VERSION 1

FEB 22, 2018

OPEN ACCESS

DOI:

dx.doi.org/10.17504/protocols.io.imrcc56

External link:

<https://doi.org/10.1093/gigascience/gix066>

Protocol Citation: Qichao Yu, Wei Zhang, Xiaolong Zhang, Yongli Zeng, Yeming Wang, Yanhui Wang, Liqin Xu, Xiaoyun Huang, Nannan Li, Xinlan Zhou, Jie Lu, Xiaosen Guo, Guibo Li, Yong Hou, Shiping Liu, Bo Li 2018.

Simulating reads for detection of transportable element insertions. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.imrcc56>

MANUSCRIPT CITATION:

Yu Q, Zhang W, Zhang X, Zeng Y, Wang Y, Wang Y, Xu L, Huang X, Li N, Zhou X, Lu J, Guo X, Li G, Hou Y, Liu S, Li B, Population-wide sampling of retrotransposon insertion polymorphisms using deep sequencing and efficient detection. *GigaScience* 6(9). doi:

[10.1093/gigascience/gix066](https://doi.org/10.1093/gigascience/gix066)

Simulating reads for detection of transportable element insertions V.1

GigaScience

Qichao Yu¹, Wei Zhang¹, Xiaolong Zhang¹, Yongli Zeng¹, Yeming Wang¹, Yanhui Wang¹, Liqin Xu¹, Xiaoyun Huang¹, Nannan Li¹, Xinlan Zhou¹, Jie Lu¹, Xiaosen Guo¹, Guibo Li¹, Yong Hou¹, Shiping Liu¹, Bo Li¹

¹BGI Shenzhen

GigaScience Press

Gigascience Database

GigaScience

ABSTRACT

We simulate paired-end reads for testing the accuracy and sensitivity of our computer program for detection of transportable element (TE) insertions (also called Mobile Element Insertions, MEIs). we named the software "Specific Insertions Detector (SID)".

BEFORE START INSTRUCTIONS

1.Setup PIRS software and SID; 2. Download the sequences of transportable elements (FASTA file) and human genome assembly sequence (hg19, FASTA file).

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

Created: Jun 26, 2017

Last Modified: Mar 02, 2018

PROTOCOL integer ID:
6545

Keywords: simulated
genomic data

- 1 Generate N random numbers from 1 to M (M is the total number of TEs) nonredundantly.

Note

In our study, N=761, M=2852.

- 2 Generate X random numbers from 1 to 22 nonredundantly.

Note

X is the number of autosomes picked up for TE insertion.

- 3 According to the lengths of the X autosomes, assign the N TEs to these X autosomes proportionably.

- 4 Generate corresponding number of random numbers for these X autosomes from 1 to the length of the chromosome.

Note

These numbers are the positions that TEs will be inserted.

- 5 Insert the TE sequence to a specific position of a specific autosome in order. Then it produces a new FASTA file (hg19_TE.fasta).

6 Run PIRS using the resulting FASTA file.

Note

Our command line is "pirs -i hg19_TE.fasta -o output.fastq.prefix -x 60".