

VCF - Processes Project

Resumo

Foi desenvolvida uma interface com objetivo de fornecer de forma facilitada a anotação de mutações, dando relevância quanto a sua patogenicidade. Além da anotação também foi construído um sistema de filtros de seleção de variantes sobre as colunas (*features*), atuando como filtro direto sobre as *features* para, por exemplo, fazer buscas por genes, posições genômicas, patogenicidades, etc. Remoção e seleção dinâmica de colunas de forma a facilitar o processo de análise em colunas de interesse de estudo, no arquivo. Por fim a ferramenta oferece a possibilidade de salvar os resultados obtidos pelas anotações e filtros utilizados, em arquivos no formato EXCEL ou CSV. O presente material visa fornecer instruções iniciais para utilização da ferramenta VCF-Processes, com exemplos de utilização.

Sumário

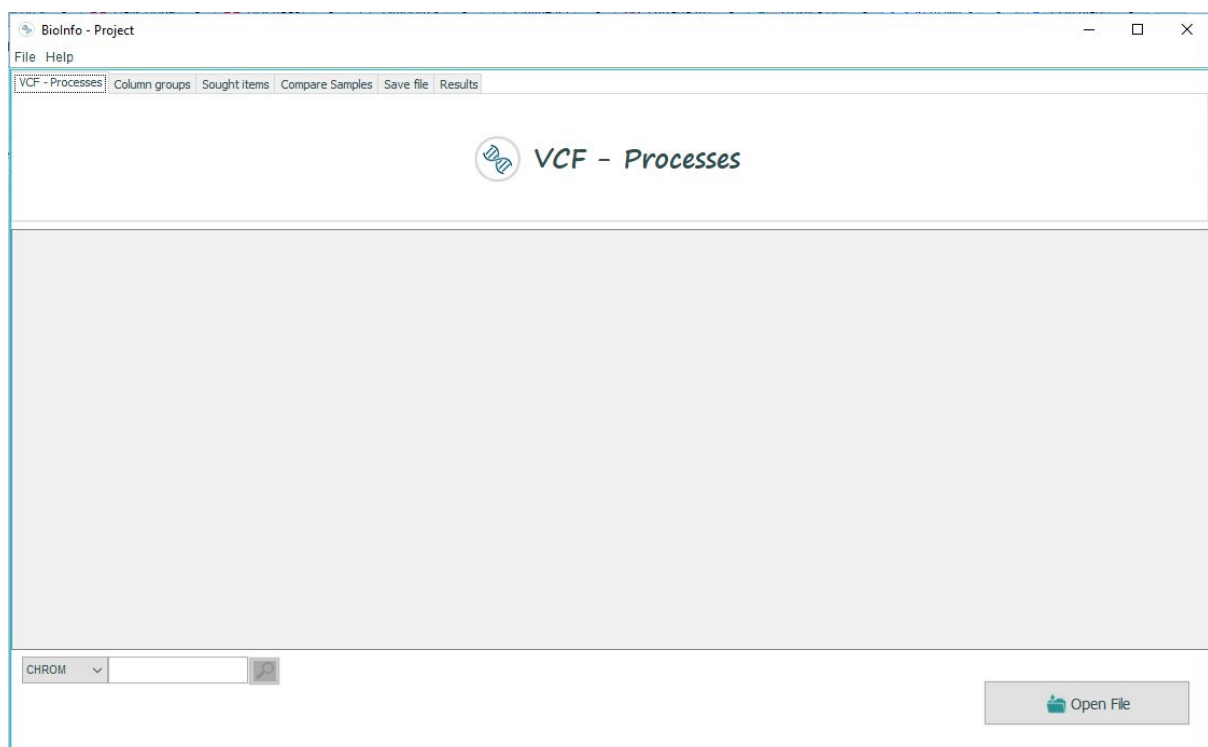
INTRODUÇÃO	2
1. ABRIR ARQUIVO VCF	3
2. ADICIONAR ANOTAÇÃO	5
3. CONSULTAS NA TABELA	7
4. SELECIONAR COLUNAS PARA VISUALIZAÇÃO	9
5. COMPARAÇÃO DE AMOSTRAS	11
6. SALVAR TABELA NOS FORMATOS EXCEL OU CSV	13
7. VISUALIZAR RESULTADOS	14
EXECUTANDO TESTE	16
Exemplo de Utilização Simples:	16
Exemplo de Utilização Simples para Trio-Analysis:	18
Exemplo de anotação com SNPEff	19
Exemplo de anotação com SNPSift	20

INTRODUÇÃO

Na atualidade a demanda por chamada de variantes de um genoma vem crescendo principalmente devido aos avanços na biologia molecular e na medicina de precisão, que visa o tratamento mais adequado e otimizado de acordo com o contexto genômico do paciente, é inegável que os estudos e análises pontuais do material genético de determinado organismo possam direcionar para um tratamento medicamentoso mais adequado e efetivo. O processo de anotação de variantes genômicas é um processo laborioso que exige muito esforço técnico/computacional, muitas vezes requerendo conhecimentos avançados de programação e sistemas operacionais como o linux. Dado e exposto o objetivo desse projeto foi o desenvolvimento de software para o sistema operacional *Windows®* com uma interface *friendly* que oferece uma forma facilitada de anotação e análise de variantes genômicas.

O procedimento de anotação e identificação de relevância de mutações, deriva dos alelos variantes identificados após os processos de alinhamento de sequências e chamada de variantes, que são padronizados em um arquivo bruto, denominado *Variant Call Format (VCF)*, que servirá de *input* para ferramenta proposta. Para a anotação dos alelos variantes a plataforma utiliza bancos de polimorfismos (*dbSNP*), preditores de efeito deletério (*Sift*, *Polyphen*, *Proven*) e anotadores de variantes como o *SnEff* e *SnSift*. Sobre os alelos variantes essas ferramentas fornecem diversas características que serão convertidas em *features* internamente em nosso sistema.

Imagem 1: Tela Inicial



Requisitos mínimos do sistema para utilização da aplicação:

- 4 Gb de RAM
- Sistema operacional Windows
- Possuir java instalado

1. ABRIR ARQUIVO VCF

Os procedimentos iniciais para a visualização do arquivo VCF podem ser observados no catálogo que se segue:

Imagem 2: Instruções de uso para abertura do arquivo

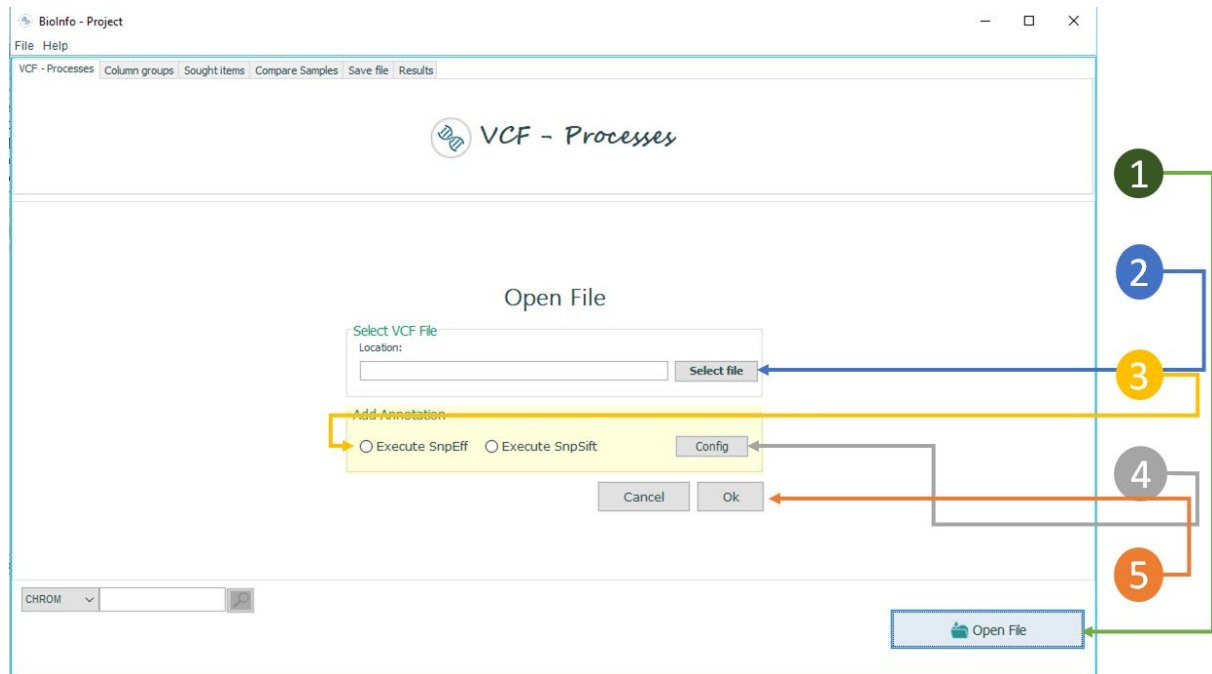


Tabela 1: Passos para Abertura do Arquivo VCF

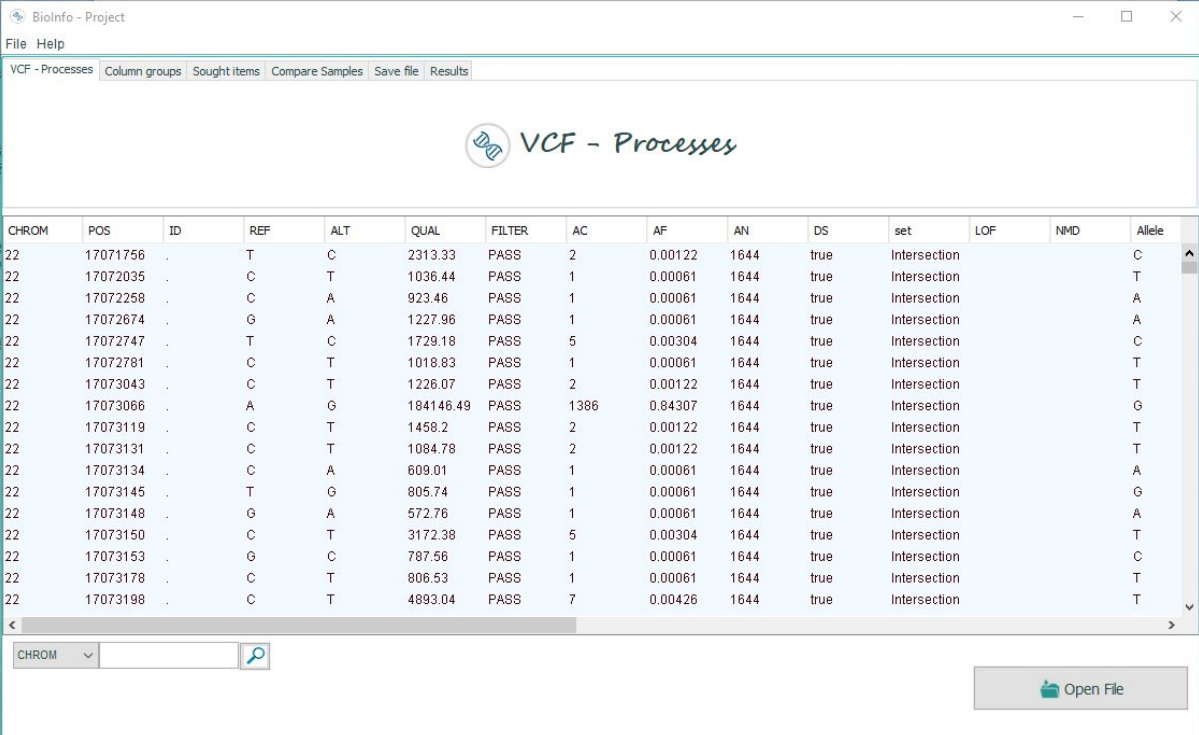
1	Selecionar “Open File” para abrir as opções de seleção do arquivo VCF que se deseja analisar
2	Selecionar onde está localizado o arquivo de variantes anotadas (“.vcf”)
3	Caso se deseje adicionar as anotações dos preditores, basta selecionar o preditor “ <i>SNPEff</i> ” ou “ <i>SNPSift</i> ” que irá adicionar as informações ao arquivo VCF original <ul style="list-style-type: none">• <i>SNPEff</i> (ferramenta de predição de efeito de variante no gene)• <i>SNPSift</i> (é uma caixa de ferramentas que permite filtrar e manipular arquivos anotados. Depois que suas variantes genômicas tiverem sido anotadas, será necessário filtrá-las para encontrar as "variantes interessantes / relevantes".)
4	Selecionar o botão “Config” para configurar as <i>databases</i> usadas na realização das anotações utilizadas pelos preditores

5

- Clicar em “ok” para que seja realizado o processamento dos dados e exibida a tabela gerada pela leitura do arquivo VCF
- Clicar em “cancel” para cancelar a abertura/processamento do arquivo

Após o arquivo ser aberto, aparecerá uma tabela, com todas as informações anotadas no arquivo VCF divididas por colunas. O passo final após o arquivo ser aberto será exibido conforme se segue:

Imagem 3: Arquivo aberto



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

2. ADICIONAR ANOTAÇÃO

Para realizar a anotação *SNPEff* é necessário realizar o download do banco que se deseja que seja feita a anotação. Ir em “config” (passo 4 das instruções de “Abrir Arquivo VCF”)

Imagem 4: Configurações dos bancos de polimorfismos SNPEff e SNPSift

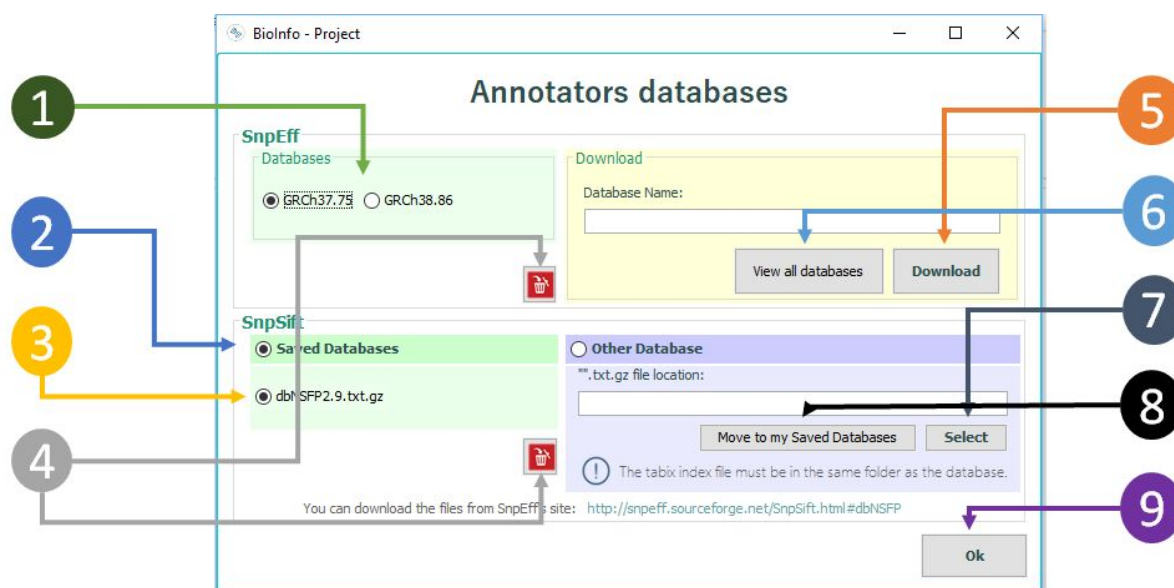


Tabela 2: Passos para configuração dos bancos de polimorfismos

1	Selecionar da lista de Databases utilizados no anotador preditor de efeitos no gene <i>SnEff</i> . Ou seja, o database selecionado nesta lista de databases será utilizado para adicionar as anotações no arquivo de variantes que se deseja analisar.
2	Os campos de seleção de <i>Databases</i> utilizadas no <i>SNPSift</i> , podem ser selecionados dentro da aplicação “ <i>Saved Databases</i> ” (<i>databases</i> que já foram adicionadas anteriormente na aplicação, do passo 8), ou fora dela “ <i>Other Database</i> ”.
3	Quando a opção “ <i>Saved Databases</i> ” está selecionada, o sistema irá considerar a <i>database</i> selecionada nesta lista de <i>databases</i> que foram inseridas anteriormente na aplicação, pelo passo 8.
4	A opção de excluir permanentemente da aplicação a <i>database</i> atualmente selecionada. Existem dois botões “excluir”, o primeiro exclui a <i>database</i> selecionada na lista do passo 1, e o segundo exclui da <i>database</i> selecionada da lista do passo 3.
5	Para realizar o <i>Download</i> de uma nova <i>database</i> utilizada para gerar novas anotações para o arquivo VCF que se deseja analisar. É necessário que o campo de “ <i>Database Name</i> ” esteja preenchido corretamente (para ver os possíveis nomes, basta ir ao passo 6).

6	Quando se deseja visualizar o arquivo que contém todos os nomes de <i>databases</i> possíveis para ser feito o <i>download</i> e aparecerem na lista de <i>databases</i> do <i>SnPEff</i> do passo 1.
7	Para selecionar o caminho da <i>database</i> que será utilizada para anotação da relevância da variação, <i>SNPSift</i> (a opção <i>Other Database</i> do passo 2 será automaticamente selecionada). Para realizar o <i>download</i> desses arquivos, pode-se acessar o link: http://snpeff.sourceforge.net/SnpSift.html#dbNSFP e baixar os arquivos do formato “ <u>nomeDatabase.txt.gz</u> ” e “ <u>nomeDatabase.txt.gz.tbi</u> ”, necessários para a execução dessa funcionalidade.
8	Após ter selecionado a <i>database</i> que será utilizada fora da aplicação, o botão “ <i>Move to my saved databases</i> ” dá a opção de mover este arquivo do formato “ <u>nomeDatabase.txt.gz</u> ”, previamente selecionado no passo 7, para dentro da aplicação. Sendo listada em seguida no passo 3. (internamente o arquivo no formato “ <u>nomeDatabase.txt.gz.tbi</u> ” também será movido para a aplicação).
9	Botão “ <i>ok</i> ” é utilizado para confirmar/finalizar as configurações setadas nos passos anteriores desta seção e retornar para a tela de abertura do arquivo VCF.

3. CONSULTAS NA TABELA

Para realizar a consulta em valores da tabela, deve-se ir de consulta, localizada na região esquerda abaixo da tabela. Selecionado:

1. o campo que se deseja consultar
2. digitando a expressão de consulta (o comportamento pode mudar de acordo com a coluna selecionada)
3. selecionar a opção “consultar” (clcando na lupa)

Imagem 5: Consultas na tabela

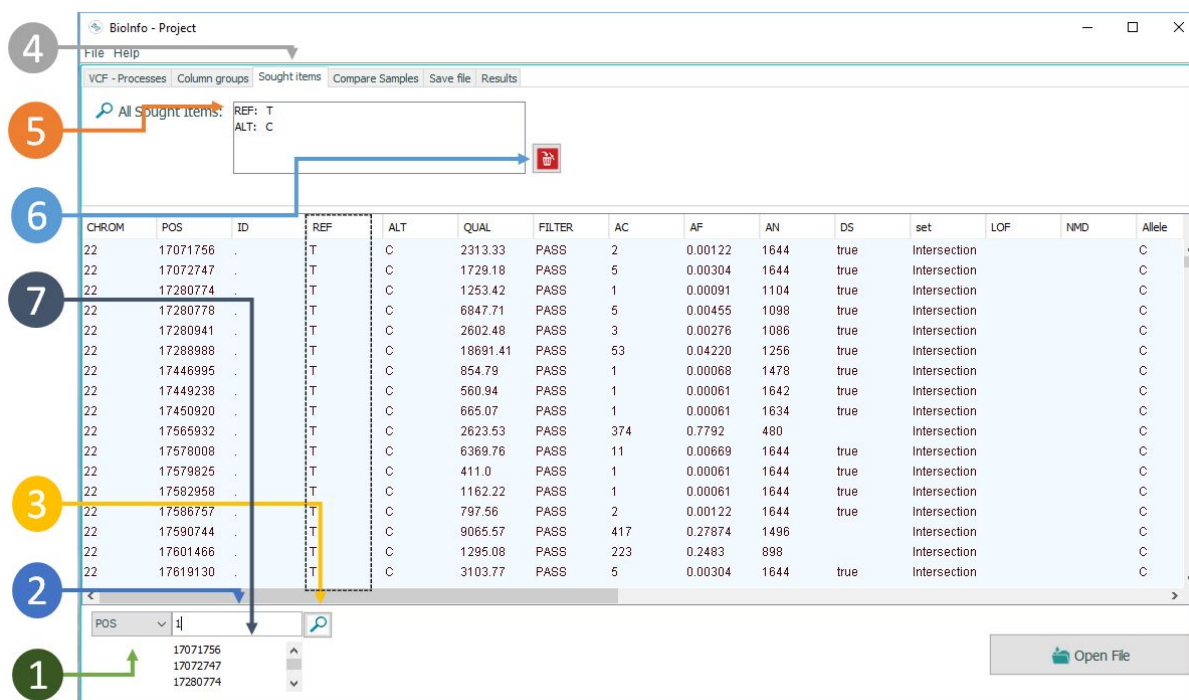


Tabela 3: Passos para realização de consulta na tabela

1	Selecionar da caixa de seleção de colunas, qual a coluna que se deseja realizar a consulta
2	Digitar o valor da expressão da consulta. Para alguns tipos, como ID e POS, por exemplo, aparecerá o autocomplete
3	Opção de consulta, que irá fazer o filtro da tabela de acordo com os parâmetros passados nos passos 1 e 2.
4	Aba “Sought items” que exhibe as informações de todas as consultas realizadas pelos passos anteriores.
5	Lista de itens consultados pela aplicação. Toda nova consulta nova realizada será adicionada nesta lista. Pode-se selecionar esses itens listados para que se possa utilizar a funcionalidade do passo 6.

6	Desfazer a consulta feita. Para que o item adicionado na lista de consultas realizadas seja removido basta selecionar o item na lista do passo 5. Isso irá atualizar a tabela com todos os outros itens anteriormente consultados, menos a consulta selecionada na lista do passo 5.
7	<i>Autocomplete</i> de expressões presentes na tabela, de forma a facilitar a seleção de consulta de expressões digitadas no passo 2.

4. SELECIONAR COLUNAS PARA VISUALIZAÇÃO

Para dar a possibilidade de visualizar menos ou mais colunas da tabela. Existe a funcionalidade de “*Column groups*”, a descrição de cada uma das funcionalidades envolvidas no processo serão mostrados a seguir:

Imagem 6: Seleção de colunas da tabela

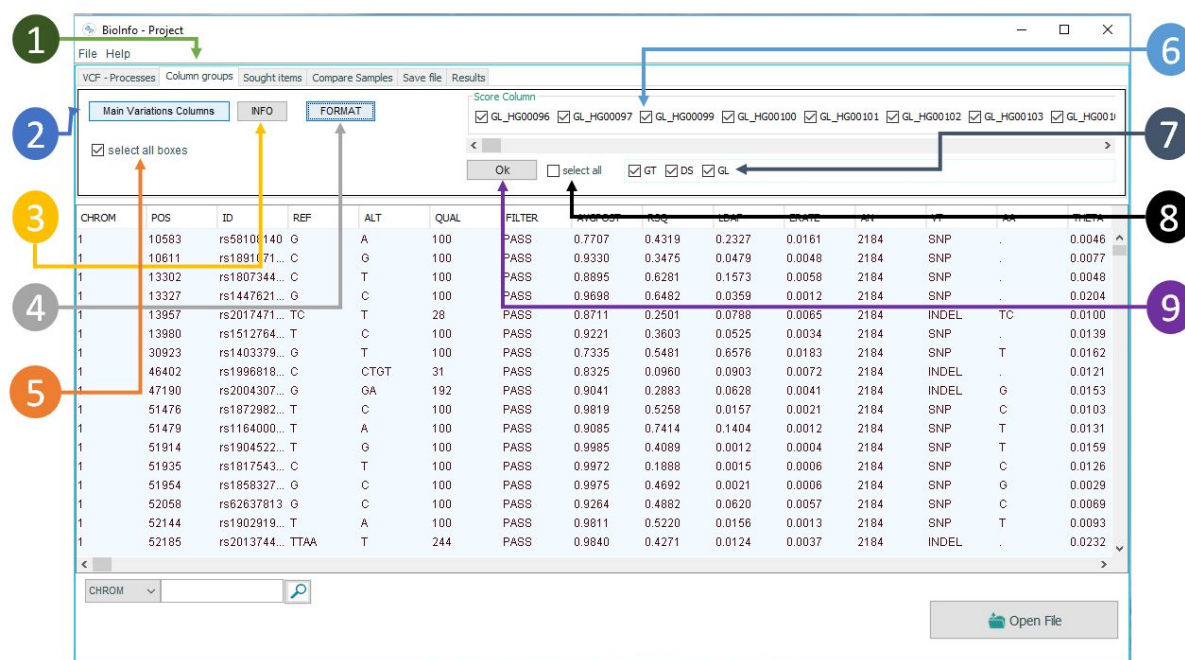


Tabela 4: Passos para seleção de exibição de colunas

1	Selecionar a aba “ <i>Column groups</i> ” para que seja exibida as funcionalidades referentes a seleção de colunas da tabela, de forma a visualizar itens de forma separada.
2	O botão “ <i>Main Variations Columns</i> ” exibe as caixas de seleção de colunas das primeiras colunas da tabela, que todo arquivo <i>VCF</i> deve possuir: <i>CHROM</i> , <i>POS</i> , <i>ID</i> , <i>REF</i> , <i>ALT</i> , <i>QUAL</i> e <i>FILTER</i>
3	O botão de seleção do campo <i>INFO</i> , exibe todas as caixas de seleção de coluna referentes ao campo <i>INFO</i> do arquivo original. Bem como as anotações realizadas pela aplicação <i>SNPEff</i> e <i>SNPSift</i> (que adicionam novas colunas no campo <i>INFO</i> do arquivo <i>VCF</i>).
4	O botão “ <i>FORMAT</i> ” exibe todas as caixas de seleção de colunas referentes ao campo <i>FORMAT</i> do arquivo <i>VCF</i> . É utilizado ao se analisar amostras diferentes num mesmo arquivo <i>VCF</i> . Podem possuir tipos como: <i>GT</i> (Genótipo), <i>DS</i> (Probabilidades do genótipo), entre outros.

5	Selecionar/desselecionar todas as colunas do arquivo. Isso fará com que a tabela exiba todas ou nenhuma coluna.
6	As caixas de seleção que devem ser exibidas ao executar os passos 2, 3 ou 4 são listados nesse painel de exibição de <i>Checkbox</i> . Caso a seleção tenha sido no passo 4, a listagem exibirá a coluna FORMAT referente a cada amostra presente no arquivo VCF (Ou seja, se o campo FORMAT possui GT(genótipo) para a amostra HG0001 e HG0002, serão exibidas as caixas de seleção GT_HG0001 e GT_HG0002.
7	Os campos gerais referentes a coluna FORMAT, exibido após a seleção do passo 4, exemplo: GT (“ <i>Genotype</i> ”), GL (“ <i>Genotype Likelihoods</i> ”), DS (“ <i>Genotype dosage from MaCH/Thunder</i> ”). Isso significa, que cada um desses pontos irá influenciar na seleção das caixas de seleção do passo 6. Ou seja, se eu tenho o campo GT(genótipo) para a amostra HG0001 e HG0002, no passo 6 serão exibidas as caixas de seleção GT_HG0001 e GT_HG0002, mas neste passo será exibida apenas a caixa de seleção “GT”, que dá a possibilidade de selecionar/desselecionar as todas as caixas GT_HG0001 e GT_HG0002, do passo 6.
8	Opção usada para selecionar todas as caixas de seleção dos passos 6 e 7 simultaneamente.
9	Botão “ok” é usado para confirmação de seleção das caixas dos passos 6 e 7. Quando se realiza a marcação das caixas de seleção nesses passos citados, é necessário que seja feita a confirmação neste botão.

5. COMPARAÇÃO DE AMOSTRAS

Esta seção da aplicação visa propor a proposta de análise entre amostras que o arquivo de anotação de variantes possuir). Esta funcionalidade somente estará disponível quando o arquivo de variantes analisado possuir várias amostras (no mínimo 3). Ela pode ser utilizada para estudar mutações que possuem caráter recessivo, ao analisar indivíduos com parentescos, quando se considera o pai (“*parent 1*”) a mãe (que então será “*parent 2*”) e o filho (“*child*”), o filtro então considerará a coluna GT(genótipo) de cada uma dessas amostras, e produzirá o resultado em que os dois pais sejam heterozigotos e o filho homozigoto (dominante ou recessivo).

Imagem 7: Análises de mutações de caráter recessivo/dominante

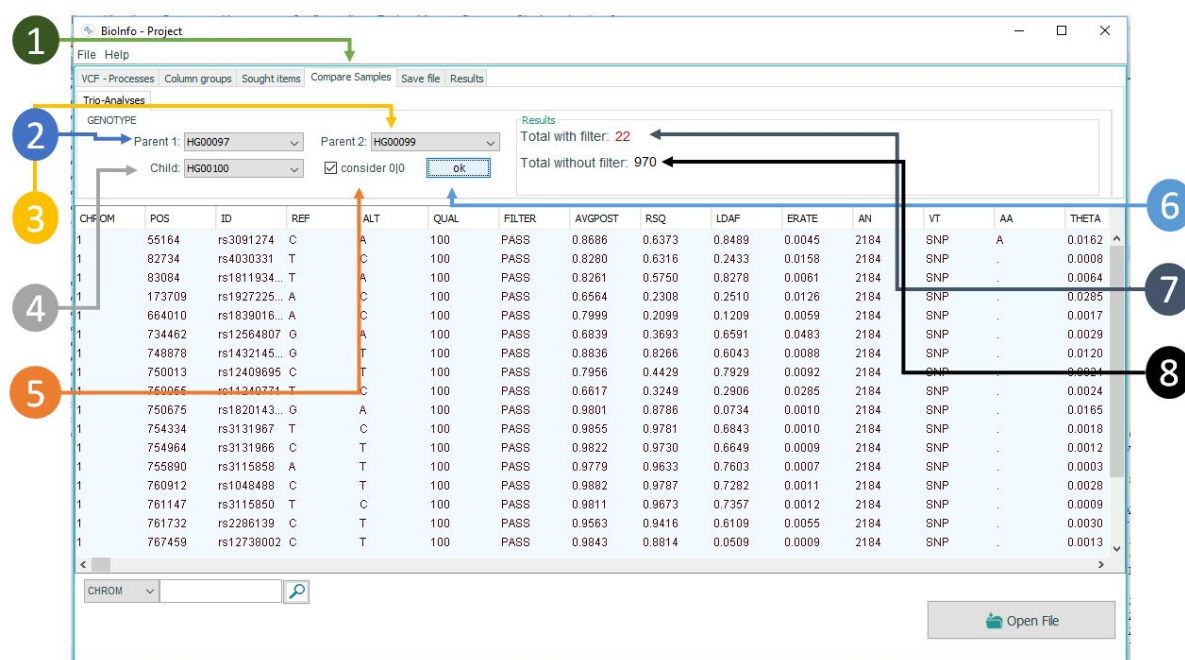


Tabela 5: Passos para comparar amostras

1	Selecionar a aba “Compare Samples” para que seja exibida as funcionalidades referentes ao filtro para análise de mutações parentais de caráter recessivo/dominante (“ <i>Trio-Analyses</i> ”)
2	Selecionar dentre as amostras, qual é a amostra do progenitor 1. (pai ou mãe)
3	selecionar dentre as amostras, qual é a amostra do progenitor 2 (uma opção diferente da do passo 2)
4	Selecionar dentre as amostras, qual é a amostra referente ao filho dos progenitores selecionados nos passos 2 e 3.
5	Selecionar/desselecionar a <i>checkbox</i> que considera(se estiver marcada) ou não(se desmarcada) a recessividade/dominância de homozigotos para alelos de referência. (do campo REF do arquivo vcf)

6	Botão “ok” inicia a consulta dos padrões de heterozigose, dos pais, e homozigose do filho, descritos nos passos 2, 3 e 4. Preenchendo como resultado os passos 7 e 8, bem como exibindo a tabela dos resultados correspondentes à análise.
7	O “ <i>Total with filter</i> ” presente no campo “ <i>results</i> ” de resultados gerais apresenta o total de variantes listadas após a análise este filtro.
8	O “ <i>Total without filter</i> ” presente no campo “ <i>results</i> ” de resultados gerais apresenta o total de variantes que estava listada, antes da análise do filtro.

Observação: Este filtro também é adicionado na lista de filtros realizados no passo 5 da seção de 3. CONSULTAS NA TABELA.

6. SALVAR TABELA NOS FORMATOS EXCEL OU CSV

Os arquivos da aplicação salva a tabela de resultados no estado atual da aplicação, ou seja, caso tenha-se realizado algum filtro, a aplicação irá salvar o arquivo considerando apenas os resultados obtidos após os filtros executados.

Imagem 8: Salvar tabela nos formatos “xls” (excel) ou “csv” (tabular)

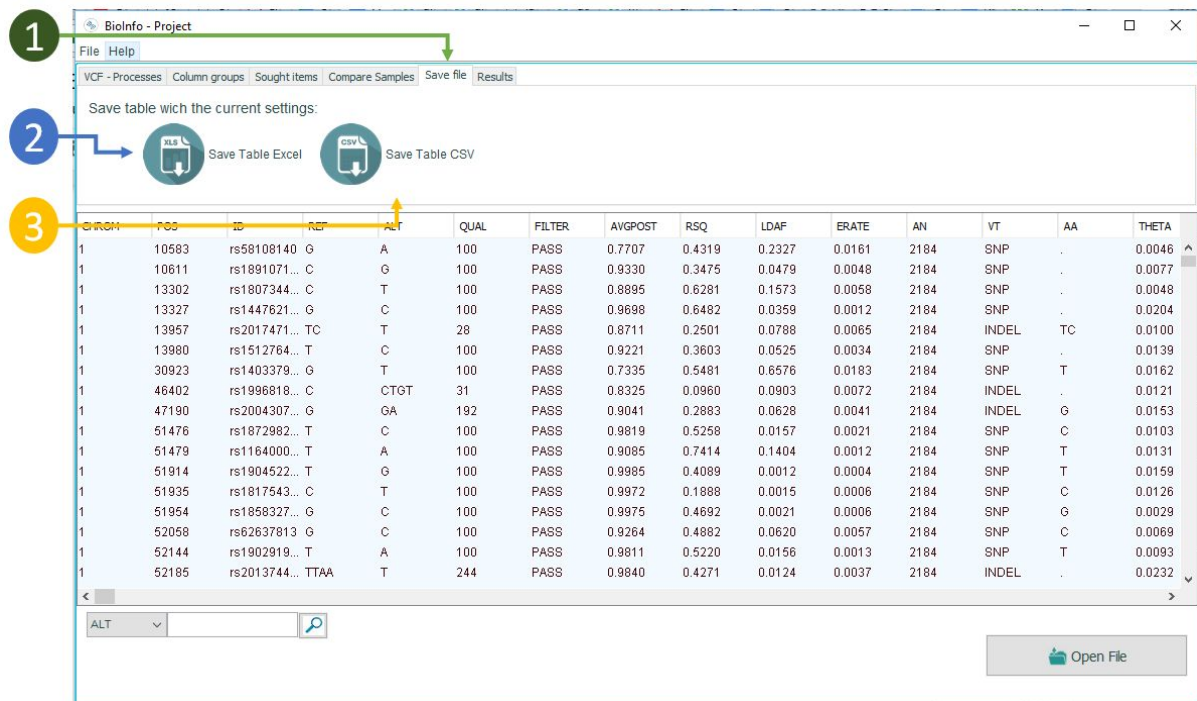


Tabela 6: Passos para salvar a tabela

1	Selecionar a aba “Save file” para que seja exibida as funcionalidades referentes ao salvamento dos dados da tabela de resultados, em seu formato atual (considerando os filtros que possam ter sido aplicados)
2	Selecionar “Save Table Excel” para salvar a tabela em seu estado atual no formato “xls” (Excel). (aparecerá em seguida a opção de seleção de repositório onde se deseja salvar o arquivo)
3	Selecionar “Save Table CSV” para salvar a tabela em seu estado atual no formato “csv” (Tabular). (aparecerá em seguida a opção de seleção de repositório onde se deseja salvar o arquivo)

7. VISUALIZAR RESULTADOS

Para que se possa visualizar alguns resultados gerais, referentes ao arquivo VCF selecionado esta seção é utilizada.

Imagem 9: Visualização de resultados gerais

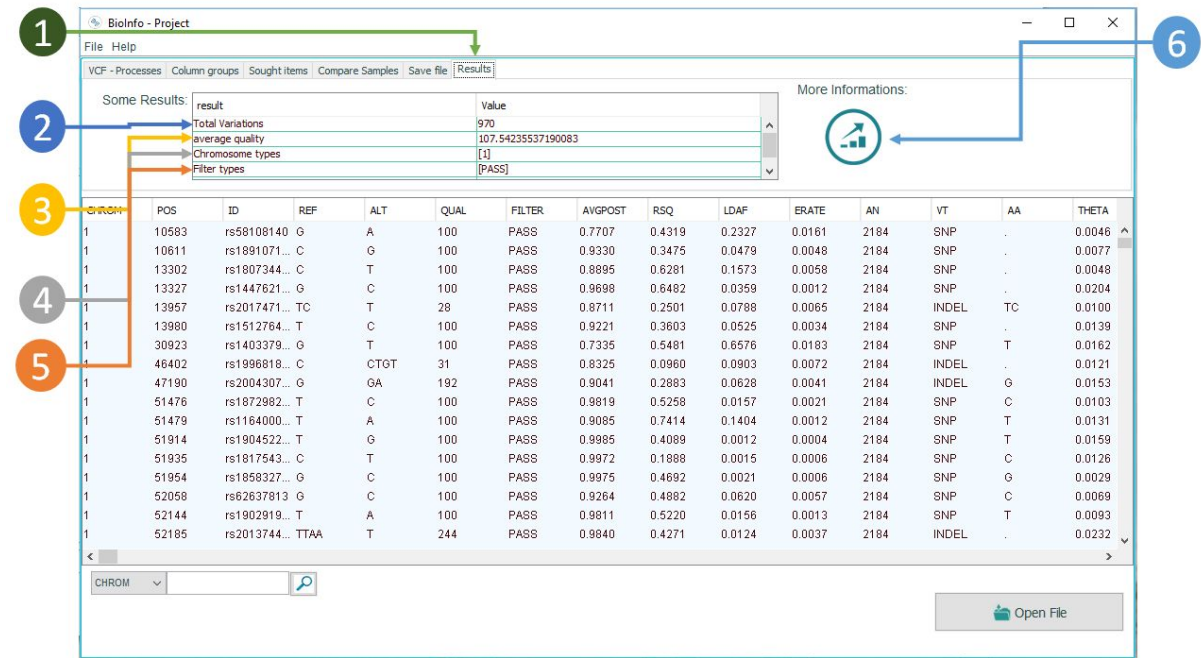


Imagem 10: Continuação de visualização de resultados gerais

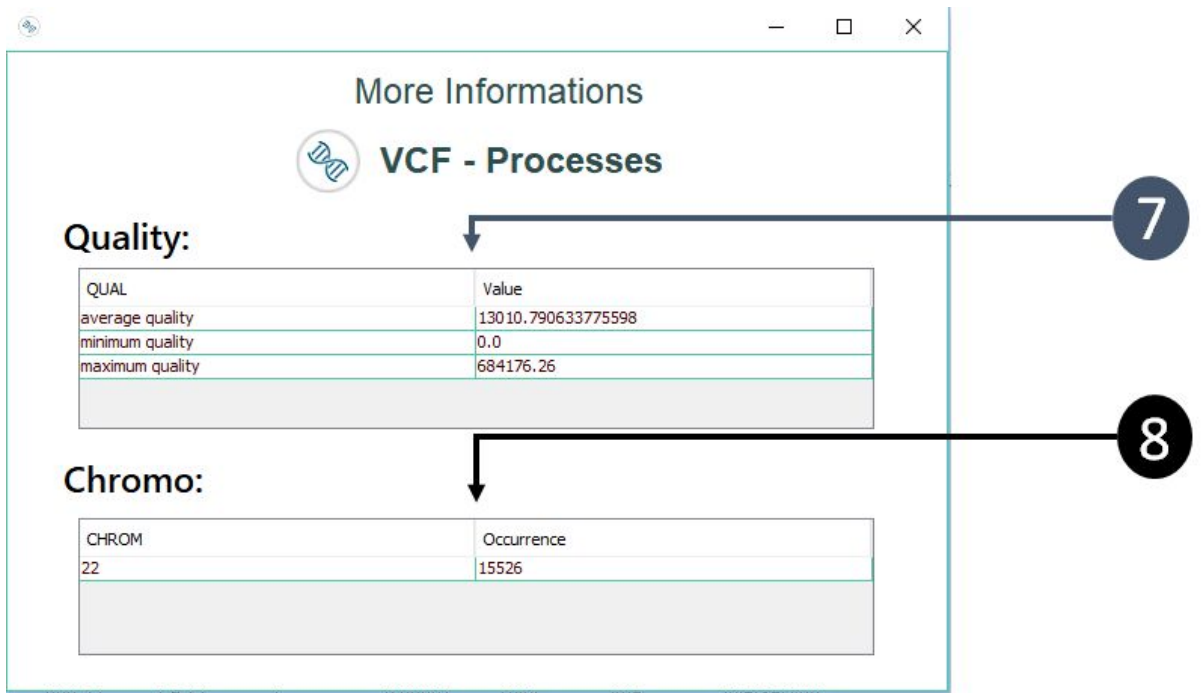


Tabela 7: Passos para visualizar resultados

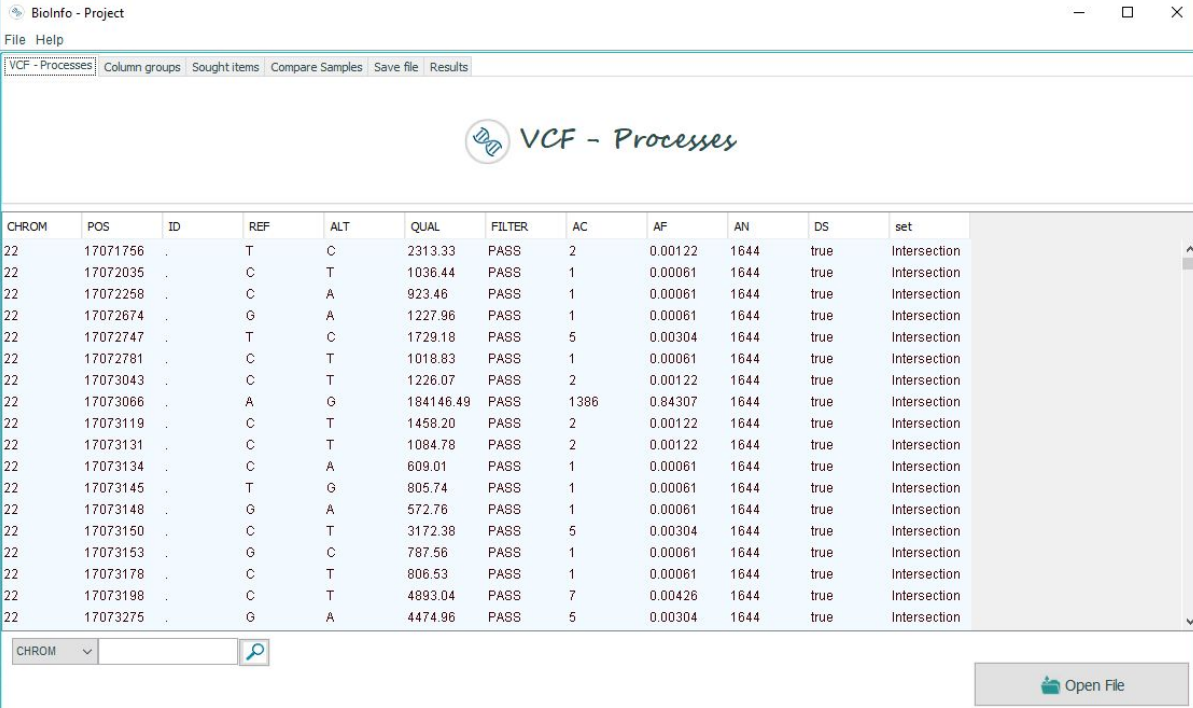
1	Selecionar a aba “Results” para que seja exibida as funcionalidades referentes aos resultados gerais do arquivo VCF selecionado
2	“ <i>Total Variations</i> ” é o total de variantes presentes no arquivo (linhas do arquivo VCF)
3	“ <i>Average Quality</i> ” é o valor da média das qualidades do arquivo (soma a qualidade de todas as variantes e divide pelo número de variantes total)
4	“ <i>Chromosome Types</i> ” o campo que exibe todas as ocorrências de tipos de cromossomos presentes no arquivo VCF. Exemplo, se existirem variantes anotadas para o cromossomo 22 e outras para o cromossomo 1, então este campo terá os valores: [1, 22] referentes a todos os tipos de cromossomos presentes no arquivo selecionado (coluna <i>CHROM</i> da tabela).
5	“ <i>Filter Types</i> ” é o campo que exibe todas as ocorrências de tipos de filtros presentes no arquivo VCF. Se existirem mais de um tipo, o resultado aparecerá: [tipo1, tipo2,...]
6	Botão de seleção para ampliar a visualização de resultados gerais. (abrirá outra tela para que se possa analisar algumas informações expandidas.
7	Informações relativas a qualidade das variantes. Com a qualidade média (“average quality”), o valor da qualidade da variante com menor valor (“ <i>minimum quality</i> ”) e o valor da variante com maior valor (“ <i>maximum quality</i> ”)
8	Exibe todas as ocorrências de tipos de cromossomos presentes no arquivo VCF e a quantidade de ocorrências de variantes para cada tipo de cromossomo presente no arquivo.

EXECUTANDO TESTE

Exemplo de Utilização Simples:

1. Abrir a aplicação e selecionar “Open File”. Depois clicar em “select file” e selecionar o arquivo “file.vcf” localizado na pasta “exemplos” (Caminho: **exemplo\file.vcf**)
2. clicar em “ok”, esperar carregar o arquivo e observar a tabela de resultados gerados pelo arquivo

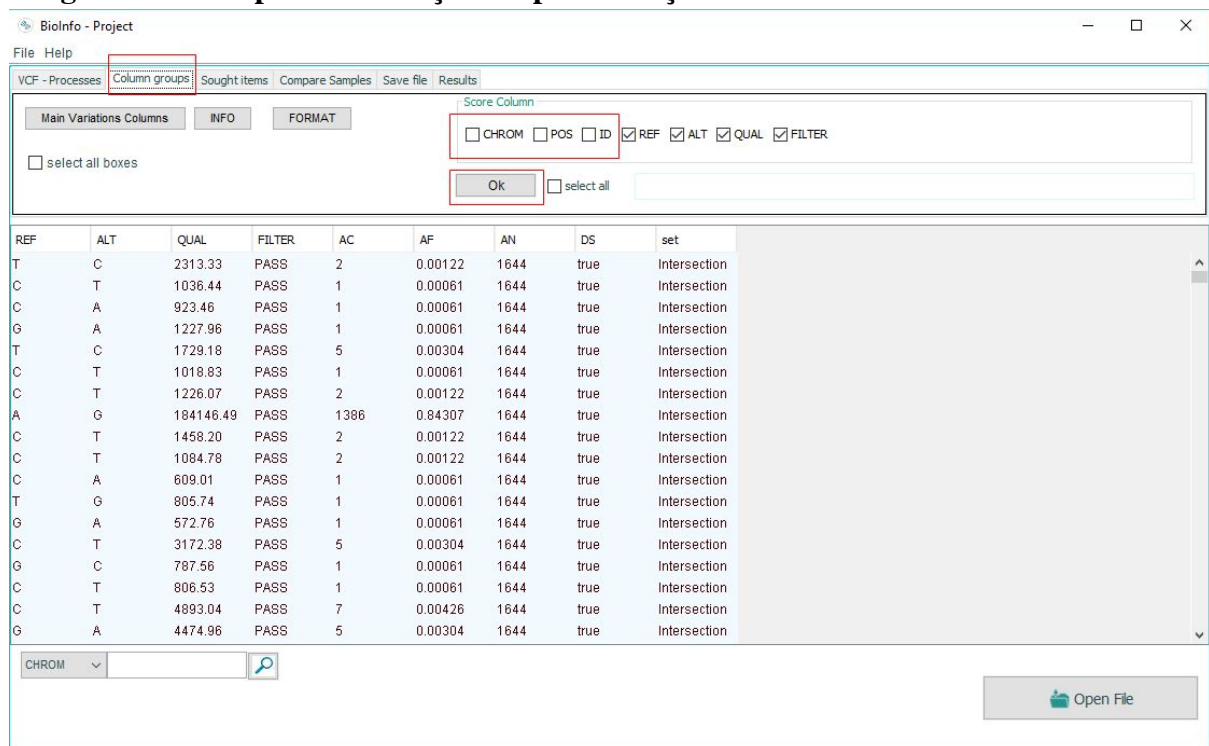
Imagem 11: Exemplo de utilização simples - arquivo file.vcf aberto



CHROM	POS	ID	REF	ALT	QUAL	FILTER	AC	AF	AN	DS	set
22	17071756	.	T	C	2313.33	PASS	2	0.00122	1644	true	Intersection
22	17072035	.	C	T	1036.44	PASS	1	0.00061	1644	true	Intersection
22	17072258	.	C	A	923.46	PASS	1	0.00061	1644	true	Intersection
22	17072674	.	G	A	1227.96	PASS	1	0.00061	1644	true	Intersection
22	17072747	.	T	C	1729.18	PASS	5	0.00304	1644	true	Intersection
22	17072781	.	C	T	1018.83	PASS	1	0.00061	1644	true	Intersection
22	17073043	.	C	T	1226.07	PASS	2	0.00122	1644	true	Intersection
22	17073066	.	A	G	184146.49	PASS	1386	0.84307	1644	true	Intersection
22	17073119	.	C	T	1458.20	PASS	2	0.00122	1644	true	Intersection
22	17073131	.	C	T	1084.78	PASS	2	0.00122	1644	true	Intersection
22	17073134	.	C	A	609.01	PASS	1	0.00061	1644	true	Intersection
22	17073145	.	T	G	805.74	PASS	1	0.00061	1644	true	Intersection
22	17073148	.	G	A	572.76	PASS	1	0.00061	1644	true	Intersection
22	17073150	.	C	T	3172.38	PASS	5	0.00304	1644	true	Intersection
22	17073153	.	G	C	787.56	PASS	1	0.00061	1644	true	Intersection
22	17073178	.	C	T	806.53	PASS	1	0.00061	1644	true	Intersection
22	17073198	.	C	T	4893.04	PASS	7	0.00426	1644	true	Intersection
22	17073275	.	G	A	4474.96	PASS	5	0.00304	1644	true	Intersection

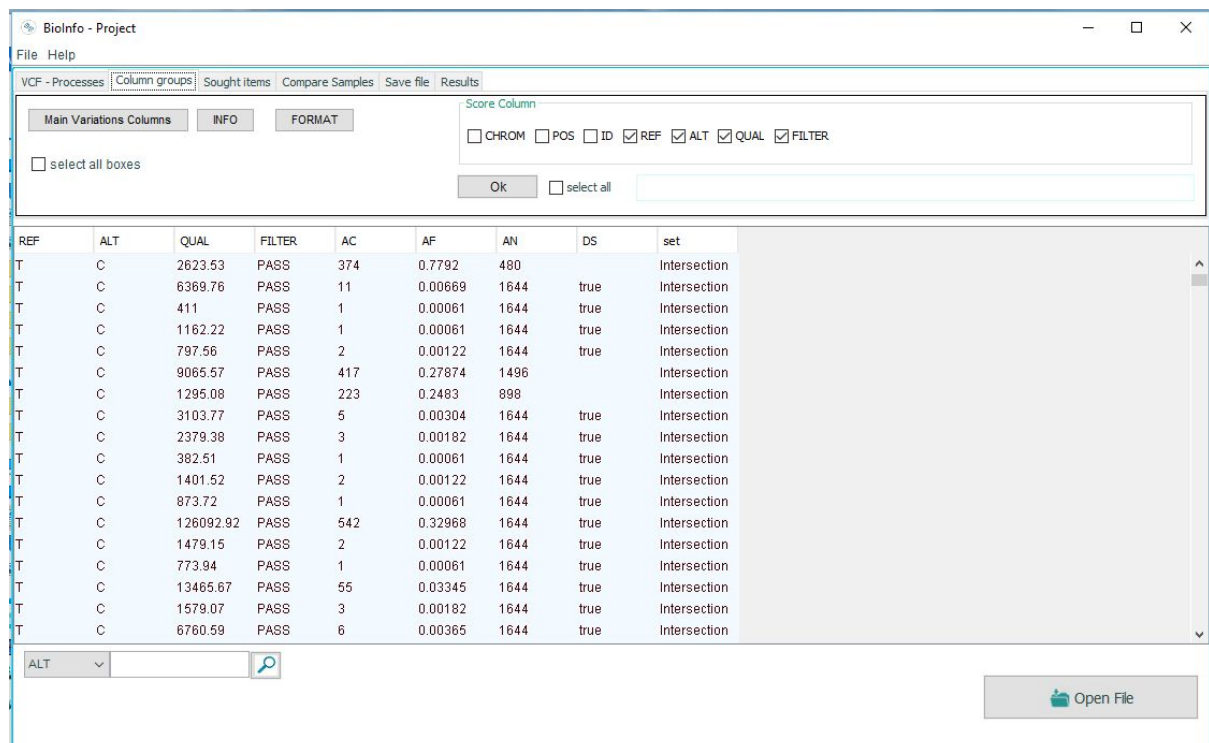
3. Selecionar a aba “column grups” e clicar no botão “Main Variations Columns”, depois desmarcar as três primeiras colunas do arquivo e clicar em “ok” em seguida (verificar que as três primeiras colunas da tabela (CHROM, POS e ID) desaparecem).

Imagem 12: Exemplo de utilização simples - seleção de colunas



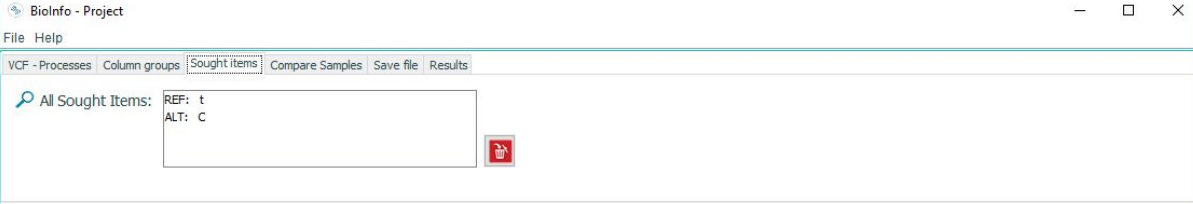
4. Realizar a consulta na tabela selecionando na lista de seleção de coluna o tipo “REF” com o valor “T” e clicar no botão de consulta (lupa) (verificar que os resultados da tabela exibem todos resultados que iniciam com a referência “T”). Depois realizar outra consulta, agora para a coluna “ALT” e colocar o valor “C” (verificar que agora são exibidos os valores da tabela com a coluna REF iniciada com valores T e a coluna ALT iniciada com valores C)

Imagem 13: Exemplo de utilização simples - Consulta REF “T” e ALT “C”



5. Clicar na aba “*Sought items*” e verificar que os dois filtros feitos no passo 4 foram listados.

Imagem 14: Exemplo de utilização simples - Consultas listadas



REF	ALT	QUAL	FILTER	AC	AF	AN	DS	set
T	C	2623.53	PASS	374	0.7792	480		Intersection
T	C	6369.76	PASS	11	0.00669	1644	true	Intersection
T	C	411	PASS	1	0.00061	1644	true	Intersection
T	C	1162.22	PASS	1	0.00061	1644	true	Intersection
T	C	797.56	PASS	2	0.00122	1644	true	Intersection

6. clique em remover a primeira consulta de REF de valor T (observar que agora a tabela lista os valores somente de alterações que iniciam com C (resultado do filtro que não foi apagado).

7. clicar na aba “Save file” e salvar o arquivo no formato excel (clicando no “*Save Table Excel*”) Verificar que o arquivo excel gerado possui exatamente os valores atuais da tabela depois do filtro (todos os valores com REF iniciados com o valor C).

8. clicar na aba “Results” e analisar os resultados gerais: total de variantes “2429”, qualidade média “21585”, ocorrências de Cromossomos [22]

Exemplo de Utilização Simples para Trio-Analysis:

1. Abrir a aplicação e selecionar “Open File”. Depois clicar em “select file” e selecionar o arquivo “1kg.head_chr1.vcf” localizado na pasta “exemplos” (Caminho: **exemplo\1kg.head_chr1.vcf**)
2. clicar em “ok”, esperar carregar o arquivo e observar a tabela de resultados gerados pelo arquivo
3. ir na aba “*Compare Samples*” e selecionar com “Parent 1” a amostra HG00097, “Parent2” para amostra HG00099 e “Child” para amostra HG00100.
 - a. caso se deseje considerar os alelos referência, deve-se selecionar “consider 0|0”
 - b. caso não se deseje considerar os casos de alelos referência, basta deixar a caixa “consider 0|0” desmarcada
4. clicar em ok e verificar que as variações listadas possuem o resultado para o “Parent 1”(HG00097) a coluna GT_HG00097 como heterozigoto e “Parent 2”(HG00099) a coluna GT_HG00097 como heterozigoto e o “Child” (HG00100) a coluna GT_HG00100 como homozigoto. Revelando assim, as amostras de mutações com caráter recessivo/dominante

Imagem 15: Exemplo de utilização simples - Trio-Analyses

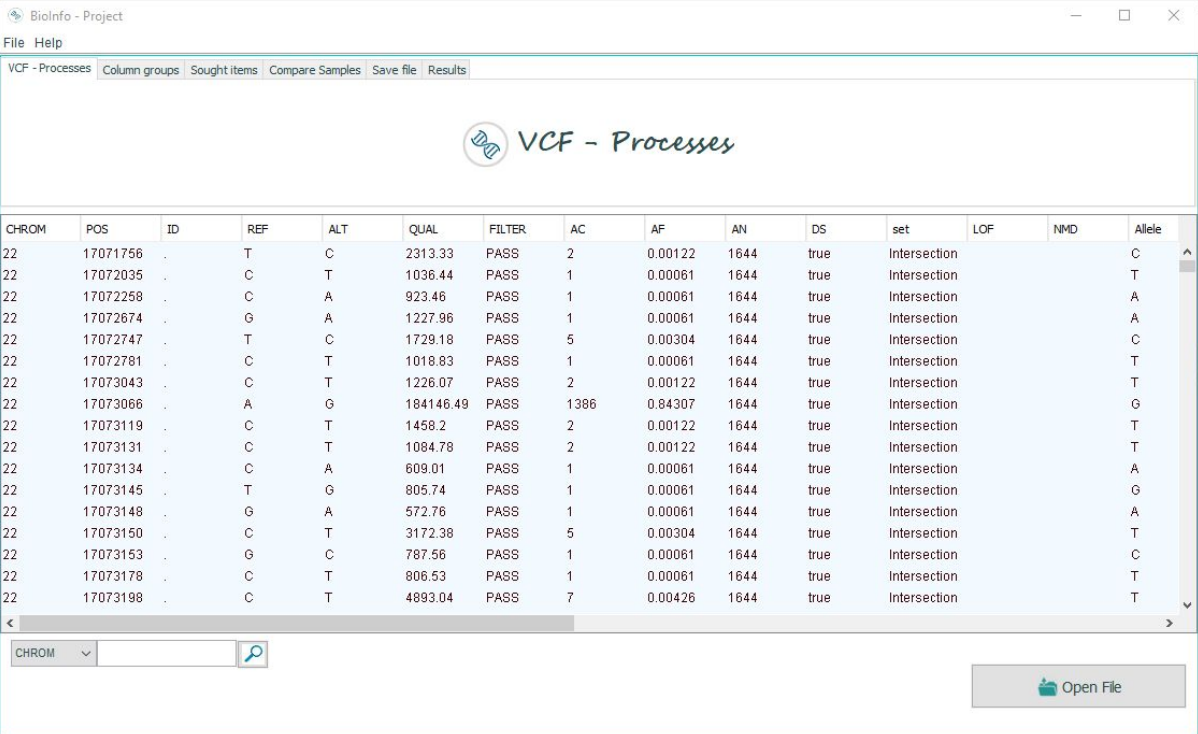
The screenshot shows the 'Trio-Analyses' window in a software application. At the top, there's a menu bar with 'File' and 'Help'. Below it, a tab bar shows 'VCF - Processes', 'Column groups', 'Sought items', 'Compare Samples', 'Save file', and 'Results'. The 'Trio-Analyses' section has a 'GENOTYPE' dropdown set to 'Parent 1: HG00097', 'Parent 2: HG00099', and 'Child: HG00100'. There's a checkbox for 'consider 0/0' and an 'ok' button. The 'Results' section shows 'Total with filter: 14' and 'Total without filter: 970'. Below this is a table with 15 columns: CHROM, POS, ID, REF, ALT, QUAL, FILTER, AVGPOST, RSQ, LDAF, ERATE, AN, VT, AA, and THETA. The table contains 15 rows of data. At the bottom, there's a 'CHROM' dropdown and a search icon. An 'Open File' button is located at the bottom right.

CHROM	POS	ID	REF	ALT	QUAL	FILTER	AVGPOST	RSQ	LDAF	ERATE	AN	VT	AA	THETA
1	55164	rs3091274	C	A	100	PASS	0.8686	0.6373	0.8489	0.0045	2184	SNP	A	0.0162
1	83084	rs1811934...	T	A	100	PASS	0.8261	0.5750	0.8278	0.0061	2184	SNP	.	0.0064
1	734462	rs12564807	G	A	100	PASS	0.6839	0.3693	0.6591	0.0483	2184	SNP	.	0.0029
1	748878	rs1432145...	G	T	100	PASS	0.8836	0.8266	0.6043	0.0088	2184	SNP	.	0.0120
1	750013	rs12409695	C	T	100	PASS	0.7956	0.4429	0.7929	0.0092	2184	SNP	.	0.0024
1	754334	rs3131967	T	C	100	PASS	0.9855	0.9781	0.6843	0.0010	2184	SNP	.	0.0018
1	754964	rs3131966	C	T	100	PASS	0.9822	0.9730	0.6649	0.0009	2184	SNP	.	0.0012
1	755890	rs3115858	A	T	100	PASS	0.9779	0.9633	0.7603	0.0007	2184	SNP	.	0.0003
1	760912	rs1048488	C	T	100	PASS	0.9882	0.9787	0.7282	0.0011	2184	SNP	.	0.0028
1	761147	rs3115850	T	C	100	PASS	0.9811	0.9673	0.7357	0.0012	2184	SNP	.	0.0009
1	761732	rs2286139	C	T	100	PASS	0.9563	0.9416	0.6109	0.0055	2184	SNP	.	0.0030
1	768253	rs2977608	A	C	100	PASS	0.9610	0.9545	0.5279	0.0035	2184	SNP	.	0.0018
1	774736	rs28830877	A	C	100	PASS	0.9492	0.7018	0.9372	0.0031	2184	SNP	.	0.0047
1	774923	rs28769076	G	C	100	PASS	0.9333	0.7068	0.9101	0.0077	2184	SNP	.	0.0010

Exemplo de anotação com SNPEff

1. Abrir a aplicação e selecionar "Open File". Depois clicar em "select file" e selecionar o arquivo "file.vcf" localizado na pasta "exemplos" (Caminho: **exemplo\file.vcf**)
2. Clicar em "Config" e verificar que a database "GRCh37.75" está selecionada para o campo SNPEff (isso significa que a anotação a ser adicionada no arquivo utilizará o banco "GRCh37.75" para gerar as anotações no arquivo atual. (clicar em "ok" na tela de "Annotators databases")
3. selecionar o botão "Execute SnpEff" e depois clicar em "ok" e aguardar o processamento de anotação do arquivo.
4. Após o processamento (pode-se passar vários minutos para isso), verificar que é exibida a tabela com as novas anotações de efeitos são geradas pelo SNPEff. (Observar também que no repositório onde foi aberto o arquivo "file.vcf" está mostrando também um novo arquivo chamado "file.SnpEffAnnotated.vcf")

Imagem 16: Anotação SNPEff - Abertura do arquivo anotado pelo SNPEff



VCF - Processes

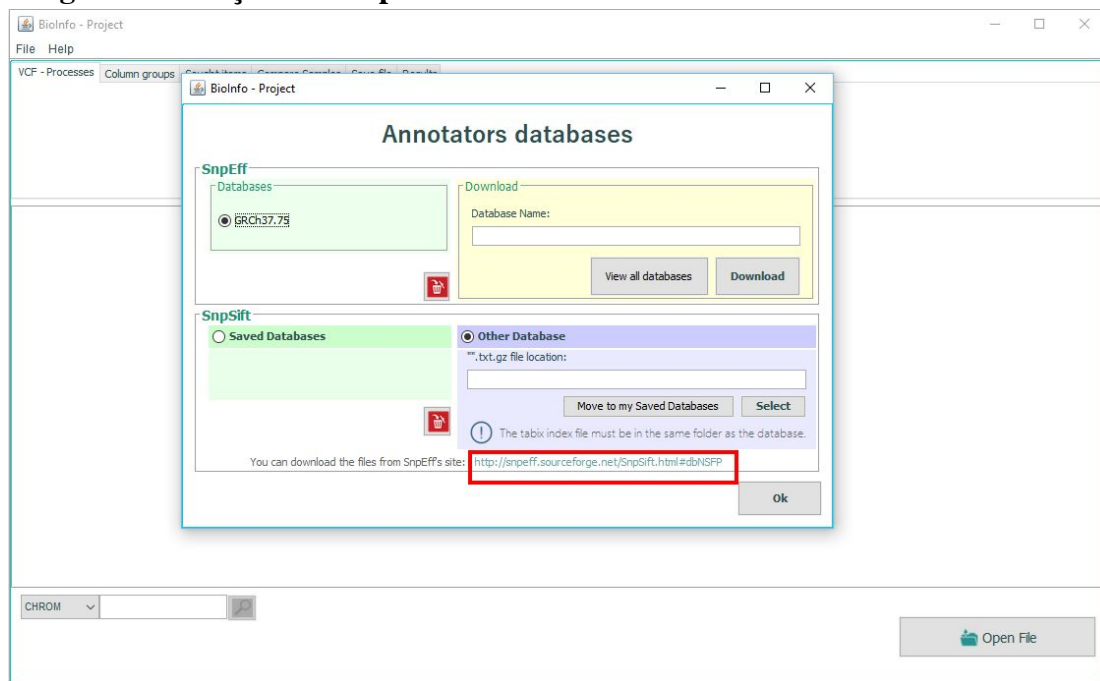
CHROM	POS	ID	REF	ALT	QUAL	FILTER	AC	AF	AN	DS	set	LOF	NMD	Allele
22	17071756	.	T	C	2313.33	PASS	2	0.00122	1644	true	Intersection			C
22	17072035	.	C	T	1036.44	PASS	1	0.00061	1644	true	Intersection			T
22	17072258	.	C	A	923.46	PASS	1	0.00061	1644	true	Intersection			A
22	17072674	.	G	A	1227.96	PASS	1	0.00061	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

Exemplo de anotação com SNPSift

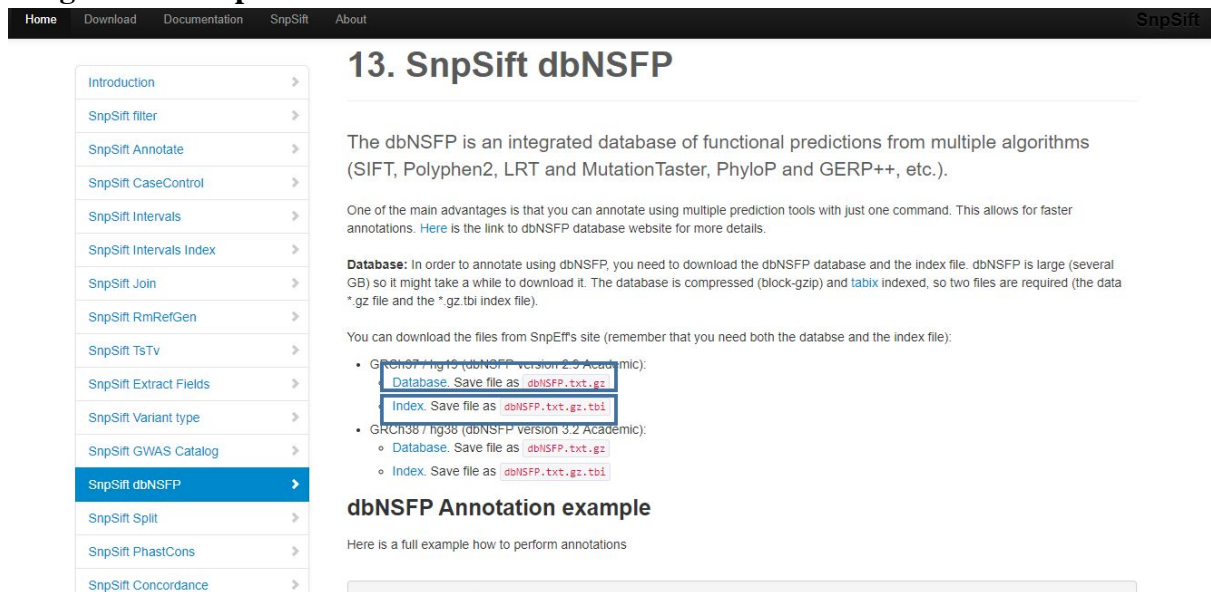
5. Abrir a aplicação e selecionar “Open File”. Depois clicar em “select file” e selecionar o arquivo “file.SnpEffAnnotated.vcf” gerado no cenário anterior, localizado na pasta “exemplos” (Caminho: **exemplo\file.SnpEffAnnotated.vcf**)
6. Clicar em “Config” e verificar que a database ainda não existem databases salvas para o SnpSift.

Imagem 17: Seleção de site para Download de database



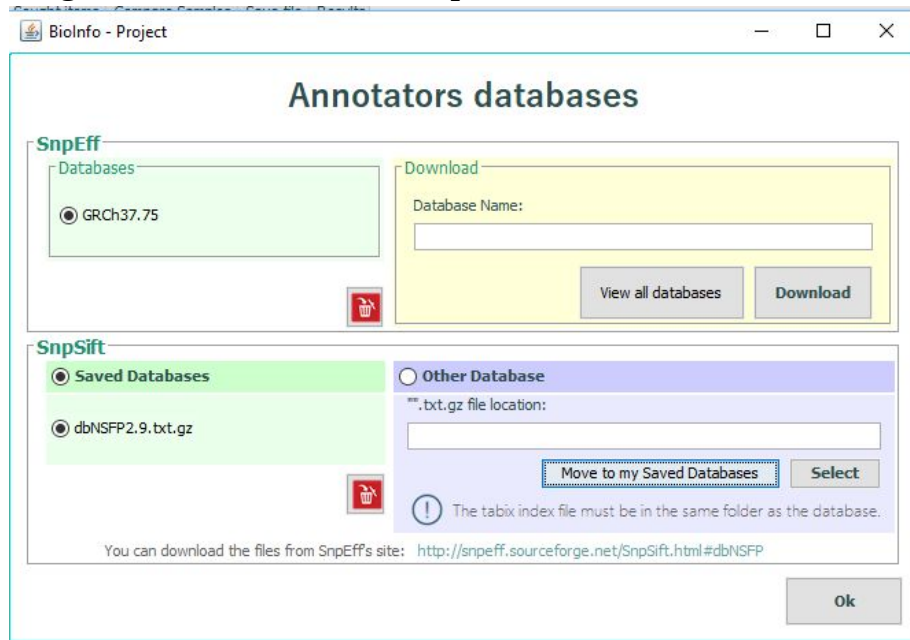
7. Clicar no link exibido na aplicação e baixar os arquivos "dbNSFP.txt.gz" e "dbNSFP.txt.gz.tbi" respectivamente.

Imagem 18: Site para download de database



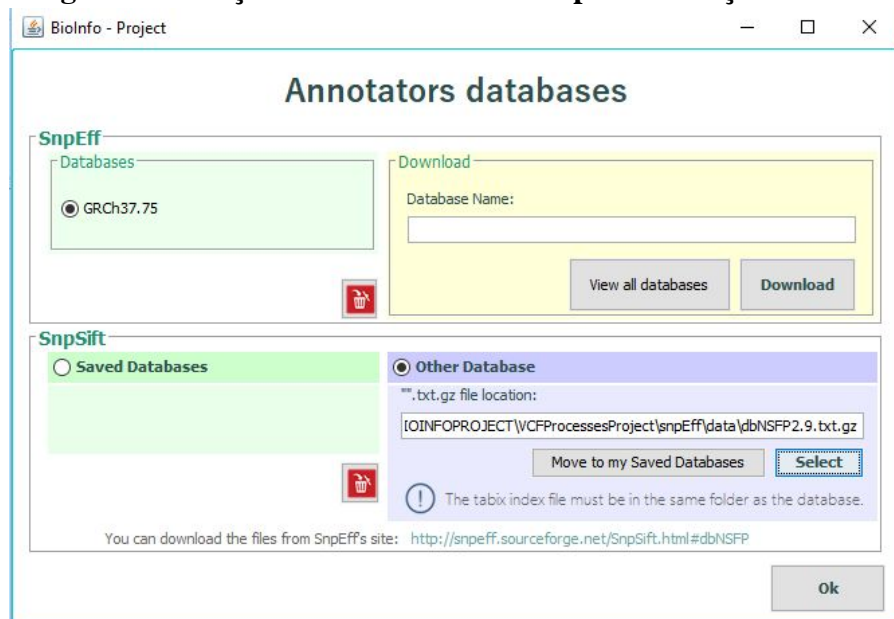
8. selecionar o botão “select” e selecionar o arquivo "dbNSFP.txt.gz" (Observe que o arquivo “dbNSFP.txt.gz.tbi” precisa estar no mesmo repositório que o arquivo "dbNSFP.txt.gz")
 - a. pode-se selecionar o botão “Move to my Saved Databases” e verificar que o arquivo baixado é adicionado na lista de databases salvas da aplicação “Saved Databases”

Imagem 19: Adicionar database para a lista de “Saved Databases”



- b. pode-se apenas manter a opção “Other Database” para que o sistema utilize a database por fora da aplicação

Imagem 20: Seleção de “Other Database” para execução do SNPSift



9. clicar em “ok” (irá retornar para o caso de uso de “Open File”). Selecionar “Execute SnpSift” e depois clicar em “ok” e aguardar o processamento de anotação do arquivo.
10. Após o processamento (pode-se passar vários minutos para isso), verificar que é exibida a tabela com as novas predições de relevância de variantes que são geradas pelo SNPSift. (Observar também que no repositório onde foi aberto o arquivo “file.SnpEffAnnotated.vcf” está mostrando também um novo arquivo chamado

“file.SnpEffAnnotated.SnpSiftAnnotated.vcf” que possui as anotações adicionadas pelo SNPSift utilizando o banco GRCh37 / hg19 do arquivo "dbNSFP.txt.gz")