

Liga Brasileira de Bioinformática

Desafio Mendelics

Dia 3 - Resultados



Ana Elisa Ribeiro Orsi

Outubro/2021

No presente relatório, serão expostos os resultados observados no terceiro dia do desafio.

1. A razão Ti/Tv foi obtida através de um script em Python (Scripts/Dia_3_p1.py). As SNVs foram selecionadas do arquivo Outputs Exigidos/Dia_3.vcf, e classificadas entre transições e transversões. Foram encontradas:

- 466 transições
- 153 transversões

Assim, a razão Ti/Tv calculada foi 3.045751633986928 (figura 1).

```
Parte 1: Transições e Transversões
Ti total: 466
Tv total: 153
Ti/Tv = 3.045751633986928
```

2. O mesmo script em Python foi utilizado para contabilizar as variantes presentes na região 16000000-20000000. Para isso, foi utilizada a coluna "POS" do arquivo VCF. Foram localizadas 65 variantes na região selecionada (figura 2).

```
Parte 2: Variantes na região de 16000000 a 20000000
Foram encontradas 65 variantes na região determinada
```

3. Para a terceira e quarta questões, foi utilizada a API do Ensembl. O script em Python utilizado pode ser verificado em Scripts/Dia_3_p2.py. A detecção das variantes não-sinônimas foi realizada pelo atributo "transcript_consequences" → "consequence_terms". A linha selecionada para impressão pode ser verificada abaixo:

```
Linha do arquivo VCF:
chr22 16591593 22:g.16591593A>G A G 225.417 PASS DP=156;VDB=0.00776903;SGB=-0.693147;FS=0;MQ0F=0;AC=2;AN=2;DP4=0,0,69,63;MQ=60 GT:PL 1/1:255,255,0
```

4. O mesmo request utilizado na questão anterior foi usado para a obtenção do atributo MAF do gnomAD. Para isso, o atributo "colocated_variants" → "frequencies" → "gnomad" foi consultado. Abaixo, a linha do arquivo VCF impressa:

```
Linha do arquivo VCF:
chr22 15820150 22:g.15820150C>A C A 221.391 PASS DP=265;VDB=0.429241;SGB=-0.693147;RPBZ=1.75647;BQBZ=-0.379966;FS=0;MQ0F=0;AC=1;AN=2;DP4=57,98,16,43;MQ=60 GT:PL 0/1:255,0,255
```