BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Software

Volume 9(14)

HOMO: A novel script for data mining of microarrays

Adda Jeanette Garcia-Chequer[1, 2*], Hueman Jaimes-Diaz[1] & Martha Verónica Ponce-Castañeda[2]

[1]Laboratory of Biotechnology and Genomic Bioinformatics, Department of Biochemistry, National School of Biological Sciences, IPN, Mexico, D.F., Mexico; [2]Laboratory of Molecular Biology, Medical Research Unit on Infectious Diseases, Pediatric Hospital, Mexican Institute of Social Security, Mexico, D.F., Mexico; Adda Jeanette García Chéquer – Email: agarciac0922@ipn.mx; *Corresponding author

Received July 15, 2013; Accepted July 16, 2013; Published August 07, 2013

Abstract:
HOMO.pl is a perl script that allows extracting important registers from an extensive data table or microarrays results. It is very useful in data mining for microarrays analysis. HOMO works as a homogenizer that converts the initial data table into more specific and manageable data according to a list of important genes or terms. This is very useful when a pathway, a condition, or a GO-Term is studied. The HOMO script has two inputs and one principal output. A table with the microarray data results is used as an input and a list of genes or important terms is used as a second input. The output is an adjusted table from the microarray results that contains only the genes included in the input list. HOMO´s principal goal is to simplify the subsequent analyses to the microarray data.

Availability: *HOMO*.pl and a suite of example files are available by electronic mail request.

Background:
Expression microarrays allow the evaluation of thousands of genes simultaneously, but the correct analysis of this amount of data remains an important challenge and depends on the particular goals of each study. There are many tools for the analysis of microarray data. Although a comprehensive analysis is desired sometimes, the most common is a specific or targeted analysis. Usually, the first treatment of the microarray data is data normalization and the extraction of differentially expressed genes. Both analyses could be easily performed using software like TM4 or Limma **[1, 2]**. Once a differentially expressed genes matrix is obtained, the next step will be the functional analysis of these data. There are other programs that help with this functional analysis like GenMAPP **[3]** or KEGG **[4]**. GenMAPP allows general analyses and gives a list of representative pathways. KEGG offers the opportunity of downloading lists of a particular condition like apoptosis, response to stress, cell cycle, insulin secretion, etc. There is also Gene Ontology (GO) where some lists of biological processes, molecular function, and cellular component lists of genes are available **[5, 6]**. Sometimes we have a list of interest genes, a condition that our group is researching about, or a combination of some cellular processes like Hematopoietic cell lineage and cellular motility. In any of these cases, the HOMO script here presented helps to adjust the microarrays results to a specific gene list and so simplify the subsequent analyses.

HOMO offers a gene specific extraction from a microarray data table using a gene list that could be obtained from any source or made by your own. The extraction that does HOMO is easily handmade when the interest data and the microarray data are small (about hundreds of registers), but when the microarray data is extensive (about thousands) this easy step becomes a trouble. HOMO´s goal is to reduce this kind of trouble.

Inputs:
The HOMO script asks for two principal inputs. The first one is a list of words; these could be genes or any other kind of words. This list must have a word by row. A pathway gene list or a GO-Term gene list would be good examples of this input. The second input is a table of microarray results. This kind of table has one or more columns assigned to gene identifiers, these could be symbols, or accession numbers, or any other. It is important that the list of the first input has the same type of identifier that the second output does. The format for both inputs must be tab delimited text. HOMO.pl as a perl script

runs in a console and the name of the results table must be captured complete including the extension of the file.

Output:
The principal HOMO´s output is an adjusted table. Using the second input (the microarray data) HOMO extracts all the registers present in the first input and creates a new data file. When a file is adjusted, HOMO allows to adjust another file to the same gene list, so you will be able to adjust many archives to the same gene list, for example when you are studying a pathway and you have many different microarrays or a course time and you need to use the same list for adjusting all the datasets.

HOMO also calculates a non-repeated gene list and the new gene list is a secondary output.

Caveat & future development:
It is preferable that the microarray data be normalized for the immediate analysis of the adjusted data. For a future development, more functions including a selection of datasets of genes will be added. We are working on a database of relevant gene lists or datasets involved with isolated interests; these include Response to Virus gene list, Immune Response gene list, Motility of tumor cells gene list, Stem cell gene list and many others. The final goal is to simplify the way to make a functional analysis of a microarray result.

References:
[1] Saeed AI *et al. Biotechniques.* 2003 **34:** 374 [PMID: 12613259]
[2] Wettenhall JM & Smyth Gk. *Bioinformatics.* 2004 **20:** 3705 [PMID: 15297296]
[3] Salomonis N *et al. BMC Bioinformatics.* 2007 **8:** 217 [PMID: 17588266]
[4] Kanehisa M & Goto S, *Nucleic Acids Res.* 2000 **28:** 27 [PMID: 10592173]
[5] Ashburner M *et al. Nat Genet.* 2000 **25:** 25 [PMID: 10802651]
[6] http://www.geneontology.org/

Edited by P Kangueane
Citation: **Chequer *et al.*** Bioinformation 9(14): 748-749 (2013)