






PhosphoSitePlus Updates


Jon Kornhauser
COMBINE 2012
August 16, 2012

Outline of talk

- Brief intro to PhosphoSitePlus (PSP)
- Improvements in HTP data quality
 - Reanalysis of in-house MS data
- New tools
 - Motif and logo analysis
- Building networks using data
 - Downloadable datasets
 - Cytoscape plugin
 - Building site-centric networks
 - Enhancing didactic pathways
 - Identifying subnetworks represented in MS datasets


PhosphoSitePlus®
Home

with grant support from





from  **Cell Signaling**
TECHNOLOGY®

[ABOUT PHOSPHOSITE](#) | [USING PHOSPHOSITE](#) | [CURATION PROCESS](#) | [CONTACT](#)

PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs). See [About PhosphoSite](#) above for more information.






Please cite the following reference for this resource: Hornbeck PV, et al. (2012) *Nucleic Acids Res.* 2012 40:D261-70. [[reprint](#)]

A PROTEIN MODIFICATION RESOURCE




PROTEIN OR SUBSTRATE SEARCH

Protein Name: **SEARCH**

ADVANCED SEARCH AND BROWSE OPTIONS

-  **Protein, Sequence, or Reference Search**
-  **Site Search**
-  **Comparative Site Search**
-  **Browse MS2 Data By Disease**
-  **Browse MS2 Data by Cell Line**
-  **Browse MS2 Data by Tissue**

DOWNLOADS, LINKS & APPLICATIONS

-  **Reprints, References, Supplemental Tables**
-  **Downloadable Datasets**
-  **Motif Analysis Tools**


WHAT'S NEW

Dec 2011 **Download "PhosphoSitePlus: a comprehensive resource..."** in January 2012 issue of *Nucleic Acids Research*.

Jul 2011 **Multiple Sequence Alignment (MSA)** added to the Protein Page.

Jul 2011 **Download PyMOL & Chimera Scripts** from the Structure Viewer window.

Jul 2011 **New Tutorial** for Navigating PhosphoSitePlus®.


Apr 2011 **New Icon** indicates links to protein-specific reagents from CST™ 

Phosphorylation Site Statistics

Non-redundant sites:	207,715
Non-redundant proteins:	19,451
Sites curated from literature:	104,199
All sites using site-specific (SS) methods:	11,396
All sites using discovery-mode MS (MS) methods:	95,459
Sites using both SS and MS methods:	5,488
MS sites observed at CST:	156,289
Number of curated papers:	14,207


Other Modification Site Statistics

Acetylation:	24,502	Caspase cleavage:	481
Di-methylation:	1,679	Methylation:	143
Mono-methylation:	3,407	Neddylation:	39
O-GlcNAc:	617	Sumoylation:	705
Tri-methylation:	231	Ubiquitination:	39,224



PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at <http://www.phosphosite.org/staticContact.do>.

[Home](#) | [My Account](#) | [Logout](#)

With enhanced literature mining using Linguamatics i2E 

Produced by 3rd Millennium | Design by Digizyme
©2003-2012 Cell Signaling Technology, Inc.

Protein modification sites in PSP: HTP and LTP

Phosphorylation Site Statistics

Non-redundant sites:	207,714
Non-redundant proteins:	19,453
Sites curated from literature:	104,191
All sites using site-specific (SS) methods:	11,386
All sites using discovery-mode MS (MS) methods:	95,462
Sites using both SS and MS methods:	5,488
MS sites observed at CST:	156,293
Number of curated papers:	14,197

Other Modification Site Statistics

Acetylation:	24,511	Caspase cleavage:	481
Di-methylation:	1,678	Methylation:	139
Mono-methylation:	3,407	Neddylaton:	39
O-GlcNAc:	617	Sumoylation:	705
Tri-methylation:	231	Ubiquitination:	39,225

Distinguishing MS data on protein page

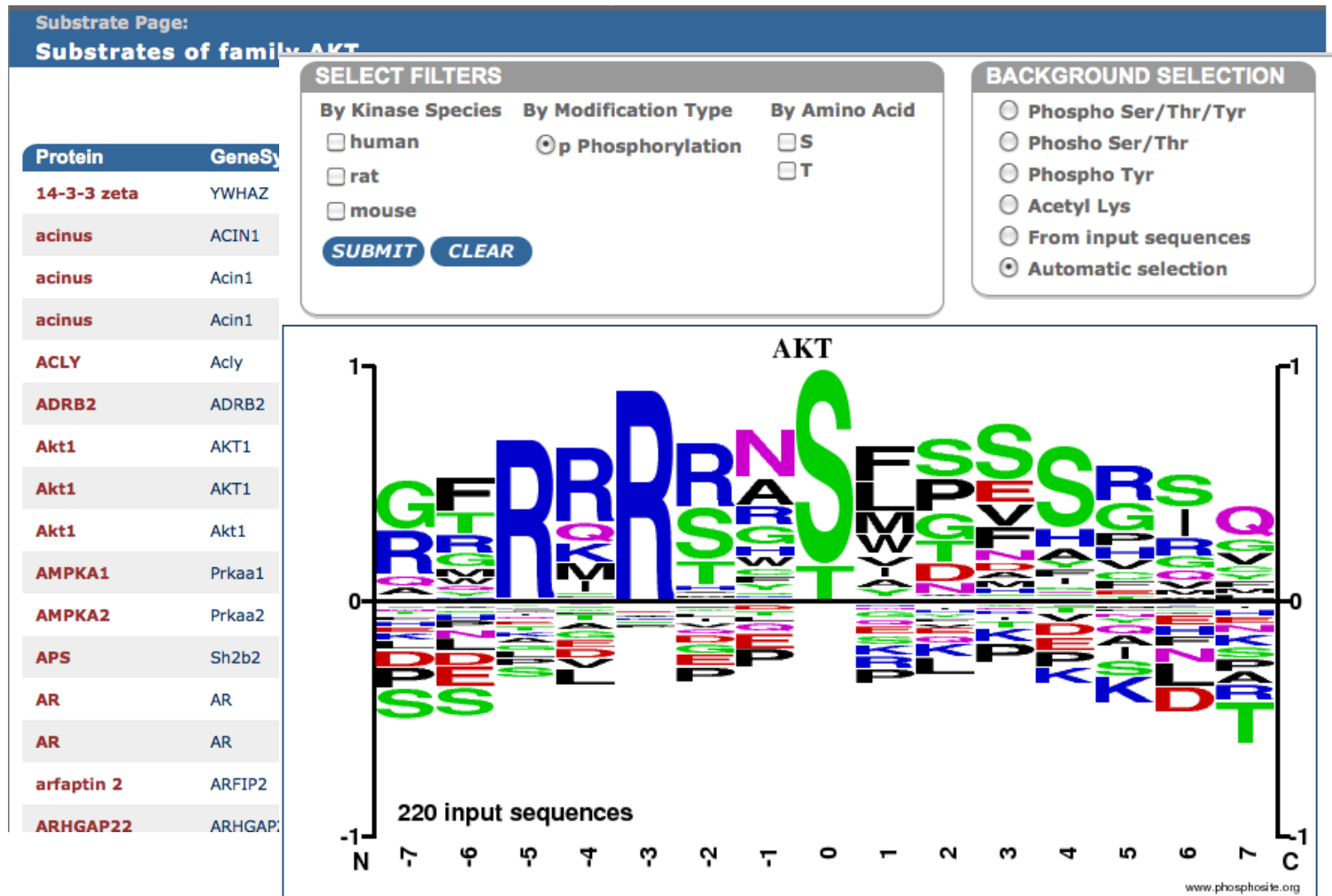
- Separate columns in site tables
- Options to hide sites with only 1 MS reference, or less than 5 references total

Reanalysis of CST MS data

- Peptide spectra being re-evaluated using improved criteria:
 - $Ascore > 13$
 - Better site localization
- Less robust sites will be removed
- Published sites are evaluated using similar criteria when available

New tools

- Motif/Logo Analysis Tools



Motif/Logo analysis

**PhosphoSitePlus®**
with grant support from    

from  **Cell Signaling TECHNOLOGY®**

[Home](#) [ABOUT PHOSPHOSITE](#) [USING PHOSPHOSITE](#) [CURATION PROCESS](#) [CONTACT](#)

PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs). See [About PhosphoSite](#) above for more information.

Please cite the following reference for this resource: Hornbeck PV. et al. (2012) *Nucleic Acids Res.* 2012 40:D261-70. [[reprint](#)]

A PROTEIN MODIFICATION RESOURCE



PROTEIN OR SUBSTRATE SEARCH

Protein Name: **SEARCH**

ADVANCED SEARCH AND BROWSE OPTIONS

-  **Protein, Sequence, or Reference Search**
-  **Site Search**
-  **Comparative Site Search**
-  **Browse MS2 Data By Disease**
-  **Browse MS2 Data by Cell Line**
-  **Browse MS2 Data by Tissue**

DOWNLOADS, LINKS & APPLICATIONS

-  **Reprints, References, Supplemental Tables**
-  **Downloadable Datasets**
-  **Motif Analysis Tools**

WHAT'S NEW

Dec 2011 **Download "PhosphoSitePlus: a comprehensive resource..."** in January 2012 issue of *Nucleic Acids Research*.

Jul 2011 **Multiple Sequence Alignment (MSA)** added to the Protein Page.

Jul 2011 **Download PyMOL & Chimera Scripts** from the Structure Viewer window.

Jul 2011 **New Tutorial** for Navigating PhosphoSitePlus®.

Apr 2011 **New Icon** indicates links to protein-specific reagents from CST™ 

Phosphorylation Site Statistics

Non-redundant sites:	207,715
Non-redundant proteins:	19,451
Sites curated from literature:	104,199
All sites using site-specific (SS) methods:	11,396
All sites using discovery-mode MS (MS) methods:	95,459
Sites using both SS and MS methods:	5,488
MS sites observed at CST:	156,289
Number of curated papers:	14,207

Other Modification Site Statistics

Acetylation:	24,502	Caspase cleavage:	481
Di-methylation:	1,679	Methylation:	143
Mono-methylation:	3,407	Neddylation:	39
O-GlcNAc:	617	Sumoylation:	705
Tri-methylation:	231	Ubiquitination:	39,224

 PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at <http://www.phosphosite.org/staticContact.do>.

[Home](#) | [My Account](#) | [Logout](#) With enhanced literature mining using Linguamatics i2E  Produced by 3rd Millennium | Design by Digizyme
©2003-2012 Cell Signaling Technology, Inc.

- Motif/Logo Analysis Tools

Motif & Logo Analysis Tools

Motif Analysis Generators

Sequence Logo Generators

Foreground data set: submit prealigned sequences of 15 residues or less (15, 13, 11, etc). The modification site needs to be centrally located for correct statistical analysis.

Paste Sequence Data:

Upload Sequence File:

Motif Analysis Generator Settings

Title:

Algorithm: ☒ [MotifAll](#) ☐ Exhaustive 1-2 AA

Background based on: ☐ [phSer](#) Proteins ☐ [phThr](#) Proteins
☐ [phTyr](#) Proteins ☐ [acLys](#) Proteins
☒ Automatic Selection ☐ [UbLys](#) Proteins

Significance:
E.g. 1e-06 or 0.000001

Support Threshold:

Upload Background:

Motif & Logo Analysis Tools

Motif Analysis Generators

Sequence Logo Generators

Foreground data set: submit prealigned sequences of 15 residues or less (15, 13, 11, etc). The modification site needs to be centrally located for correct statistical analysis.

Paste Sequence Data:

Upload Sequence File:

PSP Logo Generator Settings

Title:

Logo Size per Line: X

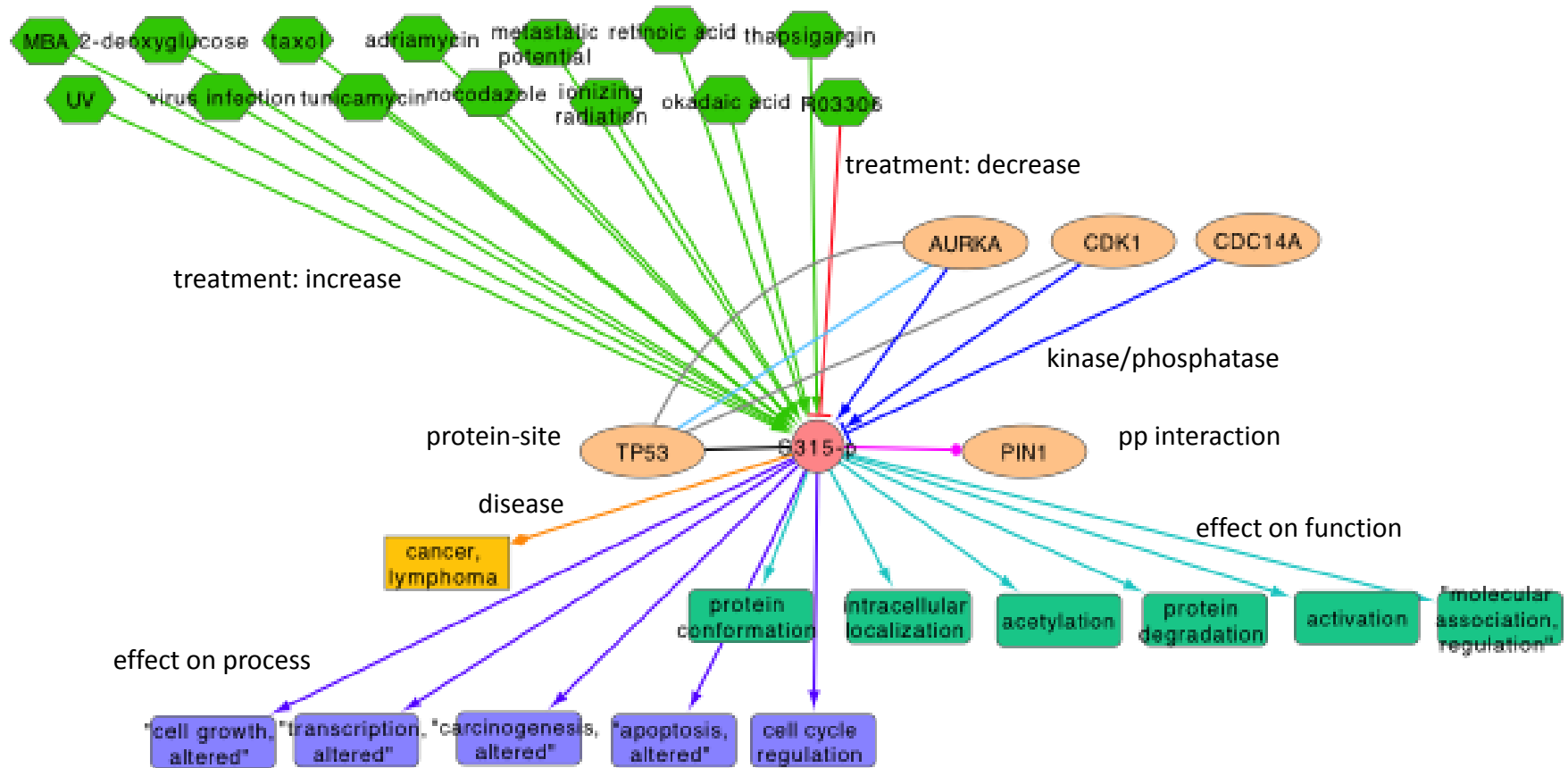
Background: ☐ [Phospho Ser](#) ☐ [Phospho Thr](#)
☐ [Phospho Tyr](#) ☐ [Acetyl Lys](#)
☐ From input sequences ☐ [Ubiquitin Lys](#)

Algorithm: ☒ [PSP Production](#) ☐ [Frequency Change](#)

First Position Number:

Logo Range: -

PSP data content provides interactions at the PTM level: Site-centric networks






Exploiting site-specific data in PSP





- Most interaction databases are protein-centric (e.g., protein-protein interactions)
- Data curated in PSP is unique in providing interactions at the modification site level
- Opportunity for more granular analysis of signaling networks; may provide greater insight into biological processes and disease


Exporting site-specific data for analysis



Home

with grant support from

from  Cell Signaling TECHNOLOGY

ABOUT PHOSPHOSITE | USING PHOSPHOSITE | CURATION PROCESS | CONTACT

PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs). See [About PhosphoSite](#) above for more information.







Please cite the following reference for this resource: Hornbeck PV. et al. (2012) *Nucleic Acids Res.* 2012 40:D261-70. [[reprint](#)]

A PROTEIN MODIFICATION RESOURCE




PROTEIN OR SUBSTRATE SEARCH

Protein Name: [SEARCH](#)


ADVANCED SEARCH AND BROWSE OPTIONS

-  Protein, Sequence, or Reference Search
-  Site Search
-  Comparative Site Search
-  Browse MS2 Data By Disease
-  Browse MS2 Data by Cell Line
-  Browse MS2 Data by Tissue


DOWNLOADS, LINKS & APPLICATIONS

-  Reprints, References, Supplemental Table
-  Downloadable Datasets
-  Motif Analysis Tools

Name	Size	Last Modified
Acetylation_site_dataset.gz	648 KB	Tue Jul 31 11:47:16 EDT 2012
Disease-associated_sites.gz	42 KB	Tue Jul 31 11:47:16 EDT 2012
Kinase_Substrate_Dataset.gz	361 KB	Tue Jul 31 11:47:17 EDT 2012
BioPAX:Kinase-substrate information	4 MB	Mon Jul 16 08:19:49 EDT 2012
Methylation_site_dataset.gz	162 KB	Tue Jul 31 11:47:18 EDT 2012
O-GlcNAc_site_dataset.gz	15 KB	Tue Jul 31 11:47:18 EDT 2012
Phosphorylation_site_dataset.gz	4 MB	Tue Jul 31 11:47:18 EDT 2012
Phosphosite_seq.txt.gz	17 MB	Tue Jul 31 11:47:19 EDT 2012
Regulatory_sites.gz	604 KB	Tue Jul 31 11:47:19 EDT 2012
Sumoylation_site_dataset.gz	20 KB	Tue Jul 31 11:47:20 EDT 2012
Ubiquitination_site_dataset.gz	1012 KB	Tue Jul 31 11:47:20 EDT 2012
phosphositeClient.jar	35 KB	Tue Jul 31 11:47:19 EDT 2012


 PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at <http://www.phosphosite.org/staticContact.do>.

[Home](#) | [My Account](#) | [Logout](#)

With enhanced literature mining using Linguistics I2E 

Produced by 3rd Millennium | Design by Digizyme

©2003-2012 Cell Signaling Technology, Inc.

<http://www.phosphosite.org/staticDownloads.do>



PhosphoSitePlus®

with grant support from



from



Cell Signaling
TECHNOLOGY®

Home

ABOUT PHOSPHOSITE

USING PHOSPHOSITE

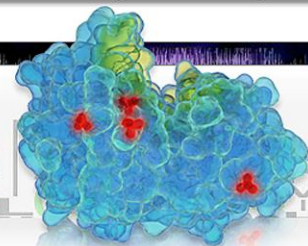
CURATION PROCESS

CONTACT

PhosphoSitePlus® (PSP) is an online systems biology resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs). See [About PhosphoSite](#) above for more information.

Please cite the following reference for this resource: Hornbeck PV. et al. (2012) *Nucleic Acids Res.* 2012 40:D261-70. [[reprint](#)]

A PROTEIN MODIFICATION RESOURCE



WHAT'S NEW

Jul 2012 Download Datasets of Regulatory or Disease-Associated Sites.

Dec 2011 Download "PhosphoSitePlus: a comprehensive resource..." in

Modification sites that regulate cellular processes, molecular functions and molecular interactions. Metadata includes sequences, PMIDs, curator's notes, etc.

Jul 2011 Download PyMOL & Chimera Scripts from the Structure Viewer

Modification sites that correlate with specific disease states. Metadata includes sequences, PMIDs, curator's notes, etc.

DOWNLOADS, LINKS & APPLICATIONS



Reprints, References, Supplemental Tables



Downloadable Datasets



Motif Analysis Tools

Number of curated papers:

14,167

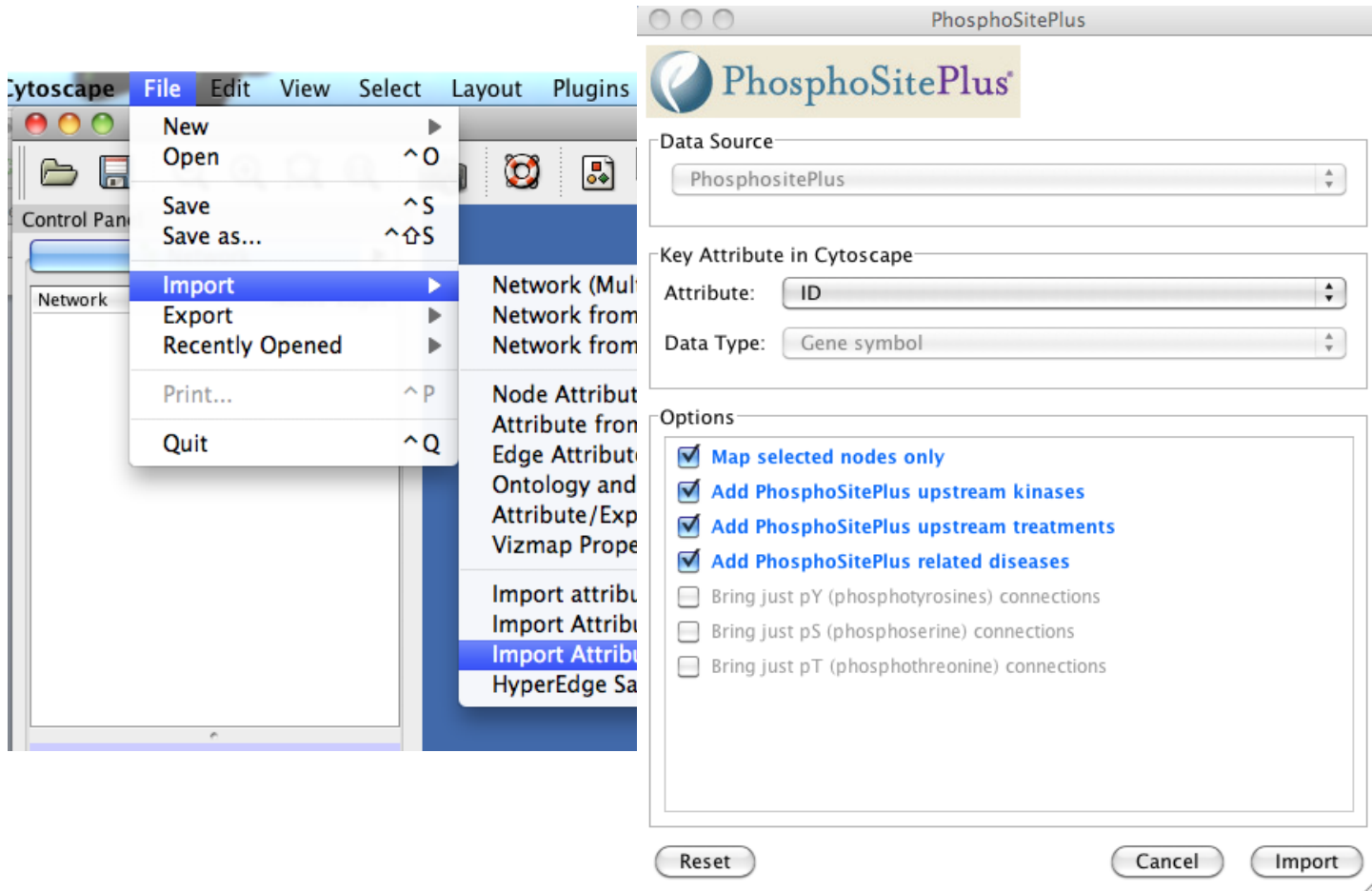
Other Modification Site Statistics

Acetylation:	23,778	Caspase cleavage:	481
Di-methylation:	1,679	Methylation:	143
Mono-methylation:	3,407	O-GlcNAc:	617
Sumoylation:	705	Tri-methylation:	231
Ubiquitination:	39,209		

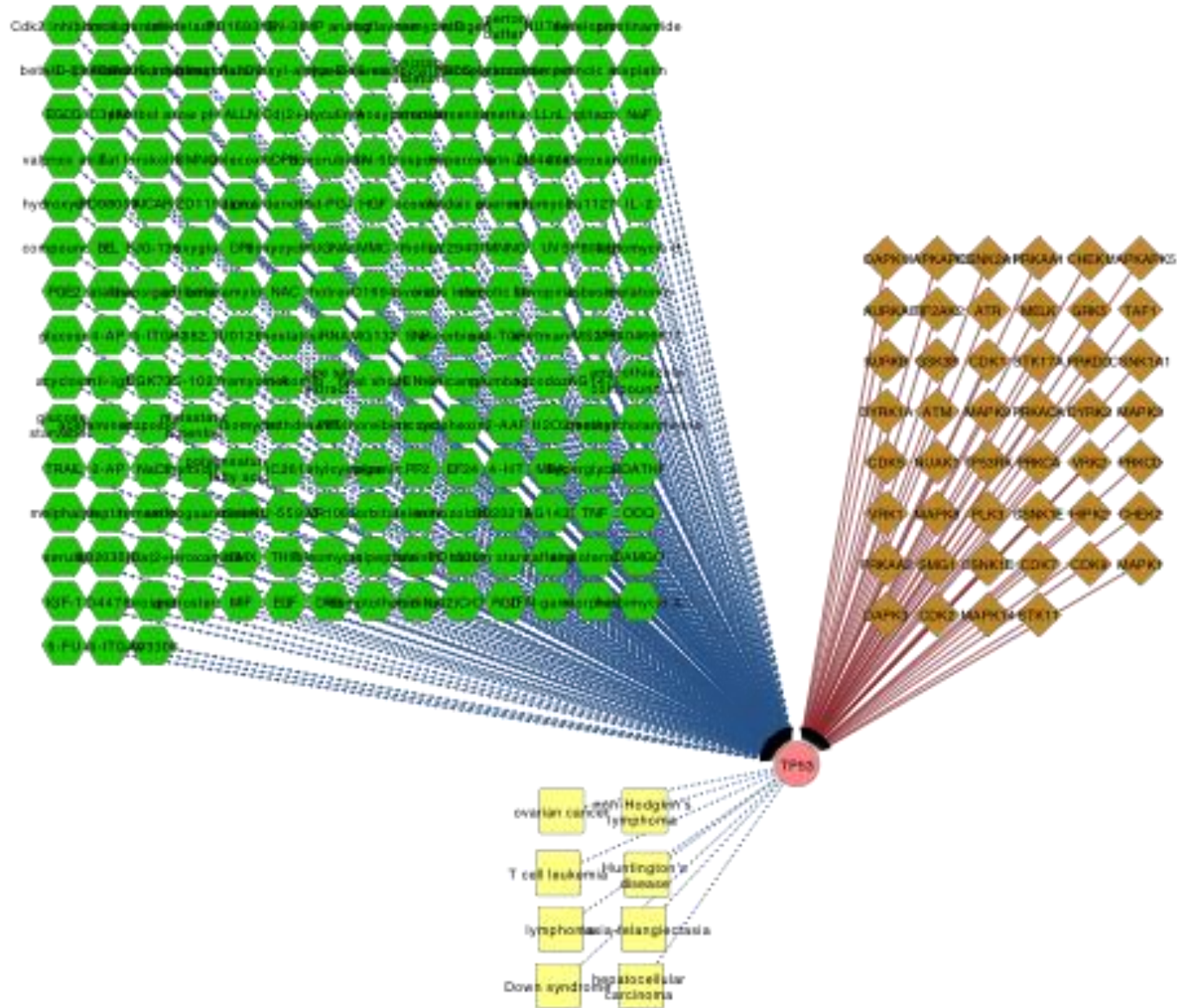


PhosphoSite, created by Cell Signaling Technology is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Information about permissions beyond the scope of this license are available at <http://www.phosphosite.org/staticContact.do>.

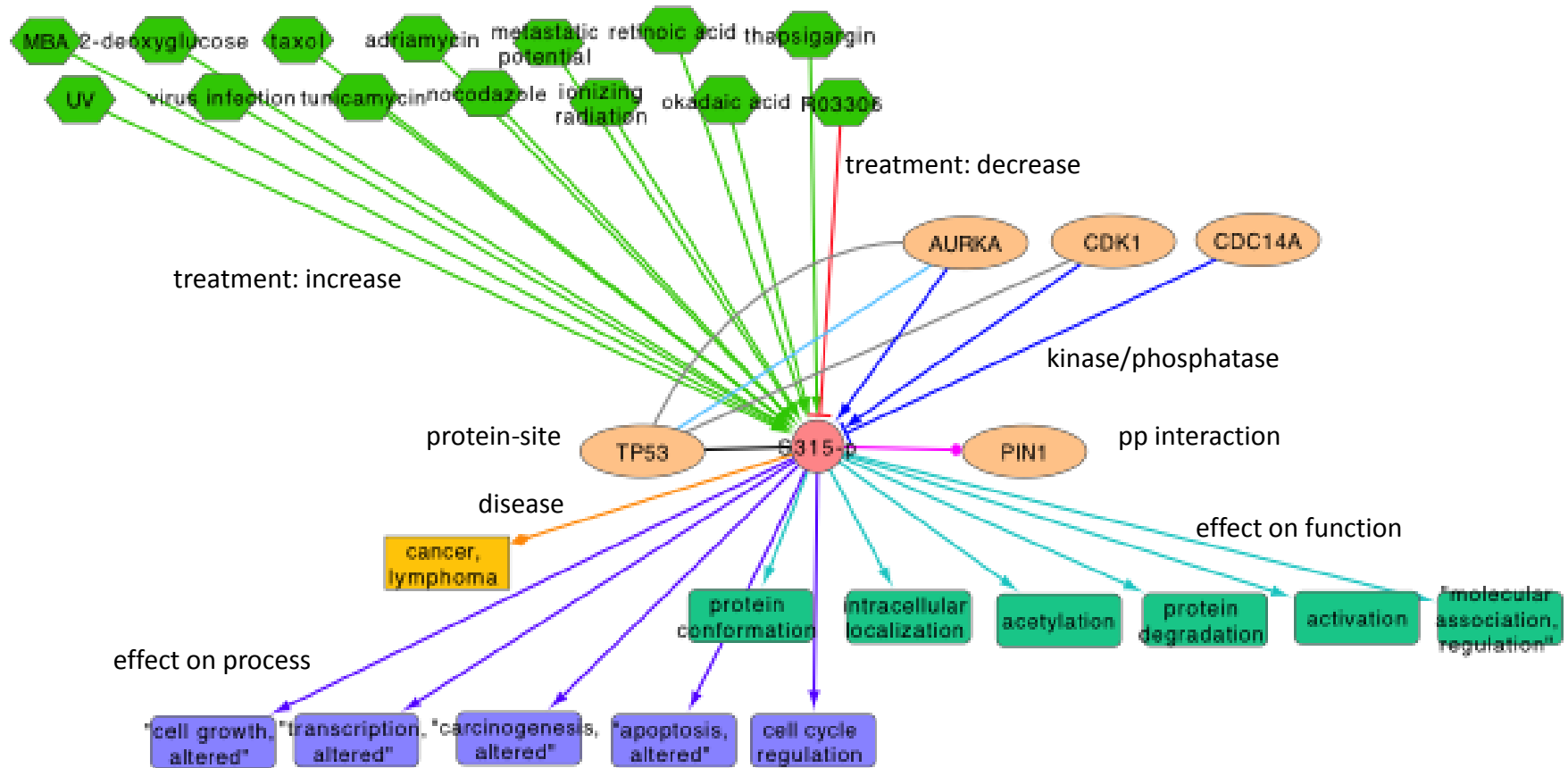
PhosphoSite plugin for Cytoscape



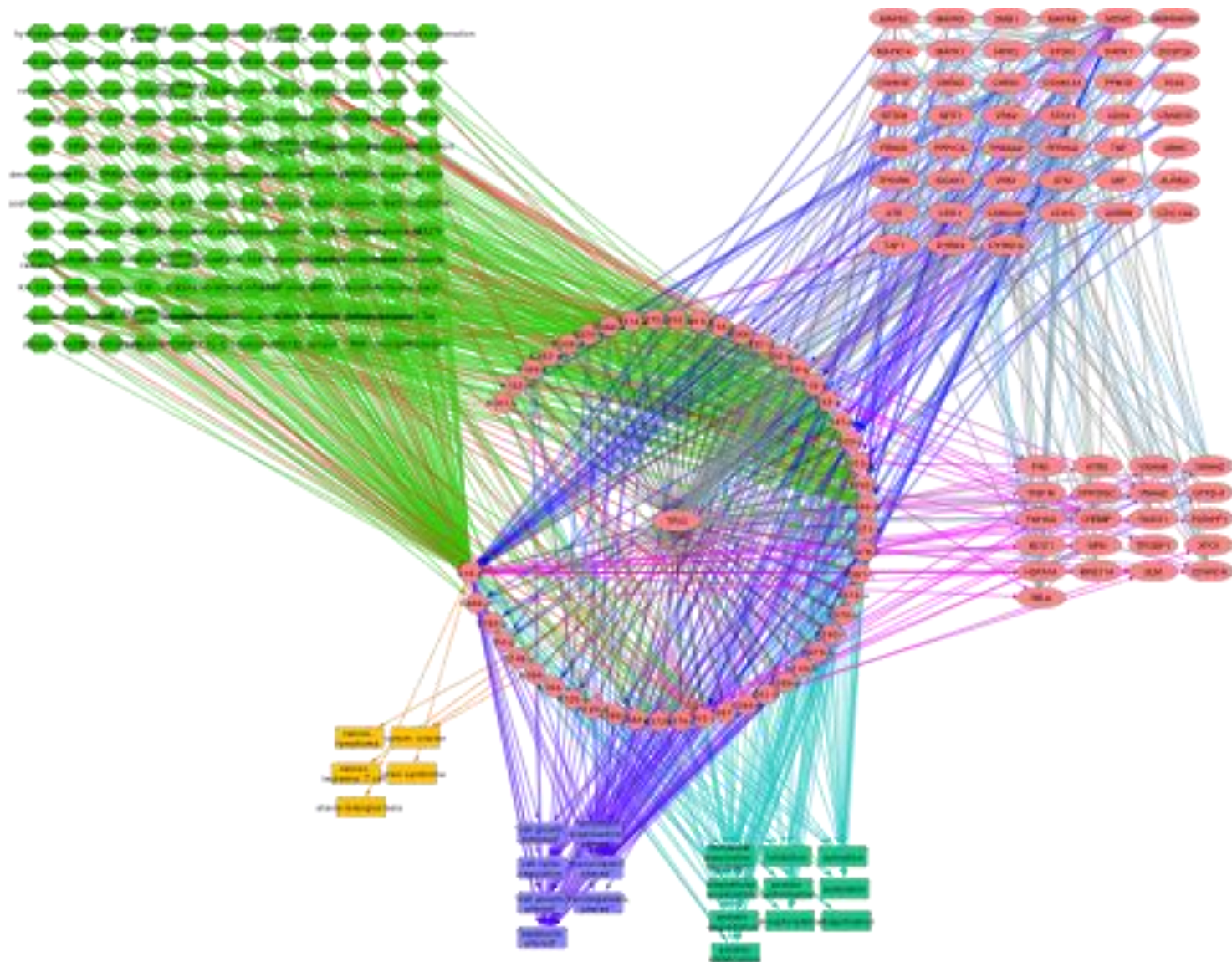
PSP plugin builds protein-centric network



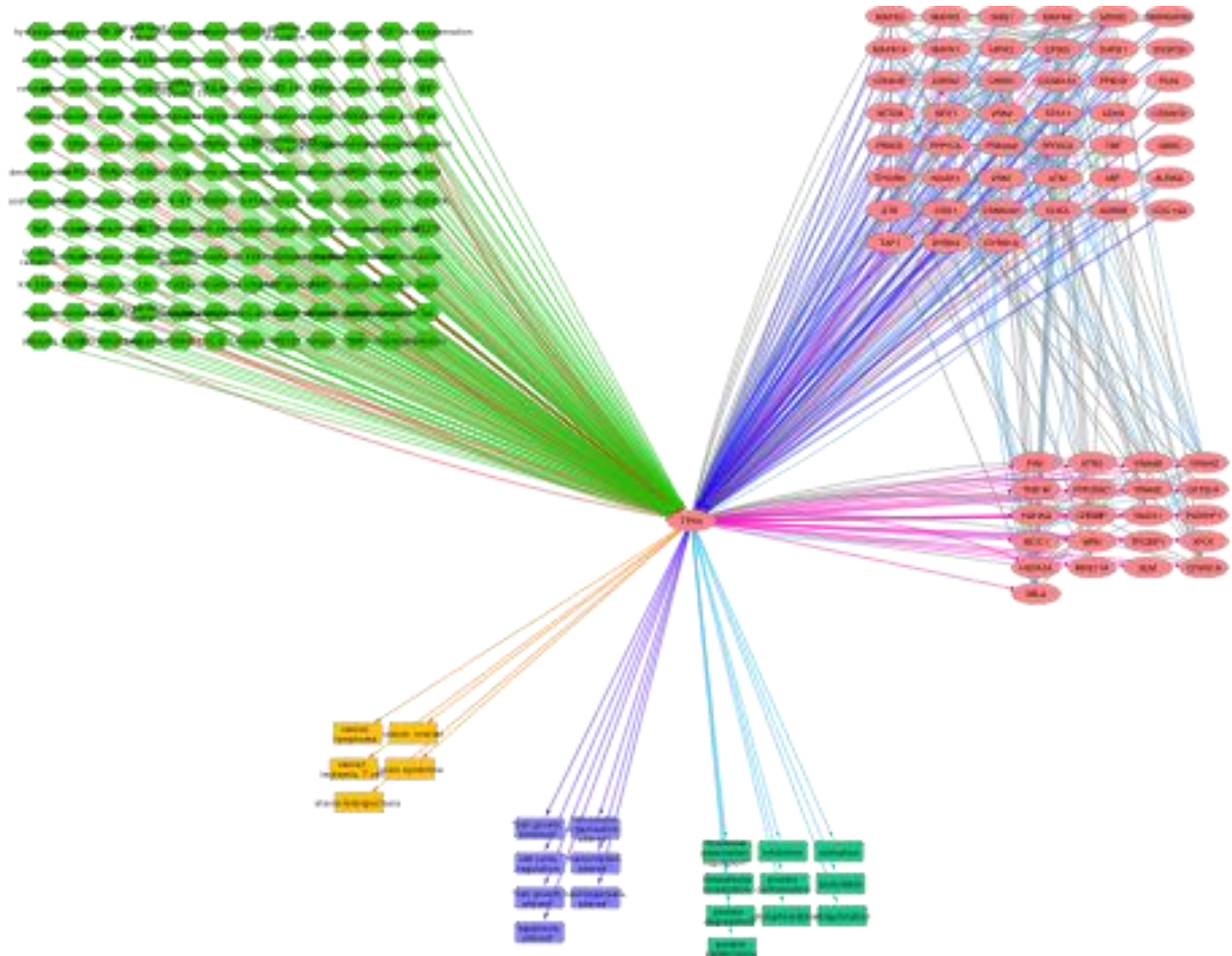
PSP data content provides interactions at the PTM level: Site-centric networks



p53 site-specific network



Using metanodes to simplify visualization



Site-centric PSP network

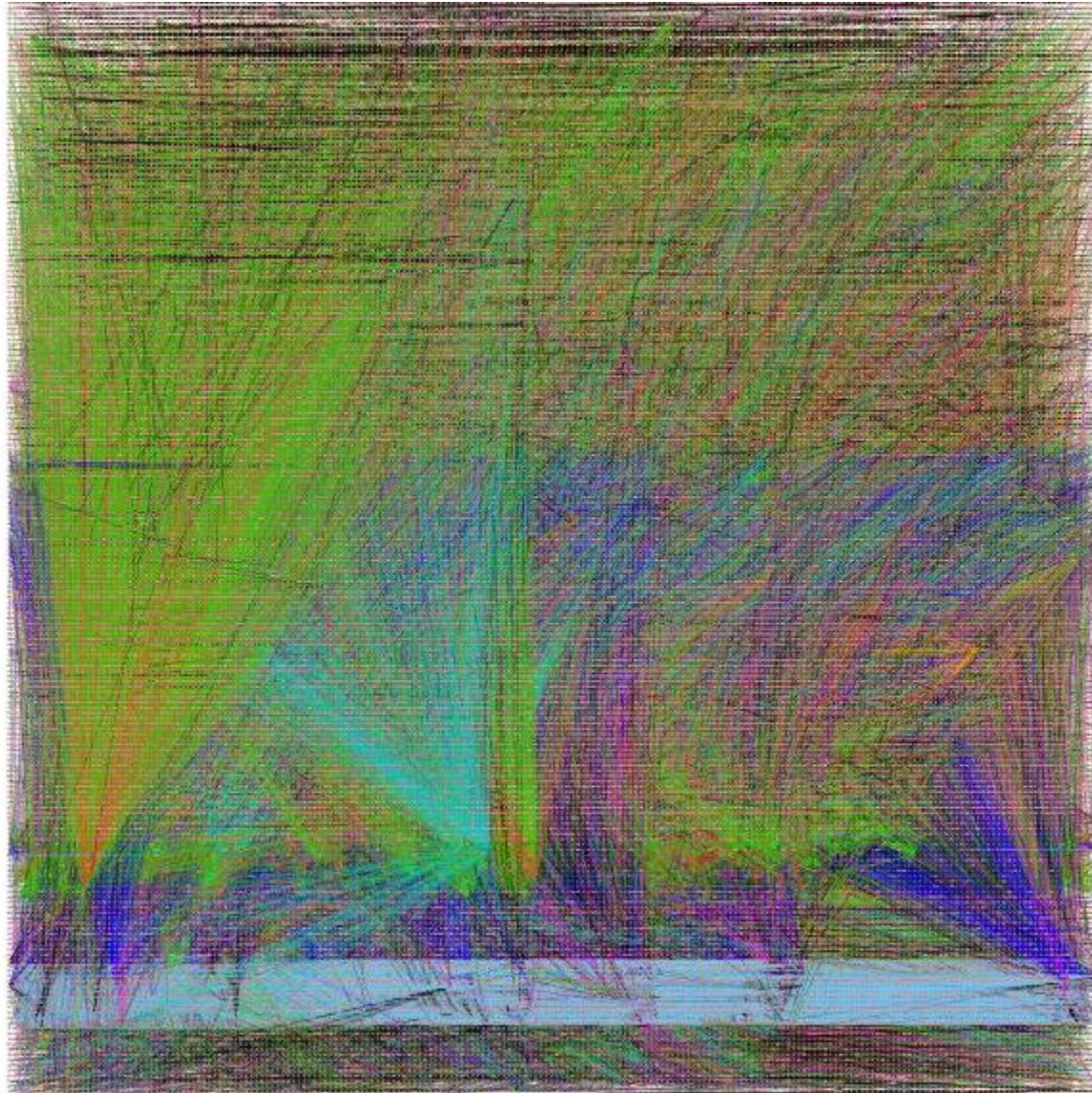
nodes:

- 23,919 sites (only “annotated” subset)
- 17,716 proteins
- 1,473 treatments
- 109 diseases
- 26 effects on function
- 34 effects on process

edges:

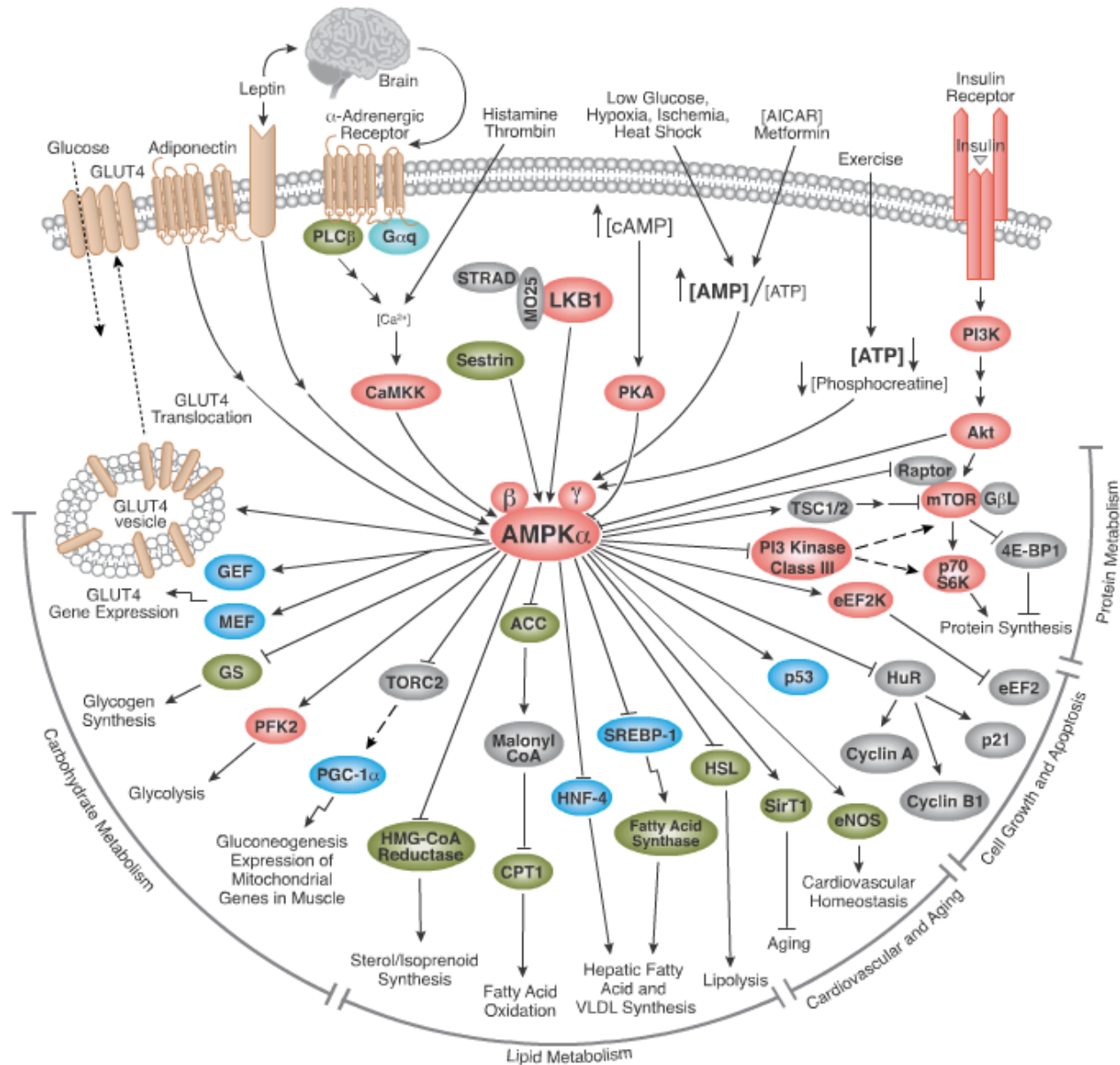
- 91,569 from PSP
- 72,174 from external DBs* (GeneMania plugin)

Entire site-centric network

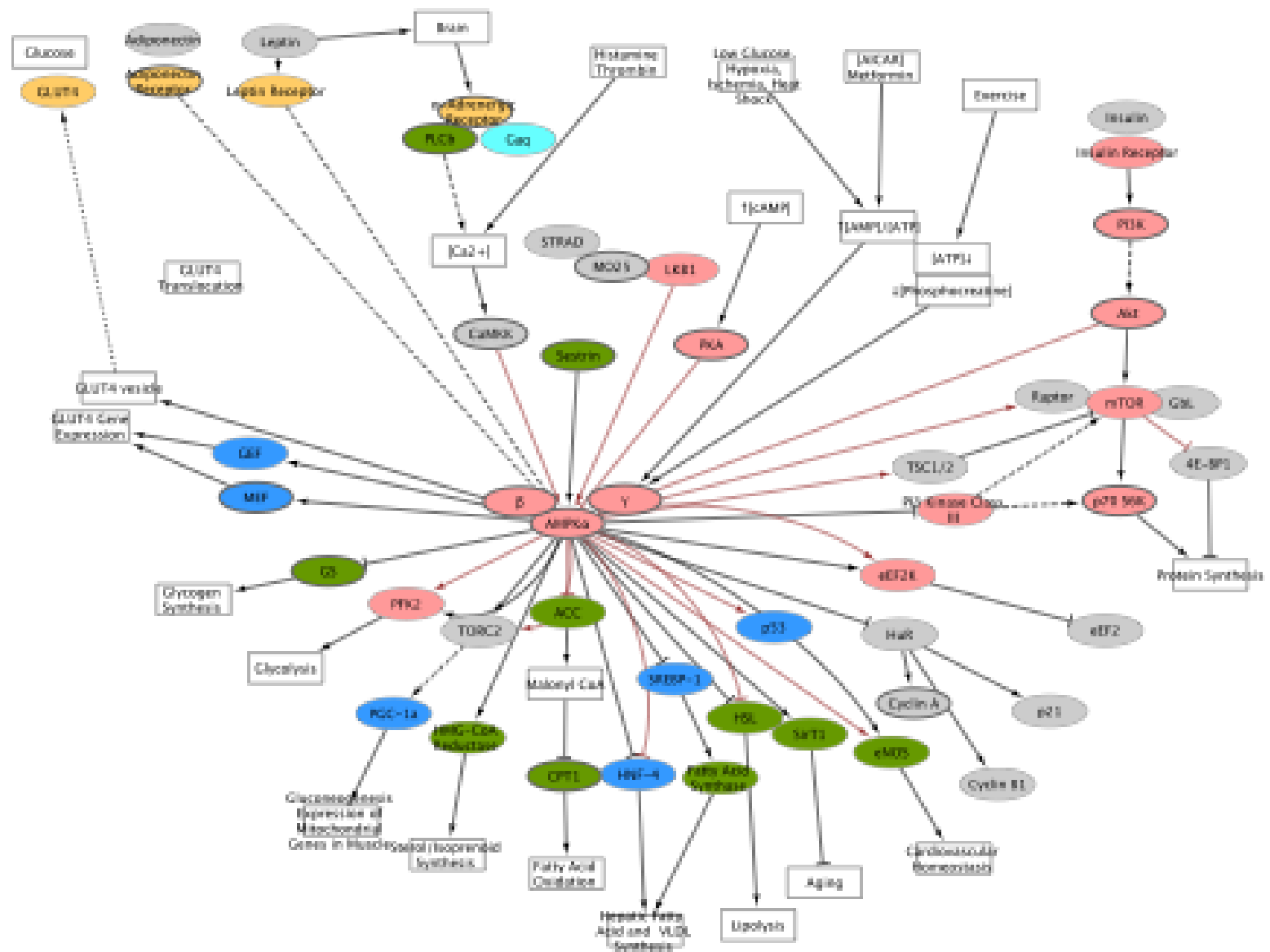


A didactic pathway

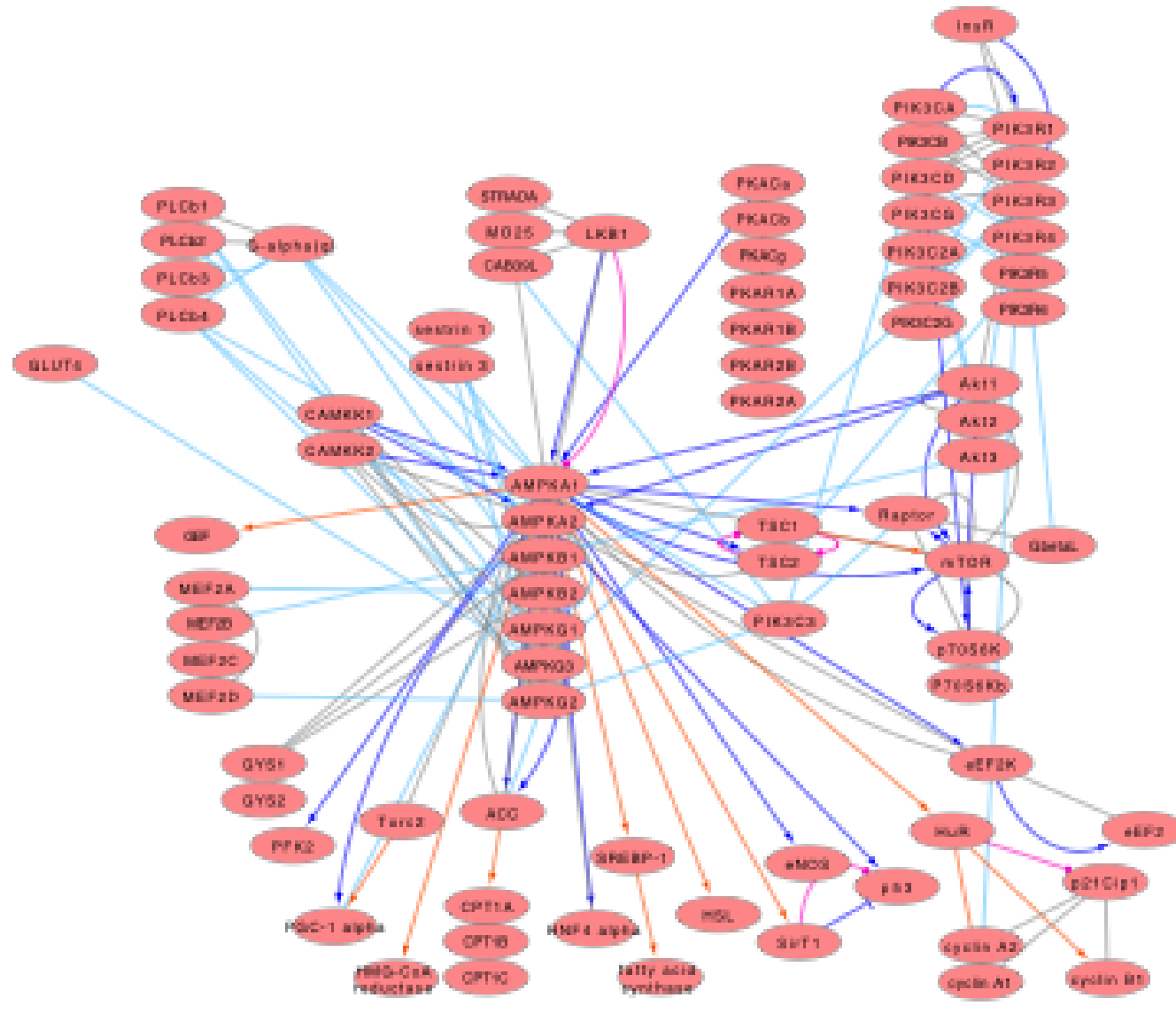
AMPK Signaling



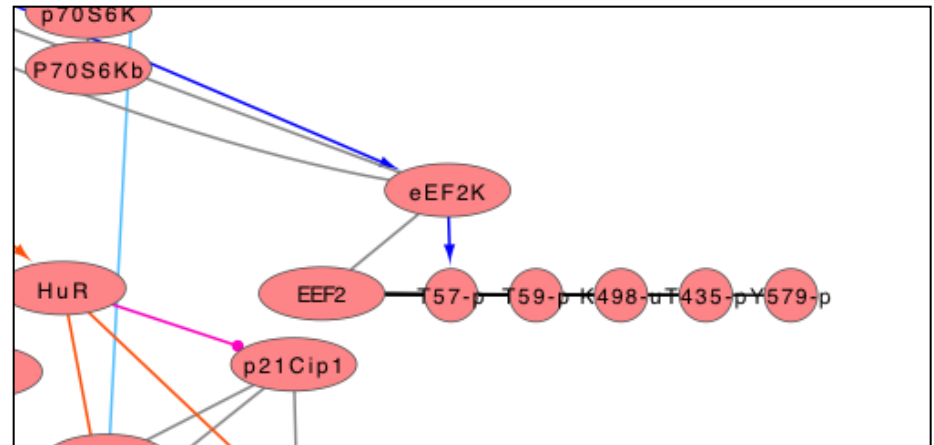
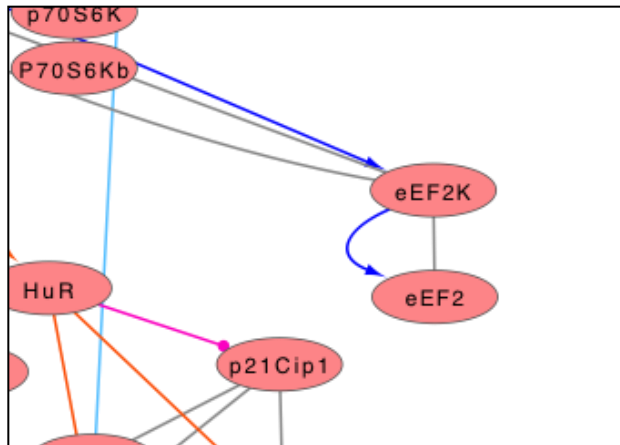
Representation in Cytoscape (protein-centric)



Site-centric representation (w/ metanodes)



Site nodes contain modsite sequences as attributes



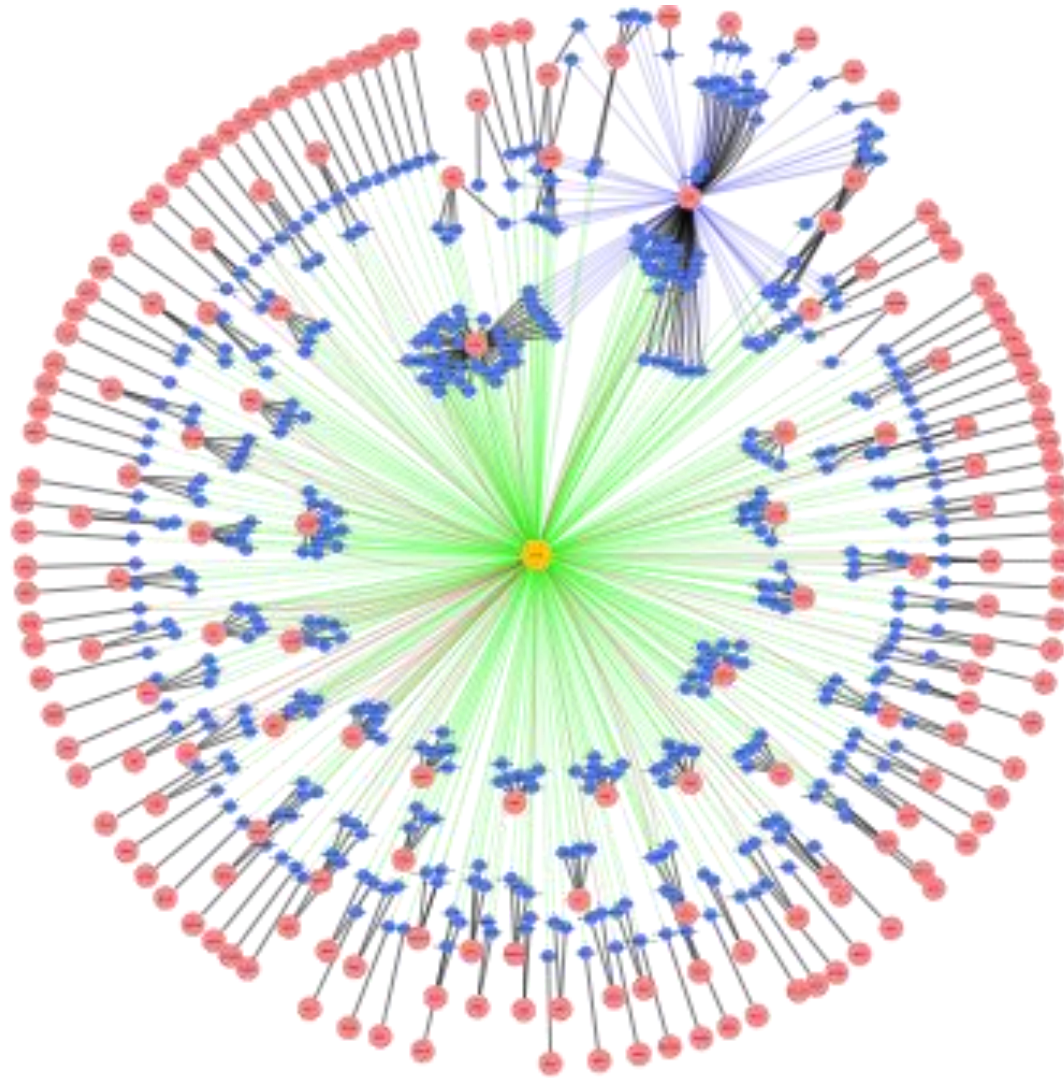
Data Panel

H_PROTEIN-MODSITE	NMER
eEF-2 T59-p	GETRFTDTRKDEQER
eEF-2 T57-p	RAGETRFTDTRKDEQ
eEF-2 K498-u	AHNMRVMKFSVSPVV
eEF-2 Y579-p	KSDPVVSYRETVSEE
eEF-2 T435-p	RIMGPNYTPGKKEDL

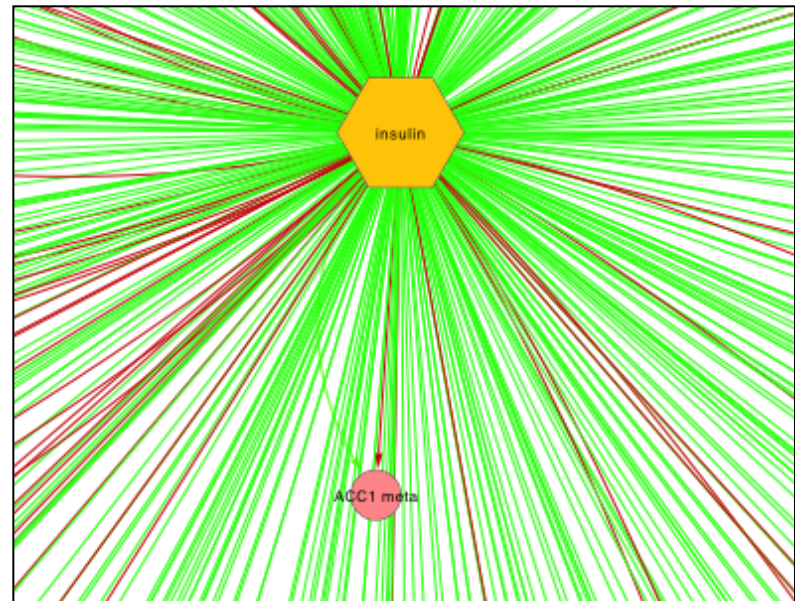
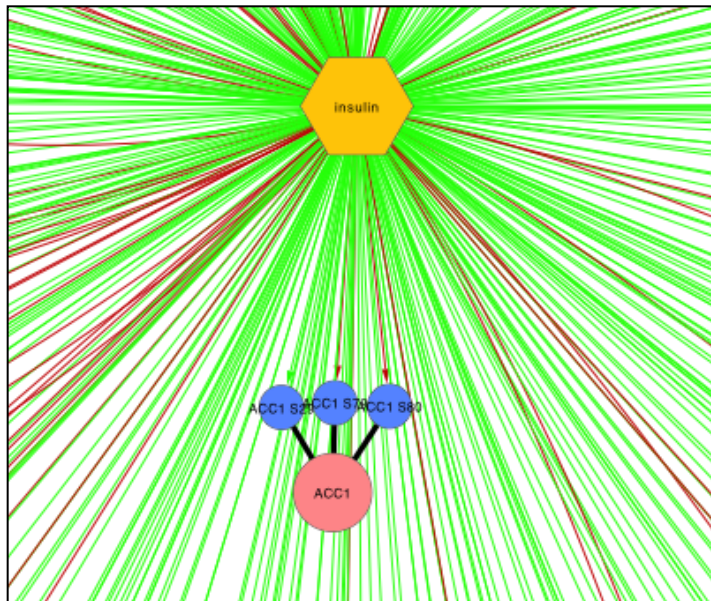
Querying network using peptide sequences in dataset

- Input: List of peptide sequences from MS dataset
- RCytoscape to select nodes with matching sequence (node attribute)
 - site nodes, or metanodes including sites
- Highlight site nodes in didactic pathway
- Identify relevant subnetworks from entire network

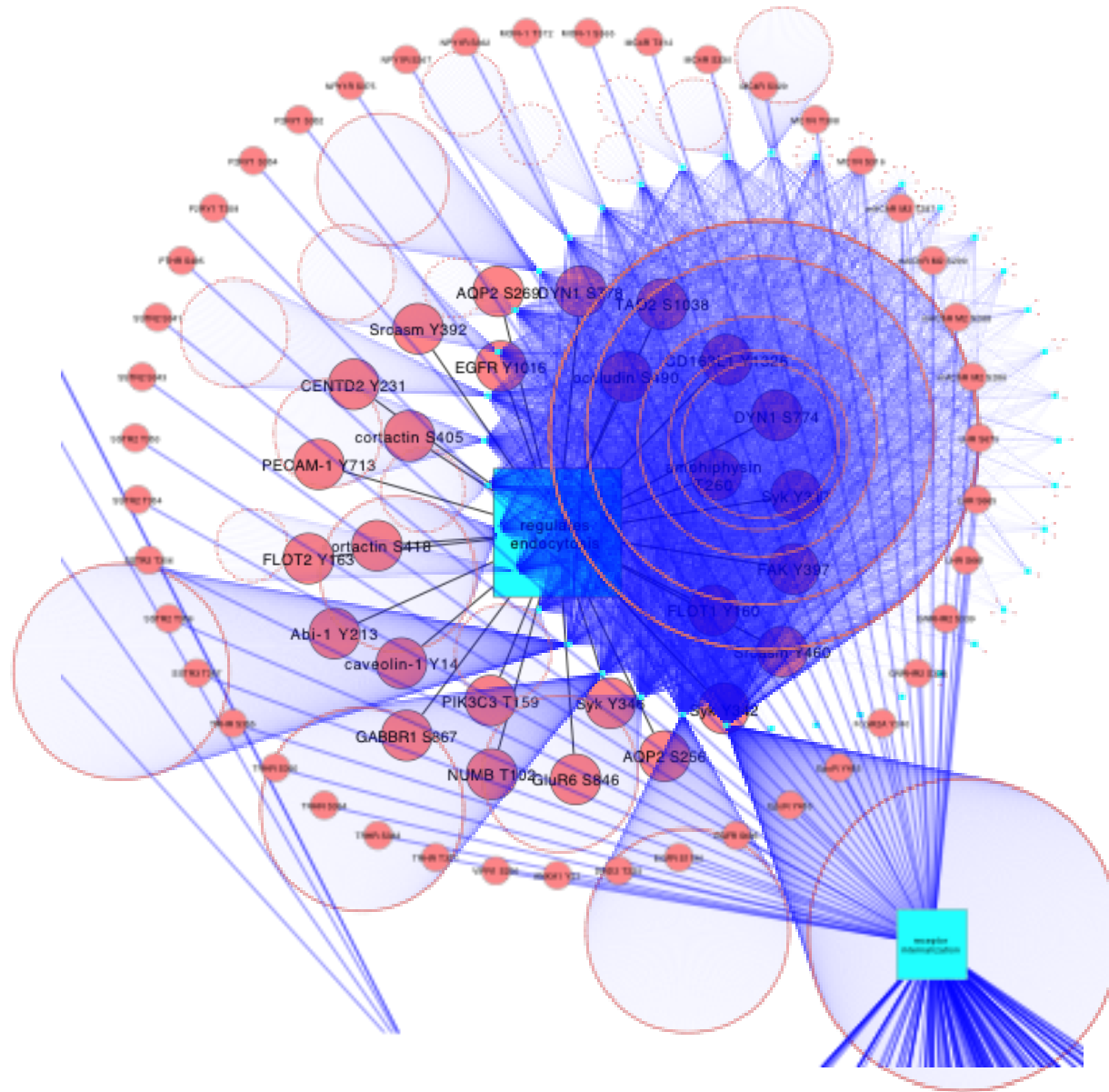
Insulin-regulated sites



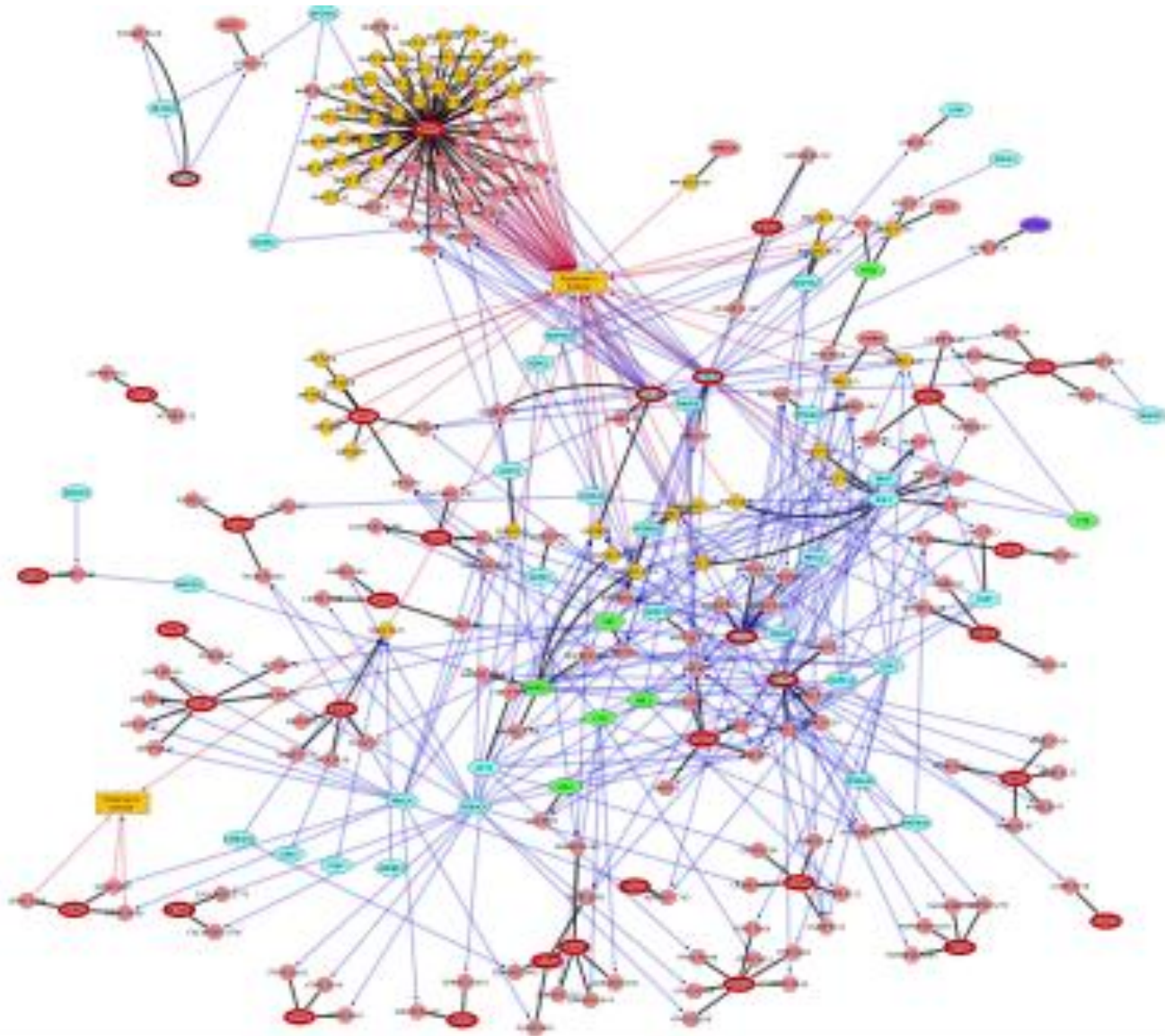
Insulin-responsive sites



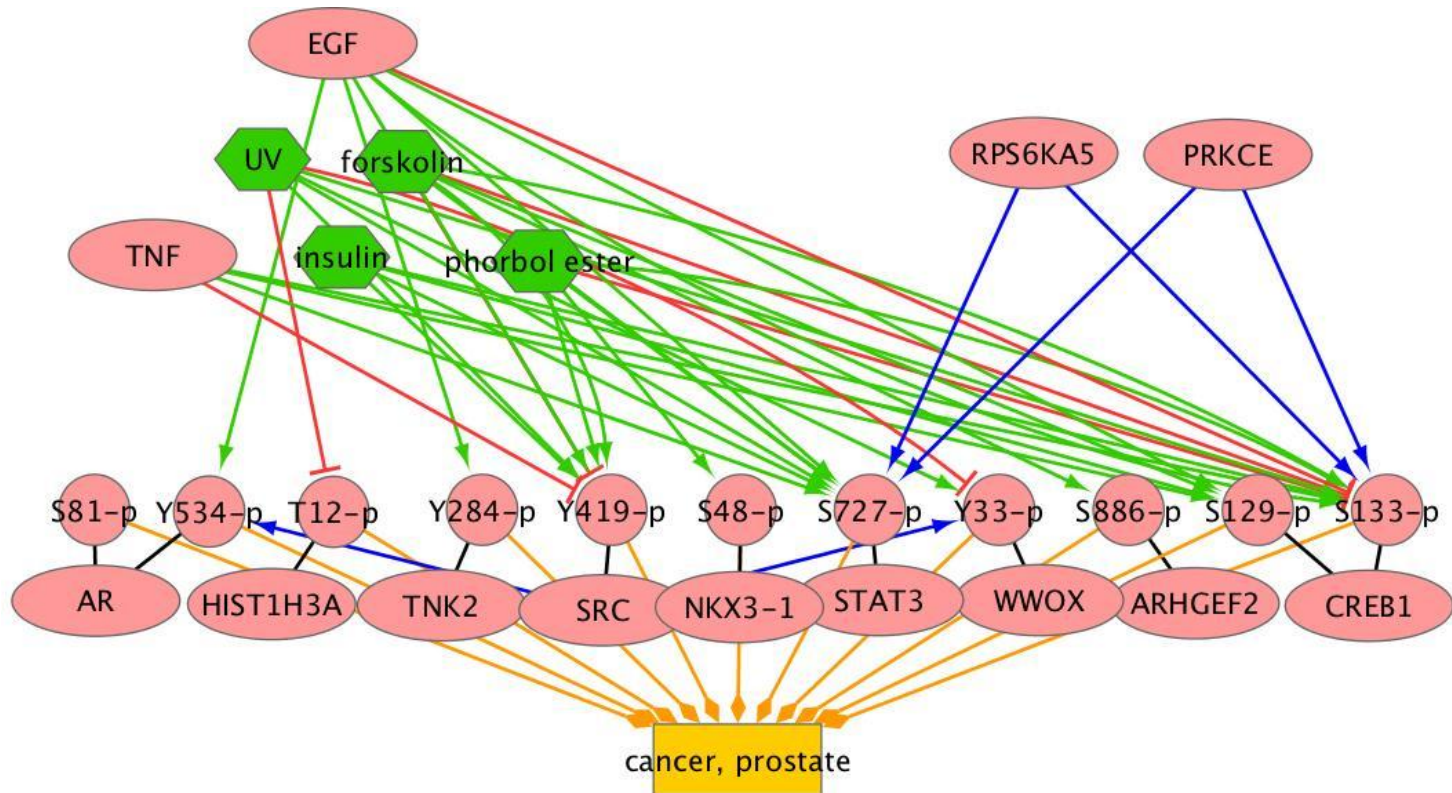
Downstream effects of PTMs



Disease subnetwork



Disease subnetwork



Challenges

- Managing visual complexity of site-centric networks
- Refined tools
 - Defining metanodes globally
 - Programmatic matching of peptide and site sequences
- Test hypothesis that site-based networks have greater analytical/predictive power
 - functional analysis (e.g., GO enrichment)
 - disease driver identification



Thanks to ...

Sasha Tkachev

Bin Zhang

Ela Skrzypek

Beth Murray

Vaughan Latham

Peter Hornbeck

Scooter Morris *UCSF*

Mark Grimes *Univ. of Montana*

with grant support from

