

# Package: GIFT

Songpeng Zu (zsp at mails.tsinghua.edu.cn)  
MOE Key Laboratory of Bioinformatics and Bioinformatics Division,  
TNLIST/Department of Automation, Tsinghua University  
October 31, 2014

## Content

Package: GIFT .....	1
Introduction .....	2
Compiling .....	2
Input File Formats .....	2
Running GIFT .....	3
Options .....	3
Matlab Scripts .....	3
Contact .....	3

## Introduction

GIFT is a C++ package to infer the chemogenomic features from drug-target interactions. Here we mainly focus on the interactions between drug chemical substructures and protein domain. Compared with previous method, GIFT predicts the interactions in a global way, which simultaneously analyzes all the substructure-domain interactions' contributions on drug-target interactions. An Expectation-Maximum (EM) method is then derived.

## Compiling<sup>1</sup>

GIFT needs two extra libraries before compiling it.

1. Boost C++ library ([www.boost.org/](http://www.boost.org/)).  
It helps us to deal with the multi-thread program to accelerate GIFT.
2. GSL-GNU Scientific Library (<http://www.gnu.org/software/gsl/>)  
It helps us to deal with the matrix operations.

The two packages should be installed first and the paths of the header files and library files should also be added into the systematic environment variables.

After this, you can use the command “make” to install the package of GIFT. We've written the file of “makefile” to compile it automatically.

## Input File Formats

Four different types of input files are needed:

1. The binary (0/1) matrix file of drug-protein interactions, and each row corresponds to one drug.
2. The binary (0/1) matrix file of drug-substructure relationship, and each row corresponds to one drug.
3. The binary (0/1) matrix file of protein-domain relationship, and each row corresponds to one protein.
4. The initial matrix file of the possibilities of substructures-domain interactions, and each row corresponds to one substructure. All the values should be around 0~1.

We provide the Matlab script as an example to show how to prepare these files. You can either directly use the script even without modification or you can prepare the files as your own way. We will introduce the Matlab scripts later.

---

<sup>1</sup> The Linux system is preferred.

## Running GIFT

Then you can type the commands in the shell environment:

---

```
$ GIFT -f 0.7 -p 0.0001 -t 8 -s drug2sub -d pro2dom -i dpinter -n inivalue -o outname.
```

---

We will explain each OPTION in the next subsection.

The running time depends on the file size, how many CPUs you use and so on. As our experience, it roughly takes around 4 hours if you use the data we provide under the normal computer with 8 CPUs used for GIFT (multi-thread version). When running GIFT, it will report automatically the time every iteration takes.

The output of our package is the matrix file of the possibilities of substructures-domain interactions, and each row corresponds to one substructure. The matrix is stored as one vector by the column order under the `gsl` function. We provide the Matlab scripts to show how to deal with it. We'll introduce the Matlab scripts later.

## Options

- -f: false negative parameter, and you can set it according to our paper.
- -p: false positive parameter, and you can set it according to our paper.
- -t : multi-threads number or the CPU number you want to use
- -s : the binary matrix file of drug-substructure relationship
- -d : the binary matrix file of protein-domain relationship
- -n : the initial matrix file of possibilities of substructures-domain interactions
- -i : the binary matrix file of drug-protein interactions with each row is the drug
- -o: the output file name.

## Matlab Scripts

We write several Matlab scripts to handle the data preprocessing, data analysis after you get the result and also the Matlab version of GIFT<sup>2</sup>. A Matlab script called “ExampleScript\_GIFT.m” in these scripts is a summary of how to use these Matlab scripts under step by step.

## Contact

Please feel free to contact us when you have questions: Songpeng Zu (zsp at mails.tsinghua.edu.cn)  
You can also fork the repository from the GitHub: <https://github.com/songpeng/GIFT>.

---

<sup>2</sup> We still suggest you use the C++ version of GIFT because GIFT is an iterative algorithm, which runs quite slow under the scripts language such as Matlab. But you can first try your data on the Matlab version, which can be used quite easily.