8 Version History
8.1 2.16 : April 30, 2019
8.2 2.15.1 : March 19, 2019
8.3 2.15.1 : March 08, 2019
8.4 2.15 : January 4th, 2018
8.5 2.14 : September 17, 2018
8.6 2.13 : April 20, 2018
8.7 2.12.1 : March 07, 2018
8.8 2.12 : February 05, 2018
8.9 2.11.1 : October 05, 2017
8.10 2.11 : October 03, 2017
8.11 2.10.2 : September 06, 2017
8.12 2.10.1 : August 25 , 2017
8.13 2.10 : August 2017
8.14 2.06 : February 2017
8.15 2.05 : November 2016
8.16 2.04 : August 2016
8.17 2.01 to 2.03
8.18 2.00 : June 2016

9 Frequently asked questions

10 Contact us
1 Introduction

VarAFT (Variant Annotation and Filter Tool) is a tool used to annotate and filter out variant files. VarAFT allows the comparison of several individuals and the collection of relevant information about variations. A direct link to IGV is available to visualize any variations using a 'BAM' file. VarAFT includes a coverage analysis module to easily visualize regions that are poorly covered through tables and dynamic charts.

2 Installation

VarAFT is available for Linux, Mac and Windows.
WARNING

JAVA version 1.8 64-bits is necessary.
The OpenJDK version of JAVA is not supported. VarAFT could be launched with JAVA 9 or higher.

2.1 Windows

For the first installation, you should use the "VarAFT_Setup_First_Install.exe" file to install 'VarAFT', 'Perl', 'Annovar', 'IGV' and 'BEDtools'. If you want to update VarAFT, you should use the 'VarAFT_Setup.exe' file.

If you have no administrator rights, you can download the 'no admin rights' version. This version will setup VarAFT in your user folder.

After a successful installation, the program can be started using the VarAFT.exe file.

NOTE

By default the tool utilize 4GB of RAM (basic configuration for a recent desktop computer). If you have a 32-bits Windows OS and less than 3GB of RAM, you must launch "C:\Program Files (x86)\VarAFT_min.exe" to start VarAFT. To use more memory you can launch VarAFT with the VarAFT_8G.exe, VarAFT_16G.exe or VarAFT_32G.exe.

2.2 MAC

Download the "*.pkg" file for MAC OS. Double click on the file. Accept the license and follow instructions. Wait until the copy finish. VarAFT will be available in your "Applications" folder.

Figure 2: Installation of VarAFT on MAC
VarAFT was built to use 8Gb of authorized max RAM. To change this value you can modify this file: "~/Applications/VarAFT.app/Contents/Java/VarAFT.cfg".

```
sudo vi ~/Applications/VarAFT.app/Contents/Java/VarAFT.cfg

[JVMUserOptions]
-Xmx=8192m <— Change this value
-Xms=128m
-XX:+UseCompressedOops
```

2.3 UNIX

**Debian**

A "*.deb" file is available to install VarAFT on any Debian based OS, such as Ubuntu, Mint and others. Double click on the file and click on the "Install" button. The application will be available in your start menu.

VarAFT was built to be launched with 8Gb of authorized max RAM. To change this value you can change this file: "~/opt/VarAFT/app/VarAFT.cfg".

```
sudo vi ~/opt/VarAFT/app/VarAFT.cfg

[JVMUserOptions]
-Xmx=8192m <— Change this value
-Xms=128m
-XX:+UseCompressedOops
```

**NOTE**

The "*.deb" file was generated and tested on Mint 17.3. It was successfully tested on Mint 18 and Ubuntu 16.06.

**RedHat**

A "*.rpm" file is available to install VarAFT for RedHat based OS such as Fedora, CentOS, Scientific Linux, and others. Double click on the file and click on the "Install" button. The application will be available in your start menu.

You can also proceed with the installation with the following command lines:
cd path_to_my_varaft_rpm_package
sudo yum install varaft-2.13-1.x86_64.rpm

VarAFT was built to be launched with 8Gb of authorized max RAM. To change this value you can edit this file: `/opt/VarAFT/app/VarAFT.cfg`.

```
sudo vi /opt/VarAFT/app/VarAFT.cfg

[JVMUserOptions]
-Xmx=8192m ← Change this value
-Xms=128m
-XX:+UseCompressedOops
```

**NOTE**
The "*.rpm" file was generated and tested on FEDORA 23. It was also successfully tested on RedHat 6.3, 7.3 and 7.4.

**Other**
On unix system, if you are not able to install VarAFT with a pre-built package, you can download the "ZIP" version and launch VarAFT from the jar file. Once downloaded, unzip the folder and type the following command lines:

```
cd path_to_my_varaft_folder
java -jar -Xmx8192m VarAFT_2.13.jar
```
3 Settings

VarAFT standard settings can be edited and adjusted to your needs.

3.1 General

- **Path to IGV**: Select the path of IGV tool. IGV allows visualization of BAM files. By default IGV is provided within VarAFT and the path is automatically set during the initial installation.

- **Path to Annovar database**: Select the directory where you want to store all needed files for the annotation process. These files need at least 50gb to be all downloaded. Please make sure you have enough free space in your computer. By default Annovar database is stored in your user folder.

- **Path to VarAFT_Project**: Select the directory where you want to store your projects. By default the folder VarAFT_Project is stored in your user folder.

![Settings Panel to configure VarAFT](image)
• **Path to dblocal default file:** Select the path of your local database. A specific VarAFT module allows for generation of this file from your VCF files.

• **Number of thread allowed:** Set the max number of allowed threads for annotation and coverage analysis. Please check your performance computer before changing this value. Note that you have less than 3gb of ram or a 32-bit system, it is advised to set this parameter to 1. An optimal value is the total number of cores - 2.

### 3.2 VCF

This section allows the configuration of VarAFT for the conversion of the VCF file.

![Settings Panel to configure VarAFT : VCF conversion](image)

- **1** Variants with an allele frequency lower than this value are considered as reference homozygous and are excluded from the analysis.

**NOTE**

Set 0 to keep all variants for example in the case of mosaicism or somatic events.
• 2 Variants with an allele frequency greater than this value are considered as homozygous.

• 3 Variants with an allele frequency greater than the first value and lower than the second are considered as heterozygous. Variants with an allele frequency lower than the first value (and greater than value of part 1) will be considered as uncertain heterozygous.

3.3 Annotation

Figure 5: Settings panel to configure VarAFT: Annotation

This section allows the configuration of VarAFT for the annotation process.

• 1 Check this box if you want to download and use KAVIAR database.

• 2 Check this box if you want to download and use HRCR1 database.

• 3 Check this box if you want to download and use gnomAD Genome database. gnomAD exome is automatically used.

• 4 Check this box if you want to download and use the Great Middlde East database.

• 5 Check this box if you want to download and use Iranome Database.
• 6 Set the distance between splicing variants and exon/intron boundaries. These variants will be flagged as splicing instead of intronic.

• 7 Check this box if you want to get all possible types of variants for all transcripts. By default only the most relevant type is displayed.

3.4 Tools

![Settings panel to configure VarAFT: Tools](image)

VarAFT uses API from UMD-Predictor and Human Splicing Finder to annotate variants. This section allows users to set login information for these 2 annotation systems.

3.4.1 UMD-Predictor

UMD-Predictor needs a token to retrieve predictions with the API. To register please go to [http://umd-predictor.eu](http://umd-predictor.eu) in order to get an unique token and paste it (1) (Ctrl-C even on MAC OS). The token in UMD-Predictor is available in the account section. The registration is free for academic and non profit use.
3.4.2 Human Splicing Finder

The HSF API requires a login and password. To register click on "Register to HSF API". Once the registration is done you have to validate your email with the mail that you will receive from HSF. Next, set your email and password used to register to HSF. To test if the connection is successful, click on "Test HSF Webservice Connection".
3.5 Table content

![Figure 7: Settings panel to configure VarAFT: Result final table content](image)

This section allows the selection of informations to include in the final result table. This part does not correspond to the settings of annotation process, but rather to which information you would like to see within the variant result table. Even if not selected, all these informations will always be available in the variant panel associated with the result table (shown in §39). Please note that, more columns you add, more memory will be required for the application to run.

The available annotations are detailed below:

**Gene Annotation**

The RefGene information is mandatory. To perform filter on Ensembl model you must select associated informations.

- **(REQUIRED) RefGene Function**: localization of the variation (exonic, splicing, intronic, UTR, ...)
- **(REQUIRED) Gene Symbol**: official gene symbol
• **(REQUIRED)RefGene Exonic Function:** impact of the variation (synonymous, non-synonymous, stop, frame shift, non frameshift ...)

• **(REQUIRED)RefGene AAChange:** impact of the variation on each transcripts and proteins according to the HGVS nomenclature (c. and p.)

• **Ensembl Function:** localization of the variation (exonic, splicing, intronic, UTR, ...) based on Ensembl data

• **Ensembl Genename**: Ensembl unique gene id (ENSG00000XXXXX)

• **Ensembl Exonic Function:** impact of the variation (synonymous, non-synonymous, stop, frame shift, non frame shift ...) based on Ensembl data

• **Ensembl AAChange:** impact of the variation on the transcripts and proteins according to the HGVS nomenclature (c. and p.) based on Ensembl data

**Prediction Annotation**

• **SIFT:** SIFT pathogenicity prediction.

• **PolyPhen2 HDIV:** Polyphen2 pathogenicity predictions based on the HDIV dataset.

• **PolyPhen2 HVAR:** PolyPhen2 pathogenicity predictions based on the HVAR dataset.

• **LRT:** Predictions from LRT.

• **MutationTaster:** Predictions from MutationTaster.

• **MetaLR:** Predictions from MetaLR system.

• **Mutation Assessor:** Predictions from Mutation Assessor.

• **FATHMM:** Predictions from FATHMM tool.

• **MetaSVM:** Predictions from MetaSVM.

• **CADD:** Phred scores from CADD.

• **UMD-Predictor:** Predictions from UMD Predictor tool

• **Provean:** Predictions based on Provean tool.

• **GERP++:** Prediction scores from GERP++ tool.

• **PhyloP7way:** PhyloP scores based on multiple alignments of 7 genomes. The larger the score, the more conserved is the site.

• **PhyloP20way:** PhyloP scores based on multiple alignments of 20 genomes. The larger the score, the more conserved is the site.

• **SiPhy29way:** SiPhy scores based on multiple alignments of 29 genomes. The larger the score, the more conserved is the site.
- **VEST3**: Prediction scores based on VEST3 tool.
- **cosmic70**: Data from the COSMIC database.
- **clinvar**: Data from ClinVar database.
- **phastCons7way**: phastCons scores based on multiple alignments of 7 genomes. The larger the score, the more conserved is the site.
- **phastCons20way**: phastCons scores based on multiple alignments of 20 genomes. The larger the score, the more conserved is the site.
- **DANN**: Prediction scores from the DANN tool.

**Frequency Annotation**

- **DbSNP ID**: rsID of dbsnp from selected version
- **1000 Genomes**: Observation frequencies from the 1000 Genomes project version of October 2014 with all ethnicities.
- **gnomAD XXX**: Frequency of observation for the Genome Aggregation Consortium. Frequency are provided for genome or exome.
- **KAVIAR**: Variant frequencies from the Known Variant Project (170 Million variants).
- **HRCR1**: Variant observation frequencies from the Haplotype Reference Consortium (40 Million variants).
- **GME**: Variant observation frequencies from the Great Middle East database.
- **Iranome**: Variant observation frequencies from the Iranome database.

**Frequency Annotation**

Change font size for the table which contains results

### 3.6 Versionning

This section displays all version of software and annotation data used by VarAFT for the current version. Informations for previous versions are available in the 'Version History' section of the documentation.
Figure 8: Settings Panel to configure VarAFT: Versionning
4 Annotation Tool

The first step to analyze a variants file is to annotate all variations. In order to accomplish this, VarAFT implements ANNOVAR, a command line tool. In the Annotation Tool Menu you have 3 options:

- **1 Variant Files Annotation**: Module for annotation of a variants file as VCF file.
- **2 UMD-Predictor**: Module to retrieve prediction scores from UMD-Predictor. These score are automatically retrieved during annotation process. However you could need to get it separately.
- **3 Manage Annovar Database**: With this module you can download the needed databases for the annotation process.
4.1 Variant Files Annotation

To annotate variants you must first select a specific project (Figure 10: 1). If the project has not been created yet, a new project will be created by clicking New Project (Figure 10: 2). A new directory is created in the VarAFT_Project folder, defined in the settings.

The second step is the selection of the file format (Figure 10: 3). There are 3 formats available: VCF, GVCF, ANN, VCF (CNV).

- **VCF** (Variant Call File) is the standard format used to list all variants into a file. Depending on the variant caller, the vcf may be not recognized or incorrectly annotated. If this occurs, please contact us with an example file. We will do our best to solve this issue. However most VCF files should be properly defined and annotated with VarAFT. Since version 2.13 VarAFT accepts gzip file.
Figure 11: Variant Call Format example

<table>
<thead>
<tr>
<th>CHROM</th>
<th>POS</th>
<th>REF</th>
<th>ALT</th>
<th>QUAL</th>
<th>FILTER</th>
<th>INFO</th>
<th>FORMAT</th>
<th>QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>17330</td>
<td>.</td>
<td>T</td>
<td>A</td>
<td>FAIL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• **gVCF** (Genomic Variant Calling Format). A gVCF is a kind of VCF, so the basic format specifications are the same as for a regular VCF, but a genomic VCF contains extra information such as sequenced regions that are identical to the reference genome. Since version 2.13 VarAFT accepts gzipped VCF file.

• **ANN** is a basic format. If you don’t have a VCF file you can create your file in an ANN format. This file is a tab delimited file with 9 columns: Chromosome, Position start, Position end, Reference Allele, Alternative Allele, Genotype (SNV/INDEL = het or hom; CNV = DEL-HTZ; DEL-HOM or DUP), Depth (Size for CNV), Frequency of alternative allele (Copy Number for CNV) and score for alternative allele.

You can see an example below:

<table>
<thead>
<tr>
<th>chr1</th>
<th>955597</th>
<th>955597</th>
<th>G</th>
<th>T</th>
<th>het</th>
<th>31</th>
<th>0.55</th>
<th>64.48</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>977031</td>
<td>977031</td>
<td>G</td>
<td>A</td>
<td>het</td>
<td>43</td>
<td>0.42</td>
<td>47.28</td>
</tr>
<tr>
<td>chr1</td>
<td>977330</td>
<td>977030</td>
<td>T</td>
<td>C</td>
<td>het</td>
<td>161</td>
<td>0.52</td>
<td>261.78</td>
</tr>
<tr>
<td>chr1</td>
<td>980824</td>
<td>980824</td>
<td>G</td>
<td>C</td>
<td>het</td>
<td>83</td>
<td>0.39</td>
<td>68.26</td>
</tr>
<tr>
<td>chr1</td>
<td>10408838</td>
<td>10408838</td>
<td>T</td>
<td>D</td>
<td>hom</td>
<td>96</td>
<td>0.98</td>
<td>459.98</td>
</tr>
<tr>
<td>chr2</td>
<td>32361482</td>
<td>32361482</td>
<td>-</td>
<td>C</td>
<td>het</td>
<td>9</td>
<td>0.71</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 1: ANN file format with SNV/INDEL data

<table>
<thead>
<tr>
<th>chr1</th>
<th>955597</th>
<th>955597</th>
<th>0</th>
<th>-</th>
<th>DEL-HTZ</th>
<th>90007</th>
<th>1</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>977031</td>
<td>977031</td>
<td>0</td>
<td>0</td>
<td>DUP</td>
<td>48521</td>
<td>6</td>
<td>150</td>
</tr>
<tr>
<td>chr1</td>
<td>977330</td>
<td>977030</td>
<td>0</td>
<td>-</td>
<td>DEL-HTZ</td>
<td>1</td>
<td>0.52</td>
<td>210</td>
</tr>
<tr>
<td>chr1</td>
<td>980824</td>
<td>980824</td>
<td>0</td>
<td>-</td>
<td>DEL-HOM</td>
<td>0</td>
<td>0.39</td>
<td>170</td>
</tr>
<tr>
<td>chr1</td>
<td>10408838</td>
<td>10408838</td>
<td>0</td>
<td>0</td>
<td>DUP</td>
<td>114029</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>chr2</td>
<td>32361482</td>
<td>32361482</td>
<td>0</td>
<td>-</td>
<td>DEL-HTZ</td>
<td>13885</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2: ANN file format with CNV data

In a third step, it is required to select the genome version (Figure 10). Currently, two versions are available on VarAFT: HG19 and HG38. For both versions, databases should be downloaded before use.

The fourth step consists in the selection of dbsnp version (Figure 10). By default, only the version 147 is downloaded. If you need to use another version of dbsnp go to the Download DBSNP module (4.4.1).

In the next step you can choose which variants to annotate. The Only PASS filter allows the annotation of only variants which have passed the variant calling filter criteria (filter flag = PASS). If you uncheck this option all variants will be annotated.
WARNING

No all VCF files have the flag filter description (PASS or other). Please check your VCF before. Annotation of a VCF with no PASS flag but with this option selected will result into an empty file and no variants will be annotated.

The Multi Samples option (Figure 10 6) allows to specify if your VCF file contains several samples into one file. If you annotate a multi sample VCF without check-in this box, only the first sample will be considered.

Once all options are set, select one or several files that you want to annotate with the Browse button (Figure 10 9). You could select several file. All selected files will be displayed in the list box (Figure 10 8). To remove one file click on Delete (Figure 10 10). To remove all files from the list click on Clear (Figure 10 11).

To launch the annotation click on the submit button. A progress bars shows you the progression of the annotation process. Depending on the file size and your hardware computer, the processing time will differ.

NOTE

With 4 threads and a VCF with 100000 variants, the process should take between 5 to 15 minutes depending of the optional database selected.

WARNING

For whole genome or whole exome with more intronic region, the HSF annotation could take a long time. HSF needs to compute all variants that could affect splicing some of those may take some time to be processed. Note If you reannotate the same VCF file the HSF annotation will be quicker.

4.2 UMD-Predictor

UMD-Predictor allows you to retrieve pathogenicity prediction scores from UMD-Predictor. (see 3.4.1).
To annotate specifically an annotated file with UMD-Predictor pathogenicity prediction scores, you have first to select a project (Figure 12: 1) with previously annotated files. To select files click on **Browse** (Figure 12: 3). To remove one file click on **Delete** (Figure 12: 4). To remove all files from the list click on **Clear** (Figure 12: 5).

To launch the annotation process click on the submit button. A new window displays the progression of the process.

**WARNING**

Depending on your network configuration, the UMD-Predictor’s webservice may not be accessible. Contact your administrator network to allow connection to umd-predictor.eu.

### 4.3 Human Splicing Finder

*Human Splicing Finder (HSF)* allows you to retrieve prediction from the API of HSF. (see 3.4.2)
To retrieve HSF predictions from a previously annotated file, you have first to select a project (Figure 13: 1). To select files click on **Browse** (Figure 12: 3). To remove one file click on **Delete** (Figure 13: 4). To remove all files from the list click on **Clear** (Figure 12: 5).

To launch the annotation process, click on the submit button. A new window displays the progression of the process.

**NOTE**

HSF predictions were retrieved during the annotation process, this module allows to reannotate a previously annotated file if an error occurs during the HSF process. This module also allows to annotate file coming from VarAFT 2.10 when HSF API annotations were not available yet.
4.4 Manage Database

To perform variant annotation, you have to download various files that will serve as database. If you want to download data for the Human Genome version 19 click on Download data for hg19. If you want to download data for the Human Genome version 38 click on Download data for hg38. These two modules automatically launch the download process et unzip downloaded files in the database directory specified in Settings (3.1). The download process takes long time (1 hours to several hours according to your network) and requires various between 30 and 50 GB free space on your storage support. If you computer/server has several hard drive, it is advice to store the database on a different disk than the VarAFT_Project folder.

Figure 14: Manage Database Module
4.4.1 Download dbSNP

The main goal of this Download dbSNP module is to automatically download a specific version of dbSNP (if the dbSNP version downloaded with the Download data for hg19 or 38 does not respond to your needs). Indeed only the dbsnp version 150 was downloaded from the two previous modules.

Figure 15: dbSNP download module

To download an other version of dbSNP, select the genome version (Figure 15 1) and the dbSNP version (Figure 15 2). Some versions exists in a flagged version (that means: no clinical flagged variant and no variant with MAF lower than 1% ). To select the flagged version check the box (Figure 15 3). The following table shows you which version are available:
<table>
<thead>
<tr>
<th>Genome Version</th>
<th>dbSNP version</th>
<th>Flagged version available</th>
</tr>
</thead>
<tbody>
<tr>
<td>HG19</td>
<td>130</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>131</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>132</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>135</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>137</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>138</td>
<td>Yes</td>
</tr>
<tr>
<td>HG19</td>
<td>144</td>
<td>No</td>
</tr>
<tr>
<td>HG19</td>
<td>147</td>
<td>No</td>
</tr>
<tr>
<td>HG19</td>
<td>150</td>
<td>No</td>
</tr>
<tr>
<td>HG38</td>
<td>144</td>
<td>No</td>
</tr>
<tr>
<td>HG38</td>
<td>147</td>
<td>No</td>
</tr>
<tr>
<td>HG38</td>
<td>150</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3: Available DBSNP version
4.4.2 Create DBLocal

![Create your Local Database](image)

Sometimes you might need to filter out your variants with specific personal data such as specific ethnic variants or known artifacts.

The aim of this module is to easily create a such file with all your variants and use it as a filter option in your variant filtration analysis.

To build this file, first provide a name (Figure 16:1). Next select the format of the input files (VCF or ANN) (Figure 16:2). Choose to use this file by default (Figure 16:3). In the next step you can decide if you want to keep variants that were able to pass quality filters or not. To do so check or uncheck the **Only PASS filter**. If you uncheck this option all variants will be considered (Figure 16:4).

To select files click on **Browse** (Figure 16:5). To remove one file click on **Delete** (Figure 16:6). To remove all files from the list click on **Clear** (Figure 16:7). If a VCF file contains several samples, each samples will be considered. Since VarAFT 2.13 gzip VCF are accepted.

Figure 16: Module for creation of local database
The output file looks like as follow:

<table>
<thead>
<tr>
<th>mutation_id</th>
<th>chromosome</th>
<th>position</th>
<th>reference</th>
<th>observed</th>
<th>nb_ind_nb</th>
<th>nb_ind_hom</th>
<th>nb_ind_all</th>
<th>nd_ind_total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>chr1</td>
<td>879676</td>
<td>G</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>290</td>
</tr>
<tr>
<td>2</td>
<td>chr1</td>
<td>879687</td>
<td>GC</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>290</td>
</tr>
<tr>
<td>4</td>
<td>chr1</td>
<td>881627</td>
<td>T</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>290</td>
</tr>
<tr>
<td>7</td>
<td>chr1</td>
<td>888639</td>
<td>TTC</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>290</td>
</tr>
</tbody>
</table>

Table 4: Example local database
5 Analysis and filter tool

![Analysis and filter tool module: Method selection](image)

Figure 17: Analysis and filter tool module: Method selection
5.1 Manual Filtering

Figure 18: Analysis and filter tool module: Project selection

This module allows you to combine and filter your annotated variant files. Several features are available. To start an analysis, select a project and click next (Figure 18).
Figure 19: Main panel of VarAFT for analysis

The main panel of VarAFT is composed of 4 parts (Figure 19):

- **Area 1**: Selection of combination mode. It is possible to combine data from multiple samples/patients (singleton, trio or any combination)
  - Autosomal Recessive Disease
    * Index Case only
    * Trio Analysis both Parents/Proband
    * Multi Analysis (vcf merged)
  - Autosomal Dominant Disease
    * Index only
    * Trio Analysis with both Parents/Proband (for de novo analysis)
  - Custom analysis
    * Cohort analysis at:
      - the variation level (variants shared by several individuals).
      - the gene level (Genes with variation shared by several individuals.)
* Somatic variation identification by comparing Normal vs Tumoral sample.
* any other type of combination between individuals

- **Area 2:** Results table. This area displays the list of selected variations with main annotations. A click on a row will update area #4. A right click on a row gives access to a contextual menu which allows to:
  - access IGV and visualize read alignment data
  - show tissues expressions
  - get data from HSF - Human Splicing Finder
  - access several links (NCBI, UCSC, OMIM, GeneCards ...)

- **Area 3:** Filtration Area. This area allows you to define and apply filtration criteria such as:
  - Variation type and localization
  - Frequency
  - Pathogenicity predictions
  - Genes information
  - Others (Clinvar, Cosmic ...)

- **Area 4:** Information Area. This area provides detailed information on the selected variant:
  - Nomenclature in all transcripts
  - Variants Information for each samples
  - Frequency in the general population
  - Pathogenicity Predictions
  - Localization in particular regions
  - Pathways (KEGG, REACTOME, PID)
  - Gene Ontology
  - OMIM
  - Tissues Expression from GTEX
  - External links

**5.1.1 Mode of inheritance**

**5.1.1.1 Autosomal Recessive Disease** Six kinds of analysis are available: Index Case Analysis, Trio Analysis and Multi Analysis. Output results depend on different possible cases (Figure 20).
**Index Case Analysis**  To perform the analysis of an index case click on .

A new panel appears (Figure 21). Click on the browse button (Figure 21: 2) to select the annotated file corresponding to the index case. To keep uncertain heterozygous variants (more information in section 3.3) let the box checked (Figure 21: 3) if not uncheck it.

To select a gene list check the box (Figure 21: 4). To add a new gene list show section 7.1.

You can use a BED file to filter your data, for that check the box (Figure 21: 5).

To launch the analysis click on the submit button. A progress bar shows you the progression of the process.

Once the analysis is completed, 3 tabs will appear: one for homozygous hypothesis, one for compound heterozygous hypothesis, and one with heterozygous one (only gene with one variant).

**Trio Analysis**  To perform a trio family analysis, click on .

A new window appears (Figure 22). Click on the browse buttons to select the following files: the annotated file corresponding to the proband (Figure 22: 1), the annotated file corresponding to
the Mother (Figure 22: 2) and the annotated file corresponding to the Father (Figure 22: 3).
To keep uncertain heterozygous variants (more information in section 3.3) let the box checked
(Figure 22: 4) if not uncheck it.
To select of a gene list check the box ((Figure 22: 5). To add a new gene list show section 7.1.
You can use a BED file to filter your data, for that check the box ((Figure 22: 6).

Once all files have been selected, click on the submit button to continue.

Once the analysis completed, 6 results tabs will appear:

- Homozygous variations (the same variation : 1 heterozygous from father, 1 heterozygous
  from mother )
- Compound heterozygous (2 variations : 1 heterozygous from father, 1 heterozygous from
  mother )
- Homozygous with only Father Heterozygous for the same variation
- Homozygous with only Mother Heterozygous for the same variation
- Heterozygous with Father Heterozygous for the same variation
Figure 22: Autosomal Recessive Disease : Trio Analysis

- Heterozygous with Mother Heterozygous for the same variation

**Family Analysis**  To perform the analysis of a family based on a multi annotated VCF file click on .
A new window appears (Figure 23). Click on the browse button to select the file (Figure 23: 1).

To keep uncertain heterozygous variants (more information in section 3.3) let the box checked (Figure 23: 2) if not uncheck it.

To select of a gene list check the box (Figure 23: 5). To add a new gene list show section 7.1. You can use a BED file to filter your data, for that check the box (Figure 23: 6).

Once all files have been selected, click on the submit button to continue.
In the next window (Figure 24), set the status for each sample. Each row correspond to one sample. For one row click on the **Sample** column to choose the status of the sample. You have the choice between 6 options:

- **Affected**: set if the sample is affected.
- **Mother**: set if the sample is the mother
- **Father**: set if the sample is the father
- **Healthy**: set if the sample is healthy
- **Healthy Carrier**: set if the sample is healthy carrier.
- **No**: set if you don’t want to use this sample in your analysis.

You have to define at least a Mother and a Father, these are mandatory.

Once the analysis is completed, 6 results tabs will appear:

- Homozygous variations (the same variation: 1 heterozygous from father, 1 heterozygous from mother)
- Compound heterozygous (2 variations: 1 heterozygous from father, 1 heterozygous from mother)
- Homozygous with only Father Heterozygous for the same variation
- Homozygous with only Mother Heterozygous for the same variation

---

<table>
<thead>
<tr>
<th>Sample List</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Name</td>
<td></td>
</tr>
<tr>
<td>464</td>
<td>Affected</td>
</tr>
<tr>
<td>465</td>
<td>Affected</td>
</tr>
<tr>
<td>466</td>
<td>Affected</td>
</tr>
<tr>
<td>467</td>
<td>Father</td>
</tr>
<tr>
<td></td>
<td>Mother</td>
</tr>
<tr>
<td></td>
<td>Affected</td>
</tr>
<tr>
<td></td>
<td>Healthy Carrier</td>
</tr>
</tbody>
</table>
• Heterozygous with Father Heterozygous for the same variation
• Heterozygous with Mother Heterozygous for the same variation
5.1.1.2 Autosomal de novo Dominant Disease  Two kinds of analysis are available: Index Case Analysis, Trio Analysis (de novo).

Index Case Analysis  To perform the analysis of an index case click on .

![Index analysis configuration panel for Autosomal dominant disease](image)

Figure 25: Index analysis configuration panel for Autosomal dominant disease

A new panel appears (Figure 25). Click on the "browse button" (Figure 25 1) to select the annotated file corresponding to the index case. To keep uncertain heterozygous variants (more information in section 3.3) let the box checked (Figure 25 2) if not uncheck it.

The last option allows you to add a selection of genes (gene list). For that check the box ((Figure 25 3) and select your list. To add a new gene list show section 7.1.

You can use a BED file to filter your data, for that check the box ((Figure 25 4).

Once all files have been selected, click on the submit button to continue.

Once the analysis is completed, 1 tab will appear with all heterozygous variant.
**Trio Analysis**  To perform the analysis of a trio family click on 📚.

![Figure 26: Autosomal Recessive Disease : Trio Analysis](image)

A new window appears (Figure 26). Click on the browse buttons to select the following files: the annotated file corresponding to the proband (Figure 26: 1), the annotated file corresponding to the Mother (Figure 26: 2) and the annotated file corresponding to the Father (Figure 26: 3). To keep uncertain heterozygous variants (more information in section 3.3) let the box checked (Figure 26: 4) if not, uncheck it.

The last option allows you to add a selection of genes (gene list). For that check the box (Figure 26: 5)) and select your list. To add a new gene list show section 7.1.

Once all files have been selected, click on the submit button to continue.

Once the analysis completed, 1 tab will appear with proband specific heterozygous variant (Not present on both parents ).
5.1.1.3 Custom Analysis  To perform a custom analysis click on .

![Custom Analysis Module](image)

Figure 27: Custom Analysis Module

A new window appears (Figure 27). Click on the browse button (Figure 27 1) to select all wanted files. Single or multi files are allowed. Annotated files with CNV can be combined with SNV/INDELs annotated files. If you want to select a genes list, check the box (Figure 27 4) and select your list. To add a new gene list show section 7.1. You can use a BED file to filter your data, for that check the box ((Figure 26 6). To continue click on submit.
You get a new window with all selected samples in a table. You can choose between the two following options:

- **Get variants from each sample with selected genotype** (Figure 28: 1): With this option all the genotypes given in part 5 will be considered.

- **Get variants with minimal number of conditions and selected genotype** (Figure 28: 2): the genotype can be chosen for all files. After choosing this option, whether you want "Variants" or "Genes" (Figure 28: 3) and the threshold (Figure 28: 4) must be chosen.

For example, you have 4 files for which you select the genotype Heterozygous. If you select "Variant" and a threshold of $\frac{3}{4}$ that means you want all the variants to be present in 3 out of the 4 files given with the selected genotype. If you select "Gene" and a threshold of $\frac{3}{4}$ that means you want all genes present in 3 files out of the 4 given whatever the variant.

Next you must set the genotype for each sample. Each row correspond to one sample.

Figure 28: Custom Analysis Module Configuration
For one row, click on the **Genotype** column (Figure 28, 7) to choose the genotype of the sample. You have the choice between 5 options for SNV/INDELs data:

- **ALL**: all variations, whatever the genotype
- **HTZ**: Heterozygous variations
- **HOM**: Heterozygous variations
- **ABS**: the variation must be absent. (this option is not available for "Select minimal number of conditions")
- **ND**: the variation may be present or absent depending on the genotype. (this option is not available for "Select minimal number of conditions")
- **NO**: Not used the corresponding sample. Very useful to analyze some samples from merges VCF.

You have the choice between 4 options for SNV/INDELs data:

- **ALL**: all variations, whatever the genotype
- **DEL-HTZ**: CNV Heterozygous deletion
- **DEL-HOM**: Heterozygous deletion
- **DUP**: CNV duplication

You can set several rows in one time. For that select rows and click on the right button of your mouse and select between:

- **Set All**
- **Set Homozygous**
- **Set Heterozygous**
- **Set Absent** (this option is not available for "Select minimal number of conditions")
- **Set Not Defined** (this option is not available for "Select minimal number of conditions")

Once you have selected your settings you can click on submit. Once the analysis completed, one result tab will appear with variations corresponding to the parameters previously selected.
5.1.2 Filtration area

The filtration area is divided into 5 tabs (Figure 29):

- Variant Type
- Frequency
- Prediction
- Genes Information
- Others

On the right, you can see 6 buttons (Figure 29):

- 1: Show/hide the filter area.
- 2: Apply saved filters for the selected project.
- 3: Save filters for the selected project.
- 4: Apply default filters
- 5: Save filters as default (available for all projects)
- 6: Reset filters
5.1.2.1 Variant Type  From the first tab (Figure 29) you can filter variants based on the type and/or the functional effect. To filter a specific item, uncheck the associated box. By right clicking you can select or unselect all items.

NOTE
Since VarAFT 2.13 you can choose between RefSeq or Ensembl gene models. To use Ensembl source, select associated columns in the setting (cf. 3.5)

Figure 30: Filter based on the population frequency data. Annotation with version < 2.10
Figure 31: Filter based on the population frequency data. Annotation with version $\geq 2.10$ and $< 2.16$. 

### Public databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Type</th>
<th>Value</th>
<th>Filter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gnomAD E</td>
<td>ALL</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>gnomAD G+</td>
<td>ALL</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>1000G</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KaViAR</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRC</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GME</td>
<td>ALL</td>
<td>$\leq$</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

*optional

### Local database

- HOM
- HET
- ALL

Filter based on File of variant
Figure 32: Filter based on the population frequency data. Annotation with version $\geq 2.16$

5.1.2.2 Frequency

**Public database** The second tab allows you to filter variants based on the population frequency. Several public databases are currently available in the new version.

- **1000G**: Data from 1000 genomes project (08-2015).
- **gnomAD**: Data from Genome Aggregation Consortium version 2.11 (since 2.16). 7 subpopulation and 8 subgroups are available. gnomAD Exome is always used. gnomAD Genome is optional and must be selected in the settings to be used.
  - **AMR**: Latino population
  - **AFR**: African population
  - **SAS**: South Asian population
  - **EAS**: East Asian population
  - **NFE**: Non Finnish European population
  - **FIN**: Finnish European population
  - **ASJ**: Ashkenazi Jewish Population
  - **OTH**: Others population.
  - **RAW**: Allele Frequency from Raw data (since 2.16)
- **MALE**: Allele Frequency for Male only (since 2.16)
- **FEMALE**: Allele Frequency for Female only (since 2.16)
- **MAX**: This annotation contains a threshold filter allele frequency for a variant. Technically, this is the highest disease-specific maximum credible population AF for which the observed AC is not compatible with pathogenicity. See gnomAD documentation for more details (since 2.16)
- **NON TOPMED**: Only samples that are not present in the Trans-Omics for Precision Medicine (TOPMed)/BRAVO release. The allele counts in this subset can thus be added to those of BRAVO to federate both datasets. (since 2.16)
- **NON NEURO**: Allele Frequency with Only samples from individuals who were not ascertained for having a neurological condition in a neurological case/control study (since 2.16)
- **NON CANCER**: Only samples from individuals who were not ascertained for having cancer in a cancer study (since 2.16)
- **CONTROLS**: Only samples from individuals who were not selected as a case in a case/control study of common disease (since 2.16)

- **KaViar**: Data from Known VARiants Project. Kaviar contains 162 million SNV sites (including 25M not in dbSNP) and incorporates data from 35 projects encompassing 77,781 individuals (13.2K whole genome, 64.6K exome). This database is optional and must be selected in the settings to be used.

- **HRC**: The Haplotype Reference Consortium. This database is optional and must be selected in the settings to be used.

- **GME**: The Great Middle East database. This database is optional and must be selected in the settings to be used.

- **Iranome**: The Iranome database. This database is optional and must be selected in the settings to be used (since 2.16).

You can also remove all variants with a dbsnp id by clicking on **Remove DBSNP**.

**Warnings**: By default DbSNP database contains data from clinical source and variants with an allele frequency above 1%. Be careful using this option.

**Local database**  Thanks to VarAFT you have the possibility to generate a local database with a selection of VCF files (show 4.4.2). If you have set a default dblocal file you can filter your variants list thanks to this database. The value **Max** will be automatically update with your dblocal. You can select between ALL, HET (Heterozygous) and HOM (Homozygous) and set the minimal threshold to perform the filter.

The button **Filter based on File of variant** offers two options (Figure 33) : **show or filter** variants. Two format file are available: a **text** file as dblocal (show 4) or a **bed** file. The bed file must have the following format :
<table>
<thead>
<tr>
<th>chr1</th>
<th>879676</th>
<th>879677</th>
<th>Variant1</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>879687</td>
<td>879689</td>
<td>Variant2</td>
</tr>
<tr>
<td>chr1</td>
<td>881627</td>
<td>881628</td>
<td>Variant3</td>
</tr>
<tr>
<td>chr1</td>
<td>888639</td>
<td>888642</td>
<td>Variant4</td>
</tr>
</tbody>
</table>

Table 5: Example bedfile

Figure 33: Filter based on a file of variant
5.1.2.3 Prediction  The third tab allows you to select variants based on the prediction scores. Several different tools are available: UMD-Predictor, Human Splicing Finder, SIFT, PolyPhen, Mutation Taster, Mutation Assessor, Provean, LRT, M-CAP, Eigen, DANN, CADD and GERP++. For the first height tools, simply uncheck the boxes of the non-desired prediction. For DANN, Eigen, CADD and GERP++, no classification exists, so a cut-off needs be selected to apply a selection.
5.1.2.4 Gene Informations

The fourth tab allows for filtration thanks to information relative to genes. Several options are available.

**Gene Name**  Select all variants localized in a specific gene by providing a gene name. It is a case sensitive research.

**Score**  Different scores gene specific are available:

- **RVIS**: RVIS score measures genetic intolerance of genes to functional mutations, as described in Petrovski et al. Original RVIS was constructed based on patterns of standing variation in 6503 samples. The authors have recently constructed scores based on the 61,000 samples from ExAC. There is high correlation, but more resolution for many genes. A gene with a positive score has more common functional variation, and a gene with a negative score has less and is referred to as “intolerant”.

- **GDI**: the gene damage index (GDI) represents the accumulated mutational damage for each human gene in the general population, and shows that highly mutated/damaged genes are unlikely to be disease-causing and yet they generate a big proportion of false positive variants harbored in such genes. Therefore removing high GDI genes is a very effective way to remove confidently false positives from WES/WGS data. The data set includes general damage prediction (low/medium/high) for different disease type (all, Mendelian, cancer, and PID).
- **LoFTool**: gene loss-of-function score percentiles. The smaller the percentile, the most intolerant is the gene to functional variation.

- **GHIS**: Genome-wide HaploInsufficiency Score. Higher is the score, the most haploinsufficiency is the gene.

Pathway & GO & HPO & OMIM  Different database are available to filter variants.

- **KEGG**: KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

- **REACTOME**: Reactome is a free, open-source, curated and peer reviewed pathway database. The goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education.

- **PID**: The Pathway Interaction Database (PID, http://pid.nci.nih.gov) is a freely available collection of curated and peer-reviewed pathways composed of human molecular signaling and regulatory events and key cellular processes.

- **GO**: Gene Ontology is the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects: molecular function (molecular activities of gene products), cellular component (where gene products are active) and biological process (pathways and larger processes made up of the activities of multiple gene products).

- **HPO**: The Human Phenotype Ontology (HPO) aims to provide a standardized vocabulary of phenotypic abnormalities encountered in human disease. Each term in the HPO describes a phenotypic abnormality, such as atrial septal defect.

- **OMIM**: OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 15,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other genetics resources.
The version 2.10 introduces a new module to combine data from KEGG, REACTOME, PID, OMIM, HPO, and GO. This module allows retrieving a list of genes from selected annotation.

To do this, first select the database (Figure 36: 6) and click on the desired item(s) in the list (Figure 36: 9). This list is populated with the gene list from the result table (after applied filters). If you want to get genes from this list, click on the ‘→’ button (Figure 36: 7). The selected items are added in the box (Figure 36: 3) and the associated genes are added in the box (Figure 36: 5). If you want to exclude genes from the final list, select an item and click on “→” button (Figure 36: 8). The selected items are added in the box (Figure 36: 4) and the associated genes are removed from the box (Figure 36: 5). You can save the obtained genes list thanks to the button (Figure 36: 1).
The Genotype-Tissue Expression (GTEx) project aims to provide to the scientific community a resource to study human gene expression and regulation and its relationship to genetic variation. This project will collect and analyze multiple human tissues from donors who are also densely genotyped, to assess genetic variation within their genomes. You can use these data to select variants in genes expressed in a particular tissue.

The figure shows the panel corresponding to GTEx. First select one or several tissue. Next select a minimal expression threshold (TPM >2) and last if you have selected more than 1 tissue set a threshold to choose the minimal number of tissue that it must expressed the gene. Click on submit. Only variant in genes expressed in the selected tissue are kept.
5.1.3 Others

- 1. Select by position. You could select variants with only a chromosome or with a specific position.

- 2. Filter variants based on the SNV score. The SNV score is a Phred quality score given by the variant caller. Set a threshold and click on the green button.

- 3. Clinvar: you can either keep or remove variants include in Clinvar. You also have the possibility to select between All, Benign or Pathogenic.

- 4. Cosmic: you can either keep or remove variants include in cosmic (v70).

- 5. Get Compound Heterozygous: In a autosomal recessive disease hypothesis you want to keep only gene with at least 2 variants. Once you performed several filtering steps, several genes are not compliant with this hypothesis. To solve this issue click on the button Get Compound Heterozygous. Click on the circle arrow to reset this filter.

- 6. Get Genes with threshold: In a cohort analysis with the custom mode, you have the possibility to select genes shared by several sample according to a threshold. Once you performed several filtering steps, several genes are not compliant with this hypothesis. To solve this issue click on the button Get Genes with threshold. Click on the circle arrow to reset this filter.
Figure 39: Panel containing variant information 1/3
Figure 40: Panel containing variant information 2/3
This area provides detailed information on the selected variant (Figure 39, 40 & 41):

- **General Information**: Gene name, position (with genome version), reference allele, alternative allele and functional impact are displayed.

- **RefGene Transcript Information**: List of all refseq id with the associated HGVS nomenclature.

- **Ensembl Transcript Information**: List of all ensembl transcript id with the associated HGVS nomenclature.
- Database frequency population: Frequency in several databases including ExAC, Exome variant server, 1000 Genomes, KaViar and HRC.
- Clinvar: List of Clinvar ID with conclusion and link to clinvar website for the variant.
- Cosmic: List of Cosmic ID with link to cosmic website for the variant.
- Prediction Tools: All prediction score for the selected variant.
- Human Splicing Finder: All prediction and interpretation from HSF foreach transcript.
- KEGG: List of KEGG pathway
- Reactome: List of Reactome pathway
- PID: List of PID pathway
- Gene Ontology: list of Gene Ontology
- OMIM: list of OMIM disease with description
- Tissue Expression from GTEx: Histogram of RPKM value for each tissue
- External links: Links to UCSC, GeneCards, NCBI, OMIM, Malacards, GTEx Portal.
### 5.1.5 Result table

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Ref</th>
<th>Alt</th>
<th>Genotype *</th>
<th>Depth *</th>
<th>SNV Score *</th>
<th>Func.refgene</th>
<th>Gene.refgene</th>
<th>ExonicFunction.refgene</th>
<th>AAChange.refgene</th>
<th>Var in File</th>
<th>Gene in File</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>444</td>
<td>445</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>0.63</td>
<td>15.01</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>502</td>
<td>503</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>0.63</td>
<td>15.01</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>573</td>
<td>574</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>0.63</td>
<td>15.01</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>EXONIC</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The result table (Figure 42) contains several fixed columns:

- **Chr**: Chromosome name
- **Start**: Start position of the variant
- **End**: End position of the variant
- **Ref**: Reference Allele
- **Alt**: Alternate Allele
- **Genotype ***: Genotype
- **Depth ***: Number of used reads at the position of the variant
- **Frequency ***: Number of reads containing the alternate allele divided by the depth.
- **SNV Score ***: quality score of the variant given by the variant caller.
- **Func.refgene**: localisation of the variation (exonic, splicing, intronic, UTR, ...)
- **Gene.refgene**: Gene Symbol
- **ExonicFunction.refgene**: impact of the variation (synonymous, non-synonymous, stop, frameshift, non frameshift ...)
- **AAChange.refgene**: detailed variation on the transcript and on the protein (c. and p.)
- **Var in File** (only available with the option "Select minimal number of conditions"): Number of files containing the variation on the total of files.
- **Gene in File** (only available with the option "Select minimal number of conditions" and "Gene"): Number of files containing the gene on the total of files.

*: Only if one sample is analyzing or if you use Trio analysis.
The table also contains annotations that have been selected in Settings (3.5).

A popup menu appears by clicking right button of your mouse on a row. In this popup menu you access to:

- **Remove Line(s)**: Remove the selected row(s).
- **Show on IGV**: Display selected bam files at the variant position on the Interactive Genome Browser.
- **Gene Expression**: show the histogram with expression level of the selected gene for each tissue.
- **Links**: Links to Malacards, GeneCards, OMIM, UCSC, NCBI, GTEx Portal.

### 5.1.6 Other Options

![Figure 43: Other options](image)

At the bottom of the main window, you can see the number of variants (Figure 43 1), the number of genes (Figure 43 2) and the project name (Figure 43 3). You have 5 buttons:

- **4**: Display selected bam files at the variant position on the Interactive Genome Browser.
- **5**: Save results table in VarAFT format. The output file could be reopen in custom mode and / or combine with other result or samples.
- **6**: Save results table in excel format. The file contains a header with information about analysis as the used samples and the used filters.
- **7**: Print results table.
- **8**: Display information about analysis.
5.2 Automatic Filtering

VarAFT introduces a new filter mode in the version 2.10. It is the automatic filtering module. This module allows to pre-filter several annotated files. For example it was very useful to prefilter big annotated files from whole genome. It can reduce the number of variants by 10,000 only with simple filter options. The process is very easy and quick. Just choose your files, select the filter criteria and apply.

INFORMATION

A annotated files of 4,781,755 variants was reduced to 178 variants in 30 seconds with filters based on the variant type (exonic), frequencies (<1%) and pathogenicity prediction.

Several pre-filtered files could combine in the manual filtering mode to explore in details the resulted variants.

Figure 44: Auto filtering Panel
To use this module, first set the filter options (Figure 44.1): As in the manual mode you have access to several options:

- **Variant type and variant impact** (Figure 45): Uncheck the box of items that you want to remove from your files.

- **Public Database** (Figure 46). Select the database that you want to use and set the threshold. "<= 0.01" means you want to keep all variants with a frequency <= 0.01 so removed variants with frequency > 0.01. For dbSNP if you check this box that means you want to remove all variants with a "rs" id.

- **Pathogenicity Prediction**: For tools with predefined class, unchecked the box to remove the associated item. For CADD, Eigen, DANN, GERP++ select a threshold to apply to remove the variants.

![Variant Type Selection](image)

**Figure 45: Auto filtering: Variant type selection**

Once filter options are set, select the project (Figure 44.2) where the output must be saved. The output file's name will be as following : Sample_XXX.autofiltered.ann.hg19_multianno_umd.txt

Next choose to keep or not uncertain heterozygous(Figure 44.3). You can chose a gene list (Figure 44.4 & 5) or a bed file to target specific region (Figure 44.6 & 7).
Next select the annotated files (Figure 44.8) and submit.

A new windows show you the progression.
Figure 47: Auto filtering : Pathogenicity Prediction
6 Coverage Analysis

The last main module of VarAFT allow for computation of coverage from BAM files.

From the menu you have 3 options (Figure 48):

- **Compute Coverage**: Module to compute the coverage from one or several BAM files.
- **Show Coverage**: Module to show output of the coverage analysis with interactive charts.
- **Create BED**: Module to create a bed file (with coding exon) from a gene list.
6.1 Compute Coverage

To perform the coverage analysis, first select a project (Figure 49 1) or create a new one (Figure 49 2). Next select the genome version between hg19 (GRCh37) or hg38 (GRCh38). (Figure 49 3). Select the type of data (Figure 49 4). Next select one or several BAM files thanks to the browse button (Figure 49 5). Before selecting a bam file check if the associated BAI file is present in the same folder. Once BAM files are selected, choice between coding-exon, all-exon or custom mode (Figure 49 8). For coding-exon mode consider only the coding exon of the transcript will be considered for the coverage analysis. For all-exon mode all the exons of the transcript will be considered for the coverage analysis. In custom mode, you must provide a bed file. Here you can compute coverage for all you want. Select if you want to include the flanking region of exon (+/- 50 bp) (Figure 49 9). The last step is the selection of a bed file. In coding-exon mode if you don’t select a bed file all transcripts will be considered. To analyze a panel of genes check the box (Figure 49 10) and select the bed file in the list. In coding-exon mode the bed file must have a specific format. You should generate it with the Create Bed module (6.3). In custom mode the bed file must be selected.

Click on submit button to launch the analysis. Two progress bars show you the advancement of the analysis.
The GRCh3X versions of the genome don’t contain "chr" in the chromosome name whereas HGXX versions contain it. Check the version used and generate the appropriate bed file. If the BAM and BED files are not compliant you will get an error message.
6.2 Show Coverage

To show the coverage analysis results, select **Show Coverage** and select the project.

![Show Coverage Analysis Module](image)

Figure 50: Show Coverage Analysis Module

In the new window select the sample in the list and click on "LOAD" (Figure 50 1). A new tab will be generated containing a table (Figure 50 2). In **coding-exon** mode each line in the table corresponds to a transcript. This table contains the following columns:

- **RefSeqName**: The RefseqMrna ID used to compute the coverage.
- **GeneSymbol**: The associated GeneSymbol.
- **Exon Number**: The total number of coding exon in the transcript.
- **Size Coding Exons**: The size of the coding part of the transcript.
- **Mean Depth**: The mean depth for the transcript.
- **SD Depth**: the standard deviation of the depth for the transcript.
• **Coverage 1x**: Coverage of transcript at 1x of depth. This means the percentage of bases read at 1x for the entire transcript (only coding exon).

• **Coverage 5x**: Coverage of transcript at 5x of depth. This means the percentage of bases read at 5x for the entire transcript (only coding exon).

• **Coverage 10x**: Coverage of transcripts at 10x of depth. This means the percentage of bases read at 10x for the entire transcript (only coding exon).

• **Coverage 20x**: Coverage of the transcript at 20x of depth. This means the percentage of bases read at 20x for the entire transcript (only coding exon).

• **Coverage 30x**: Coverage of the transcript at 30x of depth. This means the percentage of bases read at 30x for the entire transcript (only coding exon).

Several filtering options are available. It is possible to filter based on the refseq name (Figure 50, 6), the gene symbol (Figure 50, 5), and the value (Figure 50, 7). The value can be selected from a list of "Mean Depth", "Coverage 1x", "Coverage 5x", "Coverage 10x", "Coverage 20x", and "Coverage 30x". The limit can be chosen from the following list: >, <, =, >= or <=. Set a cut-off and submit. Only the transcripts corresponding to these parameters will be kept.

If you click on one cell in the table, a chart will appear in the panel (Figure 50, 3). For example if you click on a case in the column "Coverage 20x", you will get the corresponding chart for the selected transcripts. The chart contains a separate bar for each exon. A blue bar indicates an exon with 100% coverage, a yellow bar indicates coverage that is >=90% and < 100%, a red bar indicates an exon with coverage <90%. If you click on one of these bars, you will get a new chart in the right panel. (Figure 50, 4) This new chart will show a detailed chart of the exon. Each bar corresponds to one base. A blue bar indicates a base with depth >= at 20x, a yellow bar indicates a base with a depth >= at 10x and < 20x, a red bar indicates a base with a depth < 10x.

If one of the six first columns is clicked on, the chart in the left panel will correspond to the mean depth for each exon.
The "Show transcript coverage" (Figure 50) button allows you to see the coverage for one transcript in the same chart (Figure 51). Select first the transcript in the table and click on the button.

Figure 51: Coverage for all coding exons for the selected transcript
The "Show gene coverage" (Figure 50) button allows you to see the coverage for one gene with all transcripts in the same chart (Figure 52). Select first the gene in the table and click on the button.

Figure 52: Coverage for the selected gene
The "Show Summary" button shows you a bar plot with the number of transcript correlate to the coverage and the depth.

![Transcripts Coverage Distribution](image)

Figure 53: Bar plot with number of transcript correlate with the depth and coverage
The "XLS Report" (Figure 50) button generates an excel file with a summary of the table. It is advised to filter the table or to use a list of gene before using this feature.

Figure 54: Set threshold to generate xls summary report.

The excel report will contain 3 parts. For example if you set a thresholds as Figure 54, the first part contains all genes with coverage of 100% at 20x. The second part contains the genes with a coverage at 20x >= 90%. The third part contains the genes with a coverage at 20x < 90%. For the second and the third part, there is a column with all exons with a coverage 100% at 20x, a column with all exons with a coverage at 20x >= 90%, a column with all exons with a coverage at 20x < 90%, and a column with all exons not covered for coverage at 20x.

All charts generated can be copied, printed or saved in .png, .svg or .pdf format. To do so, just right click on the chart and choose the corresponding option.
6.3 Create BED

To compute coverage for a gene panel, the bed file must have a specific format. To make this process easy we create the Create Bed module. To generate the bed file select the used version genome during the mapping. Select the gene list and give a name for the output file. The name is automatically generate with the gene list name and the selected genome version. The output file is saved in the bedfiles folder localized in your VarAFT_Project folder. The list can contain only official gene symbol or the official gene symbol with a mRNA refseq id in a second column, if you want to restrict the coverage analysis only to specific transcript.

<table>
<thead>
<tr>
<th>Gene</th>
<th>mRNA refseq id</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>NM_007300</td>
</tr>
<tr>
<td>BRCA2</td>
<td>NM_000059</td>
</tr>
</tbody>
</table>

Table 6: Example gene file with official gene symbol with mRNA refseq id

7 Other

7.1 Gene list

If you want to add a gene list you can either add this from analysis module or from the top menu of the main windows "Edit > Add a gene list". In all cases you will get a new window as figure 55.

Figure 55: Module to add gene list

To add a list click on browse button and select your file. Next click on submit. VarAFT check the given list. If no official gene symbol is found, VarAFT automatically replace it by the good symbol. If a gene symbol is not recognized an error message alert you.

The file will be added in the GenesLists folder of your VarAFT_Project folder.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>mRNA Reference ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>NM_007300</td>
</tr>
<tr>
<td>BRCA2</td>
<td>NM_000059</td>
</tr>
<tr>
<td>DMD</td>
<td>NM_004006</td>
</tr>
<tr>
<td>EMD</td>
<td>NM_000117</td>
</tr>
<tr>
<td>FHL1</td>
<td>NM_001159702</td>
</tr>
<tr>
<td>LMNA</td>
<td>NM_170707</td>
</tr>
</tbody>
</table>

Table 7: Example gene file with official gene symbol with mRNA refseq id. The second column are not mandatory.
8 Version History

8.1 2.16 : April 30, 2019

Changes from previous version

- Updated annotation (clinvar, cosmic, refgene, OMIM, HPO ...)
- Add annotation from gnomad 2.11 and Iranome DB
- Improve results table content (show genotype information only if one sample is used, code color for pathogenic prediction tool, font size customizable)
- Improve IGV connection
- Add Proxy configuration (beta)
- Optimize auto-filtering module
- Solved minor issues

Dependency and Database version used

Software

- ANNOVAR : 2018-04-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72
- HSP API : v1

Variant Annotation

- DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 150 (default)
- (Updated) Clinvar 2019-04-17
- (Updated) Cosmic v88
- (Updated) GnomAD v2.11
- (New) Iranome DB 2019
- Great Middle East Database 2016-10-24
Gene Annotation

- (Updated) OMIM : 2019-04
- (Updated) HPO : 2019-04
- GTEx : 2018-12
- Gene Ontology : GSEA MSigDN 6.2
- KEGG : GSEA MSigDN 6.2
- Reactome : GSEA MSigDN 6.2
- PID : GSEA MSigDN 6.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

8.2 2.15.1 : March 19, 2019
Changes from previous version
- Solved issue with multi samples file analysis

Dependency and Database version used
see 2.12

8.3 2.15.1 : March 08, 2019
Changes from previous version
- Solved UMD-Predictor issue with latest Ensembl update
- Improve gene checking during upload
- Solved various minor issues
Dependency and Database version used

see 2.12

8.4 2.15 : January 4th, 2018

Changes from previous version

- Updated annotation (clinvar, cosmic, refgene, ensembl, GO, KEGG, GTEX, OMIM, HPO ...)
- The next update of annotation could done without VarAFT update
- Coverage files is now compressed + new index
- Coverage analysis available for all exon and/or the possibility to use a flanking region of 50bp
- Solve minors issues

Dependency and Database version used

Software

- *(Updated)* ANNOVAR : 2018-04-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72
- HSP API : v1

Variant Annotation

- DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 150 (default)
- *(Updated)* Clinvar 2018-12-25
- *(Updated)* Cosmic v87
- GnomAD v2.0
- Great Middle East Database 2016-10-24
- HRCRI 2015-12-03
- KAVIAR 2015-09-23
- *(Updated)* Refgene 2018-11-25
- *(Updated)* Ensgene 2018-05-06
Gene Annotation

- (Updated) OMIM : 2018-12
- (Updated) HPO : 2018-12
- (Updated) GTEx : 2018-12
- (Updated) Gene Ontology : GSEA MSigDN 6.2
- (Updated) KEGG : GSEA MSigDN 6.2
- (Updated) Reactome : GSEA MSigDN 6.2
- PID : GSEA MSigDN 6.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

8.5 2.14 : September 17, 2018

Changes from previous version

- Solved issue with download
- Solved issue with dblocal creation
- Solved issue with auto filtering module
- Minor issues solved

Dependency and Database version used

Software

- ANNOVAR : 2017-07-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72
- HSP API : v1
Variant Annotation
- DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 150 (default)
- Clinvar 2017-09-05
- Cosmic v82
- GnomAD v2.0
- Great Middle East Database 2016-10-24
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2017-06-01
- Ensgene 2017-06-01

Gene Annotation
- OMIM : 2018-01
- HPO : 2018-01
- GTEx : 2018-02
- Gene Ontology : GSEA MSigDN 6.1
- KEGG : GSEA MSigDN 6.1
- Reactome : GSEA MSigDN 6.1
- PID : GSEA MSigDN 6.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)
8.6 2.13 : April 20, 2018

Changes from previous version

- Updated Documentation
- Filtration based on Ensembl for the impact of variants
- Implementation of Annovar 'Separate' option
- Coverage analysis for a RNA-seq BAM
- Speed improvement for file unzipping after download
- Compliant with vcf.gz files
- Minor issues solved

Dependency and Database version used

Software

- ANNOVAR : 2017-07-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72
- HSP API : v1

Variant Annotation

- DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 150 (default)
- Clinvar 2017-09-05
- Cosmic v82
- GnomAD v2.0
- Great Middle East Database 2016-10-24
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2017-06-01
- ENSgene 2017-06-01
Gene Annotation

- OMIM : 2018-01
- HPO : 2018-01
- GTEx : 2018-02
- Gene Ontology : GSEA MSigDN 6.1
- KEGG : GSEA MSigDN 6.1
- Reactome : GSEA MSigDN 6.1
- PID : GSEA MSigDN 6.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

8.7 2.12.1 : March 07, 2018

Changes from previous version

- Solved issue for CNV Analysis
- Solved issue for dblocal filtering
- Solved issue for analysis of annotated files from version 2.06

Dependency and Database version used

see 2.12

8.8 2.12 : February 05, 2018

Changes from previous version

- Automatic Annotation with Human Splicing Finder (Need Registration to the API)
- Filtration with Human Splicing Finder
- Updated data : ClinVar, Cosmic82, OMIM, HPO, KEGG, REACTOME, PID, Gene Ontology
- New Frequency Filter Button (Graphical improvement)
- VCF from Varscan 2
- Solved several minor issues...
Dependency and Database version used

Software

- ANNOVAR : 2017-07-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72
- (New) HSP API : v1

Variant Annotation

- DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 150 (default)
- (Updated) Clinvar 2017-09-05
- (Updated) Cosmic v82
- GnomAD v2.0
- Great Middle East Database 2016-10-24
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2017-06-01
- Ensgene 2017-06-01

Gene Annotation

- (Updated) OMIM : 2018-01
- (Updated) HPO : 2018-01
- (Updated) GTEx : 2018-02
- (Updated) Gene Ontology : GSEA MSigDN 6.1
- (Updated) KEGG : GSEA MSigDN 6.1
- (Updated) Reactome : GSEA MSigDN 6.1
- (Updated) PID : GSEA MSigDN 6.1
• Gene Damaging index : 2015-10-28
• GHIS : from paper
• LofTools : from ExAC 0.3
• RVIS : v2 (ExAC)

8.9 2.11.1 : October 05, 2017
Changes from previous version
• Solved issue with multisample VCF Annotation

8.10 2.11 : October 03, 2017
Changes from previous version
• Add dbsnp 150 (from dbsnp download module)
• Solved issue with CVN annotations
• Solved issue with UMD-Predictor Web services

Dependency and Database version used
Software
• ANNOVAR : 2017-07-16
• UMD-Predictor : v2
• Bedtools : v2.25.0
• Bam Util : 1.0.14
• IGV : 2.3.72

Variant Annotation
• DbNSFP 3.3a
• 1000 Genomes 2015-08
• (Updated) DbSNP 150 (default)
• Clinvar 2017-05-01
• Cosmic v70
• GnomAD v2.0
• Great Middle East Database 2016-10-24
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2017-06-01
- Ensgene 2017-06-01

**Gene Annotation**

- OMIM : 2017-08
- HPO : 2017-08
- GTEx : 2016-06
- Gene Ontology : GSEA MSigDN 6.0
- KEGG : GSEA MSigDN 6.0
- Reactome : GSEA MSigDN 6.0
- PID : GSEA MSigDN 6.0
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

**8.11 2.10.2 : September 06, 2017**

Changes from previous version
- Solved issue with 1000G filter option

**8.12 2.10.1 : August 25 , 2017**

Changes from previous version
- Documentation Update
- Solved issue with display options
8.13 2.10 : August 2017

Changes from previous version

- Annovar Update
- New data from dbNSFP3.3, gnomAD, GME DB
- Update data (OMIM, Gene Ontology, KEGG..)
- New module with HPO
- New Pathway/HPO filter module
- Filter based on a BED file
- New Download Module (data stored on our server)
- New Auto Filtering Module
- Annotation and filtration of CNV data
- Compatibility with previous annotated files (2.06)

Dependency and Database version used

Software

- *(Updated)* ANNOVAR : 2017-07-16
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72

Variant Annotation

- *(Updated)* DbNSFP 3.3a
- 1000 Genomes 2015-08
- DbSNP 147 (default)
- *(Updated)* Clinvar 2017-05-01
- Cosmic v70
- *(Removed)* ExAC v0.3
- *(New)* GnomAD v2.0
- *(New)* Great Middle East Database 2016-10-24
• (Removed) Exome Variant Server 2014-12-22
• HRCR1 2015-12-03
• KAVIAR 2015-09-23
• (Updated) Refgene 2017-06-01
• (Updated) Ensgene 2017-06-01

Gene Annotation
• (Updated) OMIM : 2017-08
• (New HPO : 2017-08
• GTEx : 2016-06
• (Updated) Gene Ontology : GSEA MSigDN 6.0
• (Updated) KEGG : GSEA MSigDN 6.0
• (Updated) Reactome : GSEA MSigDN 6.0
• (Updated) PID : GSEA MSigDN 6.0
• Gene Damaging index : 2015-10-28
• GHIS : from paper
• LofTools : from ExAC 0.3
• RVIS : v2 (ExAC)

8.14  2.06 : February 2017

Changes from previous version
Corrects a major issue with varaft server connection, specially for the first utilisation

Dependency and Database version used

Software
• ANNOVAR : 2016-06-06
• UMD-Predictor : v2
• Bedtools : v2.25.0
• Bam Util : 1.0.14
• IGV : 2.3.72
Variant Annotation
- DbNSFP 3.1a
- 1000 Genomes 2015-08
- DbSNP 147 (default)
- Clinvar 2015-12-01
- Cosmic v70
- ExAC v0.3
- Exome Variant Server 2014-12-22
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2015-12-11
- Ensgene 2015-12-11

Gene Annotation
- OMIM : 2016-11
- GTEx : 2016-06
- Gene Ontology : GSEA MSigDN 5.1
- KEGG : GSEA MSigDN 5.1
- Reactome : GSEA MSigDN 5.1
- PID : GSEA MSigDN 5.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

8.15 2.05 : November 2016

Changes from previous version
- Update Documentation (pdf)
- Bugs Correction for export in varafat format
- OMIM update
- Use HSF webservice (beta version) on click
- Solved issues with the conf file and settings
Dependency and Database version used

Software
- ANNOVAR : 2016-06-06
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72

Variant Annotation
- DbNSFP 3.1a
- 1000 Genomes 2015-08
- DbSNP 147 (default)
- Clinvar 2015-12-01
- Cosmic v70
- ExAC v0.3
- Exome Variant Server 2014-12-22
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2015-12-11
- Ensgene 2015-12-11

Gene Annotation
- (Updated) OMIM : 2016-11
- GTEx : 2016-06
- Gene Ontology : GSEA MSigDN 5.1
- KEGG : GSEA MSigDN 5.1
- Reactome : GSEA MSigDN 5.1
- PID : GSEA MSigDN 5.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)
8.16  2.04 : August 2016

Changes from previous version

- Performance improvement for Coverage Analysis in custom mode
- Bug correction for the associated action of button "Get Compound Heterozygous" in the Filter Module

Dependency and Database version used

Software

- ANNOVAR : 2016-06-06
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72

Variant Annotation

- DbNSFP 3.1a
- 1000 Genomes 2015-08
- DbSNP 147 (default)
- Clinvar 2015-12-01
- Cosmic v70
- ExAC v0.3
- Exome Variant Server 2014-12-22
- HRCR1 2015-12-03
- KAVIAR 2015-09-23
- Refgene 2015-12-11
- Ensgene 2015-12-11
Gene Annotation

- OMIM : 2016-06
- GTEx : 2016-06
- Gene Ontology : GSEA MSigDN 5.1
- KEGG : GSEA MSigDN 5.1
- Reactome : GSEA MSigDN 5.1
- PID : GSEA MSigDN 5.1
- Gene Damaging index : 2015-10-28
- GHIS : from paper
- LofTools : from ExAC 0.3
- RVIS : v2 (ExAC)

8.17 2.01 to 2.03

Changes from previous version

Some minor fixes.

Dependency and Database version used

Software

- ANNOVAR : 2016-06-06
- UMD-Predictor : v2
- Bedtools : v2.25.0
- Bam Util : 1.0.14
- IGV : 2.3.72

Variant Annotation

- DbNSFP 3.1a
- 1000 Genomes 2015-08
- DbSNP 147 (default)
- Clinvar 2015-12-01
- Cosmic v70
- ExAC v0.3
• Exome Variant Server 2014-12-22
• HRCR1 2015-12-03
• KAVIAR 2015-09-23
• Refgene 2015-12-11
• Ensgene 2015-12-11

Gene Annotation
• OMIM : 2016-06
• GTEx : 2016-06
• Gene Ontology : GSEA MSigDN 5.1
• KEGG : GSEA MSigDN 5.1
• Reactome : GSEA MSigDN 5.1
• PID : GSEA MSigDN 5.1
• Gene Damaging index : 2015-10-28
• GHIS : from paper
• LofTools : from ExAC 0.3
• RVIS : v2 (ExAC)

8.18 2.00 : June 2016

Changes from previous version
• HG38 genome version available
• New database available : KaViar, HRCR1
• OMIM data available
• New module from local database creation
• Optimized coverage module
• Multithreaded
• VCF multi samples accepted
• New Data from GTEx for tissue expression and new module to filter gene based on expression
• Different scores for genes as GDI, RVIS, GHIS and LofTool
• Package to install VarAFT on MAC, Debian and Redhat system.
Dependency and Database version used

Software
- (Updated) ANNOVAR : 2016-06-06
- UMD-Predictor : v2
- (Updated) Bedtools : v2.25.0
- (New) Bam Util : 1.0.14
- (Updated) IGV : 2.3.72

Variant Annotation
- (Updated) DbNSFP 3.1a
- 1000 Genomes 2015-08
- (Updated) DbSNP 147 (default)
- Clinvar 2015-12-01
- Cosmic v70
- ExAC v0.3
- Exome Variant Server 2014-12-22
- (New) HRCR1 2015-12-03
- (New) KAVIAR 2015-09-23
- Refgene 2015-12-11
- Ensgene 2015-12-11

Gene Annotation
- (New) OMIM : 2016-06
- (New) GTEx : 2016-06
- (Updated) Gene Ontology : GSEA MSigDN 5.1
- (Updated) KEGG : GSEA MSigDN 5.1
- (Updated) Reactome : GSEA MSigDN 5.1
- (Updated) PID : GSEA MSigDN 5.1
- (New) Gene Damaging index : 2015-10-28
- (New) GHIS : from paper
- (New) LofTools : from ExAC 0.3
- (New) RVIS : v2 (ExAC)
9 Frequently asked questions

When I click on the IGV button on VarAFT, IGV is initiated but the bam file is not loaded. Why?

Different reasons can explain that: - Depending on your system, the time required for IGV to load differs. So if you see that the BAM file never load, launch first IGV prior to the analysis. - If you try to upload a BAM file present on a distant server you need to upload it into IGV directly on IGV. - Check if the bai file corresponding to your bam file is present in the same directory.

I’m working on Windows and when I click on VarAFT.exe, the program doesn’t start.

By default, VarAFT uses 4GB of RAM. If your system has less than 4GB of RAM and/or is a 32bits system you need to launch VarAFT with the VarAFT_min.exe file present in the VarAFT directory. However you should know that some options will be disabled such as multi samples (>6) analysis in CUSTOM. You should to set the number of thread at 1 in the settings.

I’m working on Windows and I want to use more RAM.

By default, VarAFT uses 4GB of RAM. However in the VarAFT folder you have precompiled "exe" that can used 8, 16 or 32 Gb.

Analyze and Filtering Tool: progress bar blocks at 5%.

Maybe you have a conflict in your varaft.conf file. To solve this issue, remove or rename the varaft.conf file situated in your user folder in the VarAFT_conf folder. When you will restart VarAFT, this file will be regenerated by default. So you should set again the parameters if necessary.

File empty after annotation

Check if your VCF file contains information in the FILTER column. If not uncheck the option "only pass filter" when you launch the annotation.

10 Contact us

If you encounter any problems, during download, installation or using, please contact us from web site or by email to jean-pierre.desvignes@univ-amu.fr or david.salgado@univ-amu.fr.