Description

NorahDesk is a small non-coding RNA (ncRNA) detection tool for RNA sequencing (RNA-Seq) data. It utilizes the coverage-distribution of small RNA sequence data and thermodynamic assessments of secondary structure to reliably predict and annotate ncRNA classes.

The program is implemented in Java.
It contains two packages: ncRNA_ prediction and ncRNA_annotation.
ncRNA_detection program predicts small ncRNA transcripts from RNA-Seq data.
ncRNA_annotation program categorises predicted transcripts into different classes of known ncRNAs and un-annotated transcripts (potential novel ncRNAs).

NorahDesk requires two external programs: BEDTOOLS (http://code.google.com/p/bedtools/) and Vienna RNA Package (http://www.tbi.univie.ac.at/RNA/). It also requires genome (chromosome) sequence and annotation files; the annotation files in mouse are distributed with the program.

The program is tested only on Mac OS X and Linux platforms.

Installation

Installation of external programs
BEDTOOLS and Vienna RNA Package have to be installed.

Setting the paths to all programs
NorahDesk creates output files as well as several temporary directories/files in the current working directory. Thus the easiest way to test the program is to set the temporary paths, for example,

export PATH=$PATH:/….. /BEDTools-Version-2.11.2/bin
export PATH=$PATH:/…../ViennaRNA-1.8.4/bin
export CLASSPATH=/…../NorahDesk/bin

then run NorahDesk from the directory where the output files will be created.

Commands and options

To run ncRNA_prediction:
java ncRNA_prediction.PredictRNA [options] GENOMESEQ=<directory name> INFILE=<input file>

GENOMESEQ=<directory name>	Name of a directory where the chromosome sequences are stored. If it is not in the working directory, the full path of the directory is required.
INFILE=<file name>	Input file. It has to be either ‘bam’ or ‘bed’ format with ‘.bam’ or ‘.bed’ extension.

Options:

NUMCHR=<19>	Number of chromosomes excluding X and Y, e.g. 19 for mouse and 22 for human; default is 19
MAXDG=<-5>	Maximum cut off free energy for hybridised contigs; the default is -5 (kcal/mol).
MAXDIS=<250>	Maximum distance that contigs are hybridised; default is 250 (nucleotide).

To run ncRNA_annotation:
> java ncRNA_annotation.AnnotateRNA NCRNA=<file name> ENSEMBLE=<file name> INFILE=<file name>

NCRNA=<file name>	ncRNA annotation file. If it is not in the working directory, the full path of the file is required.
ENSEMBLE=<file name>	Ensemble gene annotation file. If it is not in the working directory, the full path of the file is required.
INFILE=<file name>	Input file. This is the output file from 'ncRNA_prediction' program.

Output files

Output file of 'ncRNA_prediction' is a list of predicted transcripts in BED format.
Output files of 'ncRNA_annotation' are a list of transcripts that overlapped with known ncRNAs, a list of potential novel ncRNAs (both in BED format), and summary files for the transcripts and the known ncRNAs.