



# VCF PROCESSES: Ferramenta para análise de mutações gênicas



# Olá!



[eliseumedeiros@gmail.com](mailto:eliseumedeiros@gmail.com)

Eu sou o

**Eliseu Medeiros**

Estudante concluinte do curso de Bacharelado em  
**Tecnologia da Informação** na UFRN e iniciante como  
bioinformata.





# Início

## Projeto de IC - Universidade Federal de Ribeirão Preto - 2015

Orientador: Jorge Estefano Santana de Souza [[jorge@imd.ufrn.br](mailto:jorge@imd.ufrn.br)]



A decorative pattern of hexagons in various shades of blue and cyan, some solid and some outlined, arranged in a cluster on the left side of the slide. A small network icon is also visible near the top left.

1

# Introdução

Sequenciamento - Anotação de Variantes - VCF -  
SNPs



The diagram illustrates the process of genome assembly from overlapping segments. It is divided into four horizontal stages, connected by downward arrows:

- Cleaned genomes:** The top stage shows several long, horizontal bars of various colors (red, orange, yellow, green, blue, purple) representing individual cleaned genome sequences.
- Multiple genomes are short into variable sized segments:** The second stage shows these long bars broken down into many smaller, overlapping segments of the same colors.
- Unordered sequenced segments:** The third stage shows the segments from the previous stage rearranged into a disordered, overlapping pile.
- Computational automated assembly:** The bottom stage shows the segments being pieced back together into longer, more continuous sequences.

At the very bottom, a specific sequence of nucleotides is shown: **ATGTTCCGAGTAAGGAAAAGTCTTCTTCATCAGTAAAGGAGGAAATAGAA**. To the right of this sequence, a bracket indicates that overlapping sequence segments are combined to reconstruct the genome.



A stylized illustration of a document or folder icon. It is white with a blue border and the letters "VCF" in blue. Below the text are three horizontal lines, suggesting a list or form.

# Anotação de Variantes

## Formato de arquivo VCF

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NAO00001 NAO00002 NAO00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:..
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```



# Análise da Sequência

◇ Lugares onde a amostra difere do genoma de referência (variantes)

Tipo	O que significa	Exemplo
SNP	Polimorfismo de nucleotídeo único	Referência = 'A', amostra = 'C'
Ins	Inserção	Referência = 'A', amostra = 'AGT'
Del	Eliminação	Referência = 'AC', exemplo = 'C'
MNP	Polimorfismo de nucleotídeo múltiplo	Referência = 'ATA', exemplo = 'GTC'
MISTURADO	Múltiplos nucleotídeos e um InDel	Referência = 'ATA', exemplo = 'GTCAGT'



# Possíveis dúvidas no VCF não anotado

- As variações estão em um gene?
- Em um exon?
- Elas mudam a codificação de proteínas?
- Elas causam códons de parada prematuros?

...





A decorative pattern of hexagons in various shades of blue and cyan, some solid and some outlined, arranged in a cluster on the left side of the slide. A small network icon is also visible near the top left.

2

# SNPEff e SNPSift

Definição - Principais Funcionalidades -  
Informações Gerais da Ferramenta

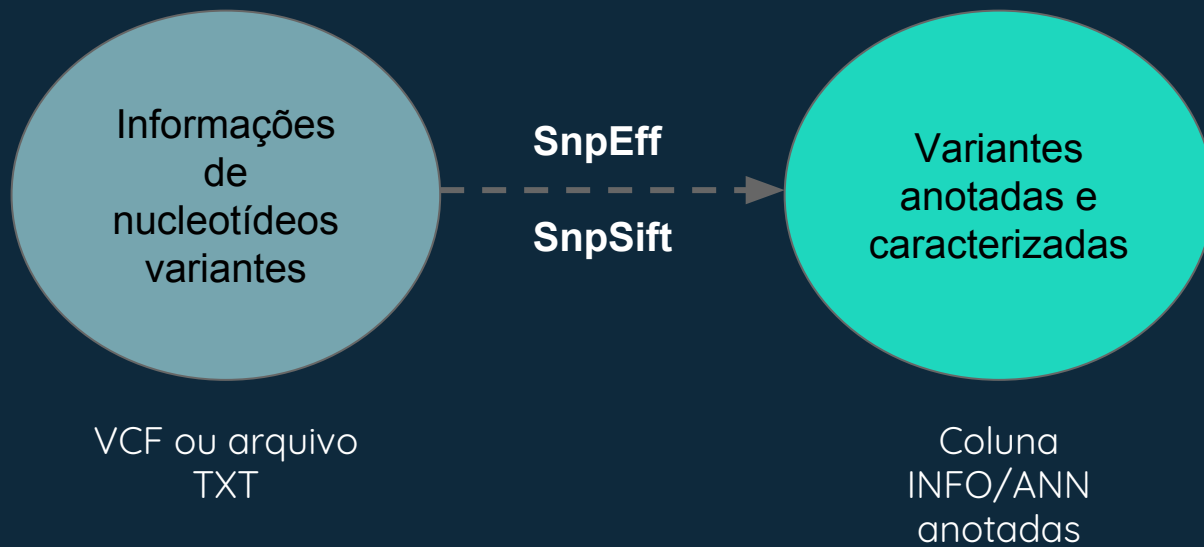


# Definição do SnpEff

- Ferramenta de anotação e previsão de efeito gênico
- “muito rápido, ele pode calcular previsões para todos os SNPs do projeto 1000 Genomes em menos de 15 minutos.”
- Open source e é integrada em outras ferramentas e pipelines



# Processo de sequenciamento e análise





# Funcionamento Básico

## ENTRADA

O resultado de um experimento de sequenciamento e geralmente está no formato de chamada variante (VCF) representa s SNPs, inserções, exclusões e MNPs.

## SAÍDA

SnEff analisa as variantes de entrada. Anota as variantes e calcula os efeitos que produzem em genes conhecidos (por exemplo, alterações de aminoácidos).

A decorative graphic on the left side of the slide, featuring a large cyan hexagon with the number '3' inside. Surrounding this central hexagon are several other hexagons in various shades of blue and cyan, some solid and some outlined. There are also small icons: a network of nodes and lines, and a magnifying glass.

3

# VCF PROCESSES

Instruções Gerais - Executando a anotação - Outras  
Informações



# Informações Gerais

- anotação de mutações, dando relevância quanto a sua patogenicidade
- filtros de seleção de variantes sobre as colunas
- filtro direto sobre as features (buscas)
- Visão geral de resultados
- Salvar em EXCEL ou CSV

Tutorial:

14 <https://github.com/eliseumedeiros/VCF-ProcessesProject/blob/master/README%20-%20VCF-ProcessesProject.pdf>





## VCF - Processes

### Open File

#### Select VCF File

Location:

#### Add Annotation

☐ Execute SnpEff☐ Execute SnpSift

CHROM



# Annotators databases

## SnpEff

### Databases

☒ ebola\_zaire



### Download

Database Name:

View all databases

Download

## SnpSift

☐ Saved Databases

☒ Other Database

"" .txt.gz file location:

Move to my Saved Databases

Select



The tabix index file must be in the same folder as the database.

\*You can download dbNSFP from the SnpSift website: <http://snpeff.sourceforge.net/SnpSift.html#dbNSFP>

## System Memory

RAM to be used when running SNP:

4 GB



Ok





## VCF - Processes

CHROM	POS	ID	REF	ALT	QUAL	FILTER	AC	AF	AN	DS	set	LOF	NMD	Allele
22	17071756	.	T	C	2313.33	PASS	2	0.00122	1644	true	Intersection			C
22	17072035	.	C	T	1036.44	PASS	1	0.00061	1644	true	Intersection			T
22	17072258	.	C	A	923.46	PASS	1	0.00061	1644	true	Intersection			A
22	17072674	.	G	A	1227.96	PASS	1	0.00061	1644	true	Intersection			A
22	17072747	.	T	C	1729.18	PASS	5	0.00304	1644	true	Intersection			C
22	17072781	.	C	T	1018.83	PASS	1	0.00061	1644	true	Intersection			T
22	17073043	.	C	T	1226.07	PASS	2	0.00122	1644	true	Intersection			T
22	17073066	.	A	G	184146.49	PASS	1386	0.84307	1644	true	Intersection			G
22	17073119	.	C	T	1458.2	PASS	2	0.00122	1644	true	Intersection			T
22	17073131	.	C	T	1084.78	PASS	2	0.00122	1644	true	Intersection			T
22	17073134	.	C	A	609.01	PASS	1	0.00061	1644	true	Intersection			A
22	17073145	.	T	G	805.74	PASS	1	0.00061	1644	true	Intersection			G
22	17073148	.	G	A	572.76	PASS	1	0.00061	1644	true	Intersection			A
22	17073150	.	C	T	3172.38	PASS	5	0.00304	1644	true	Intersection			T
22	17073153	.	G	C	787.56	PASS	1	0.00061	1644	true	Intersection			C
22	17073178	.	C	T	806.53	PASS	1	0.00061	1644	true	Intersection			T
22	17073198	.	C	T	4893.04	PASS	7	0.00426	1644	true	Intersection			T

CHROM

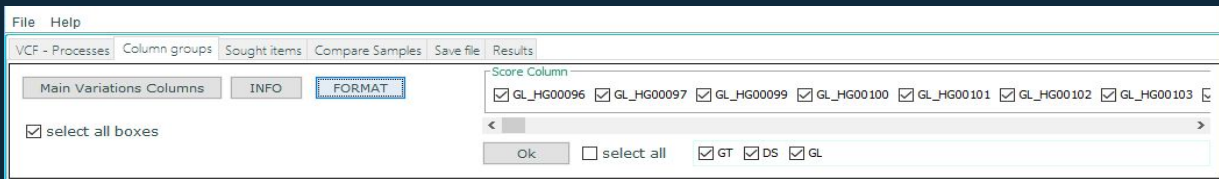


Open File

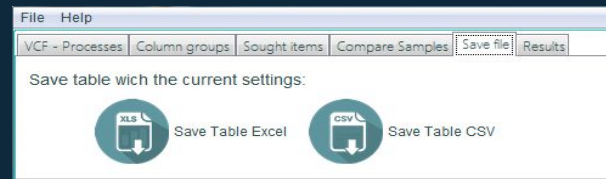


# Outras Funções

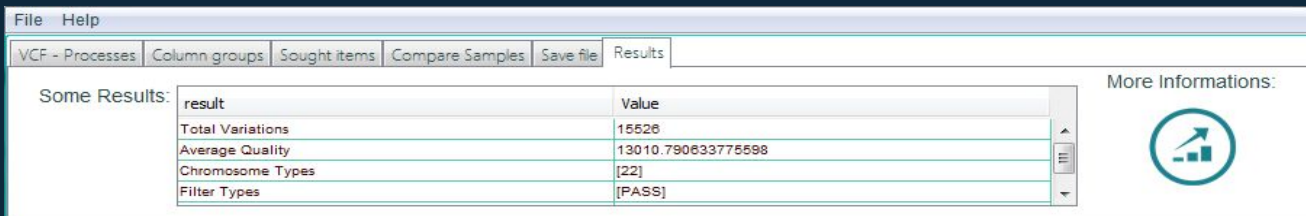
## Selecionar colunas da tabela para visualização



## Salvar em Formato xls ou csv



## Informações Gerais do Arquivo





Projeto Disponível em:

<https://github.com/eliseumedeiros/VCF-ProcessesProject.git>

