

GWAS data input

May 14, 2020

```
In [99]: import numpy as np
import pandas as pd
```

lendo arquivo do pgc

```
In [100]: GWAS = pd.read_csv ('/media/thiago/um_tera/arquivos/twins/scz.full.txt', delimiter="
conforme modelo do prs
```

```
In [101]: gwas = GWAS[['SNP', 'P', 'OR', 'A1', 'A2']]
```

```
In [102]: gwas.head()
```

```
Out[102]:
```

	SNP	P	OR	A1	A2
0	rs3131972	0.761033	10.257	A	G
1	rs3131969	0.784919	10.221	A	G
2	rs3131967	0.79352	10.227	T	C
3	rs1048488	0.761041	0.9749	T	C
4	rs12562034	0.987899	10.011	A	G

renomeando

```
In [103]: gwas.columns=['snpid', 'pval', 'or', 'a1', 'a2']
```

fazendo uma header

```
In [104]: header = """##fileformat=VCFv4.1
##fileDate=20201405
##source=gambiarra
##reference=file:///seq/references/
#CHROM POS ID REF ALT QUAL FILTER INFO
"""
```

```
In [105]: output_VCF = "gwas.vcf"
```

```
In [106]: with open(output_VCF, 'w') as vcf:
vcf.write(header)
```

criando um vcf

```
In [107]: gwas.to_csv(output_VCF, sep="\t", mode='a', index=False)
```