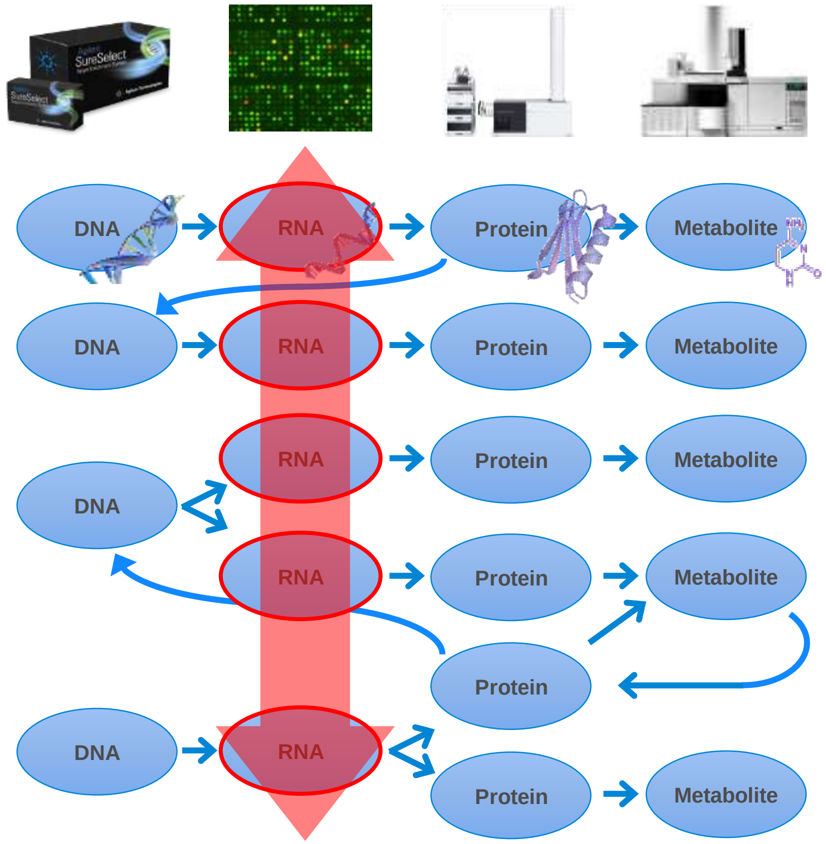
Extracting new insights from cancer genomics and proteomics data

Shweta Shukradas

Product Manager

The Cross Technology Challenge



DNA RNA Protein Metabolite

DNA RNA Protein Metabolite

RNA Protein Metabolite

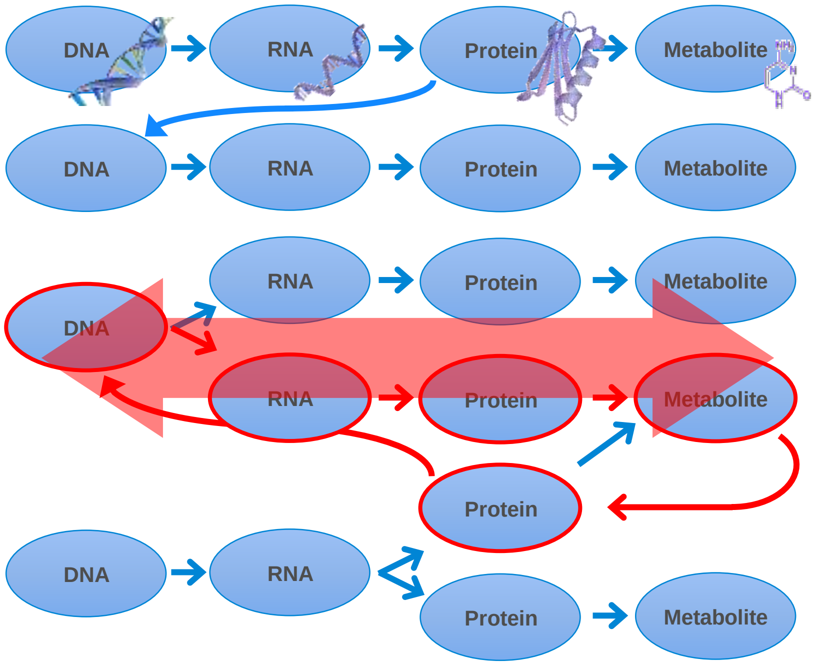
DNA

RNA Protein Metabolite

Protein

DNA RNA Protein Metabolite

"-Omics"



DNA RNA Protein Metabolite

DNA RNA Protein Metabolite

RNA Protein Metabolite

DNA

RNA Protein Metabolite

Protein

DNA RNA Protein Metabolite

Biological Processes

GeneSpring Suite of Products

GeneSpring GX

- mRNA Expression
- microRNA Expression
- Genome-wide Association
- Copy Number Variation
- Variant Analysis with vcf
- qPCR

Mass Profiler Professional

- Proteomics
- Metabolomics

Strand NGS

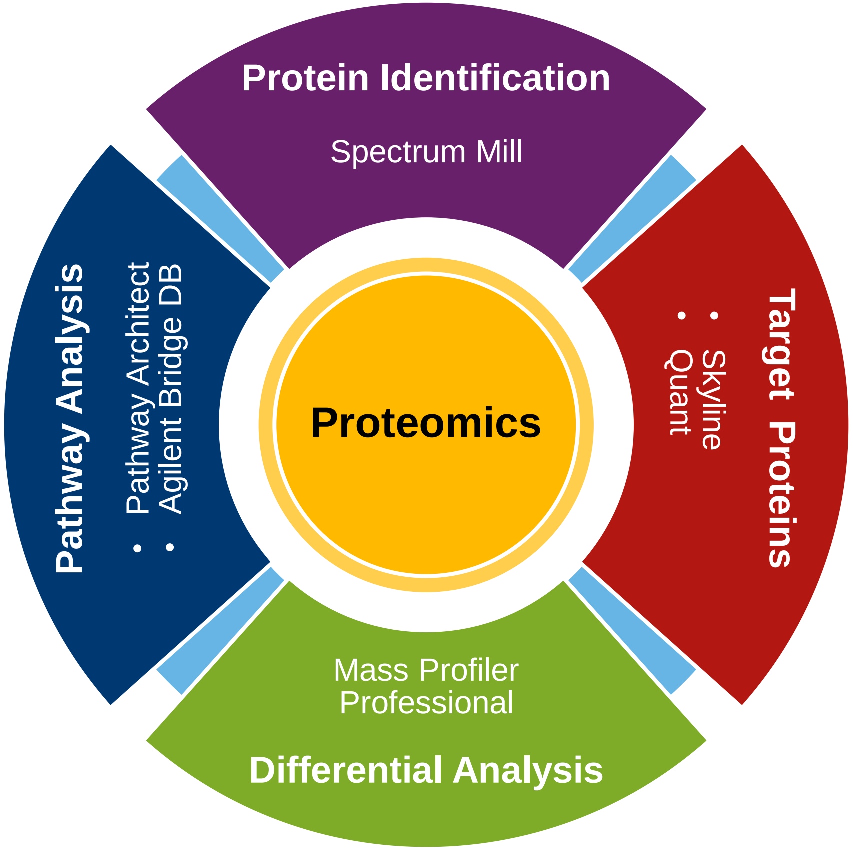
- DNA-Seq, RNA-Seq, ChIP-Seq, Methyl-Seq and small RNA-Seq

Pathway Architect

- Multi-omic Pathway Analysis
- NLP discovery

PROTEOMICS WORKFLOW IN MPP

Most Extensive Software Portfolio for Proteomics Data Analysis



Identify proteins from discovery data-dependent analysis

and/or

Target peptides for proteins of interest

Analyze data for statistically different features

Search and map annotated proteins to find interesting pathways

Visualize results on pathways

Agilent's Discovery Proteomics Workflow

Separate and Detect

Identification and Quantitation

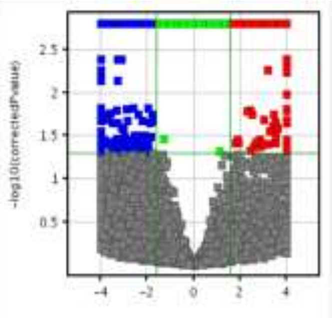
Statistical Analysis

Pathway Analysis

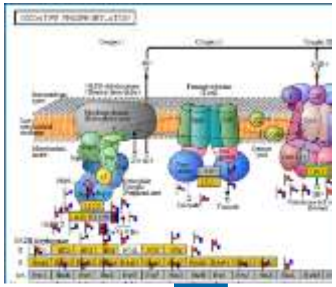
6500 Series LC/Q-TOF

Spectrum Mill

Mass Profiler Professional

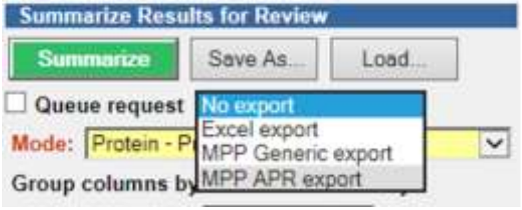


Pathway Architect



Pathway-directed targeted protein(s) experiment

Two MPP Export Modes in Spectrum Mill



Summarize Results for Review

☐ Queue request

Mode: Protein - P

Group columns by

Generic

- Support all data
- Protein abundance information only

Group (#)	Subgroup (#)	Spectra (#)	Distinct Peptides (#)	Distinct Summed MS/MS Search Score	% AA Coverage	Total Protein Spectral Intensity	Database Accession #	Protein Name
2	2.1	1377	34	518.93	87	8.13e+009	P43490	Nicotinamide phosphoribosyltransferase

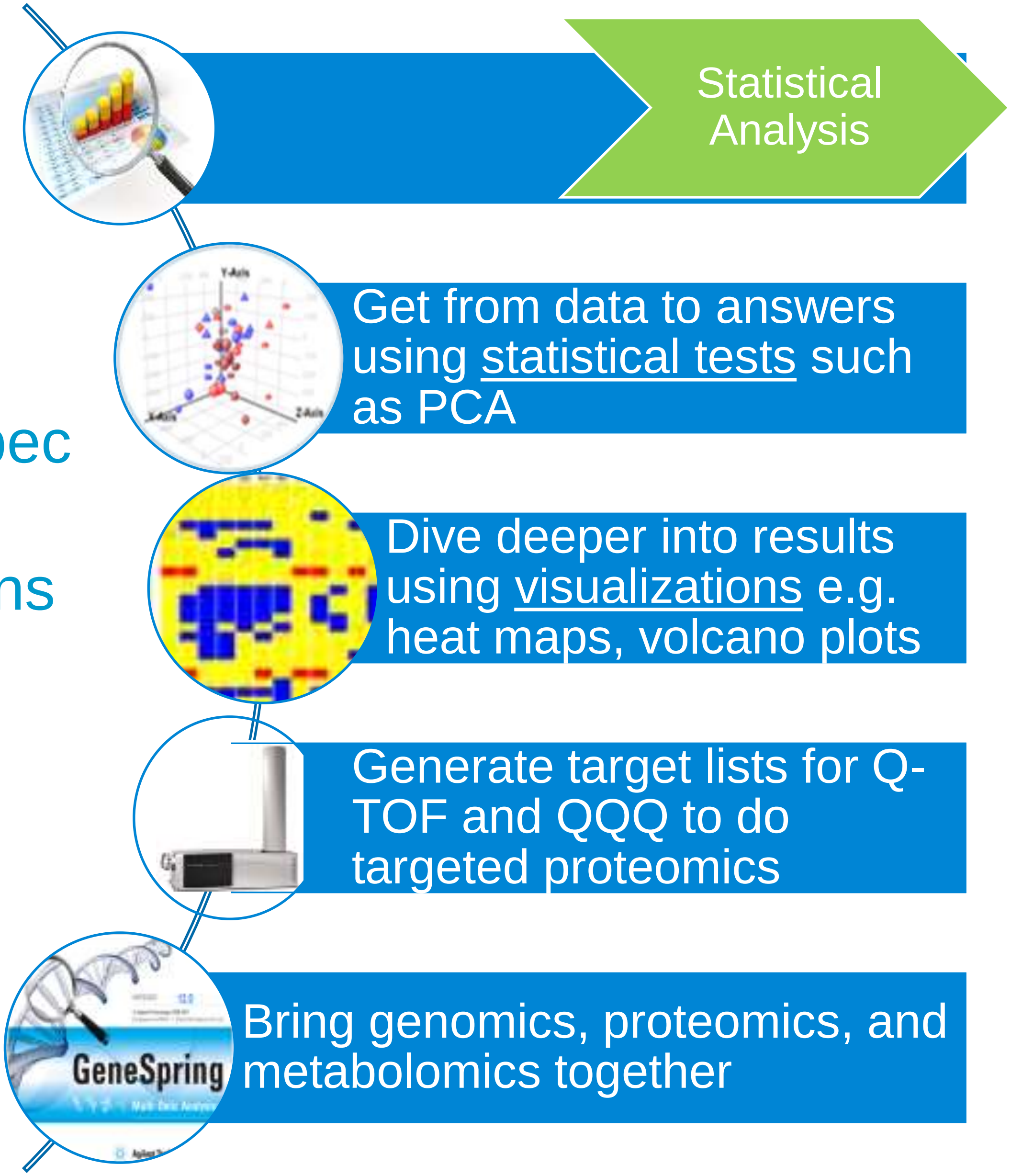
#	Filename	z	Score	SPI (%)	Spectrum Intensity	Sequence	Precursor Averagine Chi Squared	m/z Measured (Da)
1	251T-L-r002.83426.83603.2	2	25.95	95.6	3.59e+007	(R) VIQGDGVDINTLQEIVEGMK (Q)	0.99	1079.5573
2	198-T-L-r002.58496.58581.2	2	25.64	95.7	3.01e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.99	1087.5513
3	198-T-L-r002.58432.58830.2	2	25.62	96.4	3.02e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.99	1087.5513
4	245T-L-r002.38829.39180.2	2	25.22	100.0	1.38e+007	(K) YLLETSGNLDGLEYK (L)	1.00	957.9312
5	202-T-L-r001.60579.60755.2	2	25.01	96.4	4.50e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.99	1087.5538
6	251T-L-r003.82702.83139.2	2	24.97	100.0	3.32e+007	(R) VIQGDGVDINTLQEIVEGMK (Q)	0.99	1079.5567
7	202-T-L-r002.59780.59947.2	2	24.88	94.4	4.02e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.99	1087.5536
8	245T-L-r003.38592.38938.2	2	24.83	98.2	1.43e+007	(K) YLLETSGNLDGLEYK (L)	0.97	957.9308
9	245T-L-r001.54982.55241.2	2	24.72	96.5	2.67e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.99	1087.5543
10	251T-L-r001.83914.84438.2	2	24.02	97.0	3.85e+007	(R) VIQGDGVDINTLQEIVEGMK (Q)	0.99	1079.5574
11	248T-L-r003.59430.59602.2	2	24.00	92.9	4.70e+006	(R) VIQGDGVDINTLQEIVEGMK (Q)	0.99	1079.5574
12	202-T-L-r002.70029.70092.2	2	23.78	94.7	4.56e+006	(R) VIQGDGVDINTLQEIVEGMK (Q)	0.97	1079.5564
13	234T-L-r002.57546.57911.2	2	23.75	97.0	4.49e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	0.97	1087.5518
14	234T-L-r003.41804.42286.2	2	23.73	96.7	1.56e+007	(K) YLLETSGNLDGLEYK (L)	0.98	957.9298
15	234T-L-r002.42626.43180.2	2	23.58	97.9	1.59e+007	(K) YLLETSGNLDGLEYK (L)	0.96	957.9305
16	245T-L-r003.52087.52287.2	2	23.28	90.7	1.87e+006	(R) VIQGDGVDINTLQEIVEGmK (Q)	1.00	1087.5547

Agilent Proteomics Results (.apr)

- Support only Agilent data
- Protein abundance and peptide abundance included in export
- Supports label-free, SILAC, iTRAQ, TMT
- Requires MPP 14.0 release

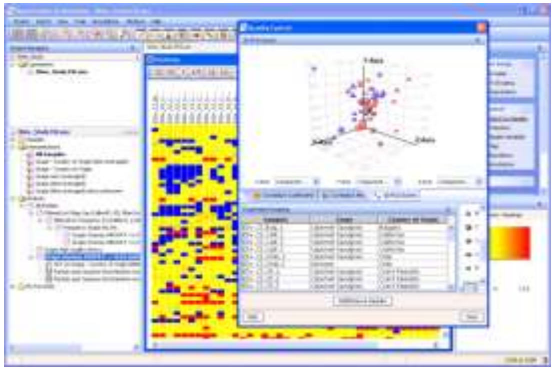
Mass Profiler Professional

Find differences in mass spec data sets and reach statistically valid conclusions

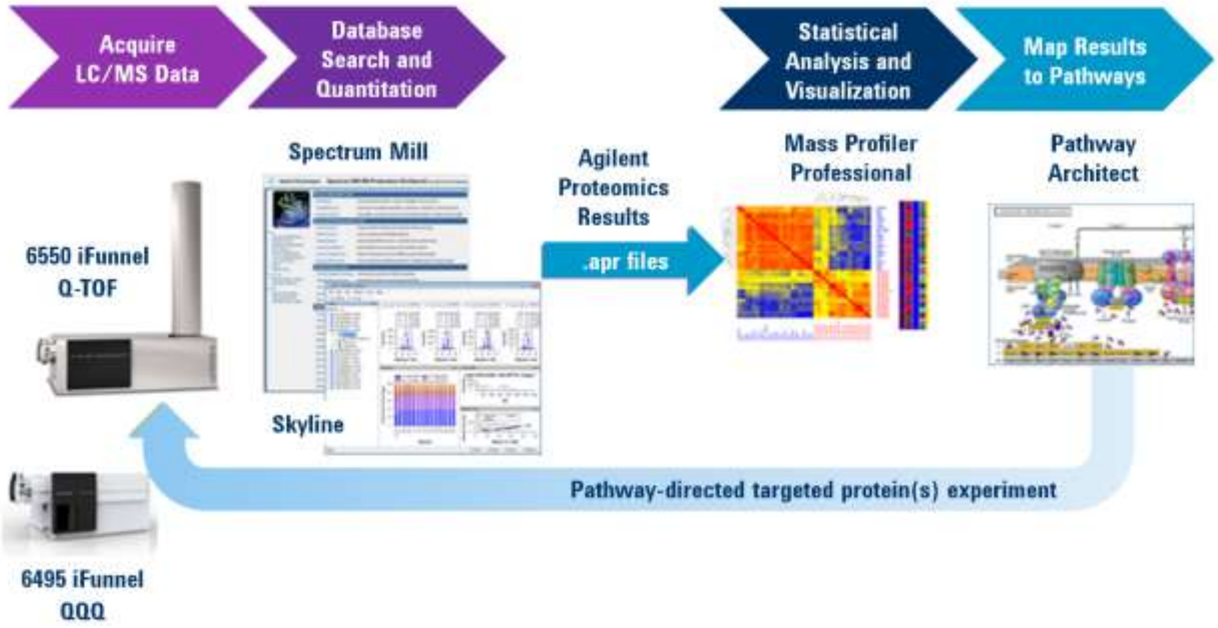


Statistical Analysis

- Get from data to answers using <u>statistical tests</u> such as PCA
- Dive deeper into results using <u>visualizations</u> e.g. heat maps, volcano plots
- Generate target lists for Q-TOF and QQQ to do targeted proteomics
- Bring genomics, proteomics, and metabolomics together



Proteomics Workflow in MPP



Acquire LC/MS Data

Database Search and Quantitation

Statistical Analysis and Visualization

Map Results to Pathways

6550 iFunnel Q-TOF

6495 iFunnel QQQ

Spectrum Mill

Skyline

Agilent Proteomics Results

.apr files

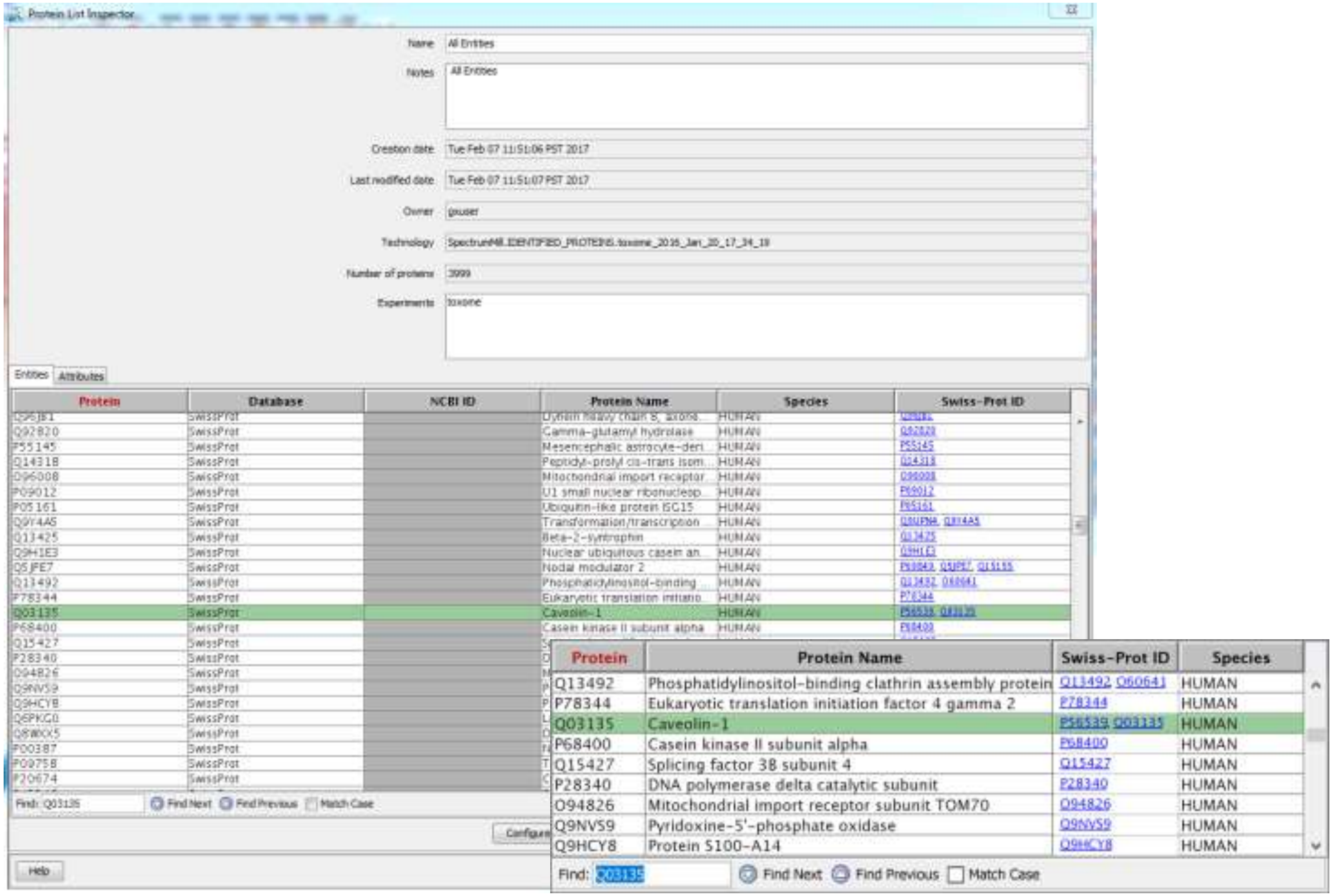
Mass Profiler Professional

Pathway Architect

Pathway-directed targeted protein(s) experiment

- Designed for new .apr format export with protein and peptide-level information
- Filters available for peptide-level and protein-level (abundance, frequency, PTMs)
- Protein/Peptide Entity Inspector

Protein List Inspector



Name: All Entities

Notes: All Entities

Creation date: Tue Feb 07 11:51:06 PST 2017

Last modified date: Tue Feb 07 11:51:07 PST 2017

Owner: pxuser

Technology: SpectrumMill.IDENTIFIED_PROTEINS.toxome_2016_Jan_20_17_34_19

Number of proteins: 3999

Experiments: toxome

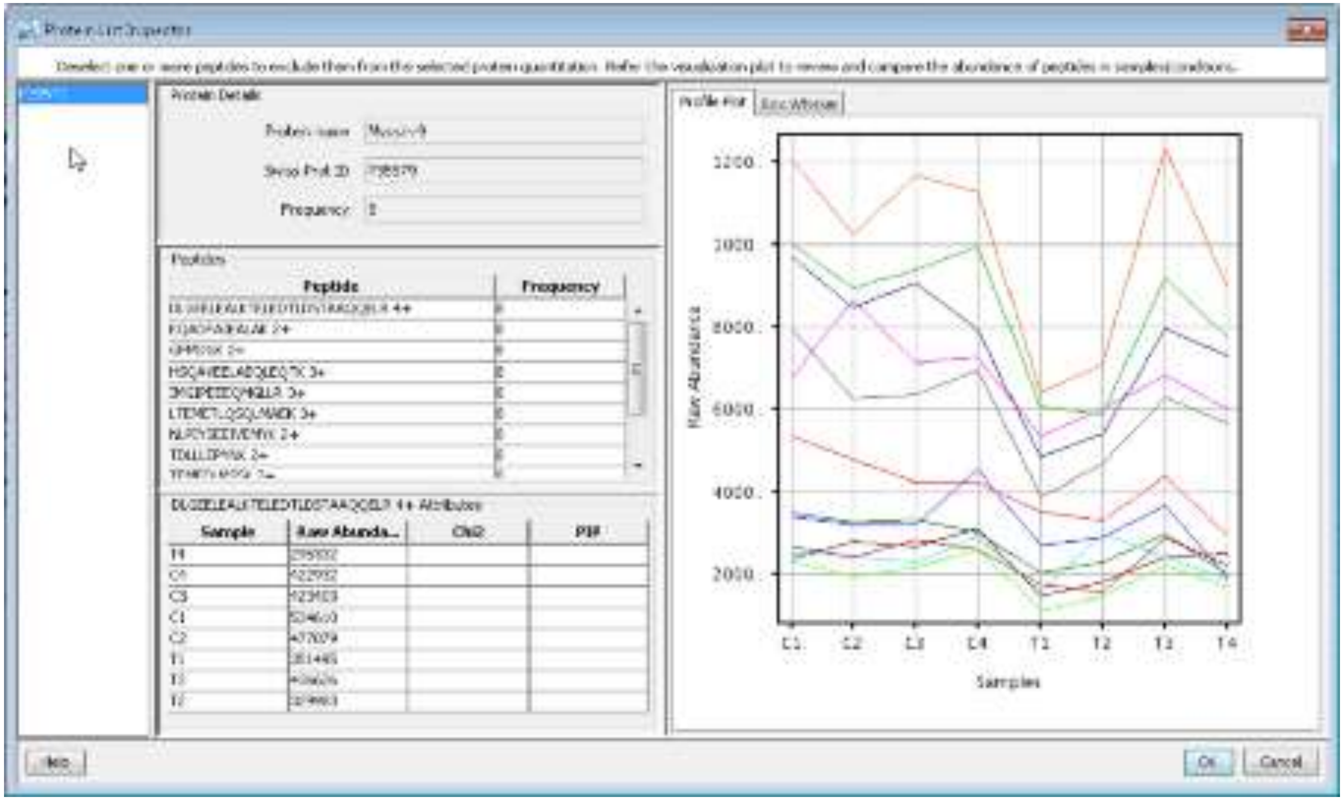
Protein	Database	NCBI ID	Protein Name	Species	Swiss-Prot ID
Q96JB1	SwissProt		Dynein heavy chain 8, axone...	HUMAN	[illegible]
Q92820	SwissProt		Gamma-glutamyl hydrolase	HUMAN	Q92820
P55145	SwissProt		Mesencephalic astrocyte-deri...	HUMAN	P55145
Q14318	SwissProt		Peptidyl-prolyl cis-trans isom...	HUMAN	Q14318
O96008	SwissProt		Mitochondrial import receptor...	HUMAN	O96008
P09012	SwissProt		U1 small nuclear ribonucleop...	HUMAN	P09012
P05161	SwissProt		Ubiquitin-like protein ISG15	HUMAN	P05161
Q9Y4A5	SwissProt		Transformation/transcription...	HUMAN	[illegible], Q9Y4A5
Q13425	SwissProt		Beta-2-syntrophin	HUMAN	Q13425
Q9H1E3	SwissProt		Nuclear ubiquitous casein an...	HUMAN	Q9H1E3
Q5JPE7	SwissProt		Nodal modulator 2	HUMAN	P69849, Q5JPE7, Q15155
Q13492	SwissProt		Phosphatidylinositol-binding...	HUMAN	Q13492, O60641
P78344	SwissProt		Eukaryotic translation initiatio...	HUMAN	P78344
Q03135	SwissProt		Caveolin-1	HUMAN	P56539, Q03135
P68400	SwissProt		Casein kinase II subunit alpha	HUMAN	P68400
Q15427	SwissProt				
P28340	SwissProt				
O94826	SwissProt				
Q9NVS9	SwissProt				
Q9HCY8	SwissProt				
Q6PKG0	SwissProt				
Q8WXX5	SwissProt				
P00387	SwissProt				
P09758	SwissProt				
P20674	SwissProt				

Find: Q03135 ○ Find Next ○ Find Previous ☐ Match Case

Protein	Protein Name	Swiss-Prot ID	Species
Q13492	Phosphatidylinositol-binding clathrin assembly protein	Q13492, O60641	HUMAN
P78344	Eukaryotic translation initiation factor 4 gamma 2	P78344	HUMAN
Q03135	Caveolin-1	P56539, Q03135	HUMAN
P68400	Casein kinase II subunit alpha	P68400	HUMAN
Q15427	Splicing factor 3B subunit 4	Q15427	HUMAN
P28340	DNA polymerase delta catalytic subunit	P28340	HUMAN
O94826	Mitochondrial import receptor subunit TOM70	O94826	HUMAN
Q9NVS9	Pyridoxine-5'-phosphate oxidase	Q9NVS9	HUMAN
Q9HCY8	Protein S100-A14	Q9HCY8	HUMAN

Find: Q03135 ○ Find Next ○ Find Previous ☐ Match Case

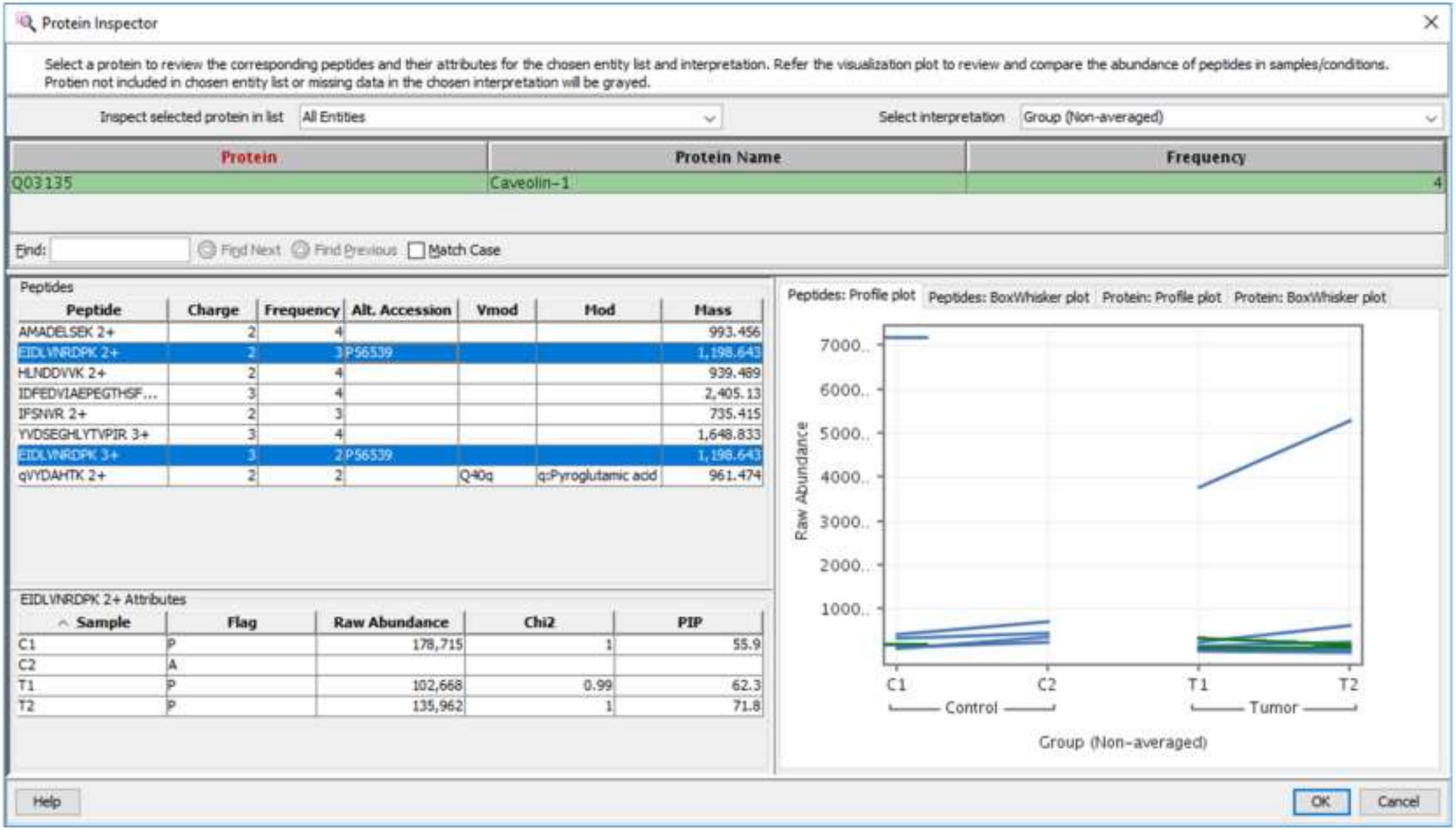
Protein Inspector Visualization



- Works **at both the protein and the peptide level simultaneously**
- This workflow design truly answers the biologist's questions of "what's changing in my proteins?" as well as "what's changing at the peptide level?"

Protein Inspector

Review peptide level information



Protein Inspector

Select a protein to review the corresponding peptides and their attributes for the chosen entity list and interpretation. Refer the visualization plot to review and compare the abundance of peptides in samples/conditions. Protien not included in chosen entity list or missing data in the chosen interpretation will be grayed.

Inspect selected protein in list: All Entities

Select interpretation: Group (Non-averaged)

Protein	Protein Name	Frequency
Q03135	Caveolin-1	4

Find: ☐ Match Case

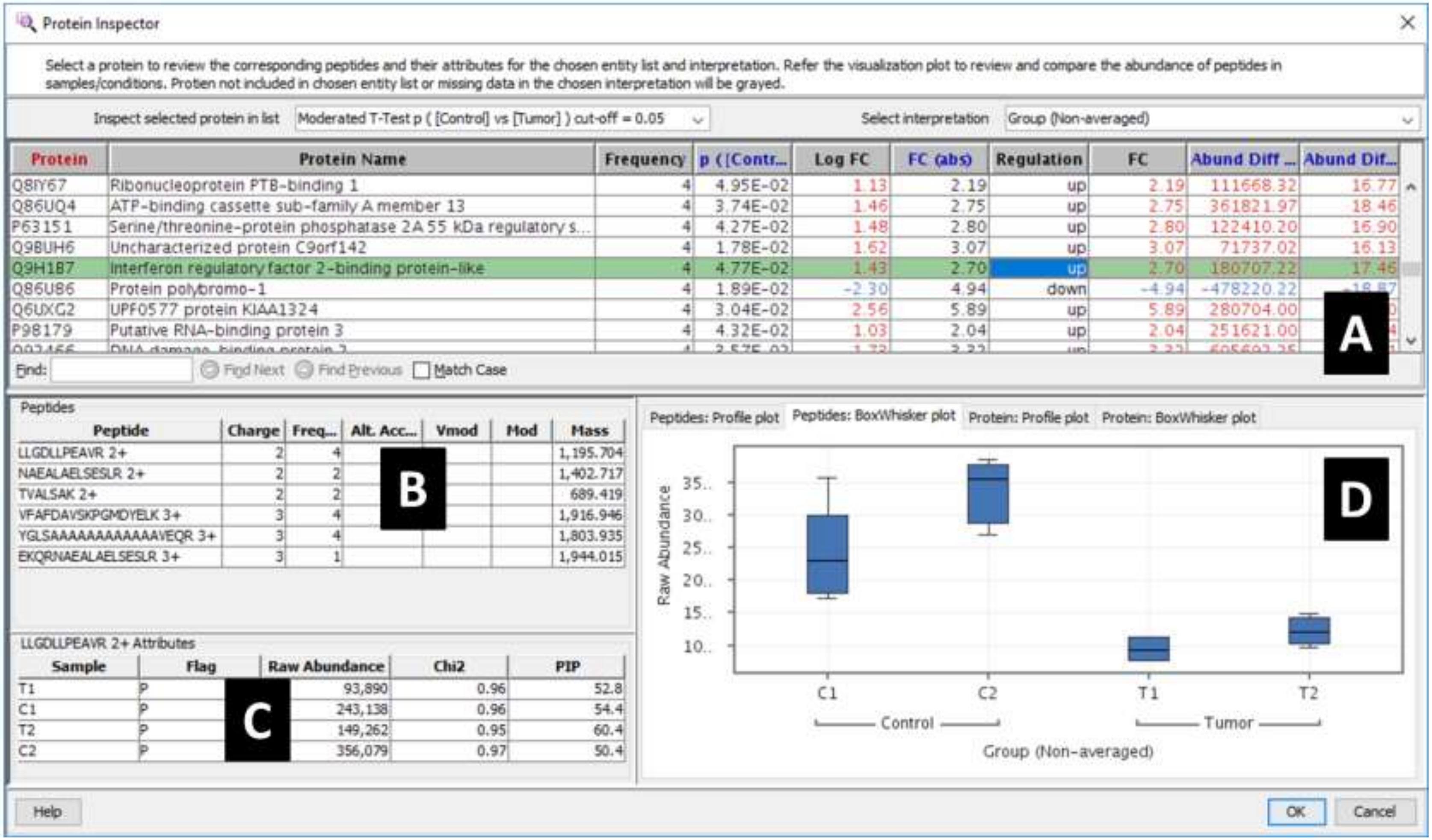
Peptides

Peptide	Charge	Frequency	Alt. Accession	Vmod	Mod	Mass
AMADELSEK 2+	2	4				993.456
EIDLVNRDPK 2+	2	3	P56539			1,198.643
HLNDDVVK 2+	2	4				939.489
IDFEDVIAEPEGTHSF…	3	4				2,405.13
IFSNVR 2+	2	3				735.415
YVDSEGHLYTVPIR 3+	3	4				1,648.833
EIDLVNRDPK 3+	3	2	P56539			1,198.643
qVYDAHTK 2+	2	2		Q40q	q:Pyroglutamic acid	961.474

EIDLVNRDPK 2+ Attributes

Sample	Flag	Raw Abundance	Chi2	PIP
C1	P	178,715	1	55.9
C2	A			
T1	P	102,668	0.99	62.3
T2	P	135,962	1	71.8

Protein Inspector Visualization



Protein Inspector

Select a protein to review the corresponding peptides and their attributes for the chosen entity list and interpretation. Refer the visualization plot to review and compare the abundance of peptides in samples/conditions. Protien not included in chosen entity list or missing data in the chosen interpretation will be grayed.

Inspect selected protein in list: Moderated T-Test p ([Control] vs [Tumor]) cut-off = 0.05

Select interpretation: Group (Non-averaged)

Protein	Protein Name	Frequency	p ([Contr...	Log FC	FC (abs)	Regulation	FC	Abund Diff ...	Abund Dif...
Q8IY67	Ribonucleoprotein PTB-binding 1	4	4.95E-02	1.13	2.19	up	2.19	111668.32	16.77
Q86UQ4	ATP-binding cassette sub-family A member 13	4	3.74E-02	1.46	2.75	up	2.75	361821.97	18.46
P63151	Serine/threonine-protein phosphatase 2A 55 kDa regulatory s...	4	4.27E-02	1.48	2.80	up	2.80	122410.20	16.90
Q9BUH6	Uncharacterized protein C9orf142	4	1.78E-02	1.62	3.07	up	3.07	71737.02	16.13
Q9H1B7	Interferon regulatory factor 2-binding protein-like	4	4.77E-02	1.43	2.70	up	2.70	180707.22	17.46
Q86U86	Protein polybromo-1	4	1.89E-02	-2.30	4.94	down	-4.94	-478220.22	-18.87
Q6UXG2	UPF0577 protein KIAA1324	4	3.04E-02	2.56	5.89	up	5.89	280704.00	[illegible]
P98179	Putative RNA-binding protein 3	4	4.32E-02	1.03	2.04	up	2.04	251621.00	[illegible]

Find: ○ Find Next ○ Find Previous ☐ Match Case

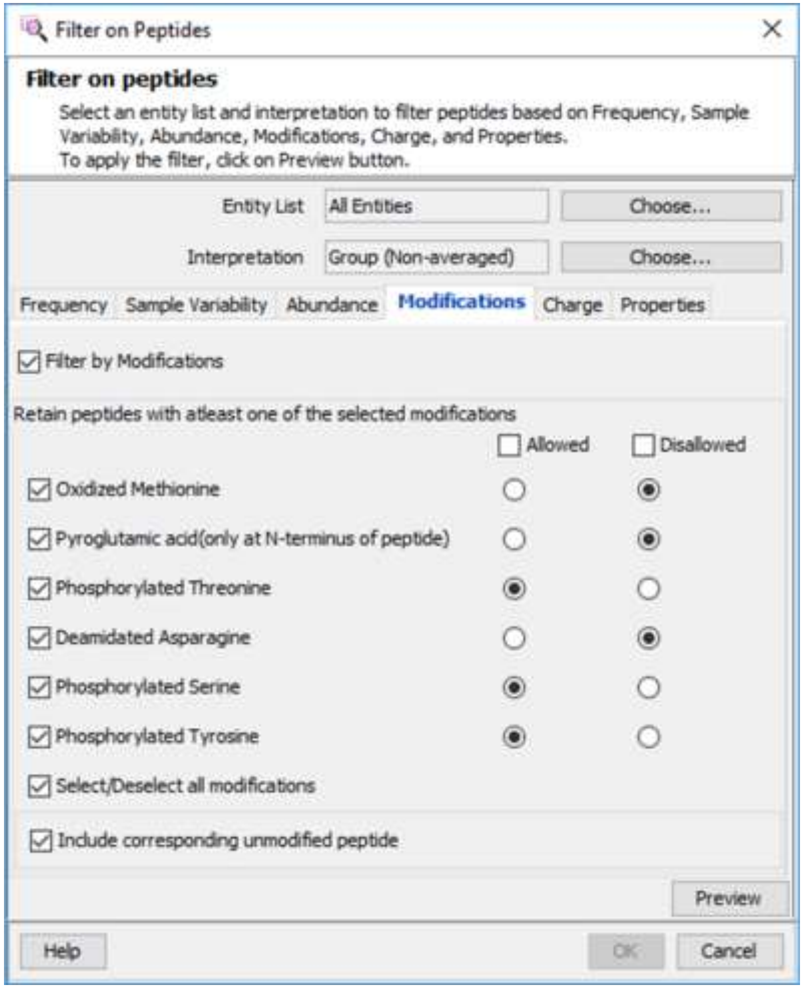
Peptides

Peptide	Charge	Freq...	Alt. Acc...	Vmod	Mod	Mass
LLGDLLPEAVR 2+	2	4				1,195.704
NAEALAELSESLR 2+	2	2				1,402.717
TVALSAK 2+	2	2				689.419
VFAFDAVSKPGMDYELK 3+	3	4				1,916.946
YGLSAAAAAAAAAAAAVEQR 3+	3	4				1,803.935
EKQRNAEALAELSESLR 3+	3	1				1,944.015

LLGDLLPEAVR 2+ Attributes

Sample	Flag	Raw Abundance	Chi2	PIP
T1	P	93,890	0.96	52.8
C1	P	243,138	0.96	54.4
T2	P	149,262	0.95	60.4
C2	P	356,079	0.97	50.4

Filtering on Peptides



Filter on peptides

Select an entity list and interpretation to filter peptides based on Frequency, Sample Variability, Abundance, Modifications, Charge, and Properties.
To apply the filter, click on Preview button.

Entity List: All Entities

Interpretation: Group (Non-averaged)

☑ Filter by Modifications

Retain peptides with atleast one of the selected modifications

	☐ Allowed	☐ Disallowed
☑ Oxidized Methionine	○	●
☑ Pyroglutamic acid(only at N-terminus of peptide)	○	●
☑ Phosphorylated Threonine	●	○
☑ Deamidated Asparagine	○	●
☑ Phosphorylated Serine	●	○
☑ Phosphorylated Tyrosine	●	○

☑ Select/Deselect all modifications

☑ Include corresponding unmodified peptide

- Modifications filter at the level of peptides
- The filter dialog lists all the available peptide modifications that are present in the selected entity list
- Filtering results in a list of proteins that only consist of specific peptides selected as allowed

Pathway Architect
Pathway analysis using multi-omic data

Pathway Analysis

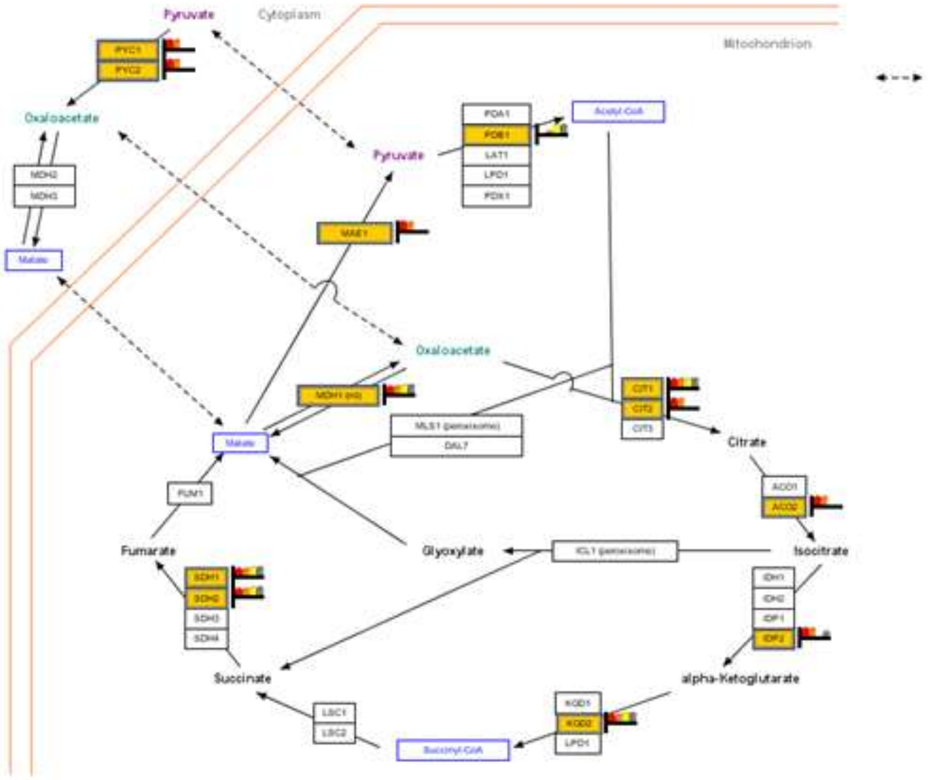
Customers can take the value of their data to another level. Pathway Architect helps them make sense of the data by:

Visualizing it on

- KEGG Pathways
- Wikipathways
- BioCyc/MetaCyc

Overlaying genomics, proteomics and metabolomics data

Designing their next experiment for MS/MS analysis or custom microarrays



TCA Cycle

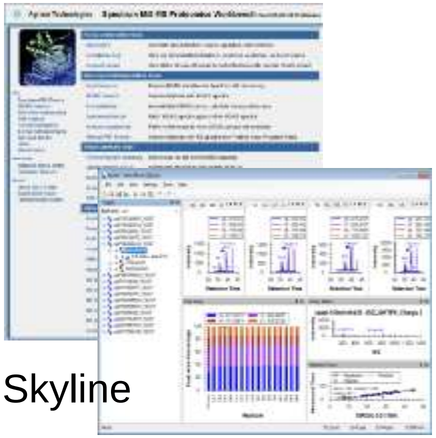
Accelerate Your Research With a Complete Pathway Centric Workflow

Acquire LC/MS data

Database Search or Quantitation

6550 iFunnel Q-TOF

Spectrum Mill



Skyline

6495 iFunnel QQQ

Pathway-directed targeted protein(s) experiment

GENOMICS WORKFLOWS IN GENESPRING GX

Genomics Research Solutions

Design → **Prepare Samples** → **Process Samples** → **Data Analysis**

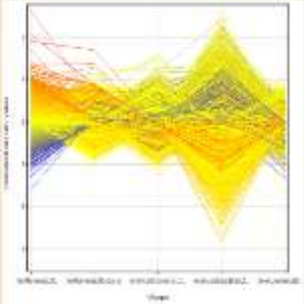
Gene Expression Analysis Workflow

Filtering

Statistical Analysis

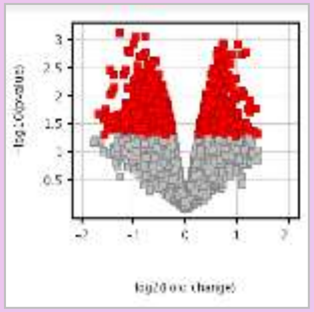
Pathway Analysis

Filtering Probe sets on expression



Probe sets filtered based on the percentile criteria

Statistical analysis

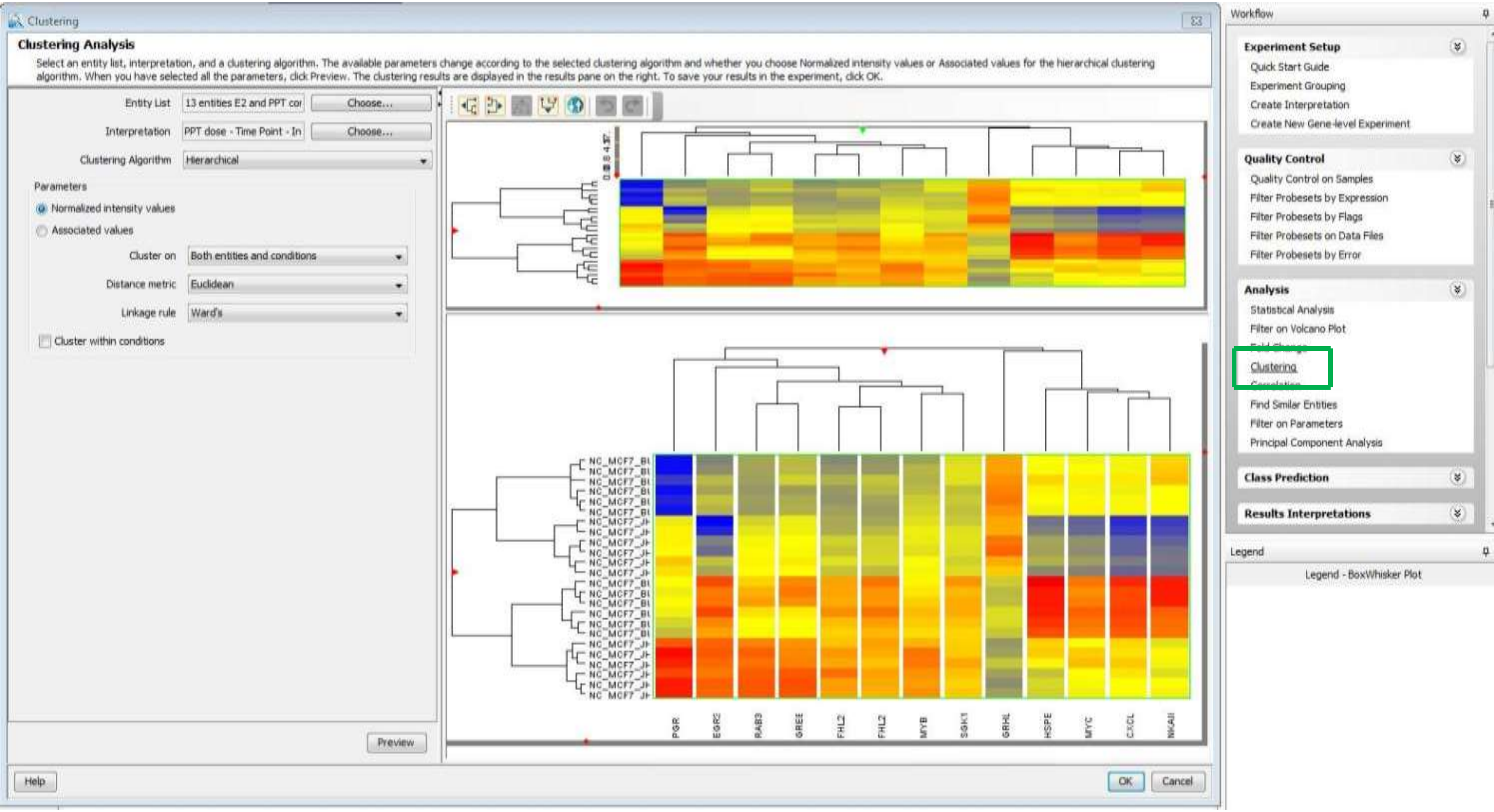


Performed to find differentially expressed probes

Pathway Analysis

Enrichment test performed with pathways of interest

Clustering - Hierarchical with preview window



Clustering

Clustering Analysis

Select an entity list, interpretation, and a clustering algorithm. The available parameters change according to the selected clustering algorithm and whether you choose Normalized intensity values or Associated values for the hierarchical clustering algorithm. When you have selected all the parameters, click Preview. The clustering results are displayed in the results pane on the right. To save your results in the experiment, click OK.

Entity List: 13 entities E2 and PPT cor

Interpretation: PPT dose - Time Point - In

Clustering Algorithm: Hierarchical

Parameters

● Normalized intensity values

○ Associated values

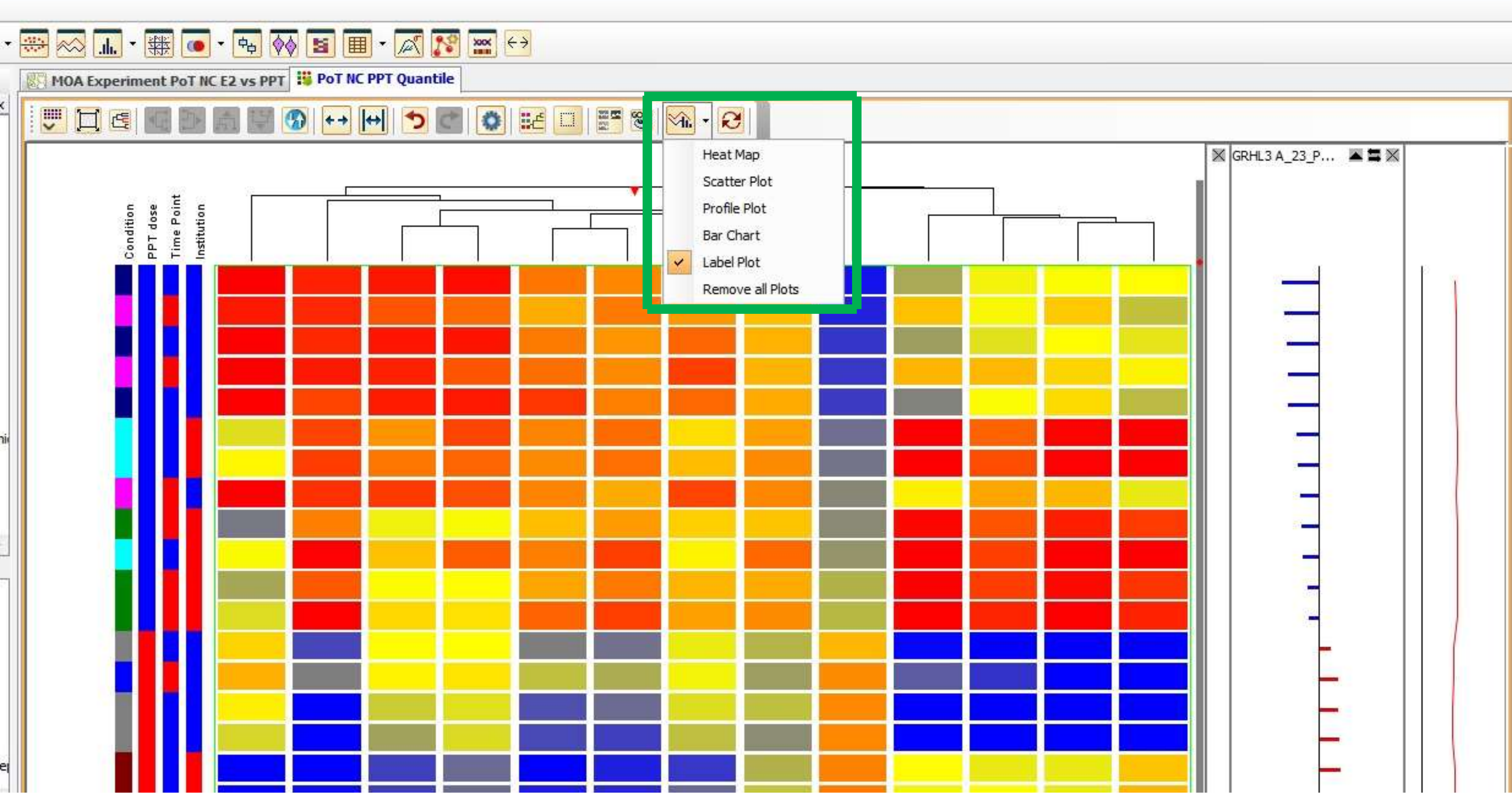
Cluster on: Both entities and conditions

Distance metric: Euclidean

Linkage rule: Ward's

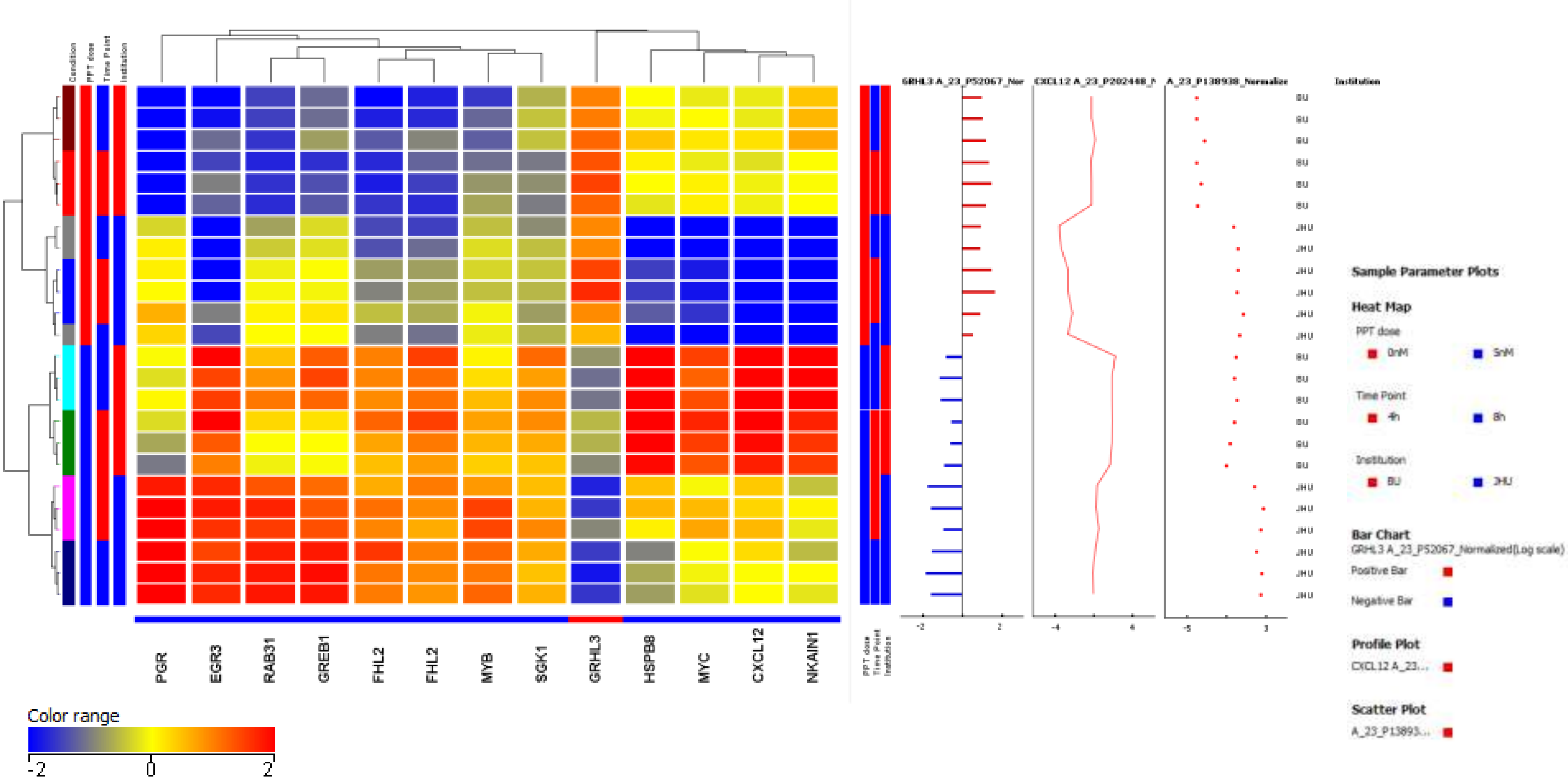
☐ Cluster within conditions

Metadata Framework: adding plots to clustered heatmap

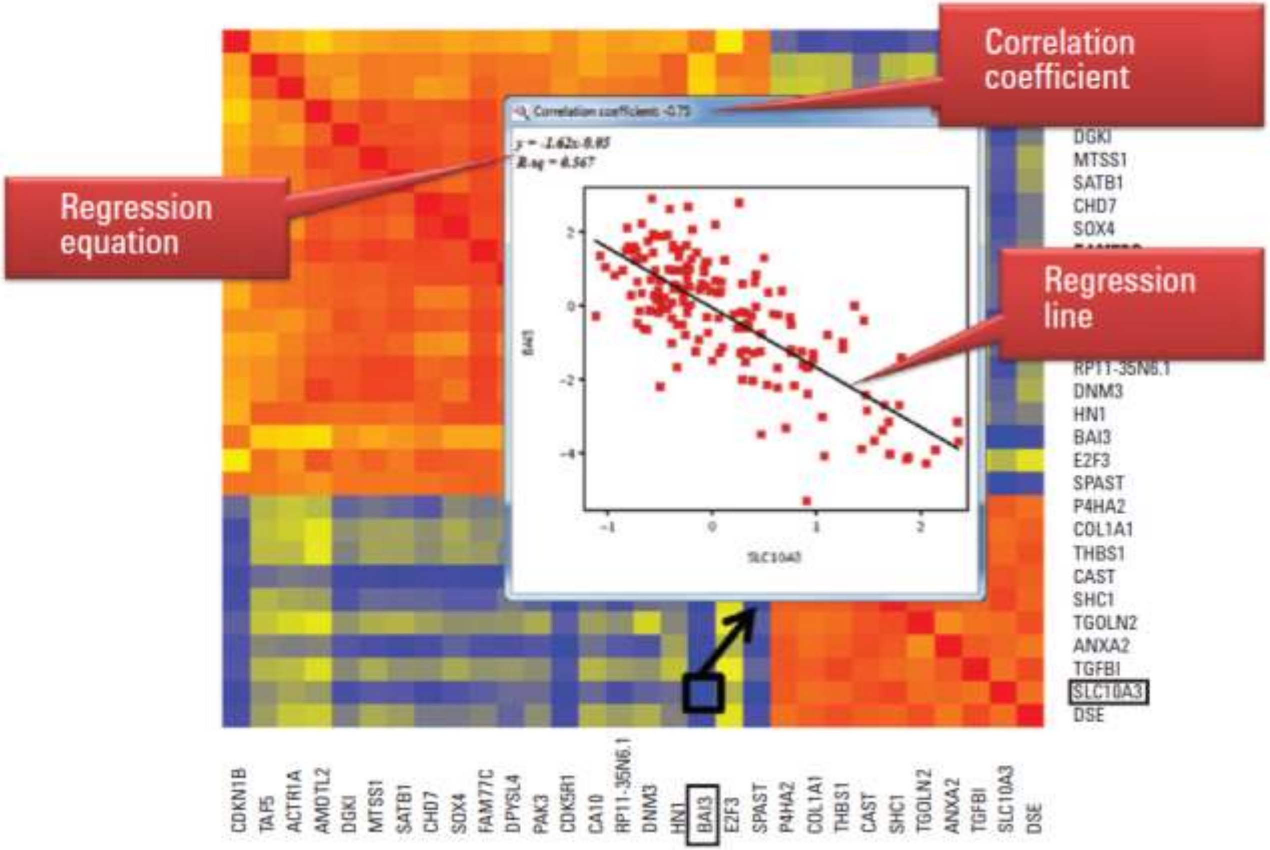


Metadata Framework

From Experimental Grouping Table information



Correlation Analysis


Correlation coefficient

Regression equation

Regression line

Correlation coefficient -0.75

y = -1.62x-0.05
R-sq = 0.567

Researchers can view relationships between entities (compounds or proteins) or samples. Clicking on a cell of a heat map, they can quickly view the specific parameters of the correlation

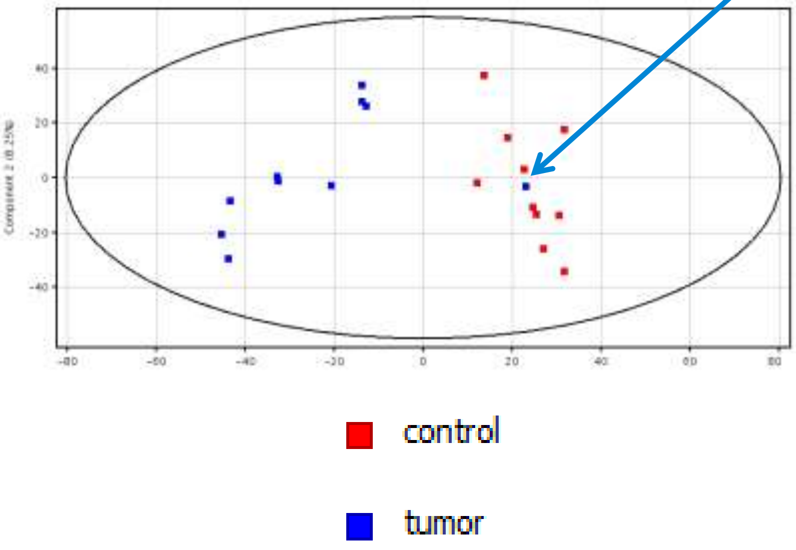
INTEGRATED ANALYSIS

Genomics and Proteomics

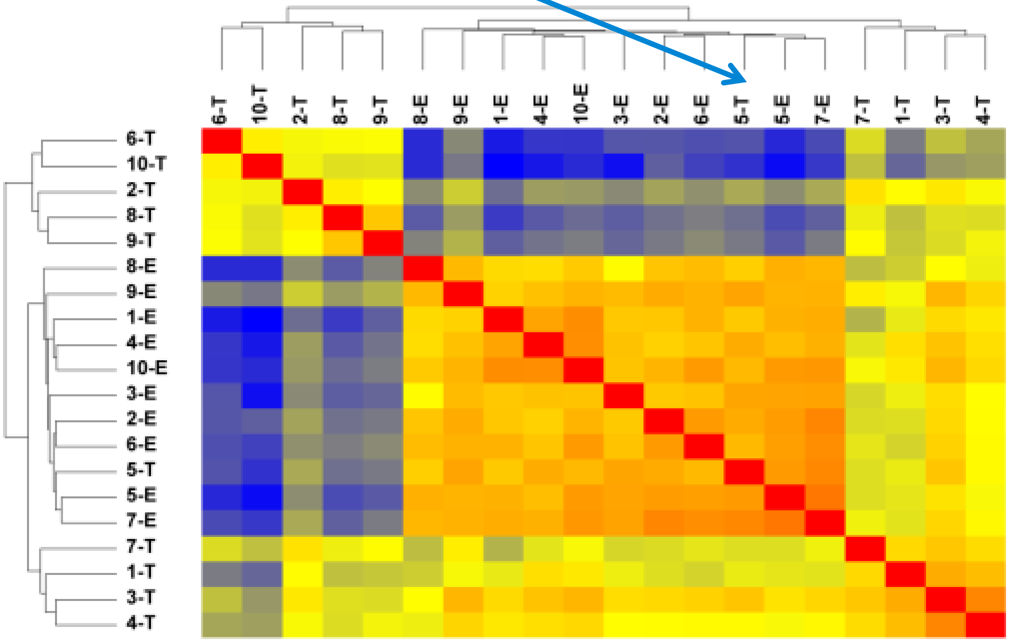
Brain Tumor Study - Quality Control of Discovery Proteomics Data

Sample 5-T

Principle Component Analysis Plot



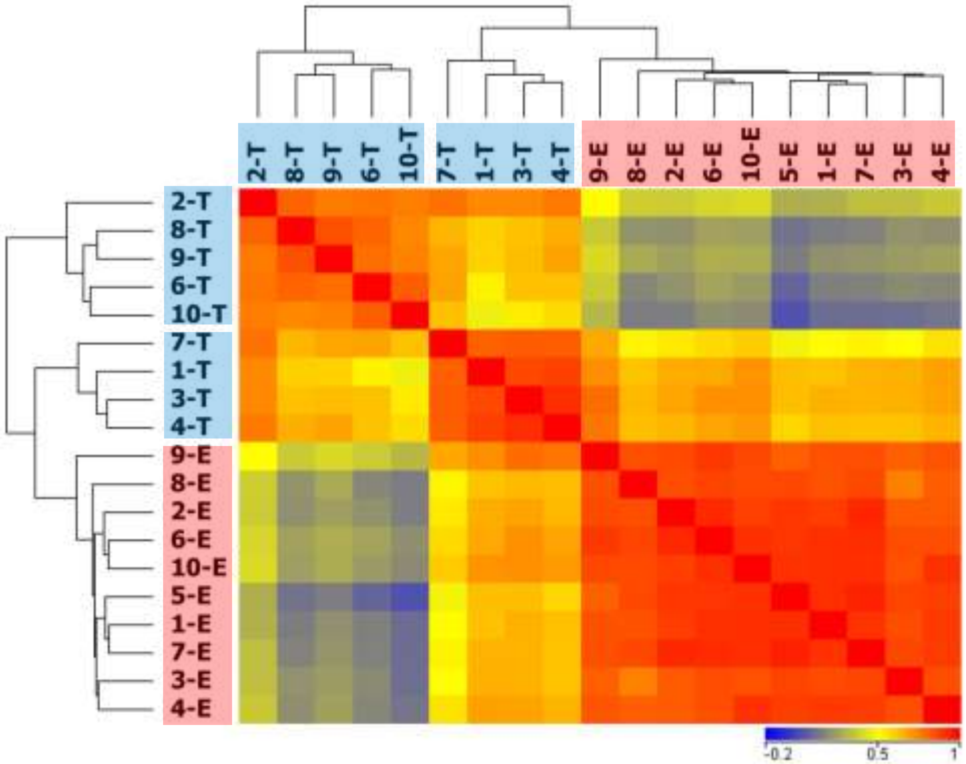
Sample Clustered Correlation Plot



T – Tumor samples   E – Epilepsy samples

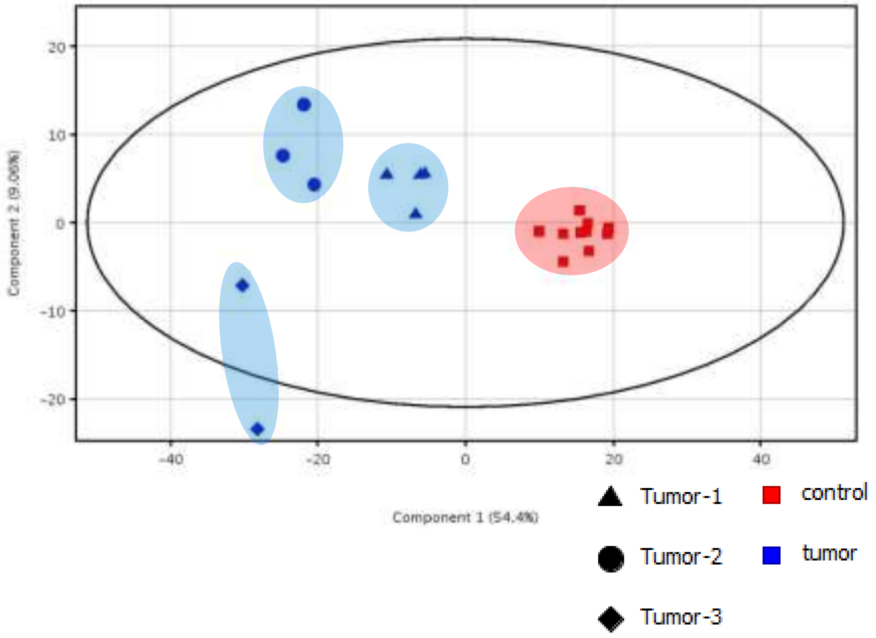
Brain Tumor Study - Sample-sample Correlation and PCA of Discovery Proteomics Data

Tumor subgroups demonstrated, using **587 differential proteins,** in both PCA and sample correlation showing the sample heterogeneity within tumors

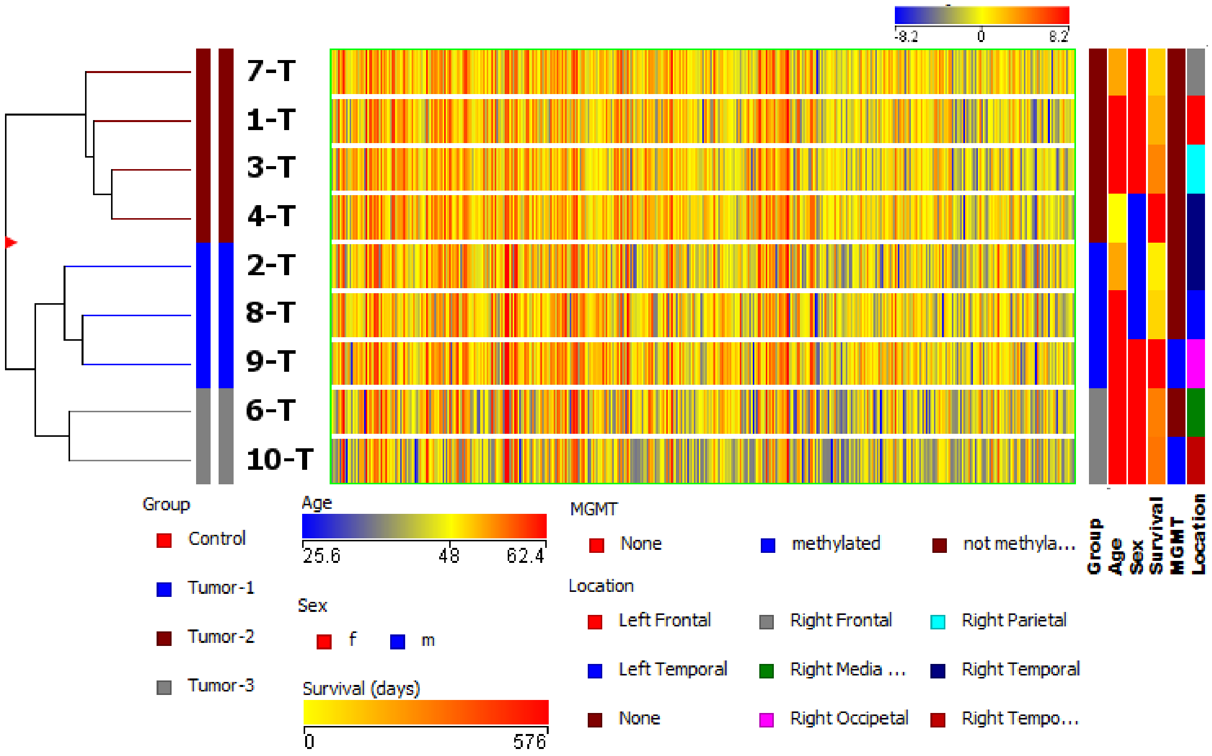


Sample-Sample Correlation

PCA

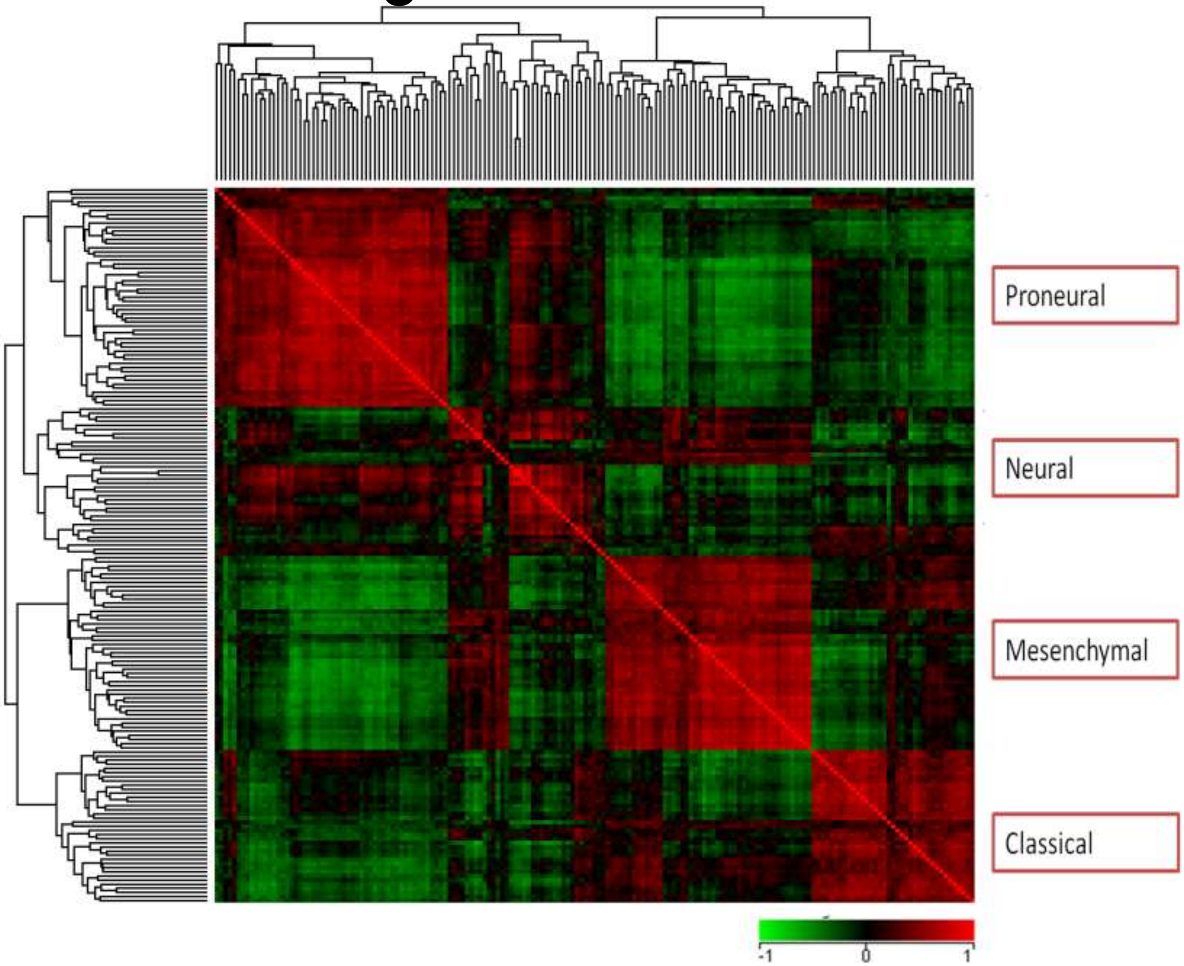


Brain Tumor Study - Metadata Analysis of Discovery Proteomics Data



- Hierarchical Clustering of samples provided similar classification as PCA
- Tumor subgroups aligned to sample metadata such as age, sex, tumor location and methylation status of MGMT
- The metadata did not explain the sub-grouping observed within the tumors based on the differentially expressed proteins

Brain Tumor Study - Hierarchical Clustering of Genomic Signatures

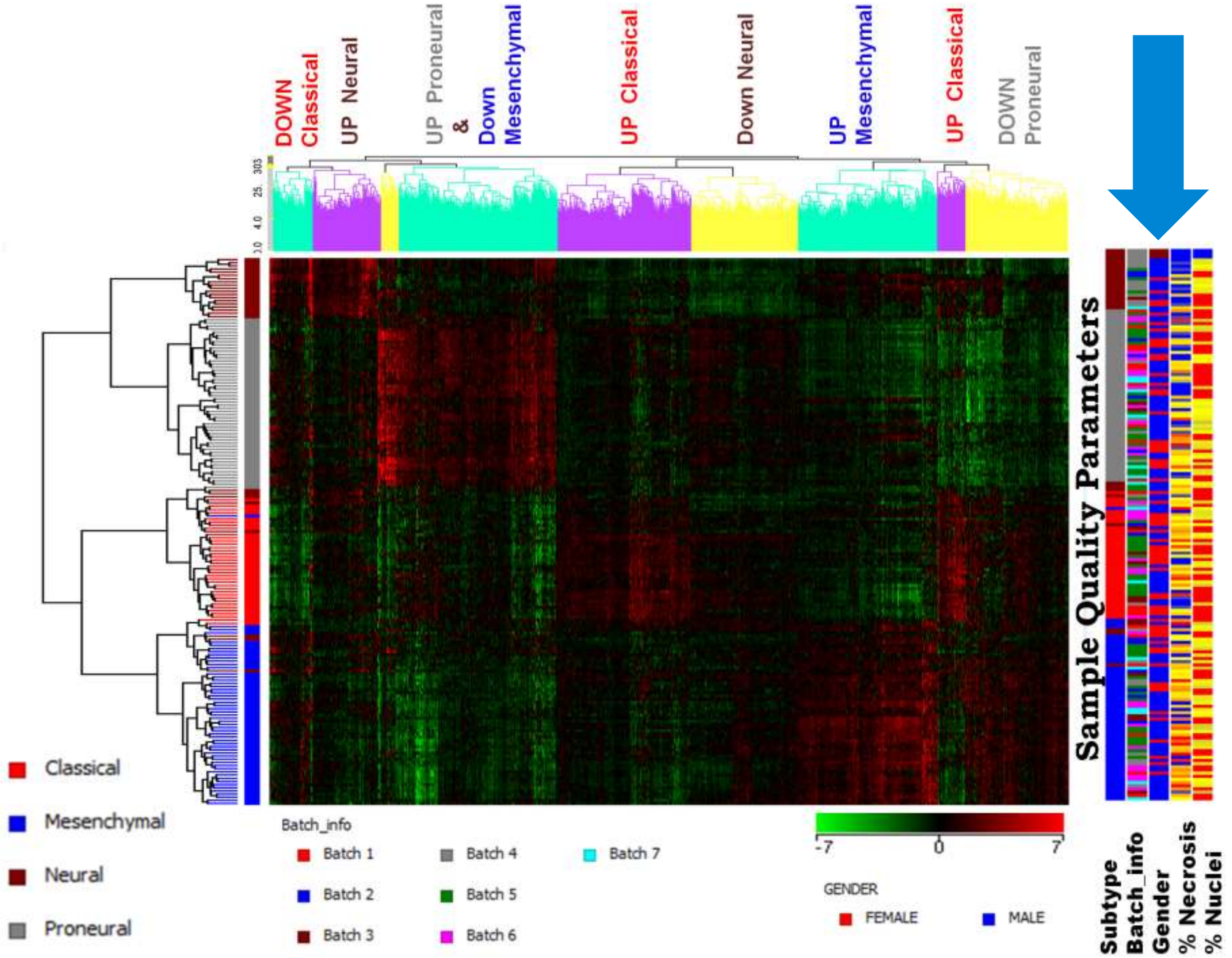


Hierarchical clustering of 840 mRNA profiles from 173 GBM tumors

The four GBM subtypes were re-created

Cancer Cell. 2010 January 19; 17(1): 98

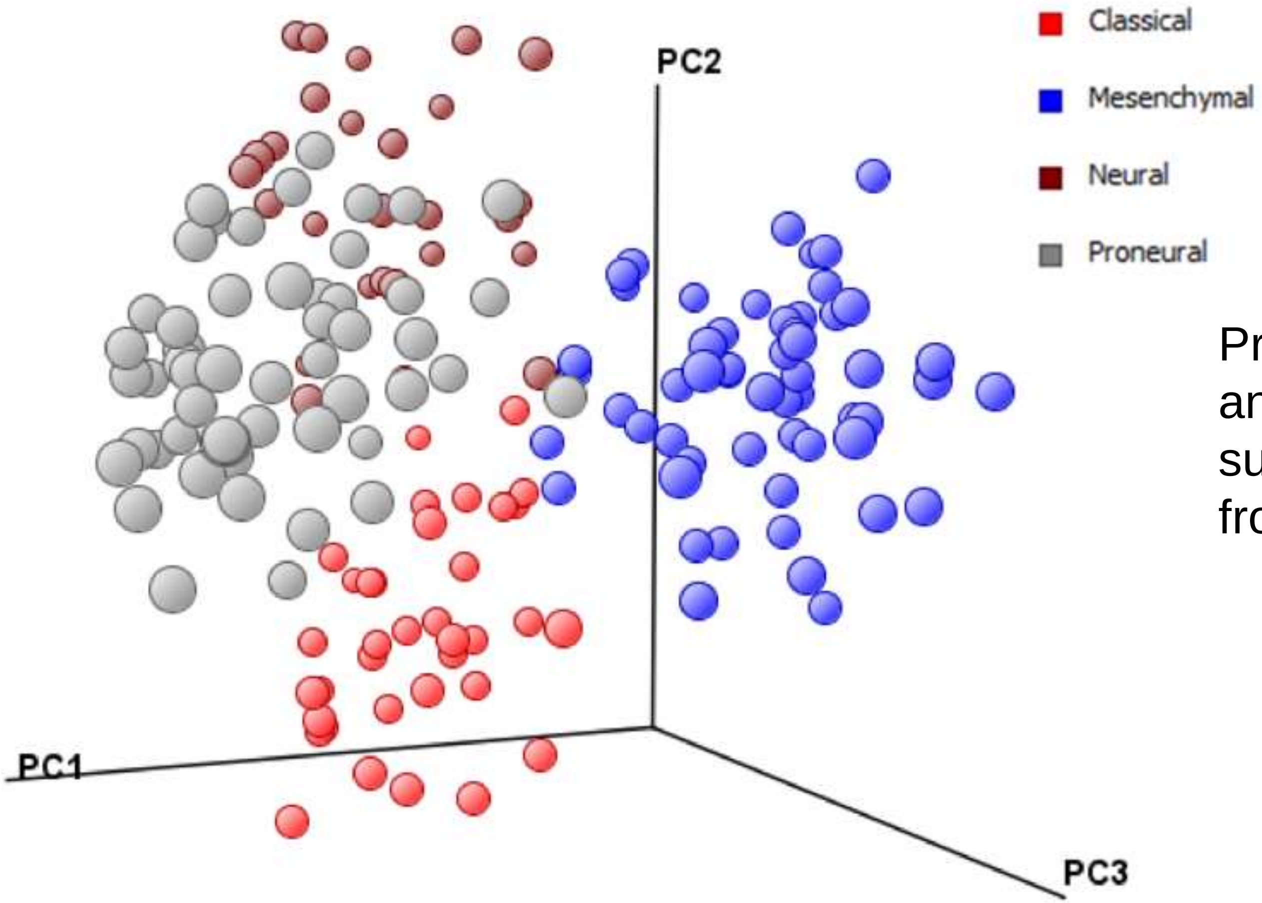
Brain Tumor Study - Metadata Analysis of Genomic Signatures



Metadata obtained from TCGA was imported into the project

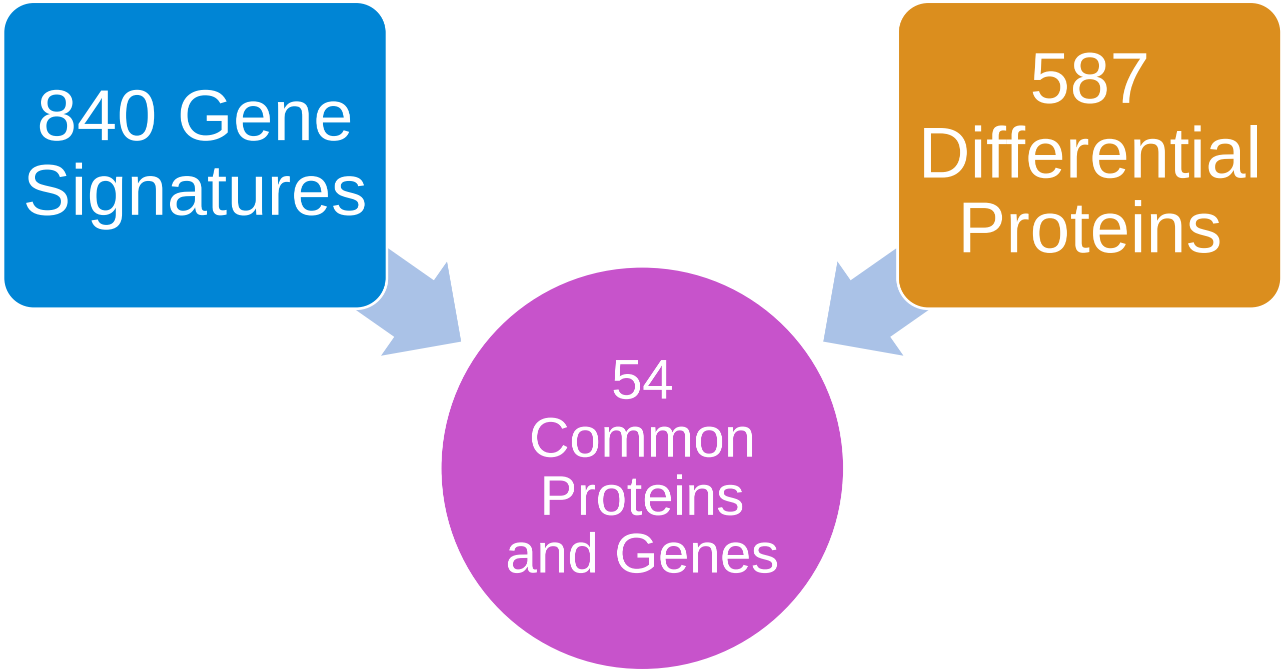
Metadata framework analysis of batch and sex was used to assess batch affects or bias

Brain Tumor Study - Principal Component Analysis of Genomic Signatures



Principal component analysis of confirms subtypes in GBM samples from TCGA

Brain Tumor Study - Venn Analysis of Genomic and Proteomic Data Set



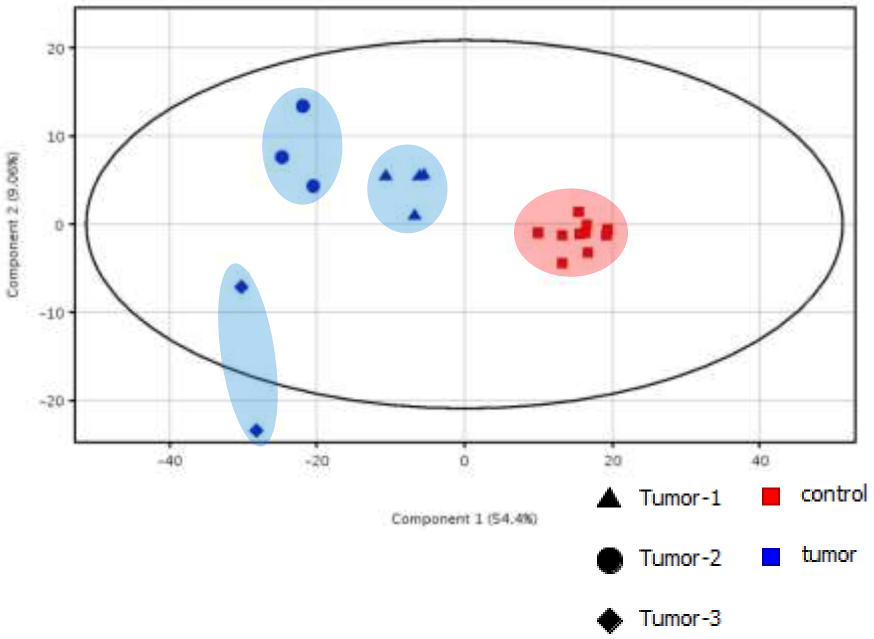
840 Gene Signatures

587 Differential Proteins

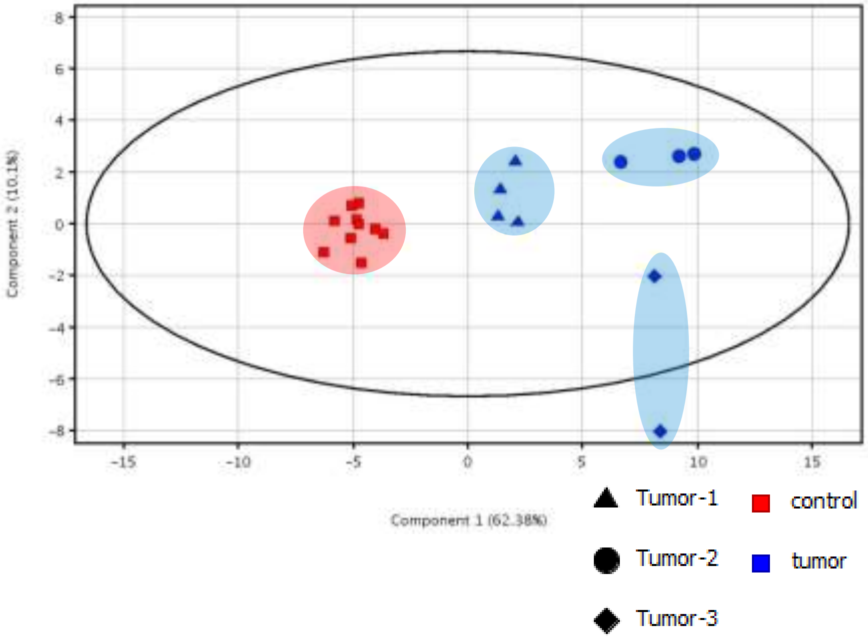
54 Common Proteins and Genes

Brain Tumor Study - PCA Comparison of Original 587 Proteins to 54 Proteins Identified Using Integration

587 Differential Proteins



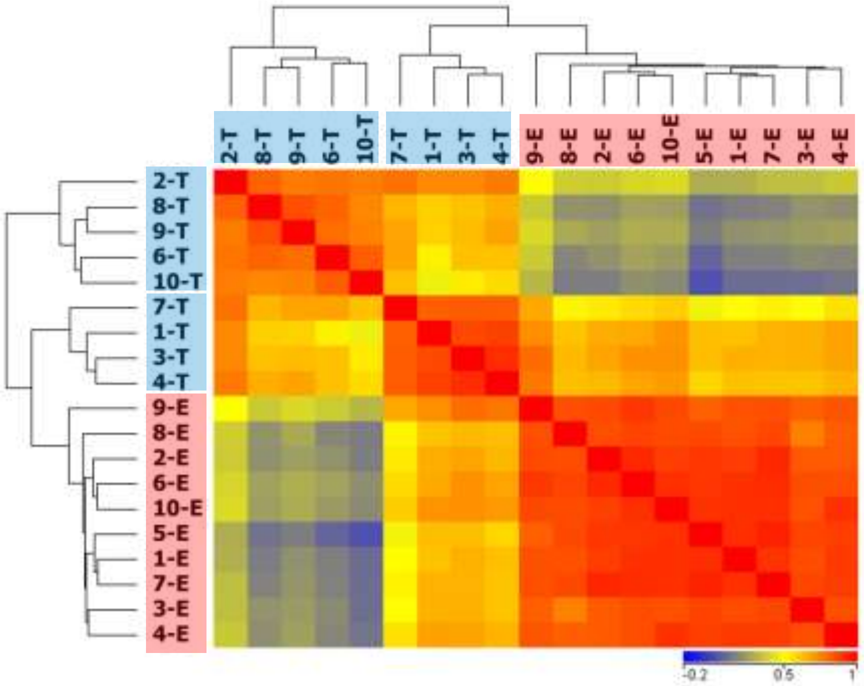
54 Proteins from Integration



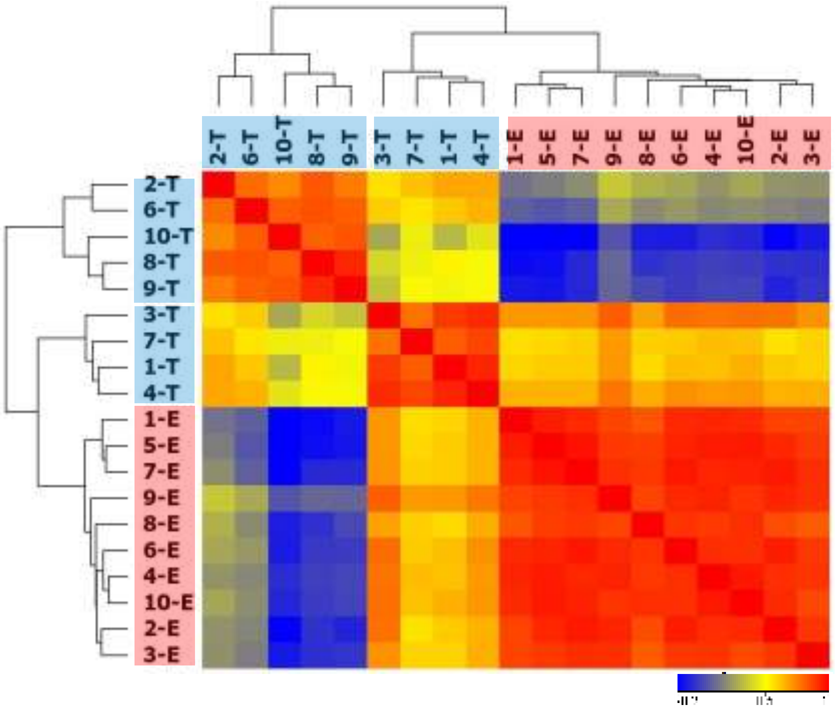
54 of 587 proteins that overlap the genomics signatures were good enough to capture the sub grouping in tumor samples, indicating the relevance between the molecular subtypes from genomics data and the subgroup pattern in proteomics data

Brain Tumor Study - Comparison of 587 Differential Proteins to 54 Proteins Identified Using Integration

587 Differential Proteins

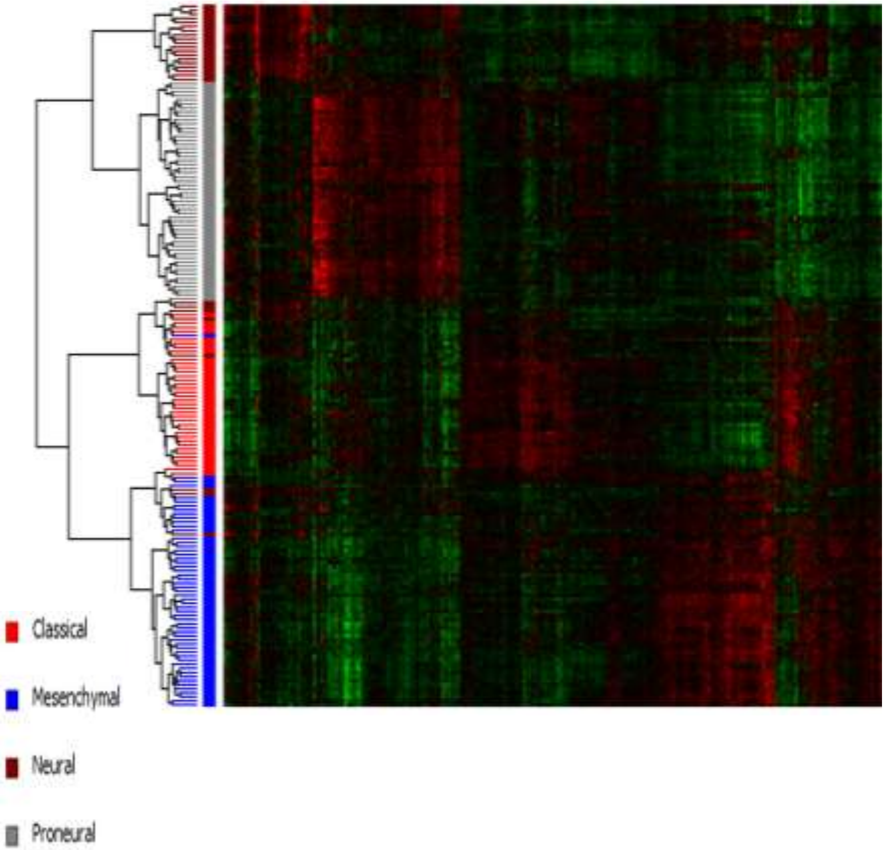


54 Proteins from Integration

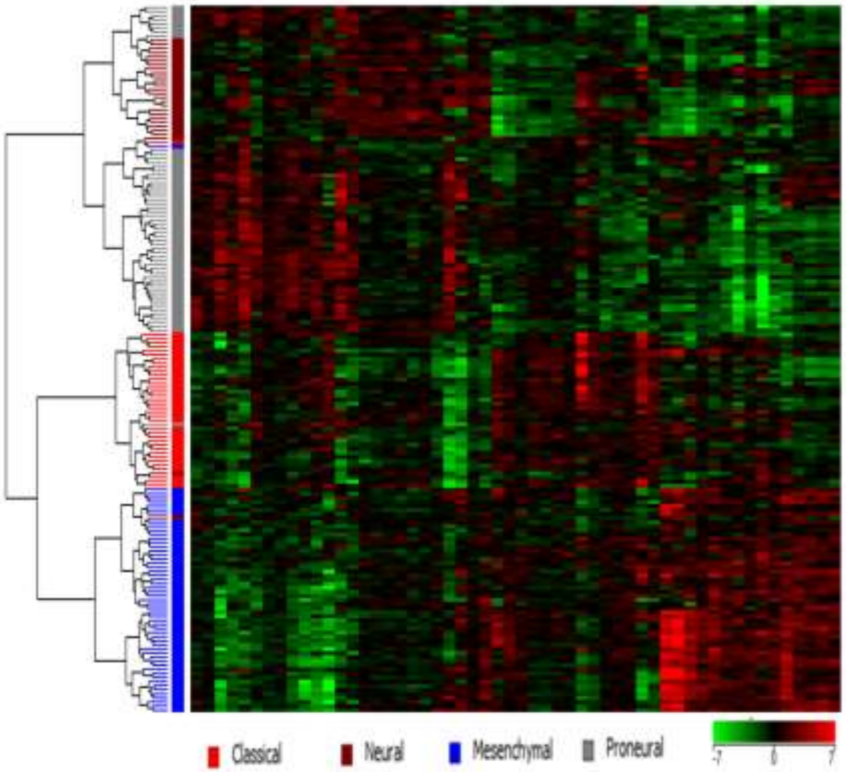


Brain Tumor Study - Performance of 54 Genes Identified Using Proteomics Integration to Original 840 Signature

Original Genomic 840 Signature



Genomic-Proteomic 54 Signature



Brain Tumor Study : Summary

Proteomics

- Use sample correlation & PCA on discovery proteomics data to identify outlier sample
- Identify a set of differential proteins for the tumor samples
- Use correlation and PCA to identify subgroups within tumor samples
- Clinical metadata visualized to reason the sub grouping within tumor samples

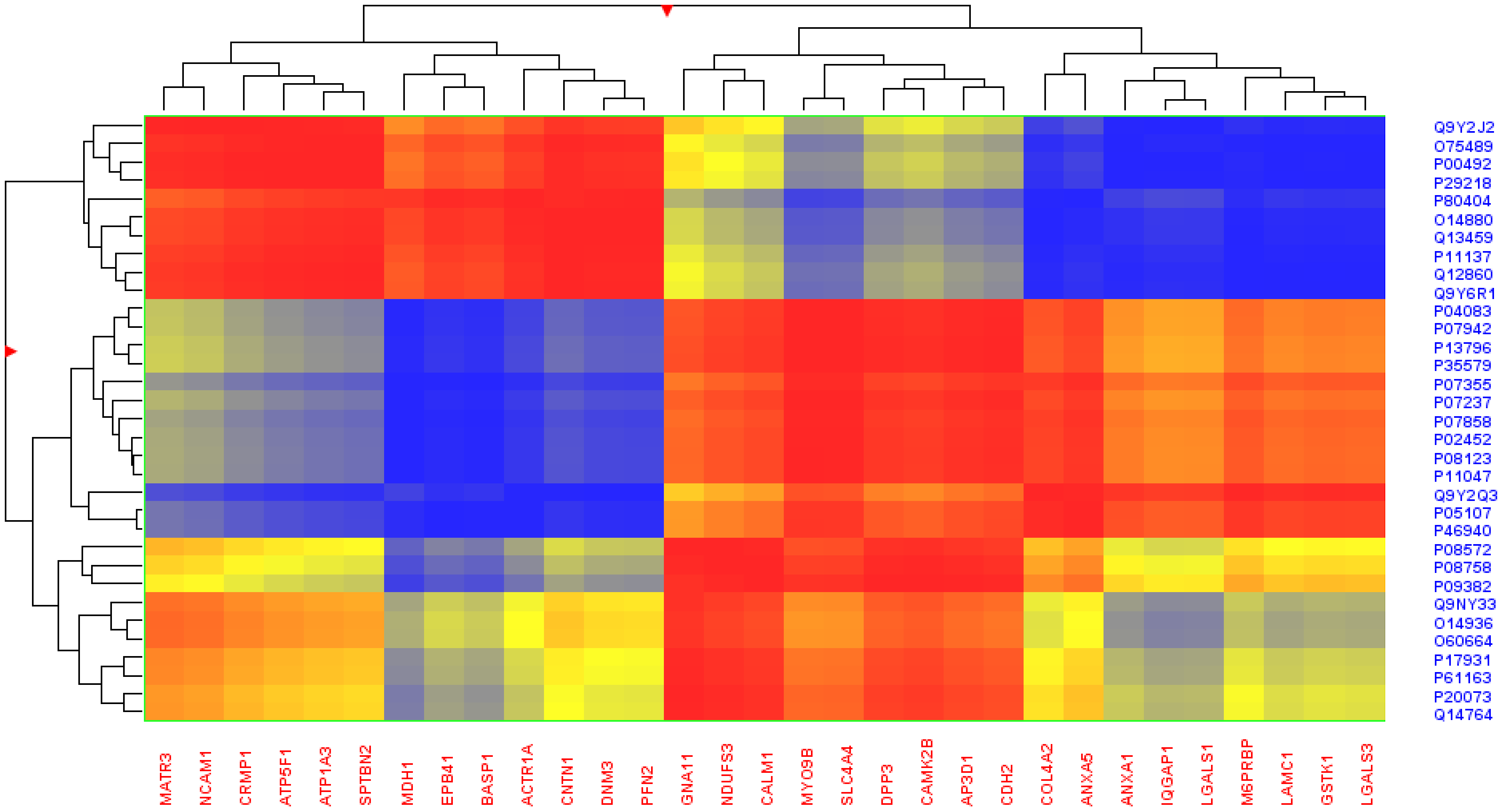
Genomics

- Use clustering and PCA on mRNA profiles to confirm subtype specific genomic signatures
- Assess data quality and bias using meta data analysis framework

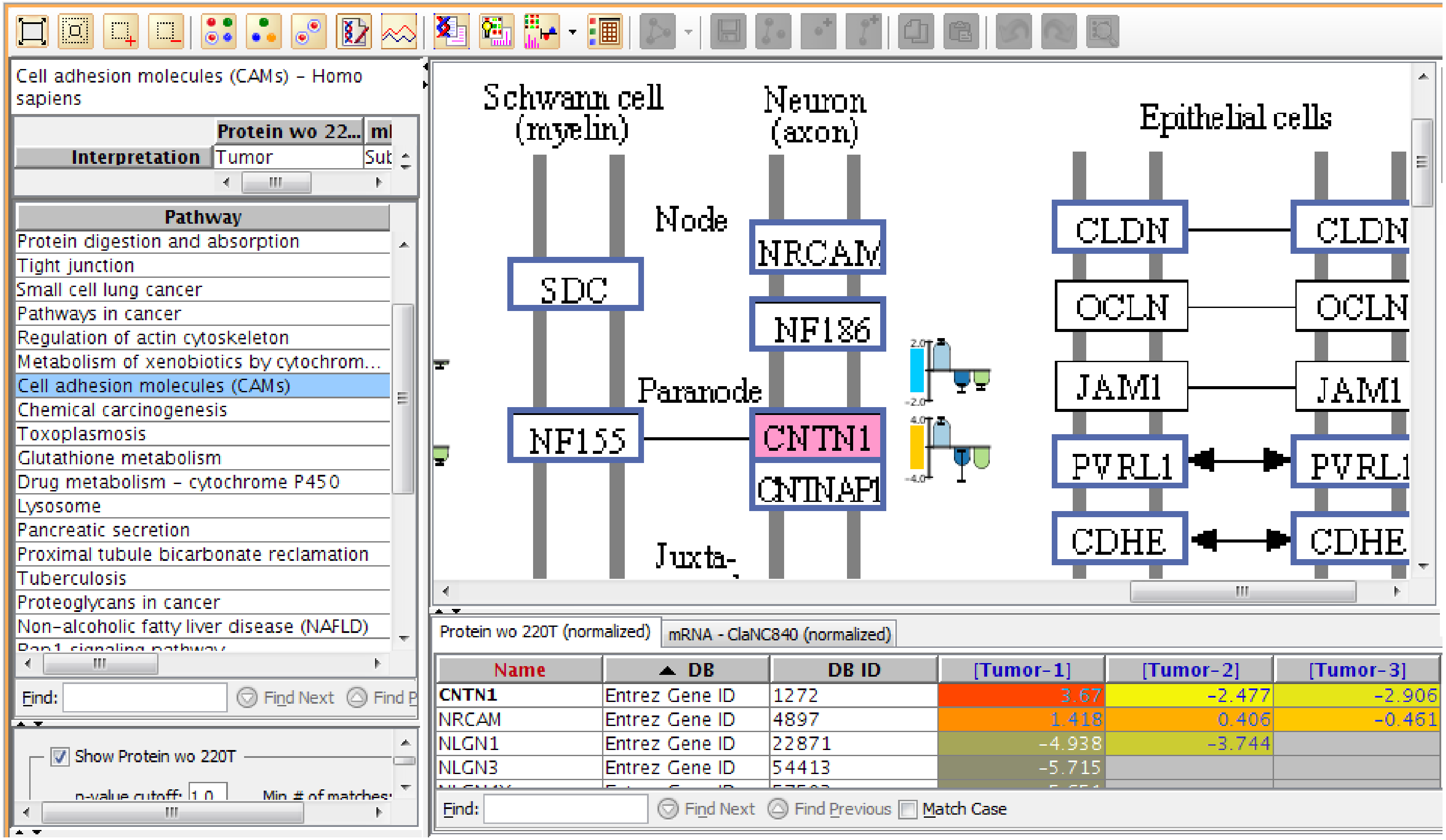
Integration

- Map gene id to protein accession number
- Perform Venn analysis of differential proteins and gene signatures
- Reduced set of gene signatures based on proteomics classified tumors in larger TCGA cohort

MOA Correlation Analysis - Concordant Genes and Proteins



Common Pathways - Concordant Genes and Proteins



Cell adhesion molecules (CAMs) - Homo sapiens

Interpretation	Protein wo 22...	ml
	Tumor	Sub

Pathway

- Protein digestion and absorption
- Tight junction
- Small cell lung cancer
- Pathways in cancer
- Regulation of actin cytoskeleton
- Metabolism of xenobiotics by cytochrom...
- Cell adhesion molecules (CAMs)
- Chemical carcinogenesis
- Toxoplasmosis
- Glutathione metabolism
- Drug metabolism - cytochrome P450
- Lysosome
- Pancreatic secretion
- Proximal tubule bicarbonate reclamation
- Tuberculosis
- Proteoglycans in cancer
- Non-alcoholic fatty liver disease (NAFLD)

Find:

☑ Show Protein wo 220T

p-value cutoff: 1.0

Min # of matches:

Protein wo 220T (normalized) | mRNA - ClaNC840 (normalized)

Name	▲ DB	DB ID	[Tumor-1]	[Tumor-2]	[Tumor-3]
CNTN1	Entrez Gene ID	1272	3.67	-2.477	-2.906
NRCAM	Entrez Gene ID	4897	1.418	0.406	-0.461
NLGN1	Entrez Gene ID	22871	-4.938	-3.744	
NLGN3	Entrez Gene ID	54413	-5.715		

Find: ○ Find Next ○ Find Previous ☐ Match Case

Proteomics

- ▨ [Tumor-1]
- ▨ [Tumor-2]
- 🟩 [Tumor-3]

Genomics

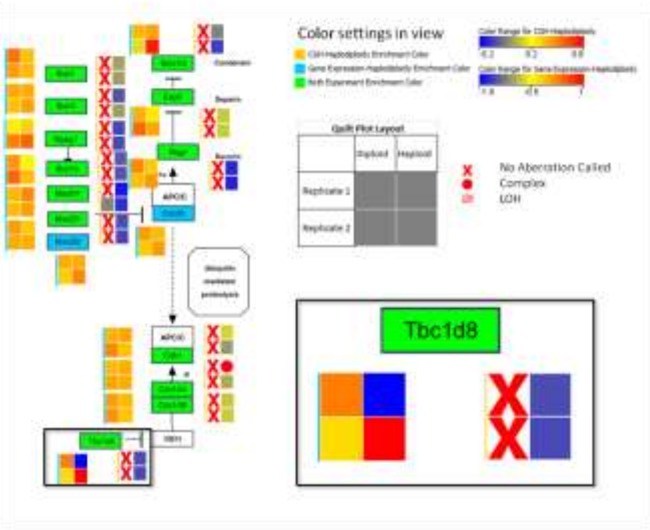
- ▨ [Proneural]
- ▨ [Classical]
- 🟩 [Mesenchymal]

DOWNSTREAM PROCESSING AND BIOLOGICAL CONTEXTUALIZATION

CGH Analysis Workflow

- Reports from Cytogenomics and AGW are imported in GeneSpring GX for further downstream analysis and data integration

- Overlay on pathways will help cancer researchers visualize aberrations with other enriched entities



- Genome browser enhancements to support CGH data

Variant Analysis with VCF Import

process

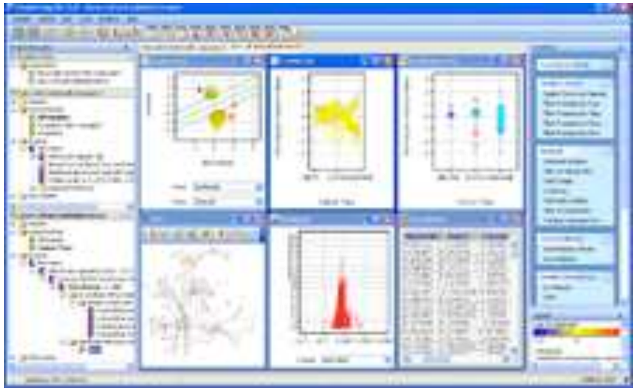
Strand NGS

SureCall

Any other NGS analysis tool

VCF files

discover



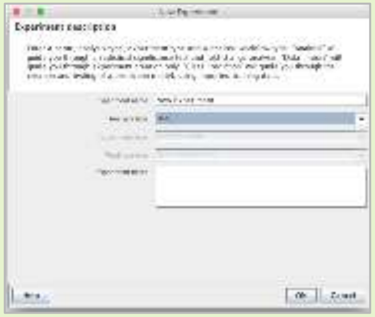
GeneSpring GX

Variant Analysis in GeneSpring GX

Data Import and Filtering

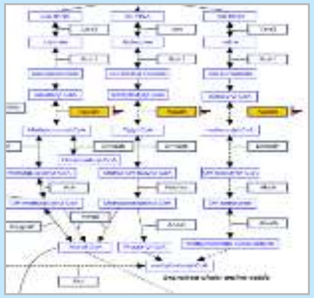
Visualization and Results Interpretation

Pathway Analysis



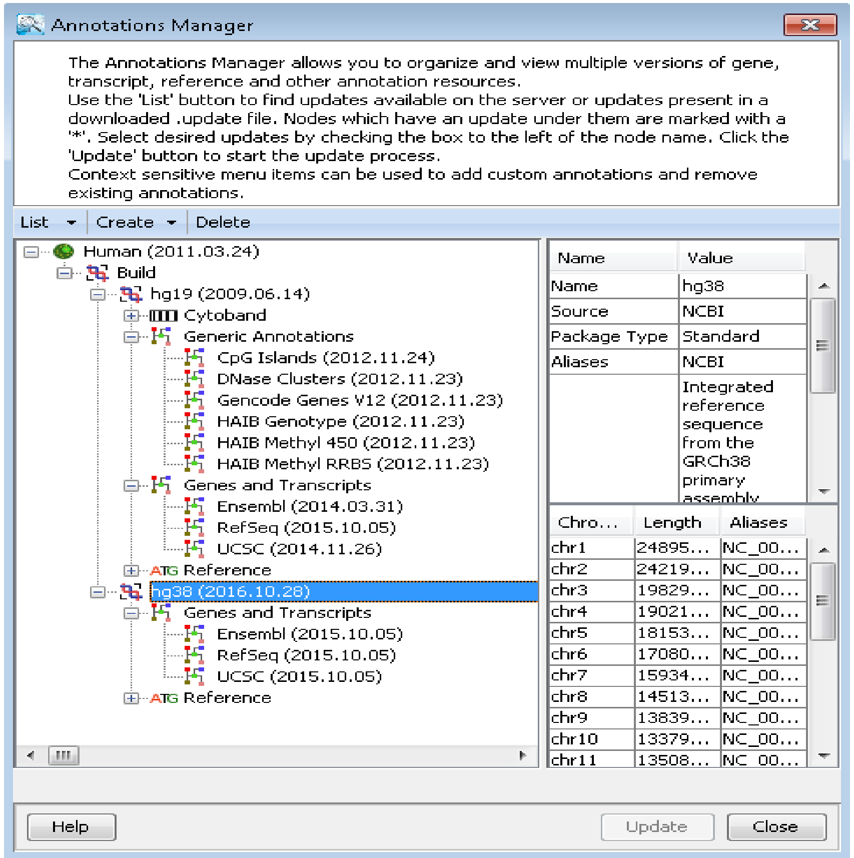
- **Import files in VCF format**
- **Annotate region list**
- **Region list operations**

- **Genome Browser**
- **Find Genic Parts**
- **Gene Ontology analysis**



- **Single experiment analysis**
- **Multi-omic analysis**

Annotations Manager – Use available annotations or import from external sources



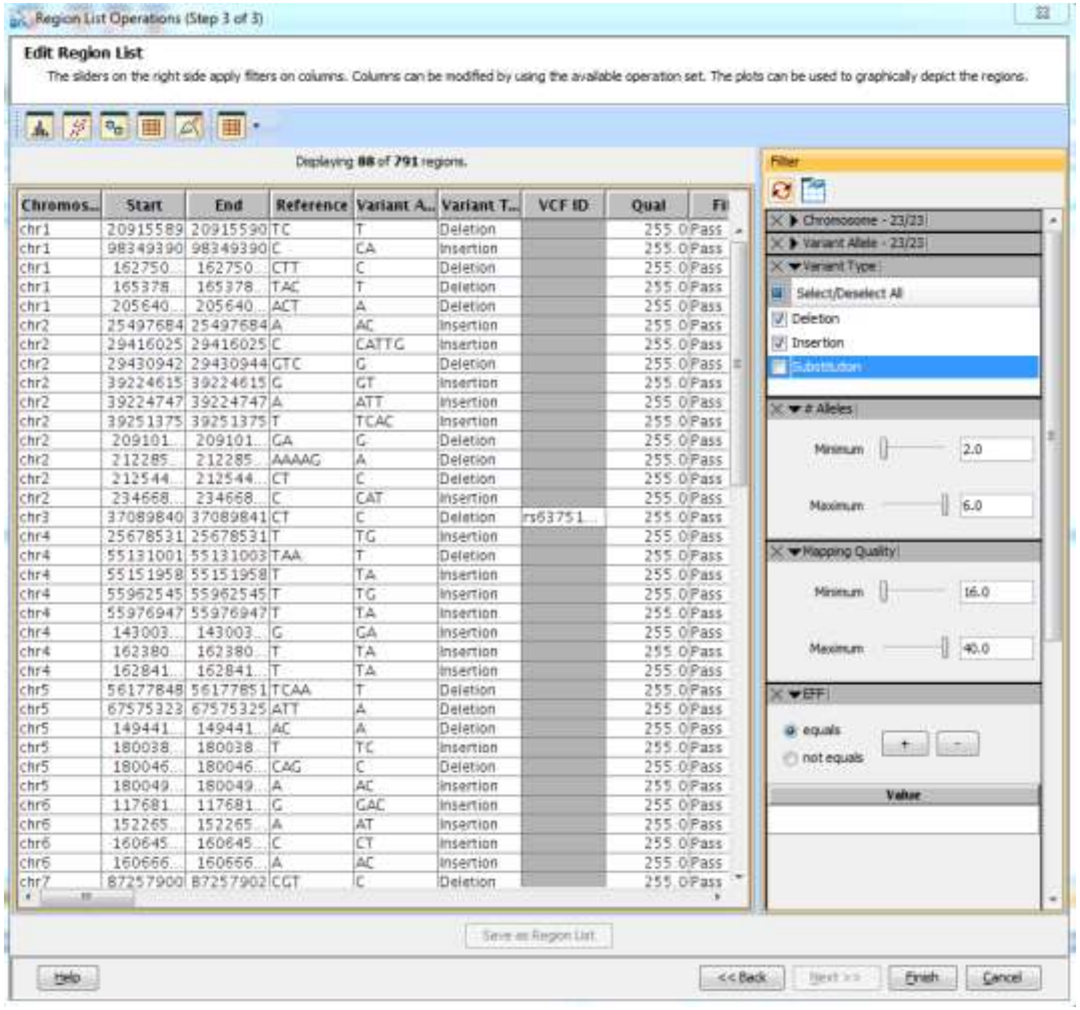
The Annotations Manager allows you to organize and view multiple versions of gene, transcript, reference and other annotation resources.
Use the 'List' button to find updates available on the server or updates present in a downloaded .update file. Nodes which have an update under them are marked with a '*'. Select desired updates by checking the box to the left of the node name. Click the 'Update' button to start the update process.
Context sensitive menu items can be used to add custom annotations and remove existing annotations.

Name	Value
Name	hg38
Source	NCBI
Package Type	Standard
Aliases	NCBI
	Integrated reference sequence from the GRCh38 primary assembly

Chro...	Length	Aliases
chr1	24895...	NC_00...
chr2	24219...	NC_00...
chr3	19829...	NC_00...
chr4	19021...	NC_00...
chr5	18153...	NC_00...
chr6	17080...	NC_00...
chr7	15934...	NC_00...
chr8	14513...	NC_00...
chr9	13839...	NC_00...
chr10	13379...	NC_00...
chr11	13508...	NC_00...

- Import annotations from ClinVar, COSMIC, dbNSFP, 1000 Genomes and more
- Annotate region list using annotations from Annotations Manager
- View annotations in Genome Browser

Filter genetic variants



Region List Operations (Step 3 of 3)

Edit Region List

The sliders on the right side apply filters on columns. Columns can be modified by using the available operation set. The plots can be used to graphically depict the regions.

Displaying **88** of **791** regions.

Chromos...	Start	End	Reference	Variant A...	Variant T...	VCF ID	Qual	Fi
chr1	20915589	20915590	TC	T	Deletion		255.0	Pass
chr1	98349390	98349390	C	CA	Insertion		255.0	Pass
chr1	162750...	162750...	CTT	C	Deletion		255.0	Pass
chr1	165378...	165378...	TAC	T	Deletion		255.0	Pass
chr1	205640...	205640...	ACT	A	Deletion		255.0	Pass
chr2	25497684	25497684	A	AC	Insertion		255.0	Pass
chr2	29416025	29416025	C	CATTG	Insertion		255.0	Pass
chr2	29430942	29430944	GTC	G	Deletion		255.0	Pass
chr2	39224615	39224615	G	GT	Insertion		255.0	Pass
chr2	39224747	39224747	A	ATT	Insertion		255.0	Pass
chr2	39251375	39251375	T	TCAC	Insertion		255.0	Pass
chr2	209101...	209101...	GA	G	Deletion		255.0	Pass
chr2	212285...	212285...	AAAAG	A	Deletion		255.0	Pass
chr2	212544...	212544...	CT	C	Deletion		255.0	Pass
chr2	234668...	234668...	C	CAT	Insertion		255.0	Pass
chr3	37089840	37089841	CT	C	Deletion	rs63751...	255.0	Pass
chr4	25678531	25678531	T	TG	Insertion		255.0	Pass
chr4	55131001	55131003	TAA	T	Deletion		255.0	Pass
chr4	55151958	55151958	T	TA	Insertion		255.0	Pass
chr4	55962545	55962545	T	TG	Insertion		255.0	Pass
chr4	55976947	55976947	T	TA	Insertion		255.0	Pass
chr4	143003...	143003...	G	GA	Insertion		255.0	Pass
chr4	162380...	162380...	T	TA	Insertion		255.0	Pass
chr4	162841...	162841...	T	TA	Insertion		255.0	Pass
chr5	56177848	56177851	TCAA	T	Deletion		255.0	Pass
chr5	67575323	67575325	ATT	A	Deletion		255.0	Pass
chr5	149441...	149441...	AC	A	Deletion		255.0	Pass
chr5	180038...	180038...	T	TC	Insertion		255.0	Pass
chr5	180046...	180046...	CAG	C	Deletion		255.0	Pass
chr5	180049...	180049...	A	AC	Insertion		255.0	Pass
chr6	117681...	117681...	G	GAC	Insertion		255.0	Pass
chr6	152265...	152265...	A	AT	Insertion		255.0	Pass
chr6	160645...	160645...	C	CT	Insertion		255.0	Pass
chr6	160666...	160666...	A	AC	Insertion		255.0	Pass
chr7	87257900	87257902	CGT	C	Deletion		255.0	Pass

Filter

Chromosome - 23/23

Variant Allele - 23/23

Variant Type

☐ Select/Deselect All

☑ Deletion

☑ Insertion

☐ Substitution

\# Alleles

Minimum 2.0

Maximum 6.0

Mapping Quality

Minimum 16.0

Maximum 40.0

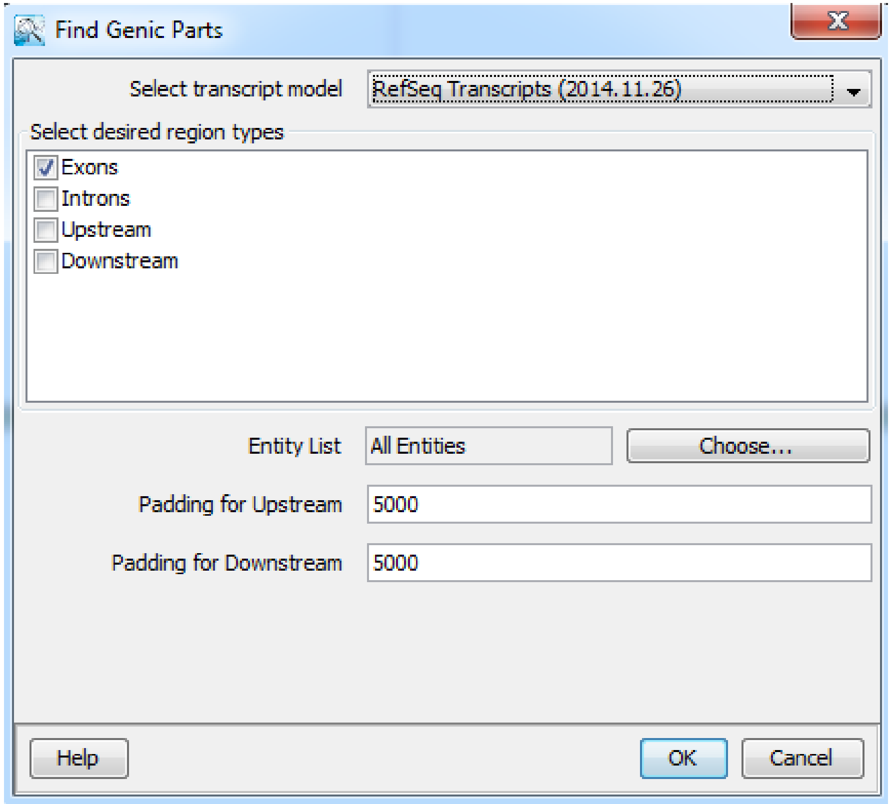
EFF

● equals

○ not equals

- Create a new region list that only contains selected columns
- Filter the input region list based on specified conditions on the retained columns
- Add columns that are functions of selected existing columns

Find Genic Parts



Select transcript model: RefSeq Transcripts (2014.11.26)

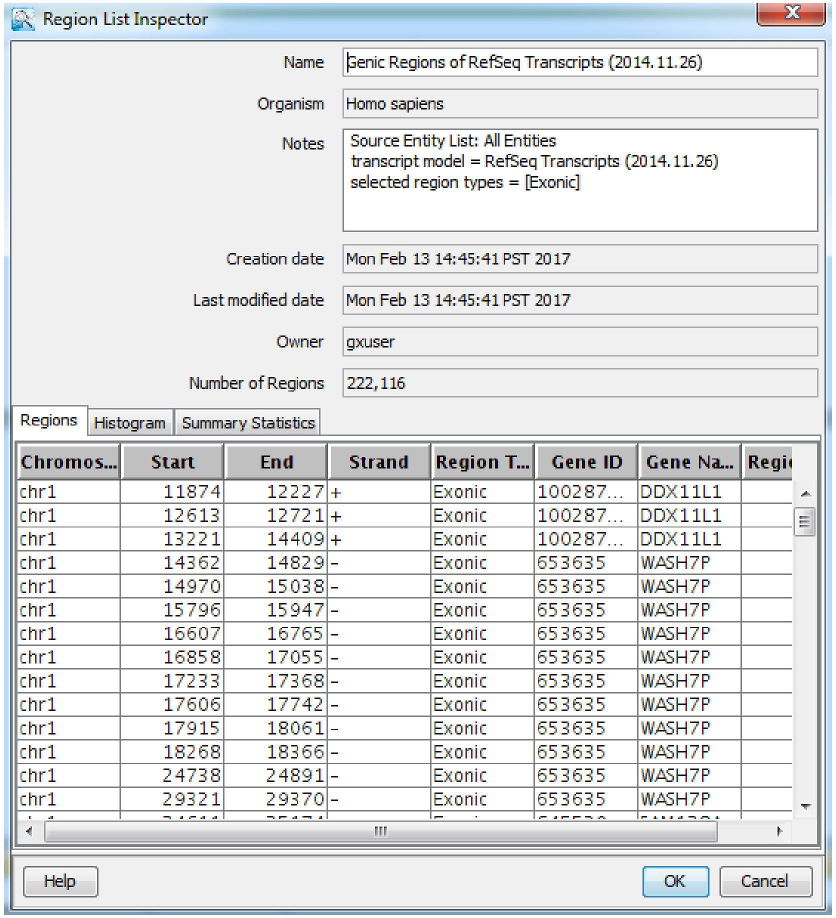
Select desired region types

☑ Exons
☐ Introns
☐ Upstream
☐ Downstream

Entity List: All Entities

Padding for Upstream: 5000

Padding for Downstream: 5000



Name: Genic Regions of RefSeq Transcripts (2014.11.26)

Organism: Homo sapiens

Notes:
Source Entity List: All Entities
transcript model = RefSeq Transcripts (2014.11.26)
selected region types = [Exonic]

Creation date: Mon Feb 13 14:45:41 PST 2017

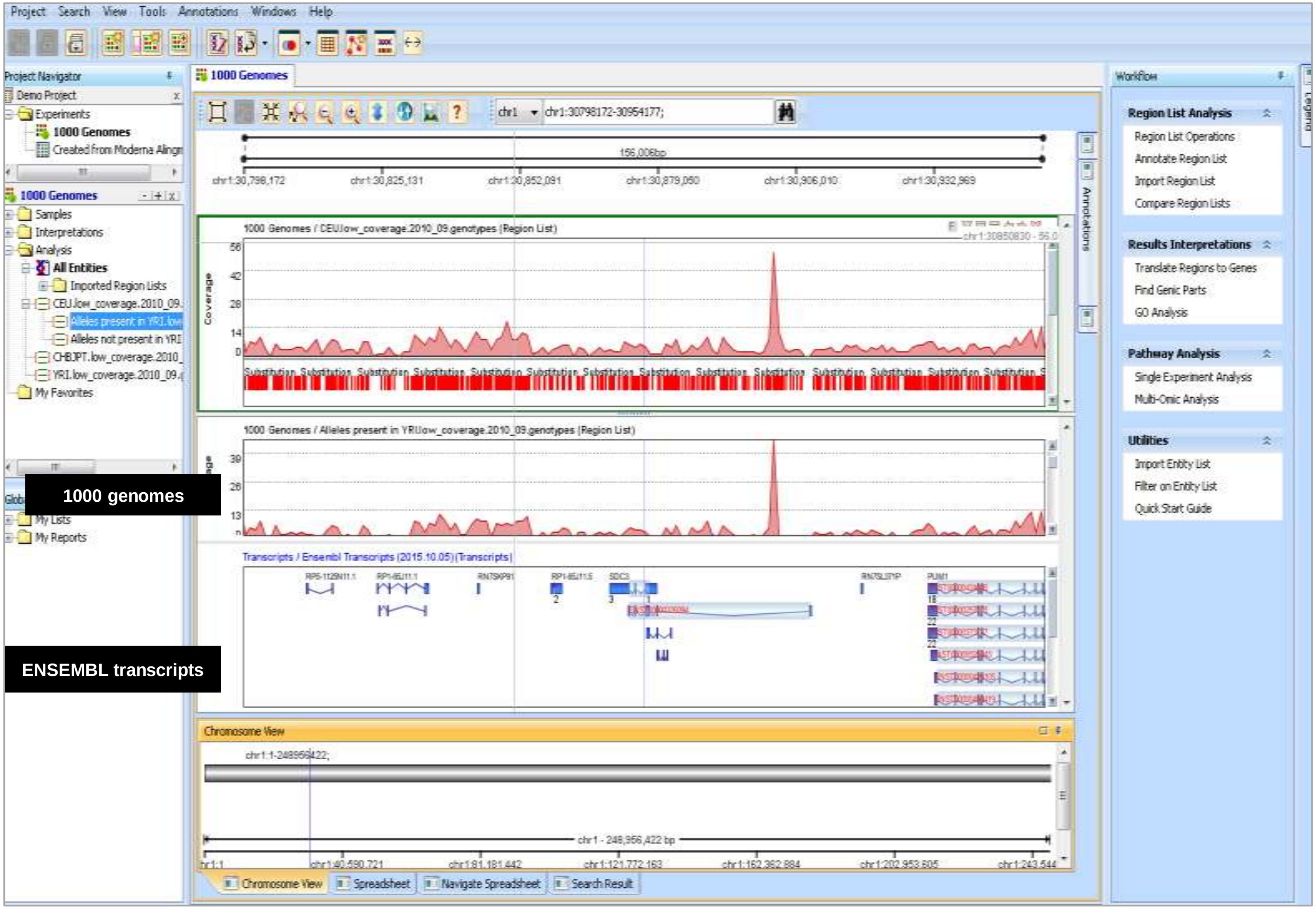
Last modified date: Mon Feb 13 14:45:41 PST 2017

Owner: gxuser

Number of Regions: 222,116

Chromos...	Start	End	Strand	Region T...	Gene ID	Gene Na...	Regi
chr1	11874	12227	+	Exonic	100287...	DDX11L1	
chr1	12613	12721	+	Exonic	100287...	DDX11L1	
chr1	13221	14409	+	Exonic	100287...	DDX11L1	
chr1	14362	14829	-	Exonic	653635	WASH7P	
chr1	14970	15038	-	Exonic	653635	WASH7P	
chr1	15796	15947	-	Exonic	653635	WASH7P	
chr1	16607	16765	-	Exonic	653635	WASH7P	
chr1	16858	17055	-	Exonic	653635	WASH7P	
chr1	17233	17368	-	Exonic	653635	WASH7P	
chr1	17606	17742	-	Exonic	653635	WASH7P	
chr1	17915	18061	-	Exonic	653635	WASH7P	
chr1	18268	18366	-	Exonic	653635	WASH7P	
chr1	24738	24891	-	Exonic	653635	WASH7P	
chr1	29321	29370	-	Exonic	653635	WASH7P	

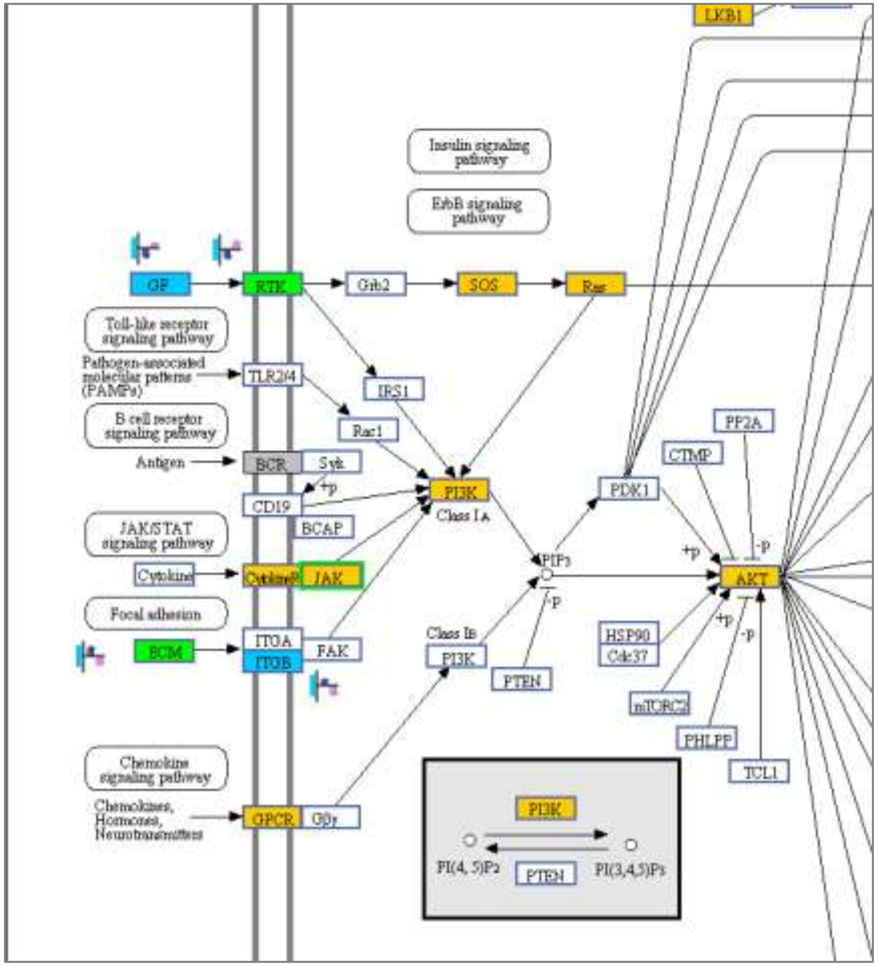
Data Visualization



1000 genomes

ENSEMBL transcripts

Multi-Omic Pathway Analysis



- Multi-omic experiment for data integration
- Overlay gene level information for variants
- Overlay other genomics, transcriptomics, proteomics or metabolomics data