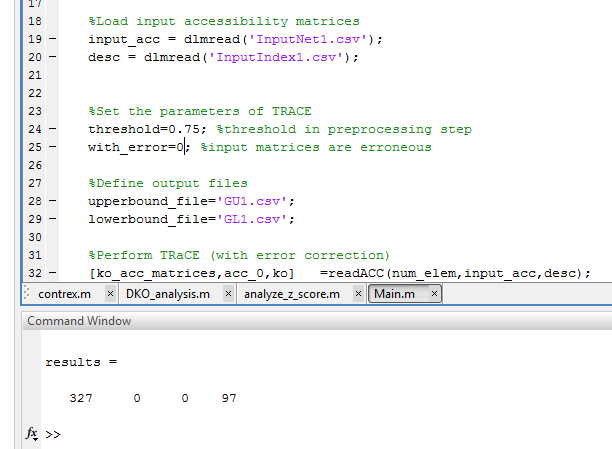
The subroutines in this package were created using MATLAB® 2011a and have also been tested on MATLAB® 2012b platform. Please refer to our manuscript titled “Ensemble Inference and Inferability of Gene Regulatory Networks” for more detailed information about the algorithm.

The package contains the following files:

1. This README file
2. The folder TRaCE\_code containing:
   1. Main.m an example of ensemble inference using TRaCE
   2. trace.m a function for ensemble inference using TRaCE (requires contrex.m and adj2acc.m)
   3. contrex.m a function for calculating ConTREx of a graph
   4. adj2acc.m a function for calculating accessibility matrix of a graph from its adjacency matrix.
   5. readAcc.m a function for reading accessibility matrices from input files
   6. Adj.csv adjacency matrix of example network containing 100 genes
   7. InputNet1.csv input data file containing error free accessibility matrices of example network
   8. InputIndex1.csv a description file of InputNet1.csv
   9. InputNet2.csv input data file containing erroneous accessibility matrices of example network
   10. InputIndex2.csv a description file of InputNet2.csv
3. The folder DREAM4 containing:
4. SKO\_analysis.m an example of ensemble inference from single knockout expression data.
5. DKO\_analysis.m an example of ensemble inference from single and double knockout expression data.
6. analyze\_z\_score.m a function for calculating z-scores for differential expression analysis and constructing accessibility matrices.
7. sort\_bounds.m a function for generating the ranked list of edges based on the ensemble bounds from trace.m and the z-scores output from analyze\_z\_score.m. (requires contrex.m and adj2acc.m).
8. read\_gold\_standard.m a function for reading gold standard network from the output of GeneNetWeaver
9. remove\_diagonal.m a function for setting the diagonal elements of a matrix to zero
10. Net1 a folder containing 10 replicates of single- and double-gene KO data for one of the gold standard networks of DREAM4 100-gene network inference subchallenge. The complete simulated gene expression data of DREAM4 gold standard networks are available from <http://www.cabsel.ethz.ch/tools/trace>.
11. License.txt modified BSD license for TRaCE

**In TRaCE\_code folder:**

Running Main.m without any modification will produce the ensemble lower and upper bounds from the accessibility matrices in InputNet1.csv, and save the bounds as the files GL1.csv and GU1.csv respectively. You should also get the following in MATLAB prompt:



These results indicate that there are:

327 edges in the upper bound but not in the true network,

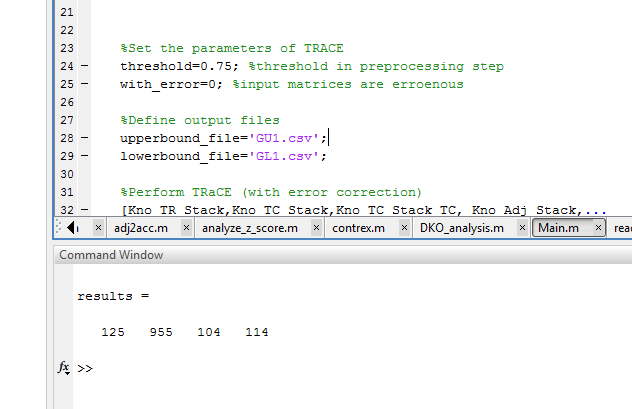
0 edge in the lower bound but not in the true network,

0 edge in the true network but not in the upper bound, and

97 edges in the true network but not in the lower bound,

respectively.

To perform TRaCE with erroneous accessibility matrices, change the variables input\_acc and desc in Main.m to InputNet2.csv and InputIndex2.csv, respectively. Running Main.m again will produce the ensemble lower and upper bounds without error correction. Note that we strongly recommend using TRaCE with error correction when dealing with erroneous accessibility matrices. The above procedure is performed only for validation test of the codes. These bounds are again saved as the files GL1.csv and GU1.csv. You should also get the following in MATLAB prompt:



These results indicate that there are :

125 edges in the upper bound but not in the true network,

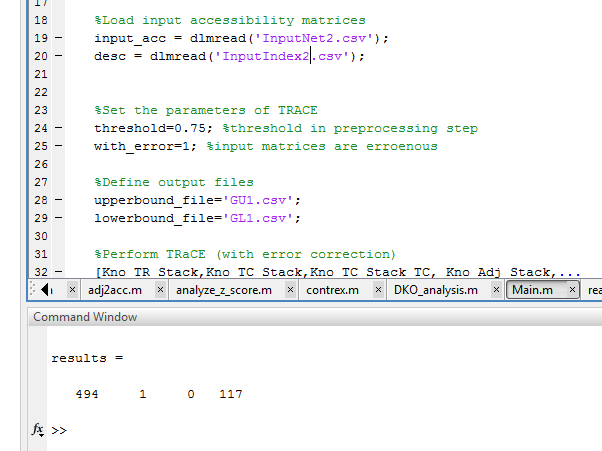
955 edges in the lower bound but not in the true network,

104 edges in the true network but not in the upper bound, and

114 edges in the true network but not in the lower bound,

respectively.

To perform TRaCE with error correction, set the variable “with\_error” in Main.m to 1 and run Main.m (using the input data file InputNet2.csv). In this case, TRaCE will run with error correction and the ensemble lower and upper bounds are saved as the files GL1.csv and GU1.csv, respectively. You should also get the following in MATLAB prompt:



These results indicate that there are :

494 edges in the upper bound but not in the true network,

1 edge in the lower bound but not in the true network,

0 edge in the true network but not in the upper bound, and

117 edges in the true network but not in the lower bound,

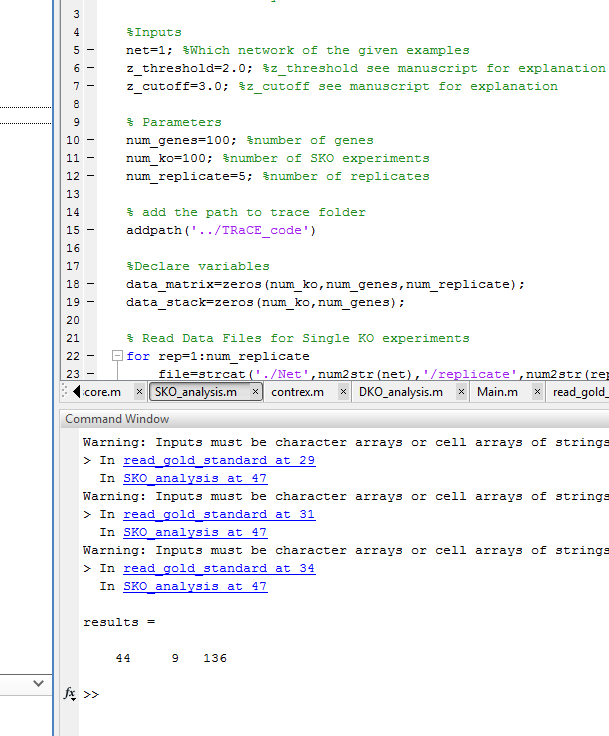
respectively.

The instructions for input file preparation can be found by typing ‘help trace’ on MATLAB command line.

**In DREAM4 folder:**

Running SKO\_analysis.m will (a) execute Z-score transformation analysis of single-gene KO data in the subfolder Net1 to produce accessibility matrices, (b) perform TRaCE with error correction to generate the ensemble lower and upper bounds from the accessibility matrices in (a), and (c) produce ranked list of edges based on confidence scores. The ranked list of edges is saved as the file DREAM4\_Example\_InSilico\_Size100\_1.txt, following the format of DREAM4 in silico network inference submission.

You should get the following in MATLAB prompt:



These results indicate that there are:

44 edges in the true network but not in the upper bound,

9 edges in the lower bound but not in the true network,

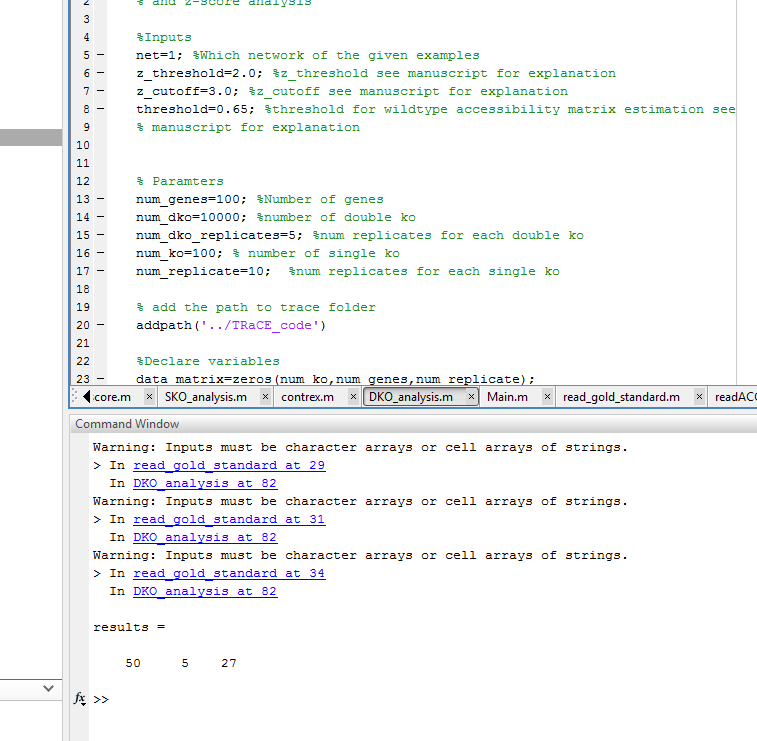
136 edges in upper bound but not in the lower bound,

respectively.

MATLAB may also produce the following warning: ‘Warning: Inputs must be character arrays or cell arrays of strings’. We can safely ignore this warning for our purpose. You could also switch off the warning using the command “warning('off','all')”.

The parameters z\_threshold and z\_cutoff are used in the construction of accessibility matrices from data (refer to our manuscript for more details). The variable num\_replicate sets the number of replicates to be used. In this case, we can set the number of replicates between 1 and 10.

Running DKO\_analysis.m will carry out the same procedure as in SKO\_analysis.m using single and double gene knockout expression data in folder Net1. The ranked list of edges is saved as DREAM4\_Example\_InSilico\_Size100\_1.txt. You should also get the following in MATLAB prompt:



These results indicate that there are :

50 edges in the true network but not in the upper bound,

5 edges in the lower bound but not in the true network,

27 edges in upper bound but not in the lower bound,

respectively.

In the manuscript, we used the ranked lists for the calculation of AUROC and AUPR. The AUROC and AUPR were evaluated using MATLAB subroutines from another publication: C. Siegenthaler and R. Gunawan, “Assessment of Network Inference Methods: How to Cope with an Underdetermined Problem”, (<http://dx.doi.org/10.1371/journal.pone.0090481>).

The parameters z\_threshold and z\_cutoff are the same as those in SKO\_analysis.m. The parameter threshold is used in the pre-processing of accessibility matrices in TRaCE with error correction (refer to our manuscript for more information). For the dataset in Net1, you can set num\_dko\_replicates between 1 and 5.

You can repeat SKO\_analysis.m and DKO\_analysis.m using other KO datasets in our manuscript by changing the variable “net” to the desired input file (from Net1 to Net5).