

A TOOL FOR PREDICTING lncRNA-RNA INTERACTIONS

# **LncTar Software Manual**

Version: 1.0  
October-15, 2014

Qinghua Cui  
&  
Jianwei Li

©Cuilab, Department of Biomedical Informatics  
Health Science Center  
Peking University  
Beijing China, 100191

Contact: Dr. Qinghua Cui, [cuiqinghua@bjmu.edu.cn](mailto:cuiqinghua@bjmu.edu.cn), [qinghua.cui@gmail.com](mailto:qinghua.cui@gmail.com)  
Dr. Jianwei Li, [lijianwei@hebut.edu.cn](mailto:lijianwei@hebut.edu.cn), [kjL205@163.com](mailto:kjL205@163.com)  
The software is only free for academic research  
The latest version of LncTar software is available from <http://www.cuilab.cn/lncTar>

# Table of Contents

<b>CHAPTER</b>	<b>1</b>	
<b>General information</b>		<b>page 3</b>
<b>CHAPTER</b>	<b>2</b>	
<b>How to use LncTar</b>		<b>page 5</b>
<b>CHAPTER</b>	<b>3</b>	
<b>Algorithms and comments</b>		<b>page 11</b>
<b>CHAPTER</b>	<b>4</b>	
<b>Examples and output</b>		<b>page 15</b>
<b>References</b>		<b>page 17</b>

## Chapter

## 1

# 1 General information

**Description:**

Long noncoding RNAs (lncRNAs) play important roles in regulation of many biological processes. LncRNAs have a diversity of functions through various mechanisms by binding diverse molecules, such as proteins, RNAs, and DNA. Among their binding molecules, RNA represents one of the most popular targets. Therefore, given the big number of lncRNAs, it becomes urgent and important to predict lncRNA-RNA interactions by developing bioinformatics methods. The sequences of lncRNAs are usually long and not fully complementary to the target sequences. Their regulatory functions are functioned by establishing stable joint structures with target sequences. This make is difficult for predicting lncRNA-RNA interactions. For this purpose, here we presented an efficient tool, LncTar.

LncTar is a software for predicting lncRNA-RNA interactions by means of free energy minimization. LncTar utilized a variation on the standard “sliding” algorithm approach to calculate the normalized binding free energy (ndG) and found the minimum free energy joint structure. The ndG was regard as a cutoff which determining the paired RNAs as either interacting or not. Of course, LncTar is not specific for lncRNAs but also can be used for predicting putative interactions among various types of RNA molecules, such as mRNA, noncoding RNAs including lncRNAs, pre-miRNAs, and other types of noncoding RNAs.

**Copyrights:** Jianwei Li and Qinghua Cui. The software and source codes can only be used for academic purpose. **For commercial usage, please contact with Dr. Qinghua Cui.**

**Version:** version 1.0 (Date: September-01, 2014)

**Availability:** Application and source code are available at

<http://www.cuilab.cn/lncTar>

**Platforms:** A PC or work station running Windows or Linux with Perl environment. For large RNA molecules (>5kb), it is recommended to run on a computer with at least 8 Gbyte RAM. For larger RNA molecules (>=50kb), larger RAM (>32Gbyte) is needed.

# Chapter

# 2

## 2 How to use LncTar?

### 2.1 Download and installation

#### a) Download

Download the `lnctar.zip` file from <http://www.cuilab.cn/lnctar> and then unzip `lnctar.zip`. The file `LncTar.pl` and the subfolder "Genecare" are the source code for running the LncTar software. `Examp1_RNA1.txt` and `Examp1_RNA2.txt` are the two files containing the RNA sequences for running sub-program 1 (type 1 task). `Examp2_RNA.txt` contains the RNA sequences for running sub-program 2 (type 2 task).

#### b) Perform a test run

LncTar can run in both Windows and Linux. Here is a test run in Windows.

1. Open a DOS window

This can be done by : Start->Program->accessories->Command Prompt

or by: Click "Start", enter "cmd" in the "Searching programs and files" box, and then enter "Return".

2. Change to the directory which contains the `LncTar.pl` file and other files.

3. Run the following command:

```
perl LncTar.pl -p 1 -l Examp1_RNA1.txt -m Examp1_RNA2.txt -d -0.1 -s F -o out.txt
```

4. An output indicating the computing process will be shown on the screen. At the same time, an output file "out.txt" will be generated in the same directory as the LncTar software.

## 2.2 Command line parameters

To use the LncTar software, it is important to understand the meaning of the command line parameters and learn to select a right parameter. Basically, LncTar has two types of command lines which are used for two types of different tasks, as follows.

1. perl LncTar.pl -p 1 -l file1 -m file2 -d ndG -s F -o out.txt
2. perl LncTar.pl -p 2 -f file -d ndG -s F -o out.txt

The parameter "-p" is used to select a sub-program will be used. Once the users enter "1", the first (1st) type of command line must be used. When "2" is entered, the second (2nd) type of command line must be used. The following are details about the two types of command lines (two types of tasks).

### (1). Type 1 command line (type 1 task)

We assume that users have two groups of RNA sequences (e.g. one group of lncRNA sequences and one group of mRNA sequences) contained in file1 and file2, respectively. The users aim to predict whether there is an interaction for each RNA in file1 and each RNA in file2 (Figure 1). That is, the users want to do a all-vs.-all prediction for all RNA sequences in file1 vs. all RNA sequences in file2. For this case, please select type 1 command line.

#### (1.1)The format of type 1 command line:

```
perl LncTar.pl -p <sub-program> -l <file1> -m <file2> -d <ndG> -s <F> -o <output>
```

**Note:** <> - means should be replaced by the relevant input.

There are six parameters for LncTar for this case. The following are the details about these parameters

-p<sub-program>: the sub-program. In this case, the input must be "1".

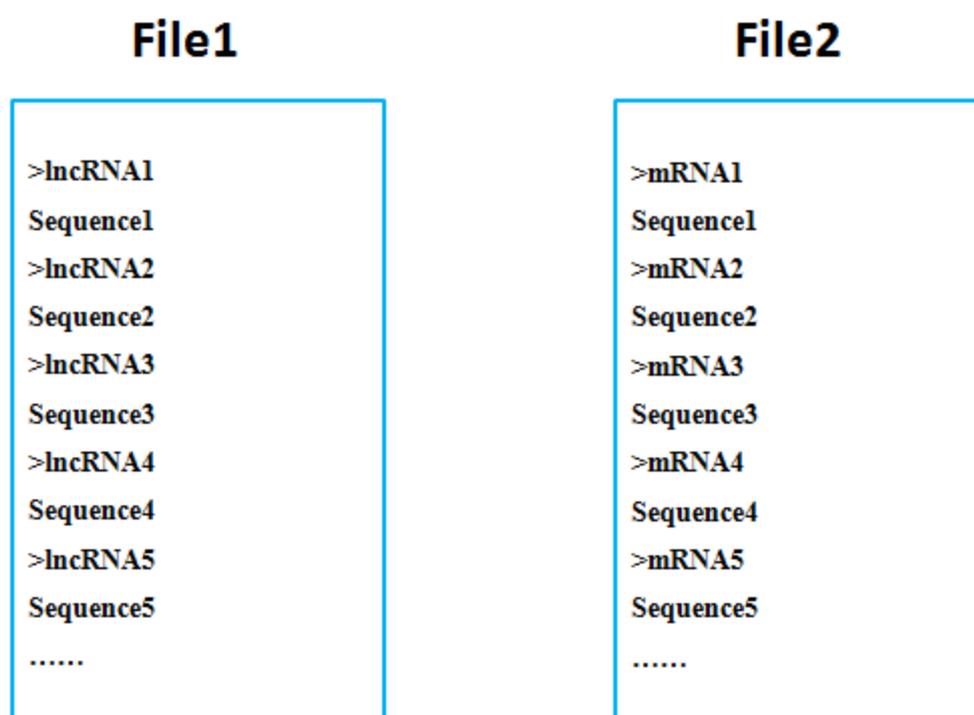
-l<file1>: the first RNA sequence file (e.g. a lncRNA sequence file ).

-m<file2>: the second RNA sequence file (e.g. a mRNA sequence file ).

-d <ndG>: the normalized deltaG (nDG), a float number that was set as a cutoff determining whether two RNA can interact with each other or not. For example, we got a ndG value of -0.12 for two RNAs, if we set the ndG cutoff as -0.08, then the two RNAs will be predicted to interact with each other. However, if we set the ndG cutoff as -0.15, then the two RNAs will be predicted to not interact with each other. Here, we suggest cutoffs as -0.08 (lowest), -0.10 (low), -0.13(medium), -0.15 (high), and -0.20 (highest).

-s <F>: a Boolean parameter indicating whether output the paired sequences. "F" means "False", "T" means "True".

-o <output>: output file name.



**Figure 1.** The file format for type 1 command line (task 1). Each file is required in FASTA format. Every two lines represents one RNA. For each one, the first line is the RNA name and the second is the RNA sequence.

## (1.2) Input file format

The two input RNA sequence files should be TEXT format. Each file is required in FASTA format. Every two lines represents one RNA. For each one, the first line is the RNA name and the second is the RNA sequence. Figure 1 shows the format.

**Note:** for the characters of RNA sequence, currently, LncTar only support four characters, "A", "C", "G", and "T". Please change "U" to "T" and remove other characters (only a limited cases contains a limited number of other characters) before running LncTar. In addition, LncTar is not case-sensitive.

## **(2). Type 2 command line (type 2 task)**

If the users want to perform predictions for pair wise RNA sequences (type 2 task, Figure 2), the LncTar software provides the type 2 command line. In this case, only 1 input file is needed. That is, the users want to do a one-vs.-one prediction for paired of RNA sequences in the input file. As shown in Figure 1, in type 2 task, LncTar only predict the interactions for lncRNA1-mRNA1, lncRNA2-mRNA2,..., and lncRNAn-mRNAn. For this case, please select type 2 command line.

### **(2.1)The format of type 2 command line:**

```
perl LncTar.pl -p <sub-program> -f <input file> -d <ndG> -s <F> -o <output>
```

**Note:** < > - means should be replaced by the relevant input.

There are five parameters for LncTar for this case. The following are the details about these parameters

-p<sub-program>: the sub-program. In this case, the input must be "2".

-f<input file>: the input paired RNA sequence file.

-d <ndG>: the normalized deltaG (ndG), a float number that was set as a cutoff determining whether two RNA can interact with each other or not. For example, we got an ndG value of -0.12 for two RNAs, if we set the ndG cutoff as -0.08, then the two RNAs will be predicted to interact with each other. However, if we set the ndG cutoff as -0.15, then the two RNAs will be predicted to not interact with each other. Here, we suggest cutoffs as -0.08 (lowest), -0.10 (low), -0.13(medium), -0.15 (high), and -0.20 (highest).

-s <F>: a Boolean parameter indicating whether output the paired sequences. "F" means "False", "T" means "True".

-o <output>: output file name.

### **(2.2) Input file format**

The input file should be formatted in TXET format. Each file has four columns separated by the key "TAB" ("\t" in computer programming). Each line represents one paired of RNA sequences. The 1st column is the name of RNA1 and the 2nd column is the sequence of RNA1. The 3rd column is the name of RNA2 and the 4th column is the sequence of RNA2. Figure 2 shows the format.

**Note:** for the characters of RNA sequence, currently, LncTar only support four characters, "A", "C", "G", and "T". Please change "U" to "T" and remove other characters before running LncTar. In addition, LncTar is not case-sensitive.

## Input File

<b>lncRNA1</b>	<b>lnc_Seq1</b>	<b>mRNA1</b>	<b>mRNA_Seq1</b>
<b>lncRNA2</b>	<b>lnc_Seq2</b>	<b>mRNA2</b>	<b>mRNA_Seq2</b>
<b>lncRNA3</b>	<b>lnc_Seq3</b>	<b>mRNA3</b>	<b>mRNA_Seq3</b>
<b>lncRNA4</b>	<b>lnc_Seq4</b>	<b>mRNA4</b>	<b>mRNA_Seq4</b>
<b>lncRNA5</b>	<b>lnc_Seq5</b>	<b>mRNA5</b>	<b>mRNA_Seq5</b>
<b>lncRNA6</b>	<b>lnc_Seq6</b>	<b>mRNA6</b>	<b>mRNA_Seq6</b>
<b>lncRNA7</b>	<b>lnc_Seq7</b>	<b>mRNA7</b>	<b>mRNA_Seq7</b>
<b>.....</b>	<b>.....</b>	<b>.....</b>	<b>.....</b>

**Figure 2.** The file format for type 2 command line (task 2). Each file contains four columns separated by TAB key. Each line represents one paired of RNAs. For each line, the first column is the RNA name and the second is the RNA sequence.

## 2.3 Output file format

Both type 1 command line (type 1 task) and type 2 command line (type 2 task) generated the same format output file. As shown in Figure 3, LncTar only output the



Chapter

3

## 3 Algorithms and comment

Basically, base pairing plays critical roles in the RNA-RNA interactions. Given that one important step in real-time polymerase chain reaction (PCR) design is detecting primer-dimer, which is also a process of base pairing. Therefore, the primer-dimer prediction algorithms shed light on the prediction of RNA-RNA interaction. For predicting lncRNA-RNA interactions, LncTar modified the primer-dimer prediction algorithm of PerlPrimer, a software for PCR primer design. PerlPrimer was an open source code software written in Perl and was initially used to design primers in standard, bisulphite and real-time polymerase chain reaction (PCR). LncTar integrated the precise melting-temperature and primer–dimer prediction algorithms in PerlPrimer for lncRNA-RNA interaction prediction. LncTar also integrated the primer-dimer checking program for calculating bimolecular secondary structures of input RNA molecules.

LncTar first takes the input lncRNA and the other RNA as the forward and reverse primers, respectively. And then LncTar predict RNA-RNA interactions by creating a binding matrix between paired RNA molecules. The complementarity of each pairing combination between RNAs is then read from the matrix. LncTar evaluates the free energy in every binding region between the two input RNA molecules. LncTar adopts the nearest-neighbor approach to maximize the number of base pairs among interacting sequences and calculates the approximate binding free energy,  $\Delta G$  (dG), of each pairing with the recent data that indicates the stability of complementarity.

In the original PerlPrimer software, two classes of primer-dimer stability are calculated. One is the extensible dimer, which will reduce the amplification of product. The other is the non-extensible dimer, which can reduce the free primer population in a reaction. In this study, our aim is to predict the interaction of lncRNAs and RNAs, only the non-extensible dimers are retained in LncTar, which can reduce running time of algorithm.

The flowchart of LncTar is shown as Figure 4.

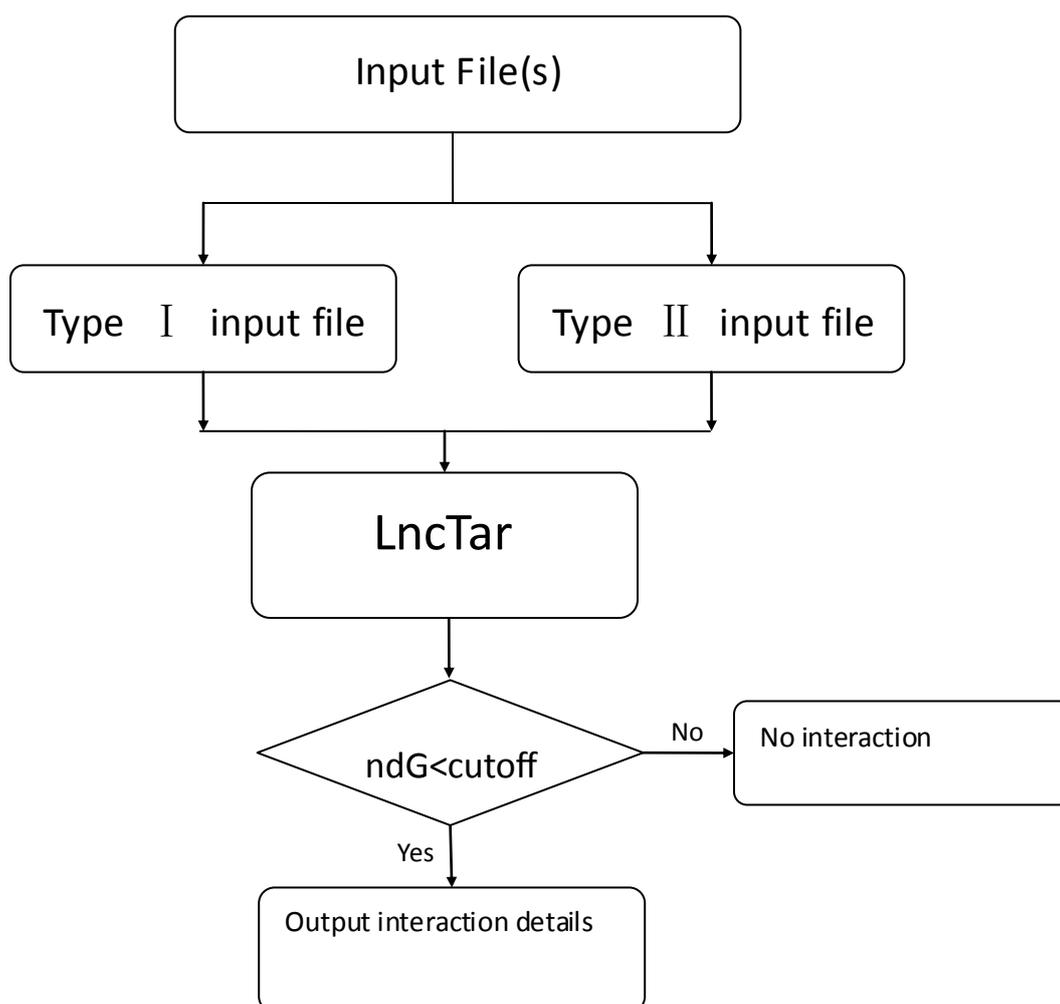
### **User-Defined Parameters**

In LncTar, the most significant modification we made was that we introduced the normalized free energy (ndG), which reflects the relative stability of internal base pairs in the paired RNAs. Given that normally longer RNA molecules often have lower  $\Delta G$ , therefore, it seems not suitable to take  $\Delta G$  as a standard to determine whether two RNA molecules interact with each other or not. A ndG by the size of RNA molecules will be better than the original  $\Delta G$ . Therefore, in LncTar, we

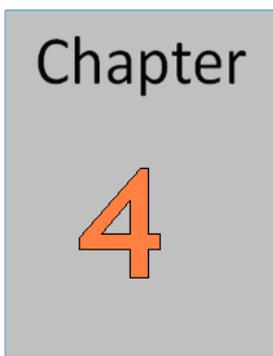
introduced normalized  $\Delta G$  (ndG), which was calculated by the following equation:

$$\text{ndG} = \Delta G / \text{length}_{\text{bindingregion}}$$

where  $\text{length}_{\text{bindingregion}}$  was the length of the binding region that lncRNA and the RNA interact with each other, respectively. The  $\text{ndG}$  was set as a float number and could be a cutoff to determine whether two RNA molecules interact with each other or not. LncTar will decide the input two RNA molecules interact with each other if it gets a ndG value equal to or less than a cutoff, for example, -0.1. Here, based on data from our experiments we suggest several cutoffs, such as -0.08 (low confidence), -0.10 (low confidence), -0.13 (medium confidence), -0.15 (high confidence), and -0.20 (very high confidence).



**Figure 4.** Flowchart of LncTar for predicting lncRNA-RNA interactions.



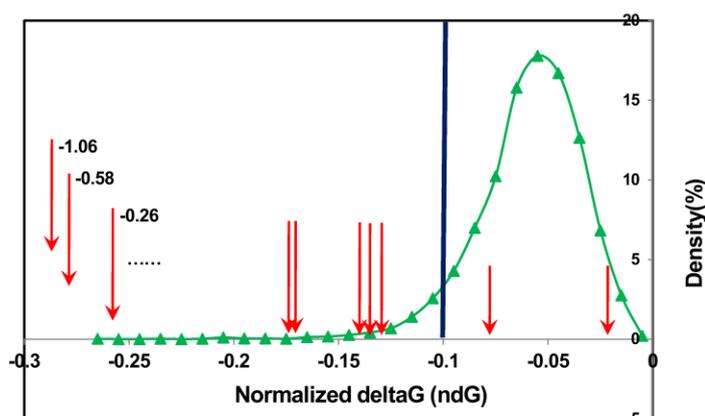
## 4 Examples and output

To test the accuracy of LncTar, we applied it to known lncRNA-mRNA interactions. For this purpose, we first curated the LncRNADisease database and the NPInter database for experimentally supported lncRNA-mRNA interactions. As a result, we obtained 10 pairs of lncRNA-mRNA interactions. We next applied LncTar to the 10 pairs of lncRNA-mRNA interactions. Table 1 shows the results of our method, it includes the free energy ( $\Delta G$ ) and the normalized deltaG (ndG) of two interacting RNAs.

**Table 1.** The ten candidate pairs of lncRNA and its target mRNA for validation experiments

lncRNA	mRNA	dG	ndG
EMX2OS	EMX2_NM_004098	-92.61	-0.0318
NPPA-AS1	NPPA	-75.92	-0.0885
BC200	CAMK2A_NM_171825	-26.75	-0.1338
BDNF-AS_NR_033312.1	BDNF_NM_170734	-277.26	-0.1362
BC200	ARC	-27.46	-0.1373
BC200	MAP1B	-34.4	-0.1720
BACE1-AS	BACE1_NM_012104	-137.79	-0.1727
HIF1A-AS1	HIF1A_NM_001530	-168.35	-0.2582
HIF1A-AS2	HIF1A_NM_181054	-1182.5	-0.5765
PEG3-AS1	PEG3_NM_006210	-1395.39	-1.0619

As shown in Figure 5, the 5% percentage of low normalized free energy (ndG) is -0.1 (Green line in Figure 5). Thus, we set -0.1 of ndG as a cutoff to assign one pair of lncRNA and mRNA to be interacting with each other or not. That is, the lncRNA-mRNA pairs with an  $\text{ndG} \leq -0.1$  will be predicted to be that they have interactions with each other. Otherwise, they do not interact with each other. As a result, 8 (80%) of the 10 experimentally supported lncRNA-mRNA interactions (red arrows in Figure 5) are successfully predicted, suggesting that LncTar has a reliable accuracy. For a comparison, we applied LncTar to 5000 random lncRNA-mRNA pairs as well, experimental results show our tool is highly accurate.



**Figure 5.** The experimental results of 10 pairs of lncRNA-mRNA interactions.