InfernoRDN
(previously DAnTE)

Ashoka Polpitiya

Where does InfernoRDN fit in?

Normalize raw abundances and rollup to proteins using RRollup, QRollup, and ZRollup

AMT Pipeline (LC-MS)

Multialign

InfernoRDN

Biological Conclusions

QRollup

Other CSV files (microarray)
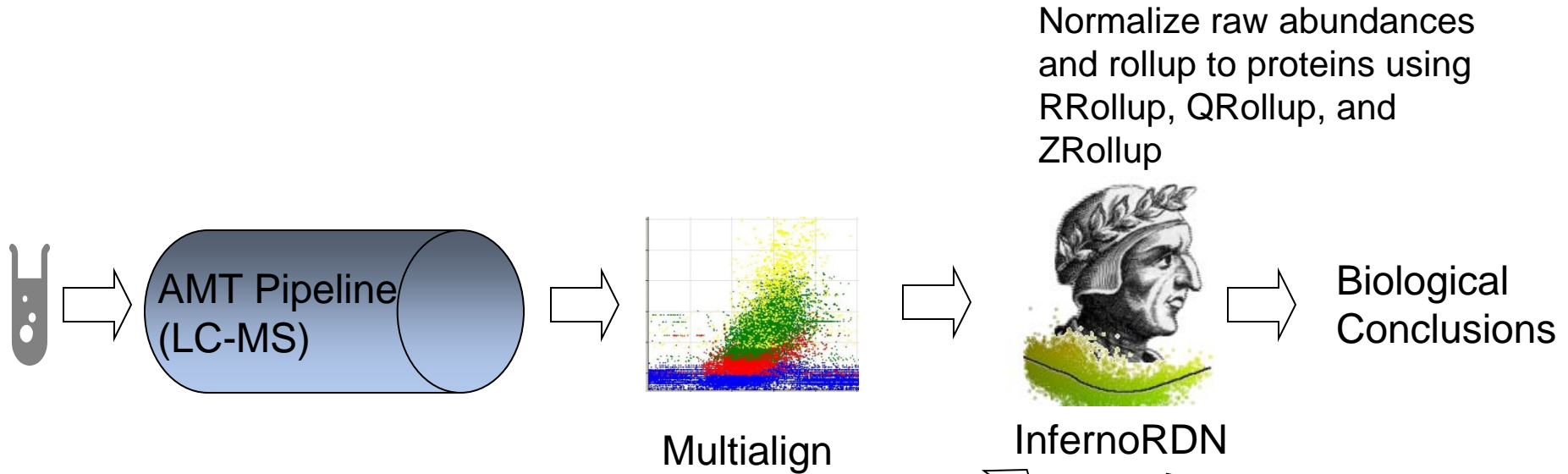
	A	B	C	D	E
1	protein name	mass tag id	Banth-002_2_(	Banth-002_3_02	Banth-002_4_03Se
2	BAS0108	232079382	502540046	589204259	254073890
3	BAS4270	232282863	502540046	589204259	254073890
4	BAS0416	232132998	292187962	530008742	215074319
5	BAS5178	232097558	292187962	530008742	215074319
6	BAS1329	232154357	228045784	364493723	213804137
7	BAS3391	232198366	228045784	364493723	213804137
8	BAS0100	232080778	144815519	279873999	161223712
9	BAS0498	232199488	144815519	279873999	161223712
10	BAS4547	232094505	144815519	279873999	161223712
11	BAS0108	232077214	139314698	231601483	169664271
12	P008|MYG_HORSE	11081	133331434	150507990	94626884
13	BAS0253	232087532	127773825	173439069	107476695

InfernoRDN

InfernoRDN is simply DAnTE re-branded

This screenshot shows the workspace after extensive data analysis

Analysis Flow in InfernoRDN
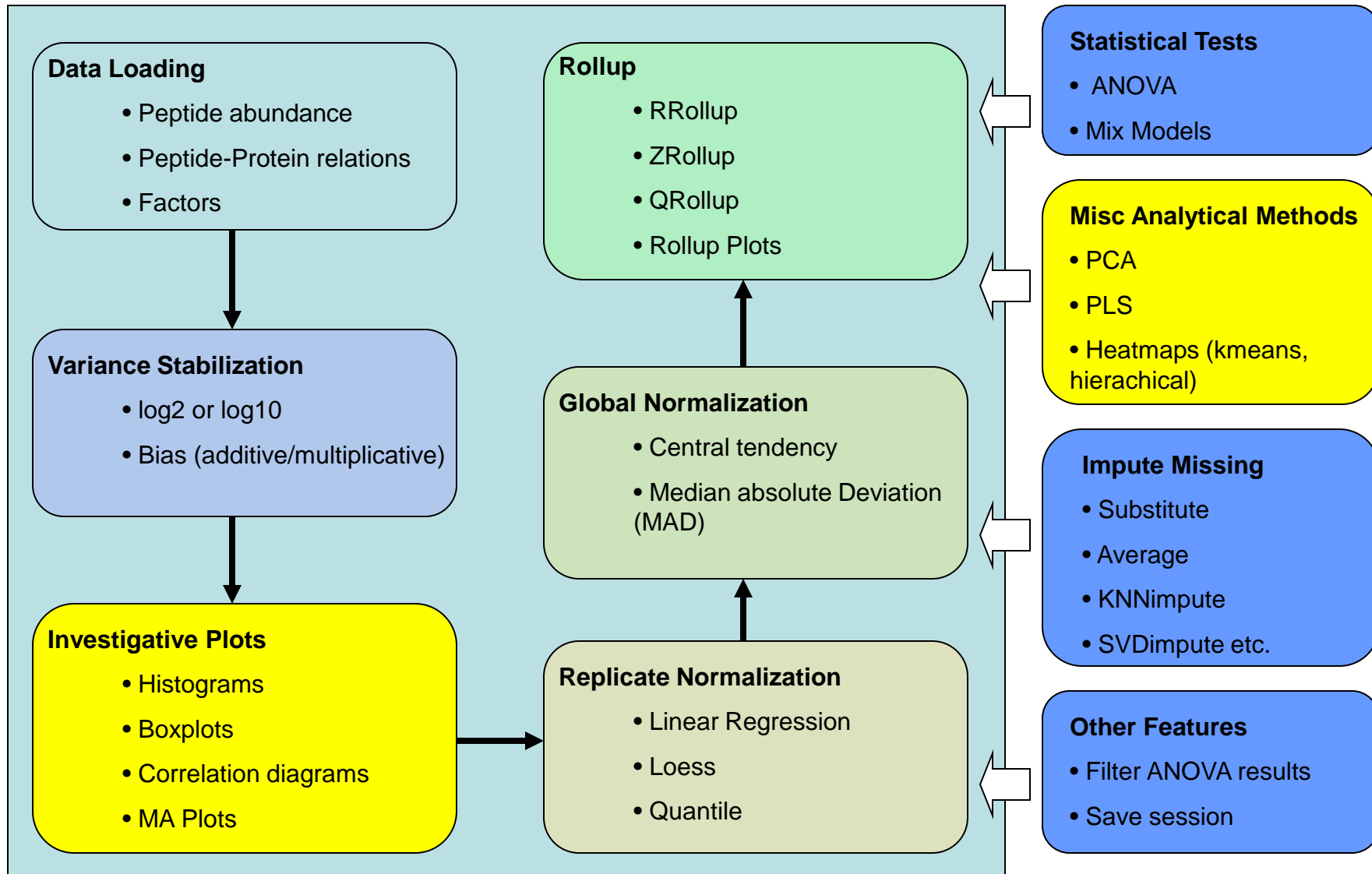
Goals of a downstream tool in Proteomics

- Identify problematic datasets
- Normalize
  - Remove systematic bias and variation due to technical artifacts
- Rolling up to proteins
- Hypothesis testing and feature discovery
  - Fixed effects (treatment)
  - Random effects (different LC columns, Batch)
  - Unbalanced data (due to missing)
  - PCA / PLS
  - Clustering (Hierarchical / K-means)

Goals of a downstream tool in Proteomics

- Identify problematic datasets
  - Correlation Plots

Outlier dataset

Color legend with overlaid histogram of correlation values

Dataset Names
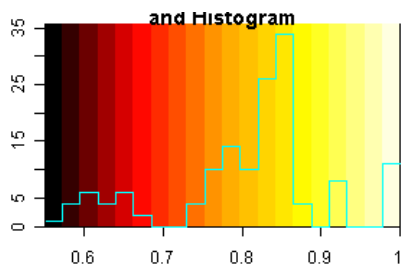
Dataset Names

Goals of a downstream tool in Proteomics

- Identify problematic datasets
- Normalize
  - Remove systematic bias and variation due to technical artifacts

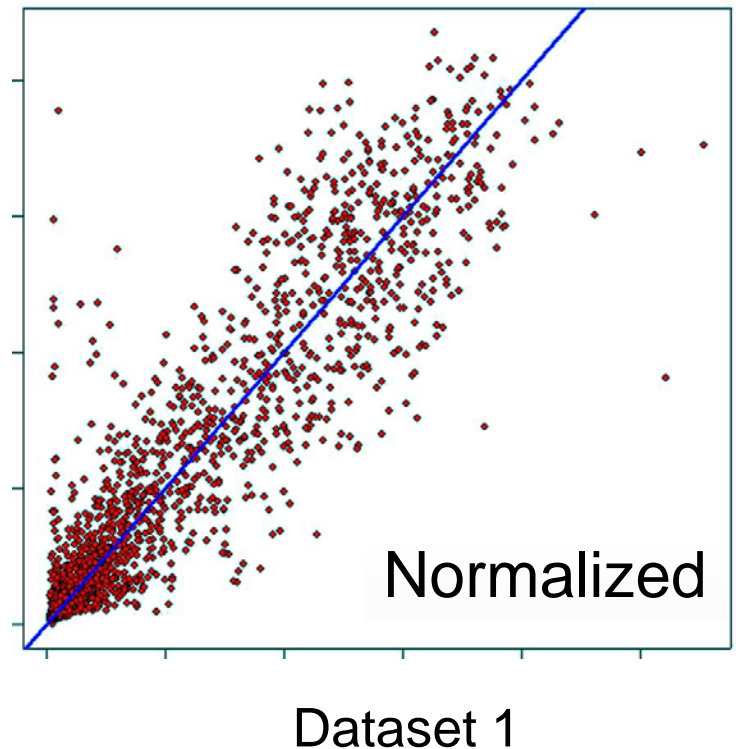# Goals of a downstream tool in Proteomics

- Identify problematic datasets
- Normalize
  - Remove systematic bias and variation due to technical artifacts
- Rolling up to proteins

IPI:IPI00003269.1

Raw

Scaled
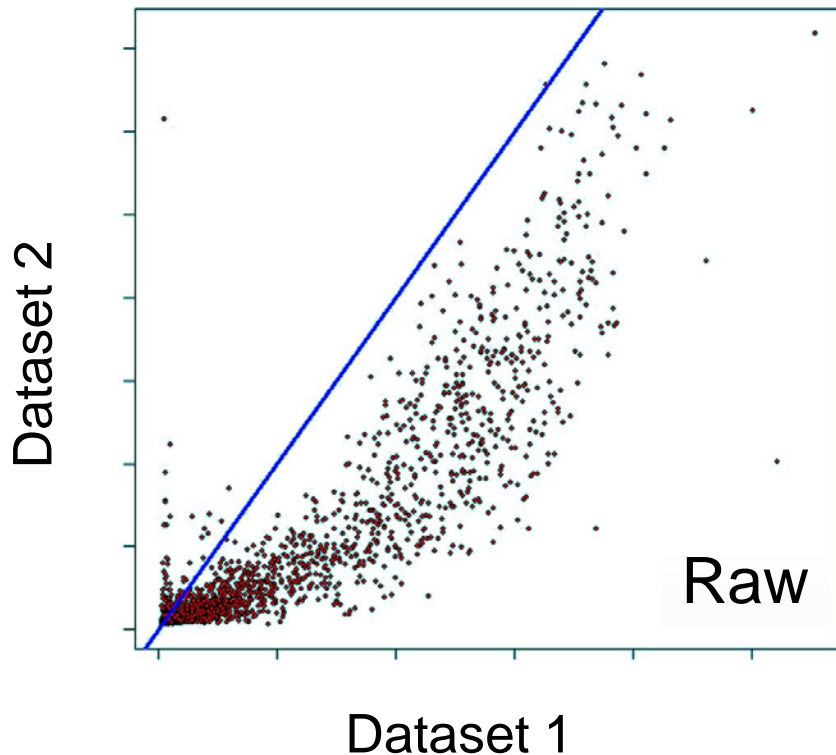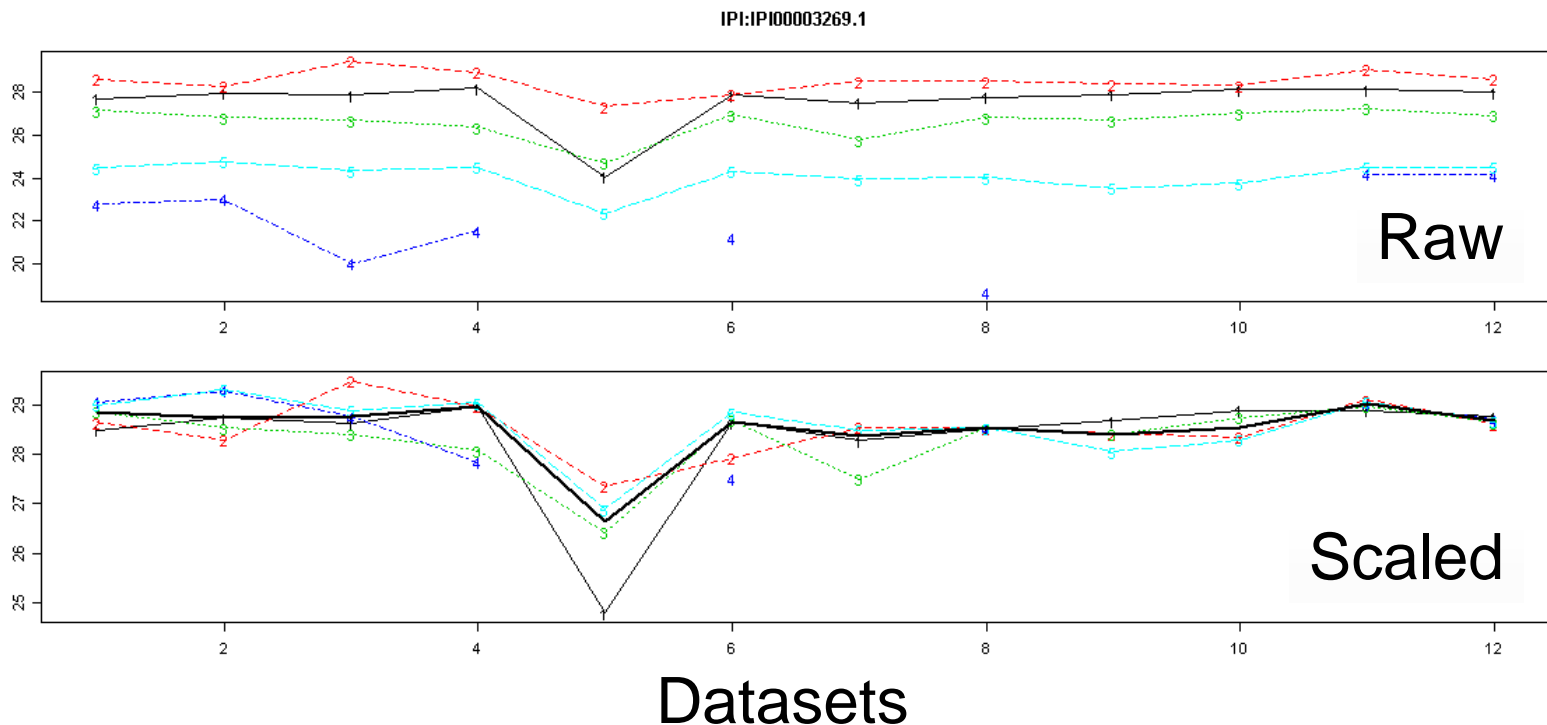
Datasets

Example dataset

- 3 Burn (human) samples and 3 Control samples.

- Each sample was run in duplicates, therefore 12 datasets.

Burn

Sham (control)
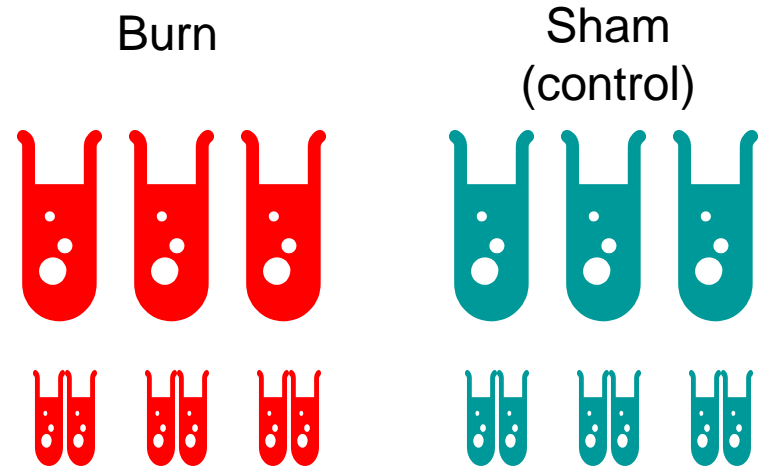
Example dataset

- Group datasets using "Factors"
  - Gender
  - Sample type
  - Technical replicate
  - Biological Replicate

- Factors for Burn data
  - Condition:
    Burn / Sham
  - Replicates:
    1,1,2,2,3,3,4,4,5,5,6,6

Burn

Sham (control)

1 1 2 2 3 3   4 4 5 5 6 6

6   6
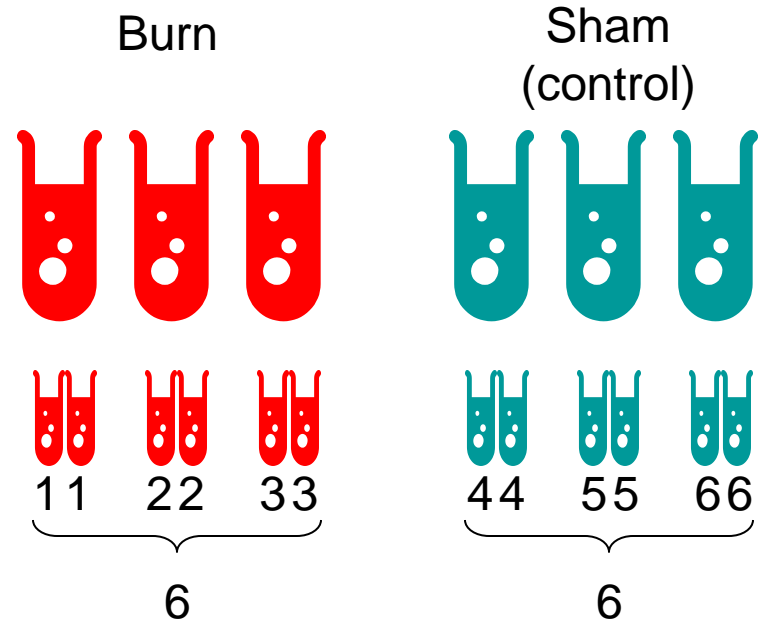
Outline of the Analysis of Data

- Load data
- Initial diagnosis with plots
- Define factors
- Normalize
  - Within a Factor
    - Linear regression
    - Loess
    - Quantile
  - Global
    - MAD
    - Mean Centering
- Rollup
- ANOVA

Data loading example
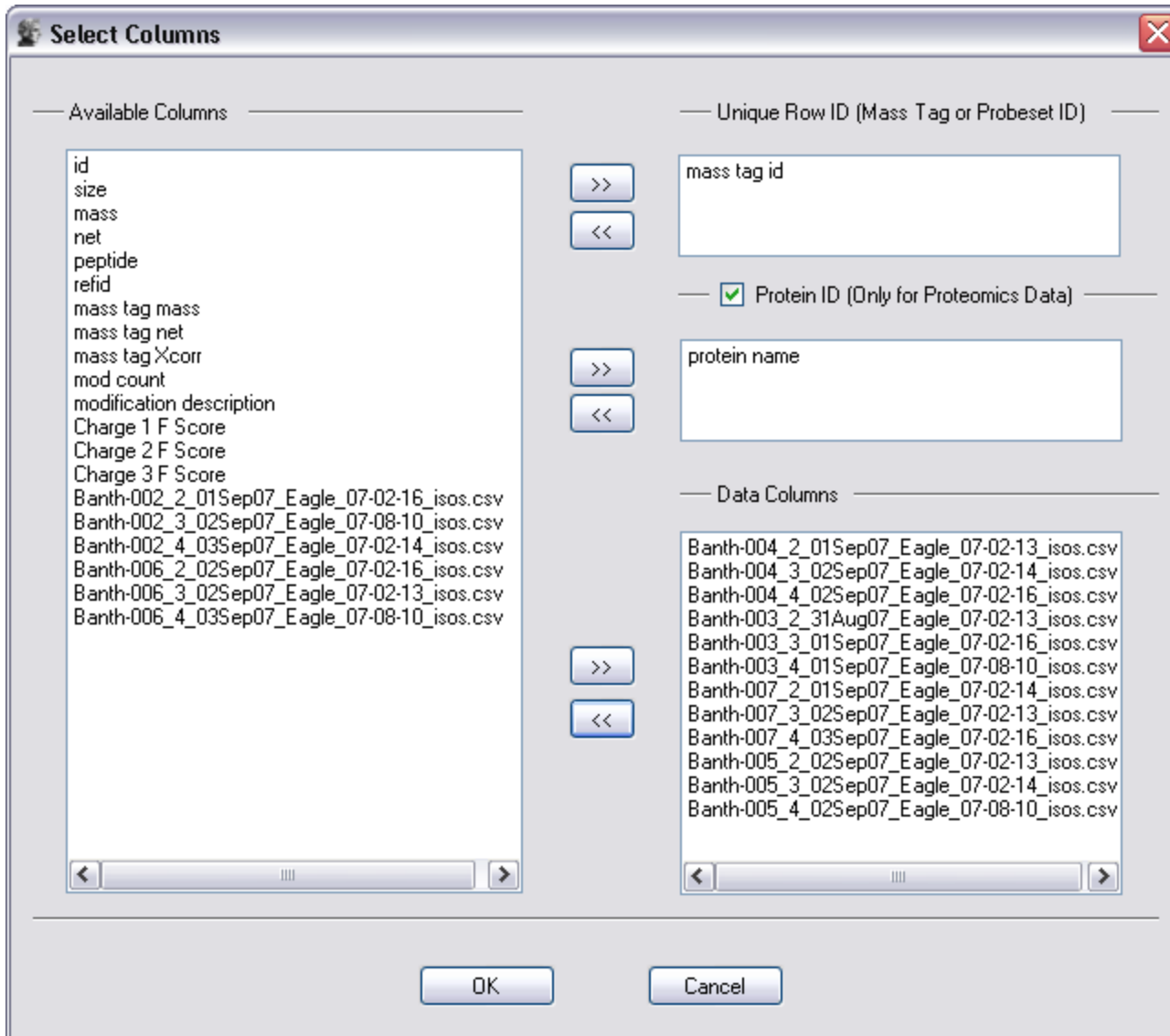
Correlations

and Histogram

Color legend with
overlaid histogram
of correlation
values

Outlier dataset

Dataset Names

Dataset Names
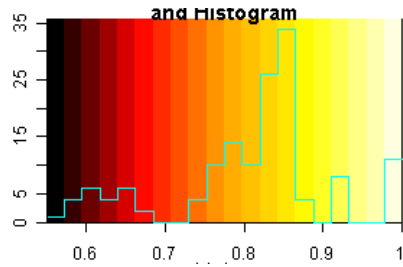
Burn_Wash_1_2
Burn_Wash_1_
Burn_Wash_2_2
Burn_Wash_2
Burn_Wash_3_2
Burn_Wash_3
Sham_Wash_1_2
Sham_Wash_1
Sham_Wash_2_2
Sham_Wash_2
Sham_Wash_3_2
Sham_Wash_3

Burn_Wash_1_2
Burn_Wash_1_
Burn_Wash_2_2
Burn_Wash_2
Burn_Wash_3_2
Burn_Wash_3
Sham_Wash_1_2
Sham_Wash_1
Sham_Wash_2_2
Sham_Wash_2
Sham_Wash_3_2
Sham_Wash_3
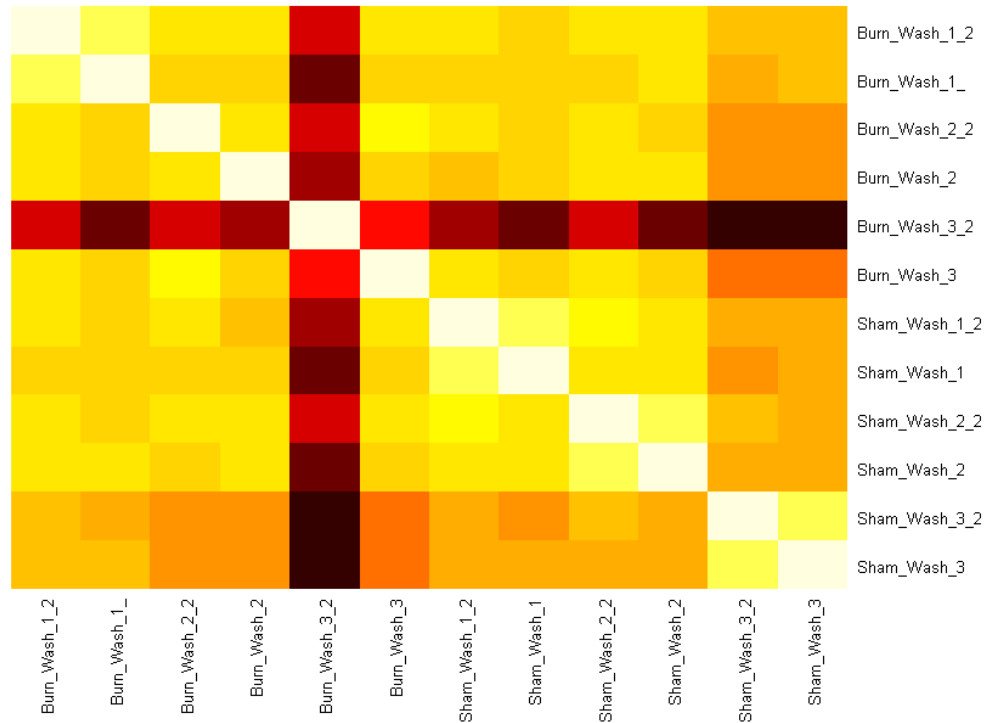
Normalizing - Box Plot Views
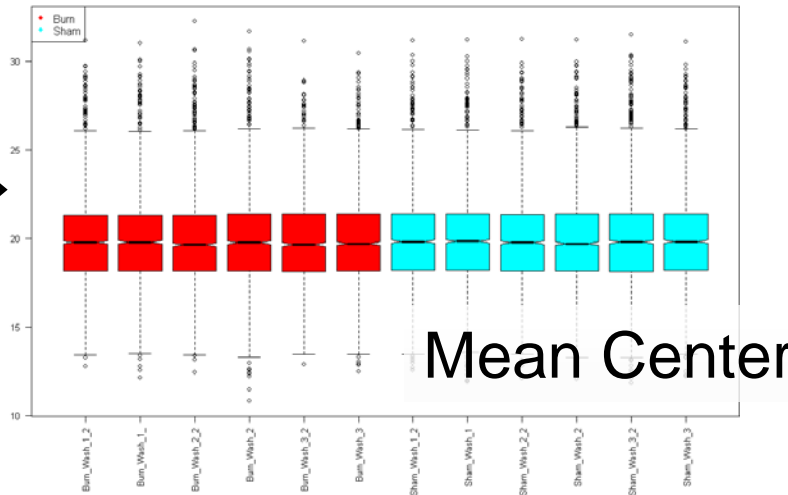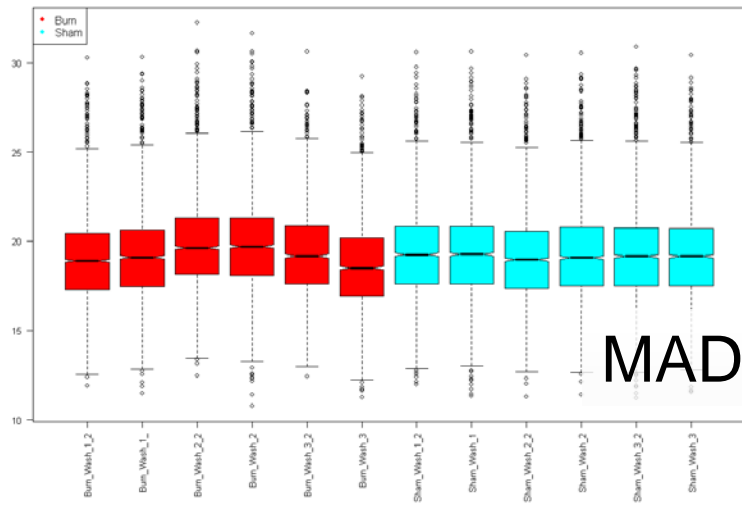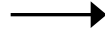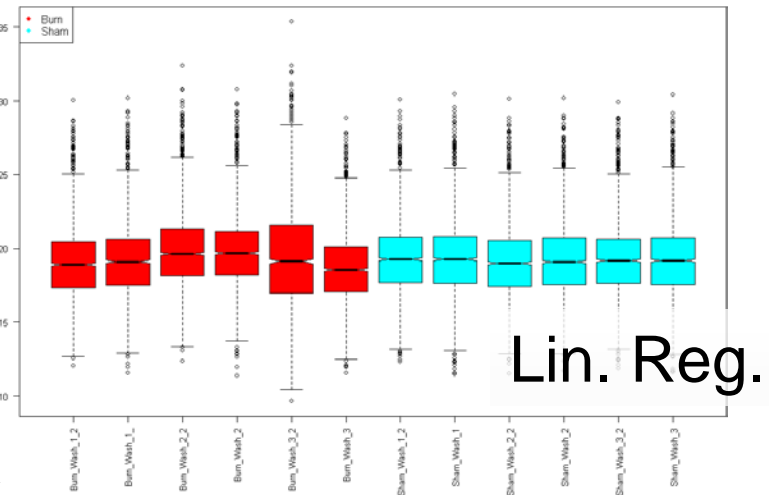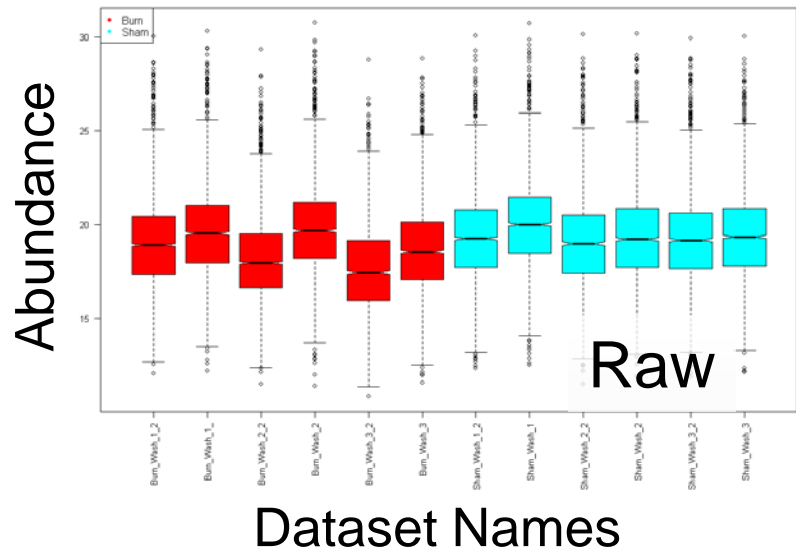
Diagnostic plots for Linear Regression

Regressing one dataset vs. second dataset
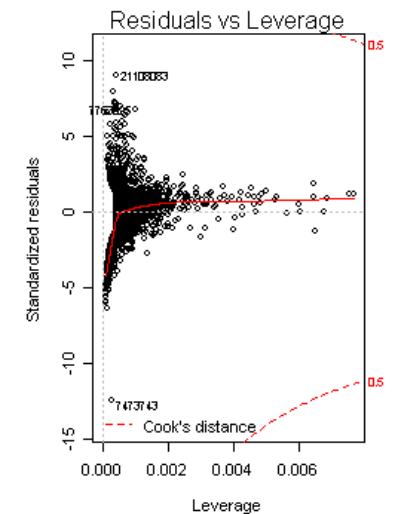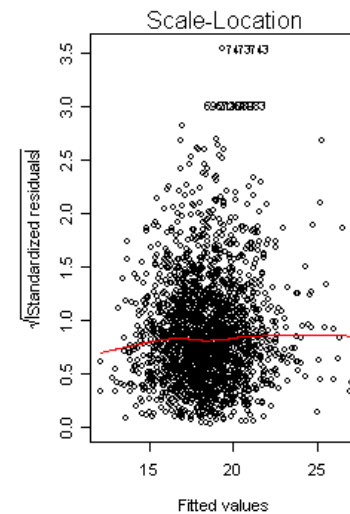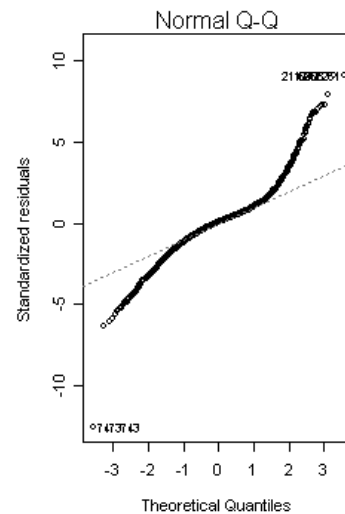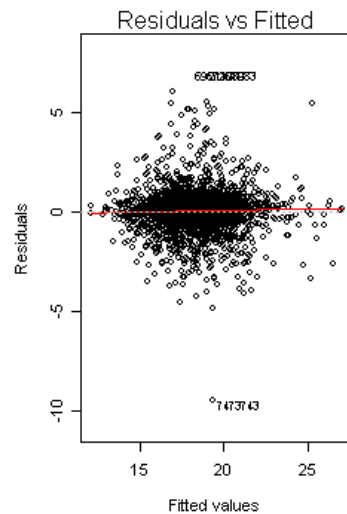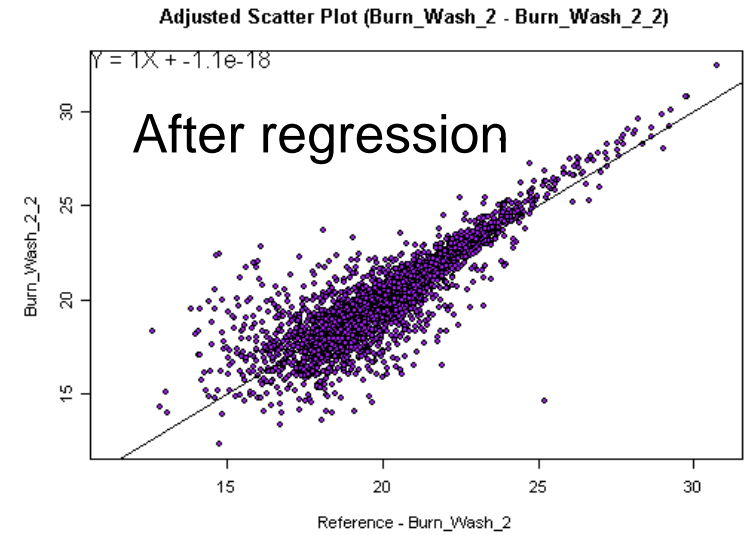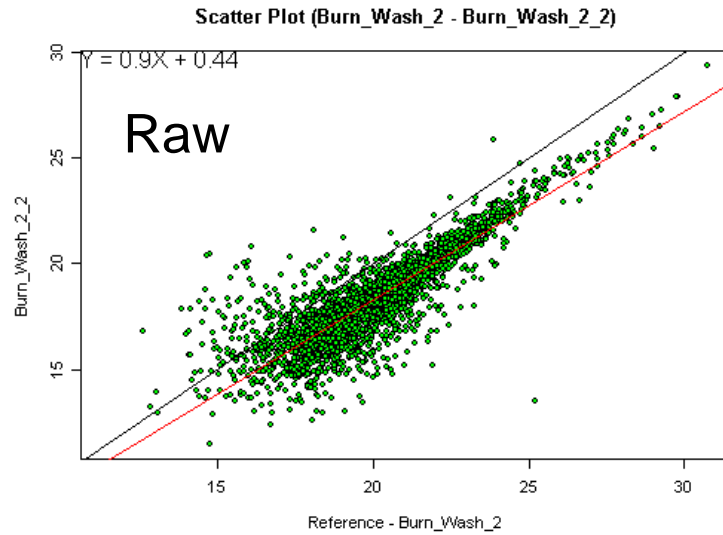
Rolling Up Peptides to Protein Abundance

IPI:IPI00003269.1

Burn   Sham
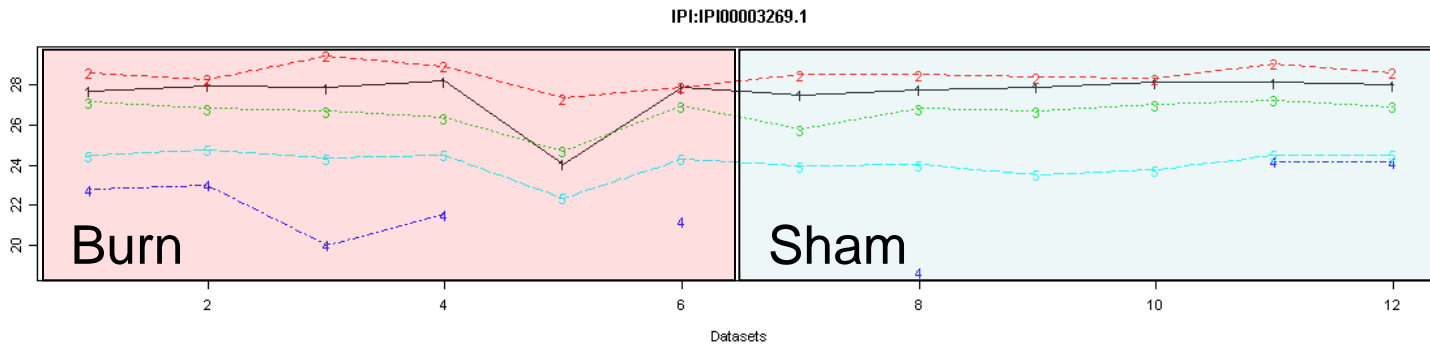
Raw peptide abundances vs. dataset (for 1 protein)

Burn   Sham
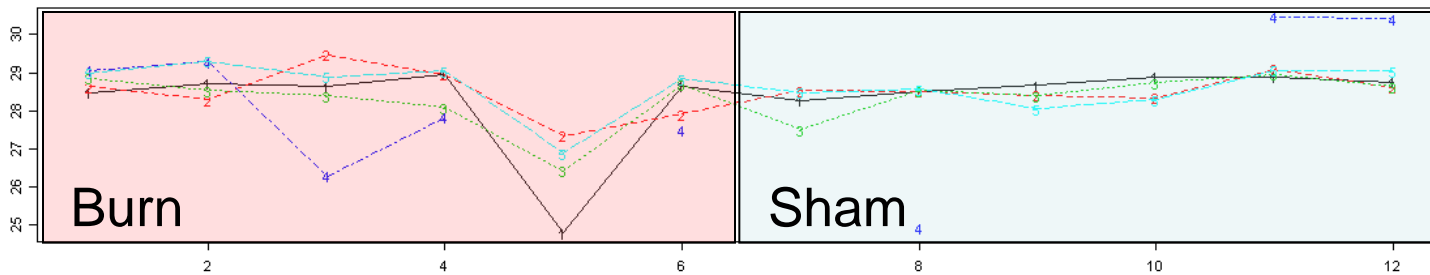
Scaled peptide abundances for this protein's 5 peptides

Burn   Sham
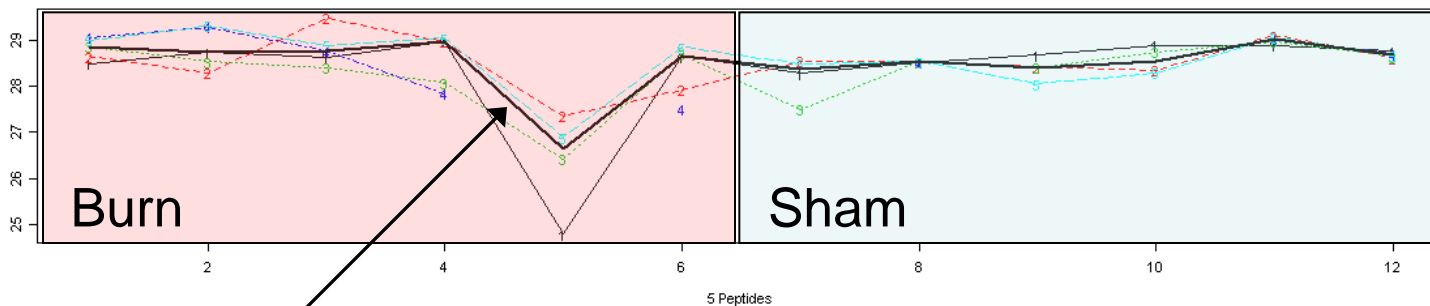
Scaled abun., outliers removed with Grubb's test

Median protein abundance (dark black line)

Heatmaps of Protein Abundance
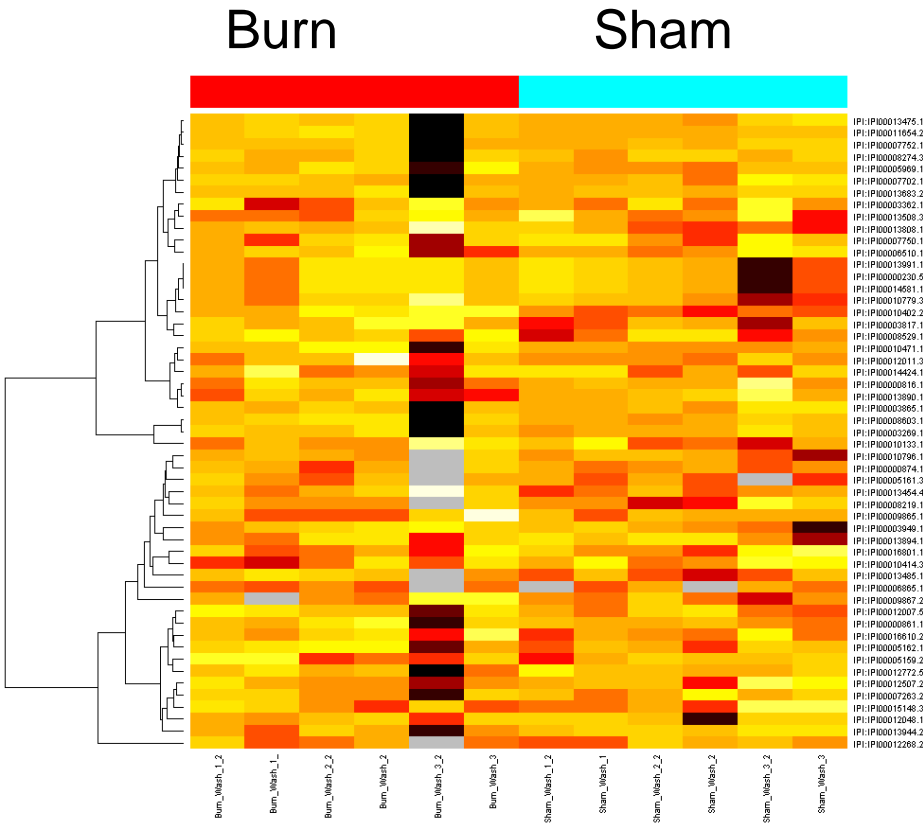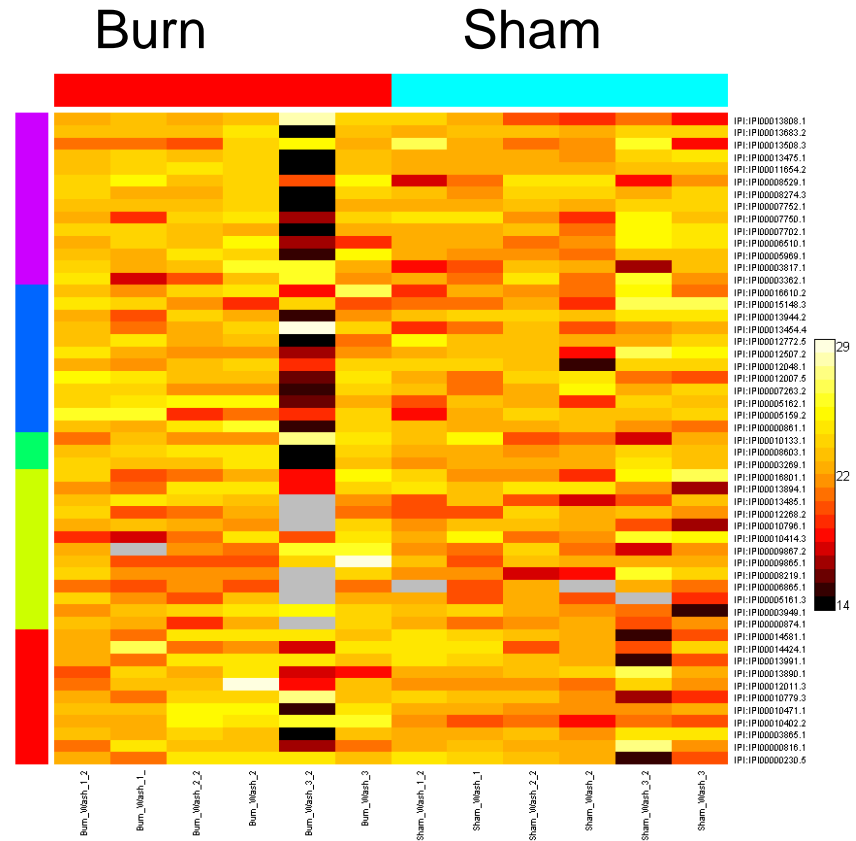
Complete InfernoRDN Feature List

- Data loading with peptide-protein group information
- Log transform
- Factor Definitions
- Normalization
  - Linear Regression
  - Loess
  - Quantile normalization
  - Median Absolute Deviation (MAD) Adj.
  - Mean Centering
- Missing Value Imputation
  - Simple
    - mean/median of the sample
    - Substitute a constant
  - Advance
    - Row mean within a factor
    - kNN method
    - SVDimpute
- Save tables / factors / session

Complete InfernoRDN Feature List

- Plots
  - Histograms
  - Boxplots
  - Correlation plots
  - MA plots
  - PCA/PLS plots
  - Protein rollup plots
  - Heatmaps
- Rolling up to Proteins
  - Reference peptide based scaling (RRollup)
  - Z-score averaging (ZRollup)
  - QRollup
- Statistics
  - ANOVA
    - Simple 1-way
    - N-Way (provisions for unbalanced data)
    - Random effects (multi level) models (REML)
  - Q-values
  - Filters
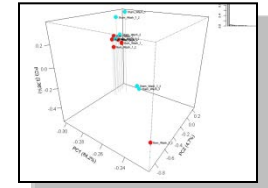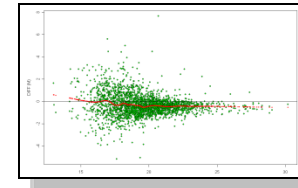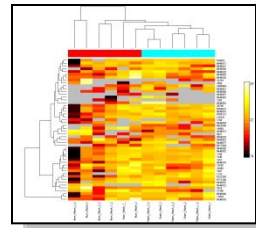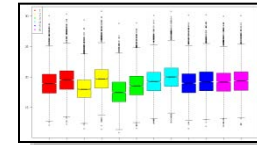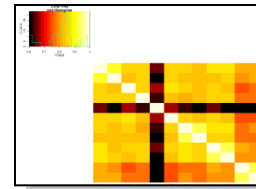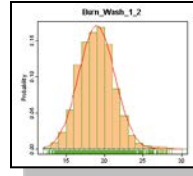
http://omics.pnl.gov/software/InfernoRDN
\\floyd\Software\Inferno

To be added…

- Tests for normality, Nonparametric tests, post-hoc tests
- Incorporating an interactive heatmap control
- SMART-AMT, peptide prophet
- Protein Quality metrics
- Improve rollup methods to cluster and differentiate protein isoforms
- Alan Dabney's work
- Network algorithms / Cytoscape

Acknowledgements

- Weijun Qian
- Deep Jaitly
- Vlad Petyuk
- Josh Adkins
- Tom Metz
- Stephen Callister
- Brian LaMarche
- Ken Auberry
- Matt Monroe
- and the rest of the informatics group

- Joel Pounds
- Susan Varnum
- Bobbie-Jo Webb-Robertson

- Active users!
  - Kim Hixson
  - Josh Turse
  - Charles Ansong
  - Feng Yang
  - Bryan Ham
  - Christina Sorensen
  - Angela Norbeck
  - Sam Purvine
  - Nathan Manes
  - Jon Jacobs

Gordon Anderson

Dick Smith

… and the group.