

Manual of MaxInfo

Introduction

MaxInfo is a software program for simultaneous isoform discovery and abundance estimation from RNA-Seq data. It is flexibly switched between the annotation-free mode and the annotation-dependent mode.

Download

The software program of MaxInfo can be downloaded [here](#). Both source codes and the executable file are provided.

Version: 03/12/2015.

Instructions

Prerequisites

MaxInfo currently runs on Windows system. The core algorithms are implemented and compiled in C++ enforcement. The implementations of MaxInfo rely on two extra software programs for raw data pre-processing.

To align the sequencing read to the genome sequences, the junction-sensitive alignment tool Tophat is required. The instructions for installing Tophat can be found [here](#). Then, the [SAMTools](#) is recommended to convert BAM files into TXT format.

Installation

The source codes of MaxInfo have been compiled to generate an executable file (**MaxInfo.exe**). The executable file can be used directly. There are no additional steps for installation.

Usage

There are some preparations to run MaxInfo for isoform discovery and abundance estimation.

Stages:

1. Download the genome sequences. They can be downloaded from public databases such as the NCBI website or the UCSC website. If you want to conduct isoform discovery and abundance estimation with genome annotations, you also need to download the genome annotations.
2. Prepare the files of sequencing reads. Use SAMTools for necessary format conversion of read data to prepare the reads for the stage of read alignment.
3. Use the junction-sensitive alignment tool to align the sequencing reads to the genome sequences. Tophat is recommended here. The results of read alignment are generally of the format of BAM. Keep the junction identification results, which are usually generated as *junctions.bed*. Use SAMTools to convert the dot BAM file to the file of TXT format.
4. Use MaxInfo for isoform discovery and abundance estimation. Step into the directory where the executable file (**MaxInfo.exe**) is kept, and use the commands as follows according to your options. The isoform identification and quantification results will be found in the file *results.gtf* in the output path.

The command can be written into a dot BAT file in the same directory with the executable file; double click the dot BAT file and MaxInfo can be run. There is an example dot BAT file in the

directory where **MaxInfo.exe** is kept. When you use MaxInfo, please just overwrite the command in the example dot BAT file with your own command, and please keep the dot BAT file in the same directory with **MaxInfo.exe**. There is example data in the folder *data* of the software package, including sequencing read data, junction identification results and gene annotation. The example sequencing read data has already been converted to TXT format. The example data can be used to test the functions of MaxInfo.

Commands:

MaxInfo -i <junctions files> <read alignment files> <output path>

This command is used for isoform discovery and abundance estimation without genome annotations.

Example:

Suppose that the sequencing read data and auxiliary data (junction identification results or genome annotations) are kept in the directory D:\\data\\, and the output path is an established folder D:\\data\\estimation. The command is as follows:

MaxInfo -i D:\\data\\junctions.bed D:\\data\\read_data.txt D:\\data\\estimation

Note that the slashes in the file directory should be double-slashes and the output path needs to be an existing directory.

MaxInfo -d <genome annotations> <junctions files> <read alignment files> <output path>

This command is used for isoform discovery and abundance estimation with genome annotations. Novel isoforms are discovered.

Example:

MaxInfo -d D:\\data\\genes.gtf D:\\data\\junctions.bed D:\\data\\read_data.txt
D:\\data\\estimation

MaxInfo -t <junctions files> <read alignment files> <output path>

This command is used for isoform identification and abundance estimation with genome annotations. In this mode, MaxInfo doesn't discover novel isoforms beyond the known isoforms. It identifies isoforms and estimates the abundances of isoforms within the provided genome annotations.

Example:

MaxInfo -t D:\\data\\genes.gtf D:\\data\\read_data.txt D:\\data\\estimation