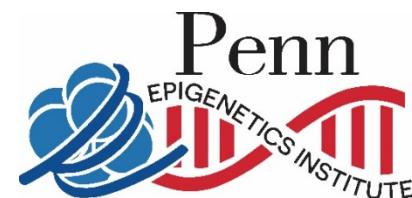
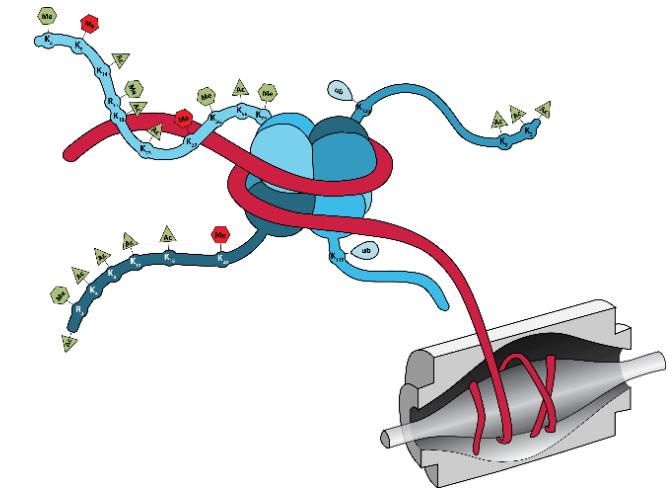


# Data independent acquisition for differential quantification of isobaric phosphopeptides and other protein post-translational modifications

Simone Sidoli, Johayra Simithy, Benjamin A. Garcia

06/04/2017, Skyline User Meeting, ASMS 2017



# The role of post-translational modifications

Advanced Review

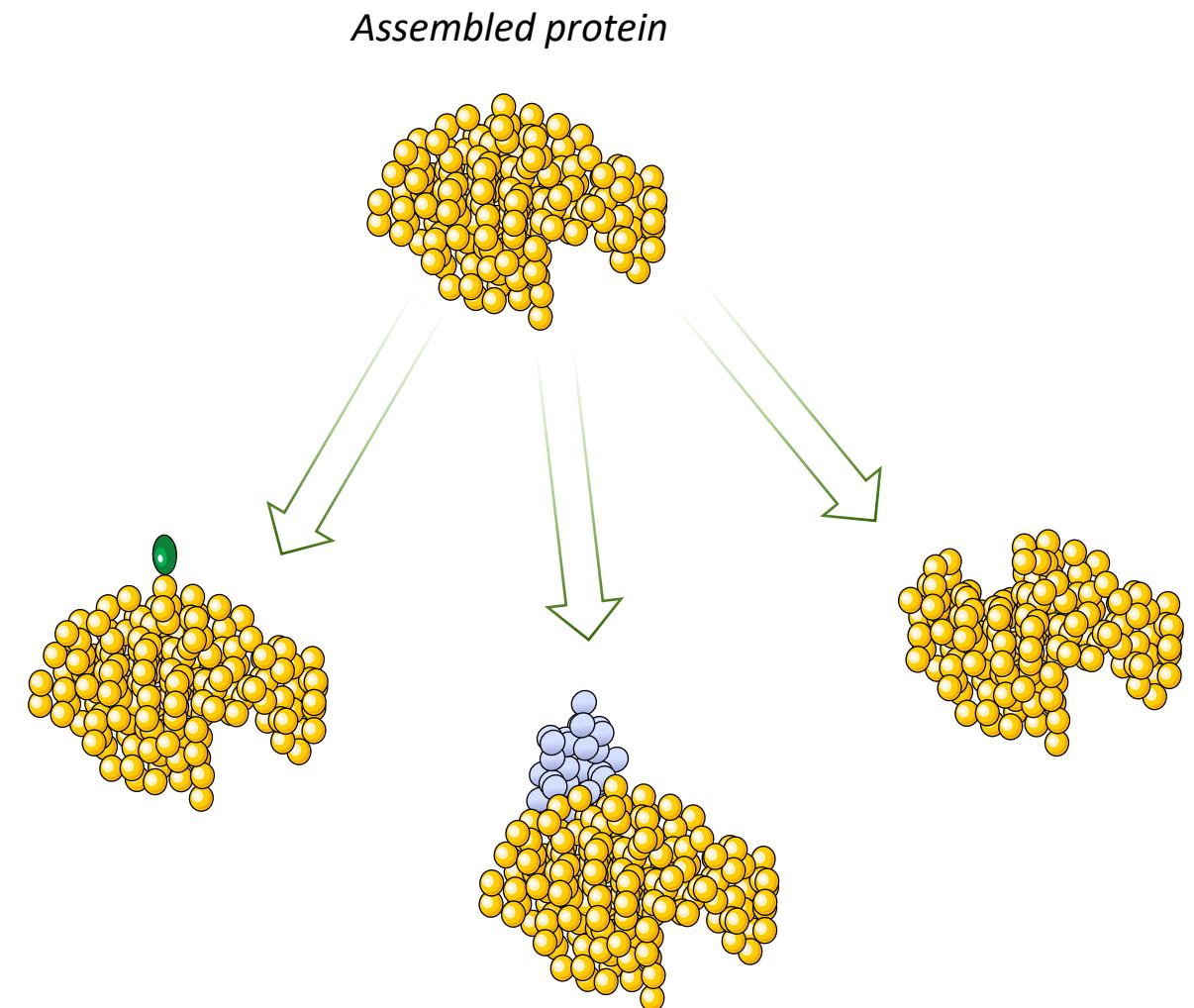
Post-translational modification:  
nature's escape from genetic  
imprisonment and the basis  
for dynamic information encoding

Sudhakaran Prabakaran,<sup>1</sup> Guy Lippens,<sup>2</sup> Hanno Steen<sup>3</sup>  
and Jeremy Gunawardena<sup>1\*</sup>



**Post-translational modifications (PTMs) are critical regulators of protein function, half-life and localization**

A post-translational modification is covalently bound to a protein after its translation. It can be a small chemical tag, or a big biomolecule, or a proteolytic cleavage

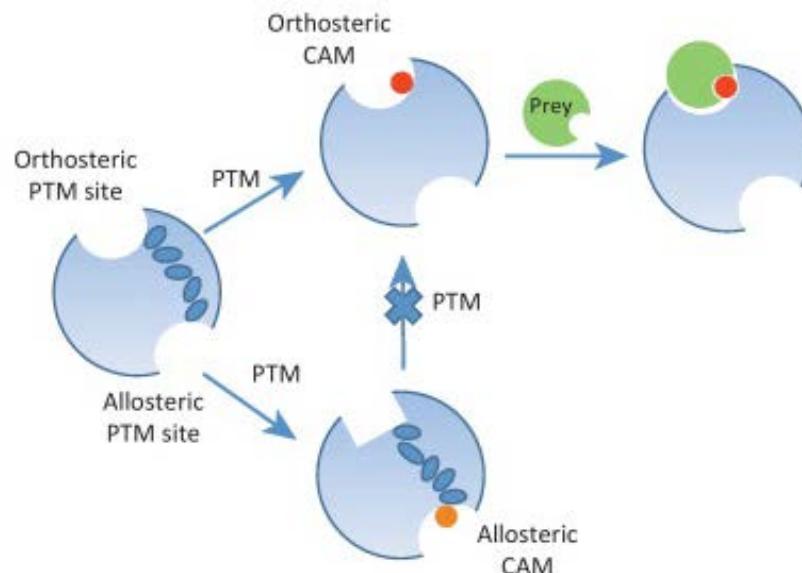


# The role of combinatorial post-translational modifications

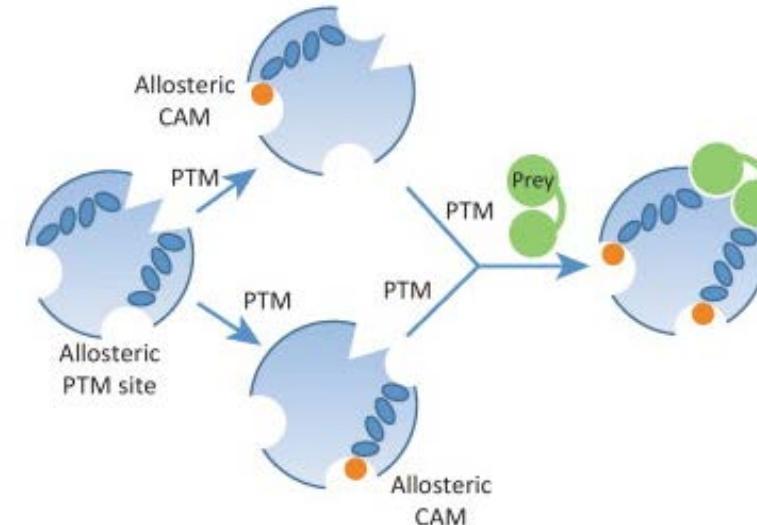
In the Garcia lab, we work with combinatorial PTMs and investigate their “cross-talk”

In protein science, a cross-talk between PTMs is an event where one PTM blocks or modifies the signal provided by another PTM. A cross-talk is “positive” if, for instance, a second PTM is required to fulfill a function that a single PTM cannot provide

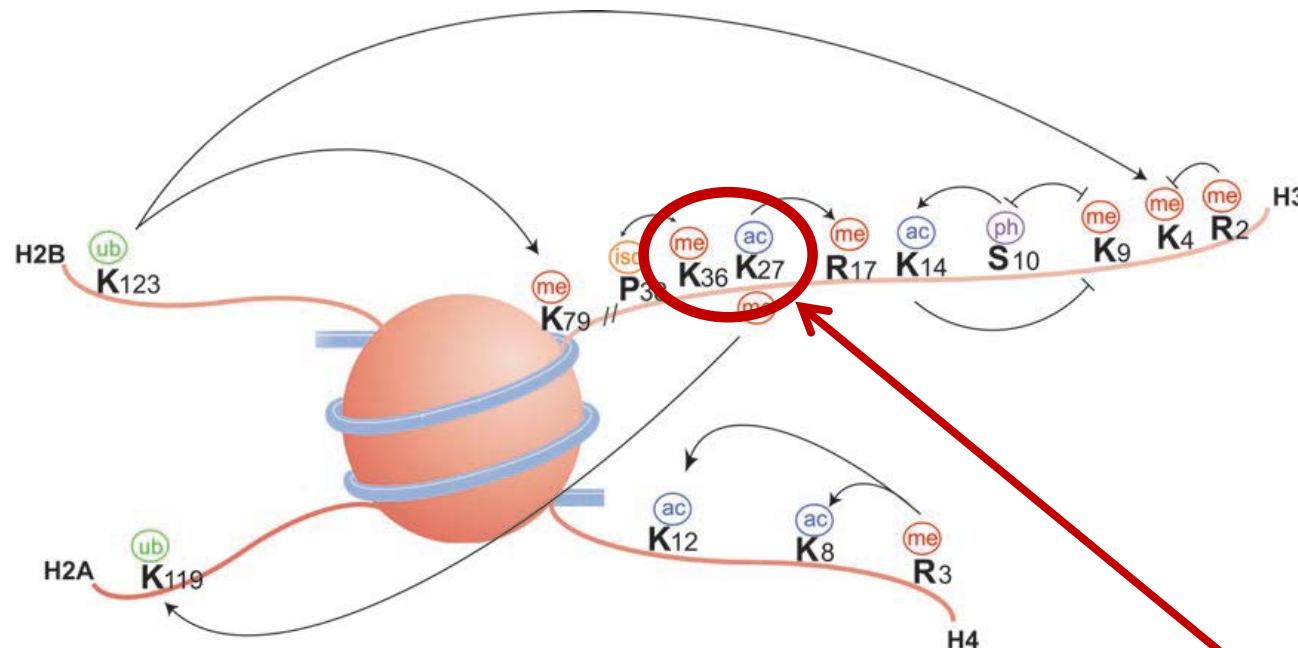
**Negative cross-talk**



**Positive cross-talk**



# Examples of histone PTM cross-talk



Numerous histone modifications are interdependent,  
i.e. the regulation of one affects the other one

## CANCER

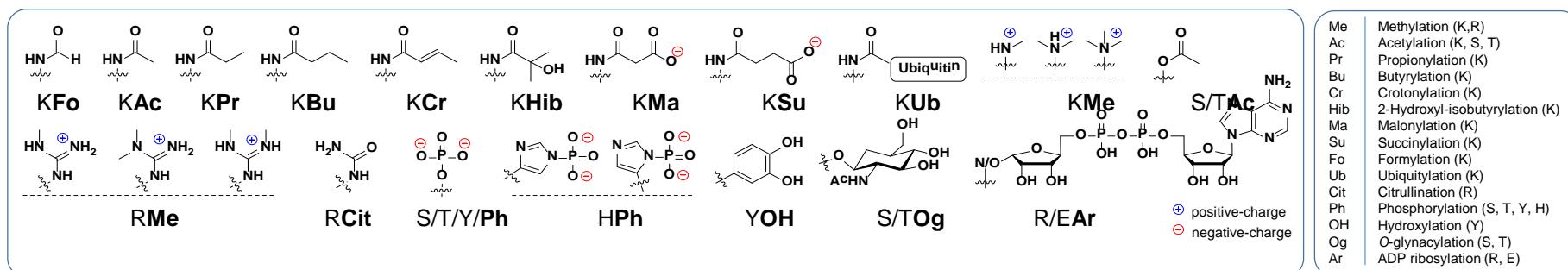
### Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape

Chao Lu,<sup>1</sup> Siddhant U. Jain,<sup>2,3</sup> Dominik Hoelper,<sup>2,3</sup> Denise Bechet,<sup>4</sup> Rosalynn C. Molden,<sup>5,6\*</sup> Leili Ran,<sup>7</sup> Devan Murphy,<sup>7</sup> Sriram Venneti,<sup>8</sup> Meera Hameed,<sup>9</sup> Bruce R. Pawel,<sup>10</sup> Jay S. Wunder,<sup>11,12</sup> Brendan C. Dickson,<sup>13,14</sup> Stefan M. Lundgren,<sup>2,3</sup> Krupa S. Jani,<sup>6</sup> Nicolas De Jay,<sup>4</sup> Simon Papillon-Cavanagh,<sup>4</sup> Irene L. Andrusis,<sup>13,14,15,16</sup> Sarah L. Sawyer,<sup>17</sup> David Grynspan,<sup>18</sup> Robert E. Turcotte,<sup>19</sup> Javad Nadaf,<sup>4</sup> Somayyeh Fahiminiyah,<sup>4</sup> Tom W. Muir,<sup>6</sup> Jacek Majewski,<sup>4</sup> Craig B. Thompson,<sup>20</sup> Ping Chi,<sup>7,21</sup> Benjamin A. Garcia,<sup>5</sup> C. David Allis,<sup>1†</sup> Nada Jabado,<sup>4,22†</sup> Peter W. Lewis<sup>2,3†</sup>

Oncohistones: histone mutations correlate with selected types of cancers

Nearby PTM sites (H3K27 and H3K36) interplay mutations and PTM levels

# Histone modifications



**H3**

+H<sub>3</sub>N-ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTV...QKST...IRKL...FKTDLRFQSS...DTN...AKR...PKD...ARRIRG ERA-COO-

10 20 30 40 50 60 70 80 90 100 110 120 130

**H4**

AcHN-SGRGKGGKGLGKGGAKRHRKVLRDN...QGITKPAIRR...VKRISGLIYEETRGVLKV...VI...RDAVTYTEHAKRKTVTAMDVVYALKRQG RTLYGF GG-COO-

10 20 30 40 50 60 70 80 90 100

**H2A**

AcHN-SGRGKQGGKARA KAKTRSSRAG LQFPVGRVHRLRKGNYAE RV...PVYL...LTA...ARDN KKT...IRNDE ELNKLLGKVTI...LLPKKTE SHH KAKGK-COO-

10 20 30 40 50 60 70 80 90 100 110 120

**H2B**

+H<sub>3</sub>N-P-EPAKSA...PKKGSKKAVTKAQQKKD...RKR SRKE SYS...YKVLK...TG I SK...S...SEASRLAHYNKRSTIT SRE...VRL...AKH...GTKAVTKY ISAK-COO-

10 20 30 40 50 60 70 80 90 100 110 120

**H1**

+H<sub>3</sub>N-SETA...EKA PVK KKA AKK AGGTPRKA SGPPV SELIT KAVA ASK ERS GVS LA AL KK AL...GY DVE KNNSRI KGL KSL VS KGT LV QT KGT GAS GS FKL NK

15 20 30 40 50 60 70 80 90 100

Hib Ac Me Ph Ph Ac/Hib Su/Fo/Ub Ac Ub Cit Ph Me Ac Cr Hib/Ub Ph Fo Cr Ub Hib Me Hib Ph Fo Ub Hib Me Hib Ub Hib Su Me Hib Ub Ac Me Hib Ub

209 190 185 180 170 160 157 144 140 130 120 110

-COO-KKKP...AKAASKA...PKTVKAKKP SKAVKK TVTAAPKK A...SK KPT A...AK KPKKAAGVPK KP KTGGAK KV KP KAEG SAAK

Boxes indicate globular domain

Almost every known PTM occurs on histones as well

The likelihood to have cross-talking PTMs is exponentially higher in hypermodified proteins

# Nearby PTMs is common in all proteins

**Resource**

**Cell**

## Systematic Functional Prioritization of Protein Posttranslational Modifications

Pedro Beltrao,<sup>1,3,\*</sup> Véronique Albanèse,<sup>4</sup> Lillian R. Kenner,<sup>1,3</sup> Danielle L. Swaney,<sup>5</sup> Alma Burlingame,<sup>2,3</sup> Judit Villén,<sup>5</sup> Wendell A. Lim,<sup>1,3,6</sup> James S. Fraser,<sup>1,3</sup> Judith Frydman,<sup>4</sup> and Nevan J. Krogan<sup>1,3,7,\*</sup>

Molecular Systems Biology 8; Article number 599; doi:10.1038/msb.2012.31  
Citation: *Molecular Systems Biology* 8:599  
© 2012 EMBO and Macmillan Publishers Limited. All rights reserved 1744-4292/12  
[www.molecularsystemsbiology.com](http://www.molecularsystemsbiology.com)

### Deciphering a global network of functionally associated post-translational modifications

Pablo Minguez<sup>1</sup>, Luca Parca<sup>2</sup>, Francesca Diella<sup>1,3</sup>, Daniel R Mende<sup>1</sup>, Runjun Kumar<sup>4</sup>, Manuela Helmer-Citterich<sup>2</sup>, Anne-Claude Gavin<sup>1</sup>, Vera van Noort<sup>1</sup> and Peer Bork<sup>1,5,\*</sup>

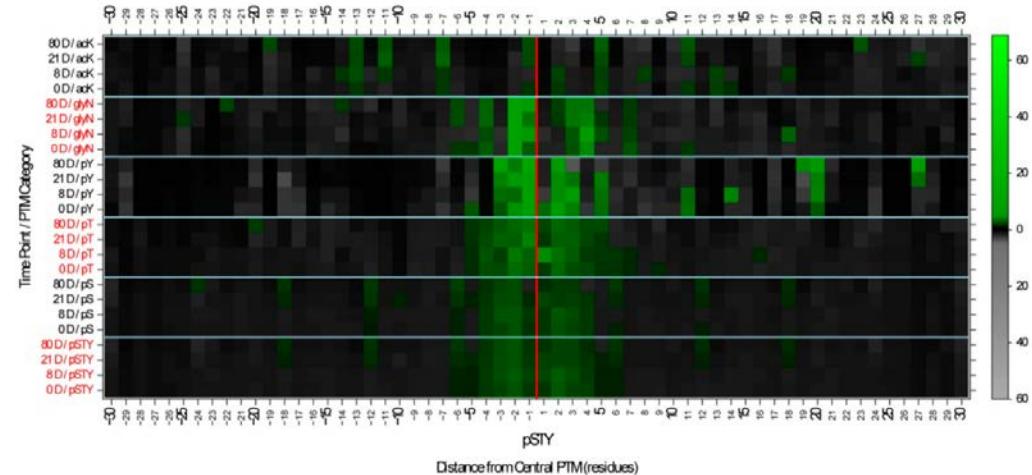
**Journal of proteome research**

**Article**

[pubs.acs.org/jpr](http://pubs.acs.org/jpr)

### Spatial and Temporal Effects in Protein Post-translational Modification Distributions in the Developing Mouse Brain

Alistair V. G. Edwards,<sup>†</sup> Gregory J. Edwards,<sup>‡</sup> Veit Schwämmle,<sup>†</sup> Henrik Saxtorph,<sup>§</sup> and Martin R. Larsen<sup>\*†</sup>



*Density plot of phosphorylation distances on a phosphoproteome. The x-axis is the distance between phosphosites (0 in the middle)*

PTMs are more frequently found nearby as compared to random. This is mostly due to enzyme docking, limitation in accessible protein surface, and evolutionary preservation of meaningful PTM domains

**Problem with mass spectrometry:**

**Nearby PTMs can easily co-localize on the same peptide**  
**i.e.**  
**Presence of isobaric peptides**

# Quantitative discrimination of isobaric peptides

ac  
|  
**K<sub>18</sub> Q L A T K A A R**

ac  
|  
**K Q L A T K<sub>23</sub> A A R**

**These PTMs have frequently different functions**

Discriminating their abundance is critical

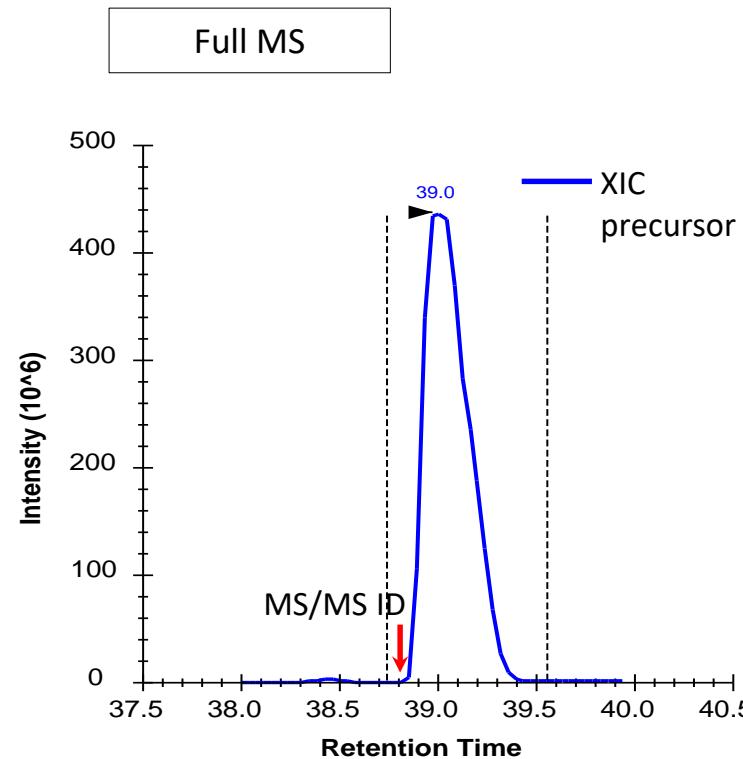
Extracting the MS ion chromatogram is insufficient

Sequence	H3K18ac		H3K23ac	
	<i>b</i>	<i>y</i>	<i>b</i>	<i>y</i>
K	227.15	1140.71	241.19	1140.71
Q	355.21	914.57	369.25	900.53
L	468.30	786.51	482.34	772.47
A	539.33	673.42	553.37	659.38
T	640.38	602.39	654.42	588.35
K	824.53	501.34	824.53	487.30
A	895.56	317.19	895.56	317.19
A	966.60	246.16	966.60	246.16
R	1122.70	175.12	1122.70	175.12

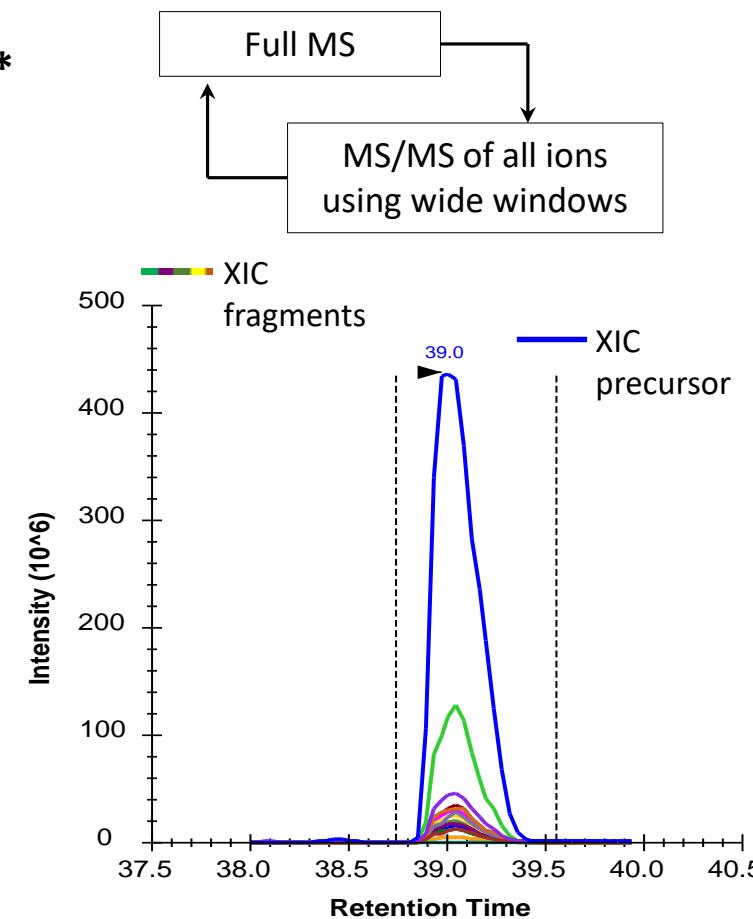
*Example with histone H3 K18/K23 acetylation*

# Data independent acquisition – profile of fragment ions

DDA

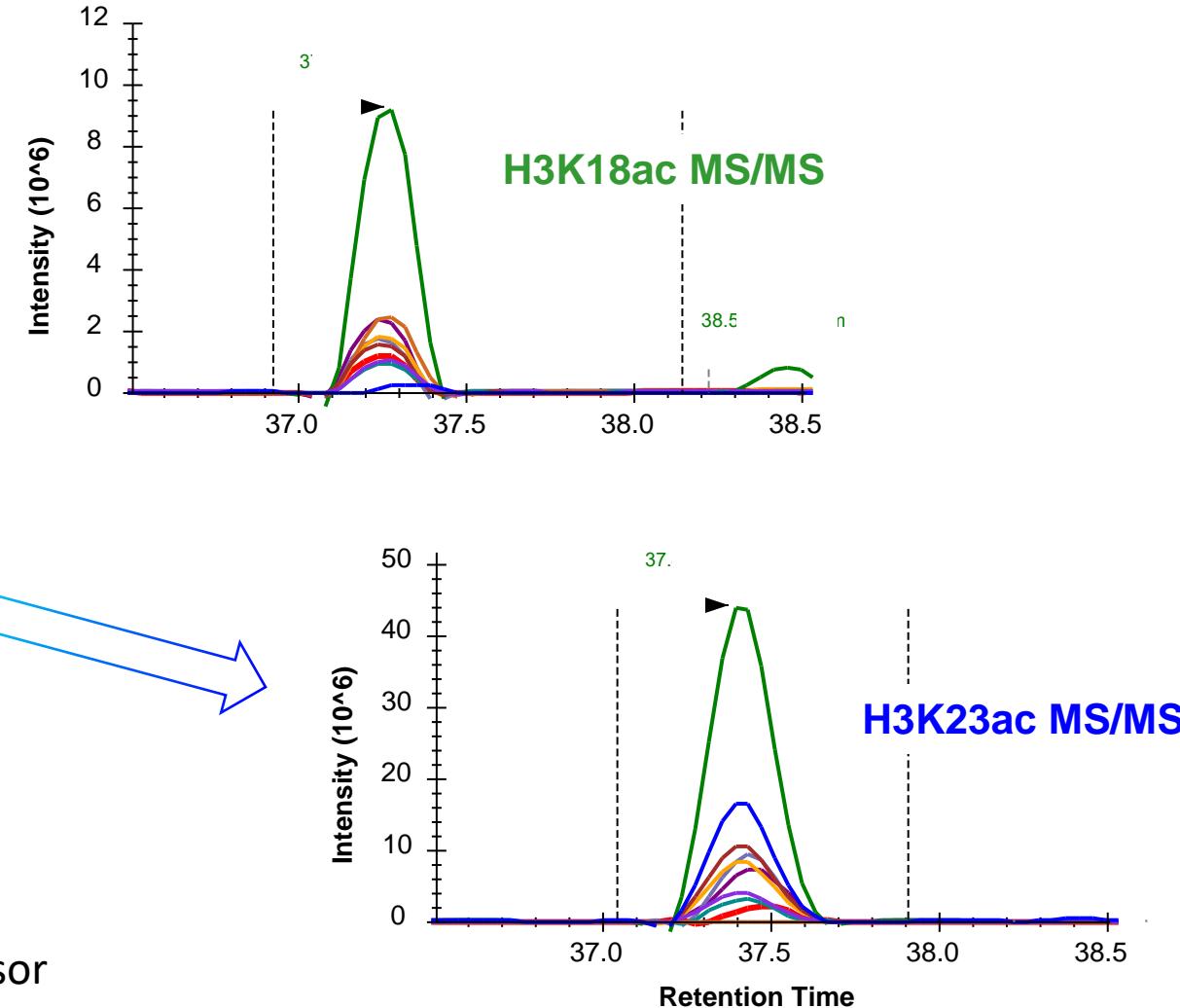
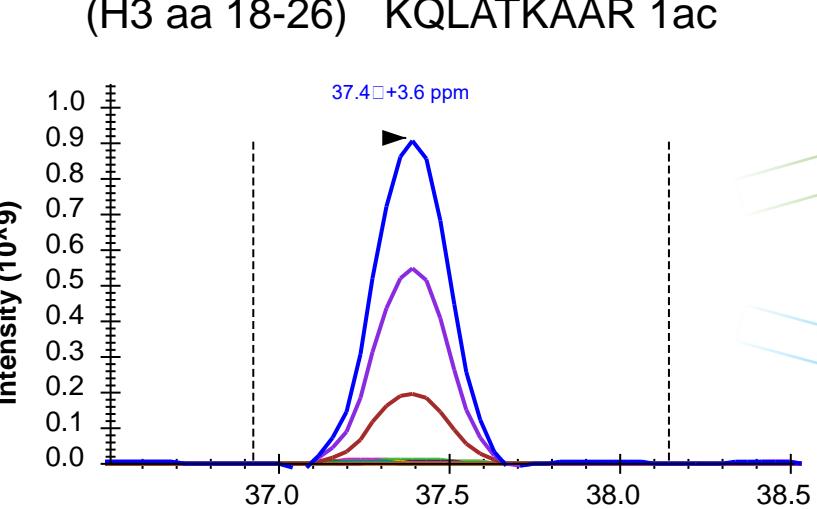


DIA\*\*\*



The Garcia lab has been committed in optimizing MS methods for discriminating isobaric peptides

# Quantitative discrimination of isobaric peptides



Fragment ions are used to divide the precursor area intensity between the two forms

## When isobaric forms are more than two

ac  
|  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac  
|  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac  
|  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac  
|  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

ac      ac  
|      |  
G K<sub>5</sub> G G K<sub>8</sub> G L G K<sub>12</sub> G G A K<sub>16</sub> R

## When isobaric forms are more than two

Diagram illustrating the calculation of medians for a set of 16 data points (y1 to y16) grouped into four categories (a, b, c, d) with keys K5, K8, K12, and K16 respectively.

Key calculations:

$$r_1 = \text{median} \left( \frac{a+b+c+d}{ab+bc+cd} \right)$$

$$r_2 = \text{median} \left( \frac{b+c+d}{bc+ca+ab} \right)$$

$$r_3 = \text{median} \left( \frac{c+d}{ca+cb+ba} \right)$$

$$a+b+c+d = 1$$

Below the diagram, the data points are listed as follows:

$y_{12}$	$y_{13}$	$y_{14}$	$y_{15}$	$y_{16}$	$y_{17}$	$y_{18}$	$y_{19}$	$y_{20}$	$y_{21}$	$y_{22}$	$y_{23}$	$y_{24}$	$y_{25}$	$y_{26}$	$y_{27}$	$y_{28}$
K5	K8	K12	K16	G	G	G	G	G	G	A	R	R	R	R	R	R
b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15	b16	b17

Below the data points, the following values are calculated:

- (a):  $k_{5ac}, k_{8ac}, k_{12ac}, k_{16ac}$
- (b):  $k_{5Do}, k_{8Do}, k_{12Do}, k_{16Do}$
- (c):  $r_1, r_2, r_3$
- (d):  $a+b+c+d=1$

On the right, a table shows the distribution of keys for each category:

LOC: K5	
a(1)	✓
b(2)	✓
c(3)	✓
d(4)	
e(5)	
f(6)	

Below the table, the following key distributions are listed:

- 1, 2 →  $\overline{ab+cd}$
- 2, 3 →  $\overline{abc+def}$
- 3, 5 →  $\overline{ab+cd+ef}$
- 5, 6 →  $\overline{abc+def}$
- 6, 3 →  $\overline{abc+def}$

```

exception: if atd > 1
            d = 1 - a
            b = 0
            c = 0
            elseif atdtb >
                  b = 1 - a - d
                  c = 0
            end

```

## 4 isobaric forms

```

        elseif  $r_1=0$  ( $r_2>0$ ,  $r_3>0$ ,  $r_4>0$ )
             $a=0$ ,  $b=r_2$ 
             $-c=r_3$ 
             $b+c+d=r_4$ 
            if  $r_2>r_4$ 
                 $d=0$ ,  $e=0$ 
                 $b=r_2$ 
                if  $r_3>r_3$ 
                     $c=0$ 
                else
                     $c=r_3-r_2$ 
             $f=1-b-c$ 
        elseif  $r_2>r_4$ 
             $c=0$ ,  $e=0$ 
             $b=r_3$ 
            if  $r_3>r_2$ 
                 $d=0$ 
            else
                 $d=r_2-r_3$ 
             $f=1-b-d$ 
        else ( $r_2<0$ ,  $r_3<0$ ,  $r_4>0$ )
            if  $r_1>r_2$ 
                 $c=0$ 
                 $b=r_3$ 
                 $d=r_1+r_3$ 
                 $e=r_4+r_2$ 
            else
                 $d=0$ 
                 $b=r_2$ 
                 $c=r_3-r_2$ 
            end
             $e=r_4-r_3$ 
        end
         $f=1-f$ 
    elseif  $r_5=0$  ( $r_1>0$ ,  $r_2>0$ ,  $r_3>0$ ,  $r_4>0$ )

```

## 6 isobaric forms

if  $R'_3 > R'_4$   
 else,  $e = 0$   
 $b = R'_3$   
 if  $a + b = 1$   
 $b = 1 - a$   
 else  
 $c = 1 - a - b$   
 end  
 elseif  $R'_1 > R'_4$   
 $c = 0, e = 0$   
 $b = R'_3$   
 if  $a + b = 1$   
 $b = 1 - a$   
 else  
 $d = 1 - a - b$   
 and  
 else ( $R'_3 < R'_4, R'_1 < R'_4$ )  
 if  $R'_3 > R'_4$   
 $c = 0$   
 $b = R'_3$   
 $d = R'_3 - R'_4$   
 $e = R'_4 - R'_3$   
 else  
 $d = 0$   
 $b = R'_3$   
 $c = R'_3 - R'_4$   
 $e = R'_4 - R'_3$   
 end  
 end

else ( $R'_2 > R'_3, R'_2 > R'_4, R'_3 > 0, R'_4 > 0$ )

$1 - f = \frac{R'_2}{R'_3}$   
 $R'_4 a + 1 - a f = R'_4$   
 or  
 $a = R'_4 / (R'_4 f + 1 / R'_1 - 1)$   
 $f = 1 - R'_4 / (R'_1 R'_4 + 1 - R'_1)$   
 if  $a f + f = 1$   
 $b = 0, c = 0, d = 0, e = 0, f = 1 - a$   
 elseif  $a > R'_2$   
 $b = a, c = 0$   
 if  $a > R'_3$   
 $c = 0, f = 1 - a - f$   
 else  
 $c = R'_3 - a$   
 if  $a f + f c = 1$   
 $c = 1 - a - f - c$   
 else  
 $e = 1 - a - f - c$   
 end  
 end  
 elseif  $a = R'_3, (a < R'_2)$

$b = 0$   
 $c = 0$   
 $d = R'_2 - a$   
 if  $a f + f d = 1$   
 $d = 1 - a - f, e = 0$   
 else  
 $e = 1 - a - f - d$   
 end  
 else ( $a < R'_2, a < R'_3, \text{ Similar to } R'_2 > 0$ )  
 $b = R'_2 - a - R'_3$   
 $b + c = 1 - a - R'_3$   
 $b + c + d + e = 1 - a - f = R'_4$

if  $R'_3 > R'_4$   
 $d = 0, e = 0$   
 $b = R'_3$   
 if  $a f + f b = 1$   
 $b = 1 - a - f$   
 $c = 0$   
 $c = 1 - a - f - b$   
 end  
 elseif  $R'_3 > R'_4$   
 $c = 0, e = 0$   
 $b = R'_3$   
 if  $a f + f b = 1$   
 $b = 1 - a - f$   
 $d = 0$   
 else  
 $d = f - a - b$   
 end  
 else ( $R'_3 < R'_4, R'_2 < R'_4$ )

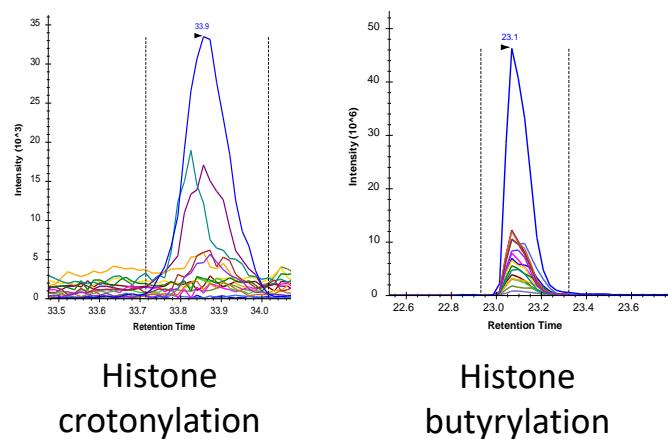
$b = R'_2 - a - R'_4$   
 $b + c = 1 - a - R'_4$   
 $b + c + d + e = 1 - a - f = R'_3$

# Software for histone PTM analysis

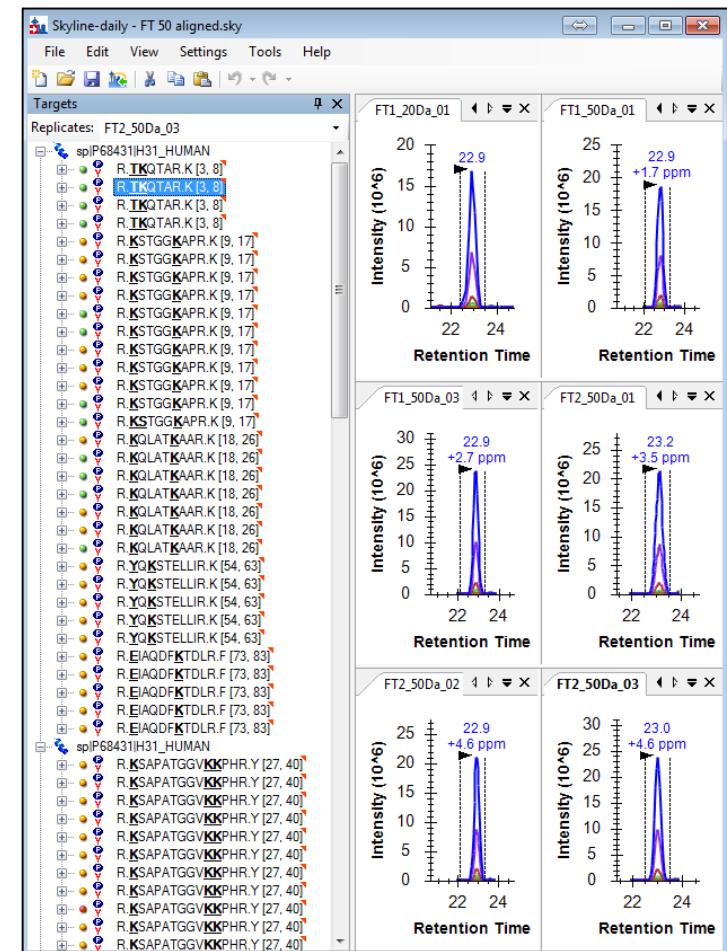
## (1) MATLAB software



## (3) New Skyline templates for rare histone PTMs



## (2) Skyline templates



# Getting there also with middle-down MS



*Example of human canonical histone H3 N-terminal tail (aa 1-50) and possible modifiable sites*

The same precursor mass can be hundreds of isobaric forms (theoretically, many more!)

For quantification, we use a combination of MS/MS ion intensity (DDA) and counting # spectra corresponding to the same identification

<http://middle-down.github.io/Software/>

Software for middle-down Proteomics

View On GitHub DOWNLOADS: ZIP TAR

Welcome to the webpage of the middle-down Proteomics software tools. The page currently contains software to validate MS/MS spectra and quantify identified polypeptides by Mascot (Matrix Science, UK) database searching engine. The tools are made in collaboration between the University of Southern Denmark and the University of Pennsylvania. The website contains Histone Coder and isoScale (peer reviewed in Proteomics 2014, see below) and a new beta version of both integrated software called isoScale slim.

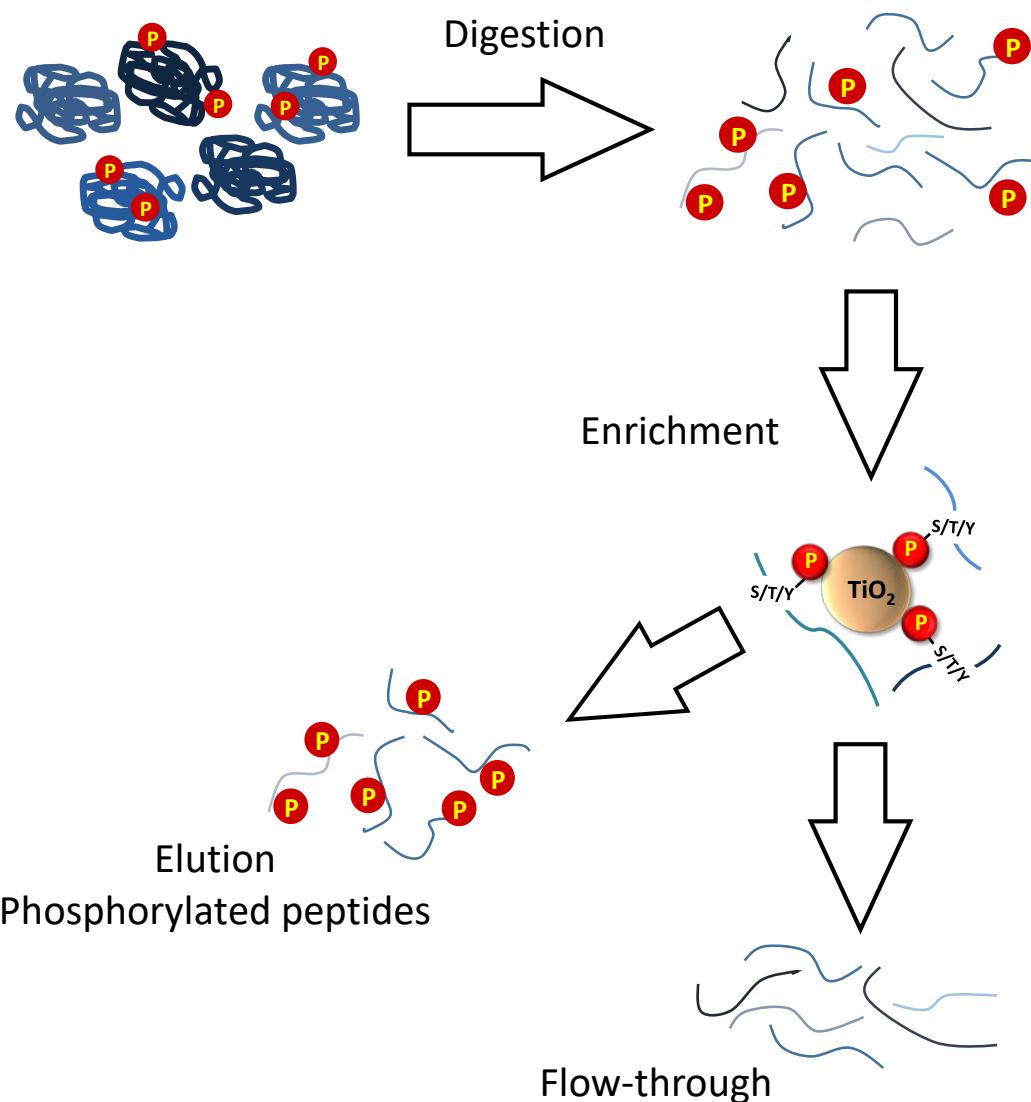
If you want to access our data repository, please visit the [Cross-talk database](#).

**Histone Coder - Download**

Histone Coder counts the number of MS/MS ions in a given spectrum to determine the unambiguous localization of a post-translational modification (PTM). The software lists number and type of site determining ions found between the assigned PTM localization by Mascot (Matrix Science) and the closest other amino acid which can host the modification. The PTMs included in

# **Isobaric forms in large-scale PTM-omics**

# Example with phosphoproteomics

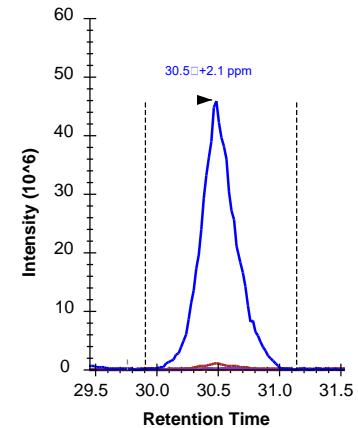
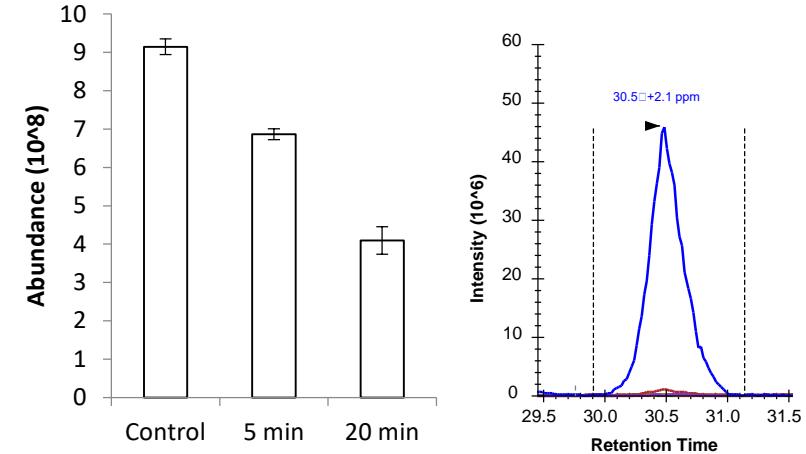
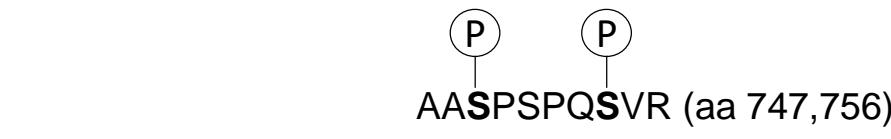
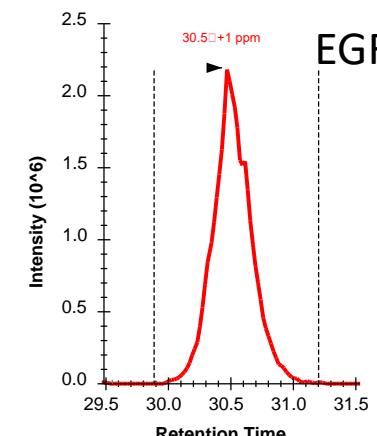
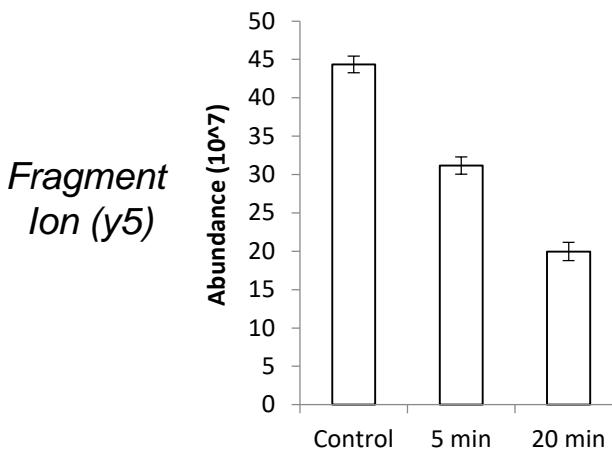
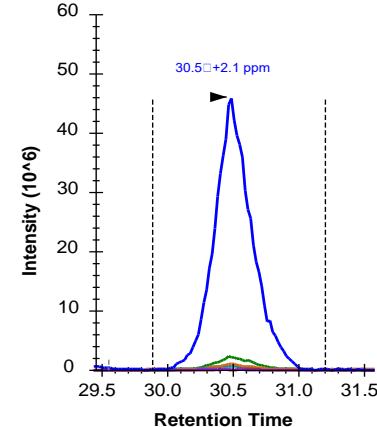
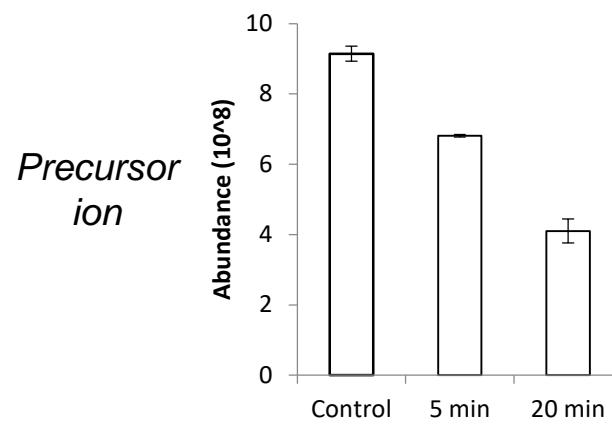
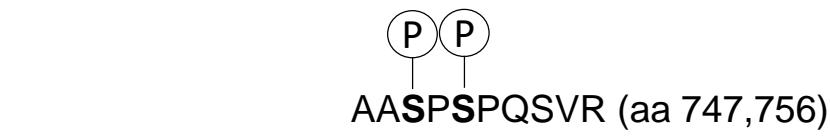


Phosphorylation is the most frequent PTM detected on proteins (>250,000 non-redundant sites, PhosphoSitePlus)

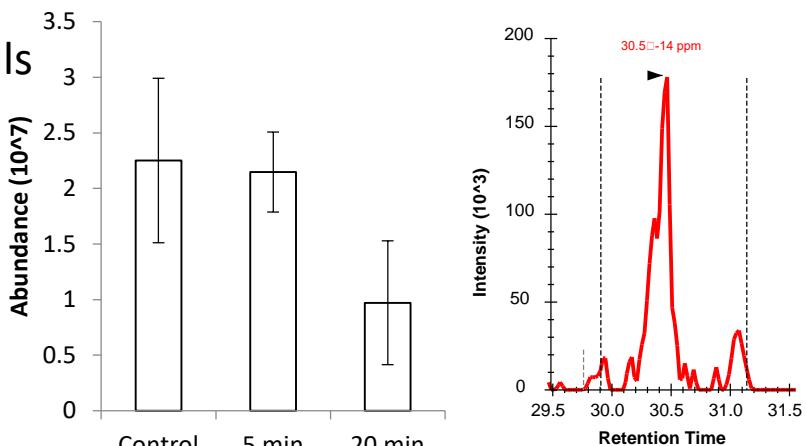
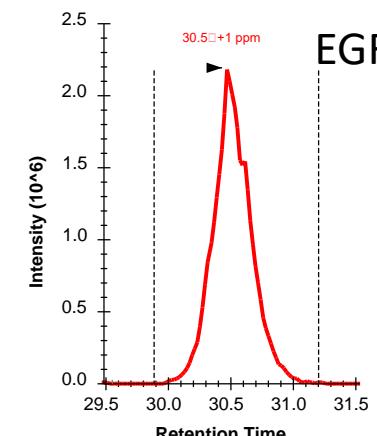
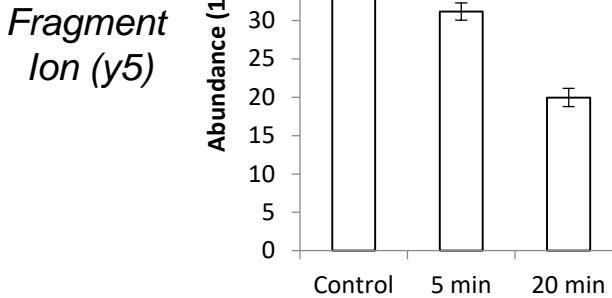
To discriminate isobaric phosphopeptides, we:

- Select isobaric phosphopeptides with same sequence and # of phospho from the spectral library (identified with DDA)
- Select fragment ions unique for each of the two forms
- Use the MS/MS extracted ion chromatogram to split the precursor MS area between the isobaric forms

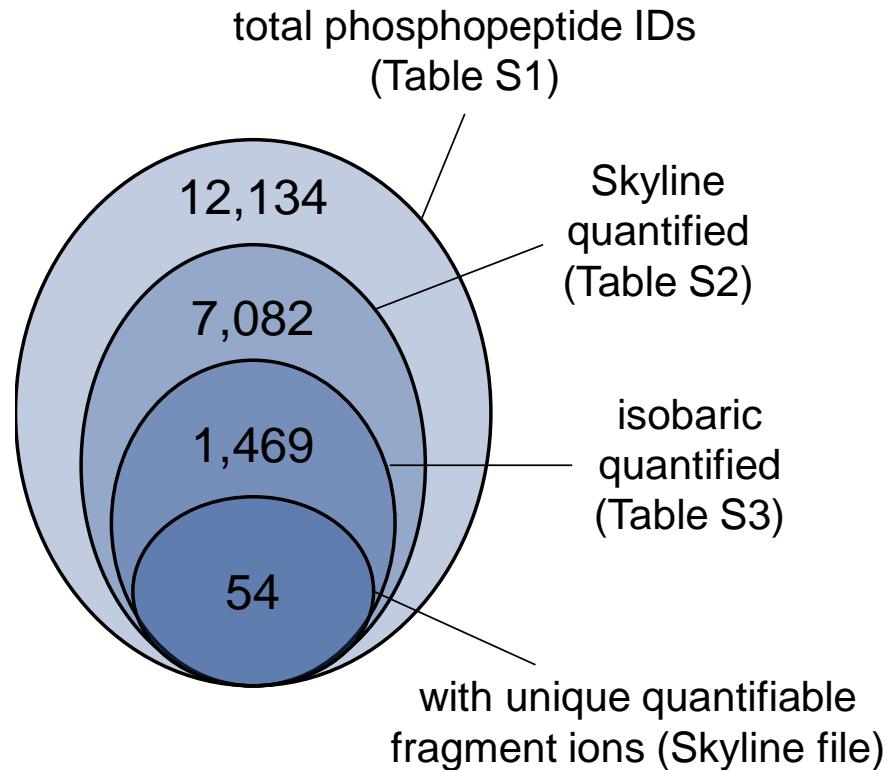
# Isobaric species in phosphoproteomics



EGF stimulation of HeLa cells



# Isobaric species in phosphoproteomics



Our first attempt was only partially successful.  
Some conclusions:

- 1) I clearly need to get better with Skyline
- 2) Standard DDA methods are insufficient to detect most isobaric phosphopeptides
- 3) Defining unique fragment ions useful for discriminating the isobaric forms is currently the major bottleneck

# We know we are not alone ☺



Article

pubs.acs.org/jpr

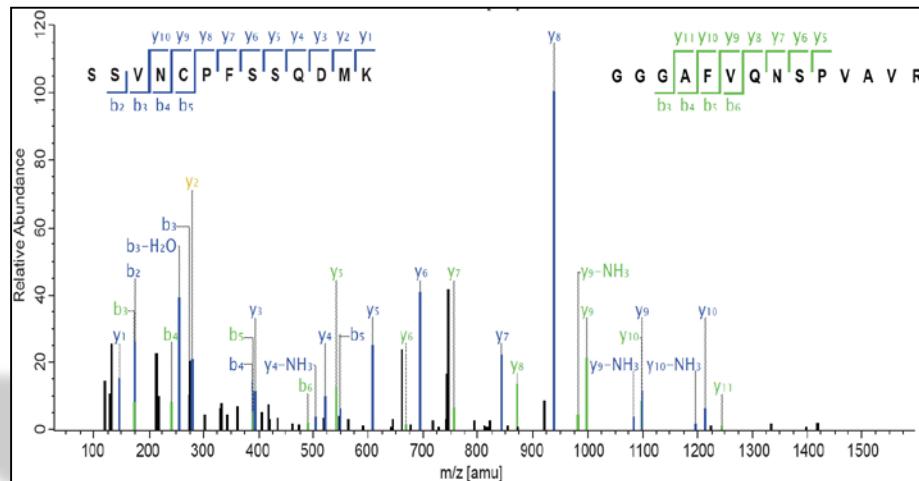
## FASIL-MS: An Integrated Proteomic and Bioinformatic Workflow To Universally Quantitate In Vivo-Acetylated Positional Isomers

Dijana Vitko,<sup>†</sup> Peter Májek,<sup>†</sup> Erika Schirghuber,<sup>†‡§</sup> Stefan Kubicek,<sup>†‡</sup> and Keiryn L. Bennett<sup>\*,†</sup>

<sup>†</sup>CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, A-1090 Vienna, Austria

<sup>‡</sup>Christian Doppler Laboratory for Chemical Epigenetics and Antiinfectives, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, A-1090 Vienna, Austria

### Second peptide search, MaxQuant



From ASMS 2017



### Quantifying phosphopeptide positional isomers in DIA experiments

#### Authors

Brian C. Searle<sup>1, 2</sup>; Michael J. MacCoss<sup>1</sup>; Judit Villén<sup>1</sup>

«... we find that in any given experiment approximately 40% of phosphopeptides exist as at least two significantly localized positional isomers.»

# Conclusions

Isobaric modified peptides have been overlooked for too long, while they hide fundamental information about biological systems and PTM cross-talk

- Independently from the quantification method (label-free, SILAC, isobaric labeling), the profile of the fragment ions is required to discriminate isobaric forms. DIA seems to be currently the only method suitable for the issue
  - *Because isobaric forms do not always completely co-elute, a single MS/MS spectrum is not sufficient to estimate their relative ratio*
- Future spectral libraries need to include isobaric peptides, and software need to cope with differential quantification of modified peptides with the same mass but different modified residues
  - *This is already a common practice in the analysis of histone peptides. It should become routine for large-scale proteomics as well*

# Acknowledgments



## The Garcia lab:

Benjamin A. Garcia

Johayra Simithy

Katarzyna Kulej

Zuo-Fei Yuan

Rina Fujiwara

## Funding:

NIH GM110174, CA196539 and  
AI118891

Leukemia & Lymphoma Society