MS Bioworks
Biological Mass Spectrometry Solutions

3950 Varsity Drive
Ann Arbor, MI 48108
734-929-5083
www.msbioworks.com
info@msbioworks.com

Project Report

Information

Client:	*Client Name*
Company:	*Company Name*
Project Number:	*MSB-1234*
Date Submitted:	*03-01-2011*
Date Completed:	*03-08-2011*

Samples

Client identifier	**MSB identifier**	**Project**
Control	12345	MSB-1234
Treated	12346	

Objective

To quantitatively profile control (light) and compound-treated (heavy – Lys 6 only) human MCF10A cell line SILAC samples using the Protein-works 40 fraction differential analysis service. Data to be processed using the MaxQuant software.

Experimental Methods

Sample Preparation

Submitted cell lysates were quantitated by Qubit fluorometry (Invitrogen), 20µg of each sample was mixed and solubilized in LDS buffer, heated at 85°C for 5 min and 20µg separated on a 4-12% Bis-Tris Novex mini-gel (Invitrogen) using the MOPS buffer system. The gel was stained with coomassie and the lane was excised into 40 equally sized segments. Gel pieces were processed using a robot (ProGest, DigiLab) with the following protocol:

- Washed with 25mM ammonium bicarbonate followed by acetonitrile.
- Reduced with 10mM dithiothreitol at 60°C followed by alkylation with 50mM iodoacetamide at RT.
- Digested with trypsin (Promega) at 37°C for 4h.
- Quenched with formic acid and the supernatant was analyzed directly without further processing.

Mass Spectrometry

Each gel digest was analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher LTQ Orbitrap Velos. Peptides were loaded on a trapping column and eluted over a 75µm analytical column at 300nL/min; both columns were packed with Jupiter Proteo resin (Phenomenex). The mass spectrometer was operated in data-dependent mode, with MS performed in the Orbitrap at 60,000 FWHM resolution and MS/MS performed in the LTQ. The fifteen most abundant ions were selected for MS/MS.

Data Processing

Data were processed through the MaxQuant software v1.0.13.13 (www.maxquant.org) which served several functions:

1. Recalibration of MS data
2. Filtering of database search results at the 1% protein and peptide false discovery rate (FDR).
3. Calculation of SILAC heavy:light ratios.

During the MaxQuant process data were searched using a local copy of Mascot with the following parameters:

Enzyme: Trypsin
Labels: Lys+6Da
Database: IPI Human v3.75 (concatenated forward and reverse plus common contaminants)
Fixed modification: Carbamidomethyl (C)
Variable modifications: Oxidation (M), Acetyl (N-term), Pyro-Glu (N-term Q), Deamidation (N,Q)
Mass values: Monoisotopic
Fragment Mass Tolerance: 0.5 Da
Max Missed Cleavages: 2

Results

A total of 5093 proteins were identified with two or more unique peptides per protein at 1% protein and peptide FDR. All data are summarized within the accompanying spreadsheet. All raw MaxQuant outputs can be accessed at:

ftp://75.144.89.5

Username: SILAC-example
Password: silac-example1234

To learn more about MaxQuant we recommend that you visit www.maxquant.org, and also read the Nature articles: Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide quantification. Nature Biotechnol 26,

1367-72 and Cox, J. et al. (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nature Protocols 4, 698-705.

The summary spreadsheet contains two worksheets:

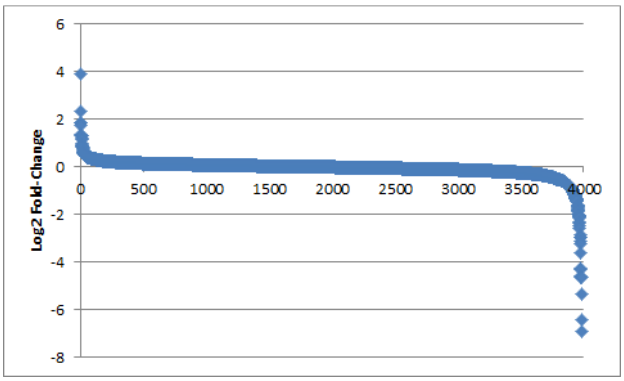
Worksheet 1: Quant

Only those proteins that had a SILAC ratio are reported in this summary. Any proteins that were identified with only Arg-containing tryptic peptides are not listed by definition; Lys peptides that were present in only one sample (or that MaxQuant could not integrate a signal in the opposite channel) are also excluded. Reverse hits and contaminant proteins (porcine trypsin, human keratins) have been removed. This leaves 3985 proteins that have been quantitated.

The column descriptions are:

Protein IDs:	IPI accession number(s)
Protein name:	Description of the protein
Gene names:	Name of the gene(s) from which the protein is transcribed
Unique peptides:	Number of unique peptides differing by sequence that are associated with the protein groups
Mol. Weight [kDa]:	Molecular weight of the leading protein in the group
Spectral Counts:	Number of times peptides differing by sequence or modification that are associated with the protein groups were observed
PEP	Posterior error probability rate of identified protein
Ratio H/L normalized:	Normalized ratio of heavy peptide divided by light peptide, calculated as the median of ratios of all peptide evidences assigned to that group
Ratio H/L significance:	p-value assigned to the significance of the change
Ratio H/L count:	Number of peptides (including multiple charge states, modifications) that were used to calculate H/L ratio
Intensity Total:	Total summed peptide intensity for the identified protein group
Intensity L	Total summed peptide intensity for the identified protein group in control (L)
Intensity H	Total summed peptide intensity for the identified protein group in treated (H)
Log2 FC	Log2 fold change of Ratio H/L normalized value

The Log2 FC is plotted below. The majority of proteins are not changing in this experiment.



Worksheet 2: ALL

Same as worksheet 1 except those proteins containing no SILAC ratio are now included. Reverse hits and contaminant proteins (porcine trypsin, human keratins) have been removed. This leaves 5093 proteins identified with two or more unique peptides. Column descriptions are as in Worksheet 1.

The default MaxQuant outputs can be downloaded as text files from the above FTP site. The two most important files are described below:

ProteinGroups

A comprehensive analysis of identified proteins, cross-references to various databases including Swissprot, Ensembl, Kegg and Gene Ontology. All reverse hits and contaminant proteins are included here.

Column header	**Description**	**External Links**
ID	Identifier of entry in the corresponding output file	
Peptide ID	Cross reference id of peptides from the peptides.txt table	
Protein IDs	List of identified protein ids which belong to the protein group	
Protein Names	Name of the protein	
Gene Names	Name of the genes from which the protein is transcribed of the protein	http://www.genenames.org/
Protein Descriptions	Description of the protein explicitly describing the isoforms	
Uniprot	Uniprot ids of proteins belonging to the protein group	http://www.expasy.org/
ENSEMBL	Ensembl ids of proteins belonging to the protein group	http://www.ensembl.org/index.html
VEGA	Vetebrate genome annotation ids corresponding to the protein group	http://vega.sanger.ac.uk/index.html
REFSEQ	REFSEQ ids of proteins present in the protein group	http://www.ncbi.nlm.nih.gov/RefSeq/
Pfam	Protein Family id	http://pfam.sanger.ac.uk/
Pfam Names	protein family names	
Pfam Descriptions	protein family name description	
GOCC	Gene ontology cellular component ids	http://www.geneontology.org/
GOCC Names	Gene ontology cellular component names	http://www.geneontology.org/
GOMF	Gene ontology Molecular function ids	http://www.geneontology.org/
GOMF Names	Gene ontology Molecular function names	http://www.geneontology.org/

GOBP	Gene ontology Biological process	http://www.geneontology.org/
GOBP Names	Gene ontology Biological process names	http://www.geneontology.org/
KEGG	Kegg gene ids	http://www.genome.jp/kegg/genes.html
KEGG Pathways	Kegg pathway ids to which the protein group belongs to	http://www.genome.jp/kegg/pathway.html
KEGG Pathway Names	Name of the kegg pathway	http://www.genome.jp/kegg/pathway.html
KEGG Ortholog	Kegg orthology id	http://www.genome.jp/kegg-bin/get_htext?htext=ko00001.keg&filedir=/files&hier=2
Proteins	Number of proteins belonging to the protein groups	
Peptides (seq)	Number of peptides differring by sequence that are associated with the protein groups	
Peptides (seq/mod)	Number of peptides differring by sequence or modification, that are associated with the protein groups	
Unique Peptides (seq)	Number of unique peptides differing by sequence that are associated with the protein groups	
Sequence Coverage [%]	Sequence coverage of the leading protein in the protein group	
Unique Sequence Coverage [%]	Sequence coverage of the leading protein in the protein group counting only unique peptides	
Mol. Weight [kDa]	Molecular weight of the leading protein in the protein group	
Sequence Length	Amino acid sequence length of the leading protein	
PEP	Posterior errror probaility rate of identified protein	
Intensity	Total summed peptide intensity for the identified protein group	
Intensity X	Total summed peptide intensity for the identified protein group in experiment X	
Reverse	plus indicates that the entry as a reverse hits	
Contaminant	plus indicates that the entry as a contaminant	

Peptides

A concise non-redundant list of identified peptide sequences: Peptide sequence, proteins groups that contain the peptide, modification, miss cleavages, length of peptide, PEP values. Mascot identification scores for the best-identified version are displayed.

Column header	**Description**
id	identifier of peptide sequence in the peptides.txt sheet
Protein Group IDs	ID of protein group to which the peptide belongs
Mod. Peptide IDs	Modification specific id for the corresponding peptides
Evidence IDs	Evidence id for the peptide including all the matched MSMS event ids
MS/MS IDs	list of MSMS events that match to the peptide
Oxidation (M) Site IDs	Modification specific id
Sequence	Amino acid sequenc of the peptide
Length	Amino acid length of the peptide
Missed Cleavages	Number of missed cleavages in the peptide
Mass	Mass of the peptide
Proteins	list of the proteins corresponding to the protein group where the peptide is assigned to
Leading Razor Protein	leading protien from the protein group
Unique	Binary "yes" or "no" stating whether the peptide is unique to the protein group
PEP	Posterior errror probaility value for the identified peptide
Mascot Score	Mascot score of the identified peptide