User Manual for CU-DREAM

Chatchawit Aporntewan[a] and Apiwat Mutirangura[b]
[a]Department of Mathematics, Faculty of Science, [b]Department of Anatomy, Faculty of Medicine
Chulalongkorn University, Bangkok 10330, Thailand
Chatchawit.A@chula.ac.th, Apiwat.M@chula.ac.th

1. System requirements

CU-DREAM requires a computer with the following settings.

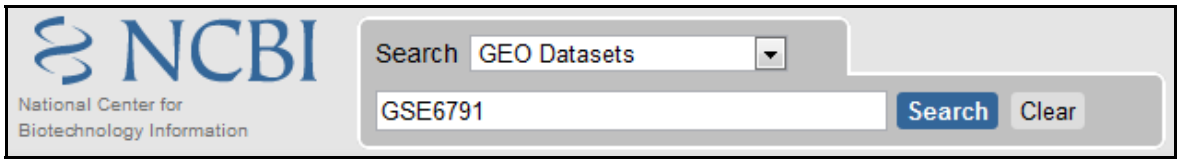
- Windows operating system.
- Microsoft .NET framework 3.5 or higher (download from the link below)
  http://www.microsoft.com/downloads/details.aspx?FamilyId=333325fd-ae52-4e35-b531-508d977d32a6&displaylang=en
- Microsoft Excel 2007 or higher.
- Microsoft Office system Primary Interop Assemblies or PIA (download from the link below)
  http://www.microsoft.com/downloads/details.aspx?FamilyID=59daebaa-bed4-4282-a28c-b864d8bfa513&displaylang=en

2. An example

An example in this section illustrates how to intersect two microarray datasets, GSE6791 and GSE7803. All steps are as follows.

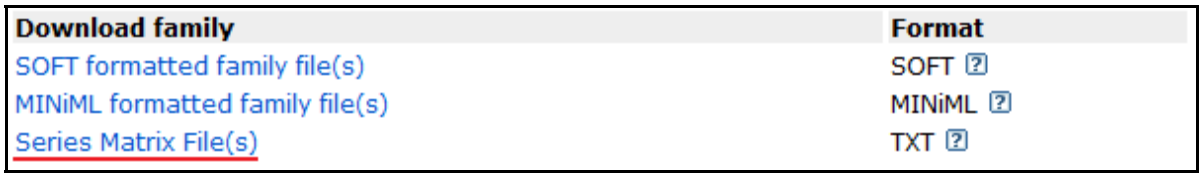
- Create a working directory, for example, C:\CU-DREAM.

- Go to NCBI website (http://www.ncbi.nlm.nih.gov), and search for datasets.

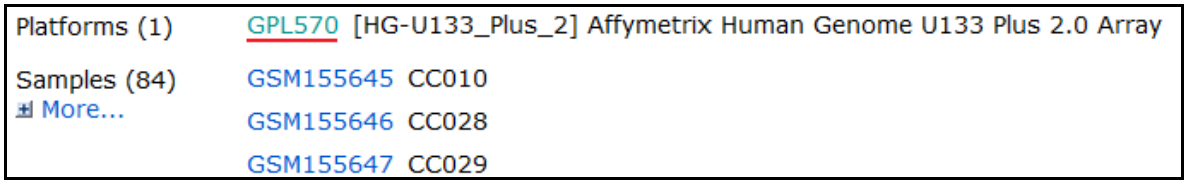


- Click at the GSE title.

- Scroll down the page, download the series matrix file, save it to the working directory, and uncompress it. Note that the link is FTP protocol (not HTTP). Your browser may not be able to download due to incorrect settings.



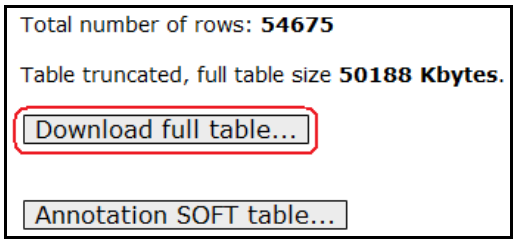
- On the same page of the previous step, click at the platform title.



- The column “Gene Symbol” is **indispensable**. CU-DREAM uses this column to convert “Probe ID” to “Gene Symbol.” If this column does not appear, please download the annotation file from ftp://ftp.ncbi.nih.gov/pub/geo/DATA/annotation/platforms, save it to the working directory, and skip the next step.

Gene Symbol
DDR1
RFC2
HSPA6
PAX8
GUCA1A
UBA7
THRA
PTPN21
CCL5

- Download full table and save it to the working directory. If the button for downloading full table is not provided, you can copy the whole table and paste it to a text file.



- Repeat the same procedure for another dataset, GSE7803.

- Download “template.xls” from our website (the URL below), make two copies, name GSE6791.xls and GSE7803.xls, and save them to the working directory.

  URL: http://pioneer.netserv.chula.ac.th/~achatcha/cu-dream/template.xls

  The file name may be GSE6791_1.xls, GSE6791_2.xls, GSE6791_3.xls, and so on if you have multiple experiments.

- Edit both GSE6791.xls and GSE7803.xls to set the parameters.

Please fill the following boxes.

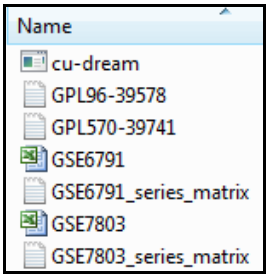
GSE file:	GSE6791_series_matrix.txt	Series matrix file downloaded from www.ncbi.nlm.nih.gov
Annotation file:	GPL570-39741.txt	Annotation file downloaded from www.ncbi.nlm.nih.gov
T-test parameter (tail):	Two-tailed distribution	Parameter for ttest() in Microsoft Excel
T-test parameter (type):	2 Series with unequal standard	Parameter for ttest() in Microsoft Excel
Differential expression:	Up	Direction of differential expression
P-value threshold:	0.01	P-value threshold for t-test

Warning for paired t-test: subjects in the same row are paired.

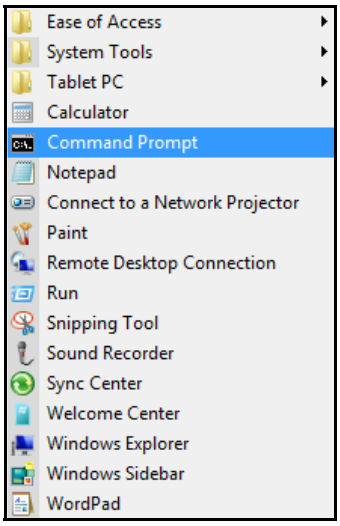
Experimental group	Note (optional)	Control group	Note (optional)
GSM155645		GSM155665	
GSM155647		GSM155666	
GSM155648		GSM155667	
GSM155651		GSM155668	
GSM155652		GSM155669	
GSM155653		GSM155670	
GSM155656		GSM155671	
GSM155660		GSM155672	
GSM155661			
GSM155662			
GSM155664			

- Download the executable file from the URL below and save it to the working directory.
  URL: http://pioneer.netserv.chula.ac.th/~achatcha/cu-dream/cu-dream.exe

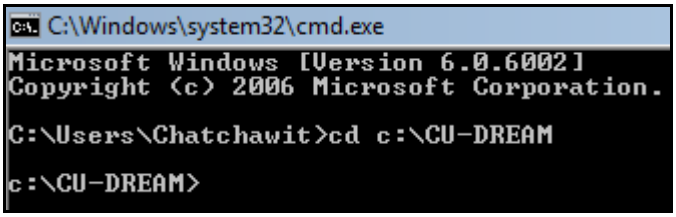
- Now there should be at least 7 files in the working directory (including the executable file).



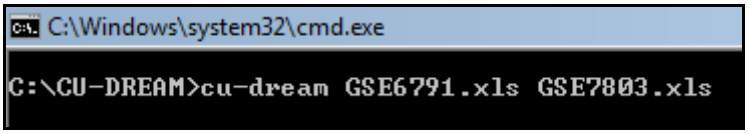
- Start the "Command Prompt" in Programs → Accessories.



- Change path to the working directory by typing "cd c:\CU-DREAM" and pressing enter.



- Start the program by typing "cu-dream GSE6791.xls GSE7803.xls" and pressing enter. The running time on a desktop with Intel Core 2 Duo E6750 and 4GB RAM is about 10 minutes.



- It is important to note that our program uses MS Excel for inputting, outputting, and statistical calculation. We strongly recommend closing all Excel applications before starting our program, and do not use Excel during the program execution.

- If the program succeeds, you will see the following message.

```
      Loading GSE from file >GSE6791.xls<.
              Loading annotation from file >GPL570-39741.txt<.
              Loading subjects from file >GSE6791_series_matrix.txt<.
              Loading array from file >GSE6791_series_matrix.txt<.
                      5000 probes read.
                      10000 probes read.
                      15000 probes read.
                      20000 probes read.
                      25000 probes read.
                      30000 probes read.
                      35000 probes read.
                      40000 probes read.
                      45000 probes read.
                      50000 probes read.
                      54675 probes read.
              Calculating t-test.
                      5000 probes calculated.
                      10000 probes calculated.
                      15000 probes calculated.
                      20000 probes calculated.
                      25000 probes calculated.
                      30000 probes calculated.
                      35000 probes calculated.
                      40000 probes calculated.
                      45000 probes calculated.
                      50000 probes calculated.
                      54675 probes calculated.
      Loading GSE from file >GSE7803.xls<.
              Loading annotation from file >GPL96-39578.txt<.
              Loading subjects from file >GSE7803_series_matrix.txt<.
              Loading array from file >GSE7803_series_matrix.txt<.
                      5000 probes read.
                      10000 probes read.
                      15000 probes read.
                      20000 probes read.
                      22283 probes read.
              Calculating t-test.
                      5000 probes calculated.
                      10000 probes calculated.
                      15000 probes calculated.
                      20000 probes calculated.
                      22283 probes calculated.
      Intersecting.
Saving file.
```

- Finally, the file "Intersect_GSE6791_GSE7803.xls" is obtained in the working directory. The first sheet shows GSE6791 array. The columns, from left to right, are probe id, gene symbol, the mean of experimental group, the mean of control group, differential mean, unadjusted p-value (t-test), experimental group, and control group. The second sheet shows GSE7803 array in the same format.

	A	B	C	D	E	F	G	H	I
1	Probe ID	Gene Symbol	Mean1	Mean2	Mean1 - Mean2	P-value	GSM155645	GSM155647	GSM155648
2									
3	1007_s_at	DDR1	12.04	12.05	-0.02	8.94E-01	11.49	11.75	12.07
4	1053_at	RFC2	7.47	7.10	0.37	3.94E-02	7.47	7.13	7.46
5	117_at	HSPA6	7.64	7.31	0.32	2.08E-01	7.81	7.40	9.93
6	121_at	PAX8	9.89	10.57	-0.68	4.98E-05	10.41	10.01	9.96

GSE6791 GSE7803 Intersection

- The third sheet shows the intersection between two arrays.

		GSE7803				
		Up (0.01)	Not up			
GSE6791	Up (0.01)	1,186	1,967	3,153	P-value	0.00E+00
	Not up	800	9,105	9,905	Odds Ratio	6.86
		1,986	11,072	13,058	Upper 95% CI	6.20
					Lower 95% CI	7.60

3. Simple Count Algorithm

It is important to address how we count the number of genes for chi-square test. A probe can be either unique (having only one corresponding gene) or homology (having multiple corresponding genes). The universe, the set of all genes, is defined as the annotated genes in both datasets, excluding the genes that possess only one homology probes. The Student's t-test is performed on all probes. A gene is counted as being up-regulated if and only if at least one unique probe is significantly up-regulated, or at least two homology probes are significantly up-regulated. Counting down-regulated genes is similar.

4. Contact

For any comments, please contact

Assist. Prof. Chatchawit Aporntewan
Department of Mathematics, Faculty of Science,
Chulalongkorn University, Bangkok 10330, Thailand
E-mail: Chatchawit.A@chula.ac.th Office: 02-218-5225 (ext 11) Cell: 081-920-1977.