

ETCHING

Version 1.3.6b

Efficient Detection of Chromosomal Rearrangements Using a Scalable k-mer Database of Multiple Reference Genomes and Variations

ETCHING takes about 3 hours for WGS data with 30X normal and 50X tumor on 30 threads on DELL 930 server.

You can also find codes, k-mer set, and DEMO files in our website.

<http://big.hanyang.ac.kr/ETCHING/>

The demo is complete within 10 min on a desktop (AMD Ryzen 7 3700X 8-Core Processor).

Requirement

System

- 64-bit LINUX with ≥ 64 GB RAM (at least ≥ 16 GB).
- Tested on Fedora workstation, Centos, and Ubuntu

Software

- g++ ($\geq 4.7.0$), make, gawk, BWA, samtools
- Python3 ($\geq 3.6.1$, < 3.8) with pandas, numpy, scikit-learn (0.23.2), skranger (≤ 0.3), and xgboost modules

Simple guide for Linux desktop beginners to install requirements.

Fedora

```
sudo yum install -y gawk gcc gcc-c++ make cmake bwa samtools
```

```
pip3 install --upgrade pip
```

```
pip3 install numpy pandas joblib scipy Cython xgboost
```

```
wget https://github.com/scikit-learn/scikit-learn/archive/refs/tags/0.23.2.tar.gz
tar zxvf 0.23.2.tar.gz
cd scikit-learn-0.23.2
python3 setup.py install
```

```
wget https://github.com/crfllynn/skranger/releases/download/0.1.1/skranger-0.1.1.tar.gz
tar zxvf skranger-0.1.1.tar.gz
cd skranger-0.1.1
python3 setup.py install
```

CentOS 8

```
sudo yum install -y epel-release
sudo yum install -y gawk gcc gcc-c++ make cmake bwa samtools

wget https://bootstrap.pypa.io/get-pip.py && python3 get-pip.py
pip3 install numpy pandas joblib scipy Cython xgboost

wget https://github.com/scikit-learn/scikit-learn/archive/refs/tags/0.23.2.tar.gz
tar zxvf 0.23.2.tar.gz
cd scikit-learn-0.23.2
python3 setup.py install

wget https://github.com/crflynn/skranger/releases/download/0.1.1/skranger-0.1.1.tar.gz
tar zxvf skranger-0.1.1.tar.gz
cd skranger-0.1.1
python3 setup.py install
```

CentOS 7

```
sudo yum install -y epel-release
sudo yum install -y gawk gcc gcc-c++ make cmake3 bwa samtools

wget https://bootstrap.pypa.io/get-pip.py && python3 get-pip.py
pip3 install numpy pandas joblib scipy Cython xgboost

wget https://github.com/scikit-learn/scikit-learn/archive/refs/tags/0.23.2.tar.gz
tar zxvf 0.23.2.tar.gz
cd scikit-learn-0.23.2
python3 setup.py install

wget https://github.com/crflynn/skranger/releases/download/0.1.1/skranger-0.1.1.tar.gz
tar zxvf skranger-0.1.1.tar.gz
cd skranger-0.1.1
python3 setup.py install
```

Ubuntu 20.04

```
sudo apt install -y gawk gcc g++ make cmake bwa samtools python3-pip

pip3 install --upgrade pip
pip3 install numpy pandas joblib scipy Cython xgboost

wget https://github.com/scikit-learn/scikit-learn/archive/refs/tags/0.23.2.tar.gz
tar zxvf 0.23.2.tar.gz
cd scikit-learn-0.23.2
python3 setup.py install

wget https://github.com/crflynn/skranger/releases/download/0.1.1/skranger-0.1.1.tar.gz
tar zxvf skranger-0.1.1.tar.gz
cd skranger-0.1.1
python3 setup.py install
```

Ubuntu 14.04, 16.04, and 18.04

```
sudo apt install -y gawk gcc g++ make bwa samtools

# Check cmake version
cmake --version

# If cmake is <3.13 or not installed,
wget https://github.com/Kitware/CMake/releases/download/v3.13.0/cmake-3.13.0.tar.gz
tar zxvf cmake-3.13.0.tar.gz
cd cmake-3.13.0
./bootstrap && make && sudo make install
cd ..

# Check python3 version
python3 --version

# If python3 is <3.6,
sudo add-apt-repository ppa:deadsnakes/ppa
sudo apt-get update
sudo apt-get install -y python3.6
sudo update-alternatives --install /usr/bin/python3 python3 /usr/bin/python3.6 2

# Install recent version of pip3 and python modules
wget https://bootstrap.pypa.io/get-pip.py
python3 get-pip.py

# Ubuntu 14.04
echo "export PATH=${HOME}/.local/bin:\$PATH" >> ~/.bashrc
source ~/.bashrc

pip3 install numpy pandas joblib scipy Cython xgboost

wget https://github.com/scikit-learn/scikit-learn/archive/refs/tags/0.23.2.tar.gz
tar zxvf 0.23.2.tar.gz
cd scikit-learn-0.23.2
python3 setup.py install

wget https://github.com/crflynn/skranger/releases/download/0.1.1/skranger-0.1.1.tar.gz
tar zxvf skranger-0.1.1.tar.gz
cd skranger-0.1.1
python3 setup.py install
```

Installation of ETCHING

```
# Download ETCHING
git clone https://github.com/ETCHING-team/ETCHING.git

# Compile
cd etching
make
```

```
ETCHING_PATH=$PWD
```

```
# Installation
```

```
# Do either
```

```
echo "export PATH=$ETCHING_PATH/bin:\$PATH" >> ~/.bashrc
```

```
echo "export LD_LIBRARY_PATH=$ETCHING_PATH/lib:\$LD_LIBRARY_PATH" >> ~/.bashrc
```

```
source ~/.bashrc
```

```
# or
```

```
sudo cp -ar $ETCHING_PATH/bin/* /usr/bin
```

```
sudo cp $ETCHING_PATH/lib/*.so /usr/lib
```

If you want to see usage,

```
etching -h
```

If you need some example,

```
etching --example
```

Demo

After installation, you can download and run demo

```
# Download and decompress DEMO
```

```
wget http://big.hanyang.ac.kr/ETCHING/DEMO.tar.gz
```

```
tar zxvf DEMO.tar.gz
```

```
# Run demo
```

```
cd DEMO
```

```
etching -1 tumor_1.fq -2 tumor_2.fq -1c normal_1.fq -2c normal_2.fq \
```

```
-g small_genome.fa -a small_genome.gtf -f demo_PGK -o example -t 8
```

Pan-Genome k-mer set

If you have no matched normal data, PGK must be helpful to select tumor specific reads.

You can download pan-genome k-mer set (PGK) from our website.

```
# Move to etching directory
cd /somewhere/you/want/

# Download
wget http://big.hanyang.ac.kr/ETCHING/PGK.tar.gz

# Decompress
tar zxvf PGK.tar.gz

# Then, you will see PGK_20200103.kmc_pre and PGK_20200103.kmc_suf in PGK:
# Here, PGK_20200103 is the name of k-mer set to be used for ETCHING.
ls PGK
```

Alternatively, you can make your own k-mer set as follows:

```
make_pgk -i reference.list -o my_pgk -v dbSNP.vcf -g hg19.fa
```

Docker

Installation of docker

```
sudo snap install docker      # version 19.03.13, or
sudo apt install docker.io    # version 20.10.2-0ubuntu1~20.04.2
sudo usermod -a -G docker $USER
```

ETCHING on a ship (for docker users)

Download our docker image using `wget` from our website (<http://big.hanyang.ac.kr/ETCHING/download.html>)

```
# Download ETCHING docker image
wget http://big.hanyang.ac.kr/ETCHING/etching_docker.tar

# Load the image
docker load -i etching_docker.tar

# Check the image
docker images
```

Output should be like below

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
etching	1.3.6b	16647cac9a99	40 hours ago	4.3 GB

Demo for docker user

Download our DEMO

```
# Download and decompress DEMO
wget http://big.hanyang.ac.kr/ETCHING/DEMO.tar.gz
tar zxvf DEMO.tar.gz
```

Run ETCHING with docker

```
docker run -i -t --rm -v /path/to/DEMO:/work/ etching:1.3.6b \
etching -1 tumor_1.fq -2 tumor_2.fq -1c normal_1.fq -2c normal_2.fq \
-g small_genome.fa -a small_genome.gtf -f /work/demo_PGK -o example_1 -t 8
```

Here, `etching:1.3.6b` is `REPOSITORY` and `TAG` of ETCHING docker image.

Replace `/path/to/DEMO` with `/your/data/path/`.

Note: Keep `/work/` in the above command line.

Alternatively, you can run ETCHING inside docker container

```
docker run -i -t --rm -v /path/to/DEMO:/work/ etching:1.3.6b /bin/bash

etching -1 tumor_1.fq -2 tumor_2.fq -1c normal_1.fq -2c normal_2.fq \
-g small_genome.fa -a small_genome.gtf -f /work/demo_PGK -o example_2 -t 8
```

Contributors

Jang-il Sohn, Min-Hak Choi, Dohun Yi, A. Vipin Menon, and Jin-Wu Nam

Bioinformatic and Genomics Lab., Hanyang University, Seoul 04763, Korea

Contact

If you have any issues, please contact us

Jang-il Sohn (sohnjangil@gmail.com)

Jin-Wu Nam (jwnam@hanyang.ac.kr)
