

# **NGS 基本フォーマットとツール 復習と補足**

基礎生物学研究所  
ゲノムインフォマティクストレーニングコース  
内山 郁夫 (**uchiyama@nibb.ac.jp**)

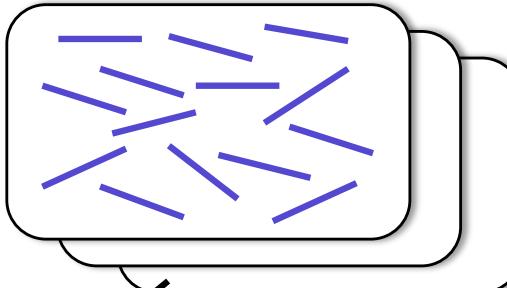
# ショートリードのマッピング

ゲノム配列  
(リファレンス reference 配列)

形式(配列)

```
>chr
AGCTTTCATCTGACTGCAACGGCAATATGTCT
CTGTGTGGATAAAAAAAGAGTGTCTGATAGCAGC
TTCTGAACGTGTTACCTGCCGTGAGTAAATTAAAA
TTTATTGACTTAGGTCACTAAATACTTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAATTACAG
AGTACACAAACATCCATGAAACGCATTAGCACCACC
ATTACCACCAACCATTACACAGGTAAACGG
```

サンプル(ゲノムDNA／RNA)  
(リード read 配列)

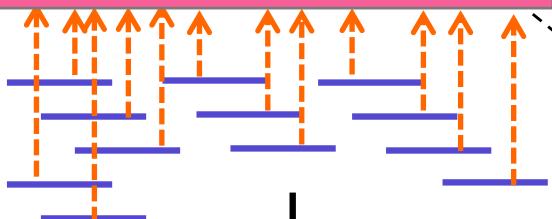


形式

(配列 + クオリティ値)

```
@SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:
ATCCGGCTGGCGACCGACTATGTTCCGGCGAATACAAGCTGG
+SRR1515276.1 HWI-ST808:151:D2D13ACXX:2:1207:
@@@AD>DDFF7DC?FFEBF@DFII<DF@AAA6AEFBDBDCA?>A?
@SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:
CACCGTGTAGTACCAAGCATCCTGCGTACAATCAGCAATCCCAGTC
+SRR1515276.2 HWI-ST808:151:D2D13ACXX:2:1207:
CCCCFDFFHDFFHIIIEGIHJJJJGFHGGHGGHGGIJJJDGIJHH
@SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:
CAGGACATCGCCTTGATCGGTTCAGACTTCGGACCAACCTGCA
+SRR1515276.3 HWI-ST808:151:D2D13ACXX:2:1207:
CCCCFFDFAFHFHIJGHIIJJHEHIIJGHIFEHIIIA@FIFH
```

リファレンス配列へのマッピング



クオリティチェック  
アダプター除去

形式(遺伝子アノテーション)

chr	RefSeq	start_codon	190	192	1.000	+	.	gene_id "b0001
chr	RefSeq	CDS	190	252	1.000	+	0	gene_id "b0001
chr	RefSeq	stop_codon	253	255	1.000	+	.	gene_id "b0001
chr	RefSeq	exon	190	255	1.000	+	.	gene_id "b0001

形式(マッピング結果)

```
@HD VN:1.0 SO:unsorted
@SQ SN:chr LN:4639675
@PG ID:bowtie2 PN:bowtie2 VN:2.2.4 CL:"/bio/bin/bowtie2-alig
SRR1515276.40 0 chr 4423609 42 51M * 0 0 GGAATTCCCTCACTGCCA
SRR1515276.158 16 chr 501700 42 51M * 0 0 ACGCACCGAGTGCAGAG
SRR1515276.212 4 * 0 0 * 0 0 GGCGCGCTTCAGCGTGT
SRR1515276.319 0 chr 2922768 42 51M * 0 0 GCTTAAGTTGATTAAGG
SRR1515276.367 16 chr 2753873 42 51M * 0 0 GCGTGTCCGTCCGCAGC
SRR1515276.411 0 chr 3440721 42 51M * 0 0 ACGGCATATAATTCTTGA
```

# 復習: **cutadapt**による アダプターの除去

実習用ディレクトリ ~/**data/IU**

入力

- リード配列(FASTQ 形式; paired-end)

**etec\_1.fq**  
**etec\_2.fq**

- アダプター配列 (それを3'端から除去)

Adapter1: AGATCGGAAGAGAGCGGTT

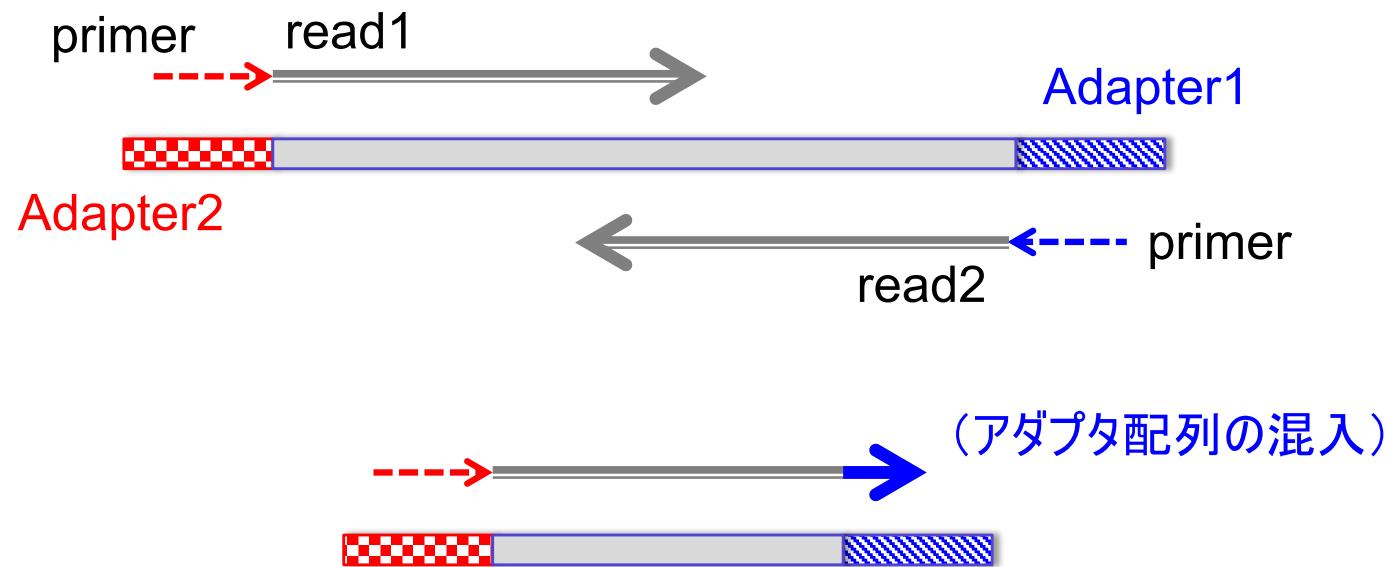
Adapter2: AGATCGGAAGAGAGCGTCG

## ◆ アダプター配列除去の実行

除去後のデータ(FASTQ形式)は **etec\_1.cut.fq**, **etec\_2.cut.fq**とする。

```
$ cutadapt  AGATCGGAAGAGAGCGGTT  AGATCGGAAGAGAGCGTCG  
       etec_1.cut.fq  etec_2.cut.fq  
      etec_1.fq    etec_2.fq
```

# Illuminaにおけるアダプター配列



Adapter1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

Adapter2: AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTA

cutadapt –a (-A) オプションでは、指定した配列とマッチした箇所以降の3'側を切り捨てるので、アダプタ配列は全長を指定しなくてもよい。

# **cutadapt** その他のオプション

- **-q [5' cutoff,] 3' cutoff** (例 : -q 20)
  - クオリティ値が指定したカットオフより低い塩基を3'端から除く(カンマ区切りでカットオフを2つ指定した場合は5'端からも除く)
- **-m min\_length** (例 : -m 30)
  - アダプター除去後の配列長が指定した長さ以下になつたら配列全体を捨てる。
  - ペアエンドの場合、ペアのどちらかが捨てられる場合は両方を捨てる。  
→2つのファイルで対応する配列の出現順が揃うようにする。
- **-O overlap\_length** (例 : -O 5)
  - アダプターとリードとの間で、マッチしたと見なす最低のオーバーラップ長を指定。デフォルトは3。



# 復習:bowtie2用インデックスの作成

実習用ディレクトリ ~/data/IU

入力

- ゲノムデータ (FASTA形式)

eco\_o139.fa 腸管毒素原性大腸菌(ETEC) O139:H28のゲノム配列

- ◆ bowtie2用インデックスの作成 (インデックス名は etec)

\$ bowtie2-build



# 復習: bowtie2の実行 (paired-end)

実習用ディレクトリ ~ / data / IU

入力

- リード配列 (FASTQ 形式; paired-end; アダプターを除去したもの)

etec\_1.cut.fq  
etec\_2.cut.fq

- リファレンス配列のインデックス名

etec (先ほど作ったもの)

- ◆ bowtie2によるマッピングの実行 (出力: etec\_bowtie2.sam)

```
$ bowtie2 [ ] etec [ ] etec_1.cut.fq [ ] etec_2.cut.fq  
[ ] etec_bowtie2.sam
```

# マッピング結果ファイル(SAMファイル)

ヘッダ(@で始まる)

```

@HD VN:1.0 SO:unsorted
@SQ SN:ETEC_chr LN:4979619
@SQ SN:pETEC_80 LN:79237
@SQ SN:pETEC_35 LN:34367
@SQ SN:pETEC_73 LN:70609
@SQ SN:pETEC_6 LN:6199
@SQ SN:pETEC_74 LN:74224
@SQ SN:pETEC_5 LN:5033
@PG ID:bowtie2 PN:bowtie2 VN:2.3.0 CL:"/bio/bin/bowtie2-align-s --wrapper basic-0 -x etec -S etec_bowtie2.sam -1 etec_1.cut.fq -2 etec_2.cut.fq
SRR345261.25 89 ETEC_chr 3758170 1 49M = 3758170 0 ACACGCGGCATGGCTG... ####ED>EBDBDDE,E...
SRR345261.25 133 ETEC_chr 3758170 0 * = 3758170 0 NNNNNNNNNNNNNNNNN... ##########
SRR345261.50 73 ETEC_chr 4361458 1 49M = 4361458 0 CAAGCGTTAACCGAA... :HEGDFHHH@BGG=B...
SRR345261.50 133 ETEC_chr 4361458 0 * = 4361458 0 NNNNNNNNATNNNNNNNN... ##########
SRR345261.75 73 ETEC_chr 4362922 1 49M = 4362922 0 CGGTGGATGCCCTGGC... DDDDBD6<B>DB>BB>1>>
SRR345261.75 133 ETEC_chr 4362922 0 * = 4362922 0 NNNNNNTTNNNTTCGG... ##########
SRR345261.100 73 ETEC_chr 679991 42 49M = 679991 0 GTGGTTAACGAGTCC... GGGGGGGGB=ED=E...
SRR345261.100 133 ETEC_chr 679991 0 * = 679991 0 NNNNNNNCACCGNTAGT... ##########
SRR345261.125 73 ETEC_chr 4376280 42 49M = 4376280 0 CTCAGGATGAGGGTCA... EEE=B<@BDEDEDE...
SRR345261.125 133 ETEC_chr 4376280 0 * = 4376280 0 NNNNNNTTTCNTTAG... ##########
SRR345261.150 89 ETEC_chr 779844 42 49M = 779844 0 TTCAAGGAACCCCTGAA... B@D>ECC?BG@ECC>...
SRR345261.150 133 ETEC_chr 779844 0 * = 779844 0 CNCNGGAGTACNTTGA... ##########
SRR345261.175 83 ETEC_chr 3605306 42 49M = 3605113 -242 CCGCTTGCGCGGGCCA... EDE<8??;?@DGGDDE...
SRR345261.175 163 ETEC_chr 3605113 42 49M = 3605306 242 CCGGGTTCTGTCGTGG... DGGDGFDGGGGGGEGD...
SRR345261.200 77 * 0 0 * 0 0 AAAAAAAAAAAAAAAA... ##########
SRR345261.200 141 * 0 0 * 0 0 AAAAAAAAAAAAAAAA... 8@######
SRR345261.225 83 ETEC_chr 2879707 1 49M = 2879600 -156 CACAACACGAGCTGAC... 8?D8BEBGD@GG8GCE...
SRR345261.225 163 ETEC_chr 2879600 1 49M = 2879707 156 CCCACCTTCTCCAGT... GGGBGDEGG@GG<G8...
SRR345261.250 99 ETEC_chr 4361346 1 49M = 4361525 228 GTACTTTCAAGCGGGGA... ECE=>EC?FDG<EGDA...
SRR345261.250 147 ETEC_chr 4361525 1 49M = 4361346 -228 CCGGGCTCAACCTGGG... #####BC...

```

リファレンス配列に  
関する情報

FLAG

マップされた染色体と位置  
(\* はマップされなかった)

同じ名前のリード  
=ペアエンドのリード対

MAPQ

CIGAR

ペアの相手がマップされた染色体(同じなら=)と位置、フラグメントの長さ(右側のリードは負値)

リード配列

配列クオリティ値

オプション

AS アライメントスコア  
XS 他の位置でのベストスコア  
YF リードがfiltering outされた理由

# 復習: SAMからBAMへの変換

実習用ディレクトリ ~ /data / IU

入力

- SAMファイル (さきほどbowtie2によって作成されたもの)  
etec\_bowtie2.sam

- ◆ SAMからBAMへ変換する (出力ファイル名 : etec\_bowtie2.bam)

```
$ samtools [ ] [ ] etec_bowtie2.sam [ ] etec_bowtie2.bam
```

- ◆ 作成したBAMファイルをヘッダ付きでSAMに変換してlessで表示する

```
$ samtools [ ] [ ] etec_bowtie2.bam [ ] less
```

# 復習:BAMファイルのインデックスづけと検索

実習用ディレクトリ ~/data/IU

入力

- BAMファイル (さきほどSAMからの変換によって作成されたもの)  
`etec_bowtie2.bam`

- ◆ リファレンス配列上の位置の順にソートする  
(出力ファイル:`etec_bowtie2_sorted.bam`)

```
$ samtools [ ] etec_bowtie2.bam [ ] etec_bowtie2_sorted.bam
```

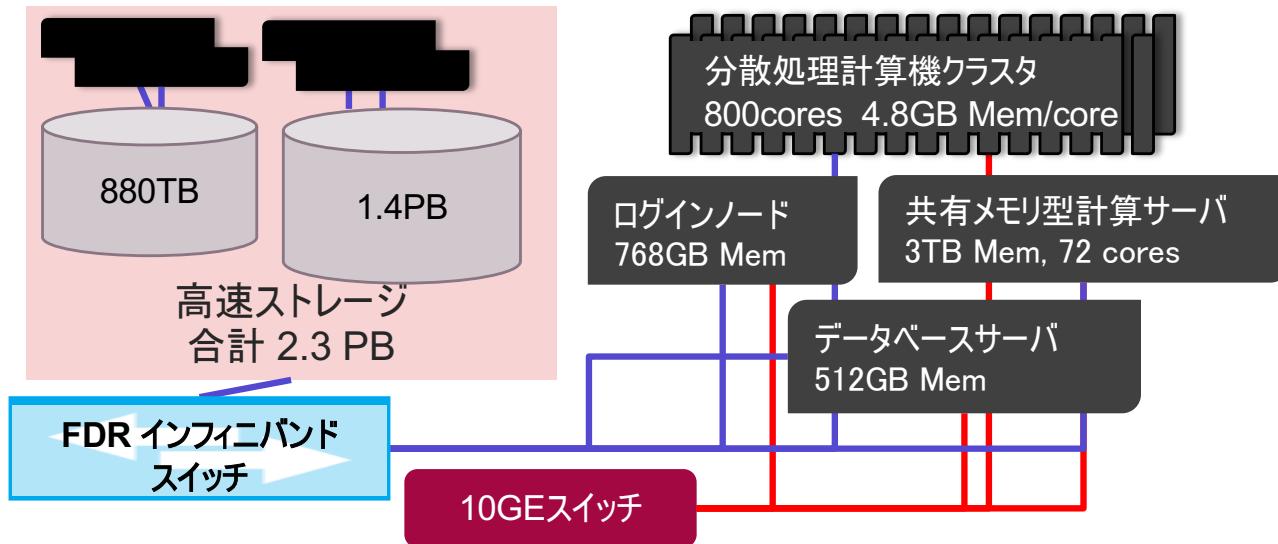
- ◆ ソートされたBAMファイルに対してインデックスを作成する

```
$ samtools [ ] etec_bowtie2_sorted.bam
```

- ◆ インデックスを使って、リファレンスの染色体配列(染色体名:ETEC\_chr)の10000-12000 の範囲にマッピングされた結果のみを表示する

```
% samtools [ ] etec_bowtie2_sorted.bam [ ]
```

# 生物情報解析システム (bias5)



メインページ 論議 開発 ソースを表示 履歴表示 検索

## 生物情報解析システム

NIBB 基礎生物学研究所 生物機能解析センター 情報管理解析室

目次 [非表示]

- 1 概要
- 2 障害・障害復旧情報
- 3 停止予定
- 4 システム構成
- 5 利用手続き
- 6 使い方

**概要**

- 生物情報解析システム Version5 は、2018年6月から運用中です。
  - 「共有メモリ型計算サーバ」メインメモリ 3TB （メモリが多く必要な解析に）
  - 「分散処理用計算機クラスタ」800コア （CPUを沢山使って分散処理したい解析に）
- 基礎生物学研究所の方はもちろん、所外の研究者の方にもお使いいただけます。ユーザー登録の取得方法
- 利用規約などは内規をご覧ください。 [生物情報解析システム利用内規](#)
- 全てのアプリケーション、データベース/ホームディレクトリを含めた全てのディスク領域は、全計算機から同様に使うことができます。
- システムをお使いいただくには「[UNIXコマンドの知識](#)」があることが前提になります。デスクトップ画面はありません。
- 2018年6月20日に実施されました「利用者説明会資料」を下記に置きました（アクセスにはVPN接続が必要です）
  - [生物情報解析システム Ver.5 利用者説明会資料](#)

**障害・障害復旧情報**

- ldas-smp 復活しました (2019/06/14~)

**停止予定**

- 停止予定はありません

**システム構成**

- マシン構成
- 分子生物学データベース

生物情報解析システムWiki  
<http://www.nibb.ac.jp/cproom/wiki3>

# リモートログインとファイル転送

## リモートログイン

- bias5 にログイン

```
% ssh username@bias5.nibb.ac.jp
```

- bias5 からログアウト

```
% exit
```

## ファイル転送

- ローカルのファイル(file) をbias5の~/directoryに転送

```
% scp file username@bias5.nibb.ac.jp/directory
```

- bias5の ~/directory/file をローカルのカレントディレクトリ(.) に転送

```
% scp username@bias5.nibb.ac.jp/directory/file .
```

# 実習

- bias5 にログインして、~/data/IUに移動
- FASTQファイル etec\_1.fq, etec\_2.fq に対して cutadaptを実行
- ゲノム配列 eco\_o139.fa に対してbowtie2用インデックスを作成
- アダプター除去した配列をインデックス付したゲノムに対してbowtie2でマッピング
- マッピング結果のSAMファイルをBAMに変換
- BAMファイルをマッピング位置でソートしてインデックスを作成
- インデックスを使って、染色体ETEC\_chrの10000-12000の範囲にマップされたリードを表示

# SAM/BAM フォーマット補足

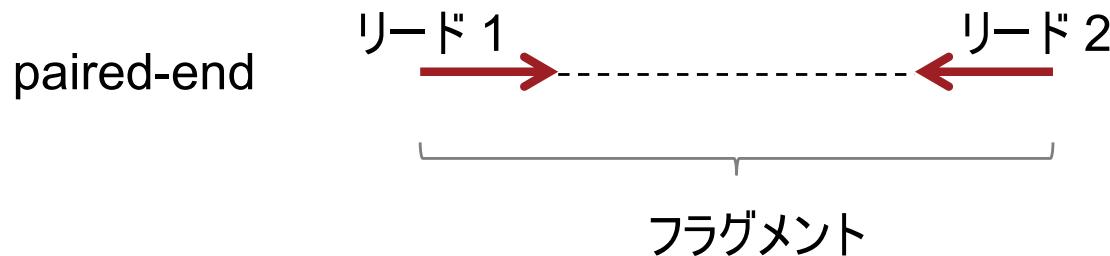
- Bowtie2のデフォルトオプションでマッピングした結果のSAM/BAMファイルは、もとのFASTQファイルに含まれている各リードの配列とクオリティデータをすべて含んでいる。以下のコマンドでSAM/BAMファイルからFASTQファイルを作成できる。

```
$ samtools fastq etec_bowtie2.bam -1 r1.fq -2 r2.fq
```

- 個々のリード配列を記録する代わりに、リファレンス配列を参照して、各リードのリファレンス上の位置とアライメント情報のみを記録することによって、さらに圧縮率を高めたバイナリ形式としてCRAM形式がある。

```
$ samtools view -C etec_bowtie2.sam -T eco_o139.fa  
-o etec_bowtie2.cram
```

# Bowtie2のオプション1 ペアエンドリード対の検索



- `-I int` フラグメント長の最小値(**default: 0**)
- `-X int` フラグメント長の最大値(**default: 500**)
- `--fr / --rf / --ff` リード1とリード2の相対的な向き (**default:fr**)



- 条件を満たさない(discordant)リード対もデフォルトでは出力される。その際、2カラム目(FLAG)の2ビット目(ペアが正しくアラインされたか？)に0がセットされる。

# フラグ(FLAG)

- True/Falseの2状態を1/0で表した変数。複数のフラグをまとめて、2進数の数値で表現される。
- フラグ値は10進数で表示されるが、2進数に変換することで解釈される。

## FLAG値

10進数	2進数	解釈
83	01010011	<p>ペアリードである</p> <p>各リードが適切にアラインされている</p> <p>逆鎖にマップされている</p> <p>1番目のリードである</p>

```
# unix コマンドによる 10進数→2進数の変換
% echo 'obase=2;83' | bc
1010011

# samtools を使ったフラグの解釈
% samtools flags 83
0x53    83      PAIRED,PROPER_PAIR,REVERSE,READ1

# 各フラグの説明を表示
% samtools flags
```

# Paired end readでのFLAG値



2進数表記 samファイルの記載は  
10進数表記

ペアリードがある	1	1	1	1	1	1	1	1	1	11111111	255
両方適切にマップされている	1	1	1	1	1	1	1	1	1	01010011	83
自分がマップされていない	0	1	1	0	0	0	1	1	1	01100011	99
ペア相手がマップされていない	1	0	0	1	0	0	1	1	1	10010011	147
ペア相手は逆鎖にマップされた	1	0	1	0	0	0	1	1	1	10100011	163
Read1の配列である	0	1	0	0	1	0	0	1	0	01001001	73
Read2の配列である	0	1	0	1	1	0	0	1	1	01011001	89
片方しかアラインしていない場合	0	1	0	0	0	1	0	0	1	01000101	69
どっちもアラインしていない場合	0	1	1	0	0	1	0	0	1	01100101	101
	1	0	0	0	1	0	0	1	0	10001001	137
	1	0	0	1	1	0	0	1	1	10011001	153
	1	0	0	0	0	1	0	0	1	10000101	133
	1	0	1	0	0	1	0	0	1	10100101	165
	0	1	0	0	1	1	0	1	0	01001101	77
	1	0	0	0	1	1	0	1	1	10001101	141

# Samtoolsを用いた フラグによるフィルタリング

- **samtools view -f フラグ値 BAMファイル**

指定したフラグ値中で1であるフラグが、BAMファイル中のフラグ値でもすべて1になっている行のみを抜き出す。

例) ペアリードでかつ両方が適切にアラインされている行のみを抜き出す

```
% samtools view -f 3 etec_bowtie2_sorted.bam
```

3は2進数で 11 だから、1番目と2番目のフラグが1である行を抜き出す(それ以外のフラグは無視する)。

- **samtools view -F フラグ値 BAMファイル**

指定したフラグ値中で1であるフラグが、BAMファイル中のフラグ値ではすべて0になっている行のみを抜き出す。

例) ペアリードの両方が適切にアラインされていない行のみを抜き出す

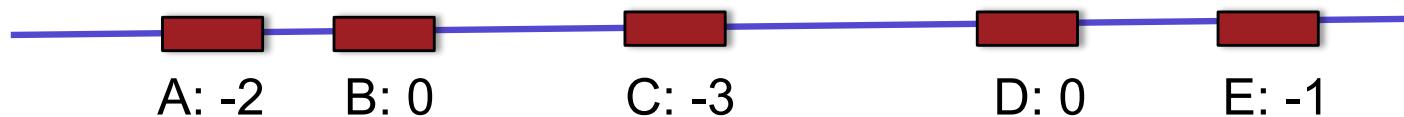
```
% samtools view -F 2 etec_bowtie2_sorted.bam
```

2番目のフラグが0である行を抜き出す。

# Bowtie2のオプション2

## アライメント出力のモード

- 一般に、1つのリードは複数の箇所にマップされる。



スコア = ミスマッチに対するペナルティ

- default (**best one mode**)

条件を満たすアライメントを検索し、最高スコアのものを1つ出力  
(ただし、検索は完全でないので、最高スコアを取りこぼす可能性はある)  
上記の例では、BまたはD（どちらかがランダムに選ばれる）

- a**

条件を満たすアライメントをすべて出力  
上記の例では、A,B,C,D,E

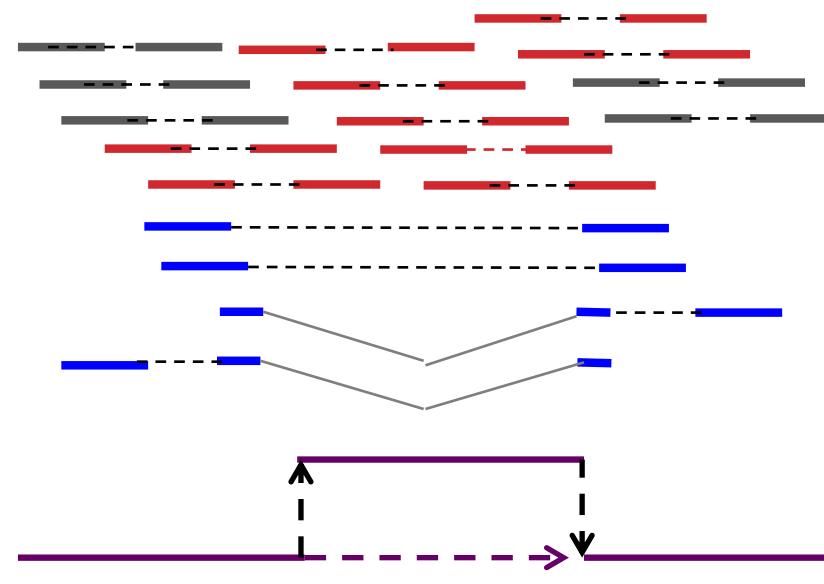
- k <int>**

条件を満たすアライメントを、見つかった順に指定した数だけ出力  
上記の例で、-k 2 のとき、左から順に見つかるとすると、AとB  
(実際には位置の順に見つかるわけではない)

- a や -k を指定したとき、最高スコアでないアライメントには9番目のフラグ  
(secondary alignment)に1がセットされる

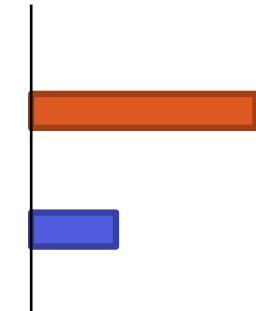
# (参考)デノボ・アセンブルによるRNA-Seq解析

## デノボ・アセンブルによる転写配列の構築



アセンブルされた  
トランスクript

**RSEM:** 各バリアントの  
存在量の推定



一つのリードを複数のトランスクriptにマップした上で存在量を推定する  
→ -a オプションを指定(または-kで大きい値を指定)

# マッピングクオリティ(MAPQ)

- マッピングクオリティ(MAPQ)値は以下の式で計算される。

$$\text{MAPQ} = -10 \log_{10}(P_e)$$

ただし、 $P_e$ はリードが間違った位置にマップされている確率の推定値。

- MAPQは、リードがその位置にどの程度ユニークにマップされたかを示す指標であり、その位置でのアライメントスコアが、他のすべての位置におけるスコアよりずっと大きいときに大きくなる。
- Bowtie2のデフォルトでは同じスコアのアライメントが複数の位置で得られた場合、ランダムに一つの位置を出力し、MAPQに低い値を設定する。
- MAPQが低いアライメントの位置は信用できないので、下流の解析の際には捨てた方が良い場合もある。

# Samtoolsを用いた MAPQによるフィルタリング

- samtools view -q 閾値 BAMファイル名

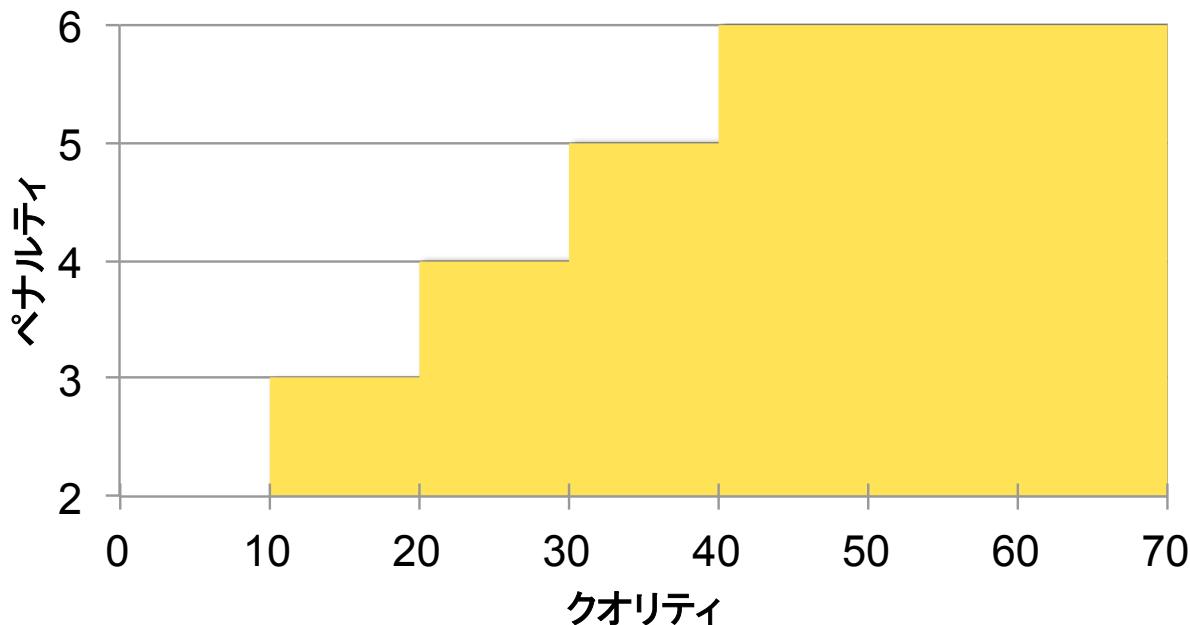
MAPQの値が閾値より小さい行を除く

例) MAPQが20以上の行のみを出力

```
$ samtools view -q 20 etec_bowtie2.bam
```

# (参考) Bowtie2におけるアライメントスコア

- マッチは0で、ミスマッチにマイナスのペナルティ(最高スコアが0点)
- ミスマッチペナルティは、クオリティ値に応じて $-2$ から $-6$ の値をとる(下図)
- あいまい塩基(N)のペナルティは $-1$
- ギャップペナルティは、ギャップの長さ $n$ に対して $-(5 + 3n)$
- スコアのカットオフは、長さ $L$ に対して $-0.6(L+1)$



# Bowtie2のオプション3

## アライメントのモード

- **--end-to-end** リード配列全長に渡るアライメント(default)

Read:        GACTGGGCGATCTGACTTCG  
              ||||||| ||||||||| |||  
Reference: GACTG--CGATCTGACATCG

- **--local** リード配列のうち、類似度の高い一部の領域のみを抜き出してアラインしたもの

Read:        ACGGTTGCGTTAA-TCCGCCACG  
              ||||||| |||  
Reference: TAACTTGCAGTAAATCCGCCTGG

# CIGAR文字列

- リードとリファレンス配列とのアライメントの詳細を表す。
- ギャップなしでアラインされている場合、 $nM$ （ $n$ はリード配列の長さ）となる。
- ギャップが入っている場合、 $nD$ （欠失）または $nI$ （挿入）（ $n$ は挿入・欠失の長さ）が入る。

**5M2D4M1I5M**

ref	AGACGAGATTA-GCATG
	: : :: : : : : :: ::
read	ACACG--ATTAGGCTTG

- ローカルアライメントのとき、両端の除かれる部分は  $nS$  で、また TopHatなどのスプライシングを考慮するアライメントにおいて、インtronとしてスキップされるリファレンス配列上の領域は  $nN$  で表される。

**5S4M1I5M**

ref	ACGGCTGATTA-GCATG
	: : : : :: ::
read	taaccATTAGGCTTG

# インデックスを使った高速検索

## ハッシュテーブル

ゲノム配列

ACACGTTACGGT.....

リード配列

CGTTGCA

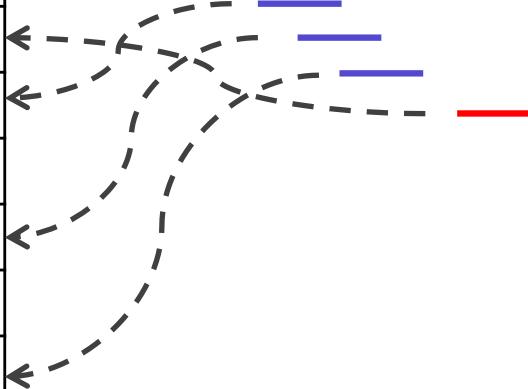
### ①インデックス作成

ハッシュテーブル  
各2-merの出現位置を記録

2-mer	positions
AC	1, 3, 8
CA	2
CG	4, 9
GG	10
GT	5, 11
TA	7
TT	6

### ②インデックスを使った初期検索(seed検索)

CGTTGCA



### ③見つかったseedを延長してアライメント

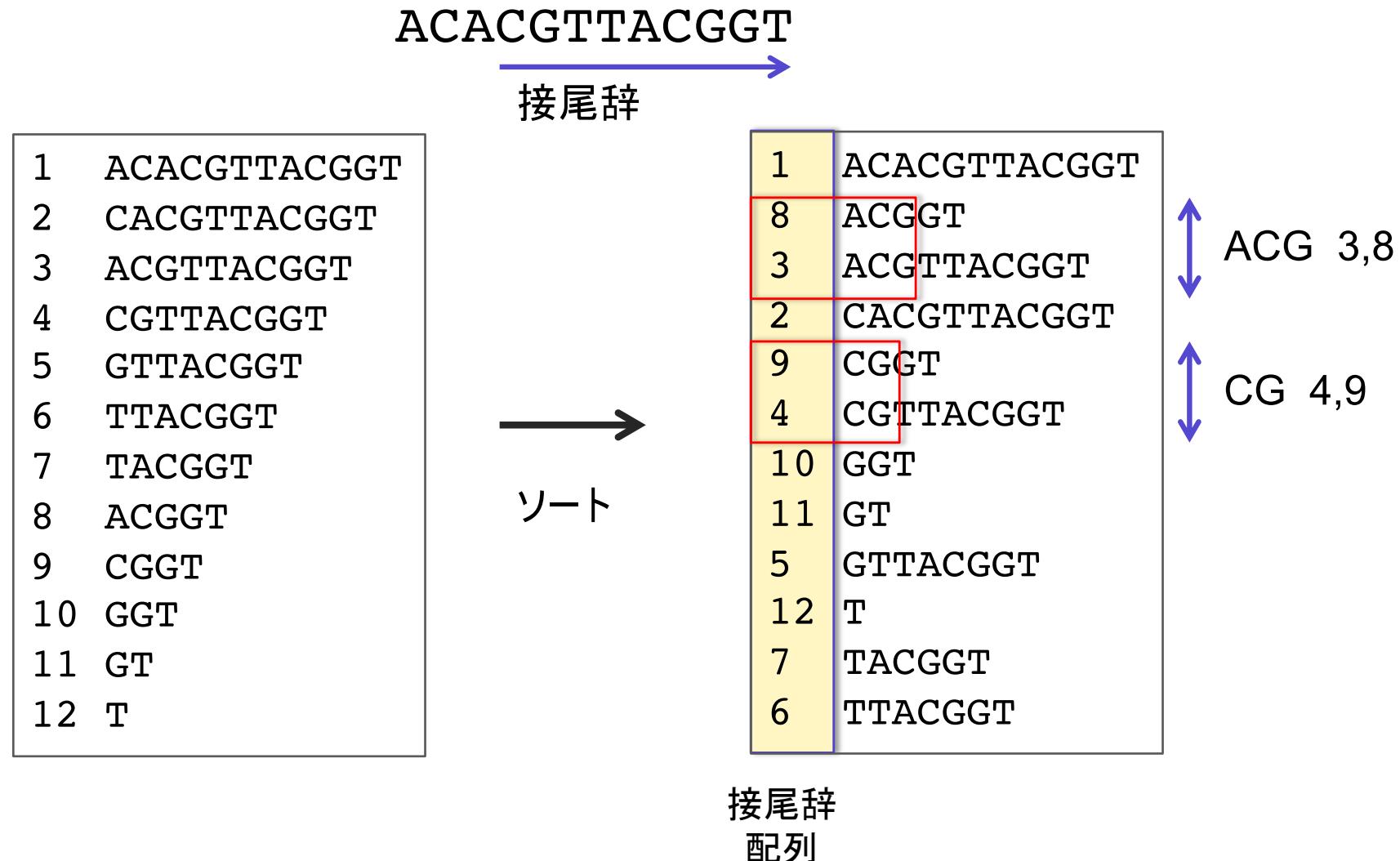
ACACGTTACGGT.....  
CGTT**GCA**

CGTTGCA

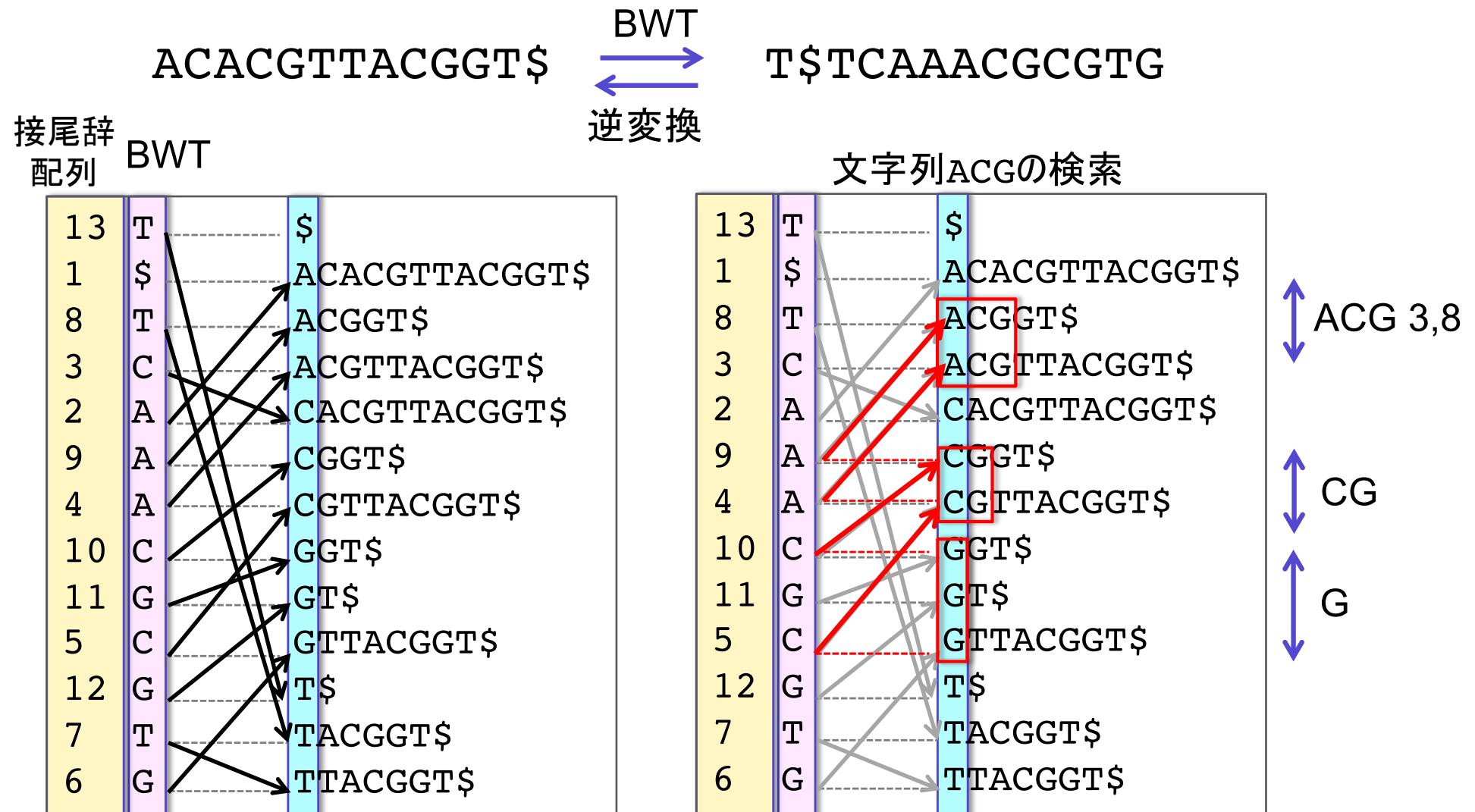
A  
C  
A  
C  
G  
T  
T  
A  
C  
G  
G  
T

エラー

# インデックスを使った高速検索 接尾辞配列(**suffix array**)



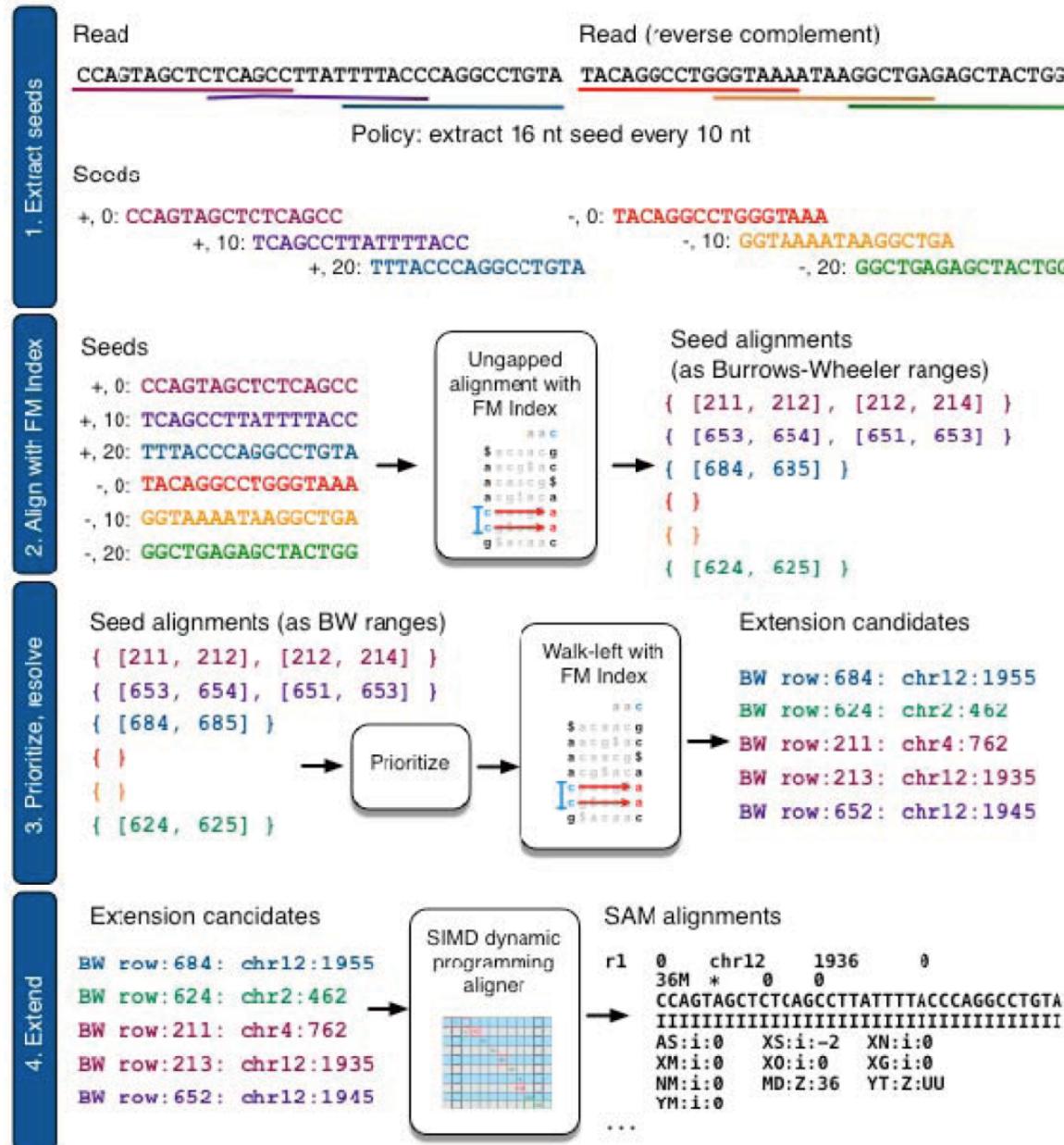
# Burrows-Wheeler 变換 (BWT)に基づく インデックス(FM-Index)



矢印(LF mapping)を辿って元の配列を再構築できる(逆変換)。

矢印を辿って接尾辞配列上での効率の良い文字列検索の実現

# Bowtie2 アルゴリズムの詳細



## 1. Seed 配列の抽出

各リード配列およびその相補配列から  $i$  塩基ごとに  $L$  塩基の配列を抽出して seed配列とする(図では $i=10$ ,  $L=16$ )。

## 2. FM index を用いた検索

各seed配列がゲノム上に出現する位置がBW rangeとして得られる。最大1つのミスマッチを考慮した検索が可能。

## 3. ヒットの優先付け、位置の取得

BW rangeの幅が小さいヒットに高い優先度をつけて、ランダムに候補をピックアップしゲノム上の位置を取得。

## 4. アライメントの計算

得られた位置の周辺で、ギャップ入りのアライメントスコアを計算。これを各候補位置について繰り返して、最高スコアを与えるゲノム上の位置を出力。

# Bowtie2のオプション4

## 検索の精度と速度に関するオプション

- **-N int** **seed** 検索時にミスマッチを許す数(0 or 1)
- **-L int** **seed** の長さ
- **-i func** **seed** をとる間隔(リード長を基に決める式を指定)
- **-D int** 最高スコアが更新されないときアライメント計算を打ち切るまでの回数
- **-R int** リードが高反復の**seed**をもつときに**re-seed**を行う最大回数

上記のオプションを同時に設定する**preset option**がある。高速(低感度)→高感度(低速)の順に4段階のオプションが用意されている。

- **end-to-end**モードの場合 (default: **sensitive**)  
**--very-fast** / **--fast** / **--sensitive** / **--very-sensitive**
- **local**モードの場合 (default: **sensitive-local**)  
**--very-fast-local** / **--fast-local** / **--sensitive-local** / **--very-sensitive-local**

# (参考) HISAT2 スプライシングを考慮した高速マッピングツール

- スプライシングを考慮して、一つのリードをゲノム上で離れた箇所にまたがってマッピングする。
- global とlocalの2つのインデックスを2段階で用いることにより、高速かつ正確にスプライスされたアライメントを実現。

