

[MANUAL](#) [INSTALL](#) [HISTORY](#) [CONTACT](#)



There are two methods (A and B) to obtain this pipeline.

## A. Install to your local machine

### Install

1. Make sure your computer has the supported CPU architecture (X86\_64) and operating system (Linux kernel 2.6.18 or later, or Mac OS 10.7 or later).
2. Decide where to put VICTOR (for example: /path/to/VICTOR/). Replace all the "/path/to/" below with the desired location.
3. Install programs.
  - 1) go to the install location (cd /path/to/)
  - 2) `curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_victor_inst.sh | bash`
  - 3) add the location to your \$PATH (add the line "PATH=/path/to/VICTOR:\$PATH" to your .profile file), or create a module file with the line "prepend-path PATH /path/to/VICTOR" and load the module before each usage.
4. Install data. Go to VICTOR's data directory (cd /path/to/VICTOR/data/) and execute one of the following:
 

```
curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_GRCh37_inst.sh | bash
curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_GRCh38_inst.sh | bash
```

 They will download and uncompress the data files to the current directory.
5. Install the latest data updates. Some data, such as ClinVar, need to be updated more frequently. I put them in a monthly update file that has a version number different from the data folder. So, even your data folder is up-to-date, you may still need to install the latest updates. Go to VICTOR's data directory (cd /path/to/VICTOR/data/) and do one the following:
 

```
curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_GRCh37_update.sh | bash
curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_GRCh38_update.sh | bash
```
6. Install R packages. Run the following commands within R:
 

```
install.packages("logistf")
install.packages("mbest")
install.packages("lme4")
install.packages("coxphf")
install.packages("meta")
install.packages("metap")
```
7. Install third-party programs. Below is a list of the programs. The minimum requirement is tabix and GNU parallel. PROVEAN, PLINK and KING are highly recommended. Make sure these programs are in your \$PATH.

| Programs     | Comment     | It will be used for        | License  | URL   |
|--------------|-------------|----------------------------|----------|---|
| tabix        | required    | retrieval of file contents | MIT      | <a href="https://sourceforge.net/projects/samtools/files/tabix/">https://sourceforge.net/projects/samtools/files/tabix/</a>                               |
| GNU parallel | required    | parallel computing         | GPLv3    | <a href="https://www.gnu.org/software/parallel/">https://www.gnu.org/software/parallel/</a>   |
| PROVEAN      | recommended | PROVEAN score calculation  | GPLv3    | <a href="http://provean.jcvi.org/">http://provean.jcvi.org/</a>   |
| blast        | recommended | PROVEAN                    | Public   | <a href="https://blast.ncbi.nlm.nih.gov/">https://blast.ncbi.nlm.nih.gov/</a>   |
| CD-HIT 4.5.8 | recommended | PROVEAN                    | GPLv2    | <a href="http://weizhongli-lab.org/cd-hit/">http://weizhongli-lab.org/cd-hit/</a>   |
| PLINK 1.9    | recommended | quality control            | GPLv3    | <a href="https://www.cog-genomics.org/plink2/">https://www.cog-genomics.org/plink2/</a>   |
| KING         | recommended | quality control            | Unknown  | <a href="http://people.virginia.edu/~wc9c/KING/">http://people.virginia.edu/~wc9c/KING/</a>   |
| ShapeIt2     | optional    | phasing                    | Academic | <a href="https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html">https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html</a> |
| BEAGLE       | optional    | phasing                    | GPLv3    | <a href="https://faculty.washington.edu/browning/beagle/beagle.html">https://faculty.washington.edu/browning/beagle/beagle.html</a>                       |
| GATK         | optional    | VCF file combining         | BSD      | <a href="https://software.broadinstitute.org/gatk/download/">https://software.broadinstitute.org/gatk/download/</a>                                       |
| gnuplot      | optional    | drawing Manhattan plots    | gnuplot  | <a href="http://www.gnuplot.info/">http://www.gnuplot.info/</a>   |

I have made a bundle file with the above software except GNU parallel and gnuplot to help you install them altogether. Go to the installation directory ("cd ~/local/" as an example) and do the following. This command will create a folder named "pkg" and put all programs and databases in it. Afterwards, you need to add PATH in the .bash\_profile file or create a module file to set PATH. License of each software is within the bundle. Note that the nr database of the bundled PROVEAN is old. You may want to upgrade that database.

```
curl -sL --insecure http://fenglab.chpc.utah.edu/download/get_pkg_inst.sh | bash
### please add "PATH=$HOME/local/pkg:$HOME/local/pkg/PROVEAN/bin:$PATH" to ~/.bash_profile, or
### make a module file with "prepend-path PATH $HOME/local/pkg" and "prepend-path PATH $HOME/local/pkg/PROVEAN/bin".
```

8. For Mac OS only, install coreutils. This package includes the timeout program that is used by VICTOR.

```
alias curl='curl -k'  
brew install coreutils  
command -v timeout # see whether you have the "timeout" program
```

## Upgrade

To check whether your programs or data are up-to-date, you can type "vQC --genome=GRCh37 --version". Change the --genome option if you want to check other genome data. If you type this command within a directory whose fullpath contains a genome name, this option can be omitted. Not only vQC, all other programs support the --version option and will give the same results. If the data is up-to-date but the programs are not, you only need to upgrade the programs. The procedure to upgrade is the same as installation.

After each upgrade, please re-make the slurm.all\_steps script again with the new template if you want to re-analyze the data using the newer version of VICTOR, because it is likely that there are important changes in the template.

## Monthly data updates

Please see the above bullet point number 5.

## B. Amazon image

Amazon Web Service (AWS) is a HIPAA-compliant hosting and cloud computing provider. The Elastic Compute Cloud (EC2) is one of the services AWS has to provide. I have installed VICTOR and third-party programs in an EC2 instance and created an Amazon Machine Image (AMI). You can either launch an EC2 instance based on this image and do cloud computing, or download and install this image in your institute. The image is free for you, but any cost for the computing and storage is between you and Amazon. I don't get any rewards from Amazon for providing this image. I don't update this image as frequently as the VICTOR bundle. So I have taken the image off the public domain. If you need the image, please contact me.

The image is stored in AWS region US West (Oregon). Username ec2-user. To do cloud computing, in the AWS Management Console (<https://aws.amazon.com/console/>) choose "Instances", then "Launch", click "Community AMIs" and search for "victor\_GRCh37". Select the image and go on with the instance launching process. You need to choose a storage capacity that can host the operating system (6G), VICTOR (14G), other programs (13G), and your data and result files (30G or more) based on the size of your study. So the total storage may be 65G. After launching, you can enter the directory \$HOME/GRCh37, upload your data, copy slurm.all\_steps to the folder and set parameters (you may need to change PRL depending on the number of CPUs), directly run the script following the instructions inside the script, get results, then terminate the instance.