Applications of Modern Proteomics in Biology and Medicine

Namrata Udeshi
Broad Institute of MIT and Harvard

"NexGen" proteomics has arrived: 4-5 fold increased detection/ quantification of proteins, PTMs in cells/tissues over past 3 yrs

- Appropriate study design
- Robust sample processing methods
- Quantitative labeling of peptides for multiplexed anal.
- Data acquired with state-of-the-art LC-MS technology
- Statistically rigorous data analysis

Unprecedented definition of proteins in cells and tissues

- 10K – 12K distinct proteins
- Precise and reproducible
- Higher throughput

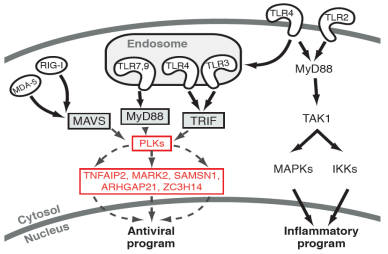
Deep and broad PTM coverage

- >25K phosphosites
- >20K ubiquitinated peps
- >10K acetylation sites

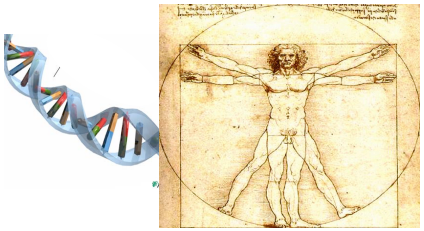
- **The number of proteins observed in tissues now begins to approximate the expressed proteome**
- **PTM analysis provide window into function and pathogenesis not accessible by genomic methods**

Precise measurement of proteins, their modifications and interaction partners is essential complement to genomics

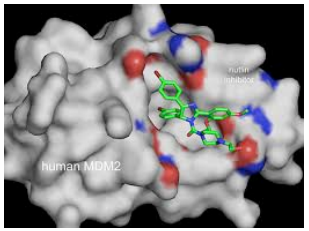
Understand disease biology, cell circuitry and signaling



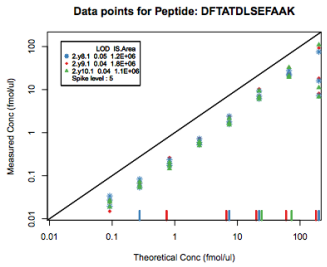
Connect genes to physiology



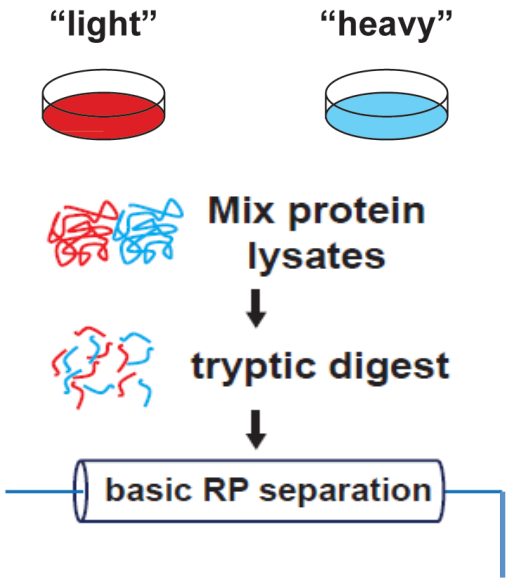
Define the targets and mechanism of action of drugs



High throughput quantitative biology (Hasmik Keshishian)



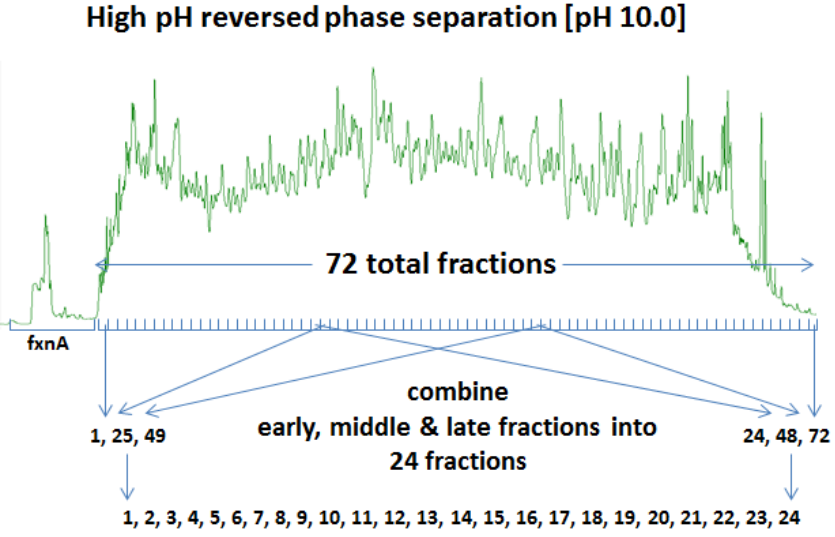
Current Large Scale Quantitative Proteomics Workflow



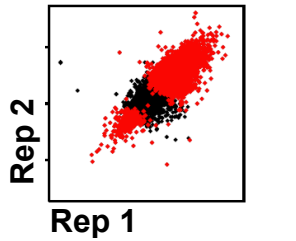
(95%) 8-12 bRP fractions PTMs

(5%) 24 bRP fractions Proteome

High Resolution UPLC-HCD-MS/MS



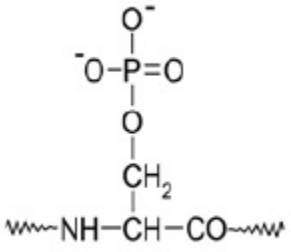
Data Analysis



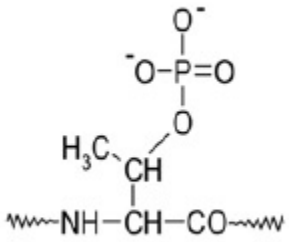
Statistical Analysis

Important PTMs currently amenable to large-scale mass spectrometry analysis

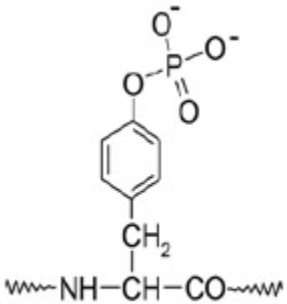
PTM	Mass shift (Δm ;Da)	Amino acids	Frequency	Enrichment methods	Coverage
Phosphorylation	79.9663	Ser, Thr, Tyr (Asp, His)	3.1%	Immobilized metal affinity chromatography (Fe-IMAC, TiO2) or antibodies	>25K sites
Ubiquitination (diGly tag)	114.0429	Lys	0.08%	Anti ε-LysGlyGly antibodies	>20K sites
Acetylation	42.0106	Lys	0.07%	Anti Acetyllysine antibodies	>10K sites



p-Serine

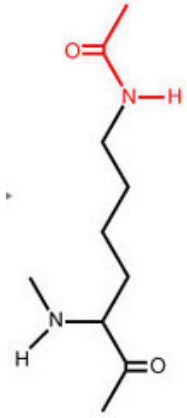


p-Threonine



p-Tyrosine

diGlycine-Lysine



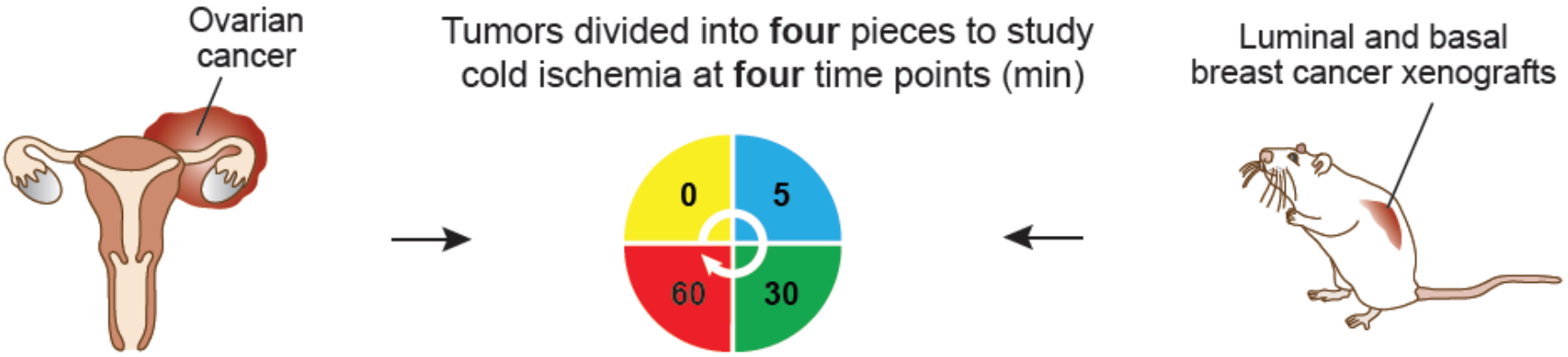
Acetyl-Lysine

Example Application: effects of post excision delay-to-freezing time on posttranslational modifications

- Large on-going effort to characterize proteomes/PTMs of genomically annotated TCGA samples
- Time between ligation, excision and freezing for the TCGA samples (post-excision delay, PDT) varied from minutes to ca. 1 hour
- Effects of ischemia and physical tissue trauma on PTM's not well studied
- Activated kinases and phosphatases can act in seconds-minutes
  - Alterations in phosphosignaling in cancer well established
- Prior studies have shown that the phosphorylation site stoichiometry can change significantly post tumor excision
  - Duration from ligation of blood flow to excision highly variable and often not taken into account (shortest time evaluated ca. 15 min.)
  - few p-sites evaluated (RPPA)

Study goal: to address concerns for larger TCGA work, evaluate changes in protein phosphorylation (<1 min and longer) induced by PDT using quantitative LC-MS/MS

Design of study to evaluate effects of cold ischemia in patient-derived xenografts and tumors



Samples: Four patient-derived ovarian cancer tumors and two patient-derived human breast cancer tumors (basal-like; luminal-like; pools of 10 tumors)

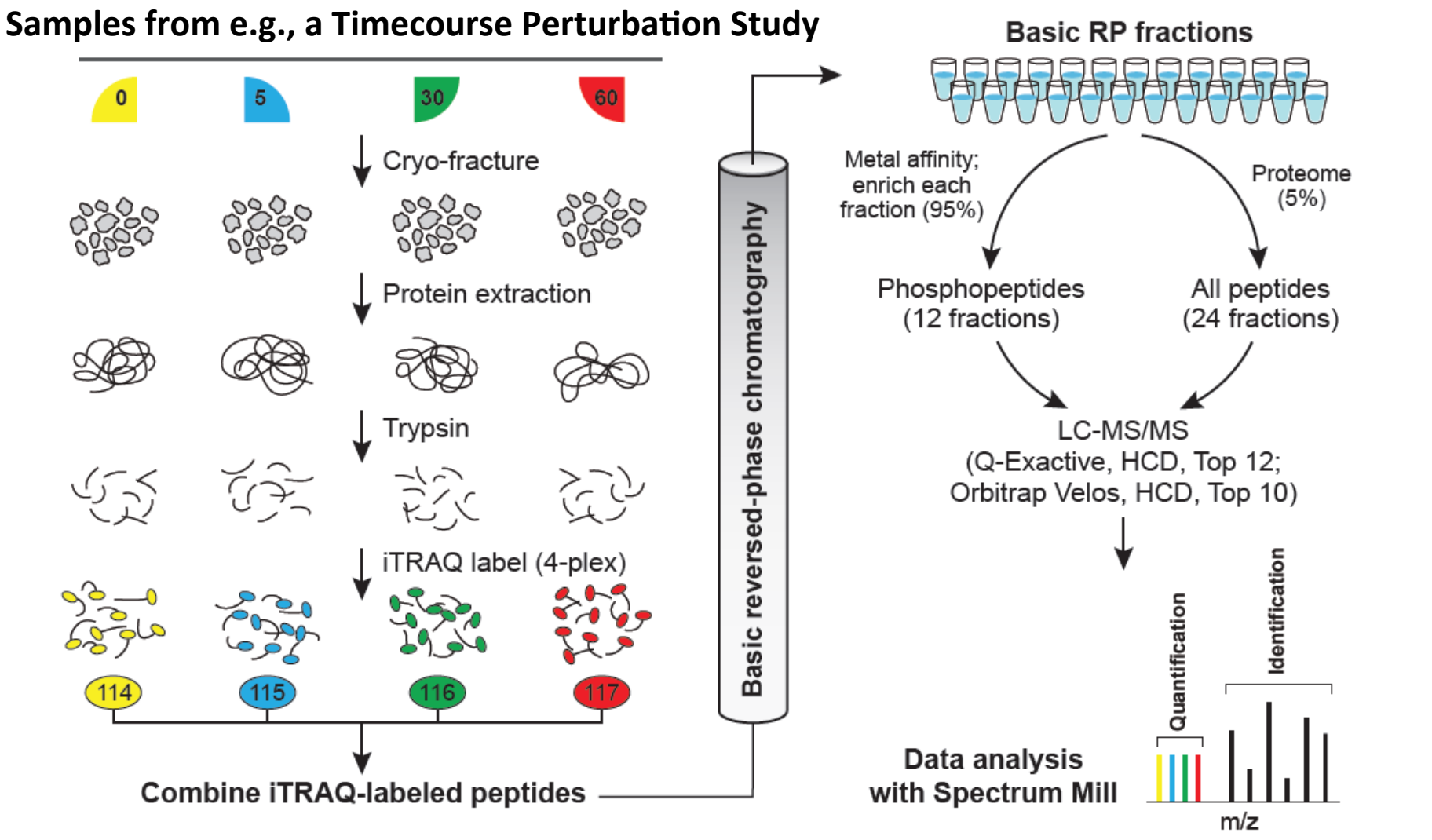
Collection: excision prior to ligation; immediate LN2 freezing

Timepoints: "0" (≤ 90s from excision to freezing); 5 minutes; 30 minutes and 60 minutes

Proteomic Data Generation: high performance instruments capable of robust iTRAQ mass-tag generation

Quantification method: 4-plex iTRAQ labeling

Integrated workflow for global proteomic and phosphoproteomic analysis in a multiplexed manner



0.5 - 1.5mg total protein per label: 95% needed for phosphoproteomics

Cold ischemia times up to 1 hour cause no change in proteome but up to 23% change in phosphoproteome

	n tumor samples	Total	average per tumor sample	overlap in at least (n-1) samples	kinetics-based regression test*	moderated F-test*	union of both tests*	% regulated of overlap**
Phosphoproteome					#up/#down	#up/#down	#up/#down	#up/#down
Ovarian Cancer	4	21792	12184	8745	274/87	387/66	427/108	4.9/1.2
Basal Breast Cancer	3	36672	26127	24760	1155/835	1056/597	1372/938	5.5/3.8
Luminal Breast Cancer	3	32524	24234	23627	3734/767	3827/927	4525/1072	19.2/4.5
Proteome								
Ovarian Cancer	4	9498	7550	6985	0/0	0/0	0/0	0/0
Basal Breast Cancer	3	18855	14989	14970	0/0	0/0	0/0	0/0
Luminal Breast Cancer	3	15753	12641	12679	0/0	0/0	0/0	0/0

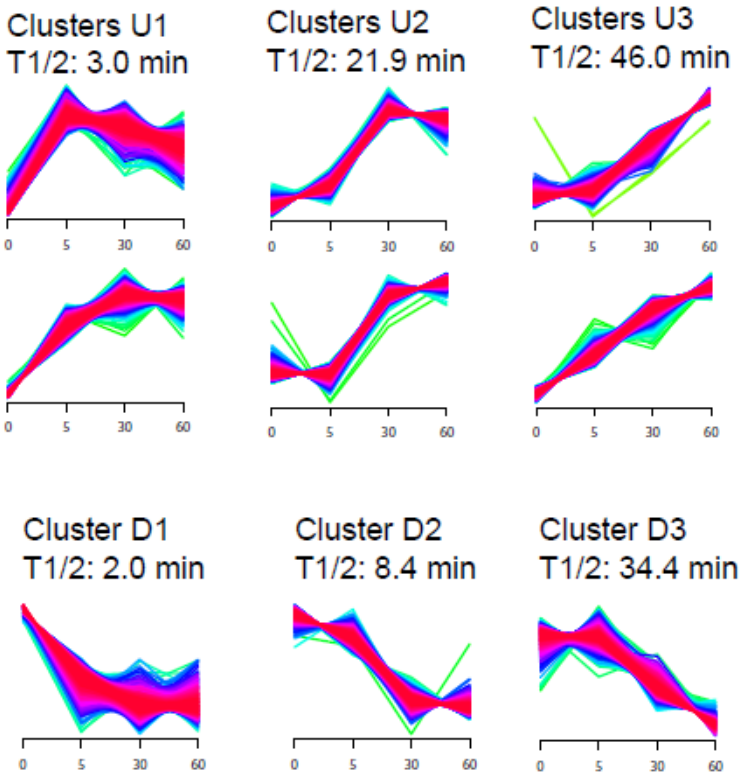
* Significant regulation at a kinetics-based regression test or moderated F-test FDR p<0.01

** Percent regulated phosphosites and proteins within overlap dataset

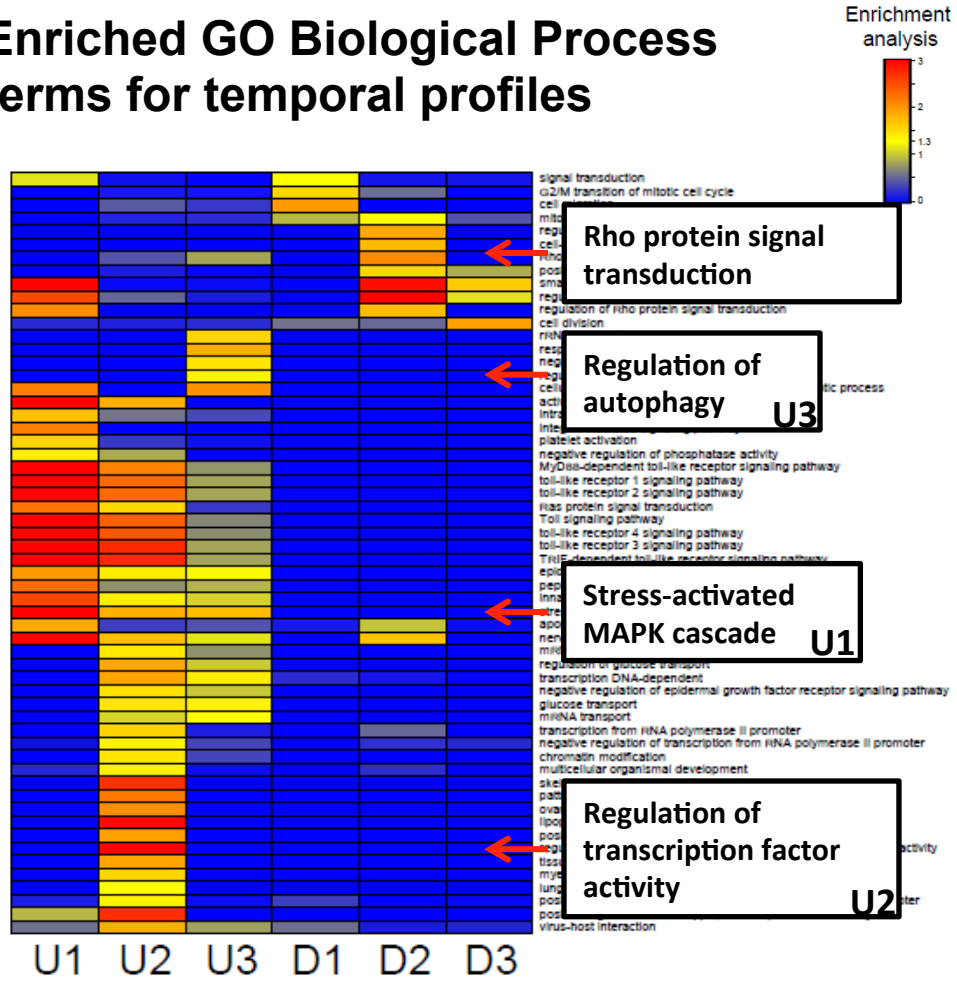
Mertins et al. *Mol Cell Proteomics* 2014

Regulated phosphoproteome (OC and BC) exhibits distinct temporal profiles with differing biological functions

Fuzzy c-means clusters of regulated phosphosites



Enriched GO Biological Process terms for temporal profiles



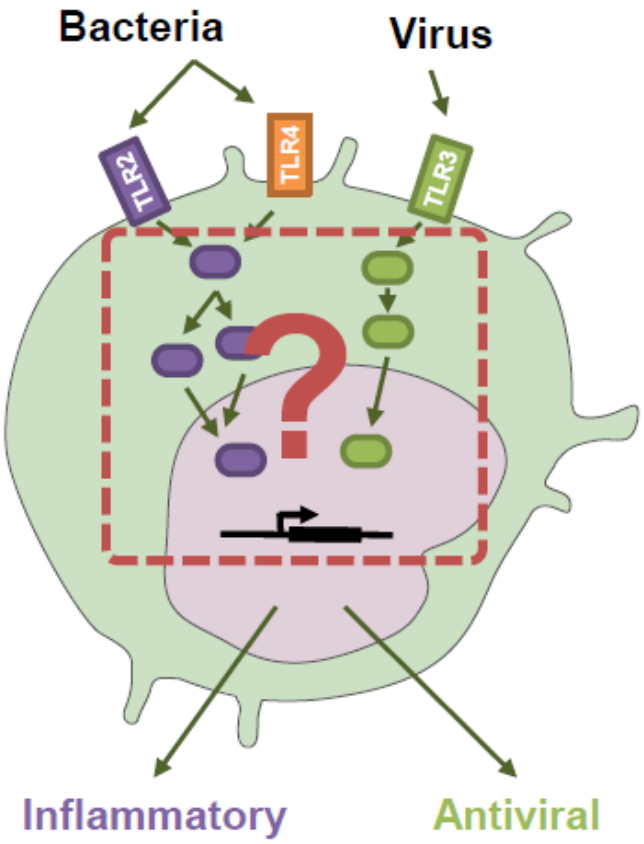
Mertins et al. *Mol Cell Proteomics* 2014

Conclusions: Cold ischemia Pilot study

- PDT/Ischemia results for both breast tumor xenografts and human ovarian cancer tumors
  - Global proteome profile is unchanged over 1 hr
  - Activation of kinases and phosphatases within minutes
  - 5 - 23% of phosphoproteome fluctuated (up and down regulated) in time-lapse studies
  - **158 phosphosites were identified that are changing due to cold ischemia in every analyzed tumor sample**.
  - Common processes affected include stress response, cell cycle regulation and cell death
  - Majority of the phosphoproteome appears stable
- Phosphoproteome-analyses were performed on TCGA samples…with caution regarding data analysis for samples prepared not necessarily with proteomics in mind.

Mertins et al; Mol Cell Proteomics 2014

Toll-like receptors are pathogen sensors on dendritic cells

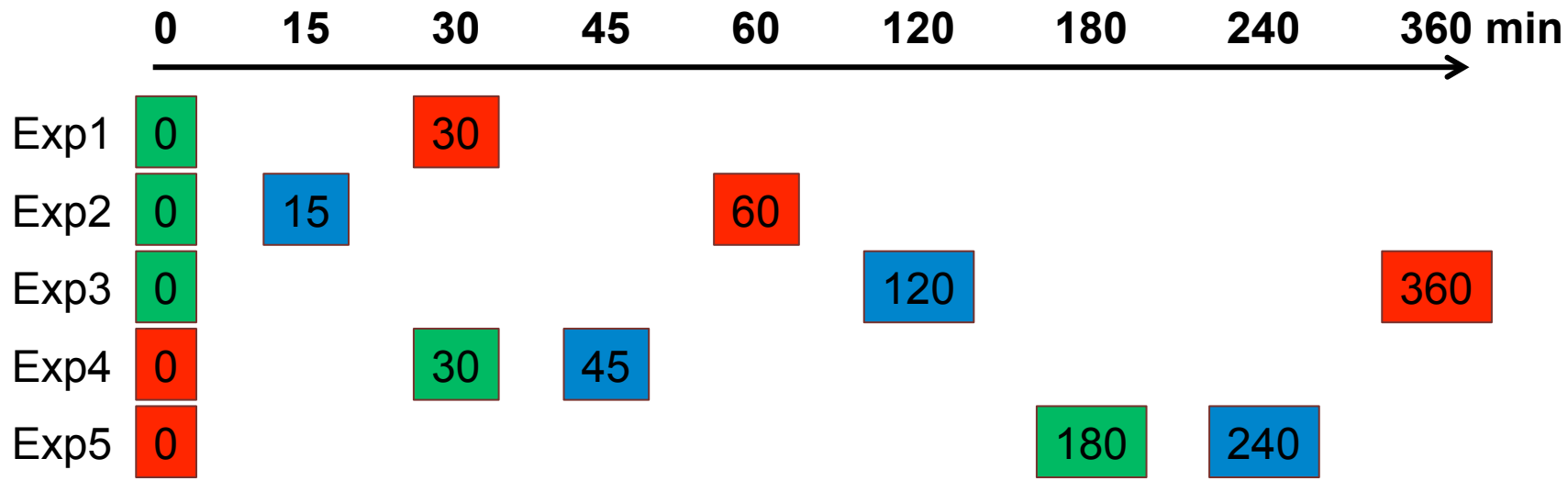


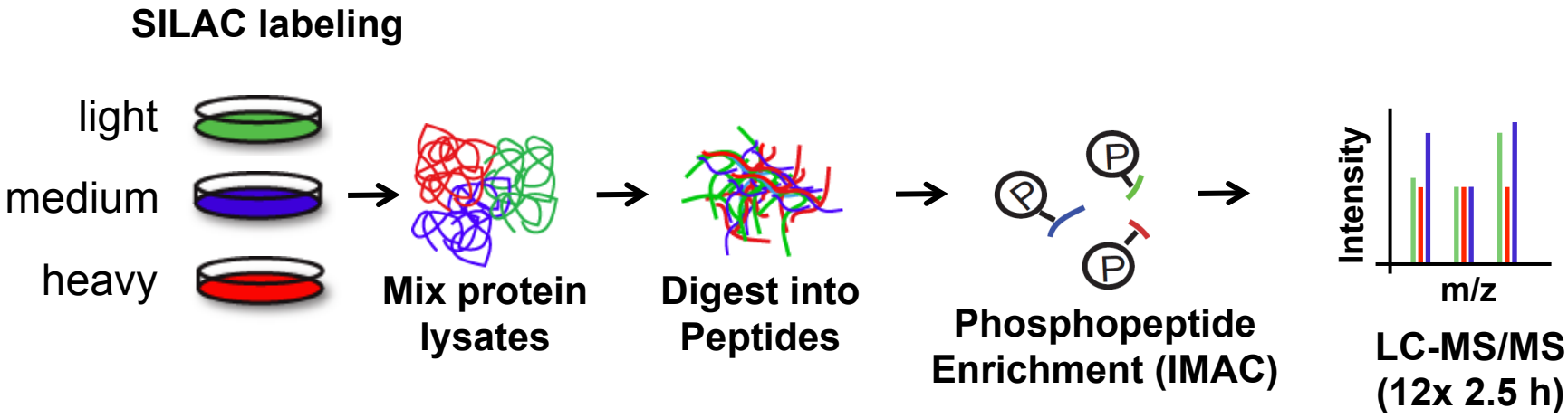
Main Questions

- Identify signaling components?
- Place components within pathways & networks?
- Connect signaling and transcriptional layers?
- Target signaling nodes to obtain desired outcomes?

Nir Hacohen, Nicolas Chevrier

Approach: monitor LPS-induced phosphorylation changes over 9 time points using quantitative proteomics





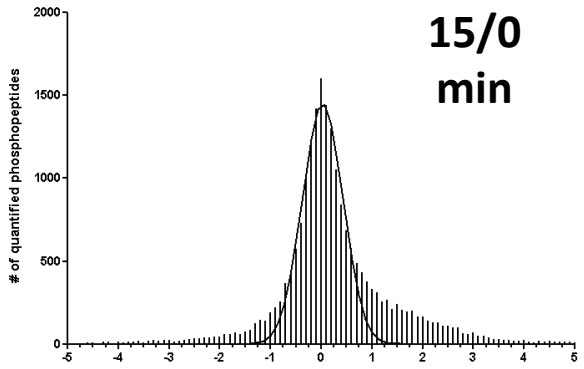
40% of all phosphoproteins are upregulated in 2 or more time points after LPS stimulation

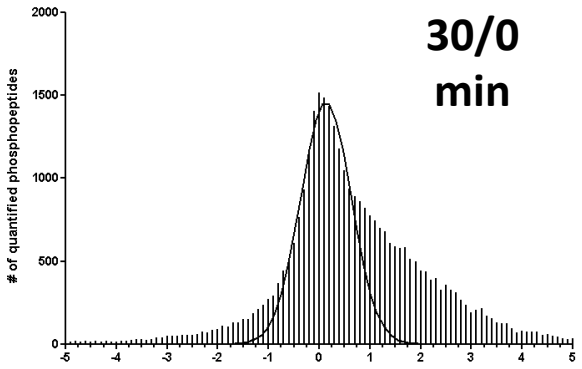
Phosphopeptides quantified in ≥7 time points	
total	12,723**
proteins	2,970
regulated p-peptides*	3,627
regulated proteins*	1,277 (40% ≥2 timepoints)
protein kinases	180
regulated kinases	86

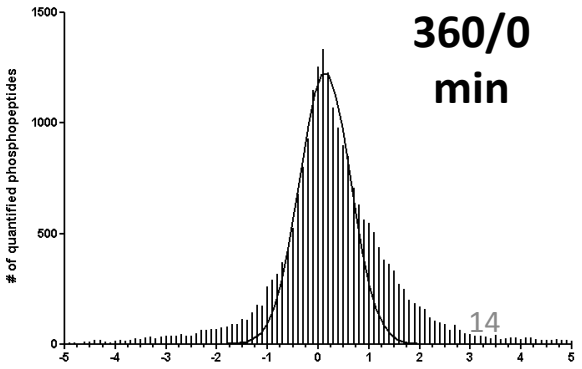
* Indicates FDR<0.01

** 7,775 localized sites

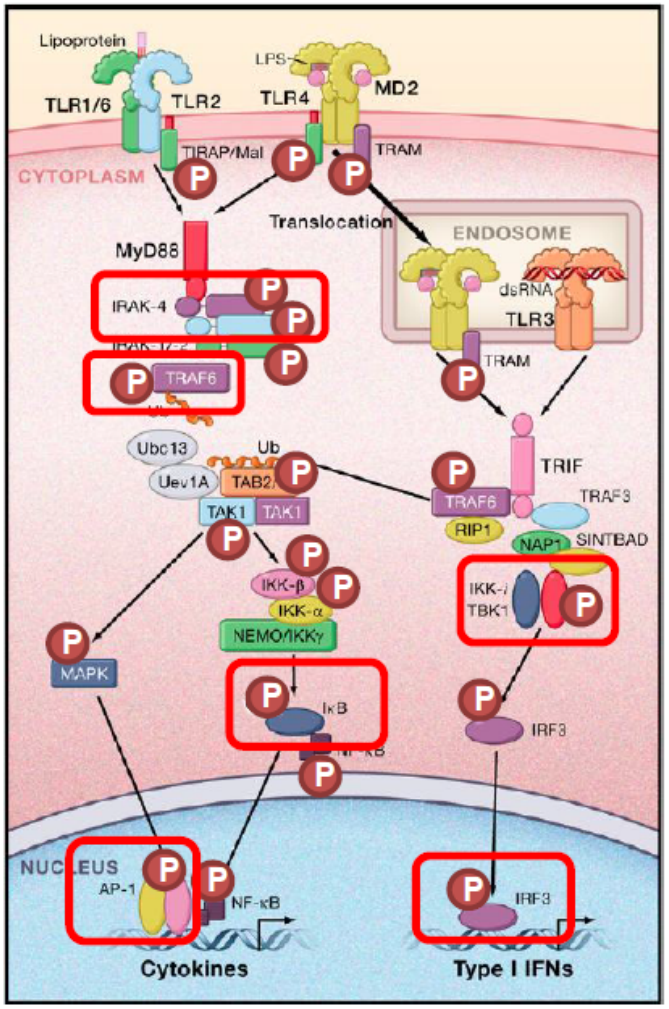
Log2 SILAC ratios:



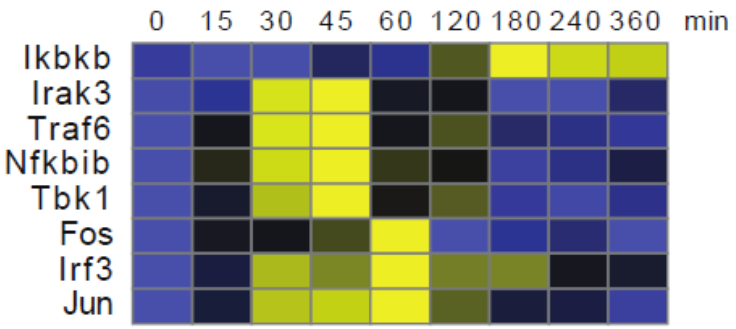


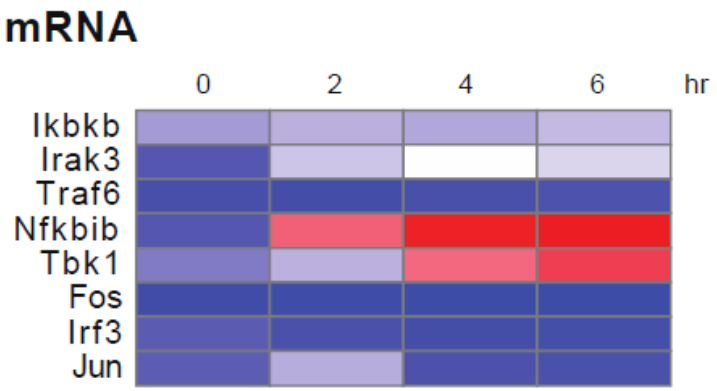


Phosphoproteomics identifies known pathway components not detected by mRNA

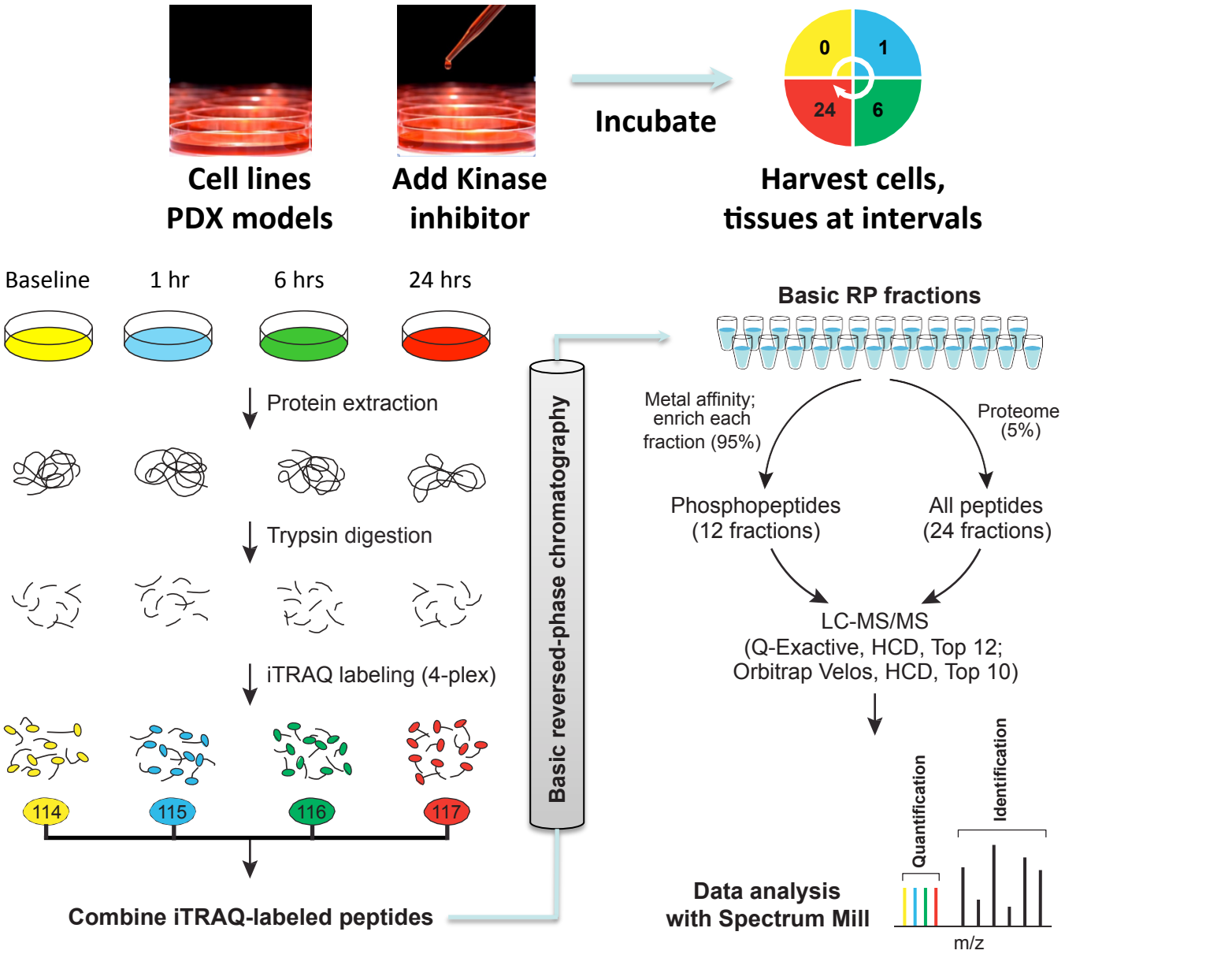


Modified from Takeuchi & Akira, Cell, 2010



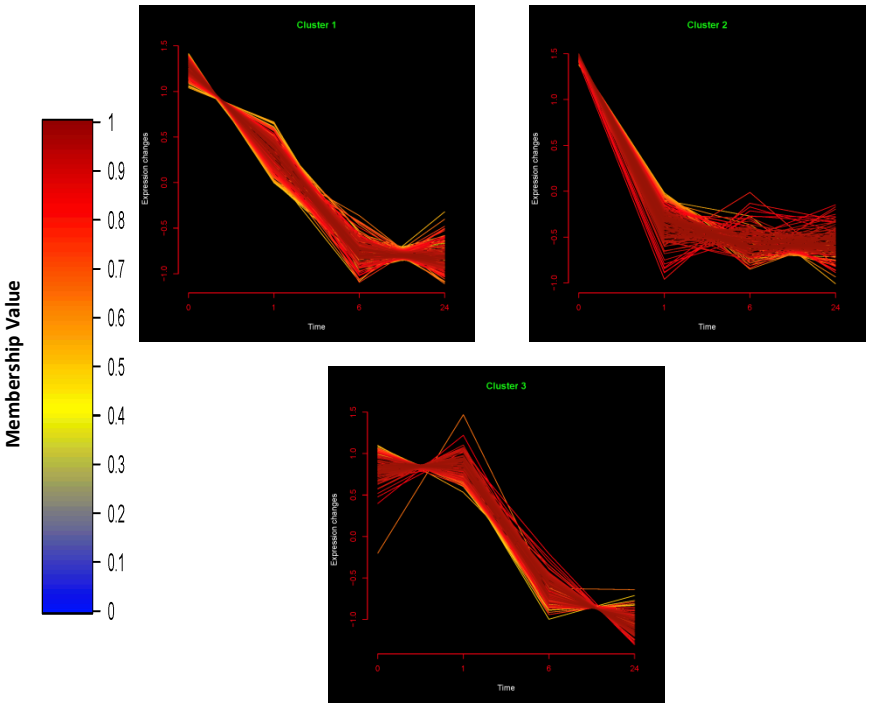


Proteomics of targeted therapeutics can provide novel insights into mechanisms of sensitivity and resistance



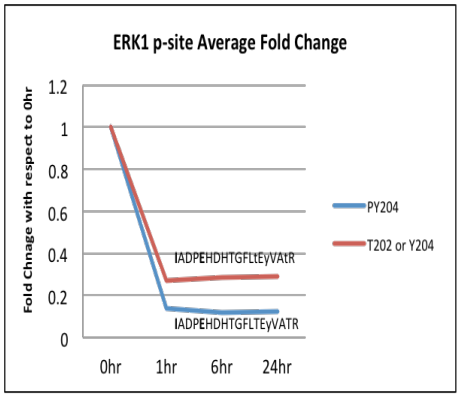
Temporal response characteristics are reproducibly observed; depth of coverage allows detailed mapping of pathway modulation

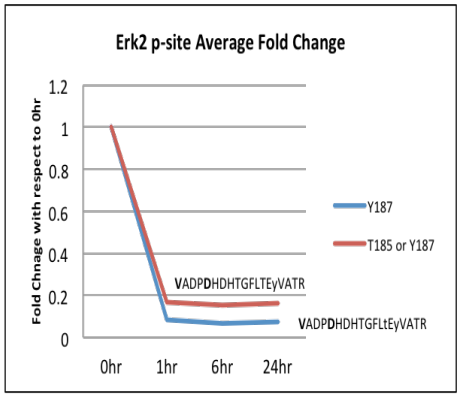
Temporal response of 423 down-regulated p-sites; 23/202 kinase p-sites regulated

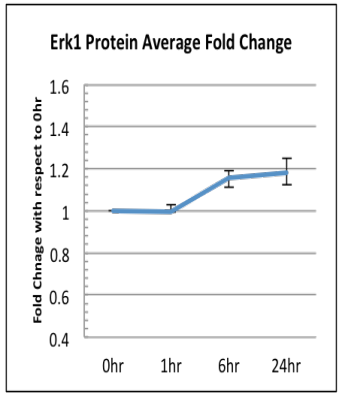


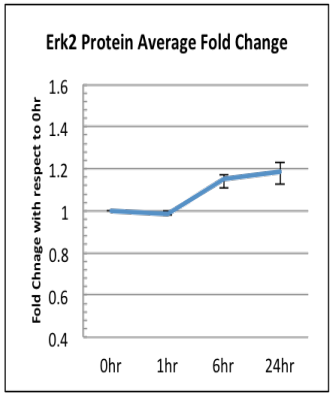
Regulated phosphosites showed reproducible temporal trends at 1, 6 and 24 hours

Measured 116 / 253 proteins (46%) in relevant canonical cancer pathway





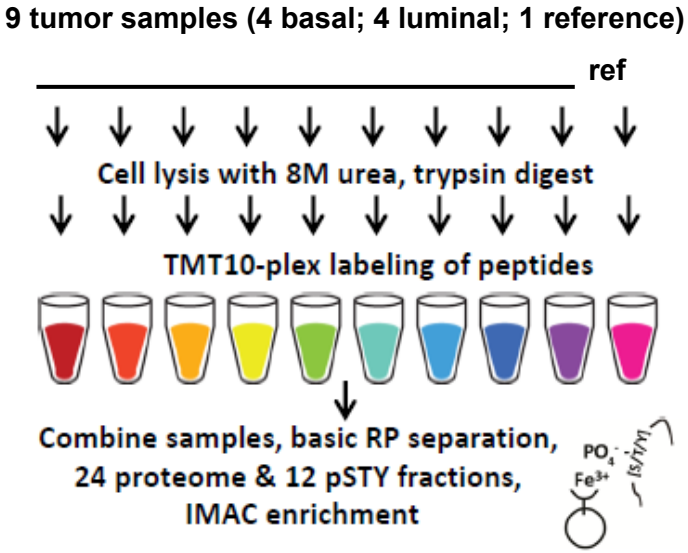




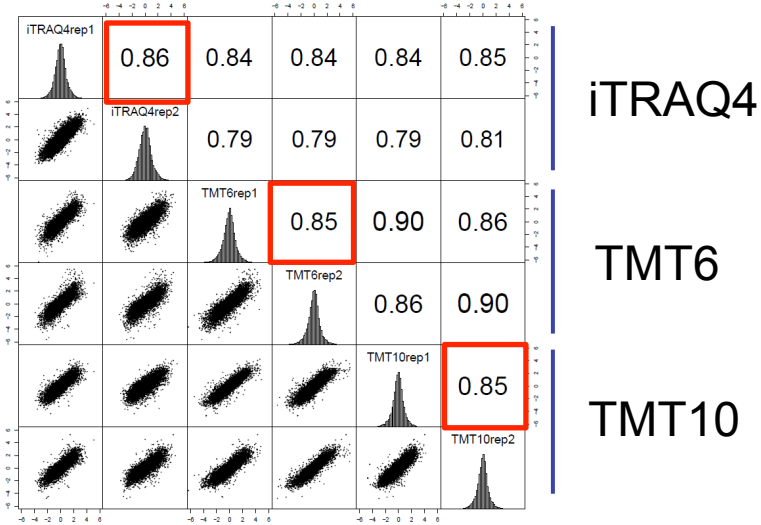
- **Phospho sites on > 200 total kinases measured; > 10% with at least one regulated phosphosite.**

Increased throughput with TMT6 and TMT10 vs. iTRAQ4 with high sensitivity and quantitative fidelity

3x increased throughput



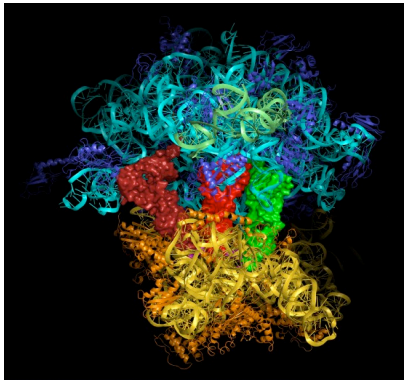
Highly consistent quantification results



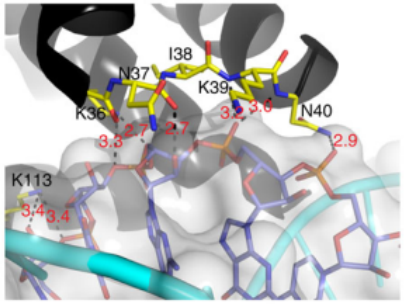
Log2 basal/luminal tumors

		Rep1/Rep2	
		Proteins/sites	Distinct Peptides
Proteome Coverage	iTRAQ4	13,201/13,101	198953/196484
	TMT6	12,839/13,839	174590/196521
	TMT10	12,624/12,908	170190/168828
Phosphoproteome Coverage	iTRAQ4	45,495/45,815	60,945/58,005
	TMT6	33,131/32,261	39,090/42,543
	TMT10	33,523/31,119	39,044/34,958

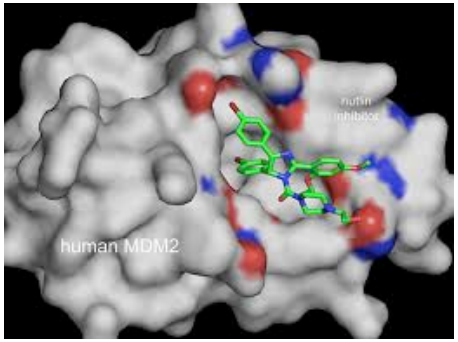
Affinity proteomics: a direct route to biological understanding through hypothesis-guided experiment



Protein-Protein (and disruption of these by small molecules)



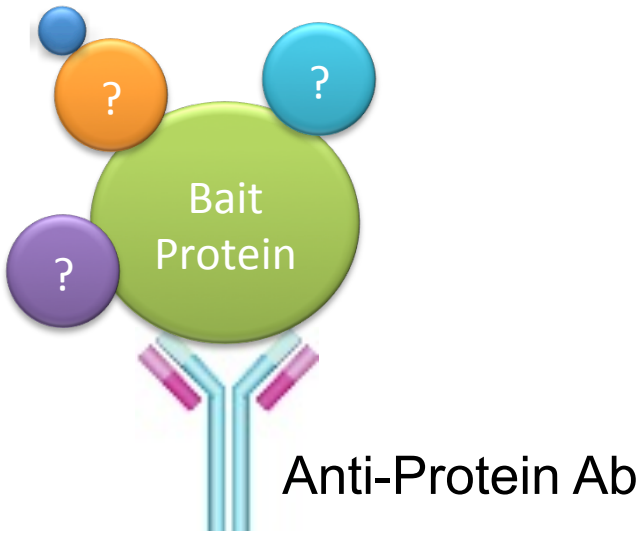
Protein-nucleic acid: DNA, RNA, lincRNA



Protein-small molecule

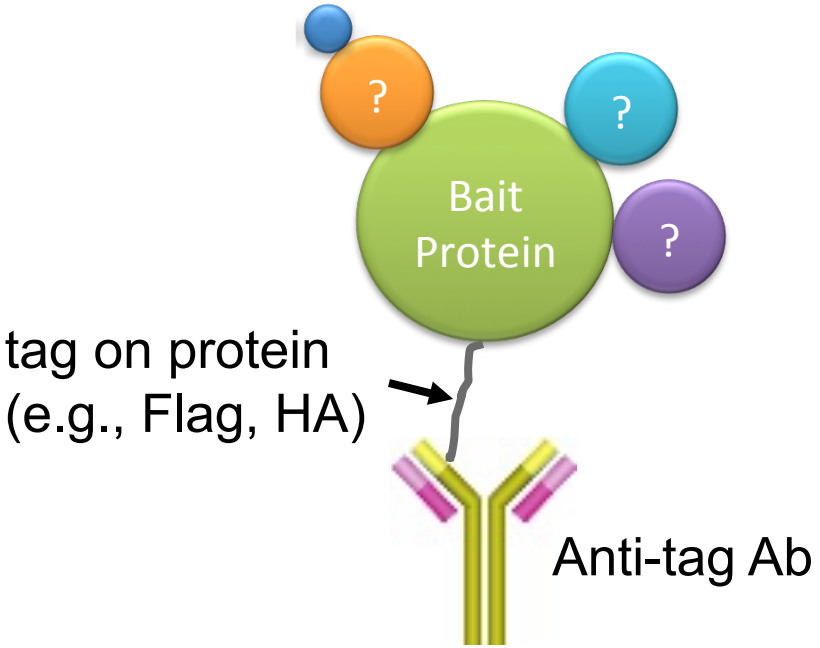
Helping to functionalize the genome: analysis of protein-protein interactions by proteomics ("guilt by association")

IP of Endogenous Protein Using IP-Competent Ab



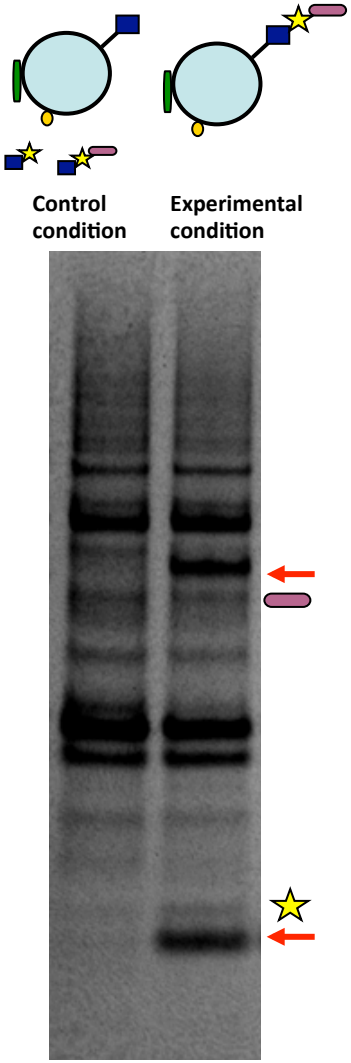
- No manipulations to cell
- Endogenous levels
- "Knockdown" control

IP of Affinity-tagged, Expressed Construct



- Generic
- Universal Control
- Easy to leverage ORF collection

“Classical” biochemical affinity enrichment



Limiting steps:

- Optimization of affinity purification conditions
  - Sensitivity or specificity?
- Each condition is handled separately
  - Manipulation artifacts
- Long lists of protein IDs requiring validation

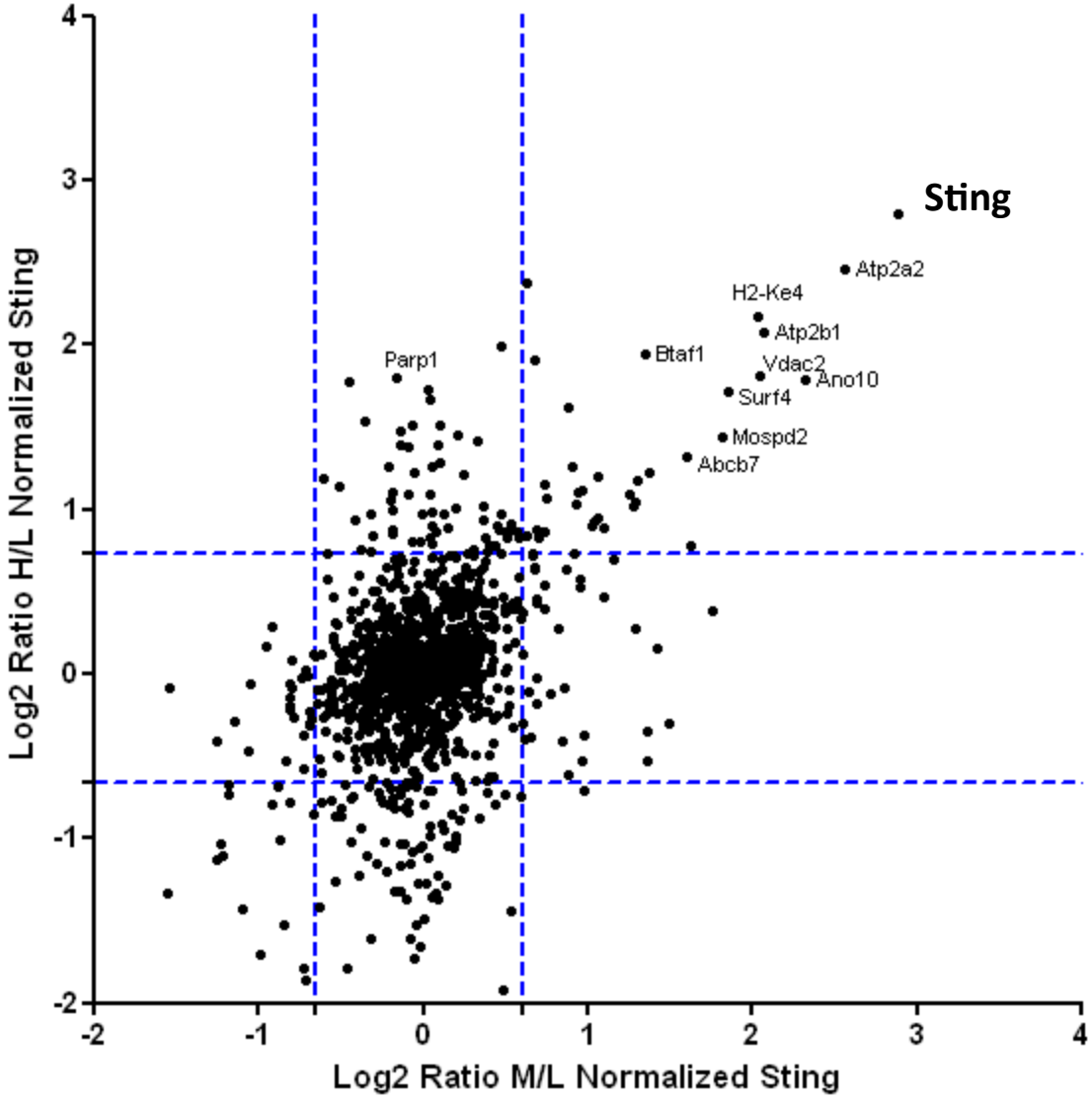
Enabling technologies:

✓ ***Quantitative*** **proteomics** by

- SILAC (metabolic labeling)

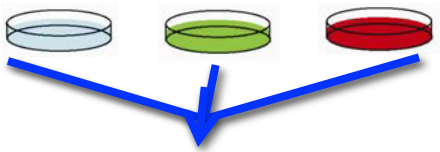
  or
- iTRAQ (chemical labeling)

Identifying candidate regulators of the ISD response by analysis of protein-protein interactors of key signal transduction proteins



Sting interaction proteome

L	M	H	SILAC label
mock	Sting-HA	Sting-HA	MEFs
(-)	(-)	(+)	Viral DNA



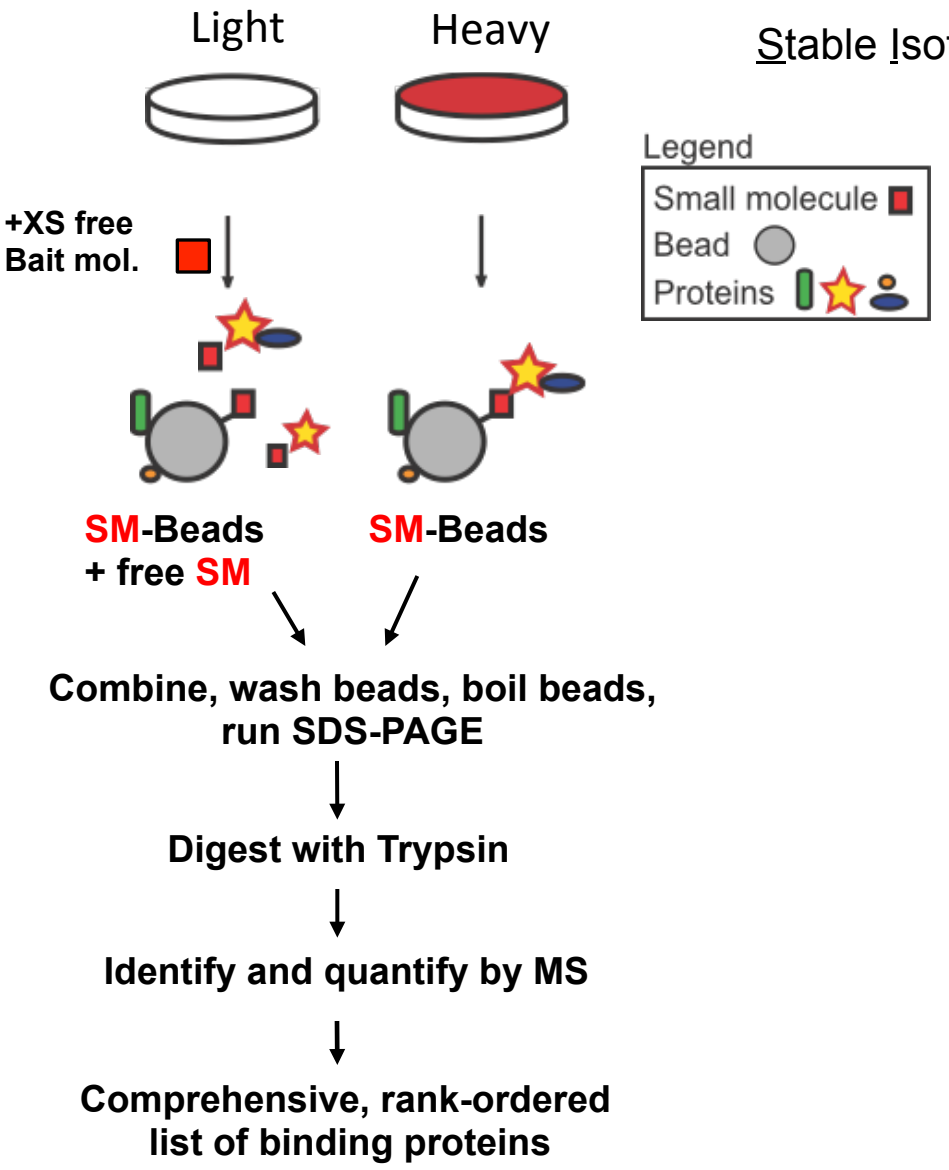
Anti-HA IP & in-gel digest

LC-MS/MS
(Q Exactive; 1 day instrument time)

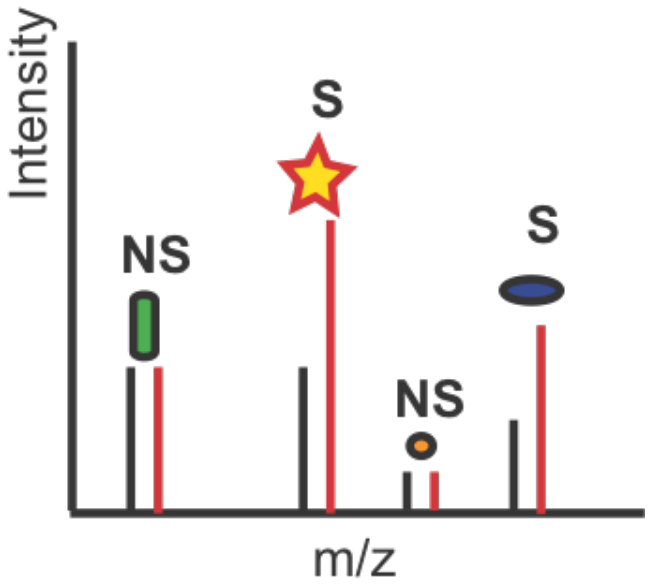
- >1400 proteins quantified
- Ca. 40 specific interactors

Lee, Hacohen et al. *Nature Immunol* 2012

Identifying targets of small molecules in cellular context with SILAC and affinity proteomics

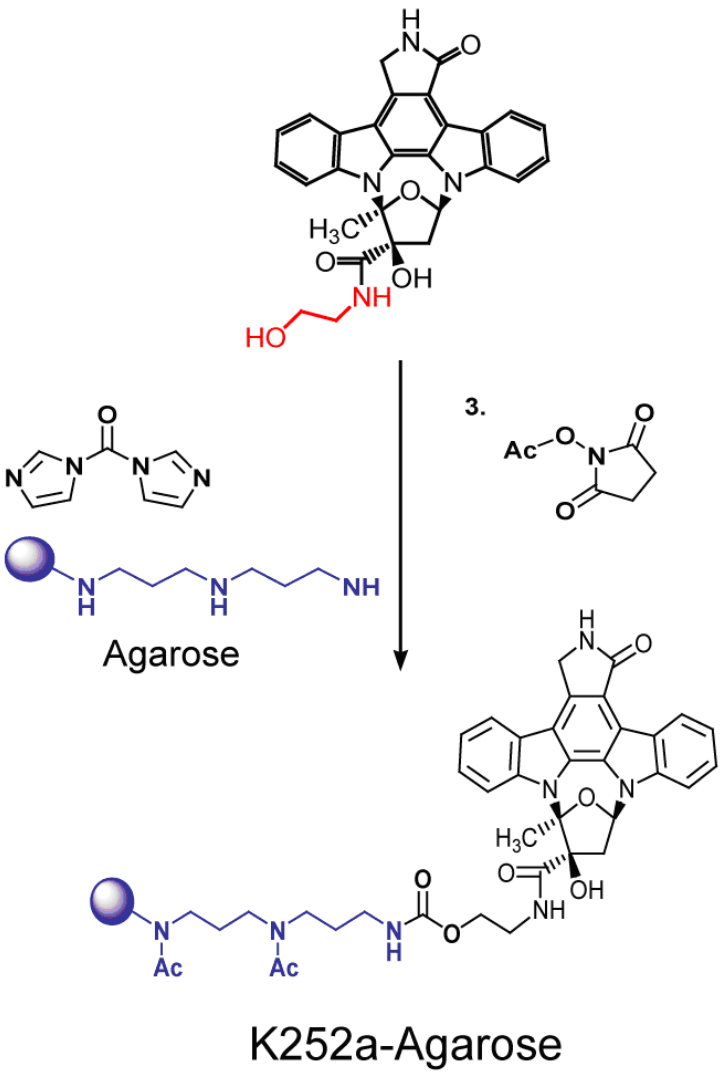


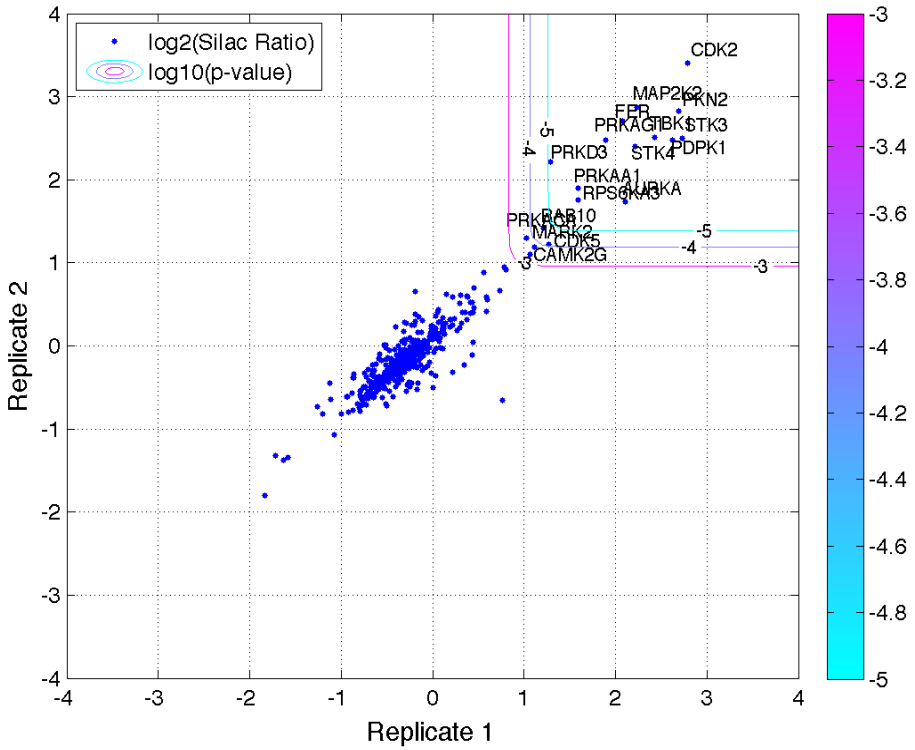
Stable Isotope Labeling by Amino acids in Cell culture



Ong, Schenone et. al., PNAS 2009

ID of Targets of K252a, a Promiscuous Kinase Inhibitor



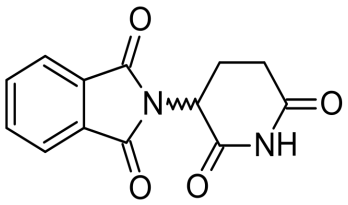


RIPK2, RHOA, ULK3,MAP2K6 CDK2, PKN2, STK3, TBK1, PDPK1, MAP2K2, STK4, FER, PRKAG1, AURKA, PRKAA1, RPS6KA3, PRKD3, CDK5, RAB10, MARK2, PRKACA, CAMK2G

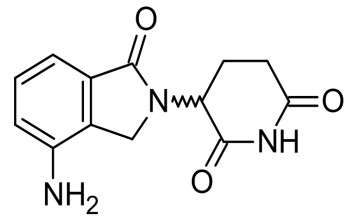
➢20 kinases identified as specific targets of k252a

Ong, Schenone et al PNAS (2009) 106: 4617-4622
Data analysis with Empirical Bayes: Margolin et al PLoS ONE 4(10):e7454

Thalidomide and its Analogs have Pleiotropic Effects – Some Catastrophic, Some Beneficial



Thalidomide



Lenalidomide

Birth Defects
1961: Contergan disaster

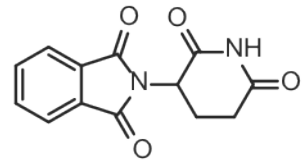
?

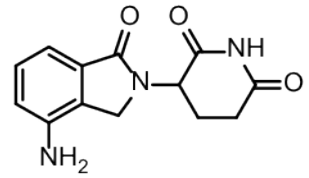
Cancer Therapies:
- Multiple Myeloma
- Lymphoma
- Myelodysplastic syndrome

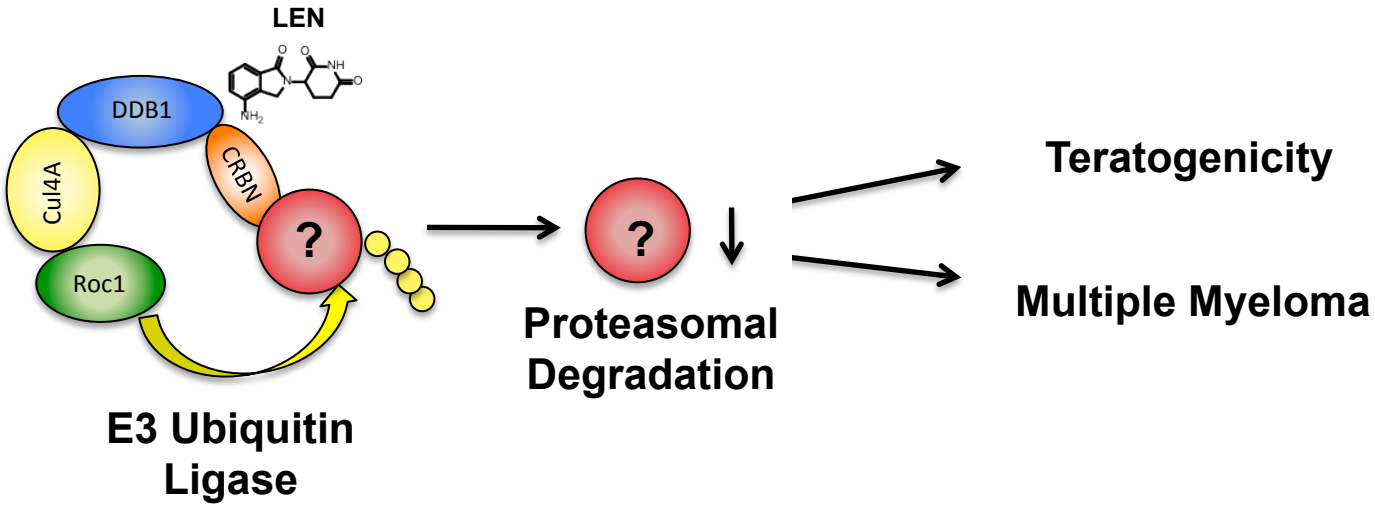
?

Immunomodulatory Drugs
Immune System Stimulation

Identification of CRBN-DDB1 as the Primary Target



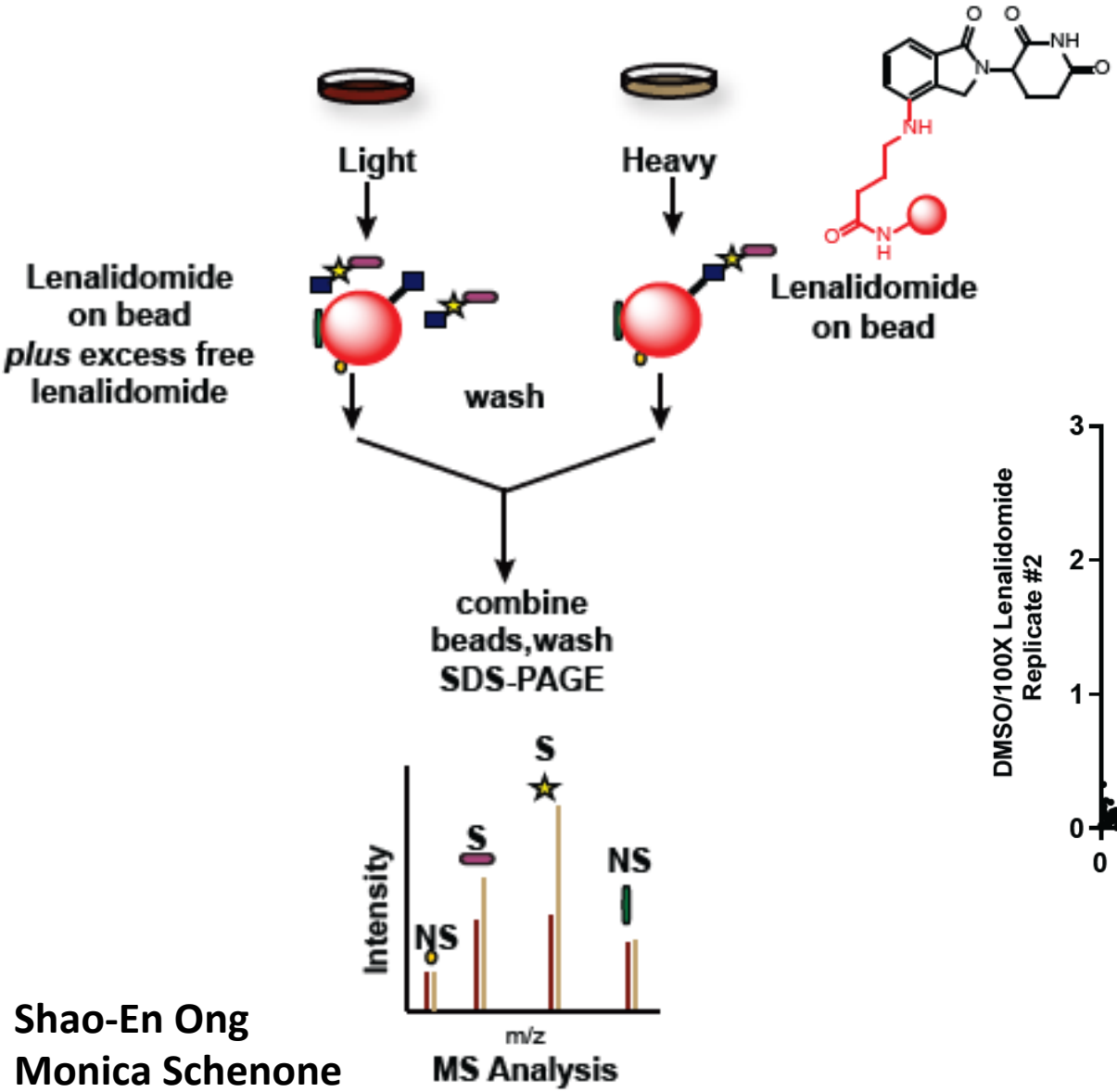


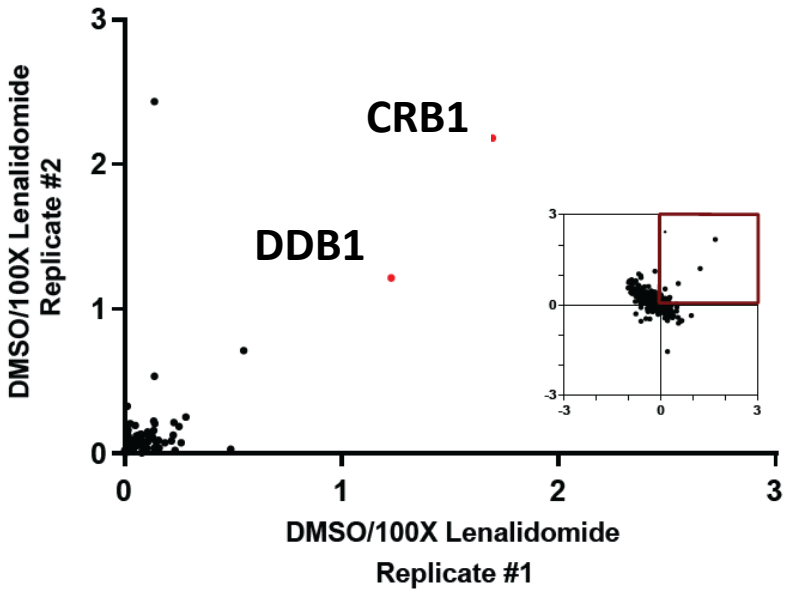


- What is the basis for the anti-myeloma activity downstream of CRBN?
- Are the pleiotropic effects of lenalidomide caused by altered ubiquitination of target proteins?
- What is the mechanism of action of lenalidomide?

Ito et al, Science 2010
Zhu et al, Blood 2011
Lopez-Girona et al, Leukemia 2012

Target ID using Lenalidomide as "bait" identified an E3 ubiquitin ligase complex as a target

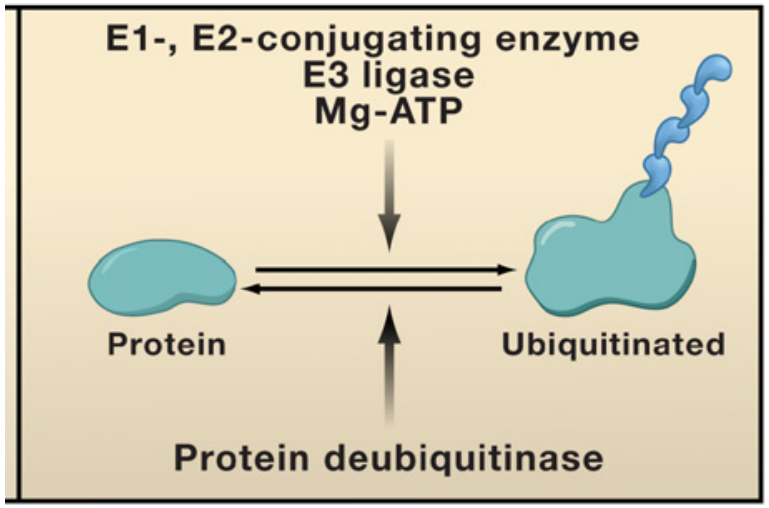




Shao-En Ong
Monica Schenone

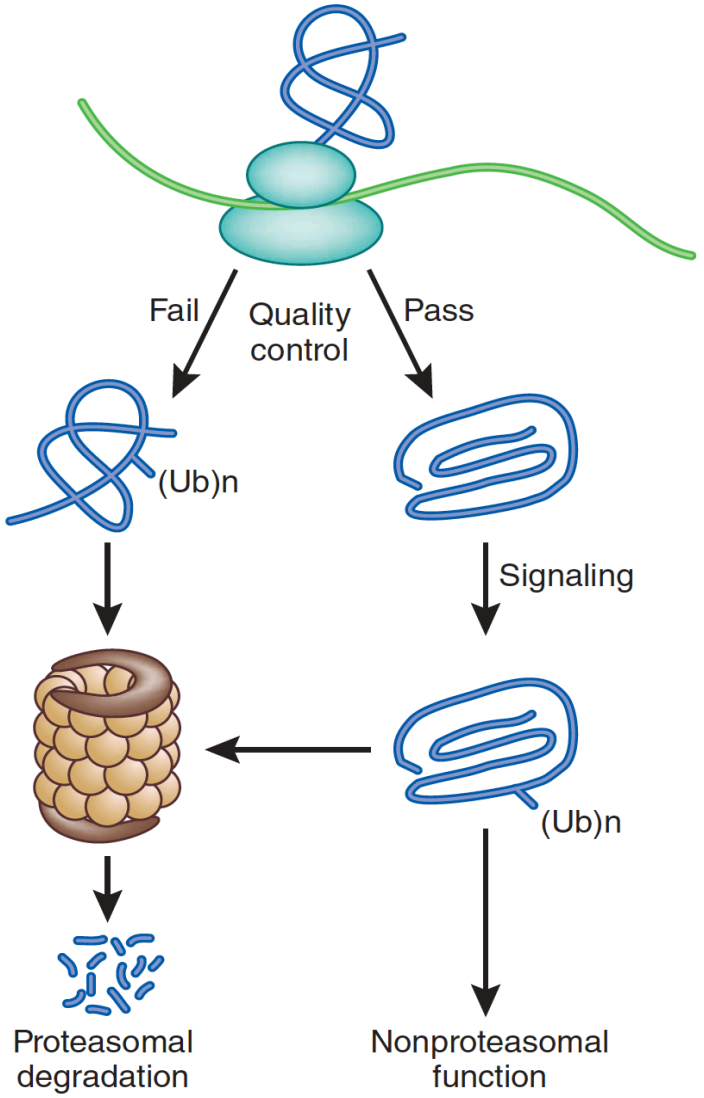
Schenone, Ebert (2010)

Ubiquitination: Another Complex System for Regulating Protein Function through Posttranslational Modification



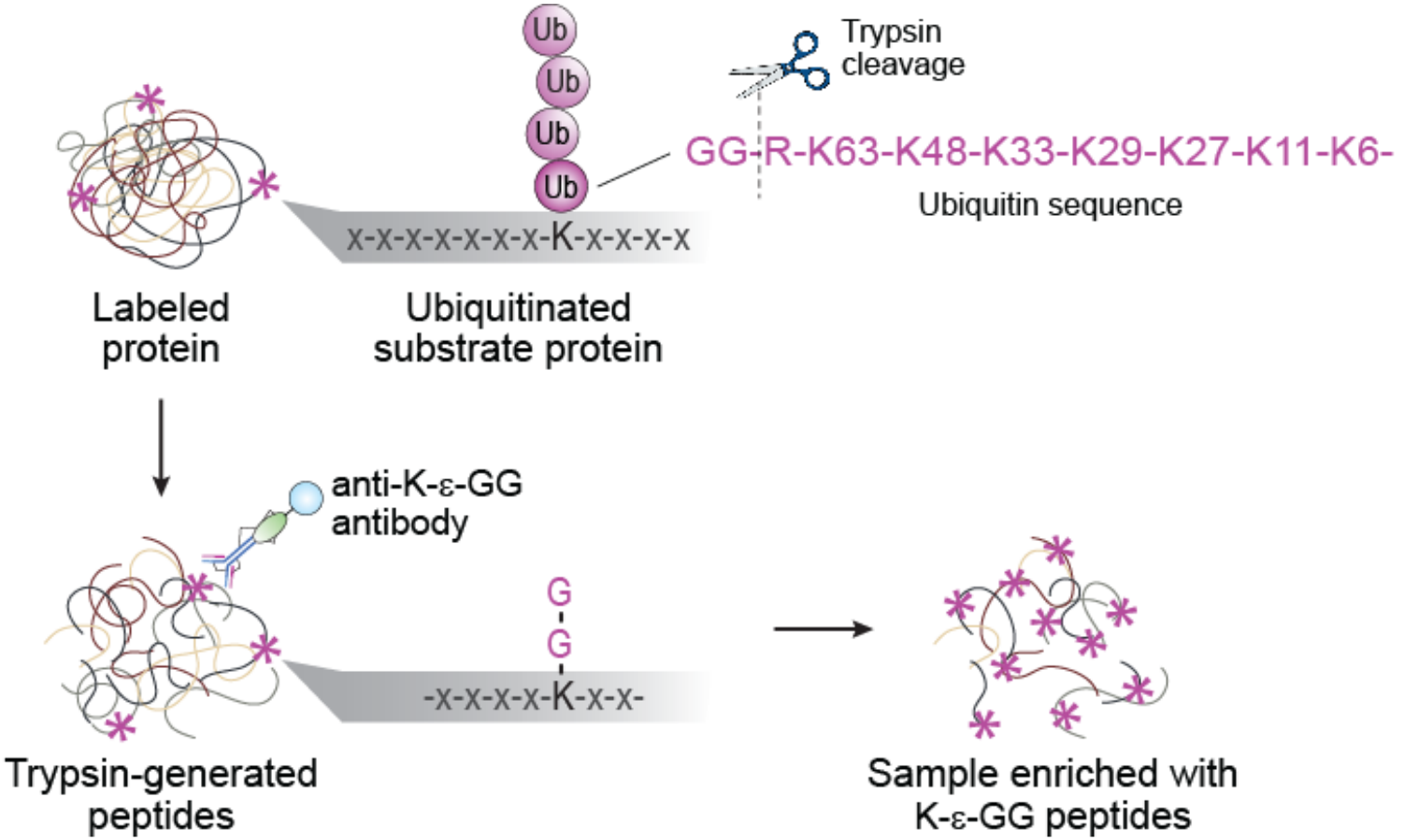
Ubiquitination
First publication 1978[b]
10 E1s[f], ~40 E2s[f], >600 E3 ligases[f]
~90 deubiquitinases[c]
Nobel Prize awarded 2004[e]
First drug approved in 2003 (Bortezomib)
One drug approved, 16 undergoing clinical trials
Current sales ~US$1.4 billion per year
<1% of pharmaceutical research and development

Cohen and Tcherpakov *Cell* 2010



Bedford et al. *Nat. Rev. Drug Discovery* 2011

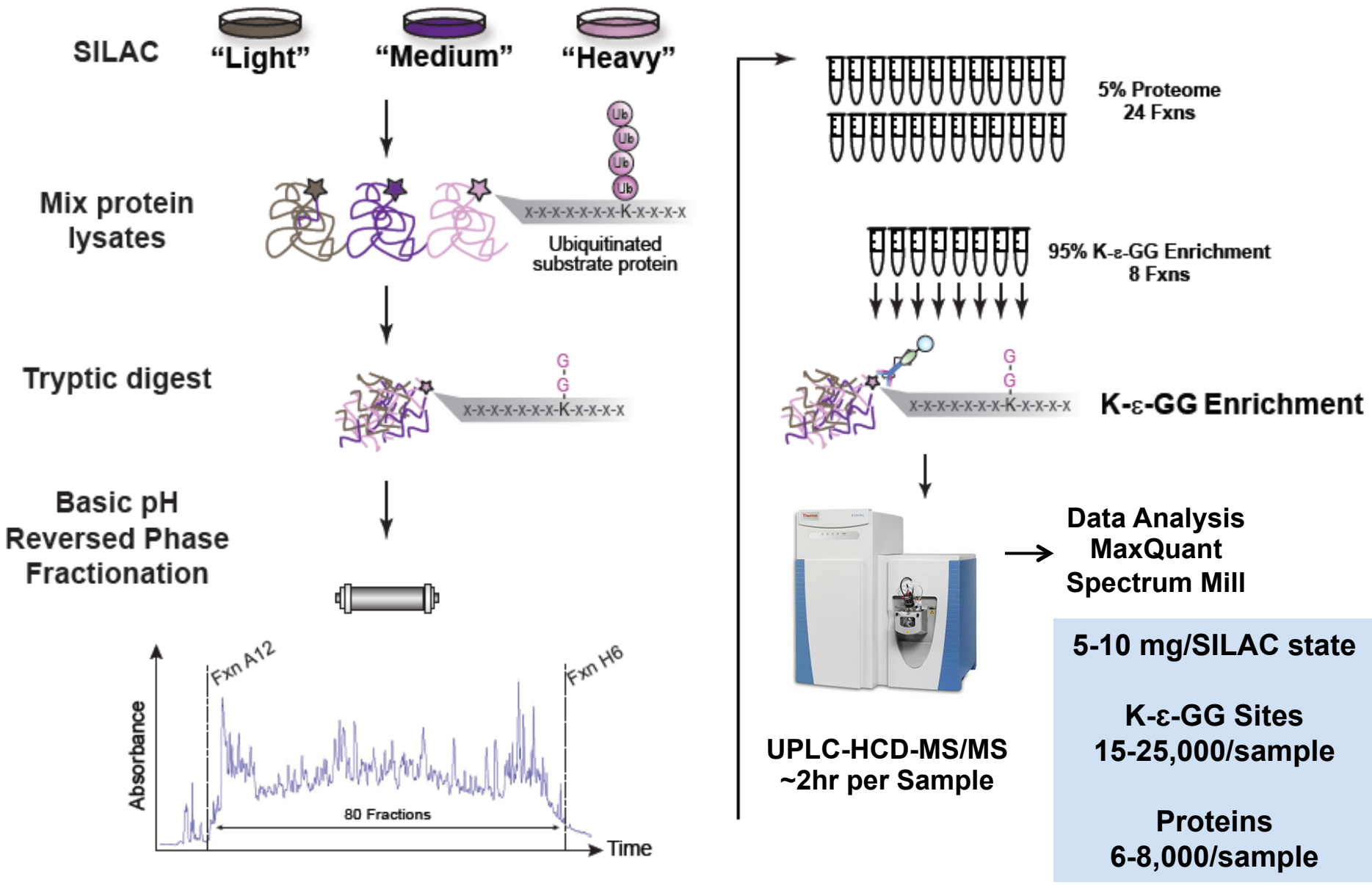
Antibodies to K-ε-GG Motif Enable Global Ubiquitin Analysis



- **>70% enrichment specificity**
- **Cannot distinguish mono vs polyubiquitination**

Udeshi et al Nature Protocols 2013
Kim et al Cell 2011
Wagner et al MCP 2011
Xu et al Nature Biotechnology 2010

Quantitative Proteomics Workflow for Ubiquitination Profiling



Lenalidomide Regulates the Ubiquitin and Protein Levels of IKZF1 and IKZF3 Transcription Factors

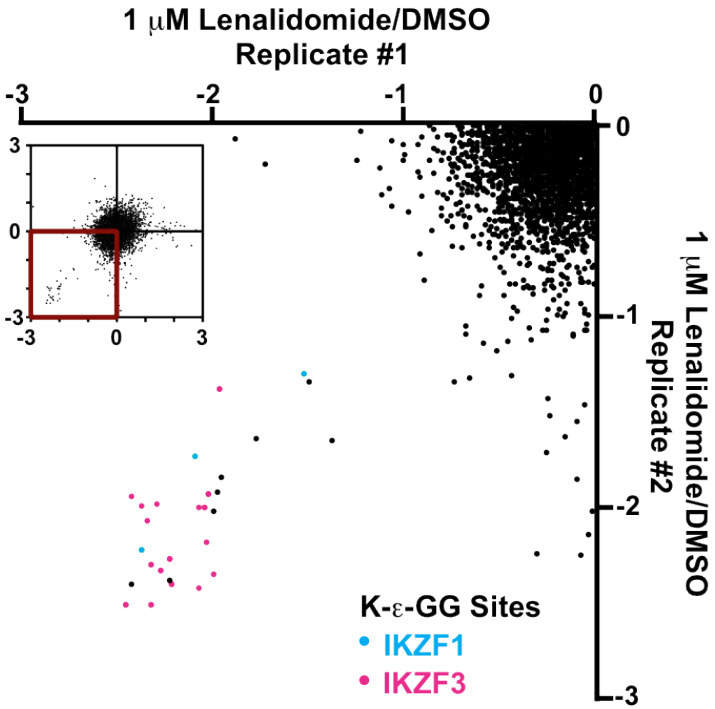
MM1S cells
Treatment Time: 12 h

DMSO
"Light"

20uM Thal
"Medium"

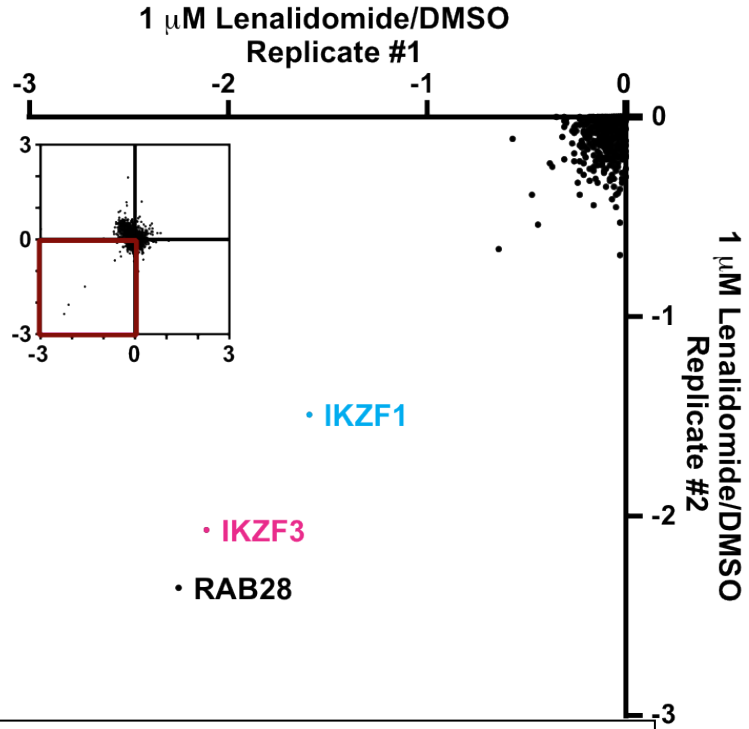
1uM Len
"Heavy"

Ubiquitin Profiling



Distinct Quantified K-ε-GG Sites
Average 3 Replicates
20,787
Overlap 3 Replicates
12,729

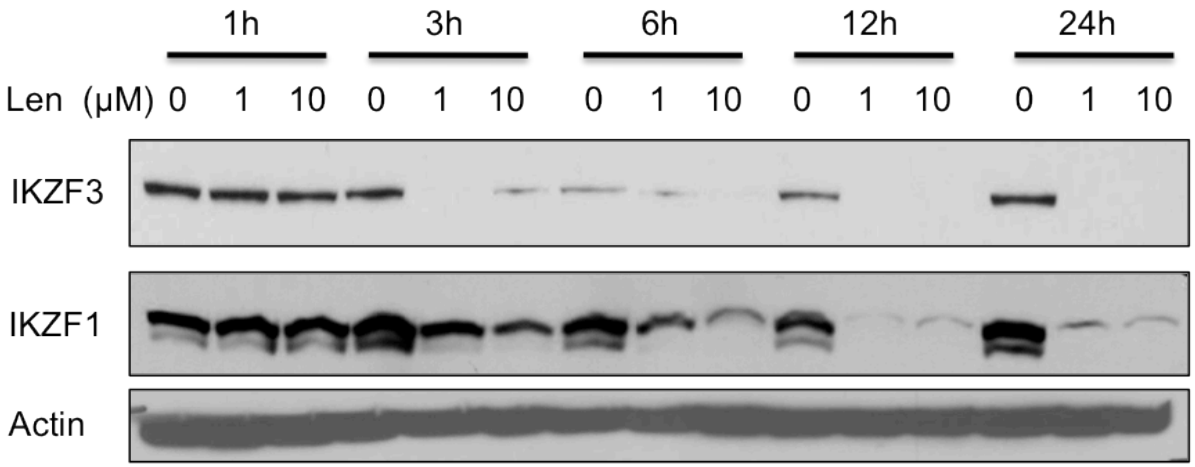
Proteome Profiling

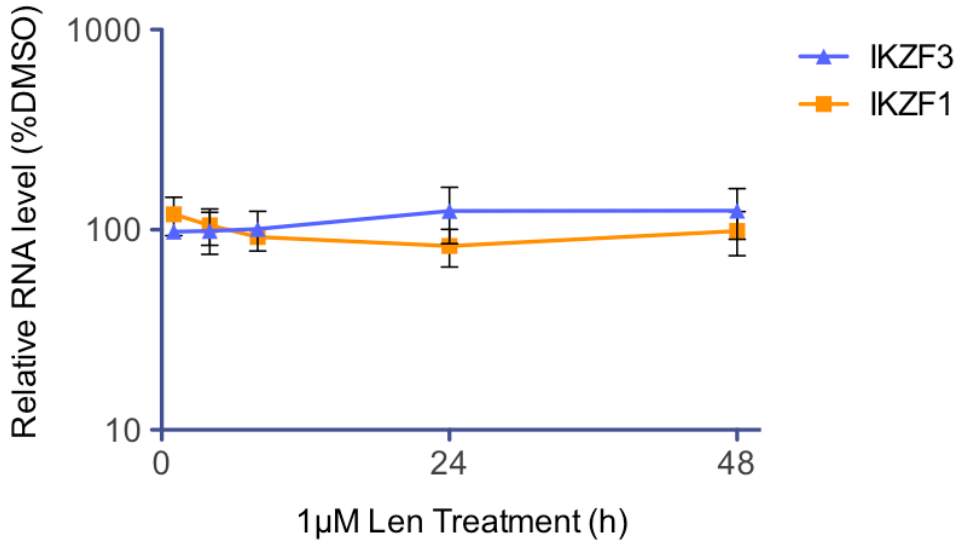


Distinct Quantified Protein Groups
Average 2 Replicates
6,012
Overlap 2 Replicates
6,102

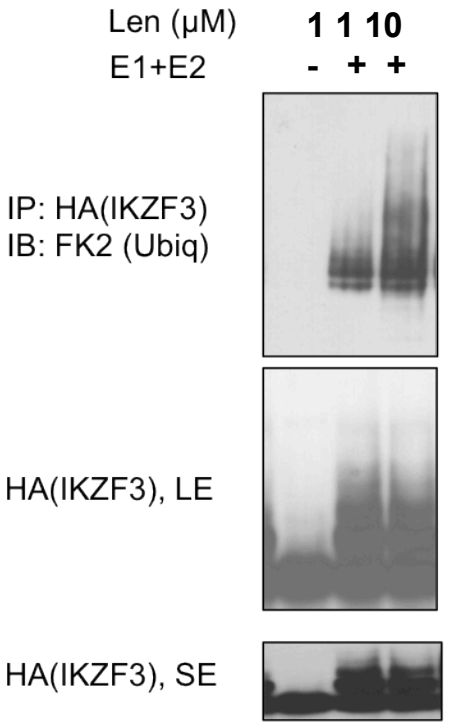
Krönke J, et al. Science. 343, 301 (2014)

Lenalidomide Decreases IKZF1 and IKZF3 Protein Levels <3h





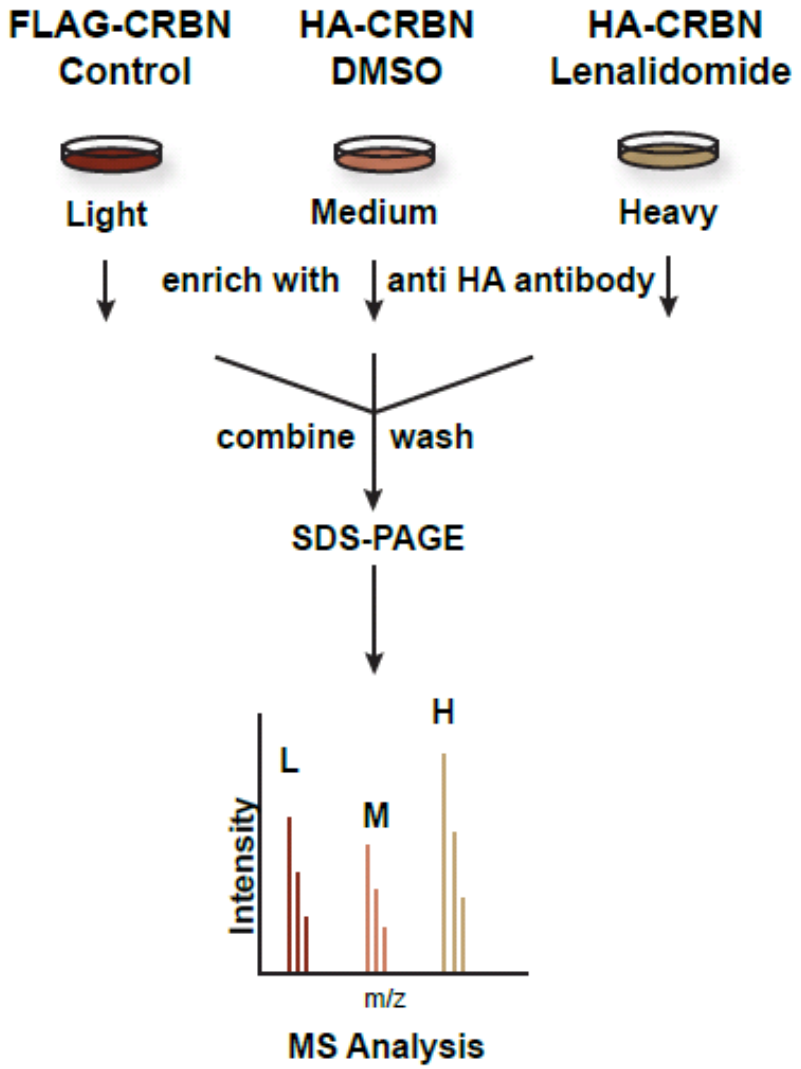
IKZF3 is ubiquitinated *in vitro* by CRBN

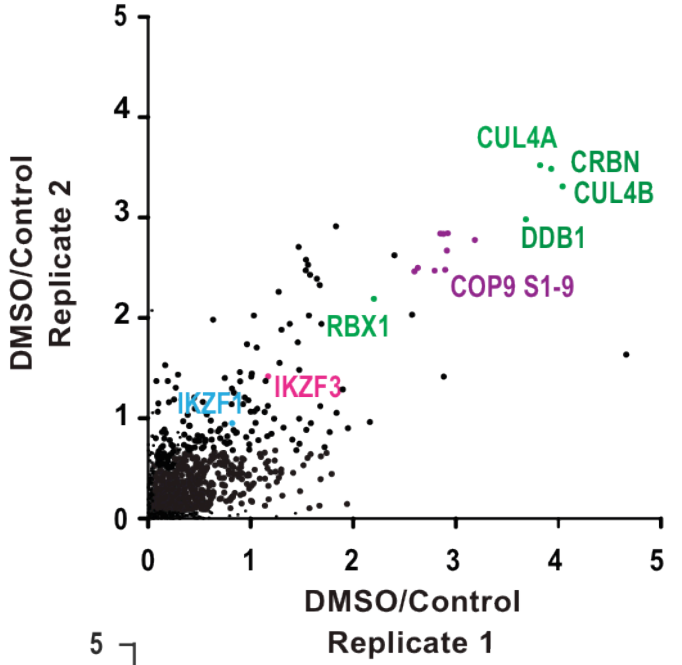


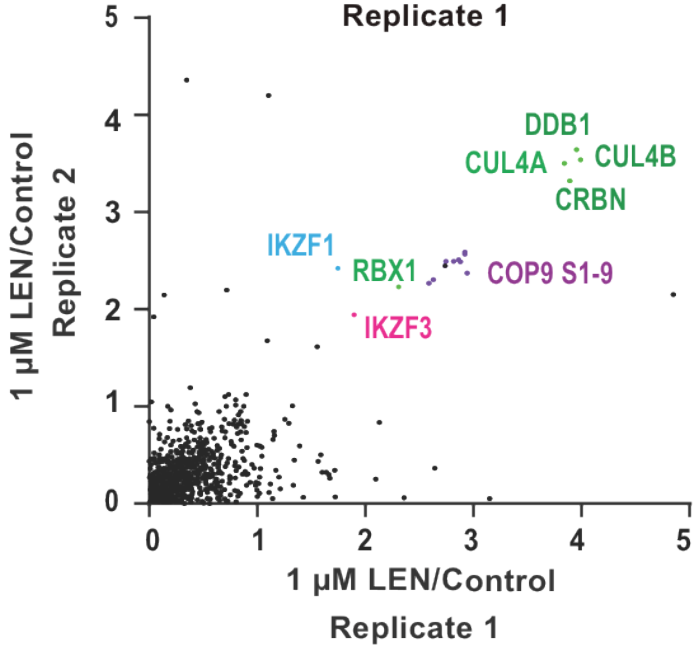
Krönke J, et al. Science. 343, 301 (2014)

Lenalidomide Increases IKZF1 and IKZF3 Binding to CRBN

What is the landscape of len-dependent CRBN protein interactions?

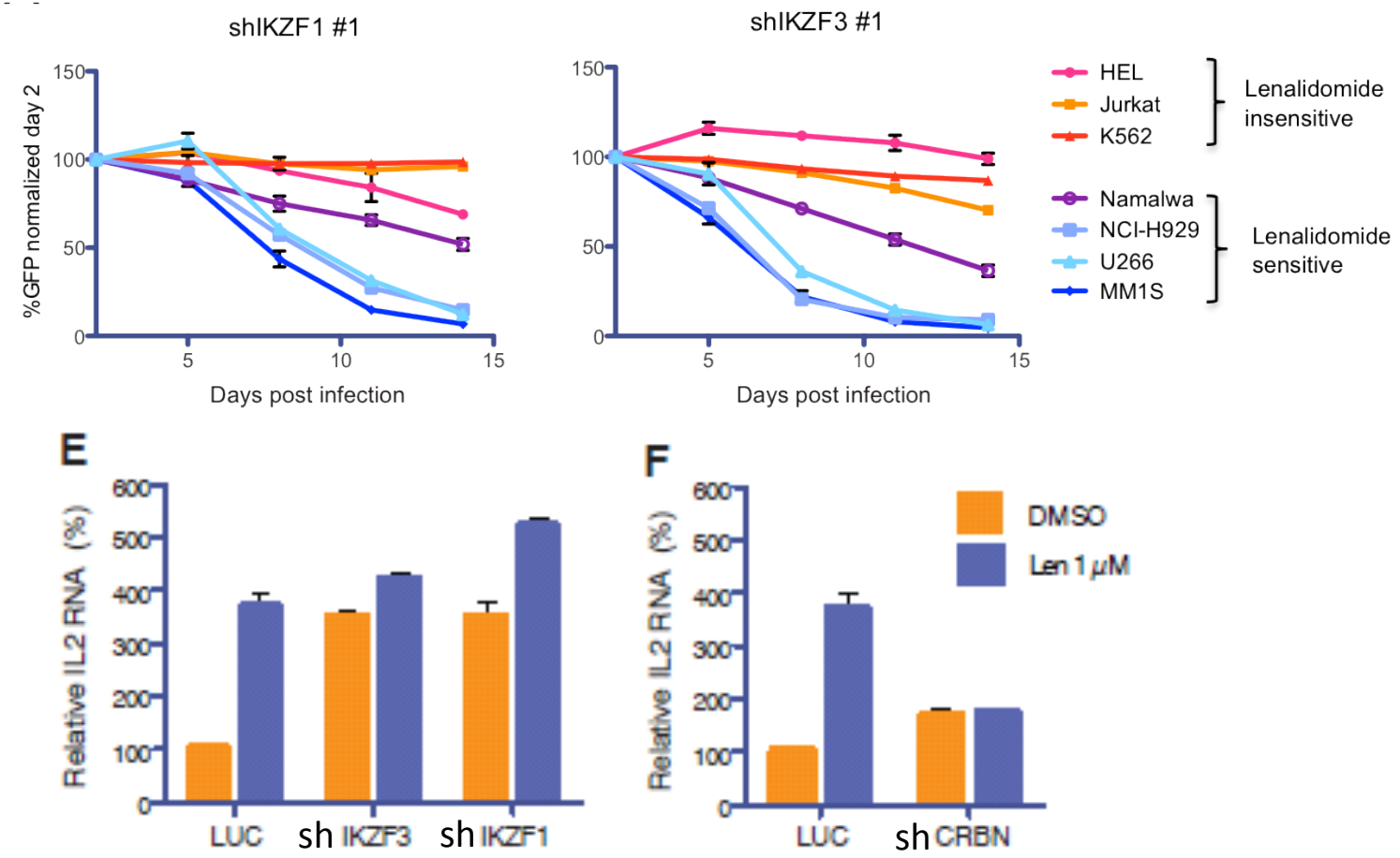






Krönke J, et al. Science. 343, 301 (2014)

Myeloma Cells are IKZF1 and IKZF3 Dependent



- **IKZF1 and 3 are essential transcription factors in lymphopoeisis**
- **KD of IKZF1/3 inhibits growth of lenalidomide sensitive cells**
- **Len regulates IL-2 levels in T cells by modulating IKZF3 expression**

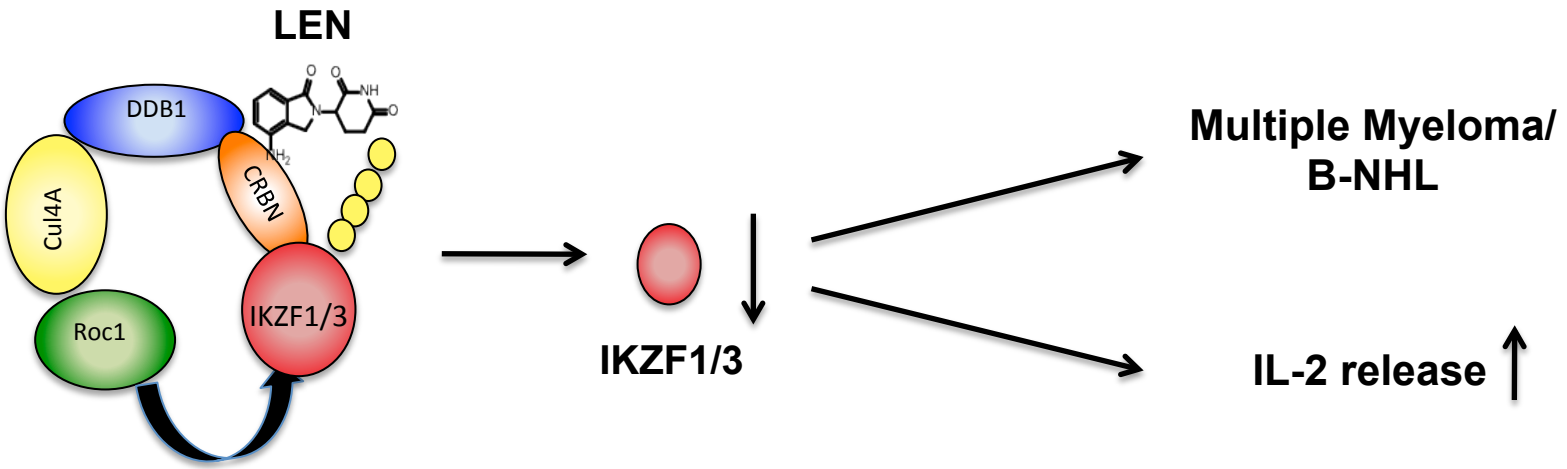
Krönke J, et al. Science. 343, 301 (2014)

Mechanism of Action of Lenalidomide

Major finding: Lenalidomide selectively induces the degradation of the Ikaros proteins IKZF1 and IKZF3. Anti-proliferative effect of lenalidomide in multiple myeloma cells is mediated by depletion of IKZF1 and IKZF3.

Mechanism: Lenalidomide promotes binding of IKZF1 and IKZF3 to CRBN, a ubiquitin ligase substrate receptor. First drug described to increase ubiquitin ligase activity.

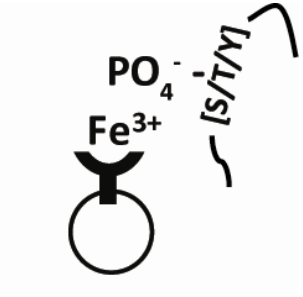
Impact: Therapeutic agents can downregulate specific targets by altering ubiquitin ligase substrate specificity.



Krönke J, et al. Science. 343, 301 (2014)
Lu G. et al Science. 343, 305 (2014)
Gandhi AK Br J Haematol. 164, 811 (2014)

Developing PTM analysis workflows: from single to serial enrichments

Historically: Single Enrichment



Phosphorylated peptides

36,000 p-sites (Huttlin, 2010)

Ubiquitinated peptides

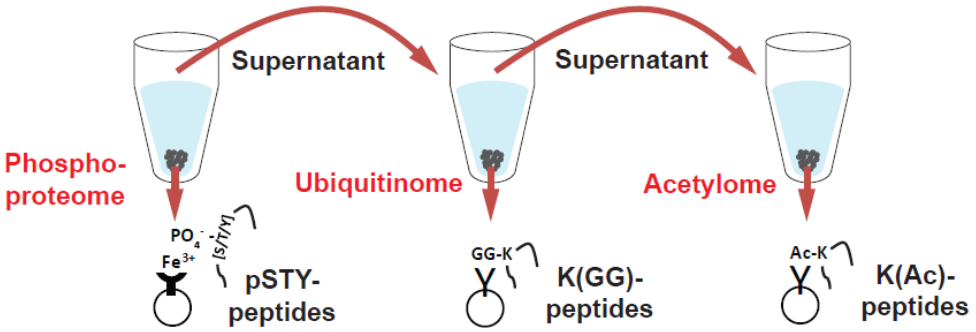
19,000 Ub-sites (Kim, 2011)

Acetylated peptides

15,000 Ac-sites (Lundby, 2012)

Goals: Serial (?)

1) Develop a pipeline that analyzes multiple PTMs from a single biological sample

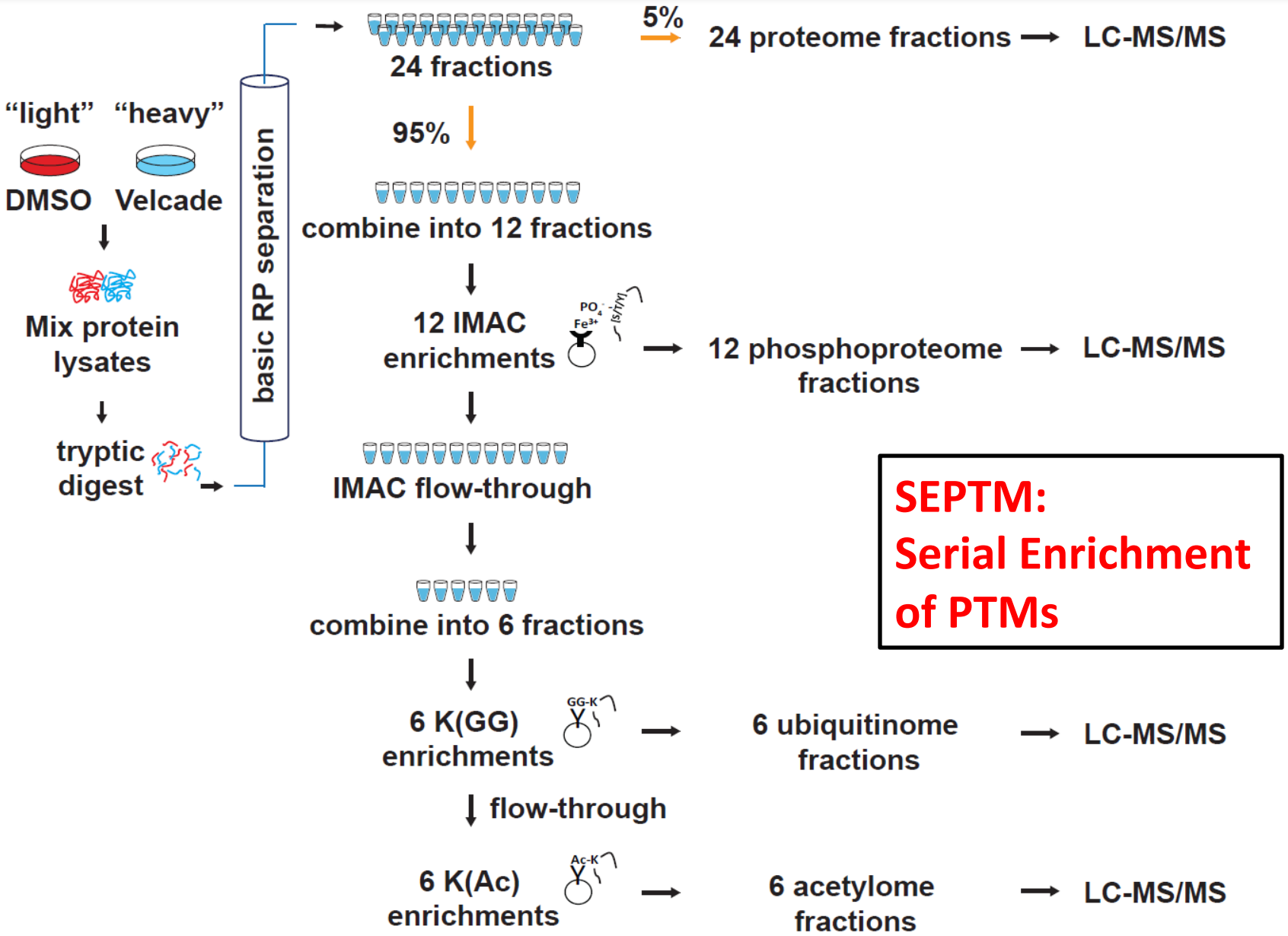


2) Obtain a similar depth of coverage as in single enrichment studies

3) Analyze PTM coverage according to fractional separation, MS time, and sample amount

Mertins et al; Nature Methods 2013

Combined workflow for proteome and PTM analysis



SEPTM:
Serial Enrichment of PTMs

"NexGen" proteomics has arrived: 4-5 fold increased detection/ quantification of proteins, PTMs in cells/tissues over past 3 yrs

- Appropriate study design
- Robust sample processing methods
- Quantitative labeling of peptides for multiplexed anal.
- Data acquired with state-of-the-art LC-MS technology
- Statistically rigorous data analysis

Unprecedented definition of proteins in cells and tissues

- 10K – 12K distinct proteins
- Precise and reproducible
- Higher throughput

Deep and broad PTM coverage

- >25K phosphosites
- >20K ubiquitinated peps
- >10K acetylation sites

- **The number of proteins observed in tissues now begins to approximate the expressed proteome**
- **PTM analysis provide window into function and pathogenesis not accessible by genomic methods**