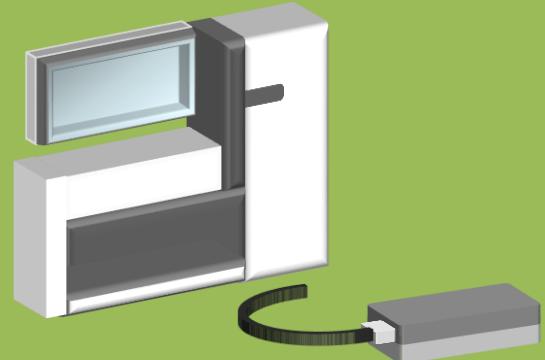
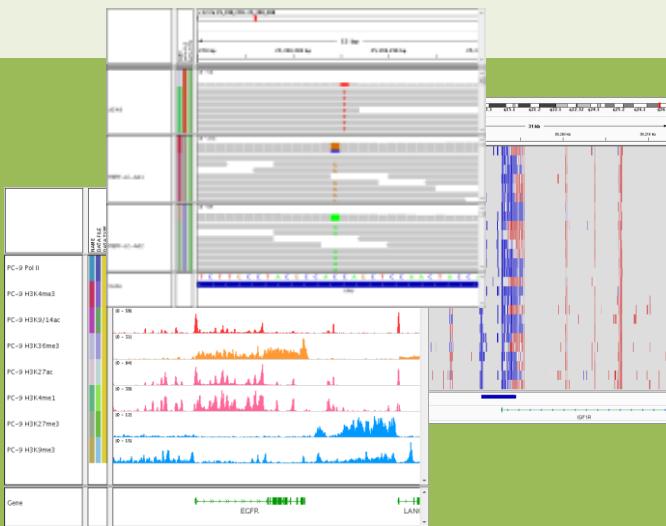


2016.09.12 AJACS東女医大



次世代シーケンスデータを用いたオミクス解析



国立がん研究センター 先端医療開発センター
ゲノムトランスレーショナルリサーチ分野
鈴木 純子

本日の予定

Session 1

シークエンスデータのマッピングと可視化
およびツールの紹介

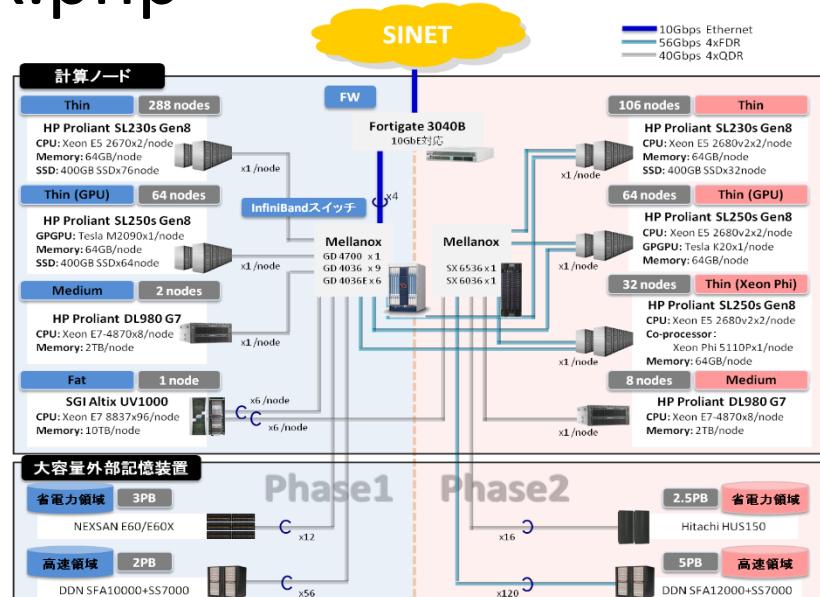
Session 2

データベース・新規技術の紹介

国立遺伝研究所 スーパーコンピュータシステム

<https://sc.ddbj.nig.ac.jp/index.php>

The screenshot shows the main page of the NIG Supercomputer Facilities website. It includes a header with the NIG logo and site navigation. Below the header, there's a "重要なお知らせ" (Important Information) section with a list of news items. A "システム構成" (System Configuration) section displays server racks and network components. A "システム使用方法" (System Usage Methods) section provides detailed instructions for various services likelustre and SGE. A "各種申請" (Various Applications) section lists various types of user requests. At the bottom, there's a "Webサービス" (Web Services) section for MGAPI and DDJ Pipeline applications.



<https://sc.ddbj.nig.ac.jp/index.php/systemconfig>

本講習会のために、アカウントをご用意いただきました。

スパコンのアカウント

ユーザ名: lect01 ~ lect55
パスワード: * * * *

一人一人にアカウントを割り当てられています。
ご確認ください。

パスワードは変更しないでください。

作業は、自分のホームディレクトリ以下で行ってください。

本日用いるソフトウェアの準備

ターミナルソフト TeraTerm

- Windows PCからLinux環境のマシンへアクセスするのに必要
※Macはターミナルでリモートアクセスできます
※その他ターミナルソフトをお持ちの方はお好きなものをお使いいただいてかまいません
- インストールは基本的にデフォルトの設定で問題ありません

<http://osdn.jp/projects/ttssh2/releases/>

ファイル転送ソフト WinSCP

Windows PC←→Linux環境のデータのやり取りに必要

※Macはターミナルでデータのアップロード/ダウンロードできます

※Cygwinインストール済みの方は不要です

<https://winscp.net/eng/download.php>

ゲノムビューアー IGV (Integrative Genomics Viewer)

リファレンスゲノム上にアライメントされたシークエンスタグを可視化するツール

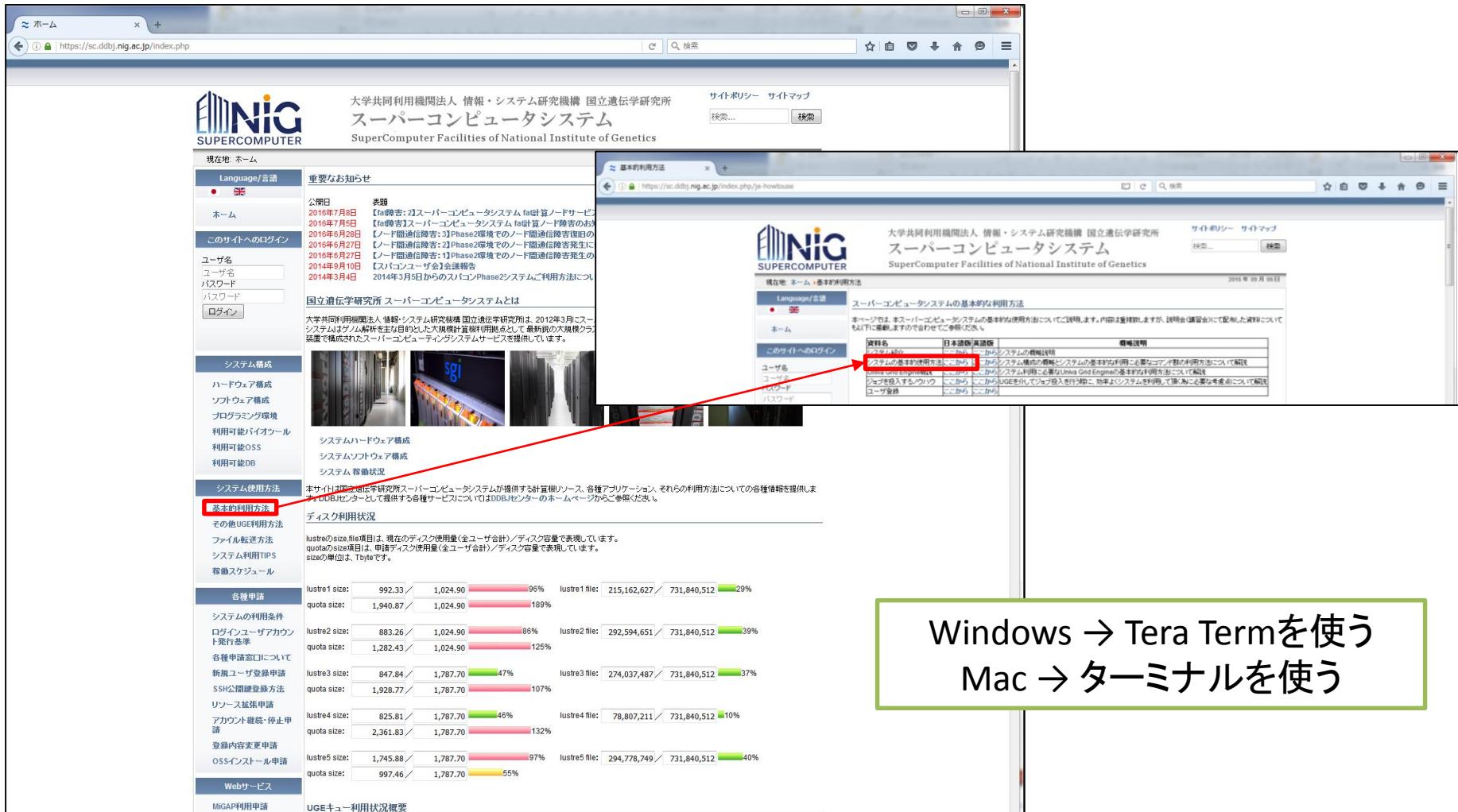
<http://software.broadinstitute.org/software/igv>

起動するか確認してください。

起動しない方はお知らせください。

PCからスパコンへログイン

<https://sc.ddbj.nig.ac.jp/index.php>

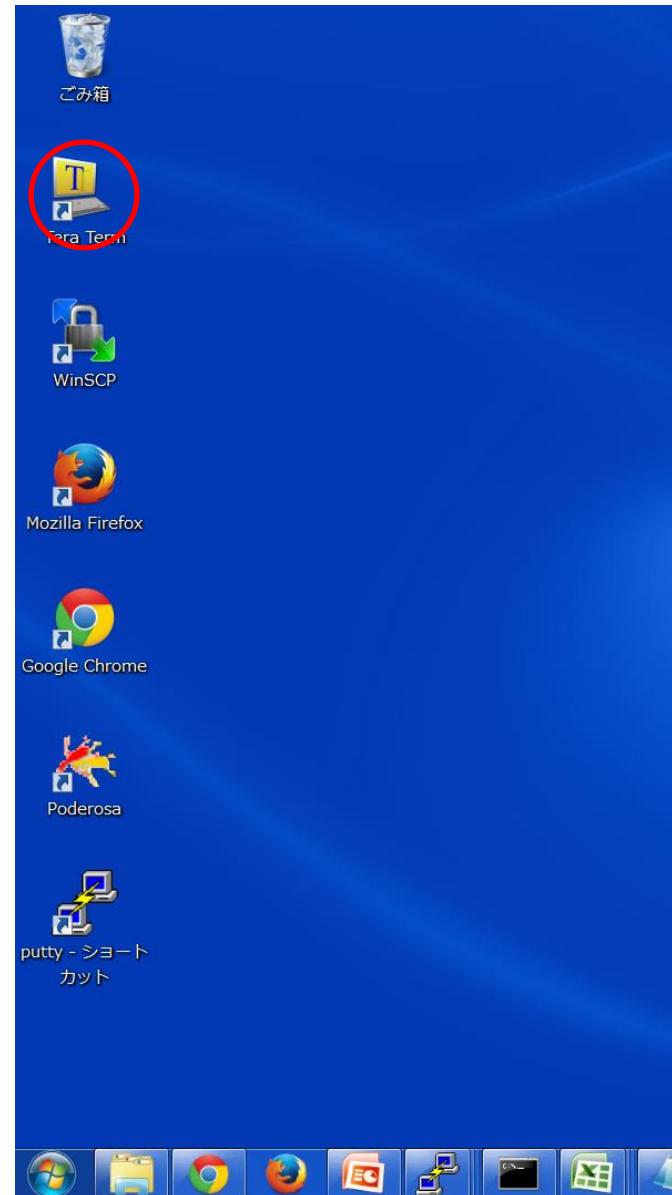


Windows → Tera Termを使う
Mac → ターミナルを使う

Mac PC: 「アプリケーション」→「ユーティリティ」→「ターミナル」

Windows PCから解析サーバへのログイン①

- TeraTermを開きます。
- 接続先の設定画面にて、ホストのボックスにIPアドレスgw.ddbj.nig.ac.jpを入力し、OKをクリックします。



Windows PCから解析サーバへのログイン②

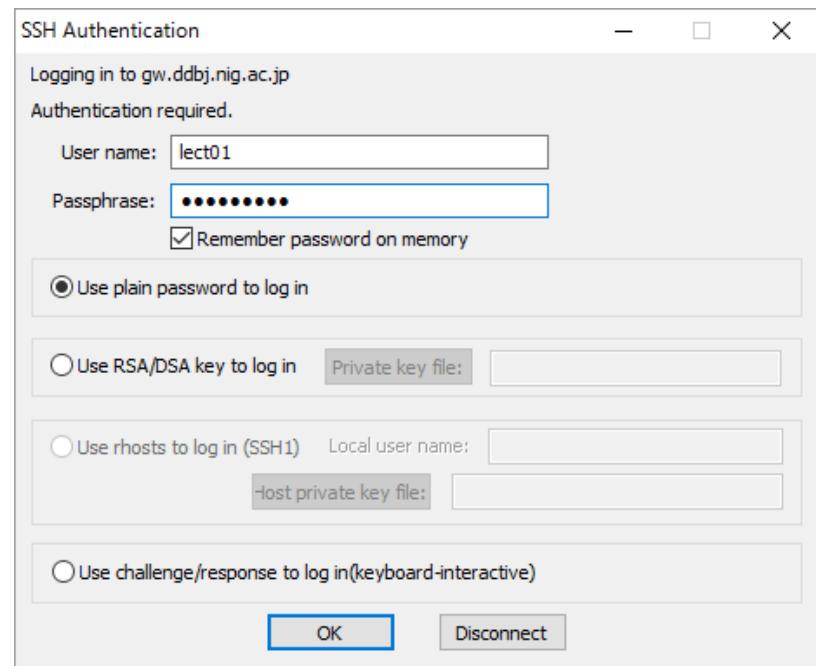
- ユーザ名とパスワードを入力します。

ユーザ名: lect01 ~ lect55
パスワード: *****

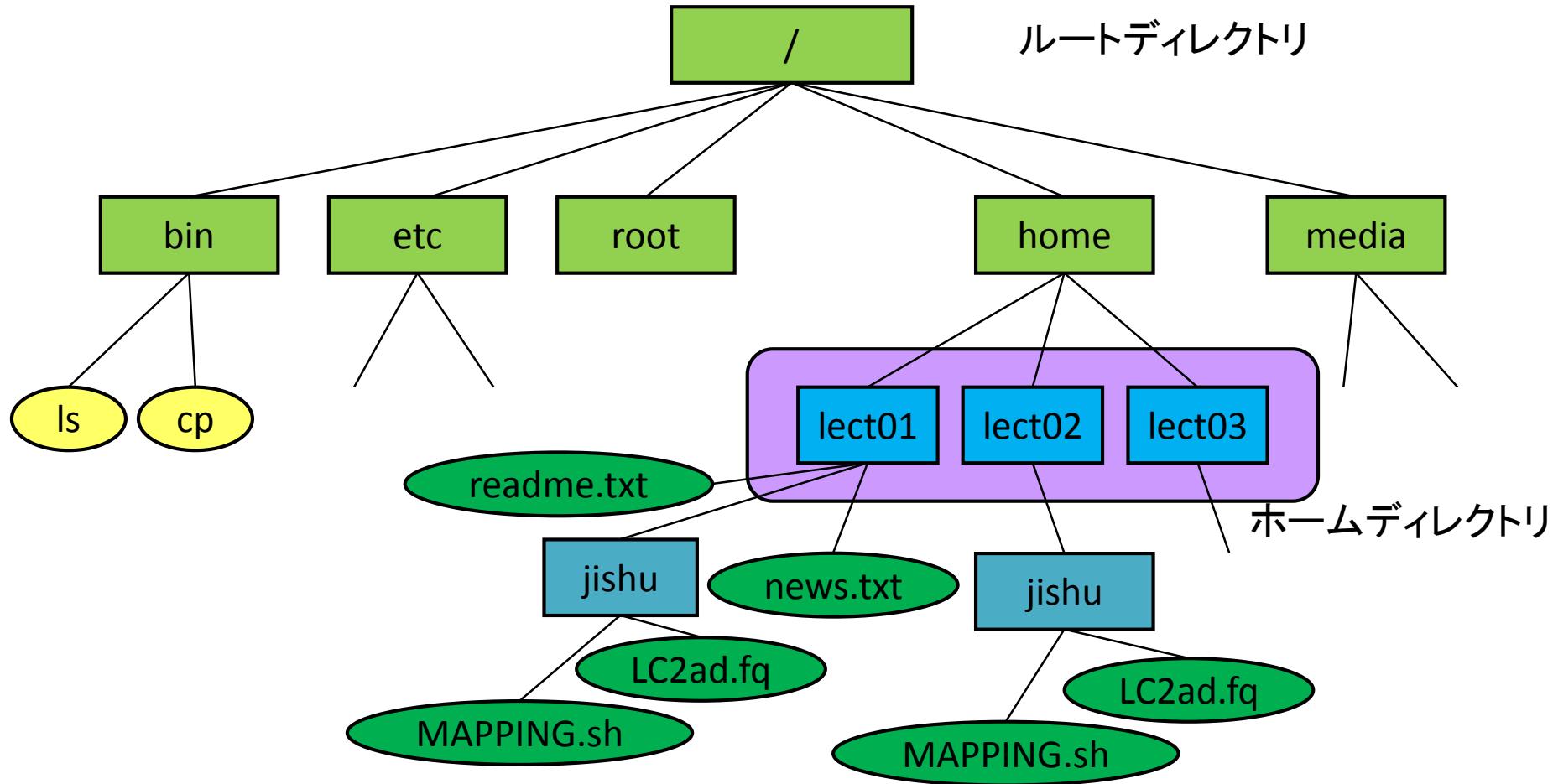
自分のユーザ名(lect01 ~ lect55)を入れてください。

- OKを押します
- 正しくログインできると、
[lect01@gw ~]\$
と表示されます。

- Macの方もWindowsの方も
次にゲートウェイノードから、ログインノードにログインします。
[lect01@gw ~]\$ qlogin
Enterキーを押すとパスワードを要求されるのでパスワードを打ってログインします。



ディレクトリ構造



- ルートディレクトリ: ツリー構造のトップとなるディレクトリ
- ホームディレクトリ: 各ユーザがログインしたときの最初のディレクトリ(/home/[user ID])
- カレントディレクトリ: 現在、作業しているディレクトリ

ディレクトリ関係の基本コマンド

- cd: 指定のディレクトリに移動する

- ディレクトリAに移動する場合

```
$ cd A
```

cdのみ入力した場合はホームディレクトリへ移動する。

cd ..と入力した場合は、一つ上のディレクトリに移動する。

- pwd: 現在のディレクトリを表示する

```
$ pwd
```

- ls: 指定のディレクトリ内のディレクトリ、ファイルを表示する

- ディレクトリAのファイルを表示する。

```
$ ls A
```

lsのみを入力した場合は、カレントディレクトリのファイルが表示される。

• -lオプションをつけると詳細表示になり、-tオプションをつけると更新日時でソートされる。

【実習】実際にコマンドを入力しましょう

```
$ pwd
```

```
$ ls
```

```
$ cp -r /home/lect01/jishu .
```

```
$ ls
```

```
$ cd jishu
```

```
$ pwd
```

```
$ ls
```

```
$ ls -lt
```

```
$ cd
```

```
$ pwd
```

```
gw.ddbj.nig.ac.jp - Tera Term VT
File Edit Setup Control Window Help
[lect02@t263 ~]$ pwd
/home/lect02
[lect02@t263 ~]$ ls
[lect02@t263 ~]$ cp -r /home/lect01/jishu .
[lect02@t263 ~]$ ls
jishu
[lect02@t263 ~]$ cd jishu
[lect02@t263 jishu]$ pwd
/home/lect02/jishu
[lect02@t263 jishu]$ ls
LC2ad fq MAPPING.sh MAPPING_pair.sh PC-9_1.fq PC-9_2.fq
[lect02@t263 jishu]$ ls -lt
合計 468004
-rw-r--r-- 1 lect02 lect 178036671 9月 7 13:38 2016 PC-9_2.fq
-rwxr-xr-x 1 lect02 lect 452 9月 7 13:38 2016 MAPPING_pair.sh
-rw-r--r-- 1 lect02 lect 178036671 9月 7 13:38 2016 PC-9_1.fq
-rw-r--r-- 1 lect02 lect 130238229 9月 7 13:38 2016 LC2ad fq
-rwxr-xr-x 1 lect02 lect 367 9月 7 13:38 2016 MAPPING.sh
[lect02@t263 jishu]$ cd
[lect02@t263 ~]$ pwd
/home/lect02
[lect02@t263 ~]$
```

ファイル内容を表示するコマンド

- cat: ファイルの全内容を表示する

```
$ cat a.txt
```

一度に全部表示されるため、大きいサイズのファイルには不向き

- head, tail: ファイルの先頭、末端を表示(デフォルトでは10行)

```
$ head a.txt
```

```
$ tail a.txt
```

100行表示したい場合は-nオプションを付ける

```
$ head -n 100 a.txt
```

- more, less: ファイルの内容をコマ送りで表示する

```
$ more a.txt
```

```
$ less a.txt
```

Enterもしくはspaceキーで進む

lessコマンドは↑↓キーでファイルを自由に見ることができる

qと打つと、more, lessコマンドを途中で中断できる

【実習】実際にコマンドを入力しましょう

```
$ cat /home/lect01/readme.txt
```

```
$ cat /home/lect01/news.txt
```

```
$ head /home/lect01/news.txt
```

```
$ tail /home/lect01/news.txt
```

```
$ more /home/lect01/news.txt
```

```
$ less /home/lect01/news.txt
```

ファイル、ディレクトリの作成、移動、削除コマンド

- cp: ファイルの複製
 - ファイルa.txtをカレントディレクトリにコピーする場合
\$ cp a.txt .
 - ファイルa.txtをディレクトリAの中にコピーする場合
\$ cp a.txt A
- mkdir: 新規にディレクトリを作成する
 - ディレクトリAを作成する
\$ mkdir A
- mv: ファイルの移動
 - ファイルa.txtをディレクトリAに移動する
\$ mv a.txt A
- rm: ファイルを削除する
 - ファイルa.txtを削除する
\$ rm a.txt
 - ディレクトリを削除するときは-rオプションを付ける
\$ rm -r A

※一度データを削除すると復元することはできません。

【実習】実際にコマンドを入力しましょう

```
$ cd  
$ cp /home/lect01/readme.txt .  
$ ls -lt
```

```
$ mkdir work  
$ ls -lt
```

```
$ mv readme.txt work  
$ ls -lt
```

```
$ cd work  
$ pwd  
$ ls -lt
```

```
$ rm readme.txt  
$ ls -lt
```

```
$ cd ..  
$ pwd  
$ ls -lt  
$ rm -r work  
$ ls -lt
```

File Edit Setup Control Window Help

```
[lect02@t263 ~]$ cd  
[lect02@t263 ~]$ cp /home/lect01/readme.txt .  
[lect02@t263 ~]$ ls -lt  
合計 4  
-rw-r--r-- 1 lect02 lect 8463 9月 7 13:42 2016 readme.txt  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:38 2016 jishu  
[lect02@t263 ~]$ mkdir work  
[lect02@t263 ~]$ ls -lt  
合計 20  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:42 2016 work  
-rw-r--r-- 1 lect02 lect 8463 9月 7 13:42 2016 readme.txt  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:38 2016 jishu  
[lect02@t263 ~]$ mv readme.txt work  
[lect02@t263 ~]$ ls -lt  
合計 8  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:43 2016 work  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:38 2016 jishu  
[lect02@t263 ~]$ cd work  
[lect02@t263 work]$ pwd  
/home/lect02/work  
[lect02@t263 work]$ ls -lt  
合計 12  
-rw-r--r-- 1 lect02 lect 8463 9月 7 13:42 2016 readme.txt  
[lect02@t263 work]$ rm readme.txt  
[lect02@t263 work]$ ls -lt  
合計 0  
[lect02@t263 work]$ cd ..  
[lect02@t263 ~]$ pwd  
/home/lect02  
[lect02@t263 ~]$ ls -lt  
合計 8  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:43 2016 work  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:38 2016 jishu  
[lect02@t263 ~]$ rm -r work  
[lect02@t263 ~]$ ls -lt  
合計 4  
drwxr-xr-x 2 lect02 lect 4096 9月 7 13:38 2016 jishu  
[lect02@t263 ~]$ █
```

ディレクトリ/home/lect-1にあるreadme.txtを
カレントディレクトリ(ホームディレクトリ)にコピー

ディレクトリworkを作成し、
readme.txtをディレクトリworkに移動

ディレクトリworkに移動する

ディレクトリworkに移動させた
readme.txtを削除

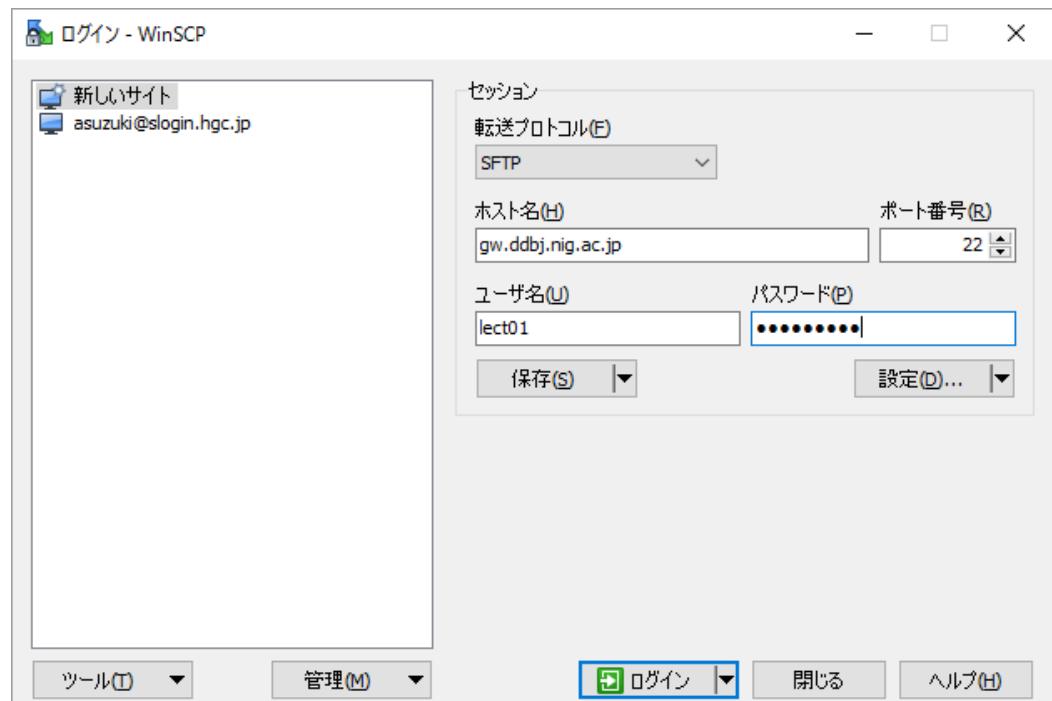
一つ上のディレクトリ
(ホームディレクトリ)に戻る

ディレクトリworkを削除

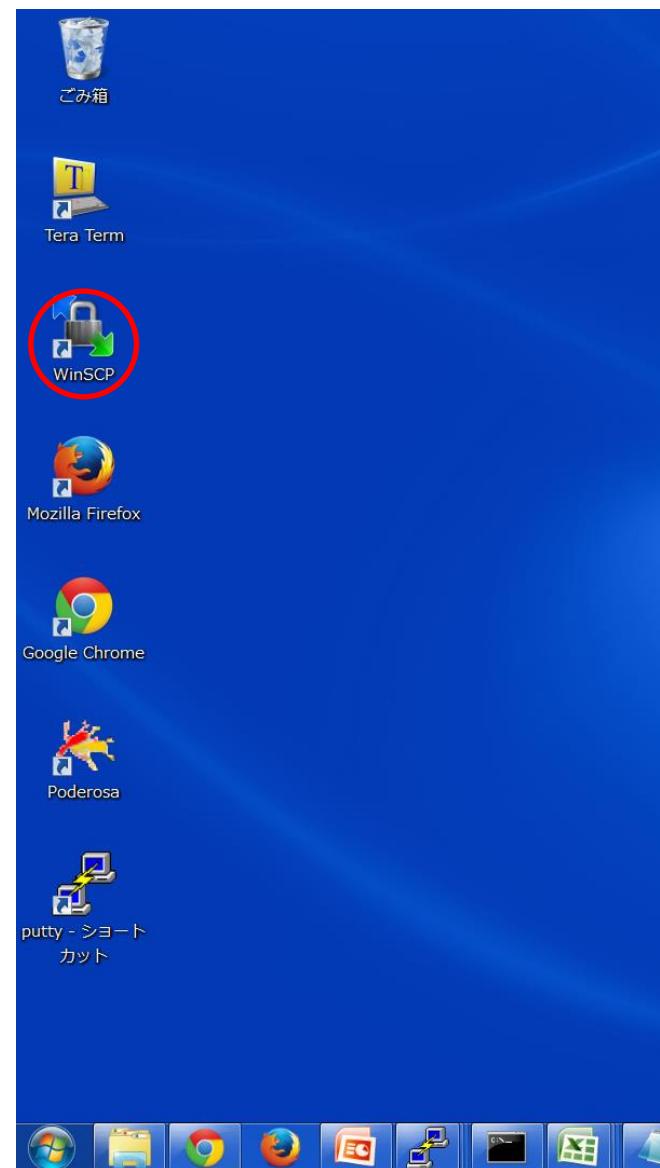
解析データのWindows PCへのダウンロード

WinSCPを開き、以下の情報を入力します

- ホスト名: **gw.ddbj.nig.ac.jp**
- ユーザ名: **lect01 ~ lect55**
- パスワード: *** * * ***



Macの方はWinSCPは使用しません。

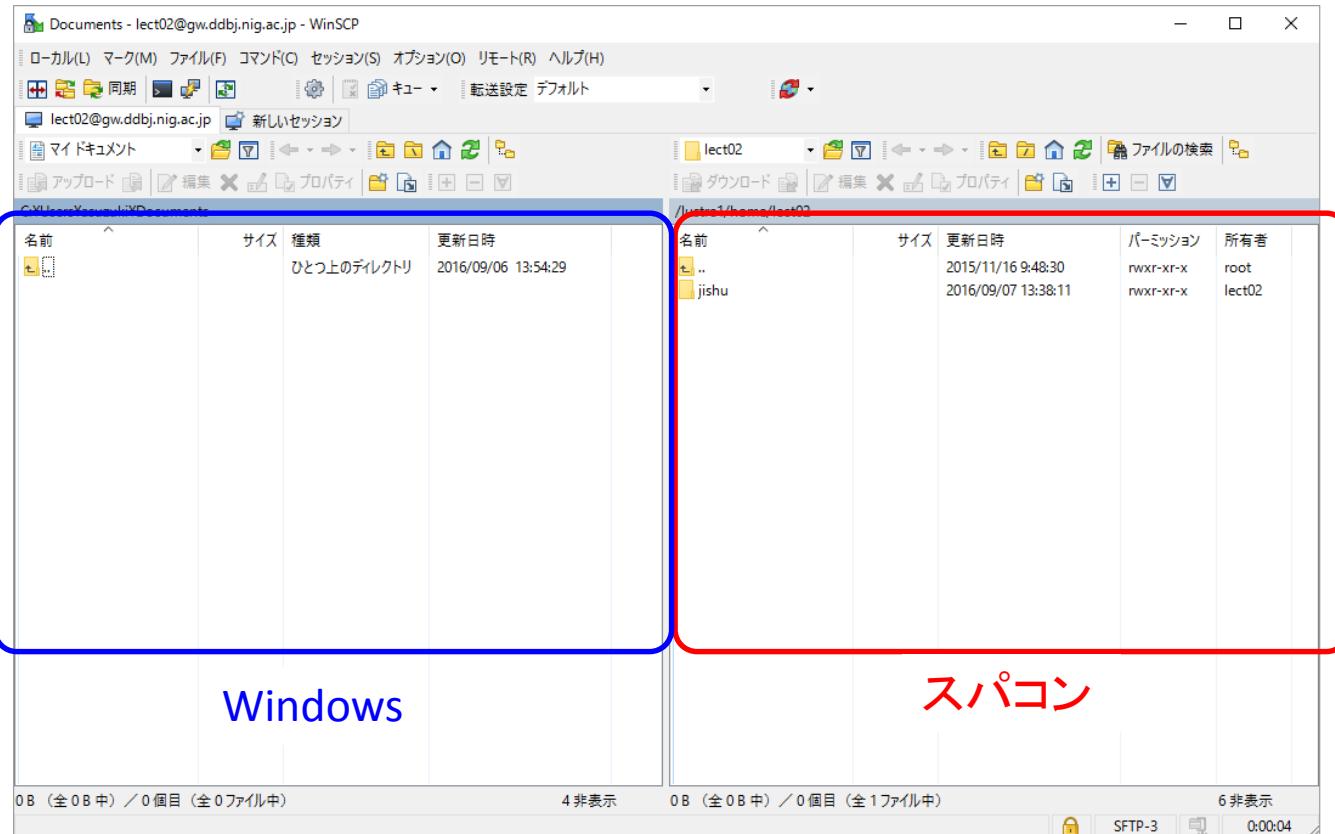


解析データのWindows PCへのダウンロード

LinuxからWindows PCへのデータコピー

後で使用します。

- 下図のようなウィンドウが表示されます。
左側がWindows PC、右側が解析サーバ(Linux)です。
- ファイルあるいはフォルダをクリックし、ドラッグ & ドロップを実行すると、
データコピーが始まります。



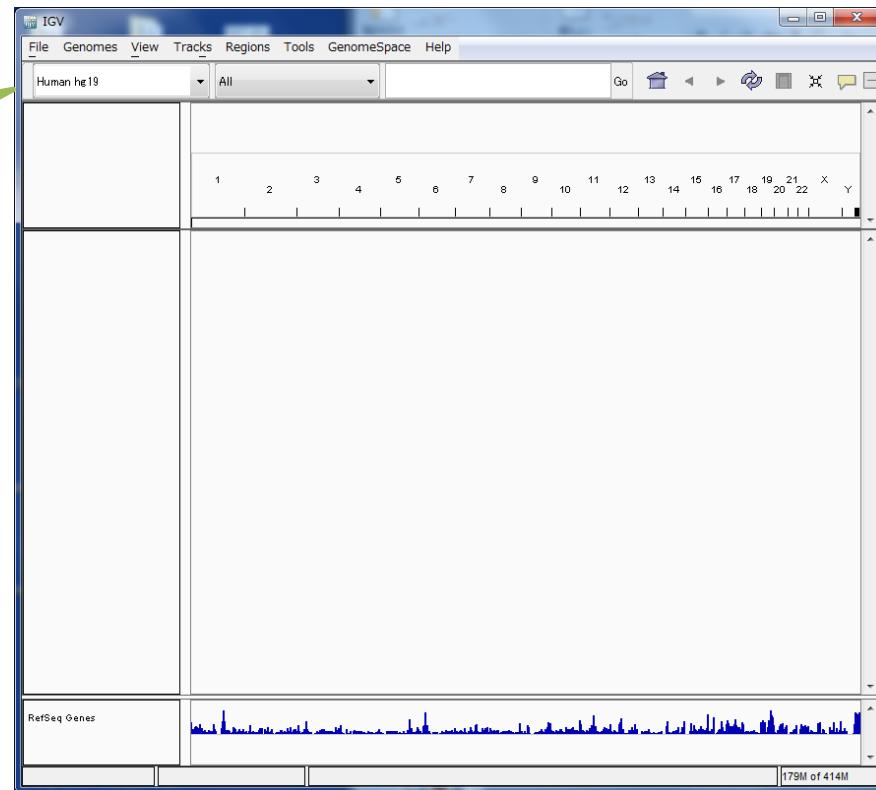
シークエンス可視化

Integrative Genomics Viewer(IGV)

シークエンステータの可視化に使用します。

IGVを起動してみてください。

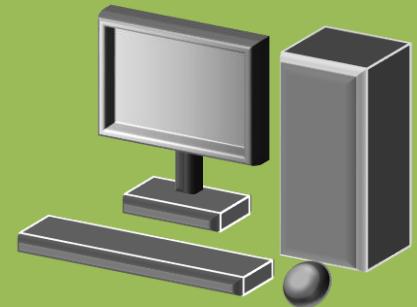
Human hg19を
選択してください



あとで使用しますので、うまく起動しない方や”Human hg19”がない方がいましたらお知らせください。

Session 1

シークエンスデータのマッピングと可視化
およびツールの紹介



課題1

新しくログインして、ホームディレクトリに入り、

\$ qlogin

ログインノードに入る
(サーバによって異なる)

\$ pwd

ディレクトリjishuに移動してファイルを確認

\$ cd jishu

\$ ls

\$ qsub -l s_vmem=8G -l mem_req=8G -l short MAPPING.sh

8Gのメモリを要求して
MAPPING.shを計算ノード
で実行する

short.qを指定

と打ってください。

MAPPING.shの中には、マッピングのコマンドが記載してあります。

あとで説明いたします。

```
gw.ddbj.nig.ac.jp - Tera Term VT
File Edit Setup Control Window Help
[lect02@gw ~]$ qlogin
Your job 19041145 ("QLOGIN") has been submitted
waiting for interactive job to be scheduled ...
Your interactive job 19041145 has been successfully scheduled.
Establishing /home/geadmin/UGER/utilbin/lx-amd64/qlogin_wrapper session to host t263i ...
lect02@t263i's password:
Last login: Wed Sep  7 13:32:48 2016 from t3511
[lect02@t263 ~]$ pwd
/home/lect02
[lect02@t263 ~]$ cd jishu
[lect02@t263 jishu]$ ls
LC2ad.fq  MAPPING.sh  MAPPING_pair.sh  PC-9_1.fq  PC-9_2.fq
[lect02@t263 jishu]$ qsub -l s_vmem=8G -l mem_req=8G -l short MAPPING.sh
Your job 19041148 ("MAPPING.sh") has been submitted
[lect02@t263 jishu]$
```

ジョブの確認と削除

状況確認

\$ qstat

```
gw.ddbj.nig.ac.jp - Tera Term VT
File Edit Setup Control Window Help
[lect02@t263 jishu]$ qstat
Job-ID prior name user state submit/start at queue jclass slots ja-task-ID
19041145 0.25001 QLOGIN lect02 r 09/07/2016 13:48:48 login.q@t263i
19041148 0.25000 MAPPING.sh lect02 r 09/07/2016 13:51:22 short.q@t102i
[lect02@t263 jishu]$
```

ジョブの投入状況の確認

qsubで投入したジョブがジョブとして投入されたかを確認します。投入したジョブの状態確認にはqstatコマンドを利用します。例えばジョブが投入されれば以下のように表示されます。

```
qstat
Job-ID prior name user state submit/start at queue
slots ja-task-ID
-----
6929724 0.00000 jobname username r 06/18/2012 13:00:37 week_hdd.q@t274i
1
6929726 0.00000 jobname username r 06/18/2012 13:00:37 week_hdd.q@t274i
1
6929729 0.00000 jobname username r 06/18/2012 13:00:37 week_hdd.q@t287i
1
6929730 0.00000 jobname username r 06/18/2012 13:00:37 week_hdd.q@t250i
1
```

この時、"state"欄の文字の意味は以下のようになります。

文字	意味
r	ジョブ実行中
qw	ジョブはキューで待機中
t	ジョブは実行ホストへ転送処理中
E	ジョブにエラーが発生
d	ジョブは削除処理中

削除

\$ qdel [job-ID]

詳しくは遺伝研スパコンの「基本的利用方法」のページへ
<https://sc.ddbj.nig.ac.jp/index.php/ja-howtouse>

次世代シークエンサー

Illumina MiSeq / HiSeqシリーズ

リード長：短鎖

主たる用途：ゲノムシークエンシング、
エキソームシークエンシング、
トランскриプトームシークエンシング



Ion PGM / Ion Proton

デスクトップ型シークエンサー

リード長：短鎖

主たる用途：アンプリコンシークエンシング、
エキソームシークエンシング

PacBio RS II/Sequel

1分子リアルタイムシークエンサー

リード長：最長>20 kbの長鎖リード

主たる用途：De novoアセンブル、

細菌ゲノムのシークエンシング、構造多型の解析

HiSeq2500/3000 (東大・新領域・鈴木穰 研)



MinION (国立がん研究センター EPOC)

ONT MinION / PromethION

ナノポアシークエンサー

リード長：短鎖～数kbの長鎖

主たる用途：DNA・RNAシークエンシング

用途も様々

Whole-genome/exome sequencing

DNA配列を解読し、SNP/SNVs やindel等を同定する

RNA-Seq、small RNA-Seq

mRNAやsmall RNAをシークエンスし、発現量の計算や新規転写産物の同定を行う

ChIP-Seq

ヒストン修飾や転写因子の結合部位を同定する

Bisulfite sequencing

DNAのメチル化のパターンを検出する

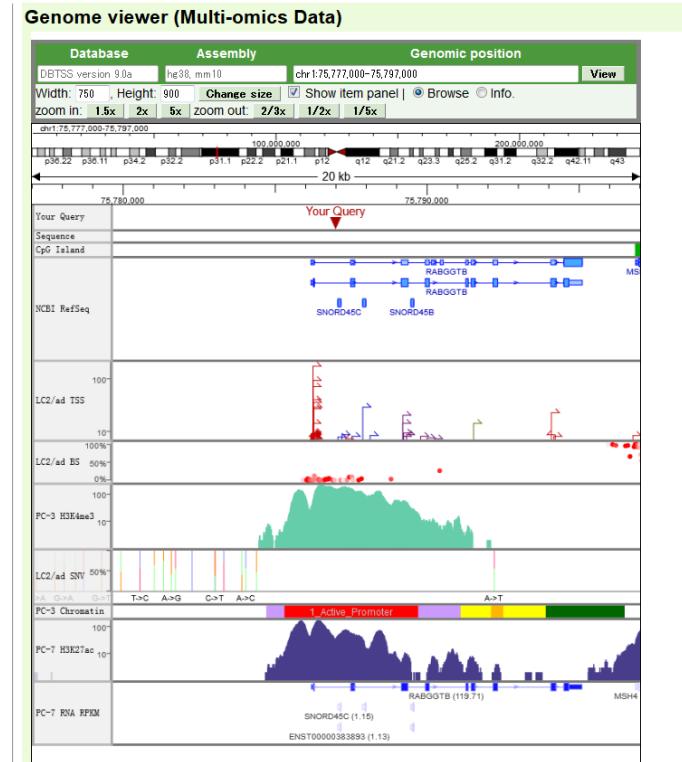
TSS-Seq

転写開始点を同定する

ATAC-Seq

オープンクロマチン領域を同定する

など



fastqファイル（シークエンスファイル）

Format [edit]

A FASTQ file normally uses four lines per sequence.

- Line 1 begins with a '@' character and is followed by a sequence identifier and an *options*/description (like a FASTA title line).
- Line 2 is the raw sequence letters.
- Line 3 begins with a '+' character and is *optionally* followed by the same sequence identifier (and any description) again.
- Line 4 encodes the quality values for the sequence in Line 2, and must contain the same number of symbols as letters in the sequence.

A FASTQ file containing a single sequence might look like this:

```
@SEQ_ID
GATTGGGGTCAAAGCAGTATCGATCAAATAGTAATCCATTGTTCAACTCACAGTTT
+
!''*((((****))%%%++)(%%%).1***-+*''))**55CCF>>>>CCCCCCCC65
```

The character '!' represents the lowest quality while '^' is the highest. Here are the quality value characters in left-to-right increasing order of quality (ASCII):

```
!"#$%&'()**,-./0123456789:;=>?@ABCDEFGHIJKLMNPQRSTUVWXYZ[¥]^_`abcdefghijklmnopqrstuvwxyz{|}~
```

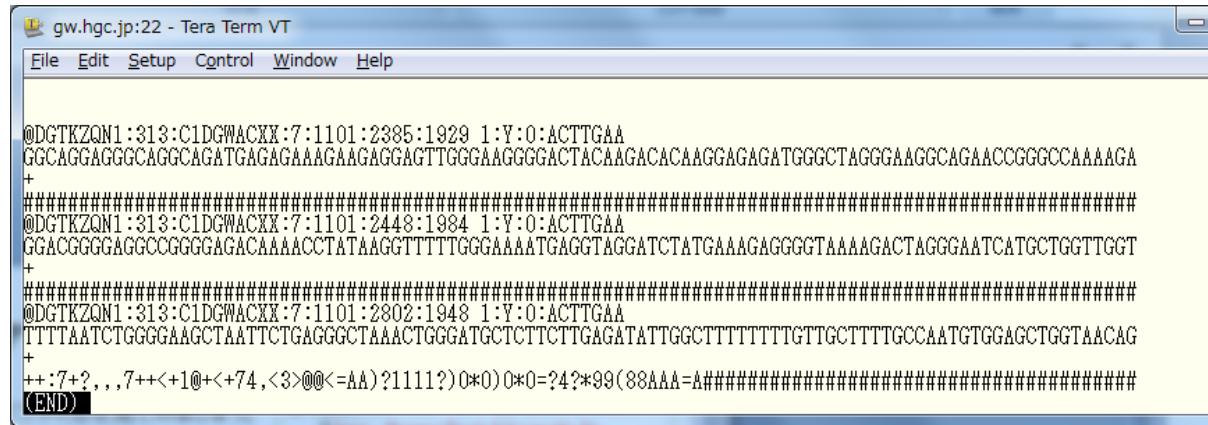


fastqファイル（シークエンスファイル）

シークエンスファイルを見てみましょう。
新しくTeraTermを立ち上げて、
ログインしてください。

【実習】実際にコマンドを入力しましょう

```
$ cd jishu  
$ less PC-9_1.fq  
$ less PC-9_2.fq
```



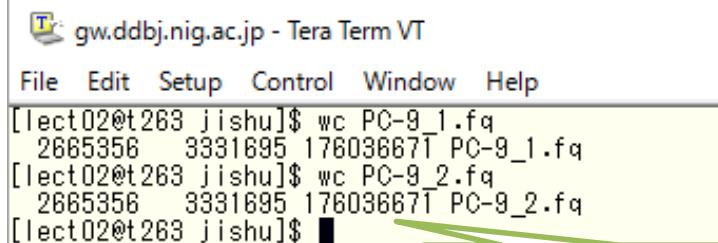
The screenshot shows a terminal window titled "gw.hgc.jp:22 - Tera Term VT". The menu bar includes File, Edit, Setup, Control, Window, and Help. The terminal output displays two fastq sequences. Sequence 1 starts with "@DGTKZQN1:313:C1DGWACXX:7:1101:2385:1929 1:Y:0:ACTTGAA" and sequence 2 starts with "@DGTKZQN1:313:C1DGWACXX:7:1101:2448:1984 1:Y:0:ACTTGAA". Both sequences are terminated with a double hash symbol "##". The command "(END)" is visible at the bottom.

lessコマンドは
qで終了します。

wcコマンドでシークエンスのリード数を
数えてみましょう。

【実習】実際にコマンドを入力しましょう

```
$ wc PC-9_1.fq
```



The screenshot shows a terminal window titled "gw.ddbj.nig.ac.jp - Tera Term VT". The menu bar includes File, Edit, Setup, Control, Window, and Help. The command "\$ wc PC-9_1.fq" is entered and its output is shown: "2665356 3331695 176036671 PC-9_1.fq". Below this, the command "\$ wc PC-9_2.fq" is entered and its output is shown: "2665356 3331695 176036671 PC-9_2.fq". The prompt "[lect02@t263 jishu]" appears at the bottom.

行数 単語数 バイト数

2665356行ありました。

fastqファイルは4行で1リード分なので、
このファイルには666339リード入ってい
ることになります

課題2

実際に今回使ったシークエンスファイル(fastqファイル)の中身を見てみましょう。

【実習】実際にコマンドを入力しましょう

\$ cd jishu
\$ ls -lt

\$ less LC2ad.fq

```
gw.hgc.jp:22 - Tera Term VT
File Edit Setup Control Window Help
@DGTZQN1:313:C1DGWACXX:7:1101:2389 1:Y:0:ACTTGAA
GGCAGGAGGGCAGGCAGATGAGAGAAAAGAAGAGGAGTTGGAAAGGGGACTACAAGACACAAGGAGAGATGGCTAGGGAAAGGCAGAACCGGGCCAAAAAGA
+
#####
@DGTZQN1:313:C1DGWACXX:7:1101:2443:1984 1:Y:0:ACTTGAA
GGACGGGAAGGCCGGGAGACAAAACCTATAAGGTTTGGAAAATGAGGTAGGATCTATGAAAGAGGGTAAAAGACTAGGGAAATCATGCTGGTTGGT
+
#####
@DGTZQN1:313:C1DGWACXX:7:1101:2802:1948 1:T:0:ACTTGAA
TTTAATCTGGGAAGCTAATTCTGAGGGCTAACTGGGATGCTCTTCTGAGATATTGGCTTTTTGTTGCTTTGCCAATGTGGAGCTGGTAACAG
+
++:7+?, +7++<+10+<+74,<3>@<=AA)<1111?)0*0)0*0=24?*99(88AAA=A##########
@DGTZQN1:313:C1DGWACXX:7:1101:3143:1935 1:N:0:ACTTGAA
CGATGCAGGATCCTATGAAATGAAACAGAACCCACGGAGTCACCCAGTCACCTGAATGTCATCTATGCCAAGATGTCCCCACC
+
?@DD?BDDHHH:D>G>9?4CF,3>FA3+2CC>EEF112?)@0@60@8?2?@FGGI5@=C3?EB).;B@>A6>>;355((;:55;A?C3<<(:@>0@3??:5
@DGTZQN1:313:C1DGWACXX:7:1101:3011:1945 1:Y:0:ACTTGAA
ATTCACAAAGTACATGTTCAAAGTAATAGAAACCTGGCCTTCATTCAAAATAACTGCAAAAGAGAGTACATGCACTGTCAGTAGCTGTTACAATTTAATA
+
??BD;,,=+<.22<AE04,,<AE<C)*:;):C*?DDEB<29)00*/?99?*)=@A##########
@DGTZQN1:313:C1DGWACXX:7:1101:3064:1958 1:N:0:ACTTGAA
CAAGACTGAGATAAAATTAAATGTTGAAATGAATTAAAGCATTTGAGGTATTTATGAGGTCAATTGCTACTTTGCAATGTGATATGGACTGCTAA
+
=?D?DFDDHHH?EBCIIIDCF<EFHGAD:F?BCDEAEC1CHHIEE8?RDGG3?@FH:<F?DF38=BPHHGIE>CGEGG>CDHHFFCHFCFFCD@>@>
@DGTZQN1:313:C1DGWACXX:7:1101:3061:1997 1:N:0:ACTTGAA
AAACGCCCTCTCGTCTGATCCGTCATAACAGCAGTCCTACTTCTCTCTCCAGTCCTAGGTGCTGGCATCACTATACTAAACAGACCG
+
@@@ADDDDAADDF@E1CF<E4EFGC)?CBB<DD9D2?B03BF49=B8,8)87,8)-7))70,.=?A3.)7.);;3;?@AA35>@@A55>B@0AA###
@DGTZQN1:313:C1DGWACXX:7:1101:3361:1936 1:N:0:ACTTGAA
LC2ad.fq
```

課題3

fastqファイルは、4行で、1リード分ということでしたが、
この LC2ad.fq のファイル中には、
何本のリードがあるか数えてみましょう

【実習】実際にコマンドを入力しましょう

\$ wc LC2ad.fq

シークエンスをヒトゲノムにマッピング

見てもらったfastqデータは、

先ほどの課題1で、ヒトの参照ゲノム配列に対して、
マッピングを実行中です。

投げたjobの中身を、みてみましょう！

【実習】実際にコマンドを入力しましょう

\$ less MAPPING.sh

```
gw.ddbj.nig.ac.jp - Tera Term VT
File Edit Setup Control Window Help
[lect01@t260 jishu]$ less MAPPING.sh
#!/bin/sh
## -S /bin/sh
## -cwd
#BWA aln
bwa aln /home/lect01/reference/all_hg19.fa LC2ad.fq > LC2ad.sai
#BWA samse
bwa samse /home/lect01/reference/all_hg19.fa LC2ad.sai LC2ad.fq > LC2ad.sam
#BWA sam -> bam
samtools view -bS -o LC2ad.bam LC2ad.sam
## samtools sort
samtools sort -o LC2ad_sorted.bam LC2ad.bam
## samtools sort
samtools index LC2ad_sorted.bam

MAPPING.sh (END)
```

おまじない
(<https://sc.ddbj.nig.ac.jp/index.php/ja-howtouse>)

BWAというソフトウェアでマッピング
bwa aln → アライメント
bwa samse → SAMファイル作成

SAMtoolsというソフトウェアで
SAMファイルをBAMファイルにする

Mapping software もさまざま

□ ELAND

Illumina社のソフトウェア

□ BWA

indelのマッピングに強く、
ゲノム・エキソーム解析に適している

□ Bowtie

少ないメモリで高速にマッピングする(indelに弱い)

□ TopHat

スプライスを考慮してマッピングする
RNA-Seqに適している

など

gw.ddbj.nig.ac.jp - Tera Term VT

[lect01@t260 ~]\$ bwa

Program: bwa (alignment via Burrows-Wheeler transformation)
Version: 0.7.13-r1126
Contact: Heng Li <lh3@sanger.ac.uk>

Usage: bwa <command> [options]

Command: index index sequences in the FASTA format
mem BWA-MEM algorithm
fastmap identify super-maximal exact matches
pemerge merge overlapping paired ends (EXPERIMENTAL)
ain sapped/unzapped alignment
samse generate alignment (single ended)
sampe generate alignment (paired ended)
bwaw index sequences in shared memory
shm convert FASTA to PAC format
pac2pac generate BWT from PAC
pac2bwt pac2bwtgen alternative algorithm for generating BWT
bwupdate update .bwt to the new format
bw2sa generate SA from BWT and Occ

Note: To use BWA, you need to first index the genome with 'bwa index'. There are three alignment algorithms in BWA: 'mem', 'bwaw', and 'ain/samse/sampe'. If you are not sure which to use, try 'bwa mem' first. Please 'man ./bwa.1' for the manual.

\$ bwa

gw.ddbj.nig.ac.jp - Tera Term VT

[lect01@t260 ~]\$ bowtie

No index, query, or output file specified!

Usage:

bowtie [options] <bowt> [<1> | <2> | --12 <r> | <s>] [<hit>]

<m1> Comma-separated list of files containing upstream mates (or the sequences themselves, if -c is set) paired with mates in <m2>
<m2> Comma-separated list of files containing downstream mates (or the sequences themselves if -c is set) paired with mates in <m1>
<r> Comma-separated list of files containing OrgSeq+5'/3' reads. Can be a mixture of paired and unpaired. Specify '-' for stdin.
<s> Comma-separated list of files containing unpaired reads, or the sequences themselves, if -c is set. Specify '-' for stdin.
<hit> File to write hits to (default: stdout)

Input:

-q query input files are FASTQ .fq/.fastq (default)
-f query input files are (multi-)FASTA .fa/.maf
-r query input files are raw one-sequence-per-line
-c query sequences given on cmd line (as <mates>, <singles>) reads and index are in colorspace
-0/-quals <file> QV file(s) corresponding to CSFASTA inputs; use with -f -c same as -Q, but for mate files 1 and 2 respectively
--Q1/-Q2 <file> skip the first <int> reads/pairs in the input
-s/-skip <int> stop after first <int> reads/pairs (excl. skipped reads)
-u/-upto <int> trim <int> bases from 5' (left) end of reads
-5/-trim5 <int> trim <int> bases from 3' (right) end of reads
-3/-trim3 <int> input quals are Phred+33 (default)

\$ bowtie

gw.ddbj.nig.ac.jp - Tera Term VT

[lect01@t260 ~]\$ tophat

TopHat maps short sequences from spliced transcripts to whole genomes.

Usage:

tophat [options] <bowtie_index> <reads1[,reads2,...]> [reads2[,reads3,...]] ¶
[quals1[,quals2,...]] [quals2[,quals3,...]]

Options:

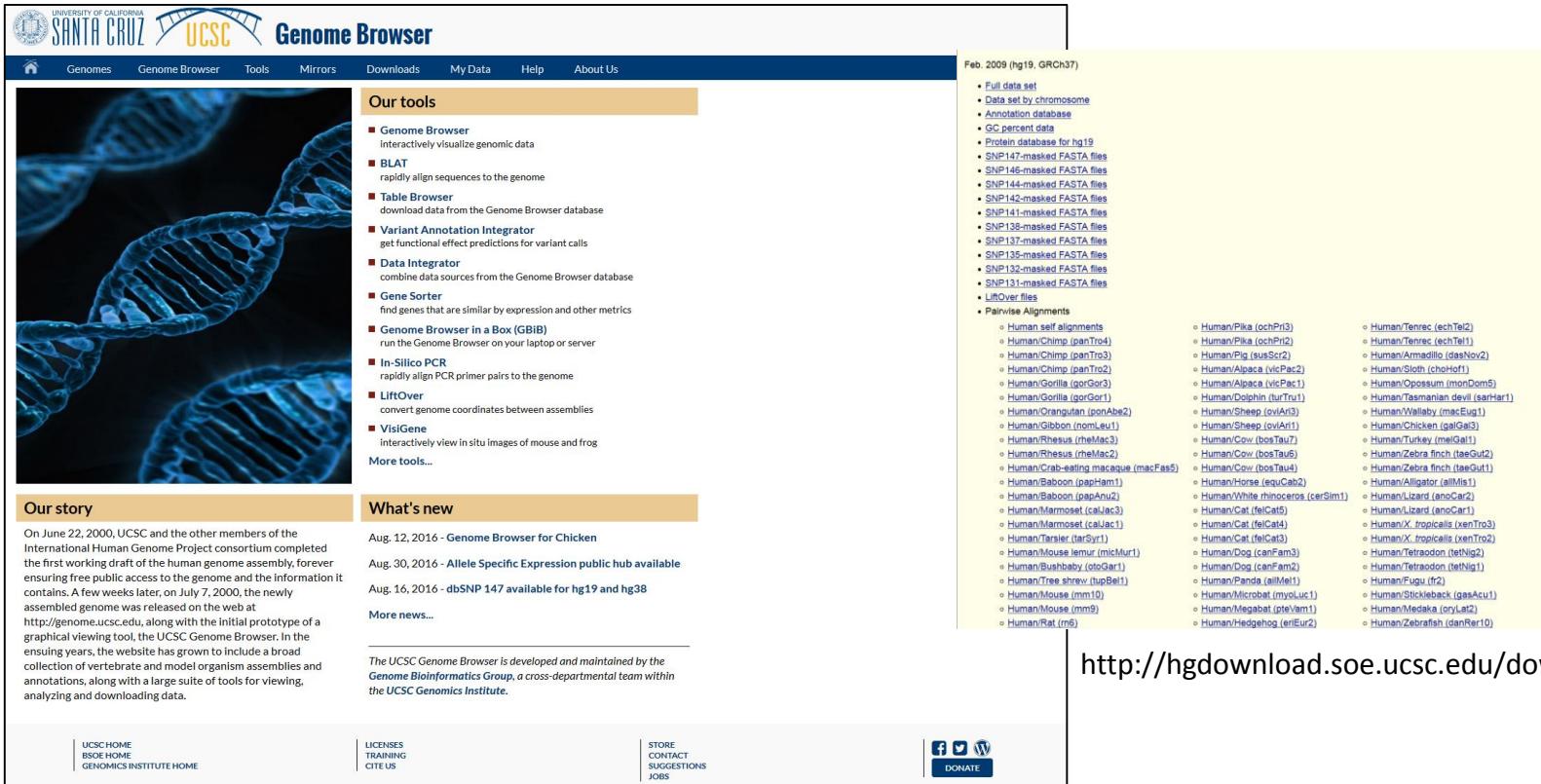
-v/-version
-o/-output-dir <string> [default: ./tophat_out]
-a/-min-anchor <int> [default: 8]
-m/-splice-mismatches <0-2> [default: 0]
-i/-min-intron-length <int> [default: 50]

\$ tophat

参照ゲノム配列

Human, mouse などの主なモデル生物の
リファレンスゲノムや遺伝子モデル等のアノテーションデータは、
UCSCやNCBIより取得できます。

今回、参照ゲノム配列として用いたのはUCSC hg19です。



The screenshot shows the UCSC Genome Browser homepage. At the top, there's a navigation bar with links to Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area features a large image of a DNA double helix. Below it, there are two sections: "Our tools" and "What's new". The "Our tools" section lists various tools like Genome Browser, BLAT, Table Browser, Variant Annotation Integrator, Data Integrator, Gene Sorter, Genome Browser in a Box, In-Silico PCR, LiftOver, VisiGene, and more. The "What's new" section lists recent updates: "Aug. 12, 2016 - Genome Browser for Chicken", "Aug. 30, 2016 - Allele Specific Expression public hub available", and "Aug. 16, 2016 - dbSNP 147 available for hg19 and hg38". A sidebar on the right is titled "Feb. 2009 (hg19, GRCh37)" and lists a comprehensive set of genomic resources, including full data sets, masked FASTA files for chromosomes 1 through 22, and various species alignments such as Human, Pika, Pig, Alpaca, Dolphin, Sheep, Gorilla, Orangutan, Gibbon, Rhesus, Macaque, Baboon, Marmoset, Tarsier, Lemur, Bushbaby, Zee shrew, Mouse, and Rat.

http://hgdownload.soe.ucsc.edu/downloads.html

先ほどの課題1の マッピングはおわりましたでしょうか？？

下記の出力ファイルが出てきているか確認してください。

【実習】実際にコマンドを入力しましょう
\$ ls -lt

出力データ

LC2ad.fq Raw データ(シークエンスタグ)

LC2ad.sai

LC2ad.sam

LC2ad.bam

LC2ad_sort.bam

LC2ad_sort.bam.bai

} マッピング結果

MAPPING.sh.eXXXXXXXX

MAPPING.sh.oXXXXXXXX

} jobのログファイル

SAM (BAM) 形式データ

<http://samtools.sourceforge.net/samtools.shtml>

SAMファイルの中身を眺めてみましょう。

【実習】実際にコマンドを入力しましょう

\$ less LC2ad.sam

\$ samtools view LC2ad.bam | more

標準フィールド

#	略号	意味	例
1	QNAME	リード名	SRR015293.3
2	FLAG	フラグ	16
3	RNAME	リファレンス名	chr3
4	POS	スタート位置	186338939
5	MAPQ	マッピングクオリティ	25
6	CIGAR	CIGAR	32M
7	RNEXT	ペアリファレンス名	*
8	PNEXT	ペアリードのスタート位置	0
9	TLEN	総断片長 (インサートサイズ+両リード長)	0
10	SEQ	リード配列	TTGTGATGATTCGACGGTAAGCCACCATGAT
11	QUAL	クオリティ	KVNKHYYQYYJSCHQYYYYYTYYYYYYYYYY
12	-	オプショナルフィールド(タグ)	XT:A:U NM:i:2 X0:i:1 X1:i:0 XM:i:2 XO:i:0 XG:i:0 MD:Z:4C7T19

<https://cell-innovation.nig.ac.jp/wiki/tiki-index.php?page=SAM>

データ可視化

<https://www.broadinstitute.org/igv/home>

The image shows a screenshot of the Integrative Genomics Viewer (IGV) website. On the left, there is a sidebar with the IGV logo and navigation links: Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide, File Formats, Release Notes, Credits, and Contact. Below the sidebar is a search bar labeled "Search website". The main content area has a header "Home" and features the text "Integrative Genomics Viewer". To the right of the text is a large image of the IGV software interface, which displays a genomic track with various data layers (e.g., tracks for genes, RNA levels, and protein expression). The overall design is clean and modern.

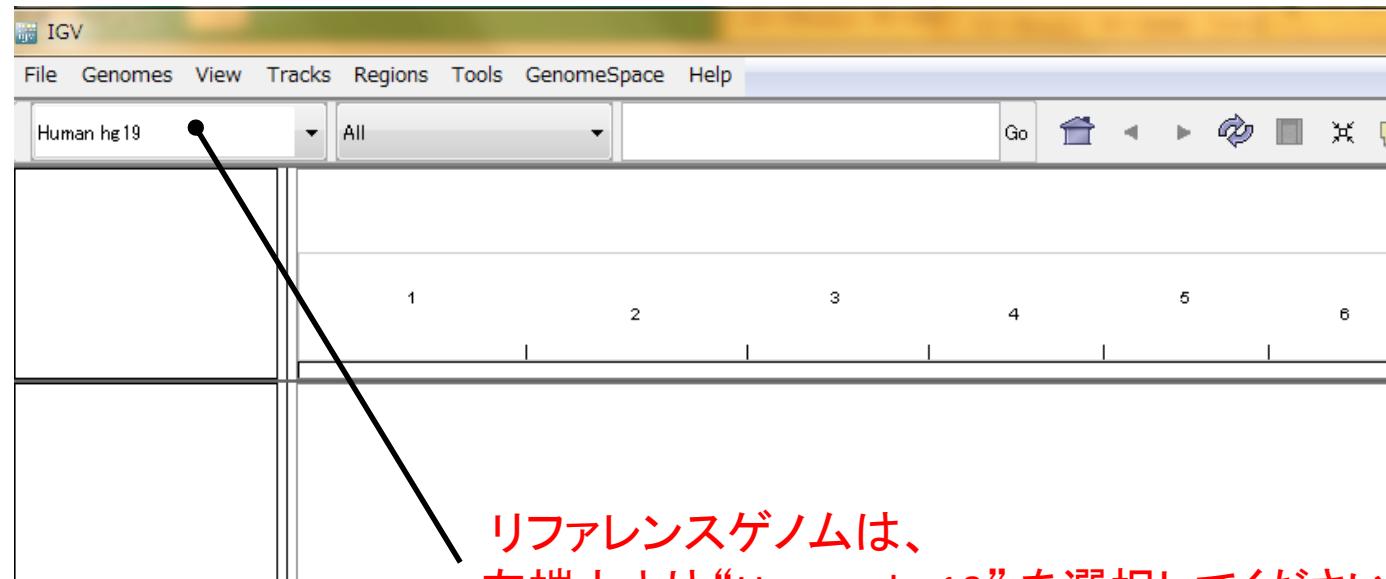
課題4

マッピングしたデータを可視化ツール(IGV)でみてみましょう。

WinSCPもしくはターミナルを用いて、PCに

LC2ad_sort.bam および LC2ad_sort.bam.bai をダウンロードしてください。

IGVを起動させて、



リファレンスゲノムは、
左端上より “Human hg19” を選択してください。

IGVのFile → Load from File... でLC2ad_sorted.bamを開いてください。

ファイル転送

Mac

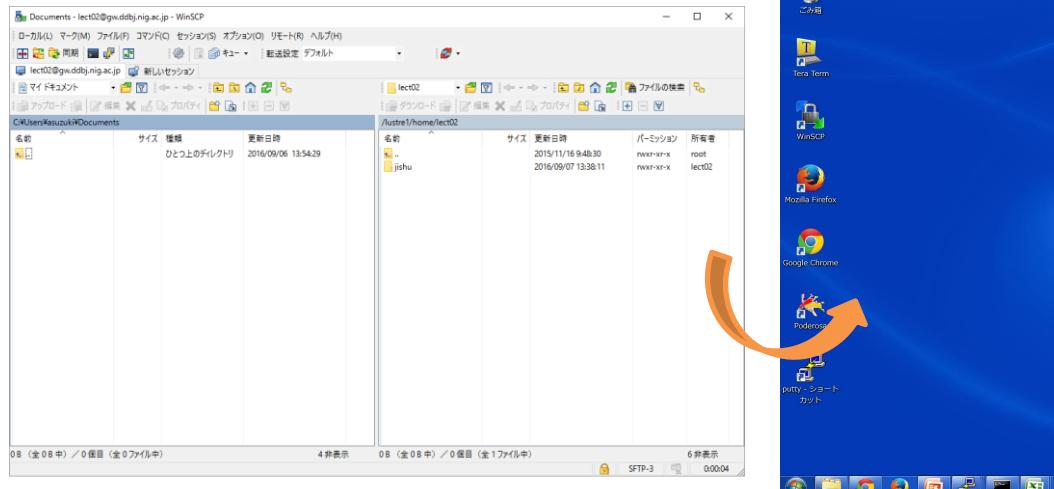
1. 新たにターミナルを起動
2. scpを利用してスパコン上のファイルをPCのデスクトップに転送

```
$ scp lect01@gw.ddbj.nig.ac.jp:/home/lect01/LC2ad_sort.bam ~/Desktop/  
$ scp lect01@gw.ddbj.nig.ac.jp:/home/lect01/LC2ad_sort.bam.bai ~/Desktop/
```

scp[スペース][アカウント名]@[サーバ名]:[転送したいファイルのパス][スペース][転送先のパス]

Windows

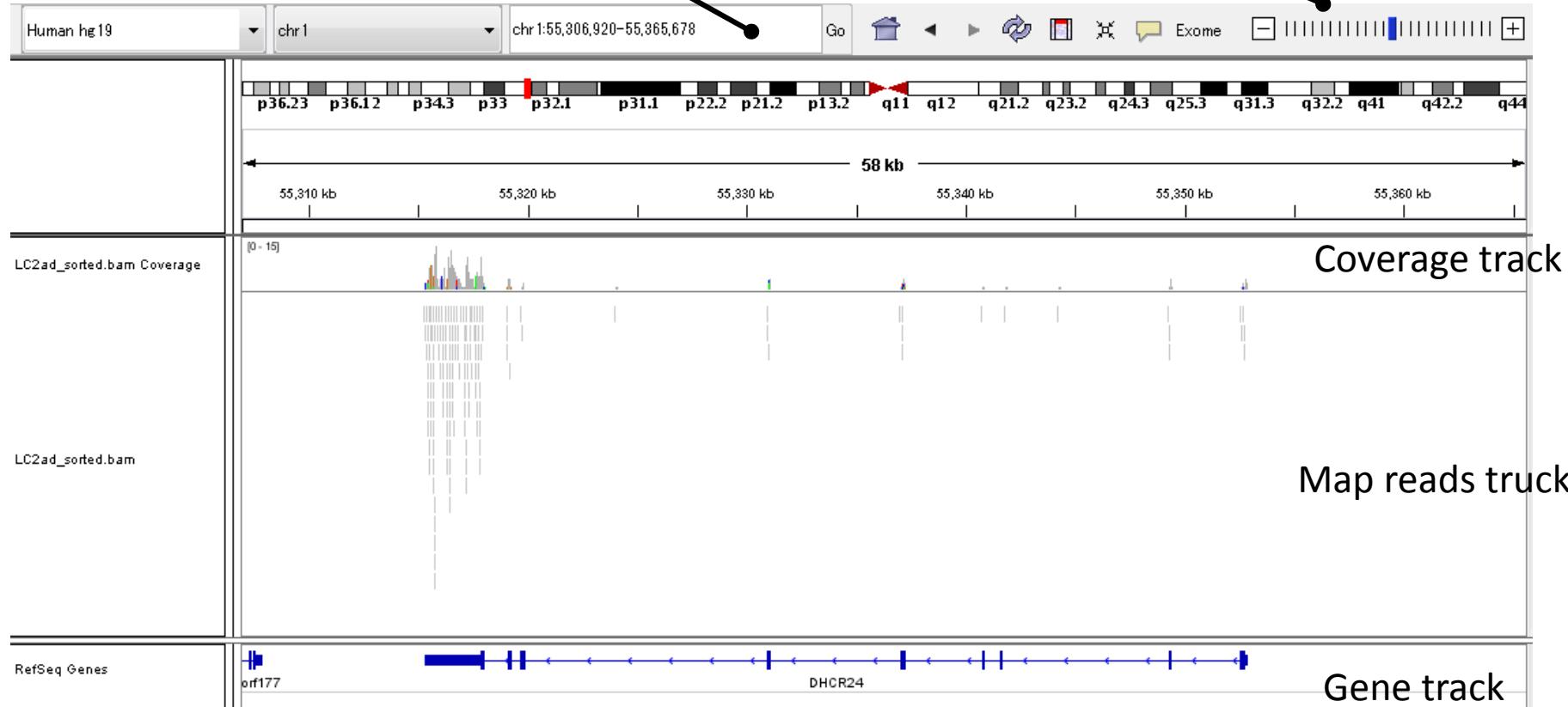
1. WinSCPを起動してログイン
2. jishuディレクトリの下にあるLC2ad_sort.bamとLC2ad_sort.bam.baiをドラッグ & ドロップでPCのデスクトップにコピー



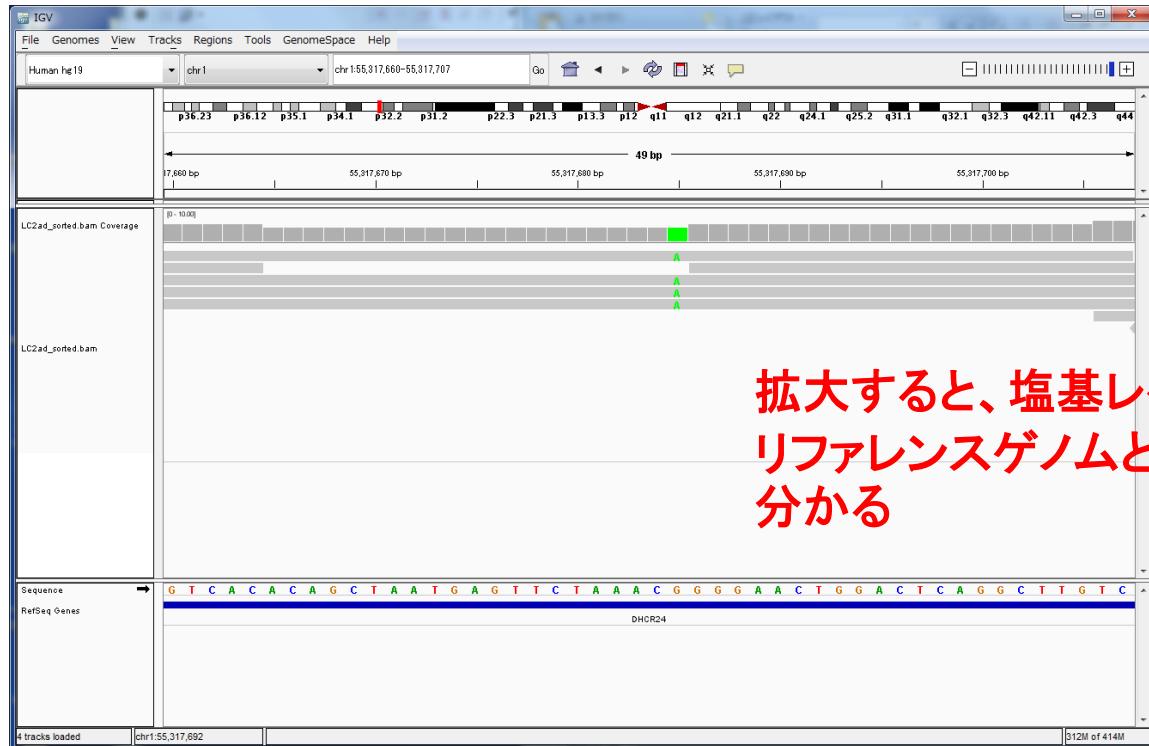
IGV 表示内容

Symbol や座標で検索可能

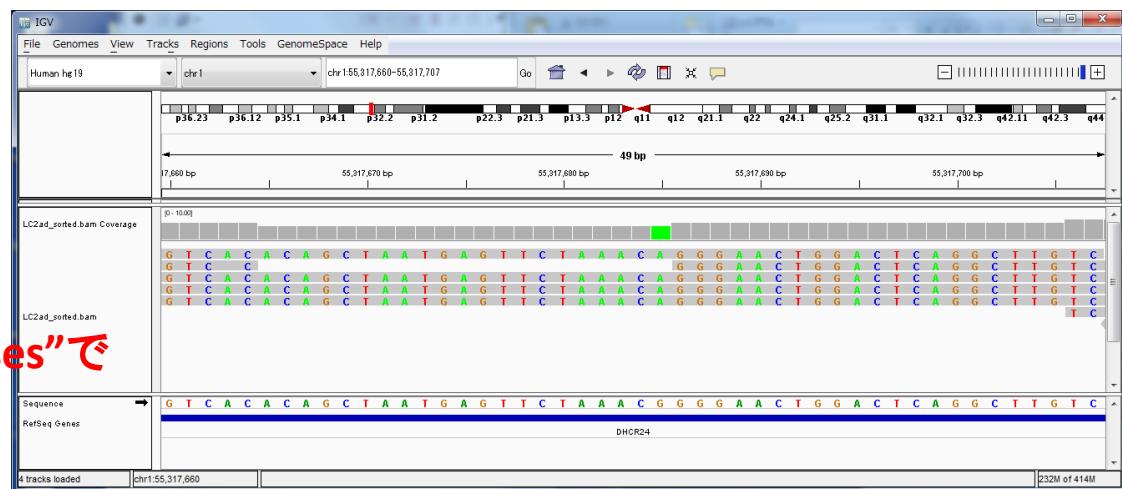
Zoom in / out



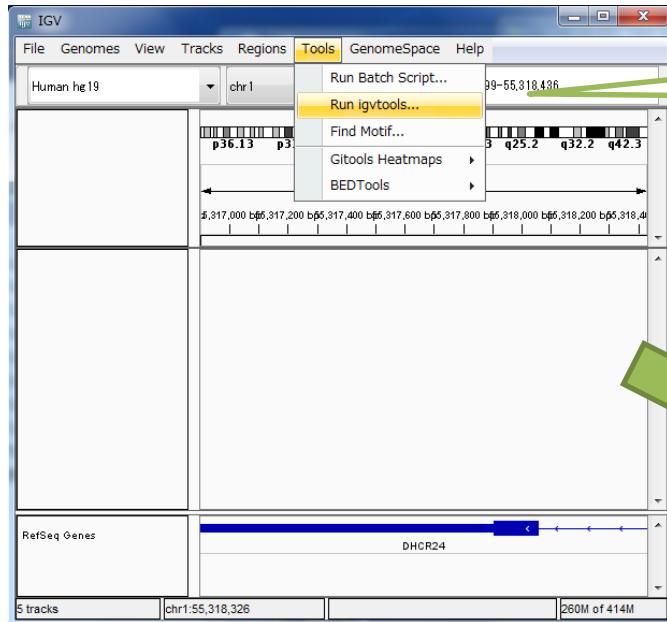
DHCR24 と検索してみましょう。



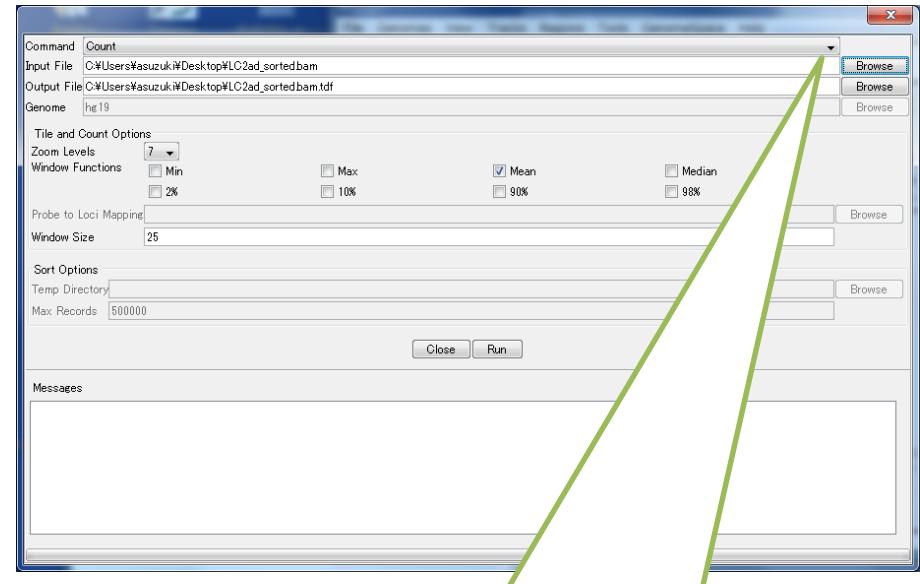
右クリック→"show all bases"で
全ての塩基を表示できる



おまけ



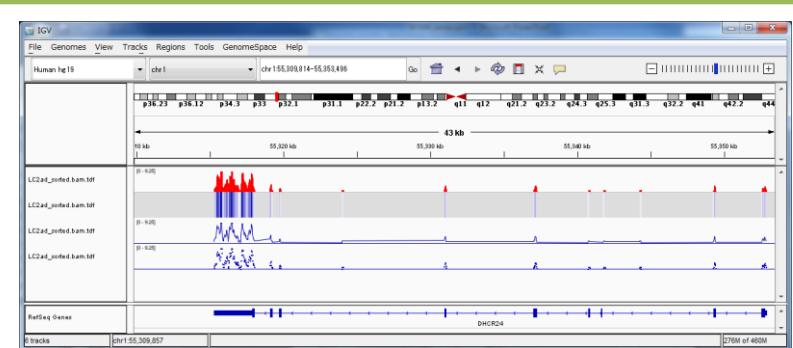
Tools → Run igvtools



Input FileにBAMファイル(LC2ad.bam)を選択
→ Run



TDFファイルをIGVで開いてみましょう



色や表示を変えることができます

課題おまけ

ペアエンドのゲノムシークエンスデータ(PC-9_1.fq, PC-9_2.fq)をマッピングしてみましょう。

新しくログインして、ホームディレクトリに入り、

```
$ qlogin  
$ pwd  
$ cd jishu  
$ ls  
$ cat MAPPING_pair.sh  
$ qsub -l s_vmem=8G -l mem_req=8G -l short MAPPING_pair.sh
```

ログインノードに入る

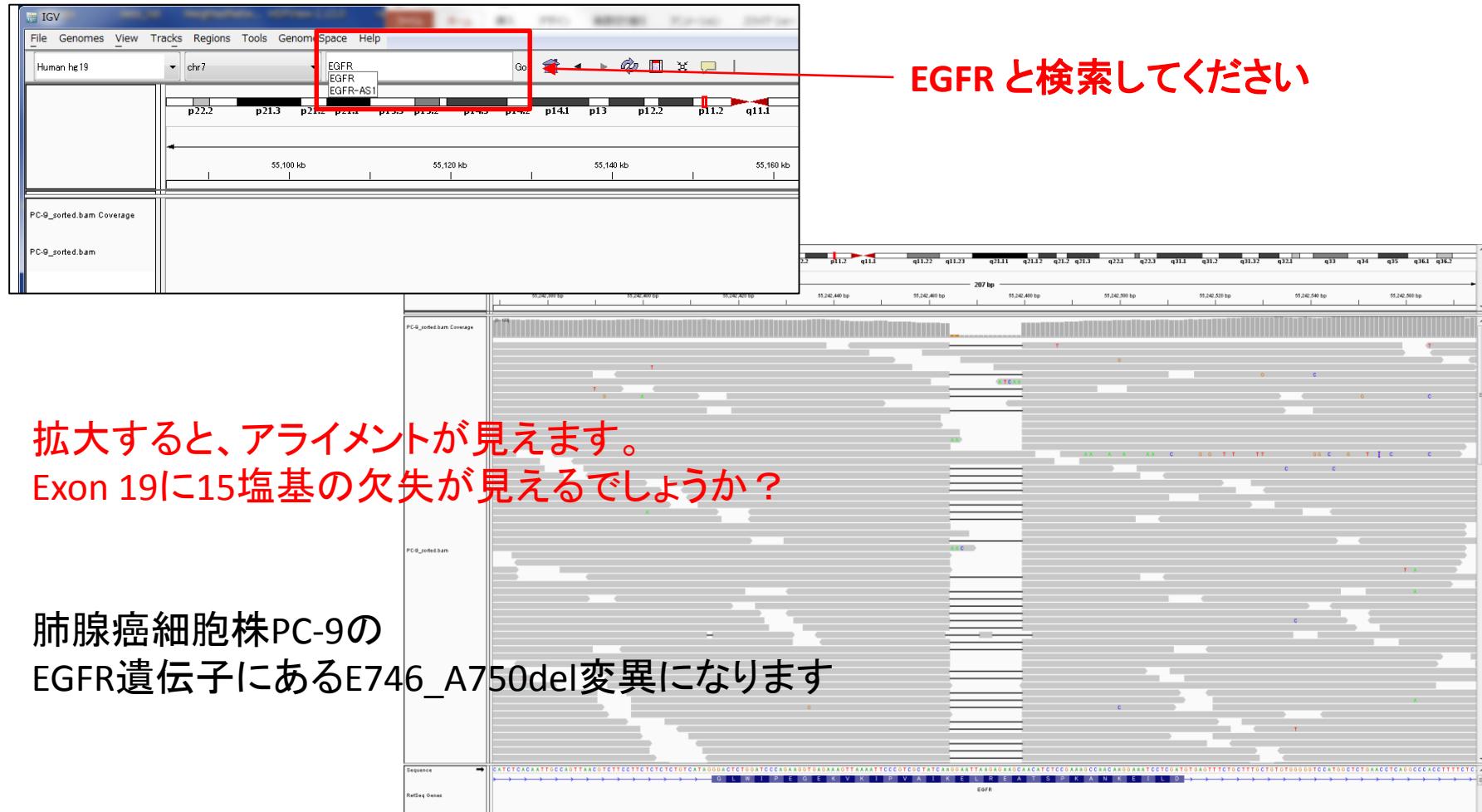
ディレクトリjishuに移動してファイルを確認

MAPPING_pair.shの中身を確認

8Gのメモリを要求して
MAPPING_pair.shを計算
ノードで実行する

課題おまけ

PC-9_1.fqとPC-9_2.fqをマッピングした結果できた
PC-9_sorted.bamとPC-9_sorted.bam.baiをPCに転送して、IGVで見てみましょう。



マッピング後の解析例①

全ゲノム・エキソーム → ゲノム多型・変異の検出

BWA-GATKの場合

BWA

参照ゲノム配列へマッピング

↓

Picard/samtools

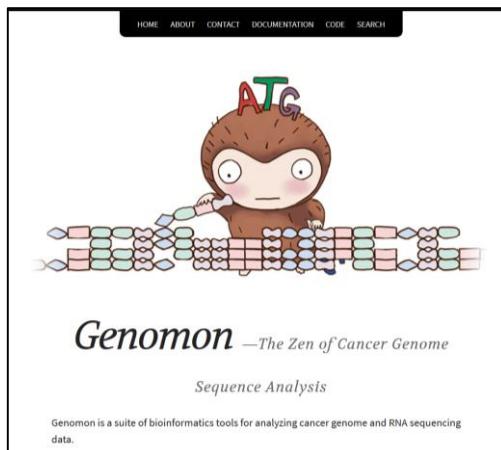
配列のソート・Duplicateの除去など

↓

GATK

多型・変異を検出

<https://software.broadinstitute.org/gatk/best-practices/>



検出した変異はすべてが正しいとは限らない！
IGVでの目視やサンガーシーケンス等での確認は必須

Genomon

東大医科研のヒトゲノム解析センタースパコンに
インストール済みのツール

<http://genomon-project.github.io/GenomonPages/>

マッピング後の解析例②

RNA-Seq → 遺伝子発現解析・融合遺伝子検出

TopHat-CuffLinksの場合

TopHat

参照ゲノム配列へマッピング

↓

CuffLinks

アセンブルによる転写産物の検出

発現量(fpkm)算出

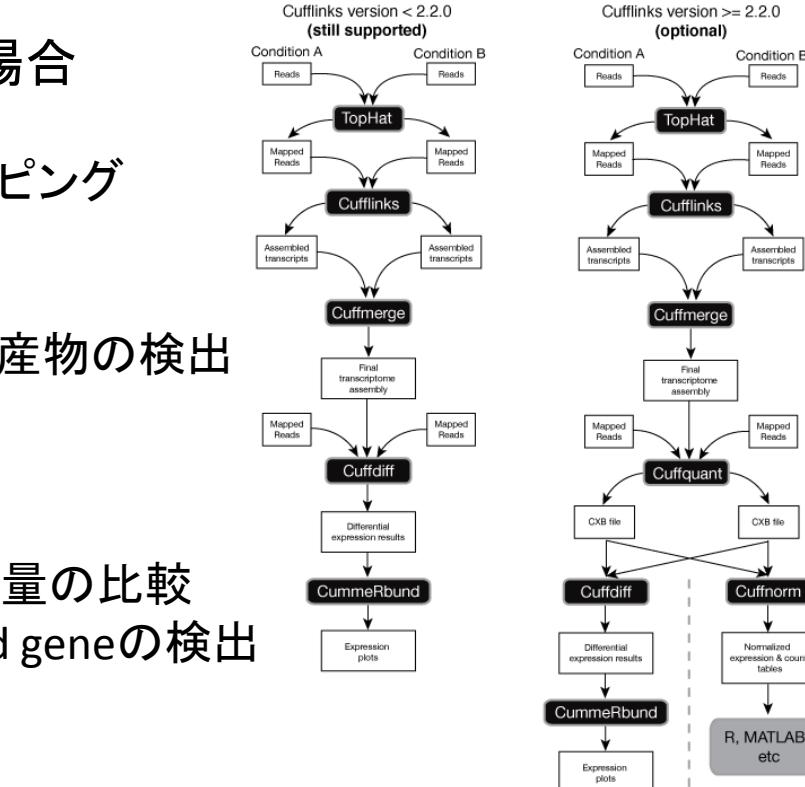
↓

Cuffdiff

サンプル(群)間の発現量の比較

Differentially expressed geneの検出

TopHat-fusion
融合遺伝子を検出



(<http://cole-trapnell-lab.github.io/cufflinks/manual/>)

tophat-fusion-postの出力例

Candidate fusion list									
The following tables show fusion candidates where fusions are grouped based on their genomic locations (table description).									
1. chr0-chr9 rf									
b.16	ENSG00000231455	chr2	9710577	[KCNIP2]	chr2	96040043	[32]	76	195
b.16	ENSG00000231455	chr2	9712139	[KCNIP2]	chr2	96040043	[42]	76	5
b.16	ENSG00000238008	chr2	9716019	[KCNIP2]	chr2	96040043	[22]	76	17
b.16	ENSG00000238008	chr2	9717040	[KCNIP2]	chr2	96040043	[188]	76	110
2. chr0-chr4 fr									
b.7	ENSG00000270106	chr1	231749471	[SHROOM3]	chr4	77852067	[302]	[23]	[142]
3. chr0-chr5 rr									
b.24	SIL1	chr5	130532129	[MEB1]	chr5	13872574	[319]	[15]	[66]
4. chr10-chr10 rf									
b.18	CCDC6	chr10	61665979	[RET]	chr10	40812031	[184]	[27]	[98]
b.18	CCDC6	chr10	61665979	[RET]	chr10	40812031	[43]	[2]	[81]
5. chr17-chr18 ff									
b.19	FOXP2	chr17	60521423	[COBE1]	chr18	57180673	[180]	[64]	[112]
b.19	FOXP2	chr17	60521423	[COBE1]	chr18	57180673	[412]	[64]	[273]
b.19	FOXP2	chr17	60521423	[COBE1]	chr18	57180673	[244]	[64]	[209]
b.19	FOXP2	chr17	60521423	[COBE1]	chr18	57180673	[106]	[64]	[76]
b.19	FOXP2	chr17	60522538	[COBE1]	chr18	57179184	[34]	[64]	[23]
6. chr9-chr9 ff									
b.11	RP51B	chr9	33240502	[VP552]	chr9	33238099	[968]	[19]	[182]
7. chr9-chr18 rr									
b.7	Cikorf79	chr9	132595725	[ENSG00000287013]	chr18	52784113	[527]	[58]	[402]
8. chr18-chr18 rr									
b.20	AFG3L2	chr18	12348271	[SPIRE1]	chr18	12494656	[40]	[5]	[15]
9. chr20-chr20 rf									
b.21	APOLCC	chr20	33981894	[PHF20]	chr20	346050405	[352]	[51]	[107]
10. chr14-chr14 rr									
b.21	PRFL4	chr14	14954593	[KATNA1]	chr14	14952917	[51]	[5]	[39]

マッピング後の解析例③

ChIP-Seq/ATAC-Seq → ピーク検出

Bowtie
参照ゲノム配列へマッピング
↓
MACS/MACS2
ピークを検出

```
README for MACS (2.1.0)

Time-stamp: <2016-02-15 15:31:42 Tao Liu>

Introduction

With the improvement of sequencing techniques, chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) is getting popular to study genome-wide protein-DNA interactions. To address the lack of powerful ChIP-Seq analysis methods, we present a novel algorithm, named Model-based Analysis of ChIP-Seq (MACS), for identifying transcription factor binding sites. MACS is a fast and powerful tool for identifying significant peaks from ChIP-Seq data. MACS improves the spatial resolution of binding sites through combining the information of both sequencing tag position and orientation. MACS can be easily used for ChIP-Seq data alone, or with control sample with the increase of specificity.

Install

Please check the file 'INSTALL' in the distribution.

Usage of MACS2

macs2 [-h] [--version]
      [callpeak,filterdup,bdgpeakcall,bdgrep,randsample,bdgdiff,bdgbroadcast]

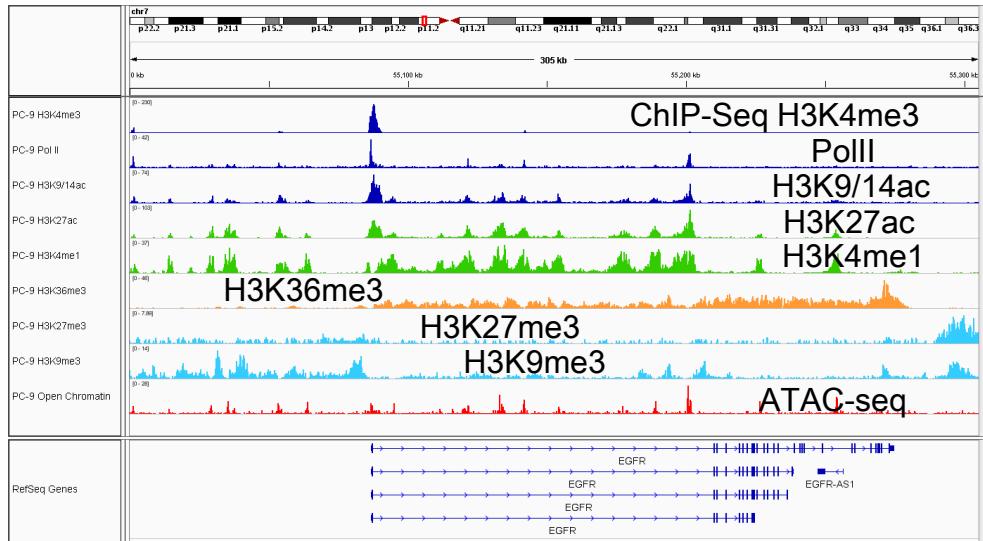
Example for regular peak calling:
  macs2 callpeak -t ChIP.bam -c Control.bam -f BAM -g hs -n test -B -q 0.01

Example for broad peak calling:
  macs2 callpeak -t ChIP.bam -c Control.bam --broad -g hs --broad-cutoff 0.1

There are seven major functions available in MACS serving as sub-commands.

callpeak: Main MACS2 Function to Call peaks from alignment results.
bdgpeakcall: Call peaks from bedGraph output.
bdgbroadcast: Call broad peaks from bedGraph output.
bdgdiff: Deduct noise by comparing two signal tracks in bedGraph.
bdgrep: Differential peak detection based on paired four bedGraph files.
filterdup: Remove duplicate reads at the same position, then convert acceptable format to BED format.
predictd: Predict d or fragment size from alignment results.
pileup: Pileup aligned reads with a given extension size (fragment size or d in MACS language). Note there will be no step for duplicate reads filtering or sequencing depth scaling, so you may need to do certain post-processing.
randsample: Randomly sample number/percentage of total reads.
refinepeak: (Experimental) Take raw reads alignment, refine peak summits and give scores measuring balance of forward- backward tags. Inspired by SPP.

We only cover 'callpeak' module in this document. Please use 'macs2 COMMAND -h' to see the detail description for each option of each module.
```



```
gw.ddbj.nig.ac.jp - Tera Term VT
File Edit Setup Control Window Help
[lect01@1260 ~]$ macs14
Usage: macs14 <-t tfile> [-n name] [-g genomysize] [options]
Example: macs14 -t ChIP.bam -c Control.bam -f BAM -g hs -n test -w --call-subpeaks

macs14 -- Model-based Analysis for ChIP-Sequencing
$ macs14
```

Options:

- version show program's version number and exit
- h, --help show this help message and exit.
- t TFILE, --treatment=TFILE ChIP-seq treatment files. REQUIRED. When ELANDMULTIPET is selected, you must provide two files separated by comma, e.g. s_1_1_eland_multi.txt,s_1_2_eland_multi.txt
- c CFILE, --control=CFILE Control files. When ELANDMULTIPET is selected, you must provide two files separated by comma, e.g. s_2_1_eland_multi.txt,s_2_2_eland_multi.txt
- n NAME, --name=NAME Experiment name, which will be used to generate output file names. DEFAULT: "NA"
- f FORMAT, --format=FORMAT Format of tag file, "AUTO", "BED" or "ELAND" or "ELANDMULTI" or "ELANDMULTIPET" or "ELANDEXPORT" or "SAM" or "BAM" or "BOWTIE". The default AUTO option will let MACS decide which format the file is. Please check the definition in NORFADMF file if you choose F

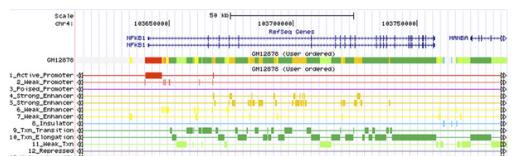
(<https://github.com/taoliu/MACS>)

細胞工学別冊「次世代シーケンサー DRY解析教本」秀潤社 参考

ChIP-Seq → ChromHMM

ChromHMM: Chromatin state discovery and characterization
<http://compbio.mit.edu/ChromHMM/>

ChromHMM: Chromatin state discovery and characterization



ChromHMM is software for learning and characterizing chromatin states. ChromHMM can integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to discover de novo the major re-occurring combinatorial and spatial patterns of marks. ChromHMM is based on a multivariate Hidden Markov Model that explicitly models the presence or absence of each chromatin mark. The resulting model can then be used to systematically annotate a genome in one or more cell types. By automatically computing state enrichments for large-scale functional and annotation datasets ChromHMM facilitates the biological characterization of each state. ChromHMM also produces files with genome-wide maps of chromatin state annotations that can be directly visualized in a genome browser.

- [ChromHMM software v1.11 \(version log\)](#)
- [ChromHMM manual](#)

Quick instructions on running ChromHMM

1. Install Java 1.5 or later if not already installed.
2. Unzip the file ChromHMM.zip
3. To try out ChromHMM learning a 10-state model on the sample data enter from a command line in the directory with the ChromHMMjar file the command:

```
java -mx1600M -jar ChromHMMjar LearnModel SAMPLERDATA_HG18 OUTPUTSAMPLE 10 hg18
```

After termination in ~5-10 minutes a file in OUTPUTSAMPLE/webpage_10.html will be created showing output images and linking to all the output files created. If a web browser is found on the computer the webpage will automatically be opened in it.

In general binarized input for the *LearnModel* command can be generated by first running the *BinarizeBed* command on bed files with coordinates of aligned reads or the *BinarizeBam* command on bam files with the coordinates of aligned reads.

New in version 1.11: ChromHMM has a *BinarizeBam* command which allows binarizing bam files of aligned reads.

New in version 1.10: ChromHMM has the option for parallel training with multiprocessors leading to significantly reduced training times. Add the '-p 0' option to the *LearnModel* command to have ChromHMM to try to use as many processors as available or specify the maximum it should use.

- The ChromHMM software is described in: Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9:215-216, 2012.
- Here are links to some existing ChromHMM annotations in hg19 available for 127 Reference Epigenomes (Roadmap Epigenomics), 9-ENCODE cell types (from Ernst et al. *Nature* 2011) and 6-ENCODE cell types (from ENCODE Integrative Analysis).
- Contact: Jason Ernst (jason.ernst@ucdavis.edu) with any questions, comments, or bug reports.
- Subscribe to a [mailing list for announcements of new versions](#).
- ChromHMM is released under a [GPL 3 license](#).
- ChromHMM source code is available on GitHub [here](#).
- Funding for ChromHMM provided by NSF Postdoctoral Fellowship 0905968 to JE and grants from the National Institutes of Health (NIH 1-RC1-HG005334 and NIH 1 U54 HG004570).

```
$ cd  
$ cp /home/lect01/ChromHMM/LC2ad_dense.bed .
```

でホームディレクトリにLC2/adのChromHMMの結果がコピーされます。
WinSCPでPCIにダウンロードして、IGVで見てみてください。

様々なヒストン修飾のChIP-Seqデータなどから、
クロマチンの状態をパターン化してくれるソフトウェア

Ernst J and Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9:215-216, 2012.

ENCODEやRoadmap Epigenomics projectでも ChromHMMデータが公開されている。

Session 2

データベース・新技術の紹介

The screenshot shows the DBTSS (Database of Transcriptional Start Sites) homepage. At the top, it says "Release 9.0 Updated (July 9, 2015) Based on UCSC hg38, mm10". The main header features the text "~-Database of Transcriptional Start Sites~-" and "DBTSS" in large, stylized letters. Below the header, there's a "Database Search" form with fields for "Species" (set to "H. sapiens"), "Keyword" (set to "NM_*"), and "Lift over" (set to "hg38"). A "Search" button is present. To the right of the search form, there's a "About this database" section with a detailed description of the database's purpose and content. This section includes a paragraph about the database's history and its integration with other genomic data. Further down, there's a "Human Chromatin Features" section. On the far right, there's a "News" sidebar with a list of recent updates, such as "09 Jul 2015: New T helper cell data of mouse are now available" and "30 Jun 2015: New BRIC-Seq data (UPF1 RNA) are now available".

データベースDBTSS

DBTSS (<http://dbtss.hgc.jp/>)

The screenshot shows the DBTSS homepage with a green header bar containing the title "DBTSS Home". Below the header is a search bar with the URL "dbtss.hgc.jp". The main content area features the DBTSS logo and the text "- DataBase of Transcriptional Start Sites -". A banner indicates "Release 9.0 Updated (July 9, 2015) Based on UCSC hg38, mm10". On the left, there is a "Database Search" sidebar with fields for "Species" (set to "H. sapiens"), "Keyword" ("NM_*"), and "Lift over" ("hg38"). The main content area includes sections for "About this database" (describing the database's purpose and data), "Human Chromatin Features" (linking to external resources), and a "News" section listing recent updates.

About this database

Welcome to DBTSS (DataBase of Transcriptional Start Sites).

To support transcriptional regulation studies, we have constructed the DBTSS (DataBase of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

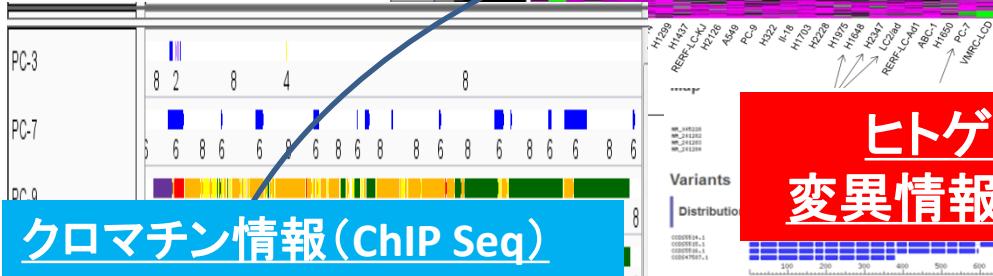
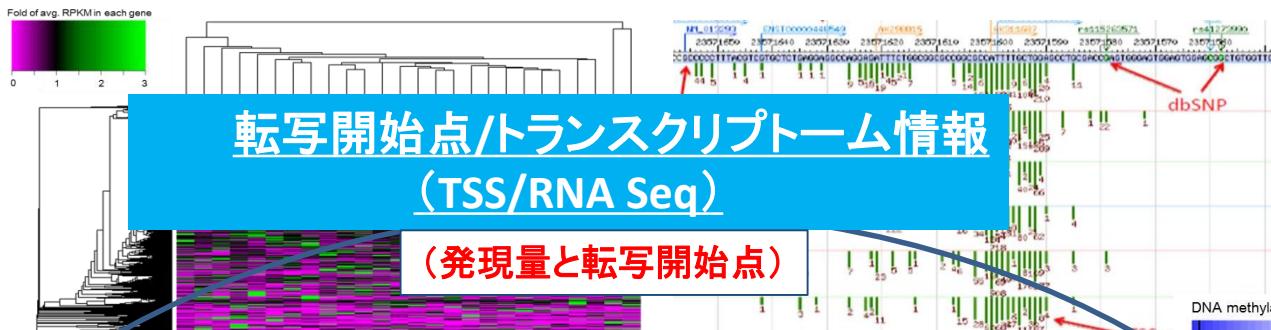
This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS tag sequences for collected from a total of 20 tissues and 7 cell cultures. We also integrated our newly generated RNA-seq data of subcellular-fractionated RNAs and ChIP-seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project.

In this update, we further associated those TSS information with public and original SNV data, in order to

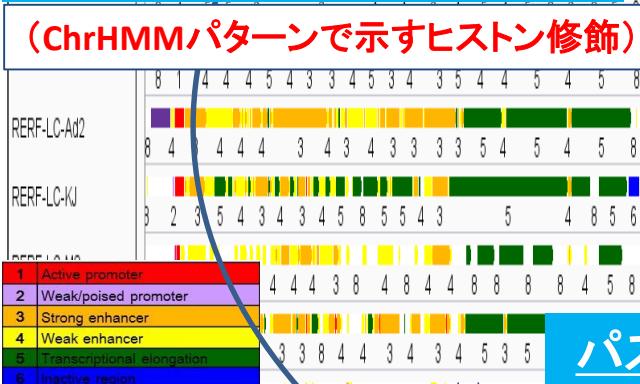
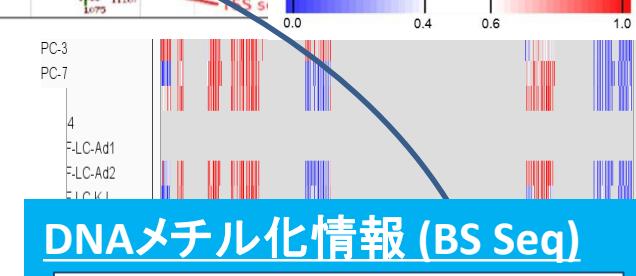
News

- 09 Jul. 2015: New T helper cell data of mouse are now available. Raw data accession: [DRA000928](#), [DRA001102](#) and [SRP007894](#) (Genome Shien).
- 30 Jun. 2015: New BRIC-Seq data (UPF1 RNAi) are now available. Raw data accession: [DRA000591](#) (Genome Shien).
- 15 Sep. 2014: [New DBTSS](#) opened.

ヒト応用研究を志向したオミクス情報の統合



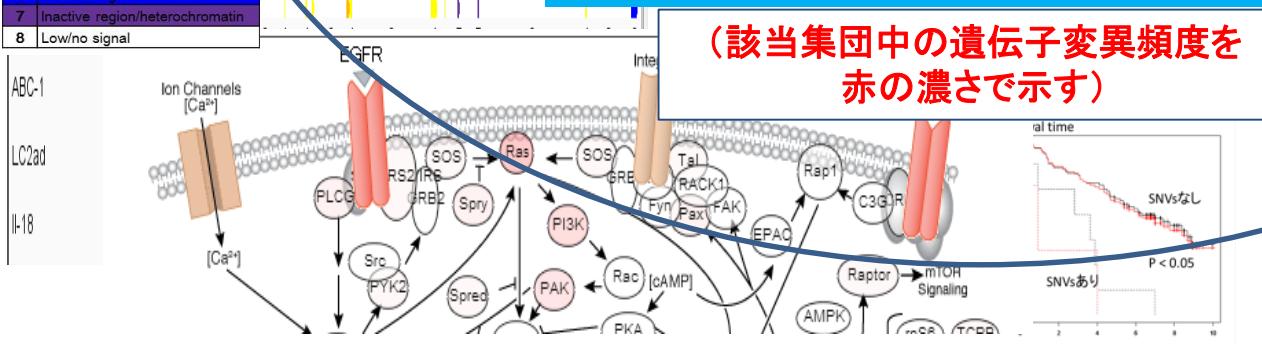
ヒトゲノム 変異情報の統合



(それぞれの検体での変異部位)

パスウェイマップ(文献情報)からの検索

(該当集団中の遺伝子変異頻度を
赤の濃さで示す)



データコンテンツ

Topページ → “Data Contents”

The screenshot shows the DBTSS Home page with a sidebar on the left containing links for 'Lung Adenocarcinoma', 'SNV Summary in Cancers', 'Pathway Map', and 'Documents'. The main content area displays a contact form and a message about genome-wide characterization of transcriptional start sites in humans by integrative transcriptions. A logo for NEDO and the Japanese Ministry of Education, Culture, Sports, Science and Technology is visible.

Data contents

Number of dataset

	Human	Mouse	Malaria	Chyzon	Rat	Chimpanzee	Macaque
TSS-seq	73	7	1	1	1	1	2
RNA-seq	42	0	0	0	0	0	0
ChIP-seq	255	0	0	0	0	0	0
RIP-seq	12	0	0	0	0	0	0
BS-seq	26	0	0	0	0	0	0
ChromHMM	36	0	0	0	0	0	0
SNV	49	0	0	0	0	0	0

Data contents

Cell lines													Memo	
Cell line		General information*					Cell culture		Sequencing dataset					
Name	Type	in-house analysis number	Distributor	Catalogue number	Ethnicity	Gender	Age	Smoking status	Medium	Dish	Whole-genome Seq	RNA-Seq	BS-Seq	TSS-Seq
SAEC (control)	Small airway epithelial cell	s_35	TAKARA (Lonza)	-	-	M	-	-	-	collagen Type I-coated	-	♦	-	♦
PC-9	lung adenocarcinoma	s_9	RIKEN BRC	RCB4455	Japanese	-	-	-	RPMI	-	♦	♦	♦	♦
PC-14	lung adenocarcinoma	s_10	IBL	-	Japanese	-	-	-	RPMI	-	♦	♦	♦	♦
RERF-LC-KJ	lung adenocarcinoma	s_13	RIKEN BRC	RCB1313	Japanese	M	78	-	RPMI	-	♦	♦	♦	♦
RERF-LC-Ad1	lung adenocarcinoma	s_11	JCRB	JCRB1020	Japanese	M	70	-	RPMI	-	♦	♦	♦	♦
RERF-LC-Ad2	lung adenocarcinoma	s_12	JCRB	JCRB1021	Japanese	M	-	-	RPMI	-	♦	♦	♦	♦
LC2/ad	lung adenocarcinoma	s_18	RIKEN BRC	RCB0440	Japanese	F	51	-	DMEM	collagen Type I-coated	♦	♦	♦	♦
RERF-LC-MS	lung adenocarcinoma	s_14	JCRB	JCRB0081	Japanese	-	-	-	EMEM	-	♦	♦	♦	♦
VMRC-LCD	lung adenocarcinoma	s_16	JCRB	JCRB0814	Japanese	M	-	-	EMEM	-	♦	♦	♦	♦
ABC-1	lung adenocarcinoma	s_17	JCRB	JCRB0815	Japanese	M	47	-	EMEM	-	♦	♦	♦	♦
PC-7	lung adenocarcinoma	s_8	IBL	-	Japanese	-	-	-	RPMI	-	♦	♦	♦	♦
PC-3	lung adenocarcinoma	s_7	JCRB	JCRB0077	Japanese	F	48	-	RPMI	collagen Type I-coated	♦	♦	♦	♦
II-18	lung adenocarcinoma	s_19	RIKEN BRC	RCB2093	Japanese	-	-	-	RPMI	-	♦	♦	♦	♦
RERF-LC-OK	lung adenocarcinoma	s_15	JCRB	JCRB0811	Japanese	-	-	-	RPMI	-	♦	♦	♦	♦
A549	lung adenocarcinoma	s_1	ATCC	CCL-185	Caucasian	M	58	-	DMEM	-	♦	♦	♦	♦
A427	lung adenocarcinoma	s_20	ATCC	HTB-53	Caucasian	M	52	-	RPMI	-	♦	♦	♦	♦
H322	lung adenocarcinoma	s_21	ATCC	CRL-5806	Caucasian	-	-	-	RPMI	-	♦	♦	♦	♦
H1648	lung adenocarcinoma	s_25	ATCC	CRL-5882	Black	M	39	Y	RPMI	collagen Type I-coated	♦	♦	♦	♦
H1650	lung adenocarcinoma	s_26	ATCC	CRL-5883	Caucasian	M	27	Y	RPMI	-	♦	♦	♦	♦
H1975	lung adenocarcinoma	s_29	ATCC	CRL-5908	-	F	-	N	RPMI	-	♦	♦	♦	♦

TSS-Seq, ChIP-Seq(ヒストン修飾), RNA-Seq, RIP-Seq, BS-Seq, SNPs/SNVs data

細胞株を中心に。

ヒトの多型・変異情報も格納(>15000)

Cancer genomes

ICGC: International Cancer Genome Consortium

TCGA: The Cancer Genome Atlas

Gastric adenocarcinoma, Urothelial bladder carcinoma, Glioblastoma,
Clear cell renal cell carcinoma, Endometrial carcinoma,
Acute myeloid leukemia, Breast tumors, Squamous cell lung cancers,
Colorectal cancer, Ovarian carcinoma...

National Cancer Center Hospital East:

Lung adenocarcinoma (2013 *PLoS ONE*) ← NBDCヒトデータベース: JGA00000000001

Small cell lung cancer (2014 *J Thorac Oncol*)

Others:

Myelodysplasia (2011 *Nature*), Clear-cell renal cell carcinoma (2013 *Nat Genet*),

Lung adenocarcinoma (2012 *Cell*)...

Japanese genomes (SNPs)

HGVD: The Human Genetic Variation Database

ToMMo: Tohoku Medical Megabank Organization

JPDSC: The Japan PGx Data Science Consortium

Release 9.0 Updated July 9, 2015
Based on UCSC hg38 mm10

Top

Database Search

About this database

Welcome to DBTSS (Database of Transcriptional Start Sites). To support transcriptional regulation studies, we have constructed the DBTSS (Database of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS tag sequences for collected from a total of 20 different cell types. We also integrated our newly developed TSS-seq data of a subset of tissues, fragmentary TSS-seq data of histone modifications, RNA polymerase II, and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external genomic data, such as chromatin map of the ENCODE project.

In this update, we further associated those TSS information with public and original SNV data, in order to identify single nucleotide variations (SNVs) in the regulatory regions.

It is believed that single nucleotide variations (SNVs) in the transcriptional regulatory regions are responsible for many human diseases, including cancers. However, it remains difficult to identify functionally relevant SNVs from those having no evident biological consequences. In this version of DBTSS, we attempted to associate SNVs with biological inferences of transcriptional regions. We used SNVs which we identified from genomic analyses of various types of cancers, including somatic mutations of 100 lung adenocarcinoma and lung small cell carcinoma. For germline variants, we used SNVs in dbSNP as well as our unique dataset obtained in the TSS-seq analysis. We also integrated those SNV data with our original datasets of TSS-seq, RNA-seq, ChIP-seq of representative histone modifications and Bisulfite Sequencing of cytosine methylation of DNA. Particular, we present multi-omics data of 26 lung adenocarcinomas cells and 26 lung small cell carcinomas (TSS-seq, RNA-seq, ChIP-seq and TSS tag) and perform sequence analysis of these samples and their corresponding materials.

We further connected the multi-omics data of model organisms by genome-genome alignment. We provide a unique data resource to investigate what genomic features are observed in a particular genomic coordinates in a wide variety of samples.

These data can be browsed in our new viewer which also supports versatile search conditions of users. We believe new DBTSS is helpful to understand biological consequences of the massively identified TSSs and identify human genetic variations which are associated with diseased transcriptional regulations.

Human Chromatin Features

Search

Search from Genomic Position:
chr1:75:767,000

Search

Search from SNP (dbSNP rsID):
rs1023965

Search

Search from SNV (COSMIC:
somatic mutation):
BSAF

Search

Search from SNV-enriched Gene
in Cancers

Lung adenocarcinoma

Search

SNV Summary in Cancers

SNV

Search

References

Suzuki A, Wakiguni H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, Suzuki Y, Nakai K. DBTSS as an integrative platform for transcriptome-wide genomic sequence variation data. *Nucleic Acids Res* 2015; (Database): D11-D16.

Suzuki A, Mizaki S, Yamane H, Kawase A, Matsushima K, Suzuki M, Sugano S, Esaki H, Suzuki Y, Tsuchihara K. Identification and characterization of mutations in Japanese lung adenocarcinoma without sequenced tissue counterparts. *PLoS One*. 2013; See 12(6).

Yamashita R, Saitoh NP, Kanai A, Tamimoto K, Arauchi T, Tanaka Y, Sugano S, Esaki H, Suzuki Y, Tsuchihara K. Identification and characterization of mutations in Japanese lung adenocarcinoma without sequenced tissue counterparts. *PLoS One*. 2013; See 12(6).

Tsuchihara K, Suzuki Y, Wakiguni H, Itoh T, Tanimoto K, Hashimoto M, Matsushima K, Mizaki S, Sugano S, Arauchi T, Tanaka Y, Sugano S. (2009) Massive transcriptional start site analysis of hypoxia cells. *Nucleic Acids Res* 2009 Feb 22;

[Japanese]
[Database Manual] (Yodoh)

Contact us

We welcome your comments and feedback about our database.
Please feel free to contact us... yazuh@ims.u-tokyo.ac.jp

Copyright © 2015

Pathway Map

Species:
H. sapiens

Search

Documents

- Experimental Procedures

- Data Contents

- Help

- Download

- Previous version

Release 9.0 Updated July 5, 2016
Based on UCSC hg38 mm10

Top

Database Search

Species:
H. sapiens

Keyword:
NM_*

Search

Utr over
hg38

chr1:99,950,000-100,050,000

Search

News

- 09-Jul-2015 New T helper cell mouse of now available. Raw data accession: DBA000262, DBA001112 and DBA002053 (Genome Sheet).
- 30-Jun-2015 New BRS-Seq data (U1FF RNA) are now available. Raw data accession DBA00581 (Genome Sheet).
- 15-Sep-2014 New DBTSS opened.

Search

Font size: 15 Change font size

Show 50 entries

Term: T

Genomic position:

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

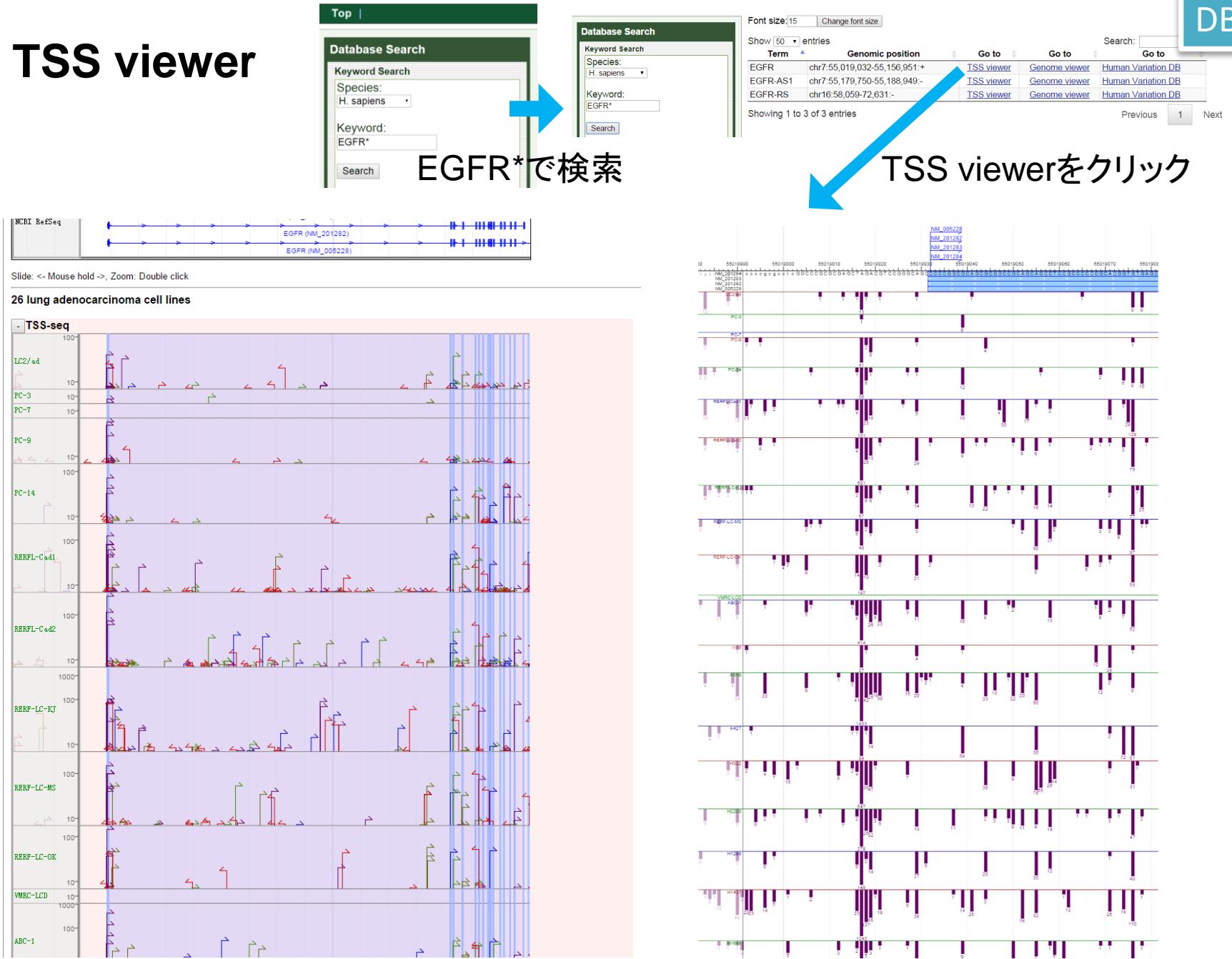
Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB

Go to: TSS viewer Genome viewer Human Variation DB</p

TSS viewer



Genome viewer

DBTSS

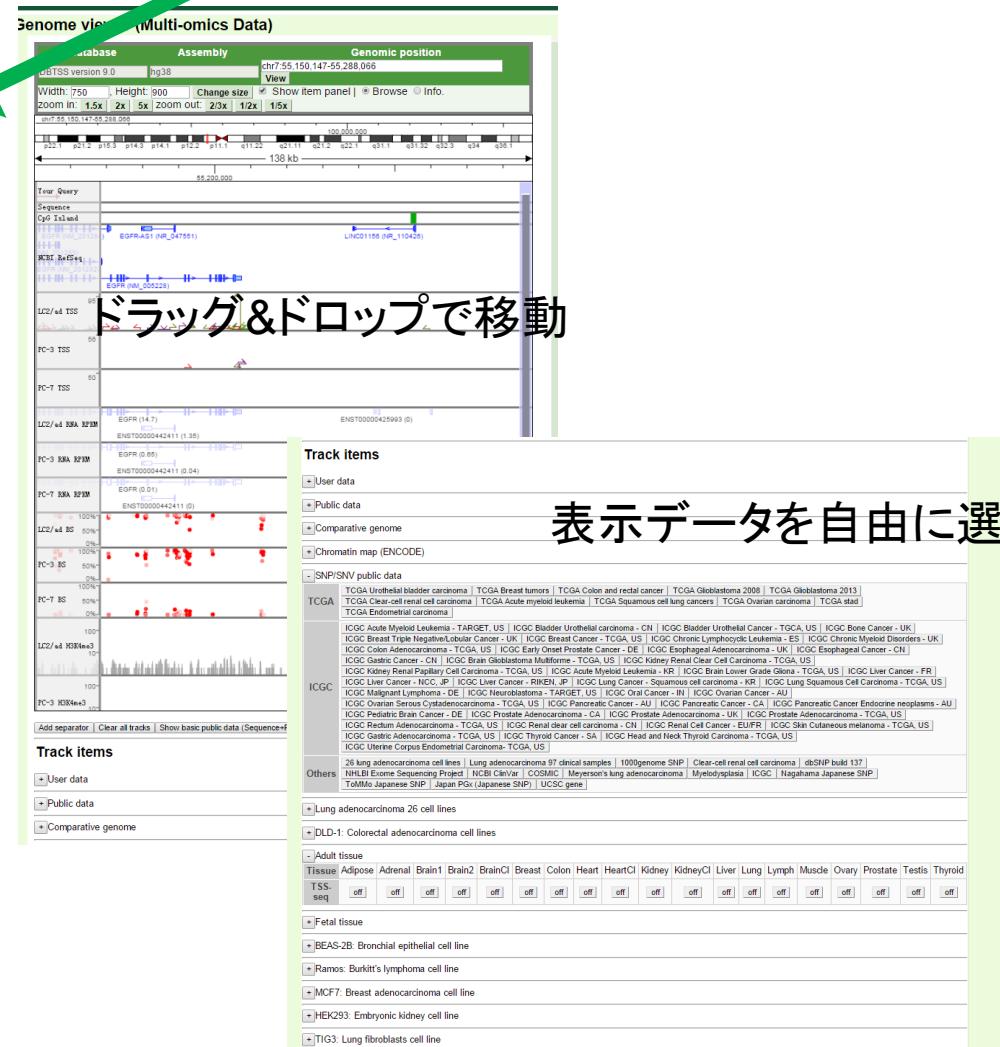
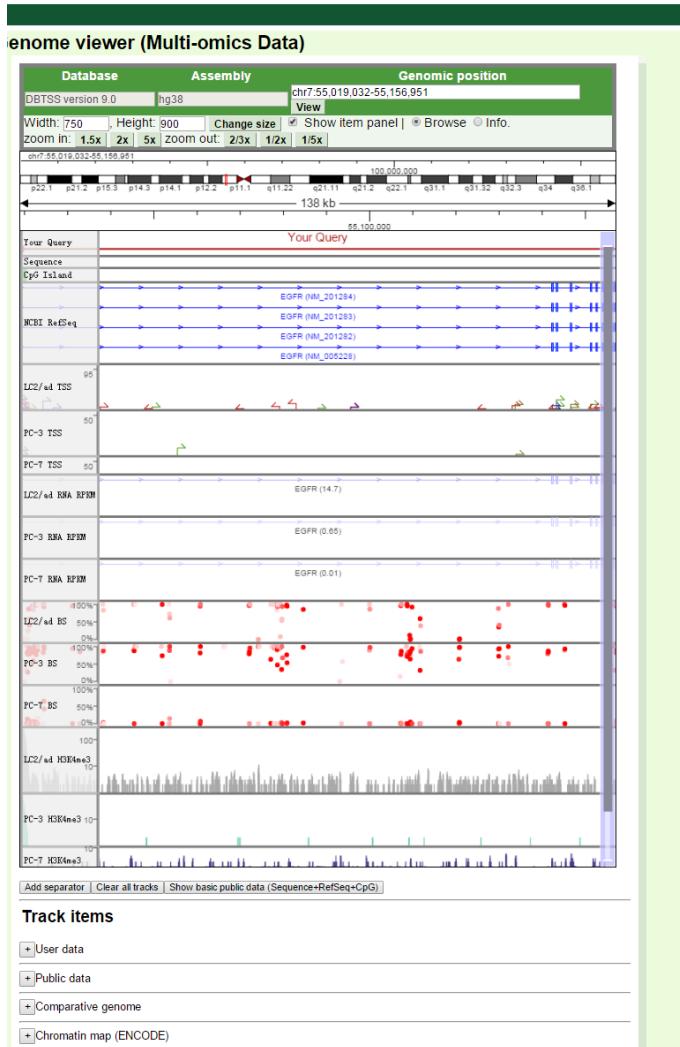
The screenshot shows the DBTS interface. On the left, a green arrow points from the 'Database Search' section to the 'Database Search' panel. The 'Database Search' panel has a 'Keyword Search' section with 'Species: H. sapiens' and 'Keyword: EGFR*'. Below it is a 'Search' button. The main area shows a table of genomic entries:

Term	Genomic position	Go to	Go to	Go to
EGFR	chr7:55,019,032-55,156,951:+	TSS viewer	Genome viewer	Human Variation DB
EGFR-AS1	chr7:55,179,750-55,186,943:-	TSS viewer	Genome viewer	Human Variation DB
EGFR-RS	chr16:58,059-72,631:-	TSS viewer	Genome viewer	Human Variation DB

At the bottom, there are links for 'Previous' and 'Next' entries, and a large green arrow points from the table towards the 'Genome viewer' link.

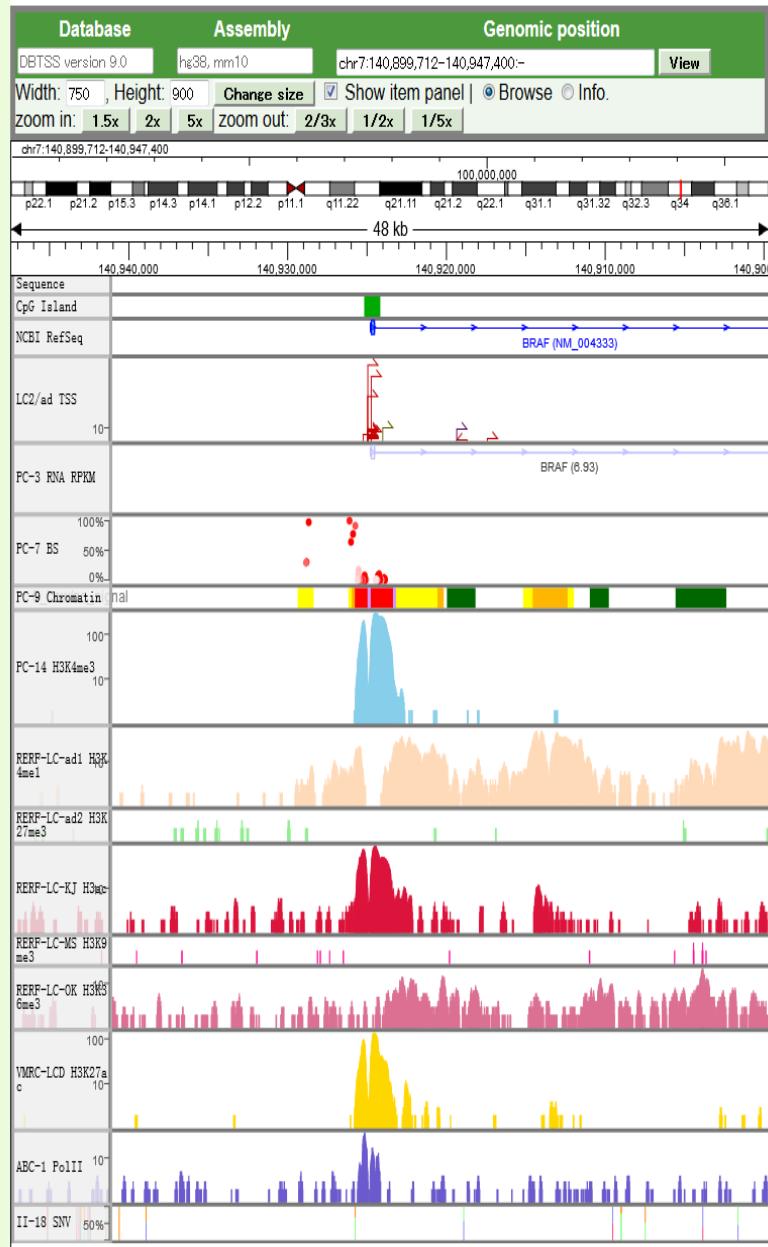
EGFR*で検索

Genome viewerをクリック



Genome viewer (BRAF遺伝子を例に)

Genome viewer (Multi-omics Data)



Gene model

TSS-Seq

RNA-Seq

BS-Seq

ChromHMM

H3K4me3

H3K27me3

H3ac

H3K9me3

H3K36me3

H3K9ac

H3K27ac

Pol II

SNV

ChIP-Seq

Pathway Map

Top |

Database Search

Species: H. sapiens

Keyword: NM_*

Search

Lift over: hg38

chr1:99,950,000-100,050,000

Human Chromatin Features

Search from Genomic Position: chr1:75,787,000

Search

Search from SNP (dbSNP rsID): rs375229869

Search

Search from SNV (COSMIC: somatic mutation): BRAF

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma ...

Search

Search from SNV-enriched Gene in Cancers: Lung adenocarcinoma ...

Search

SNV Summary in Cancers: NM_*

Search

Pathway Map

Species: H. sapiens

Search

Documents

- Experimental Procedures
- Data Contents

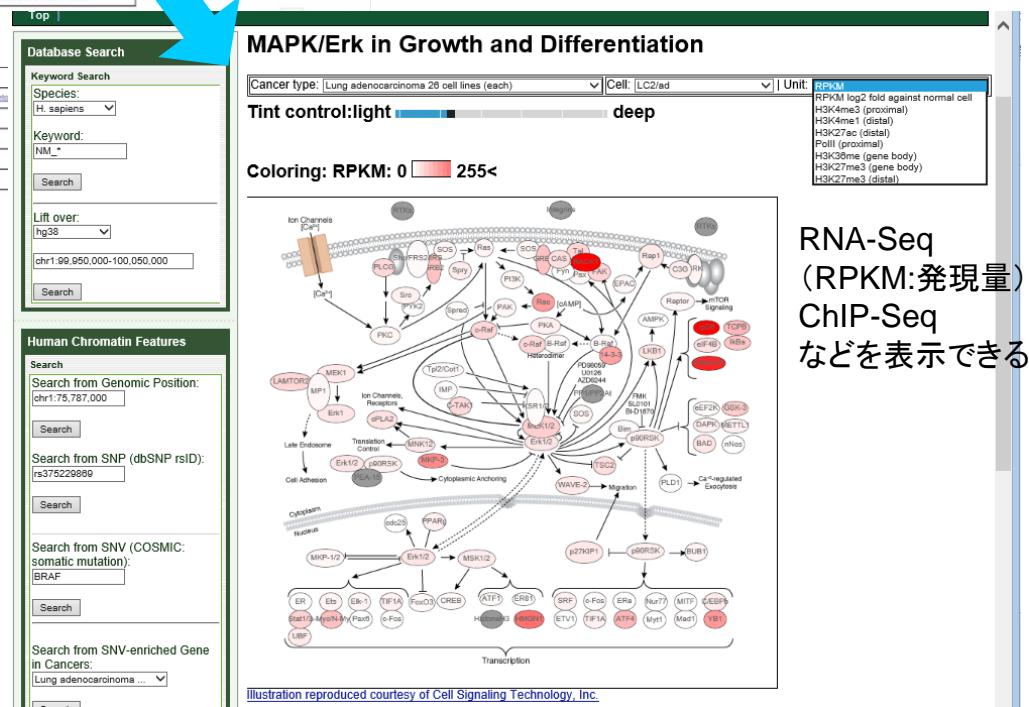
Release 9.0 Updated (July 9, 2015)
Based on UCSC hg18, mm10

CSTやKEGGのpathwayリスト

MAPK/Erk in Growth and Differentiationをクリック



Mapの
Searchを
クリック



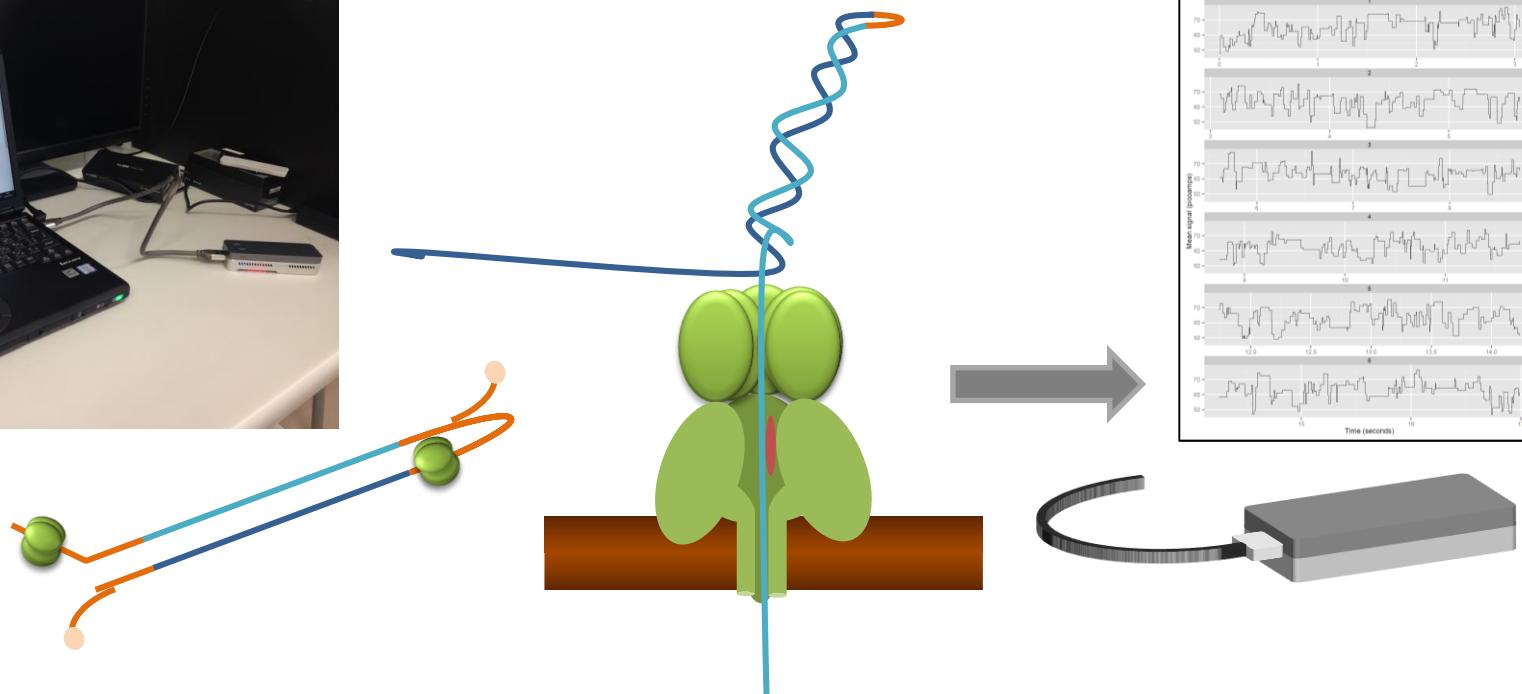
ぜひDBTSSを使ってください

ポータブル&ロングリードシークエンス

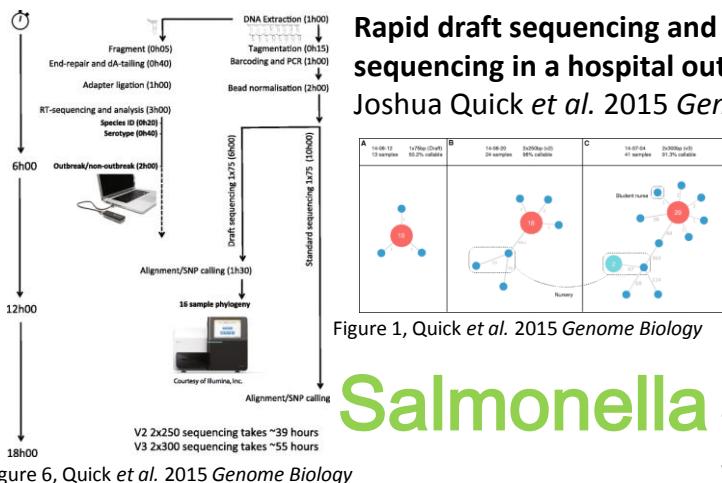
MinION (Oxford Nanopore Technologies)

- ナノポアシークエンサー
- 安価な初期投資 (\$1,000スターターキット)
- ポータブル (103g: flow cell込)
- ロングリード
- 10,000-100,000リード

<https://www.nanoporetech.com/>



ナノポアシークエンサーを用いた研究



Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*
Joshua Quick et al. 2015 *Genome Biology*

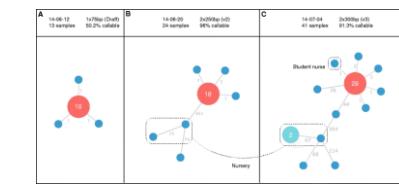


Figure 1, Quick et al. 2015 *Genome Biology*

Salmonella

MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island

Philip M Ashton et al. 2015 *Nature Biotechnology*



Figure 2, Ashton et al. 2015 *Nature Biotechnology*

Nanopore sequencing detects structural variants in cancer

Alexis L Norris et al. 2016 *Cancer Biology & Therapy*

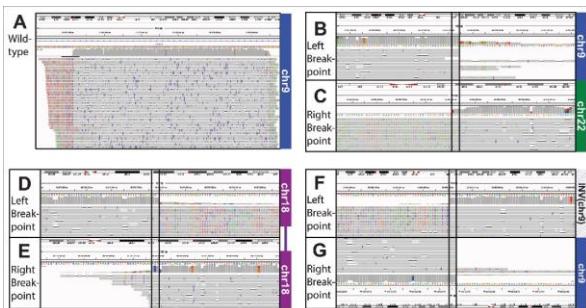


Figure 3, Norris et al. 2016
Cancer Biology & Therapy

Structural variants in cancer

Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis
Alexander L. Greninger et al. 2015 *Genome Medicine*

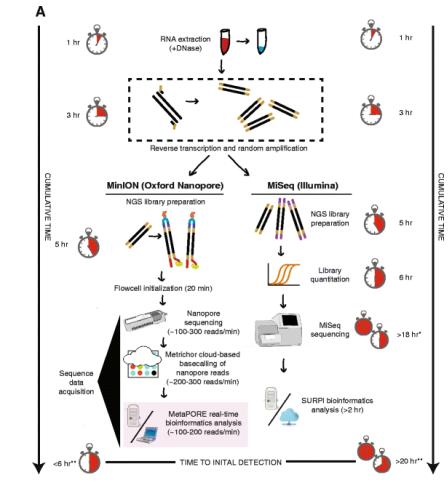


Figure 1, Greninger et al. 2015 *Genome Medicine*

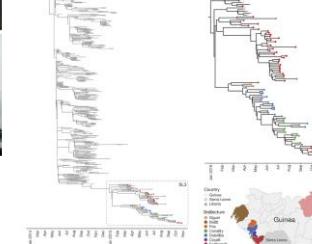


Figure 2, Quick et al. 2016 *Nature*

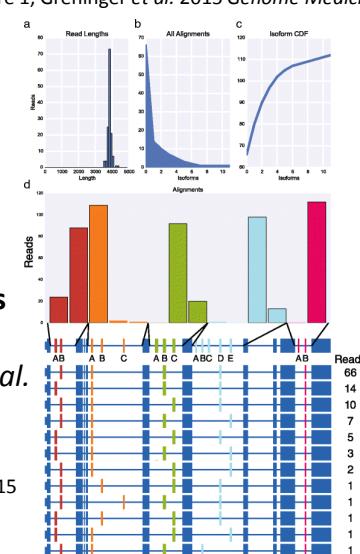


Figure 3, Bolisetty and Rajadinakaran et al. 2015
Genome Biology

Splicing patterns

Determining exon connectivity in complex mRNAs by nanopore sequencing

Mohan T. Bolisetty and Gopinath Rajadinakaran et al.
2015 *Genome Biology*

合成ロングリードシークエンス GemCode (10x Genomics)

Linked read @ GemCode技術

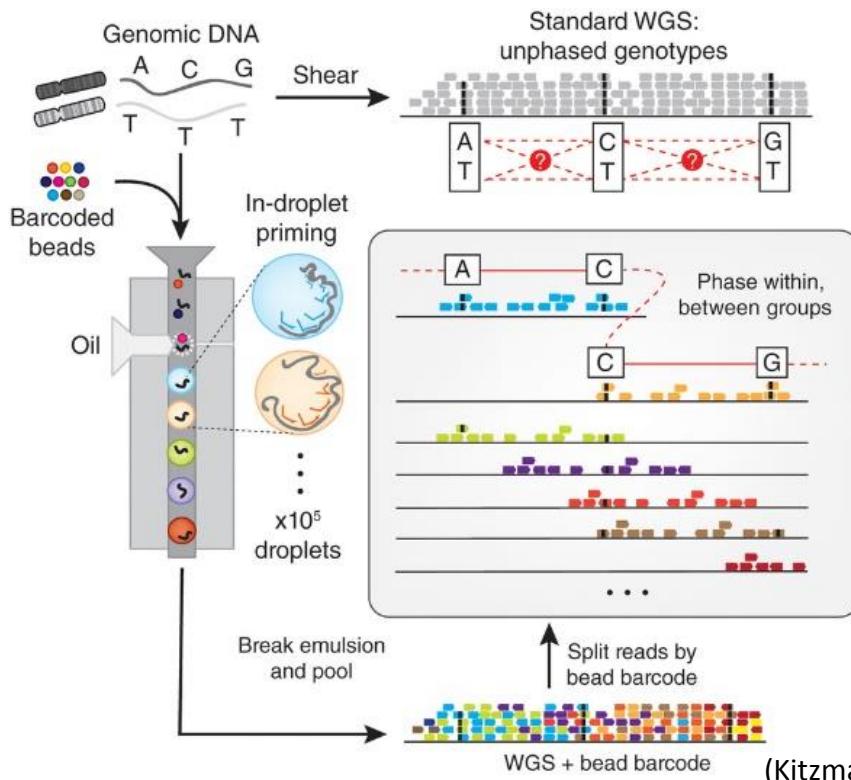
分子バーコーディング + ショートリードシークエンシング

- 750,000 beads
- Long genomic DNA fragments (~ 100 kb)

Chromium@10X Genomics

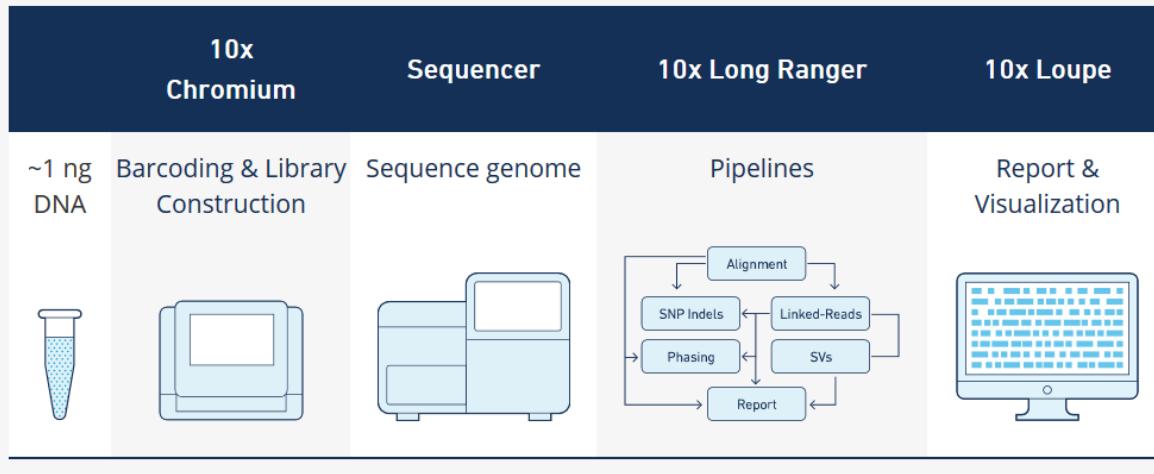


<http://www.10xgenomics.com/instrument/>

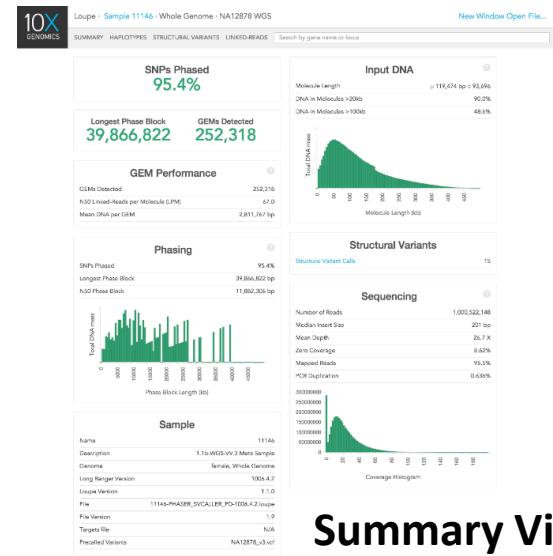
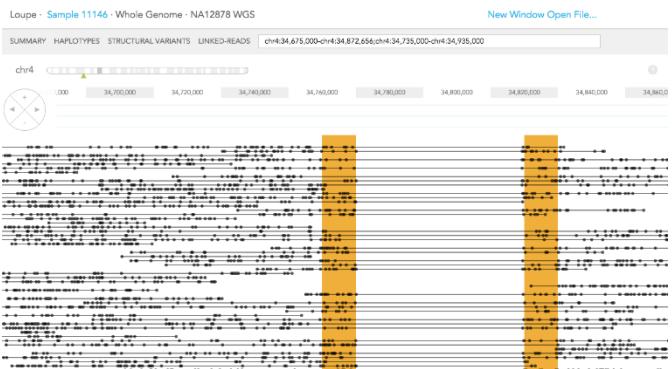


GemCode@10x Genomics

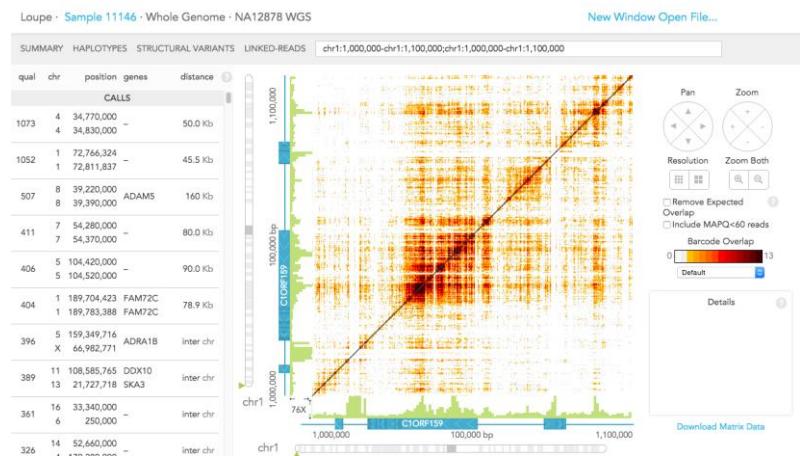
<http://software.10xgenomics.com/genome-exome/visualization/latest/what-is-loupe> より抜粋



Haplotype Browser



Summary View



Linked-Reads

Structural Variants

シングルセルシークエンス

さまざまなプラットフォームが開発されている

C1 (Fluidigm社)

- RNA-Seq
- ATAC-Seq
- Whole-exome sequencing

C1 system@Fluidigm



Nx1-Seq

- RNA-seq

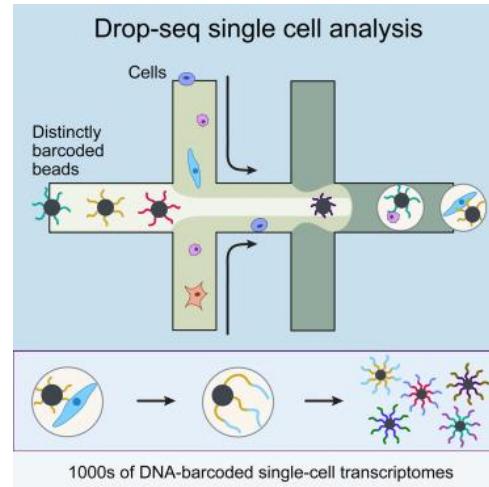
Drop-seq

- RNA-Seq

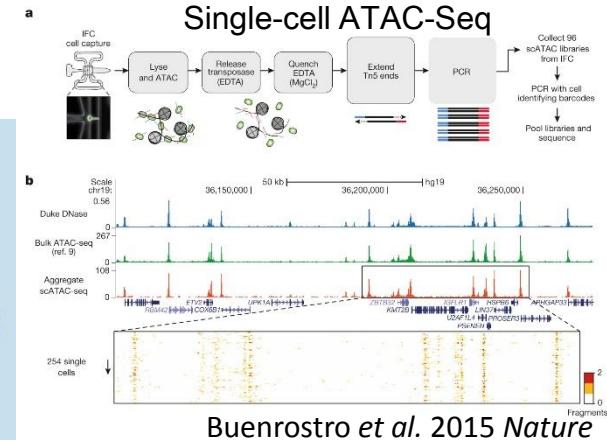
Chromium Single Cell 3' Solution (10x Genomics社)

- RNA-Seq

など



Macosko et al. 2015 Cell



Chromium@10x Genomics



謝辞

国立遺伝学研究所
遺伝研スーパーコンピュータシステムのご担当者様
講習会用のアカウントをご用意いただきました。

東京女子医科大学のご担当者様
スパコン接続のためにご助力くださりありがとうございました。

ありがとうございました
ご質問・コメント等がありましたら
遠慮なくお願ひします。