Identify putative gene signatures distinguishing wild type and susceptible mice exposed to an environmental stressor.

Objective: Explore a Study of the effects of hyperoxia on mice with or without a disruption in the Nrf2 gene. The Nrf2 gene product contributes to protection against oxidative stress.

Detailed work flow:
Go to the CEBS home page (www.cebs.niehs.nih.gov/prototype/)
Select the “Display All Studies” and click on the Investigation “Hyperoxia_Mouse_2004. Expand this investigation by clicking on it and select the associated study which has the same title.

Under the “Design Information” column located towards the right, select “Details” to view the objective and background of the Study. Select “Design” to view the experimental groups. This reveals a parallel study of mice carrying a disrupted Nrf2 gene (“Mut”) and mice without the disruption. This information is available by clicking on a group name and viewing the information for individual mice in the group.

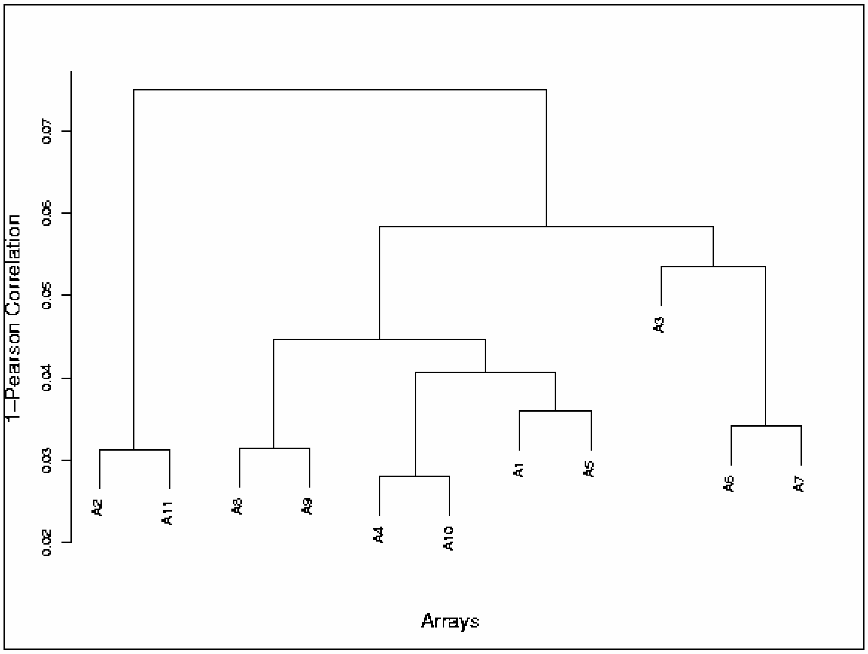
Return back to the “Display All Studies” page (click the back button twice). Select “Timeline” to view the schedule of events and the protocols applied to the animals. Clicking on “protocol” link on the treatment bar brings up the details of both the genetic stressor treatment and the environmental stressor treatment. Navigating between “Design” and “Timeline” permits the user to understand the experiment from the depositor’s viewpoint. In this case the depositor intended that each arm, the knockout and the wild type, would be compared to an untreated control group of genetically similar animals.

This example illustrates that it is also possible to use CEBS to carry out an analysis with a different design than that intended by the depositor.

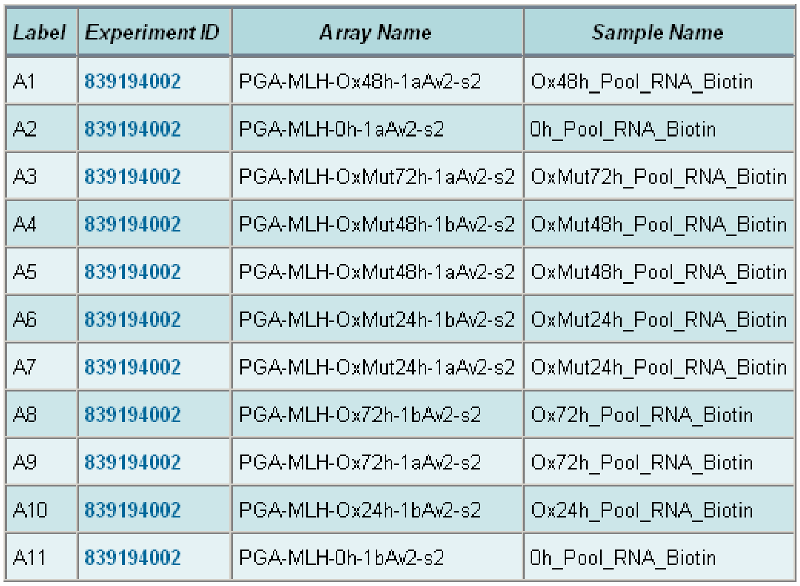
Return back to the “Display All Studies” page. Tick the check box next to the Study, and scroll down to the bottom of the page and select “Experiments”. This brings the user to the “CEBS Experiment Selection” page. Tick the box by the experiment and select “View Details about Selected Experiment(s)”. Scrolling down to the detailed view at the bottom, the user can view details of the experiment, download the data, and also inspect the QC data for the arrays and download the original data files. <NB: the QC data feature is not active in CEBS 2.0β; QC data can be retrieved from CEBS 1.6 http://cebs.niehs.nih.gov/>.

Tick the experiment check box and select “View arrays from the selected experiments”. Once the list of arrays returns, select “Retrieve data” at the bottom of the page to go to the “Data Preprocessing Options” page. You can accept the manufacturer-recommended preprocessing steps or a method of your choice. Click “Continue”.

At the “CEBS Microarray Data Analysis Options” page select “Visualize Preprocessed Data” and view the dendrogram:



The sample names (from the legend at the bottom of the page) are given below:



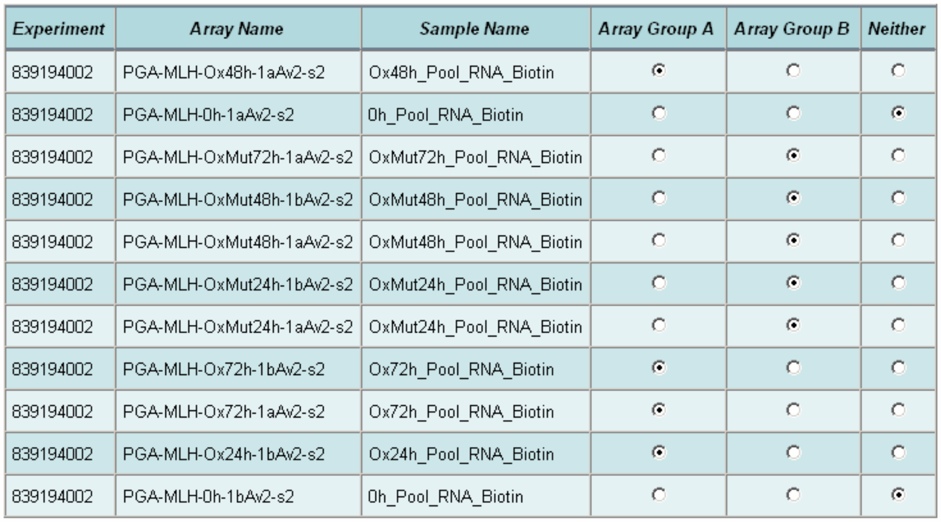
Label	Experiment ID	Array Name	Sample Name
A1	839194002	PGA-MLH-Ox48h-1aAv2-s2	Ox48h_Pool_RNA_Biotin
A2	839194002	PGA-MLH-0h-1aAv2-s2	0h_Pool_RNA_Biotin
A3	839194002	PGA-MLH-OxMut72h-1aAv2-s2	OxMut72h_Pool_RNA_Biotin
A4	839194002	PGA-MLH-OxMut48h-1bAv2-s2	OxMut48h_Pool_RNA_Biotin
A5	839194002	PGA-MLH-OxMut48h-1aAv2-s2	OxMut48h_Pool_RNA_Biotin
A6	839194002	PGA-MLH-OxMut24h-1bAv2-s2	OxMut24h_Pool_RNA_Biotin
A7	839194002	PGA-MLH-OxMut24h-1aAv2-s2	OxMut24h_Pool_RNA_Biotin
A8	839194002	PGA-MLH-Ox72h-1bAv2-s2	Ox72h_Pool_RNA_Biotin
A9	839194002	PGA-MLH-Ox72h-1aAv2-s2	Ox72h_Pool_RNA_Biotin
A10	839194002	PGA-MLH-Ox24h-1bAv2-s2	Ox24h_Pool_RNA_Biotin
A11	839194002	PGA-MLH-0h-1bAv2-s2	0h_Pool_RNA_Biotin

Study of the dendrogram shows that the room air controls (A2 and A11) are clearly distinct from the animals exposed to hyperoxia; there is also a difference in disrupted (A3, A6, A7) animals compared to the animals with wild type Nrf2.

Select “Continue with analysis” and then select “Comparison of Two groups of Arrays” to compare two groups of arrays. Click “Continue”.

Recall that the design of the experiment was to compare the time course samples of disrupted strain to the 0 hour, room air control, and similarly for the wild type Nrf2 strain. This is not possible in CEBS. At the moment CEBS supports pair-wise comparisons of groups containing at least two arrays. Given that the hierarchical clustering suggests that animals with Nrf2-disruptions respond differently to hyperoxia than animals with wild type Nrf2, we can use CEBS to identify genes with differential levels of expression between these two groups.

Set up two groups: Group A with 4 arrays from exposed mice with wild type Nrf2 and Group B with 5 arrays from exposed mice with disrupted Nrf2. Do not include the “room air” controls. You should see this:

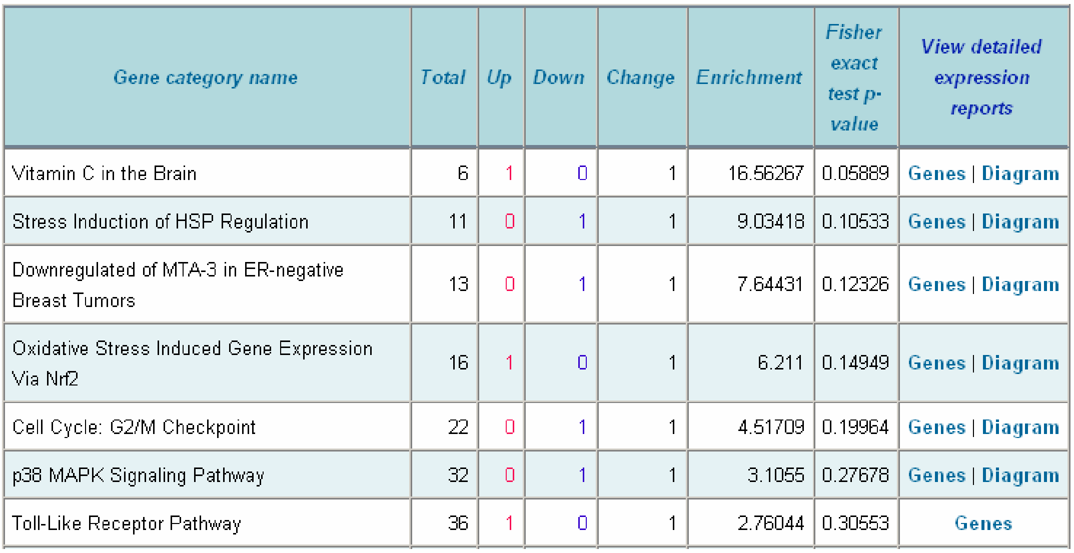


Experiment	Array Name	Sample Name	Array Group A	Array Group B	Neither
839194002	PGA-MLH-Ox48h-1aAv2-s2	Ox48h_Pool_RNA_Biotin	●	○	○
839194002	PGA-MLH-0h-1aAv2-s2	0h_Pool_RNA_Biotin	○	○	●
839194002	PGA-MLH-OxMut72h-1aAv2-s2	OxMut72h_Pool_RNA_Biotin	○	●	○
839194002	PGA-MLH-OxMut48h-1bAv2-s2	OxMut48h_Pool_RNA_Biotin	○	●	○
839194002	PGA-MLH-OxMut48h-1aAv2-s2	OxMut48h_Pool_RNA_Biotin	○	●	○
839194002	PGA-MLH-OxMut24h-1bAv2-s2	OxMut24h_Pool_RNA_Biotin	○	●	○
839194002	PGA-MLH-OxMut24h-1aAv2-s2	OxMut24h_Pool_RNA_Biotin	○	●	○
839194002	PGA-MLH-Ox72h-1bAv2-s2	Ox72h_Pool_RNA_Biotin	●	○	○
839194002	PGA-MLH-Ox72h-1aAv2-s2	Ox72h_Pool_RNA_Biotin	●	○	○
839194002	PGA-MLH-Ox24h-1bAv2-s2	Ox24h_Pool_RNA_Biotin	●	○	○
839194002	PGA-MLH-0h-1bAv2-s2	0h_Pool_RNA_Biotin	○	○	●

Clicking “Continue” brings up the “Define Criteria for Differentially Expressed Gene(s)”. Click “Continue” to apply these thresholds. Setting the thresholds at the default values (2-fold minimum change and $p < 0.05$) returns 125 genes with altered expression between the two groups.

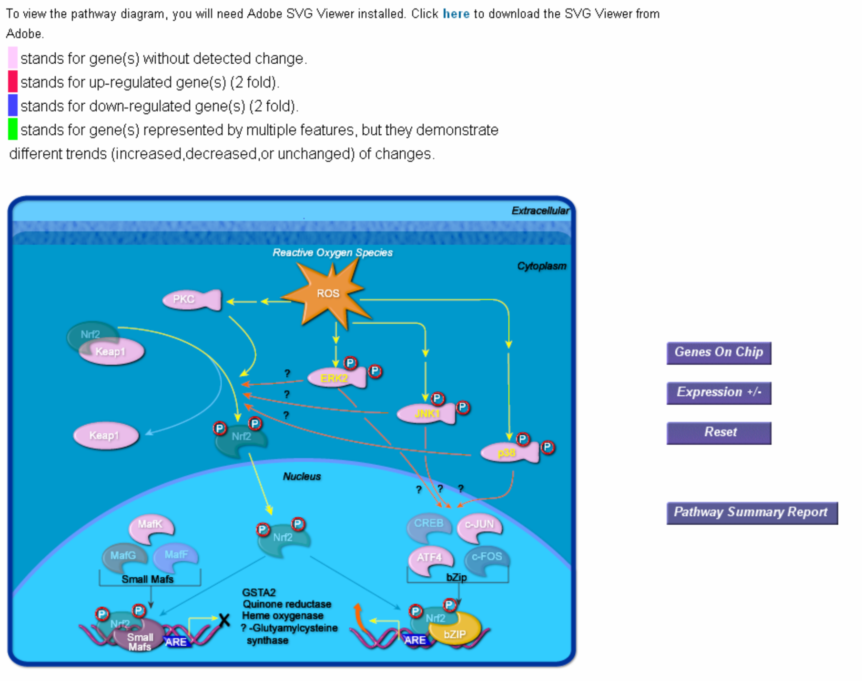
The annotations for individual genes can be retrieved from the “Expression Report” by clicking “View Expression Report” and then clicking on the name of genes of interest. This list can be sorted by clicking on any column header.

Return to the “Biological Analysis of Gene Expression Data” page via the back button. Selecting “Perform Gene Category Analysis by BioCarta Pathways” brings up the BioCarta Gene Categories page; using the enrichment to sort the pathways (click on the “Enrichment” column header) identifies the following pathways.



Gene category name	*Total*	*Up*	*Down*	*Change*	*Enrichment*	*Fisher exact test p-value*	*View detailed expression reports*
Vitamin C in the Brain	6	1	0	1	16.56267	0.05889	Genes | Diagram
Stress Induction of HSP Regulation	11	0	1	1	9.03418	0.10533	Genes | Diagram
Downregulated of MTA-3 in ER-negative Breast Tumors	13	0	1	1	7.64431	0.12326	Genes | Diagram
Oxidative Stress Induced Gene Expression Via Nrf2	16	1	0	1	6.211	0.14949	Genes | Diagram
Cell Cycle: G2/M Checkpoint	22	0	1	1	4.51709	0.19964	Genes | Diagram
p38 MAPK Signaling Pathway	32	0	1	1	3.1055	0.27678	Genes | Diagram
Toll-Like Receptor Pathway	36	1	0	1	2.76044	0.30553	Genes

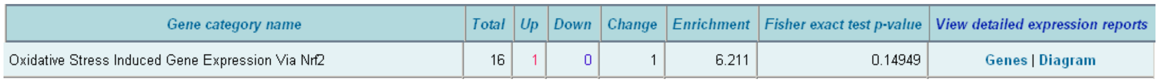
Clicking on the “Diagram” link to the right of the pathway “Oxidative Stress Induced Gene Expression Via Nrf2” brings up a pathway diagram. Recall that the Nrf2 gene was knocked out in half the subjects. You can overlay the expression results using the “Expression +/-“ link to the right of the diagram.



To view the pathway diagram, you will need Adobe SVG Viewer installed. Click here to download the SVG Viewer from Adobe.

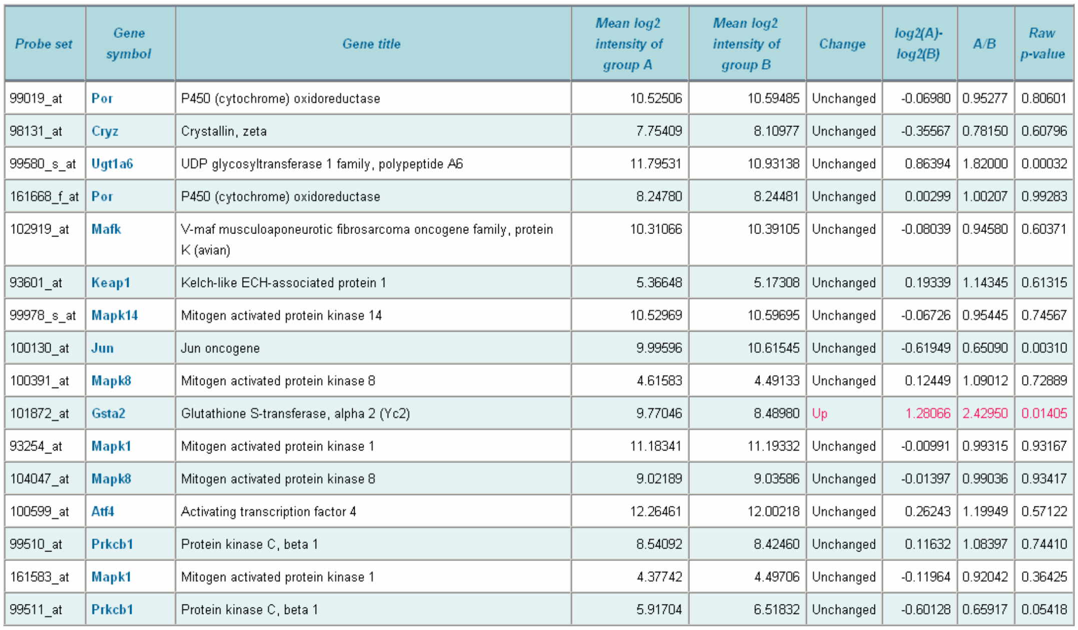
- stands for gene(s) without detected change.
- stands for up-regulated gene(s) (2 fold).
- stands for down-regulated gene(s) (2 fold).
- stands for gene(s) represented by multiple features, but they demonstrate different trends (increased,decreased,or unchanged) of changes.

Note that the original pathway summary showed one gene with increased expression:



Gene category name	Total	Up	Down	Change	Enrichment	Fisher exact test p-value	View detailed expression reports
Oxidative Stress Induced Gene Expression Via Nrf2	16	1	0	1	6.211	0.14949	Genes | Diagram

However there are no “red” gene symbols on the diagram. Returning to the “Annotated BioCarta Gene Categories” page, select the “Genes” link to the right of the “Oxidative Stress Induced Gene Expression Via Nrf2” pathway. This brings up the list of genes associated in BioCarta with this pathway. The list is updated more frequently than the diagram is. In this case, the gene with increased expression (Gsta2) is not included in the diagram.



Probe set	Gene symbol	Gene title	Mean log2 intensity of group A	Mean log2 intensity of group B	Change	log2(A)-log2(B)	A/B	Raw p-value
99019_at	Por	P450 (cytochrome) oxidoreductase	10.52506	10.59485	Unchanged	-0.06980	0.95277	0.80601
98131_at	Cryz	Crystallin, zeta	7.75409	8.10977	Unchanged	-0.35567	0.78150	0.60796
99580_s_at	Ugt1a6	UDP glycosyltransferase 1 family, polypeptide A6	11.79531	10.93138	Unchanged	0.86394	1.82000	0.00032
161668_f_at	Por	P450 (cytochrome) oxidoreductase	8.24780	8.24481	Unchanged	0.00299	1.00207	0.99283
102919_at	Mafk	V-maf musculoaponeurotic fibrosarcoma oncogene family, protein K (avian)	10.31066	10.39105	Unchanged	-0.08039	0.94580	0.60371
93601_at	Keap1	Kelch-like ECH-associated protein 1	5.36648	5.17308	Unchanged	0.19339	1.14345	0.61315
99978_s_at	Mapk14	Mitogen activated protein kinase 14	10.52969	10.59695	Unchanged	-0.06726	0.95445	0.74567
100130_at	Jun	Jun oncogene	9.99596	10.61545	Unchanged	-0.61949	0.65090	0.00310
100391_at	Mapk8	Mitogen activated protein kinase 8	4.61583	4.49133	Unchanged	0.12449	1.09012	0.72889
101872_at	Gsta2	Glutathione S-transferase, alpha 2 (Yc2)	9.77046	8.48980	Up	1.28066	2.42950	0.01405
93254_at	Mapk1	Mitogen activated protein kinase 1	11.18341	11.19332	Unchanged	-0.00991	0.99315	0.93167
104047_at	Mapk8	Mitogen activated protein kinase 8	9.02189	9.03586	Unchanged	-0.01397	0.99036	0.93417
100599_at	Atf4	Activating transcription factor 4	12.26461	12.00218	Unchanged	0.26243	1.19949	0.57122
99510_at	Prkcb1	Protein kinase C, beta 1	8.54092	8.42460	Unchanged	0.11632	1.08397	0.74410
161583_at	Mapk1	Mitogen activated protein kinase 1	4.37742	4.49706	Unchanged	-0.11964	0.92042	0.36425
99511_at	Prkcb1	Protein kinase C, beta 1	5.91704	6.51832	Unchanged	-0.60128	0.65917	0.05418

In this case study the depositor suggested one plan for the microarray analysis, while an alternative analysis workflow was used here, leading to different but biologically sensible gene patterns being detected.