Contents

Welcome to the MoBiDiC Prioritizing Algorithm (MPA) !

Overview

The MPA is a prioritizing algorithm for Next Generation Sequencing molecular diagnosis. We propose an open source and free for academic user workflow.

Variant ranking is made with a unique score that take into account curated database, biological assumptions, splicing predictions and the sum of various predictors for missense alterations. Annotations are made for exonic and splicing variants up to +300nt.

We show the pertinence of our clinical diagnosis approach with an updated evaluation of in silico prediction tools using DYSF, DMD, LMNA, NEB and TTN variants from the human expert-feeded Universal Mutation Database [1] with courtesy regards of curators for pathogenic variants and from the ExAc database [2] to define the dataset of neutral variants.

Citing MPA

> **Yauy et al.** An algorithm based on an updated evaluation of prediction tools to prioritize variants with a unique score for TTN and NGS molecular diagnosis. **(2017)**

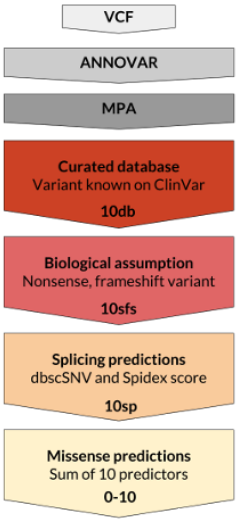

Figure 1: *MPA diagram*

Installation

To download MPA, please use git to download the most recent development tree. Currently, the tree is hosted on github, and can be obtained via:

```
$ git clone https://github.com/mobidic/MPA.git
```

Requirements

- Linux or macOS
- Python 3.5
- Standard Library:
  - argparse
  - csv
  - os
  - re
  - sys
- Librairies:
  - PyVCF

  To install python librairies see Dependencies section.

Annotation The MPA uses, as input, an annotated VCF file with Annovar [3] and the following databases :

- Curated database: ClinVar [4]
- Biological assumption : refGene [5]
- Splicing predicition : Spidex [6], dbscSNV [7]
- Missense prediction : dbNSFP [8]

  In addition you will need to annotate your VCF with Annovar [8] (cf. Quick guide for Annovar).

Check installation

You can check if all depencies are installed. Launch the script *install.py* :

```
$ python mpa.py -e
```

If everything is good you will see something like that :

```
$ python mpa.py -e
> Launch data test.
> ...
> Everything is OK !
```

Troubleshooting

Python version If you see this message :

```
$ python mpa.py -e
> [ERROR] Requires Python 3.x, not Python 2.x
```

This message means your default python is not python3. In most cases this issue can be solved by specifying wich version of python you wish to use:

```
$ python3 mpa.py -e
```

In other cases you probably need to install python3.x : documentation here.

Dependencies Maybe some dependencies are not installed, in these cases an error will show up:

```
$ python mpa.py -e
> Launch data test.
> [ERROR] Missing module: vcf
> ...
> Error !
```

In these cases you will need to install a module like *PyVCF* (module vcf). You can use *pip* to install it like this in the most cases :

```
$ pip install PyVCF
> Collecting PyVCF
> Requirement already satisfied: setuptools in /home/foo/miniconda3/lib/
    python3.6/site-packages/setuptools-27.2.0-py3.6.egg (from PyVCF)
> Installing collected packages: PyVCF
> Successfully installed PyVCF-0.6.8
```

More details for *pip* : documentation here.

Quick start

To run the MPA script, use this command line :

```
python MPA.py -i path/to/input.vcf -o path/to/ouptut.csv
```

Quick guide for Annovar

This algorithm introduce here need some basics annotation. We introduce here a quick guide to annotate your VCF files with Annovar.

Install Annovar Follow instruction to download Annovar at : > http://www.openbioinformatics.org/annovar/annovar_download_form.php

Unpack the package by using this command :

```
tar xvfz annovar.latest.tar.gz
```

Download all databases In Annovar folder, download all database needed with annotate_variation.pl:

```
perl annotate_variation.pl -buildver hg19 -downdb -webfrom annovar refGene
    humandb/
perl annotate_variation.pl -buildver hg19 -downdb -webfrom annovar
   clinvar_20170130 humandb/
perl annotate_variation.pl -buildver hg19 -downdb -webfrom annovar
   dbnsfp33a  humandb/
perl annotate_variation.pl -buildver hg19 -downdb -webfrom annovar
   dbscsnv11 humandb/
```

For Spidex database, follow instruction here :

http://www.openbioinformatics.org/annovar/spidex_download_form.php

Annotate a VCF The following command line annotate a VCF file :

```
perl path/to/table_annovar.pl path/to/example.vcf humandb/ -buildver hg19
   -out path/to/output/name -remove -protocol refGene,refGene,
   clinvar_20170130,dbnsfp33a,spidex,dbscsnv11 -operation g,g,f,f,f,f,f -
   nastring . -vcfinput -otherinfo -arg '-splicing 20','-hgvs',,,,,,
```

Montpellier Bioinformatique pour le Diagnostique Clinique (MoBiDiC)

CHU de Montpellier

France

1. Béroud, C. et al. UMD (Universal Mutation Database): 2005 update. *Hum. Mutat.* **26**, 184–191 (2005).
2. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
3. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).

4. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2015).
5. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).
6. Xiong, H. Y. et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
7. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
8. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **37**, 235–241 (2016).