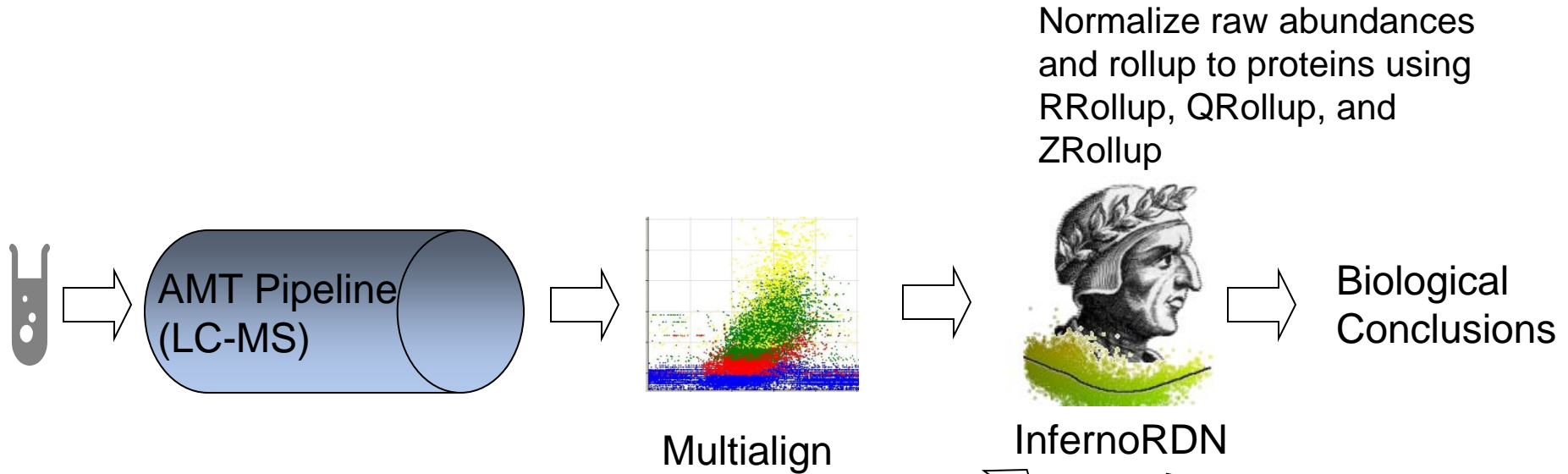


InfernoRDN

(previously DAnTE)

Ashoka Polpitiya

Where does InfernoRDN fit in?



	A	B	C	D	E
1	protein name	mass tag id	Banth-002_2	(Banth-002_3_02	Banth-002_4_03Sej
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	P008IMYG_HORSE	11081	133331434	150507990	94626884
13	BAS0253	232087532	127773825	173439069	107476695

QRollup

Other CSV files (microarray)

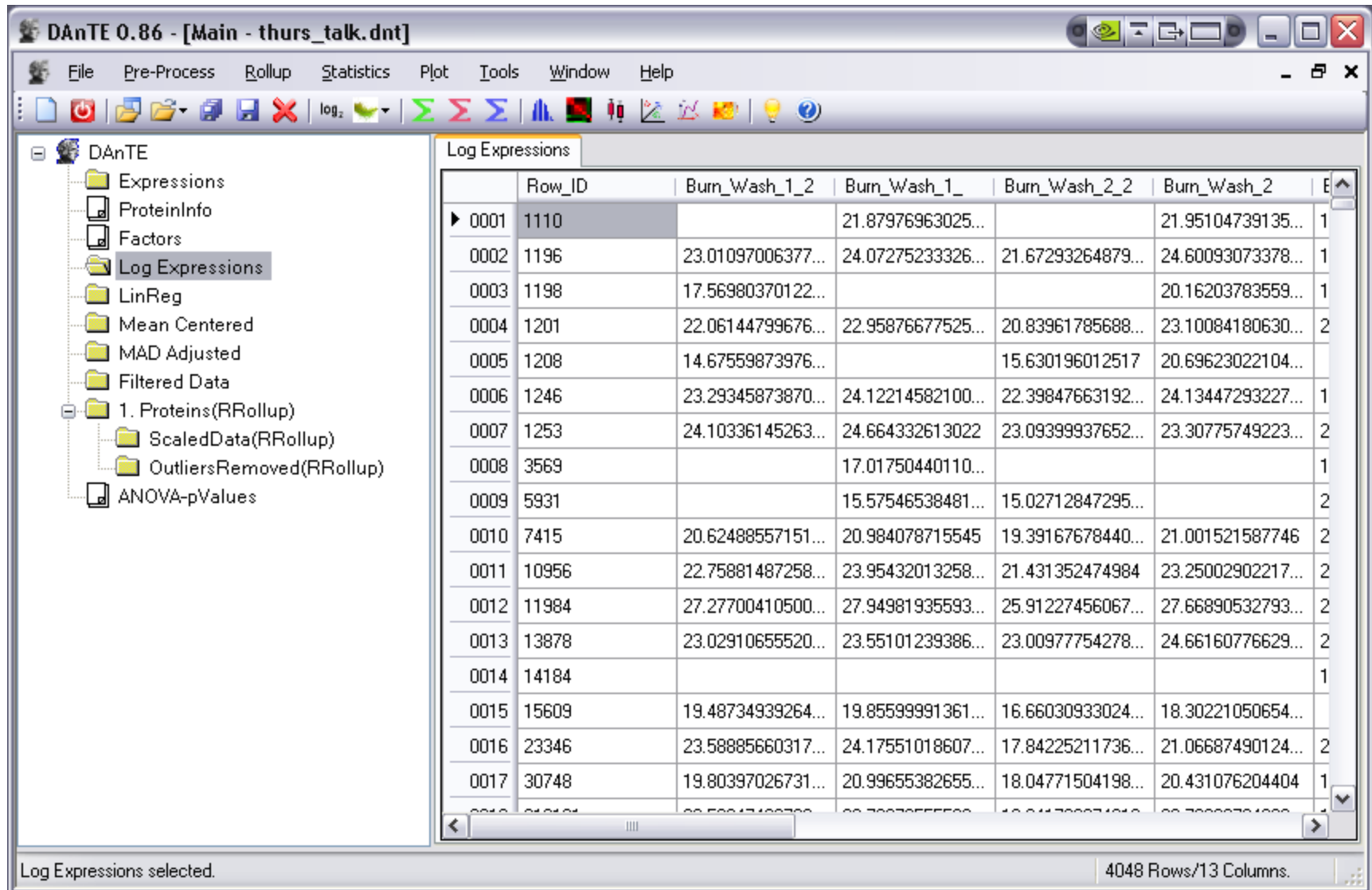


InfernoRDN

The screenshot displays the InfernoRDN software interface. The main window title is "InfernoRDN - [Main - bioinformatics_supplement.dnt]". The menu bar includes File, Data, Plot, Proteins, Grouping, Compare, Explore, Tools, Window, and Help. The left sidebar shows a tree view with the following items: Inferno, Expressions, p-Values, Protein Info, Factors, Log Expressions, Filtered Data, Mean Centered, **RRollup** (selected), OutliersRemoved, and ScaledData. The main area shows a table titled "RRollup" with the following columns: Protein, PepCount, A1_run1, A1_run2, A2_run1, A2_run2, A3_run1, and a scroll arrow. The table contains 26 rows of data, with the first row (001) highlighted. The status bar at the bottom indicates "RRollup selected." and "448 Rows/14 Columns."

	Protein	PepCount	A1_run1	A1_run2	A2_run1	A2_run2	A3_run1
▶ 001	IP:IP0000023...	6	23.589333917...	23.308916520...	23.808591173...	23.772222559...	23.88833568
002	IP:IP0000081...	9	24.004968082...	24.337455946...	23.974373502...	24.161457305...	23.48958090
003	IP:IP0000086...	4	21.130037596...	20.776856627...	21.414246384...	21.764591817...	18.43321481
004	IP:IP0000087...	2	19.575598967...	19.403337385...	18.495120098...	19.459416859...	
005	IP:IP0000326...	5	28.749813159...	28.516128714...	27.940977814...	28.470915635...	26.46727996
006	IP:IP0000336...	16	22.930962200...	22.371827041...	22.589313837...	22.905366926...	23.13381013
007	IP:IP0000381...	4	23.243145918...	22.961854695...	22.998190600...	23.727499261...	23.64507057
008	IP:IP0000386...	23	24.385509965...	24.120007843...	24.248738934...	24.215698391...	22.54774317
009	IP:IP0000394...	2	19.075790726...	19.745908867...	20.509124470...	20.638087078...	21.47162923
010	IP:IP0000515...	4	21.190575746...	21.202763036...	19.868544140...	20.204063545...	19.88294013
011	IP:IP0000516...	2	20.474623180...	19.396842178...	18.764439740...	19.974999010...	
012	IP:IP0000516...	2	21.297606438...	21.312656830...	21.492895387...	21.539715265...	18.93936469
013	IP:IP0000596...	2	23.382748773...	23.117167789...	23.516249898...	23.417977755...	21.56372487
014	IP:IP0000651...	2	22.377489168...	22.767199588...	22.395326116...	23.109935685...	20.59108145
015	IP:IP0000686...	2	17.561745688...	17.060509658...	17.888613505...	16.935163108...	
016	IP:IP0000726...	3	21.743791941...	21.745081062...	21.358536754...	21.371665287...	20.62276847
017	IP:IP0000770...	10	23.655277293...	23.687463495...	23.254415256...	23.336652711...	21.79386622
018	IP:IP0000775...	11	22.548051989...	21.840988269...	22.637373860...	22.895160899...	21.40210703
019	IP:IP0000775...	11	23.290664732...	23.347349238...	23.252506178...	23.429840905...	21.06929061
020	IP:IP0000821...	2	20.673991758...	19.639534656...	19.808830042...	19.733962840...	
021	IP:IP0000827...	5	23.555178189...	23.113063122...	23.036518432...	23.479617000...	21.33300087
022	IP:IP0000852...	3	23.428124485...	23.784980817...	23.553251496...	23.284080529...	22.36696173
023	IP:IP0000860...	12	28.599741032...	28.648238069...	28.297479286...	28.492998781...	26.78820618
024	IP:IP0000986...	5	19.983794065...	18.462256998...	18.664647152...	18.630630552...	20.41381082
025	IP:IP0000986...	2	19.711975866...		19.067197542...	18.307396113...	23.34367179
026	IP:IP0001013...	5	26.289455636...	26.484587503...	26.218479710...	26.238076575...	26.79389134

InfernoRDN is simply DAnTE re-branded

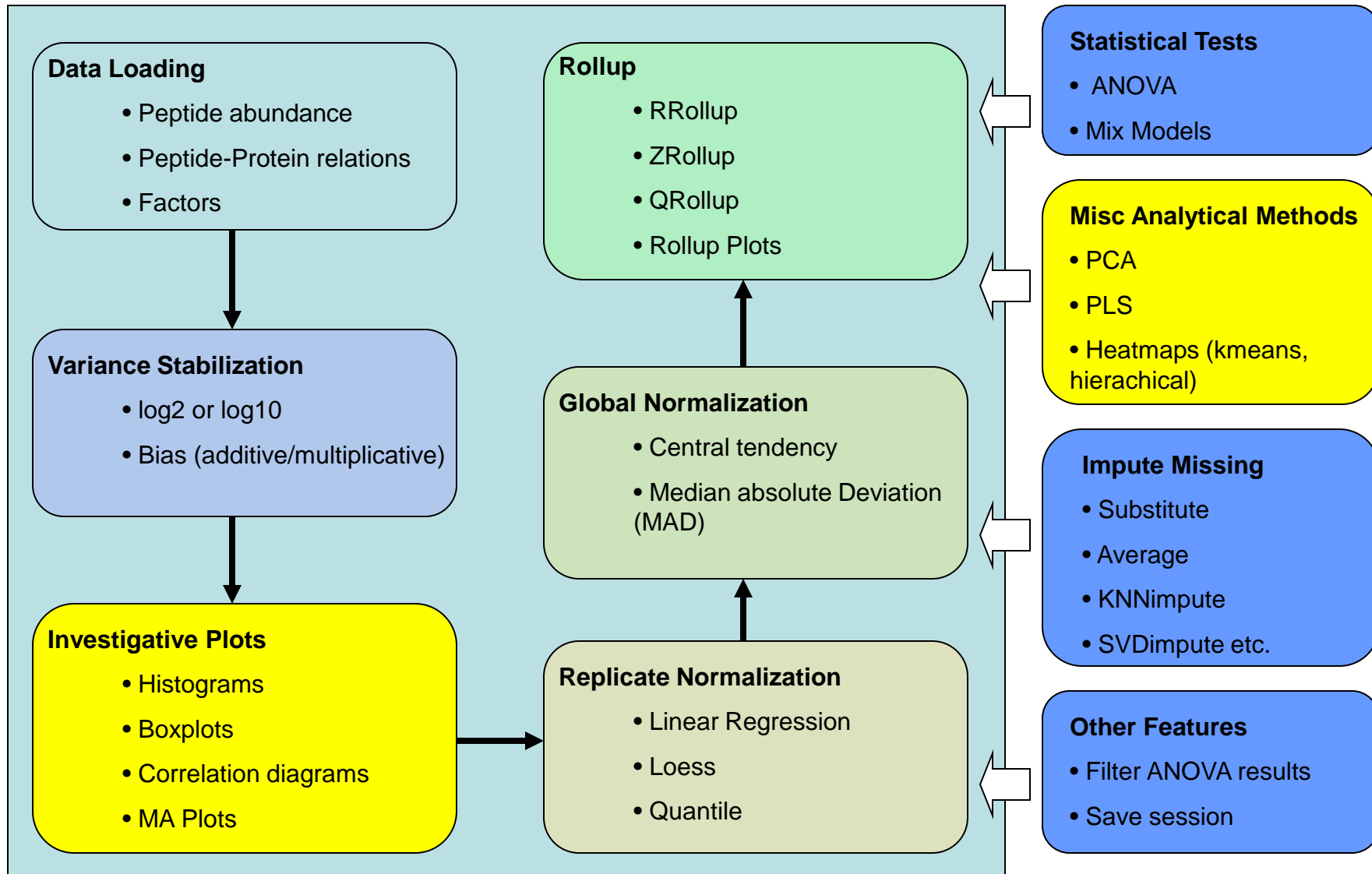


The screenshot displays the DAnTE 0.86 software interface. The window title is "DAnTE 0.86 - [Main - thurs_talk.dnt]". The menu bar includes File, Pre-Process, Rollup, Statistics, Plot, Tools, Window, and Help. The toolbar contains icons for file operations, mathematical functions, and data visualization. The left sidebar shows a tree view of the workspace with folders for Expressions, ProteinInfo, Factors, Log Expressions (selected), LinReg, Mean Centered, MAD Adjusted, Filtered Data, 1. Proteins(RRollup), ScaledData(RRollup), OutliersRemoved(RRollup), and ANOVA-pValues. The main area shows a table titled "Log Expressions" with columns: Row_ID, Burn_Wash_1_2, Burn_Wash_1_, Burn_Wash_2_2, Burn_Wash_2, and E. The table contains 17 rows of data, with the first row (0001) highlighted. The status bar at the bottom indicates "Log Expressions selected." and "4048 Rows/13 Columns."

	Row_ID	Burn_Wash_1_2	Burn_Wash_1_	Burn_Wash_2_2	Burn_Wash_2	E
▶ 0001	1110		21.87976963025...		21.95104739135...	1
0002	1196	23.01097006377...	24.07275233326...	21.67293264879...	24.60093073378...	1
0003	1198	17.56980370122...			20.16203783559...	1
0004	1201	22.06144799676...	22.95876677525...	20.83961785688...	23.10084180630...	2
0005	1208	14.67559873976...		15.630196012517	20.69623022104...	
0006	1246	23.29345873870...	24.12214582100...	22.39847663192...	24.13447293227...	1
0007	1253	24.10336145263...	24.664332613022	23.09399937652...	23.30775749223...	2
0008	3569		17.01750440110...			1
0009	5931		15.57546538481...	15.02712847295...		2
0010	7415	20.62488557151...	20.984078715545	19.39167678440...	21.001521587746	2
0011	10956	22.75881487258...	23.95432013258...	21.431352474984	23.25002902217...	2
0012	11984	27.27700410500...	27.94981935593...	25.91227456067...	27.66890532793...	2
0013	13878	23.02910655520...	23.55101239386...	23.00977754278...	24.66160776629...	2
0014	14184					1
0015	15609	19.48734939264...	19.85599991361...	16.66030933024...	18.30221050654...	
0016	23346	23.58885660317...	24.17551018607...	17.84225211736...	21.06687490124...	2
0017	30748	19.80397026731...	20.99655382655...	18.04771504198...	20.431076204404	1

This screenshot shows the workspace after extensive data analysis

Analysis Flow in InfernoRDN

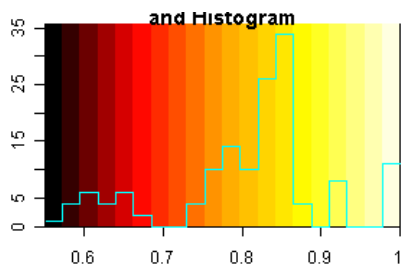


Goals of a downstream tool in Proteomics

- 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)

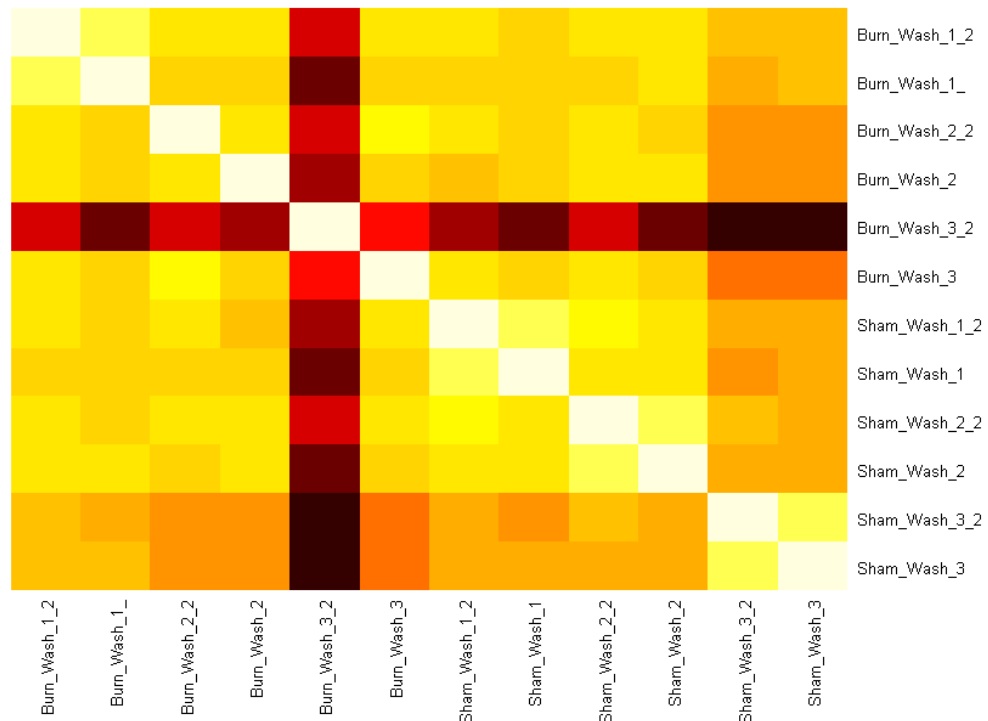
Goals of a downstream tool in Proteomics

- Identify problematic datasets
 - Correlation Plots



Color legend with overlaid histogram of correlation values

Outlier dataset

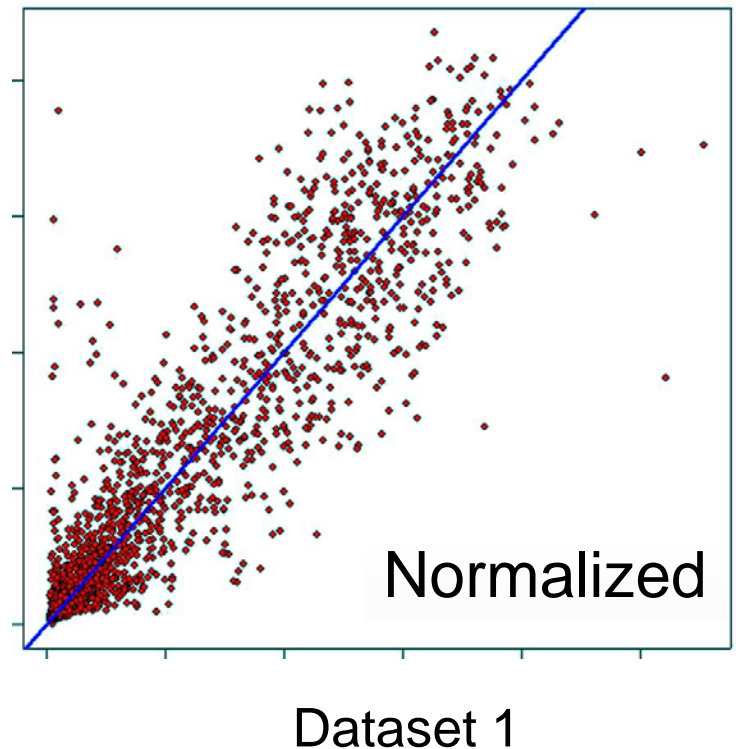
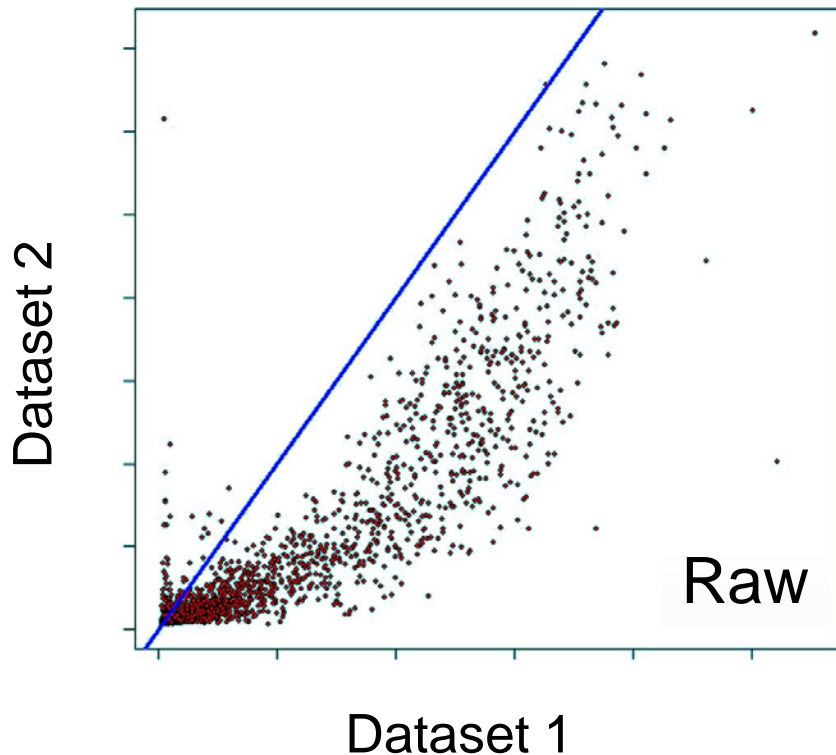


Dataset Names

Dataset Names

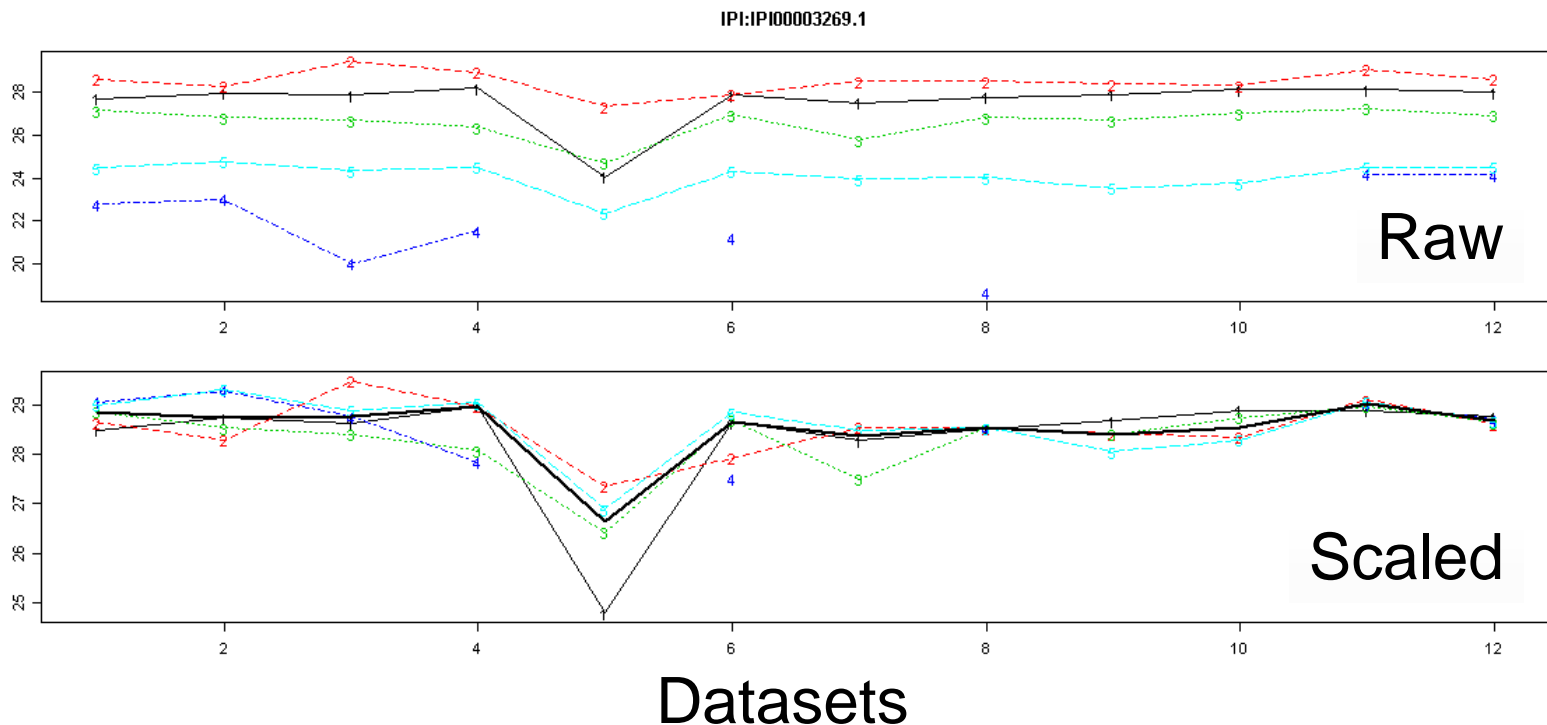
Goals of a downstream tool in Proteomics

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



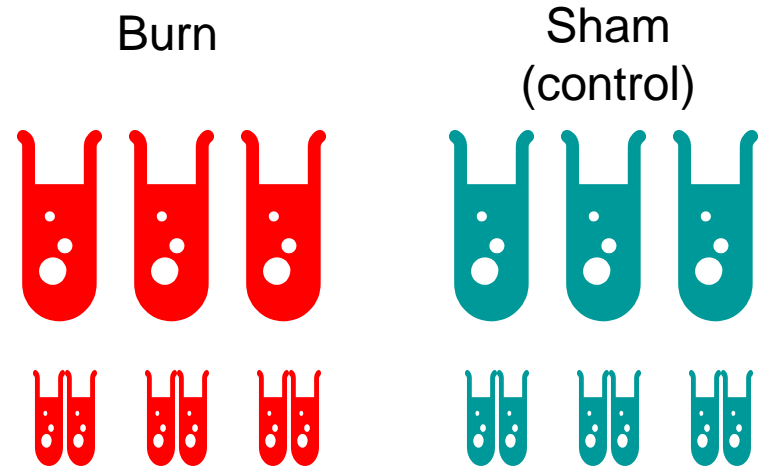
Goals of a downstream tool in Proteomics

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



Example dataset

- 3 Burn (human) samples and 3 Control samples.
- Each sample was run in duplicates, therefore 12 datasets.



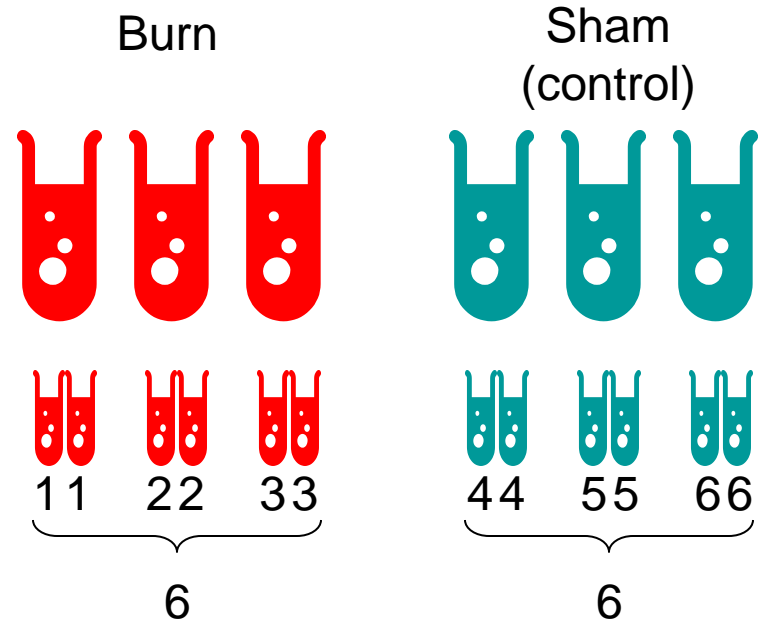
Example dataset

- Group datasets using “Factors”

- Gender
- Sample type
- Technical replicate
- Biological Replicate

- Factors for Burn data

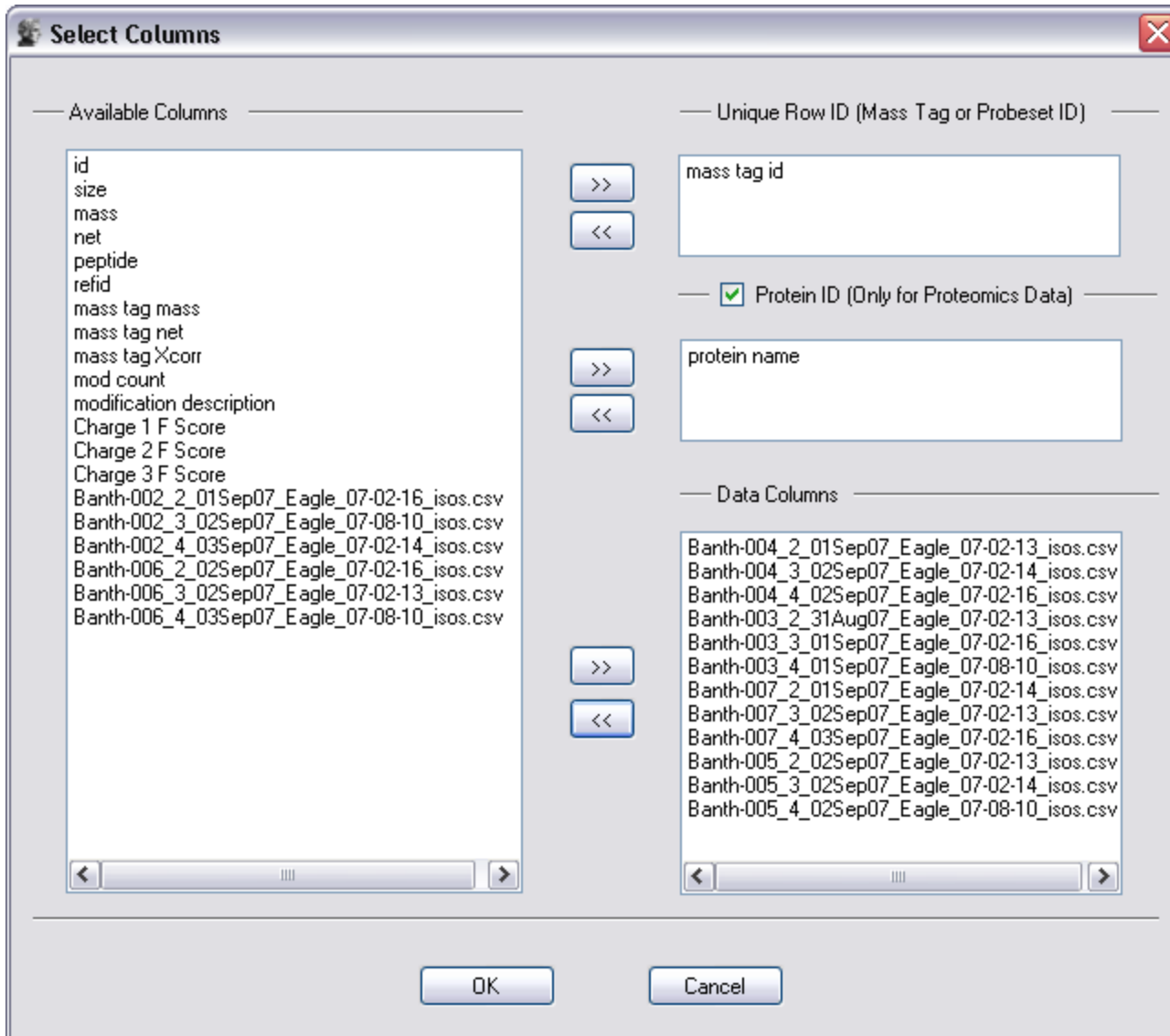
- Condition:
Burn / Sham
- Replicates:
1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6



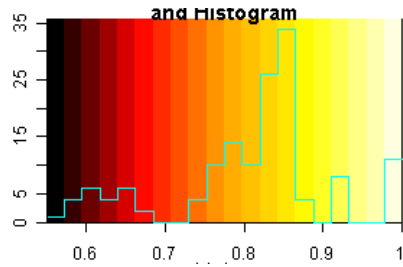
Outline of the Analysis of Data

- 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

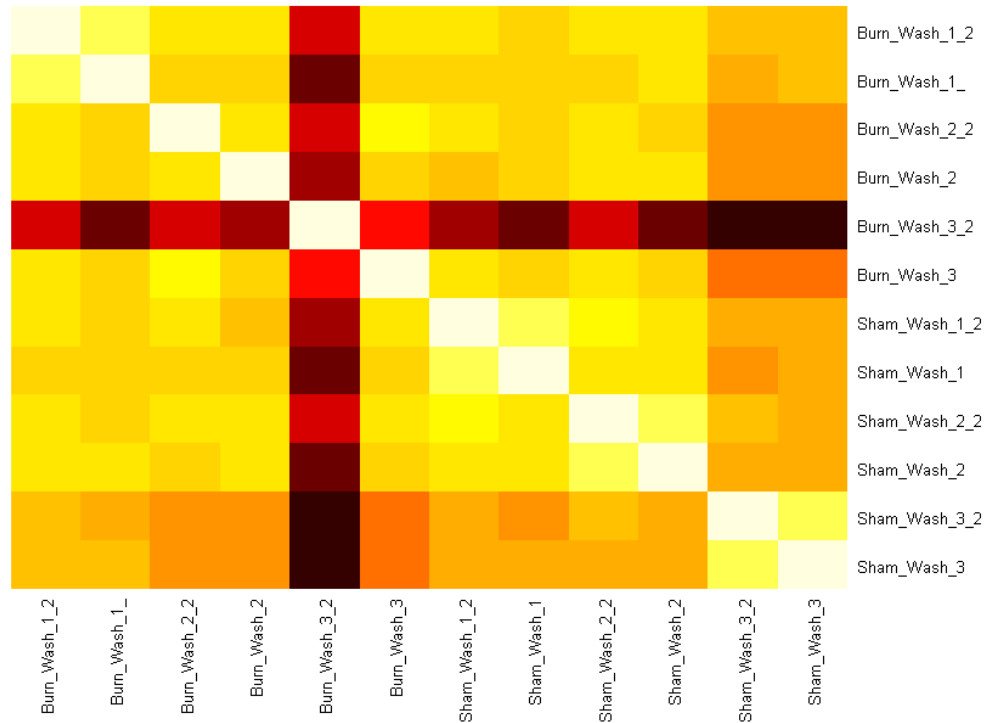


Correlations



Color legend with overlaid histogram of correlation values

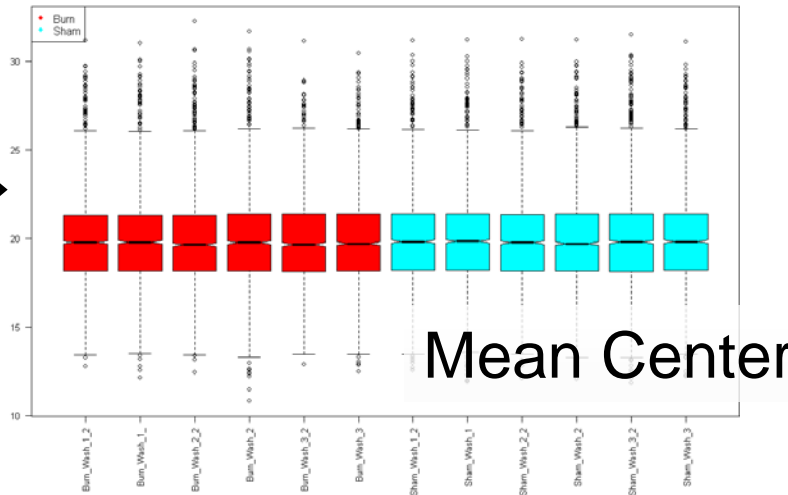
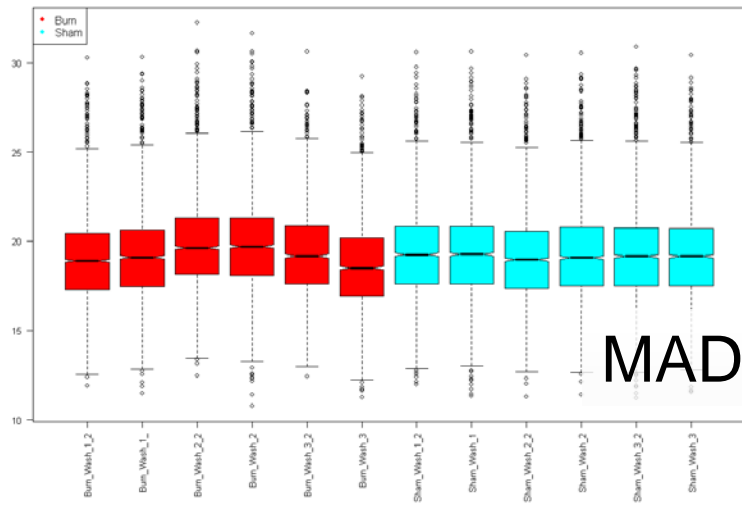
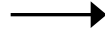
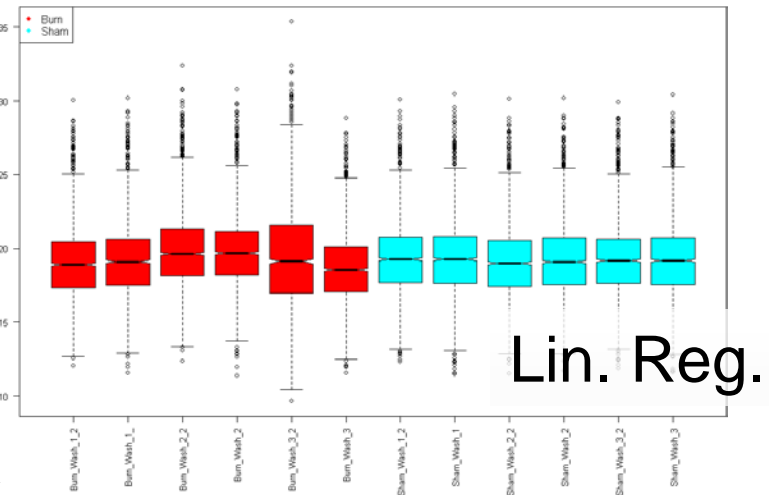
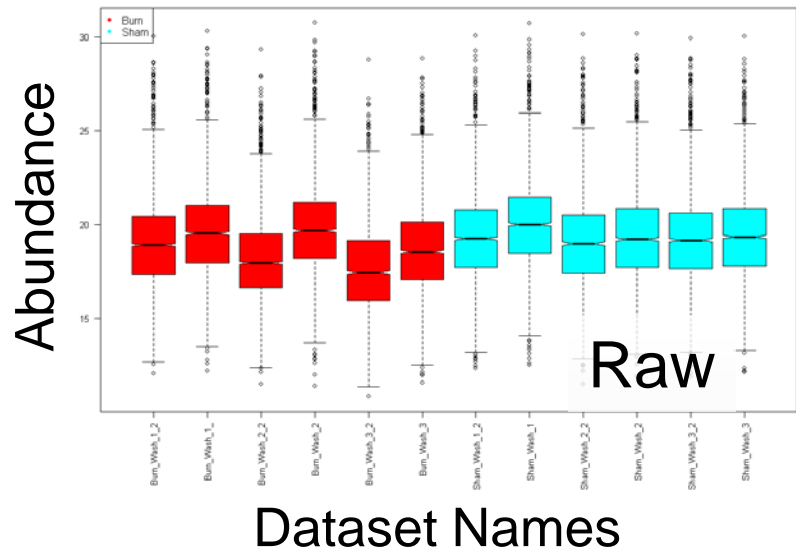
Outlier dataset



Dataset Names

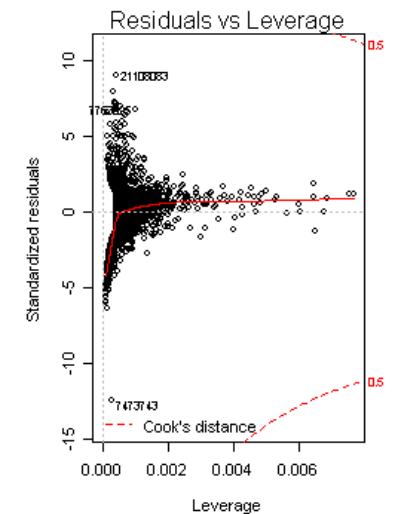
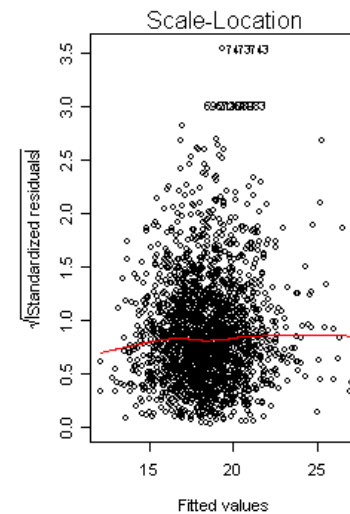
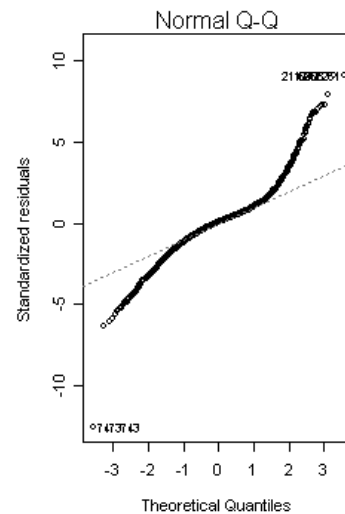
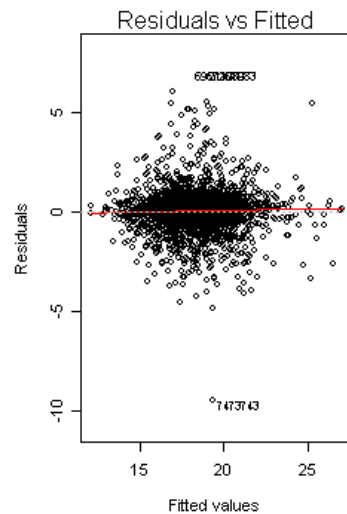
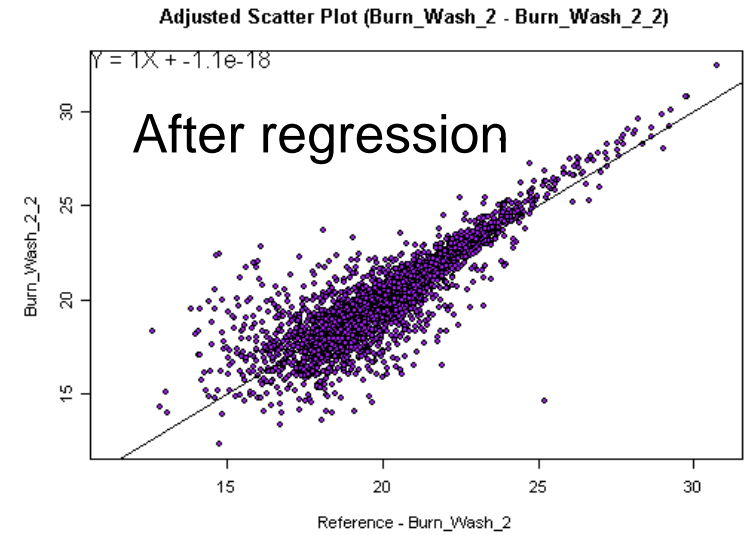
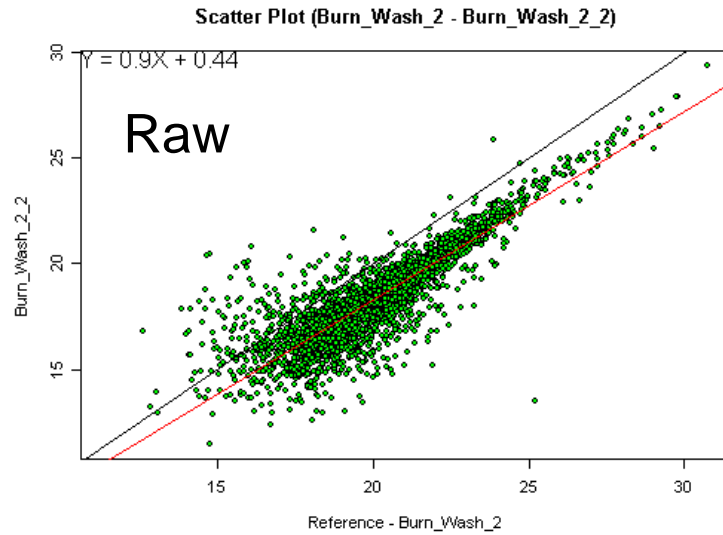
Dataset Names

Normalizing - Box Plot Views

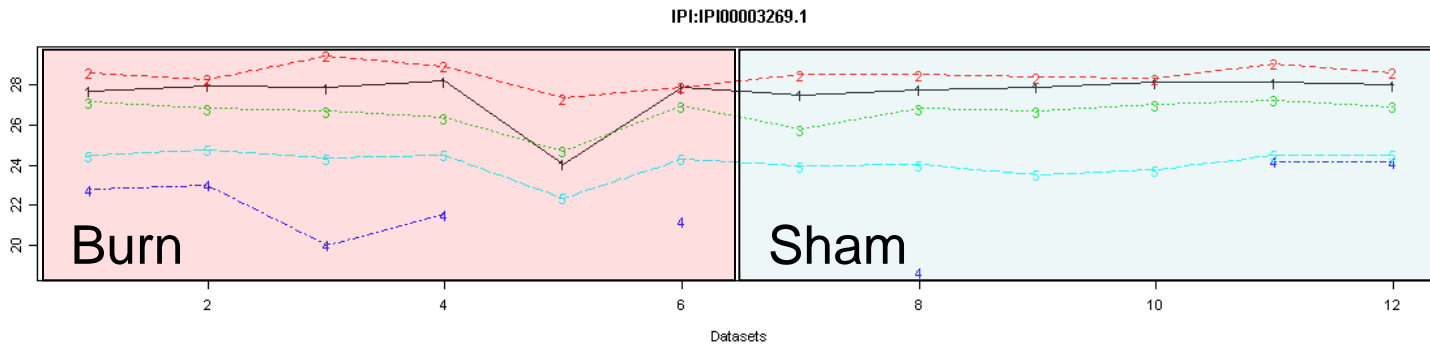


Diagnostic plots for Linear Regression

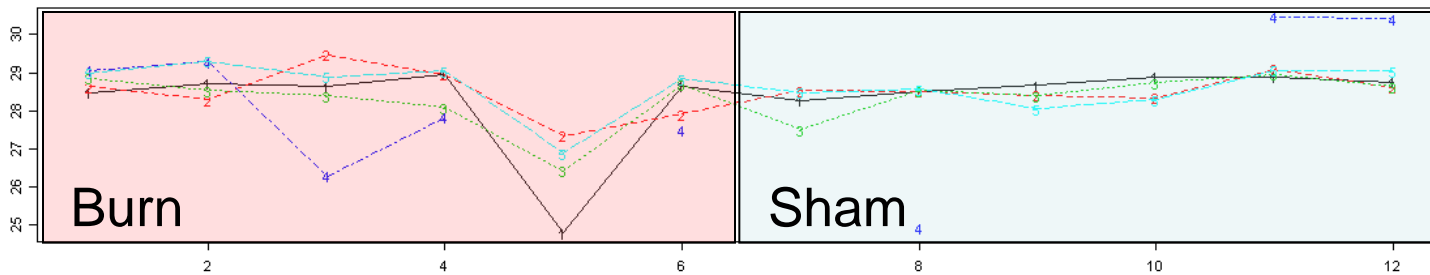
Regressing
one dataset
vs. second
dataset



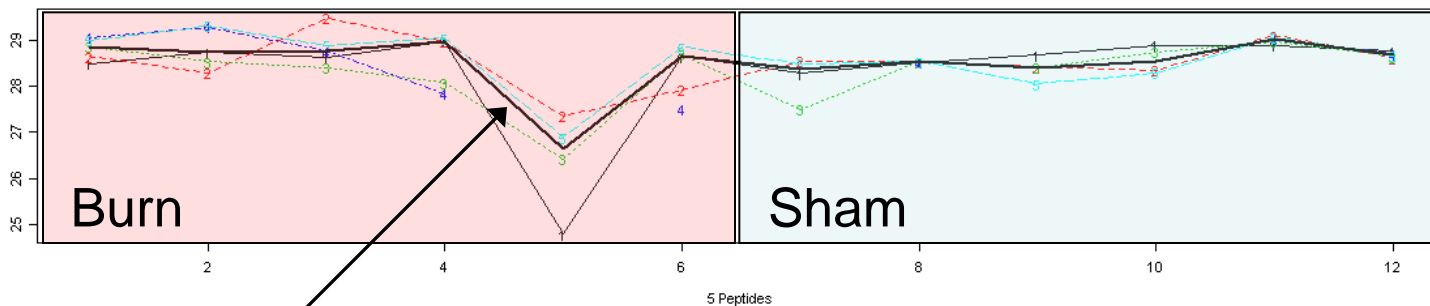
Rolling Up Peptides to Protein Abundance



Raw peptide abundances vs. dataset (for 1 protein)



Scaled peptide abundances for this protein's 5 peptides



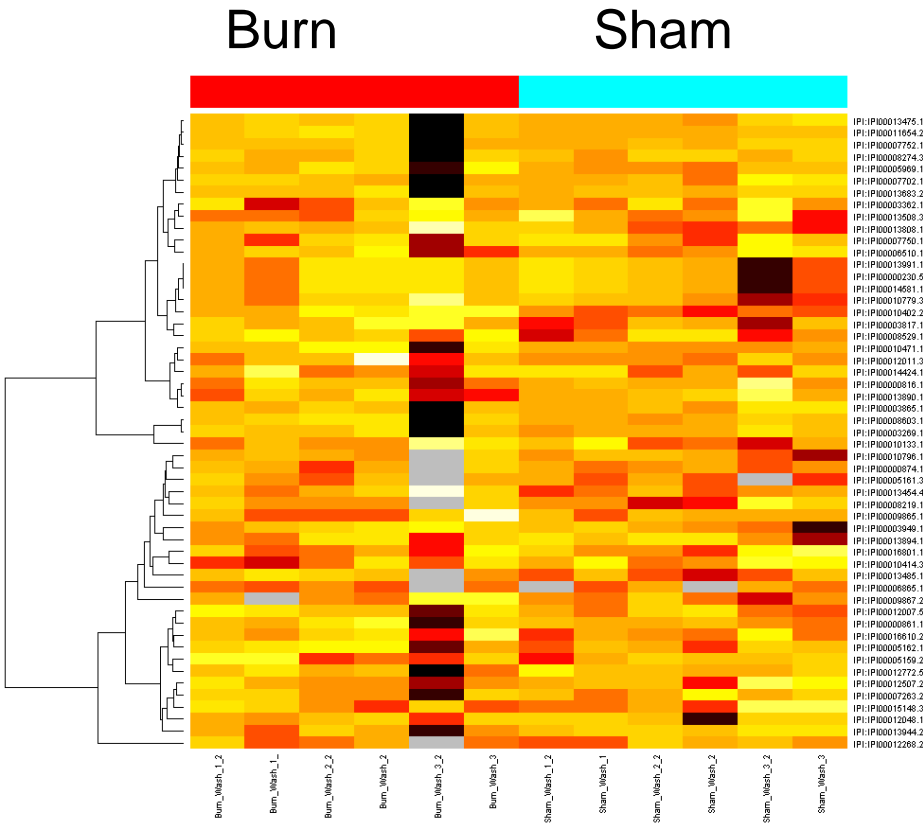
Scaled abund., outliers removed with Grubb's test

Median protein abundance (dark black line)

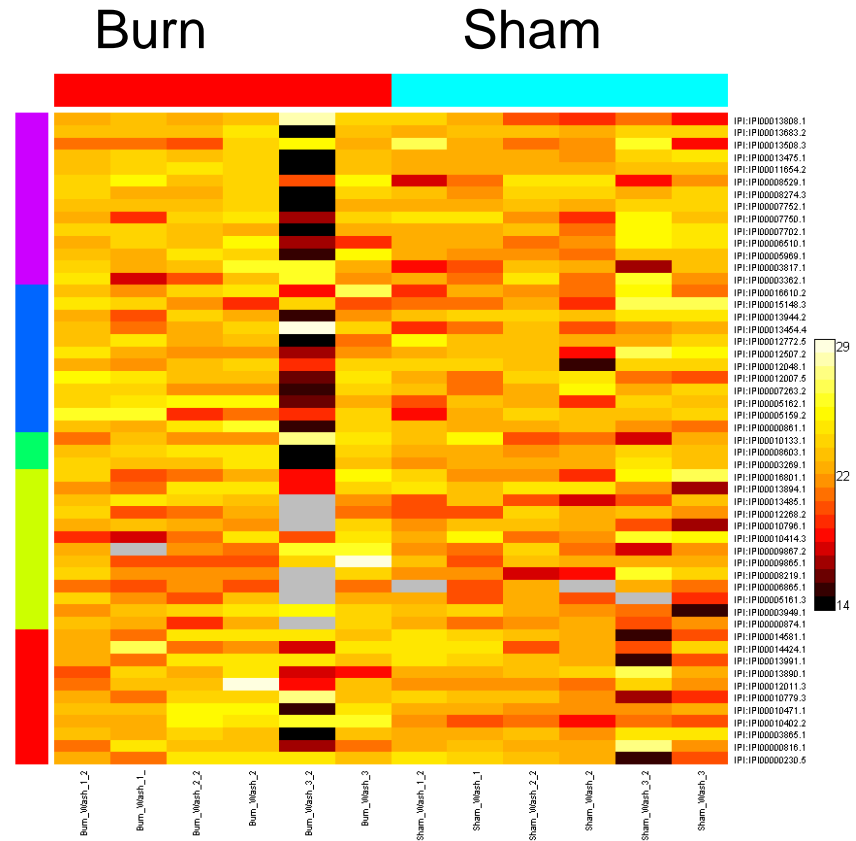
Heatmaps of Protein Abundance

Hierarchical clustering of rows

K-means clustering of rows
(using 5 clusters)



Datasets



Datasets

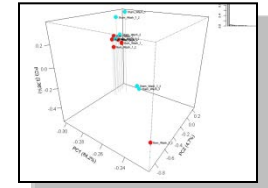
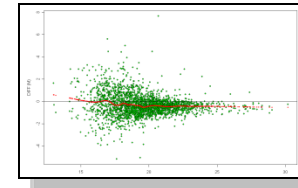
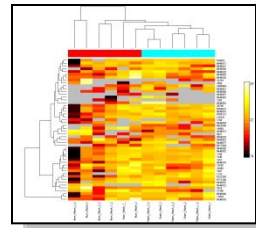
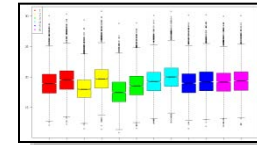
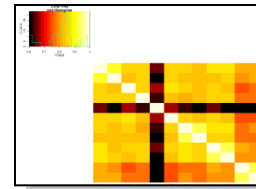
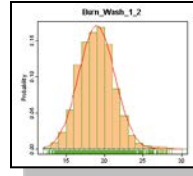
Complete InfernoRDN Feature List

- 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

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.