Scallop User Reference

Mingfu Shao[1] and Carl Kingsford[1]

[1]Computational Biology Department, Carnegie Mellon University
{mingfu.shao, carlk}@cs.cmu.edu

1 Installation

To install Scallop, you need to first download/compile a few sofware packages (Samtools, Boost, and GUROBI), setup the corresponding environmental variables, and then compile the source code of Scallop.

1.1 Install Samtools

Download Samtools from `http://www.htslib.org/` with verion 1.2 or higher. Compile it to generate the htslib file `libhts.a`. Set environment variable `HTSLIB` to indicate the directory of `libhts.a`. For example, for Unix plateforms, add the following statement to the file `~/.bash_profile`:

```
export HTSLIB="/directory/to/your/htslib/htslib-1.2.1"
```

1.2 Install Boost

Download Boost from `http://www.boost.org`. Uncompress it somewhere (compiling and installing are not necessary). Set environment variable `BOOST_HOME` to indicate the directory of Boost. For example, for Unix plateforms, add the following statement to the file `~/.bash_profile`:

```
export BOOST_HOME="/directory/to/your/boost/boost_1_60_0"
```

1.3 Install GUROBI

Download GUROBI from `http://www.gurobi.com/` and uncompress the package somewhere (compiling and installing are not required). You need to apply an academic license to use the full features of GUROBI (Please refer to the GUROBI documentation for more information.) After that, set two environment variables, `GUROBI_HOME` and `GRB_LICENSE_FILE`, which indicates the directory of GUROBI, and the location of your license file, respectively. For example, for Unix plateforms, add the following two statements to the file `~/.bash_profile`:

```
export GUROBI_HOME="/directory/to/your/gurobi/linux64"
export GRB_LICENSE_FILE="/location/of/your/license/gurobi.lic"
```

1.4 Compile Scallop

Get the source code of Scallop through `git`:

```
$git clone git@github.com:shaomingfu/scallop.git .
```

Execute the following commands to generate `Makefile` and compile:

```
$cd src
$aclocal
$autoconf
$autoheader
$automake -a
$./configure
$make
```

The executable file `scallop` will be present at `src/src`. You might want to link it into `bin` through

```
$cd bin
$ln -sf ../src/src/scallop .
```

2 Command line

The usage of Scallop is as follows:

```
$./scallop -c <config> -i <input.gtf> -a <algo> -o <output.gtf>
```

Parameter `config` configures the behavior of the algorithm. There is such an example configure file at `bin/example.config`. Currently we work on perfectly estimated splice graph, represented in a `gtf` file with augmented expressions. One such example can be found at `bin/example.expression.gtf`. There are three options for `algo` parameter: `scallop1`, `scallop2`, and `greedy`. With option of `scallop1`, the program will only run the core algorithm to partly decompose the given splice graph, which will predicte fewer transcripts but with higher accuracy. With option of `scallop2`, the program will completely decompose the given splice graph, using greedy algorithm following the core part of the algorithm. With option of `greedy`, the program will only use greedy algorithm to fully decompose the given splice graph. The returned transcripts will be written in the file specified by parameter `output.gtf`.