

Calculating the rank products using Excel

Imagine a two-color microarray experiment for x genes measured in m replicates. To determine the rank products using Microsoft Excel it is convenient to prepare a spreadsheet that contains three parts (see the example tables, where $x=9$, $m=3$): The first part (columns A and B in the example) contains the gene names, and any annotations, identifiers and additional data that might be associated with each gene. The next part (columns C to E in the example) consists of m columns containing the normalized log-ratios for each gene in the m replicate arrays. The third part (columns F to H in the example) contains the same number of columns as the second part, i.e. one per replicate. Each of these columns just contains the number $1..x$ (the number of genes) in ascending order (these are easily prepared using the auto-fill function of Excel).

Now select columns A...E and sort them by the log-ratios of the first replicate (column C) in descending order (for the calculation of up-regulation rank products). Note that column F now shows the rank of each gene in the first replicate. Next, select columns A...F and sort them again by the log-ratios, this time of the second replicate (column D). After this, column F still contains the ranks for the first replicate, while column G now gives the ranks for the second replicate. Extend the selection to include columns A...G and repeat this procedure for all replicates.

After determining the ranks of each gene in all replicates the rank product (RP -value) can be calculated in an additional column using the product function of Excel, i.e. $\text{PRODUCT}(\text{column } F...H)$ divided by x^m , for each gene. Then sort all the data by increasing RP -value; the most significantly up-regulated genes will be at the top of the list.

To calculate the down-regulation rank products, it is not necessary to repeat the whole sorting procedure, but instead the ranks can just be subtracted from $x+1$ before taking the product.

After the first preparation of the data file, this procedure requires only m sorting steps and one calculation of the products and should not take more than about one minute per replicate when done manually using a standard desktop computer.

Original data table

	A	B	C	D	E	F	G	H
	Unique ID	Name	log-ratio 1	log-ratio 2	log-ratio3	rank 1	rank 2	rank 3
1	123	Enzyme X	2.77	-1.38	2.39	1	1	1
2	234	Protein Y	2.04	2.67	2.04	2	2	2
3	345	Gene Z	-0.55	2.92	1.91	3	3	3
4	456	ORF A	-2.66	2.57	-2.68	4	4	4
5	567	EST B	0.45	0.86	2.05	5	5	5
6	678	Gene Product G	0.06	0.63	-1.87	6	6	6
7	789	Homologue H	1.54	-0.73	1.24	7	7	7
8	876	Prediction P	2.12	1.67	0.18	8	8	8
9	987	Control C	-0.68	0.61	-0.83	9	9	9

Table after first sorting step (sorted by log-ratio 1, descending)

	Unique ID	Name	log-ratio 1	log-ratio 2	log-ratio3	rank 1	rank 2	rank 3
	123	Enzyme X	2.77	-1.38	2.39	1	1	1
	876	Prediction P	2.12	1.67	0.18	2	2	2
	234	Protein Y	2.04	2.67	2.04	3	3	3
	789	Homologue H	1.54	-0.73	1.24	4	4	4
	567	EST B	0.45	0.86	2.05	5	5	5
	678	Gene Product G	0.06	0.63	-1.87	6	6	6
	345	Gene Z	-0.55	2.92	1.91	7	7	7
	987	Control C	-0.68	0.61	-0.83	8	8	8
	456	ORF A	-2.66	2.57	-2.68	9	9	9

Final table, after three sorting steps, to be used for the rank product calculation

	Unique ID	Name	log-ratio 1	log-ratio 2	log-ratio3	rank 1	rank 2	rank 3	rank product
	123	Enzyme X	2.77	-1.38	2.39	1	9	1	=PRODUCT(F1:H1)
	567	EST B	0.45	0.86	2.05	5	5	2	=PRODUCT(F2:H2)
	234	Protein Y	2.04	2.67	2.04	3	2	3	=PRODUCT(F3:H3)
	345	Gene Z	-0.55	2.92	1.91	7	1	4	=PRODUCT(F4:H4)
	789	Homologue H	1.54	-0.73	1.24	4	8	5	=PRODUCT(F5:H5)
	876	Prediction P	2.12	1.67	0.18	2	4	6	=PRODUCT(F6:H6)
	987	Control C	-0.68	0.61	-0.83	8	7	7	=PRODUCT(F7:H7)
	678	Gene Product G	0.06	0.63	-1.87	6	6	8	=PRODUCT(F8:H8)
	456	ORF A	-2.66	2.57	-2.68	9	3	9	=PRODUCT(F9:H9)