

## 1. MAF アライメント相同配列抽出プログラム

### (1) ファイル名

maf\_allele.pl

### (2) 目的

1000 Genomes から取得したヒトの VCF ファイルでは、REF フィールドに reference 配列が、ALT フィールドに alternative 配列が記載されている。しかし、これらから ancestral allele が何であるかを判別できない。INFO フィールドの AA サブフィールドに ancestral allele が記載されている変異サイトもあるが、すべての変異サイトを網羅しているわけではない。

そこで、UCSC Table Browser から Multiz Align (multiz46way) の MAF ファイル<sup>[1]</sup>をダウンロードし、オランウータンの相同配列とチンパンジーの相同配列から ancestral allele を推定する。

このプログラムでは、VCF ファイルに記載されている変異サイトについて、MAF ファイルに記載されている指定した種の相同領域の配列をタブ区切りテキスト形式で出力する。

### (3) 使用方法

```
$ ./maf_allele.pl [VCF ファイル名] [MAF ファイル名] [染色体名] [reference 名] [出力種名]
```

[VCF ファイル名] は入力 VCF ファイル名を指定する。

[MAF ファイル名] は入力 MAF ファイル名を指定する。

[染色体名] は染色体名を指定する。

[reference 名] は VCF ファイルの reference 配列に相当する配列の、MAF ファイルにおける名称を指定する。

[出力種名] は相同領域の配列を出力したい種の、MAF ファイルにおける名称を指定する。「,」で区切って複数指定することができる。

#### 《例》

```
$ ./maf_allele.pl nh1-1.vcf nh1-1_multiz46way.maf 1 hg19.chr1 panTro2.chr1,ponAbe2.chr1
```

この例では“nh1-1.vcf”に記載された 1 番染色体の変異サイトについて、“nh1-1\_multiz46way.maf”の panTro2.chr1 および ponAbe2.chr1 の相同領域の配列を出力する。

[1] MAF ファイル … multiple alignment format (<http://genome.ucsc.edu/FAQ/FAQformat.html#format5>)

## 2. RepeatMask 領域座標出力プログラム

### (1) ファイル名

rmsk\_pos.pl

### (2) 目的

1000 Genomes から取得したヒトの VCF ファイルによってヒトゲノム情報を解析する際、リピート領域を除いて解析を行いたい。

そこで、UCSC Table Browser から Repeating Elements by RepeatMasker の RMSK テーブル<sup>[2]</sup>をダウンロードし、Simple repeats に含まれる変異サイトを解析対象から除外する。

このプログラムでは、VCF ファイルに記載されている変異サイトについて、RMSK テーブルの Simple repeats に含まれている変異サイトの染色体名と座標を vcftools 入力形式で出力する。

### (3) 使用方法

```
$ ./rmsk_pos.pl [VCF ファイル名] [RMSK 名]
```

[VCF ファイル名] は入力 VCF ファイル名を指定する。

[RMSK 名] は入力 RMSK テーブルのファイル名を指定する。

《例》

```
$ ./rmsk_pos.pl nh1-1.vcf nh1-1.rmsk > nh1-1_rep.txt
```

この例では“nh1-1.vcf”に記載された変異サイトのうち、“nh1-1.rmsk”の Simple\_repeat の範囲内にあるサイトの染色体番号と座標を“nh1-1\_rep.txt”に出力する。

この出力結果を使用して、vcftools で当該変異サイトを除いた VCF ファイルを作成することができる。

《例》

```
$ vcftools --vcf nh1-1.vcf --out nh1-1_norep --exclude-positions nh1-1_rep.txt  
--recode --recode-INFO-all
```

この例では“nh1-1.vcf”から“nh1-1\_rep.txt”に記載された変異サイトを除いたものを“nh1-1\_norep.record.vcf”に出力する。

[2] RMSK テーブル … RepeatMasker (<http://www.repeatmasker.org/>) の出力結果