V5: Protein phosphorylation during cell cycle

Protein **phosphorylation** and **dephosphorylation** are highly controlled biochemical processes that respond to various intracellular and extracellular stimuli.
They belong to post-translational modifications (PTMs).

Note: phosphorylation of histone tails also belongs to this class of PTMs.

Phosphorylation status modulates protein activity by
- influencing the tertiary and quaternary **structure** of a protein,
- controlling its **subcellular distribution (**e.g cytoplasm ⇔ nucleus for Per/Cry**)**, and
- regulating its **interactions** with other proteins.

Regulatory protein phosphorylation is a **transient modification**
that is often of low occupancy or “stoichiometry”

Low occupancy means that only a fraction of the copies of a particular protein may be phosphorylated on a given site at any particular time, or it occurs on regulatory proteins of low abundance, such as protein kinases and transcription factors.

Cell Cycle and the Phosphoproteome

CELL CYCLE

Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis

Jesper V. Olsen,[1,2*] Michiel Vermeulen,[1,3*] Anna Santamaria,[4*] Chanchal Kumar,[1,5*] Martin L. Miller,[2,6] Lars J. Jensen,[2] Florian Gnad,[1] Jürgen Cox,[1] Thomas S. Jensen,[7] Erich A. Nigg,[4] Søren Brunak,[2,7] Matthias Mann[1,2†]

(Published 12 January 2010; Volume 3 Issue 104 ra3)

www.SCIENCESIGNALING.org 12 January 2010 Vol 3 Issue 104 ra3

Aim: Analyze all proteins that are modified by phosphorylation during different stages of the cell cycle of human HeLa cells.

Ion-exchange chromatography + HPLC + MS + sequencing led to the identification of 6695 phosphorylated proteins („the phospho-proteome“). From this, 6027 quantitative cell cycle profiles were obtained.

A total of 24,714 phosphorylation events were identified.
20,443 of them were assigned to a specific residue with high confidence.

Finding: about **70%** of all proteins get phosphorylated.

Review: protein quantification by SILAC

ARTICLE

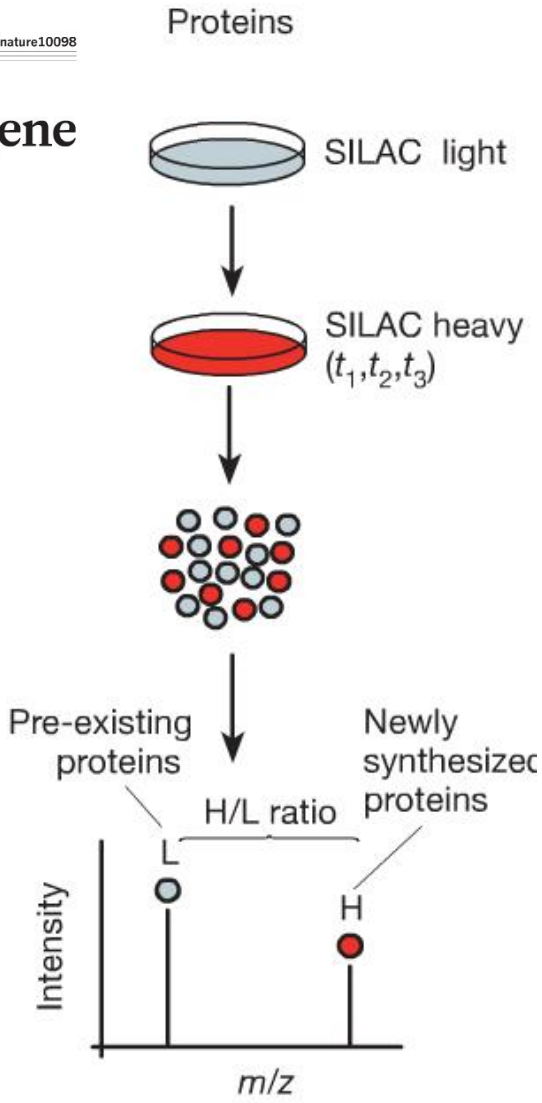
doi:10.1038/nature10098

Global quantification of mammalian gene expression control

Björn Schwanhäusser[1], Dorothea Busse[1], Na Li[1], Gunnar Dittmar[1], Johannes Schuchhardt[2], Jana Wolf[1], Wei Chen[1] & Matthias Selbach[1]

SILAC: „**s**table **i**sotope **l**abelling by **a**mino acids in **c**ell culture“ means that cells are cultivated in a medium containing heavy stable-isotope versions of essential amino acids.

When non-labelled (i.e. light) cells are transferred to heavy SILAC growth medium, newly synthesized proteins incorporate the heavy label while pre-existing proteins remain in the light form.



Protein turnover is quantified by mass spectrometry and next-generation sequencing, respectively.

Schwanhäuser et al. Nature 473, 337 (2011)

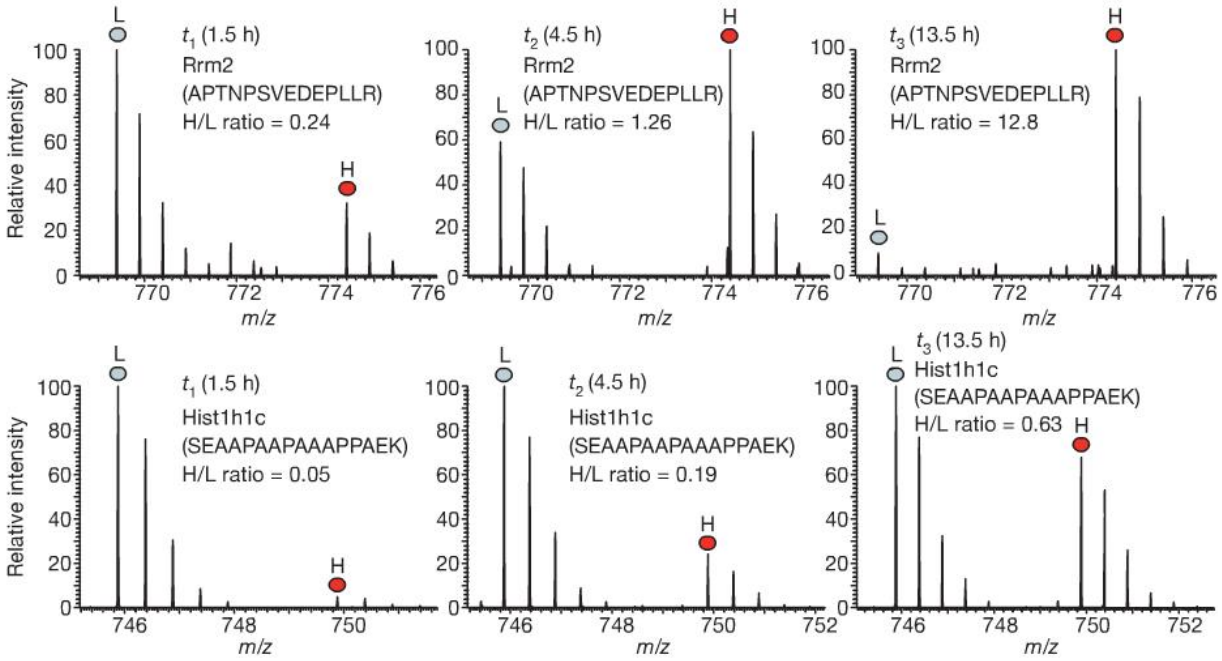
H/L ratios of individual proteins

Mass spectra of peptides for two proteins.

Top: **high-turnover protein**
Bottom: **low-turnover protein**.

Over time, the heavy to light (H/L) ratios increase.

H-concentration of high-turnover protein saturates.
That of low-turnover protein still increases.

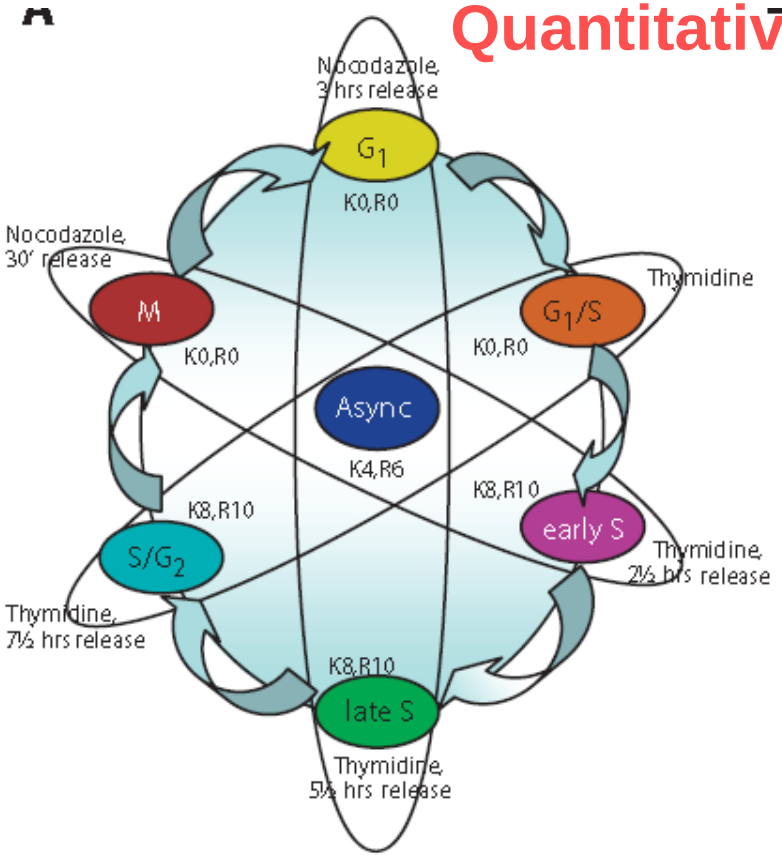


This example illustrates the principles of SILAC and mass spectroscopy signals (peaks).
m/z: mass over charge ratio of a peptide fragment

In the Olson et al. study, the authors used H and L forms to label different stages of the cell cycle.

Schwanhäuser et al. Nature 473, 337 (2011)

Quantitative proteomic analysis



Center: asynchronously growing cell population as internal standard to allow normalization between experiments.

HeLa S3 cells were SILAC-labeled with 3 different isotopic forms (light – medium – heavy) of arginine and lysine.

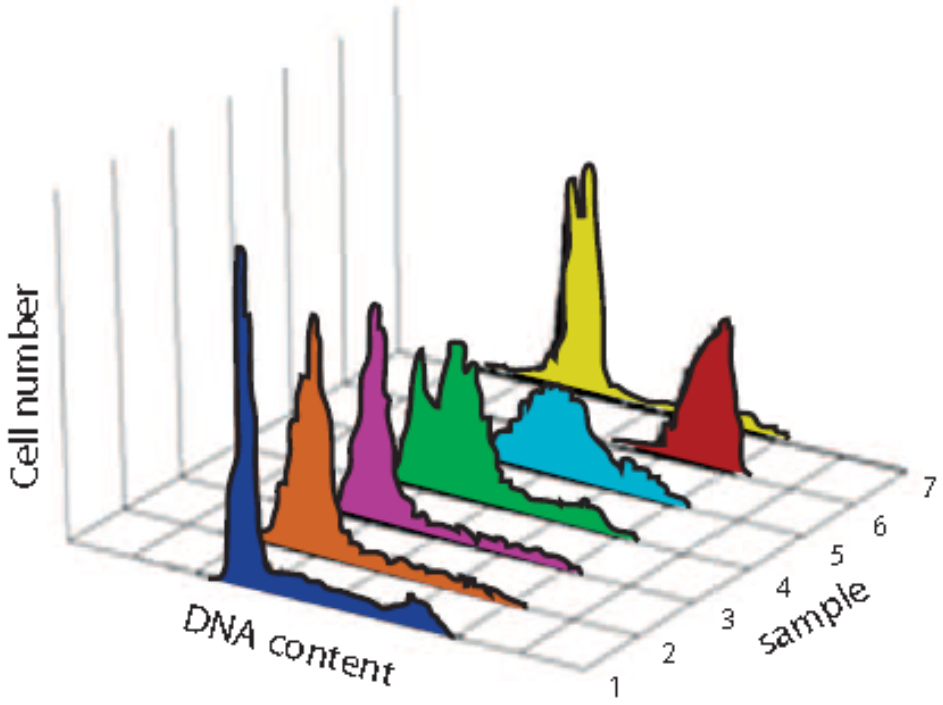
3 individual populations of heavy and light SILAC cells were synchronized with a **thymidine** block (analog of thymine, blocks entry into S phase).

Cells were then collected at 6 different time points across the cell cycle after release from the **thymidine arrest**.

Out of this, 2 samples were collected after a further **cell cycle arrest** with **nocodazole** and release. (Nocodazole interferes with polymerization of microtubules.)

FACS profiles of individual HeLa populations

	% Cells		
	G_1	S	G_2/M
1. Asynchronous	64	27	9
2. Thymidine block	50	46	4
3. Thymidine block + release 2½ h	36	60	4
4. Thymidine block + release 5½ h	23	70	7
5. Thymidine block + release 7½ h	15	70	15
6. Nocodazole block + release ½ h	1	11	88
7. Nocodazole block + release 3 h	82	12	6



Cells were fixed and collected by centrifugation.

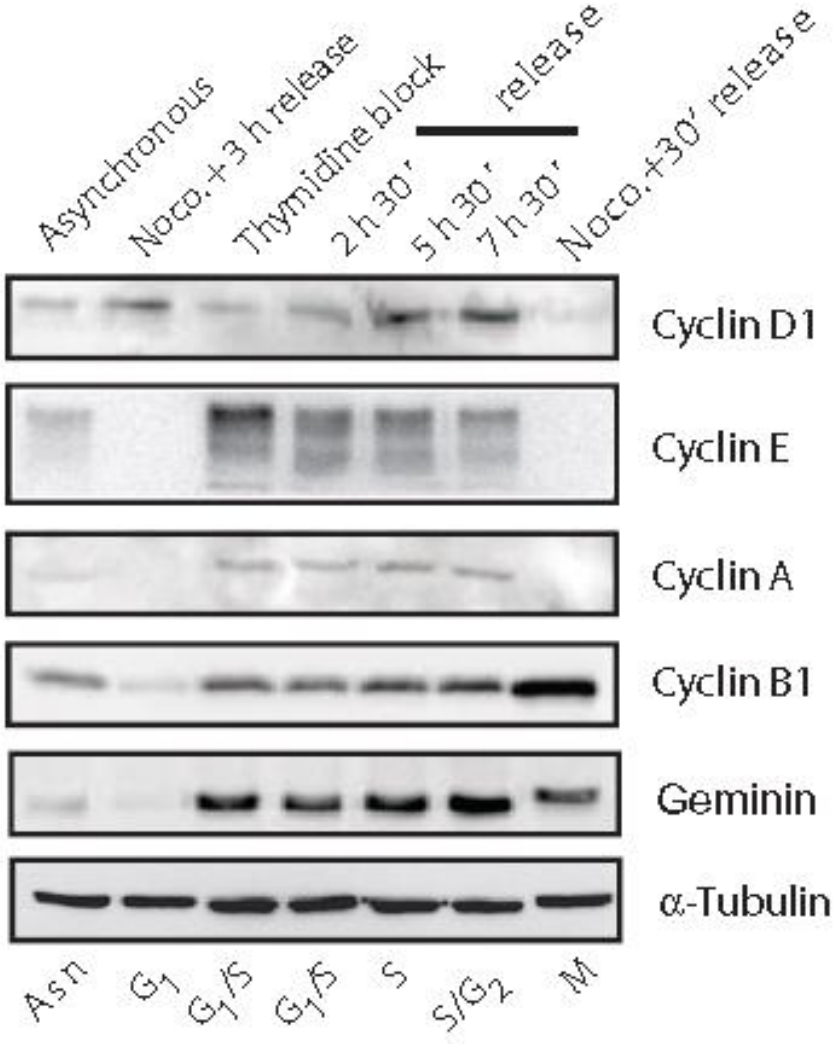
Then the **DNA content** of the cells was determined with propidium iodide.

The DNA content is the basis for classifying the **state** along the cell cycle.

→ Samples 1 – 5 are not pure states, but **mixtures**.
Nocodazole block is quite efficient in synchronizing cells (samples 6 and 7).

Olsen Science Signaling 3 (2010)

Quantification of cell cycle markers



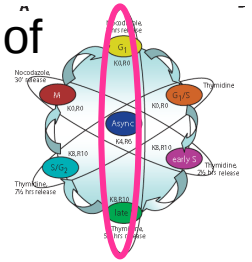
Immunoblot analysis of known cell cycle marker proteins in the different cell populations (α-tubulin is a control).

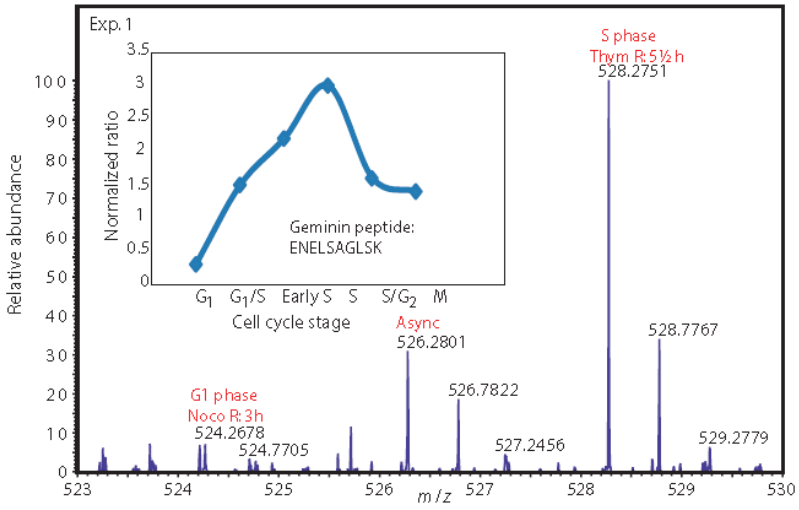
The abundance of 20% of the proteome changed by at least 4-fold throughout the cell cycle (difference between lowest and highest abundance).

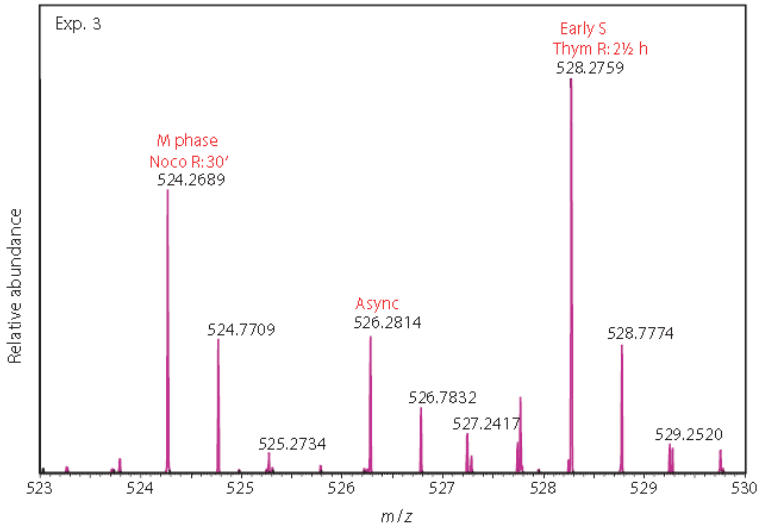
Because a **fourfold change** best accounted for the dynamics of already described cell cycle components, this ratio was used as a threshold for subsequent analysis.

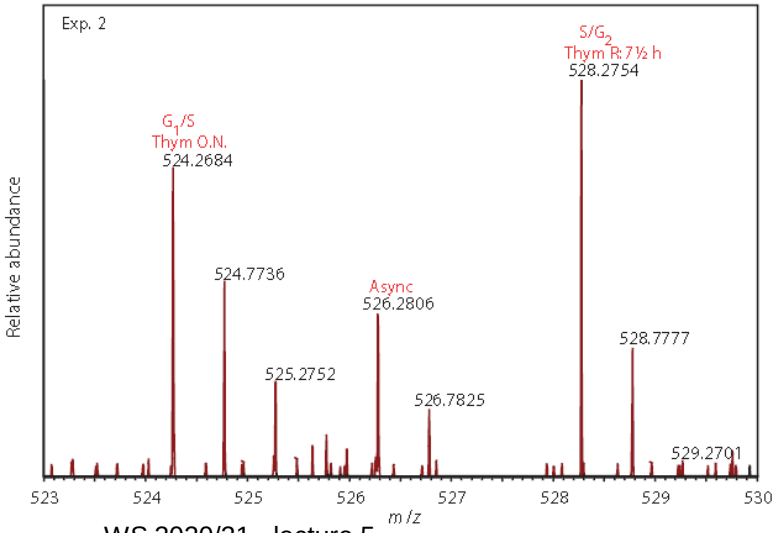
Monitor protein abundance by MS

Experiment 1: mixture of
L = G1 phase
M = Async
H = S phase









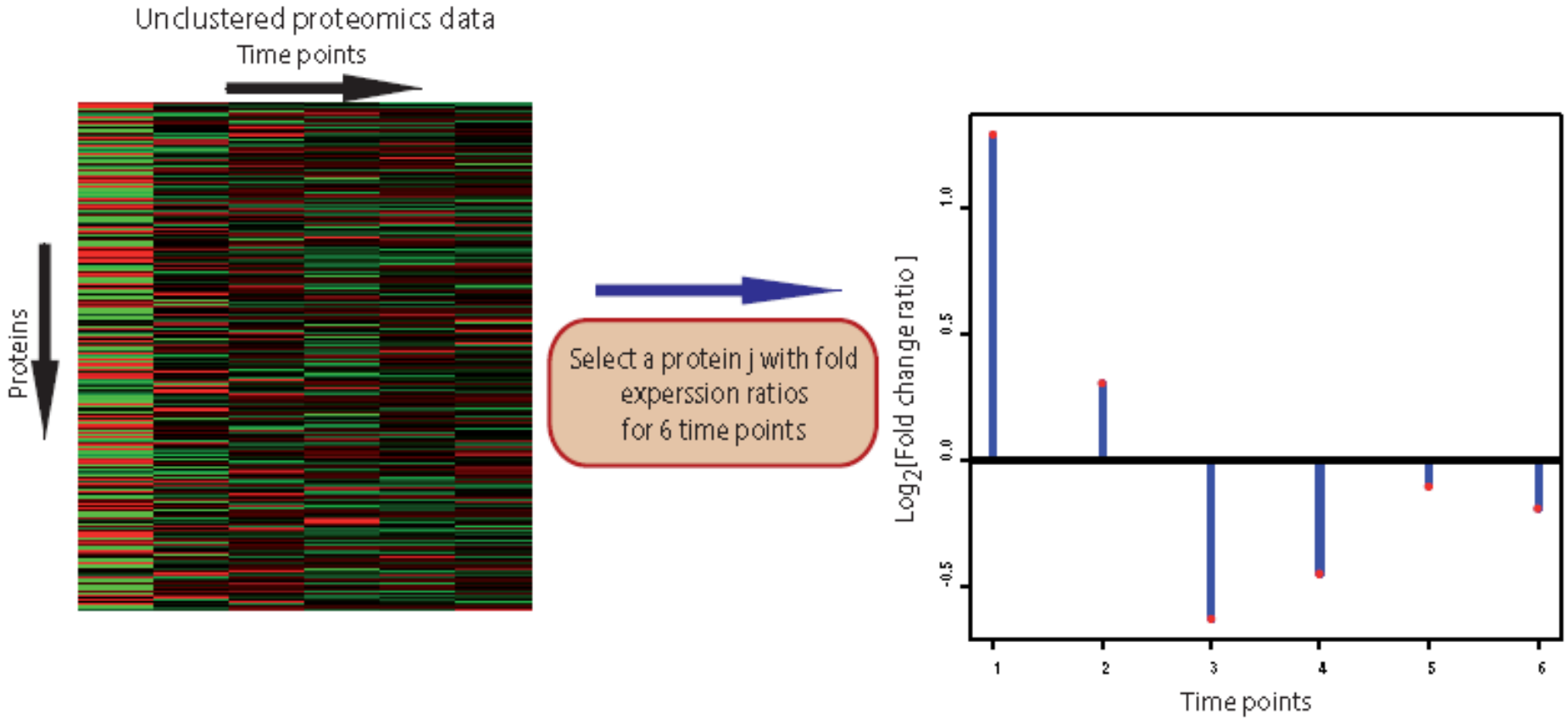
Representative MS data showing how the abundance of the proteins was monitored in 3 experiments to obtain information from the 6 stages of the cell cycle.

The data show the MS analysis of a tryptic SILAC peptide triplet derived from the cell cycle marker protein **Geminin**.

Relative peptide abundance changes were normalized to the medium SILAC peptide derived from the asynchronously grown cells in all three experiments.
The inset of Exp. 1 shows the combined six-time profile of Geminin over the cell cycle.

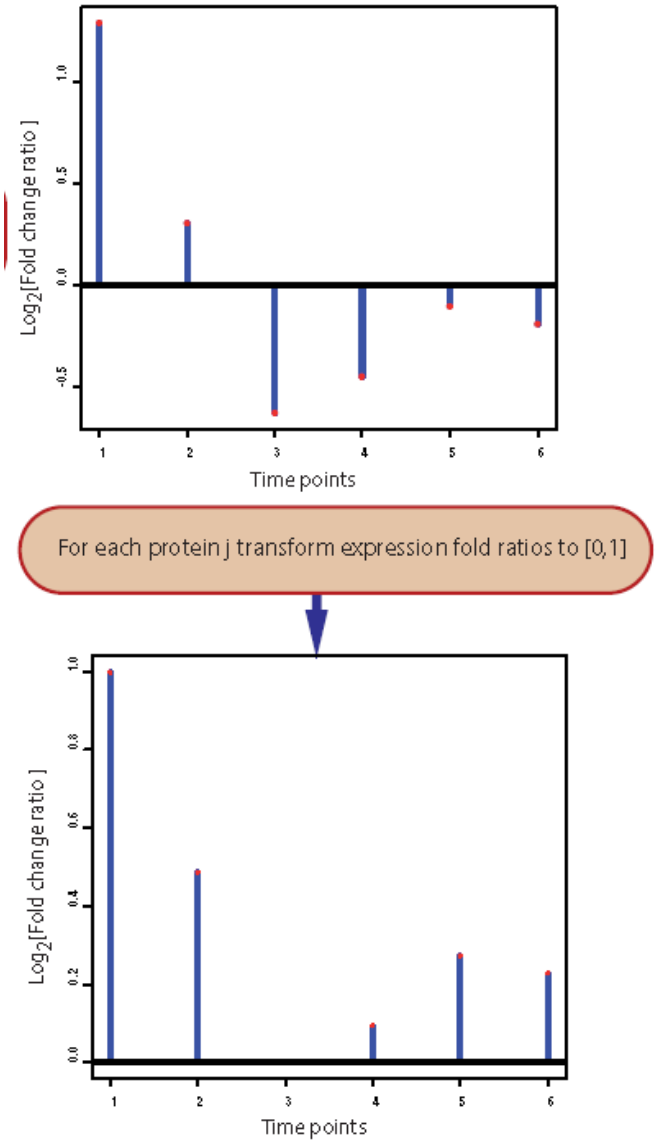
Olsen Science Signaling 3 (2010)

Bioinformatics Workflow (1)



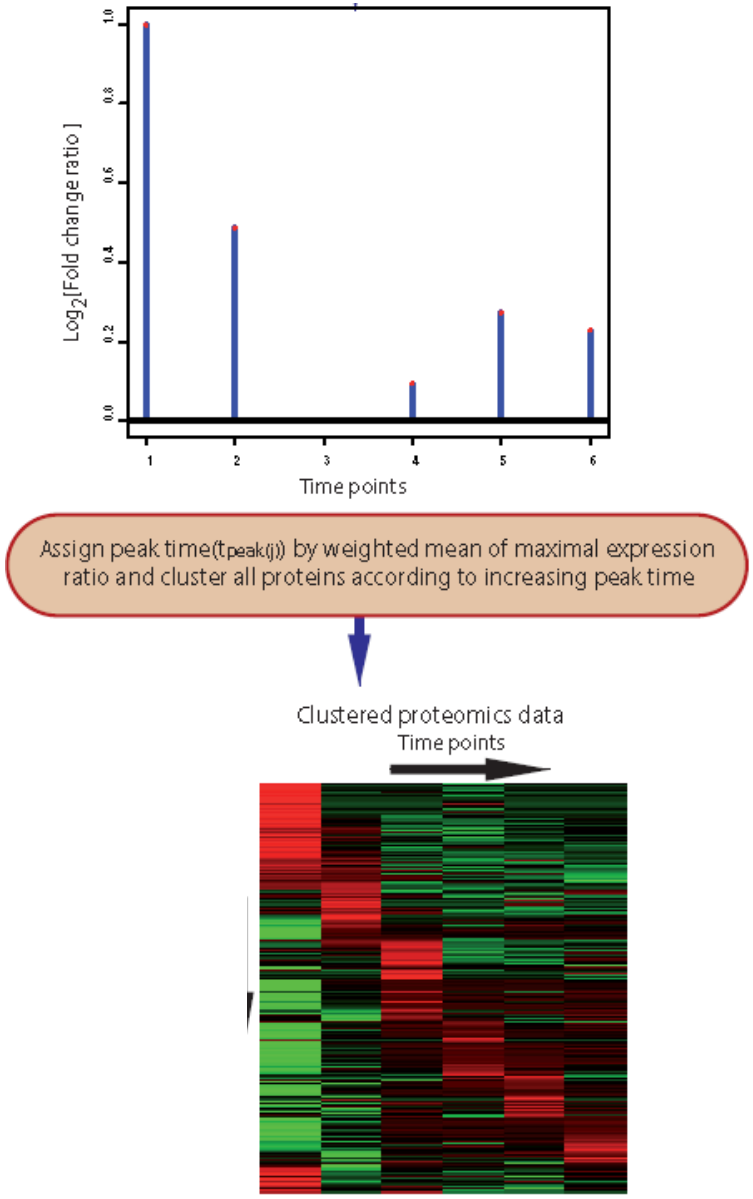
Olsen Science Signaling 3 (2010)

Bioinformatics Workflow (2)



Olsen Science Signaling 3 (2010)

Bioinformatics Workflow (3)

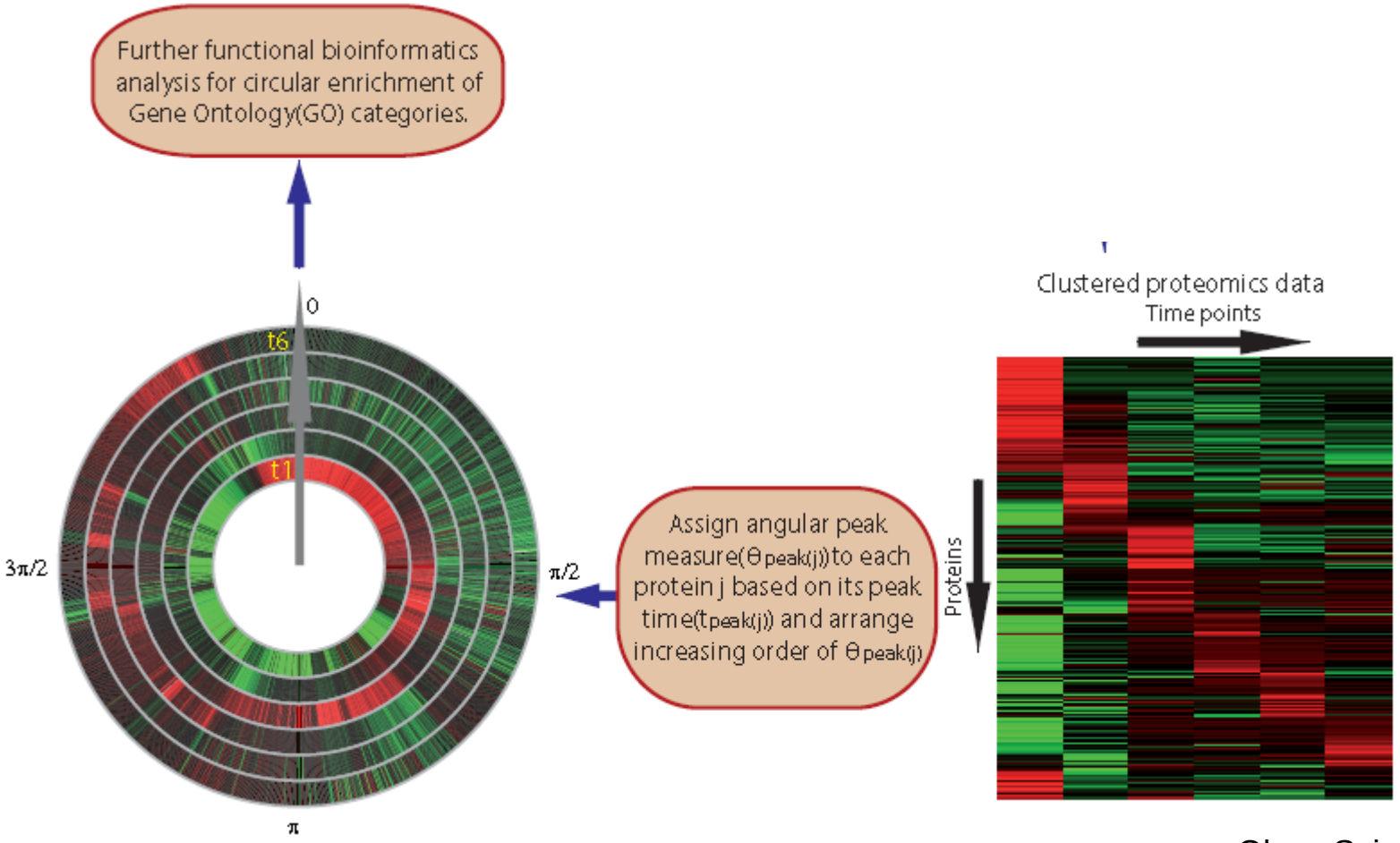


For each protein a peak time index was calculated by weighted mean of its maximal expression at time point t_i w.r.t its adjacent time points t_{i-1} and t_{i+1}.

The proteins were then clustered according to their increasing peak time indices.

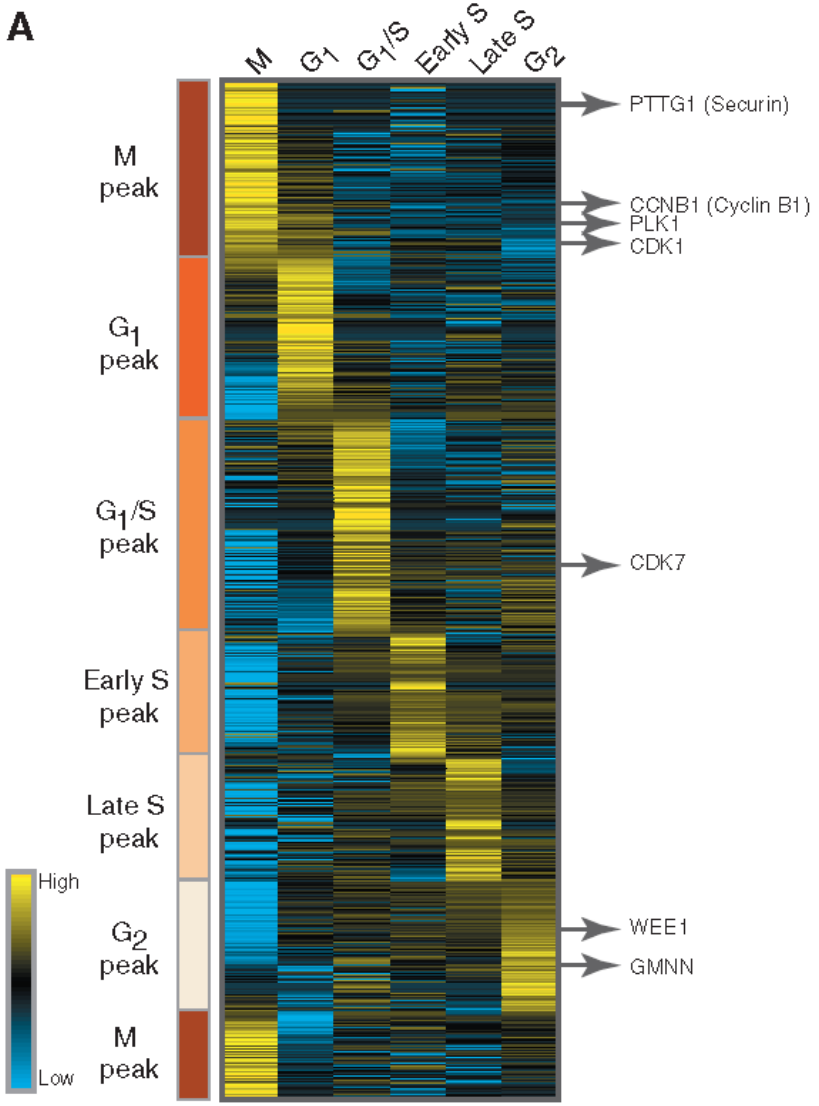
Olsen Science Signaling 3 (2010)

Bioinformatics Workflow (4)



Olsen Science Signaling 3 (2010)

Dynamics of the proteome during the cell cycle

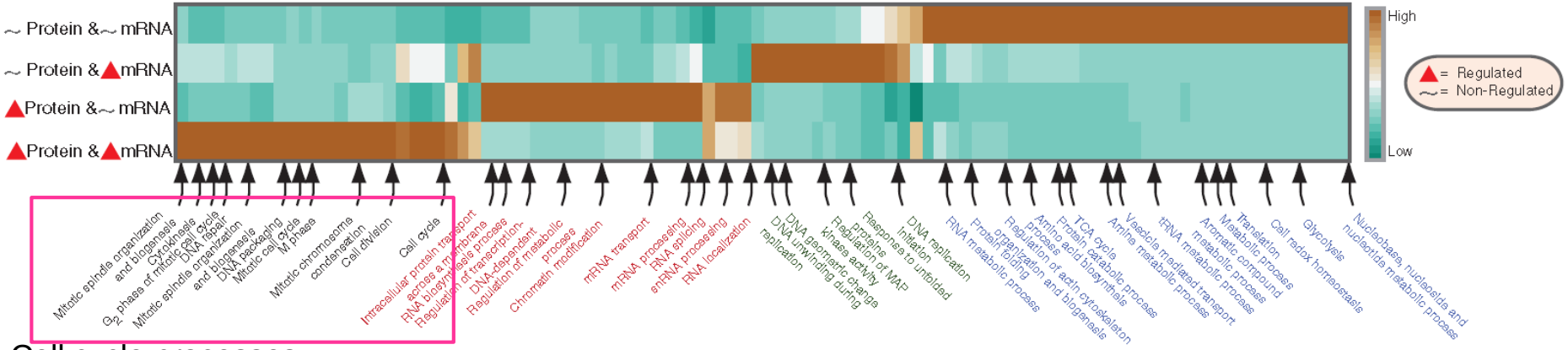


Proteins whose abundance changed at least fourfold during the cell cycle were clustered in all cell cycle stages by calculating a time peak index by weighted mean of the ratio of maximal abundance.

For each cell cycle stage, there are clear patterns of up- and down-regulation.

Olsen Science Signaling 3 (2010)

Comparison of mRNA and protein dynamics



Cell cycle processes

Comparison of mRNA and protein dynamics during the cell cycle.
Measured protein dynamics were correlated to published mRNA data.

Proteins were grouped on the y axis in 4 categories from top to bottom:

- unchanging mRNA and protein
- changing mRNA and unchanging protein
- unchanging mRNA and changing protein
- and changing mRNA and changing protein.

The x axis shows clustered gene ontology (GO) biological process terms enriched in at least one of the above 4 categories.

High and **low** represent statistical over- or underrepresentation, respectively.

Absolute phosphorylation site stochiometry

Now we want to derive the phosphorylation state of **individual protein residues** during the cell cycle. We need to substract out the changes of protein abundance.
-> we want to know (1) and (2) below

(1) Proportion of phosphorylated to unphosphorylated peptide in Light SILAC state: $\frac{N_L^{PHOS}}{N_L^{NonP}} = a$

N_L^{PHOS} is the total copy number of a given phosphopeptide in the light SILAC state, and N_L^{NonP} is the total copy number the corresponding unphosphorylated peptide in the light SILAC state

(2) Proportion of phosphorylated to unphosphorylated peptide in Heavy SILAC state: $\frac{N_H^{PHOS}}{N_H^{NonP}} = b$

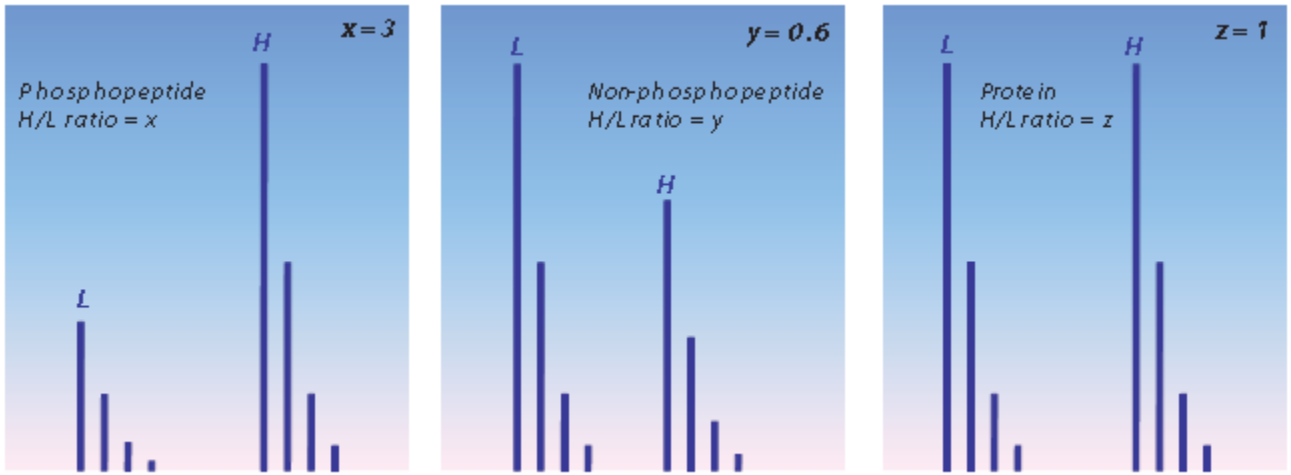
N_H^{PHOS} is the total copy number of a given phosphopeptide in the heavy SILAC state, and
N_H^{NonP} is the total copy number the corresponding unphosphorylated peptide in the heavy SILAC state

(3) We expect that $\frac{N_H^{PHOS} + N_H^{NonP}}{N_H^{PROTEIN}} = \frac{N_L^{PHOS} + N_L^{NonP}}{N_L^{PROTEIN}}$

$N_L^{PROTEIN}$ is the total copy number of the phosphoprotein in the light SILAC state, and
$N_H^{PROTEIN}$ is the total copy number the phosphoprotein in the heavy SILAC state

Olsen Science Signaling 3 (2010)

Available experimental data



To determine phosphorylation sites that show dynamic profiles due to changes in phosphorylation state rather than due to changes in protein abundance, we consider the measured phosphopeptide H/L ratios.

From the experiment we have:

- the SILAC ratio *x* for phosphopeptide
- the SILAC ratio *y* for non-phosphopeptide (the unphosphorylated version of the phosphopeptide),
- and protein ratio *z* (the total amount of the protein, i.e. the sum of its phosphorylated and nonphosphorylated forms).

Absolute phosphorylation site stochiometry

From the MS data we know:

(4) Relative phosphopeptide ratio = $\frac{N_H^{PHOS}}{N_L^{PHOS}} = x$

(5) Relative unphosphorylated peptide ratio = $\frac{N_H^{NonP}}{N_L^{NonP}} = y$

(6) Relative total phosphoprotein ratio = $\frac{N_H^{PROTEIN}}{N_L^{PROTEIN}} = z$

If we know x, y and z then we can solve equations 1 and 2 by substituting in equations 3:

(1) Occupancy rate in Light SILAC state: $\frac{N_L^{PHOS}}{N_L^{NonP}} = a = \frac{z-y}{x-z}$

(2) Occupancy rate in Heavy SILAC state: $\frac{N_H^{PHOS}}{N_H^{NonP}} = b = \frac{x \cdot (z-y)}{y \cdot (x-z)}$

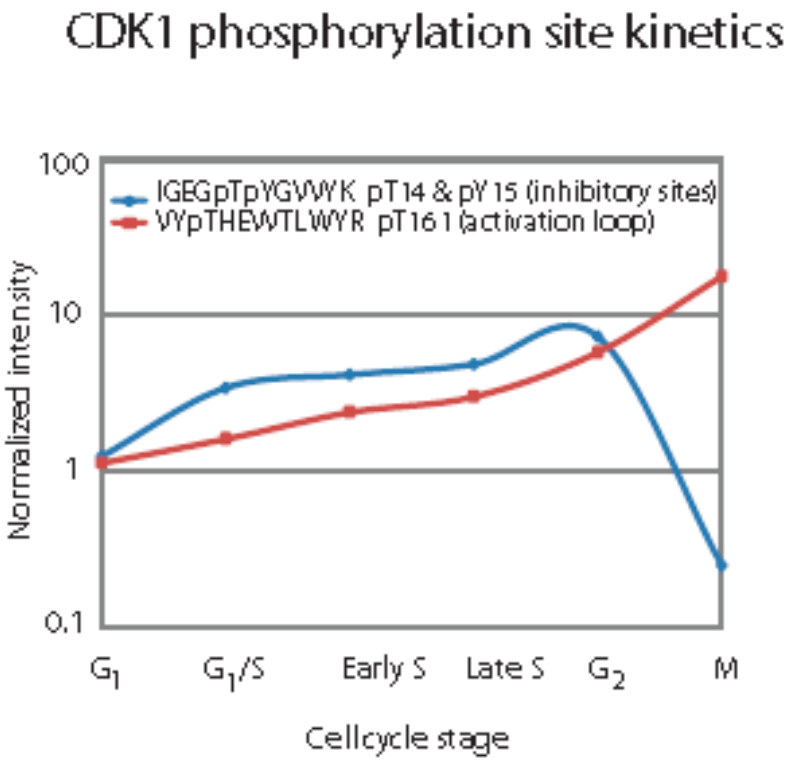
We expect that $N_L^{PHOS} + N_L^{NonP} = N_H^{PHOS} + N_H^{NonP} = 100\% = 1$

and can therefore calculate the phosphorylation site occupancy in the Light and Heavy SILAC state as:

(3) Light SILAC occupancy: a/(1+a) and Heavy SILAC occupancy: b/(1+b)

Olsen Science Signaling 3 (2010)

Example: Dynamic phosphorylation of CDK1

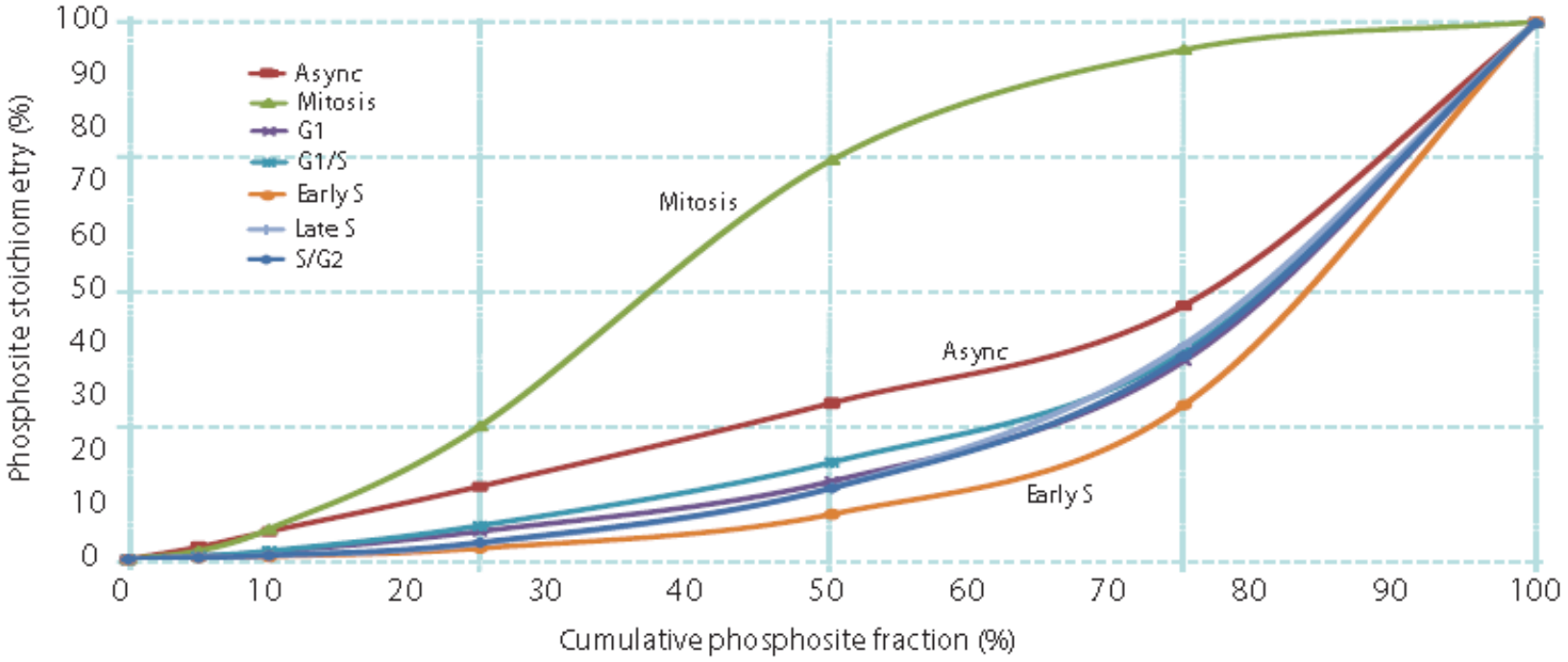


Dynamic profile of two CDK1 phosphopeptides during the cell cycle.

The activating site Thr161 (red) peaks in mitosis, whereas phosphorylation of the inhibitory sites Thr14 and Tyr15 (blue) is decreased in mitosis

Olsen Science Signaling 3 (2010)

Total phosphosite occupancy in different stages of cell cycle

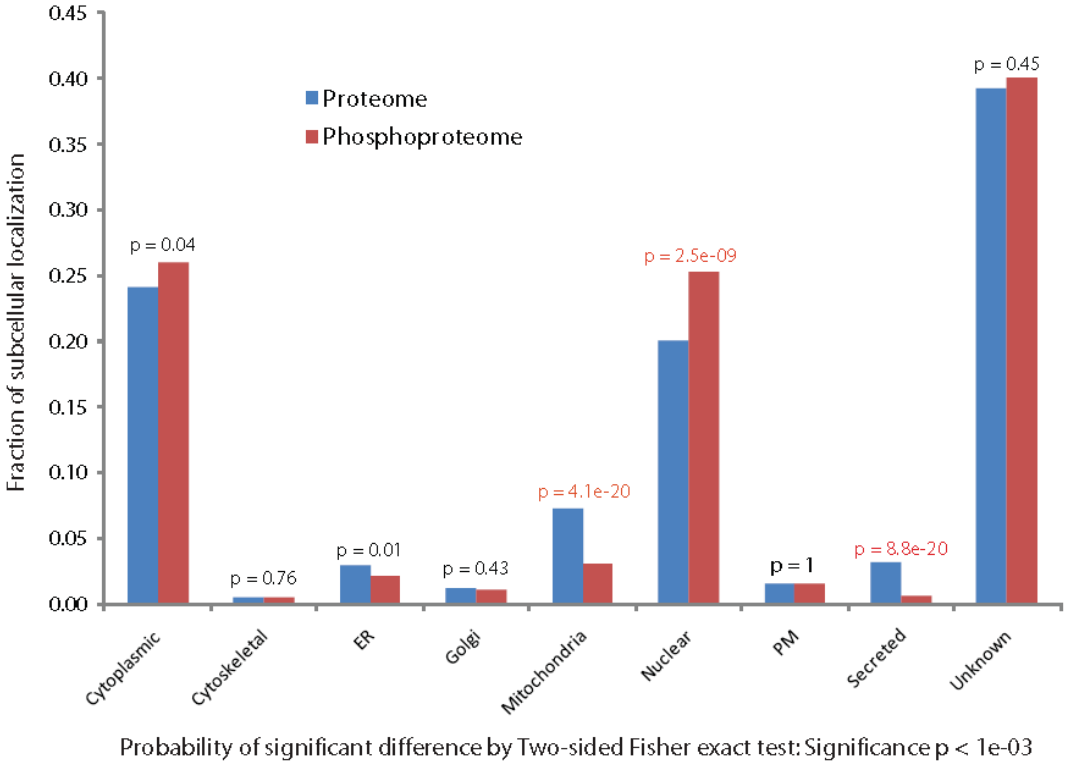


Fifty percent of all mitotic phosphorylation sites have occupancy of 75% or more.

Olsen Science Signaling 3 (2010)

Differential phosphorylation

Gene ontology (GO) analysis of protein and phosphoproteins subcellular localization. All proteins identified by MS were clustered according to their GO annotation for sub-cellular localization (Blue bars). The same clustering was done for all phosphoproteins (Red bars).

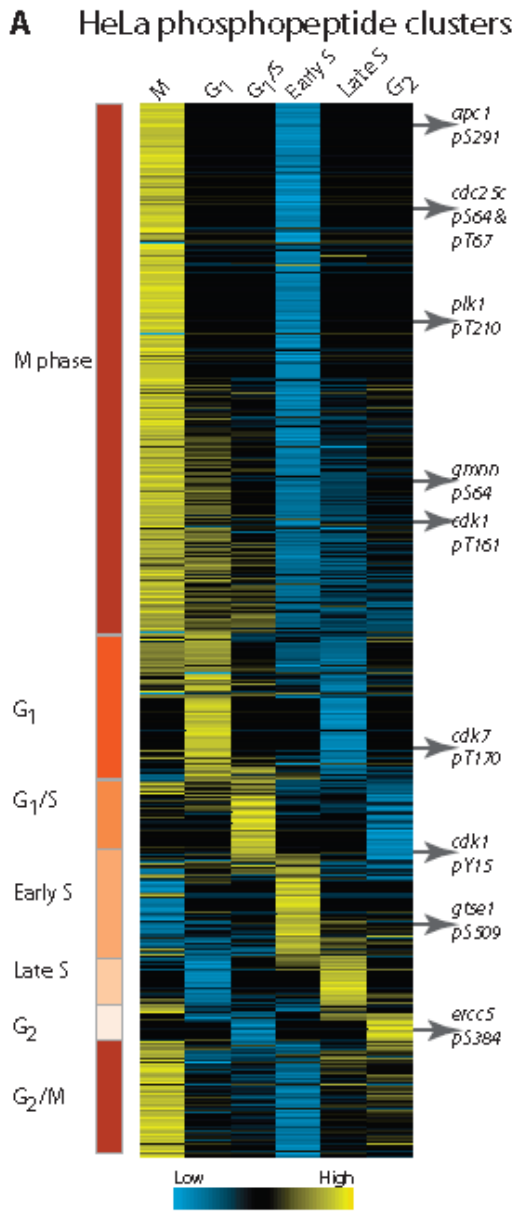


y-axis : percentage of the indicated sub-cellular fractions from the total.

Compared to the proteome distribution, phosphorylated proteins are over-represented in the nucleus and under-represented amongst mitochondrial and secreted proteins.

Olsen Science Signaling 3 (2010)

Dynamics of the Phosphoproteome



Dynamics of the phosphoproteome during the cell cycle.

Clustering of regulated phosphorylation sites in all cell cycle stages.

More than half of all identified regulated phosphorylation sites peak in mitosis.

Olsen Science Signaling 3 (2010)

Who phosphorylates? -> NetPhorest algorithm

COMPUTATIONAL BIOLOGY

Linear Motif Atlas for Phosphorylation-Dependent Signaling

Martin Lee Miller,[1,2*] Lars Juhl Jensen,[2,3*] Francesca Diella,[3] Claus Jørgensen,[4] Michele Tinti,[5] Lei Li,[6] Marilyn Hsiung,[4] Sirlester A. Parker,[7] Jennifer Bordeaux,[7] Thomas Sicheritz-Ponten,[1] Marina Olhovsky,[4] Adrian Pasculescu,[4] Jes Alexander,[8] Stefan Knapp,[9] Nikolaj Blom,[1] Peer Bork,[2,10] Shawn Li,[6] Gianni Cesareni,[5] Tony Pawson,[4] Benjamin E. Turk,[7] Michael B. Yaffe,[8†] Søren Brunak,[1,2†] Rune Linding[4,8,11†]

(Published 2 September 2008)

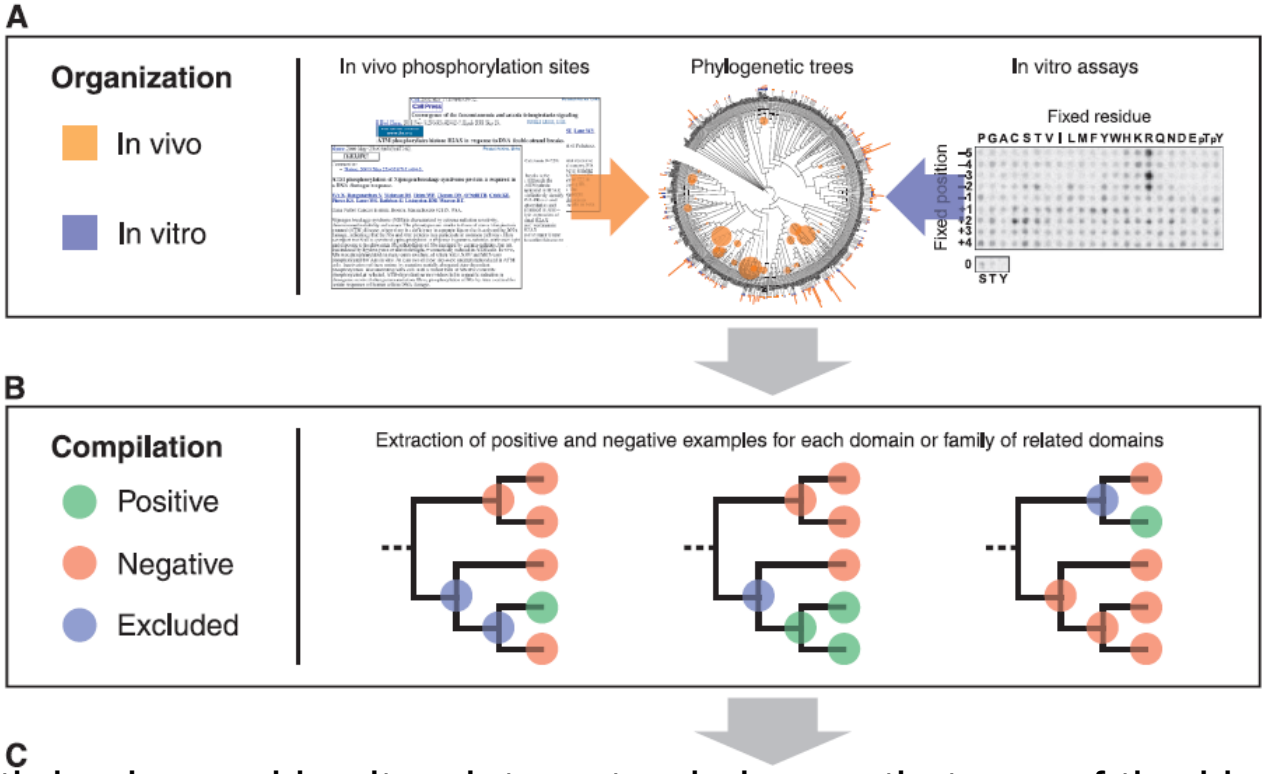
Systematic and quantitative analysis of protein phosphorylation is revealing dynamic regulatory networks underlying cellular responses to environmental cues. However, matching these sites to the kinases that phosphorylate them and the phosphorylation-dependent binding domains that may subsequently bind to them remains a challenge. NetPhorest is an atlas of consensus sequence motifs that covers 179 kinases and 104 phosphorylation-dependent binding domains [Src homology 2 (SH2), phosphotyrosine binding (PTB), BRCA1 C-terminal (BRCT), WW, and 14–3–3]. The atlas reveals new aspects of signaling systems, including the observation that tyrosine kinases mutated in cancer have lower specificity than their non-oncogenic relatives. The resource is maintained by an automated pipeline, which uses phylogenetic trees to structure the currently available in vivo and in vitro data to derive probabilistic sequence models of linear motifs. The atlas is available as a community resource (http://netphorest.info).

Miller Science Signaling 1 (2008)

NetPhorest algorithm

Analyze *in vivo* protein-phosphorylation sites that are linked to at least one kinase [Phospho.ELM] or phospho-binding domain [DOMINO].

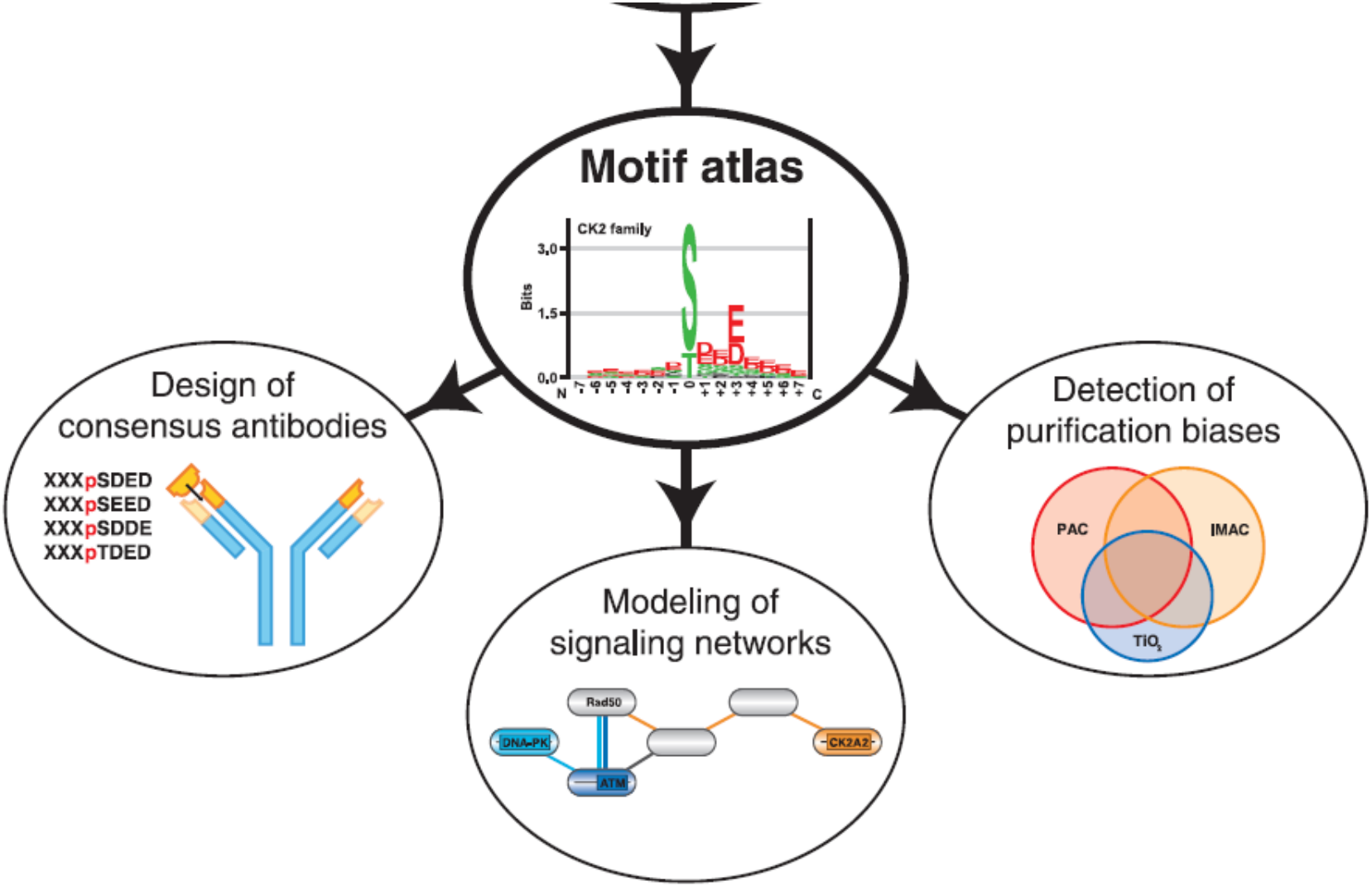
Analyze *in vitro* assays that interrogate kinase specificity by degenerate peptide libraries



map both in vivo and in vitro data onto phylogenetic trees of the kinase and phospho-binding domains, which capture how similar the domains are to one another and thereby how likely they are to have similar substrate specificities.

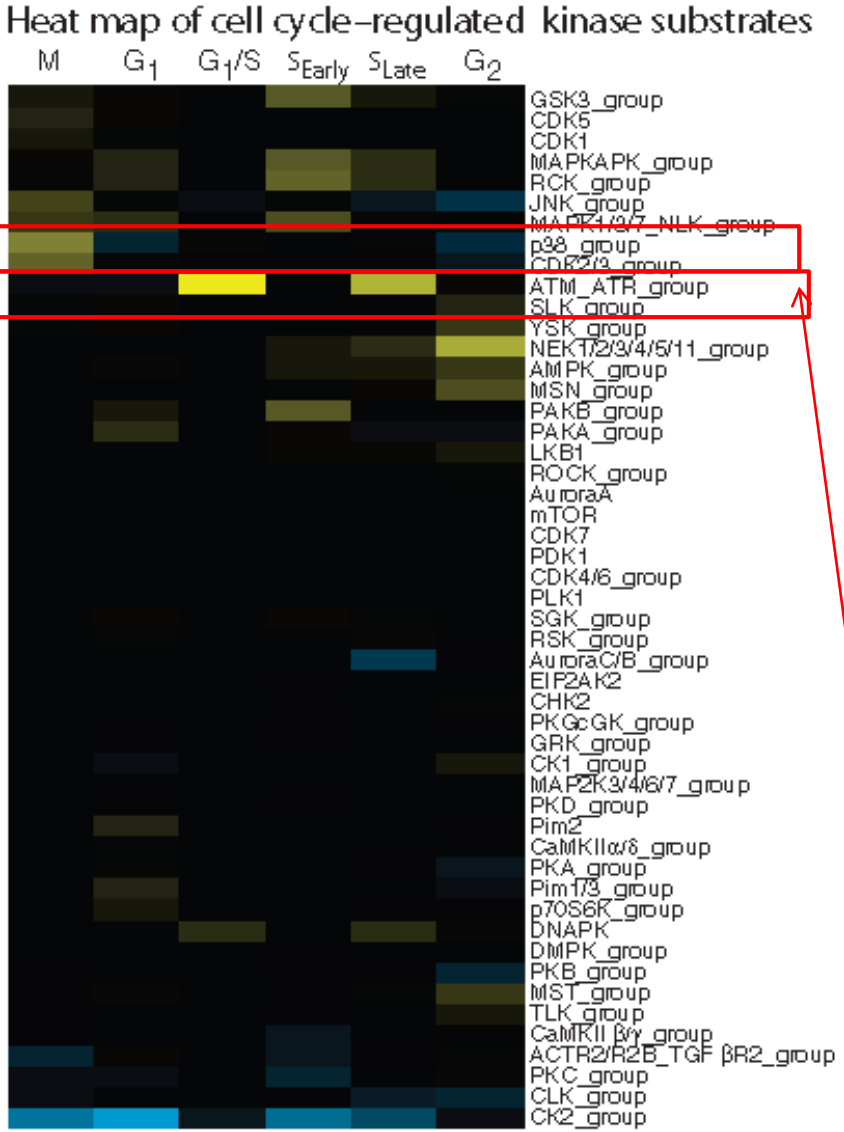
Miller Science Signaling 1 (2008)

NetPhorest algorithm: applications



Miller Science Signaling 1 (2008)

Cell Cycle-regulated kinase substrates



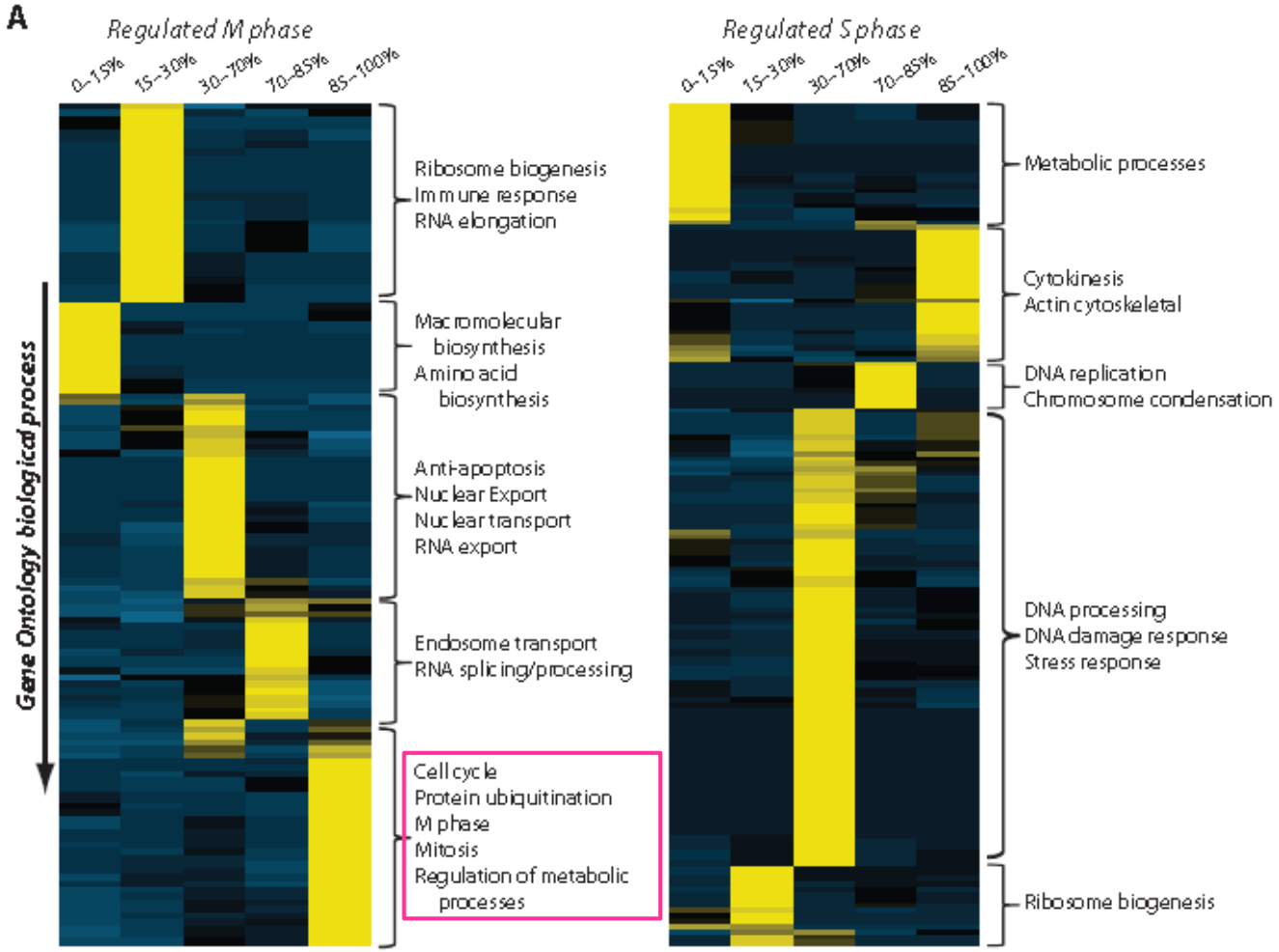
The NetPhorest algorithm was used to predict kinase-substrate relationships of all serine and threonine phosphorylated proteins.

The heat map shows over- (yellow) and underrepresentation (blue) of predicted kinase substrates during different stages of the cell cycle compared to a background of phosphorylation sites that did not change with the cell cycle.

Predicted CDK2 and CDK3 substrates were most highly phosphorylated in M phase.

ATM_ATR substrates are high in S phase.

Proteomic phenotyping of phosphorylation site stochiometry



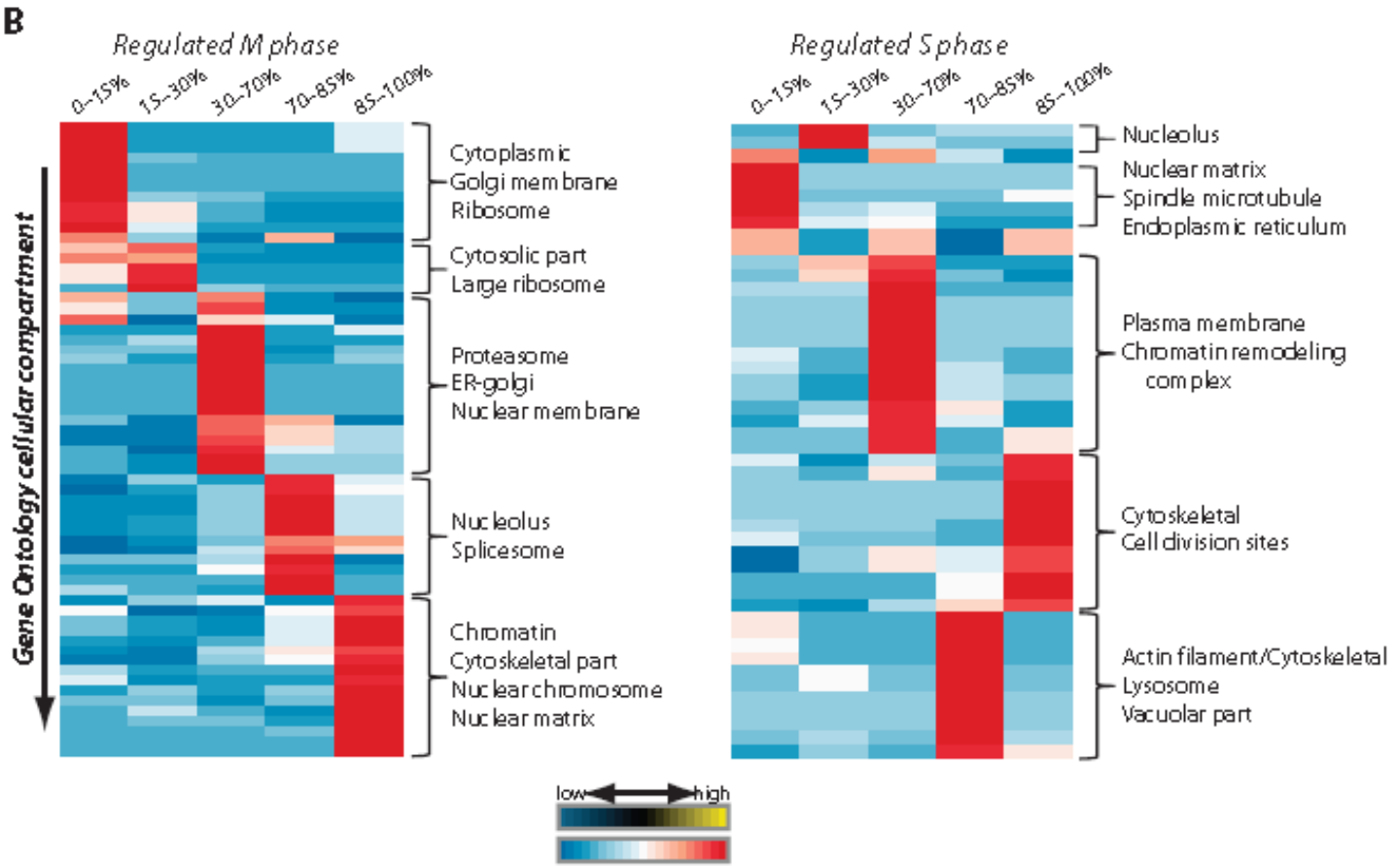
Phenotypic phosphoproteome comparison organized by GO biological process for mitotic (left) and S phase (right) cells.

Proteins involved in metabolic processes have high-occupancy phosphorylation sites during mitosis, but low-occupancy sites during S phase

color scale: yellow, high overrepresentation; dark blue, high underrepresentation.
The phospho proteins were divided into five quantiles on the basis of their maximum phosphorylation-site occupancy and analyzed for GO category (biological process and cellular compartment) enrichment by hypergeometric testing.

Olsen Science Signaling 3 (2010)

GO cellular compartment analysis



Proteomic phenotype analysis of GO cellular compartment level.

Olsen Science Signaling 3 (2010)

Summary

Phosphorylation of protein residues is an important mechanism to regulate protein structure, protein activity, protein localization, and protein interactions.

About 70% of all cellular proteins are phosphorylated to some extent.

Phosphorylation is a dynamic state variable during the cell cycle.

Phosphorylation levels are controlled by the ca. 518 different human kinases as well as by phosphatases.

-> these are important potential drug targets (problem is achieving specificity)

Paper 5 / Assignment 3

Mihkel Örd, Kaidi Möll, Alissa Agerova, Rait Kivi, Ilona Faustova, Rainis Venta, Ervin Valk & Mart Loog, Nature Structural & Biology Biology 26, 649–658 (2019)
Multisite phosphorylation code of CDK
https://www.nature.com/articles/s41594-019-0256-4