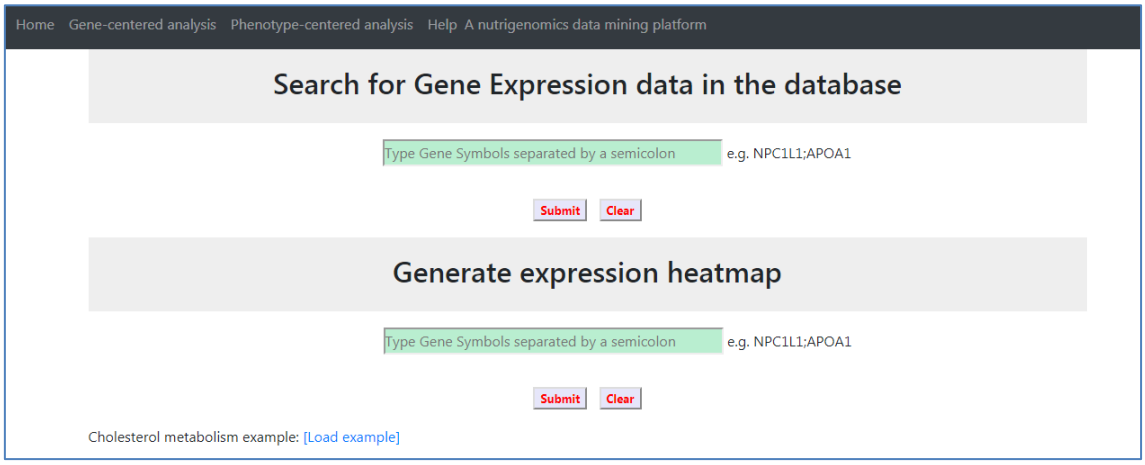
NutriGenomeDB User's Manual

Gene-centered analysis module:

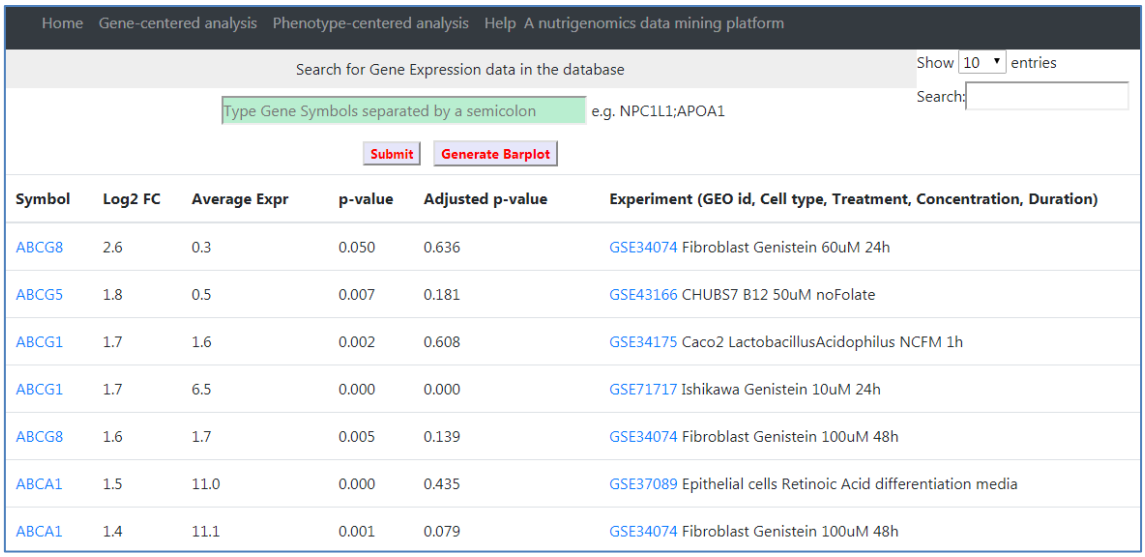
This module allows the exploration of differentially expressed genes accross nutrigenomics experiments. An exhaustive search of microarray gene expression assays on human cells in response to nutrients and bioactive food compounds was performed on GEO database. All the experiments included in NutriGenomeDB are defined by a list of the top 10% differentially expressed genes, sorted by adjusted p-value. NutriGenome DB contains data generated among 32 different microarray platforms. As those platforms differ widely in characteristics such as the number of targets and available annotations, the number of genes defining each experiment may vary between 1500 and 3000. Two different types of queries can be launched from the main screen.



Main user interface of the Gene-centered analysis module: An example data input can be automatically loaded by clicking on Load Example. Those data correspond to genes implied in cholesterol metabolism.

Search for Gene Expression data in the database:

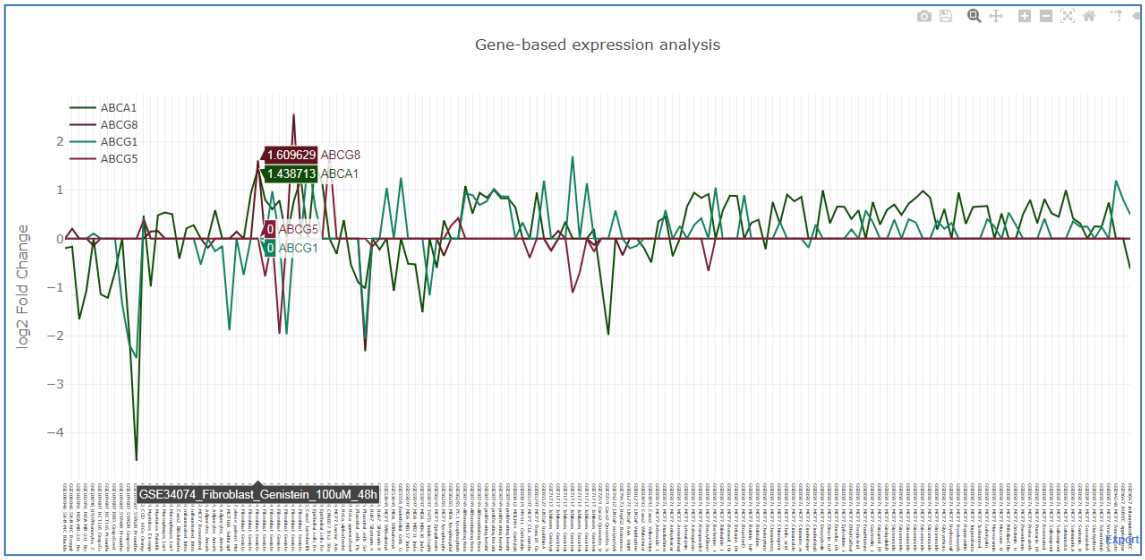
Users can explore whether a set of genes appear as differentially expressed among our collection of nutrigenomics experiments by introducing in the search box one or multiple human Gene Symbols. It is a exact match search, please pay attention to **spelling and capitalization.** Results are presented as a sortable data table with log2 fold changes expression values relative to the control, average expression levels, statistics and information about the target experiment, which is directly linked to the data source in GEO database.



Symbol	Log2 FC	Average Expr	p-value	Adjusted p-value	Experiment (GEO id, Cell type, Treatment, Concentration, Duration)
ABCG8	2.6	0.3	0.050	0.636	GSE34074 Fibroblast Genistein 60uM 24h
ABCG5	1.8	0.5	0.007	0.181	GSE43166 CHUBS7 B12 50uM noFolate
ABCG1	1.7	1.6	0.002	0.608	GSE34175 Caco2 LactobacillusAcidophilus NCFM 1h
ABCG1	1.7	6.5	0.000	0.000	GSE71717 Ishikawa Genistein 10uM 24h
ABCG8	1.6	1.7	0.005	0.139	GSE34074 Fibroblast Genistein 100uM 48h
ABCA1	1.5	11.0	0.000	0.435	GSE37089 Epithelial cells Retinoic Acid differentiation media
ABCA1	1.4	11.1	0.001	0.079	GSE34074 Fibroblast Genistein 100uM 48h

Search results table: The introduced genes are described by their expression level, statistical information and the source experiment. The experiment column contains the corresponding GEO id (link for further examination) as well as a brief summary; human cell type, compound and treatment duration. Each column is sortable, and the Search box on the top right allows to look up for a specific gene or compound. Number of shown entries can also be tuned.

A visual inspection of the returned differential expression values can be performed by clicking on the Generate plot button, available at the top center of the table.

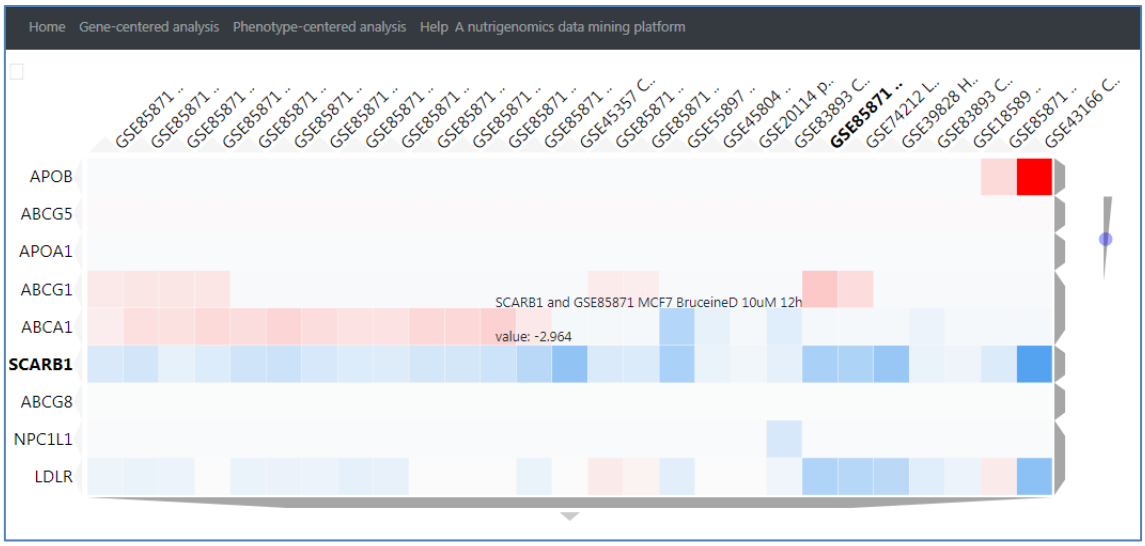


Line chart visualization: From the previously introduced genes, a visual representation of the differential expression level accross nutrigenomics experiments can be obtained after clicking on "Generate plot" button. The obtained visualization is completely interactive. Expression values of the introduced genes, and the corresponding experiment are highlighted when the mouse cursor is placed above them. This example allows to easily identify the **co-expression trend of the ATP binding cassette subfamily genes**. The visualization can be downloaded as an image in png format.

Generate expression heatmap:

Differential expression levels of a set of genes accross nutrigenomics experiments can also be visualized as an interactive hierarchically clustered heatmap. Depending on the introduced set of genes, such a visualization is useful to identify clusters of nutrients or food compounds which trigger a potentially interesting gene signature. Differential expression data is shown as row normalized z-score values. This visualization makes use of Clustergrammer tool (Fernandez, N. F. et al. **Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data**. Sci. Data 4:170151 doi: 10.1038/sdata.2017).

Clustered heatmap visualization: Each cell shows information about the corresponding gene, experiment and row z-score normalized differential expression level. Number of columm clusters (experiments) and row clusters (genes) can be tuned based on their similarity. **Only the experiments where at least one of the introduced genes was differentially expressed are shown**. Clusters of experiments and genes can be zoomed for further inspection by clicking on the corresponding arrow.

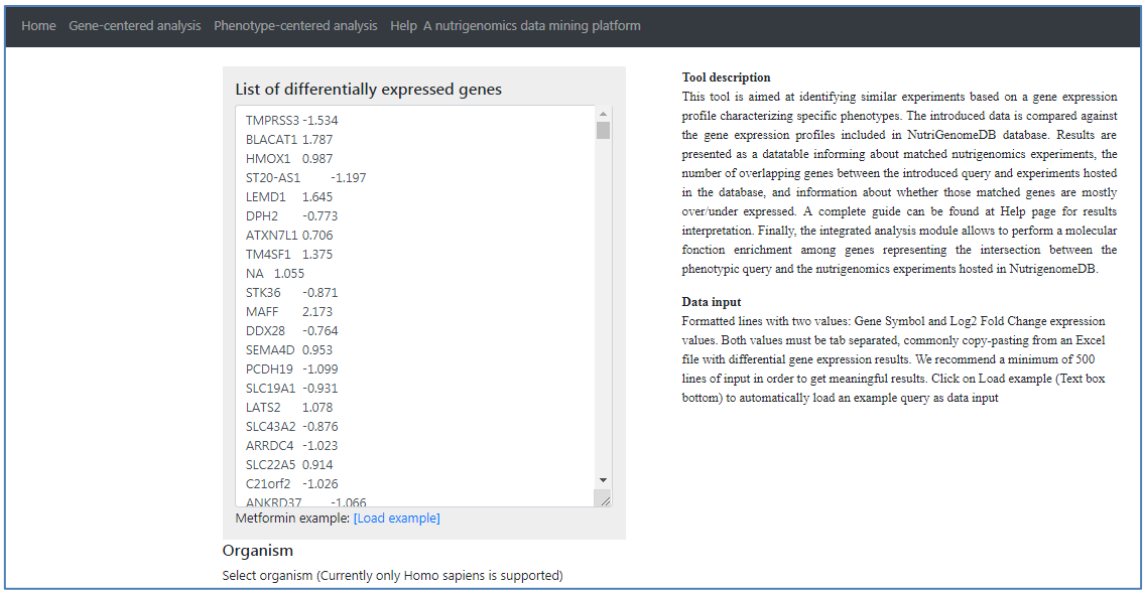


Zoom visualization of an identified experimental cluster: Only the set of experiments identified within the clustered are shown. Values of the expressed genes can be easily inspected.

Phenotype-centered analysis module:

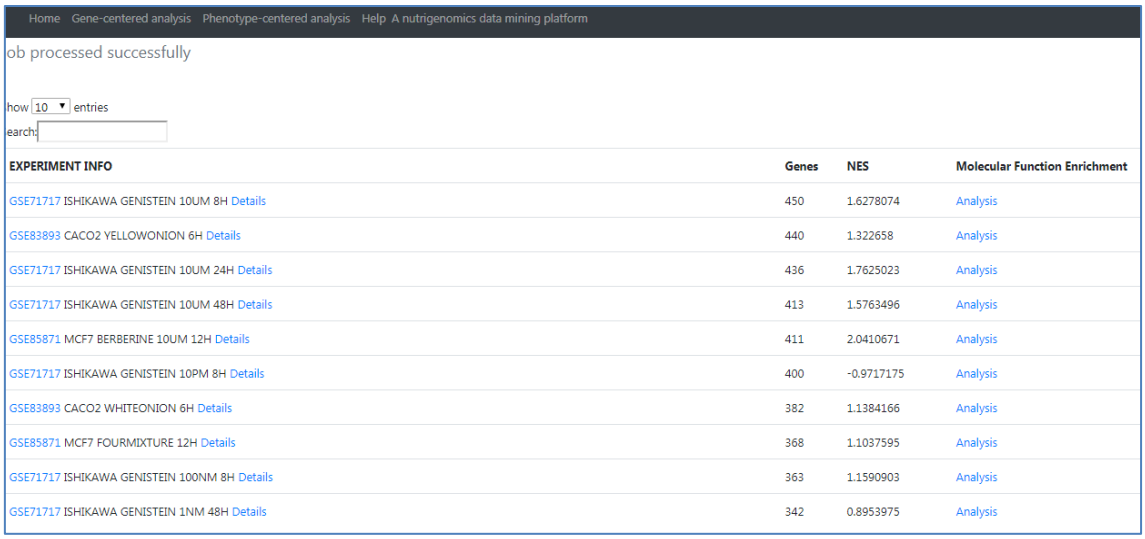
This module allows users to find functional connections between their phenotype of interest and experiments included in NutriGenomeDB by comparing gene expression profiles. NutriGenomeDB (v.1.1) contains 231 gene expression profiles from human cells exposed to more than 140 different nutrients and bioactive food compounds.

This analysis module is built upon the widely used pattern-matching algorithm Gene Set Enrichment Analysis (GSEA). The Nutrigenomic Gene Sets included in NutriGenomeDB are defined as the top 10% differentially expressed genes in the included nutrigenomics experiments (sorted by adjusted p-value), which are mostly statistically significant. Therefore this tools allows users to find trustable functional connections between a list of differentially expressed genes resulting from a phenotype of interest, with the Nutrigenomic Gene Sets defining different treatments with nutrients and bioactive food compounds available in NutriGenomeDB.



Main user interface of the Phenotype-centered analysis module: An example data input can be automatically loaded by clicking on Load Example: Those data correspond to a gene signature after treatment of MCF7 cells with the antidiabetic drug Metformin (GSE36847).

Required data input: this module expects as input a list of human Gene Symbols and their differential expression values (in log2 scale) **separated by tabulator**. For numeric data, the **decimal separator must be a dot (.)**. This is the typical output from a gene differential expression experiment; from an Excel file, the corresponding Gene symbol and logFC columns should be copy-pasted directly into the application text box. Example data can be loaded by clicking at the bottom of the text box (Load Example). The application automatically handles missing values (NA) and duplicated gene symbols in the data input. We recommend to perform queries with a minimum number of 500 genes and their fold change values in order to obtain meaningful results. An example output result can be found here.



EXPERIMENT INFO	Genes	NES	Molecular Function Enrichment
GSE71717 ISHIKAWA GENISTEIN 10UM 8H Details	450	1.6278074	Analysis
GSE83893 CACO2 YELLOWONION 6H Details	440	1.322658	Analysis
GSE71717 ISHIKAWA GENISTEIN 10UM 24H Details	436	1.7625023	Analysis
GSE71717 ISHIKAWA GENISTEIN 10UM 48H Details	413	1.5763496	Analysis
GSE85871 MCF7 BERBERINE 10UM 12H Details	411	2.0410671	Analysis
GSE71717 ISHIKAWA GENISTEIN 10PM 8H Details	400	-0.9717175	Analysis
GSE83893 CACO2 WHITEONION 6H Details	382	1.1384166	Analysis
GSE85871 MCF7 FOURMIXTURE 12H Details	368	1.1037595	Analysis
GSE71717 ISHIKAWA GENISTEIN 100NM 8H Details	363	1.1590903	Analysis
GSE71717 ISHIKAWA GENISTEIN 1NM 48H Details	342	0.8953975	Analysis

Analysis results table (example output): The resulting experiments are defined by their GEO ID (link for further examination) as well as a brief description; human cell type, compound and treatment information (compound concentration and duration). **By default, the output result is sorted by the number of overlapping genes**. It highlights that a treatment with 10uM of Genistein for 8h on Ishikawa cells triggers a gene signature closely related to the one produced by a treament with metformin. Indeed the scientific literature provides strong evidences of the antidiabetic properties of Genistein.

By clicking on "Details" link, users can download the list of overlapping genes between the query and matching experiments as well as further experimental information. Each column is sortable by Genes (number of overlapping genes), NES (Normalized Enrichment score; a high NES means that among the overlapping genes, a high portion of those are highly overexpressed in the query and the matching experiment, viceversa for a negative NES). Search box on the top left allows users to browse for specific compounds or GEO ID's among the results.

	A	B	C	D	E	F	G	H
1	Symbol	logFC Query	logFC	AveExpr	t	P.Value	adj.P.Val	B
2	PDGFC	1.147	-0.72321138	8.76963641	-5.05830995	0.00040748	0.006931	0.22724342
3	ADRB1	1.004	0.67988642	6.41738485	5.33429336	0.00020752	0.0044617	0.69084808
4	FAM46B	0.944	-0.51189798	7.06944209	-5.32090967	0.00021195	0.0045311	0.66921893
5	SLC22A5	0.914	-0.50477989	8.30506507	-4.83348166	0.00046522	0.0079134	-0.13550002
6	LARP6	0.91	1.2530745	8.08806732	12.4443478	5.18E-08	2.34E-05	9.01042649
7	SAT1	1.014	0.86662302	10.6936816	8.43410438	1.11E-05	0.00049235	4.92013652
8	SAT1	0.888	0.86662302	10.6936816	8.43410438	1.11E-05	0.00049235	4.92013652
9	JADE2	0.844	-1.00546947	7.41601081	-5.82954276	9.67E-05	0.0026352	1.47314144
10	TUFT1	0.836	0.99324427	9.3380301	9.59243484	8.03E-07	0.00011114	6.33229477
11	YY1	0.828	-0.58590917	9.54434227	-5.47181752	0.00080713	0.01013064	0.78216123
12	YPEL2	0.808	0.8012732	7.13554402	4.9367681	0.00039274	0.00703573	0.03773036
13	PHLPP1	0.794	-0.78464931	7.22186926	-4.76830537	0.00051809	0.00854492	-0.24553661
14	LATS2	1.078	1.02420495	6.15164959	8.9856197	1.57E-06	0.00016953	5.66483532
15	LATS2	0.792	1.02420495	6.15164959	8.9856197	1.57E-06	0.00016953	5.66483532
16	ERBB4	0.769	-0.88160535	8.45537804	-8.06479652	4.63E-06	0.00034311	4.57376605
17	SALL2	0.761	-0.42846221	8.95162794	-4.34081097	0.0010645	0.01445638	-0.98037999
18	ABCC5	0.753	0.89014961	6.92736073	5.51428156	0.00015661	0.00366078	0.9792412
19	DDIT4	0.738	1.16952241	11.3917062	8.63937121	2.33E-06	0.00022069	5.26606693
20	TMCC1	0.735	0.15981408	7.72664864	-0.27660509	0.00031516	0.00589334	0.36084582
21	HOTAIRM1	0.732	0.65696824	5.72751982	4.92045017	0.0007226	0.01000291	-0.02207353
22	WDR47	0.727	1.03882578	8.27691718	7.66553674	7.61E-06	0.00047304	4.06903399
23	CDKN1A	0.717	0.63167593	9.2220775	4.94454401	0.00038779	0.00697902	0.0507143

Example details of a matched experiment: The downloaded list of overlapping genes between the query and matching experiments shows details about the expression level of the introduced gene signature (logFC Query column) and differential expression information of the matched nutrigenomic experiment. This example corresponds to overlapping genes of the treatment with 10uM of Genistein for 8h on Ishikawa cells.

By clicking on the Analysis link, users are allowed to perform a functional enrichment analysis of the list of overlapping genes that connect the gene expression query and matched experiments from NutriGenomeDB. Via a java web-service, Panther database is interrogated for overrepresented molecular functions that characterize those overlapping genes. In this way users can get further insights into the functional connections highlighted by this analysis tool at the molecular level.

A	B	C	D	E
Id	Name	GeneId	P-value	FDR
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	YY1	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	NFIL3	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	MAX	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	FOXK2	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	E2F2	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	HES1	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	PAXBP1	1.75E-04	0.08932599
GO:0001078	transcriptional repressor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	IRF2BPL	1.75E-04	0.08932599
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	YY1	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	NFIL3	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	MAX	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	FOXK2	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	E2F2	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	HES1	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	PAXBP1	4.80E-04	0.12273896
GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	IRF2BPL	4.80E-04	0.12273896
GO:0061733	peptide-lysine-N-acetyltransferase activity	ING3	5.72E-03	0.19479923
GO:0061733	peptide-lysine-N-acetyltransferase activity	EPC1	5.72E-03	0.19479923
GO:0061733	peptide-lysine-N-acetyltransferase activity	RSF1	5.72E-03	0.19479923
GO:0061733	peptide-lysine-N-acetyltransferase activity	KANSL1L	5.72E-03	0.19479923
GO:0003700	DNA-binding transcription factor activity	PHTF2	6.35E-03	0.20288372
GO:0003700	DNA-binding transcription factor activity	MAX	6.35E-03	0.20288372

Example of a functional enrichment output: The downloaded file informs about the overrpesented molecular functions among the identified overlapping genes between the query and matched experiment. This example corresponds to overlapping genes of the treatment with 10uM of Genistein for 8h on Ishikawa cells. Genes implied in a transcriptional repressor activity are statistically significant under a $p<0.1$ threshold (lowest False Discovery Rate).

Submitting new data to NutriGenomeDB:

The aim of NutriGenomeDB is to become a central hub of information related to nutrigenomics, by expanding its available gene expression data in order to create an useful and growing ressource for the cientific community. Therefore we encourage researchers to submit their data related to nutrigenomics experiments performed on human cell for inclusion in NutriGenomeDB. Please write us an email for further enquiries.