

# EpiMut

Copyright (c) "2019, by Institute of Nuclear Sciences VINCA, Serbia  
Contributors: Vladimir Perovic, Branislava Gemovic, Radoslav Davidovic and  
Nevena Veljkovic  
Affiliation: Laboratory for Bioinformatics and Computational Chemistry  
URL: <http://www.vin.bg.ac.rs/180/tools/epimut.php>

## 1. Description

EpiMut is software for prediction of functional effects of amino acid substitutions (AAS) in epigenetic factors. It is based on new approach for mathematical conversion of protein sequences that uses digital signal processing of numerical representations of protein sequences. This is accomplished through encoding of amino acids using various biochemical and biophysical amino acid indices. In the EpiMut workbench, Naïve Bayes models were generated by machine learning (ML) algorithms for each gene separately.

## 2. Requirements

EpiMut is written in JAVA language and is available as standalone application, which can be executed on any operating system containing Java Virtual Machine.

Minimum system requirements for EpiMut are:

RAM: 1 GB

Disk space: 1 GB

## 3. Installation

EpiMut software is available for download at:

<http://www.vin.bg.ac.rs/180/tools/epimut.php>

In order to run the program it is necessary to install Java Runtime Environment 8 (JRE), which can be found for Windows, Linux, Mac OS and Solaris systems at:

<https://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>.

## 4. Usage

EpiMut requires input file ('input\_gene\_name.aas') to be located in the folder where executable and additional required files are. The input file must contain the amino acid substitution in each line in the following format:

original\_amino\_acid position\_in\_protein substite\_amino\_acid (example: G187V )

The position and the original amino acid should correlate to the protein wild type for specific gene for which the prediction is to be calculated.

Format of the command to execute the EpiMut is:

java -jar EpiMut.jar gene\_name input\_gene\_name.aas output\_gene\_name.txt

Where 'gene\_name' is the name of the gene for which the prediction should be calculated. List of all gene names is listed at the end of this document.

The program can be invoked from the command line as (example for gene aridla):

java -jar EpiMut.jar aridla input\_aridla.aas output\_aridla.txt

or to increase the memory needed for the java VM:

```
java -Xmx2G -jar EpiMut.jar arid1a input_arid1a.aas output_arid1a.txt
```

The calculated prediction values with tested AAS will be written in 'output\_arid1a.txt' file.

## 5. Notice:

EpiMut machine learning classifiers were generated using the H2O platform, developed by H2O.ai, Inc. licensed under the Apache License, Version 2.0.

## 6. Files

This release of the EpiMut program contains following files:

'EpiMut.jar': java jar file for calculating prediction values

'decoders.zip' and 'epimut\_decoder\_lists.zip': required additional file with physico-chemical codes and selected features

'epimut\_fasta\_sequences.zip': required additional files with reference protein sequences in fasta format

'input\_arid1a.aas': example of input file with AASs; each line have one AAS for evaluation

'output\_arid1a.txt': output of the calculation; it can be used to check and compare the results of the calculations for the input example; each line have the tested AAS and the predicted MUTATION probability [0-1]

'EpiMut.bat': batch file for Windows system which execution will automatically generate predictions for the data in input file 'input\_arid1a.aas' and write them in 'output\_arid1a.txt' file

'EpiMut.sh': script for Linux system

'LICENSE.txt': license for use of EpiMut tool as a Free Software, under the Apache License, Version 2.0

'NOTICE.txt': notice of using the libraries from H2O platform developed by the H2O.ai Inc.

## 7. List of genes

This is the list of genes (gene\_name) for which the prediction can be calculated:

arid1a	asx11
atm	atrx
bcor	crebbp
dnmt3a	ep300
ezh2	jak2
kmt2a	kmt2c
kmt2d	nsd1
setd2	sf3b1
spen	tet2
tp53	