

Simulation of Physiological Systems

Computational Biology

Kontopoulos Stefanos

Question 3.1

The INS gene is insulin and is a protein product. From the previous name by which it was characterized, but also based on the source, [OMIM](#), it is related to insulin-dependent type 2 diabetes (INSULIN-DEPENDENT DIABETES MELLITUS 2; IDDM2).

The CLOCK gene is a protein product and is related to circadian rhythms, that is, how cells perceive the passage of time. Disorders in the external environment, ie change of time zone, more hours of light than darkness, affect the gene and therefore the "biological clock". In constant environmental conditions, the rhythms of the biological clock tend to have a period of 24 hours.

Question 3.2

The CLOCK gene belongs to chromosome 4q12 and the helix-loop-helix protein structural form, which characterizes one of the largest families of bilateral transcription factors.

Question 3.3

Figure [1](#).

Question 4.1

The word '*atc*' in this genome occurs 21 times. The matlab code is in the ex6.m file and this function is in *find_str.m*.

Question 5.1

For this experiment, measurements were made in 31 samples, which were divided into 2 conditions. In one the 12 samples did a low-calorie diet consuming 5MJ (approximately 1200 calories) (diet intervention group) and in the other the remaining 19 samples continued their diet normally, without losing weight (control group).

Question 5.2

i)

The title and corresponding ID of the samples are shown in the following table 1:

Title	Sample_geo_accession
Liver 1	GSM162954
Liver 2	GSM162956
Liver 3	GSM162957
Liver 9	GSM162958
Liver 12	GSM162959
Liver 25	GSM162960
Liver 26	GSM162961
MHLiv	GSM162962

Table 1: Samples.

ii)

The samples are from a Homo sapiens organism, according to line 43: *Sample_organism_ch1*.

iii)

The sample correspondence and experiment status are shown in the table below 2:

iv)

Read.table () is written as follows: read.table (file.path ("GSE7117_series_matrix.txt"), skip = 68, header = TRUE, sep = " t", row.names = 1), where file.path ("") specifies the name

Sample_geo_accession	Group
GSM162954	control
GSM162956	diet intervention
GSM162957	diet intervention
GSM162958	control
GSM162959	diet intervention
GSM162960	control
GSM162961	diet intervention
GSM162962	control

Table 2: Groups.

of the file to be read and passed as a table, `skip = 68` (integer) instructs to ignore the first 68 lines of the file, `header = TRUE` (logical) specifies that the file contains the values of the variables in its first row, `sep = "t"` sets as the field change character in the table to be TAB and finally, `row.names = 1` defines the first column of the table that contains the names of the series.

v)

The dimensions of table `x` are 54678 rows on 8 columns.

vi)

The different columns are assigned to `Sample_geo_accession` id.

vii)

The value of `x[3.5] = 42.91798` represents the concentration of RNA extracted from the liver.

ix)

With the command `x[200,]` we see the levels of the `1552540_s_at` genome in all the samples. With the command `x[, 7]` we see the levels of all genomes for the 7th sample GSM162961.

Question 5.3

The round (apply (x, 2, summary), digits = 2) command rounds the two decimal places in the x table after applying a margin to the second line.

Question 5.5

i)

The table 3 shows the results obtained for the most important LUM and SH3YL1 genes. P.Value values of both genes are acceptable according to the theory, while the logFC of SH3YL1 is much lower than the theoretical value ($|\logFC| > 1.5$) and therefore is not acceptable.

ID	adj.P.Value	P.Value	t	B	logFC	Gene.symbol	Gene.title
201744_s_at	0.4	0.00000731	9.4	0.8482	1.751	LUM	lumican
204019_s_at	0.544	0.00002013	8.27	0.5363	0.883	SH3YL1	SH3 and SYLF domain containing 1

Table 3: Genes results.

The figure 2 for the first gene is shown below and shows the levels of expression of this gene in all samples. The lumican gene encodes a member of the SLRP (small leucine-rich proteoglycan) family. In its bifunctional molecules, the protein entity binds collagen fibres and highly charged hydrophilic glycosaminoglycans regulate endoviral distances. The lumican is also the major corneal keratin proteoglycan, but is distributed throughout the body in collagen matrices and may regulate collagen fibril organization, peripheral growth, corneal transparency, epithelial migration, and epithelial regurgitation.

Question 5.6

i)

The thermal map consists of 54675 lines and 8 columns.

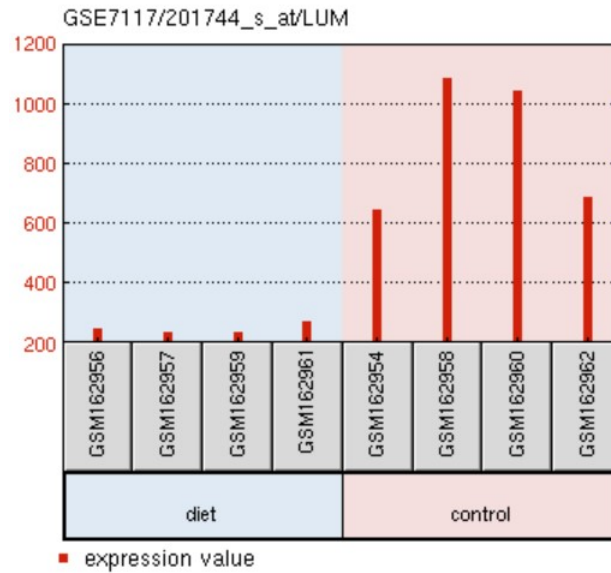


Figure 2: Expression levels.

ii)

The following detail of the map 3 shows the level of appearance of each gene in the 8 samples of the experiment. There is a colour gradient that defines pink as the maximum expression and green as the minimum.

iii)

The expression value of the PTPN14 gene for sample GSM162954 is moderate.

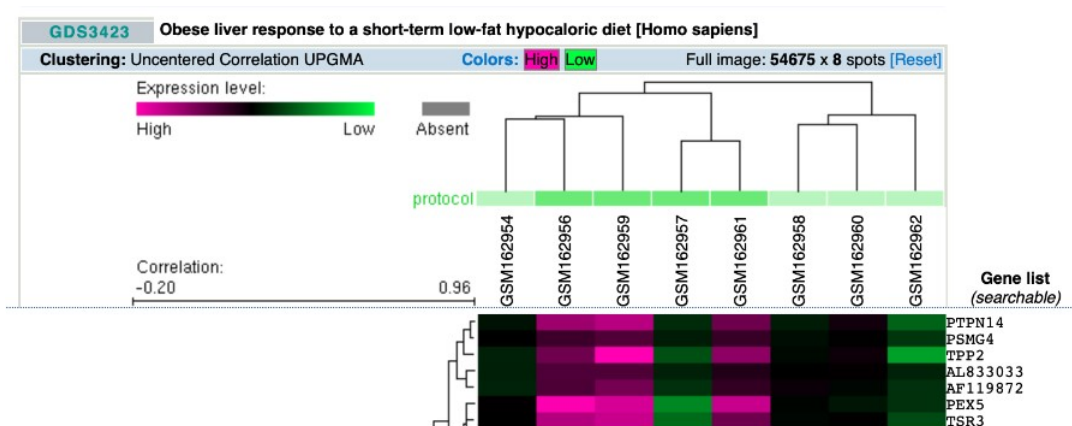


Figure 3: Obese liver response.