

Sample Report



Monarch Life Sciences™
Advancing proteomics • Transforming biomarker research

Statistical Summary for Project 000-0024-T-A01

Project Goal:

The goal of this study was to analyze the differential protein expression induced by the GFAT inhibitor, azaserine, on the 3T3 adipocyte differentiation program.

Project Background:

This study was carried out on mouse 3T3.L1 fibroblasts, induced into differentiated adipocytes.

In group A, the 3T3.L1 fibroblasts were treated with azaserine. After initiating differentiation, samples from both the Treated (A) and Control(C) groups were collected at different time points (0, 6, 12, 24, 48 and 72hrs) and analyzed by mass spectrometry.

Statistical Summary Report for Project 000-0024-T-A01:

This report and 6 additional files are available on a DVD

These attached files are explained in this report:

1. **SEchart**-000-0024-T-A01.doc (Open with Word)
(Plot of mean and standard error for individual proteins)
2. **ProteinMeans**-000-0024-T-A01.xls (Open with Excel)
(Means and other summary data on all individual proteins)
3. **VARchart**-000-0024-T-A01.doc (Open with Word)
(Plot of intensity levels showing sources of variability for individual proteins)
4. **ProteinIntensities**-000-0024-T-A01.csv (First 65536 rows with Excel)
(Protein level intensities, text file, comma delimited)
5. **ClusterPlots**-000-0024-T-A01.pdf (Open with Adobe Reader)
(Plots of 2 to 8 proteins per page with similar profiles)
6. **PeptideIntensities**-000-0024-T-A01.csv (First 65536 rows with Excel)
(Peptide level intensities and sequence data, text file, comma delimited)

Overall Summary:

There were 2569 proteins identified and quantified. Of these 2569 proteins there were 1108 proteins identified with high confidence (Priority 1). Of the 1108 priority 1 proteins there were 984 proteins that had at least one significant change between treatment groups (there were 16 two group comparisons selected among the 12 groups). The significance threshold is set to control the False Discovery Rate (FDR) at less than 5%. A False Discovery is a protein declared significant when it isn't. The replicate median % Coefficient of Variation (%CV) for the Priority 1 proteins was 6.64% and the combined replicate plus sample median %CV was 7.31%. The %CV is the Standard Deviation divided by the Mean on a % scale. There were a total of 1954 proteins that had significant changes among all 2569 proteins identified. **Many experts would view any protein identification outside of priority 1 as questionable (Carr, S., Molecular & Cellular Proteomics 3.6, 2004, pp. 531-533).**

Experimental Design:

The experimental design is described below:

Treatment	Time (Hours)	Group Abbreviation	Number of Samples Per Group	Number of Replicates Per Sample
A	0	A00	3	2
A	6	A06	3	2
A	12	A12	3	2
A	24	A24	3	2
A	48	A48	3	2
A	72	A72	3	2
C	0	C00	3	2
C	6	C06	3	2
C	12	C12	3	2
C	24	C24	3	2
C	48	C48	3	2
C	72	C72	3	2

There were a total of $2 \times 6 \times 3 \times 2 = 72$ injections. The injection order was randomized on the LTQ.

Of the 66 possible two group comparisons the following 16 were tested for significance:

Comparison	Control/Reference Group	Treatment/Comparator Group
1	A00	A06
2	A00	A12
3	A00	A24
4	A00	A48
5	A00	A72
6	C00	C06
7	C00	C12
8	C00	C24
9	C00	C48
10	C00	C72
11	A00	C00
12	A06	C06
13	A12	C12
14	A24	C24
15	A48	C48
16	A72	C72

Protein Identification:

The proteins quantified were classified according to identification quality (Priority) as shown in the table below:

Protein Priority	Peptide ID Confidence	Multiple Sequences	Median Number of Sequences	Number of Proteins
1	HIGH	YES	4	1108
2	HIGH	NO	1	637
3	LOW	YES	2	25
4	LOW	NO	1	799
<i>Overall</i>			<i>1</i>	<i>2569</i>

Priority assignments reflect our level of confidence in the protein identification. Priority 1 proteins would have the highest likelihood of correct identification and Priority 4 the lowest likelihood of correct identification. This priority system is based on the quality of the amino acid sequence identification (Peptide ID Confidence) and whether one or more sequences were identified (Multiple Sequences). **Many experts would view any protein identification outside of priority 1 as questionable (Carr, S., *Molecular & Cellular Proteomics* 3.6, 2004, pp. 531-533).** The Peptide ID Confidence assigns a protein into a ‘High’ or ‘Low’ classification. This is based on the peptide with the highest peptide ID Confidence (the best peptide). Proteins with best peptide having a confidence between 90-100% are assigned to the ‘HIGH’ category. Proteins with best peptide having a confidence between 75-89% are assigned to the ‘LOW’ category. All peptides with confidence less than 75% were pre-filtered out. For priority 1 proteins all peptides with ID Confidence <90% are also pre-filtered out. Sequest and Tandem database search algorithms are used for amino acid sequence identification. Each algorithm compares the observed peptide MS/MS spectrum and theoretically derived spectrums from the data base to assign quality scores (Xcorr in Sequest and e-Score in Tandem). These quality scores and other important predictors are combined in a proprietary algorithm that assigns an over all score, %ID Confidence, to each peptide. The assignment was based on a random forest recursive partition supervised learning algorithm.

The %ID Confidence score is calibrated so that approximately X% of the peptides with %ID Confidence > X% are correctly identified. Higgs, R. E., et. al., *Journal of Proteome Research*, 2005, Vol. 4, pp. 1442-1450 gives a brief description and reference for this machine learning approach in the section on peptide identification, starting at the bottom of page 1444. A detailed description will be published in the near future.

It is also intuitively understood that our confidence in protein identification is increased with the number of distinct amino acid sequences identified. Therefore we also categorize proteins depending on whether they have only one or multiple sequences of the required confidence. A protein is classified as ‘YES’ in the ‘Multiple Sequences’ column if it has at least two distinct amino acid sequences with the required ID confidence; otherwise it is classified as ‘NO’.

Protein Quantification:

Every peptide quantified has an intensity measurement for every sample. The intensity measurement is a relative quantity giving the area under the curve (auc) from the select ion chromatogram (XIC) after background noise removal. The auc is measured at the same retention time for each sample after the sample chromatograms have been aligned (Higgs, R. E., et. al., *Journal of Proteome Research*, 2005, Vol. 4, pp. 1442-1450). The intensities are then transformed to the log scale where base 2 has become customary. The log transformation serves two purposes. First, relative changes in protein expression are best described by ratios. However ratios are difficult to statistically model and the log transformation converts a ratio to a difference which is easier to model. Second, as is frequently the case in biology the data better approximate the normal distribution on a log scale (Limpert, E., et. al., *BioScience*, 2001, Vol. 51, No. 5, pp. 341-352) which is important because normality is an assumption of the ANOVA models used to analyze this data. The base of the log transform is arbitrary with base 2 the most common with genomic data. Base 2 is popular because a two fold change (or doubling, or 100% increase) yielding an expression ratio of 2 is transformed to 1 on a log base 2 scale (i.e. a two fold change is a unit change on the log base 2 scale). After log transformation the data are then quantile normalized (Bolstad, B. M., et. al., *Bioinformatics*, 2003, Vol. 19, No. 2, pp. 185-193). Quantile normalization is a method of normalization that essentially ensures that every sample has a peptide intensity histogram of the same scale, location and shape. This normalization removes trends introduced by sample handling, sample preparation, possible total protein differences as well as changes in instrument sensitivity while running multiple samples.

If multiple peptides have the same protein identification then their quantile normalized log base 2 intensities are averaged to obtain log base 2 protein intensities. The average of the normalized log peptide intensities is a weighted average. A peptide is weighted proportional to the peptide ID Confidence. The log base 2 protein intensity is the final quantity that is fit by a separate ANOVA statistical model for each protein. The ANOVA (Analysis of Variance) is a statistical model that separates the variation due to groups, samples and replicates and constructs the appropriate statistics for discovering group differences. The statistical model is covered in more detail at the end of the report.

Summary of Significant Results:

The number of significant changes, the maximum absolute fold changes and the exhibited variability (coefficient of variation) for each priority group are described in the table below:

Protein Priority	Peptide ID Confidence	Multiple Sequences	Number of Proteins	Number Significant Changes	Max Absolute Foldchange	Median %CV replicate	Median %CV rep + sample
1	HIGH	YES	1108	984	4.9845395	6.64	7.31
2	HIGH	NO	637	418	25.804793	13.97	15.11
3	LOW	YES	25	18	2.0360905	10.86	11.89
4	LOW	NO	799	534	11.49045	15.35	17.45
<i>Overall</i>			2569	1954	25.804793	10.85	11.93

The table above gives the number of proteins with significant changes for each Priority level. The threshold for significance is set to control the **False Discovery Rate (FDR)** for each two group comparison at 5% (Reiner, et. al., *Bioinformatics*, 2003, Vol. 19, No. 3, pp. 368-375). The FDR is estimated by the q-value which is an adjusted p-value. The FDR is the proportion of significant changes that are false positives. If proteins with a q-value $\leq .05$ are declared significant it is expected that 5% of the declared changes will be false positives. It is a misconception that the p-value estimates the FDR. The p-Value estimates the False Positive Rate (FPR) which is the proportion of false positives among the proteins that in reality did not change. The FPR = 1- Specificity and FDR = 1 – Positive Predictive Value in the language of medical diagnostics. (Note: the p-Value to q-Value adjustment is done separately for priority 1, priority 2 and the LOW confidence categories.)

The maximum observed absolute Fold Change is also given for each Priority Level.

Fold Change is computed as follows:

Fold Change = Mean Treated Group / Mean Control Group
When Mean Treated Group \geq Mean Control Group

Fold Change = \square Mean Control Group / Mean Treated Group
When Mean Control Group $>$ Mean Treated Group

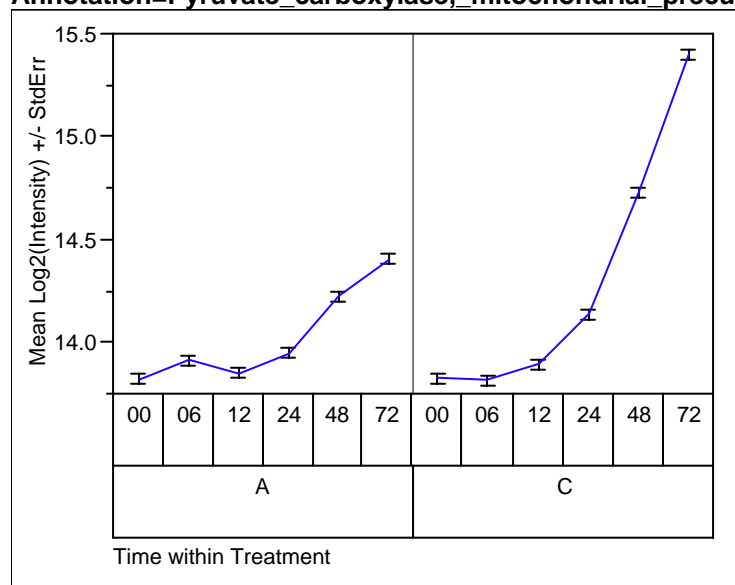
Absolute Fold Change = | Fold Change | = absolute or positive value of the Fold Change
A Fold Change of 1 means there is no change.

Also in the table is the Median % Coefficient of Variation (%CV) for each Priority Level. The %CV is the standard deviation / mean on a % scale. The %CV is given both for the replicate variation as well as the combined replicate plus sample variation.

1. Standard Error charts for proteins (SEchart.doc):

This file contains separate standard error charts for proteins with ranks 1-984. These are all of the priority 1 proteins with significant changes. For each of the proteins there is a plot of the group mean intensity levels on the log base 2 scale plus or minus the standard error. The standard error is computed from the statistical model and is a measure of the precision of the mean. The blue line connects the group means and helps to visualize the treatment trend. A change in 1 unit on the log base 2 scale represents a doubling or a two fold change. The rank 1 protein plot is shown as an example below. The maximum fold change (maxFC), minimum q-Value (q_min) and a brief annotation are displayed. Rank is assigned by sorting all the proteins in the following order: Significant Change (Yes, No), Priority (1-4) and q_min. The plotted values are found in the protein means spread sheet described in the next section.

**Rank=1, protein_id=IPI00114710.1, maxFC=2.980572, q_min=0,
Annotation=Pyruvate_carboxylase,_mitochondrial_precursor**



2. Protein Means Spread Sheet (ProteinMeans.xls):

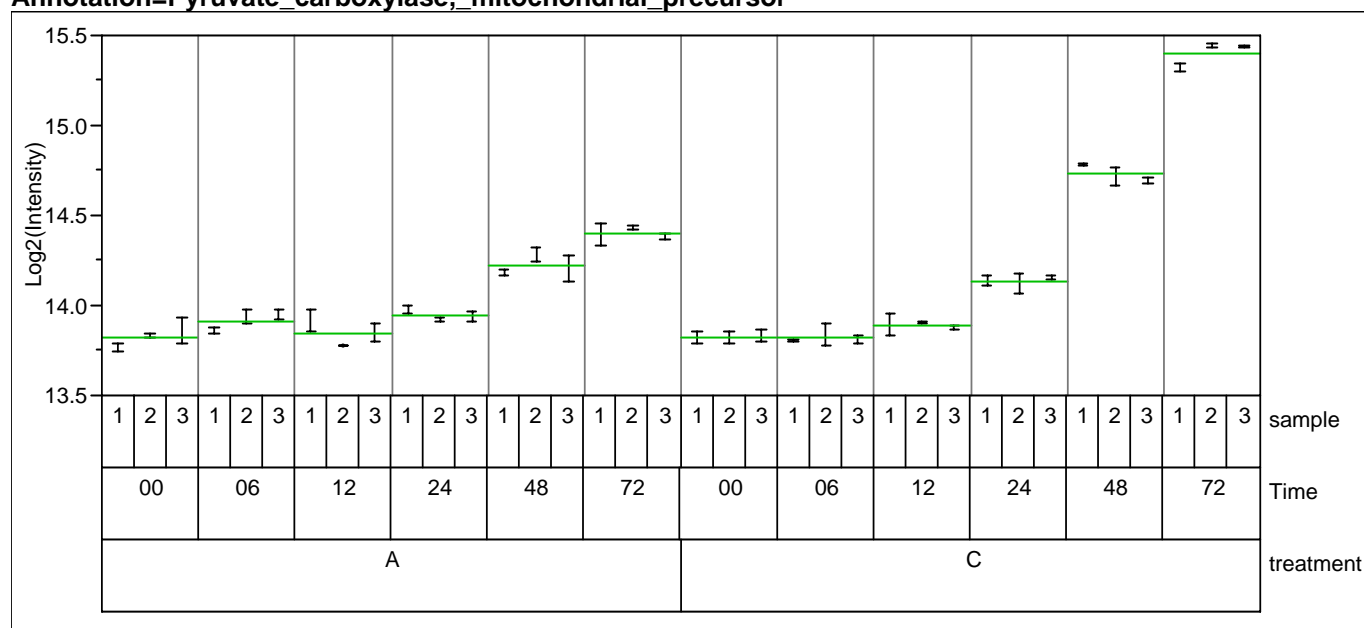
The proteins in the spread sheet are ordered by Rank. Rank is assigned by sorting all the proteins in the following order: Significant Change (Yes, No), Priority (1-4), q_min. There is a single row for each protein quantified with summary information as described below:

Column Name	Description
Rank	Ranked by Significant Change, Priority, q_min
Sig_Change	YES if q_min <= .05 otherwise NO (among 16 comparisons)
Priority	1-4 based on peptide ID confidence and number of sequences
Protein_ID	IPI or GI data base number
Annotation	Available Annotation
Best_Sequence	Amino Acid Sequence of peptide with highest ID Confidence
ID_Confidence	% ID Confidence of Best Sequence
Max_Fold_Change	Maximum absolute fold change among selected 16 two group comparisons
q_min	Minimum q-Value among selected 16 two group comparisons
q_int	q-Value for testing different shape for A and C time profiles
q_A00_A06, ... (16 cols)	q-Values for selected 16 two group comparisons
p_min	Minimum p-Value among selected 16 two group comparisons
p_int	p-Value for testing different shape for A and C time profiles
p_A00_A06, ... (16 cols)	p-Values for selected 16 two group comparisons
FC_A00_A06, ... (16 cols)	Fold Change for selected 16 two group comparisons (1 st group in pair is control for calculation; see statistics section)
mean_A00, ... (12 cols)	The mean protein intensity for each group
%CV_Rep	% Coefficient of Variation for replicate variation
%CV_Rep+Sample	% Coefficient of Variation for replicate plus sample variation
mean_log2_A00, ... (12 cols)	The mean of the log base 2 protein intensities for each group (plotted in standard error charts)
se_log2_A00, ... (12 cols)	The ANOVA standard error of log base 2 mean for each group (plotted in standard error charts)
Number_Sequences	The number of distinct amino acid sequences for this protein
Number_Peptides	The number of peptides quantified (not all distinct sequences)
ClusterRank	Rank of cluster group (see description in report)
ClusterMetric	Metric describing cluster similarity (see description in report)

3. Variability charts for proteins (VARchart.doc):

This file contains separate variability charts for proteins with ranks 1-984. These are all of the priority 1 proteins with significant changes. For each of the proteins there is a plot of the individual protein intensity levels on the log base 2 scale (in sample number order). The horizontal green line is the group mean. Intensities for the two replicates are joined by a vertical line and each sample is plotted separately. This plot allows the replicate, sample and group variation to be visually assessed. A change in 1 unit on the log base 2 scale represents a doubling or a two fold change. The rank 1 protein plot is shown as an example below. The maximum fold change (maxFC), minimum q-Value (q_min) and a brief annotation are displayed. Rank is assigned by sorting all the proteins in the following order: Significant Change (Yes, No), Priority (1-4) and q_min. The plotted values are found in the protein intensity spread sheet described in the next section.

Rank=1, protein_id=IPI00114710.1, maxFC=2.980572, q_min=0,
Annotation=Pyruvate_carboxylase,_mitochondrial_precursor



4. Protein Intensities Spread Sheet (ProteinIntensities.csv):

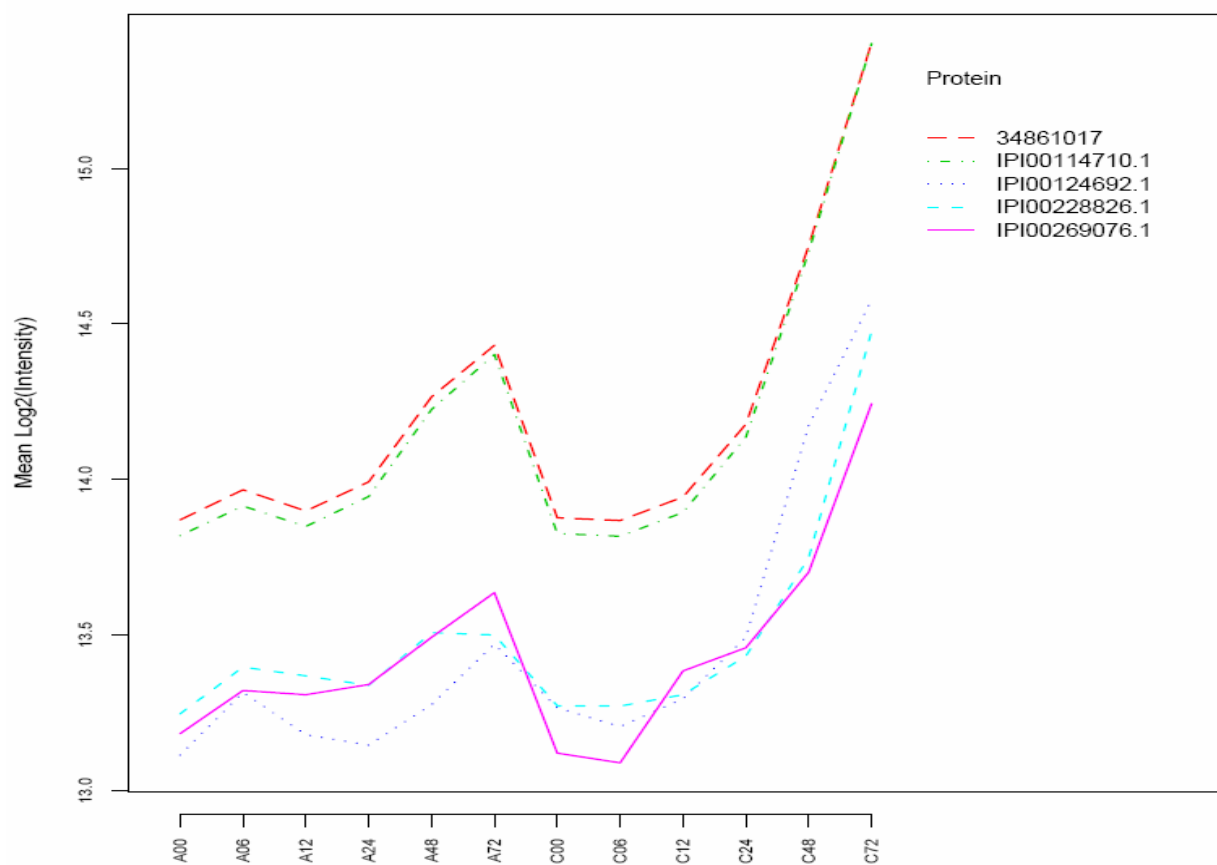
This is a comma delimited text file with protein intensity information. If it is double clicked the first 65536 rows will open in Excel. Excel can open a maximum of 65536 rows. For larger data sets you need a high quality text editor like UltraEdit. If you want to process larger data sets (sort it, plot it, etc.) you will need a statistical package like JMP or SAS. JMP is recommended because of the GUI interface and other features. We use JMP to validate our SAS analysis. The data in the text file are sorted by rank, group, sample, rep. The following table explains the column headings:

Column Name	Description
rank	Protein rank; ranked by Significant Change (YES/NO), Priority, q_min
protein_id	IPI or GI data base number
sample_name	Complete sample name
group	Group identification (e.g treatment condition, phenotype, etc.)
sample	Abbreviated sample name
rep	Sample replicate number
injection_order	Number indicating sequential randomly assigned injection order
auc	Protein level area under the curve (protein intensity) {peptide weighted average}
qauc	Quantile normalized auc
log2auc	Log base 2 scale of auc
qlog2auc	Log base 2 scale of qauc

5. Cluster Plots of Expression Profiles (ClusterPlots.pdf):

This file contains cluster plots for proteins with ranks 1-984. These are all of the priority 1 proteins with significant changes. These are all the proteins with significant changes or from priority 1. The proteins were grouped into 194 clusters with the constraint that the clusters have from 2 to 8 proteins per cluster. The criterion for grouping was similarity in expression profiles. (For a two group study this simply means similar fold changes). The small cluster size allows the user to assess the profile similarity visually. Cluster Rank 69 is shown as an example below because the Rank 1 protein is in this cluster. The clusters are ranked by the Cluster Metric which is a similarity measure for the profiles in a single cluster (the bigger the more similar). The INCAPSulate clustering algorithm that defines the clusters and the cluster metric are proprietary pending publication. Because the protein profiles for each cluster are directly visualized the abstract nature of the algorithm is not important. There is no claim that the clusters imply anything biological about the data. The only value is if the user finds the clusters useful for their interpretation of the data. Regarding profile similarity consider the following suggestions: “Profile similarity is a concept underlying many microarray elucidation procedures. It has been thought that genes with similar expression profiles are likely to be functionally associated. The encoded proteins may participate in the same pathway, form a common structural complex, or be regulated by the same mechanism.” (Li, et. al., *PNAS*, 2004, Vol. 101, No. 44, pp. 15561-15566)

Cluster Rank = 69 , Cluster Metric = 719.08978979



6. Peptide Intensities Spread Sheet (PeptideIntensities.csv):

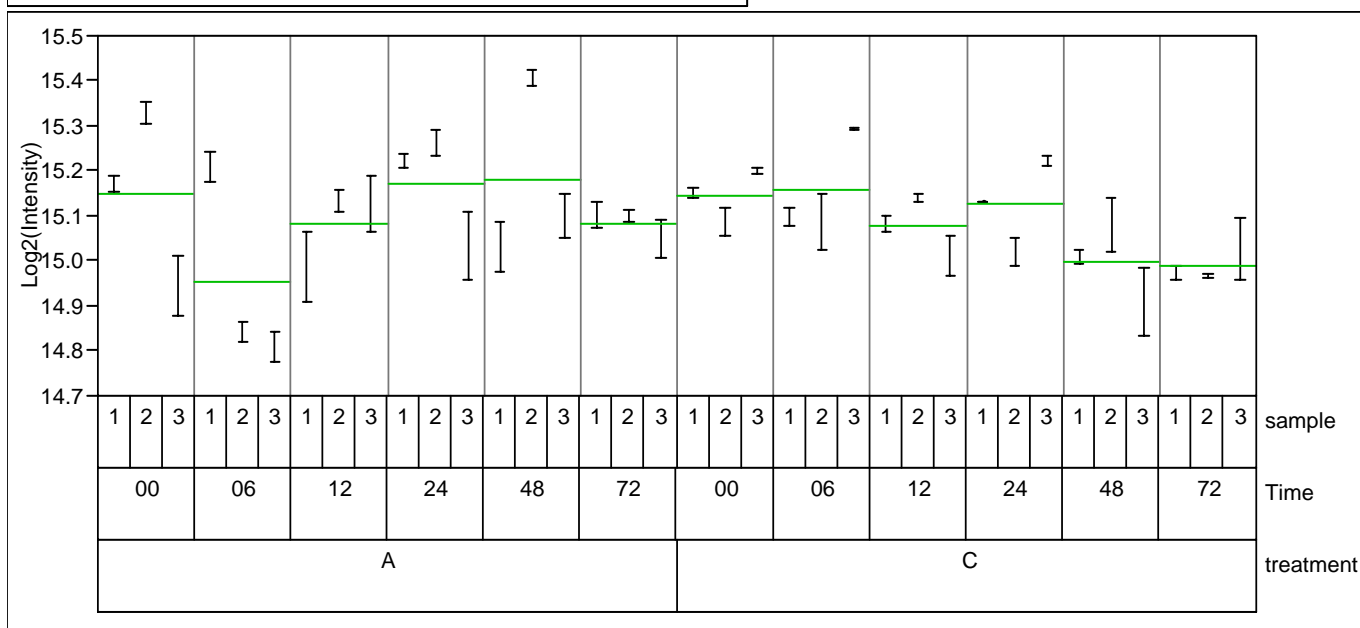
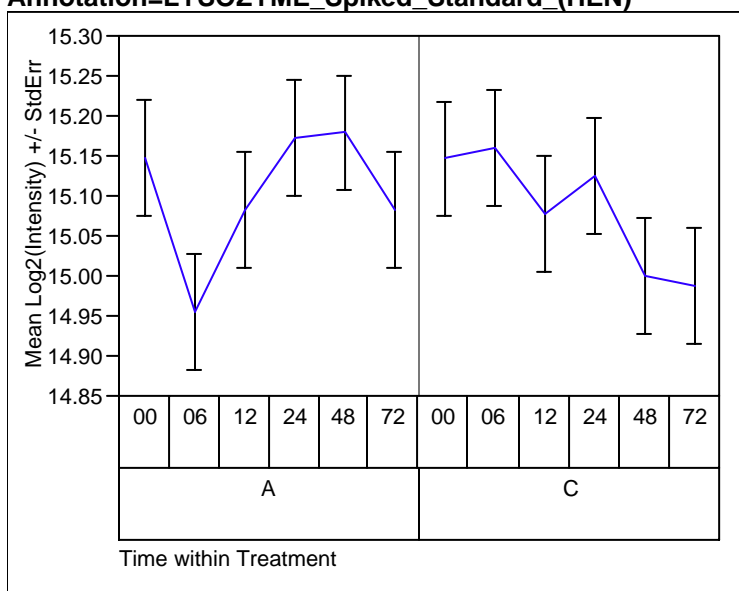
This is a comma delimited text file with peptide level information including peptide amino acid sequence and peptide intensity (area under the curve). If it is double clicked the first 65536 rows will open in Excel. Excel can open a maximum of 65536 rows. For larger data sets you need a high quality text editor like UltraEdit. If you want to process larger data sets (sort it, plot it, etc.) you will need a statistical package like JMP or SAS. JMP is recommended because of the GUI interface and other features. We use JMP to validate our SAS analysis. The data in the text file are sorted by rank, peptide_id, group, sample, rep. The following table explains the column headings:

Column Name	Description
rank	Protein rank; ranked by Significant Change (YES/NO), Priority, q_min
protein_id	IPI or GI data base number
peptide_id	Unique peptide identification in our output file
sequence	Amino acid sequence
num_tryp_cuts	Number of trypsin cleavage sites
zstate	Peptide charge state
IDconf	Peptide % ID confidence
sample_name	Complete sample name
group	Group identification (e.g treatment condition, phenotype, etc.)
sample	Abbreviated sample name
rep	Sample replicate number
injection_order	Number indicating sequential randomly assigned injection order
auc	Area under the select ion chromatogram curve (XIC), i.e. peptide intensity
gauc	Quantile normalized auc (the antilog of qlog2auc)
log2auc	Log base 2 of auc
qlog2auc	Quantile normalized log2auc

Spiked Constant Internal Standard:

Chicken lysozyme was spiked at a constant amount of total protein before tryptic digestion. Below are the SEchart and VARchart for chicken lysozyme. This protein should show no significant change between groups. If there is a significant group effect (i.e. if q-Value < .05) then you should discard significant changes in other proteins with smaller fold changes.

Rank=2042, protein_id=126608, maxFC=1.152568, q_min=0.15863,
Annotation=LYSOZYME_Spiked_Standard_(HEN)



Statistical Methods:

For each protein a separate analysis of variance (ANOVA) model is fit:

$$\text{Log}_2(\text{Intensity}) = \text{Overall Mean} + \underset{\text{(Fixed)}}{\text{Group Effect}} + \underset{\text{(Random)}}{\text{Sample Effect}} + \underset{\text{(Random)}}{\text{Replicate Effect}}$$

Log₂(Intensity) is the protein intensity based on the weighted average of the quantile normalized log base 2 peptide intensities with the same protein identification.

Group Effect refers to the fixed effects (not random) caused by the experimental conditions or treatments that we want to compare.

$$\text{Group} = \text{Treatment} + \text{Time} + \text{Interaction}$$

The group effect can be partitioned into treatment, time and interaction effects. A significant interaction indicated that the time profiles have different shapes for the two treatment groups A and C.

Sample Effect (nested within group) refers to the random effects from individual biological samples. It also includes the random effects from the individual sample preparations.

Replicate Effect (nested within sample) refers to the random effects from replicate injections from the same sample preparation.

All of the injections from one experiment are run in random order on the same LTQ by the same operator.

When an ANOVA model has two or more random effects (such as Sample and Replicate here) and at least one fixed effect (such as Group here) it is called a Mixed Model.

These models were fit using PROC MIXED in SAS (version 9) for each protein. The REML method was used as a fit mechanism and degrees of freedom were computed using the Satterthwaite method. The RANDOM statement was used to model the covariance with the NOBOUND parameter option in the PROC statement. This method is the default method for fitting random effects mixed models in JMP (version 6). JMP is used to check/validate the SAS analysis for at least one protein. The information from the model fit was used to construct the protein means spread sheet.

Because protein intensity is on a log base 2 scale the group means and their differences are converted to arithmetic means and fold change by the following example formulas:

$$\begin{aligned} T &= \text{Treatment group average of log base 2 protein intensities} \\ C &= \text{Control group average of log base 2 protein intensities} \end{aligned}$$

We first take antilogs for base 2.

$$\text{Mean}_T = 2^T$$

$$\text{Mean}_C = 2^C$$

Finally Fold Change is computed

$$\text{Fold Change} = \text{Mean}_T / \text{Mean}_C \quad \text{when } \text{Mean}_T \geq \text{Mean}_C$$

$$\text{Fold Change} = \text{Mean}_C / \text{Mean}_T \quad \text{when } \text{Mean}_C > \text{Mean}_T$$

Variance components for the CV calculation were computed on the log scale and then converted to a CV on the arithmetic scale corresponding to the original auc or 'area under the curve'. This transformation was done assuming the original scale has the log normal distribution.

$$\%CV = 100 * \sqrt{\exp((\ln(2) * \text{standard deviation on log 2 scale})^2) - 1}$$

Ln is natural Log; ln(2) converts scale from log base 2 to natural log; exp is the exponential function; (see Limpert, 2001 for lognormal and CV formula)

Primary References:

B. M. Boldstad, R.A. Irizarry, M. Astrand and T.P. Speed, (2003), "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias", *Bioinformatics*, Vol. 19, No. 2, pp. 185-193.

This paper discusses various methods for normalizing 'omics data. In particular it highlights the advantages of quantile normalization which has been recommended by the paper's authors. We have of course chosen quantile normalization as our method of choice.

Steven Carr, Ruedi Aebersold, Michael Baldwin, Al Burlingame, Karl Clauser, and Alexey Nesvizhskii, (2004), "The Need for Guidelines in Publication of Peptide and Protein Identification Data", *Molecular & Cellular Proteomics*, Vol. 3, No. 6, pp. 531-533.

This paper discusses the issues concerning clarity in describing the quality of protein identification from MS/MS spectra, especially as it relates to publication. We feel our current method of defining priorities 1-4 using Peptide ID Confidence (High, Low) and Multiple Sequences (Yes, No) follows the spirit of their guidelines. Our recommendation to limit attention to priority 1 also follows their guidelines for reporting identification criteria.

Richard E. Higgs, Michael D. Knierman, Valentina Gelfanova, Jon P. Butler, and John E. Hale, (2005), “Comprehensive Label-Free Method for the Relative Quantification of Proteins from Biological Samples”, *Journal of Proteome Research*, Vol. 4, pp. 1442-1450.

This paper summarizes the proprietary methodology we use for quantification. It covers in detail the methods used for zoom scan assessment, MS/MS filtering, chromatographic alignment, peptide quantification and various validation studies.

Ker-Chau Li, Ching-Ti Liu, Wei Sun, Shinsheng Yuan, and Tianwei Yu, (2004), “A System for enhancing genome-wide coexpression dynamics study”, *PNAS*, Vol. 101, No. 44, pp. 15561-15566.

I quote from the second paragraph of this paper concerning the value of profile similarity. “Profile similarity is a concept underlying many microarray elucidation procedures. It has been thought that genes with similar expression profiles are likely to be functionally associated. The encoded proteins may participate in the same pathway, form a common structural complex, or be regulated by the same mechanism.” It is with these possibilities in mind that we implemented the INCAPSulate clustering algorithm. Although not implemented by us the paper also suggests methods for elucidating protein functional association that does NOT result in profile similarity.

Eckhard Limpert, Werner A. Stahel, and Markus Abbt, (2001), “Log-normal Distributions across the Sciences: Keys and Clues”, *BioScience*, Vol. 51, No. 5, pp. 341-352.

This paper show how common the log-normal distribution is for biology and other sciences. This helps motivate our use of the log transformation of the peptide intensity. If the data have a log-normal distribution then the log transformed data are normally distributed and amenable to common ANOVA statistical models.

Anat Reiner, Daniel Yekutieli and Yoav Benjamini, (2003), “Identifying differentially expressed genes using false discovery rate controlling procedures”, *Bioinformatics*, Vol. 19, No. 3, pp. 368-375.

This paper explains the most important concept of setting the significance threshold to control for the False Discovery Rate. The False Discovery Rate (FDR) is the proportion of false positives (discoveries) among those changes declared significant. The most common way of estimating the FDR is by an adjustment to the p-Value. The resulting adjustment in the p-Value sometimes called the q-Value was first discovered by Yoav Benjamini in 1995. This paper describes this adjustment and other approaches to estimating the FDR.

Monarch LifeSciences is a contract research organization specializing in protein biomarker discovery, development and validation.

Our clients know that there is a difference between finding protein biomarkers and finding the *right* biomarkers. **Monarch LifeSciences** sets the standard in protein biomarker discovery and validation by emphasizing the statistical significance of each discovery, leading to high confidence biomarkers. Biotech and pharmaceutical companies across the US and Europe increasingly rely on **Monarch's** protein biomarker discovery and validation services to support their research goals.

- We employ state-of-the-art mass spectrometry technologies along with proprietary sample preparation methods and quantification algorithms.
- We have an active internal research program in collaboration with the Indiana University School of Medicine. We continually improve our techniques and publish in peer-reviewed journals to expand the proteomics biomarker knowledge base.
- Our *LifeMarker*[™] mass-spec based assays support both pre-clinical and clinical applications of biomarkers. These assays are comparable in speed, accuracy and cost to traditional antibody based assays.
- **Monarch** does not hold client IP – your data belongs to you.