

Venn Mapper

Software and manual written By Marcel Smid.

This manual covers Venn Mapper version 1.01

Software copyright by Marcel Smid/Erasmus-MC 2002

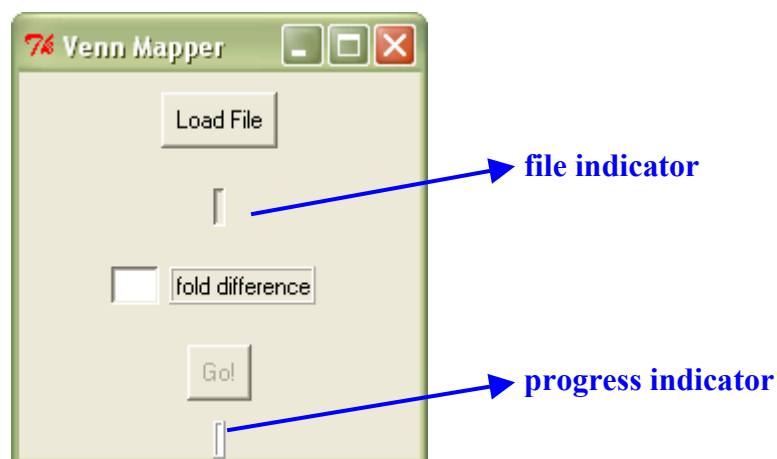
*Original concept: Marcel Smid, Lambert CJ Dorssers and Guido Jenster,
Bioinformatics, 2003; vol.19 nr. 16 p2065-2071*

Introduction

Venn Mapper is a program that compares heterologous microarray data sets, based on the number of common, differentially expressed genes. The application loads microarray data (gene expression ratios) and determines which genes are up- or down-regulated by a user-defined ratio cut-off level. For each experiment, lists of differentially expressed genes are computed. Every list will be compared to every other list, and the number of co-occurring genes will be calculated. With the use of the binomial distribution, so called z-values can be assigned to the overlap found between two lists. Different output files are generated by the Venn Mapper, reporting: (1) the number of up- and down-regulated genes per array and in the comparisons, (2) the z-values, and (3) the gene-IDs.

This manual is intended as a reference for using the software, and not as a comprehensive introduction to the methods employed. For more information on those methods, see reference 1.

Starting Venn Mapper



(this picture was taken in Windows XP, other operating systems may show a different 'phenotype' of this window)

The workflow is as follows :

Load file :

A standard browse window appears so the user may select the appropriate file.

Venn Mapper will only work properly with tab-delimited text files in the following format:

gene-ID	exp A	exp B	exp C
Hs.1	ratio	ratio	ratio
Hs.2	ratio	ratio	ratio
Hs.3	ratio	ratio	ratio

The first row must contain the identifiers of the columns. The first record describes the gene-identifier used, and can be any alpha-numeric value; this field is not used in the application. Codes for the microarray experiments (such as exp A) and gene identifiers (here Unigene cluster numbers) are required. The ratio-fields contain the microarray data; the Venn Mapper program will only work with $^2\log$ ratio-values. Missing ratio values (empty cells) are allowed.

After selecting the file, the **file indicator** window will expand to show the filename.

In order to perform Venn Mapper on arrays from different sources or formats, the ratio files of those arrays need to be combined first, for example using Microsoft Access or SRS. Only the genes present on all arrays should be joined and the ratios of each unique gene from the different arrays should be within a single row.

Choose the fold difference cut-off.

Most microarray experiments measure dual-probe fluorescent intensities and return the data as ratios. This data is best evaluated by transforming the data into log-values; in case of Venn Mapper a \log_2 transformation.

However, for ease of use, Venn Mapper allows the user to select the fold difference value as a normal value (such as 2-fold difference). Venn Mapper will calculate the $^2\log$ value internally. To get a feel for the $^2\log$ scale see this table.

fold-difference	2 x down	1.7 x down	No change	1.7 x up	2 x up
$^2\log$	-1	-0.765	0	0.765	1

Venn Mapper will only need the fold-difference value (allowed are 0-9 and decimal point).

The **GO** button will only be clickable after a value has been typed in the fold-difference field.

When **GO** has been pressed, the **progress indicator** will inform the user of the progress of the program. When the indicator says **Done!**, Venn Mapper is done (;-) You will need to exit the application to gain full access to the output files.

The Venn Mapper will generate three (tab-delimited) files in the same directory as the input file :

- 1) fact_x.x_numbers.txt
- 2) fact_x.x_zvalue.txt
- 3) fact_x.x_genes.txt

x.x will be the fold difference value selected by the user.

NOTE : If these files already existed, they will be overwritten. Furthermore if (one of) these files were beforehand loaded in another program, *e.g.* Microsoft Excel, Venn Mapper will not be able to write to these files.

ad 1) The number output file will look similar to this :

		Exp A up	Exp A down	Exp B up	Exp B down	Exp C up	Exp C down
		524	386	772	686	524	542
Exp A up	524			294	4	224	
Exp A down	386			2	156		130
Exp B up	772	294	2			336	
Exp B down	686	4	156				292
Exp C up	524	224		336			
Exp C down	542		130		292		

Of the **524** genes up-regulated (by a factor defined in the fold-difference window) in Exp A, **294** are also up-regulated in Exp B (wherein **772** genes in total are up-regulated).

Using the binomial distribution (see reference 1), *z*-values can be calculated of the overlaps found. The *z*-value indicates the number of SD (standard deviations) between the expected and observed values. To get a feel for the values:

<i>z</i> -value	0	0.674	1.96	2.58	3.29
<i>p</i> -value	1	0.5	0.05	0.01	0.001

(two-sided *p*-values)

ad 2) Table of *z*-values.

UniqID	Name	Exp A up	Exp A down	Exp B up	Exp B down	Exp C up	Exp C down
1	Exp A up			4.3	-2.3	4.7	0
2	Exp A down			-1.3	2.2	0	2.1
3	Exp B up	4.3	-1.3			3.7	0
4	Exp B down	-2.3	2.2			0	2.7
5	Exp C up	4.7	0	3.7	0		
6	Exp C down	0	2.1	0	2.7		

This table can be directly imported in Cluster for implementing various methods of clustering and/or TreeView for visualisation of the *z*-profile (reference 2).

Self/self comparisons are set to empty fields. Furthermore, in case the observed value is higher than the expected value, overlaps of exactly 1 gene will not be used for *z*-value calculations, but are also set to an empty field. It is highly improbable that such

an overlap will provide biological relevance. Negative z -values indicate that the observed number of overlapping genes is less than the expected number.

ad 3) Table of the genes found in the overlaps.

Of each overlap in the table from numbers.txt, the genes are listed per field (genes within a field are delimited by a | sign). The gene-ID used in the input table (in example Hs.1, Hs.2 and Hs.3) are used in this table.

The use of the “|” delimiter is based on its frequent use as a logical **OR**, for instance in SRS which makes querying the gene-string easier.

This completes the manual. I hope it has been sufficiently clear to get working with the program. Remarks, suggestions, and program improvements are welcome through e-mail to Marcel Smid: m.smid@erasmusmc.nl

This manual and the Venn Mapper software can be downloaded from :

<http://www.erasmusmc.nl/gatcplatform/>

References :

1) Marcel Smid, Lambert C.J. Dorssers and Guido Jenster (2003). Venn Mapping : clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. *Bioinformatics* **19** (16) 2065-2071.

2) Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**(25): 14863-14868.