

Report on Gene Expression Analysis for Customer Samples

Sample processing and Quality Control

Samples were received on 16 February 2007. Initial sample QC performed and summarized below. Sample concentration was determined by UV absorption. The 260/280 and 260/230 ratios were calculated to assess protein and organic solvent contaminations respectively.

SAMPLE	DATE	ng/ μ l	260/280	260/230
<i>CUST1</i>	2/16/2007	398.55	2.04	2.06
<i>CUST2</i>	2/16/2007	395.69	2.04	2.06
<i>CUST3</i>	2/16/2007	450.64	2.04	2.02
<i>CUST4</i>	2/16/2007	333.63	1.95	1.95
<i>CUST5</i>	2/16/2007	327.36	1.94	1.94
<i>CUST6</i>	2/16/2007	497.97	2.14	2.14

Additional sample QC was performed after the IVT reaction and yield determined. 1.0 μ gs of labeled cRNA were hybridized to Affymetrix U133plus2 chips (lot#: 1234567). Chips were scanned on 16 February 2007. The file names of the resulting .DAT, .CEL and .CHP files are listed in the table below for each sample. All Affymetrix data files are included on the enclosed CD or DVD.

SAMPLE	Affy - .DAT, .CEL and .CHP
<i>CUST1</i>	07-0010-H
<i>CUST2</i>	07-0011-H
<i>CUST3</i>	07-0012-H
<i>CUST4</i>	07-0013-H
<i>CUST5</i>	07-0014-H
<i>CUST6</i>	07-0015-H

Global Chip Parameters

The Affymetrix Expression Report (attachment A) indicated the scaling factors, average signal and average background all consistent with high quality comparable

data. The 3'/5' ratios for the “housekeeping” genes B-Actin and GAPDH were close to 1.0 for all chips consistent with good RNA quality.

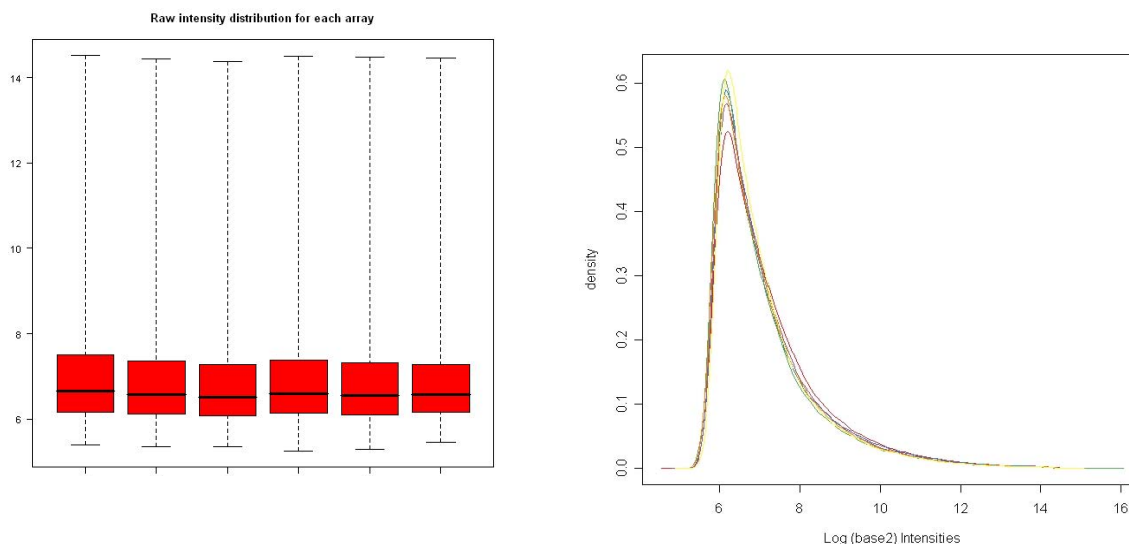
Affymetrix Global Chip Parameters

SAMPLE	CUST1	CUST2	CUST3	CUST4	CUST5	CUST6
Scale Factor (to TGT of 100)*	1.099450	1.156290	1.393676	1.302371	1.335922	1.392445
Avg. Background	49.11904	49.22829	47.92639	49.42028	48.12601	50.93617
% Present	40.14266	40.45908	38.69227	39.52995	39.86100	39.53909

*TGT set to 2500 in the Affymetrix Expression Report (attachment A)

Distribution of Fluorescence Intensities

In addition to the Affymetrix Expression Report, Precision Biomarker Resources, Inc. uses it's own parameters to determine overall chip quality. The un-scaled distributions of raw intensity values (log2) provide a good indicator of how similar or dissimilar chip data are within an experiment. Two views of this data are provided below. On the left are Box Plots of the raw intensities for each chip. The red box is the range of the intensities from the middle 50% of data points. The dark black line in the red box is the median intensity. The grey “whiskers” extending above and below the red box indicate the overall range of the intensities. The line graph on the right shows the distribution of intensities. What is important to note is how similar the distributions are. In fact they are almost completely super imposable.

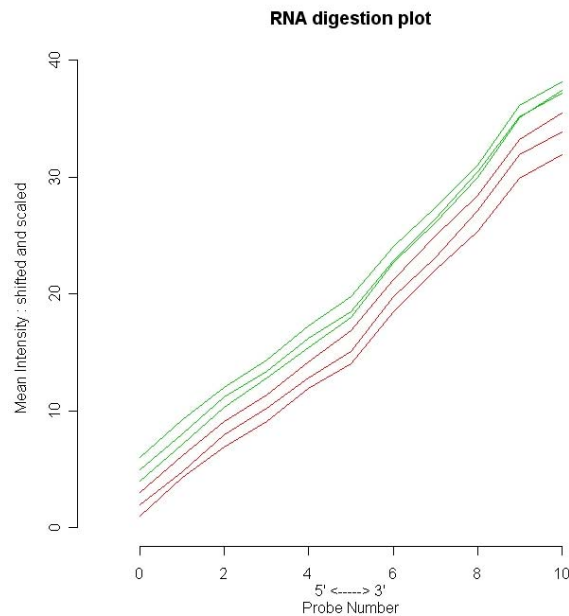


Raw Intensity Data for Chips Run 16 February 2007: *Left Panel* – Box plots for individual chips. From left to right; 1)CUST1, 2)CUST2, 3)CUST3, 4)CUST4, 5)CUST5, 6)CUST6. *Right Panel* – Histogram of log2 Intensities for individual chips. Each line is color coded for an individual sample but the overlap is too high to distinguish individual chips.

RNA Quality

The 3'/5' ratios of B-Actin and GAPDH are typically used for RNA quality assessment. However, these 3'/5' ratios can be quite variable even with high quality data. Therefore, a more global assessment of RNA quality is desirable.

The individual probes of a GeneChip probeset are numbered sequentially from the 5' end of the targeted transcript. On the U133v2 chip there are 11 PM probes per targeted transcript. Theoretically, a 3'/5' ratio can be determined for each individual probeset. The behavior of individual probes and RNA quality are the two most important factors in this. Probe effects dominate over RNA quality for any single probeset even with highly degraded RNA. When the average of 54,000 probes at each position are compared, a 3'/5' trend can be visualized. In practice, the slope of the line is the most important indicator. If all the arrays have similar slopes then it is valid to make comparisons within genes across arrays. The plot and summary of slopes below indicate comparable RNA was hybridized with each chip.



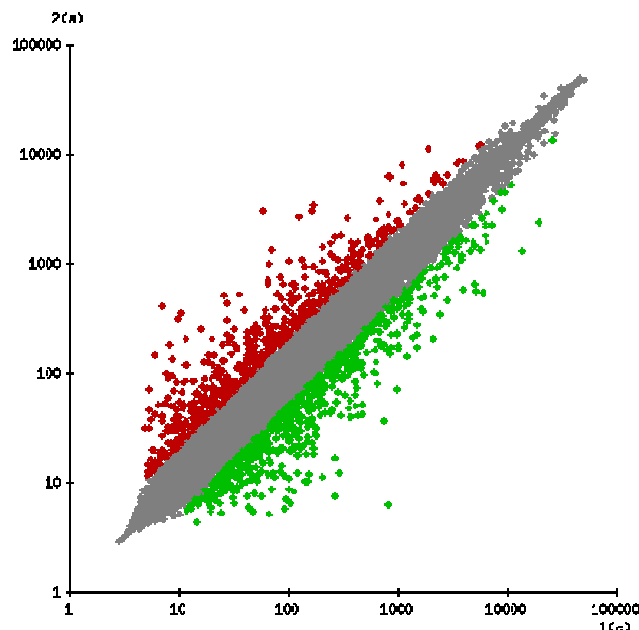
Global Assessment of RNA Quality. Each line represents one chip. Red lines are from samples labeled "CUST 1,2,and3" and green lines are from labeled "CUST4,5,and6". Plotted on the Y axis is the chip mean intensity by probeset position. Intensities have been shifted from original data for a clearer view. The slope is unchanged indicating similar RNA quality across all samples and the slope data are summarized in the table below.

SAMPLE	CUST1	CUST2	CUST3	CUST4	CUST5	CUST6
Slope	3.13	3.25	3.30	3.38	3.27	3.25
<i>pValue</i>	9.75e-11	2.30e-10	7.05e-11	1.17e-10	1.61e-10	8.78e-11

Pair-Wise Comparisons of Case and Control Samples

Two approaches to probe-level analysis and normalization were used in the analysis of these data. The first was MAS 5.0. The second was by GCRMA in either GeneSifter or R statistical software. GCRMA is a model based algorithm that weights PM-MM within a probeset and across the arrays in a data set. The resulting expression values were averaged for the Control (CUST1,2,and3) and Case (CUST4,5,and6) samples. The two groups were then compared by Student's t-test and a p-value of ≤ 0.01 was used to determine significance. A second filter of 2.0 fold difference was applied to eliminate small differences. Both the MAS 5.0 and GCRMA approaches returned similar gene lists. The scatter plot below, from GCRMA analysis, graphically summarizes comparisons of the Control and Case groups. Red dots represent gene with relative over expression in the Cases and green dots those genes with relative over expression in the Controls.

Scatter Plot



A total of 1,288 genes were determined to be differentially expressed by this method. When the Control group is used as a baseline, 629 of these genes were found to

be relatively up-regulated in the Cases and 659 relatively down-regulated in the Cases group. These gene lists can be found in an Excel spreadsheet on the accompanying CD.

A major concern in the analysis of differential gene expression is the multiplicity issue which can lead to a large amount of false positive data. Several approaches have been described to account for this. The p-values for these data were adjusted by the Bonferroni procedure. This procedure reduced the number of differentially expressed genes to 340. Tables of the gene lists can be found as attachments B and C.

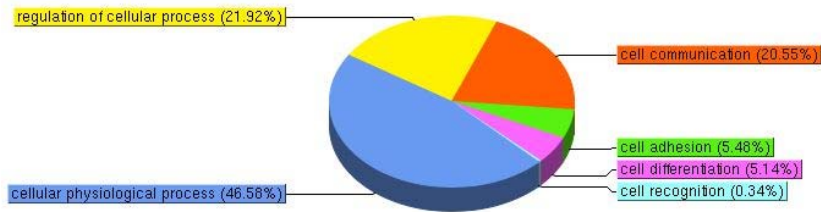
Gene Ontology and KEGG Pathway Analysis of the Differentially Expressed Genes

Closing the gap between the differentially expressed gene lists described above and knowledge of the biological function requires integration of additional biological data and databases. The Gene Ontology (GO) Consortium (<http://www.geneontology.org/>) defines three ontologies or structured vocabularies that assist in characterizing knowledge of gene lists. These ontologies are biological process (BP), molecular function (MF), and cellular component (CC). Placing the gene lists in the context of a biological pathway is another approach to gain functional knowledge. The Kyoto Encyclopedia of Genes and Genomes (KEGG) provides canonical pathways that the gene lists can be compared against (<http://www.genome.jp/kegg/pathway.html>).

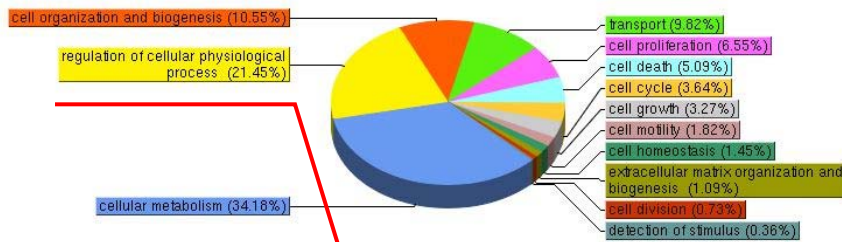
Gene Ontology Report

Gene ontologies are hierarchical. The pie charts below start at the highest level ontology. The wedges represent the fraction of differentially expressed genes associated with a particular ontology. In the example below, the ontologies have been drilled down two additional levels to “regulation of physiological process”. Each of the “regulation.....” sub-headings have been listed along with the number of genes changed and the “z” score for that category. A “z” score greater than 2.0 suggests that the probability of change for this ontology is significantly greater than chance.

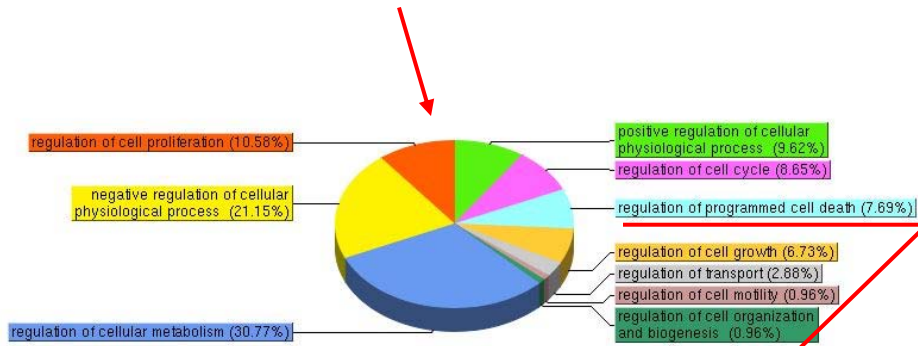
Physiological Process



Cellular Physiological Process



Regulation of Cellular Physiological Process



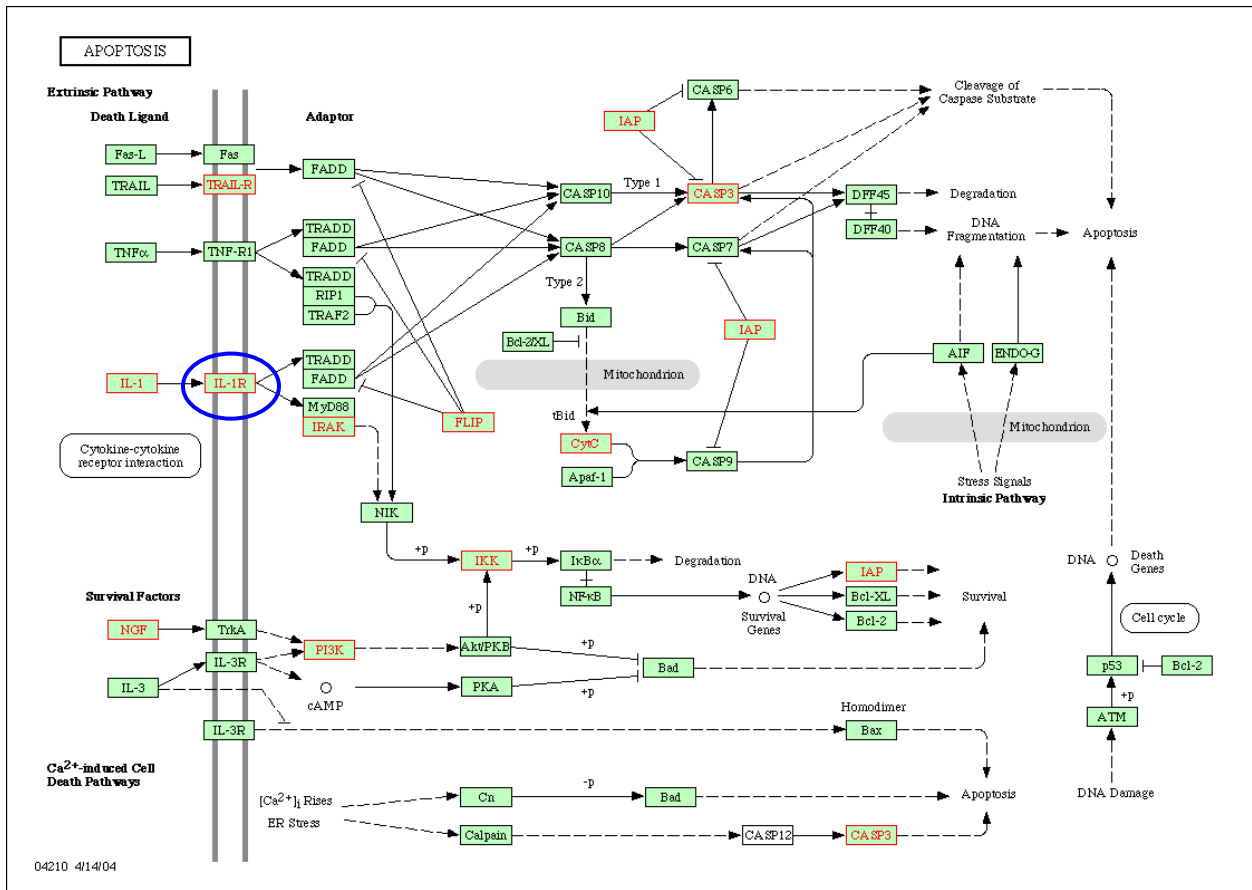
Regulation of Programmed Cell Death

Ontology	Genes	GO	Totals			z-score	
			List	↑	↓	Array	↑
regulation of cellular metabolism	32	22	10	2075	1.21	-1.71	
negative regulation of cellular physiological process	22	12	10	640	2.95	2.32	
regulation of cell proliferation	11	5	6	287	1.69	2.56	
positive regulation of cellular physiological process	10	1	9	484	-1.56	2.76	
regulation of cell cycle	9	2	7	425	-0.85	2.07	
regulation of programmed cell death	8	3	5	357	-0.00	1.36	
B-cell translocation gene 1, anti-proliferative /FL=gb:NM_001731.1							
ESTs							
ESTs							
GCN1 (general control of amino-acid synthesis 1, yeast)-like 1							
Homo sapiens B-cell translocation gene 1, anti-proliferative (BTG1), mRNA. /PROD							
Homo sapiens mRNA for kainate receptor subunit (GRIK2 gene). /PROD=kainate recep							
Homo sapiens serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), mem							
Homo sapiens TNF-induced protein (GG2-1), mRNA. /PROD=TNF-induced protein /FL=gb							
tumor necrosis factor, alpha-induced protein 3 /FL=gb:NM_006290.1 gb:M59465.1							
regulation of cell growth	7	4	3	109	3.25	2.35	
regulation of transport	3	1	2	57	0.76	2.35	
regulation of cell motility	1	1	0	33	1.38	-0.51	
regulation of cell organization and biogenesis	1	1	0	56	0.78	-0.66	
regulation of cell activation	-	0	0	40	-0.58	-0.56	
regulation of cell division	-	0	0	2	-0.13	-0.13	
regulation of cell redox homeostasis	-	0	0	5	-0.21	-0.20	
regulation of cellular defense response	-	0	0	3	-0.16	-0.15	
regulation of receptor recycling	-	0	0	1	-0.09	-0.09	
regulation of synaptic growth at neuromuscular junction	-	0	0	2	-0.13	-0.13	

KEGG Pathway Report

Significantly up-regulated KEGG pathways. Significance is determined by comparing the ratio of the genes from a given pathway differentially expressed vs. the total number of pathway genes represented on the chip. A “z” score of greater than 2.0 suggests the change in the pathway genes is likely real and not due to chance.

Pathway	List	Totals		Array	z-score	
TGF-beta signaling pathway	15	11	4	77	6.18	0.80
ECM-receptor interaction	15	10	5	85	5.07	1.19
Cell Communication	14	10	4	106	4.21	0.14
Atrazine degradation	2	2	0	7	4.14	-0.51
Peptidoglycan biosynthesis	1	1	0	2	4.05	-0.27
Arachidonic acid metabolism	8	6	2	53	3.79	0.10
O-Glycan biosynthesis	5	3	2	22	3.09	1.42
Focal adhesion	22	12	10	192	2.98	1.30
Nitrogen metabolism	3	3	0	24	2.89	-0.94
Neuroactive ligand-receptor interaction	7	1	6	286	-2.61	-1.36
Complement and coagulation cascades	15	5	10	65	2.42	5.23
Axon guidance	13	7	6	126	1.91	0.76



Example KEGG Pathway. Genes from the list of differentially expressed genes are highlighted in red