

## Tutorial: Análise de Variantes (*Variant Calling*)

Frederico Schmitt Kremer

**Objetivos:** Neste tutorial mostraremos como é feita uma análise de SNPs e INDELS a partir de dados de sequenciamento de nova geração. Utilizaremos como modelo dados gerados pela plataforma Ion Torrent PGM, mas estas mesmas abordagens podem ser utilizadas para qualquer plataforma de sequenciamento, sendo necessário apenas a verificação de quais programas de gerenciamento são mais apropriados para cada caso.

Programas necessários:

- **segemehl.x.**  
Disponível em:  
<http://www.bioinf.uni-leipzig.de/Software/segemehl/>
- **SAMtools.**  
Disponível em:  
<http://samtools.sourceforge.net/>  
apt-get: `sudo apt-get install samtools`
- **BCFtools.**  
Disponível em:  
<https://samtools.github.io/bcftools/bcftools.html>  
apt-get: `sudo apt-get install bcftools`
- **vcfutils.pl.**  
Disponível em:  
<https://raw.githubusercontent.com/lh3/samtools/master/bcftools/vcfutils.pl>

**Arquivos necessários**

- Sequência de referência (formato FASTA).  
Disponível em: [http://200.132.101.131/aulas/aula\\_7/dados/referencia.fasta](http://200.132.101.131/aulas/aula_7/dados/referencia.fasta)
- Leituras de sequenciamento (formato FASTQ)  
Disponível em: [http://200.132.101.131/aulas/aula\\_7/dados/reads.fastq](http://200.132.101.131/aulas/aula_7/dados/reads.fastq)

## Rodando a análise de variantes

### 1. Criando um arquivo de *index* para a referência (para o **segemehl.x**)

```
$ segemehl.x -x referencia -d referencia.fasta
```

### 2. Alinhando as leituras contra a sequência de referência

```
$ segemehl.x -i referencia -d referencia.fasta -q reads.fastq > alinhamento.sam
```

### 3. Converter SAM para BAM

```
$ samtools view -Sb alinhamento.sam > alinhamento.bam
```

### 4. Ordenar o arquivo BAM

```
$ samtools sort -f alinhamento.bam alinhamento.sorted.bam
```

### 5. Indexar o arquivo BAM

```
$ samtools index alinhamento.sorted.bam
```

### 6. Indexar a sequência de referência (para o **samtools**)

```
$ samtools faidx referencia.fasta
```

### 7. Converter o formato BAM para BCF

```
$ samtools mpileup -g -f referencia.fasta alinhamento.sorted.bam > vars.raw.bcf
```

### 8. Computar a probabilidade de cada base ser um SNP ou INDEL

```
$ bcftools view -bvcg vars.raw.bcf > vars.filtered.bcf
```

### 9. Filtrar as variantes com base no seu valor de qualidade

```
$ bcftools view vars.filtered.bcf | vcfutils.pl varFilter - > vars.final.vcf
```