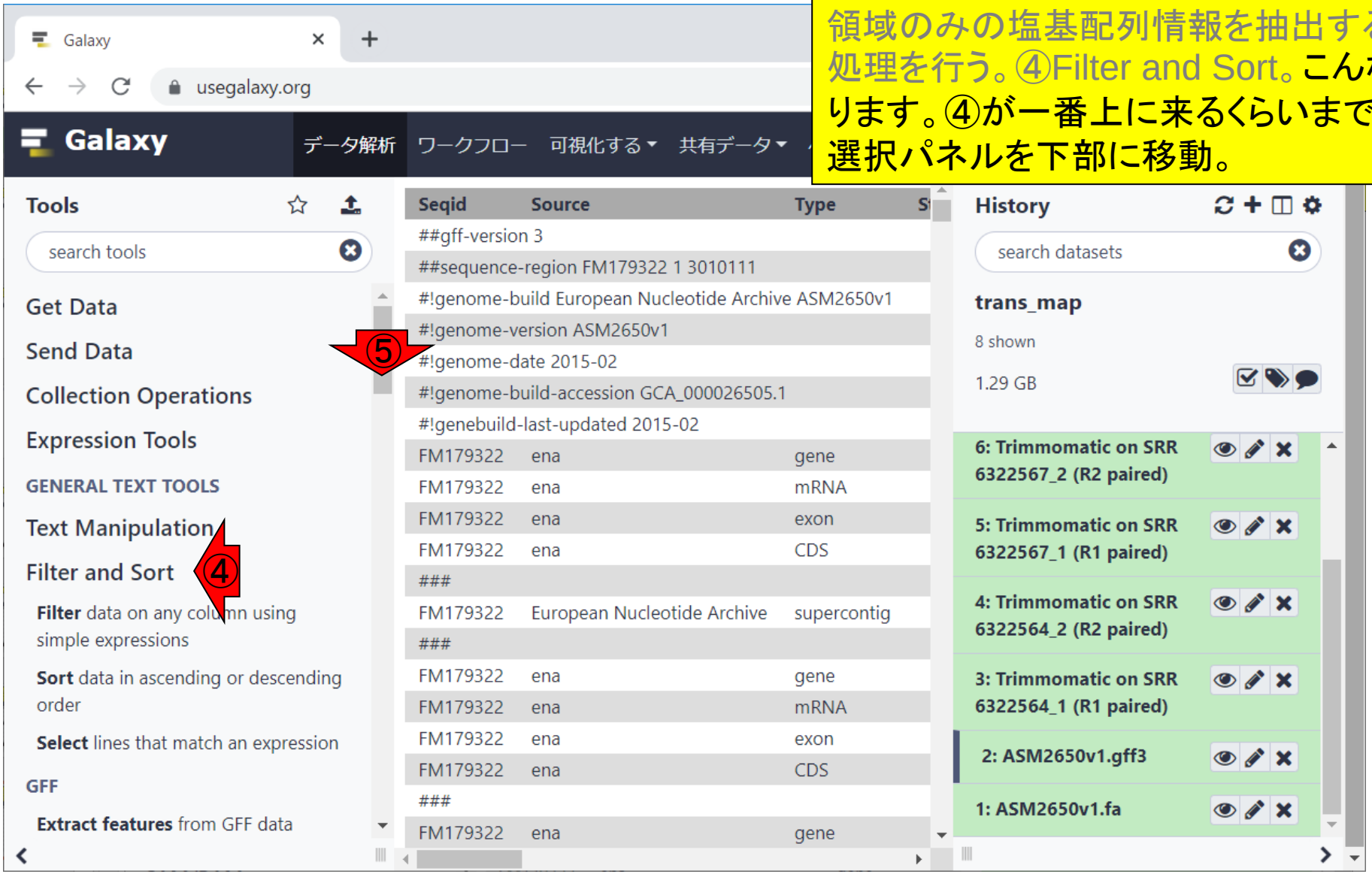
W5：GFFの前処理

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。



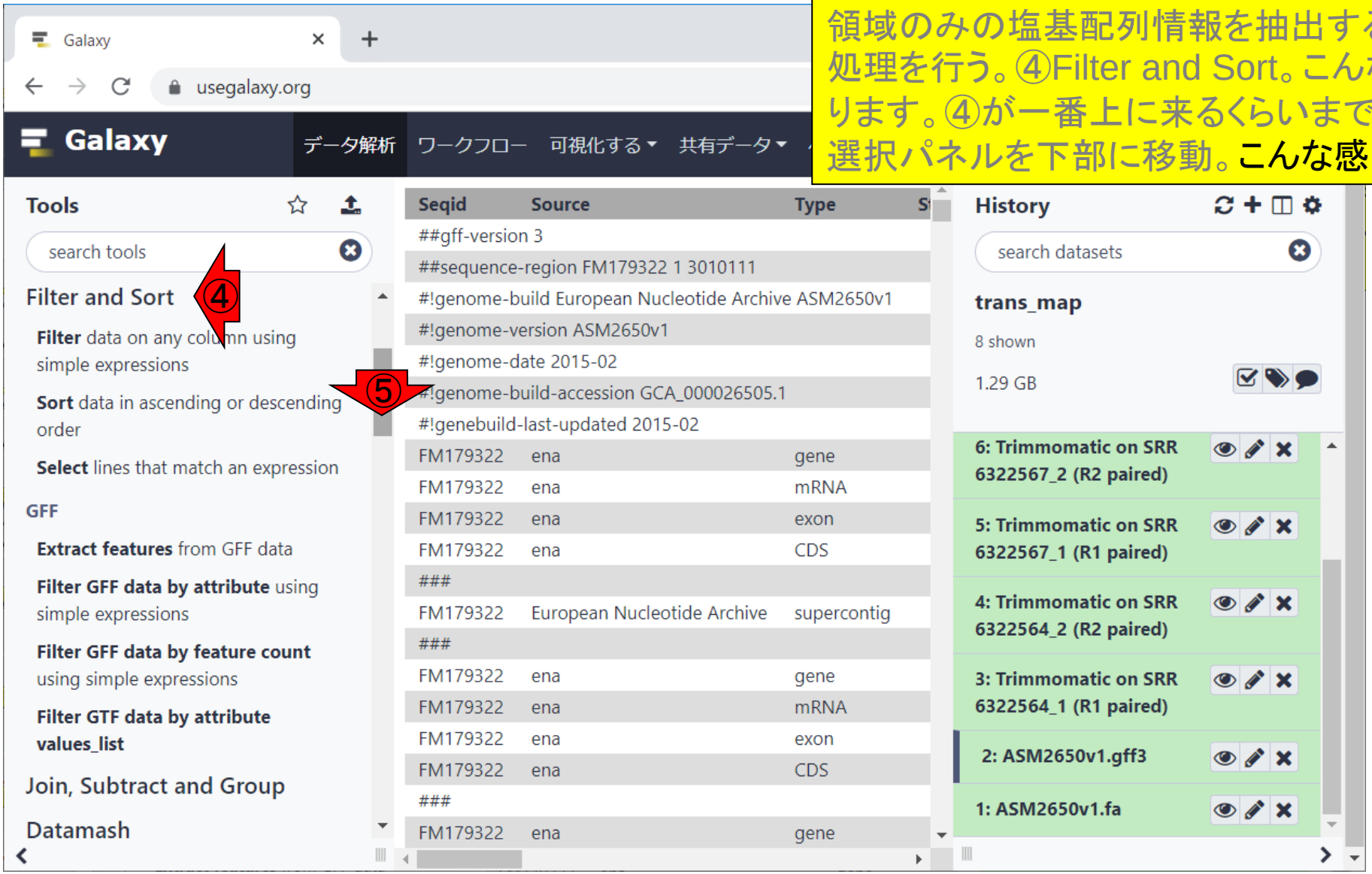
W5：GFFの前処理

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。



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。

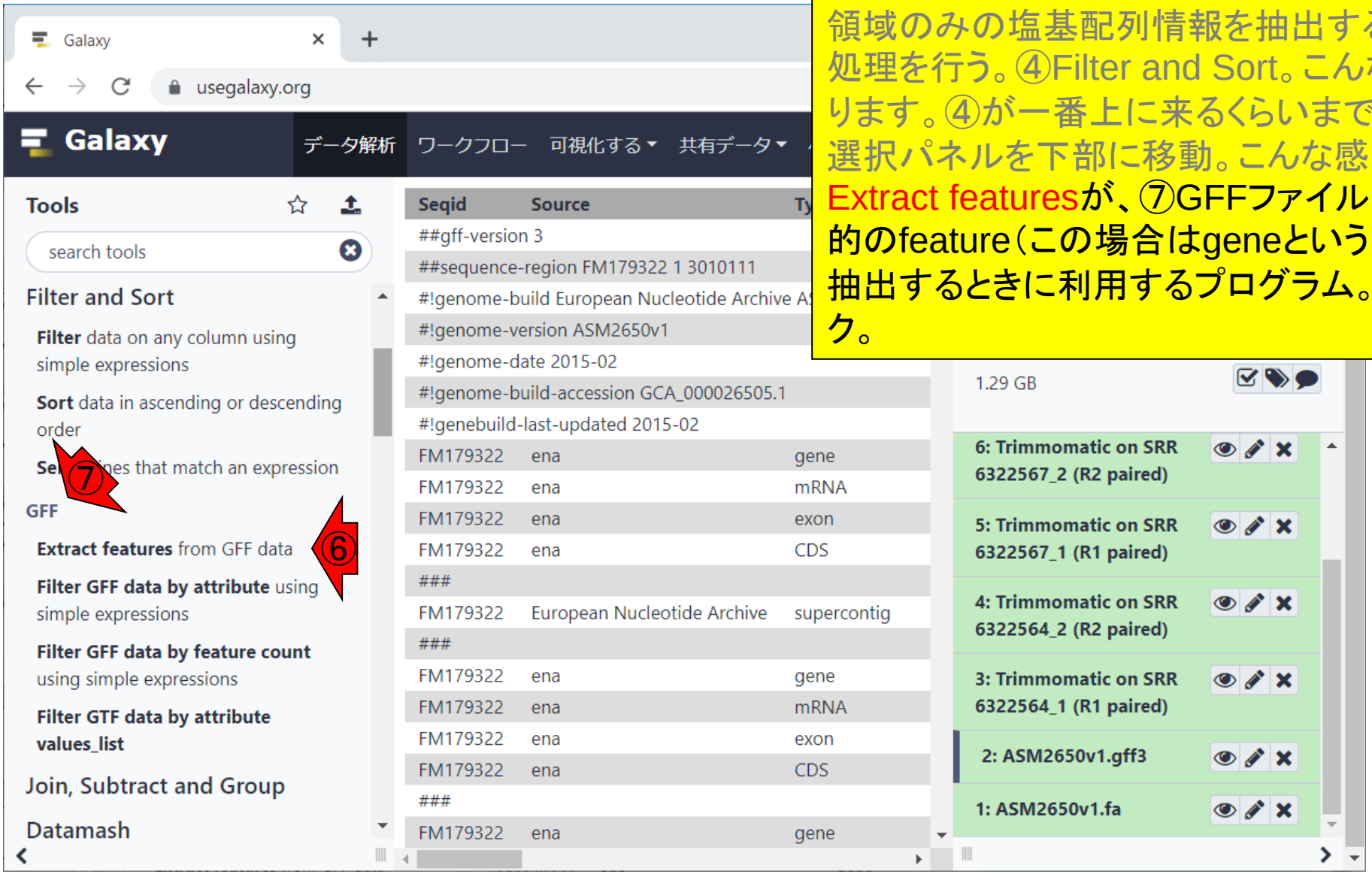


Contents

- 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

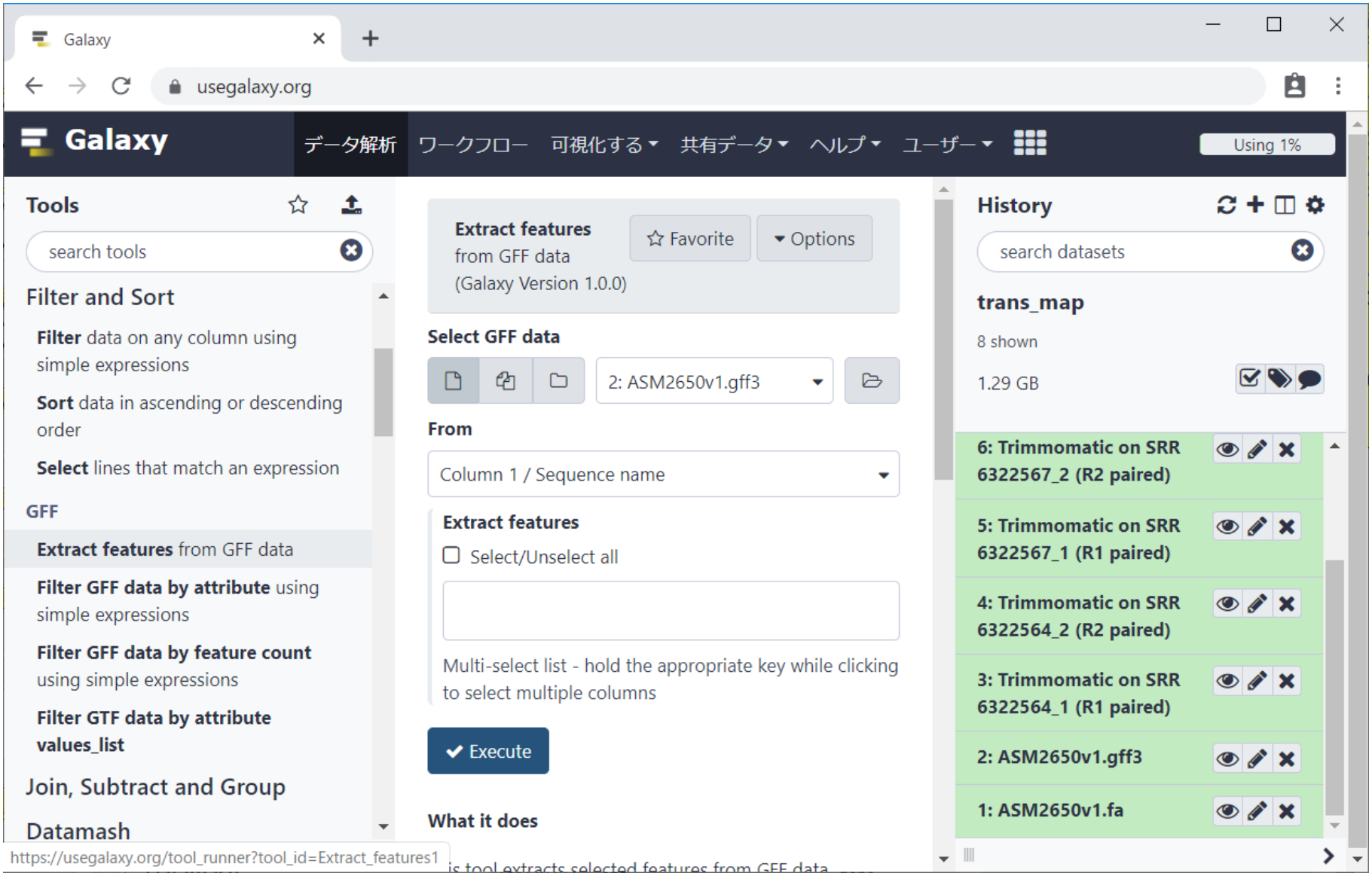
W6-1：Extract feature

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。⑥Extract 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。



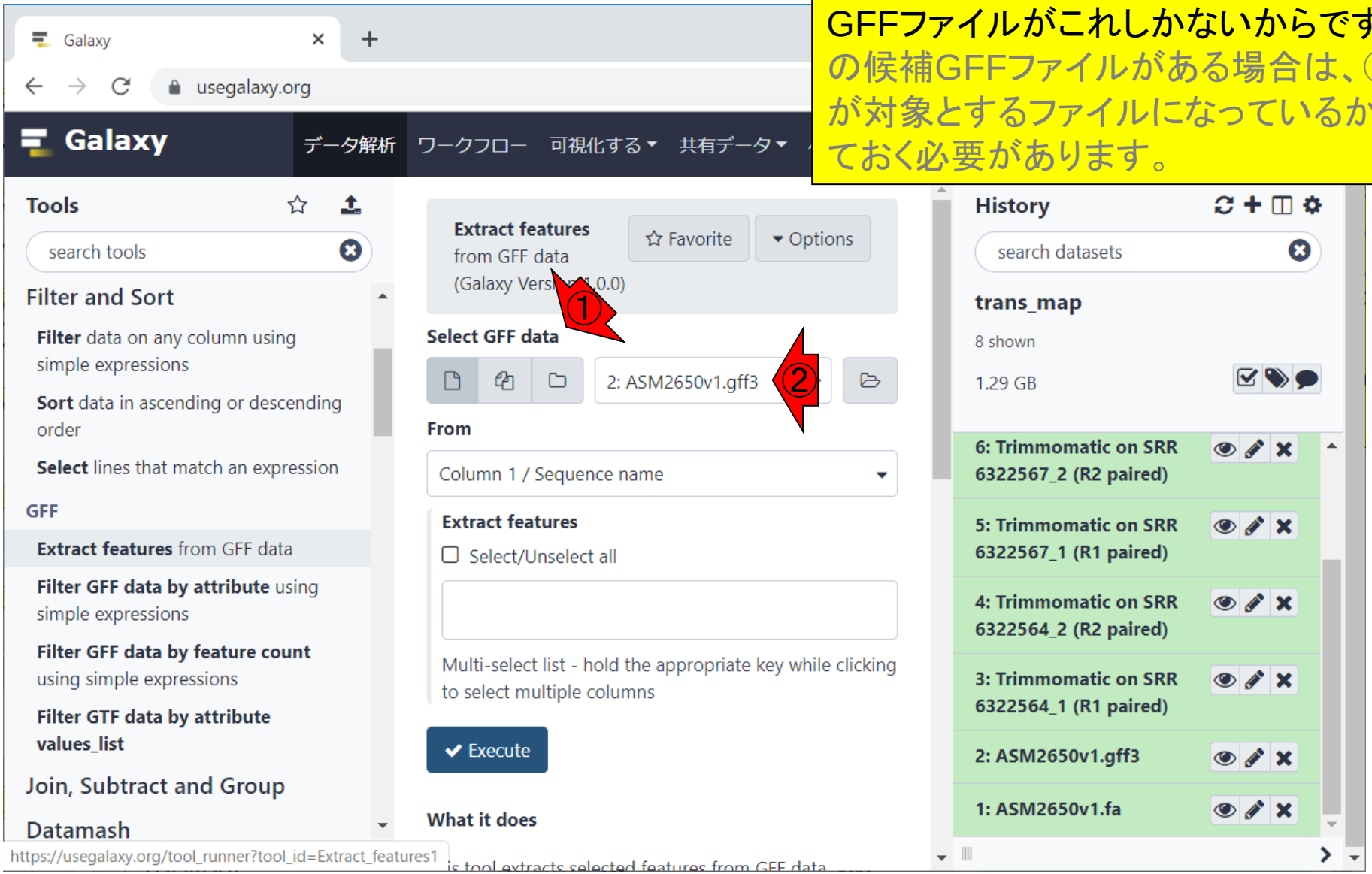
こんな感じになります。

W6-2：Extract features



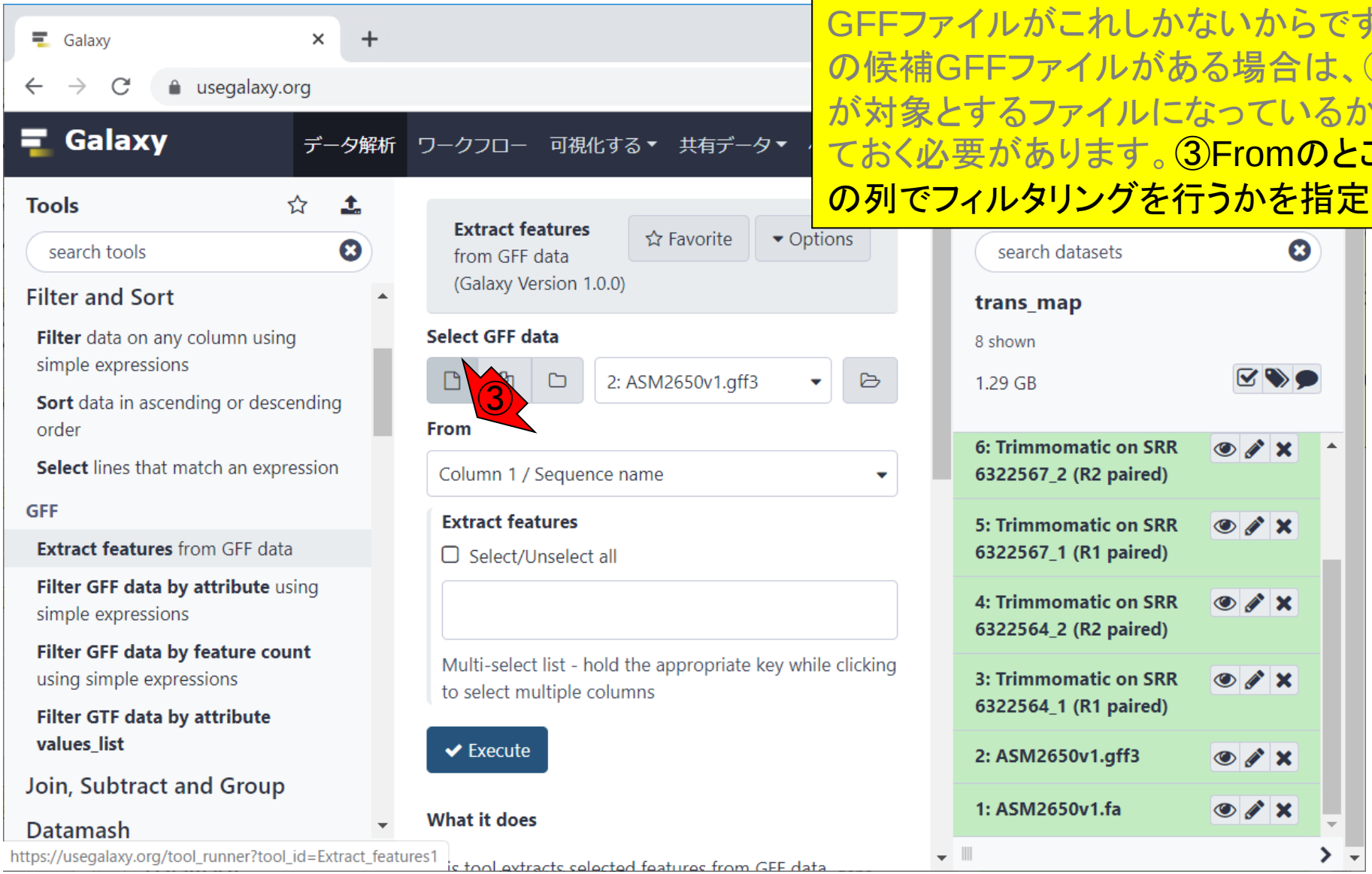
W6-3：Extract feature

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。



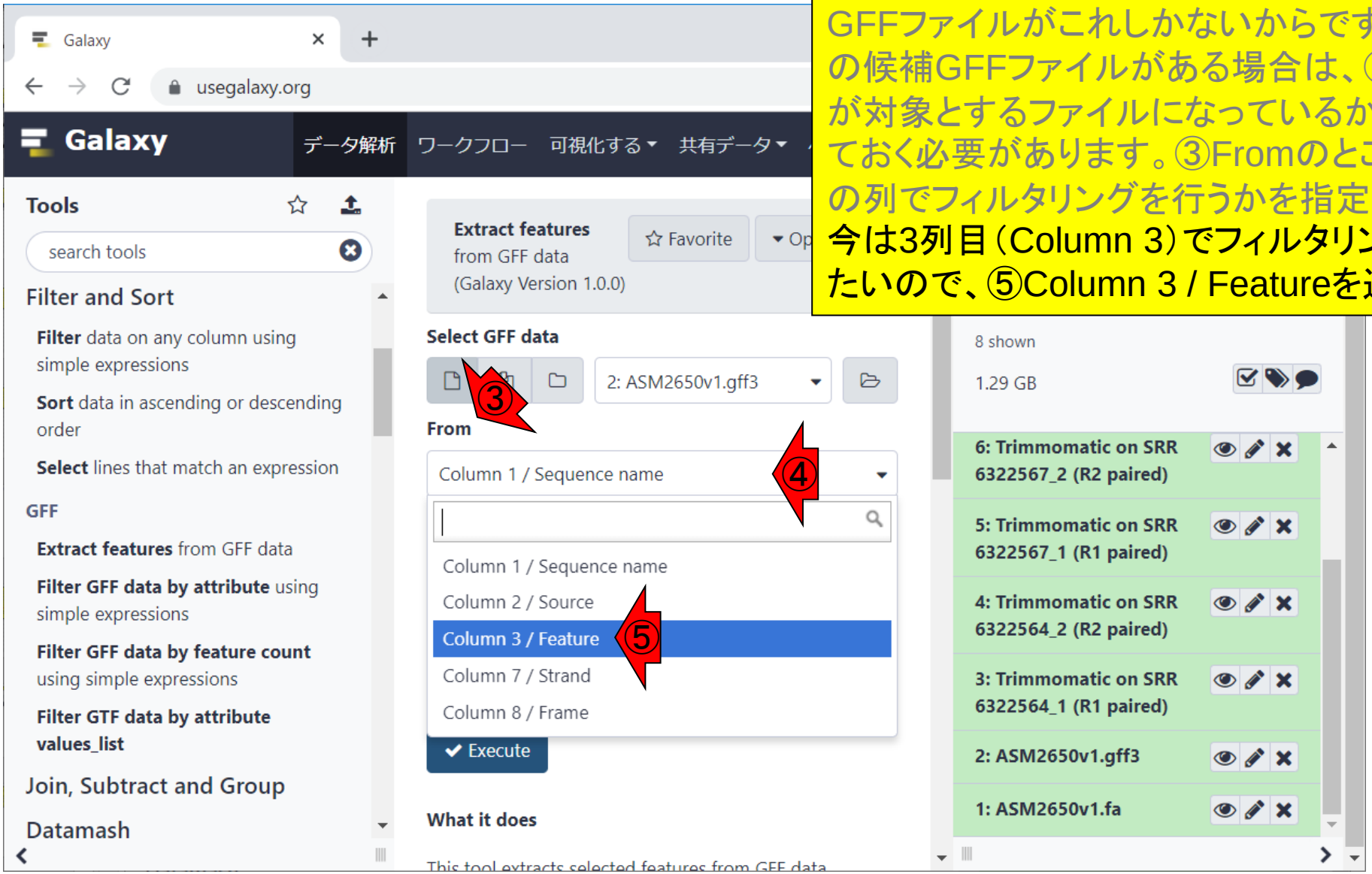
W6-4：Extract feature

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。



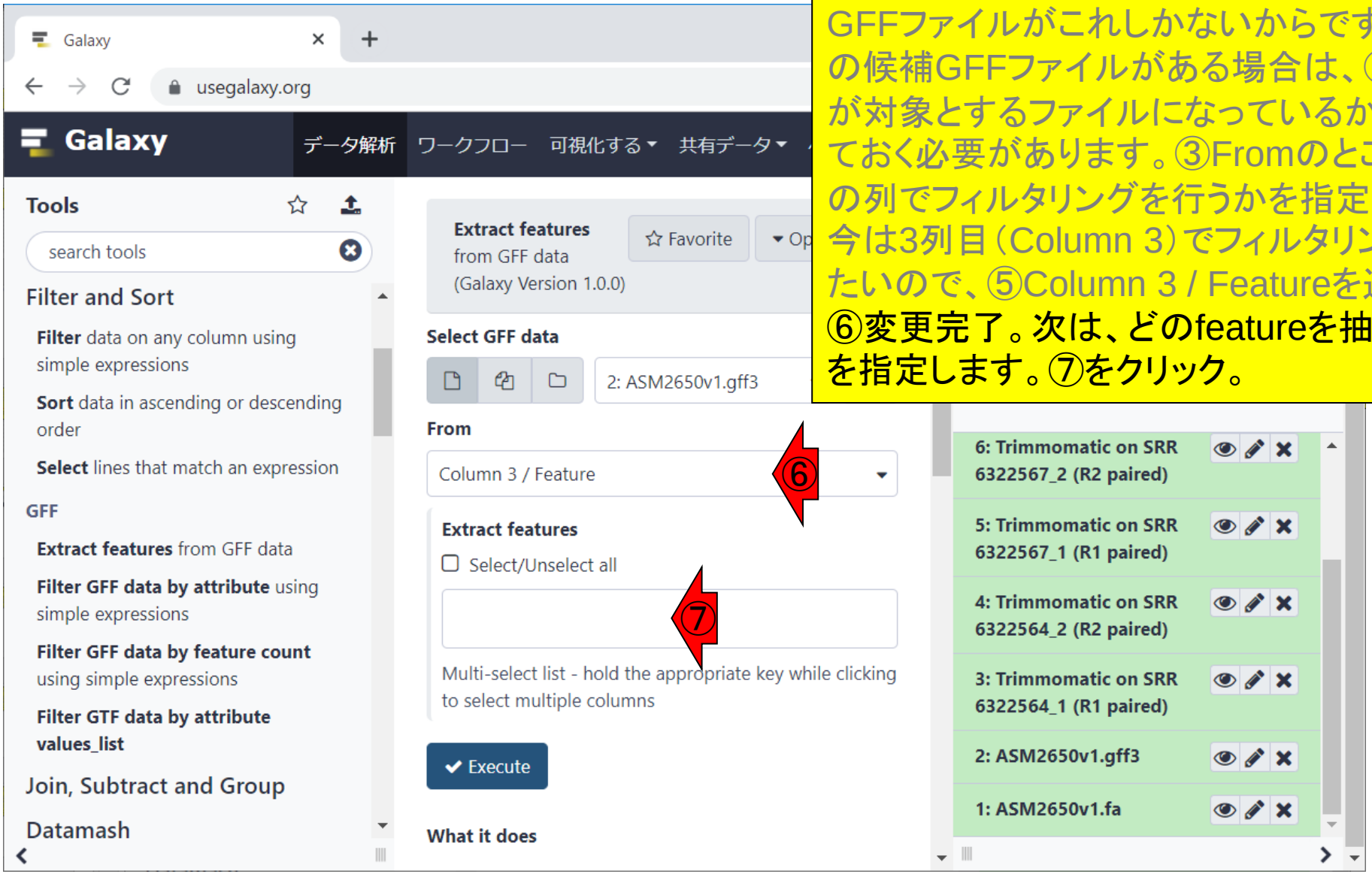
W6-5：Extract feature

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のところで、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。

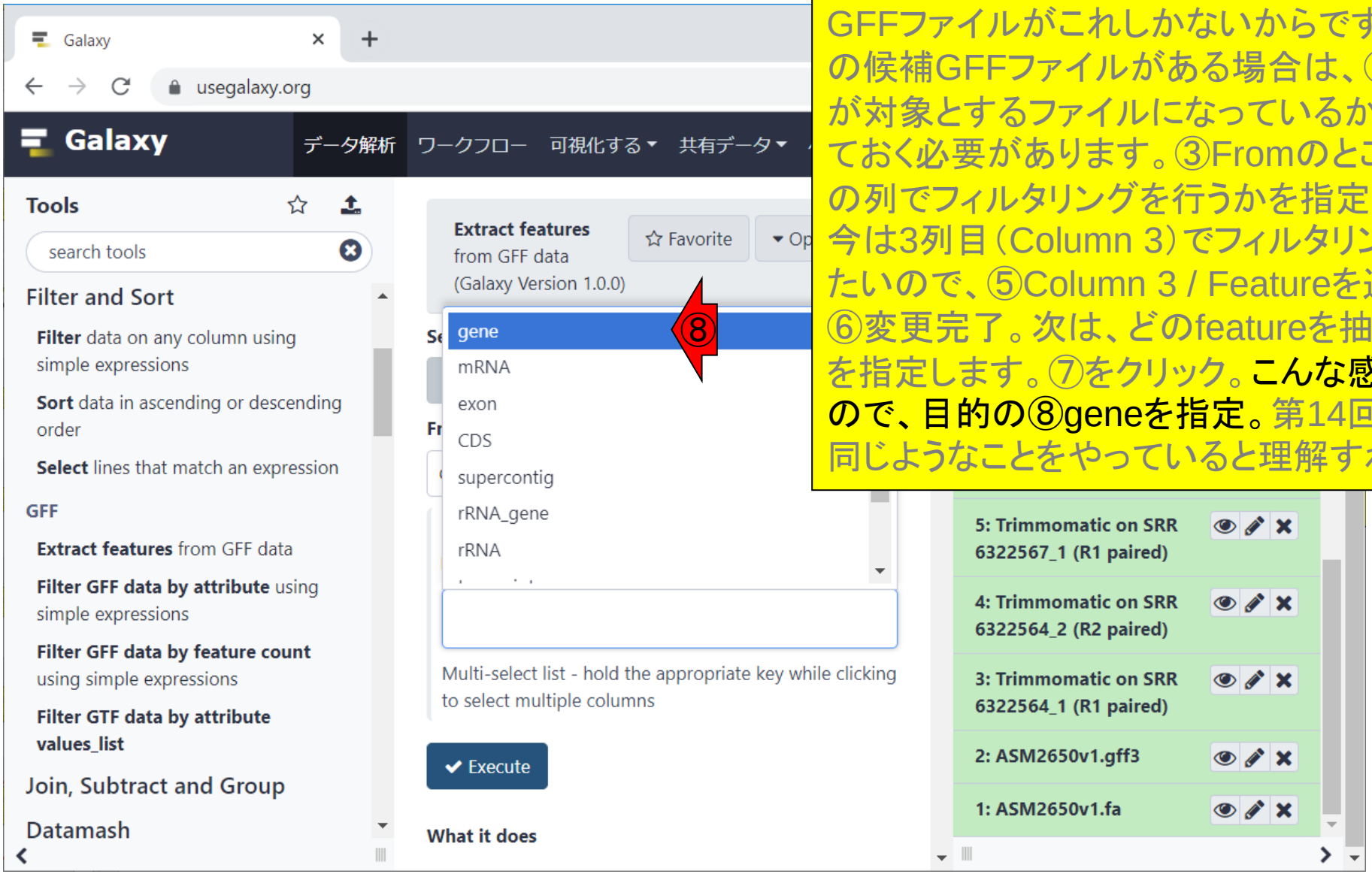


W6-6：Extract feature

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のところで、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。



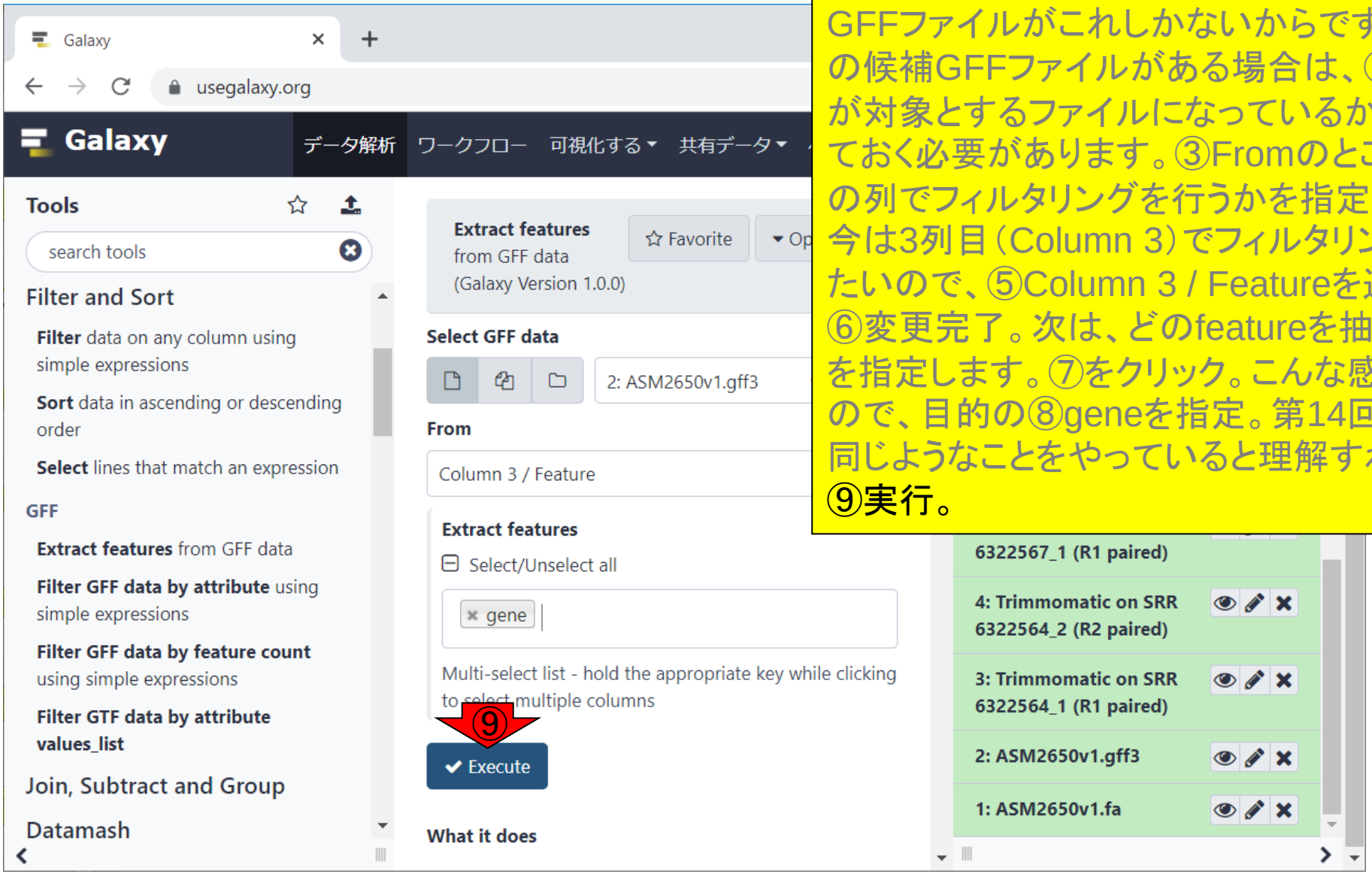
W6-7 : Extract feature



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。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。

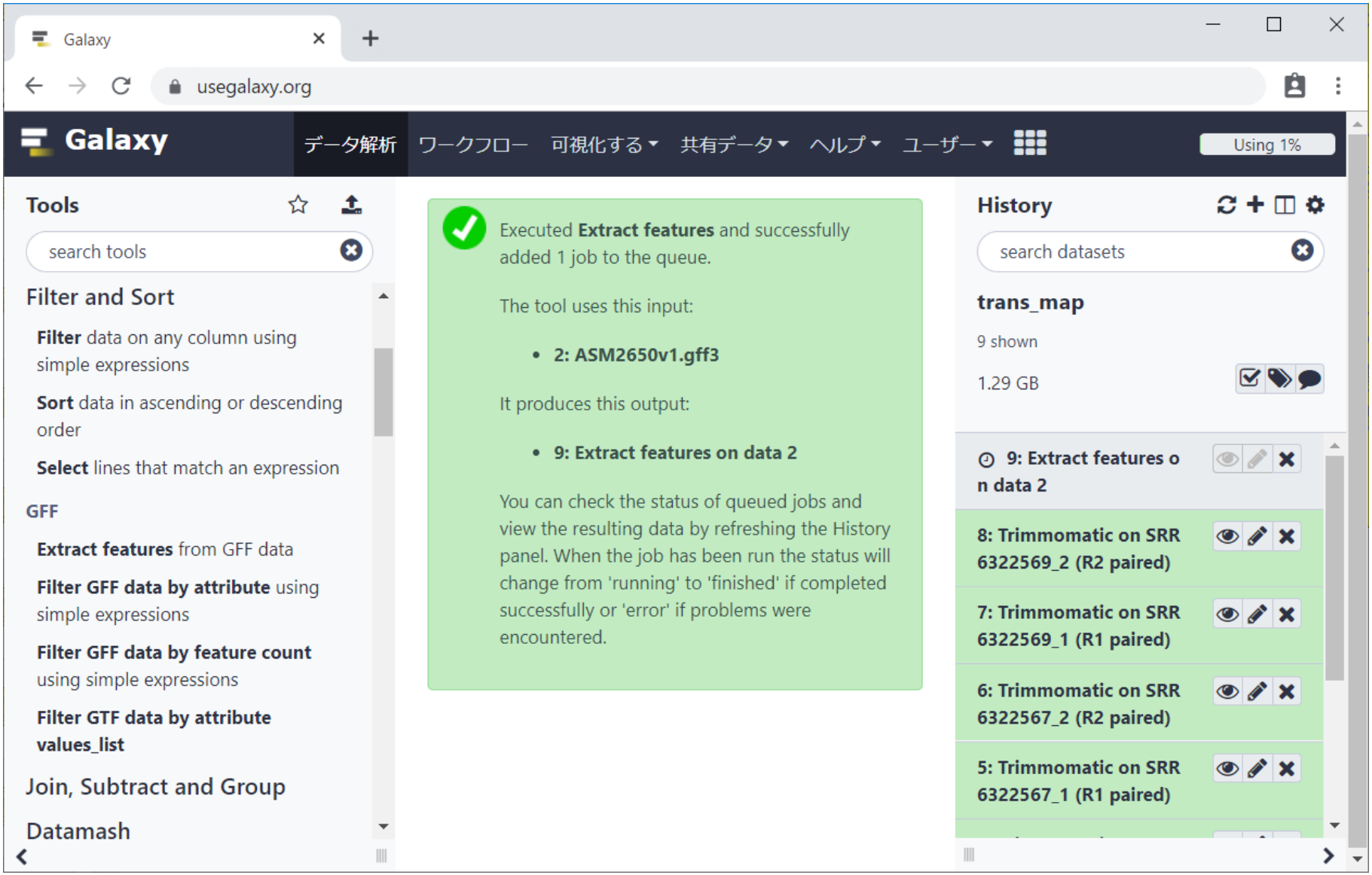
W6-8：Extract feature

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のところで、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理解すればよい。⑨実行。



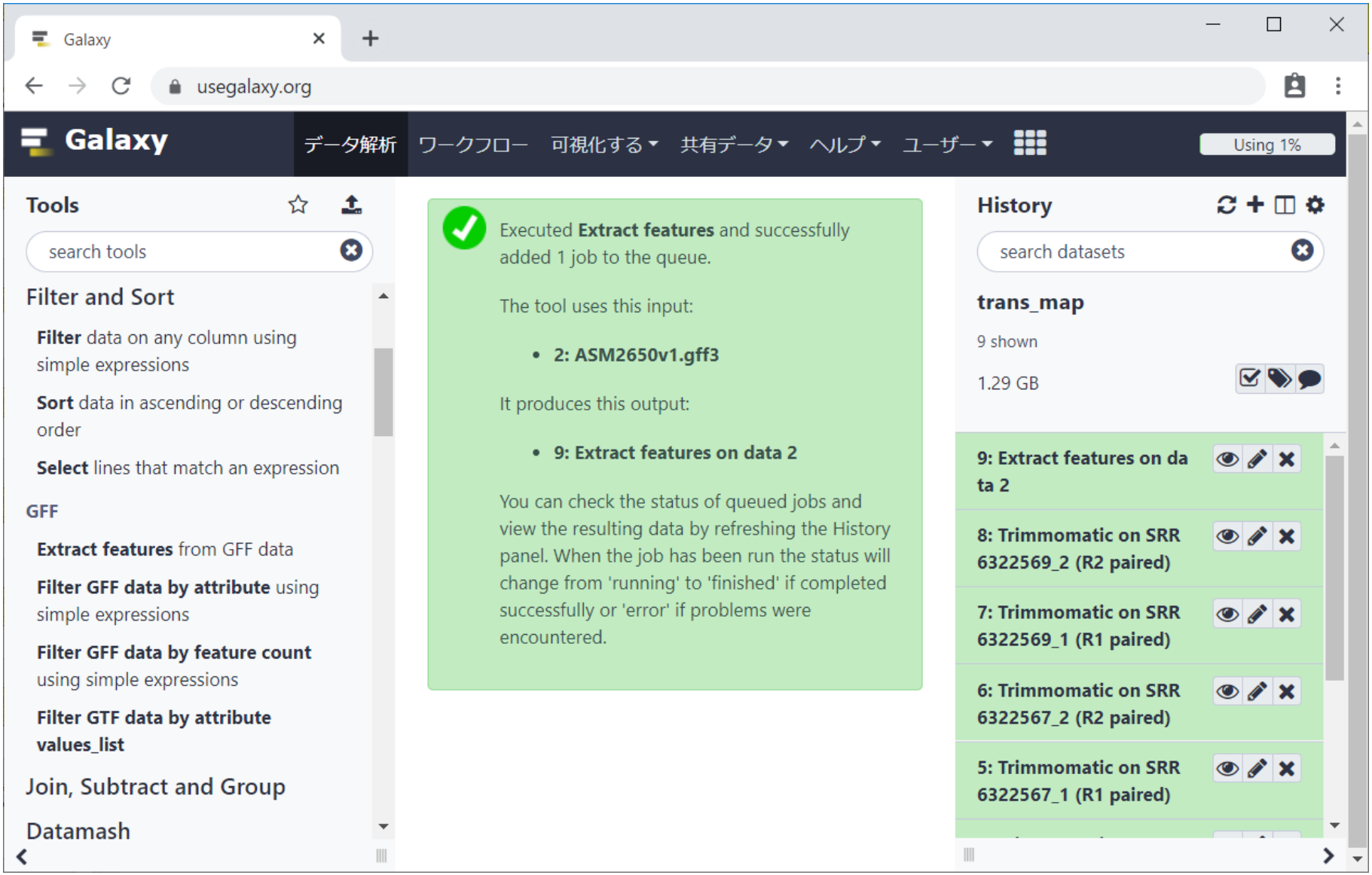
実行中…。

W6-9：Extract features



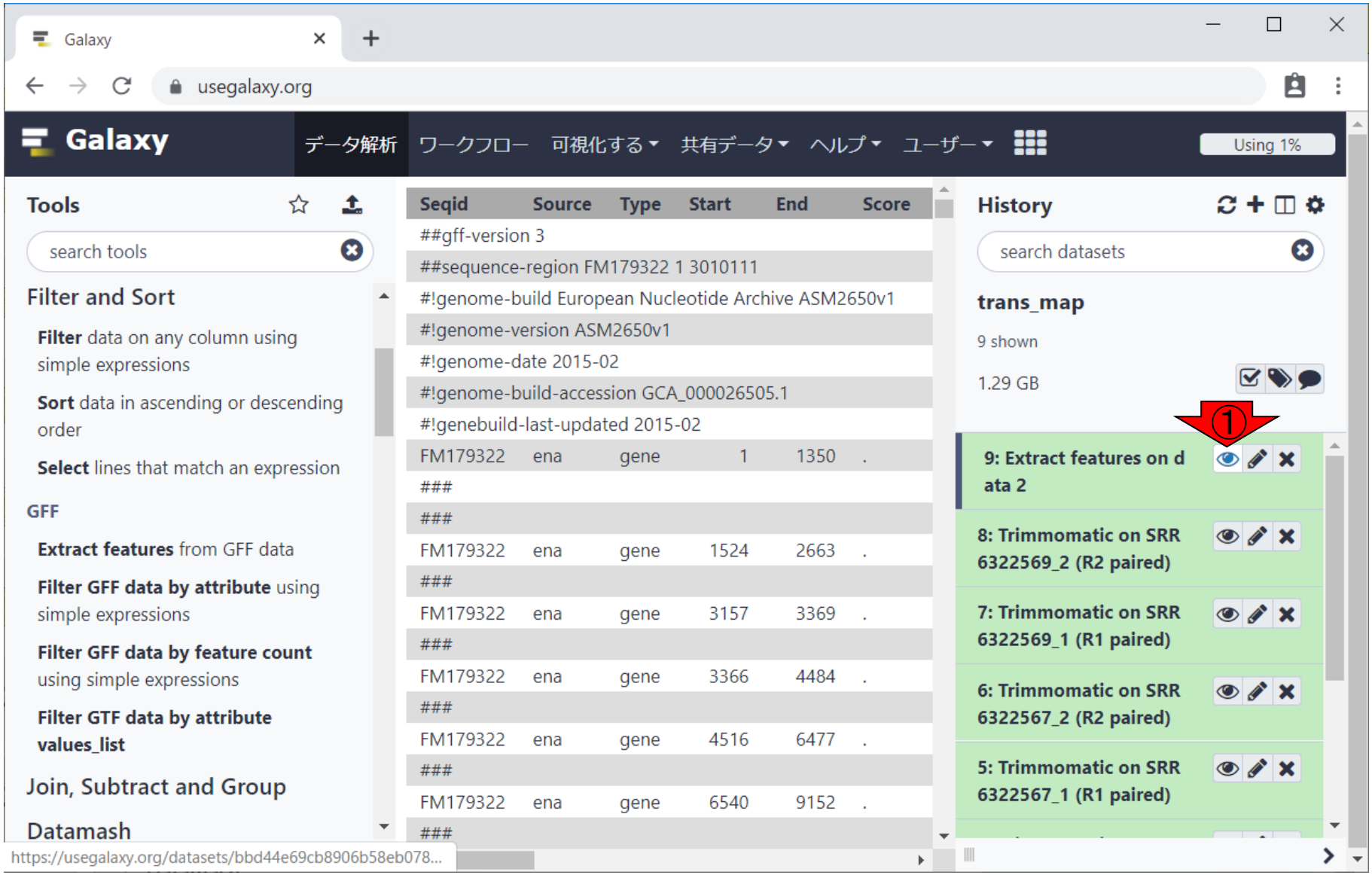
実行完了。

W6-10：Extract features



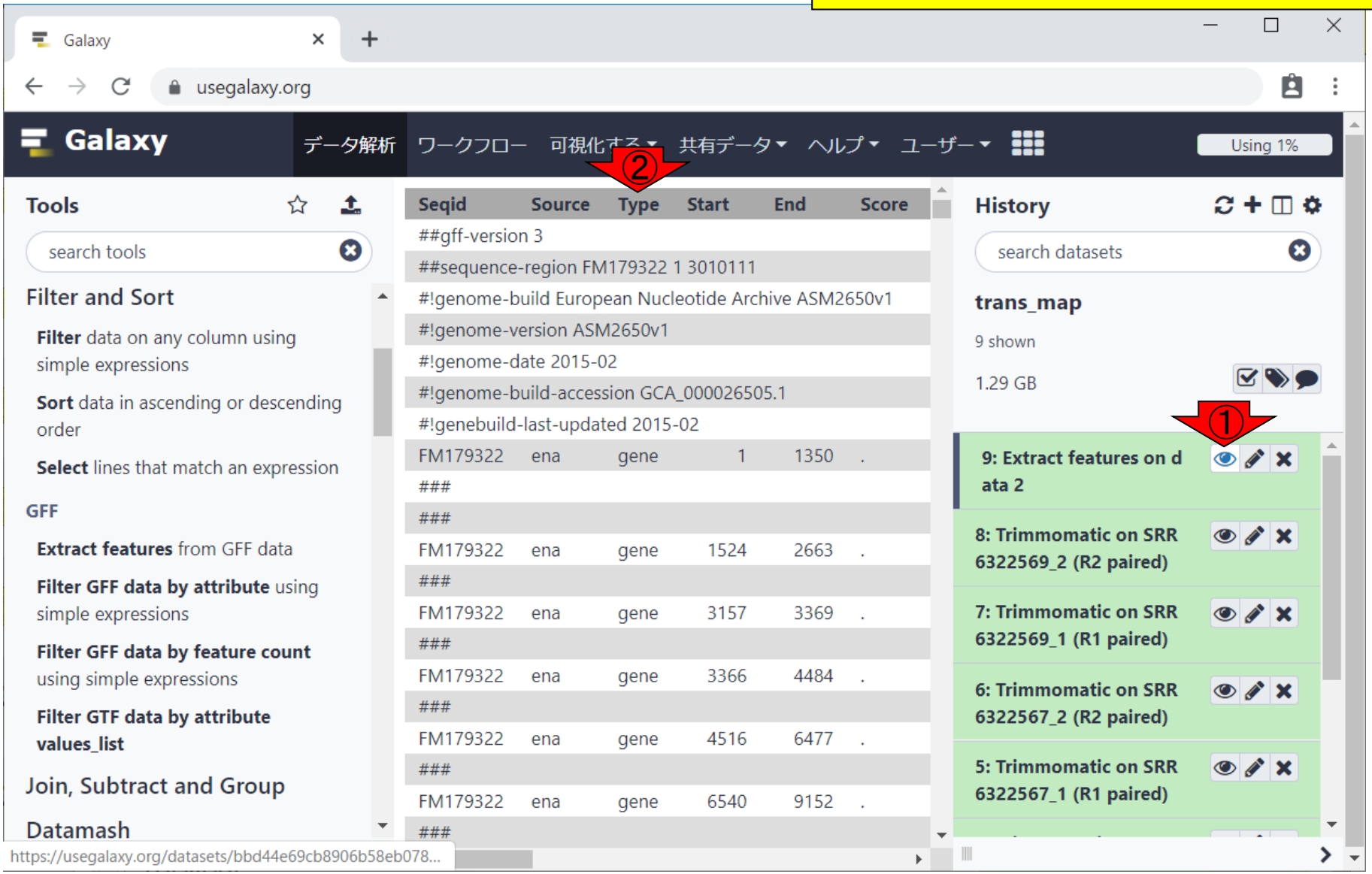
W6-11：Extract features

実行完了。①出力結果の中身を中央パネル上で表示。



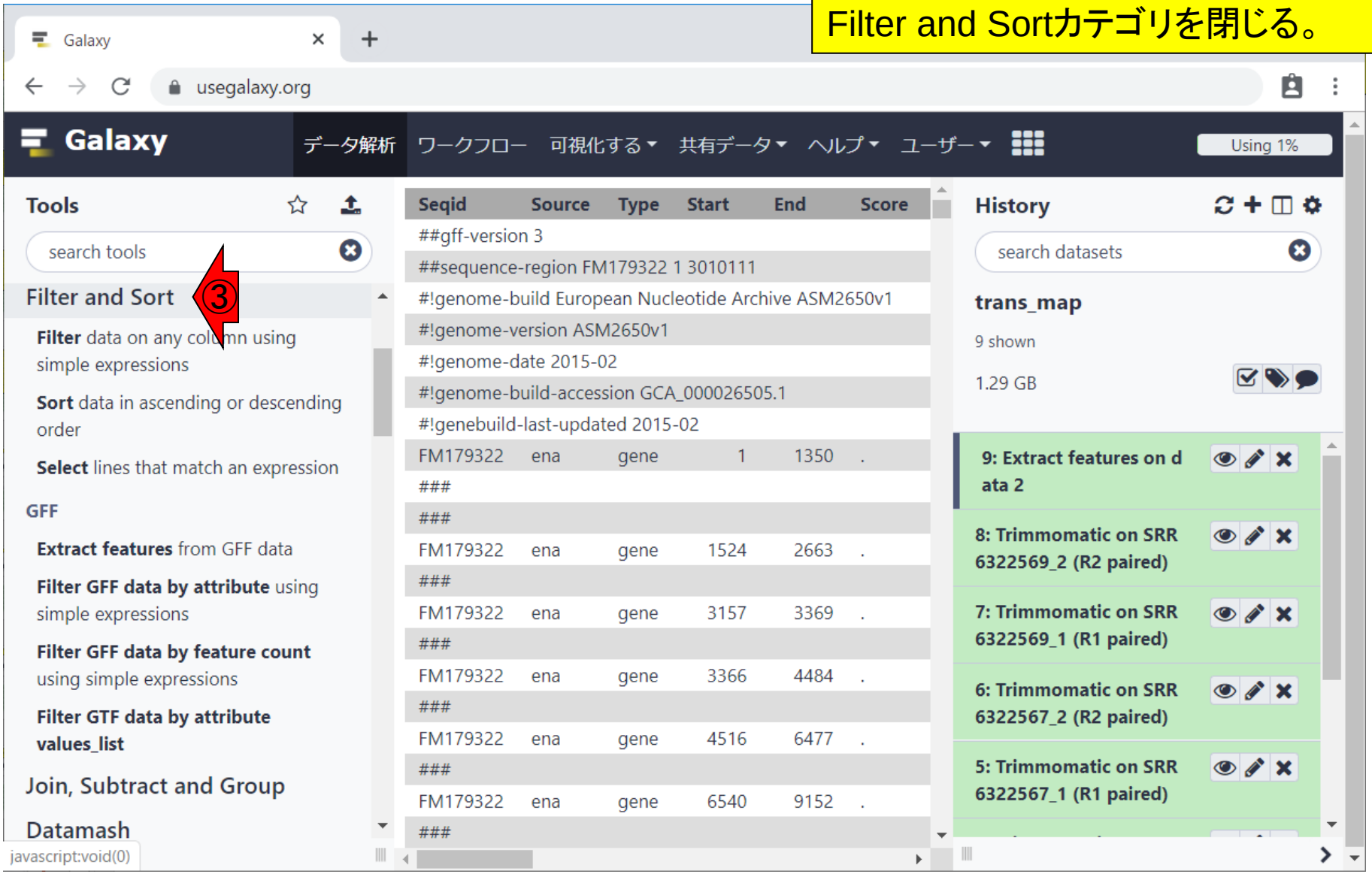
W6-12：Extract featur

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。



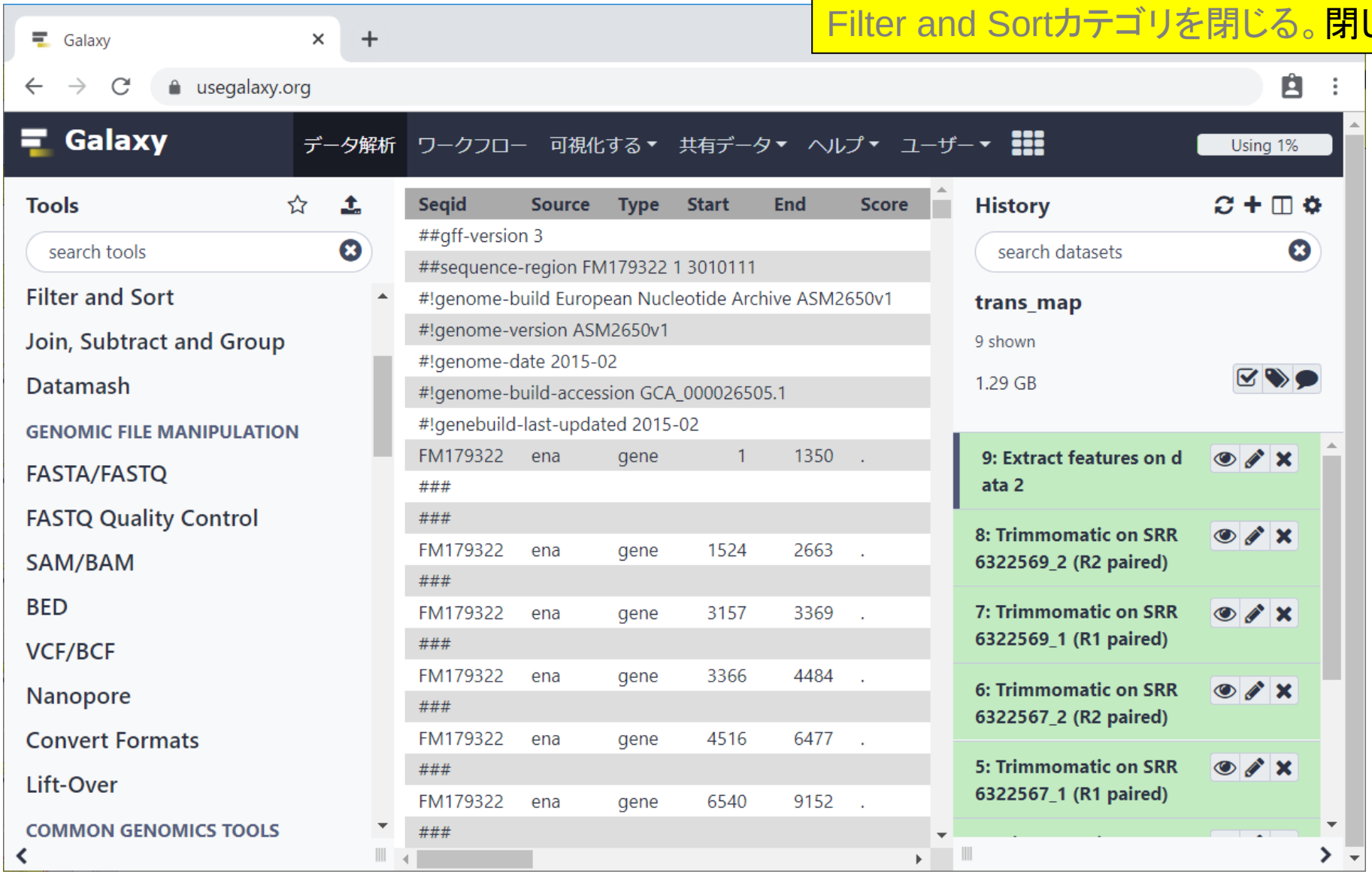
W6-13：Extract featur

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。



W6-14：Extract featur

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。

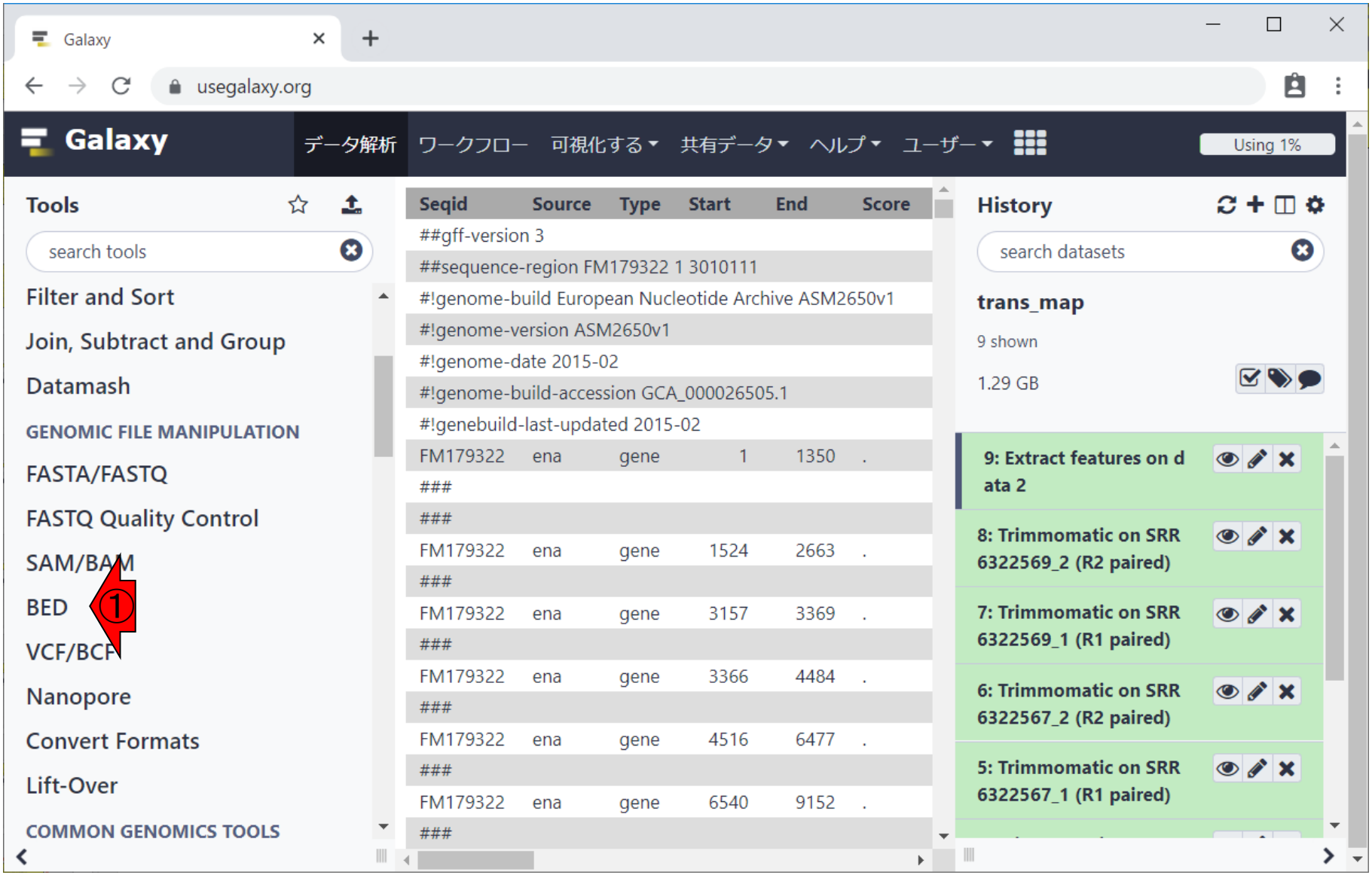


Contents

- 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

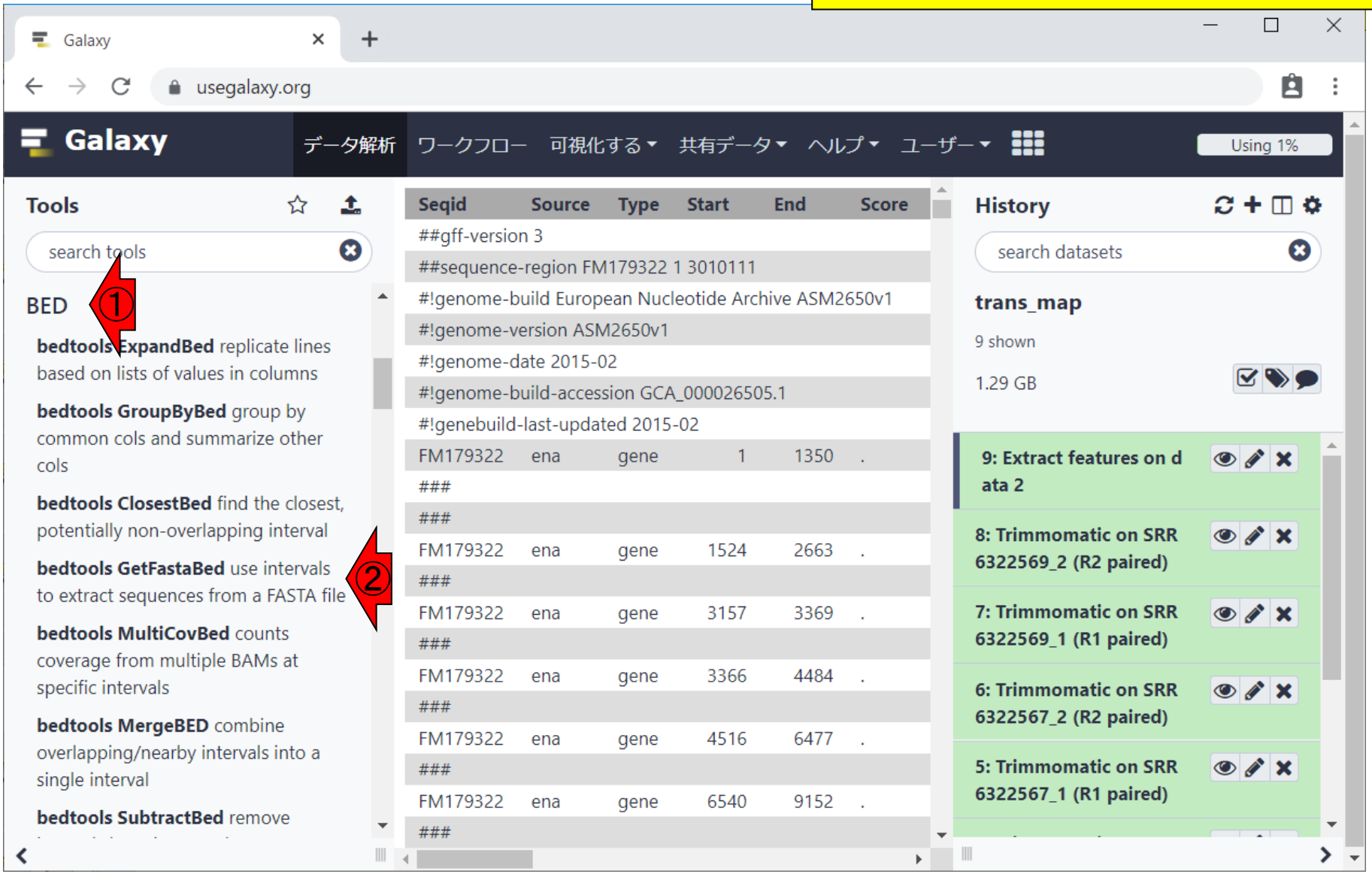
①次はBED。

W7-1：bedtools GetFastaBed



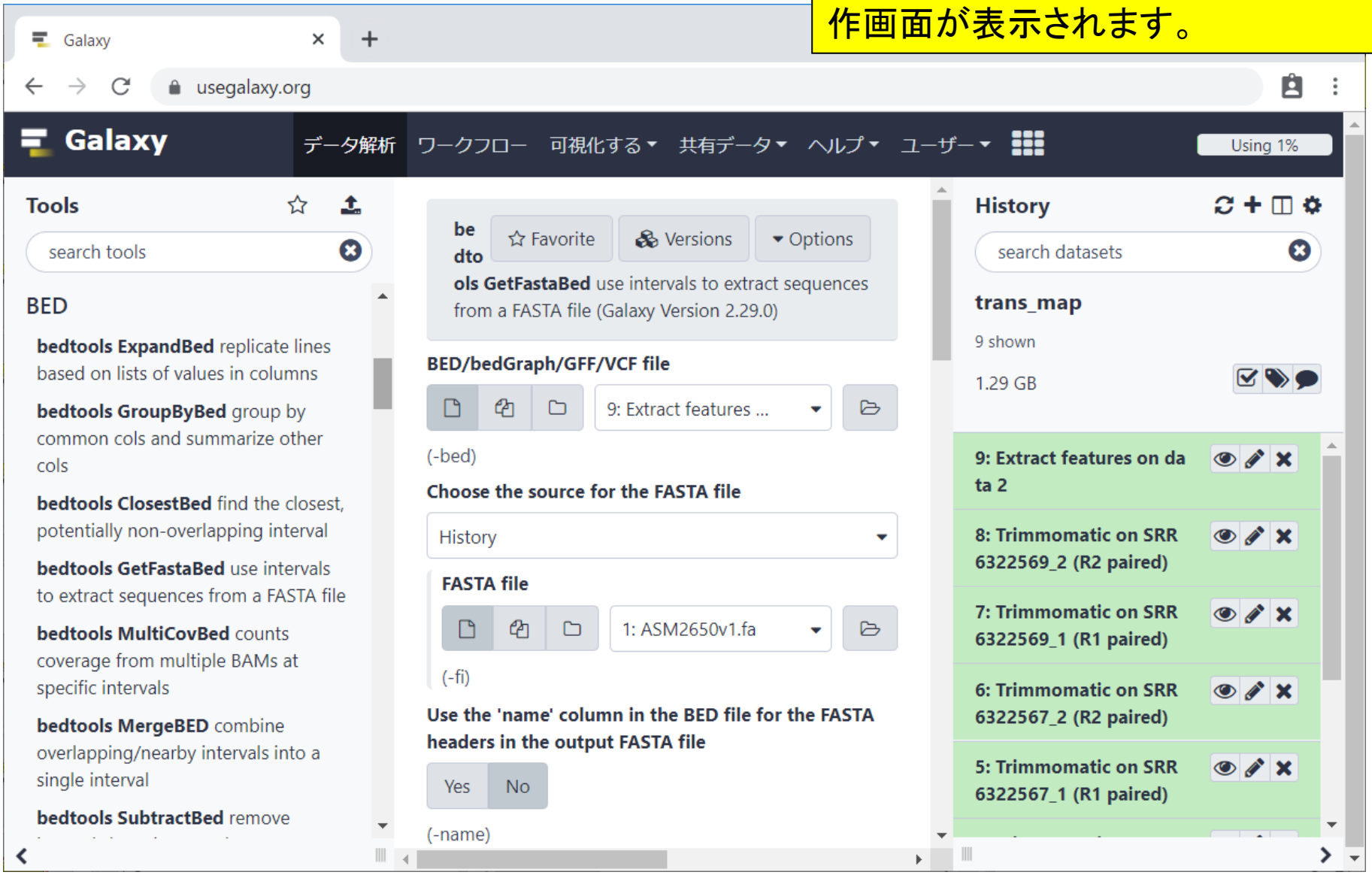
W7-2：bedtools GetFa

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。



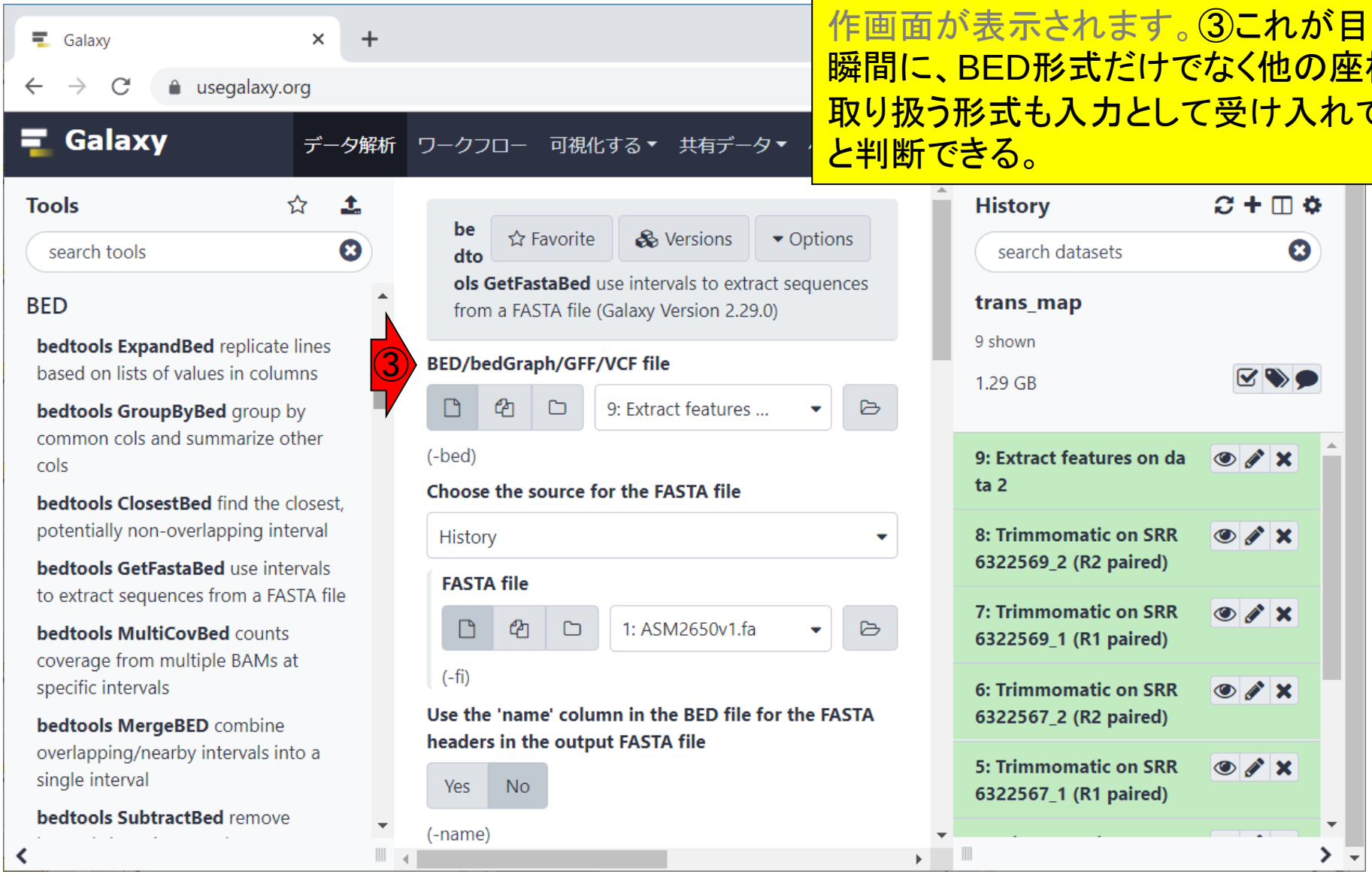
W7-3：bedtools GetFa

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。



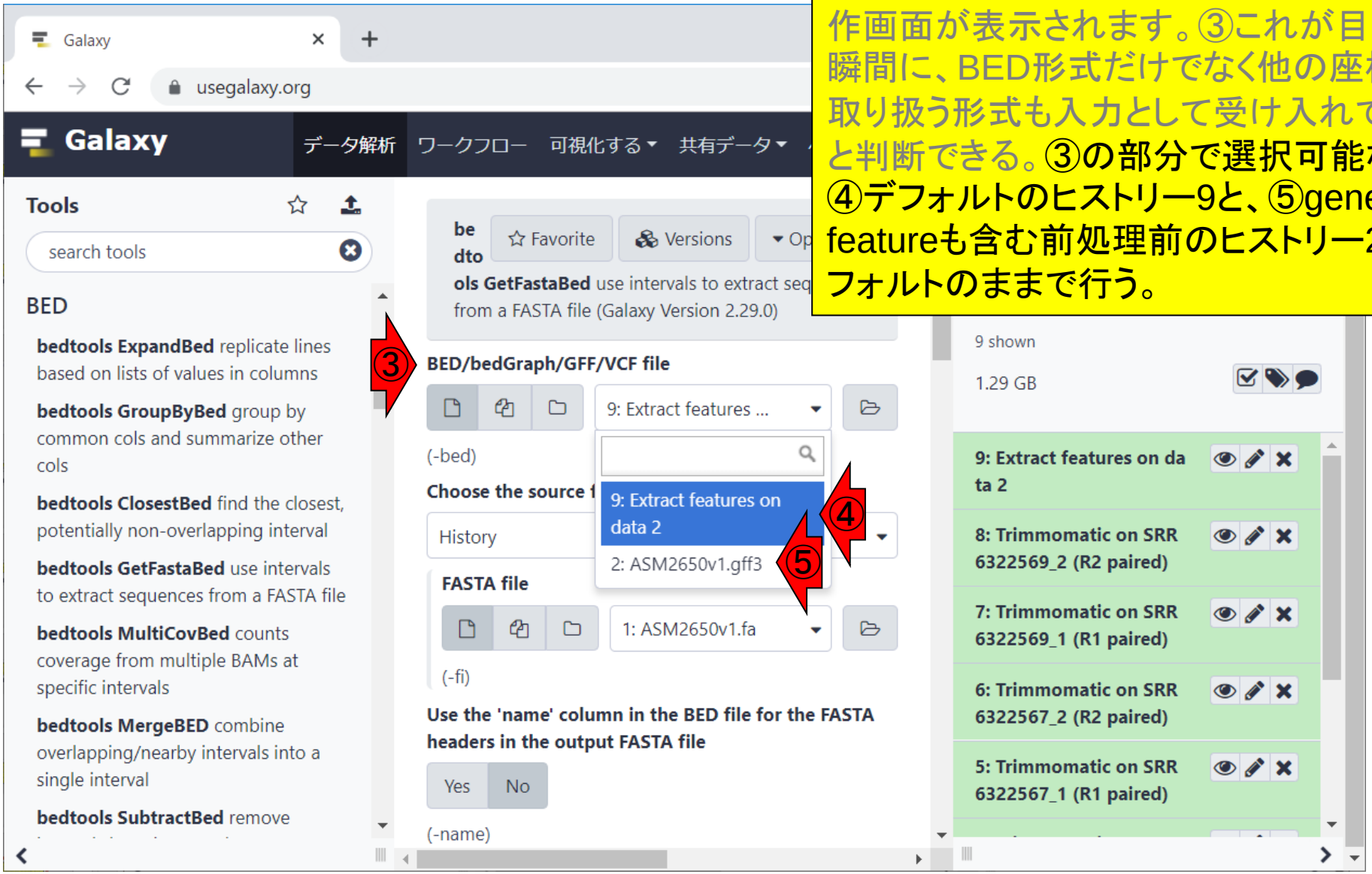
W7-4：bedtools GetFa

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。



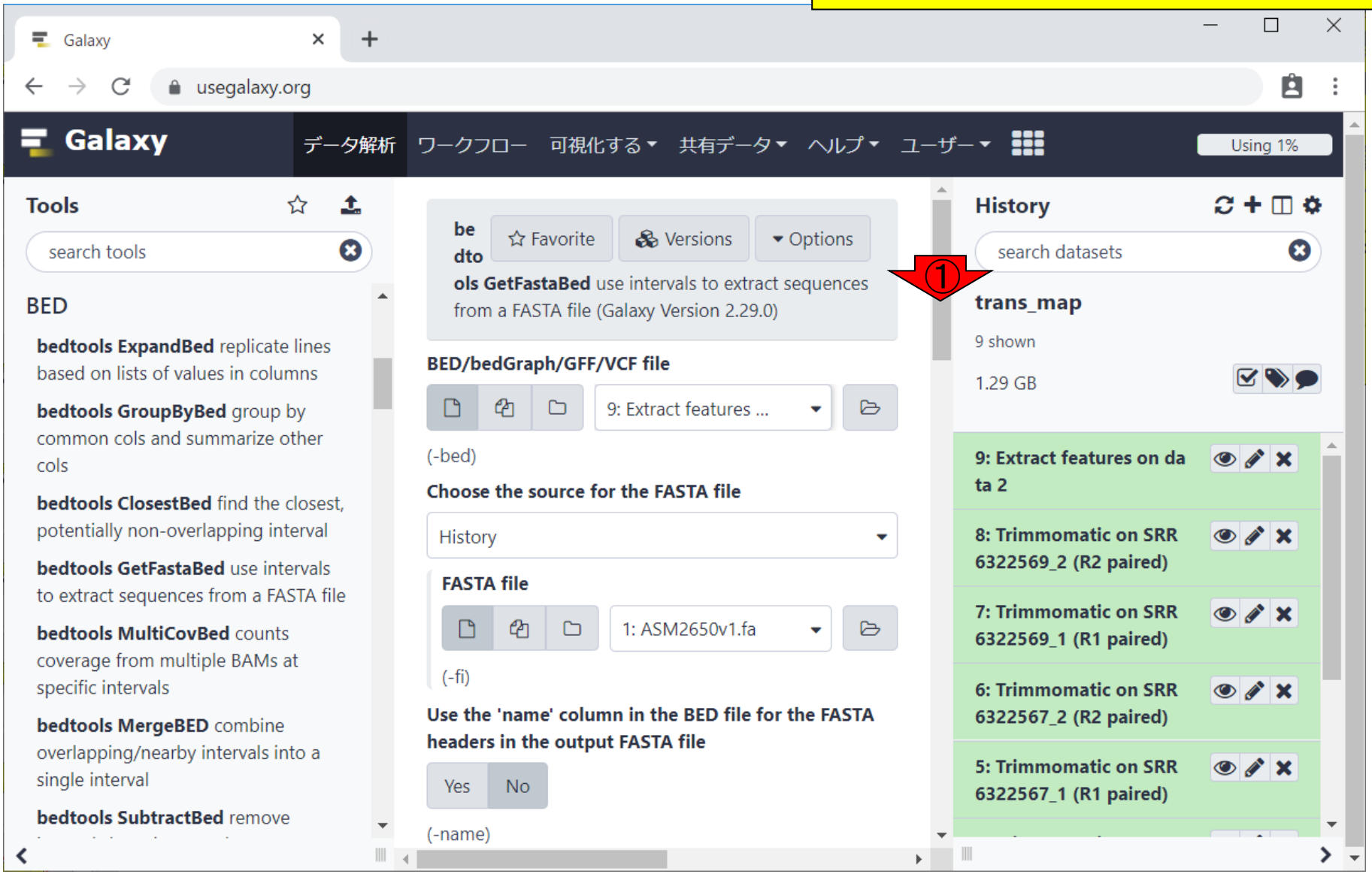
W7-5：bedtools GetFa

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。



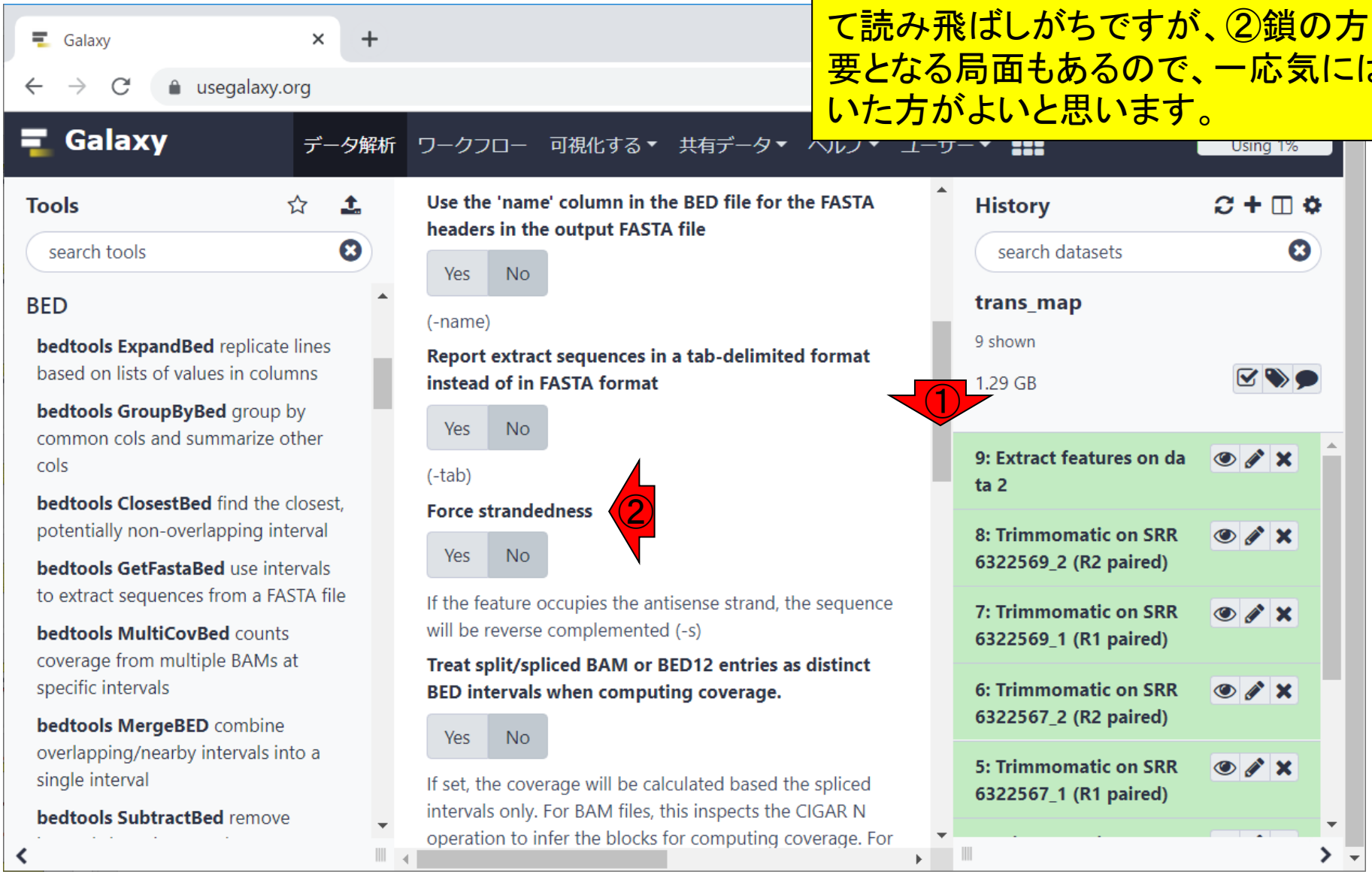
W7-6：bedtools GetFa

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。



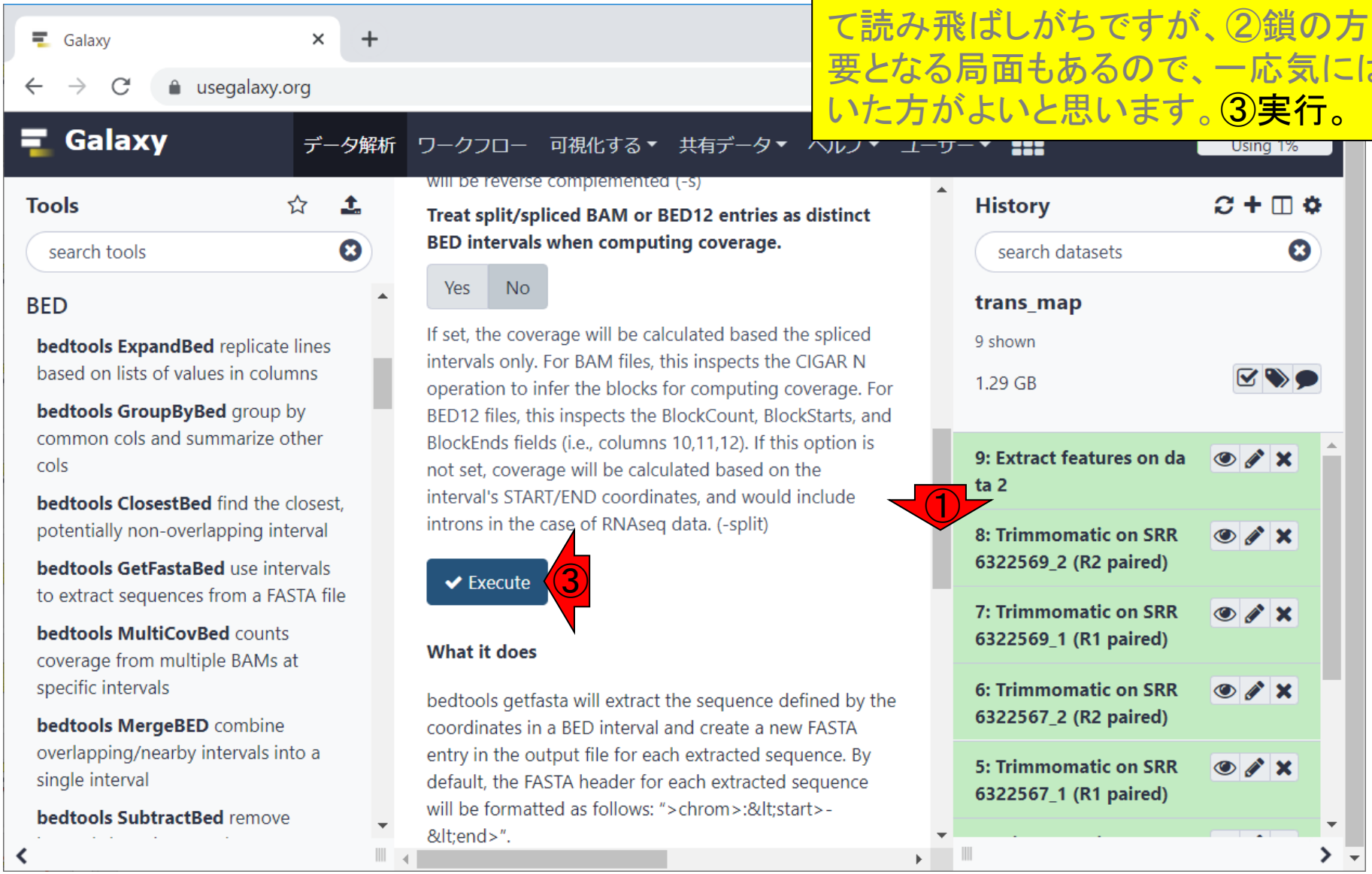
W7-7：bedtools GetFa

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。



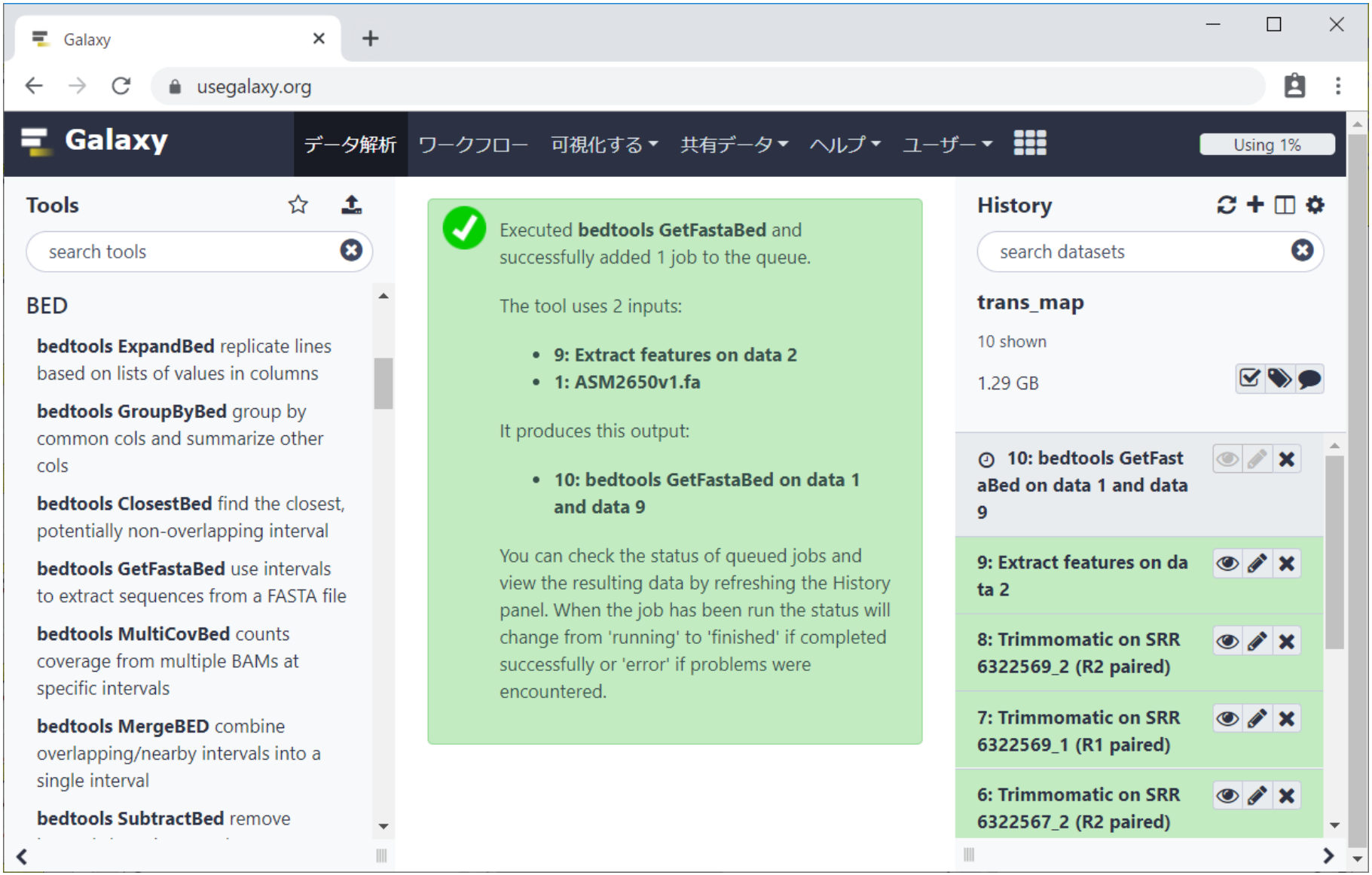
W7-8：bedtools GetFa

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。



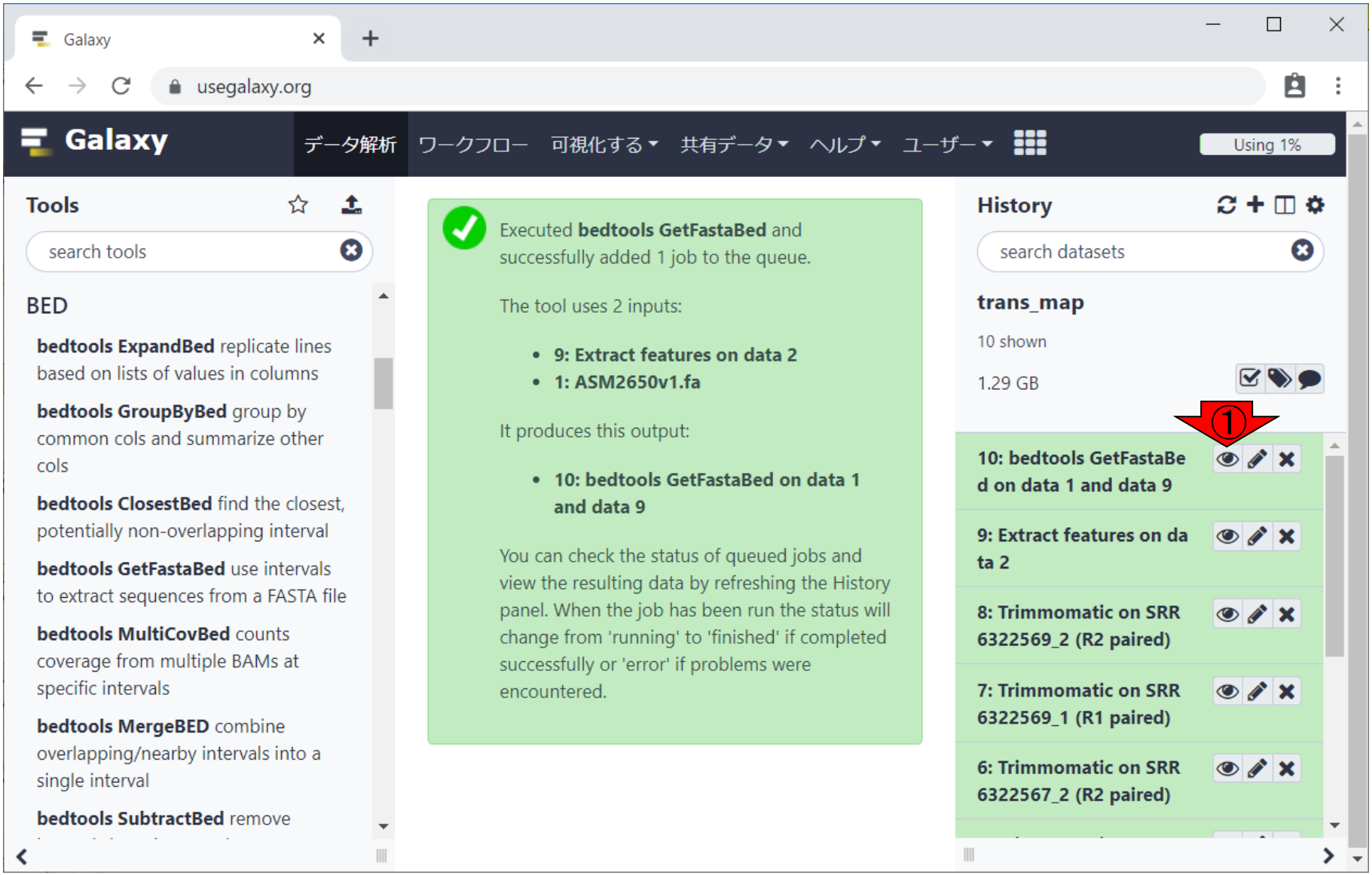
実行中…

W7-9：bedtools GetFastaBed



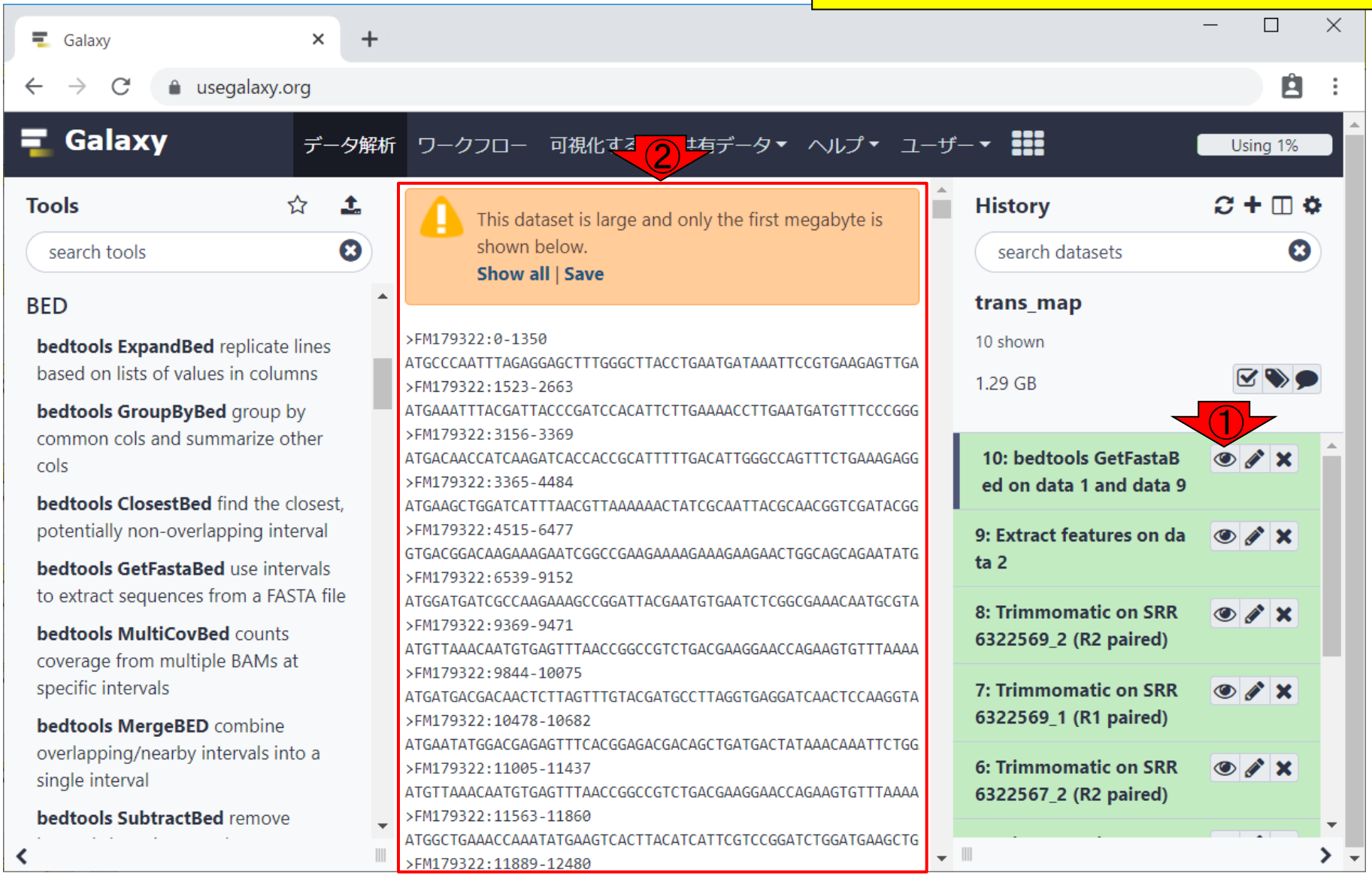
実行完了。①ヒストリー10の中身を中央パネル上に表示。

W7-10：bedtools GetFastaBed



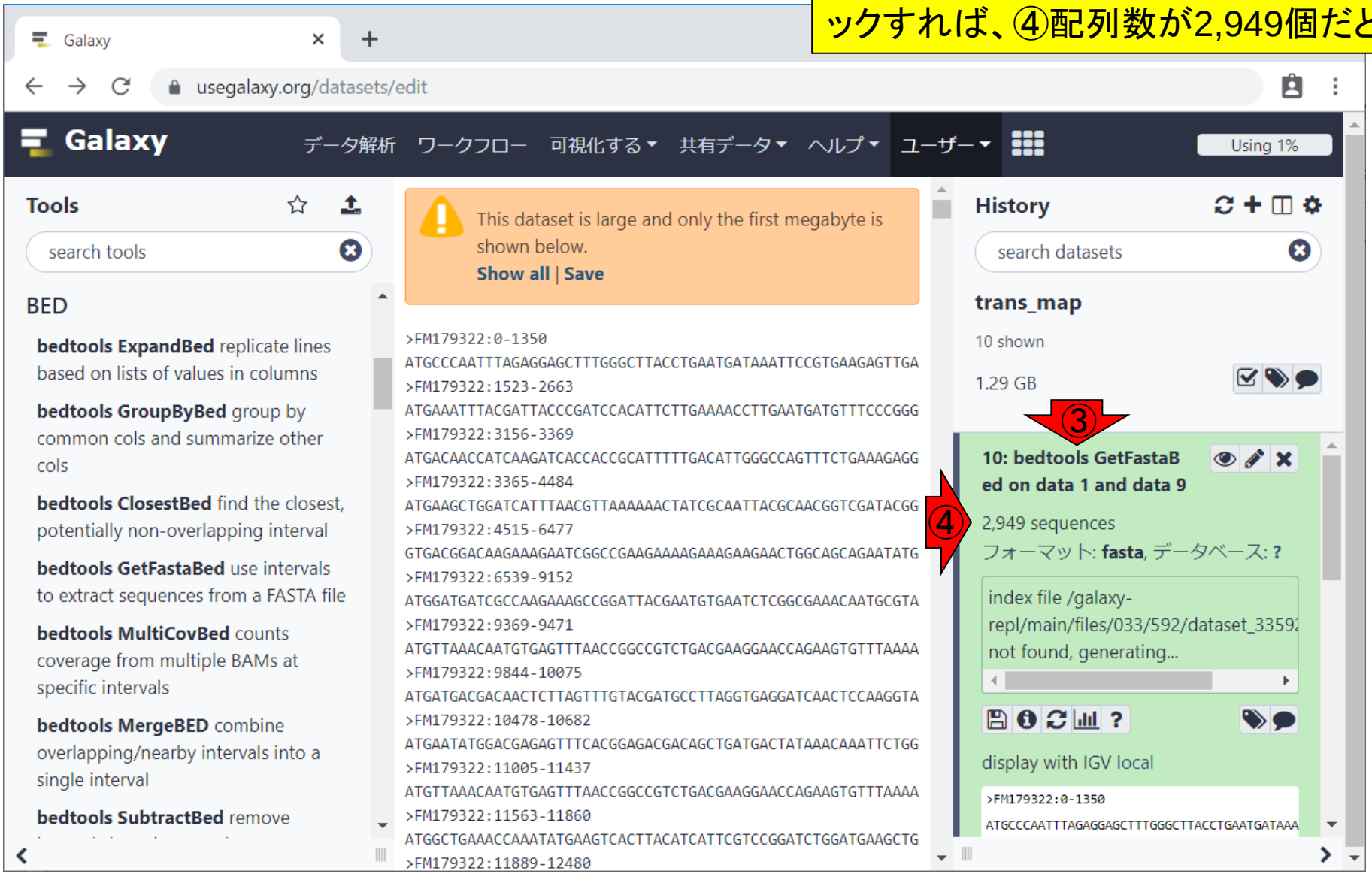
W7-11：bedtools GetF

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。



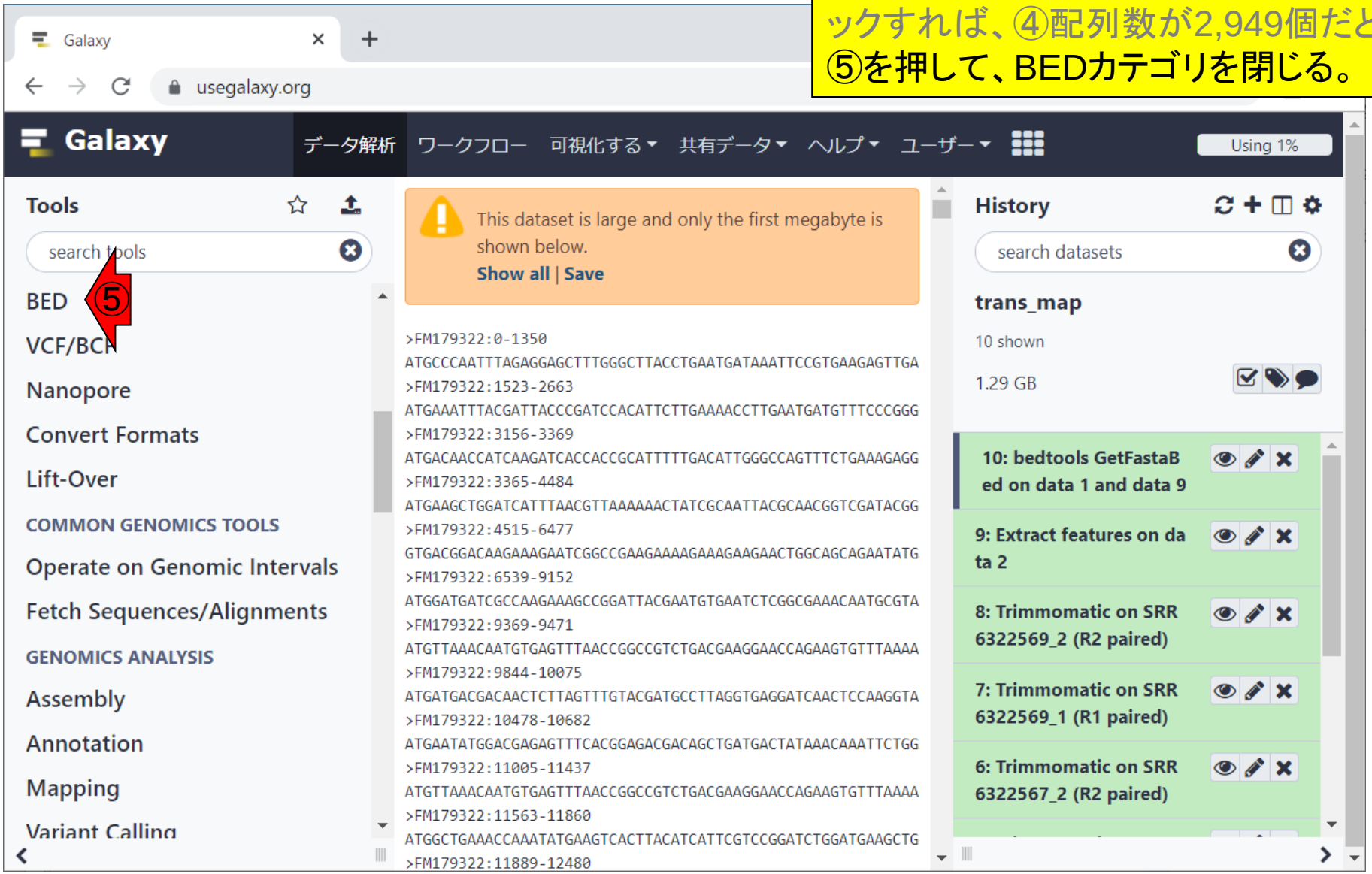
W7-12：bedtools GetF

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。



W7-13：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。
⑤を押して、BEDカテゴリを閉じる。

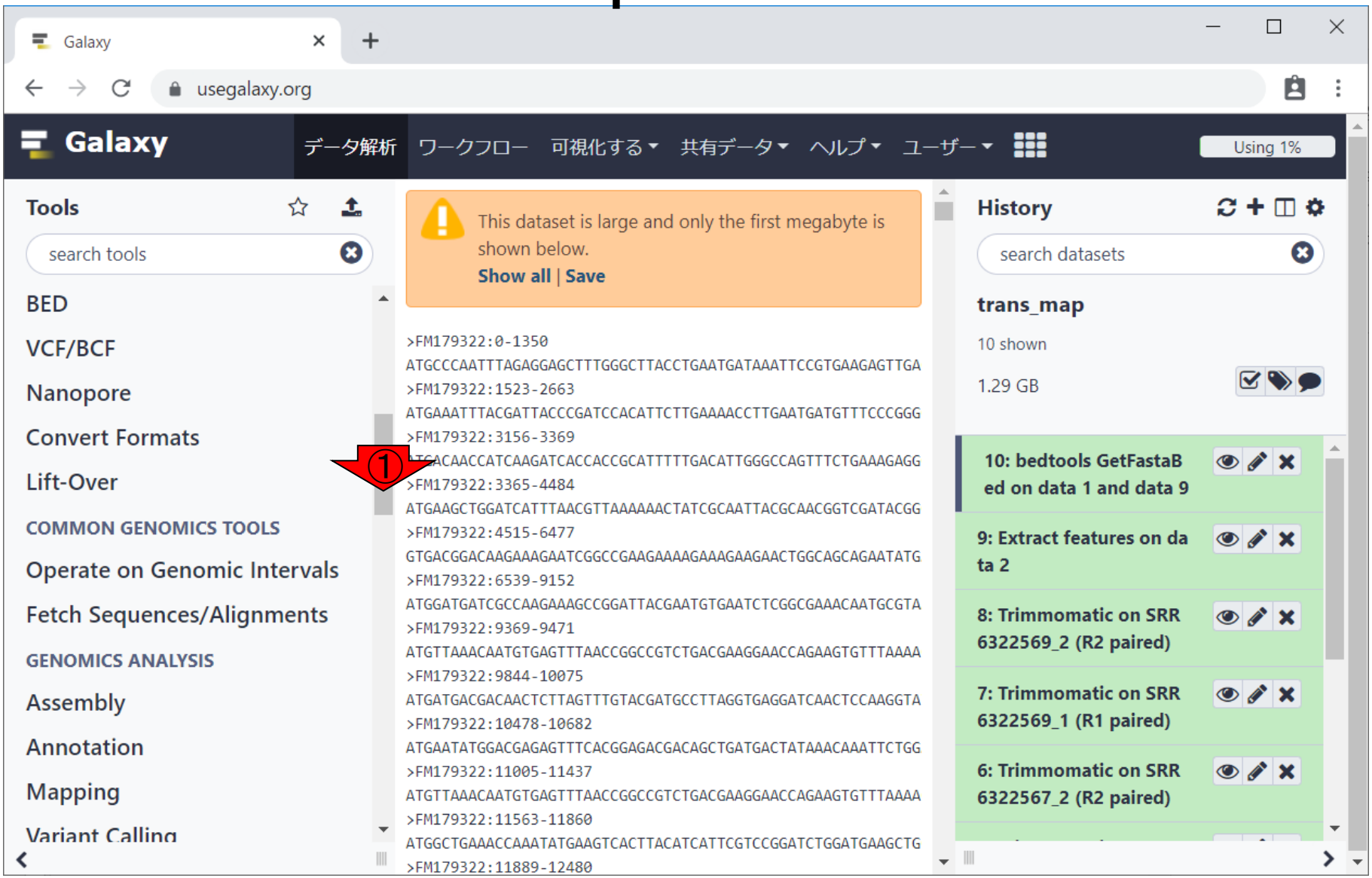


Contents

- 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

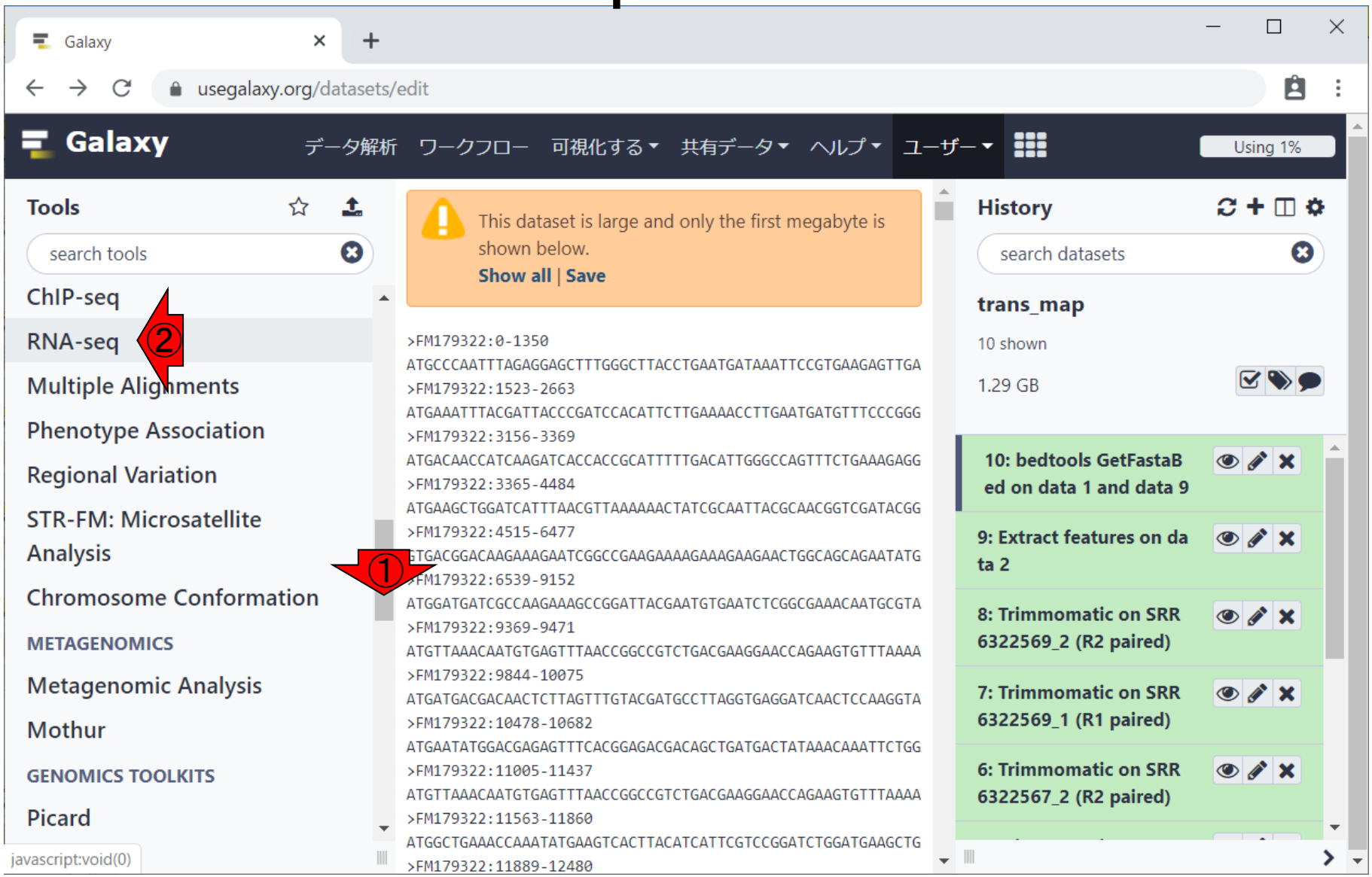
①ツール選択パネル下部に移動。

W8-1：Kallisto quant



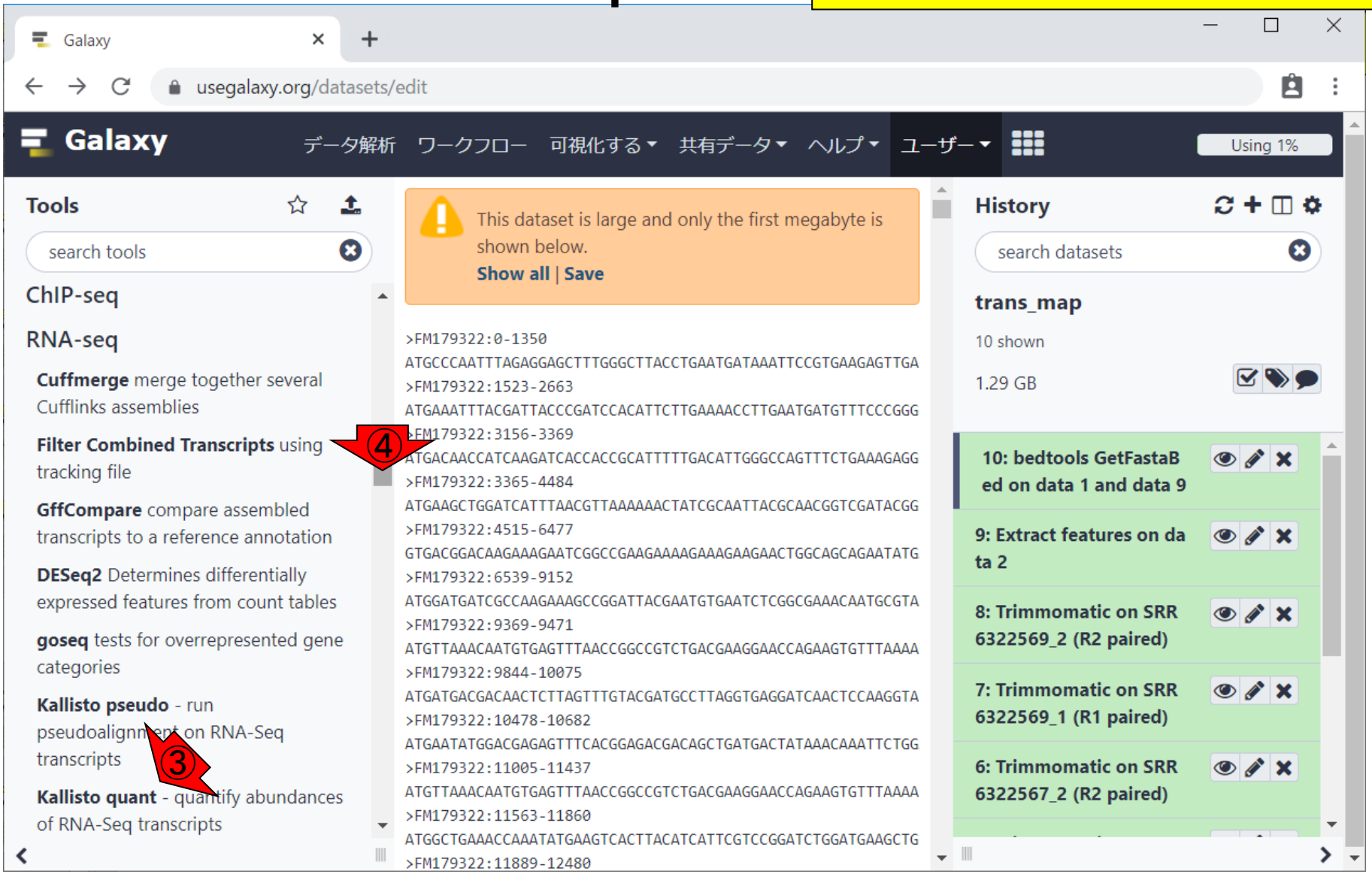
W8-1：Kallisto quant

①ツール選択パネル下部に移動。②RNA-seq。



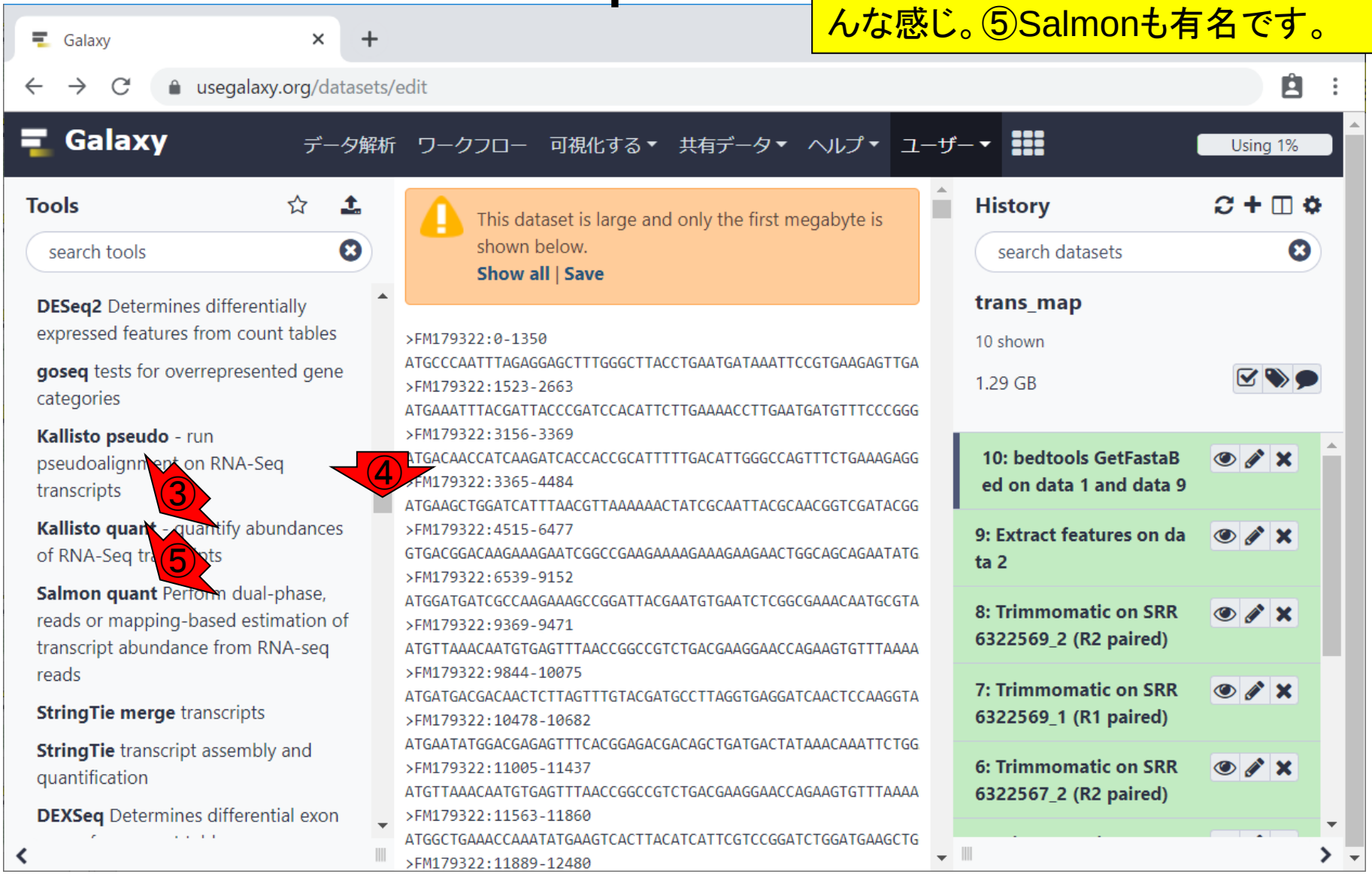
W8-1：Kallisto quant

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。



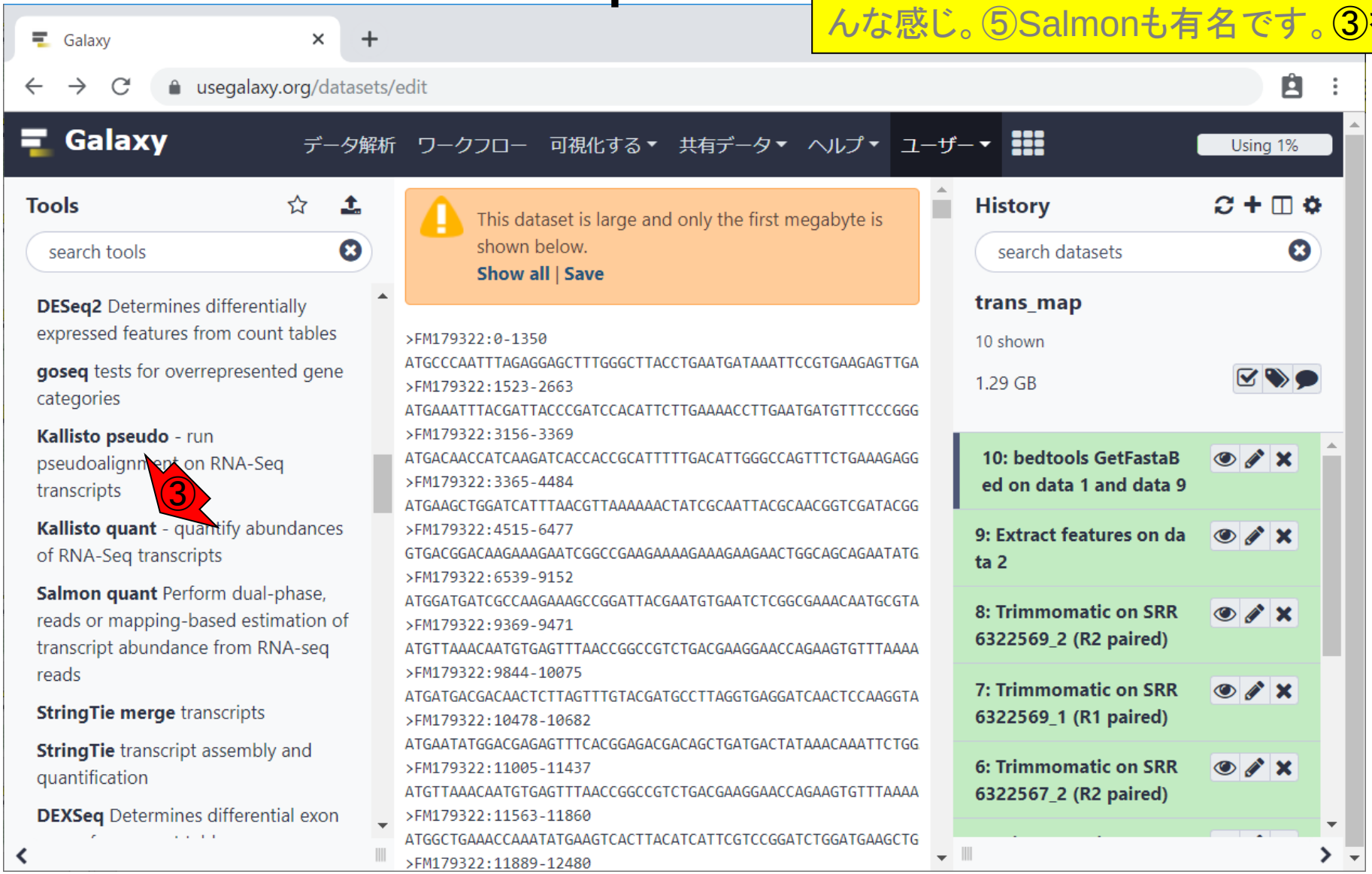
W8-1：Kallisto quant

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。



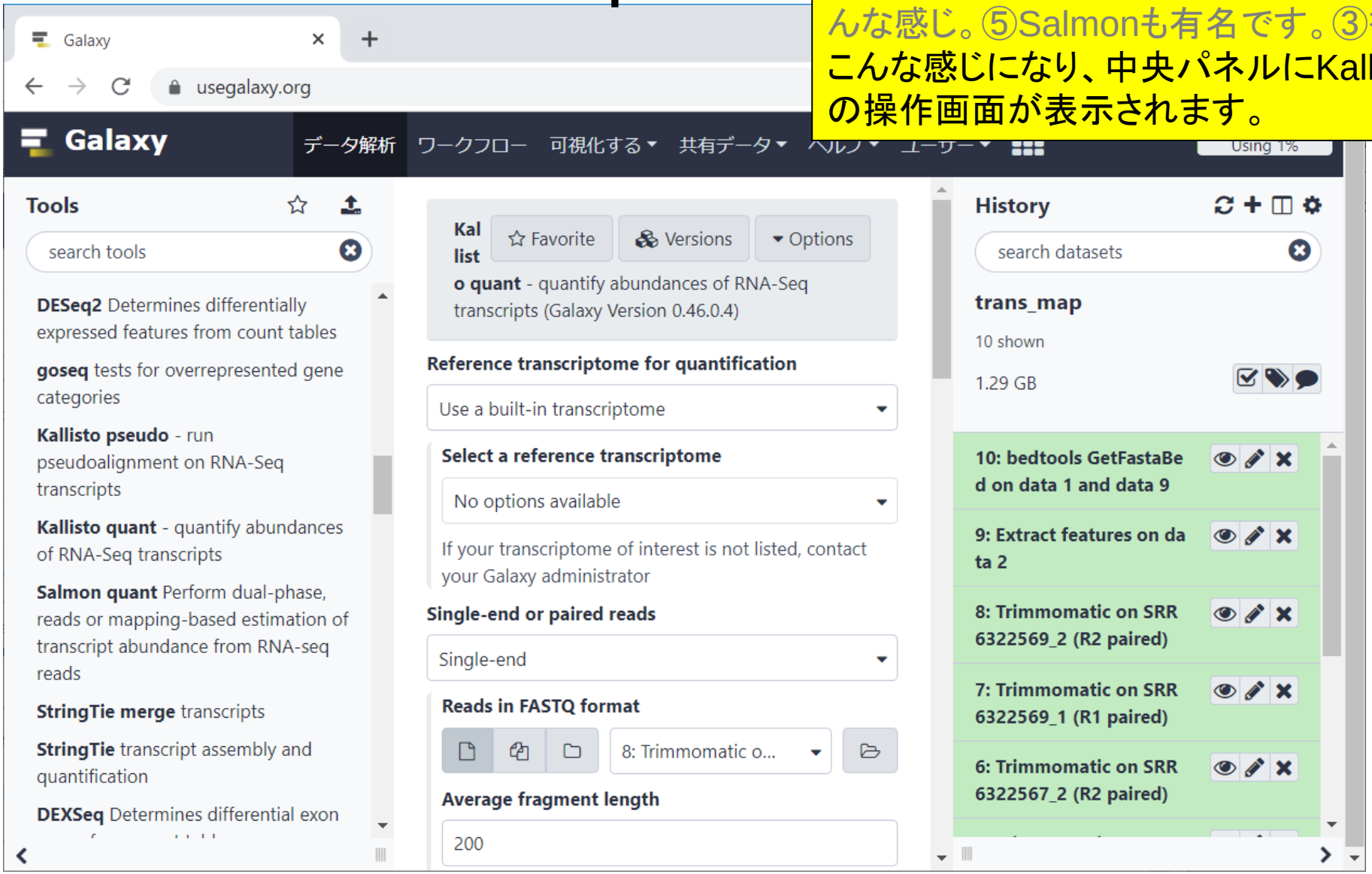
W8-1：Kallisto quant

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。



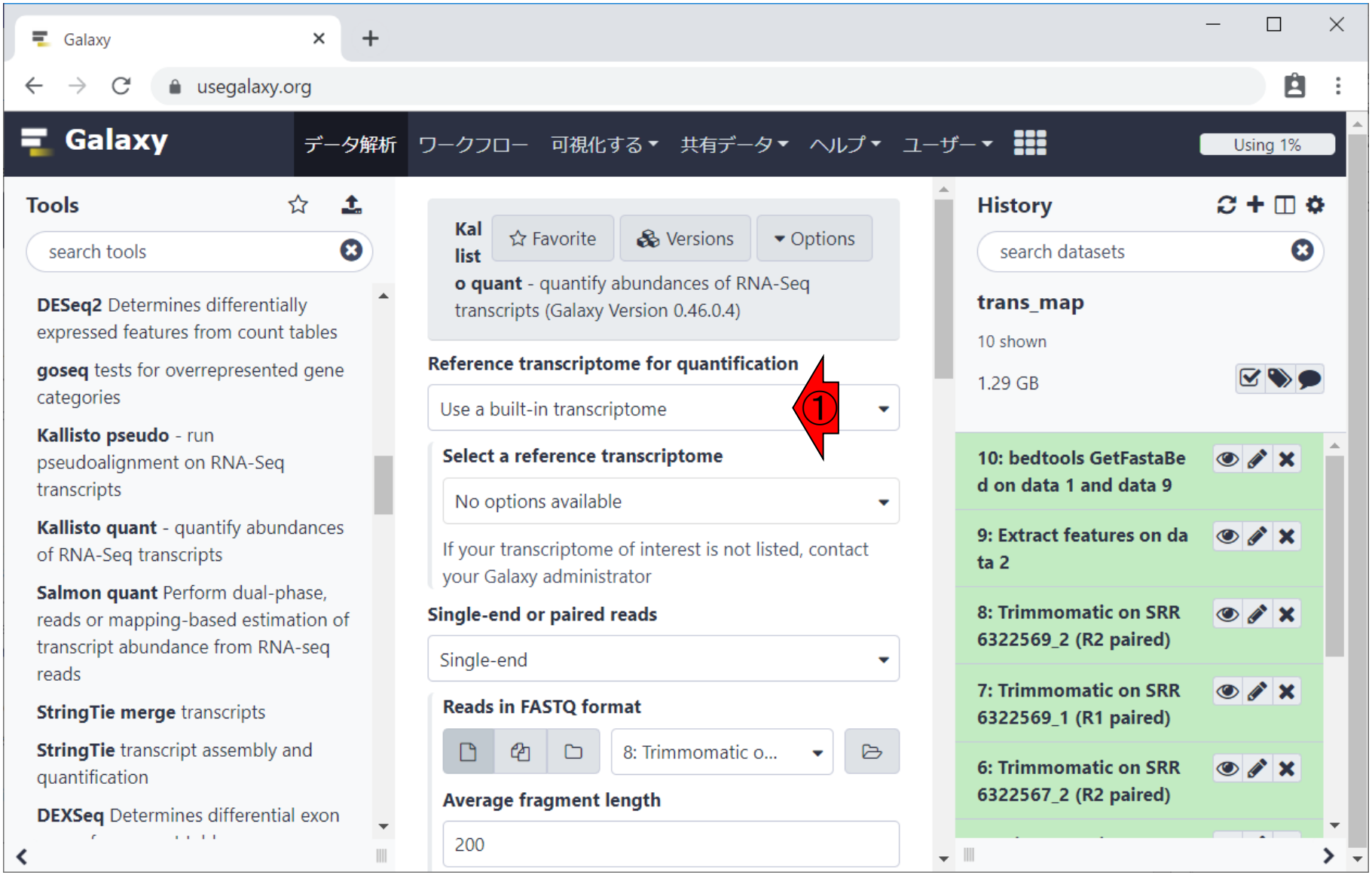
W8-1：Kallisto quant

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。



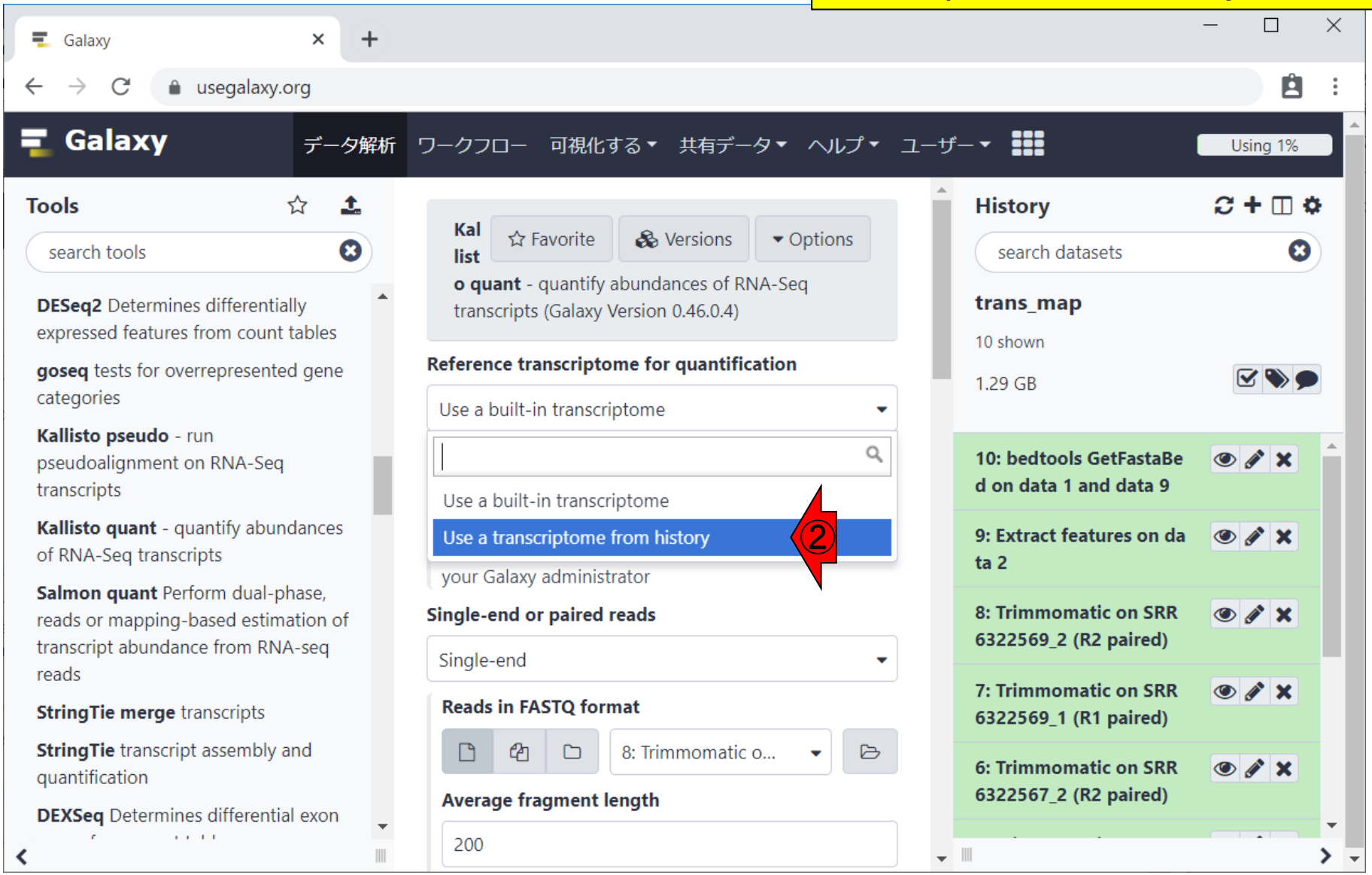
W8-2：リファレンス

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…



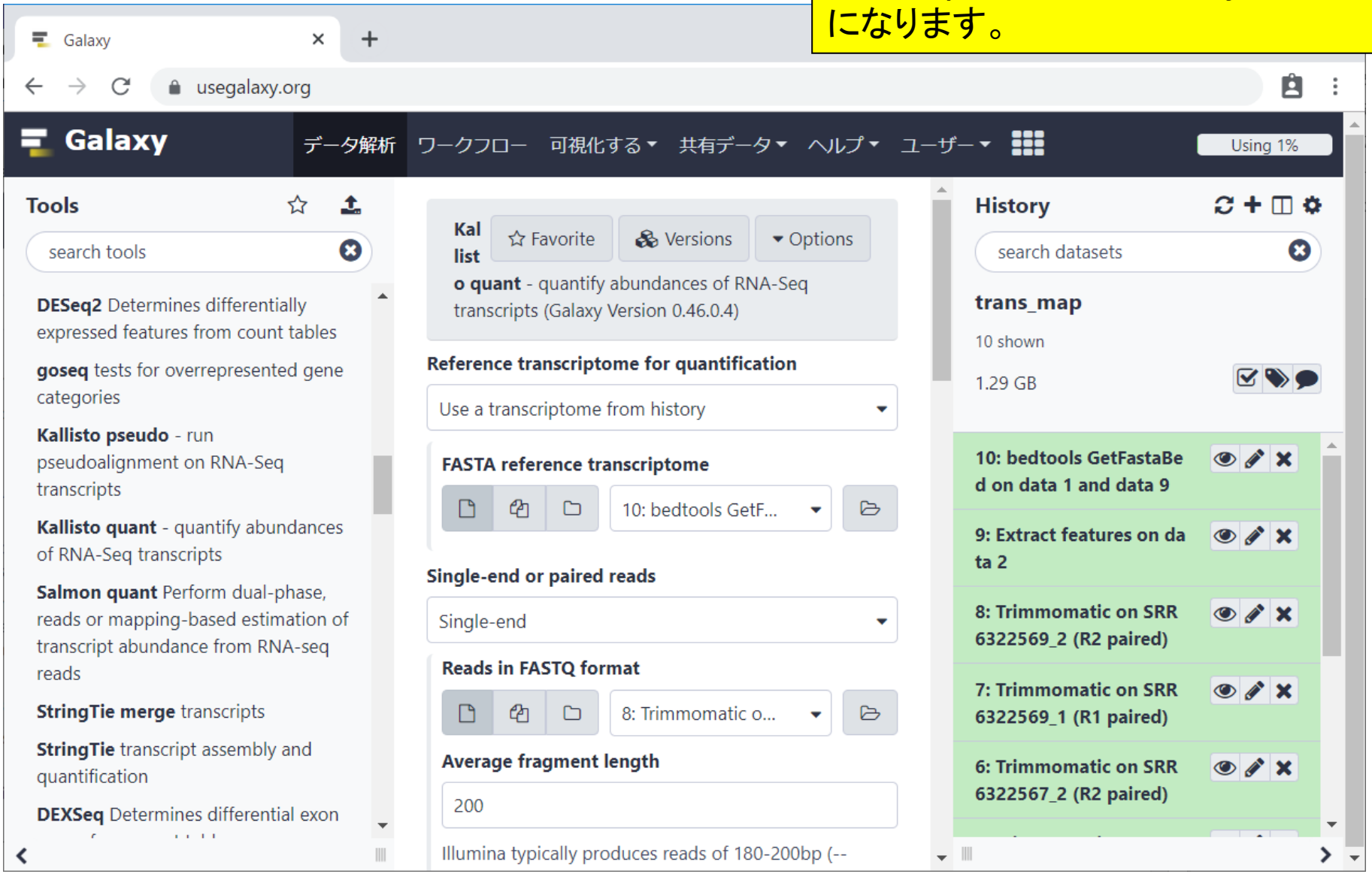
W8-2：リファレンス

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



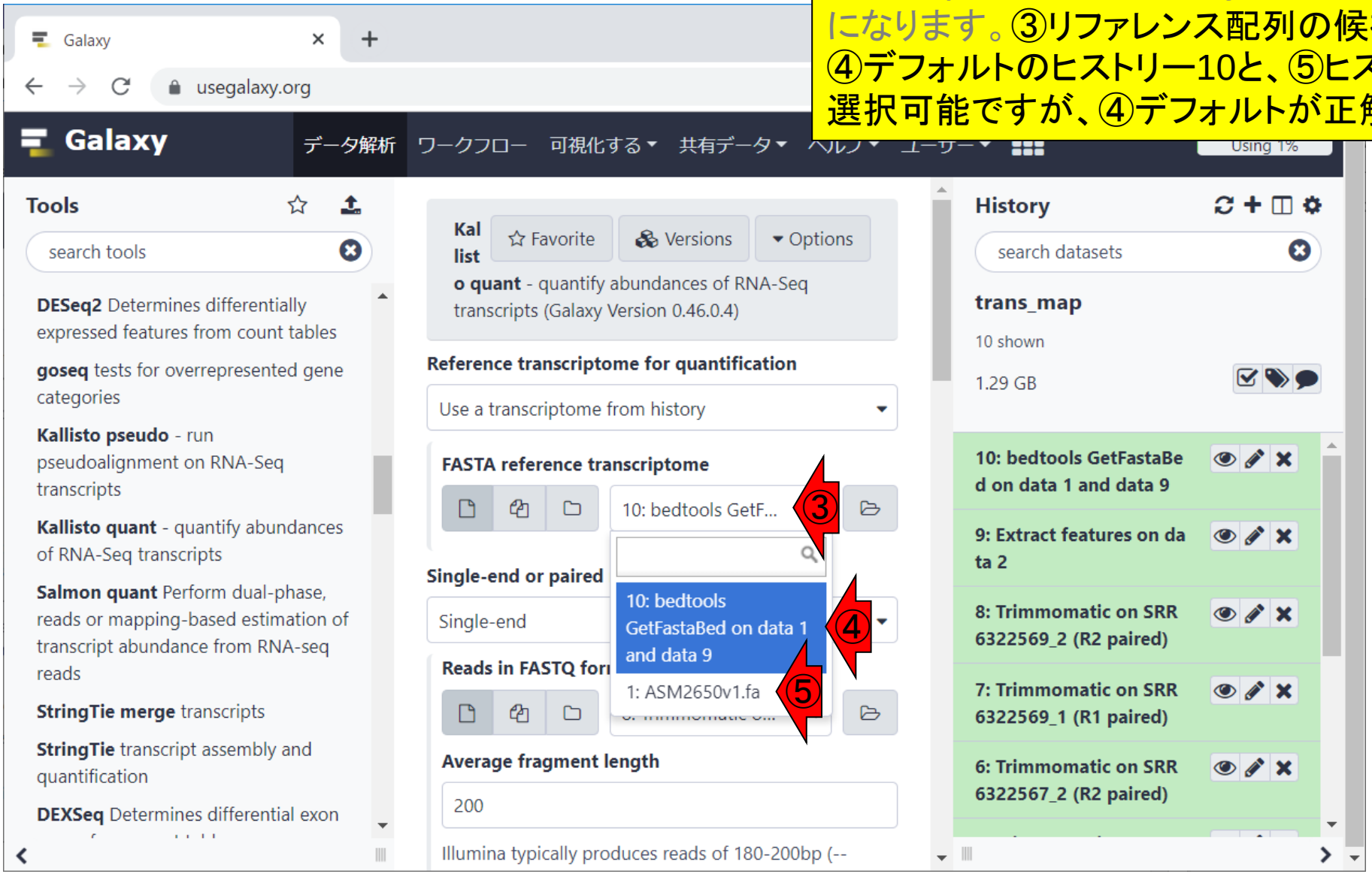
W8-2：リファレンス

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



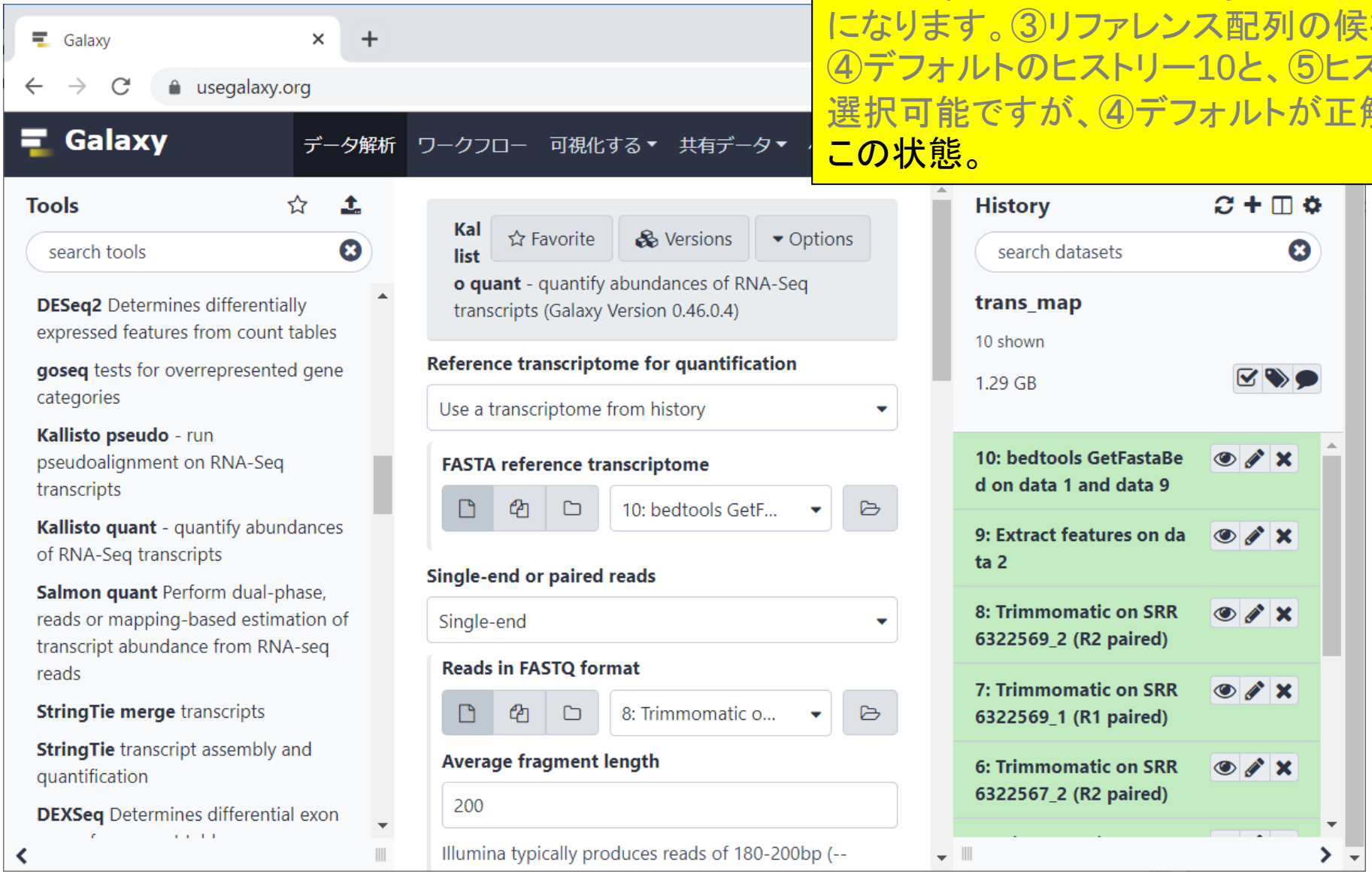
W8-2：リファレンス

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



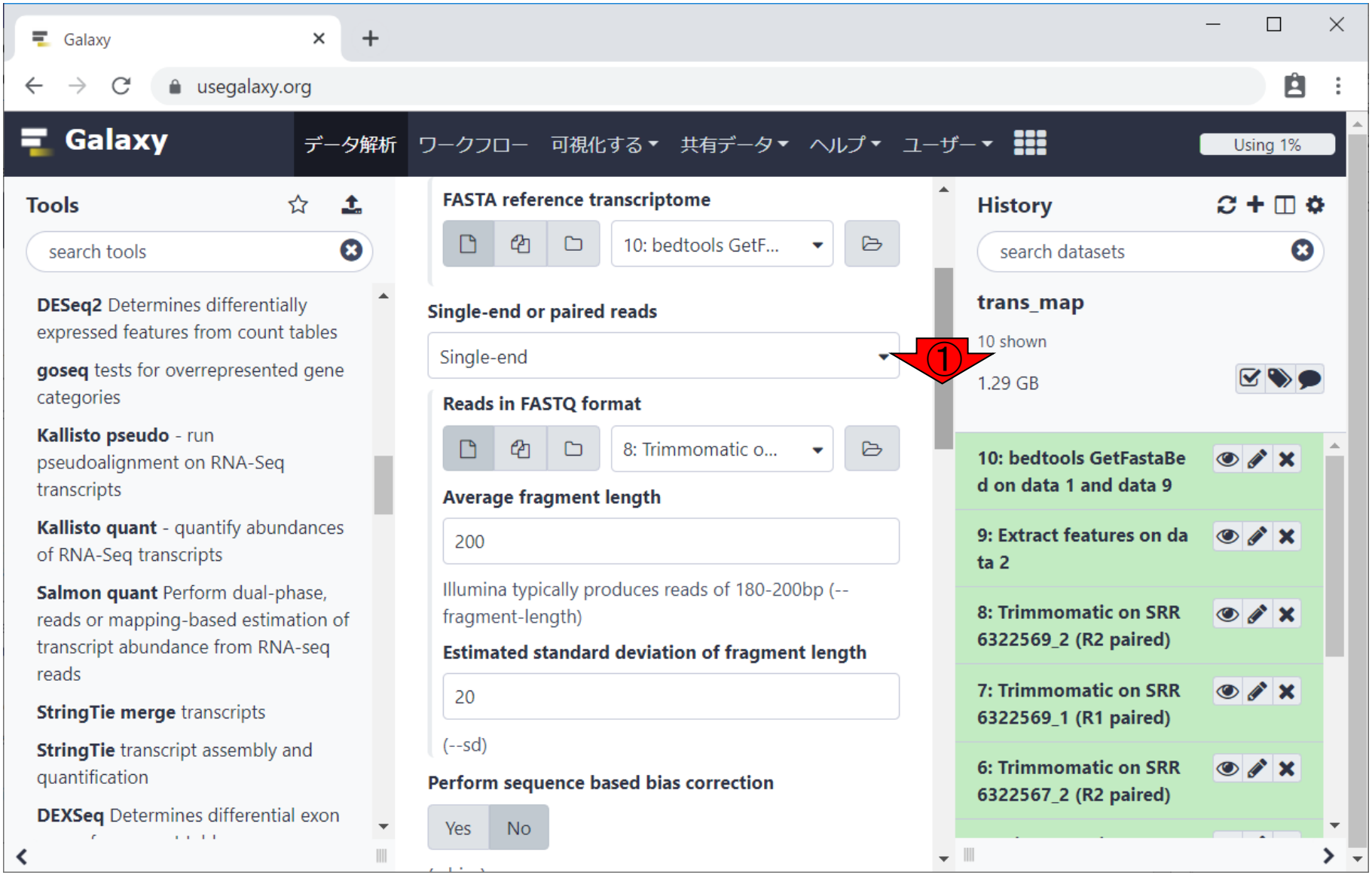
W8-2：リファレンス

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



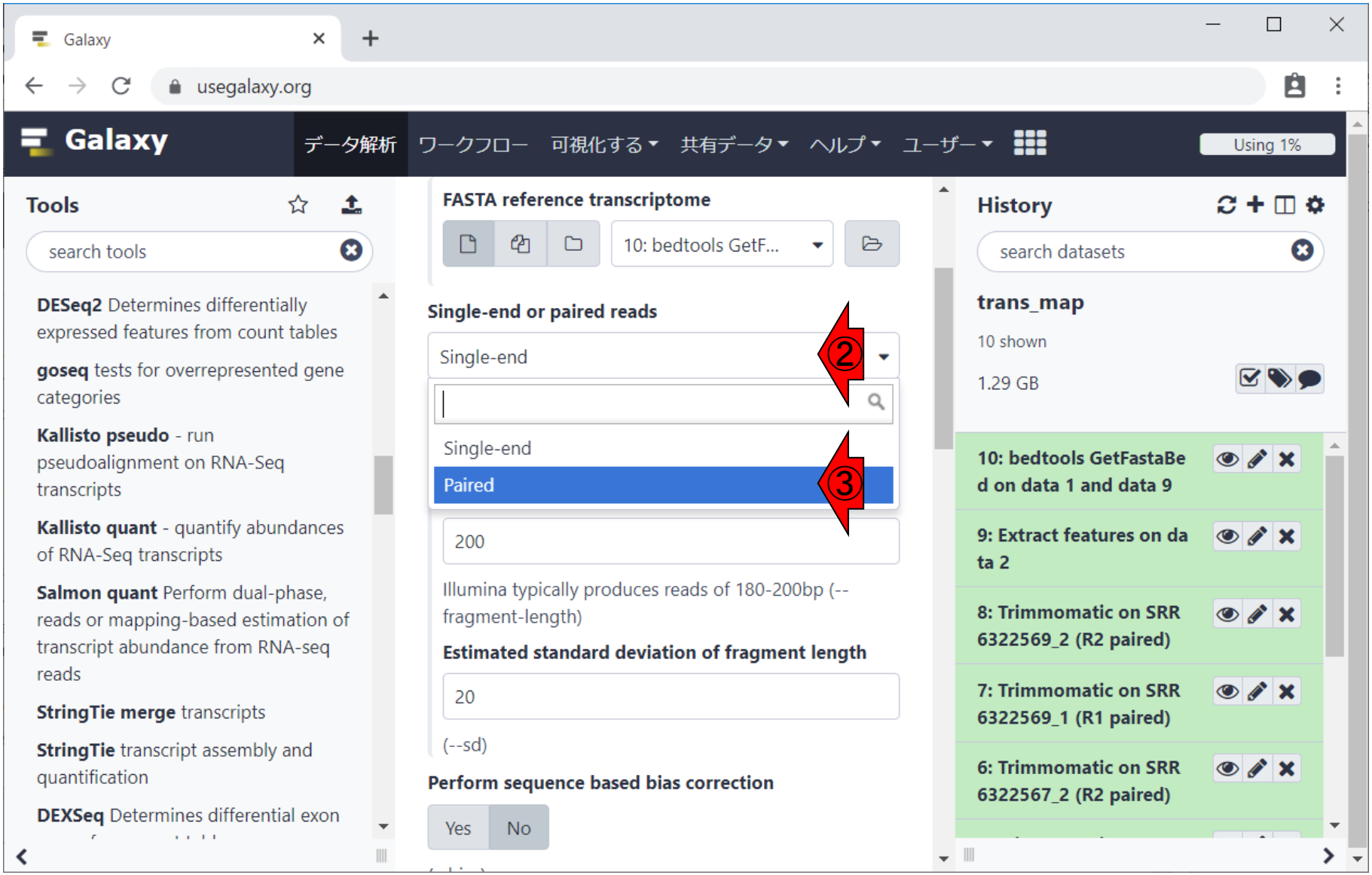
①少しページ下部に移動。

W8-3：Pairedに変更



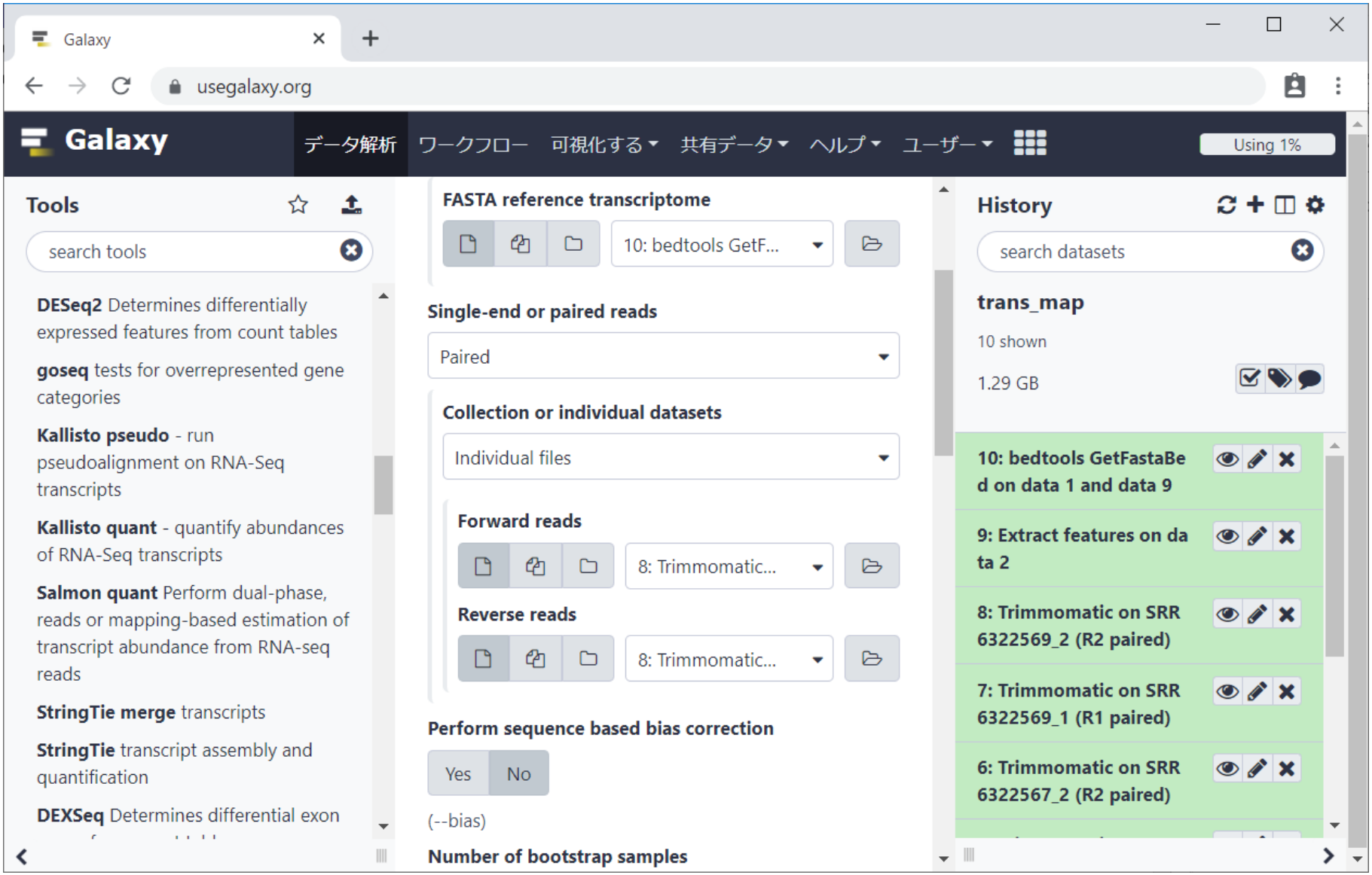
W8-3：Pairedに変更

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。



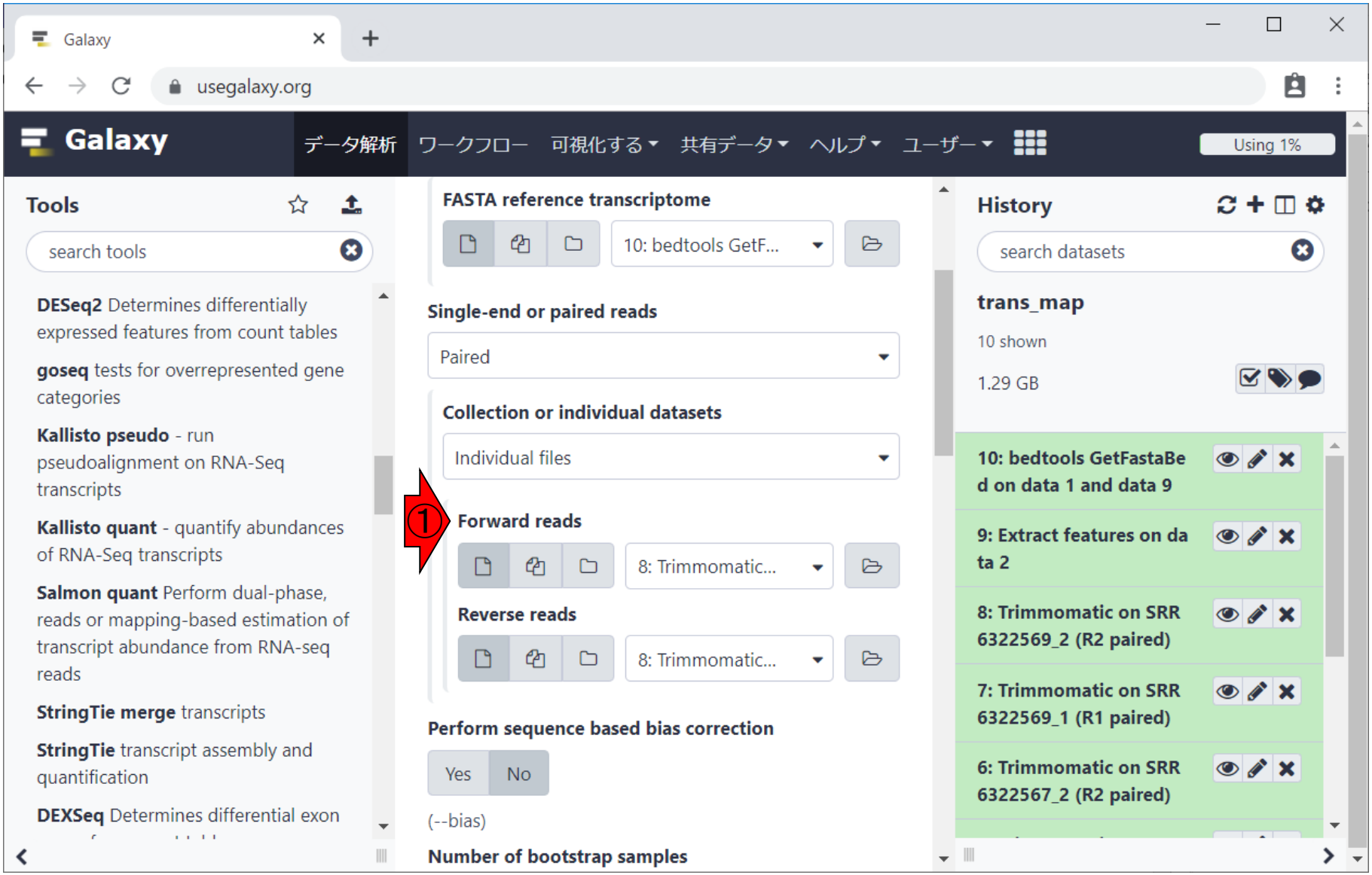
W8-3：Pairedに変更

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。



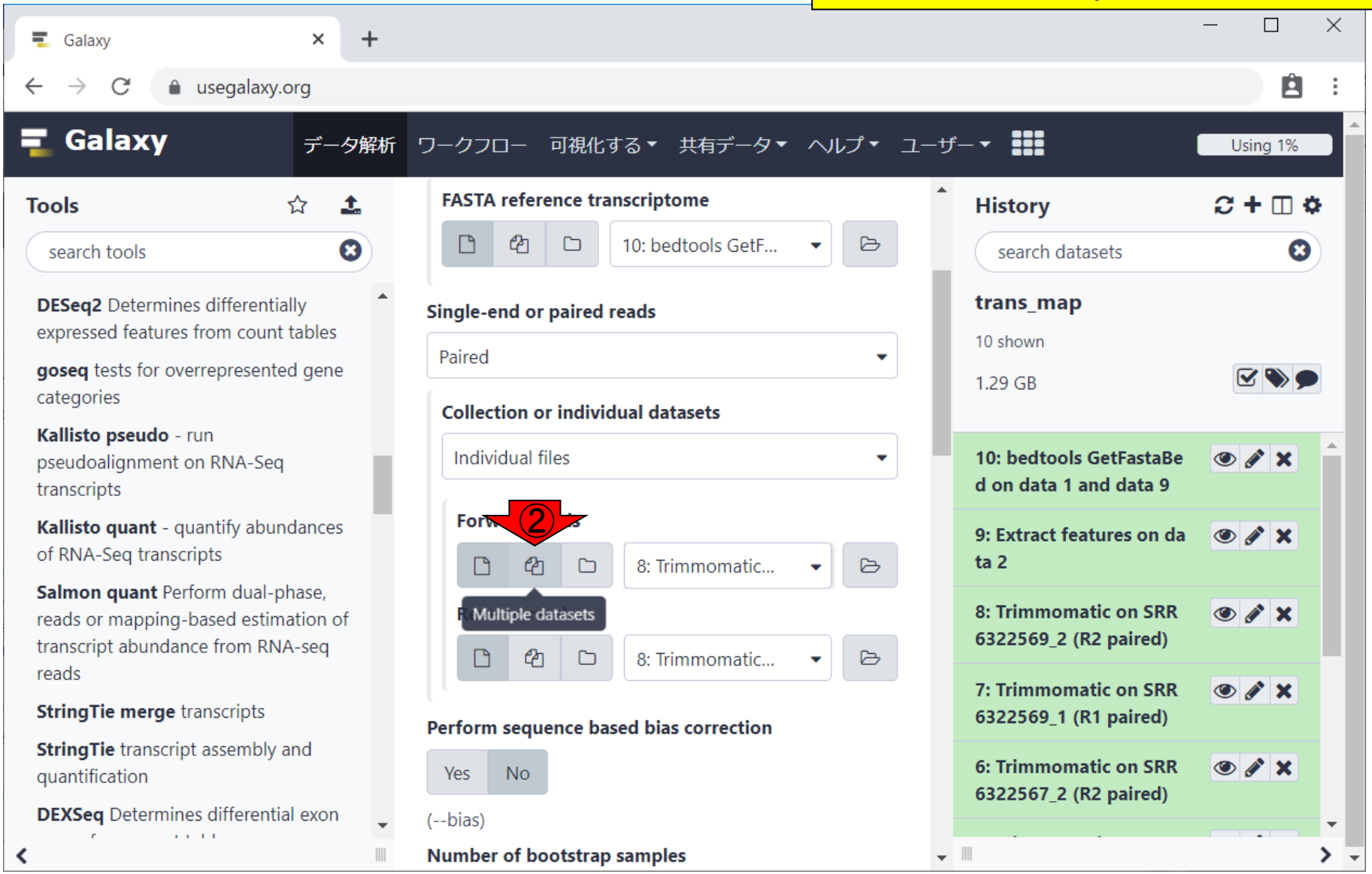
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward側。



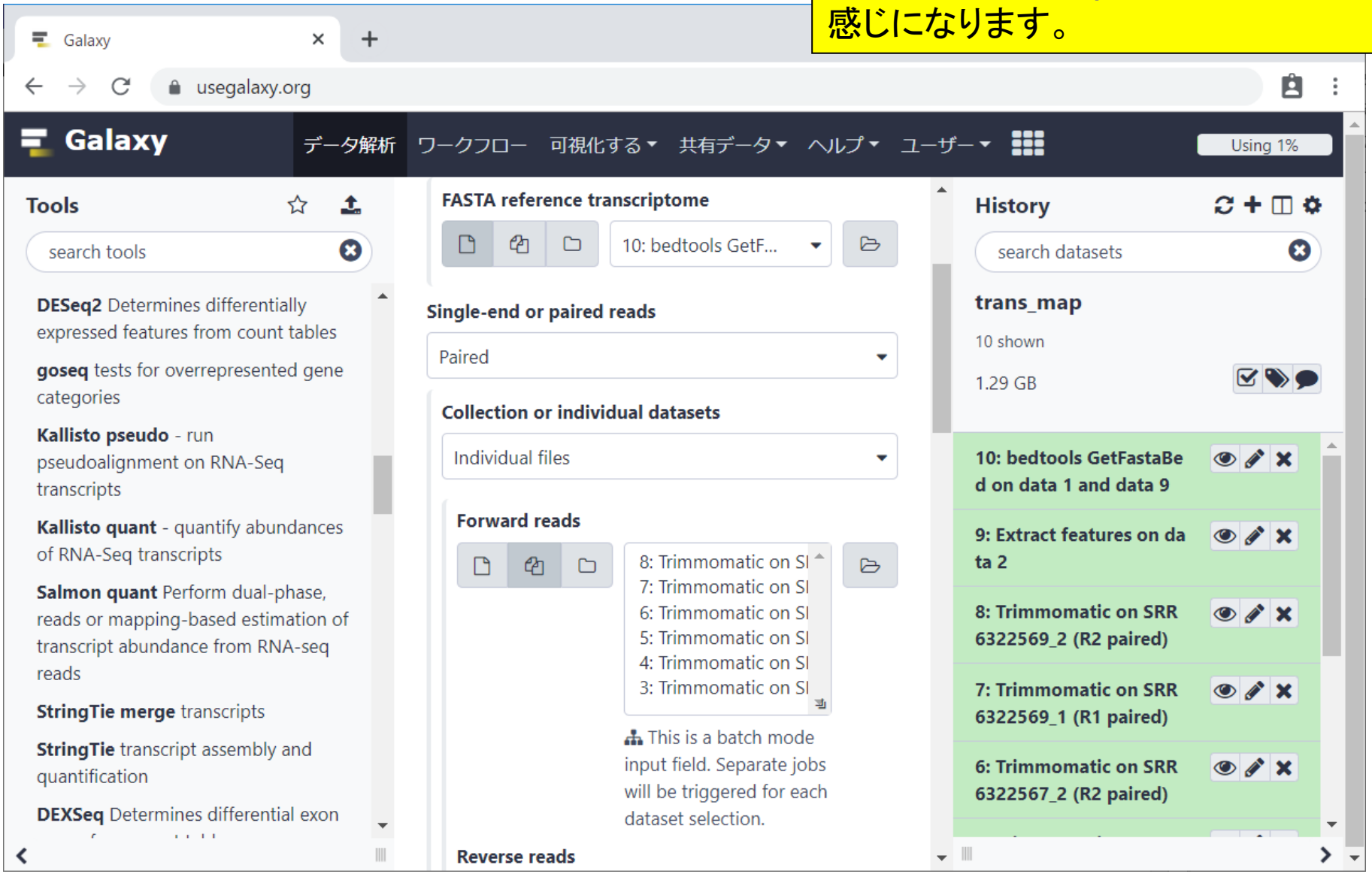
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



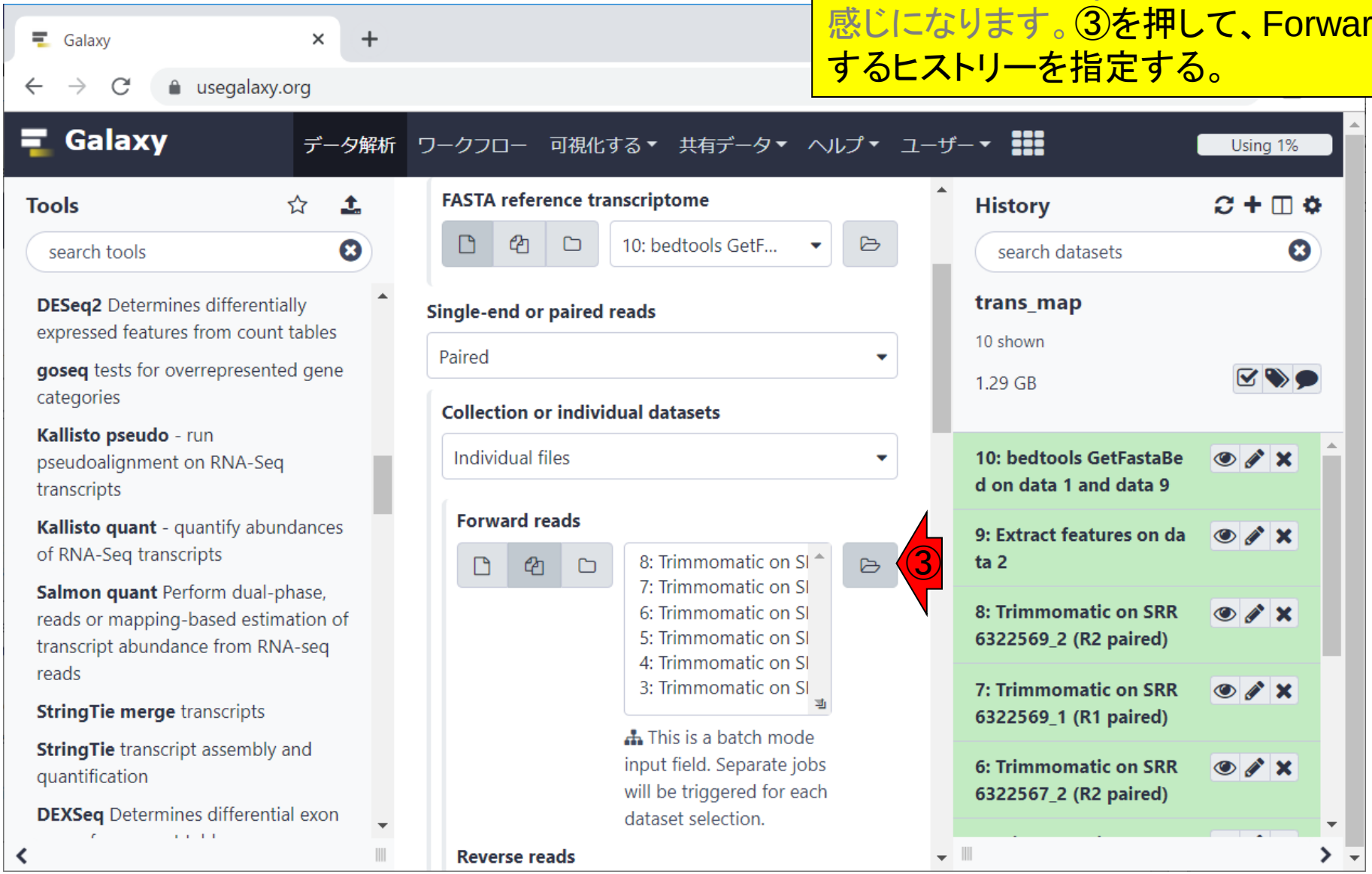
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



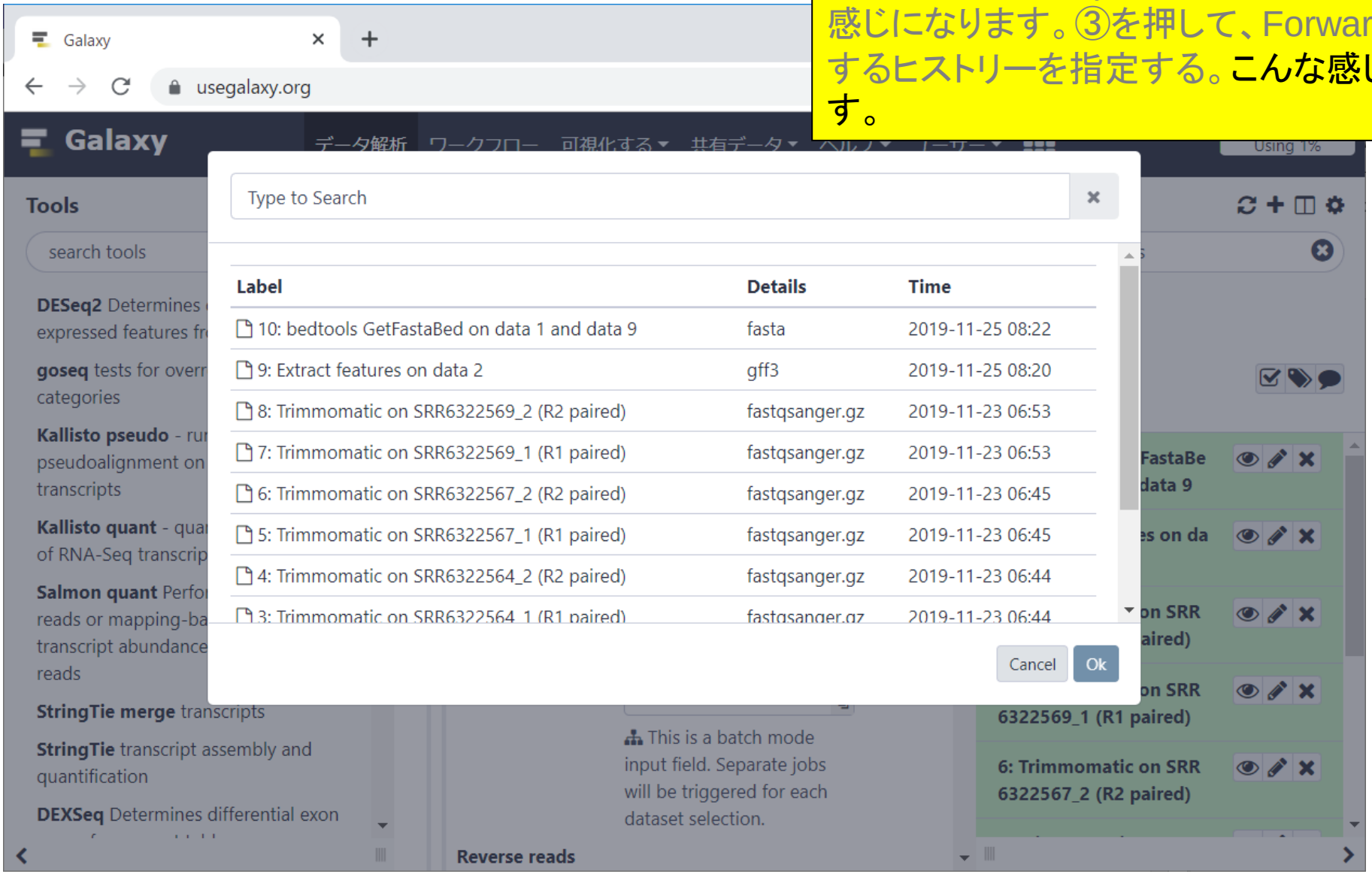
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



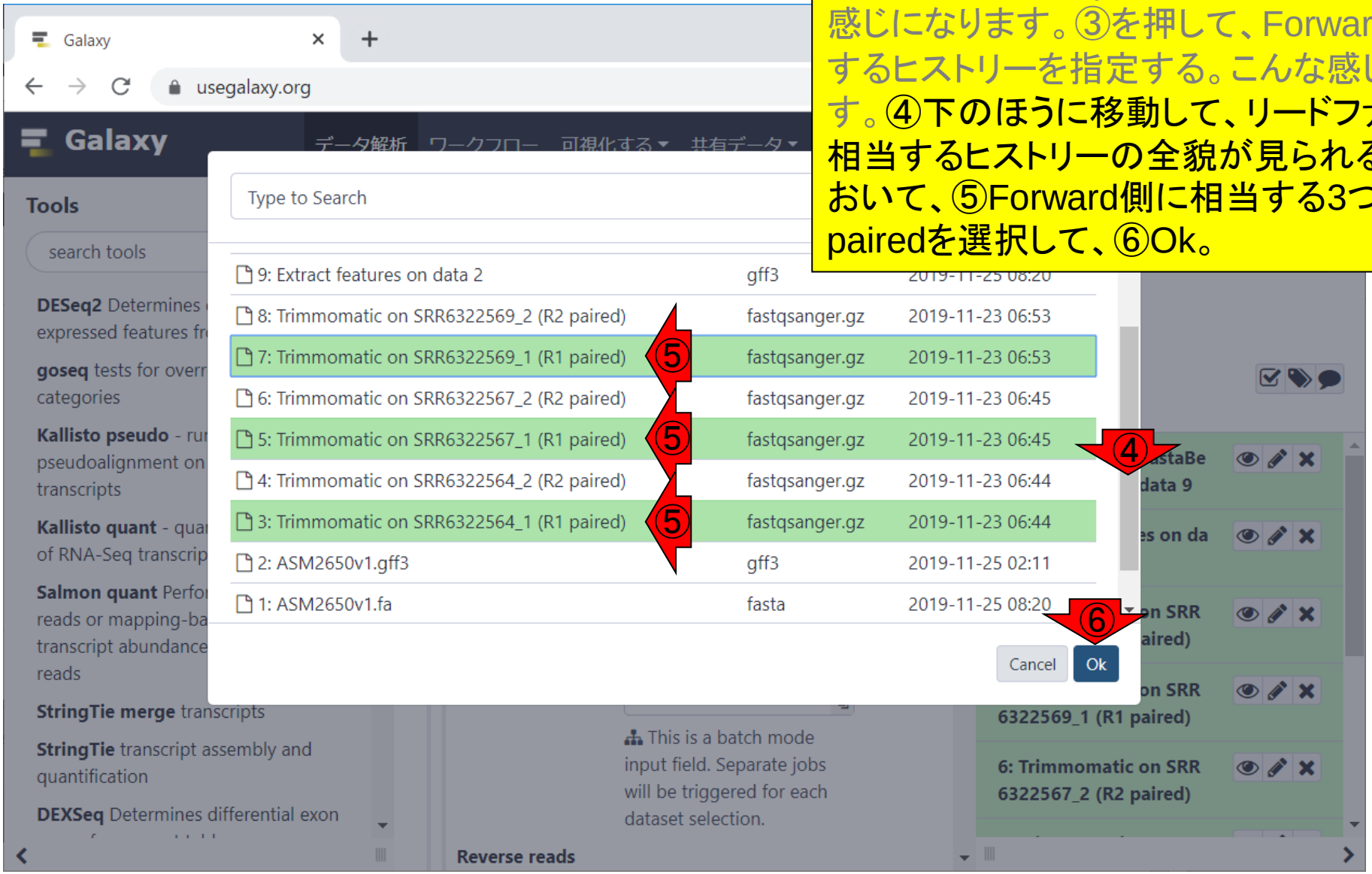
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



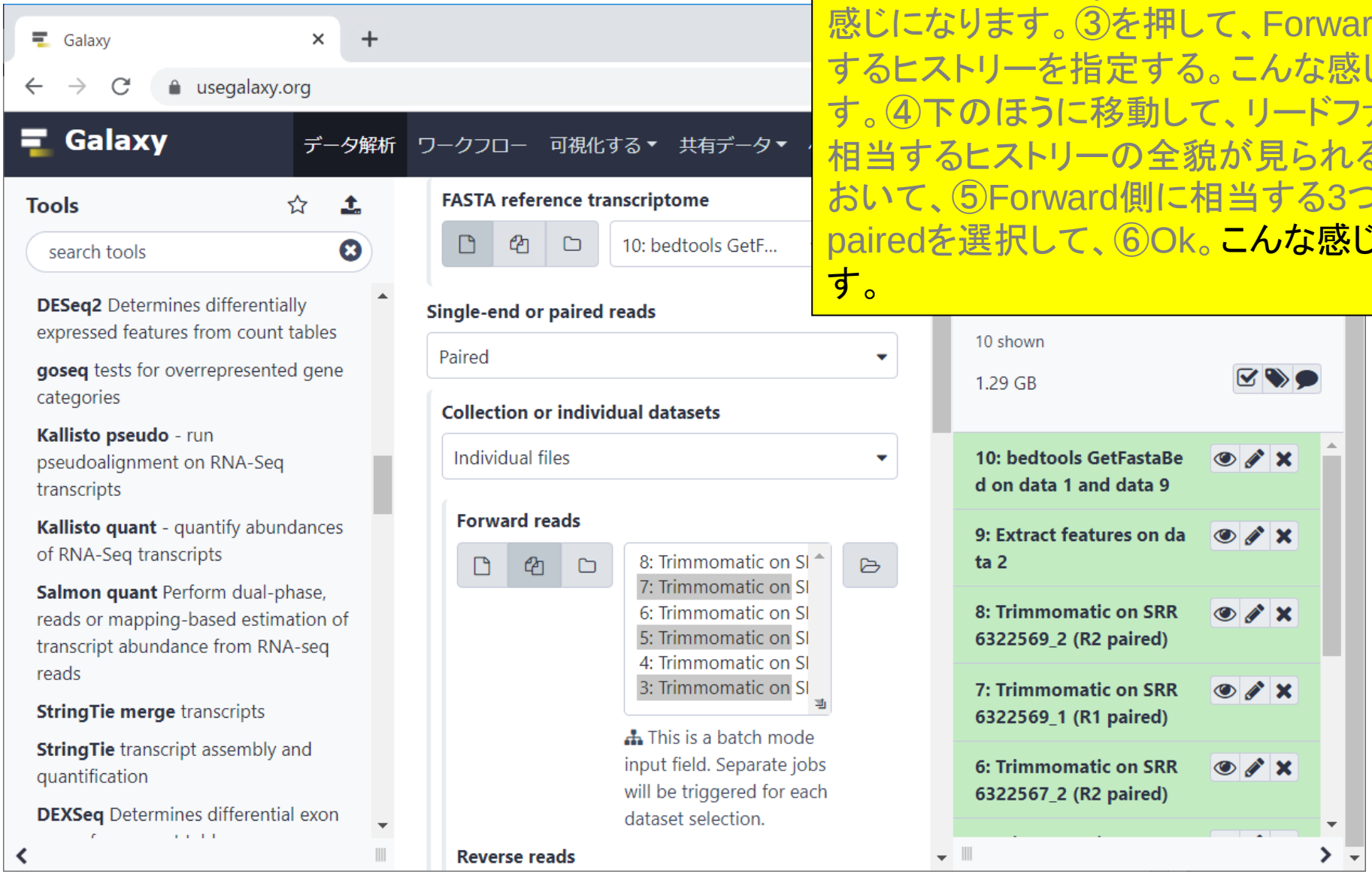
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



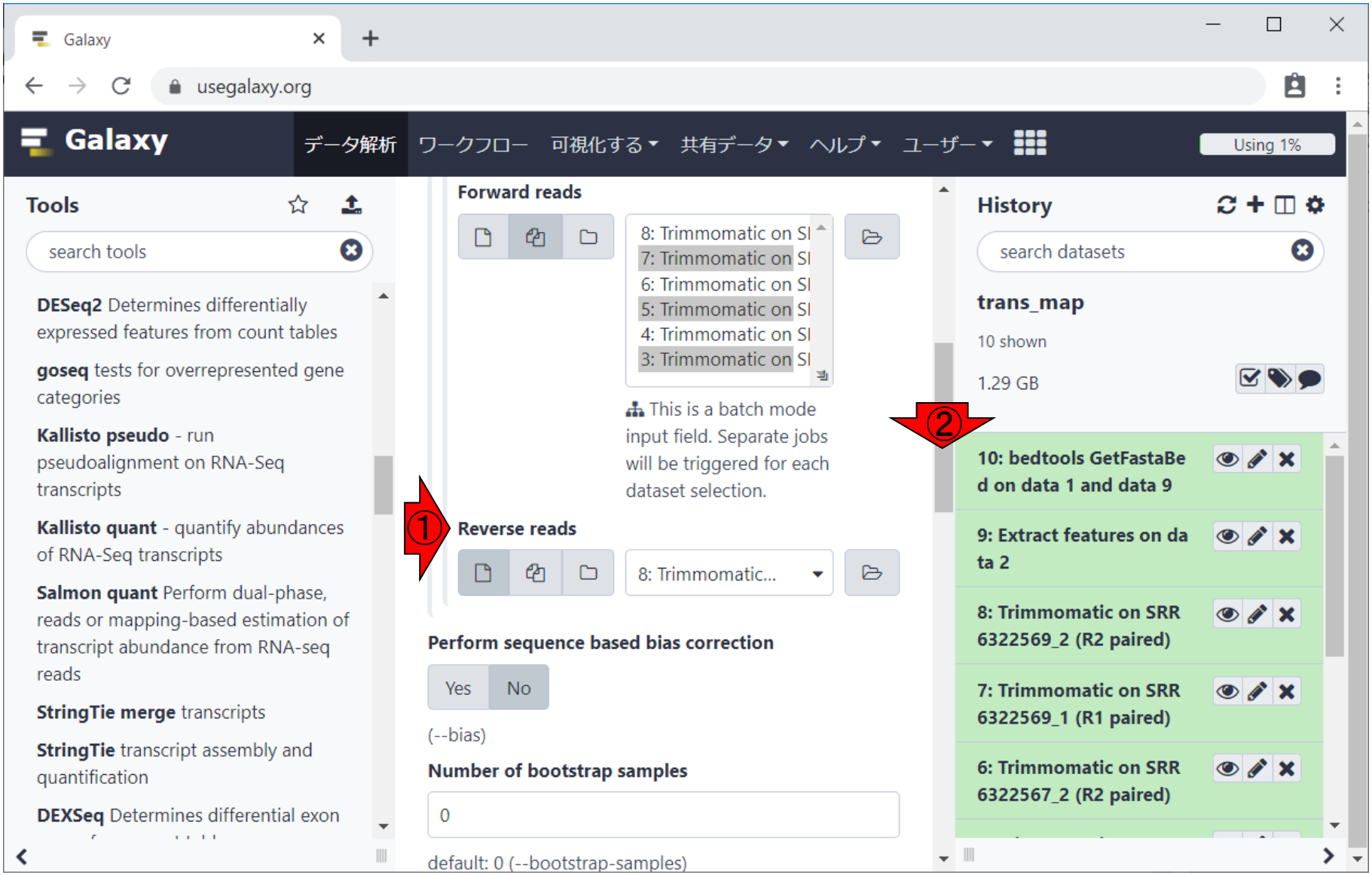
W8-4：Forward側

次が、リードファイル指定の本番。まずは①Forward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。



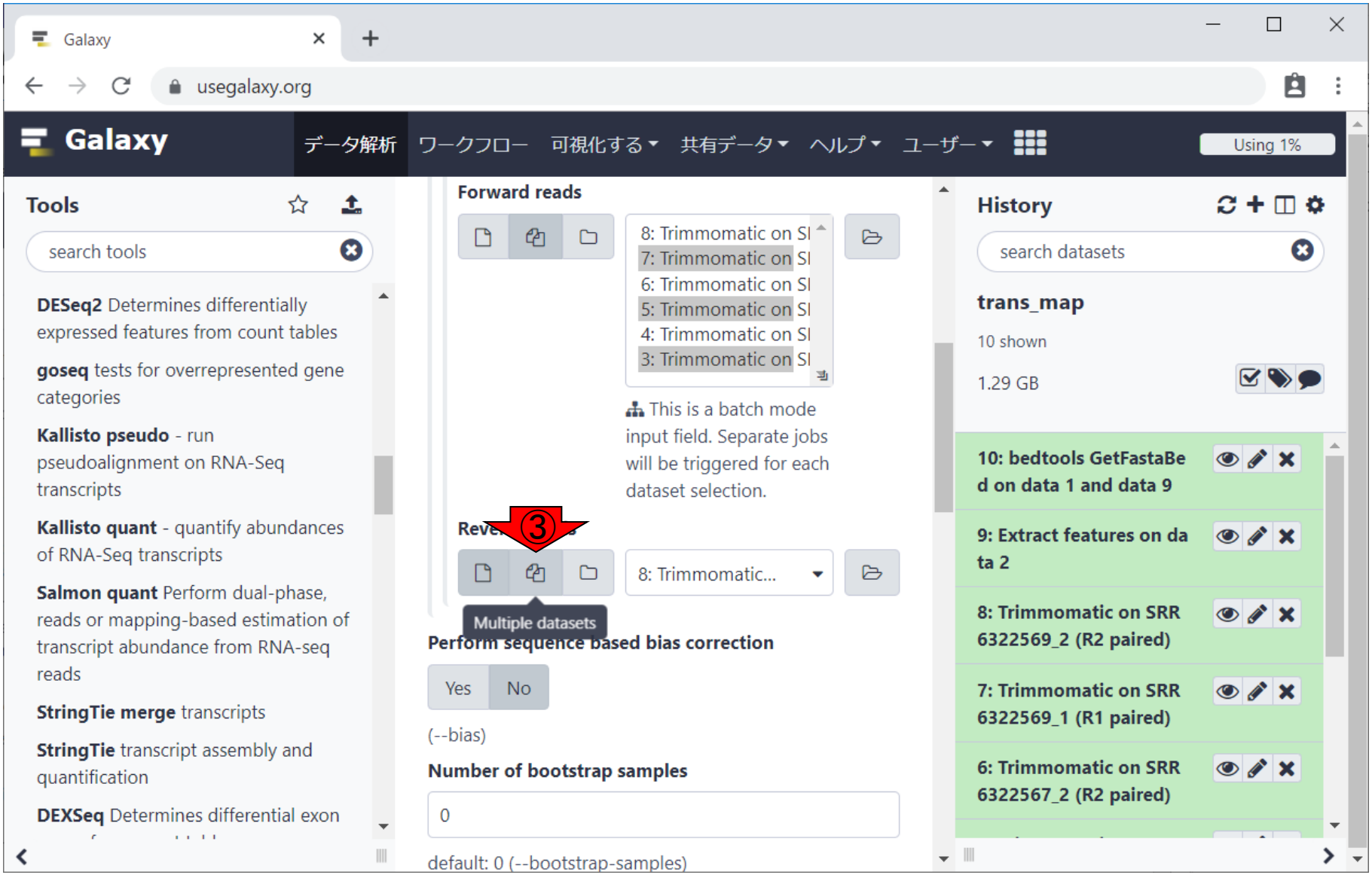
W8-5：Reverse側

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。



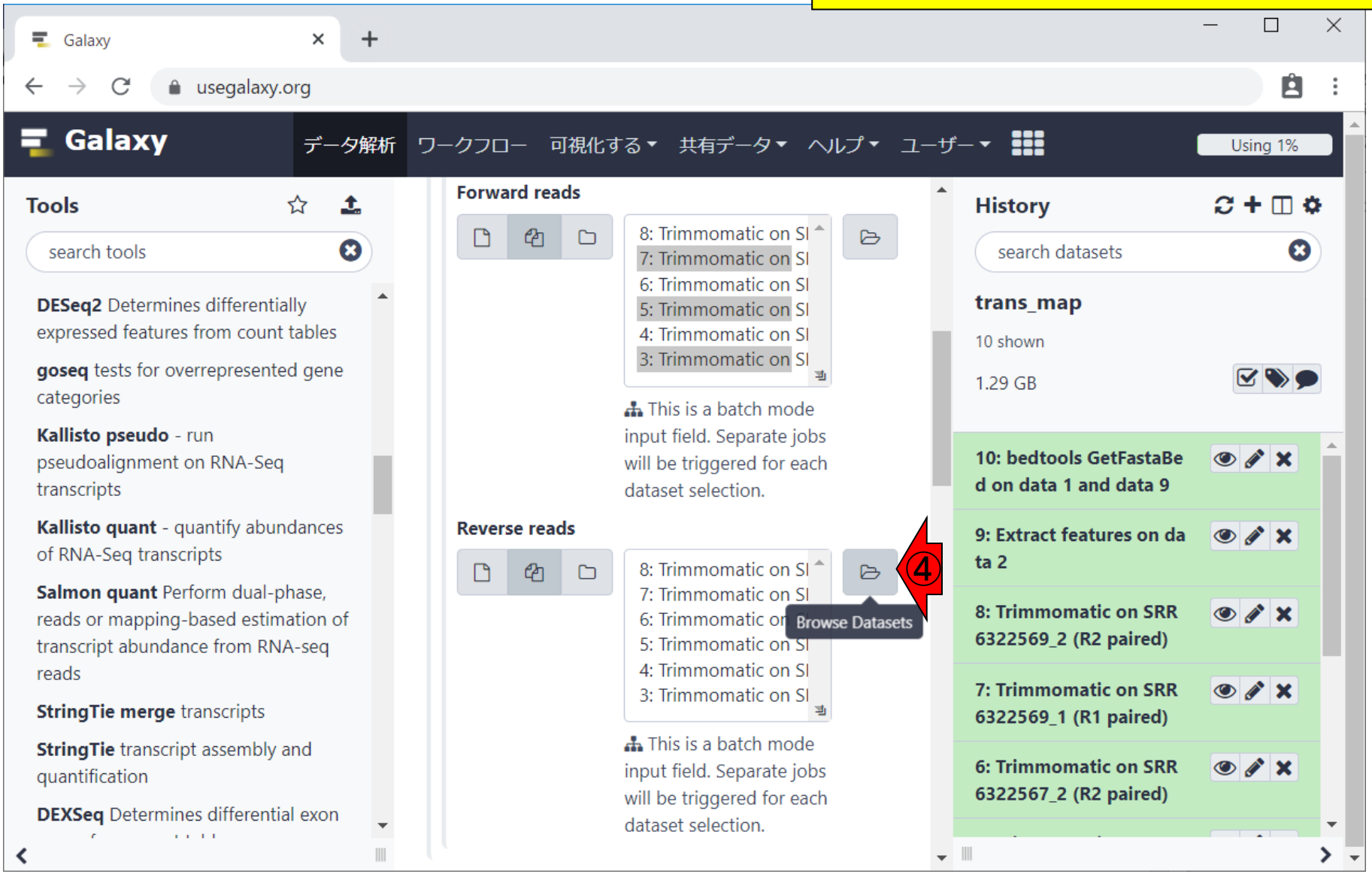
W8-5：Reverse側

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。



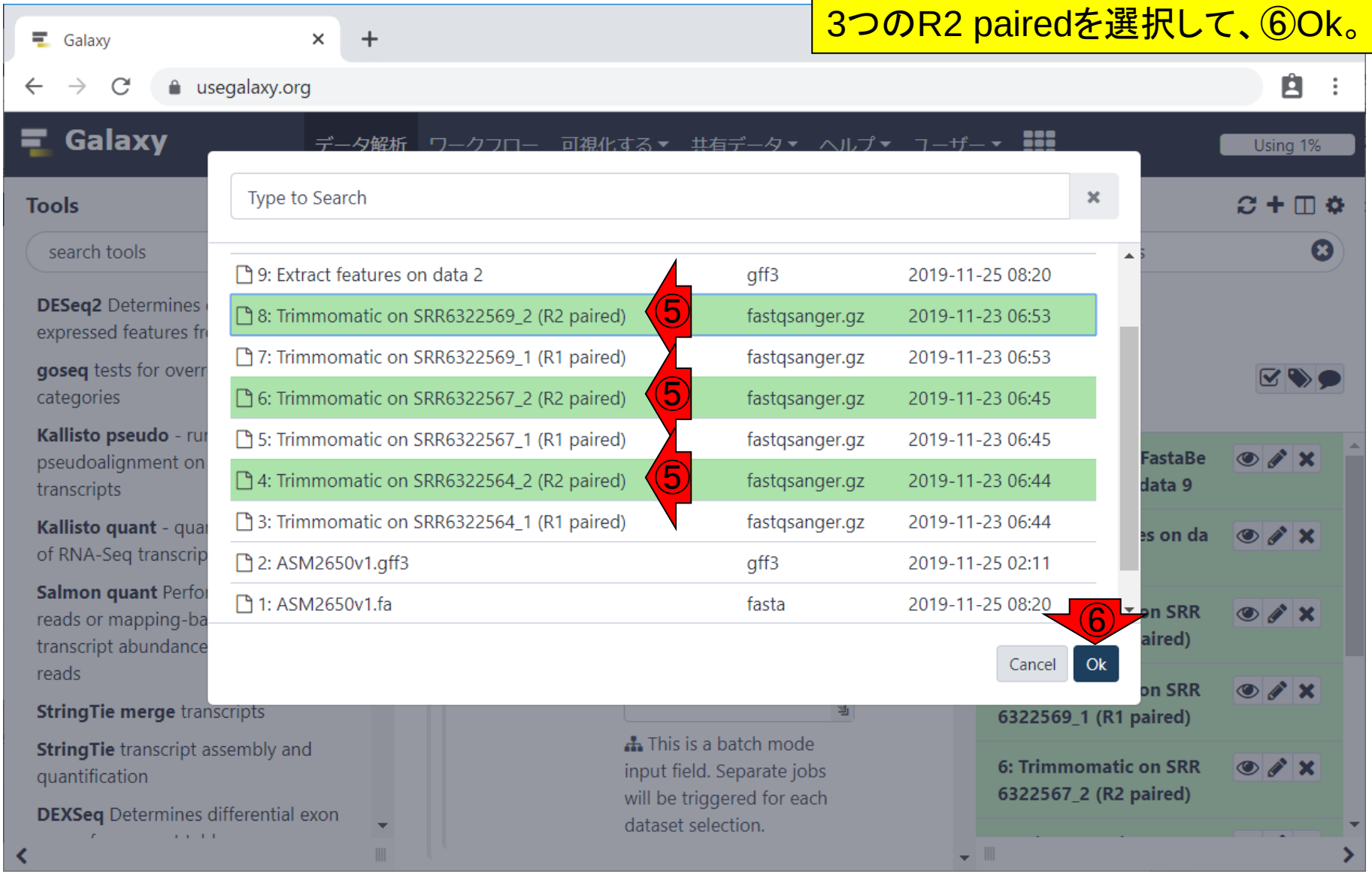
W8-5：Reverse側

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。



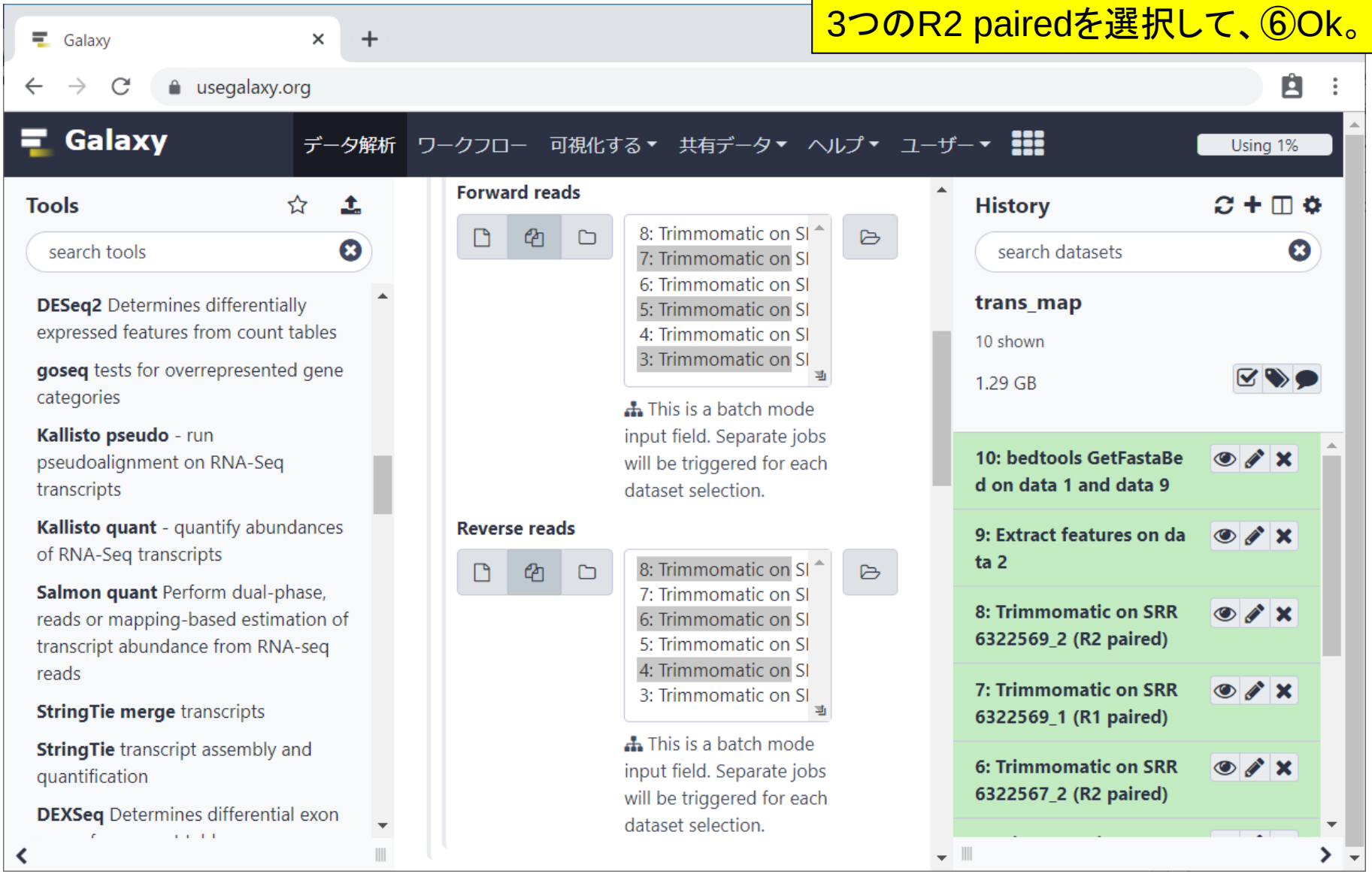
W8-5：Reverse側

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。



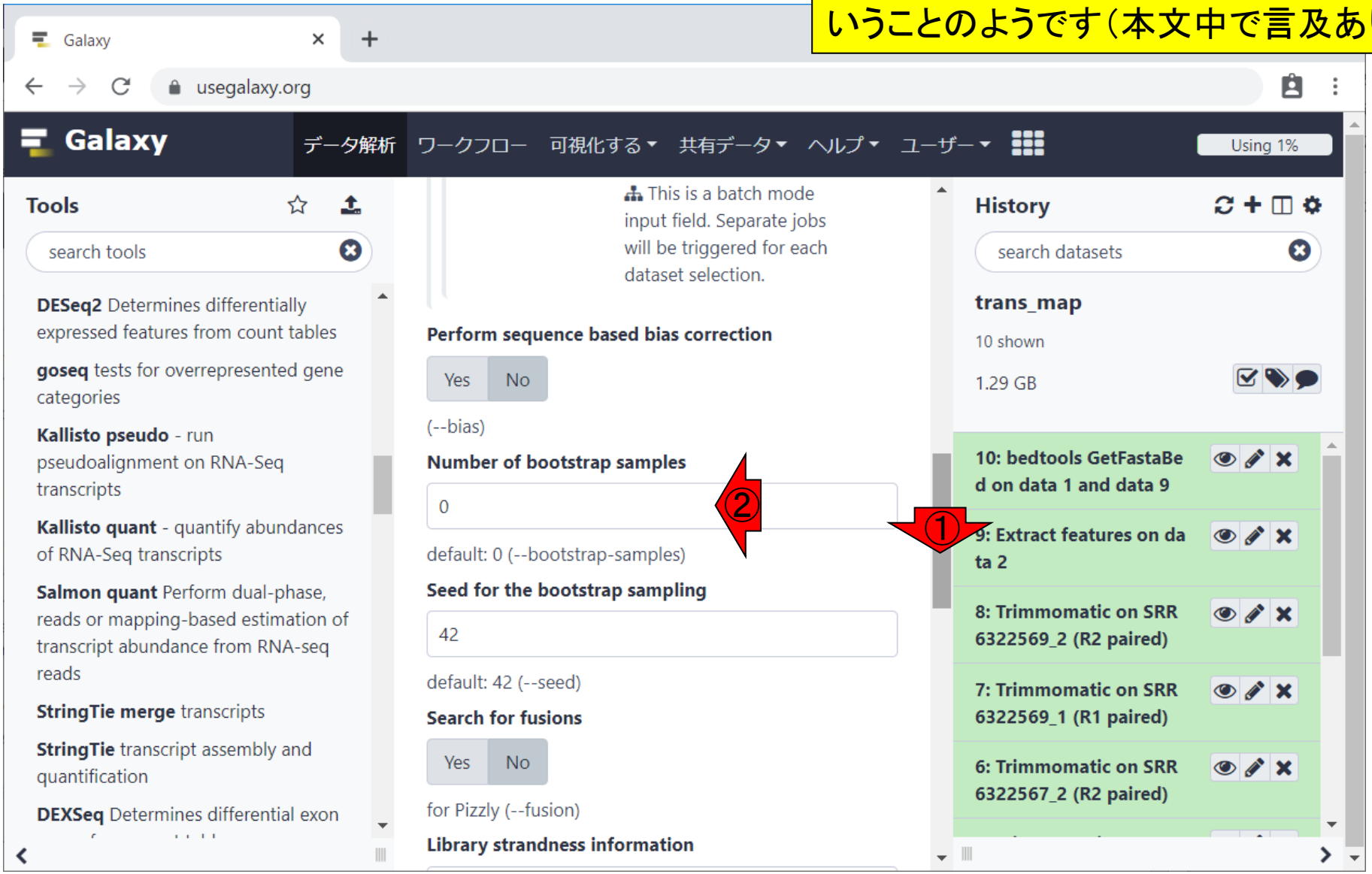
W8-5：Reverse側

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。



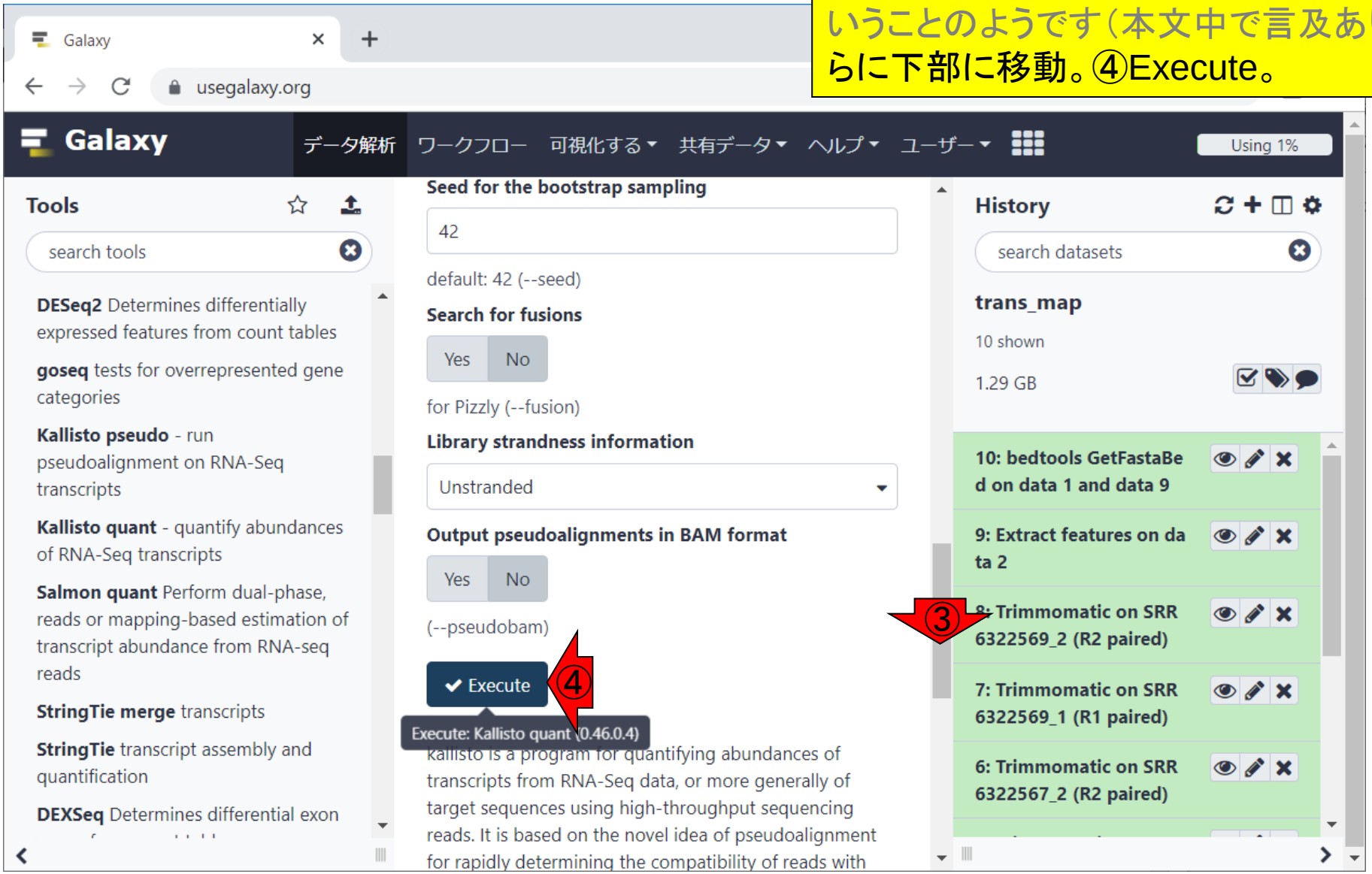
W8-6：実行

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）。



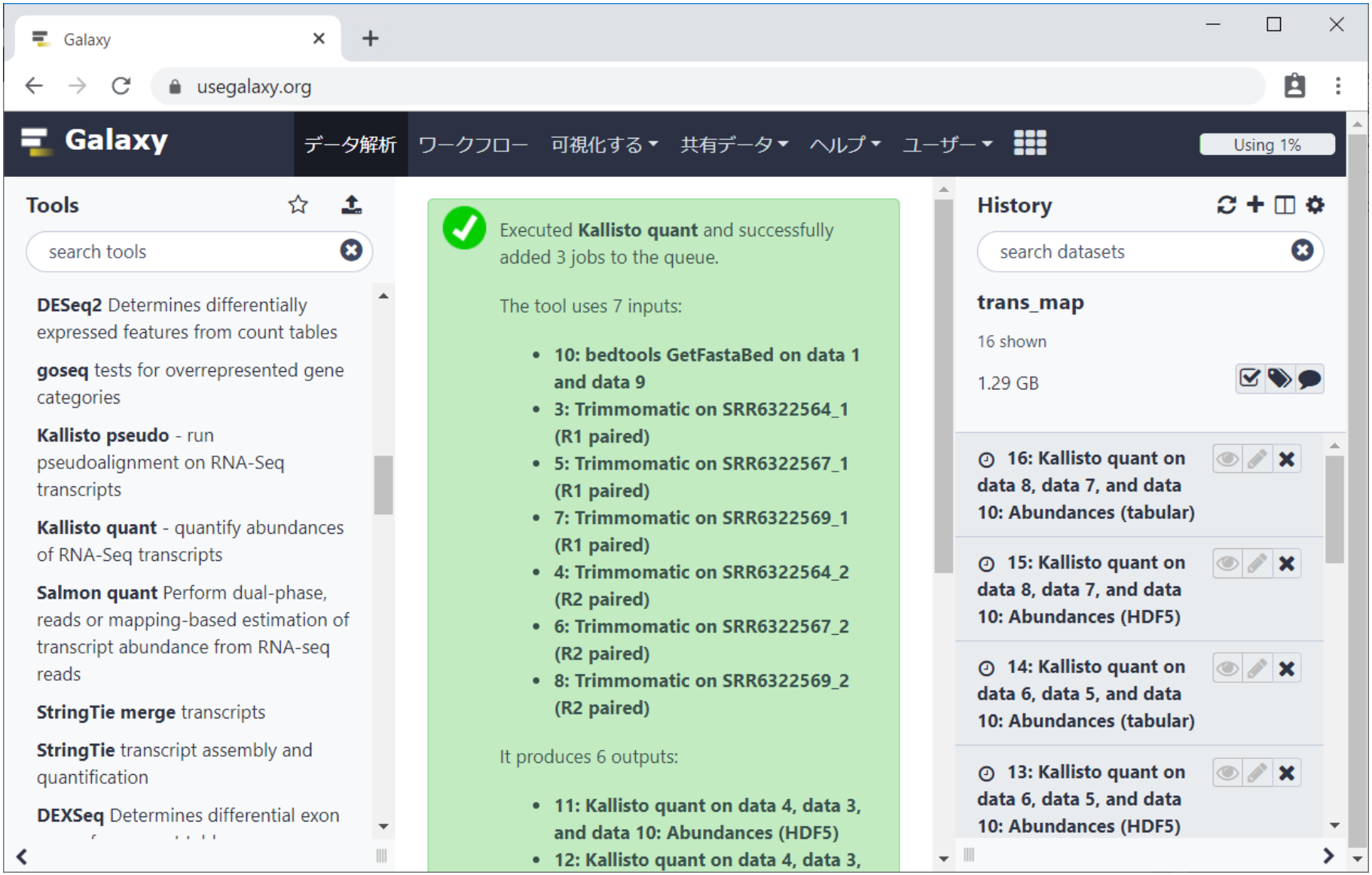
W8-6：実行

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。



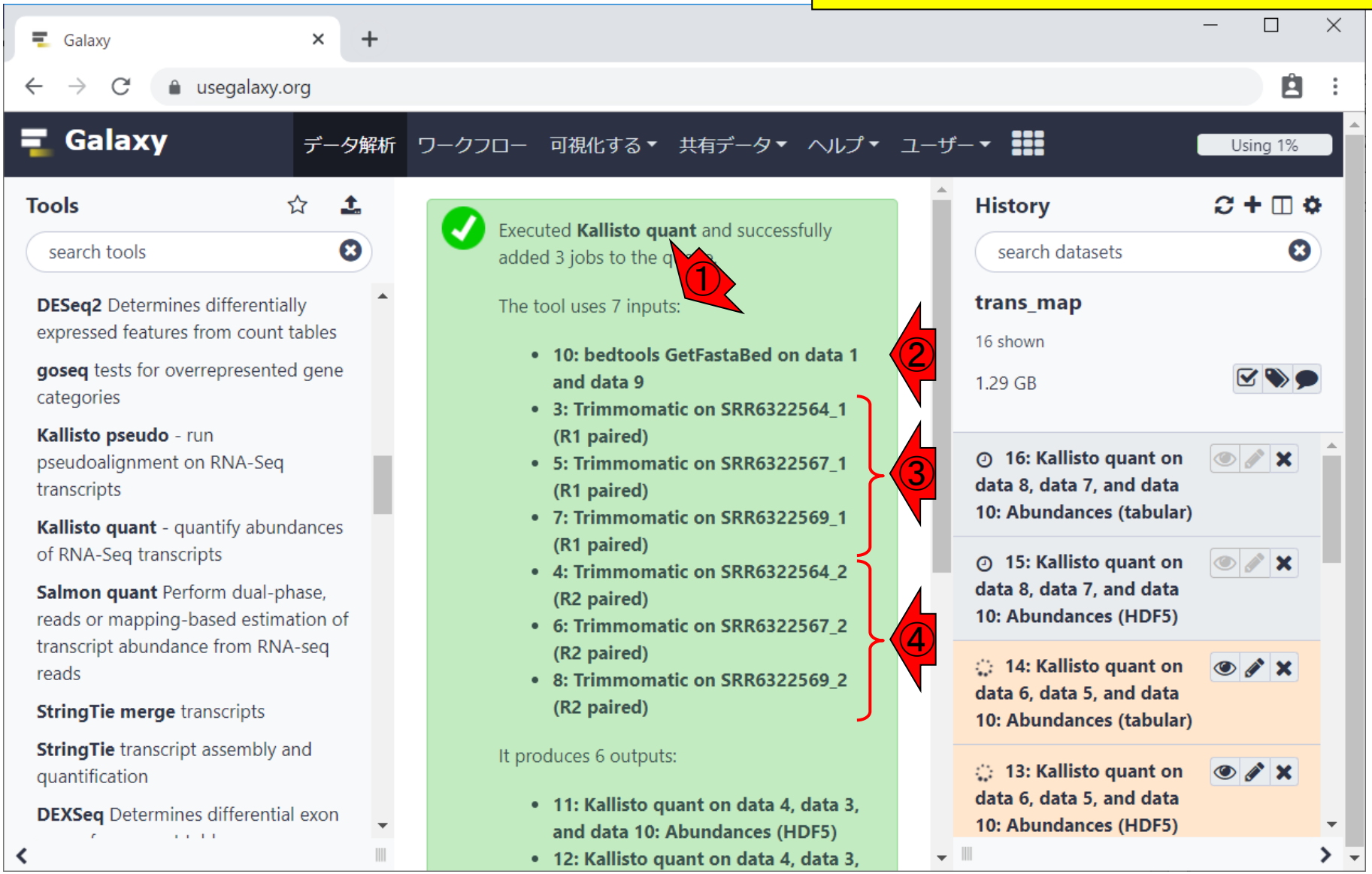
実行中…。

W8-7：実行中…


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)

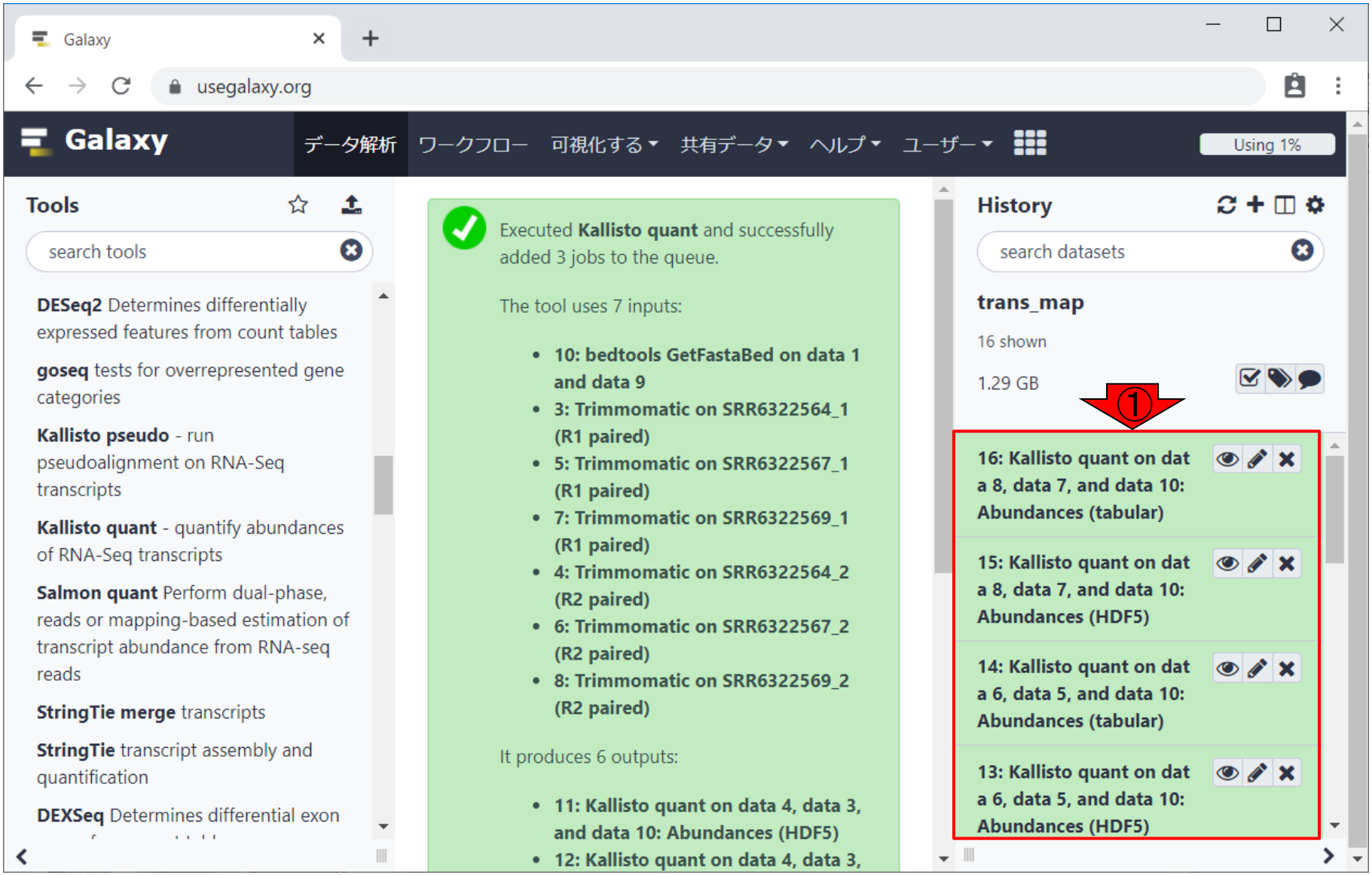
W8-7：実行中…

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。



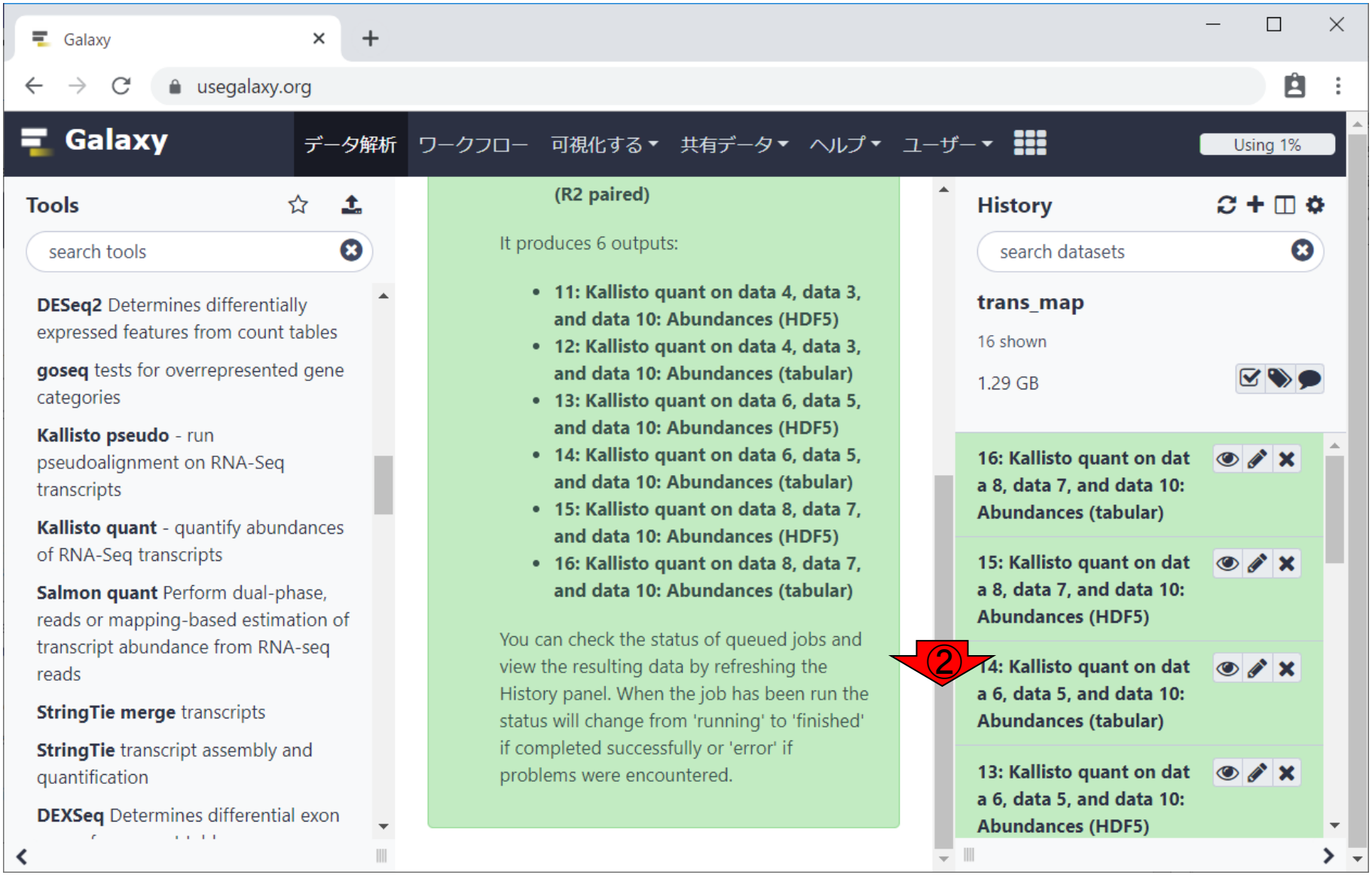
①実行完了。このときは約5分。

W8-8：実行完了



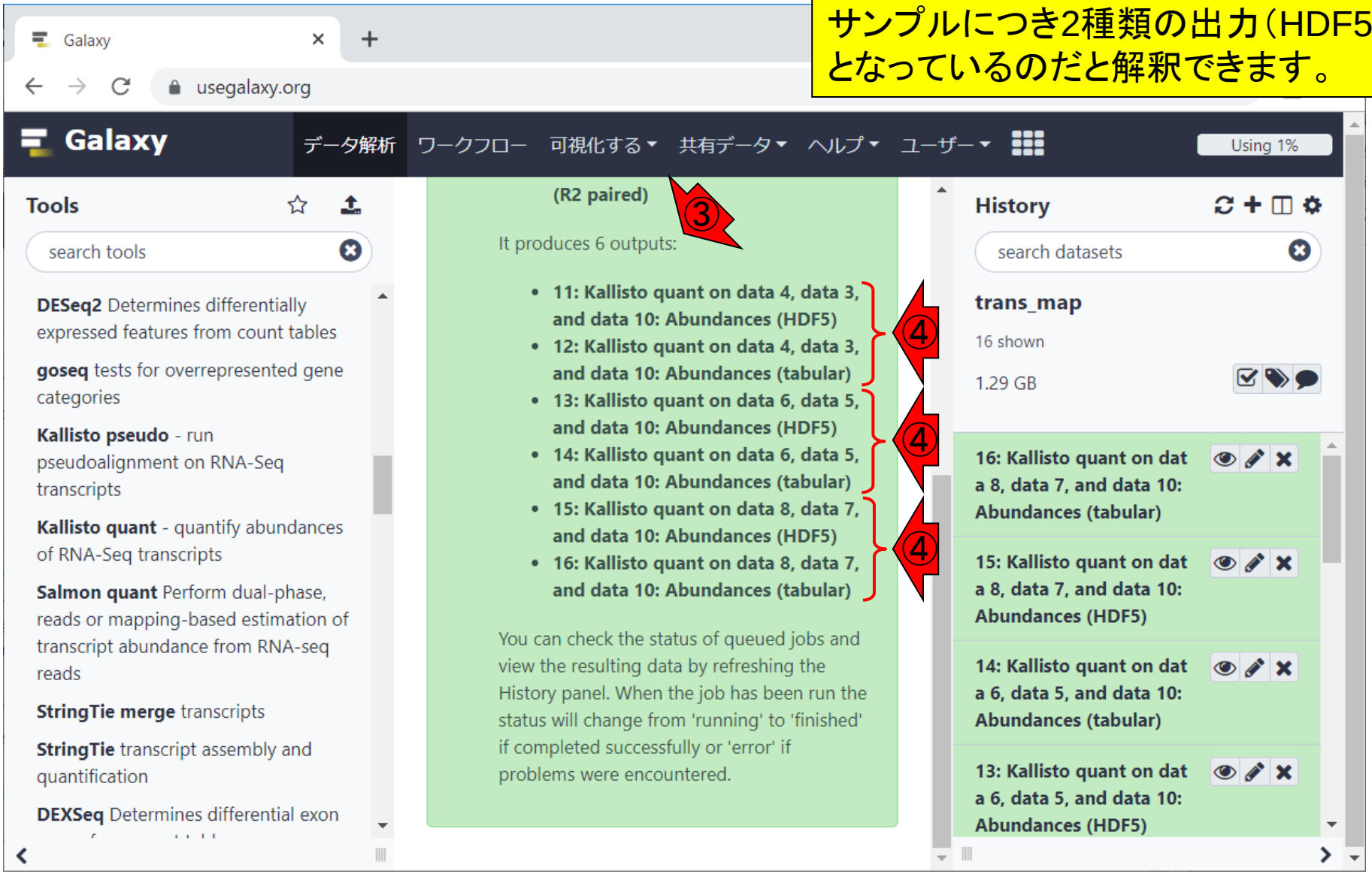
W8-8：実行完了

①実行完了。このときは約5分。②中央パネル下部に移動。



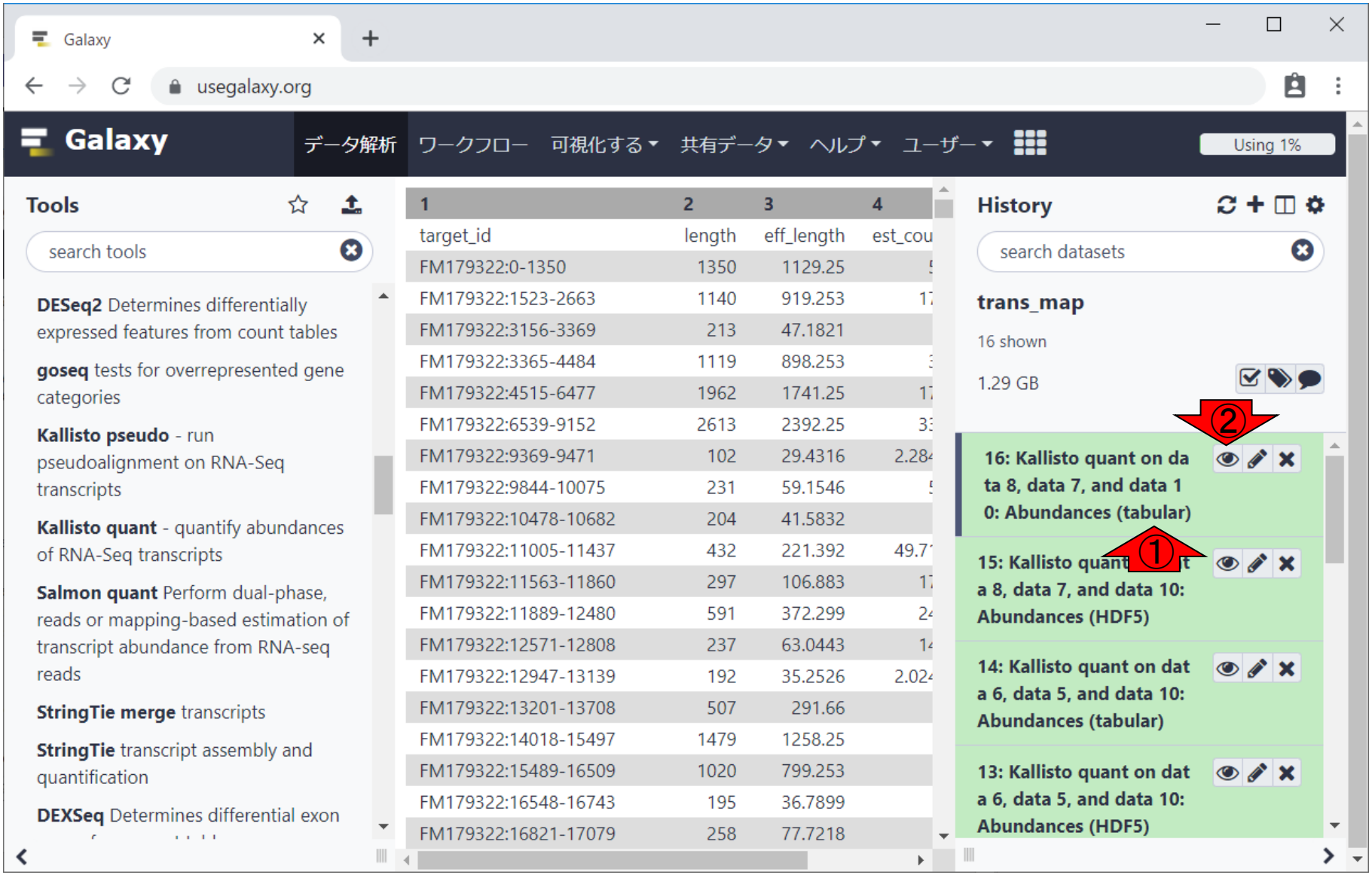
W8-8：実行完了

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。



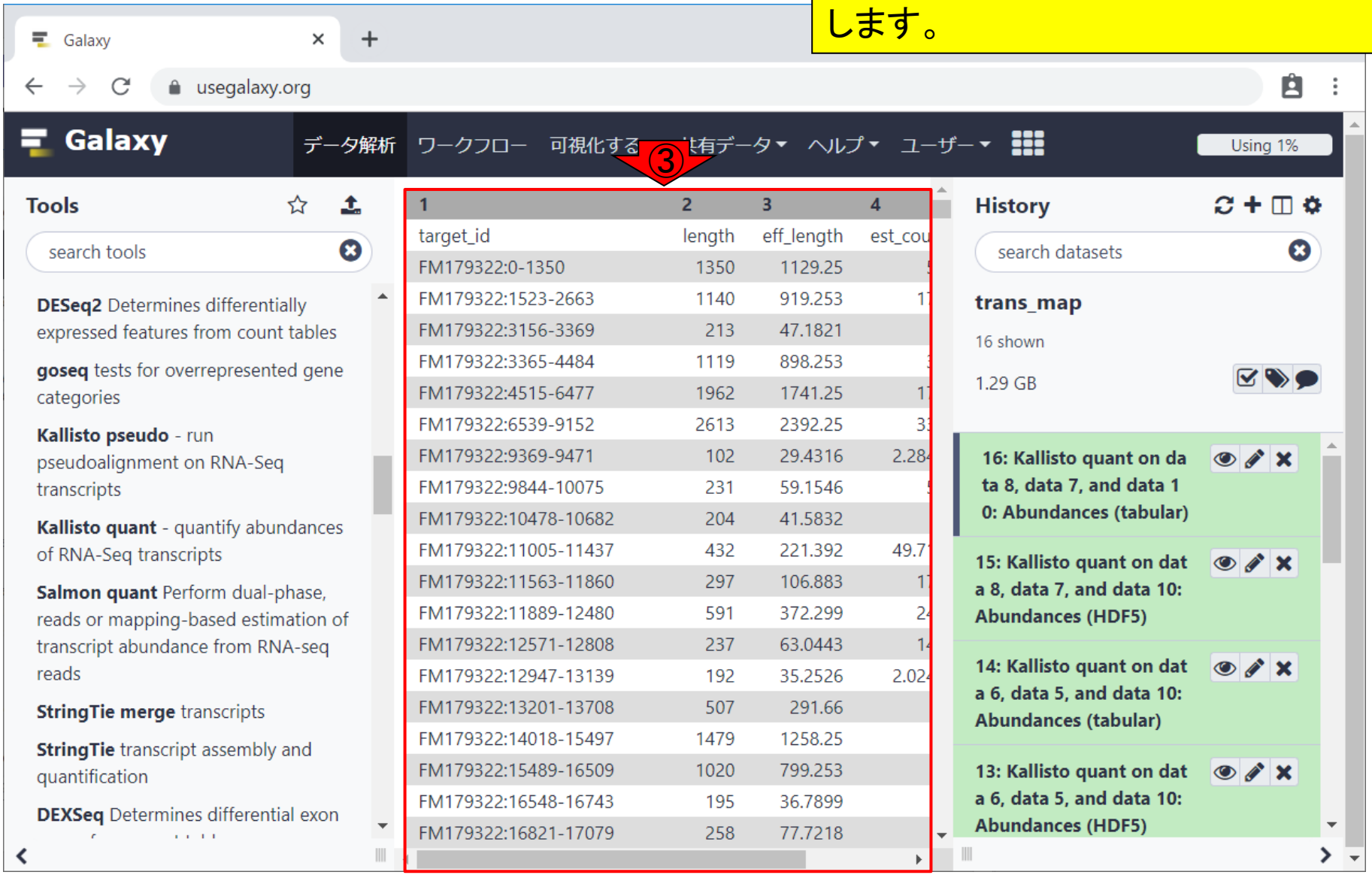
W8-9：tabular

①tabularのほうのデータを、②中央パネル上に表示させた状態。



W8-9：tabular

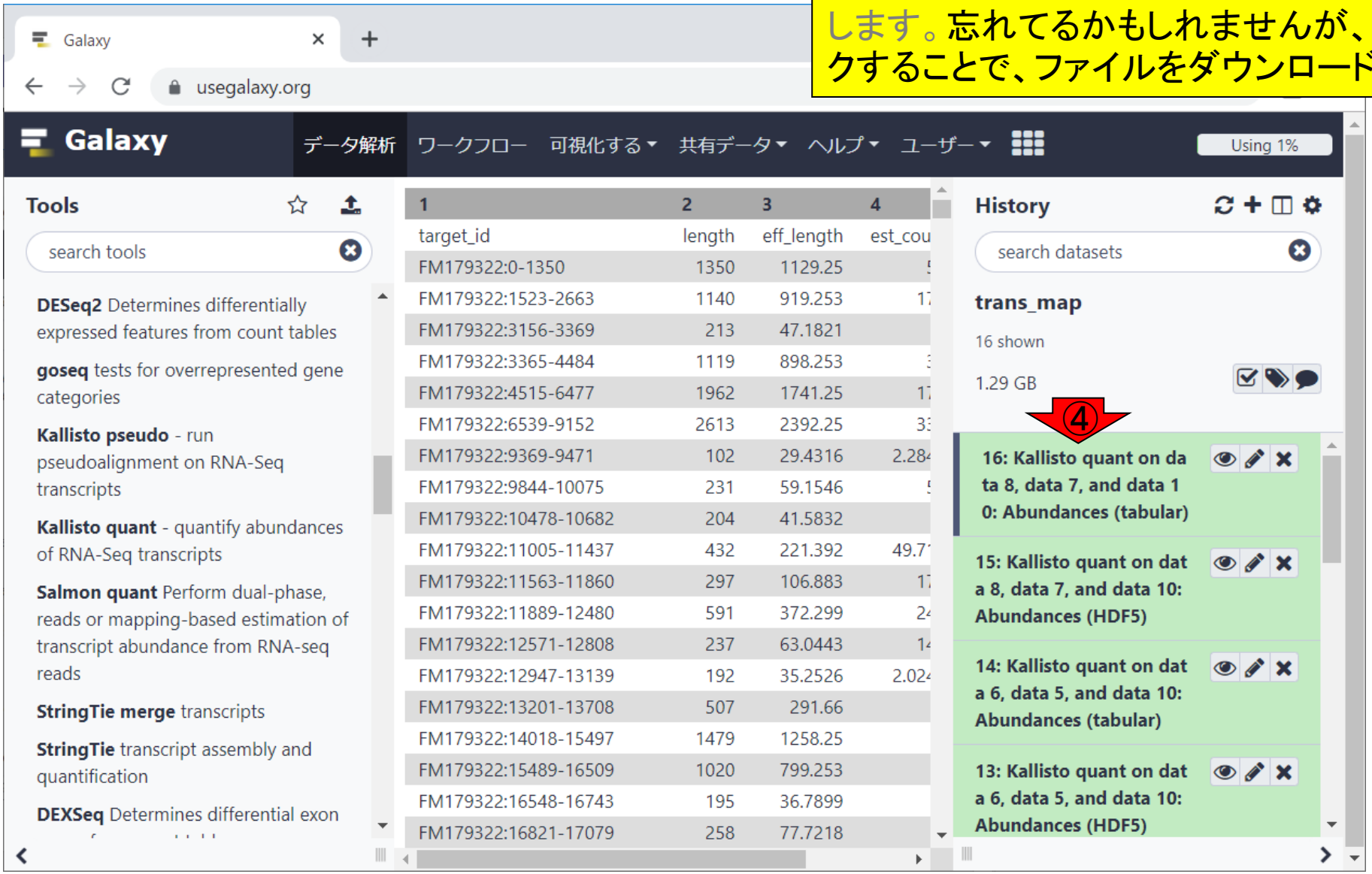
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。



参考

W8-9：tabular

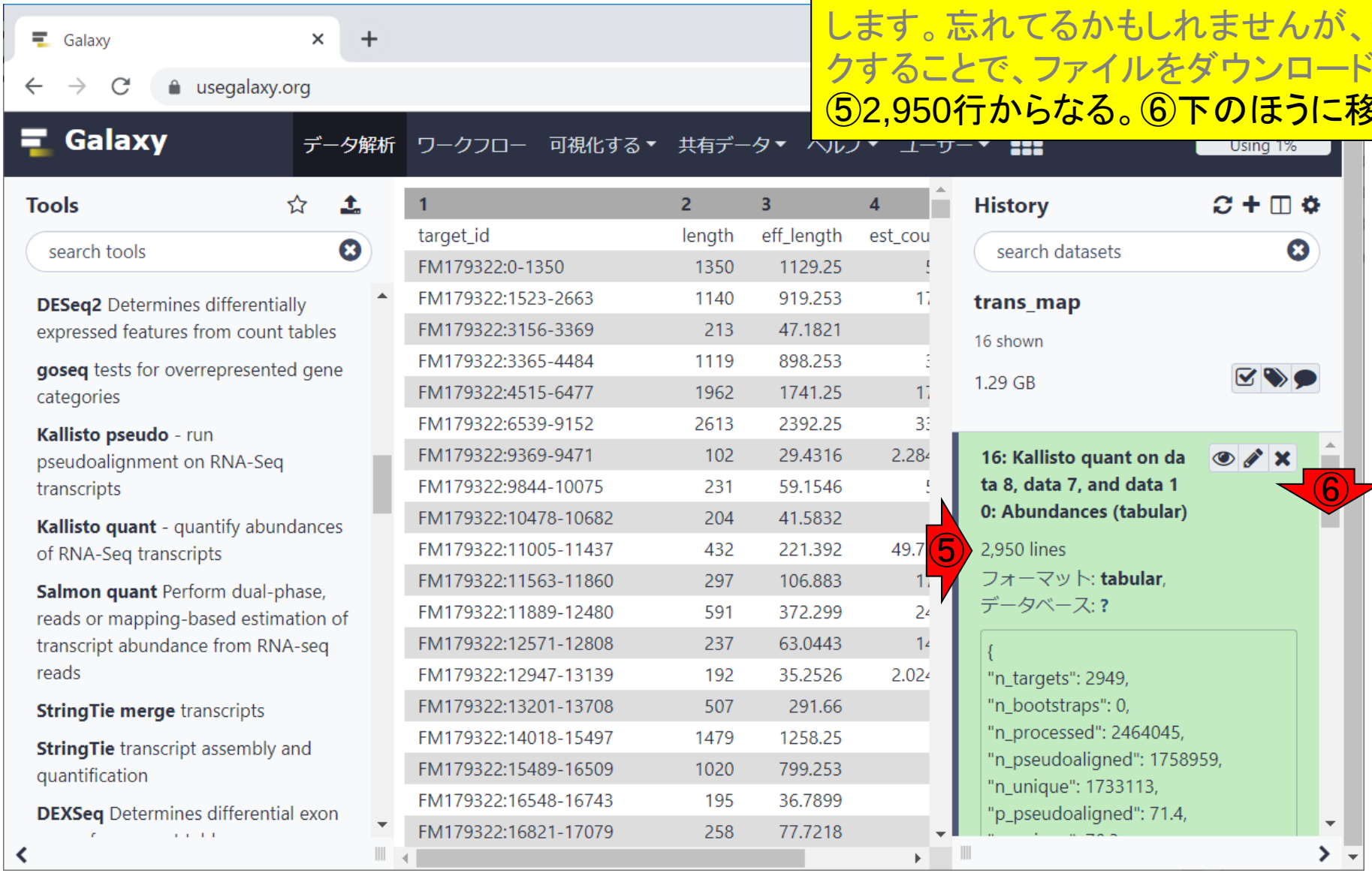
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。



参考

W8-9：tabular

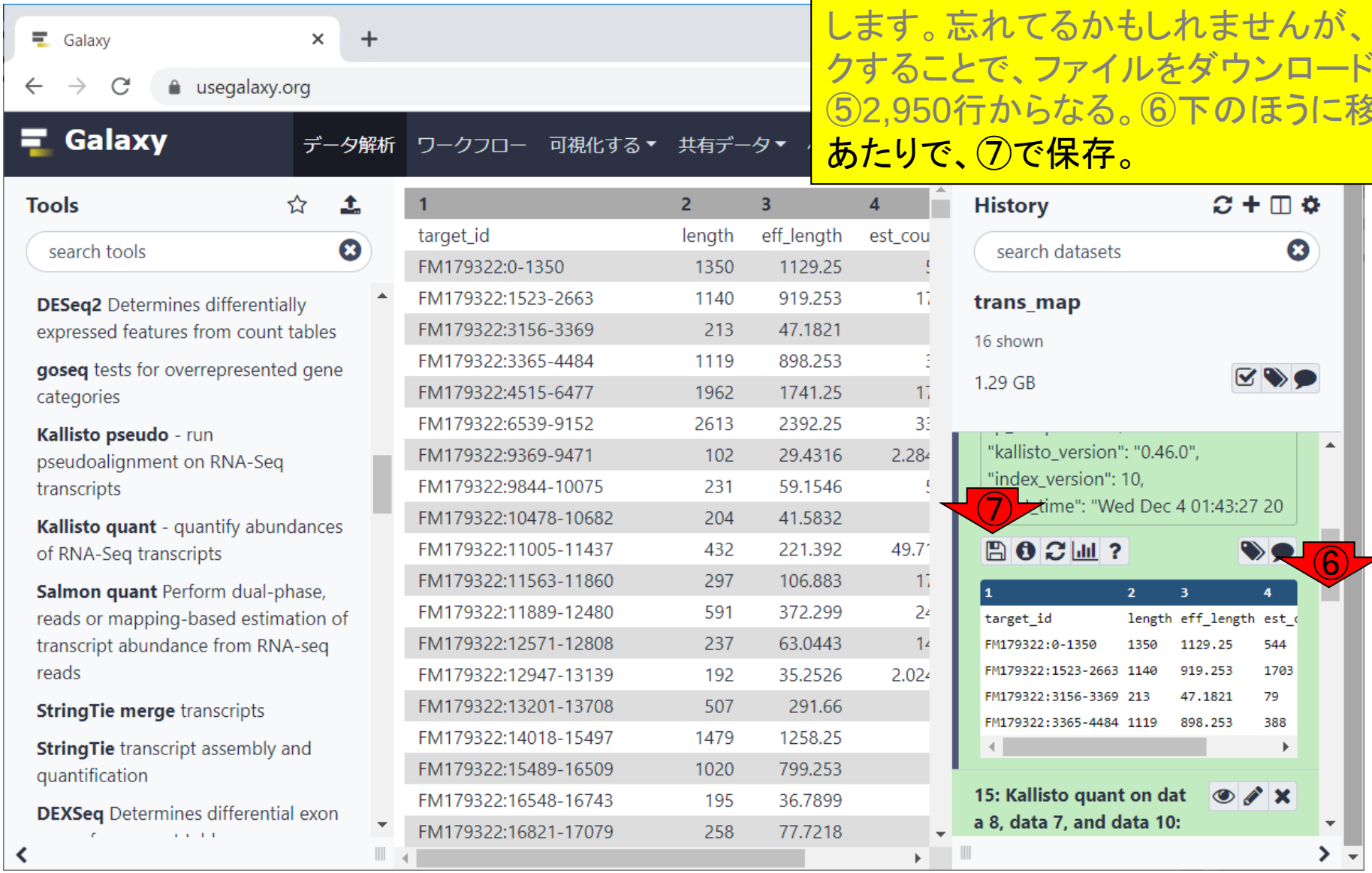
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。



参考

W8-9：tabular

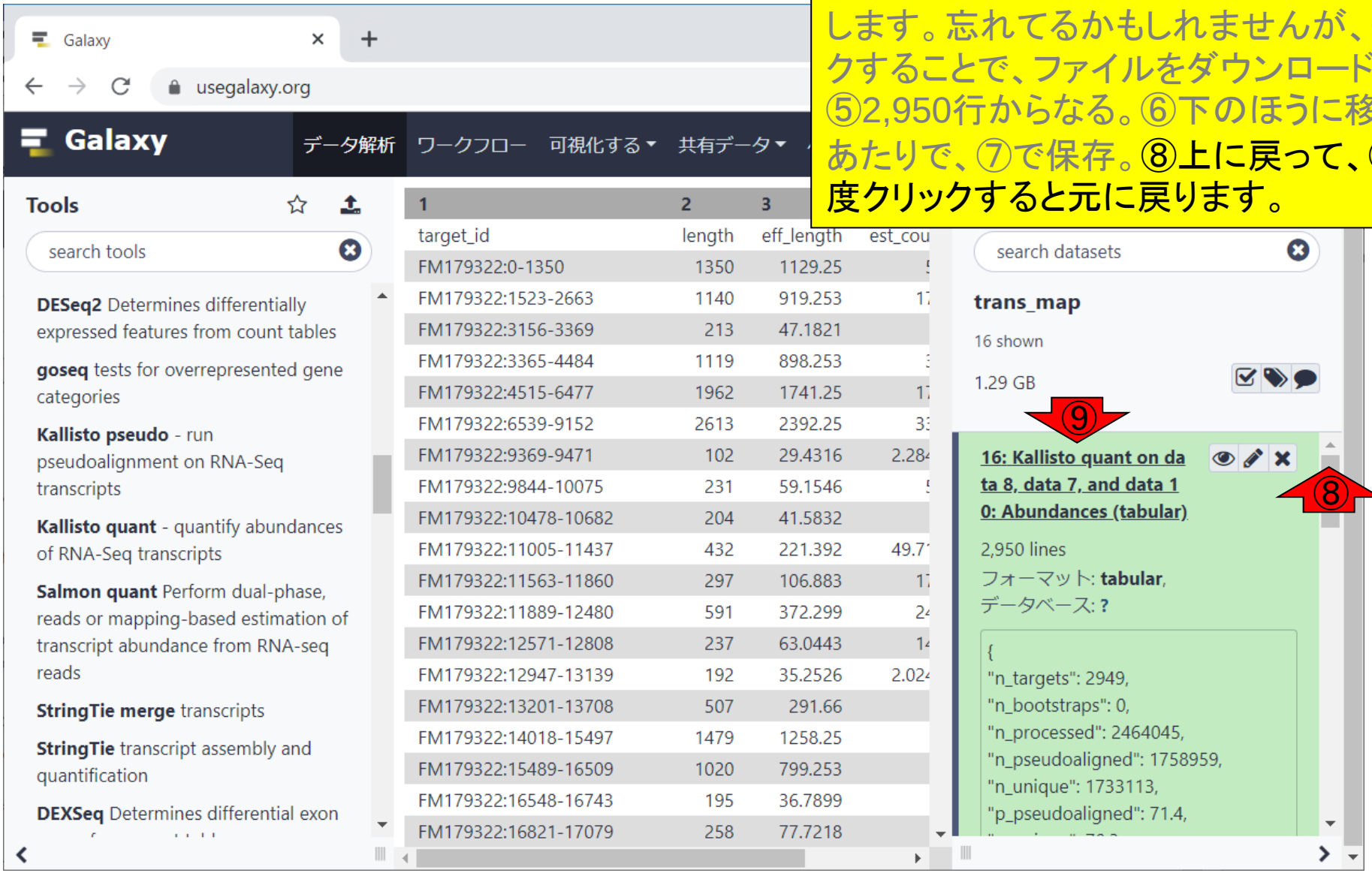
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。



参考

W8-9：tabular

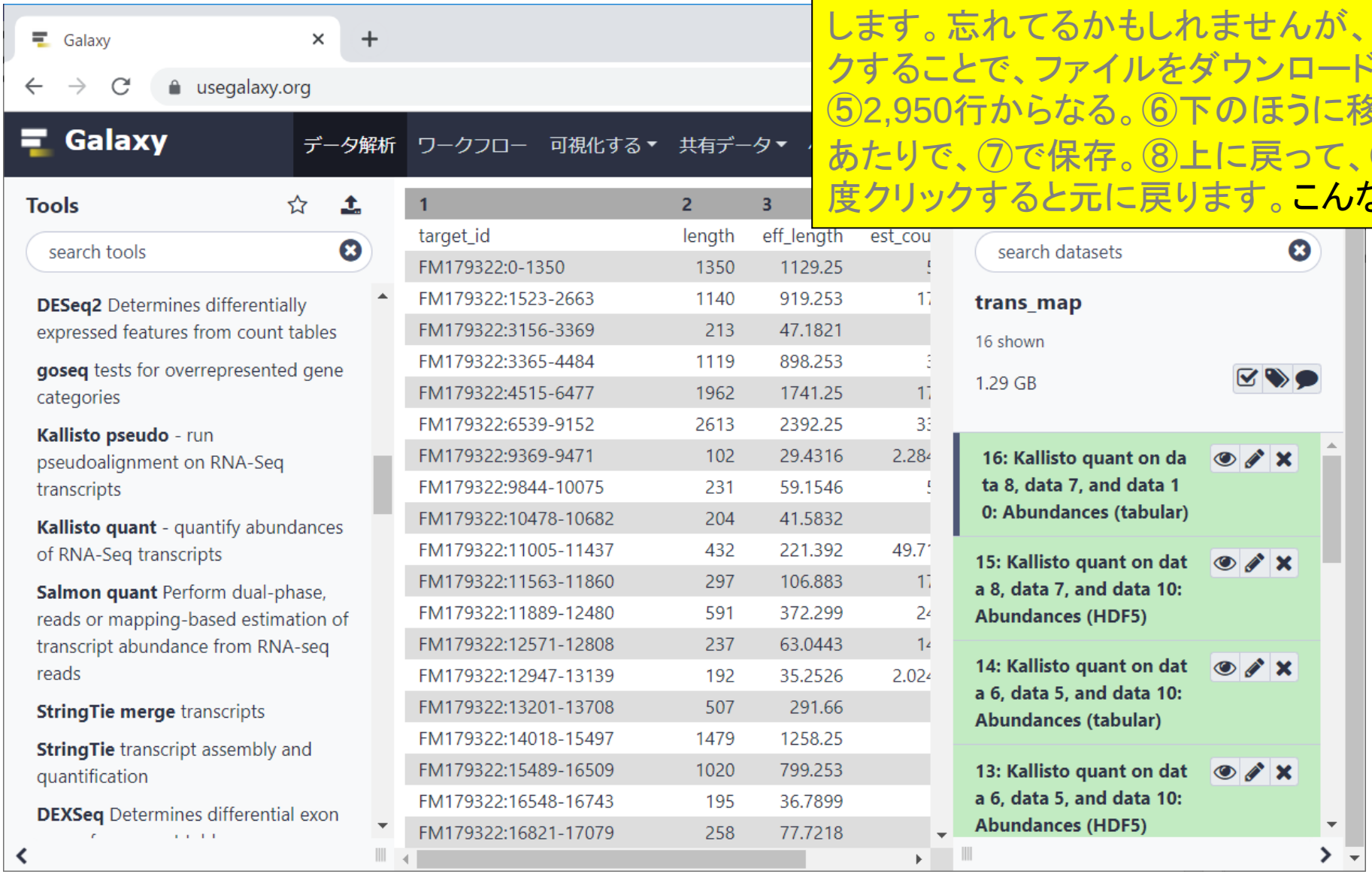
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。⑧上に戻って、⑨をもう一度クリックすると元に戻ります。



参考

W8-9：tabular

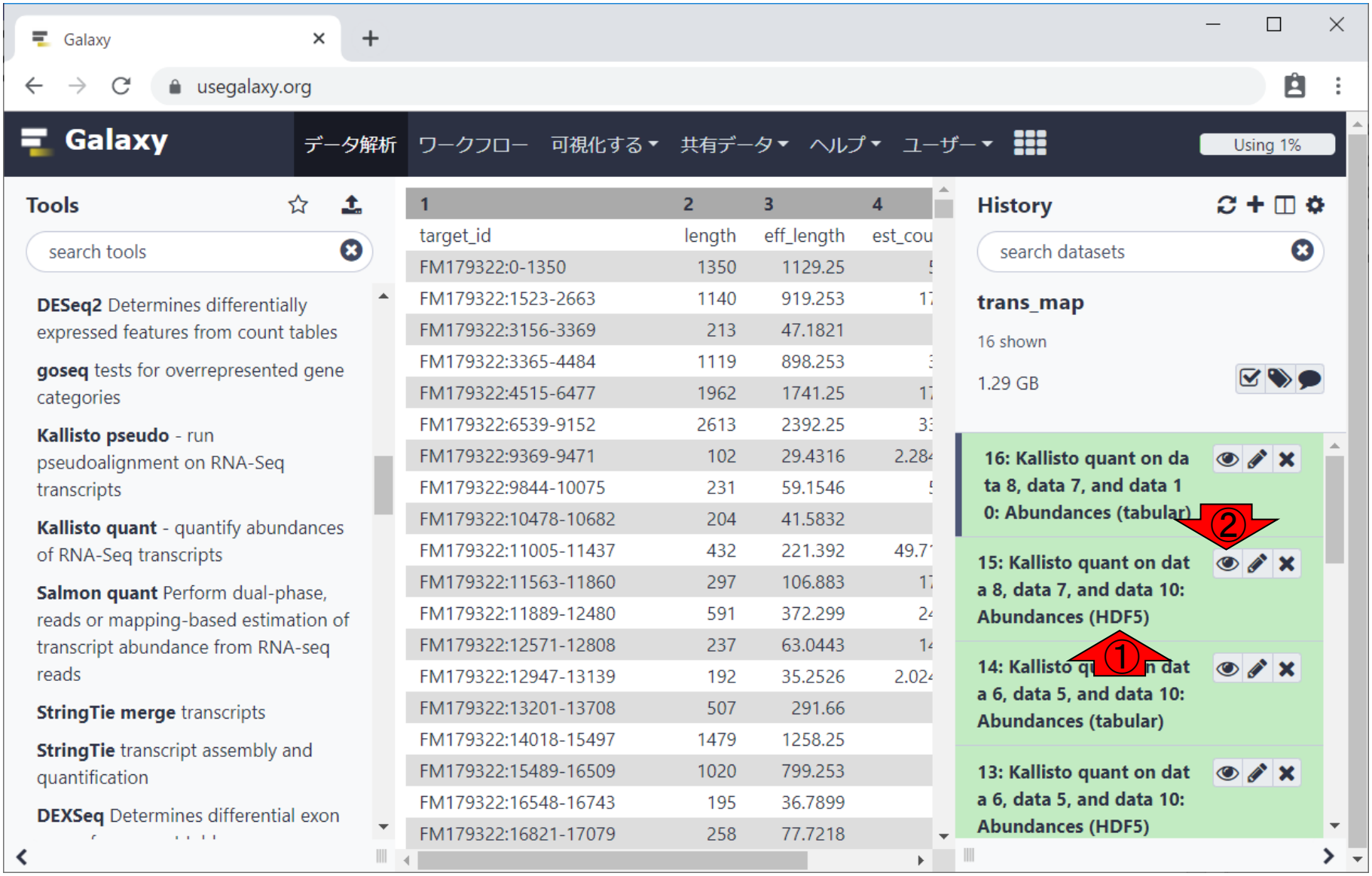
①tabularのほうのデータを、②中央パネル上に表示させた状態。③のように見えた段階で、「tabularってタブ区切りのことなのだろう」と解釈します。忘れてるかもしれませんが、④をクリックすることで、ファイルをダウンロードできます。⑤2,950行からなる。⑥下のほうに移動。このあたりで、⑦で保存。⑧上に戻って、⑨をもう一度クリックすると元に戻ります。こんな感じ。



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…。

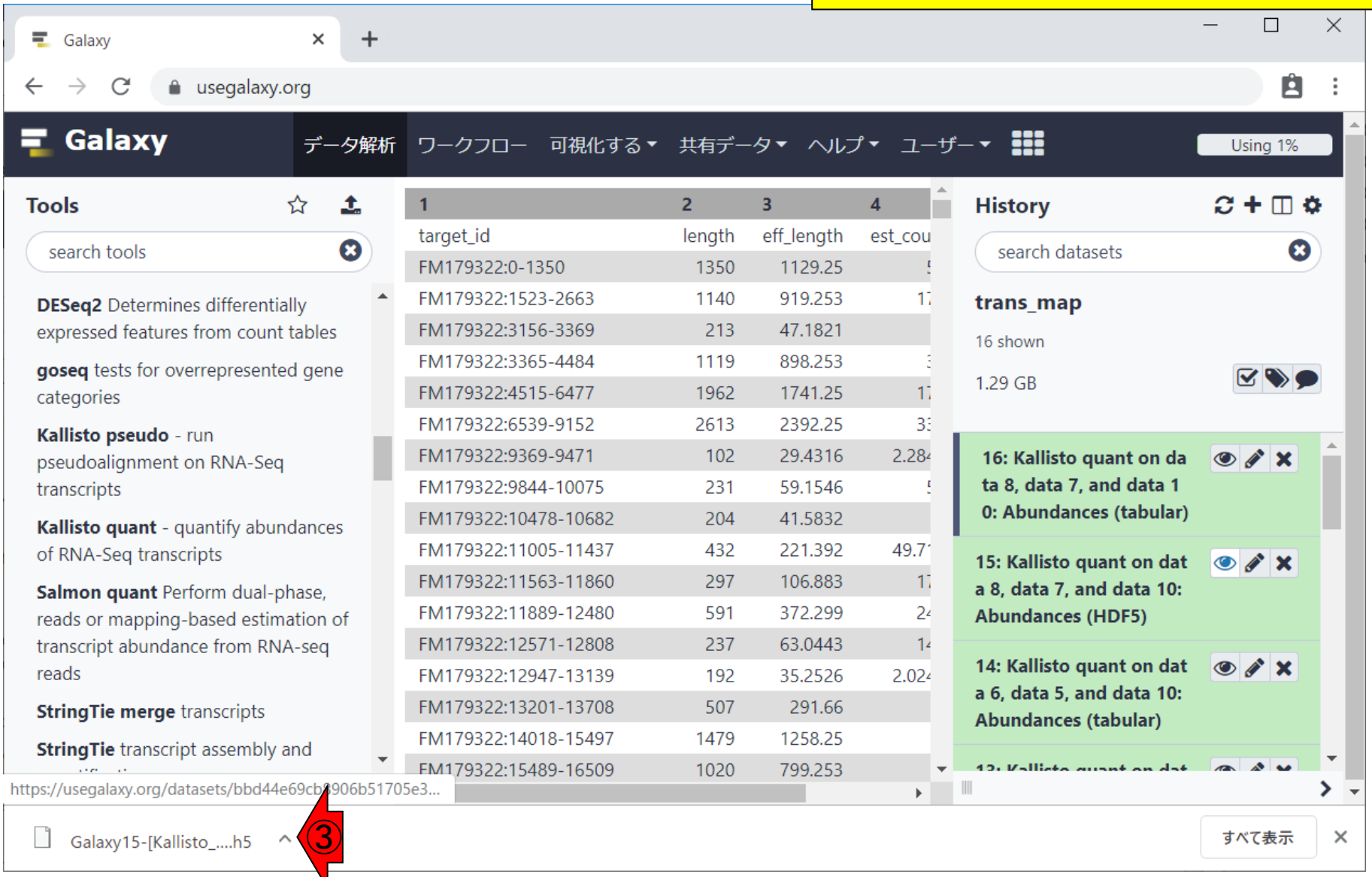
W8-10：HDF5



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

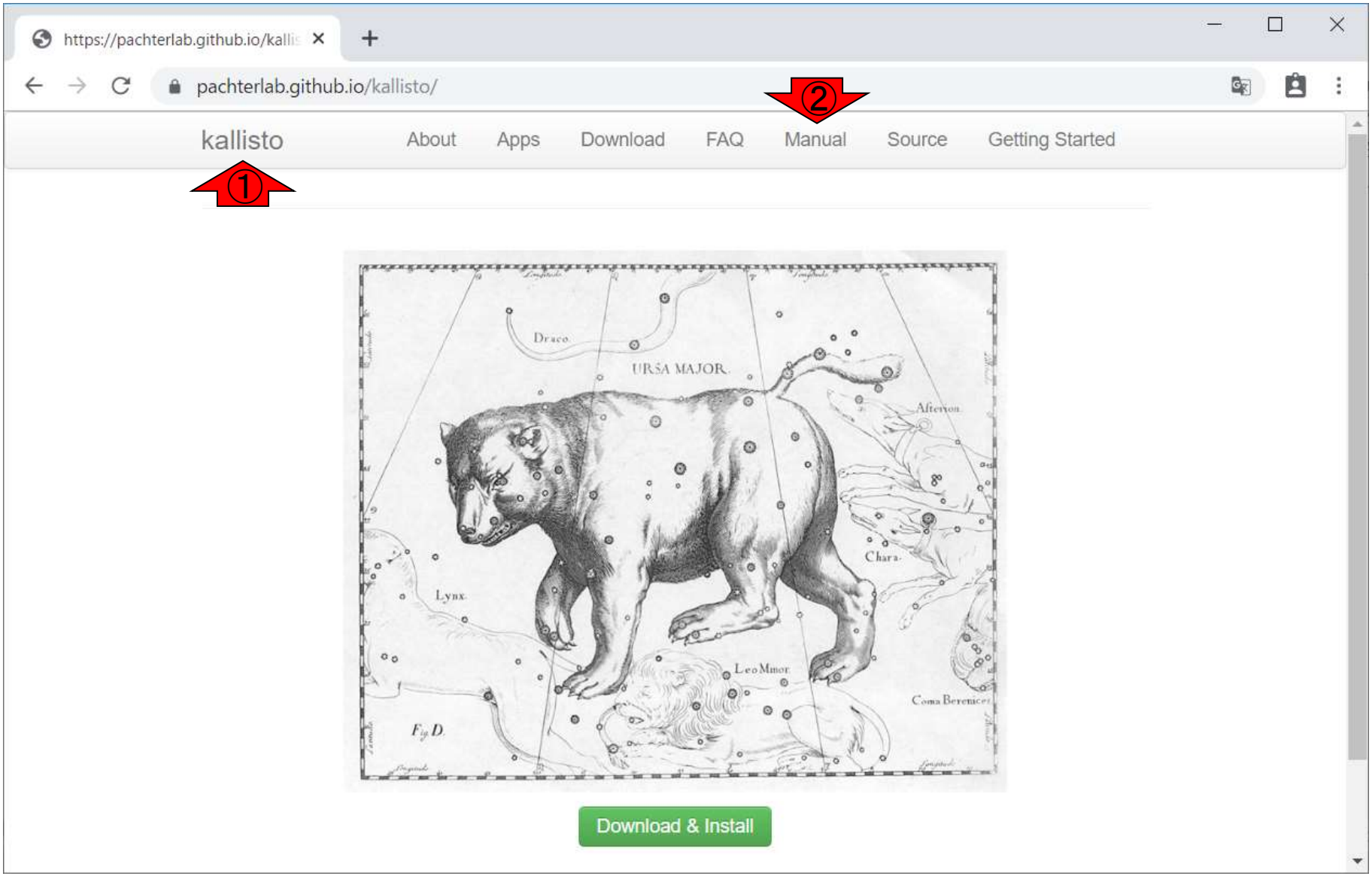


Contents

- 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

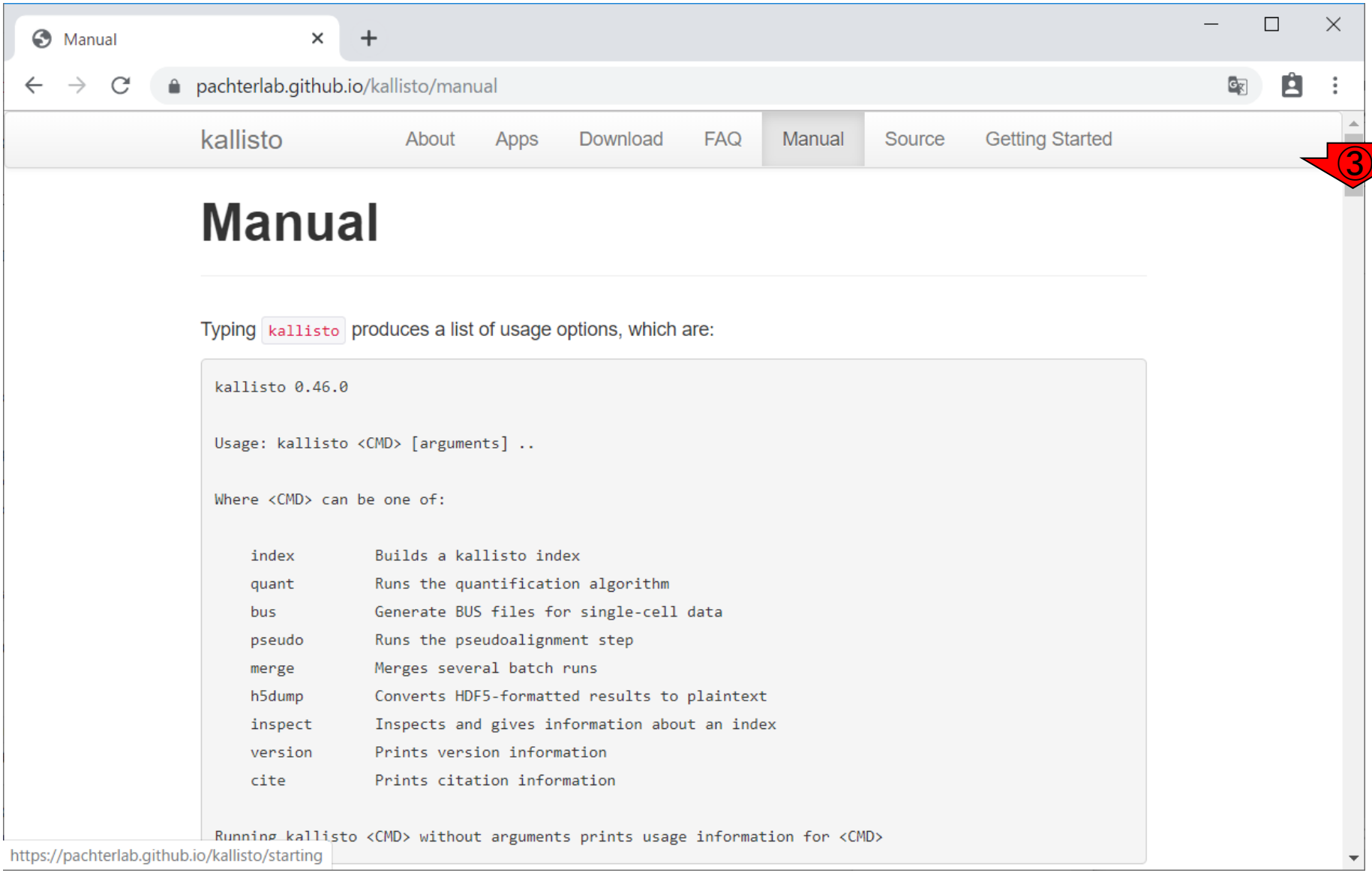
①Kallistoのページ。②Manual。

W9-1：マニュアル



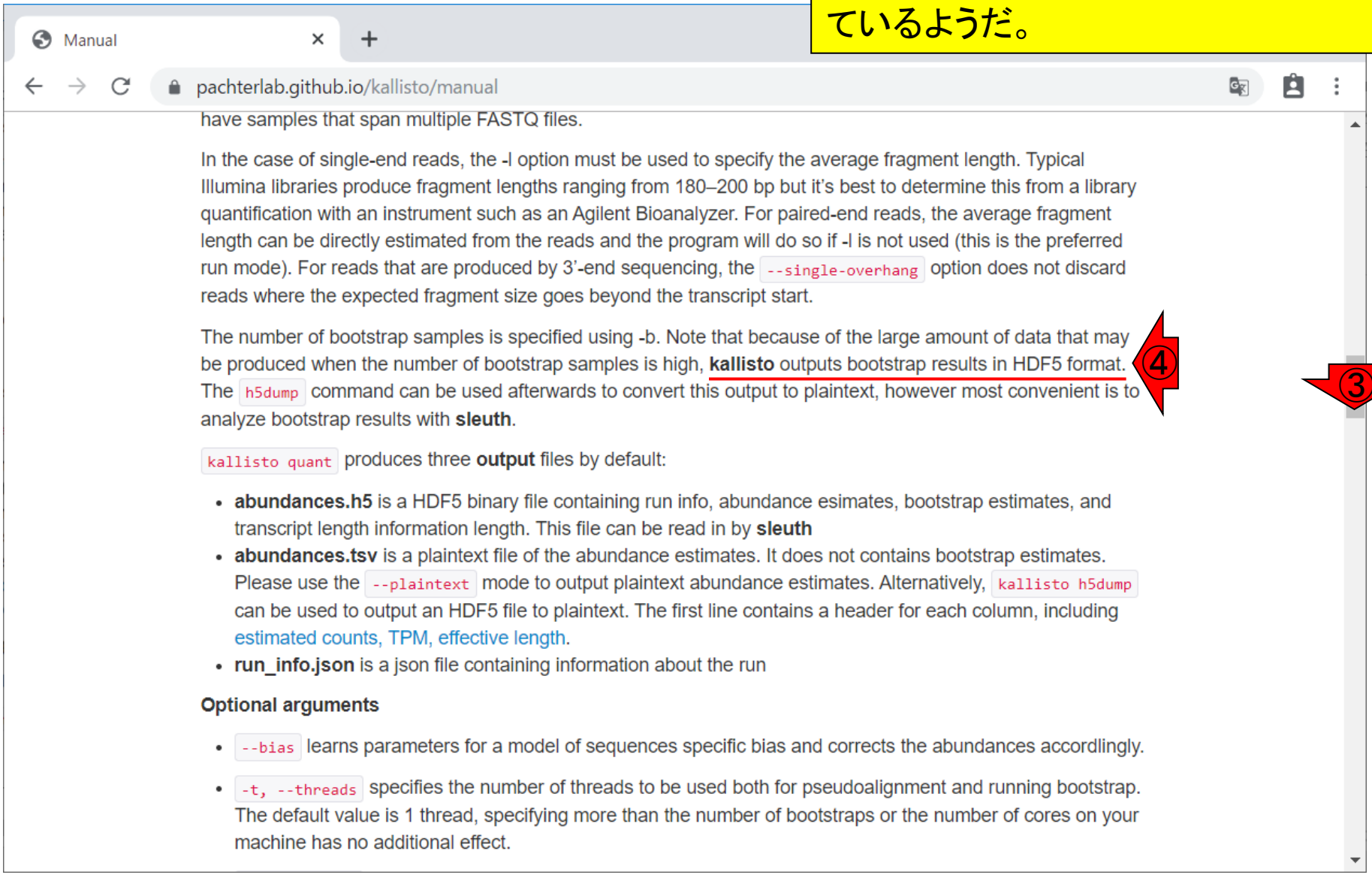
W9-1：マニュアル

①Kallistoのページ。②Manual。③ページ下部に移動。



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。



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。

Manual

pachterlab.github.io/kallisto/manual

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**.

⑤

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

- **abundances.h5** is a HDF5 binary file containing run info, abundance esimates, 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 accordlingly.
- `-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。

Manual

pachterlab.github.io/kallisto/manual

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 esimates, 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 accordlingly.
- `-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。

Manual

pachterlab.github.io/kallisto/manual

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 esimates, 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.

②

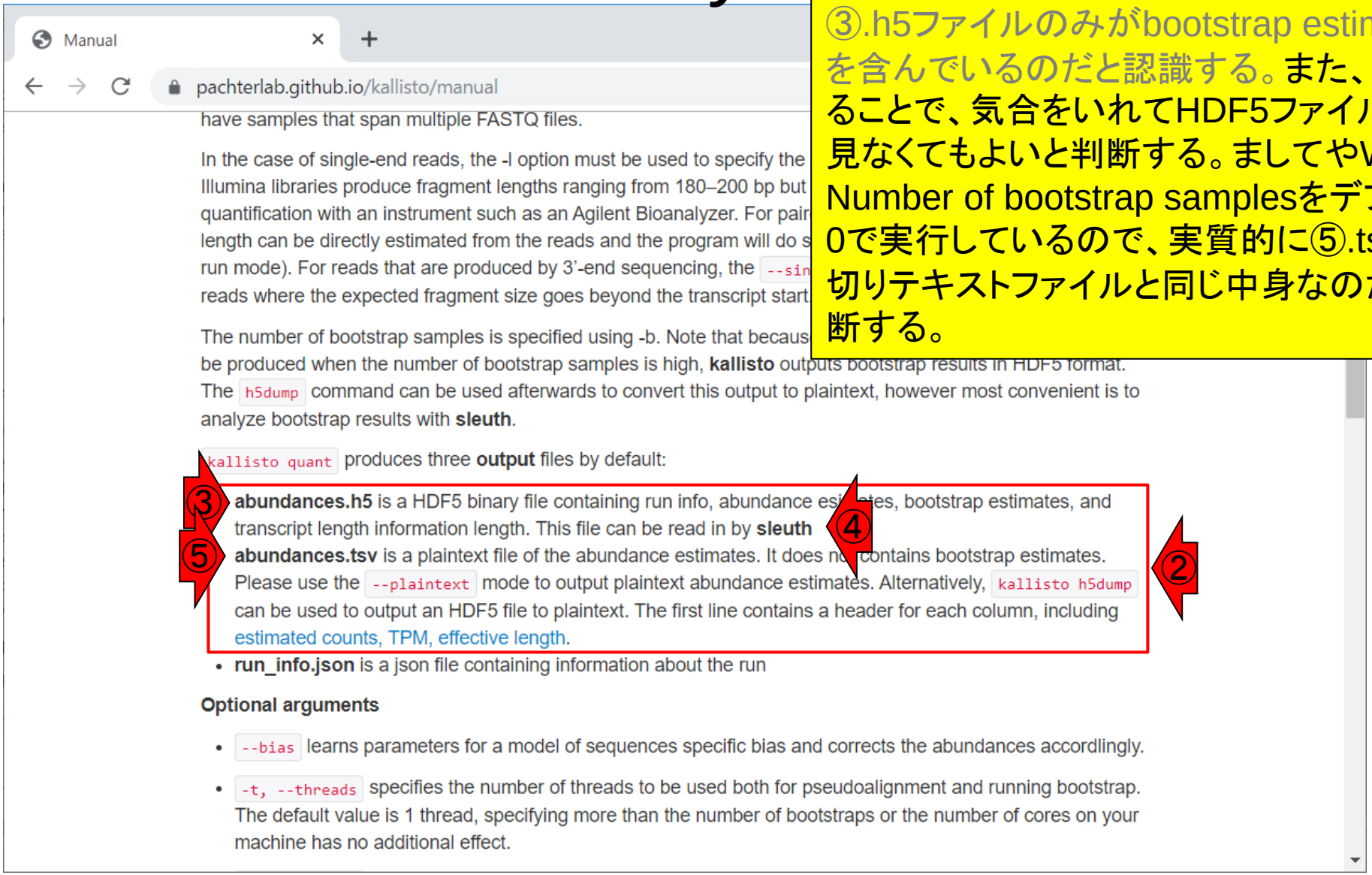
- **run_info.json** is a json file containing information about the run

Optional arguments

- `--bias` learns parameters for a model of sequences specific bias and corrects the abundances accordlingly.
- `-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。

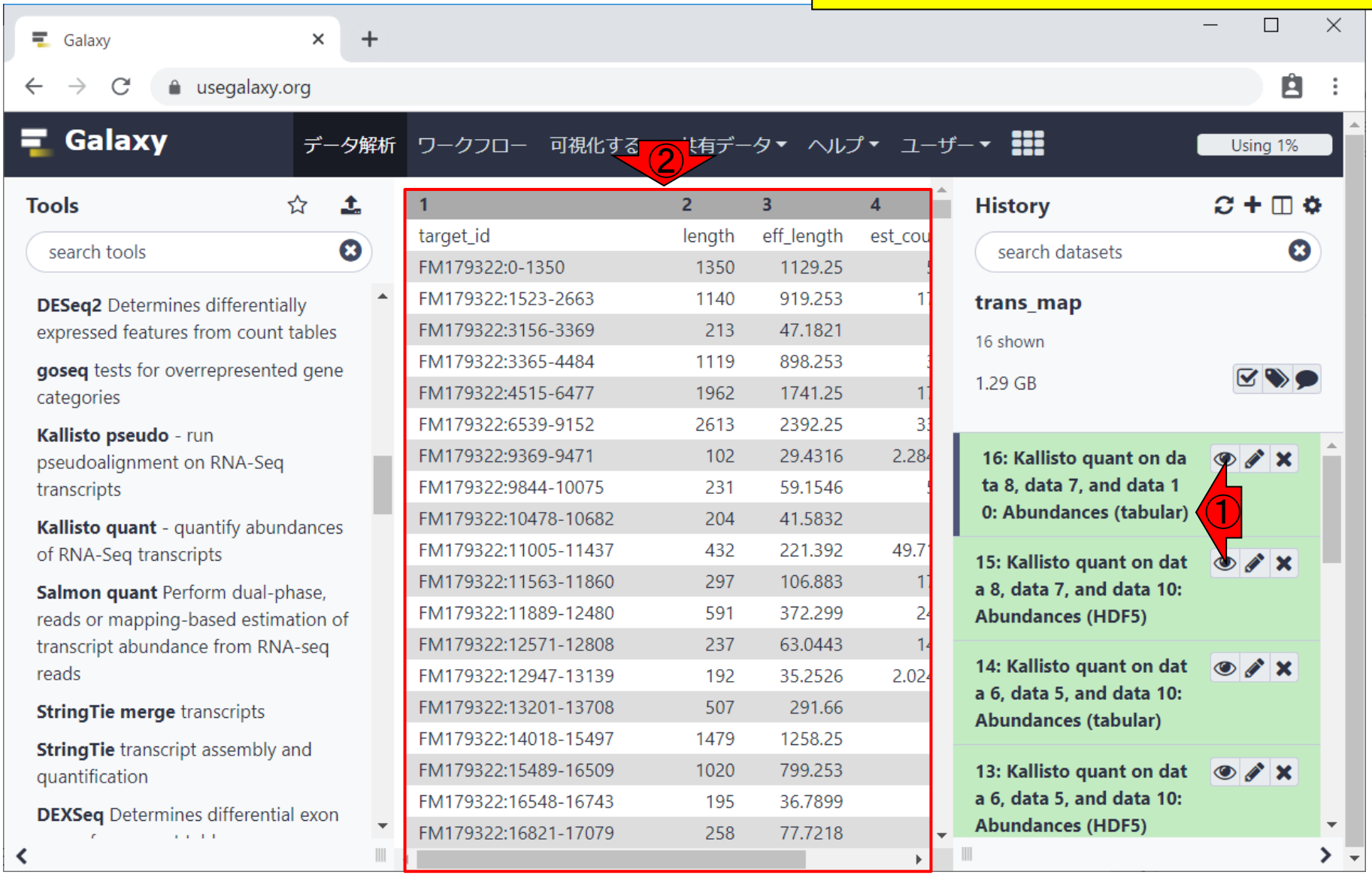


Contents

- 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

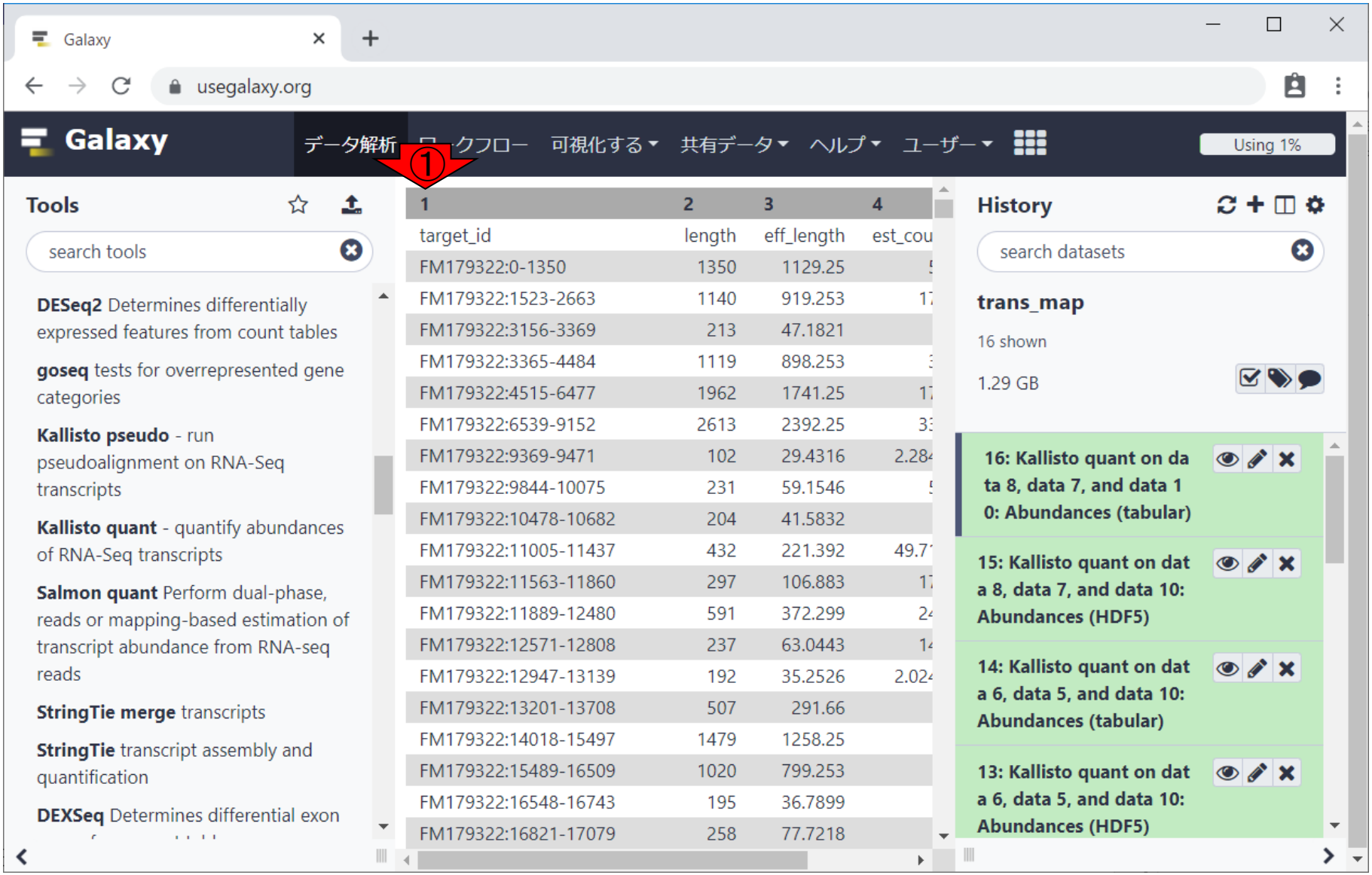
W10-1：結果の解説

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。



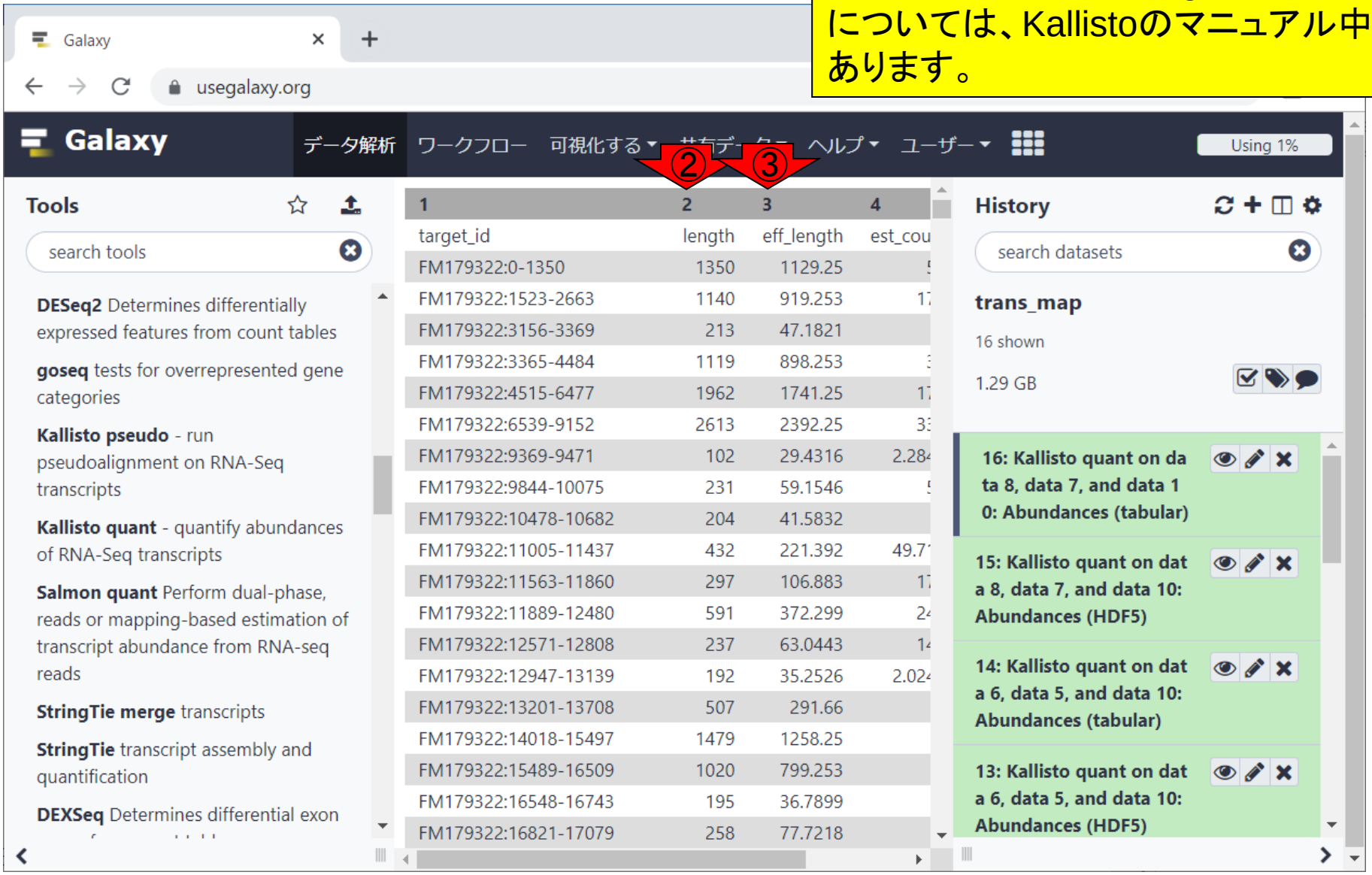
①1列目（target_id）は配列名。

W10-2：1列目



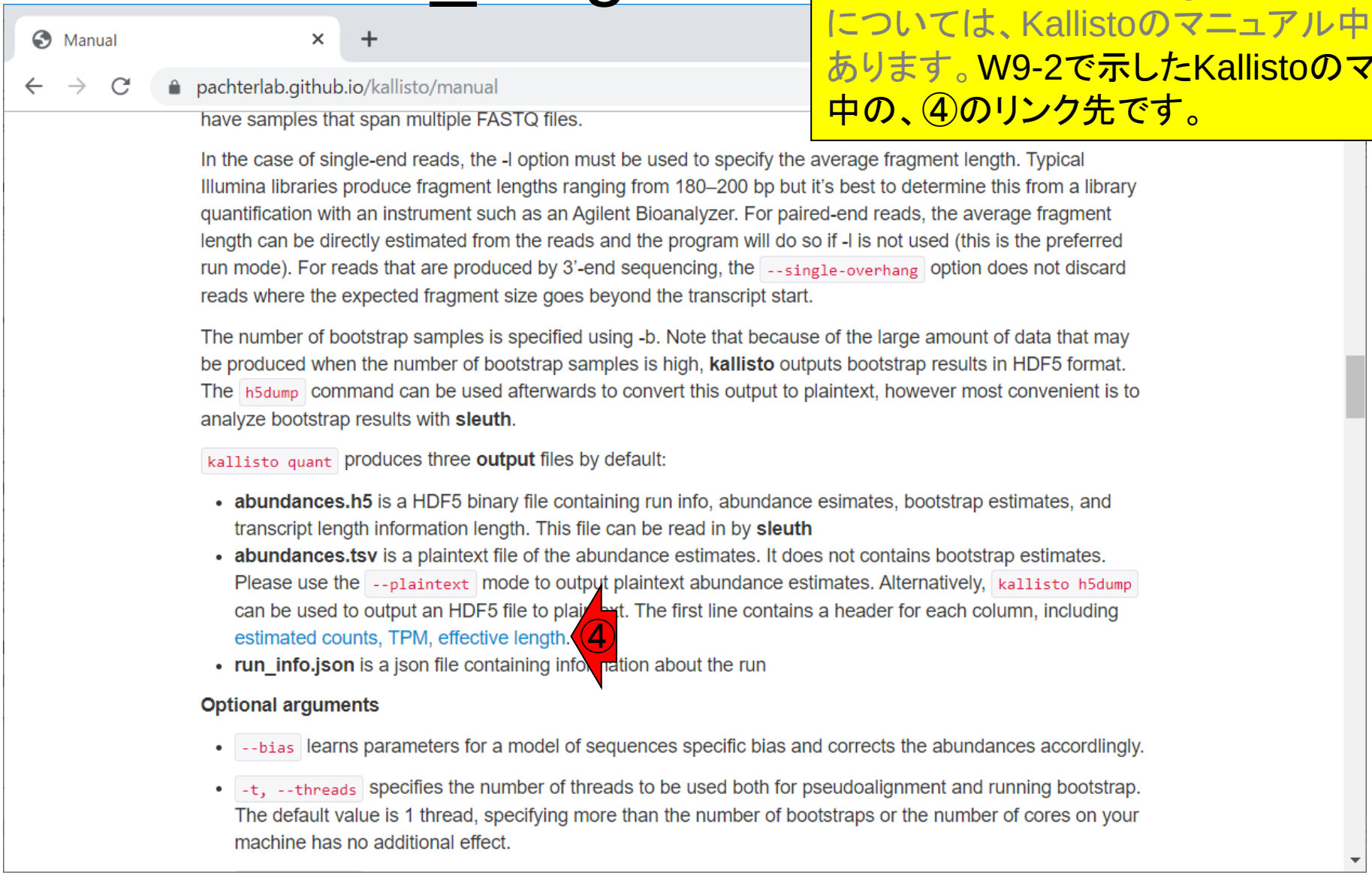
W10-3：2-3列目

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。



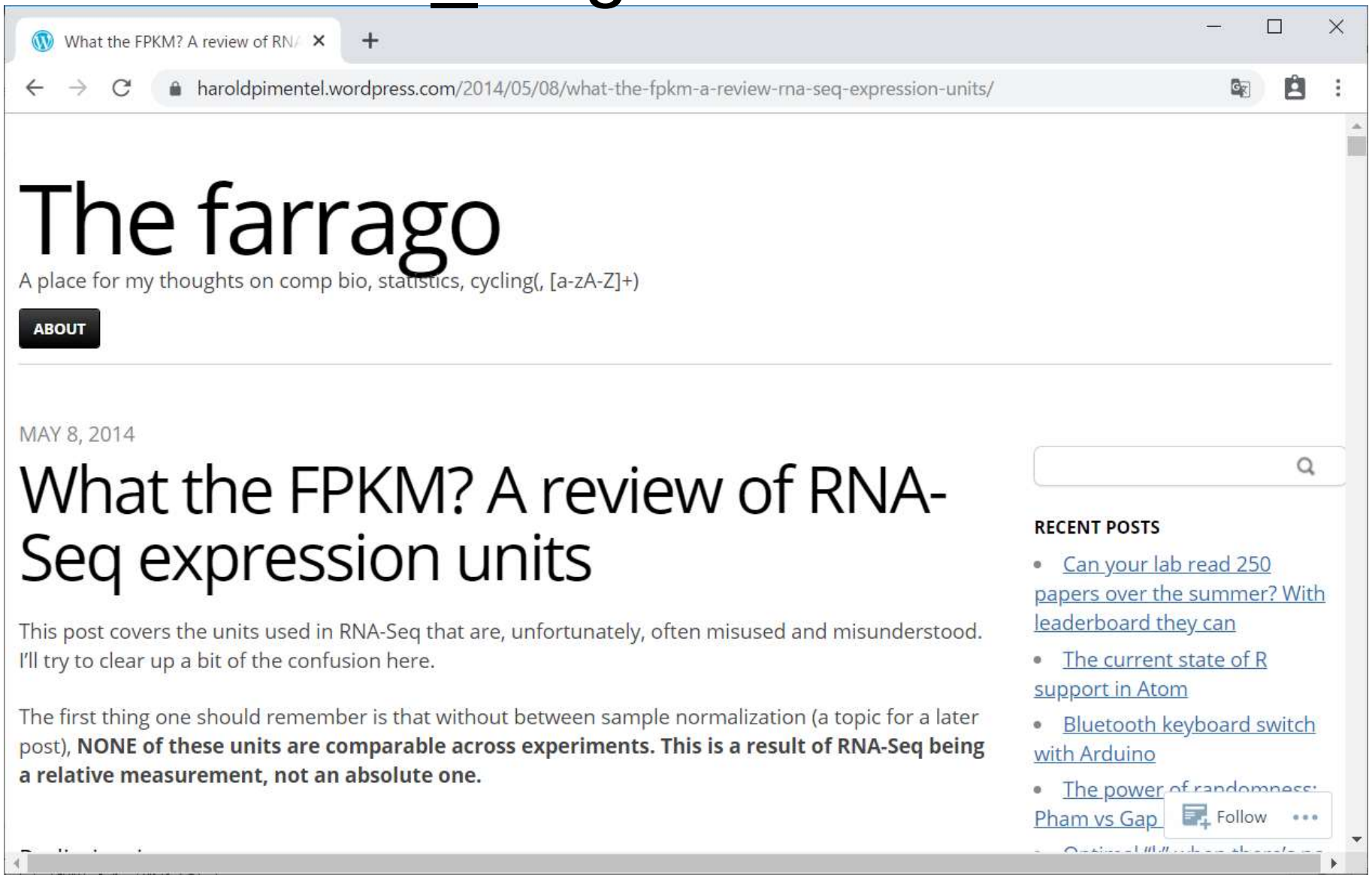
W10-4：eff_length

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。



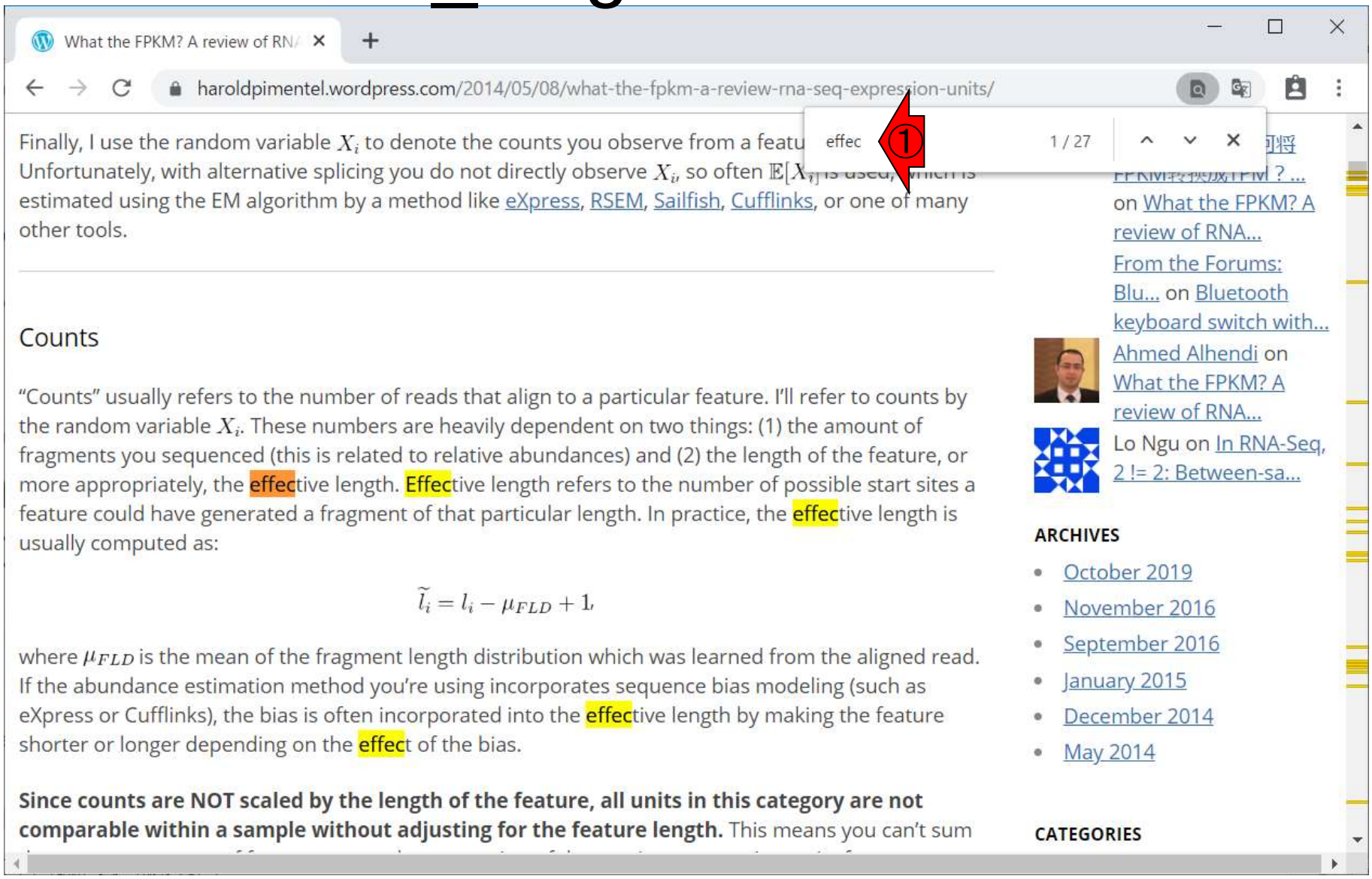
（リンクをクリックすると）こんな感じになります。

W10-4：eff_length



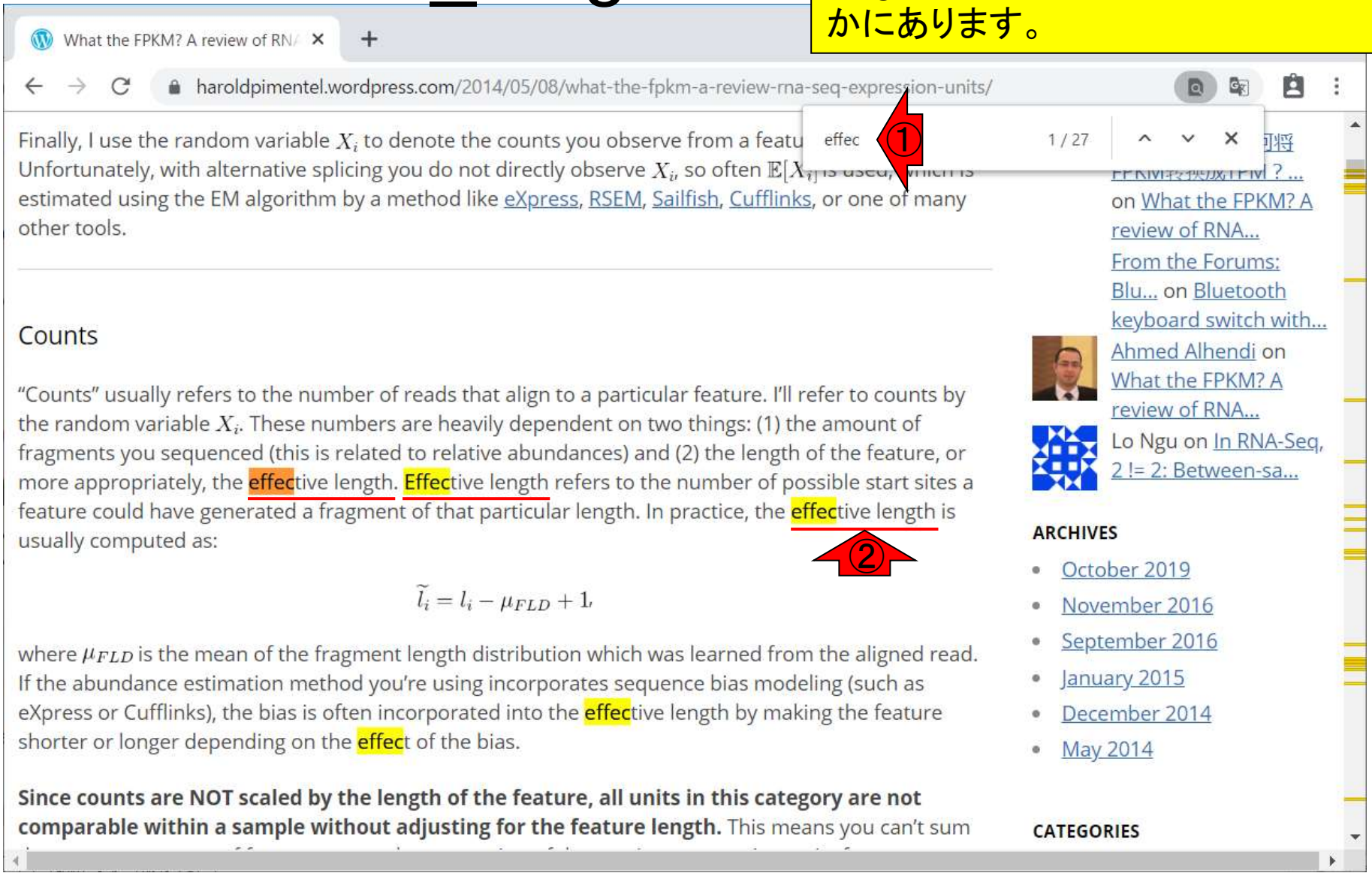
W10-4：eff_length

（リンクをクリックすると）こんな感じになります。
①「effec」でページ内検索したところ。



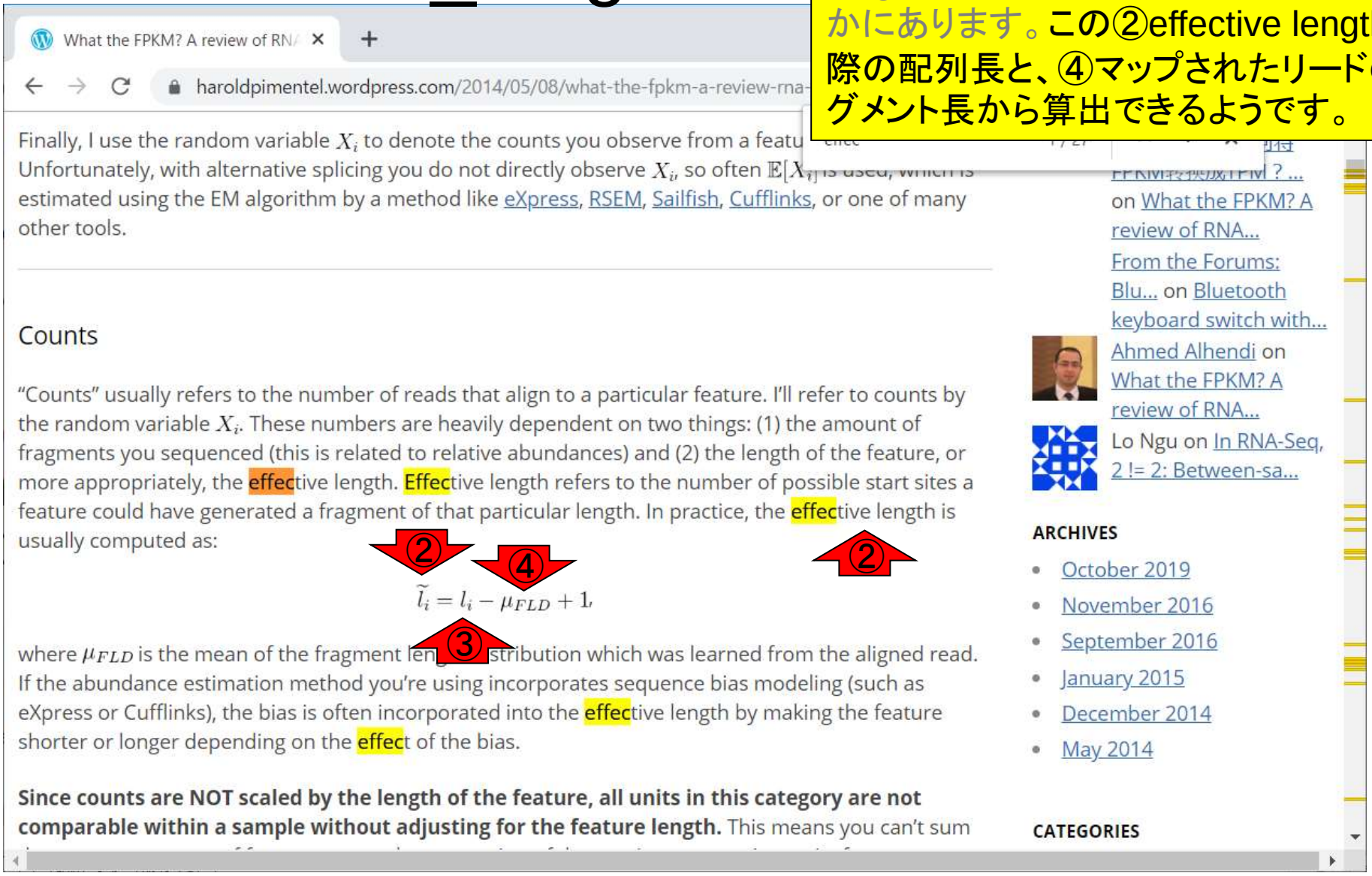
W10-4：eff_length

（リンクをクリックすると）こんな感じになります。①「effec」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。



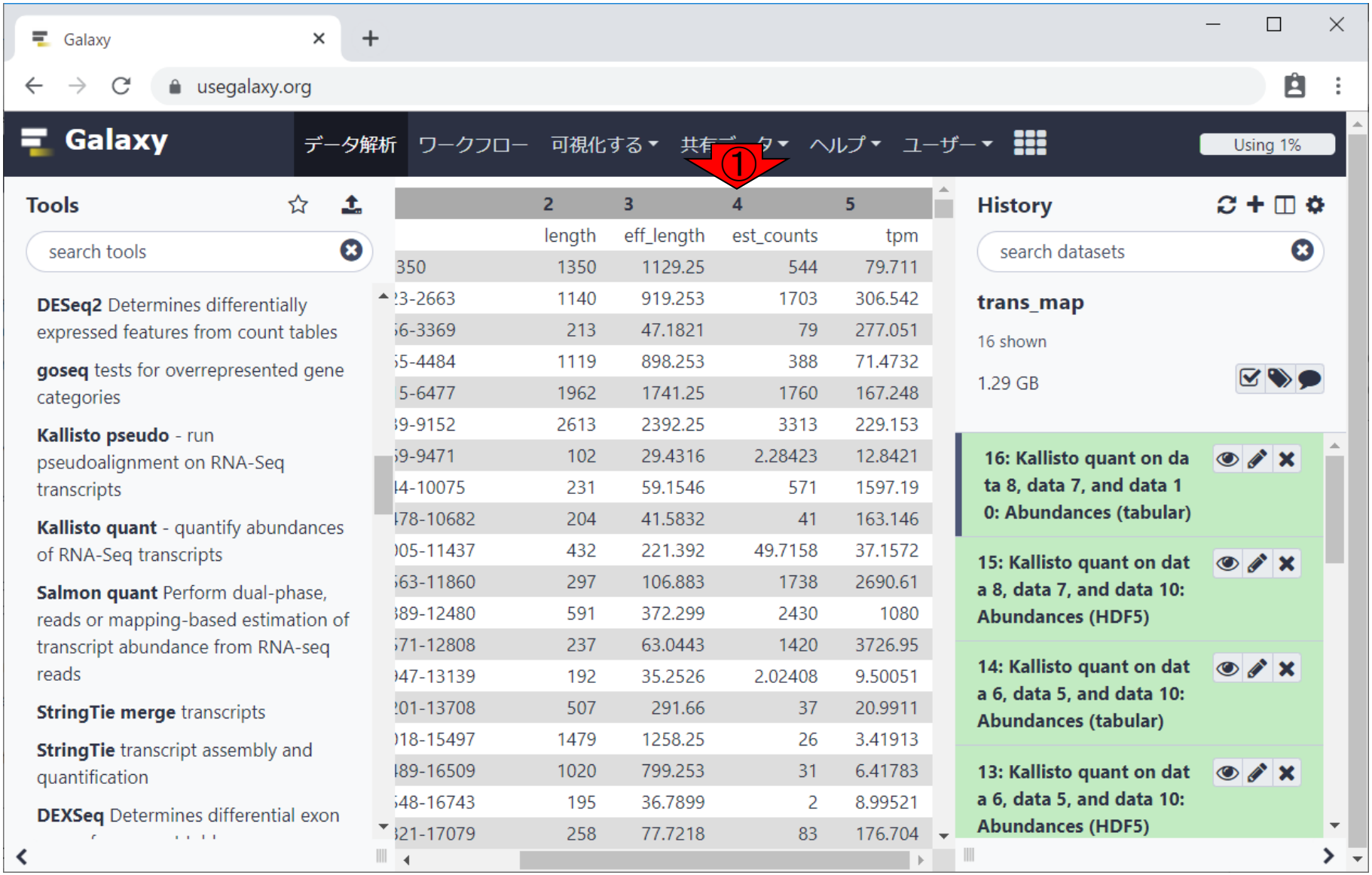
W10-4：eff_length

（リンクをクリックすると）こんな感じになります。①「effec」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。



W10-5：4列目

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

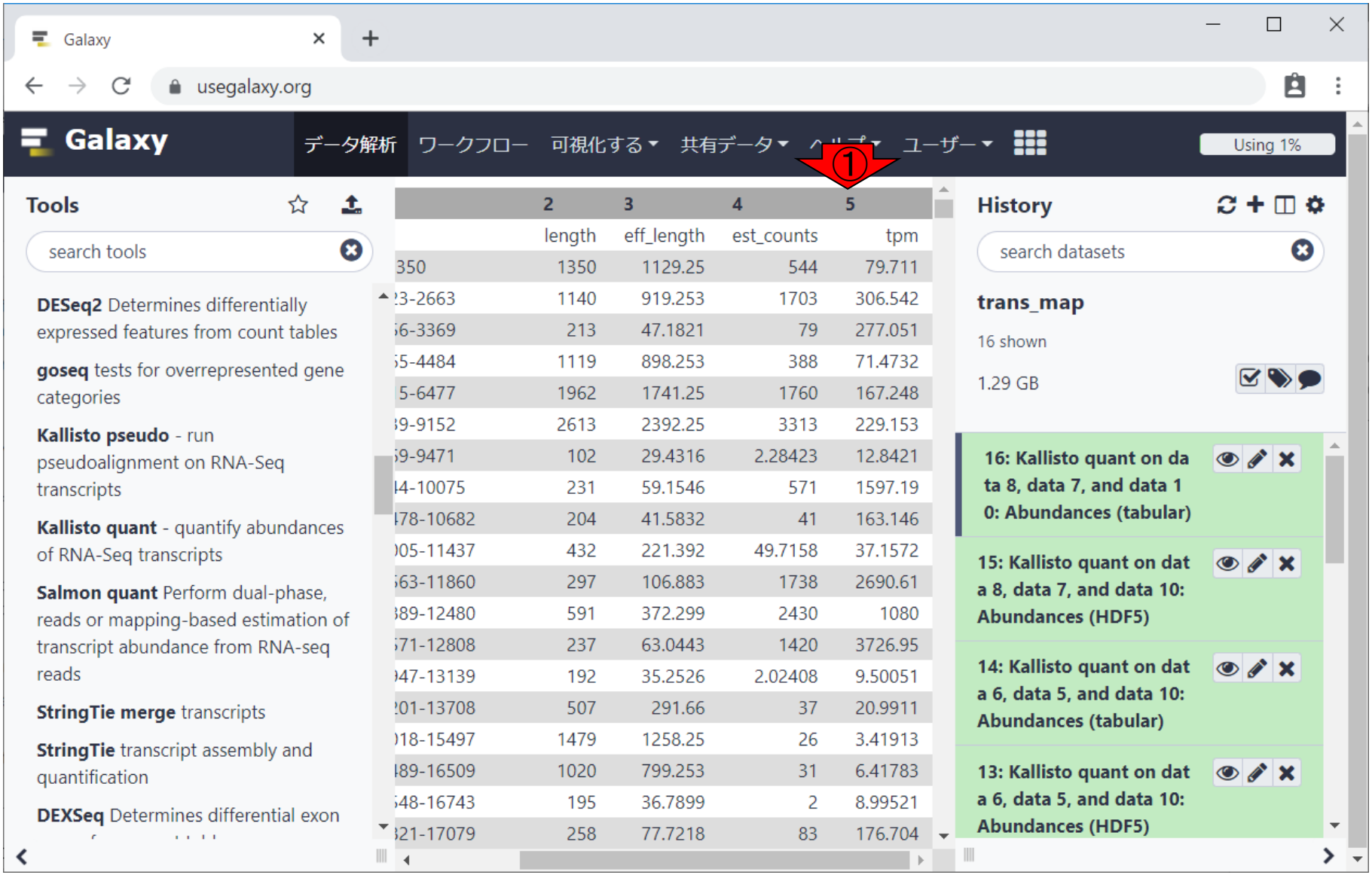


Contents

- 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。



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 RN

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

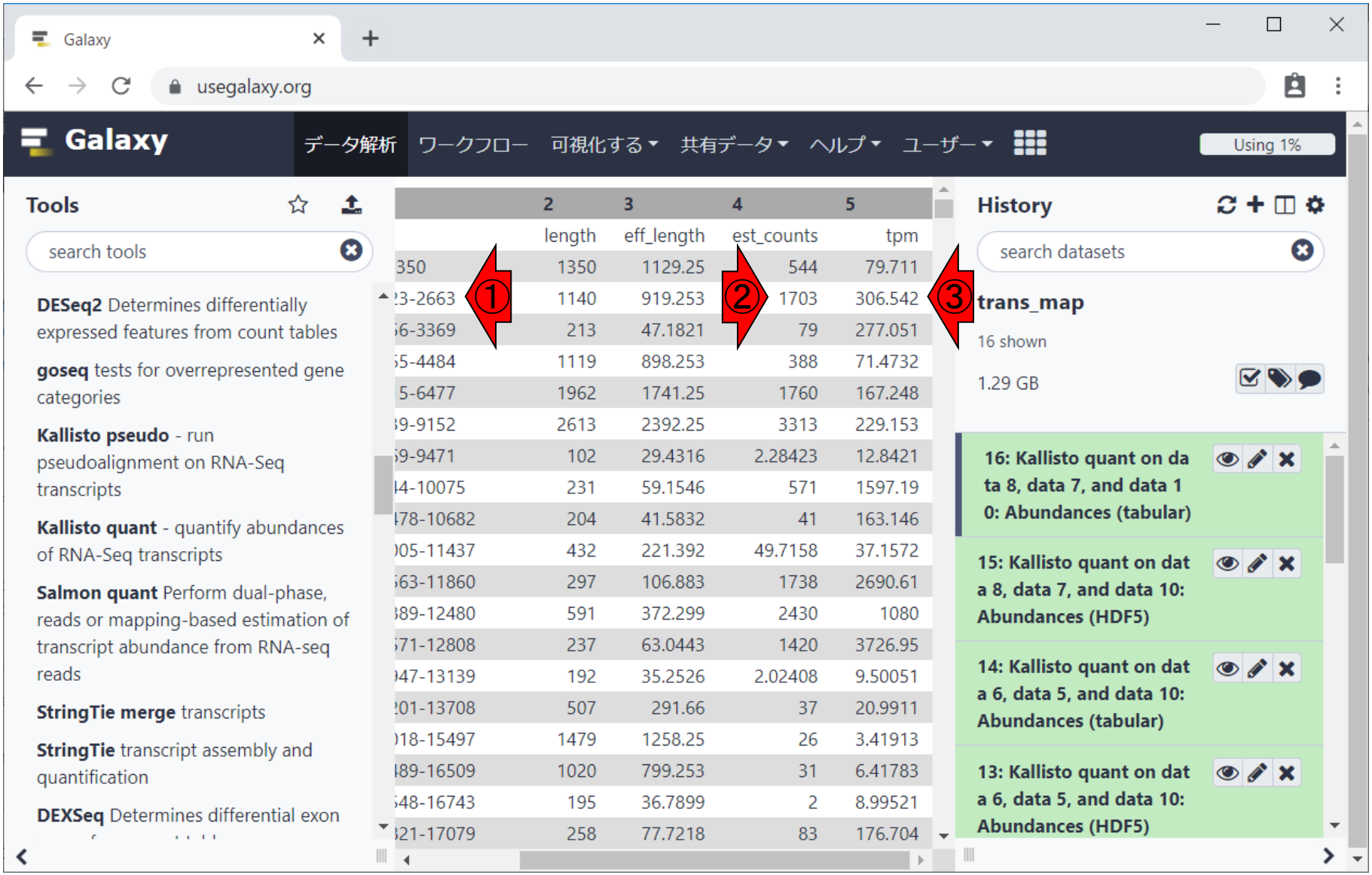
Since we are interested in taking the length into consideration, a natural measurement is the rate, counts per base ($X_i/\widetilde{l}_i$). As you might immediately notice, this number is also dependent on the total number of fragments sequenced. To adjust for this, simply divide by the sum of all rates 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:

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

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.

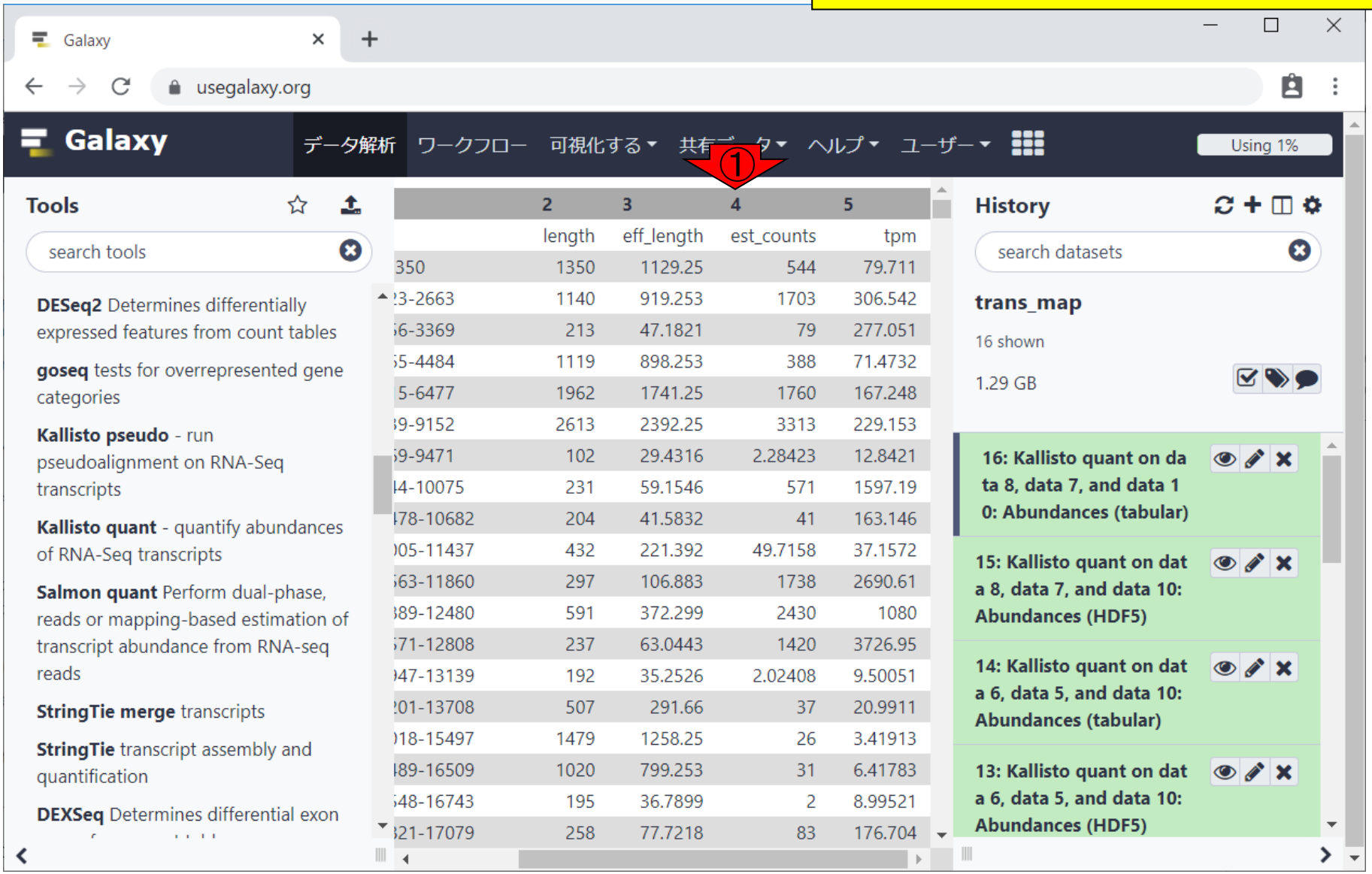
W11-3：2行目の転写物

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



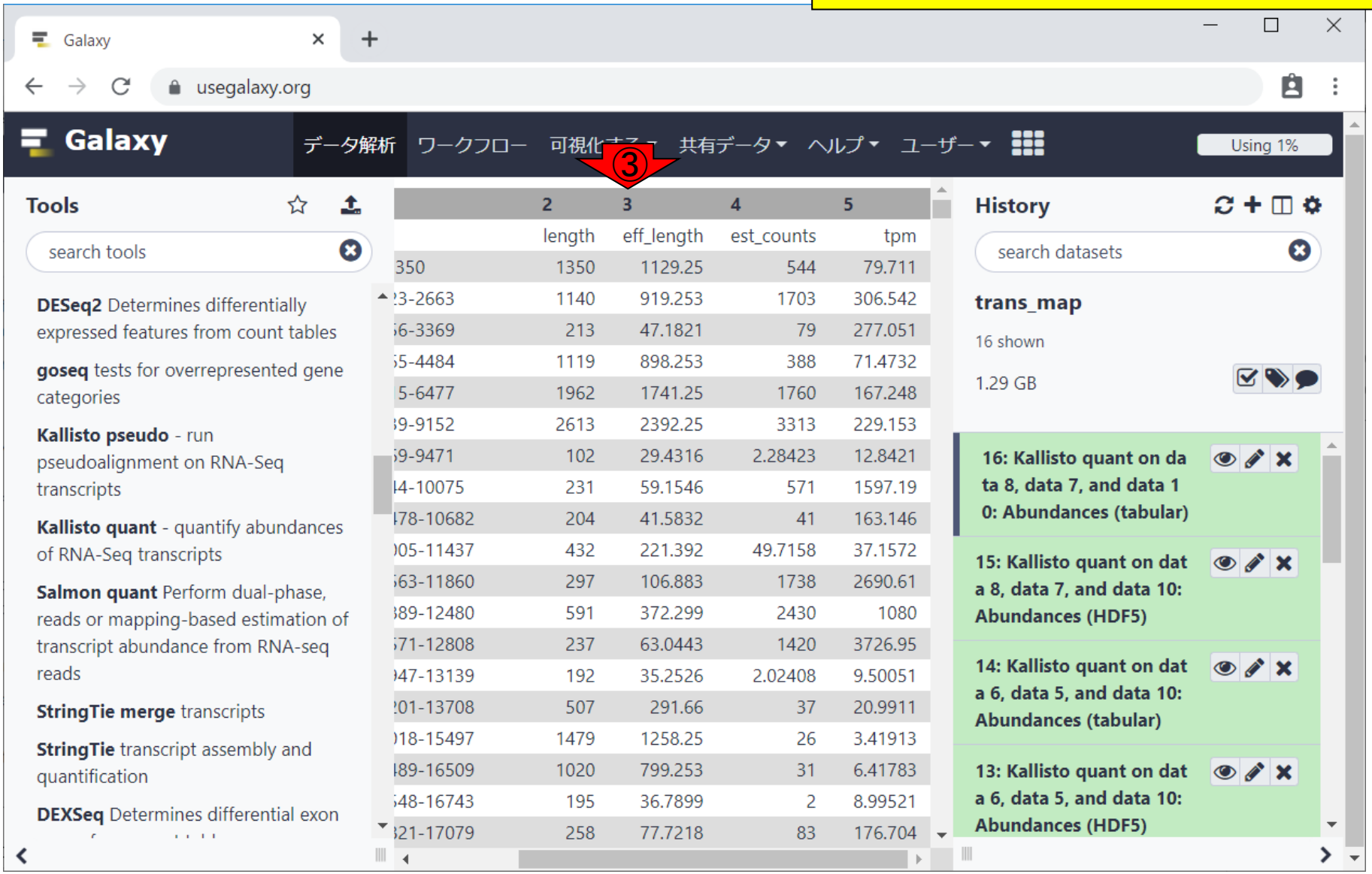
W11-4：4列目

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ところがポイント。



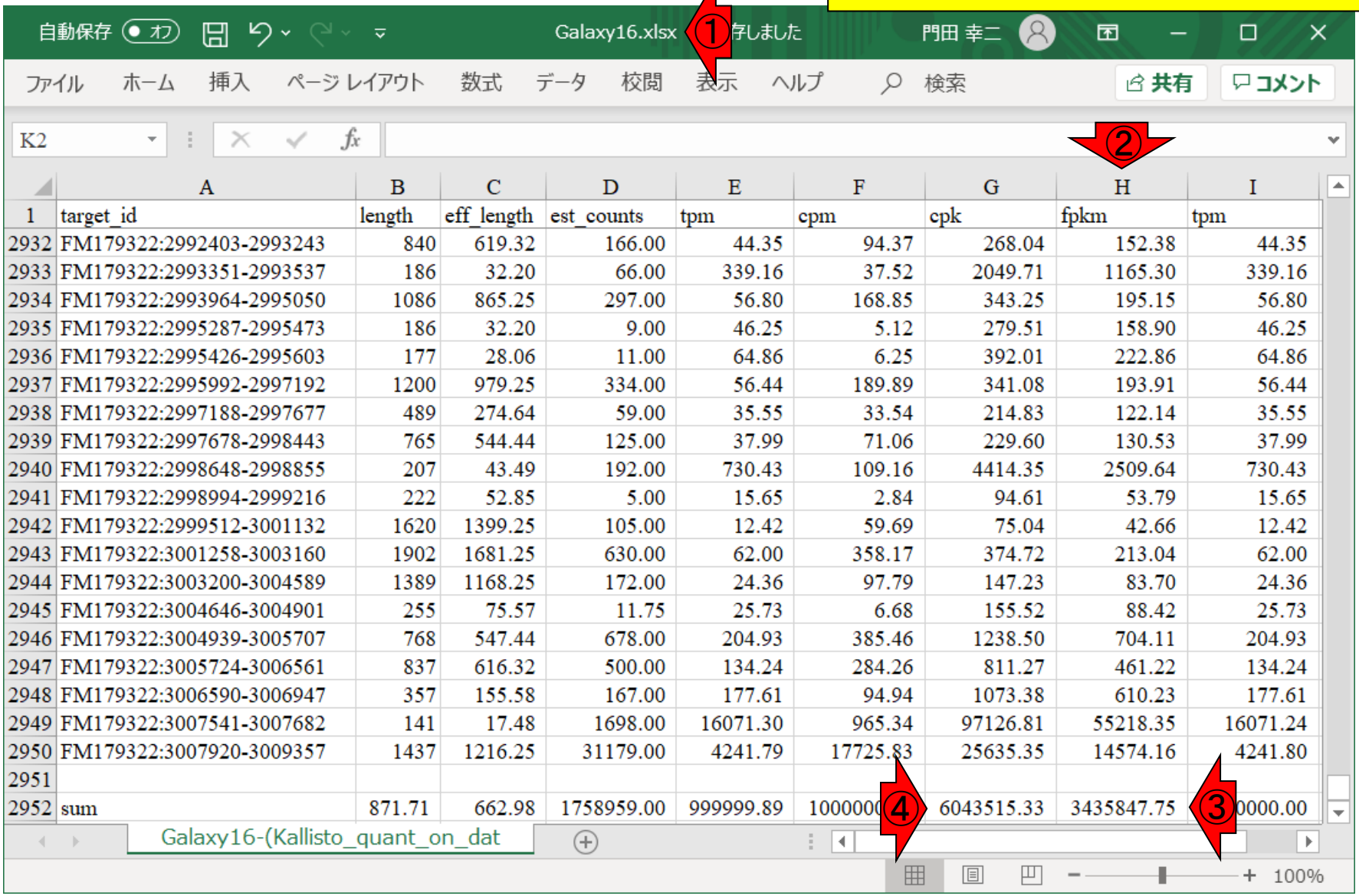
W11-5：3列目

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ところがポイント。



W11-6：ヒストリー16

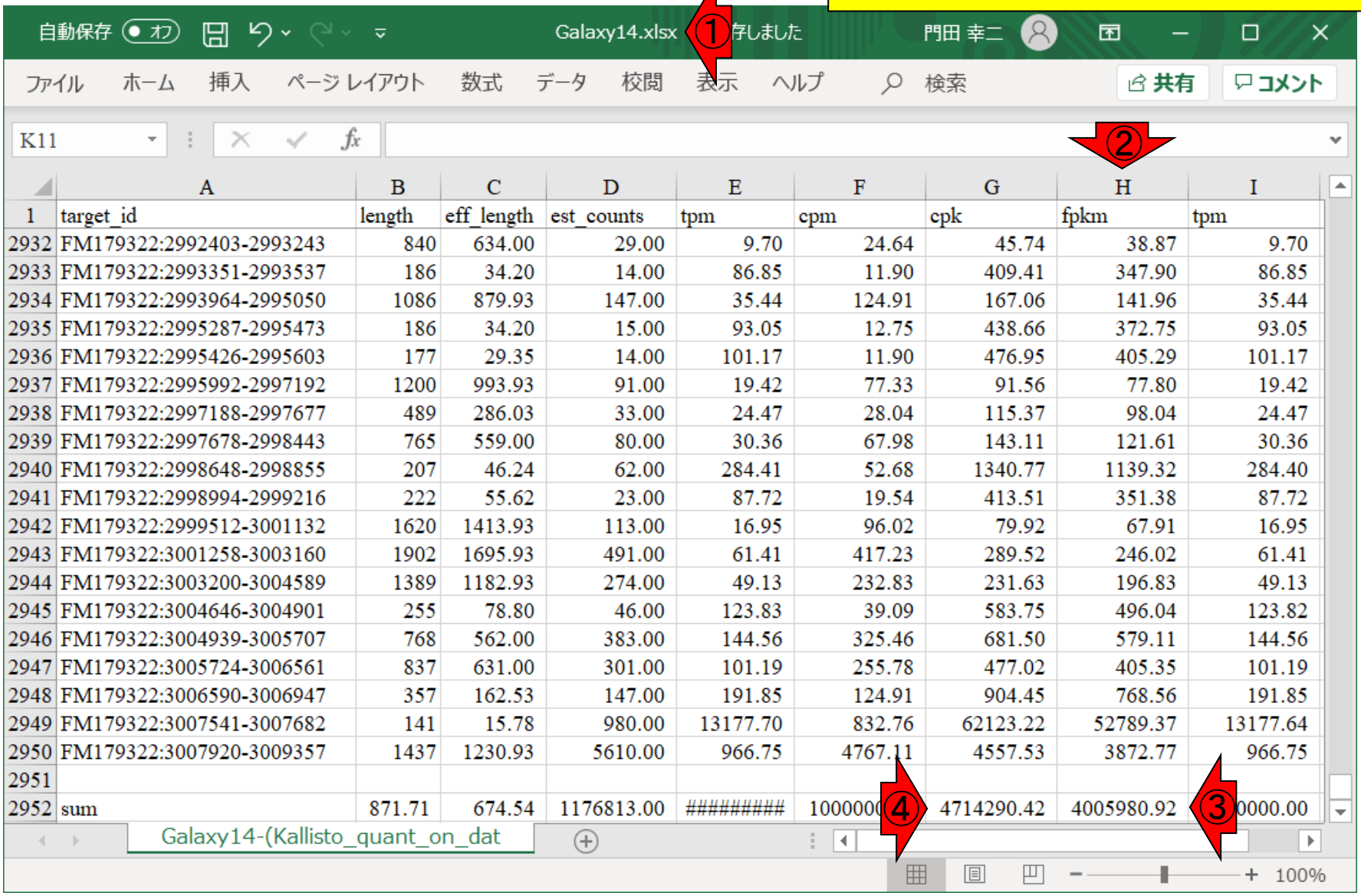
①ヒストリー16の、②FPKM値の、③総和は3,435,847.75。④CPK値の総和は6,043,515.33。



	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	6043515.33	3435847.75	0000.00

W11-7：ヒストリー14

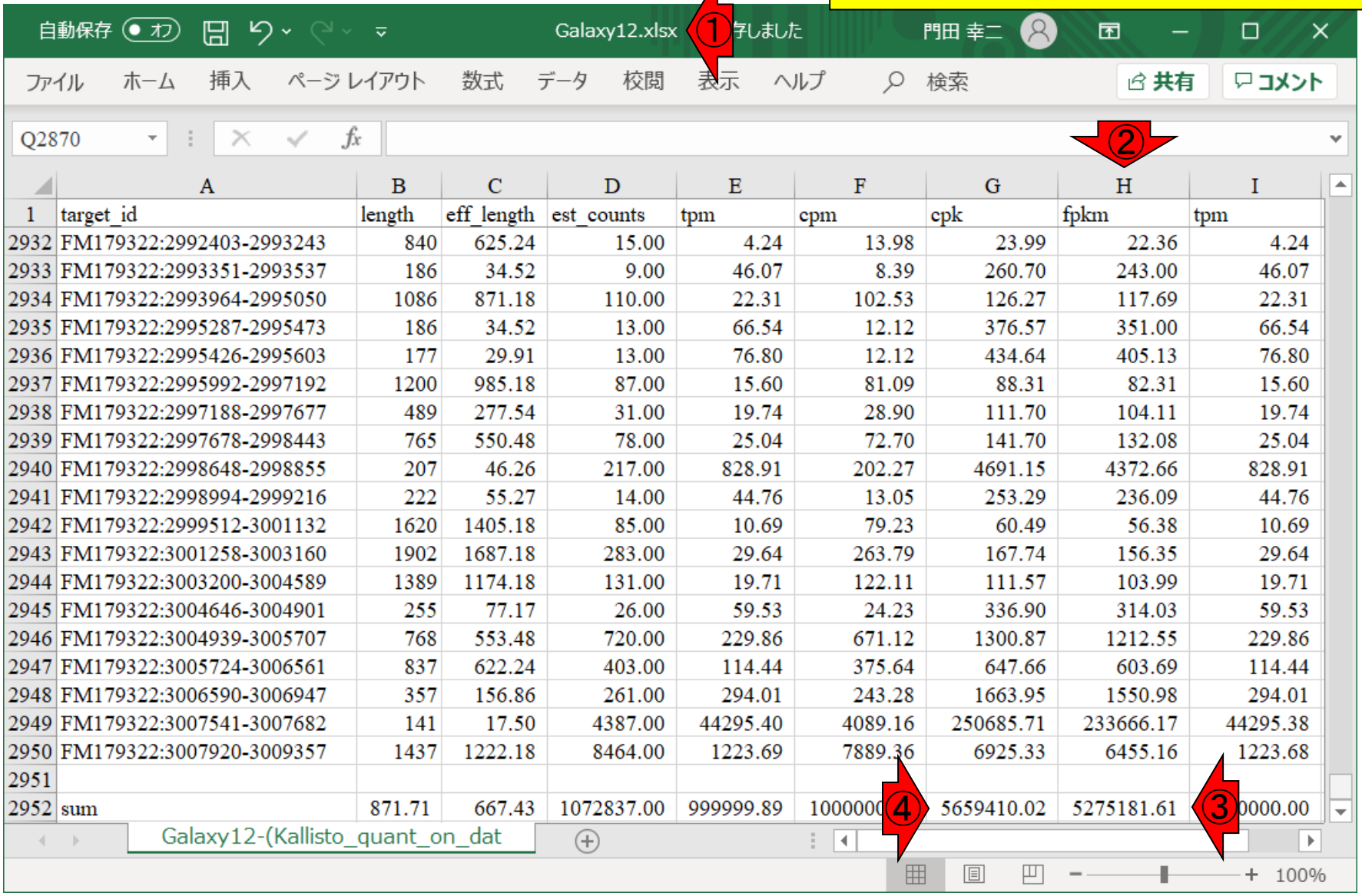
①ヒストリー14の、②FPKM値の、③総和は4,005,980.92。④CPK値の総和は4,714,290.42。



	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	0000.00

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。



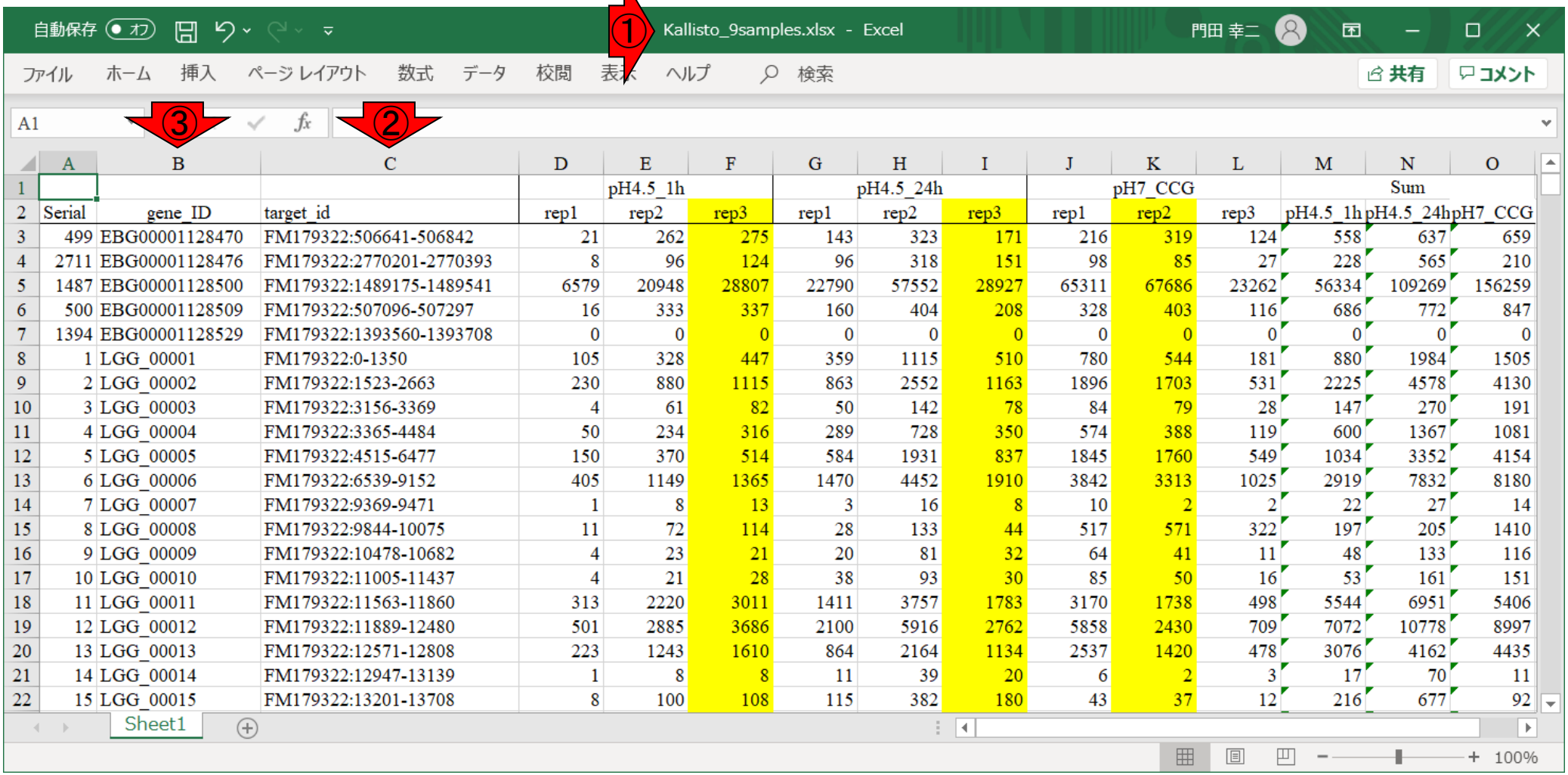
	A	B	C	D	E	F	G	H	I
1	target_id	length	eff_length	est_counts	tpm	cpm	cpk	fpkm	tpm
2932	FM179322:2992403-2993243	840	625.24	15.00	4.24	13.98	23.99	22.36	4.24
2933	FM179322:2993351-2993537	186	34.52	9.00	46.07	8.39	260.70	243.00	46.07
2934	FM179322:2993964-2995050	1086	871.18	110.00	22.31	102.53	126.27	117.69	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	5659410.02	5275181.61	0000.00

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

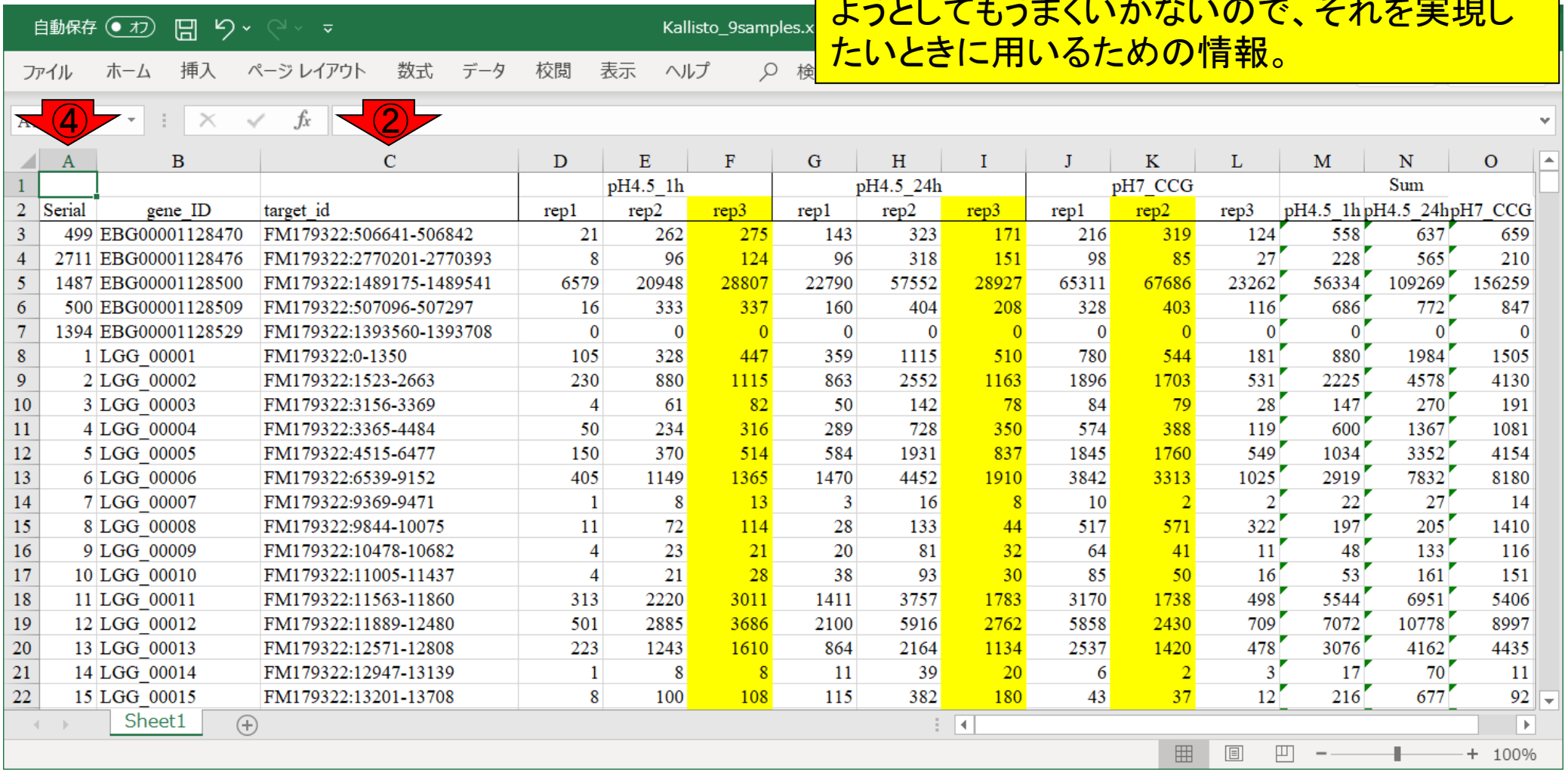
①Kallisto_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene_IDと対応づけている。



	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

W12-1：9サンプル

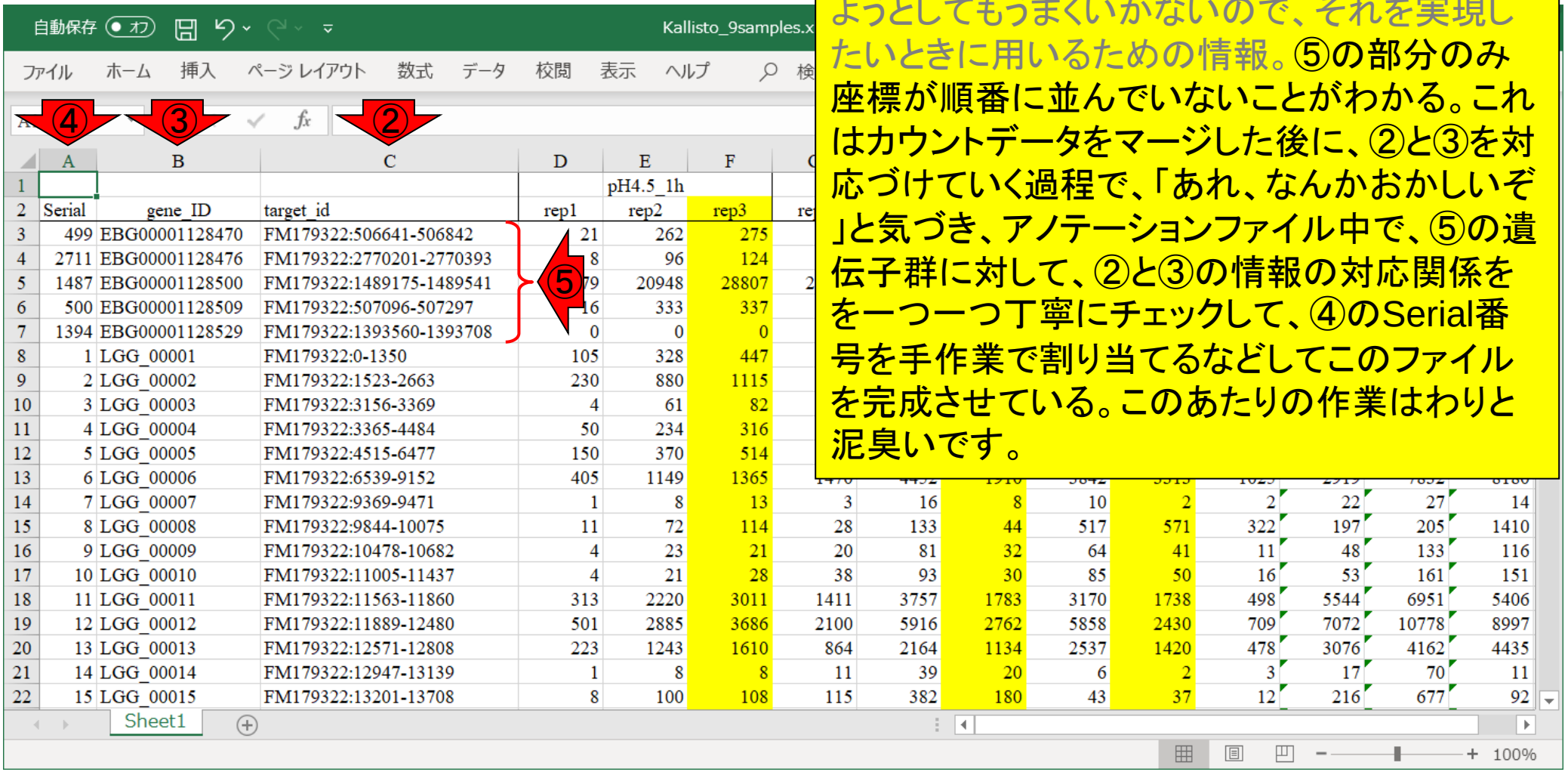
①Kallisto_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene_IDと対応づけている。④の列は、②の領域で座標順に並べようとしてもうまくいかないので、それを実現したいときに用いるための情報。



	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

W12-1：9サンプル

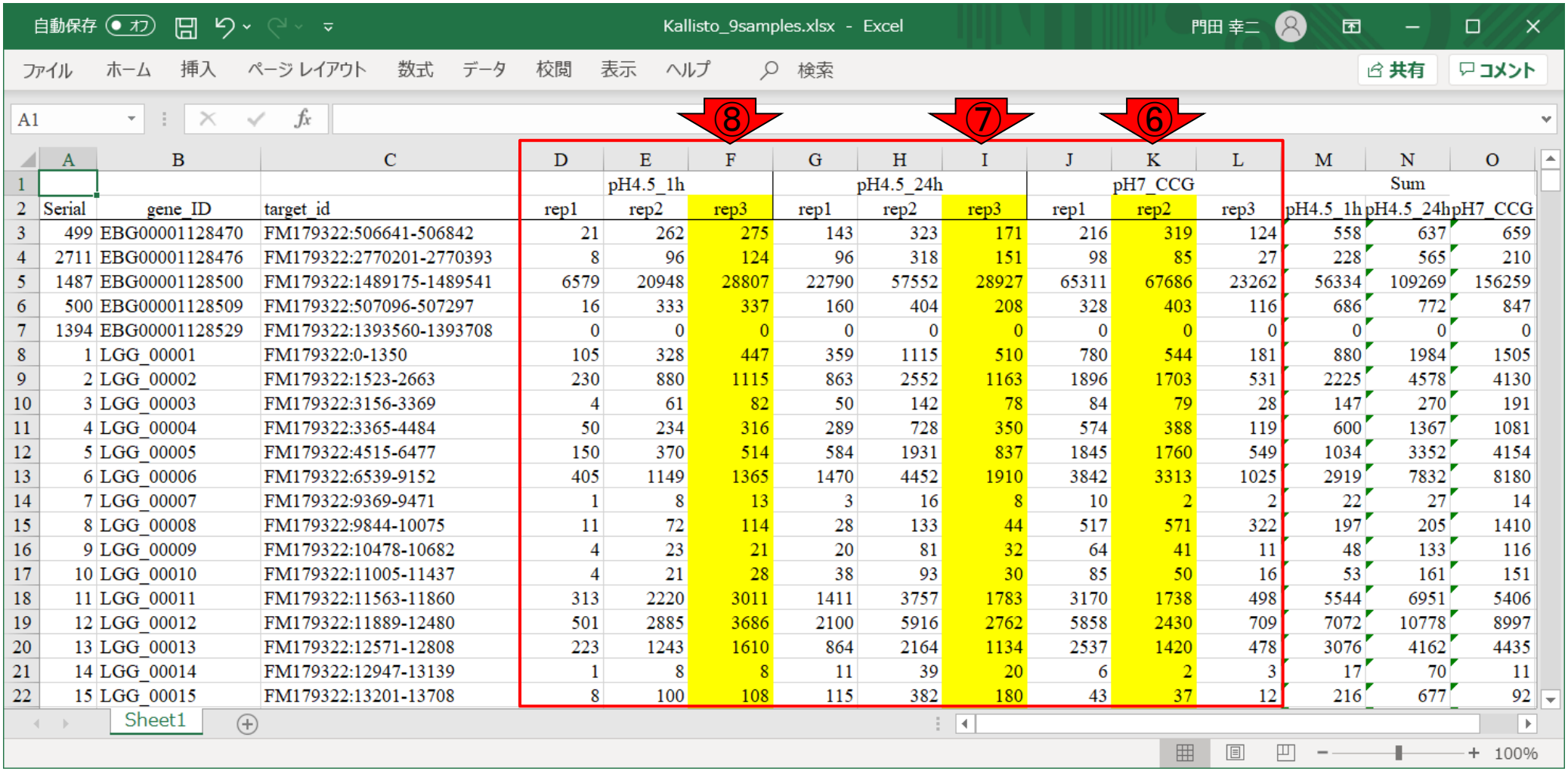
①Kallisto_9samples.xlsx。②今回用いたリファレンストランスクリプトーム配列中のdescription情報が転写物領域の座標情報のみなので、試行錯誤して③gene_IDと対応づけている。④の列は、②の領域で座標順に並べようとしてもうまくいかないので、それを実現したいときに用いるための情報。⑤の部分のみ座標が順番に並んでいないことがわかる。これはカウントデータをマージした後に、②と③を対応づけていく過程で、「あれ、なんかおかしいぞ」と気づき、アノテーションファイル中で、⑤の遺伝子群に対して、②と③の情報の対応関係をを一つ一つ丁寧にチェックして、④のSerial番号を手作業で割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。



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1					pH4.5_1h									
2	Serial	gene_ID	target_id	rep1	rep2	rep3	re							
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	2							
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	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

W12-1：9サンプル

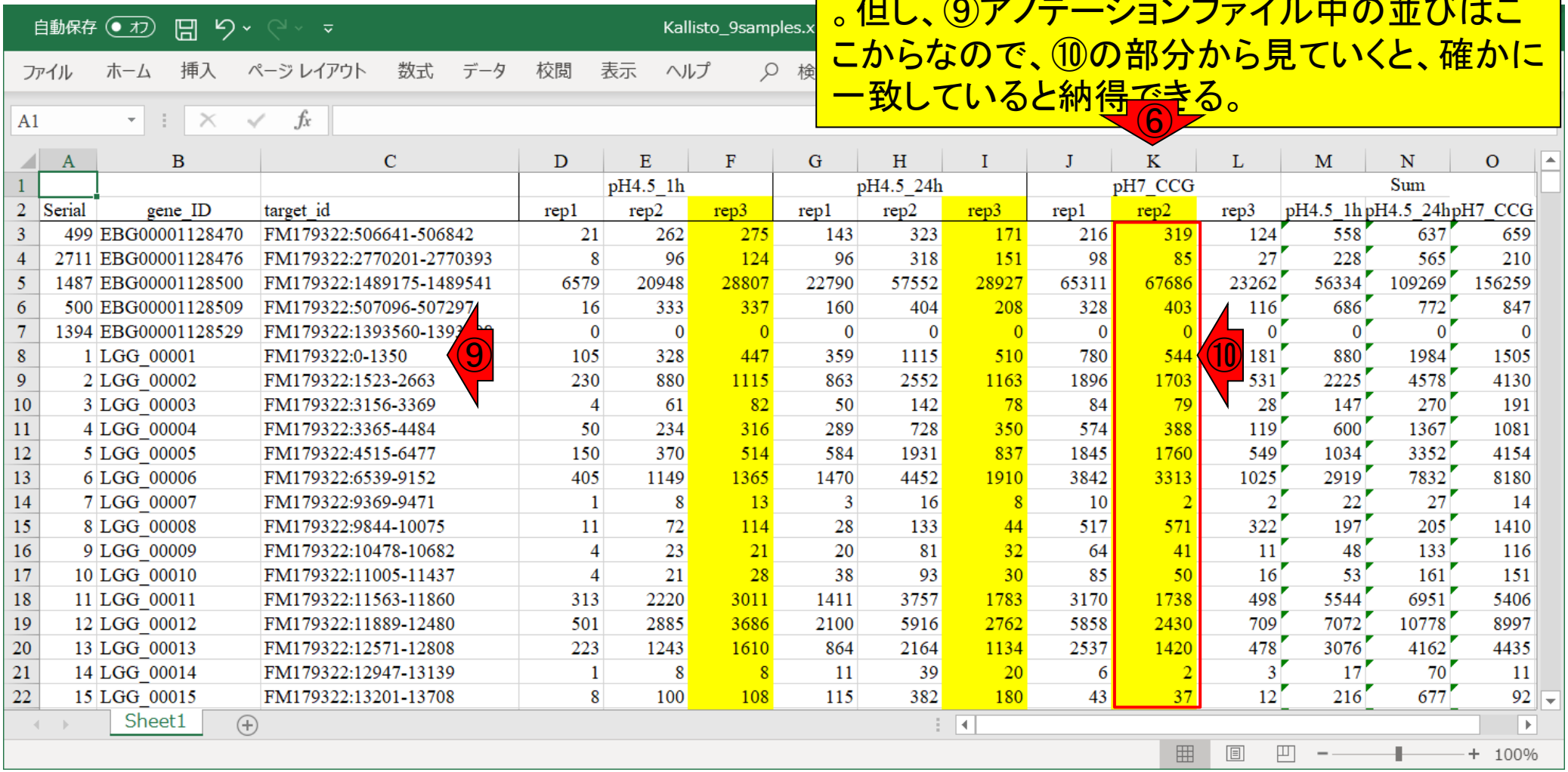
赤枠内が9サンプル分のカウント情報をマージした結果。黄色部分が第15回のGalaxy上で実際に行ったものに対応する。⑥ヒストリー16、⑦ヒストリー14、⑧ヒストリー12。



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

W12-1：9サンプル

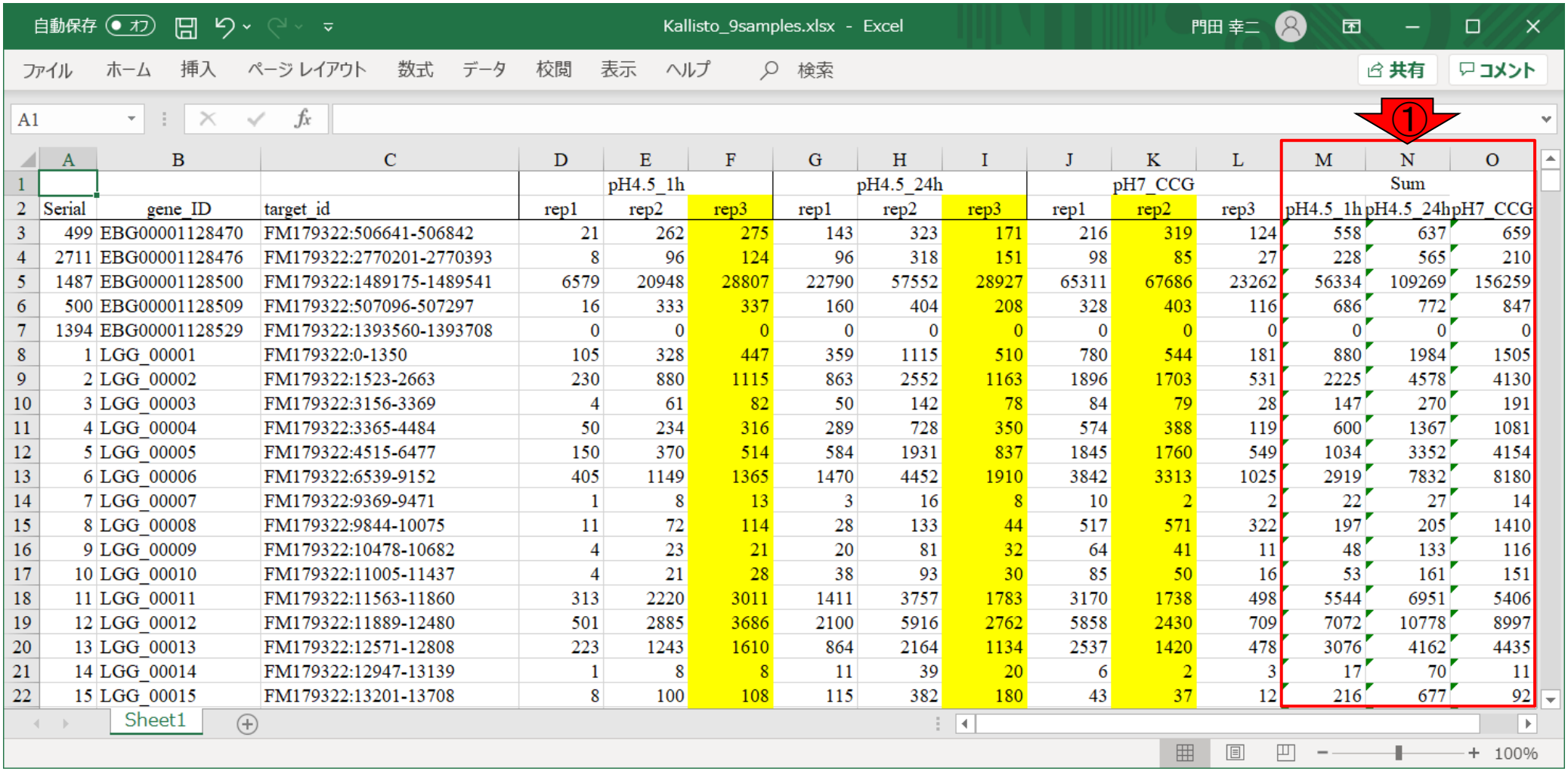
赤枠内が9サンプル分のカウント情報をマージした結果。黄色部分が第15回のGalaxy上で実際に行ったものに対応する。⑥ヒストリー16、⑦ヒストリー14、⑧ヒストリー12。例えば⑥のヒストリー16のカウント情報はW10-5でも見られる。但し、⑨アノテーションファイル中の並びはここからなので、⑩の部分から見ていくと、確かに一致していると納得できる。



				pH4.5_1h			pH4.5_24h			pH7_CCG			Sum		
Serial	gene_ID	target_id	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3	pH4.5_1h	pH4.5_24h	pH7_CCG
499	EBG00001128470	FM179322:506641-506842	21	262	275	143	323	171	216	319	124	558	637	659
2711	EBG00001128476	FM179322:2770201-2770393	8	96	124	96	318	151	98	85	27	228	565	210
1487	EBG00001128500	FM179322:1489175-1489541	6579	20948	28807	22790	57552	28927	65311	67686	23262	56334	109269	156259
500	EBG00001128509	FM179322:507096-507297	16	333	337	160	404	208	328	403	116	686	772	847
1394	EBG00001128529	FM179322:1393560-139	0	0	0	0	0	0	0	0	0	0	0	0
1	LGG_00001	FM179322:0-1350	105	328	447	359	1115	510	780	544	181	880	1984	1505
2	LGG_00002	FM179322:1523-2663	230	880	1115	863	2552	1163	1896	1703	531	2225	4578	4130
3	LGG_00003	FM179322:3156-3369	4	61	82	50	142	78	84	79	28	147	270	191
4	LGG_00004	FM179322:3365-4484	50	234	316	289	728	350	574	388	119	600	1367	1081
5	LGG_00005	FM179322:4515-6477	150	370	514	584	1931	837	1845	1760	549	1034	3352	4154
6	LGG_00006	FM179322:6539-9152	405	1149	1365	1470	4452	1910	3842	3313	1025	2919	7832	8180
7	LGG_00007	FM179322:9369-9471	1	8	13	3	16	8	10	2	2	22	27	14
8	LGG_00008	FM179322:9844-10075	11	72	114	28	133	44	517	571	322	197	205	1410
9	LGG_00009	FM179322:10478-10682	4	23	21	20	81	32	64	41	11	48	133	116
10	LGG_00010	FM179322:11005-11437	4	21	28	38	93	30	85	50	16	53	161	151
11	LGG_00011	FM179322:11563-11860	313	2220	3011	1411	3757	1783	3170	1738	498	5544	6951	5406
12	LGG_00012	FM179322:11889-12480	501	2885	3686	2100	5916	2762	5858	2430	709	7072	10778	8997
13	LGG_00013	FM179322:12571-12808	223	1243	1610	864	2164	1134	2537	1420	478	3076	4162	4435
14	LGG_00014	FM179322:12947-13139	1	8	8	11	39	20	6	2	3	17	70	11
15	LGG_00015	FM179322:13201-13708	8	100	108	115	382	180	43	37	12	216	677	92

①が反復データをマージした結果。

W12-2：Table1の前説



自動保存 オフ　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

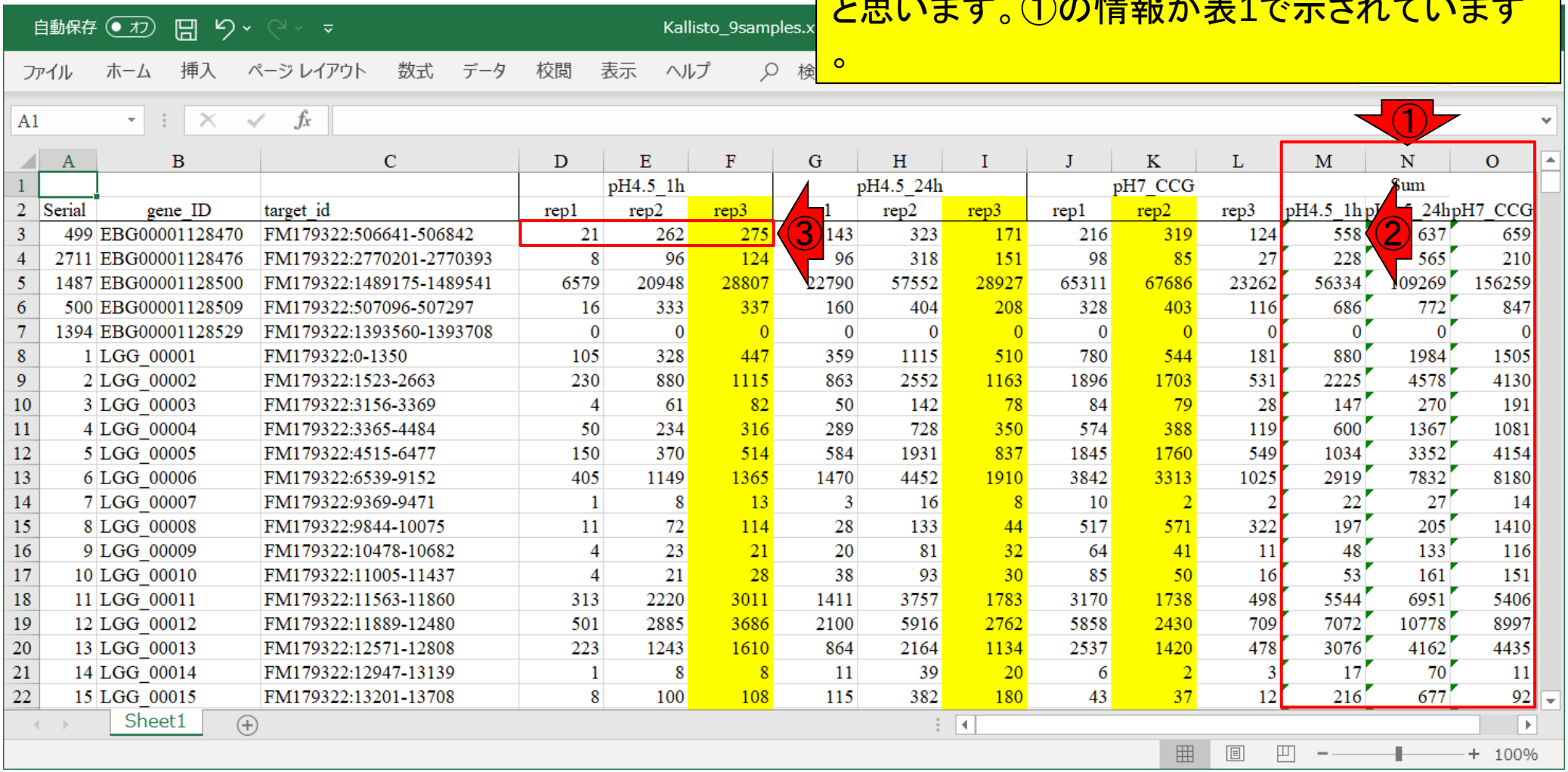
①

	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%

W12-2：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。



	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