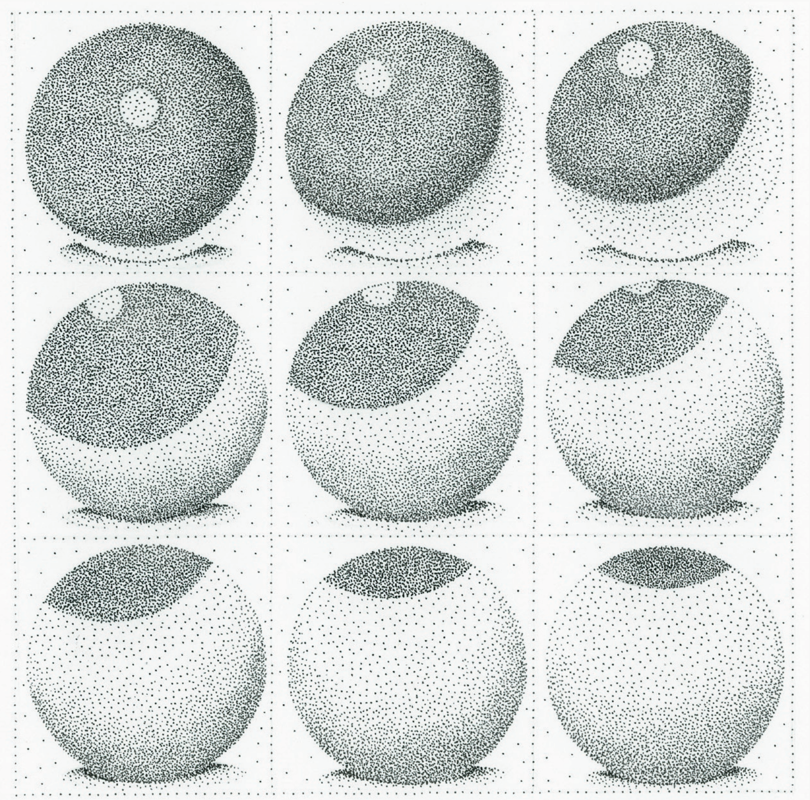
Exploring protein phosphorylation dynamics by mass spectrometry:

pushing the boundaries of time resolved phosphoproteomics



Juan Manuel Valverde Barrantes

Exploring protein phosphorylation dynamics by mass spectrometry: pushing the boundaries of time resolved phosphoproteomics

Juan Manuel Valverde Barrantes

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group, Utrecht University, The Netherlands.

ISBN: 978-94-6423-930-0

Cover artwork by: Marlen Espeleta

Printed by: Proefschriftmaken | www.proefschriftmaken.nl

Exploring protein phosphorylation dynamics by mass spectrometry: pushing the boundaries of time resolved phosphoproteomics

De plasticiteit van eiwit fosforylering ontrafeld: innovaties in tijd gereguleerde fosfoproteomics

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 13 september 2022 des middags te 2.15 uur

door

Juan Manuel Valverde Barrantes

geboren op 23 november 1988
te San José, Costa Rica

Promotors:

Prof. dr. A.F.M. Altelaar
Prof. dr. A.J.R. Heck

On the cover:

Xenopus laevis eggs illustrated by stippling technique. In the latter, layers of small dots are used to create patterns with varying degrees of shading. The *X. laevis* unfertilized egg shows a darkly pigmented animal hemisphere and a lightly pigmented vegetal hemisphere. After fertilization, an 'activation wave' caused by an influx of Ca^{2+}induces cytoplasmic and cortical movements that reorient the vegetal hemisphere downwards and the animal hemisphere upwards. This is accompanied by a contraction of the pigment in the animal hemisphere. The sequence in the cover starts with an unfertilized egg, and finishes with a reoriented fertilized egg, just before the first embryonic cell division. *X. laevis* is a powerful animal model used in a variety of fields, from cell cycle research to developmental biology, and it was used for the experiments described in chapter IV of this thesis.

Table of contents

Chapter I **9**

Introduction to MS based proteomics and its application to the study of protein phosphorylation

Chapter II **51**

Capturing the signalling dynamics of the MAPK-AKT-mTOR pathway in a single targeted phosphoproteomics assay

Chapter III **73**

Phosphoproteome profiling reveals differences in ERK1 and ERK2 dependent phosphorylation

Chapter IV **113**

A CDK-mediated phosphorylation switch of disordered protein condensation

Chapter V **165**

Combined quantitative (phospho)proteomics and mass spectrometry imaging reveal temporal and spatial protein changes in human intestinal ischemia-reperfusion

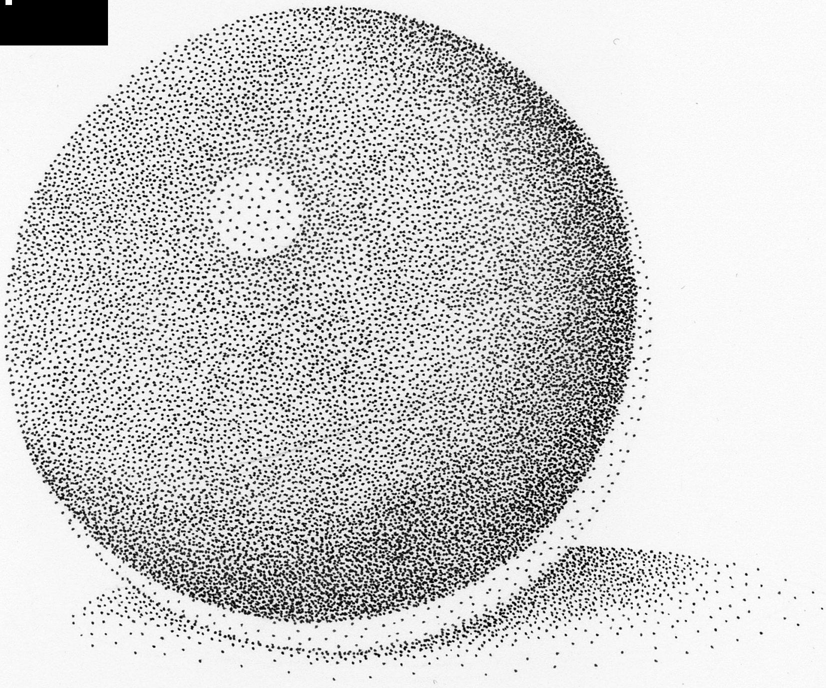
Chapter VI **213**

Conclusions and outlook

Chapter VII **227**

Summary, Nederlandse samenvatting, About the author, List of publications, Acknowledgements

CHAPTER I



Introduction to MS based proteomics and its application to the study of protein phosphorylation

Mass spectrometry based proteomics

At the epilogue of the 20th century, the study of proteins changed drastically due to rise of mass spectrometry-based technologies. This started thanks to the development of the so-called 'soft' ionization techniques, which permitted the study of analytes which before were off limits due to their labile nature when subjected to other ionization strategies. This revolution allowed the application of MS to the study of biomolecules such as proteins, lipids and carbohydrates. Eventually, these techniques would evolve rapidly, making possible the study of proteomes, and hence giving rise to the now established field of proteomics.

A brief look into the history of mass spectrometry

At the beginning of the 20th century, physicist J.J. Thomson and colleagues developed an instrument capable of determining the mass of the particles present in cathode rays, hence discovering the first subatomic particle, the electron (*1*). Thomson was awarded the Nobel Prize in Physics in 1906 due this discovery, and without knowing, he also gave birth to a new field of experimental physics: mass spectrometry (MS) (*2*). Together with Francis Aston, Thomson continued using this technology to measure the mass to charge (m/z) ratio of atoms and molecules, by means of electric and magnetic fields that affected the trajectory of these molecules under vacuum. Their efforts permitted the first measurements of the different masses of stable isotopes (*3, 4*). Other scientist followed the ground laid by Thomson and Aston to expand the use MS to further characterize atomic isotopes. A glaring example of this would be the work done by Alfred Neir, who devoted his life to the development of new MS instruments, which he eventually would use to isolate the uranium isotope U^{235} (*5*). His work was part of a joint effort to conclude that U^{235} underwent nuclear fission, findings that would give rise to the nuclear era, largely impacting the history of humankind throughout the 20th century (*6*).

During the decade of the 1940s MS found its way to the chemical lab, where it was mostly used by industrial chemists interested in determining the composition of a variety of mixtures. Especially in the oil industry, MS was used quantitatively to determine the ratios of different small molecules found in crude oil (*5*). This laid the ground for the application of MS to other organic components. Fred McLafferty, Klaus Biemann and Carl Djerassi were pioneers who started to use MS to identify unknown compounds, building a bridge between MS and structural chemistry. These advances opened the door for MS to be widely used for the determination of small organic compounds (*7*).

Despite the major advances made, certain analytes remained elusive for MS. Since mass spectrometers only detect charged molecules, these have to be ionized before entering the instrument (*8*). This was largely achieved by a variety of techniques that required sample volatilization and ionization in the gas phase (*9, 10*). This was not amenable for large biomolecules, like complex carbohydrates, proteins and nucleic acids, which fragmented upon such treatment. It was not until the late 1980s that two 'soft' ionization techniques were introduced to the scientific community. Namely, electrospray ionization (ESI) and soft laser desorption would make macromolecules such as proteins, available to MS analysis. These seminal discoveries, made in part by John B. Fenn and Koichi Tanaka, were awarded with the Nobel Prize in Chemistry 2002 (*11*). In the following section, I will briefly discuss these techniques and the major role they have played during the development of MS based proteomics.

Soft ionization: a stepping stone towards MS based proteomics

Despite the promising potential displayed by MS during the first half of the 20th century, its full potential was still far from being achieved. The so called 'ionization problem' rendered the technique incompatible with the large molecules present in living organisms. In order to address this issue, John B. Fenn and others decided to dissolve the analyte of interest in a volatile solvent. The application of a voltage to such solution would induce the formation of highly charged droplets that upon further evaporation of the solvent should render intact ions of the analyte in the gas phase (Figure 1A). In the groundbreaking work published by Fenn and colleagues at the end of the 1980s, they showed how organic labile compounds, like peptides and even proteins could be electrosprayed and detected by MS (*12*).

Using a different approach, Tanaka and collaborators set out to induce analyte ionization by means of soft laser desorption (*13*). In his seminal discovery, Tanaka showed that a low energy nitrogen laser could ionize large biomolecules without fragmentation. This prompted the development of other laser desorption techniques that would become widespread. The most popular one, matrix assisted laser desorption (MALDI), was developed by Michael Karas and Franz Hillenkamp shortly after Tanaka's discovery (*14*). Here, the sample is placed in a nitric acid matrix, which is then bombarded with a laser of a specific wavelength that matches the absorption maximum of the matrix. This way the energy is absorbed by the matrix and not directly by the analyte (Figure 1B). Energy transfer to the analyte via the matrix eventually leads to volatilization and prevents fragmentation. MALDI is the most widely used laser desorption technique, and is usually coupled to Time of flight (TOF) instruments (described in upcoming sections). Even though this setup became common and is considered to be a corner stone of proteomics, ESI became the more popular ionization technique among the scientific community, very likely due to is compatibility with liquid separation techniques, which enables the analysis of complex mixtures with high sensitivity and relatively high throughput.

Just as described by Fenn in his seminal publication in 1989, the eventual marriage of ESI with high resolution MS would change the world of protein analysis (*12*). In the upcoming paragraphs

I will describe the basic workflow and techniques used during a MS based proteomics experiment, from sample preparation to data analysis.

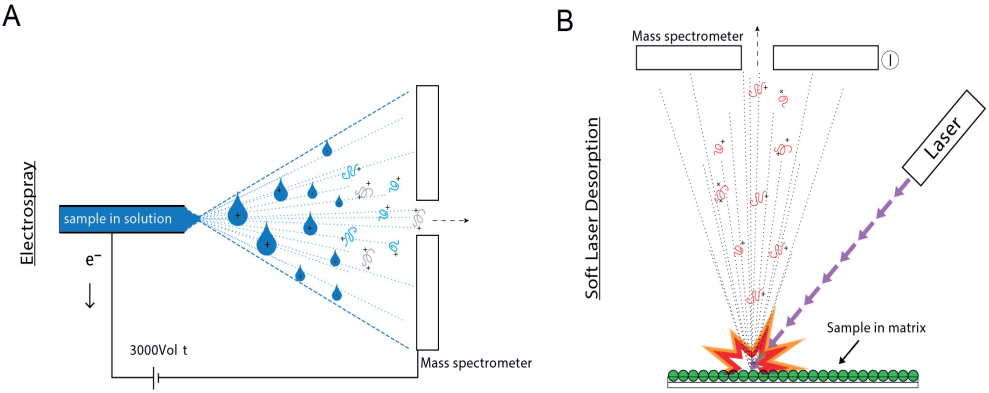


Figure 1. Soft ionization techniques used in proteomics. (A) Electrospray ionization (ESI). (B) Soft laser desorption. Modified from *Advanced information on the Nobel Prize in Chemistry, 2002.*

The proteomics workflow

Sample preparation

Every proteomics experiment starts with a biological sample, which can range from cultured cells or tissue biopsies to bio-fluids like plasma or saliva. Due to the scope of this thesis, I will focus on the sample preparation techniques used for cells and tissues (Figure 2). First, cells must be lysed to make proteins available. Most of the time a combination of mechanical and chemical lysis is used, which consists of maceration or sonication, followed by treatment with detergents and/or chaotropic agents (*15, 16*). The latter allow the disruption of lipid membranes and aid in the solubilization of proteins. Next, soluble biomolecules can be separated from cell debris by centrifugation, followed by enzymatic digestion of proteins, which renders a complex mixture of peptides. Trypsin is the most widely used enzyme, due to its high specificity and wide availability (*17*). This enzyme cleaves after arginine and lysine residues, and even though most research questions can be approached with tryptic digestion, certain applications benefit from the use of complementary enzymes with different specificities (*18*).

Protein digestion is a key step in the MS-proteomics workflow, since bulk separation of peptides is easier and more robust than separation of proteins (*19*). Furthermore, peptide sequencing permits access to localization of post-translational modifications (PTMs), which is not readily available when studying full intact proteins. This peptide-centric approach is known as bottom-up proteomics and is widely used for the study of full proteomes. Its counterpart, top-down proteomics, focuses on the characterization of intact proteins, and even though advances have been made to study complex mixtures of proteins (*20*), so far the technique is mostly applied to purified proteins and is not as robust and widespread as bottom-up proteomics (*19*). Importantly, top-down proteomics can discriminate different proteoforms, enabling the estimation of a PTM stoichiometry, which is not easily determined by bottom-up approaches (*21*).

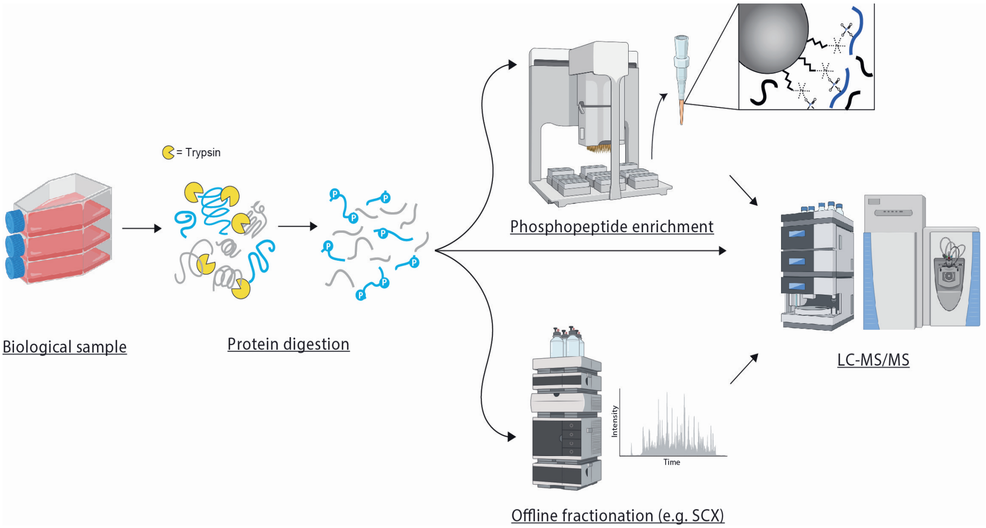


Figure 2. Classic workflow of proteomics experiments. Biological samples are submitted to cell lysis followed by protein digestion. Peptides can be further fractionated by High-pH or SCX off-line fractionation followed by sample measurement with LC-MS/MS. For phosphoproteomics experiments, phosphorylated peptides can be enriched before data acquisition. The figure shows a diagram of Agilent AssayMap Bravo, which was used for phosphopeptide enrichment in the chapters of this thesis. Bravo, LC and MS illustrations were obtained from BioRender.com.

Reducing complexity: from fractionation to enrichment techniques

After protein digestion, peptides can be purified using a variety of techniques, from C18 reversed-phase chromatography to strong cationic exchange (SCX) and hydrophilic interaction liquid chromatography (HILIC) (*22, 23*). The choice of method usually depends on the application and the next steps of the sample preparation workflow. C18 resins are widely used for peptide cleanup and can be found in a variety of different formats (*24, 25*). This step is necessary to get rid of other biomolecules present in the sample (e.g. nucleic acids and lipids) and chemicals introduced during the lysis and digestion (e.g. detergents and chaotropes). The latter might be particularly problematic since many of these molecules are not amicable to liquid chromatography and mass spectrometry analysis.

Once peptides have been cleaned, other steps to reduce sample complexity are usually introduced. Cell digests contain a large variety of peptides spanning a wide dynamic range of concentrations (*26, 27*). This makes the detection of less abundant peptides quite challenging, although it can be partially circumvented by reducing the sample complexity (*28*). Chromatographic methods for sample fractionation present the best current method to separate complex mixtures of peptides based on their physicochemical properties. A popular tactic is to use an orthogonal fractionation approach, which combines chromatographic separation methods that rely on different physicochemical properties, with the attempt to reduce complexity as much as possible (*29, 30*). A widely used 'off-line' separation technique is high-pH reversed phase fractionation: here peptides are dissolved at a high pH (≈10) and are separated based on hydrophobicity on a C18 column (*31, 32*). This method is highly orthogonal to the online low-pH reversed phase separation commonly used just before peptide measurement in the mass spectrometer.

Depending on the research question, measuring the full proteome might not be necessary, and other purification techniques might be required. For instance, antibody based purifications are widely used to study protein interactions (*33*). Here, an antibody directed against a specific protein is used to 'fish out' the target and its interactors. This approach drastically reduces the sample complexity enabling the measurement of peptides only from the proteins of interest. This immuno-purification must be done before sample digestion, before epitopes lose their conformational integrity. In addition, other techniques to reduce complexity can be applied after protein digestion. For example, when studying PTMs, modified peptides are usually less abundant than the non-modified counterparts, hence its necessary to use enrichment methods (*34*). A great example are phosphorylated peptides, which are rarely detected when measuring full proteomes. These low abundant analytes are enriched based on the negative charge of the phosphate group, subsequently enabling their detection by the mass spectrometer (*35*) (Figure 2). The field of phosphoproteomics will be discussed in depth in the upcoming sections.

Liquid chromatography coupled to mass spectrometry (LC-MS)

Peptides are dissolved in a low-pH solvent just before online liquid chromatography (LC) reversed phase separation. Here a high-pressure liquid chromatography (HPLC) system is coupled to the mass spectrometer. The HPLC usually incorporates an automated sampler that injects the sample into a C18 analytical column, where peptides interact with a hydrophobic stationary phase, and are eluted out of the column by a gradient of increasing concentration of non-polar solvent (commonly used is a combination of high purity water and acetonitrile). Here, less hydrophobic peptides will elute first, while highly hydrophobic peptides elute later. Most proteomic applications use flow rates in the nano liter scale ('nano-flow') which confers great sensitivity for low stoichiometry targets. The end of the column either incorporates or is attached to a needle, upon which an electric current is applied to generate the voltage needed for electrospray ionization of the peptides in the sample. Here, droplets containing ionized peptides with a variety of charges are formed. Solvent evaporation increases repulsion of the charged peptides which repel each other until the surface tension is broken and free ions transfer into the gas phase (Figure 1A) (*8, 36*).

As ions enter the vacuum of the mass spectrometer, they are stabilized and moved by electric and/or magnetic fields (*37*). Most instruments have a set of optics that focus and guide the ions towards the mass analyzer, where the m/z value is measured. High-resolution mass analyzers have enough resolving power for peptides to be isotopically resolved, that is to separate the different peptide isotopes (*37*). These occur due to the natural presence of stable isotopes of carbon (^{13}C) and nitrogen (^{15}N). Measurement of a set of isotopes is usually described as an 'isotope envelope', where adjacent isotopes are separated by a mass of 1 Da. However, since mass spectrometers measure m/z values, the measured difference between each isotope depends on the charge state of the peptide. Singly charged species are separated by a value of 1 m/z (or 1 Th), while doubly charged peptides are separated by 0.5 m/z (0.5 Th), and so on. It is by measuring the difference between isotopes that charge states are assigned to peptides, and that mass values are subsequently derived from m/z measurements (*38, 39*).

Despite the high resolving power of current mass analyzers, this measurement is not enough to accurately identify peptides in such complex mixtures. Most mass spectrometers measure the mass of the precursor peptide and subsequently fragment this peptide and measure the masses of the fragments (*40*). These experiments, referred to as tandem MS (MS/MS), provide the information necessary to determine the sequence of the peptide, which then is computationally assigned to a protein, based on a database search against the full proteome of the organism being studied (*26*). The proteome database is obtained by translating the sequenced genome, highlighting the importance genomic technologies have had on the development of proteomics (*41*).

Before diving any deeper into peptide sequencing and quantification, it is necessary to dwell on the different types of mass analyzers currently used, and how these are combined into hybrid instruments that facilitate proteomic measurements at unparalleled speed and precision.

Types of mass spectrometers: from mass analyzers to hybrid mass instruments

All mass analyzers measure the m/z of ions in the gas phase, using diverse physical principles resulting in various resolving powers. Next, I will briefly describe the mass analyzers used in proteomics, starting with low-resolution analyzers like quadrupoles and ion traps, followed by the high-resolution mass spectrometers.

a. Quadrupole: It consists of four rod-shaped electrodes, to which a combination of direct current (DC) and alternate current (AC) is applied, generating electromagnetic fields which are capable of moving and focusing ions of different masses. Both AC and DC amplitudes can be manipulated to filter out certain masses while isolating others (Figure 3A) (*42*). Quadrupoles are widely used in tandem, in instruments called triple quadrupoles or in combination with other mass analyzers. In the former, it functions as mass filter and mass analyzer, while in the latter it is mainly used as a mass filter upstream of the mass analyzer. These hybrid instruments will be discussed in the next section.
b. Ion trap: the most common type of ion trap, also called linear ion trap (LIT), has a similar configuration as the quadrupole, but it can store ions axially, by applying different DC currents to different sections of the rods. The latter creates a 'potential well', in which ions are trapped. LIT can scan for different m/z values by applying two AC currents. First, an exit AC current is applied steadily, while another AC, known as the 'main RF' is ramped up until ions start to resonate and exit the trap. The latter happens in an m/z dependent fashion, so it also permits to isolate ion populations inside the trap according to their mass (*42*).
c. Time of flight (TOF): this type of mass analyzer relies on the principle that heavier ions travel at slower velocities than smaller ions under vacuum (*37, 43*). By measuring the time it takes an ion to 'fly' through a fixed distance, is possible to obtain information about its m/z (Figure 3B). This analyzer has been combined with MALDI to enable mass spectrometry imaging (MSI), a technique in which precursors are directly ionized from pretreated tissue samples, conferring spatially resolved peptide intensities that correlate with protein distribution in the tissue (*44-46*). An example of this specific application can be found in **chapter V**.
d. Orbitrap: here ions are trapped inside a cell by an electric field; the walls of the cell function as an outer electrode, while a spindle shaped electrode sits in the middle. The ions oscillate along the axis of the inner electrode and the frequency of these oscillations is recorded as an image current by the outer electrode. The frequency of these oscillations is dependent on the mass of the ion, and they can be Fourier transformed to obtain m/z values (*47, 48*) (Figure 3C).

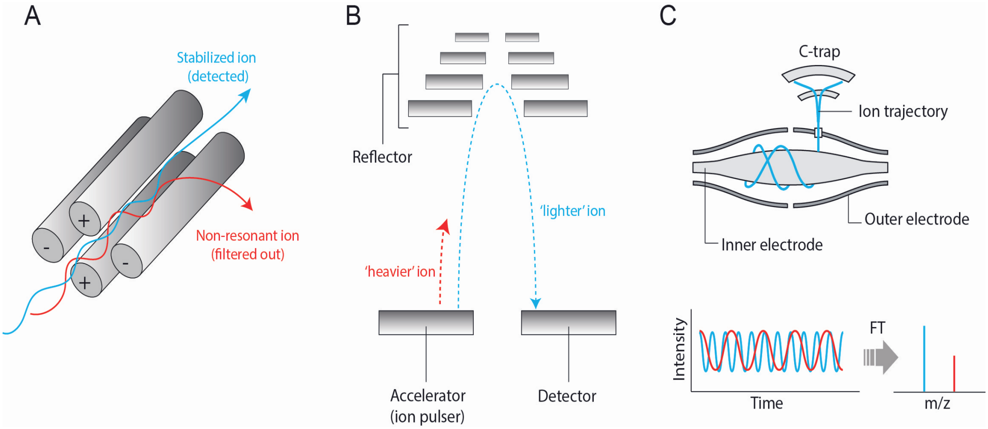


Figure 3. Types of mass analyzers. (A) Quadrupole showing the trajectories of a stabilized ion and a non-resonant ion. (B) Time of flight (TOF) showing the flight trajectories of two ions with different masses. (C) Orbitrap, showing the ion trajectory and a representative diagram of Fourier transformation of ion oscillations into m/z values. Orbitrap diagram was modified from Aebersold & Mann, 2016 (*33*).

These mass analyzers have been combined in a variety of ways in so-called 'hybrid' instruments. Here I will describe some of the instruments that were used for the experimental chapters of this thesis, alongside other popular hybrid mass spectrometers used nowadays.

a. Quadrupole-Orbitrap instruments: by combining a quadrupole up-front of the Orbitrap these instruments are able to acquire full scans to determine peptide masses at high-resolution (MS1), followed by precursor isolation, fragmentation and measurement of peptide fragments (MS2). The first version of this type of hybrid, the Q-Exactive®, has an S-lens that provides improved ion transmission from atmosphere to vacuum. This is followed by a series of flatapoles that guide the ion beam towards the quadrupole while getting rid of uncharged molecules. During the full scan, the quadrupole transmits all ions towards the C-trap, where ions are stabilized and injected into the Orbitrap for high-resolution mass analysis. During an MS2 scan, the quadrupole functions as a mass filter, selecting a specific precursor, which is moved to the high-energy collision induced dissociation cell for fragmentation (discussed in the upcoming section). These ions are then moved to the C-trap and injected into the Orbitrap (*49*) (Figure 4A). Different upgrades have been made to this line of instruments, namely the introduction of an ultra-high-field Orbitrap mass analyzer in the Q-Exactive HF® (*50, 51*) and the improvements in ion transferring and peak picking algorithms introduced in the Q-Exactive HF-X® (*52*). A newer version of this hybrid, the Orbitrap Exploris 480®, has many of the ion optics used in the Q-Exactive HF-X®, but with a layout reconfiguration rendering a more compact and robust instrument. Improvements in the bent flatapole, quadrupole and pumping system allow for a faster scanning speed and higher resolution than previous quadrupole-Orbitrap hybrids (Figure 4B) (*53*).

b. Ion trap-Orbitrap: the first generation of this hybrid allowed high-resolution precursors mass detection in the Orbitrap with subsequent precursor isolation and fragmentation in the LIT (*54, 55*). The next generation of instruments included a quadrupole for mass filtering while keeping the Orbitrap and LIT. By having two mass analyzers, this highly versatile instrument is capable of parallelizing MS1 and MS2 scans, increasing acquisition rate and proteome coverage (*56*).
c. Triple quadrupole: as the name implies, here three quadrupoles are used in tandem with the first and last quadrupoles used as mass filters (*57, 58*). The first quadrupole is used for precursor isolation, while the second quadrupole is used as a collision cell for analyte fragmentation. The last quadrupole is used to filter for specific fragments, which are then detected by an electron multiplier (Figure 4C). This instrument has been widely adopted by the field of targeted proteomics (addressed in an upcoming section), where it confers high sensitivity and selectivity with unparalleled robustness for the measurement of low abundant targets in complex mixtures (*59, 60*).
d. Quadrupole-TOF: also called Q-TOF or Triple-TOF® since it has a similar configuration to the one of the triple-quadrupole, although here the third quadrupole is replaced by a TOF (Figure 4D). This combination of mass filters up front of the high-resolution mass analyzer provide a fast scanning instrument capable of tandem MS, enabling deep proteome coverage (*61*).

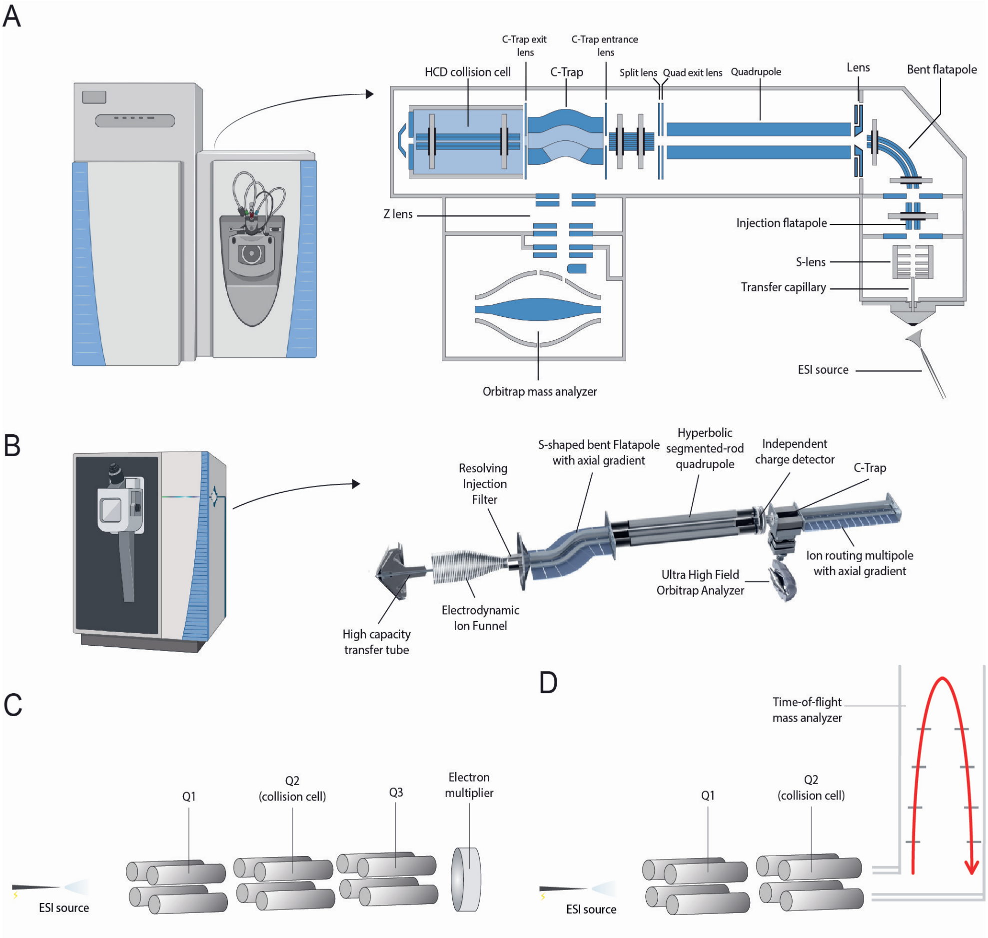


Figure 4. Hybrid instruments. (A) Q-Exactive and a diagram of the instrument internal configuration, modified from Michalski *et al*, 2011 (*49*). (B) Orbitrap Exploris and diagram of the internal components of the instrument, modified from Bekker-Jensen *et al*, 2020 (*53*). (C) Diagram of a triple quadrupole. (D) Diagram of a quadrupole-TOF instrument. Illustrations of the external appearance of the mass spectrometers were obtained from BioRender.com.

Peptide fragmentation

As previously addressed, measurement of the precursor m/z is not enough for exact peptide identification. Structural information in the form of amino acid sequence is necessary, and for this, the peptide has to be fragmented followed by measurement of the fragment ions. The m/z values of the fragments and the precursor are used to puzzle together the peptide sequence. Two main types of fragmentation techniques are used for proteomics applications:

a. Collision induced dissociation (CID): here peptides are accelerated to undergo collision with an inert gas, like nitrogen, argon or helium. The energy generated in these collisions is transformed into vibrational energy that causes peptide bond breakage (*62*). There are two types of CID fragmentation:
    - Resonance-excitation CID, which takes place in a LIT. Here, an m/z specific energy is applied, causing the precursor ion to resonate and collide with the gas atoms. Since this type of fragmentation takes place in the LIT, fragments are measured with a low resolution. Furthermore, low mass fragments are usually not acquired and breakage of lowest energy bonds, like those of certain PTMs (e.g. phosphorylation), leads to so-called neutral losses that translate into uninformative MS2 spectra (*63*).
    - Beam-type CID, is used in triple-quadrupoles and quadrupole-TOF instruments. It is also known as high-energy C-trap dissociation (HCD), as first coined by Olsen and collaborators when it was introduced into an Ion trap-Orbitrap hybrid (*64*). Specialized HCD cells were subsequently incorporated into other Orbitrap hybrids, since it enables full-mass fragment acquisition, as opposed to fragmentation and MS2 acquisition in LIT. Furthermore, due to its fast activation time and high energy, peptide bond breakage is favored, instead of the neutral losses frequently observed in phosphopeptides during resonance-excitation CID (*65, 66*).

    Both types of CID based fragmentations induce the appearance of y- and b-type ions (Figure 5) (*67, 68*). Most of the instruments used in this thesis apply HCD, which is known to be one of the best fragmentation methods for phosphoproteomics experiments (*66, 69*).

b. Electron transfer dissociation techniques: even though this type of fragmentation techniques were not used in the chapters of this thesis, they deserve attention since they have proven to be useful in phosphoproteomics experiments (*63, 70, 71*). Two main types of electron based dissociation techniques exist: electron capture dissociation (ECD) and electron transfer dissociation (ETD). The latter has become more popular in the proteomics field since it is incorporated in a variety of instruments (*72-74*). In ETD a radical anion is used as a vehicle to transfer the electron to the positively charged peptide. Electron acquisition leads to peptide backbone fragmentation that is independent of vibrational energy, hence avoiding breakage of lower energy bonds (*75*). The latter makes it ideal for fragmentation of peptides carrying PTMs, such as the phosphate group. Moreover, ETD can be used complementary to HCD, since it generates c- and z-ions that have proven to be useful for applications that require exact PTM localization on the peptide backbone, such as phosphoproteomics (Figure 5) (*68, 76, 77*). Still, the combination of these fragmentation methods takes a toll on the instrument scanning speed, and since HCD has been shown to provide high identification rates for phosphopeptides (*69*), we decided to use this fragmentation technique for the phosphoproteomics experiments reported in this thesis.

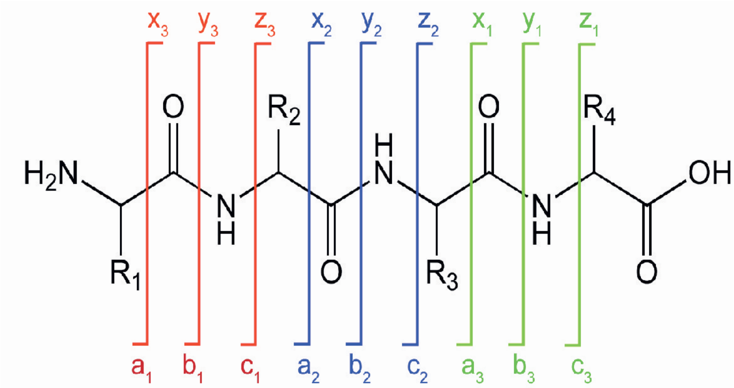


Figure 5. Peptide fragment notation. Based on nomenclature suggested by Roepstorff & Fohlman, 1984 (*68*). Image obtained from 'Peptide sequence tag', *Wikipedia, The Free Encyclopedia,* 2021.

MS acquisition methods and data analysis

With the development of hybrid instruments, researchers find themselves with a variety of options when it comes to choosing a measurement or acquisition method. These strategies can be divided into two large categories. First, the so-called shotgun proteomics or data-dependent acquisition methods, which are mainly used in discovery proteomics experiments. These are preferred for hypothesis-free projects in which the goal is to measure as many analytes as possible. Next are the targeted proteomics methods, which focus on measuring a subset of analytes, hence, they are preferred for hypothesis-driven projects. Both strategies will be described in depth in the upcoming sections, highlighting their corresponding strengths and shortcomings.

Data dependent acquisition (DDA) or shotgun proteomics

DDA is the most popular acquisition strategy in the field of proteomics. Here, a full MS scan of all precursors co-eluting out the LC column is acquired at high resolution (MS1). Next, a defined number of precursors, usually those with the highest intensities (also referred as a 'top N'), are isolated separately and subjected to fragmentation and subsequent fragment mass acquisition (Figure 6A) (*19, 78*). To maximize the number of MS/MS events of different peptides, once a precursor is fragmented it is placed in an exclusion list for a determined amount of time (usually a few seconds), to avoid repeated fragmentation of the same precursor (*79*). Here it is important to highlight that the main goal is to optimize the time usage of the instrument, to maximize the number of precursors measured and fragmented (*80, 81*). The time it takes the instrument to do an MS1 scan and the subsequent MS2 scans is known as the cycle time. Ideally, cycle times should be short, to ensure constant sampling of the precursors eluted from the LC at the MS1 level, while fragmenting as many precursors as possible to increase the number of peptide identifications.

The acquisition process previously described is repeated constantly during the whole LC-gradient, which generates thousands of MS1 and MS2 spectra. Next, the goal is to match the precursor masses to the theoretical masses of peptides obtained from an *in silico* digestion of the proteome being studied. As previously mentioned, this information is obtained from the sequenced genome of the organism of interest. Once the measured peptide masses are matched to theoretical masses, the acquired fragmentation spectra are compared to the expected fragment masses (*82*). When both the precursor mass and the spectra coincide with the *in silico* predictions, this is considered a positive spectral match (PSM) (*83*). Matches are usually scored, depending on the degree of similarity between the obtained spectra and the theoretical one. These scores can be used to subsequently filter out the poor matches (*84*). Furthermore, since this has to be done for thousands of MS2 spectra, it is likely that false-positive matches occur. To control this, a decoy database of scrambled peptide sequences is combined with the database of interest. After PSM scoring the number of decoy PSMs is used to estimate the false discovery rate (FDR) (*85, 86*).

When it comes to quantification strategies, different approaches can be used when applying shotgun proteomics. Label-free quantification (LFQ) has emerged as a straightforward option to measure relative changes in peptide intensities across a group of samples (*87*). This relies on the MS1 signal of precursor peptides, which is obtained in the form of an ion extracted chromatogram and used for quantification (Figure 6A) (*88*). Even though it is a cheap and relatively straightforward approach, several criteria must be met to ensure good quality results: all samples must be of a similar type and handled with the exact same sample preparation workflow. This entails that samples to be compared amongst each other should be measured in the same batch, trying to assure the same measuring conditions for all. Moreover, it is key that chromatographic separation remains stable during the data acquisition (*89, 90*). Changes during any of these steps can translate into loss of sample or bias introduction that would hamper accurate quantification.

To avoid some of the pitfalls of LFQ, labelling strategies can be implemented. There are two types of labeling strategies, namely stable isotope labelling of amino acids in cell culture (SILAC), in which the experimental model is grown in the presence of either heavy or light amino acids. Peptides derived from these samples can be differentiated at the MS1 level in the mass spectrometer. This aspect is key, since samples with different labels can be treated differently (e.g. growth factor stimulated vs non-stimulated), and subsequently mixed, in order to apply the exact same sample preparation workflow (*91, 92*). Furthermore, this also reduces the measurement time, since with LFQ samples must be measured separately, while with SILAC they are acquired at the same time. When it comes to chemical labelling at the peptide level, isobaric tags for relative and absolute quantification (iTRAQ) (*93*) or tandem mass tags (TMT) (*94*) have become popular alternatives. Here peptides from different samples are labeled with different isobaric tags, hence the precursor mass of the same peptide is identical across all samples. Samples can be combined, and upon fragmentation, different tags release reporter ions of different masses. The intensity of the reporter ion correlates with the peptide concentration, allowing quantification at the MS2 level.

Despite having different strategies for quantification, DDA measurements usually suffer from the stochastic nature of peak picking methods used to determine which precursors are fragmented after each MS1 scan (*95, 96*). The latter leads to the so-called missing values problem: some peptides are identified in certain samples while they are not detected in others. The latter is a pitfall of shotgun proteomics, since posterior data analysis and interpretation is difficult, hampering the ability to extract biological insight from the measurements (*97*). Some options to address this issue have been developed, such as the match-between-runs algorithm included in the MaxQuant software (*90, 98*). Here, the peptide identification in one sample is used to look for precursors with the same m/z and charge state in the other samples, within a specific retention time window. If the precursor is found, the identification is 'transferred' and this MS1 signal can now be used for quantification, hence minimizing the number of missing values (*99*). Another approach relies on the assumption that most missing values are due to a decrease in the peptide concentration in the sample, hence falling below the limit of detection (*100*). Based on this, missing values can be replaced by a variety of methods, for example, by substituting with random low values from the distribution of intensities of the sample (*101, 102*). The latter facilitates downstream data analysis and extraction of relevant biological information.

Data independent acquisition (DIA)

To circumvent some of the pitfalls of DDA, new DIA methods have been developed in recent years. Here, the instrument samples all precursors eluting out of the LC column and these are systematically fragmented. For this, the MS1 range is divided into windows of a predetermined size (e.g. 20 m/z), and all precursors within each of these windows (e.g. 200-220 m/z) are co-fragmented and measured simultaneously (*103, 104*). The latter avoids semi-stochastic peak picking at the MS1 level, hence reducing the number of missing values in large datasets. This strategy also permits detection of low abundant peptides that would not be selected for fragmentation using DDA. In addition, the quantification is done at the MS2 level, in clear contrast of DDA-LFQ measurements, in which quantification is based on the precursor signal. Altogether, this translates into an increase in measurement sensitivity.

Despite the prospective benefits, DIA brought its own set of challenges, since by fragmenting several precursors at the same time, the output is a complex MS2 spectra not amicable with peptide identification using the regular database searches previously described. For this reason, new computational tools had to be developed to extract the information from these rich MS2 spectra (*105-107*). Many of them relied on the previous acquisition of a peptide library representative of the set of samples to be analyzed. A popular way of doing this is by pooling the samples and performing a high-pH reverse phase fractionation, followed by measurement of each fraction by DDA and standard database search. Next, the samples are individually measured using DIA and specialized software is used to extract ion chromatograms of the precursors and fragments found in the peptide library (*108*).

The dependence of DIA on previous DDA measurements hampered the popularization of this acquisition strategy during its early days. However, recent developments in computational tools have enabled DIA to outmatch DDA, both in terms of sensitivity and reproducibility, all this without the need of previously acquiring a spectral library (*109*). The latter is indicative of the promising potential of DIA, which is likely to become more popular in the upcoming years, as faster instruments and better computational tools are developed.

Targeted proteomics

Depending on the research question, researchers might not be interested in measuring as many proteins as possible, as is the case for DDA or DIA, but instead would prefer to reproducibly measure a smaller group of targets across different samples. For this goal, mass spectrometers can be programmed to follow only a predefined list of m/z values. To achieve this, it is necessary to have mass filtering capacity to exclusively select the predefined list of precursors. This is achieved by using hybrid instruments that incorporate a quadrupole up-front of the mass analyzer. Triple-quadrupoles are the preferred option for targeted proteomics measurements: they are fast, robust and cheaper than high-resolution mass spectrometers (*26*). Still, hybrids with high-resolution mass analyzers have also been used for targeted acquisition with great success, as will be discussed in the next sections.

Targeted proteomics presents some clear advantages when compared to DDA approaches: targeted acquisition can provide higher sensitivity, since quantification is done at the MS2 level (addressed in the upcoming section), while in LFQ shotgun-proteomics this is done at the MS1

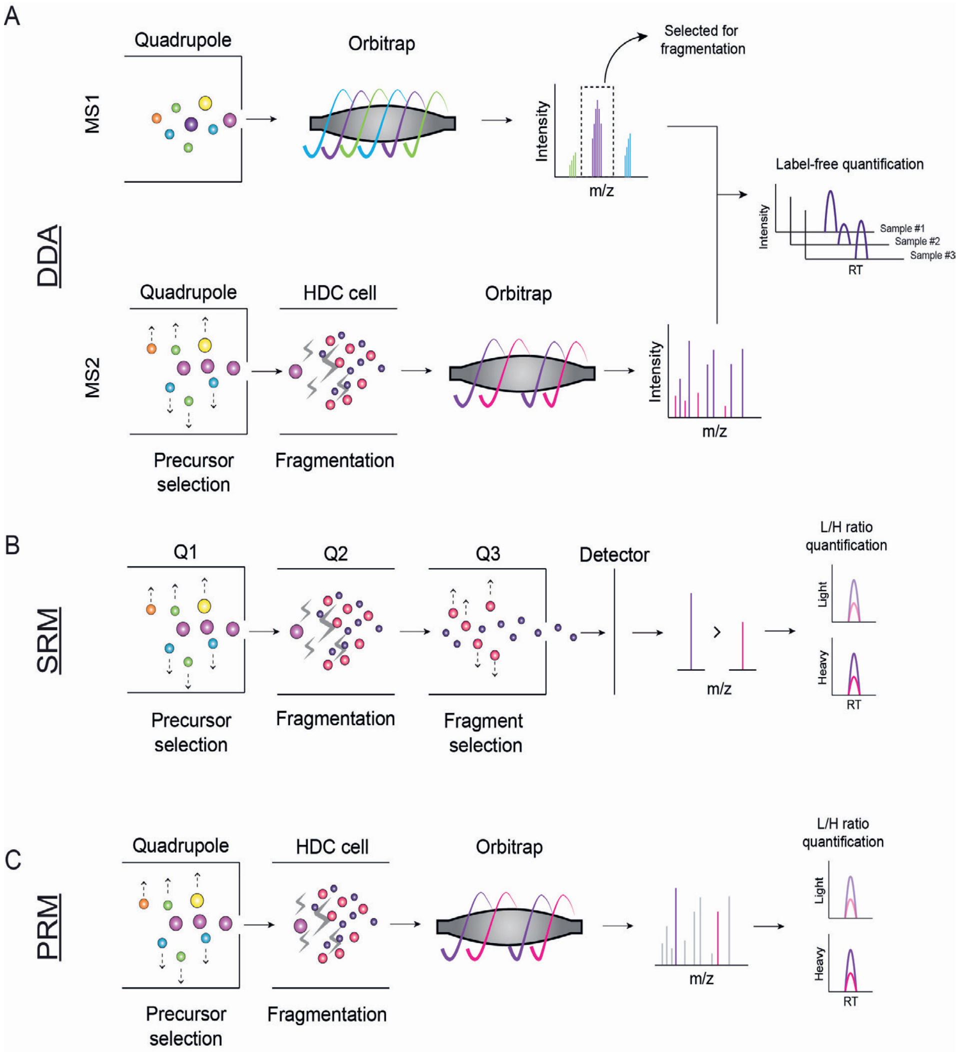


Figure 6. Types of data acquisition. (A) Shotgun-proteomics or data-dependent acquisition (DDA). MS1 scan detects the most abundant precursors, which are isolated for fragmentation followed by measurement of fragments in the MS2 scan. Information from the MS2 scan is used for peptide sequencing and identification, while the MS1 signal can be used for label-free quantification. (B) Selected reaction monitoring (SRM). Precursors are isolated in Q1, followed by fragmentation (Q2) and sequential isolation of peptide fragments in Q3. (C) Parallel reaction monitoring (PRM). Precursors are isolated in the quadrupole followed by fragmentation and measurement of all peptide fragments in the Orbitrap. For both SRM and PRM, quantification can be based on light to heavy ratios calculated for each transition.

level. The MS1 scan is known to be noisier due to the high background found in complex samples, which is partially avoided after specific precursor filtering done in targeted proteomics (*110*). Moreover, tailored optimization of measurement conditions for each analyte of interest can translate into increased sensitivity, which is not possible in DDA methods (*59*). Finally, the use of standards, consisting of heavy-labeled versions of the peptides of interest is a great tool for method development and quantification, leading to more accurate and reproducible results, as will be described in the next paragraphs.

Despite clear benefits, targeted proteomics also presents certain drawbacks when compared to shotgun proteomics. First, assay development is more time consuming than for DDA, since measuring conditions must be optimized for each specific analyte. In addition, the use of heavy-labeled standards can steeply increase the cost of applying this technology, especially when compared to LFQ shotgun-proteomics. Finally, since previous knowledge is necessary for target selection and assay development, this technique is mainly limited to hypothesis-driven projects. There are different types of targeted proteomics approaches, in the next sections, I will describe the main two, which are also the ones used in some of the chapters of this thesis.

Selected reaction monitoring (SRM)

This type of measurement is done on a triple-quadrupole. Here, the first quadrupole functions as a mass filter, selecting only the precursor of interest. The second quadrupole is used for CID fragmentation of individual precursors. The third quadrupole filters for specific fragments of each precursor, which eventually hit a detector, namely an electron multiplier used to record the signal (Figure 6B). The pair of a precursor and a specific fragment is known as a transition. Several transitions can be measured for a given precursor, which is the reason why the term multiple reaction monitoring (MRM) has also been widely used as a synonym to SRM (*59, 60, 110*).

When SRM is used to measure peptides to infer protein expression, the correct selection of the peptide is crucial to guarantee selectivity. It is important to select peptides that are specific for the protein of interest, these are known as proteotypic peptides. Next, another important factor to consider in the peptide selection process is that different peptides coming from the same protein can ionize differently, translating into different intensities in the MS measurement. To achieve the highest possible sensitivity, peptides that ionize well and are detected with a relatively high intensity should be prioritized. Furthermore, it is important to avoid the selection of peptides with PTMs, either naturally occurring ones or as an artefact due to sample preparation (e.g. methionine oxidation) (*59, 60*). If the interest lies in measuring a specific PTM, like a biologically relevant phosphosite, the peptide options drastically reduce. If a PTM cannot be accessed by mass spectrometry using tryptic digestion, one might consider using an alternative protease to obtain a more suitable peptide for measurement (*18*).

After peptide selection, it is important to choose the best transitions. These should be the ones that have the highest intensities and the least interference. This can be done with aid of online tools that suggest the best transitions based on publicly available data (*111-113*), or by direct measurement of the targets of interest. For this, is important to have access to a sample in which the target peptides are easily measured. To avoid the challenges involved with using a biological sample for method development, it is ideal to synthesize heavy-labeled versions of the targets of interest (*114*). These molecules behave identically at the chromatographic level when compared to the endogenous peptides, and fragment in the same way as well. They can be used for transition selection and for collision energy (CE) optimization. The latter can be optimized for each individual transition, to further increase the sensitivity of the method (*59, 60*).

During the method development stage, it is also important to determine the exact retention time of each precursor. This information can be used to schedule the measuring of the transitions to their corresponding expected retention time windows. The latter is particularly relevant when measuring several peptides and their corresponding transitions in a single method (*115*). The instrument allocates a given amount of time to measure each transition, also known as the dwell time. The latter depends on the number of transitions being concurrently measured at any given moment; the more transitions, the less dwell time the instrument can allocate to each transition. The time it takes the instrument to cycle through the whole list of transitions is known as the cycle time, and it should be kept relatively short since quantification will be based on the integration of peak areas from extracted ion chromatograms (*59*). For this reason, the cycle time is usually fixed, depending on the average chromatographic peak width. If all transitions were to be measured during the entire method, dwell times could be potentially low, hence compromising the data quality. If dwell times were fixed to a desired value, it could cause prolonged cycle times, hence compromising the quality of the extracted ion chromatograms for quantification. By scheduling the measurement of transitions these situations are avoided, keeping the dwell times above the desired minimum, and a cycle time that assures accurate peak integration (*59*).

For the assessment of data quality, several criteria can be used. First, the exact co-elution of different transitions strongly suggests that the signal originates from the same precursor and not from interferences. Moreover, by spiking the heavy-labeled standards into the samples, it is possible to compare the signal of the light or endogenous version versus the signal of the heavy. Co-elution of both versions is used to confirm that the light signal detected corresponds to the target of interest (*60*). The latter also helps in the interpretation of missing values, since the absence of endogenous signal with the presence of the heavy labeled standard, suggests that the target fell below the limit of detection.

The same heavy-labeled standards used for method development and assessment of data quality can be employed for quantification. Since the same amount of standards is spiked in all samples, a ratio of the light to heavy signal can be used to quantify expression changes across different samples (Figure 6B) (*59, 60, 114*). Both assessment of data quality by visual inspection of ion chromatograms, and extraction of light to heavy ratios can be done with the aid of specialized software like Skyline® (*116*).

Parallel reaction monitoring (PRM)

In PRM the measurement is done on a hybrid instrument with a high-resolution mass analyzer, namely an Orbitrap or ToF. The quadrupole is used for precursor isolation, followed by fragmentation and subsequent parallel measurement of all fragments in the Orbitrap at high-resolution (Figure 6C). This technique has been shown to perform similarly to SRM, with comparable reproducibility, dynamic range and accuracy (*117, 118*). In addition, PRM requires less time investment in method development, since no optimization per-transition is done. The only information needed is the m/z values of the precursors of interest and their expected elution time (if scheduled measurement is to be applied) (*119, 120*).

Among the parameters to be optimized are the CE, which can be set individually for each precursor, the Orbitrap resolution and the C-trap fill time. First, the higher resolving power of the Orbitrap translates into a higher selectivity (*120*). Resolution can be increased by augmenting the transient time, which is the time ions oscillate around the Orbitrap (*121*). This translates into a theoretical increase in the signal-to-noise ratio after Fourier transformation (*122*), while also conferring higher discrimination from interfering fragments generated by co-isolation of other molecular species with the same m/z as the precursor (*121*). Next, the C-trap fill time has a direct impact on the sensitivity and the dynamic range of the measurement; at higher fill times it is possible to quantify analytes at lower concentrations (*119*). Here it is important to highlight that optimization of the instrument measurement time is key; since high-resolution measurement and fill time of the next precursor take place in parallel, the fill time should be adjusted according to the resolution used (e.g. transient time). In other words, while the Orbitrap is performing a high-resolution measurement, the instrument should be filling the C-trap for the next measurement. If these two parameters are not coordinated (e.g. C-trap fill time different from the Orbitrap transient time), the instrument will not operate at its full capacity. If specific applications require prioritizing either sensitivity or specificity, the necessary adjustments can be made to the fill time and resolution, respectively (*119, 120*).

In terms of multiplexing the measurement of several targets, PRM also benefits from scheduled acquisition. Here, the cycle time is defined by the number of targets and the transient time. By lowering the number of targets measured concomitantly it is possible to keep fill times high with cycle times that ensure proper sampling of the precursor elution profile (*119*). For assessment of data quality, raw data can also be imported into Skyline® software. Here, the signal from MS2 spectra is extracted in the form of ion chromatograms, so that co-elution of different fragments can be visually inspected, just as with SRM data. Moreover, PRM data can also be searched against proteome databases, to further confirm the identification of the target of interest, providing yet another level at which measurement selectivity can be assessed (*119*). Finally, PRM also benefits from the use of heavy labeled standards, which can be used for method development and quantification based on light-to-heavy ratios, as it is done for SRM measurements (Figure 6C).

Overall, PRM has been a nice addition to the toolbox of targeted proteomics. Due to its straightforward method development, great sensitivity and selectivity, it has been used in a wide range of samples, from whole cell lysates (*123-125*) to clinical specimens (*118, 126-128*). In addition, the popularity of the benchtop quadrupole-Orbitrap instruments (e.g. Q-Exactive® series), gives an edge to PRM over other targeted methods. Labs that only have this type of high-resolution mass analyzer for shotgun-proteomics, can now implement targeted acquisition using the exact same instrument. Importantly, even though I focused on PRM applied in a quarupole-Orbitrap instrument, as was done in **chapter IV**, this type of targeted acquisition can also be done on other hybrid instruments with high-resolving power, like the quadrupole-TOF (Q-TOF) (*129*).

Protein phosphorylation

Regulation of cellular processes has proven to be highly complex, involving different layers of regulation, relying on more than just control of gene expression. One of the main regulators of protein functionality are post-translational modifications (PTMs) (*130*). Many of these chemical modifications are reversibly added to the polypeptide backbone, providing a fast mechanism to convey messages within the intracellular environment. Up to date, there are more than 200 different PTMs described (*131*), providing the cell with a diverse pool of modifications to regulate cellular processes at different levels. Despite this overwhelming variety, protein phosphorylation has been the most studied PTM thus far, proven to be involved in almost every cellular process studied up to date (*132, 133*). In the next section, I will address the reasons why protein phosphorylation is so prevalent in living organisms and why it evolved as a major regulator of cellular signaling.

Phosphate as a key building block of life

Phosphorus-containing molecules have proven to be important players in the evolution of living organisms, very likely due to their abundance and chemical versatility (*134*). Phosphate salts were available during the first stages of life, providing evolution with an important piece for the first self-replicating molecules. The phosphate ester bond was first introduced during the development of the genetic code, since phosphate esters are responsible for joining bases and forming chains of nucleic acids (e.g. DNA and RNA). In addition, the phosphate group is present in the high-energy molecule adenosine triphosphate (ATP), which is very stable under physiological conditions, but can also be hydrolyzed by enzymatic activity. Due to these qualities, ATP is found at high concentrations in cells and is able to provide free phosphate groups, which were then used by evolution as a key molecule to modify proteins (*135*).

The introduction of a phosphate group in a protein backbone adds a negative charges at pH 7.0 and can drastically impact the structure and chemical environment of the protein (*136*). The higher charge density of phosphorylated residues allows them to easily form hydrogen bonds or salt bridges within the same molecule or with other molecules (*137*). These features are useful to regulate protein-protein interactions, enzymatic activity, intracellular localization and even protein degradation (Figure 7A) (*132, 138*).

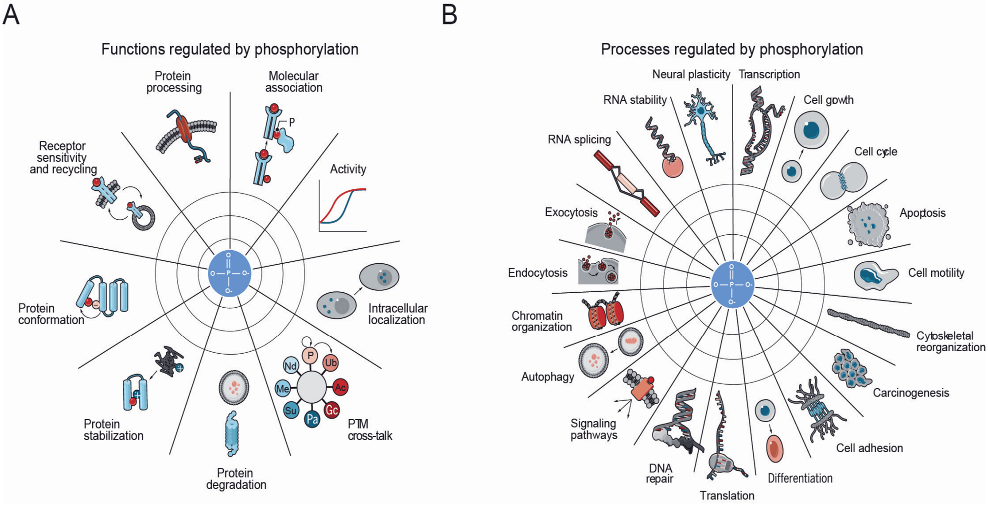


Figure 7. The versatility of protein phosphorylation. (A) Molecular functions regulated by protein phosphorylation. (B) Biological processes known to be modulated by protein phosphorylation. Modified from Needham *et al*, 2019 (*138*).

The robustness and versatility of the phosphate group allowed it to become the main PTM across the three domains of life. From bacteria, which use histidine phosphorylation to regulate their two component systems (*139, 140*), to archaea which also rely on histidine phosphorylation but also exhibit higher levels of tyrosine, serine and threonine phosphorylation (*141*). The latter three are the most abundant modified residues in eukaryotes, where protein phosphorylation has been described as a major regulator of signal transduction from unicellular organisms like yeast and pathogens like *Plasmodium (142-144)*, to multicellular organisms like *Caenorhabditis elegans* (*145*) and *Homo sapiens (146, 147)*. Due to the widespread distribution of protein phosphorylation, it is possible to find evidence of this PTM regulating a myriad of cellular processes, from cell cycle progression to DNA transcription and cytoskeleton reorganization (Figure 7B).

The success of protein phosphorylation as a PTM was accompanied by the evolution of enzymes that catalyze the addition and removal of phosphate groups, allowing phosphorylation to further position itself as the main driver of rapid cellular changes. The switch-like behavior of protein phosphorylation is possible thanks to kinases, which add phosphate groups, and phosphatases, which catalyze the reverse reaction (*133*). As will be addressed in the next section, these enzymes are of paramount importance since they control the balance of protein phosphorylation inside the cell. Many pathologies, including cancer (*148, 149*) and neurodegenerative diseases (*150, 151*), are often associated with unbalanced protein phosphorylation, highlighting the importance of understanding the effects of this PTM on proteins and their functions.

A brief look into the kinome and kinase specificity

Eukaryotic protein kinases are responsible for transferring the γ-phosphate of ATP molecules mainly to serine, threonine and tyrosine residues of the peptide backbone. The enzymes that carry this task contain the eukaryotic protein kinase catalytic domain, which consists of approximately 250-300 amino acids that fold into a core structure with catalytic capacity (*152*). The domain consists of two lobes with a cleft in the middle, which lodges the ATP molecule leaving the γ-phosphate exposed on the outer part of the cleft, to facilitate the transfer of the phosphoryl group to the substrate (Figure 8A) (*132*). Due to the high conservation of the kinase catalytic domain, Manning and colleagues mapped this domain with great sensitivity to a large collection of human genomic data, which allowed them to identify 518 protein kinases, just shy of 2% of all human genes (*153*).

All known kinases have been classified into different groups, based on the sequence comparison of their catalytic domain, sequence similarity outside of the catalytic domain and previously known biological functions. Despite the high conservation of the catalytic domain, changes in the hydrophobicity and charge in the rest of the protein sequence determine the substrate specificity of each kinase (Figure 8B). One of the main differences used for grouping is the nature of the phosphorylated residue, with tyrosine kinases grouped separately from serine/threonine kinases. Tyrosine kinases have a deeper cleft capable of accommodating a more bulky residue when compared to serine/threonine (Figure 8C). This difference in the catalytic domain is one of many that aids to dictate protein kinase specificity. Here we have to consider that the number of potentially phosphorylatable amino acids (hundred thousands) is far greater than the number of those actually detected to be phosphorylated by any given kinase (from just a few to a few hundred) (*132*).

So how does a kinase phosphorylate only a specific set of substrates among a sea of potential targets? Evidence suggests that a combination of multiple mechanisms permits this important selectivity. As previously mentioned, the catalytic cleft determines what type of residue is phosphorylated, but it also interacts with consensus sequences surrounding the phosphorylation site. The latter determines the so-called phosphorylation motif of the kinase (Figure 8D) (*154-156*). Unfortunately, these consensus sequences can be shared between several kinases, while there can also be overlap of different consensus motifs for the same phosphorylation site. In addition, sequences outside the catalytic domain, known as distal docking motifs are able to improve the local concentrations of targets around the kinase, hence increasing the chances of phosphorylation (e.g. D and DEF motif for MAP kinases) (*157, 158*). Moreover, other factors such as targeting subunits (e.g. cyclins for CDKs) increase the affinity for certain substrates (Figure 8E). These targeting subunits are also thought to aim the kinase to specific locations in the cell, which very likely removes many potential targets that compete with *bona fide* substrates (*159, 160*). Finally, interaction between kinase and substrate can be mediated by other proteins known as adaptors or scaffolds. These scaffolds act in a similar way to targeting subunits, with the difference that the interaction between kinase and adaptor is thought to be more dynamic (*161, 162*). In addition, scaffolds are usually enriched in certain locations of the cell, while also having the ability to recruit other kinase regulators, from activators to inhibitors, usually in the form of other kinases or phosphatases (*162, 163*). This permits a fine tuned control of the enzyme activity in both space and time, highlighting the potential complexity of kinase regulation.

What about phosphatases?

Historically, phosphatases have gathered less attention than their counterparts, despite the fact that both enzymes control the levels of protein phosphorylation that determine key cellular decisions, like cell cycle progression (*164*). The number of serine and threonine phosphatases in the human genome is almost one order of magnitude lower when compared to the number of kinases (*165*). So how is specificity achieved in the realm of phosphatases? Many of them form holoenzymes, composed of a shared catalytic core and regulatory subunits that confer specificity (*166*). A recent study tried to map all protein phosphatase genes, finding a total of 189 genes, of which only 105 had a previously documented protein phosphatase activity. Moreover, of phosphosites with a known upstream kinase, only 7% of them have an assigned phosphatase (*165*). The latter shows that substrate assignment for phosphatases has lagged behind that of kinases, which is a challenge that must be addressed by the scientific community if a comprehensive understanding of protein phosphorylation is to be achieved.

MS-based phosphoproteomics

The development of high-end mass spectrometers during the last decades and the evident relevance of protein phosphorylation on a variety of biological processes prompted the rise of MS-based phosphoproteomics as a field of study. In this section, I will describe some of the main challenges faced when studying protein phosphorylation by MS, how the scientific community has addressed many of them, and how some are still relevant hurdles. Moreover, I will also highlight some seminal studies in the field that influenced the work presented in this thesis.

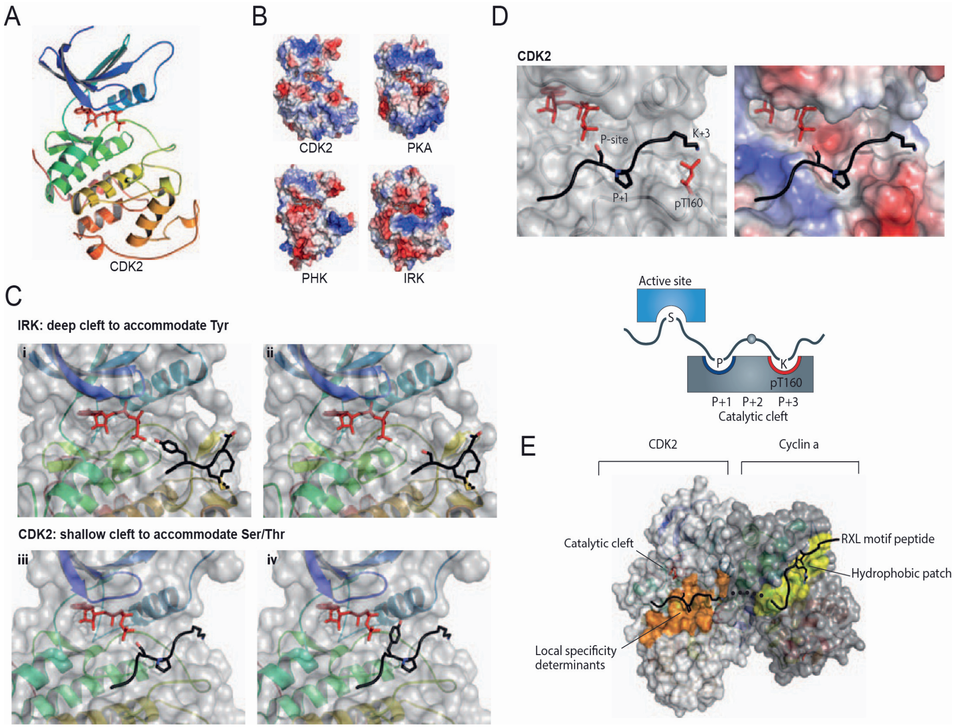


Figure 8. Kinase structure and determinants of kinase specificity. (A) Structure of CDK2 (PDB ID: 1QMZ), showing a first lobe comprised mostly of β-sheets (blue) and a second lobe formed by α-helices (green, orange and yellow). This two-lobe structure forms an ATP-binding cleft with catalytic activity. (B) Four different kinases, showing a similar fold but with differences in the charge and hydrophobicity of the surface residues. Positively charged residues are shown in blue, while negatively charged residues are shown in red. (C) The depth of the catalytic cleft determines the specificity of the phosphorylation site residue: i) IRK deep cleft accommodates the bulky Tyr residue, ii) while the Ser residue does not extend deep enough into the cleft. iii) Ser residue nicely fits into the narrow cleft of a Ser/Thr kinase (CDK2), iv) while the Tyr residue is too large to fit into the catalytic cleft. (D) Local interactions aid in determining specificity. The Pro residue at the +1 position is greatly favored because the nitrogen atom is unable to form a hydrogen bond, unlike any of the other amino acids. The positively charged residue at +3 interacts with the phosphorylated T160 of CDK2. If this residue is not phosphorylated, substrates show less binding affinity with no preference for a positively charged amino acid at +3 position. (E) Targeting subunits enhance substrate recognition. Cyclin A (right) bound to CDK2 (left). The hydrophobic patch of cyclin A (shown in yellow) recognizes a docking motif in the substrate peptide (shown in black). Local interactions near the catalytic cleft determine substrate specificity (shown in orange). All panels are modified from Ubersax & Ferrell, 2007 (*132*).

The many challenges of phosphoproteomics

Protein phosphorylation is known to occur at relatively low stoichiometry, which hampers the detection of phosphorylated peptides in complex samples. The latter magnifies the stochastic nature of precursor selection observed during DDA measurements, leading to highly incomplete datasets that hinder comparison between samples and proper analyte quantification. To partially circumvent this, it is necessary to enrich the phosphopeptides, getting rid of the unmodified counterparts. The latter can be achieved by a variety of methods, including immuno-based purification (e.g. purification of phospho-Tyrosine with a specific antibody) (*167*) and chromatographic methods, like strong cation exchange (SCX) (*168, 169*) or strong anion exchange (SAX) chromatography (*170, 171*). In addition to the latter, immobilized metal ion affinity chromatography (IMAC) approaches have become widely popular due to their elevated performance in phosphopeptide enrichment (*172-174*). In the latter, positively charged ions in the stationary phase interact with the negatively charged phosphopeptides. This interaction is used to retain phosphopeptides on the column while non-modified peptides flow through (*174, 175*). In this thesis we used either Ti^{4+}-IMAC or Fe^{3+}- IMAC, the latter on an automated platform that has shown unparalleled reproducibility and sensitivity, allowing for high-throughput enrichment of phosphopeptides from many samples (*176*).

Another technical hurdle of protein phosphorylation analysis by MS is the accuracy of phosphorylation site localization. The information necessary to determine which residue of the peptide backbone carries the phosphate group is obtained from the MS2 fragment spectra (*177-179*). As previously discussed, the fragmentation method is what determines the type of fragments obtained, and in the case of phosphoproteomics, beam-type CID fragmentation is preferred due to its availability in a wide range of current hybrid instruments, but mainly due to its ability to produce fragment-rich spectra that aid in phosphosite localization (*69*). On instruments that use ion trap CID fragmentation, neutral loss of the phosphoric acid molecule leads to poor fragmentation of the peptide backbone and compromises the capacity to accurately determine site localization (*63, 179, 180*).

Even with the best fragmentation technique, phosphopeptide measurements still poses an important challenge. For example, two peptides with the same sequence, but carrying a phosphate group on different residues (Figure 9A) have the exact same m/z value, hence they cannot be differentiated solely at the MS1 level. Moreover, if these phospho-isomers are not chromatographically separated, they will be fragmented concomitantly, compromising phosphosite localization and leading to inaccurate MS1-based quantification (Figure 9B,C) (*181*). The latter poses a great hurdle, since phospho-isomer separation by liquid chromatography is highly dependent on the physicochemical properties of the peptide and the conditions of the analytical column (*69, 182*). Here, MS2 quantification has an edge, since quantification can be based on the signal of phospho-isomer specific fragments. Overall, this shows how targeted phosphoproteomics could be a great complement to shotgun phosphoproteomics, since it has greater sensitivity for low stoichiometry targets such as phosphopeptides, and it can also discriminate between phospho-isomers if specific transitions are measured (*183-186*).

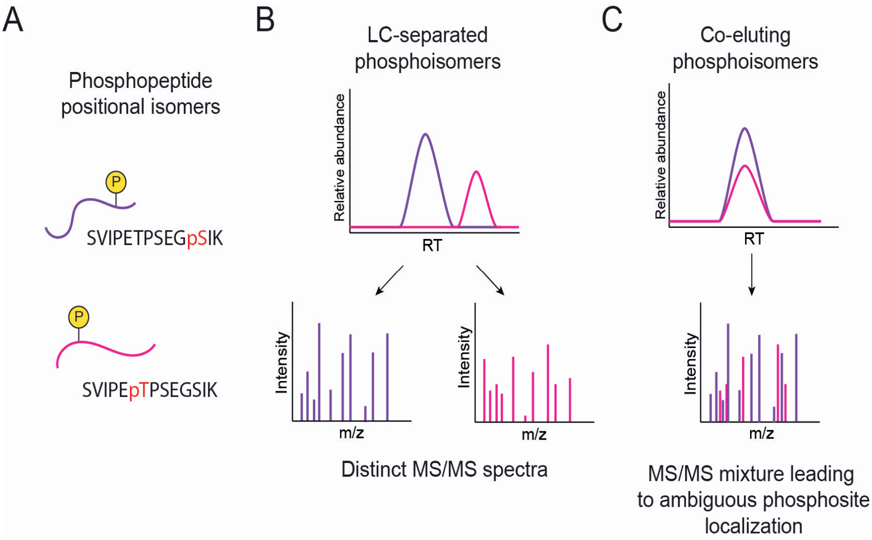


Figure 9. Measurement of phospho-isomers by LC-MS. A, diagram of two phosphopeptide positional isomers. B, isomers resolved by LC are fragmented separately, enabling accurate peptide sequencing and phosphosite localization. C, Co-eluting phospho-isomers cannot be separated at the MS1 level, hence they are fragmented together leading to a mixed MS/MS spectra which hampers accurate phosphosite localization. Modified from Courcelles *et al*, 2012 (*181*).

In addition to the technical hurdles described, the dynamic nature of protein phosphorylation further increases the analytical challenge of measuring this PTM. Phosphorylation is known to be highly dependent on stimulus and can change in very short time frames (*146*). Moreover, if changes in the overall proteome are not considered, up/down regulation of a phosphopeptide after a prolonged period of time could be caused by underlying changes in the protein levels (Figure 10A,B) (*187*). Furthermore, it is important to consider that due to the principle of matter conservation, down-regulation of mono-phosphorylated peptides could be caused by the increase of multi-phosphorylated forms or by the appearance of another PTM on the same peptide (Figure 10C,D) (*188*). All of the latter are to be taken into consideration when planning a phosphoproteomics experiment and during data analysis, and highlight the importance of studying protein phosphorylation in a time-resolved fashion (*189*). To achieve this, previous knowledge of the time-scale of the biological process being studied is key to gathering samples at relevant time points, which will permit to capture of true dynamical behavior (*190*).

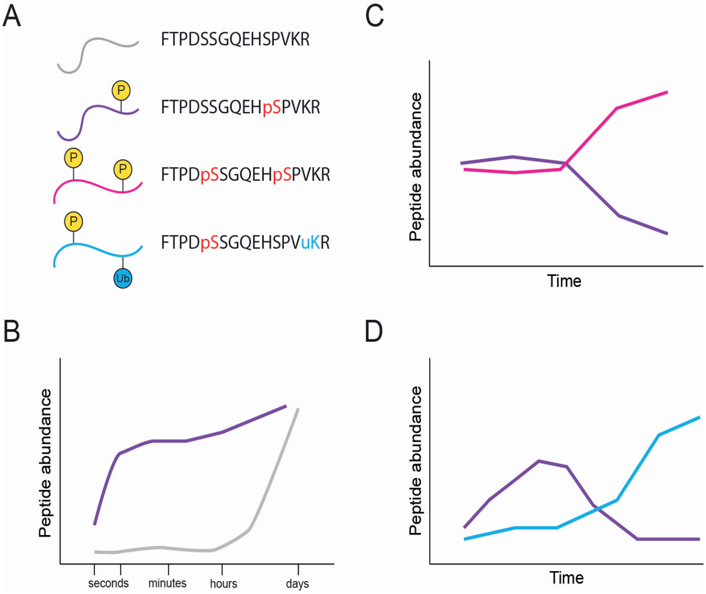


Figure 10. Potential ambiguity during data interpretation. A, diagrams of four different species originating from the same peptide backbone; unmodified peptide (gray), singly-phosphorylated peptide (purple), doubly-phosphorylated peptide (pink) and singly-phosphorylated and ubiquitinated peptide (light blue). B, increase in phosphorylated peptide after a prolonged time could be due to a global increase in the protein levels. C, downregulation of a singly-phosphorylated peptide could be due to an increase in the doubly-phosphorylated form of the same peptide. D, downregulation of a singly-phosphorylated peptide could be due to the appearance of another PTM in the peptide backbone. Modified from Solari *et al*, 2015 (*188*).

Time-resolved phosphoproteomics

After more than two decades of MS-based phosphoproteomics, the number of detected phosphorylation sites keeps increasing. Still, of all phosphosites reported in the database PhosphoSitePlus (*191*), only 2.8% have an assigned function. Several studies have shown that changes in protein phosphorylation through time strongly suggests functionality (*138*). Hence, the study of protein phosphorylation dynamics is of paramount importance to aid in the understanding of how this PTM regulates cellular processes (*189*).

Pioneering work by Blagoev and colleagues revealed the dynamics of phospho-tyrosine (p-Tyr) modification after treatment with epidermal growth factor (EGF). Affinity purification with an anti-p-Tyr specific antibody was used to purify proteins phosphorylated after stimulus. By collecting different time points, up to 20 minutes after treatment, the authors could show the dynamic nature of tyrosine phosphorylation (*192*). Interestingly, this study did not employ a phosphopeptide enrichment strategy, hence phosphopeptides were not measured and it was not possible to determine which sites were actually phosphorylated. A few years later, Olsen & Blagoev would incorporate SCX and TiO_2 chromatography for phosphopeptide enrichment, which enabled them to identify around 6,000 phosphosites, of which, close to 13% showed a dynamic trend after stimulus with EGF (*146*).

Several studies would continue to follow this trend and use MS-based technologies to study the dynamic behavior of phosphoproteomes in response to a variety of different stimuli and in different biological contexts (*144, 193-195*). For example, de Graaf and colleagues studied the effect of prostaglandin E2 on Jurkat T cells, which revealed a dynamic response comprised of phosphosites with very different behaviors. Moreover, they showed that this complex response is given by a variety of kinases, and not only by protein kinase A, as previously thought (*173*). Next, two other studies showed that phosphorylation dynamics can happen extremely fast, with a sub-minute response time, highlighting how this PTM permits fast signal transmission in the cellular milieu. Namely, the work by Kanshin & Bergeron-Sandoval showed that dynamic phosphosites, and especially those reacting early to osmotic stress (in less than 60 seconds), show a degree of conservation similar to that of known functional phosphosites in yeast, while static phosphosites tend to be less conserved (*196*). Next, the work by Humphrey and colleagues showed phosphorylation dynamics in the liver of rats after stimulus with insulin, which revealed that differential phosphorylation can be observed in an early (within seconds) and intermediate fashion (within minutes). Moreover, this work also showed the importance of reducing the lengthy workflow of phosphoproteomics experiments, by implementing sample preparation on a 96 well plate format which allowed sample processing at high-throughput (*197*).

Phosphoproteomics applied to the study of the cell cycle

The work described in the previous paragraph shows that many of the first applications of phosphoproteomics revolved around the study of signaling pathways stimulated after specific treatments (e.g. insulin, EGF, osmotic stress). Nevertheless, changes in protein phosphorylation are known to be the master regulator of probably one of the most vital cellular processes: cell cycle progression. Deregulation in cell cycle control has been linked to a variety of diseases, with cancer being the most notorious (*198*). The latter posed MS-based phosphoproteomics as a powerful tool to better understand how protein phosphorylation regulates the complex process in which a cell duplicates and divides its contents into two daughter cells.

Work done by Dephoure and colleagues showed that a wide range of proteins are phosphorylated during mitosis (>1,000), many of which are likely targets of cyclin dependent kinases (CDKs), due to the large proportion of CDK minimal consensus motifs detected (S/T-P). Interestingly, they also reported on the high number of proteins that are phosphorylated on multiple residues in mitosis, with Ki-67 showing >100 mitotically regulated sites (*199*). This protein was used as a model CDK target to validate some of the results shown in **chapter IV.** Complementary to the findings of Dephoure and colleagues, another group reported that many kinases are phosphorylated during mitosis, which supported the idea of cell cycle progression being strongly regulated by protein phosphorylation, very likely by the concerted activity of a variety of kinases (*200*).

Expanding on the previous discoveries, the Mann group analyzed the proteome and phosphoproteome of cells throughout different phases of the cell cycle. By combining cell synchronization techniques (e.g. thymidine block and nocodazole block), they managed to provide a large dataset of phosphoprotein dynamics during the cell cycle of HeLa S3 cells. One of their key findings is that up to 70% of all detected proteins seem to be phosphorylated, which is substantially larger than previous estimations (~30%). In addition, they show that there is a drastic change in phosphosite stoichiometry from interphase to mitosis, with mitotic sites showing high levels of site occupancy. Potential CDK sites displayed the highest mean occupancy (~88%), which is strongly suggestive of functionality (*201*).

The studies analyzing cell cycle phosphorylation mentioned thus far have generally used cells blocked at different stages of the cell cycle to generate "snapshots" of the phosphorylation landscape. A more recent study by Swaffer and colleagues used a similar approach to study phosphorylation dynamics in *Saccharomyces pombe* during cell cycle progression. Here, cells expressing an ATP-analogue sensitive CDK were arrested at G2, with subsequent release and sampling every 5-10 minutes. This revealed a progressive increase in CDK mediated phosphorylation as the cells progressed from interphase to mitosis (*202*). Still, this study also relied on cell synchronization, which is known to generate artefacts due to the mere perturbation of the cell cycle (*203, 204*). In addition, others have previously raised the issue of how studies on cell populations cannot provide the resolution needed to detect the real phosphorylation dynamics from single cells (*190*). In **chapter IV** we tackled these problems by taking advantage of the highly synchronous cell divisions of *Xenopus laevis* early embryos. By measuring phosphorylation dynamics at high temporal resolution on single embryos, we managed to measure phosphorylation dynamics in an eukaryote organism *in vivo.* Due to the high synchronization of the cell cycles in these early embryos, we think our measurements provide a close view of what protein phosphorylation dynamics in single cells look like.

References

1. J. J. Thomson, XL. Cathode Rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **44**, 293-316 (1897).
2. K. S. Sharma, Mass spectrometry—The early years. *International Journal of Mass Spectrometry* **349-350**, 3-8 (2013).

3. F. W. Aston, LXXIV. A positive ray spectrograph. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **38**, 707-714 (1919).
4. F. W. Aston, Isotopes and Atomic Weights. *Nature* **105**, 617-619 (1920).
5. J. Griffiths, A brief history of mass spectrometry. *Anal Chem* **80**, 5678-5683 (2008).
6. P. W. Beck, NUCLEAR ENERGY IN THE TWENTY-FIRST CENTURY: Examination of a Contentious Subject. *Annual Review of Energy and the Environment* **24**, 113-137 (1999).
7. F. W. McLafferty, A century of progress in molecular mass spectrometry. *Annu Rev Anal Chem (Palo Alto Calif)* **4**, 1-22 (2011).
8. L. Konermann, E. Ahadi, A. D. Rodriguez, S. Vahidi, Unraveling the mechanism of electrospray ionization. *Anal Chem* **85**, 2-9 (2013).
9. M. S. B. Munson, F. H. Field, Chemical Ionization Mass Spectrometry. I. General Introduction. *Journal of the American Chemical Society* **88**, 2621-2630 (1966).
10. M. A. Baldwin, F. W. McLafferty, Direct chemical ionization of relatively involatile samples. Application to underivatized oligopeptides. *Organic Mass Spectrometry* **7**, 1353-1356 (1973).
11. B. G. Levi, Chemistry Nobel Laureates Helped Develop Tools to Study Large Biological Molecules. *Physics Today* **55**, 19-22 (2002).
12. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
13. K. Tanaka *et al.*, Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **2**, 151-153 (1988).
14. M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **60**, 2299-2301 (1988).
15. Y. Pasing, S. Colnoe, T. Hansen, Proteomics of hydrophobic samples: Fast, robust and low-cost workflows for clinical approaches. *Proteomics* **17**, (2017).
16. M. Kuljanin, D. Z. Dieters-Castator, D. A. Hess, L. M. Postovit, G. A. Lajoie, Comparison of sample preparation techniques for large-scale proteomics. *Proteomics* **17**, (2017).
17. E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev* **32**, 453-465 (2013).
18. P. Giansanti, L. Tsiatsiani, T. Y. Low, A. J. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **11**, 993-1006 (2016).
19. Y. Zhang, B. R. Fonslow, B. Shan, M. C. Baek, J. R. Yates, 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **113**, 2343-2394 (2013).
20. J. C. Tran *et al.*, Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254-258 (2011).
21. T. K. Toby, L. Fornelli, N. L. Kelleher, Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual Review of Analytical Chemistry* **9**, 499-519 (2016).
22. N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324 (2014).
23. A. J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr* **499**, 177-196 (1990).
24. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
25. J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).
26. B. Domon, R. Aebersold, Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).

27. L. M. de Godoy *et al.*, Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* **7**, R50 (2006).
28. L. Wu, D. K. Han, Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Review of Proteomics* **3**, 611-619 (2006).
29. M. Gilar, P. Olivova, A. E. Daly, J. C. Gebler, Orthogonality of separation in two-dimensional liquid chromatography. *Anal Chem* **77**, 6426-6434 (2005).
30. M. S. Ritorto, K. Cook, K. Tyagi, P. G. Pedrioli, M. Trost, Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. *J Proteome Res* **12**, 2449-2457 (2013).
31. H. Kim *et al.*, An efficient method for high-pH peptide fractionation based on C18 StageTips for in-depth proteome profiling. *Analytical Methods* **11**, 4693-4698 (2019).
32. T. S. Batth, C. Francavilla, J. V. Olsen, Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* **13**, 6176-6186 (2014).
33. R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
34. S. Doll, A. L. Burlingame, Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol* **10**, 63-71 (2015).
35. M. W. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. Heck, Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **76**, 3935-3943 (2004).
36. M. Wilm, Principles of electrospray ionization. *Mol Cell Proteomics* **10**, M111 009407 (2011).
37. A. G. Marshall, C. L. Hendrickson, High-Resolution Mass Spectrometers. *Annual Review of Analytical Chemistry* **1**, 579-599 (2008).
38. J. Colinge, J. Magnin, T. Dessingy, M. Giron, A. Masselot, Improved peptide charge state assignment. *Proteomics* **3**, 1434-1440 (2003).
39. D. L. Tabb *et al.*, Determination of peptide and protein ion charge states by fourier transformation of isotope-resolved mass spectra. *Journal of the American Society for Mass Spectrometry* **17**, 903-915 (2006).
40. R. Aebersold, D. R. Goodlett, Mass Spectrometry in Proteomics. *Chemical Reviews* **101**, 269-296 (2001).
41. R. Aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
42. J. P. Savaryn, T. K. Toby, N. L. Kelleher, A researcher's guide to mass spectrometry-based proteomics. *Proteomics* **16**, 2435-2443 (2016).
43. A. Radionova, I. Filippov, P. J. Derrick, In pursuit of resolution in time-of-flight mass spectrometry: A historical perspective. *Mass Spectrometry Reviews* **35**, 738-757 (2016).
44. A. R. Buchberger, K. DeLaney, J. Johnson, L. Li, Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal Chem* **90**, 240-265 (2018).
45. J. Han *et al.*, Imaging of protein distribution in tissues using mass spectrometry: An interdisciplinary challenge. *TrAC Trends in Analytical Chemistry* **112**, 13-28 (2019).
46. K. Huber *et al.*, Approaching cellular resolution and reliable identification in mass spectrometry imaging of tryptic peptides. *Analytical and Bioanalytical Chemistry* **410**, 5825-5837 (2018).
47. R. A. Zubarev, A. Makarov, Orbitrap mass spectrometry. *Anal Chem* **85**, 5288-5296 (2013).
48. A. Makarov, Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **72**, 1156-1162 (2000).
49. A. Michalski *et al.*, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **10**, M111 011015 (2011).

50. R. A. Scheltema *et al.*, The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics* **13**, 3698-3708 (2014).
51. C. D. Kelstrup *et al.*, Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J Proteome Res* **13**, 6187-6195 (2014).
52. C. D. Kelstrup *et al.*, Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J Proteome Res* **17**, 727-738 (2018).
53. D. B. Bekker-Jensen *et al.*, A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol Cell Proteomics* **19**, 716-729 (2020).
54. A. Makarov *et al.*, Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry* **78**, 2113-2120 (2006).
55. J. R. Yates, D. Cociorva, L. Liao, V. Zabrouskov, Performance of a Linear Ion Trap-Orbitrap Hybrid for Peptide Analysis. *Analytical Chemistry* **78**, 493-500 (2006).
56. M. W. Senko *et al.*, Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Analytical Chemistry* **85**, 11710-11714 (2013).
57. R. A. Yost, C. G. Enke, Selected ion fragmentation with a tandem quadrupole mass spectrometer. *Journal of the American Chemical Society* **100**, 2274-2275 (1978).
58. R. A. Yost, C. G. Enke, Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Analytical Chemistry* **51**, 1251-1264 (1979).
59. V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4**, 222 (2008).
60. P. Picotti, R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9**, 555-566 (2012).
61. G. L. Andrews, B. L. Simons, J. B. Young, A. M. Hawkridge, D. C. Muddiman, Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Analytical Chemistry* **83**, 5442-5446 (2011).
62. D. F. Hunt, J. R. Yates, 3rd, J. Shabanowitz, S. Winston, C. R. Hauer, Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **83**, 6233-6237 (1986).
63. P. J. Boersema, S. Mohammed, A. J. Heck, Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* **44**, 861-878 (2009).
64. J. V. Olsen *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712 (2007).
65. C. K. Frese *et al.*, Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Journal of Proteome Research* **10**, 2377-2388 (2011).
66. N. Nagaraj, R. C. J. D’Souza, J. Cox, J. V. Olsen, M. Mann, Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. *Journal of Proteome Research* **9**, 6786-6794 (2010).
67. D. M. Waldera-Lupa, A. Stefanski, H. E. Meyer, K. Stühler, The fate of b-ions in the two worlds of collision-induced dissociation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 2843-2848 (2013).
68. P. Roepstorff, J. Fohlman, Letter to the editors. *Biomedical Mass Spectrometry* **11**, 601-601 (1984).
69. H. Marx *et al.*, A large synthetic peptide and phosphopeptide reference library for mass spectrometry–based proteomics. *Nature Biotechnology* **31**, 557-564 (2013).

70. A. Stensballe, O. N. Jensen, J. V. Olsen, K. F. Haselmann, R. A. Zubarev, Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom* **14**, 1793-1800 (2000).
71. J. Wiesner, T. Premsler, A. Sickmann, Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **8**, 4466-4483 (2008).
72. G. C. McAlister *et al.*, A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res* **7**, 3127-3136 (2008).
73. S. J. Pitteri, P. A. Chrisman, J. M. Hogan, S. A. McLuckey, Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2*. *Anal Chem* **77**, 1831-1839 (2005).
74. Y. Xia *et al.*, Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Anal Chem* **78**, 4146-4154 (2006).
75. J. E. P. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9528 (2004).
76. C. K. Frese *et al.*, Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **84**, 9668-9673 (2012).
77. C. K. Frese *et al.*, Unambiguous Phosphosite Localization using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *Journal of Proteome Research* **12**, 1520-1525 (2013).
78. L. C. Gillet, A. Leitner, R. Aebersold, Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem (Palo Alto Calif)* **9**, 449-472 (2016).
79. K. Hodge, S. T. Have, L. Hutton, A. I. Lamond, Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics* **88**, 92-103 (2013).
80. A. Michalski, J. Cox, M. Mann, More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *Journal of Proteome Research* **10**, 1785-1793 (2011).
81. N. W. Bateman *et al.*, Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol Cell Proteomics* **13**, 329-338 (2014).
82. A. I. Nesvizhskii, O. Vitek, R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* **4**, 787-797 (2007).
83. J. S. Cottrell, Protein identification using MS/MS data. *Journal of Proteomics* **74**, 1842-1851 (2011).
84. M. Y. Hein, K. Sharma, J. Cox, M. Mann, in *Handbook of Systems Biology,* A. J. M. Walhout, M. Vidal, J. Dekker, Eds. (Academic Press, San Diego, 2013), pp. 3-25.
85. A. I. Nesvizhskii, A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics* **73**, 2092-2123 (2010).
86. J. E. Elias, S. P. Gygi, in *Proteome Bioinformatics,* S. J. Hubbard, A. R. Jones, Eds. (Humana Press, Totowa, NJ, 2010), pp. 55-71.
87. K. A. Neilson *et al.*, Less label, more free: Approaches in label-free quantitative mass spectrometry. *PROTEOMICS* **11**, 535-553 (2011).
88. D. Chelius, P. V. Bondarenko, Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **1**, 317-323 (2002).
89. S. Nahnsen, C. Bielow, K. Reinert, O. Kohlbacher, Tools for Label-free Peptide Quantification *. *Molecular & Cellular Proteomics* **12**, 549-556 (2013).

90. J. Cox *et al.*, Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ * . *Molecular & Cellular Proteomics* **13**, 2513-2526 (2014).

91. S.-E. Ong *et al.*, Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics*. *Molecular & Cellular Proteomics* **1**, 376-386 (2002).

92. M. Krüger *et al.*, SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134**, 353-364 (2008).

93. P. L. Ross *et al.*, Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents*. *Molecular & Cellular Proteomics* **3**, 1154-1169 (2004).

94. A. Thompson *et al.*, Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* **75**, 1895-1904 (2003).

95. H. Liu, R. G. Sadygov, J. R. Yates, A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry* **76**, 4193-4201 (2004).

96. D. L. Tabb *et al.*, Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography−Tandem Mass Spectrometry. *Journal of Proteome Research* **9**, 761-776 (2010).

97. D. A. Stead *et al.*, Information quality in proteomics. *Briefings in Bioinformatics* **9**, 174-188 (2008).

98. J. Cox *et al.*, Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research* **10**, 1794-1805 (2011).

99. M. Y. Lim, J. A. Paulo, S. P. Gygi, Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model. *Journal of Proteome Research* **18**, 4020-4026 (2019).

100. Y. Karpievitch *et al.*, A statistical framework for protein quantitation in bottom-up MS-based proteomics. *Bioinformatics* **25**, 2028-2034 (2009).

101. C. Lazar, L. Gatto, M. Ferro, C. Bruley, T. Burger, Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research* **15**, 1116-1125 (2016).

102. L. Jin *et al.*, A comparative study of evaluating missing value imputation methods in label-free proteomics. *Scientific Reports* **11**, 1760 (2021).

103. L. C. Gillet *et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717 (2012).

104. C. Ludwig *et al.*, Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology* **14**, e8126 (2018).

105. A. Doerr, DIA mass spectrometry. *Nature Methods* **12**, 35-35 (2015).

106. C.-C. Tsou *et al.*, DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods* **12**, 258-264 (2015).

107. H. L. Röst *et al.*, OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology* **32**, 219-223 (2014).

108. K. Barkovits *et al.*, Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Molecular & cellular proteomics : MCP* **19**, 181-197 (2020).

109. D. B. Bekker-Jensen *et al.*, Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Communications* **11**, 787 (2020).

110. E. Borràs, E. Sabidó, What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *PROTEOMICS* **17**, 1700180 (2017).

111. U. Kusebauch *et al.*, Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **166**, 766-778 (2016).

112. J. A. Mead *et al.*, MRMaid, the web-based tool for designing multiple reaction monitoring (MRM) transitions. *Mol Cell Proteomics* **8**, 696-705 (2009).
113. C. A. Sherwood *et al.*, MaRiMba: a software application for spectral library-based MRM transition list assembly. *J Proteome Res* **8**, 4396-4405 (2009).
114. P. Picotti *et al.*, High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods* **7**, 43-46 (2010).
115. J. Stahl-Zeng *et al.*, High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites*. *Molecular & Cellular Proteomics* **6**, 1809-1817 (2007).
116. B. MacLean *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
117. A. C. Peterson, J. D. Russell, D. J. Bailey, M. S. Westphall, J. J. Coon, Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics *. *Molecular & Cellular Proteomics* **11**, 1475-1488 (2012).
118. S. Gallien *et al.*, Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer*. *Molecular & Cellular Proteomics* **11**, 1709-1723 (2012).
119. S. Gallien, A. Bourmaud, S. Y. Kim, B. Domon, Technical considerations for large-scale parallel reaction monitoring analysis. *Journal of Proteomics* **100**, 147-159 (2014).
120. A. Bourmaud, S. Gallien, B. Domon, Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *PROTEOMICS* **16**, 2146-2159 (2016).
121. S. Gallien, E. Duriez, K. Demeure, B. Domon, Selectivity of LC-MS/MS analysis: Implication for proteomics experiments. *Journal of Proteomics* **81**, 148-158 (2013).
122. A. Makarov, E. Denisov, O. Lange, S. Horning, Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer. *Journal of the American Society for Mass Spectrometry* **17**, 977-982 (2006).
123. H. Tsuchiya, K. Tanaka, Y. Saeki, The parallel reaction monitoring method contributes to a highly sensitive polyubiquitin chain quantification. *Biochemical and Biophysical Research Communications* **436**, 223-229 (2013).
124. R. T. Lawrence, B. C. Searle, A. Llovet, J. Villén, Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nature Methods* **13**, 431-434 (2016).
125. H.-J. Kim, D. Lin, H.-J. Lee, M. Li, D. C. Liebler, Quantitative Profiling of Protein Tyrosine Kinases in Human Cancer Cell Lines by Multiplexed Parallel Reaction Monitoring Assays *. *Molecular & Cellular Proteomics* **15**, 682-691 (2016).
126. G. E. Ronsein *et al.*, Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of Proteomics* **113**, 388-399 (2015).
127. Y. J. Kim *et al.*, Quantification of SAA1 and SAA2 in lung cancer plasma using the isotype-specific PRM assays. *PROTEOMICS* **15**, 3116-3125 (2015).
128. S. N. Thomas *et al.*, Multiplexed Targeted Mass Spectrometry-Based Assays for the Quantification of N-Linked Glycosite-Containing Peptides in Serum. *Analytical Chemistry* **87**, 10830-10838 (2015).
129. B. Schilling *et al.*, Multiplexed, Scheduled, High-Resolution Parallel Reaction Monitoring on a Full Scan QqTOF Instrument with Integrated Data-Dependent and Targeted Mass Spectrometric Workflows. *Analytical Chemistry* **87**, 10222-10229 (2015).
130. Y. L. Deribe, T. Pawson, I. Dikic, Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17**, 666-672 (2010).

131. M. Mann, O. N. Jensen, Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261 (2003).

132. J. A. Ubersax, J. E. Ferrell, Jr., Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8**, 530-541 (2007).

133. T. Hunter, Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**, 225-236 (1995).

134. T. Hunter, Why nature chose phosphate to modify proteins. *Philos Trans R Soc Lond B Biol Sci* **367**, 2513-2516 (2012).

135. P. Cohen, The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-130 (2002).

136. Samuel M. Pearlman, Z. Serber, James E. Ferrell, A Mechanism for the Evolution of Phosphorylation Sites. *Cell* **147**, 934-946 (2011).

137. L. N. Johnson, R. J. Lewis, Structural basis for control by phosphorylation. *Chem Rev* **101**, 2209-2242 (2001).

138. E. J. Needham, B. L. Parker, T. Burykin, D. E. James, S. J. Humphrey, Illuminating the dark phosphoproteome. *Sci Signal* **12**, (2019).

139. R. V. Swanson, L. A. Alex, M. I. Simon, Histidine and aspartate phosphorylation: two-component systems and the limits of homology. *Trends Biochem Sci* **19**, 485-490 (1994).

140. M. C. Pirrung, Histidine kinases and two-component signal transduction systems. *Chem Biol* **6**, R167-175 (1999).

141. D. Esser *et al.*, Protein phosphorylation and its role in archaeal signal transduction. *FEMS Microbiol Rev* **40**, 625-647 (2016).

142. C. Doerig, J. C. Rayner, A. Scherf, A. B. Tobin, Post-translational protein modifications in malaria parasites. *Nat Rev Microbiol* **13**, 160-172 (2015).

143. M. M. Alam *et al.*, Phosphoproteomics reveals malaria parasite Protein Kinase G as a signalling hub regulating egress and invasion. *Nat Commun* **6**, 7285 (2015).

144. B. M. Invergo *et al.*, Sub-minute Phosphoregulation of Cell Cycle Systems during Plasmodium Gamete Formation. *Cell Rep* **21**, 2017-2029 (2017).

145. D. F. Zielinska, F. Gnad, M. Jedrusik-Bode, J. R. Wisniewski, M. Mann, Caenorhabditis elegans has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J Proteome Res* **8**, 4039-4049 (2009).

146. J. V. Olsen *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).

147. K. Sharma *et al.*, Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep* **8**, 1583-1594 (2014).

148. P. Blume-Jensen, T. Hunter, Oncogenic kinase signalling. *Nature* **411**, 355-365 (2001).

149. K. Rikova *et al.*, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190-1203 (2007).

150. I. Grundke-Iqbal *et al.*, Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci U S A* **83**, 4913-4917 (1986).

151. J. Eidenmuller *et al.*, Phosphorylation-mimicking glutamate clusters in the proline-rich region are sufficient to simulate the functional deficiencies of hyperphosphorylated tau protein. *Biochem J* **357**, 759-767 (2001).

152. S. S. Taylor, M. M. Keshwani, J. M. Steichen, A. P. Kornev, Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 2517-2528 (2012).

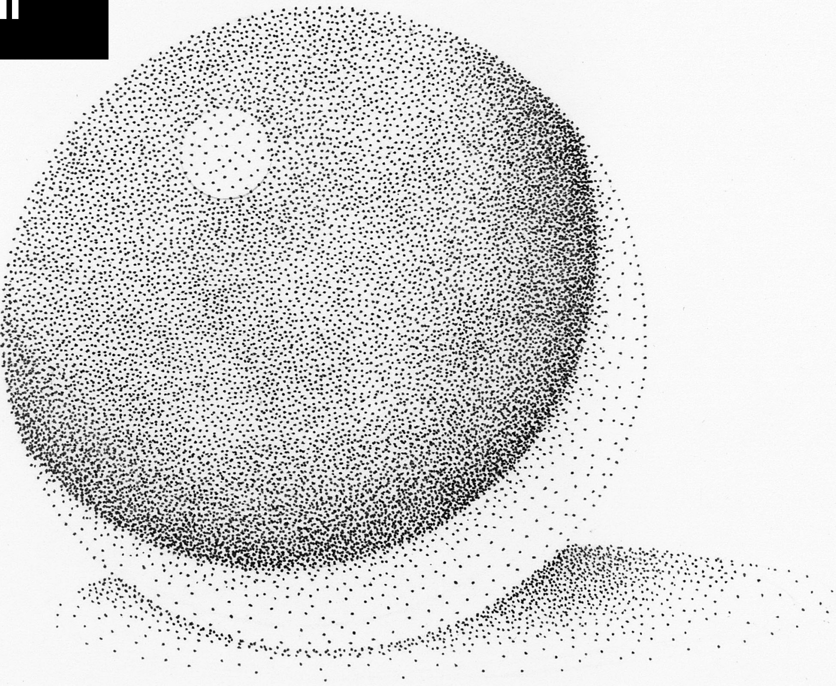
153. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).

154. J. E. Hutti *et al.*, A rapid method for determining protein kinase phosphorylation specificity. *Nature Methods* **1**, 27-29 (2004).
155. M. B. Yaffe *et al.*, A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nature Biotechnology* **19**, 348-353 (2001).
156. Z. Songyang *et al.*, Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Current Biology* **4**, 973-982 (1994).
157. R. M. Biondi, A. R. Nebreda, Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *Biochemical Journal* **372**, 1-13 (2003).
158. T. Tanoue, M. Adachi, T. Moriguchi, E. Nishida, A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nature Cell Biology* **2**, 110-116 (2000).
159. M. E. Miller, F. R. Cross, Cyclin specificity: how many wheels do you need on a unicycle? *Journal of Cell Science* **114**, 1811-1820 (2001).
160. K.-Y. Cheng *et al.*, The Role of the Phospho-CDK2/Cyclin A Recruitment Site in Substrate Recognition *. *Journal of Biological Chemistry* **281**, 23167-23179 (2006).
161. R. P. Bhattacharyya, A. Reményi, B. J. Yeh, W. A. Lim, Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annual Review of Biochemistry* **75**, 655-680 (2006).
162. T. Pawson, D. Scott John, Signaling Through Scaffold, Anchoring, and Adaptor Proteins. *Science* **278**, 2075-2080 (1997).
163. F. Witzel, L. Maddison, N. Blüthgen, How scaffolds shape MAPK signaling: what we know and opportunities for systems approaches. *Frontiers in Physiology* **3**, (2012).
164. I. Nasa, A. N. Kettenbach, Coordination of Protein Kinase and Phosphoprotein Phosphatase Activities in Mitosis. *Frontiers in Cell and Developmental Biology* **6**, (2018).
165. J. Chen Mark, E. Dixon Jack, G. Manning, Genomics and evolution of protein phosphatases. *Science Signaling* **10**, eaag1796 (2017).
166. Y. Shi, Serine/Threonine Phosphatases: Mechanism through Structure. *Cell* **139**, 468-484 (2009).
167. J. Rush *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnology* **23**, 94-101 (2005).
168. J. Villén, S. P. Gygi, The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **3**, 1630-1638 (2008).
169. B. A. Ballif, J. Villen, S. A. Beausoleil, D. Schwartz, S. P. Gygi, Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics* **3**, 1093-1101 (2004).
170. S. B. Ficarro *et al.*, Online Nanoflow Multidimensional Fractionation for High Efficiency Phosphopeptide Analysis*. *Molecular & Cellular Proteomics* **10**, O111.011064 (2011).
171. M. Dong *et al.*, Depletion of Acidic Phosphopeptides by SAX To Improve the Coverage for the Detection of Basophilic Kinase Substrates. *Journal of Proteome Research* **11**, 4673-4681 (2012).
172. H. Zhou *et al.*, Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment *. *Molecular & Cellular Proteomics* **10**, (2011).
173. E. L. de Graaf, P. Giansanti, A. F. M. Altelaar, A. J. R. Heck, Single-step Enrichment by Ti4+-IMAC and Label-free Quantitation Enables In-depth Monitoring of Phosphorylation Dynamics with High Reproducibility and Temporal Resolution*. *Molecular & Cellular Proteomics* **13**, 2426-2434 (2014).
174. H. Zhou *et al.*, Specific Phosphopeptide Enrichment with Immobilized Titanium Ion Affinity Chromatography Adsorbent for Phosphoproteome Analysis. *Journal of Proteome Research* **7**, 3957-3967 (2008).
175. H. Zhou *et al.*, Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nature Protocols* **8**, 461-480 (2013).

176. H. Post *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **16**, 728-737 (2017).
177. S. A. Beausoleil, J. Villén, S. A. Gerber, J. Rush, S. P. Gygi, A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature Biotechnology* **24**, 1285-1292 (2006).
178. B. E. Ruttenberg, T. Pisitkun, M. A. Knepper, J. D. Hoffert, PhosphoScore: An Open-Source Phosphorylation Site Assignment Tool for MSn Data. *Journal of Proteome Research* **7**, 3054-3059 (2008).
179. T. E. Thingholm, O. N. Jensen, M. R. Larsen, Analytical strategies for phosphoproteomics. *PROTEOMICS* **9**, 1451-1468 (2009).
180. A. M. Palumbo *et al.*, Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass Spectrometry Reviews* **30**, 600-625 (2011).
181. M. Courcelles, G. Bridon, S. Lemieux, P. Thibault, Occurrence and Detection of Phosphopeptide Isomers in Large-Scale Phosphoproteomics Experiments. *Journal of Proteome Research* **11**, 3753-3765 (2012).
182. L. Otvos, I. A. Tangoren, K. Wroblewski, M. Hollosi, V. M. Y. Lee, Reversed-phase high-performance liquid chromatographic separation of synthetic phosphopeptide isomers. *Journal of Chromatography A* **512**, 265-272 (1990).
183. T. Schmidlin *et al.*, High-Throughput Assessment of Kinome-wide Activation States. *Cell Syst* **9**, 366-374 e365 (2019).
184. T. Schmidlin *et al.*, Assessment of SRM, MRM3, and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *PROTEOMICS* **16**, 2193-2205 (2016).
185. M. P. Y. Lam *et al.*, An MRM-based workflow for quantifying cardiac mitochondrial protein phosphorylation in murine and human tissue. *Journal of Proteomics* **75**, 4602-4609 (2012).
186. M. P. Y. Lam *et al.*, Site-specific quantitative analysis of cardiac mitochondrial protein phosphorylation. *Journal of Proteomics* **81**, 15-23 (2013).
187. R. Wu *et al.*, Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes*. *Molecular & Cellular Proteomics* **10**, M111.009654 (2011).
188. F. A. Solari, M. Dell'Aica, A. Sickmann, R. P. Zahedi, Why phosphoproteomics is still a challenge. *Mol Biosyst* **11**, 1487-1493 (2015).
189. A. F. M. Altelaar, J. Munoz, A. J. R. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics* **14**, 35-48 (2013).
190. J. E. Purvis, G. Lahav, Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945-956 (2013).
191. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520 (2015).
192. B. Blagoev, S.-E. Ong, I. Kratchmarova, M. Mann, Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnology* **22**, 1139-1145 (2004).
193. T. S. Batth *et al.*, Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling. *Cell Reports* **22**, 2784-2796 (2018).
194. T. G. Rigbolt Kristoffer *et al.*, System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation. *Science Signaling* **4**, rs3-rs3 (2011).
195. Sean J. Humphrey *et al.*, Dynamic Adipocyte Phosphoproteome Reveals that Akt Directly Regulates mTORC2. *Cell Metabolism* **17**, 1009-1020 (2013).

196. E. Kanshin, L.-P. Bergeron-Sandoval, S. S. Isik, P. Thibault, Stephen W. Michnick, A Cell-Signaling Network Temporally Resolves Specific versus Promiscuous Phosphorylation. *Cell Reports* **10**, 1202-1214 (2015).
197. S. J. Humphrey, S. B. Azimifar, M. Mann, High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nature Biotechnology* **33**, 990-995 (2015).
198. M. Malumbres, M. Barbacid, Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer* **9**, 153-166 (2009).
199. N. Dephoure *et al.*, A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences* **105**, 10762 (2008).
200. H. Daub *et al.*, Kinase-Selective Enrichment Enables Quantitative Phosphoproteomics of the Kinome across the Cell Cycle. *Molecular Cell* **31**, 438-448 (2008).
201. V. Olsen Jesper *et al.*, Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Science Signaling* **3**, ra3-ra3 (2010).
202. M. P. Swaffer, A. W. Jones, H. R. Flynn, A. P. Snijders, P. Nurse, CDK Substrate Phosphorylation and Ordering the Cell Cycle. *Cell* **167**, 1750-1761 e1716 (2016).
203. S. Cooper, The synchronization manifesto: a critique of whole-culture synchronization. *The FEBS Journal* **286**, 4650-4656 (2019).
204. T. Ly, A. Endo, A. I. Lamond, Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**, e04534 (2015).

CHAPTER II



Capturing the signalling dynamics of the MAPK-AKT-mTOR pathway in a single targeted phosphoproteomics assay

Juan Manuel Valverde[1,2†], Donna O. Debets[1,2†], Maarten Altelaar[1,2]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, 3584 CH Utrecht, Netherlands.
[2]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, Netherlands.

[†]Equal contributions

Manuscript under review

Abstract

The MAPK-AKT-mTOR protein network integrates extra- and intracellular signals to determine cellular fate, regulating pivotal biological processes such as cell growth and metabolism. Due to this crucial role, pathway dysregulation has been implicated in multiple diseases, such as metabolic disorders and cancer. The MAPK-AKT-mTOR pathway consists of dozens of proteins and signal transduction is primarily driven by protein phosphorylation. Here, we present a targeted phosphoproteomics assay to study the phosphorylation dynamics of the MAPK-AKT-mTOR pathway in detail with high sensitivity and in a high throughput manner. By using a multi-protease approach, we increased the pathway coverage with phosphosites that were previously inaccessible. This novel approach yields the most comprehensive method for the detailed study of mTOR signalling to date (covering 150 phosphopeptides on more than 70 phosphoproteins), which can be applied to *in vitro* and *in vivo* systems and has the sensitivity to be compatible with small sample amounts. We demonstrate the feasibility of this assay to monitor the plasticity of MAPK-AKT-mTOR phosphorylation dynamics in response to cellular stimuli with high temporal resolution and amino acid residue specificity. We found highly dynamic phosphorylation events upon treatment with growth factors, revealing the sequential nature of phosphosites in this signalling pathway. Furthermore, starvation of glucose and amino acids showed upregulation of AKT-targets PRAS40^{T246} and FOXO3^{T32}, highlighting the role of AKT in cellular response to starvation. These findings illustrate the potential of this assay to obtain new biological insight when monitoring dynamics of functional phosphosites.

Introduction

Cells continuously sense extra- and intracellular stimuli to regulate cell growth and metabolism. Under healthy circumstances, proliferation is stimulated when conditions are favourable and halted under stress. To achieve this, a wide variety of stimuli, from growth factors and nutrients to DNA damage, are sensed and integrated by signalling pathways to determine cellular fate (*1*). These pathways are driven by protein phosphorylation, which allows for fast and dynamic transmission of information throughout the cell (*2, 3*)

In eukaryotes, mitogen-activated protein kinase (MAPK) cascades are key players to relay information from growth factor stimuli at the cell surface to the interior of the cell (*4*). The kinases in these pathways have been heavily studied and are known to be key regulators of processes ranging from embryonic development to cell proliferation and growth (*5-7*). In addition to MAPKs, the phosphoinositide 3-kinase (PI3K)-AKT signalling network is another major integrator of growth factor stimuli, like insulin and cytokines (*8*), making it another essential regulator of cell fate. Hence, proteins of both of these pathways are often found mutated in malignancies (*9*). Interestingly, these pathways converge into two protein complexes, mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2).

Both mTORC1 and mTORC2 contain the mTOR serine/threonine protein kinase, which functions as the catalytic subunit responsible for phosphorylation of other protein kinases within the signalling cascade and effector proteins that directly regulate cell metabolism (*10*). Besides the mTOR kinase, each complex has additional subunits that function as either enhancers or inhibitors of its kinase activity. Some of these subunits are directly regulated by MAPK and PI3K-AKT activity (*11*). The mTOR network controls processes such as ribosome biogenesis, mRNA translation, cytoskeleton organisation and both catabolic and anabolic programs, like autophagy and lipid biosynthesis, respectively (*1, 10*). Together, the MAPK-AKT-mTOR signalling network converts extracellular signals (mainly via MAPK and AKT signalling) and intracellular environmental cues (*e.g.* DNA damage and ATP levels) to determine cellular fate.

The MAPK-AKT-mTOR pathway is composed of dozens of proteins, which transmit signals via phosphorylation of specific amino acid residues on target proteins. Therefore, to elucidate the complex regulation of this intricate signalling network, analytical methods to monitor the dynamics of these phosphosites are required. Protein phosphorylation events can be studied using antibody-based techniques, such as western blotting. These are generally very sensitive, yet their use is restricted by the availability of adequate phosphosite-specific antibodies. Furthermore, western blots are merely semi-quantitative and are difficult to multiplex. These limitations can be overcome by the application of targeted mass spectrometry (MS), namely selected reaction monitoring (SRM) (*12*). By selectively measuring peptides and peptide fragments of interest, this technique can monitor phosphosites with amino acid residue specificity (*13*). Furthermore, combined with highly selective phosphopeptide enrichment, SRM provides high sensitivity.

Quantitative reproducibility is ensured using heavy labelled standards. Targeted MS methods are easily multiplexed and allow for hundreds of phosphopeptides to be measured in a single MS run.

Here, we developed a targeted MS assay to study the phosphorylation dynamics of the MAPK-AKT-mTOR pathway in a high throughput manner, selecting phosphorylation sites with annotated biological function. We make use of multiple proteases for protein digestion, which increases the pathway coverage with phosphosites that were previously inaccessible. This novel approach yields the most comprehensive method for the detailed study of MAPK-AKT-mTOR signalling to date, which can be applied to *in vitro* and *in vivo* human samples and has the sensitivity to be compatible with small amounts of starting material. In a proof-of-principle experiment, we demonstrate that we can monitor the great plasticity of MAPK- AKT-mTOR phosphorylation dynamics with high temporal resolution and amino acid residue specificity. We show unique phosphorylation patterns for different stimuli and find sequential phosphorylation events that correlate with their known functionality. Overall, we think this assay could be a great resource for scientists who are interested in obtaining mechanistic insight in how phosphorylation via the MAPK- AKT-mTOR network controls cell growth and metabolism.

Results

Selection of phosphopeptides for assay development

We first performed a literature search to identify the main phosphoproteins in the MAPK-AKT-mTOR pathway and subsequently selected the principal phosphosites on these proteins, focussing on sites with known functionality. Next, we aimed to choose the most suitable phosphopeptides for LC-MS/MS analysis for each phosphosite of interest (Figure 1A, top). We therefore generated an extensive phosphopeptide library using BJ-EHT fibroblast cells transduced with an inducible RAS mutation. Induction of RAS activity in this cell line results in hyperphosphorylation of the MAPK pathway and hence enhanced our ability to detect phosphosites of interest. To extend the coverage of the phosphoproteome beyond sites accessible by trypsin digestion, proteins were digested in parallel by four proteases: Trypsin/LysC, AspN, GluC or Chymotrypsin. The generated peptides were fractionated by high pH fractionation prior to phosphopeptide enrichment and LC-MS/MS analysis (Figure 1A, bottom).

In total, our phosphopeptide library consisted of 46,080 unique phosphosites on 9,065 phosphoproteins (Figure S1). Importantly, each proteolytic enzyme generated peptides rendering unique phosphosites and phosphoproteins that were not identified by any of the other proteases. Protein digestion with Trypsin/LysC generated the highest number of (unique) phosphosites, followed by GluC, AspN and Chymotrypsin (Figure 1B).

To generate our targeted phosphoproteomics assay, we extracted the biologically relevant phosphosites from our phosphopeptide library. Since merely three phosphosites of interest were uniquely identified with AspN and Chymotrypsin, these proteases were excluded for further use.

Proteolytic digestion with Trypsin/LysC and GluC contrarily generated unique phosphopeptides covering many more phosphosites within the pathway. For example, FOXO3^{T32} was accessible with GluC digestion, whereas tryptic digestion did not generate a peptide suitable for LC-MS/MS analysis (Figure 1C). Additionally, the use of multiple proteases allowed us to select peptides with advantageous physiochemical properties. Phosphopeptides containing a methionine were disfavoured since methionine oxidation can occur *in vitro*, complicating method development and potentially reducing sensitivity since the signal is split over two peaks (Figure 1C). Lastly, we took advantage of the differences in detectability of specific phosphopeptides generated by the two proteases. Endogenous EIF4B^{S406} was detected at a much higher intensity (13x) after GluC digestion compared to Trypsin/LysC. This quantitative benefit resulted in increased sensitivity of our targeted phosphoproteomics assay (Figure 1D). Overall, the application of multiple proteases for protein digestion increased the coverage of biologically relevant phosphosites and allowed for the selection of peptides with suitable chemical and quantitative properties.

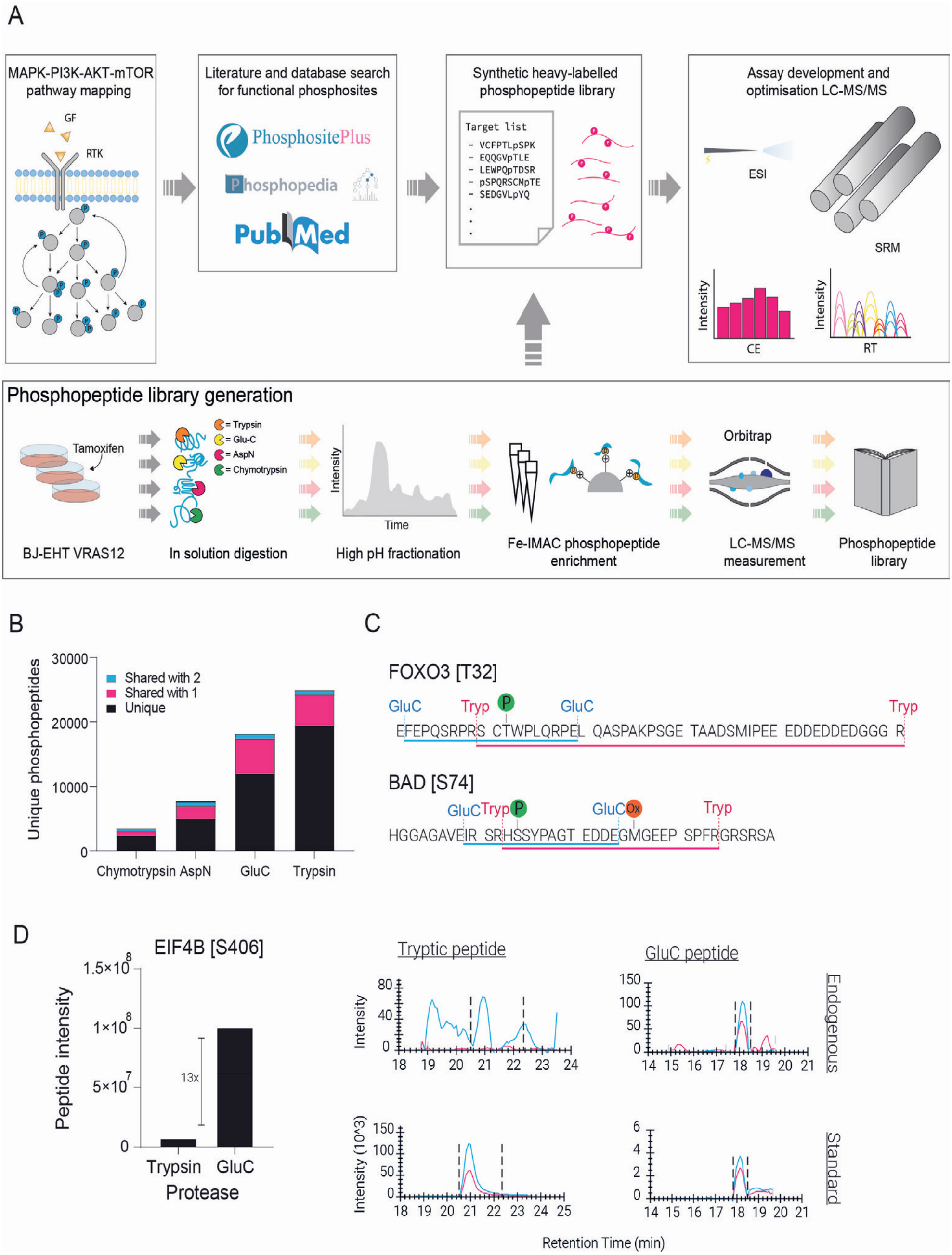


Figure 1. Development of targeted phosphoproteomics assay and multi-protease approach. (A) Top: Mapping of proteins in the MAPK-AKT-mTOR network was done by comprehensive literature mining. Functional phosphosites were selected and information on the corresponding phosphopeptides was obtained from our phosphopeptide library (bottom panel) combined with Phosphopedia. **Legend continues on next page.**

Figure 1 legend continued. The selected sequences were synthesised, incorporating heavy labelled amino acids, and used as internal standards and for assay development. Collision energy (CE) was optimised for each transition individually to increase sensitivity. Retention time scheduling was performed to measure hundreds of peptides in one MS run. (A) Bottom: An in-house phosphopeptide library was created using cells with inducible RAS activity, which were subjected to lysis and protein digestion with multiple enzymes in parallel (Trypsin/LysC, GluC, AspN and Chymotrypsin). High pH fractionation was used to maximize phosphopeptide detection, followed by Fe-IMAC enrichment and MS acquisition. The generated phosphopeptide library was mined for phosphopeptides of interest within the MAPK-AKT-mTOR pathway. (B) Number of phosphopeptides per protease. (C) Example of a phosphosite that was less favourable by trypsin digestion (top): tryptic peptide for FOXO3^{T32} was too long for efficient MS detection. Example of a phosphosite with favourable peptide properties using GluC for digestion (bottom): The GluC peptide for BAD^{S74} lacks the methionine, which is prone to oxidation during sample preparation. (D) GluC digestion improved detection of endogenous EIF4B^{S406}. Higher intensity for endogenous EIF4B^{S406} by GluC compared to trypsin digestion was found in our peptide library (left). This resulted in higher sensitivity in the SRM assay (right).

Assay development

In total, we included 150 phosphopeptides covering 72 phosphoproteins in the MAPK-AKT-mTOR pathway, covering many of the biologically relevant phosphosites within this signalling network (Figure 2A). All selected phosphopeptides were synthesised incorporating a heavy isotopic label and were used as reference standard. We observed good sensitivity and a large linear range of quantification as demonstrated by dilution series of our synthetic peptides (Figure S2A). The quantitative reproducibility was further determined by replicate measurements, which showed good coefficient of variation (Figure S2B). To measure all phosphopeptides in a single LC-MS/MS run, we optimised retention time scheduling using a dimensionless value that reflects chromatographic retention relative to a mix of commercial standards (iRTs) (*14*). Therefore, robust LC separation was critical (Figure S2C).

In brief: for our targeted phosphoproteomics assay, first cells were lysed and proteins were digested in parallel with either GluC or Trypsin/LysC. GluC and Tryptic peptides were pooled prior to automated sample clean-up. Heavy standards were added prior to phosphopeptide enrichment, and the enriched samples were measured on a nanoLC coupled to a triple-quadrupole MS (Figure 2B).

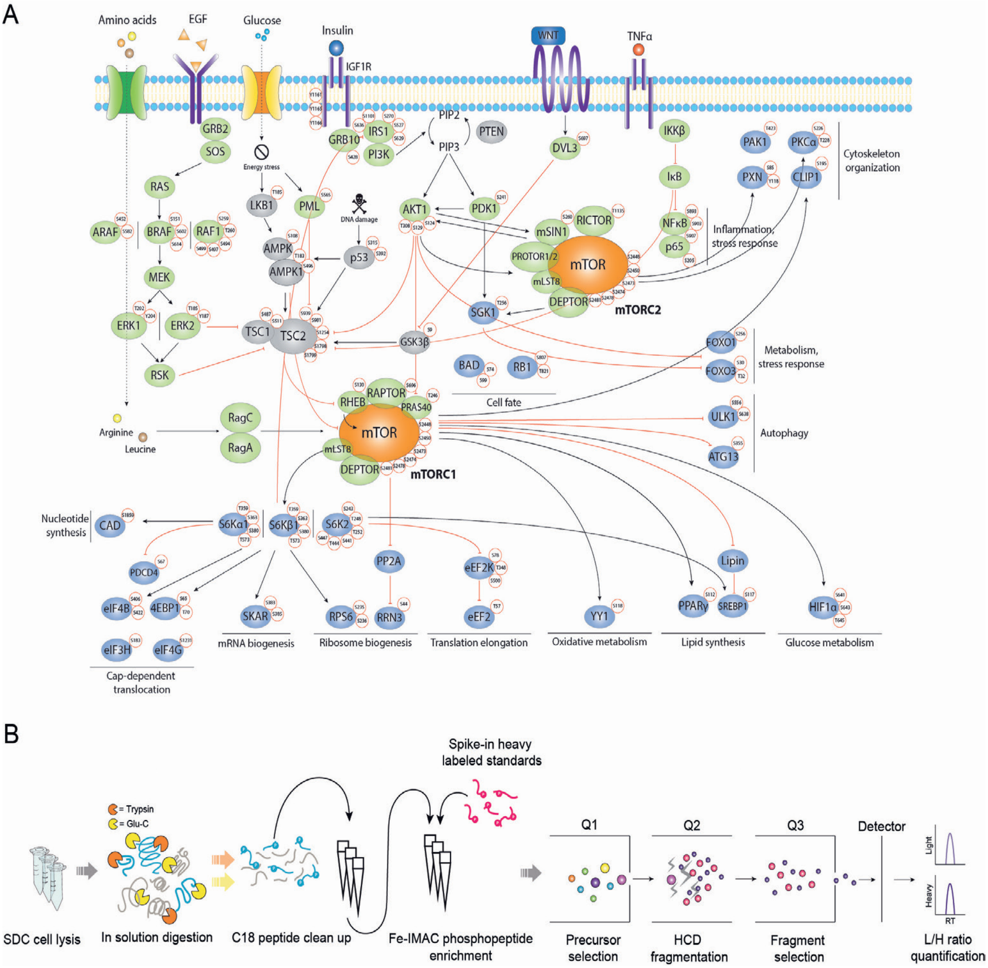


Figure 2. Targeted phosphoproteomics assay extensively covers the biologically relevant phosphosites in the MAPK-AKT-mTOR pathway. (A) Comprehensive view of the protein network, in red circles the phosphosites included in our assay. Green nodes represent activators of mTOR kinase activity, grey nodes represent inhibitors, blue nodes are effector proteins. Black arrows suggest activation, red arrows indicate inhibition. (B) Targeted assay workflow.

MAPK-AKT-mTOR phosphorylation dynamics

To assess the performance of our targeted phosphoproteomics assay, we analysed the MAPK-AKT-mTOR phosphorylation dynamics of MCF10A cells stimulated with EGF and insulin. Cells were starved for 24 hours before stimulation with EGF and insulin, and samples were harvested after 24h starvation and 1 min, 30 min and 60 min after stimulation (Figure 3A). We also assessed MAPK-AKT-mTOR phosphorylation status under normal growth conditions (24 hours in fully supplemented media). Cell growth was halted under starvation, but rescued upon addition of EGF and insulin (Figure S3). In total, we quantified 55 phosphorylation events across the different conditions. These phosphorylation events displayed distinct trends, grouping phosphopeptides into five different clusters (Figure 3B). The first three clusters represented phosphosites that were downregulated during normal growth conditions and starvation but were clearly responsive to addition of EGF and insulin after starvation. Within these three clusters, we could discern phosphorylation events that were instantly upregulated after growth factor addition (within a minute) and swiftly downregulated (cluster 1, 'early response'), behaving almost in a switch-like fashion. Furthermore, we identified a group of phosphosites that were quickly upregulated and remained upregulated over a longer period (cluster 2, 'sustained response'). Lastly, cluster 3 represented phosphorylation events that were upregulated upon growth factor addition at a later stage (30-60 min) and remained upregulated during the time course of the experiment ('delayed response'). These data highlight the feasibility of our assay to detect dynamic phosphorylation events upon stimulus.

In contrast to cluster 1, 2 and 3, the phosphorylation events in clusters 4 and 5 were not responsive to addition of EGF and insulin. Interestingly, the phosphosites in cluster 4 were upregulated during normal growth conditions (full media), decreased upon starvation and did not recover after replenishing with EGF and insulin (Figure 3B). Multiple phosphosites in this cluster were directly involved in protein translation, namely EIF4B^{S406}, EIF3H^{S183} and 4EBP1^{S65}. EIF4B and EIF3H are crucial components of the translation initiation complex, and their functionality is phosphorylation-driven; increased phosphorylation is positively correlated with induced translation (*15, 16*). Furthermore, the hypophosphorylated form of 4EBP1 represses translation initiation by binding to initiation factor EIF4E and hence inhibiting its assembly into the translation initiation complex (*17*). Downregulation of these three phosphosites results in reduced translation, which is anticipated upon starvation. Interestingly, phosphorylation translation initiation factor eIF4G1^{S1231} was found upregulated in cluster 3 ('late response'). This phosphosite is known to be regulated by ERK and correlates with formation of the initiation complex (*18*). Overall, our data indicate that addition of EGF and insulin after prolonged starvation did not restore protein translation within this relatively short timeframe (60 min).

Next, we set out to establish whether we could identify any kinase-substrate pairs from the dynamic phosphorylation events in cluster 1 ('early response'), cluster 2 ('sustained response') and cluster 3 ('delayed response'). Indeed, in cluster 1 we identified phosphorylation of mitogen-activated protein kinases, namely BRAF^{S121}**,** ERK1$^{T202/Y204}$ and ERK2$^{T185/Y187}$, these kinases are known to be readily activated upon growth factor stimulation. Phosphorylation of BRAF^{S121} by ERK activates a negative feedback loop inhibiting the interaction between RAS and BRAF, resulting in reduced phosphorylation of ERK (*19*). This could explain the transient phosphorylation of ERK we observed (Figure 3C).

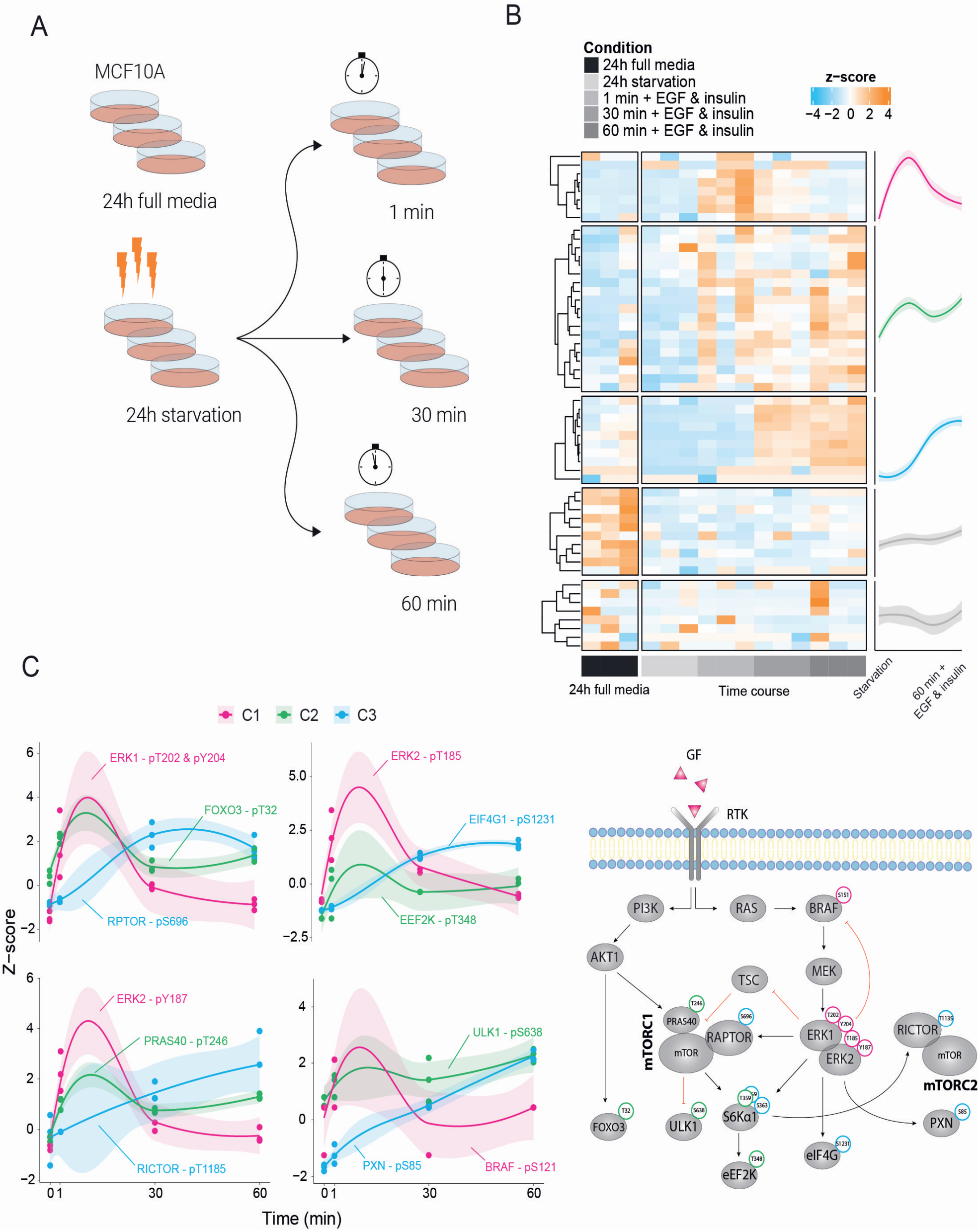


Figure 3. Probing MAPK-AKT-mTOR activity after growth factor stimuli. (A) Experimental setup: MCF10A cells were grown under normal conditions for 24 hours, followed by starvation in additive-free medium for 24 hours, after which cells were stimulated with EGF and insulin. Cells were harvested at indicated time points. (B) Unsupervised clustering of quantified phosphopeptides highlights five clusters with distinct phosphorylation patterns. (C) Profile plots of the dynamics of regulated phosphosites (left) correlating with suggested cascade of signaling events (right). Colour represents the clusters in Figure 3B.

Increased ERK activity directly following stimulus (cluster 1), resulted in phosphorylation of ERK substrates at later time points (cluster 3) (Figure 3C). Amongst these was RPTOR^{S696}, which upon upregulation activates mTORC1, resulting in further downstream phosphorylation of kinase S6KA1$^{T359/S363}$ (Figure 3C). S6KA1 has been reported to phosphorylate RICTOR^{T1135}, bridging the mTORC1 pathway with mTORC2 signalling. These phosphorylation dynamics are in concordance with the cascade of phosphorylation events previously reported in literature (Figure 3C).

In cluster 2, we identified phosphorylation events indicative of AKT activation upon insulin stimulation: upregulation of FOXO3^{T32} and PRAS40^{T246} (Figure 3C) (*20-22*). FOXO3 triggers apoptosis in the absence of survival factors, whereas upon phosphorylation by AKT, FOXO3 is retained in the cytosol, restricting its pro-apoptotic function (*23*). PRAS40 on the other hand is a negative regulator of mTORC1. Upon phosphorylation, PRAS40 dissociates from mTORC1, activating mTOR signalling (*24*). The increased and lasting upregulation of FOXO3 and PRAS40 phosphorylation upon growth factor stimulation found in this experiment shows activation of the PI3K-AKT pathway resulting in reduced pro-apoptotic signals and increased mTORC1 activity. This heightened mTOR activity can subsequently lead to phosphorylation of its downstream substrate, the kinase ULK1^{S638} (Figure 3C). ULK1 activation upon starvation initiates an autophagy response, whereas phosphorylation of ULK1 by mTOR results in inhibition of autophagy when nutrients are replenished (*25*). Our data suggests that activation of mTORC1 upon EGF and insulin stimulation also results in ULK^{S638} phosphorylation. Together, these results highlight the feasibility of our assay to disentangle patterns of sequential phosphorylation within the MAPK-AKT-mTOR pathway.

After highlighting phosphorylation dynamics caused by stimulus with growth factors, we set out to assess the effect of other nutrients known to modulate MAPK-AKT-mTOR activity. For this, we starved cells of either glucose or amino acids, using nutrient free media. Subsequently, we reincorporated each nutrient and collected samples at different time points. Since these experiments involved complete exchanges of media, we performed a control experiment changing media with the same composition and collected samples at the same time points (Figure 4A). Overall, we quantified 39 and 49 phosphopeptides upon stimulation with amino acids and glucose, respectively. Stimulus with these nutrients did not reveal a strong dynamic phosphorylation response, in clear contrast with growth factor treatment observed before (Figure 3B). Interestingly, glucose and amino acid deprivation caused increased phosphorylation of PRAS40^{T246} and FOXO3^{T32}, which were both downregulated upon reincorporation of each nutrient. As mentioned previously, both phosphosites were clearly upregulated after treatment with growth factors (Figure 4B). If these two sites are used as proxy of AKT activity, our results suggest that nutrient starvation causes upregulation of PI3K-AKT pathway. Others have previously suggested that upon depletion of glucose and amino acids cells engage on a survival program by increasing the activity of the MAPK-PI3K-AKT pathways, in order to ensure mTOR activation and promote cell survival (Figure 4C) (*26*). Overall, these results show that our assay can be used to discriminate differences in MAPK-AKT-mTOR signalling upon treatment with different stimuli.

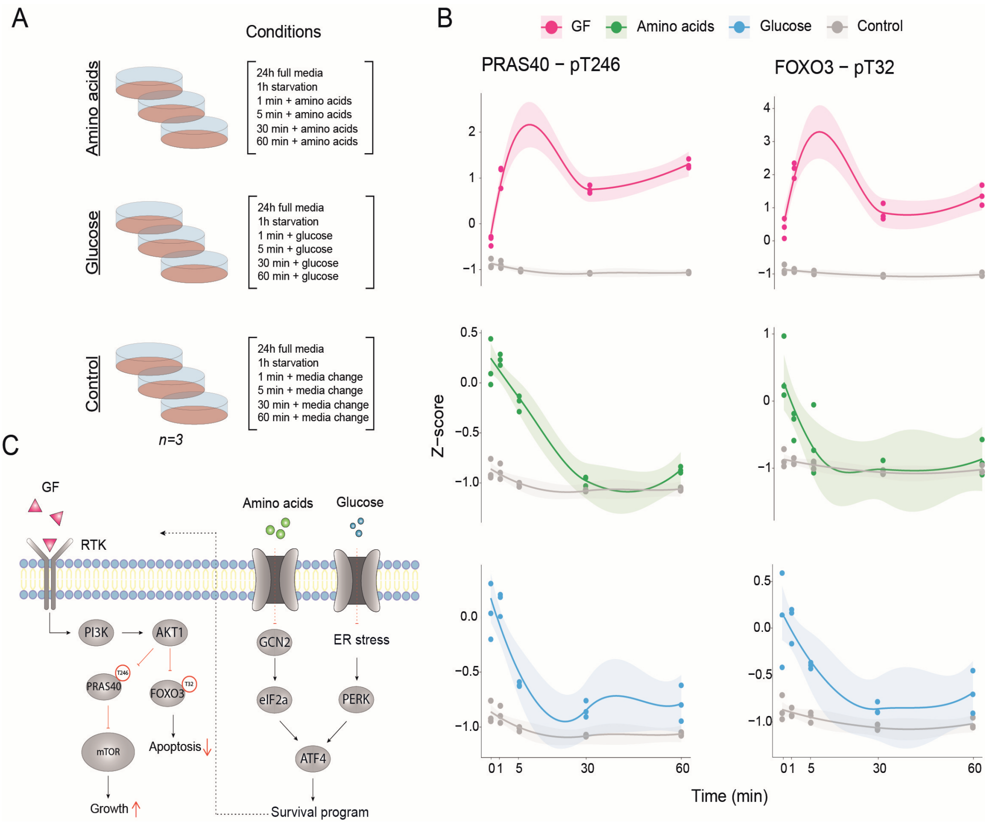


Figure 4. Studying the phosphorylation dynamics of MAPK-AKT-mTOR after stimulus with glucose and amino acids following starvation. (A) Experimental setup: MCF10A cells were starved in glucose-free or amino acid-free media for 1 hour, followed by stimulus with glucose or amino acids respectively. Cells were harvested after 1 hour of starvation and at different time points following stimulus. (B) Response of PRAS40^{T246} and FOXO3^{T32} shows differential response after starvation and stimulus with glucose and amino acids, when compared to growth factor (GF) treatment. Time point 0 represents starvation. (C) PRAS40^{T246} and FOXO3^{T32} illustrate AKT activation due to a cellular survival program induced by glucose or amino acid deprivation (*26*).

Conclusion & Discussion

In this study, we developed a targeted MS assay to study functional phosphorylation dynamics of the MAPK-AKT-mTOR pathway in detail, with high reproducibility and sensitivity. Even though trypsin is the most commonly used digestive enzyme in proteomics studies due to its favourable characteristics (*27*), our study highlights the advantage of the use of multiple proteases to extensively cover the important phosphosites within the pathway, which here can be combined in a single MS analysis. The benefit of this approach is especially high for the study of posttranslational modifications since this requires the measurement of specific peptide sequences. This was evident from our phosphopeptide library which included unique phosphosites and phosphoproteins for each protease. Overall, our assay highlights the potential of custom-made targeted proteomics assays to measure specific phosphosites of interest.

In our experiments, we demonstrated the ability to monitor the plasticity of the phosphorylation dynamics in the MAPK-AKT-mTOR network with high temporal resolution and amino acid residue specificity, under diverse external stimuli. We found that most proteins involved in translation initiation were not phosphorylated upon EGF and insulin stimulus, contrary to previous studies (*15, 16*). Only eIF4G1^{S1231} was found upregulated in cluster 3 ('delayed response'). This suggests that translation initiation factors are not phosphorylated at the same rate. The lack of increased phosphorylation of other initiation factors might be due to the relatively short time course of our experiment or could indicate that replenishing with EGF and insulin (in absence of serum) is inadequate to fully rescue translation initiation after prolonged starvation. Further experiments are needed to fully elucidate the phosphorylation dynamics of translation initiation factors after starvation.

We found dynamic ULK1 phosphorylation upon stimulation with EGF and insulin after starvation. Increased mTOR activity has been shown to phosphorylate ULK^{S638}, inhibiting autophagy upon replenishing with full media (*25*). Our data shows that stimulation with EGF and insulin (and the resulting mTOR activation) is sufficient to induce ULK1^{S638} phosphorylation. Additionally, we found upregulation of ULK1^{S556} upon growth factor stimulus. This phosphosite is responsive to AMPK activity upon starvation and increased ULK1^{S556} has been suggested to induce autophagy (*28, 29*). Therefore, increased ULK^{S556} upon EGF and insulin stimulation is surprising. This discrepancy highlights that the ULK1 phosphorylation dynamics are not fully elucidated.

Next, by using PRAS40^{T246} and FOXO3^{T32} as a proxy of AKT activity, we found evidence in our data of AKT activation upon amino acid and glucose starvation. This is in line with previous research which has proposed increased AKT activity as a survival mechanism upon depletion of these nutrients (*26, 30, 31*). We showed that AKT activation of mTORC1 via PRAS40^{T246} and transcriptional regulation via FOXO3^{T32} are potential effectors of this survival program. Interestingly, our results did not show upregulation of the MAPK pathway upon nutrient starvation, which has also been proposed as a survival mechanism (*32, 33*). This indicates that PI3K-AKT has a more prominent role than MAPK in this type of survival response. This is of interest since both amino acid and glucose starvation have been proposed as therapeutic strategies against certain malignancies (*34, 35*). Overall, these findings suggest that PRAS40 and FOXO3 could be potential therapeutic targets, which could be exploited by combining nutrient starvation with AKT inhibitors.

In conclusion, the experiments described here highlight the potential of our targeted phosphoproteomics assay to monitor the highly dynamic phosphorylation events of the MAPK-AKT-mTOR pathway. Our method extensively covers the important phosphosites in the network (including 150 phosphopeptides on more than 70 phosphoproteins), is sensitive, high throughput and can be applied to *in vivo* and *in vitro* human samples. Therefore, we believe that this assay is a great resource for scientists to study the integration of environmental cues by the MAPK-AKT-mTOR pathway to regulate cellular fate in health and disease.

Methods

Generation of phosphopeptide library

BJ-EHT vRAS12 cells were cultured in DMEM and treated with 100 nM tamoxifen for 1hr, or 24hr to induce RAS activity. Cells were washed with ice cold PBS, detached from the cell culture surface using Trypsin (Lonza) and stored at -80 °C until further use.

BJ-EHT vRAS12 cells, treated with tamoxifen, were lysed using 1% sodium deoxycholate lysis buffer as described previously (Post et al., 2017). Proteins were digested overnight in parallel using one of the four following proteases: Trypsin/LysC, GluC, Chymotrypsin or AspN at 1:100 (w/w) at 37 °C. Samples were acidified and desalted using Sep-Pak C18 cartridges (Waters), eluted with 80% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) and dried down. Samples were fractionated on a high-pH reversed-phase C18 column (Kinetex 5u Evo C18 100A, 150 x 2.1mm, Phenomenex) coupled to an Agilent 1100 series HPLC over a 60 min gradient. Fractions were concatenated to five samples before phosphopeptide enrichment. Enrichment of phosphorylated peptides was performed on the AssayMap BRAVO platform (Agilent Technologies) using Fe(III)-IMAC cartridges following the method described previously (*36*).

Samples were analysed by nanoLC-MS/MS on a Q Exactive HF-X mass spectrometer (Thermo Scientific) equipped with an Agilent 1290 LC system with an LC gradient of 115 min gradient (9% to 35% B). MS settings were as follows: full MS scans (375-1600 m/z) were acquired at 60,000 resolution with an AGC target of 3e6 charges and max injection time of 20 msec. HCD MS2 spectra were generated for the top 12 precursors using 30,000 resolution, 1e5 AGC target, a max injection time of 50 msec, a scan range of 200-2000m/z and normalised collision energy of 27%. MS2 isolation windows were 1.4m/z.

Raw data files were processed with MaxQuant v1.6.3.4 using a Uniprot human database. Carbamidomethyl (C) was set as a fixed modification and Oxidation (M), Acetylation (protein N-terminus) and Phosphorylation (STY) were set as variable modifications. Each proteolytic enzyme was searched individually. Max missed cleavages was set to four for GluC and Trypsin/LysC and 5 for Chymotrypsin and AspN. Results were filtered using a 1% FDR cut off at the PSM level.

Selection of phosphopeptides for SRM assay

Relevant phosphoproteins and phosphosites within the mTOR pathway were identified by literature search. Phosphopeptides that represented these biologically relevant phosphosites were selected in the following manner: firstly, phosphopeptides were selected from our phosphopeptide library. If multiple phosphopeptides were available for the same phosphosite, the most suitable phosphopeptide was chosen. Firstly, peptides without methionine were favoured over peptides including a methionine. Furthermore, the phosphopeptide with the highest intensity was chosen. If our phosphopeptide library did not yield a phosphopeptide to cover a phosphosite of interest we extracted suitable peptides from publically available databases (*37, 38*).

Generation of spectral library

Selected phosphopeptides were synthesised incorporating an isotopically labelled amino acid. These heavy labelled peptides were used to generate a spectral library. For this, peptides were dissolved and mixed with iRT peptides for retention time alignment (*14*). Peptides were analysed by nanoLC-MS/MS on a Q Exactive HF-X mass spectrometer (Thermo Scientific) as described previously with the following changes: HCD MS2 spectra were generated for the top 15 precursors and max injection time was set to 54 msec. Raw files were subjected to database search using Proteome Discoverer (2.2). Tryptic or GluC were selected as proteolytic enzyme, max 3 missed cleavages were allowed, Carbamidomethyl (C) was set as a fixed modification and Oxidation (M), Acetylation (protein N-terminus) and Phosphorylation (STY) were set as variable modifications. Results were filtered using Percolator, FDR was set to 1%. The resulting files were used to build a spectral library in Skyline.

SRM assay development

Heavy labelled standard peptides were used to develop the SRM assay. For each precursor, at least three fragments were measured. The topmost intense fragment ions were chosen and complemented with an additional transition if Y1-ion was amongst the top three most intense ions or if an extra transition was needed to distinguish phospho-isomers. Collision energy optimisation was performed for each transition individually as described previously (*39*).Retention time scheduling was performed using iRT as described previously (*14*).

SRM assay characteristics

The SRM assay we developed had the following characteristics: LC-MS/MS analysis was performed on an Ultimate 3000 RSLCnano System (Thermo Scientific) coupled to a TSQ Altis Triple Quadrupole (Thermo Scientific). Peptides were reconstituted in 20mM citric acid, loaded on a pre-column (C18 PepMap100, 5 µm) and separated on a PepMap RSLC C18 column (2µm, 75µm x 25cm) using a 100 min gradient (2.2% to 34% Buffer B 100% ACN + 0.1% FA) at a flow of 300 nl/min. Retention time windows were set to 5 minutes, Q1 and Q3 resolution was set to 0.7 and a cycle time of 5 seconds was used.

Sensitivity test

Sensitivity of the method was determined by a dilution series of heavy labelled standards spiked into a HELA phosphopeptide background. HELA cells were cultured in DMEM Cells were washed with ice cold PBS, detached from the cell culture surface using Trypsin (Lonza) and stored at -80 °C until further use. Cell lysis, protein digestion (using Trypsin/ LysC) and sample clean-up was performed as described previously for BJ-EHT cells. After sample clean-up, 200µg of cell digest was enriched for phosphorylated peptides (as described previously) and spiked-in with decreasing amount of heavy labelled standards (100 fmol, 50 fmol, 10 fmol, 5 fmol, 1 fmol, 0.1 fmol). This was performed in triplicates. Samples were subjected to the SRM assay described previously. Raw data were analysed using skyline and peaks were individually assessed. Quantification was based on the total peak area.

MCF10A experiments

MCF10A cells were cultured in DMEMF12 medium, supplemented with 5% horse serum, 1% penicillin/streptomycin, EGF (20ng/ml), hydrocortisone (0.5 mg/ml), cholera toxin (100ng/ml) and insulin (10µg/ml) (normal growth conditions). Cells were then starved from either growth factors (using regular DMEMF12 without additives), amino acids (using DMEMF12 without amino acids), or glucose (using DMEMF12 without glucose) for 24 hours. Afterwards, cells were treated with either growth factors (EGF 20 ng/ml and insulin 10 µg/ml), amino acids or glucose for different time points. Cells were washed with ice cold PBS, detached from the cell culture surface using Trypsin (Lonza) and stored at -80 °C until further use. Cells were lysed as described previously for BJ-EHT cells. Protein digestion was performed using Trypsin/LysC or GluC (1:75 w/w enzyme: substrate ratio) overnight at 37 °C. Samples were acidified and tryptic and GluC peptides were combined. Samples were desalted using the AssayMAP Bravo Platform (Agilent Technologies) using C18 cartridges (Agilent Technologies). Cartridges were primed using 100 µl priming solution (80% ACN 0.1% FA) and subsequently equilibrated using 50 µl equilibration solution (0.1% FA in miliQ). Samples were loaded at 10 µl/min and eluted with 50 µl of priming solution at 10 µl/min. After desalting, heavy labelled standards were spiked into each sample. Subsequently, enrichment of phosphorylated peptides was performed on the AssayMap BRAVO platform (Agilent Technologies) using Fe(III)-IMAC cartridges following the method described previously (*36*). Samples were dried down using and stored at -80°C until they were subjected to LC-MS analysis (as described previously under SRM assay characteristics).

Supplementary material

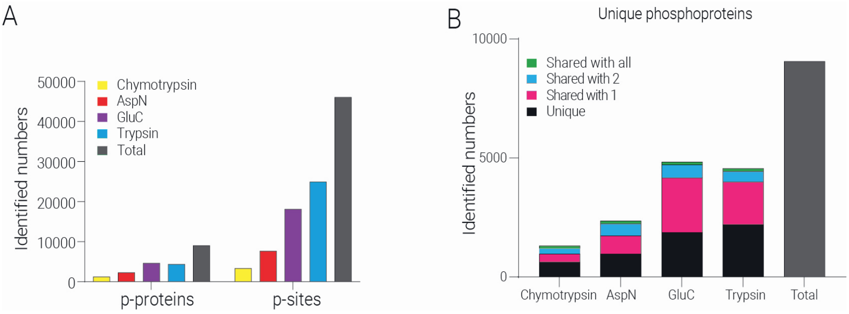


Figure S1. In-house built phosphopeptide library numbers. (A) Number of identified phosphoproteins, phosphopeptides and phosphosites per protease. (B) Phosphoproteins identified per protease coloured by uniqueness.

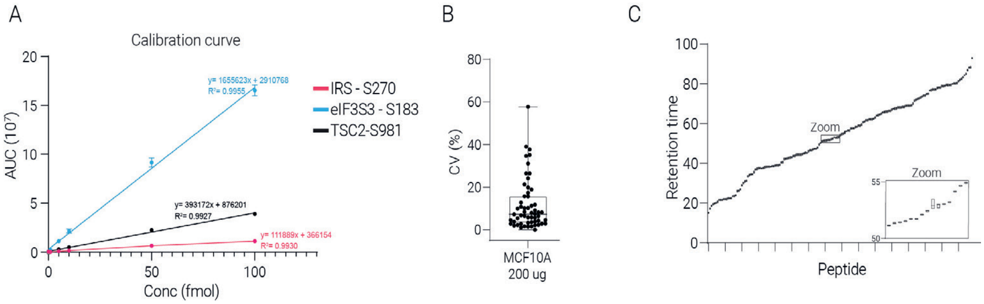


Figure S2. Analytical characteristics of the targeted MS assay. (A) Calibration curve of three phosphopeptides titrated down in a phosphopeptide background illustrates large dynamic range. (B) Coefficient of variation for replicate measurements illustrates quantitative reproducibility. (C) Retention time reproducibility is high, demonstrating robust LC setup.

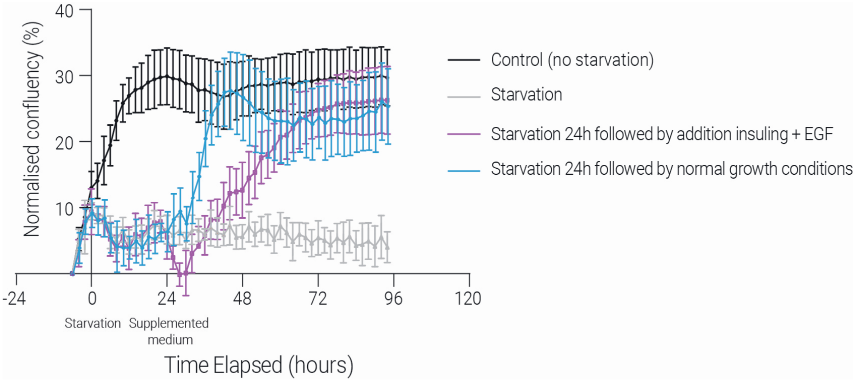


Figure S3. Cellular growth curves. Cellular growth curves under normal growth conditions (black), continues starvation (grey), 24h starvation followed by addition of EGF and insulin (pink) and 24h starvation followed by normal growth conditions (blue).

Acknowledgements

We would like to thank Pieter C. van Breugel for providing us with the BJ-EHT fibroblast cells and Mirjam J. Damen for help with the high pH fractionation. This work has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the European Union and the NWO funded Netherlands Proteomics Centre through the National Road Map for Large-scale Infrastructures program X-Omics, Project 184.034.019. JMV is supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR).

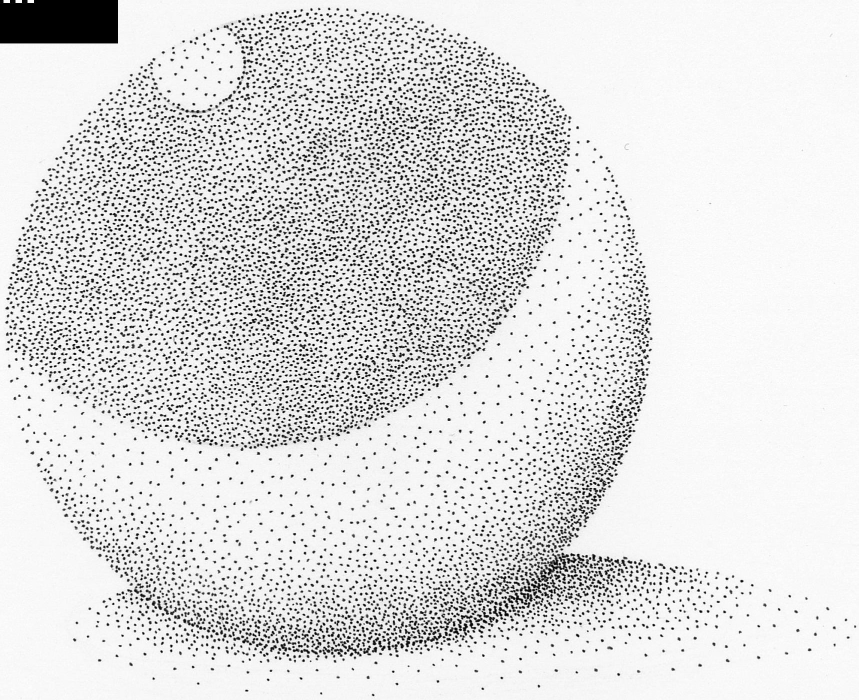
References

1. G. Y. Liu, D. M. Sabatini, mTOR at the nexus of nutrition, growth, ageing and disease. *Nat Rev Mol Cell Biol* **21**, 183-203 (2020).
2. T. Pawson, J. D. Scott, Protein phosphorylation in signaling--50 years and counting. *Trends Biochem Sci* **30**, 286-290 (2005).
3. T. Hunter, Why nature chose phosphate to modify proteins. *Philos Trans R Soc Lond B Biol Sci* **367**, 2513-2516 (2012).
4. R. Seger, E. G. Krebs, The MAPK signaling cascade. *FASEB J* **9**, 726-735 (1995).
5. W. B. Hamilton, J. M. Brickman, Erk signaling suppresses embryonic stem cell self-renewal to specify endoderm. *Cell Rep* **9**, 2056-2070 (2014).
6. M. Burotto, V. L. Chiou, J. M. Lee, E. C. Kohn, The MAPK pathway across different malignancies: a new perspective. *Cancer* **120**, 3446-3456 (2014).
7. M. Drosten, M. Barbacid, Targeting the MAPK Pathway in KRAS-Driven Tumors. *Cancer Cell* **37**, 543-550 (2020).
8. G. Hoxhaj, B. D. Manning, The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. *Nat Rev Cancer* **20**, 74-88 (2020).

9. M. S. Lawrence *et al.*, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501 (2014).
10. R. A. Saxton, D. M. Sabatini, mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **169**, 361-371 (2017).
11. M. C. Mendoza, E. E. Er, J. Blenis, The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci* **36**, 320-328 (2011).
12. J. Adachi, R. Narumi, T. Tomonaga, Targeted Phosphoproteome Analysis Using Selected/Multiple Reaction Monitoring (SRM/MRM). *Methods Mol Biol* **1394**, 87-100 (2016).
13. E. L. de Graaf *et al.*, Signal Transduction Reaction Monitoring Deciphers Site-Specific PI3K-mTOR/MAPK Pathway Dynamics in Oncogene-Induced Senescence. *J Proteome Res* **14**, 2906-2914 (2015).
14. C. Escher *et al.*, Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111-1121 (2012).
15. M. K. Holz, B. A. Ballif, S. P. Gygi, J. Blenis, mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* **123**, 569-580 (2005).
16. A. G. van Gorp *et al.*, AGC kinases regulate phosphorylation and activation of eukaryotic translation initiation factor 4B. *Oncogene* **28**, 95-106 (2009).
17. A. C. Gingras *et al.*, Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev* **13**, 1422-1437 (1999).
18. M. I. Dobrikov, E. Y. Dobrikova, M. Gromeier, Dynamic regulation of the translation initiation helicase complex by mitogenic signal transduction to eukaryotic translation initiation factor 4G. *Mol Cell Biol* **33**, 937-946 (2013).
19. D. A. Ritt, D. M. Monson, S. I. Specht, D. K. Morrison, Impact of feedback phosphorylation and Raf heterodimerization on normal and mutant B-Raf signaling. *Mol Cell Biol* **30**, 806-819 (2010).
20. E. Vander Haar, S. I. Lee, S. Bandhakavi, T. J. Griffin, D. H. Kim, Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40. *Nat Cell Biol* **9**, 316-323 (2007).
21. K. S. Kovacina *et al.*, Identification of a proline-rich Akt substrate as a 14-3-3 binding partner. *J Biol Chem* **278**, 10189-10194 (2003).
22. W. H. Zheng, S. Kar, R. Quirion, Insulin-like growth factor-1-induced phosphorylation of the forkhead family transcription factor FKHRL1 is mediated by Akt kinase in PC12 cells. *J Biol Chem* **275**, 39152-39158 (2000).
23. G. Tzivion, M. Dobson, G. Ramakrishnan, FoxO transcription factors; Regulation by AKT and 14-3-3 proteins. *Biochim Biophys Acta* **1813**, 1938-1945 (2011).
24. L. Wang, T. E. Harris, R. A. Roth, J. C. Lawrence, Jr., PRAS40 regulates mTORC1 kinase activity by functioning as a direct inhibitor of substrate binding. *J Biol Chem* **282**, 20036-20044 (2007).
25. L. Shang *et al.*, Nutrient starvation elicits an acute autophagic response mediated by Ulk1 dephosphorylation and its subsequent dissociation from AMPK. *Proc Natl Acad Sci U S A* **108**, 4788-4793 (2011).
26. G. Pathria, Z. A. Ronai, Harnessing the Co-vulnerabilities of Amino Acid-Restricted Cancers. *Cell Metab* **33**, 9-20 (2021).
27. E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom Rev* **32**, 453-465 (2013).
28. J. E. Jang *et al.*, AMPK-ULK1-Mediated Autophagy Confers Resistance to BET Inhibitor JQ1 in Acute Myeloid Leukemia Stem Cells. *Clin Cancer Res* **23**, 2781-2794 (2017).
29. C. Lin *et al.*, Inhibition of CAMKK2 impairs autophagy and castration-resistant prostate cancer via suppression of AMPK-ULK1 signaling. *Oncogene* **40**, 1690-1705 (2021).

30. D. M. Gwinn *et al.*, Oncogenic KRAS Regulates Amino Acid Homeostasis and Asparagine Biosynthesis via ATF4 and Alters Sensitivity to L-Asparaginase. *Cancer Cell* **33**, 91-107 e106 (2018).
31. S. Shin *et al.*, ERK2 Mediates Metabolic Stress Response to Regulate Cell Fate. *Mol Cell* **59**, 382-398 (2015).
32. M. M. Thiaville *et al.*, MEK signaling is required for phosphorylation of eIF2alpha following amino acid limitation of HepG2 human hepatoma cells. *J Biol Chem* **283**, 10848-10857 (2008).
33. G. Pathria *et al.*, Translational reprogramming marks adaptation to asparagine restriction in cancer. *Nat Cell Biol* **21**, 1590-1603 (2019).
34. S. Ganapathy-Kanniappan, J. F. Geschwind, Tumor glycolysis as a target for cancer therapy: progress and prospects. *Mol Cancer* **12**, 152 (2013).
35. J. S. Kang, Dietary restriction of amino acids for Cancer therapy. *Nutr Metab (Lond)* **17**, 20 (2020).
36. H. Post *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **16**, 728-737 (2017).
37. P. Giansanti *et al.*, An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep* **11**, 1834-1843 (2015).
38. R. T. Lawrence, B. C. Searle, A. Llovet, J. Villén, Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nat Methods* **13**, 431-434 (2016).
39. B. MacLean *et al.*, Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry. *Analytical Chemistry* **82**, 10116-10124 (2010).

CHAPTER III



Phosphoproteome profiling reveals differences in ERK1 and ERK2 dependent phosphorylation

Juan Manuel Valverde[1,2], Pieter van Breugel[1,2], Maarten Altelaar[1,2]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, 3584 CH Utrecht, Netherlands.
[2]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, Netherlands.

Manuscript in preparation

Abstract

Extracellular cues driving cell proliferation are largely transmitted by signaling of mitogen activated protein kinases (MAPKs). In most vertebrates, two enzymes locate at the core of these phosphorylation cascades, namely ERK1 and ERK2. These are responsible for phosphorylating substrates related to a wide variety of processes, like cell migration, regulation of gene expression, and control of cell proliferation and cell differentiation. Deregulation of these enzymes has been observed in a wide range of human diseases, from developmental syndromes to cancer. Even though a variety of studies have tried to assess differences in functionality between ERK1 and ERK2, there is still an ongoing debate about the potential redundancy between these two enzymes. Here we used a combination of CRISPR-Cas9 gene deletion technology and mass spectrometry based phosphoproteomics to determine differences in substrate phosphorylation between ERK1 and ERK2. We find that more substrates are downregulated in the absence of ERK2, and time resolved phosphoproteomics showed that ERK2 seems to be a more efficient kinase than ERK1. We also provide evidence showing that absence of either enzyme causes changes to the overall proteome, while ERK2 deletion caused clear upregulation of alternative signaling pathways, which was not observed in the absence of ERK1. This work shows that despite having a degree of redundancy, ERK1 and ERK2 preferentially phosphorylate a different subset of targets, with lack of ERK2 causing larger changes to the overall phosphoproteome. We think these differences might help explain why lack of ERK2 has been linked to more severe phenotypes than ERK1.

Introduction

Signal transduction of extracellular cues is largely driven by protein phosphorylation (*1*). Cascades of kinases relay information sensed by a variety of membrane receptors to effector proteins. Some of these effectors are kinases themselves, which phosphorylate a myriad of targets involved in diverse cellular processes (*2*). Mitogen activated protein kinases (MAPKs) locate at the core of some of the most studied protein phosphorylation cascades in eukaryotes. There are different families of MAPKs, mainly determined by the stimuli to which they respond, such as growth factors (GFs), cytokines and stress signals (*3, 4*). MAPKs phosphorylate proteins involved in gene expression, mRNA translation, cytoskeleton dynamics, cell metabolism and cell cycle progression (*2*). Overall, MAPK signaling is of paramount importance for cellular homeostasis and organism development.

Work done during the last three decades revealed that two enzymes are largely responsible for intracellular protein phosphorylation upon stimulus with epidermal growth factor (EGF) and insulin (*5, 6*). Both enzymes, extracellular signal-regulated kinase 1 (ERK1, also known as *MAPK3*) and extracellular signal-regulated kinase 2 (ERK2, or *MAPK1*) have molecular masses of around 42-45 kDa. They have up to 85% sequence homology, with differences locating outside of their catalytic domains (*7*). Presence of ERK can be traced down all the way to yeast, with existence of an ancestral version of ERK in invertebrate animals and most vertebrates having both isoforms (*8*). These findings highlight the relevance of these enzymes, but have also raised questions about their specificity (do they phosphorylate the same targets?) and their functionality (do they regulate different processes?).

From the first description of ERK1 and ERK2, several strategies have been used to challenge their apparent redundancy, seeking to find differences between the two enzymes. Strains of knock-out mice of either ERK1 or ERK2 have made some of the most relevant findings so far; *erk1* -/- mice are viable and fertile, while *erk2* deletion is lethal for the early embryo (*9-13*). These findings strongly suggest these enzymes play different roles during development. Using a different model, morpholino experiments in zebrafish embryos showed a higher lethality while reducing ERK2 activity and highlight that ERK1 knock-down can be 'cross-rescued' by treatment with ERK2 RNA, while the opposite is not true (*14*). Furthermore, a study from the same group shows that ERK2 knockdown has a higher impact on the gene expression profile of the embryo when compared to ERK1 depletion (*15*). While these studies point towards ERK2 playing a more relevant role during development, other reports have shown that while *erk1* -/- mice are viable, adults display defects in adipogenesis, erythropoiesis and skin homeostasis (*16-18*). Opposite to these findings, which suggest specific roles for each kinase, a recent study showed that transgenic overexpression of ERK1 can alleviate the lack of ERK2 during development, suggesting redundancy between the enzymes, and that overall ERK activity is what determines the phenotype (*19*). Here it is important to highlight that most studies do not asses overall ERK activity, so it is difficult to interpret if the phenotypic differences are due to kinase specificity or due to differences in overall kinase activity upon depletion of either one of the isoforms (*7*). In addition, these enzymes are known to be differentially expressed among tissues (*8*), increasing the complexity when trying to assess their specificity and functionality.

Tissue specific gene disruption experiments have been used to assess the functionality of these enzymes, since they allow to bypass the embryonic lethality observed in *erk2* -/- mice models. For instance, in macrophages, ERK2 removal does not influence proliferation, as long as ERK1 is expressed. On the other hand, disruption of ERK1 in the absence of ERK2 is lethal, and only a small population of cells, that retained some ERK2 activity due to incomplete gene deletion, were able to proliferate (*20*). In osteoclasts, which are a class of specialized macrophages, ERK1, but not ERK2, has been found to regulate its development and bone resorption activity (*21*). Importantly, in both studies, the authors also showed that these cells express higher levels of ERK1 when compared to ERK2. Analogously, studies that have shown higher expression of ERK2, usually find specific phenotypes upon deletion of this enzyme. This has been observed in brain mouse tissue, in which ERK2 depletion causes anomalies in social and cognitive behaviors (*22*). The same experimental model also showed that ERK2 deficient mice displayed altered pain responses. These mice also presented compensatory hyperactivation of ERK1, which upon inhibition did not aggravate alterations in pain response, suggesting ERK2 plays a more important role than ERK1 in nociception (*23*). Finally, another study in embryonic stem cells showed that ERK2 deletion causes hyperphosphorylation of ERK1, however, also showed downregulation of overall ERK activity. Still, transgenic overexpression of either of the two isoforms rescued substrate phosphorylation and the gene expression program, suggesting there is a degree of redundancy between the enzymes (*24*).

Overall, this paints a picture in which assessment of isoform differences is highly dependent on cell type due to variation in enzyme expression and activity. Hence, relying solely on phenotype observation to determine differences between the ERK isoforms only answers part of the question. A complementary approach would be to assess global changes in protein phosphorylation by mass spectrometry upon enzyme inhibition. Unfortunately, most of the phosphoproteomics work done so far has focused on overall ERK substrate identification without differentiating between the two enzymes (*25-28*). While some attempts have been made to determine specific substrates of ERK2 (*29*), a side-by-side comparison of the effect on the phosphoproteome of each enzyme is currently lacking.

Here we attempt to shed light on the differences in substrate phosphorylation between ERK1 and ERK2. For this, we combined CRISPR-Cas9 technology with mass spectrometry-based phosphoproteomics to study changes in protein phosphorylation of epithelial cells after deletion of ERK1 or ERK2. This revealed that disruption of these kinases already has a drastic and differential impact on the proteome. Moreover, growth factor treatment revealed further differential changes in the phosphoproteome response to the silencing of each enzyme, with the more drastic changes observed in ERK2-depleted cells. Targeted phosphoproteomics revealed activation of auxiliary kinases almost exclusively in response to ERK2 reduced activity. By exploring phosphorylation dynamics after growth factor treatment in a time-course experiment, we show that ERK2 downregulation translates into a lower phosphorylation rate for most substrates, while ERK1 downregulation does not cause the same effect. Overall, our work expands the number of potential ERK substrates known to date, while revealing differences in substrate phosphorylation efficiency between ERK1 and ERK2. We think the higher overall efficiency of ERK2 is (at least in part) the reason why most phenotypical changes reported to date have been observed in the absence of this enzyme.

Results

ERK1/2 gene knock-out validation

We used CRISPR-Cas9 to genetically silence either ERK1 or ERK2 on epithelial MCF10A cells. We did dilution cloning attempting to obtain isogenic knock-out cell lines of each isoform. After genetic screening (see methods), we selected two cell lines for each enzyme and a control cell line treated with nonsense sgRNA, hence considered wild type (WT). For each of the isoforms we had one cell line strongly suggesting deletion of either ERK1 or ERK2 (cell lines 1.2 and 2.2, respectively). On the other hand, we also had cell lines for which genetic evidence was not conclusive, suggesting a polyclonal population or partial gene deletion (disruption of only one of the two alleles) (Figure S1A). Still, we decided to take them along to compare with the other cell lines and see if our phosphoproteomic measurements detected any differences.

Cells were starved for 24 hours followed by GF treatment with supplemented medium (including EGF, insulin and serum) (Figure 1A), which revealed differences in the phosphoproteome of the selected cell lines. First, principal components analysis (PCA) showed that all cell lines are highly responsive to the GF treatment after starvation (Figure S1B). PCA analysis of only GF treated samples revealed differences between the selected cell lines, showing that those with the higher chances of being isogenic according to our genetic screening are the ones that differ the most from the control (Figure S1C). We looked at the significantly changing phosphosites among all the GF treated samples (ANOVA, Benjamini-Hochberg $p < 0.05$), which showed that indeed cell lines with no conclusive evidence of being isogenic behave more similar to the control (Figure S1D, red frames). To confirm the depletion of ERK1/2 we used targeted phosphoproteomics to measure phosphorylation sites known to activate these enzymes. This showed that both enzymes were clearly upregulated after GF treatment (Figure S1E), and confirmed that cell lines 1.2 and 2.2 display the lowest levels of ERK1 and ERK2, respectively (Figure S1F). Based on these results we decided to only use these cell lines for further data analysis.

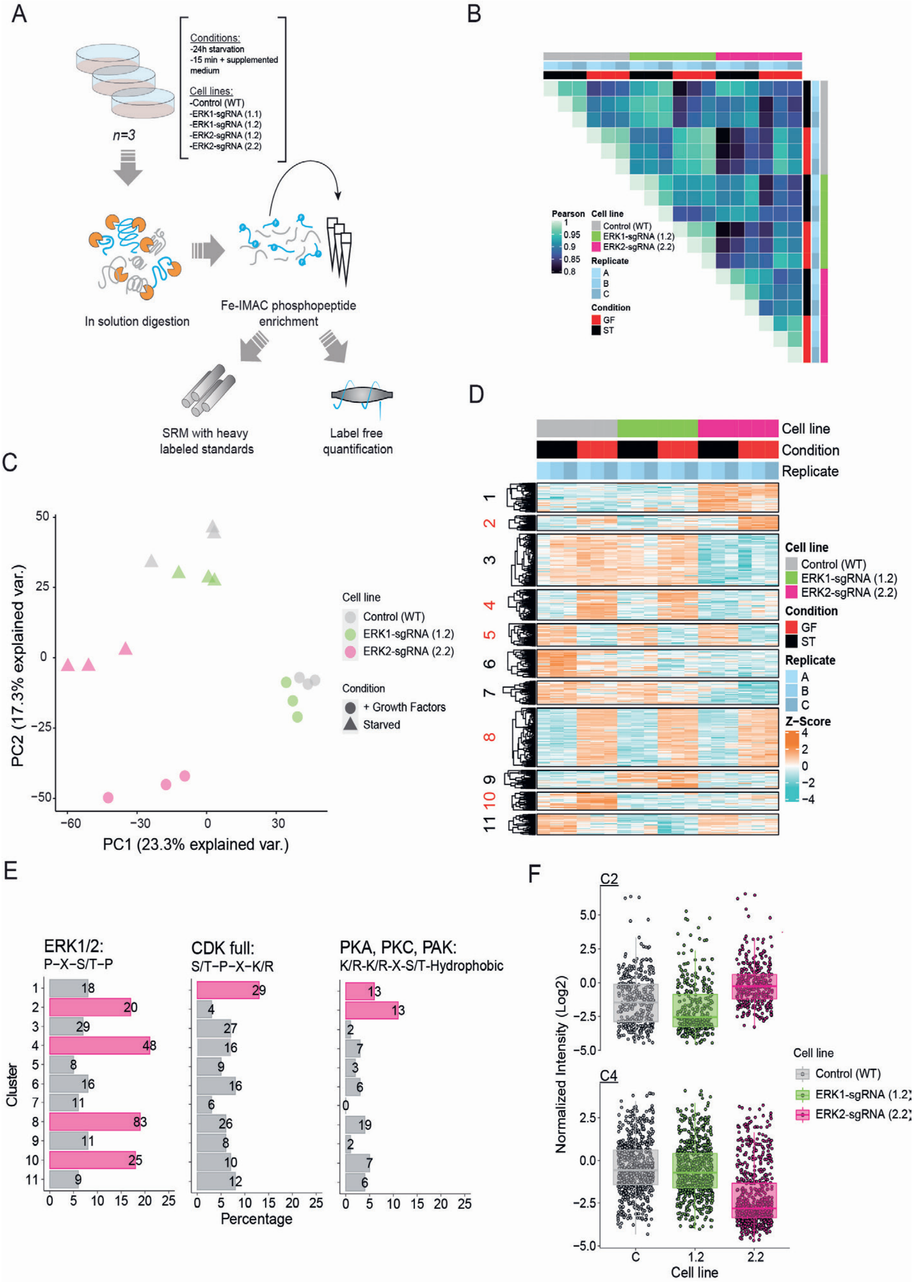


Figure 1. Phosphoproteomics reveal differences in substrate phosphorylation between ERK1 and ERK2. (A) Workflow: different cell lines were starved and subsequently treated with supplemented medium (including GFs). Cells were lysed followed by protein digestion (400 µg per sample). Each sample was divided in two equal parts and subjected to automated phosphopeptide enrichment. **Legend continues on next page.**

Figure 1 legend continued. The phosphopeptides recovered from each column were used for either label free quantification or SRM. We analyzed a biological triplicate of each condition. (B) and (C) Pearson correlation matrix and PCA analysis for all samples based on the expression levels of all quantified phosphosites (*n*= 5,082). (D) Hierarchical clustering of significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05) (*n*= 2,402). Clusters highlighted in red are the ones responsive to GF treatment. (E) Bar plots showing the proportion of specific motifs per cluster (number at the end of each bar shows the actual number of phosphosites with a given motif per cluster). (F) Box plots of all quantified phosphorylation sites in C2 and C4.

Impact of ERK1 and ERK2 on the phosphoproteome

When focusing on ERK1-sgRNA (1.2) and ERK2-sgRNA (2.2), we could see a good overall correlation between biological replicates (Figure 1B). Furthermore, PCA analysis confirmed that samples are separated by their condition (GF treated or starved), but more importantly, it seems like ERK2-sgRNA (2.2) differs the most when compared to the control (Figure 1C). Next, significantly changing phosphosites (ANOVA, Benjamini-Hochberg $p < 0.05$) were subjected to unsupervised clustering, which revealed a variety of dynamic trends (Figure 1D). We detected phosphosites that are clearly responsive to the GF treatment (C2, C4, C5, C8, C10) while others seem to be cell line dependent (C1, C3, C6, C7, C9, C11). First, we focused on those that are GF dependent, since we already showed that both ERK1 and ERK2 were highly responsive to this treatment. Motif analysis revealed that the four clusters showing upregulation after GF treatment are the ones with the higher prevalence of the ERK1/2 full consensus motif (P-X-S/T-P) (Figure 1E). The phosphosite intensities of these clusters suggest that C2 contains targets preferentially phosphorylated by ERK1, while preferred targets of ERK2 locate to C4 (Figure 1F). The other two clusters show phosphorylation sites that could be phosphorylated by any of the two enzymes (C8) or that needed presence of both isoforms to be upregulated (C10). When looking at other kinase motifs, we also detected overrepresentation of the shared motif for PKA, PKC and PAK kinases in C1 and C2, with the latter showing a little over 11% of phosphosites presenting this recognition sequence. When looking at motifs overrepresented in any of the other clusters, the CDK full motif (S/T-P-X-K/R) stood out, since it comprises 13% of the phosphosites in C1 (Figure 1E).

Next, we explored differences in the proteins located in the clusters with the potential ERK1 and ERK2 targets. Network analysis of C2 revealed proteins previously linked to ERK1/2, like STMN1, a protein known to regulate microtubule filaments (Figure 2A). This protein is phosphorylated on S38 by ERK1/2 (*30, 31*). The same was observed for the transcription factor STAT3 (Figure 2A), which is known to be phosphorylated on S727 by a variety of kinases, including ERK1/2 (*32*). Our results suggest that these phosphorylation sites are preferentially phosphorylated by ERK1 (Figure 2B). We also detected phosphorylation of a few proteins related to chromatin remodeling, Golgi apparatus and DNA replication (Figure 2A). Overall, GO analysis of the whole cluster revealed that the majority of proteins in C2 are related to cell junctions and cytoskeleton (Figure S2A, left). C4 contained proteins related to a variety of processes, where terms related to the nuclear pore complex stood out, since they were not enriched in the other clusters (Figure S2A, right). For this cluster, network analysis revealed a few groups of highly interconnected proteins related to different biological processes (Figure 2C). One of these clusters contained proteins involved in RNA transport, including several nucleoporins. Phosphosites from NUP98, NUP35, NUP153 and TPR were clearly downregulated in the ERK2 silenced cells (Figure 2D). Since nucleoporins are known to be phosphorylated by cell cycle related kinases during mitosis (*33*), we wondered if the differences observed are caused by differences in proliferation between cell lines. We assessed this by microscopy, which revealed that indeed ERK2-sgRNA (2.2) cell line proliferate slower when compared to the control and the ERK1-sgRNA (1.2) cell line (Figure S2B). These differences in proliferation were accounted for during the sample preparation process, by including a protein quantification assay to ensure that the same amount of input material was used for all samples. C10 did not show a clear enrichment of a specific term not observed in the other clusters, while C8 presented enrichment of chromatin binding proteins and histone modifiers (Figure S2A).

To further explore if the phosphosites on the clusters of interest (C2, C4, C8, C10) are targets of ERK1/2 we looked at the overlap of each of them against the most recent database of ERK1/2 we could find in the literature (*34*). Briefly, this database contained information from different sources, one of them being PhosphositePlus database (*35*), and other publications that have tried to directly determine substrates of these kinases (*29*). We focused only on the phosphosites reported in human (n=1,979 phosphorylation sites). As we suspected, above 23% of the phosphosites in each of the clusters of interest are known ERK1/2 targets, with C8 having the highest number of known substrates (35% of the total phosphosites in C8) (Figure S2C). Other clusters did not contain more than 8% of known substrates of ERK1/2, clearly showing that we can distinguish potential ERK1/2 substrates based on their motif and response to GF treatment (Figure S2C). Next, we explored those phosphosites who have the ERK full motif but have not been yet reported as ERK1/2 targets. We focused on C2 and C4, which contain the potential ERK1 and ERK2 substrates, and found 14 and 38 phosphosites that have not been previously linked to these kinases, respectively (Figure 2E). These potential new targets are summarized in Supplementary tables 1 and 2.

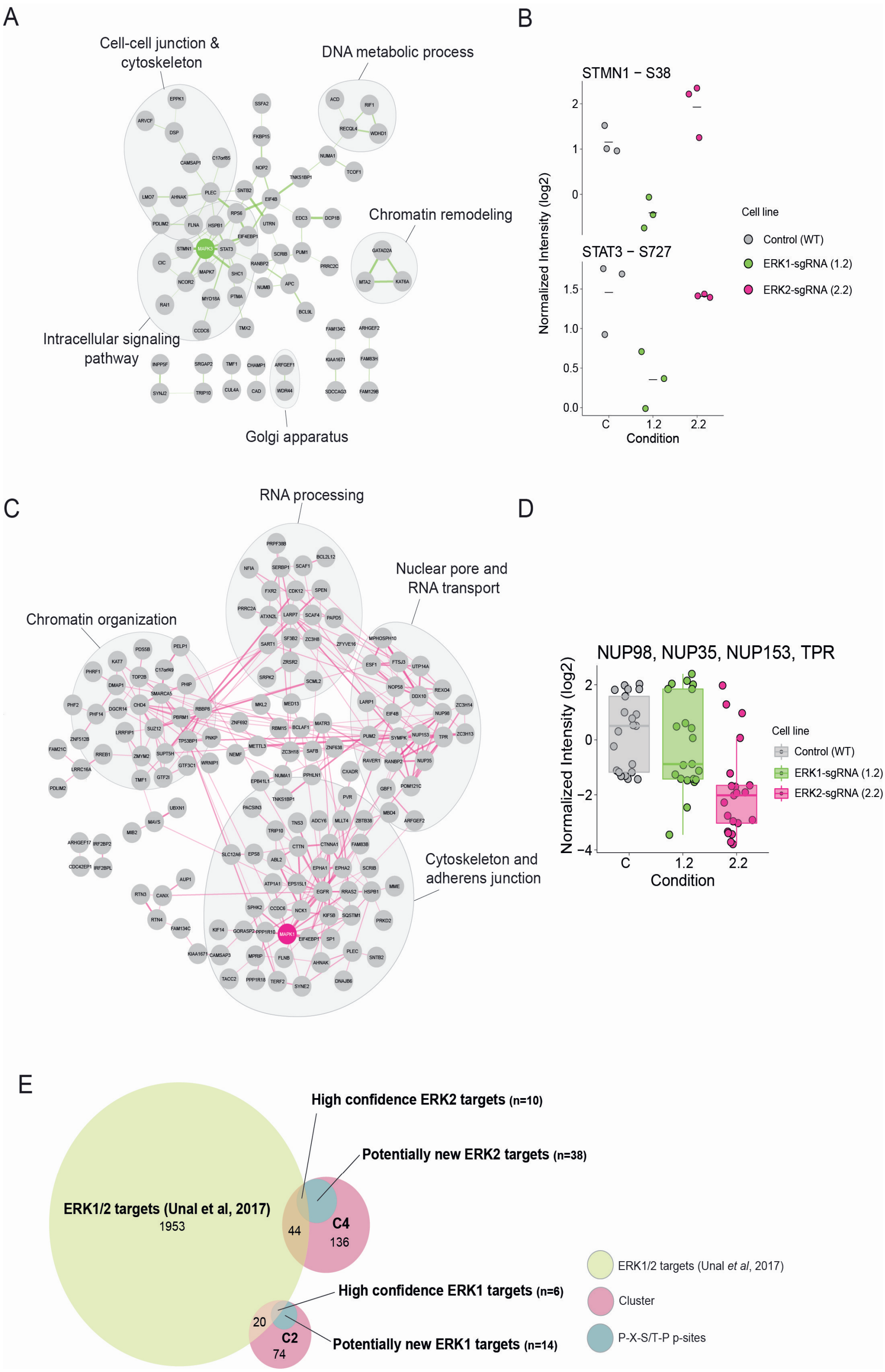

A
Cell-cell junction & cytoskeleton
DNA metabolic process
Chromatin remodeling
Intracellular signaling pathway
Golgi apparatus
MAPK3
B
STMN1 – S38
STAT3 – S727
Normalized Intensity (log2)
Condition
C
1.2
2.2
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
C
RNA processing
Nuclear pore and RNA transport
Chromatin organization
Cytoskeleton and adherens junction
MAPK1
D
NUP98, NUP35, NUP153, TPR
Normalized Intensity (log2)
Condition
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
E
High confidence ERK2 targets (n=10)
Potentially new ERK2 targets (n=38)
ERK1/2 targets (Unal et al, 2017)
1953
44
C4
136
High confidence ERK1 targets (n=6)
20
C2
74
Potentially new ERK1 targets (n=14)
ERK1/2 targets (Unal et al, 2017)
Cluster
P-X-S/T-P p-sites

Figure 2. Overview of targets preferentially phosphorylated by either ERK1 or ERK2. (A) STRING network of proteins from C2 (phosphosites preferentially phosphorylated by ERK1). GO analysis revealed sub clusters of proteins with different biological functions. Width of the edges correlates with the confidence of the interactions. (B) Examples of two phosphorylation sites with known biological function that are upregulated in an ERK1 dependent manner. (C) STRING network of proteins from C4 (phosphosites preferentially phosphorylated by ERK2). GO analysis revealed proteins related to a variety of processes (highlighted GO term corresponds to the most representative term for the given sub cluster). Width of the edges correlates with the confidence of the interactions. (D) Boxplot of phosphorylation sites from nucleoporins in C4. (E) Venn diagram showing overlap of C2 and C4 versus a previously published database of ERK1/2 targets.

Targeted phosphoproteomics reveal kinase activity rewiring after ERK2 deletion

To better understand the differences in the phosphoproteome caused by the lack of ERK1 or ERK2, we used targeted phosphoproteomics to assess the phosphorylation state of phosphosites formerly known to be biologically relevant. We used previously developed single reaction monitoring (SRM) assays to measure relevant targets of the MAPK-AKT-mTOR pathway (*36*), and the phosphorylation state of the activation loop of a variety of kinases (*37*). Our goal was to reveal if the deletion of either one of these enzymes, which locate at the heart of the MAPK-AKT-mTOR network, affected the phosphorylation state of other members and the overall activity of this pathway. Furthermore, by assessing activation state of different kinases, we could determine the enzymes responsible for the phosphorylation observed in clusters upregulated after deletion of either ERK1 or ERK2 (C9 and C1, respectively).

In total, we quantified 39 phosphopeptides from the MAPK-AKT-mTOR pathway (Figure S3A). First, we found evidence suggesting that the PDK1-AKT1 axis is activated upon starvation. Interestingly, this only happened for the control cell line and not for the cells that lacked either one of the ERK isoforms (Figure S3A). Next, we also found that PRAS40, one of the regulating subunits of mTOR complex 1 (mTORC1), which is known to be phosphorylated on T246 by AKT1 (*38, 39*), was downregulated in the ERK1-sgRNA cell line. The same was true for kinase GSK3β-S9, which is also supposed to be under control of AKT1 (*40*), and for eIF4G1-S1231 and TSC2-S1254. The latter two are phosphosites involved in translation initiation and regulation of mTOR activation, respectively (*41-44*). All these phosphosites are downregulated in ERK1-silenced cells (Figure 3A, top). On the other hand, we found phosphosites downregulated in both ERK1-sgRNA and ERK2-sgRNA cell lines, but that clearly show a stronger downregulation in the absence of ERK2 (Figure 3A, bottom). Importantly, many of these are located downstream of mTORC1. One case that stood out was BAD-S99, which is a phosphosite involved in anti-apoptotic response, which is clearly upregulated in the ERK2-sgRNA cell line (Figure 3B). A similar case was SIN1-S260, which showed upregulation in the ERK2-sgRNA cell line during starvation (Figure S3A). This phosphosite is under control of mTORC2 and regulates the activity of this same complex (*45*). Many of these phosphosites regulate the output of the MAPK-AKT-mTOR network and control processes like autophagy, mRNA biogenesis and protein translation, therefore determining cellular fate (Figure 3C). These results reveal that silencing of either ERK1 or ERK2 has differential impact on the MAPK-AKT-mTOR pathway, but overall, suggest that absence of ERK2 causes downregulation of important effectors downstream of mTORC1.

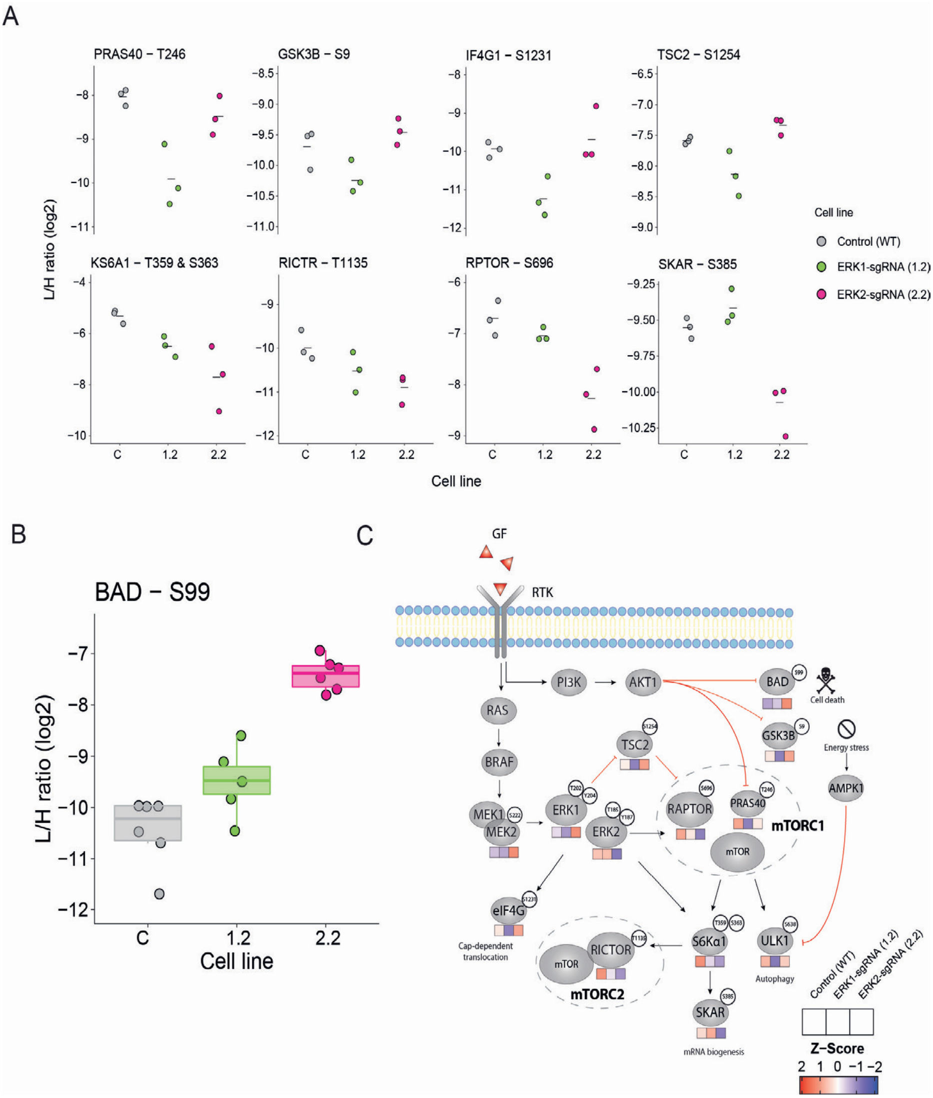


Figure 3. Targeted phosphoproteomics reveal differential effect of ERK1 or ERK2 deletion on MAPK-AKT-mTOR pathway. (A) Dot plots showing biologically functional phosphosites downregulated in the absence of ERK1 (top) and downregulated mainly after ERK2 depletion (bottom). **Legend continues on next page**

Figure 3 legend continued. (B) Boxplot showing upregulation of BAD-S99 after depletion of ERK1 or ERK2. (C) Simplified network with some members of the MAPK-AKT-mTOR pathway showing differential regulation in the absence of ERK1 or ERK2. Heat maps below each node show the average intensity for the phosphosites shown in the network after GF treatment. Black arrows represent phosphorylation driven activation while red blunt arrows represent phosphorylation driven inhibition.

Next, we assessed the activating phosphorylation of 22 kinases, which revealed different trends (Figure S3B). There were kinases which appear to be strongly upregulated after GF treatment exclusively in ERK2-silenced cells, such as MEK2, ERK5 and AKT3 (Figure 4A, top). On the other hand, CDK11A, CDK11B, GSK3α and PKA catalytic subunit beta (PKACB) are also upregulated in the ERK2-sgRNA cell line, but independently of the stimulus (Figure 4A, bottom). On the contrary, a kinase like CDK12 seems to be overall downregulated after ERK2 deletion (Figure 4B). Some of these enzymes are known to regulate activation of other kinases, moreover, enzymes like PKA and AKT have been suggested to phosphorylate targets that directly control cell metabolism and cellular fate (Figure 4C). Overall, these results suggest upregulation of alternative signaling pathways, especially in response to ERK2 deletion.

Proteome response to the lack of ERK1 or ERK2

Since our results so far suggest drastic changes to the overall phosphoproteome upon deletion of either ERK1 or ERK2, we wondered if some of these changes could also be caused by overall changes in the proteome. We measured the proteome of these cell lines, with high correlation between biological replicates (Figure 5A), while PCA analysis revealed that differences between samples were mainly cell line dependent (Figure 5B). The latter was confirmed by unsupervised clustering of significantly changing proteins (ANOVA, Benjamini-Hochberg $p < 0.05$), which revealed that protein levels are not changing after 15 minutes of GF treatment (Figure S4A). This confirmed that changes observed in phosphorylation sites are not caused by modulation of total protein levels. Still, comparison of each cell line against the control revealed clear changes of the proteome on absence of either isoform. ERK1-sgRNA cell line showed downregulation of proteins related to carbohydrate metabolism (*e.g.* glycolytic process, ATP generation from ADP) with upregulation of proteins related to RNA binding and ribosome biogenesis (Figure 5C). In comparison, the ERK2-silenced cell line also showed downregulation of proteins involved in carbohydrate metabolism, but also downregulation of proteins related to ribonucleoprotein complex assembly and translation initiation. This cell line also displayed upregulation of proteins related to a variety of processes, from ribosomal subunits to mitochondrial proteins and components of focal adhesions (Figure 5D). A deeper GO analysis of the proteome can be found on Figure S4B. Overall, these results show that changes in protein phosphorylation are not a byproduct of modulation of protein levels. Having said that, deletion of these enzymes clearly affects the proteome, with the more drastic differences observed in the ERK2-silenced cell line.

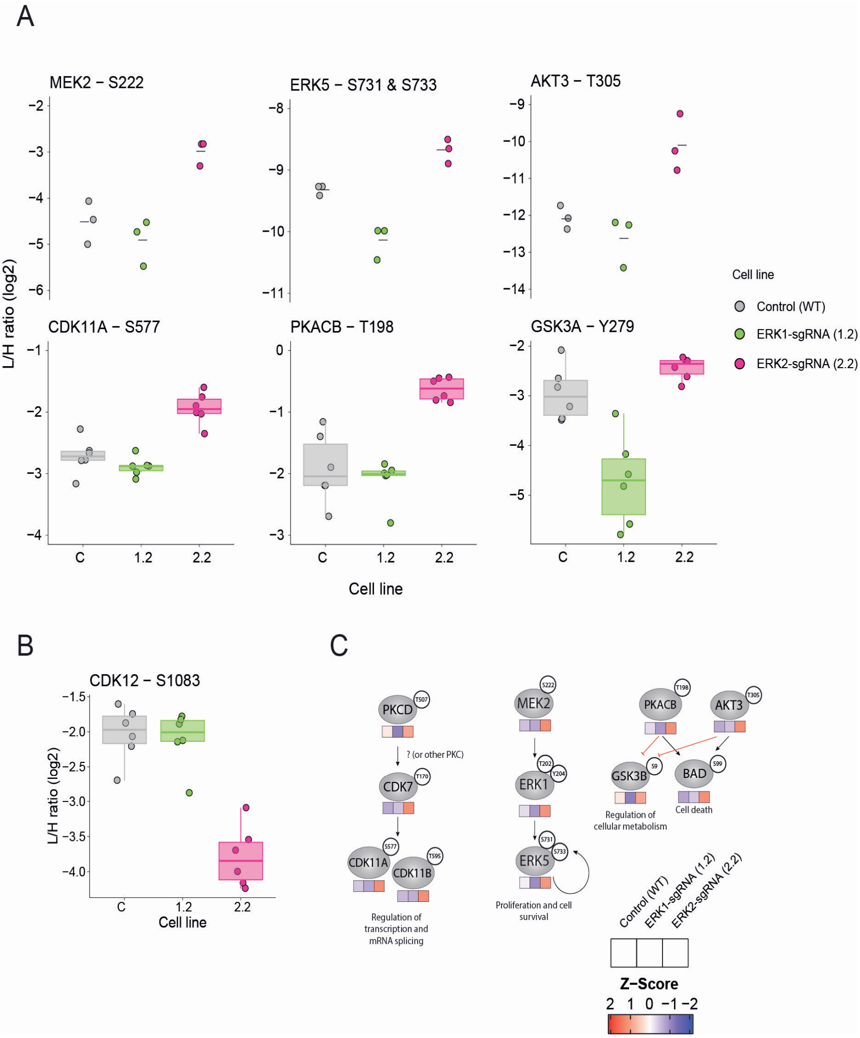


Figure 4. Assessment of kinome activation state reveals differences between ERK1 and ERK2 silenced cells. (A) Dot plots showing upregulation of activating phosphorylation sites after GF treatment in ERK2-silenced cells (top). On the bottom, boxplots show kinases which are upregulated after ERK2 depletion regardless of the condition (starved or GF treated). (B) CDK12-S1083 was clearly downregulated in the ERK2-sgRNA (2.2) cell line. (C) Signaling axis likely to be activated upon ERK2 deletion. Heat maps below each node show the average intensity for the phosphosites shown in the network after GF treatment. Black arrows represent phosphorylation driven activation while red blunt arrows represent phosphorylation driven inhibition.

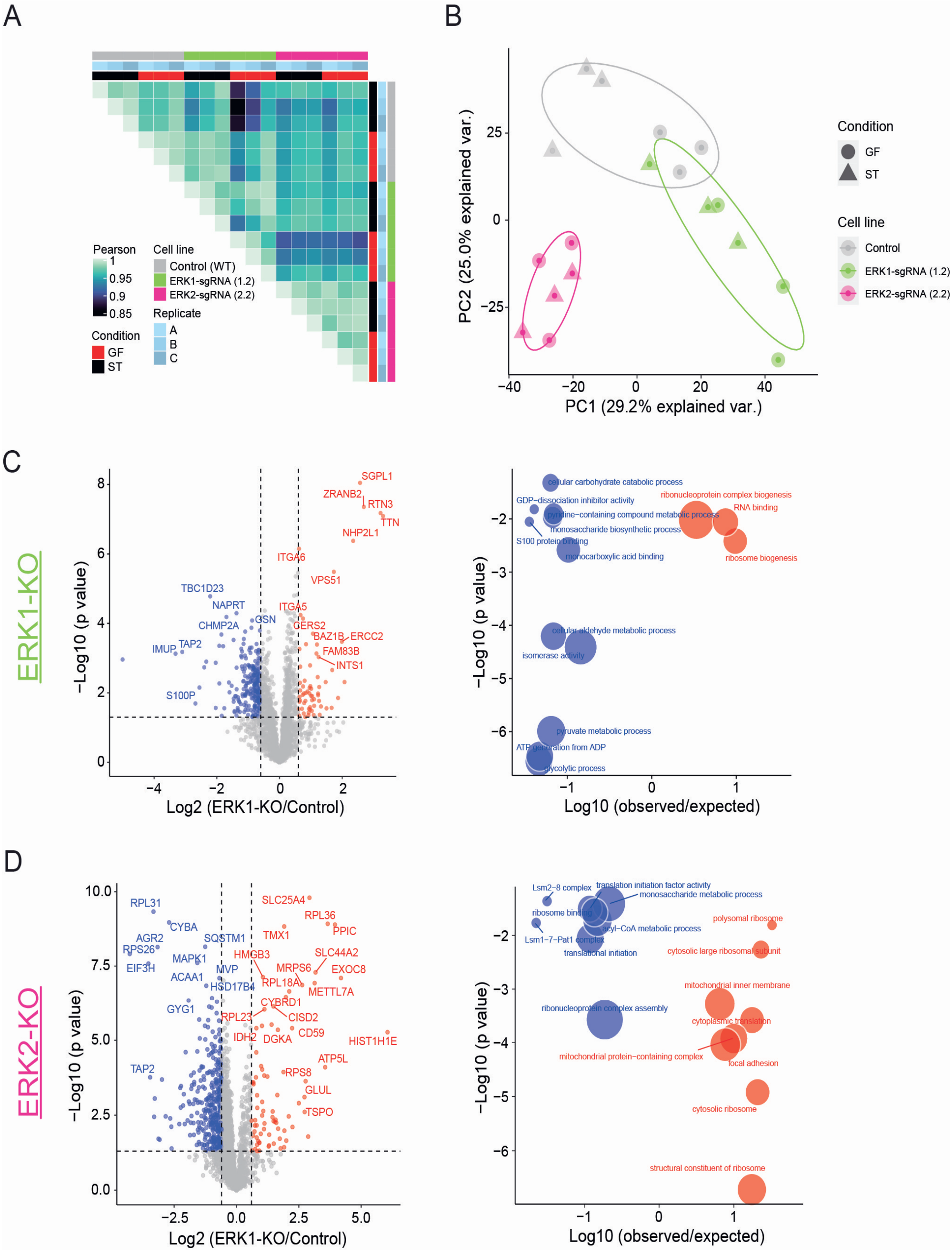

A
Pearson
1
0.95
0.9
0.85
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
Replicate
A
B
C
Condition
GF
ST
B
PC2 (25.0% explained var.)
PC1 (29.2% explained var.)
25
0
−25
−40
−20
0
20
40
Condition
GF
ST
Cell line
Control
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
C
ERK1-KO
−Log10 (p value)
Log2 (ERK1-KO/Control)
SGPL1
ZRANB2
RTN3
TTN
NHP2L1
ITGA6
VPS51
ITGA5
CERS2
BAZ1B
ERCC2
FAM83B
INTS1
TBC1D23
NAPRT
CHMP2A
GSN
IMUP
TAP2
S100P
cellular carbohydrate catabolic process
GDP−dissociation inhibitor activity
ribonucleoprotein complex biogenesis
RNA binding
pyrimidine−containing compound metabolic process
monosaccharide biosynthetic process
S100 protein binding
monocarboxylic acid binding
ribosome biogenesis
cellular aldehyde metabolic process
isomerase activity
pyruvate metabolic process
ATP generation from ADP
glycolytic process
Log10 (observed/expected)
D
ERK2-KO
−Log10 (p value)
Log2 (ERK1-KO/Control)
RPL31
CYBA
AGR2
RPS26
EIF3H
MAPK1
ACAA1
SQSTM1
MVP
HSD17B4
GYG1
TAP2
SLC25A4
RPL36
PPIC
TMX1
HMGB3
SLC44A2
MRPS6
EXOC8
RPL18A
METTL7A
CYBRD1
RPL23
CISD2
CD59
IDH2
DGKA
HIST1H1E
ATP5L
RPS8
GLUL
TSPO
Lsm2−8 complex
translation initiation factor activity
monosaccharide metabolic process
ribosome binding
acyl−CoA metabolic process
Lsm1−7−Pat1 complex
translational initiation
polysomal ribosome
cytosolic large ribosomal subunit
mitochondrial inner membrane
ribonucleoprotein complex assembly
cytoplasmic translation
mitochondrial protein−containing complex
focal adhesion
cytosolic ribosome
structural constituent of ribosome
Log10 (observed/expected)

Figure 5. Deletion of ERK1 or ERK2 has a differential impact on overall protein expression. (A) and (B) PCA analysis and Pearson correlation matrix of all quantified proteins (*n*=2,478). (C) Volcano plot showing proteins differentially expressed in ERK1-silenced cells when compared to the control (left). GO analysis of reduced (blue) and increased (red) proteins, showing the most representative GO terms of each group of proteins (right). The size of the dot reflects the number of proteins related to the GO term (see materials and methods for details). (D) Volcano plot showing proteins differentially expressed in ERK2-silenced cells when compared to the control (left). GO analysis of reduced (blue) and increased (red) proteins, showing the most representative GO terms of each group of proteins (right). The size of the dot reflects the number of proteins related to the GO term (see materials and methods for details).

ERK2 depletion has a larger effect on phosphorylation dynamics

To explore the phosphorylation dynamics of both of these enzymes, we did another phosphoproteomics experiment, this time harvesting cells at different time points after GF treatment (Figure 6A). Biological replicates showed high correlation (Figure 6B), which was also evident by PCA analysis, with samples from the same time points clustering together (Figure 6C). We focused on significantly changing phosphosites (ANOVA, Benjamini-Hochberg $p < 0.05$), which revealed six clusters (Figure 6D), of which C3 had the highest percentage of ERK full motifs, followed by C1 (Figure 6E). Furthermore, the highest number of previously known ERK substrates was found on C3 (Figure S5A). Hence, we think this cluster has the likely targets of ERK1/2. While C1 seems to have substrates that could be phosphorylated by both enzymes, C3 shows phosphosites with a slower phosphorylation rate in the ERK2-depleted cells (Figure 6F). GO analysis showed that C1 has substrates related to chromatin binding and histone modification, while C3 contains structural components of the nuclear pore complex (Figure S5B). These findings correlate with what we previously described (Figure S2A), and support that while some nuclear proteins are phosphorylated by either one of the isoforms, nucleoporins in particular seem to be preferentially regulated by ERK2 (Figure S5B, C). Furthermore, when we subjected C3 to another round of clustering, we could see phosphosites with different phosphorylation rates (Figure S6A). We wondered if these were caused by the differences in the phosphorylation motif, since one of the clusters clearly displayed a higher proportion of ERK full canonical motifs (Figure S6B, C). For this, we explored the trend in time of phosphosites with the full and minimal ERK motif, which revealed no clear differences between the two (Figure S6D). Finally, we also explored the overlap of C3 against the database of ERK1/2 substrates, which revealed 140 potentially new targets with the ERK full motif, greatly expanding our knowledge of residues phosphorylated by these enzymes (Figure 6G).

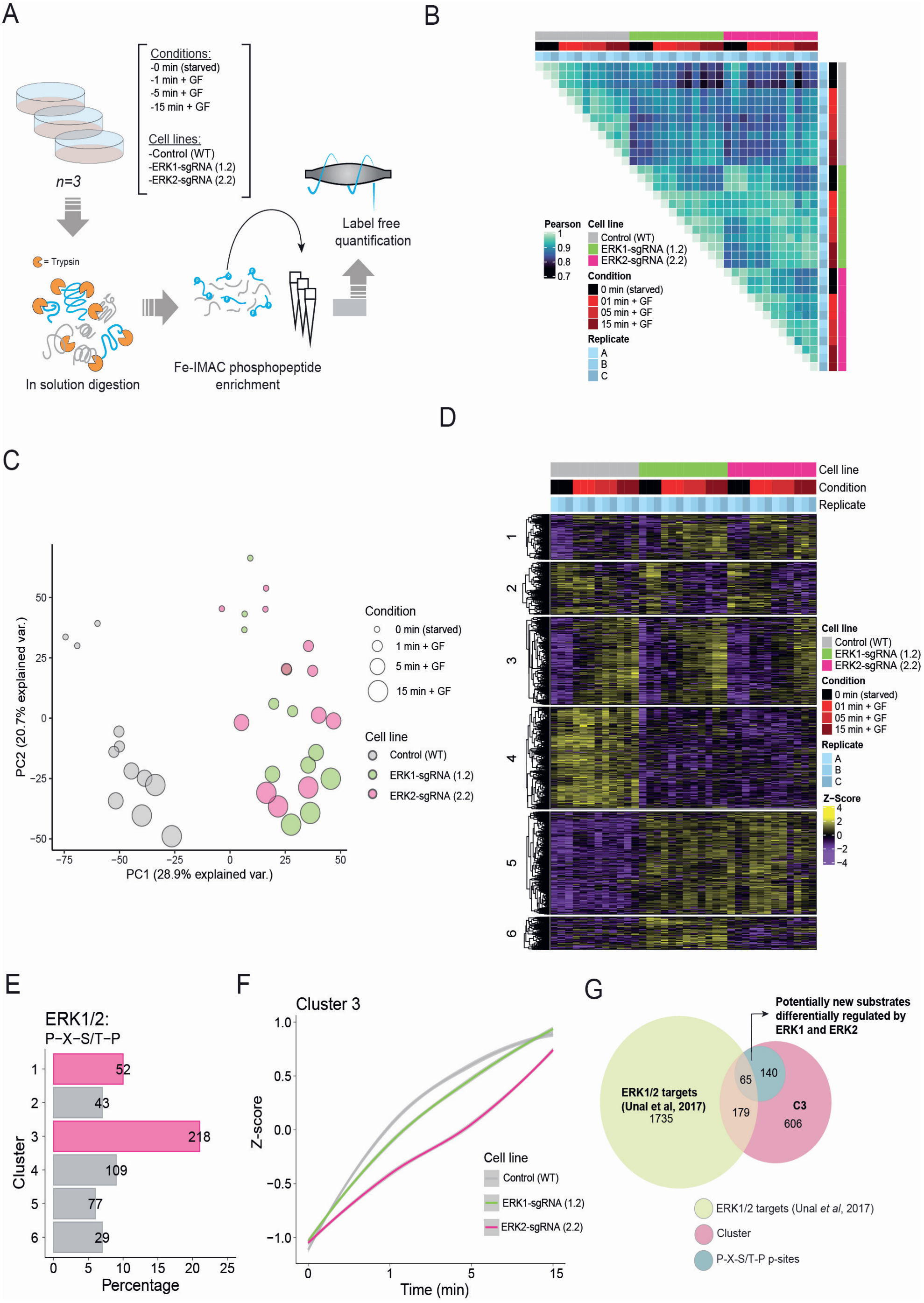

A
Conditions:
-0 min (starved)
-1 min + GF
-5 min + GF
-15 min + GF
Cell lines:
-Control (WT)
-ERK1-sgRNA (1.2)
-ERK2-sgRNA (2.2)
n=3
= Trypsin
In solution digestion
Fe-IMAC phosphopeptide enrichment
Label free quantification
B
Pearson
1
0.9
0.8
0.7
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
Condition
0 min (starved)
01 min + GF
05 min + GF
15 min + GF
Replicate
A
B
C
C
PC2 (20.7% explained var.)
PC1 (28.9% explained var.)
50
25
0
−25
−50
−75
−50
−25
0
25
50
Condition
0 min (starved)
1 min + GF
5 min + GF
15 min + GF
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
D
Cell line
Condition
Replicate
1
2
3
4
5
6
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
Condition
0 min (starved)
01 min + GF
05 min + GF
15 min + GF
Replicate
A
B
C
Z−Score
4
2
0
−2
−4
E
ERK1/2:
P−X−S/T−P
Cluster
1
2
3
4
5
6
52
43
218
109
77
29
0
5
10
15
20
25
Percentage
F
Cluster 3
Z-score
1.0
0.5
0.0
−0.5
−1.0
0
1
5
15
Time (min)
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
G
Potentially new substrates differentially regulated by ERK1 and ERK2
ERK1/2 targets (Unal et al, 2017)
1735
65
140
179
C3
606
ERK1/2 targets (Unal et al, 2017)
Cluster
P-X-S/T-P p-sites

Figure 6. Time-resolved phosphoproteomics reveal differences between ERK1 and ERK2. (A) Workflow: cells were starved and subsequently treated with GFs (EGF, insulin and serum). Next, cells were lysed and proteins digested, followed by automated phosphospeptide enrichment. Phosphopeptides were measured on an Orbitrap Exploris 480. (B) and (C) Pearson correlation matrix and PCA analysis of all quantified phosphorylation sites (*n*=8,516). (D) Hierarchical clustering of significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05) (*n*= 4,976). (E) Bar plot showing the proportion of ERK full motifs per cluster (number at the end of each bar shows the actual number of phosphosites with the motif). (F) Smoothed curve showing the trend of phosphorylation sites in C3. (G) Venn diagram showing overlap of cluster 3 versus a previously published database of ERK1/2 targets.

Discussion

Here we explored the differences in substrate phosphorylation between cells lacking either ERK1 or ERK2.. Hierarchical clustering of phosphosite intensities revealed differences in the phosphoproteome, some of which were independent of GF treatment. The latter suggests that lack of either one of the isoforms might induce a rewiring of signaling networks as compensation, as has been shown by others (*46*). Our data clearly shows compensation in the ERK2-sgRNA (2.2) cell line, which displays increased phosphorylation of ERK1 and of potential ERK1 targets. Others that have depleted cells of ERK2 have also reported this phenomenon (*47*). The combination of hierarchical clustering with phosphorylation motif analysis, allowed us to differentiate between what we believe is compensatory phosphorylation (cell line dependent) and potential targets of either ERK1 or ERK2 (GF treatment dependent). Furthermore, we also observed that phosphosites upregulated after ERK2 depletion displayed higher number of phosphorylation motifs from other kinases (*e.g.* CDK, PKA, PAK), suggesting these are likely to be more active in these cells. Overall, we observed that depletion of ERK2 induces larger changes to the phosphoproteome when compared to depletion of ERK1. The latter is already interesting since most phenotypical changes reported in the literature have also been observed in the absence of ERK2.

Next, we showed that indeed certain phosphosites are preferentially phosphorylated by either ERK1 or ERK2. We found that ERK1 dependent phosphosites are mostly related to cytoskeleton and cell junctions, including STMN1-S38. The latter is involved in regulation of the microtubule cytoskeleton dynamics (*48*). Here, we also found STAT3-S727, which is a highly studied phosphosite known to regulate the activity of this transcription factor (*49*). In contrast, ERK2 dependent phosphosites were related to a variety of GO terms, where constituents of the nuclear pore showed the highest enrichment. Interestingly, others have shown that ERK phosphorylation of nucleoporins regulates nucleocytoplasmic transport (*27*). Moreover, phosphorylation of nucleoporins by ERK also mediates its own nuclear translocation, which is dependent on Nup153 (*50*). Our data nicely shows that nucleoporin phosphorylation is downregulated in the absence of

ERK2, including at least one phosphosite located in Nup153 (S614). This is particularly relevant since others have demonstrated that ERK2 shuttles between the nucleus and the cytoplasm at a higher rate than ERK1 (*51*). So far, this behavior has been attributed to differences in the sequence of the N-terminus of these enzymes, although our data suggests nucleoporin phosphorylation might also be playing a role in this process.

By assessing the phosphorylation state of the MAPK-AKT-mTOR we found that ERK1 and ERK2 might exert differential regulation over this network. First, we confirmed that in the absence of ERK2, ERK1-T202 & Y204 are upregulated, and so is MEK2-S222. The latter suggest a compensation mechanism by which cells aim to upregulate overall ERK activity in the absence of ERK2. Interestingly, the opposite is not true; when ERK1 is depleted, levels of ERK2-T185 & Y187 are the same as in the control. Next, we found that certain targets were specifically downregulated in the ERK1-silenced cells, specifically EIGF1-S1231, which is involved in translation initiation and was previously reported as an ERK1/2 substrate (*41*). Similarly, TSC2-S1254 was downregulated in absence of ERK1 and upregulated in the ERK2-sgRNA cell line. TSC2 is known to be phosphorylated in order to remove its inhibitory effects on mTORC1 (*42-44*). This, added to the fact that another inhibitor of mTOR, namely PRAS40 (*38, 39*), is downregulated in T246 after ERK1 depletion, would suggest that ERK1 plays an important role in mTORC1 activation. However, regulation of this complex is likely to be more nuanced, since we also detected that RPTOR-S696, a previously reported substrate of ERK1/2 (*52*), which is phosphorylated to induce mTORC1 activation (*53*), displayed the lowest levels after ERK2 depletion. In addition, a similar trend was observed for KS6A1-T359 & S363 and SKAR-385. All these are phosphosites directly linked to the activation of these kinases (*54-56*), which are downstream of mTORC1. The same was true for RICTR-T1135, which regulates the activity of mTORC2 and is also downstream of mTORC1 (*57*). Overall, these results show that despite differences in the phosphorylation state of direct regulators of mTORC1, absence of ERK2 seems to have stronger effect on the overall activity of this mTOR complex.

By exploring global kinome activation, we found that AKT3-T305 was upregulated in the ERK2-sgRNA cell line, which might explain why known targets of AKT1 were also upregulated. Interestingly, the upregulation of SIN-S260 after ERK2 depletion could be related to upregulation of mTORC2 activity, which controls the AKT pathway (*58*). If these findings translate into overall increased AKT activation as compensation for the lack of ERK2, it would also explain why this cell line also showed upregulation of PRAS40-T246. Interestingly, upregulation of PRAS40-T246 does not seem to translate into higher mTORC1 activity, as we previously discussed. We also see upregulation of ERK5-S731 & S733, which is a less known MAP kinase also involved in proliferation and cell survival (*59*). Both of these phosphosites have been reported as ERK5 auto phosphorylation targets (*60*), although S733 has also been shown to be phosphorylated by ERK1/2

(*61*). This could indicate that upon lack of ERK2, upregulation of MEK-S222 and ERK1-T202 & Y204 could also induce upregulation of ERK5. Furthermore, other kinases like PKCD, CDK7 and both CDK11A and CDK11B, all show upregulation in the absence of ERK2. We think these three kinases might be activated in tandem, since PKC isoforms have been shown to regulate CDK7 (*62*), which is a CDK activating kinase known to regulate CDK11 activation (*63*). The latter is involved in transcription and mRNA splicing, and is upregulated in a variety of cancers, where it seems to drive cell growth and proliferation (*64*). We found that depletion of ERK2 induces upregulation of this signaling axis, revealing another way in which cells respond to the lack of this kinase. On the contrary, CDK12 is downregulated in the ERK2 depleted cells. This kinase has tumor suppressing activity in ovarian, breast and prostate cancer, where it controls expression of genes related to DNA damage response (*65-67*). Its downregulation could be a response to induce cell proliferation in ERK2-depleted cells. Finally, we also observed upregulation of widely expressed, constitutive kinases, like PKACB and GSK3A. We think these kinases are also partially responsible for the survival of the ERK2-silenced cells. For example, BAD-S99, which has been reported as target of PKA or AKT (*68*), is upregulated in the absence of ERK2. This phosphosite is known to determine the pore-forming activity of BAD; phosphorylation inhibits its pro-apoptotic activity (*68*). Altogether, we think deletion of ERK2 causes drastic rewiring of kinase activity seemingly responsible to enhance cell survival and proliferation, which is not observed in the ERK1-sgRNA cell line.

By assessing the full proteome of both ERK1-sgRNA (1.2) and ERK1-sgRNA (2.2) cell lines, we found that deletion of either one of the enzymes induced differences in protein expression. Still, it was evident that the more drastic changes were observed in ERK2-silenced cells. Regardless of having proteome dependent changes, we concluded from our time course phosphoproteomic experiment that absence of ERK2 also has a differential impact on phosphorylation dynamics when compared to the control and ERK1-depleted cells. Our data showed that ERK2 is a more efficient kinase when compared to ERK1, which might explain why most phenotypes are observed in absence of ERK2. It could also explain why in certain scenarios, overexpression of ERK1 manages to rescue the lack of ERK2. Importantly, most drastic phenotypes have been observed after ERK2 depletion during embryonic development. The latter consists of a sequential set of accurately timed events that take place for proper organism development. We wonder if the lack of a 'fast' kinase able to efficiently phosphorylate substrates might be partly the cause for the severe phenotypes. Even though we showed that phosphosites with a faster phosphorylation rate have a higher proportion of ERK full motifs, when we compared the phosphorylation rate of phosphosites with the ERK full motif against those with the ERK minimal motif side by side, we could not find any real differences in the timing of these phosphorylation events. The latter would suggest other substrate features different from the motif determine the phosphosite phosphorylation rate.

In conclusion, our work adds to the evidence suggesting that ERK1 and ERK2 exert differential control over the phosphoproteome, and ultimately the proteome, with ERK2 seemingly having a more relevant role in cellular metabolism and proliferation (*69*). Moreover, we also provide evidence of the high plasticity of phosphorylation driven signaling networks, since despite the absence of either one of the core kinases of the MAPK pathway, cells managed to adapt and continued to proliferate. We think further efforts are needed to understand the mechanistic role of each of these enzymes on cellular signaling and their respective impact on cellular homeostasis.

Methods

CRISPR-Cas9 gene disruption and cell line screening

Lentiviral vector production

To produce the lentiviral vectors, HEK293T cells were transfected with pLentiCRISPR-v2 (Addgene, 52961) plasmids containing sgRNA sequences against ERK1 or ERK2 or a non-targeting sgRNA as negative control (see Table 1) together with packaging plasmids psPAX2 (Addgene, 12260) and pMD2G VSV-G (Addgene, 12259). Media containing lentiviral vectors were harvested 2 and 3 days post-transfection, pooled and clarified by centrifugation. pLentiCRISP-v2 plasmids were a kind gift of Professor Dr. Daniel Peeper from Netherlands Cancer Institute (NKI).

Table 1. sgRNA sequences used in this study.

Number	Name	sgRNA sequence	Cell line
PvB883	pLentiCRISPR-v2 MAPK1-sgRNA#1	CAACCTCTCGTACATCGGCG	ERK2-sgRNA (2.1)
PvB884	pLentiCRISPR-v2 MAPK1-sgRNA#2	CGCGGGCAGGTGTTCGACGT	ERK2-sgRNA (2.2)
PvB886	pLentiCRISPR-v2 MAPK3-sgRNA#3	GCGTAGCCACATACTCCGTC	ERK1-sgRNA (1.2) ERK1-sgRNA (1.2)
PvB887	pLentiCRISPR-v2 Ctrl-sgRNA	-	Control (WT)

Cell line production

MCF10A cells were infected with viral vectors in the presence of 8 µg/mL polybrene (Sigma-Aldrich, H9268). Two days after infections, cells were selected by media containing 2.5 ug/mL Puromycin Dihydrochloride (Thermo Fisher Scientific/Gibco, A1113803, 10 mg/mL). After selection was complete, isogenic cell lines were obtained by single-cell plating in a 48-well cluster dish. For each construct around 20 MCF10A clones were selected and expanded for gDNA isolated (Isolation II Genomic DNA kit, Bioline, BIO-52067). To determine gene disruption by InDel formation, genomic PCR reactions were performed using primers as indicated in Table 2. Resulting PCR products were subcloned into pJET1.2/blunt cloning vector (Thermo Scientific CloneJET PCR Cloning Kit, #K1231) and transformed into competent *E.coli* cells (Bioline, compentent cells, Silver Efficiency, BIO-85026). For each MCF10A clone, several *E.coli* clones were expanded, plasmid DNA isolated and verified by Sanger sequencing (Macrogen). For both ERK1 and ERK2 we selected clones that showed disruption of the gene by InDels (Figure S1A).

Table 2. Primer sequences used in this study.

Number	Sequence	Name
2218	GCAAAGGGAACAGTGGAGGC	Fw2_MAPK3/ERK1-sgRNA#3
2220	AGGGTAGAATTCCTGTGTCATGGG	Rv2_MAPK3/ERK1-sgRNA#3
2390	TCTCACTTCAGCCCATGAGTTCTG	Fw1_MAPK1/ERK2_sgRNA#1/#2
2394	CAACAACACCTAAACCTCAACCGG	Rv1_MAPK1/ERK2_sgRNA#1/#2
2135	CGACTCACTATAGGGAGAGCGGC	pJET1.2 forward sequencing primer
2136	AAGAACATCGATTTTCCATGGCAG	pJET1.2 reverse sequencing primer

Cell culture

MCF10A epithelial non-transformed cells were grown in DMEM-F12 supplemented medium (0.5 mg/mL hydrocortisone, 10 ug/mL insulin, 20 ng/mL EGF, 100 ng/mL cholera toxin, 5% horse serum and Pen-Strep 1%). For starvation, cells were kept in non-supplemented medium for 24 hours before harvesting. After starvation, cells were treated with fully supplemented DMEM F-12. For the time course experiment, cells were starved of growth factors, which were then reintroduced after 24 hours (10 ug/mL insulin, 20 ng/mL EGF and 5% horse serum). All experiments were done in 15 cm diameter dishes seeded with 8e10^{6} cells 24 hours before inducing starvation. Cells were scrapped from plates and washed once with ice-cold PBS. Next, dry cell pellets were snap frozen in liquid nitrogen and stored at -80°C until further processing.

Sample preparation for mass spectrometry measurements

Cells were lysed with sodium deoxycholate (SDC) 1%. Lysis buffer was also supplemented with 10 mM tris(2-carboxyethyl)- phosphine hydrochloride (TCEP), 40 mM chloroacetamide, 100 mM TRIS pH 8.0, protease inhibitor (cOmplete mini ethylenediaminetetraacetic acid (EDTA)-free; Roche, Basel, Switzerland) and a phosphatase inhibitor (PhosSTOP, Roche). Next, samples were sonicated using a Bioruptor Plus (Diagenode, Liege, Belgium) for 15 cycles of 30 seconds. The protein amount in each sample was quantified by a Bradford protein assay. Proteins were digested overnight at 37 °C with Lys-C (FUJIFILM Wako pure chemical corporation, Osaka, Japan) and trypsin (Sigma-Aldrich, Zwijndrecht, The Netherlands), with enzyme to protein ratios of 1:75 and 1:50, respectively. SDC was precipitated by the addition of 2% formic acid (FA), and peptides were vacuum dried and stored at -80°C.

Peptides were reconstituted in 2% FA and loaded into C18 stage-tips (Agilent Technologies) for peptide cleanup using the automated AssayMAP Bravo Platform (Agilent Technologies). Stage-tips were first washed with 80% acetonitrile (ACN) and 0.1% FA. Next, they were equilibrated with 0.1% FA before peptide loading at a 10 uL/min speed. Peptides were eluted with 80% ACN, 0.1% FA. Before phosphopeptide enrichment, samples used for targeted phosphoproteomics measurements were spiked with heavy labeled standards of phosphopeptides of interest. Peptides were subjected to phosphopeptide enrichment using Fe(III)-NTA 5µL cartridges (Agilent Technologies) in the AssayMAP Bravo Platform, as described by Post *et al* (*70*). Both flow through (peptides) and eluates (phosphopeptides) were dried down and stored at -80°C until further use.

LC-MS/MS analysis

For the screening phosphoproteomics experiment, samples were analyzed using an Ultimate 3000 uHPLC system coupled to an Orbitrap HF-X (Thermo Fisher Scientific, Waltham, Massachusetts). Peptides were separated using a nanoflow rate of 300 nL/min on an analytical column (ID of 75 µm and 50 cm length; packed in-house with 2.7 µm Poroshell EC-C18 particles [Agilent]). We used a two-system buffer consisting of solvent A (0.1% FA in water) and B (0.1% FA in 80% ACN). We used a 98 minutes gradient, from 9% to 32% of solvent B. This was followed by a ~5 minutes wash with 99% solvent B and a 10 minutes column equilibration with 8% solvent B.

The mass spectrometer was operated in data dependent acquisition (DDA) mode. Full scan MS was acquired from 375-1600m/z with a 60,000 resolution at 200m/z. Automotatic gain control (AGC) target was set to 3e6 ions with a maximum injection time of 20ms. Up to 15 of the most intense precursor ions were isolated (1.4m/z window) for fragmentation using high energy collision induced dissociation (HCD) with a normalised collision energy of 27. For MS2 scans an accumulation target value of 1e5 ions and a maximum injection time of 50ms were selected. Scans were acquired from 200-2000m/z with a 30,000 resolution at 200m/z. Dynamic exclusion was set at 12 seconds.

The proteome samples were analyzed using an Ultimate 3000 uHPLC system coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific, Waltham, Massachusetts). For these samples we used a 157 minute gradient, from 9% to 44% of solvent B. This was followed by a ~5 minutes wash with 99% solvent B and a 10 minutes column equilibration with 9% solvent B. The mass spectrometer was operated DDA mode, with a full scan from 375 to 1600 m/z at 60,000 resolution (at 200 m/z). RF lens (%) was set to 40 and the AGC target was set to 'standard' with the maximum injection time mode set to 'auto'. The latter two are supposed to maximize sensitivity while maintaining a maximum scan rate. A minimum intensity threshold of 50,000 was used to trigger a data dependent scan. Precursors were selected for fragmentation with an isolation window of 1.4 m/z. AGC target and injection window were also set as 'standard' and 'auto'. Precursors were fragmented with an HCD collision energy of 28%. MS2 scans started at 120 m/z with a 30,000 resolution at 200 m/z. Precursors were added to the dynamic exclusion list for 24 seconds after being fragmented once.

Samples from the time course phosphoproteomics experiment were measured using the same setup for the proteome measurements described before, with a few exceptions; we used a 98 minutes gradient from 9% to 36% solvent B. Next we did a ~5 minutes wash with 99% solvent B followed by a 10 minute column equilibration with 9% solvent B. The mass spectrometer was also operated in DDA mode with the same parameters for the full MS, fragmentation and MS2 scans as described before, with the only exception that precursors were kept in the dynamic exclusion list for a shorter time (16 seconds).

Single reaction monitoring (SRM) measurements were done on an Ultimate 3000 uHPLC system coupled to a TSQ Altis Triple Quadrupole (Thermo Fisher Scientific, Waltham, Massachusetts). Peptides were reconstituted in 20mM citric acid, loaded on a trap-column (C18 PepMap100, 5 µm) and separated on a PepMap RSLC C18 column (2µm, 75µm x 25cm) using a 100 min gradient (2.2% to 34% solvent B) at a flow rate of 300 nl/min. Retention time windows were set to 5 minutes, with Q1 and Q3 resolution set to 0.7 m/z and a cycle time of 5 seconds.

Data processing

Raw files were searched against the human SwissProt database (accessed in March 2021) with Maxquant (*71*) (v1.6.17.0, v1.6.8.0, v2.0.3.0). For the proteome database search, the default settings were used, with the following exceptions: methionine oxidation and protein N-term acetylation were set as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. 'Match between' runs was enabled with the default parameters, and fractions were set so that matching was only done between samples from the same cell line. For the phosphoproteome database search similar conditions were used, including variable phosphorylation of serine, threonine and tyrosine. For the time course experiment, the same parameters were used, with the exception that this time fractions were set so that matching was only done between bioreplicates of the same time point and the same cell line.

Results were further analysed using the Perseus platform (*72*). Here, decoy sequences, potential contaminants and proteins only identified by peptides that carry one or more modifications were removed. Intensities were log2 transformed and normalized by subtracting the median intensity of each sample. For the time course experiment no median normalization was done since this proved to be unnecessary (data already displayed normal distribution). Samples were grouped by condition (starved or GF treated) and cell line (sgRNA-ERK1, sgRNA-ERK2, sgRNA-Ctrl). Data was subsequently filtered using this grouping, so that only proteins or phosphosites detected in at least all three bio replicates of one of the groups were kept for quantification. For the time course experiment we used a more stringent strategy for data filtering, in which only phosphosites detected in at least 70% of all samples of one cell line were kept for further analysis. In all experiments missing values were replaced from a random normal distribution using a downshift of 1.8 times the standard deviation of the dataset, and a width of 0.3 times the standard deviation. The latter effectively replaced missing values at the lower end of the intensity distribution. Next, ANOVA (Benjamini-Hochberg FDR, <0.05) test was used to determine which phosphosites (or proteins) changed between samples.

Data was imported to the R environment for hierarchical clustering using k-means in the ComplexHeatmap package (*73*) and for visualization using ggplot2 package (https://ggplot2.tidyverse.org/). Gene ontology (GO) analysis was done using Panther website (*74*). Results from either GO complete or GO slim annotation sets were manually filtered to show only the top hits on the list (for full list of enriched GO terms check supplementary material).

Sequence windows centered on the detected phosphosites were used for motif analysis. This was done with an in house script using regular expressions applying the following rules:

-Polo like kinase: [D/N/E/Y]-X-[S/T]-[Hydrophobic / ^P]

-Aurora kinase A/B: [K/R]-X-[S/T]*[^P]

-Casein kinase 1: [D/E]-[D/E]-[D/E]-X-X-[S/T]* or [S/T]-X-X-[S/T]*

-Casein kinase 2:[S/T]-[S/T]*-X-[E/D/S]

-DNA dependent kinase (DDK): [S/T]*-[E/D]-X-[E/D] or [S/T]*-[S/T]-P

-Cyclin dependent kinase full consensus motif: [S/T]*-P-X-[K/R]

-Cyclin dependent kinase and ERK minimal consensus motif: [S/T]*-P

-ERK: P-X-[S/T]*-P

-Glycogen synthase kinase 3 (alpha/beta): S-P-X-X-[S/T]*-P or [S/T]*-P-X-X-S-P

-Shared motif for AKT (isoforms 1-3), p70RSK and p90S6K: R-X-R-X-X-[S/T]*

-Shared motif for Protein kinase A, Protein kinase C and PAK (isoforms 1-4): [R/K]-[R/K]-X-[S/T]*-[Hydrophobic]

Rules were defined using the substrate logo described for each enzyme of interest in PhosphoSitePlus (*35*). Sequence logos of the phosphorylation sites were obtained for each cluster using all detected phosphorylation sites for each experiment as a background (*75*).

Lists of proteins from certain clusters were uploaded to Cytoscape to obtain interaction information from STRING database. We used a medium confidence cut off and removed singletons from the final figures. GO analysis for sub clusters inside each network was done with BiNGO.

Raw files from SRM measurements were loaded into Skyline (*76*). Signal quality for each target of interest was assessed visually for all samples. Quality control of endogenous signals was done by confirming the perfect co-elution of both peptide forms (heavy and light), assessing their retention time and peak shape. Light to heavy (L/H) ratios were loaded into R for data visualization using ComplexHeatmap package and ggplot 2 package.

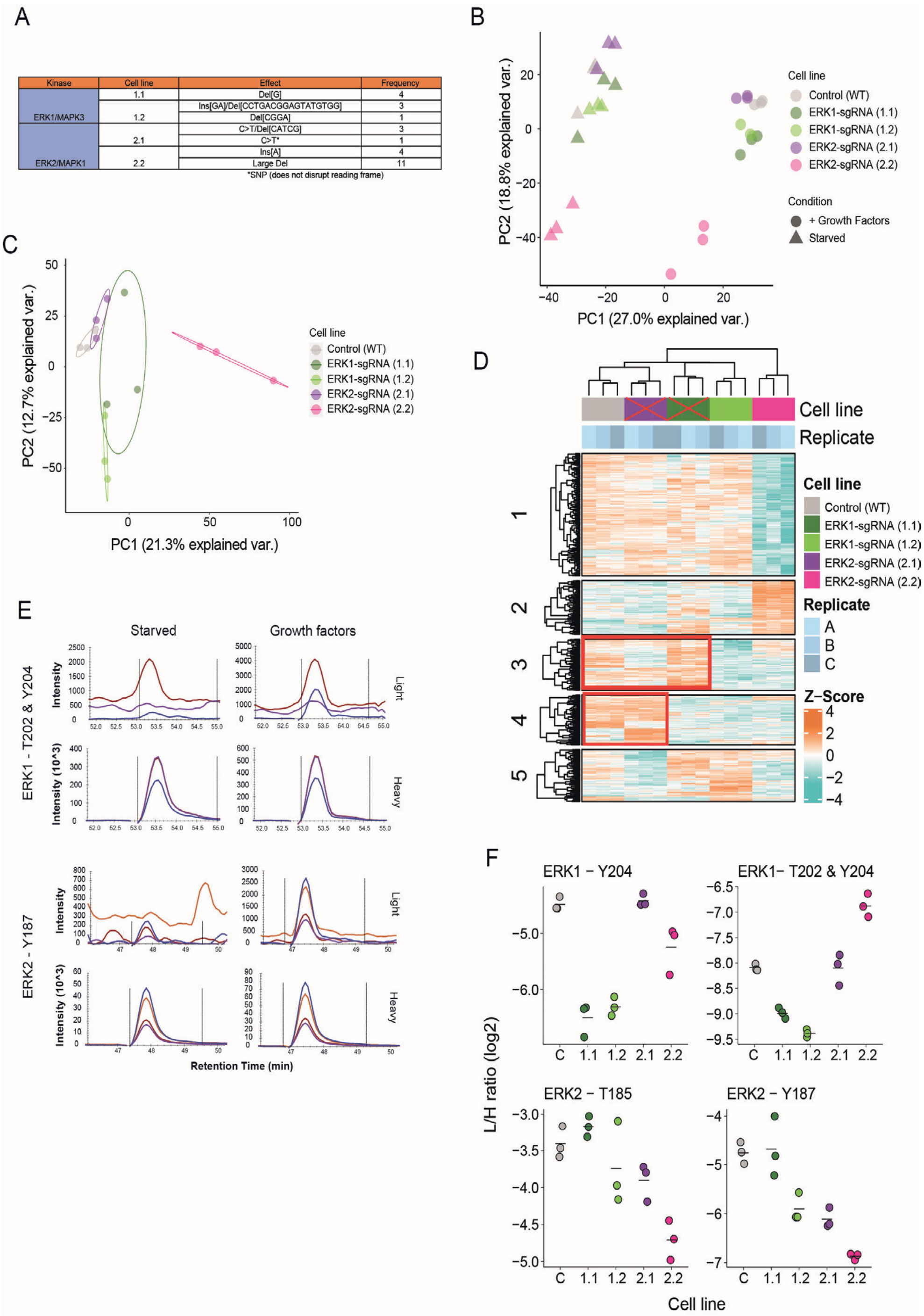
Supplementary material

Supplementary table 1. Potential new targets of ERK1.

Potential new ERK1 substrates		
Uniprot ID	**Gene name**	**Phosphosite**
Q96AP0	ACD	S339
Q96JM3	CHAMP1	S427
Q96TA1	FAM129B	S665
Q9P206	KIAA1522	S612
Q9P206	KIAA1522	S929
Q9Y618	NCOR2	S2054
Q7Z5J4	RAI1	T1068
Q96C92	SDCCAG3	S243
Q13425	SNTB2	S95
O75044	SRGAP2	S994
P16949	STMN1	S38
P82094	TMF1	S72
Q15642	TRIP10	S335
Q15642	TRIP10	S351

Supplementary table 2. Potential new targets of ERK2.

Potential new ERK2 substrates		
Uniprot ID	**Gene name**	**Phosphosite**
O43306	ADCY6	S54
Q9Y679	AUP1	S288
Q16204	CCDC6	S367
Q96JM3	CHAMP1	S319
Q96JM3	CHAMP1	S427
P78310	CXADR	S332
Q9NPF5	DMAP1	T445
Q86VR2	FAM134C	T10
Q5T0W9	FAM83B	T583
Q5T5P2	KIAA1217	S1794
Q5T5P2	KIAA1217	S313
Q5T5P2	KIAA1217	S326
Q9P206	KIAA1522	S979
Q9BY89	KIAA1671	S1757
Q9BY89	KIAA1671	S1760
Q6PKG0	LARP1	S774
Q5VZK9	LRRC16A	S1094
Q7Z434	MAVS	S152
Q14980	NUMA1	S1757
Q96JY6	PDLIM2	S137
A1L390	PLEKHG3	S76
Q96T60	PNKP	T122
Q6NYC8	PPP1R18	S432
Q9GZR2	REXO4	S15
P62070	RRAS2	S186
Q92766	RREB1	S161
O95197	RTN3	S246
Q9UQR0	SCML2	S511
Q14160	SCRIB	S1306
Q14160	SCRIB	S1309
P04920	SLC4A2	S173
O60264	SMARCA5	S116
Q13425	SNTB2	S95
Q8WXH0	SYNE2	S6389
O95359	TACC2	S2317
Q9C0C2	TNKS1BP1	T131
Q68CZ2	TNS3	S866
Q04323	UBXN1	S200



Kinase	Cell line	Effect	Frequency
ERK1/MAPK3	1.1	Del[G]	4
	1.2	Ins[GA]/Del[CCTGACGGAGTATGTGG]	3
		Del[CGGA]	1
ERK2/MAPK1	2.1	C>T/Del[CATCG]	3
		C>T*	1
	2.2	Ins[A]	4
		Large Del	11

*SNP (does not disrupt reading frame)

Figure S1. ERK1 or ERK2 knock-out cell line generation. (A) Results of the genetic screening used to determine presence of mutations in ERK1 and ERK2 genes. (B) and (C) PCA analysis of all quantified phosphorylation sites among all samples and among GF treated samples only (*n*=5, 705). (D) Hierarchical clustering of significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05) in GF treated samples (*n*=1,170). Red frames show similitudes between samples with no clear genetic evidence of a complete ERK1 or ERK2 deletion (ERK1-sgRNA [1.2] and ERK2-sgRNA [2.1], respectively) and the control. Red crosses depict cell lines that were discarded and not considered in any further data analysis. (E) Extracted ion chromatograms showing the upregulation of ERK1 and ERK2 phosphorylation sites after GF treatment. Top panels show the signal of endogenous phosphopeptides while bottom panels show the signal of heavy labeled standards. (F) Dot plots showing the levels of ERK1 and ERK2 phosphorylation sites among the different cell lines. ERK1-sgRNA (1.2) and ERK2-sgRNA (2.2) showed the lowest levels of activating phosphorylation sites of ERK1 and ERK2, respectively. SRM confirmation of effective kinase deletion was used as another criterion to pick these cell lines for further data analysis.

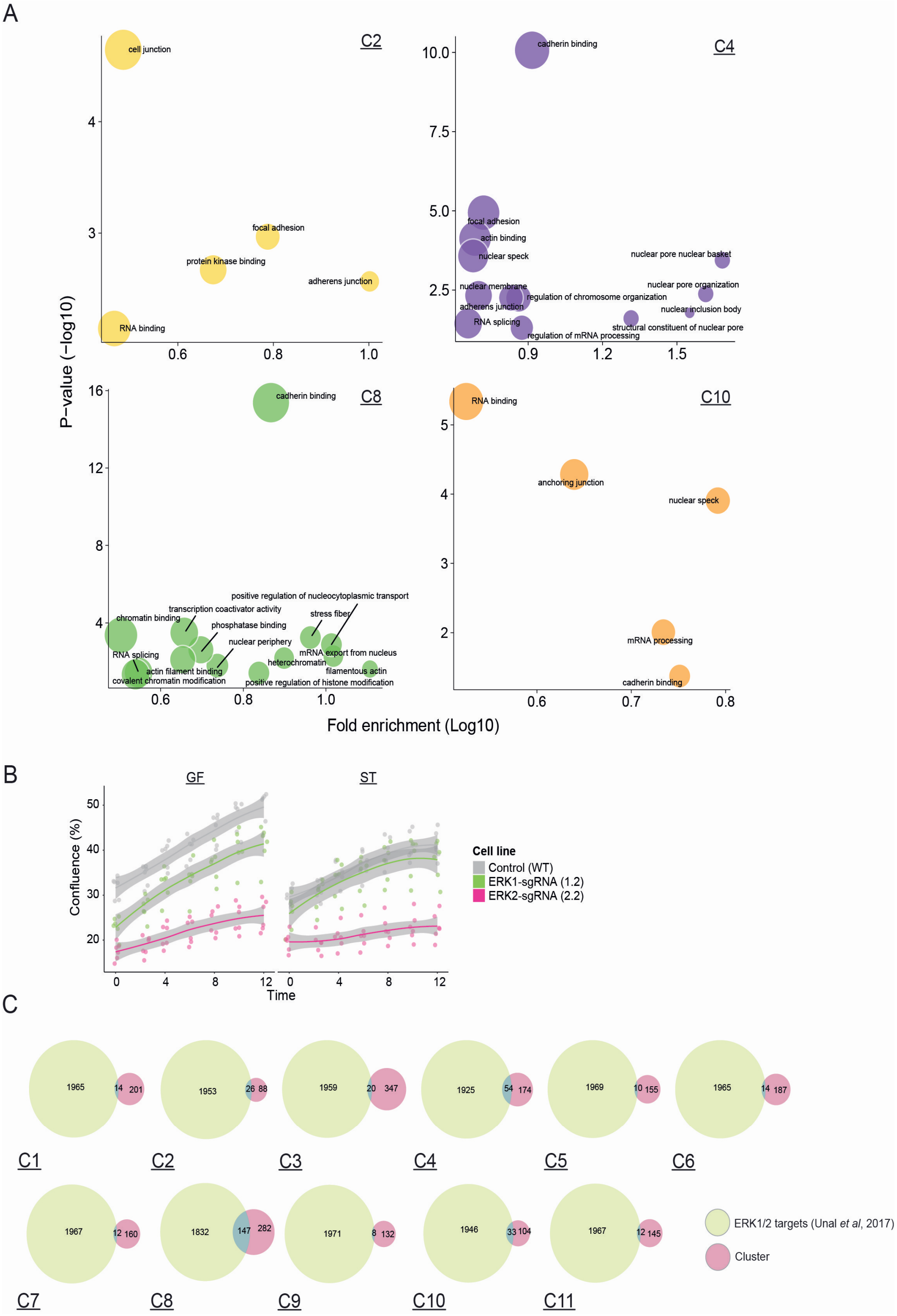

A
C2
cell junction
focal adhesion
protein kinase binding
adherens junction
RNA binding
4
3
0.6
0.8
1.0
C4
cadherin binding
focal adhesion
actin binding
nuclear speck
nuclear pore nuclear basket
nuclear membrane
nuclear pore organization
regulation of chromosome organization
adherens junction
nuclear inclusion body
RNA splicing
structural constituent of nuclear pore
regulation of mRNA processing
10.0
7.5
5.0
2.5
0.9
1.2
1.5
P−value (−log10)
C8
cadherin binding
positive regulation of nucleocytoplasmic transport
transcription coactivator activity
stress fiber
chromatin binding
phosphatase binding
nuclear periphery
RNA splicing
mRNA export from nucleus
heterochromatin
actin filament binding
filamentous actin
covalent chromatin modification
positive regulation of histone modification
16
12
8
4
0.6
0.8
1.0
C10
RNA binding
anchoring junction
nuclear speck
mRNA processing
cadherin binding
5
4
3
2
0.6
0.7
0.8
Fold enrichment (Log10)
B
GF
ST
Confluence (%)
50
40
30
20
0
4
8
12
Time
Cell line
Control (WT)
ERK1-sgRNA (1.2)
ERK2-sgRNA (2.2)
C
1965 14 201
C1
1953 26 88
C2
1959 20 347
C3
1925 54 174
C4
1969 10 155
C5
1965 14 187
C6
1967 12 160
C7
1832 147 282
C8
1971 8 132
C9
1946 33 104
C10
1967 12 145
C11
ERK1/2 targets (Unal et al, 2017)
Cluster

Figure S2. Differences between ERK1 and ERK2 depleted cells and their substrates. (A) GO analysis of clusters that contained a high proportion of ERK full motifs. The size of the dot reflects the number of proteins related to the GO term (see materials and methods for details). (B) Growth curves for control (WT), ERK1-sgRNA (1.2) and ERK2-sgRNA (2.2) cell lines during 12 hours after starvation or treatment with GFs. (C) Overlap of proteins from each of the clusters shown on Figure 1D versus a database of previously known ERK1/2 substrates.

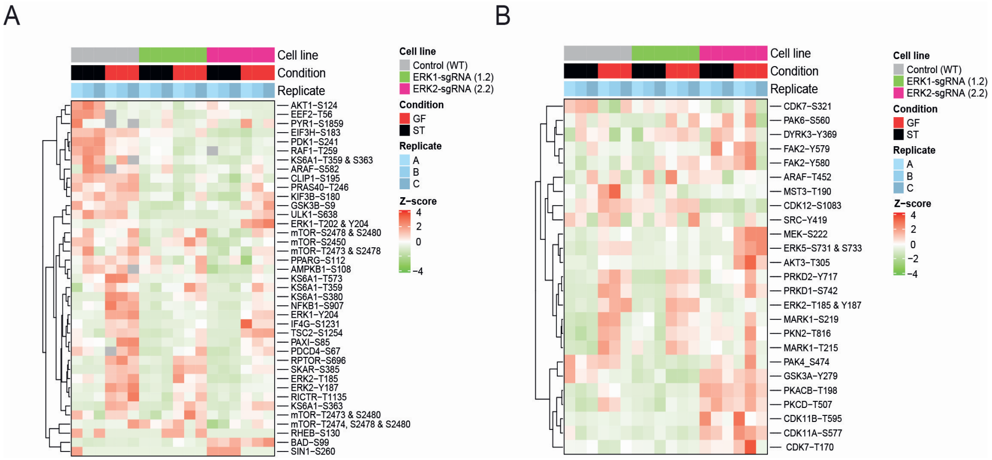


Figure S3. Targeted phosphoproteomics reveal differences between ERK1 and ERK2 impact on cellular signaling. (A) Heat map showing the behavior of the 39 phosphopeptides of the MAPK-AKT-mTOR network quantified among samples. (B) Heat map showing the behavior of the 25 phosphopeptides with regulatory phosphorylation sites of 22 different kinases.

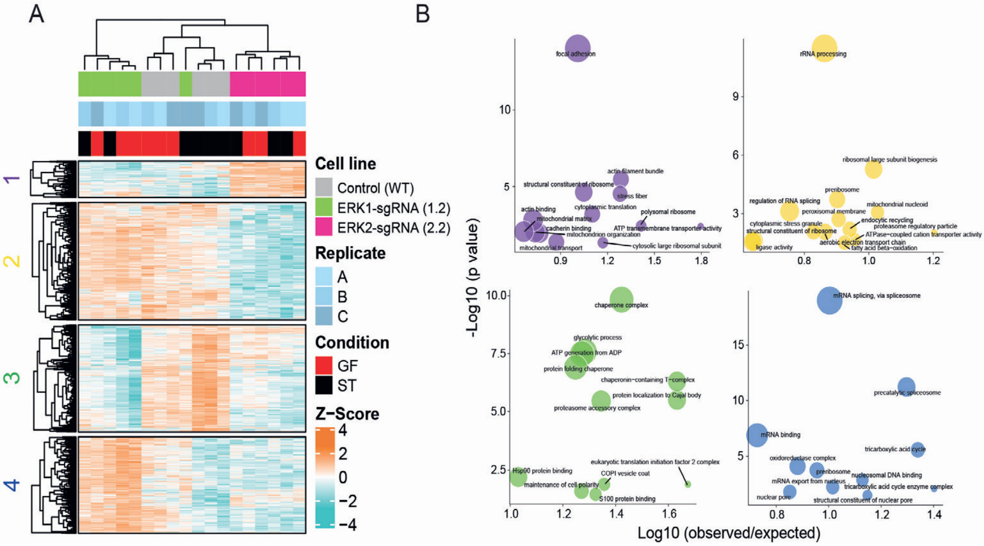


Figure S4. Changes in the proteome after depletion of either ERK1 or ERK2. (A) Hierarchical clustering of significantly changing proteins (ANOVA, Benjamini-Hochberg correction, FDR 0.05). (B) GO analysis of the clusters shown in (A).

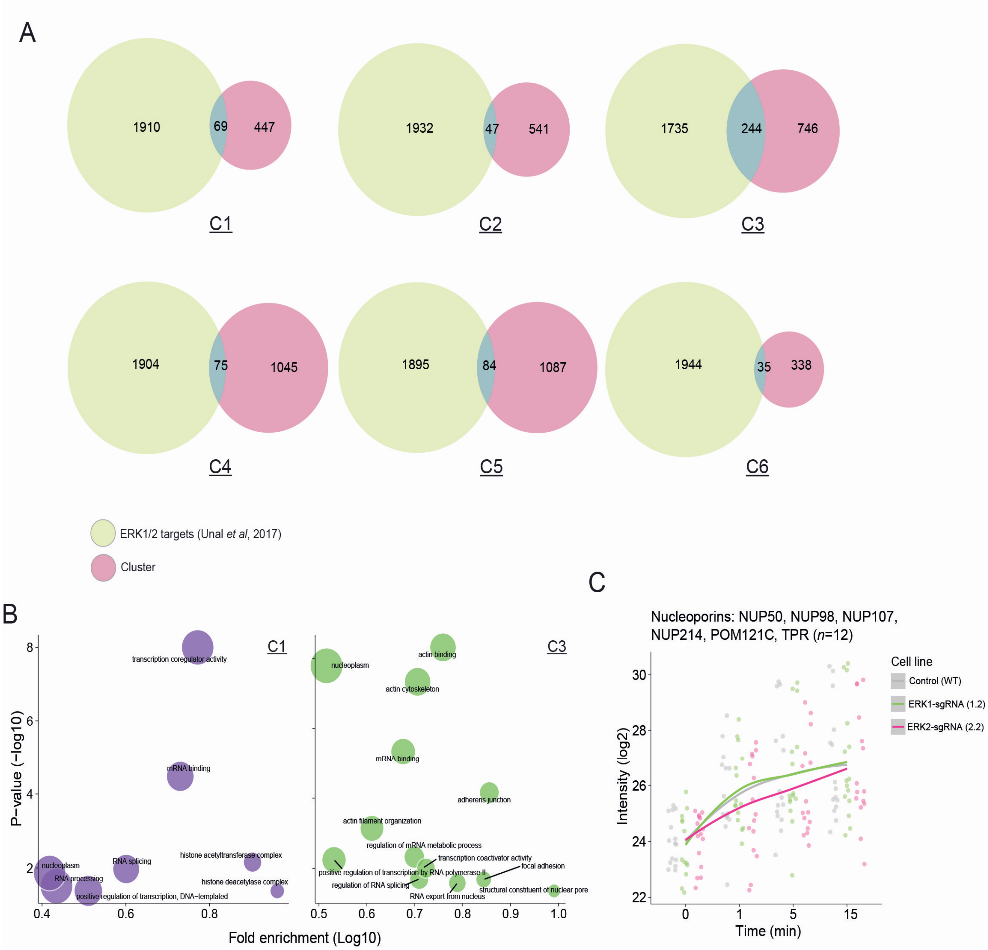


Figure S5. Comparison between dynamic phosphorylation sites. (A) GO analysis of Clusters 1 and 3 of the time course experiment shown on Figure 6. (B) Venn diagrams showing the overlap of each cluster from Figure 6D against a database of ERK1/2 substrates. (C) Smoothed curve showing the trend of phosphorylation sites located in nucleoporins.

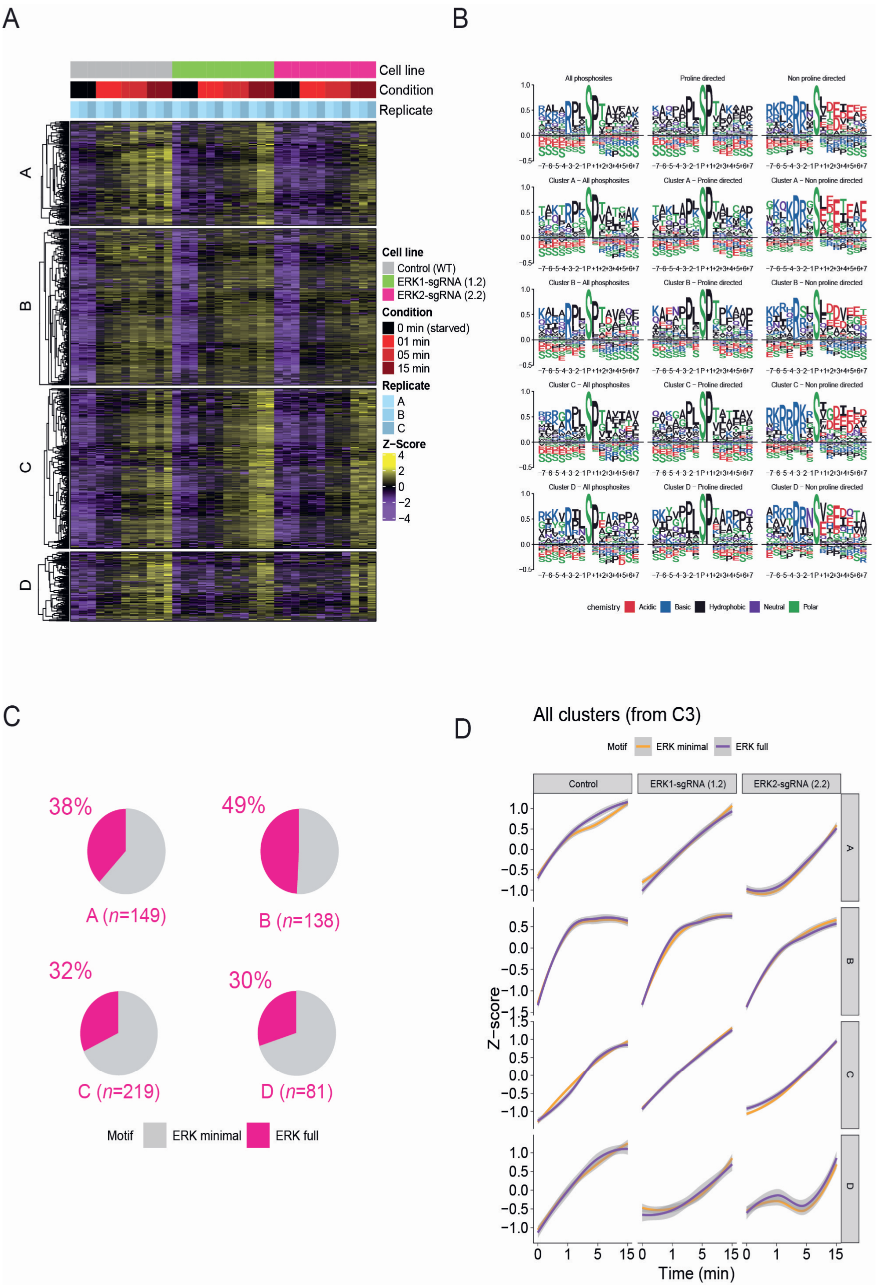


Figure S6. Motif analysis of phosphosites with a slower phosphorylation rate upon ERK2 depletion. (A) Hierarchical clustering of phosphosites located in C3 (Figure 6D). (B) and (C) Phosphorylation logo and pie charts showing increased overrepresentation of ERK1/2 full motifs in cluster B when compared to the other clusters. **Legend continues on next page.**

Figure S6 legend continued. Logos on (D) are shown separately for all phosphorylation sites, proline directed phosphosites and non-proline directed phosphosites. (D) Smoothed curves for all phosphosites with the minimal ERK motif or full ERK motif. Trends are shown for separately for each cell line and each cluster from (A).

Acknowledgements

This work has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the European Union and the NWO funded Netherlands Proteomics Centre through the National Road Map for Large-scale Infrastructures program X-Omics, Project 184.034.019. JMV is supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR).

References

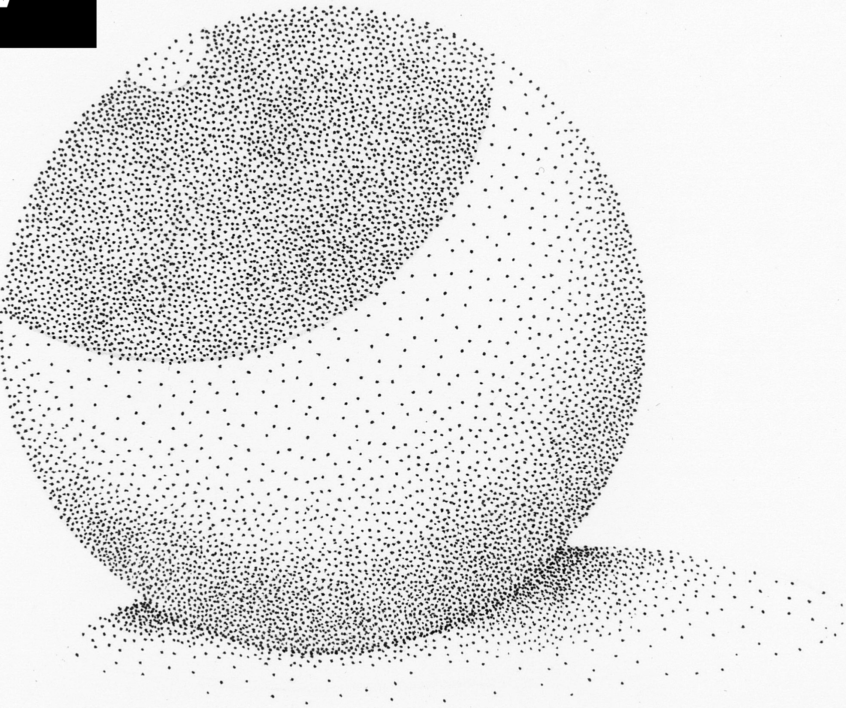
1. T. Pawson, J. D. Scott, Protein phosphorylation in signaling--50 years and counting. *Trends Biochem Sci* **30**, 286-290 (2005).
2. H. Lavoie, J. Gagnon, M. Therrien, ERK signalling: a master regulator of cell behaviour, life and fate. *Nat Rev Mol Cell Biol* **21**, 607-632 (2020).
3. W. Zhang, H. T. Liu, MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res* **12**, 9-18 (2002).
4. D. K. Morrison, MAP kinase pathways. *Cold Spring Harb Perspect Biol* **4**, (2012).
5. M. Hoshi, E. Nishida, H. Sakai, Activation of a Ca2+-inhibitable protein kinase that phosphorylates microtubule-associated protein 2 in vitro by growth factors, phorbol esters, and serum in quiescent cultured human fibroblasts. *J Biol Chem* **263**, 5396-5401 (1988).
6. L. B. Ray, T. W. Sturgill, Rapid stimulation by insulin of a serine/threonine kinase in 3T3-L1 adipocytes that phosphorylates microtubule-associated protein 2 in vitro. *Proc Natl Acad Sci U S A* **84**, 1502-1506 (1987).
7. R. Busca, J. Pouyssegur, P. Lenormand, ERK1 and ERK2 Map Kinases: Specific Roles or Functional Redundancy? *Front Cell Dev Biol* **4**, 53 (2016).
8. R. Busca *et al.*, ERK1 and ERK2 present functional redundancy in tetrapods despite higher evolution rate of ERK1. *BMC Evol Biol* **15**, 179 (2015).
9. G. Pages *et al.*, Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* **286**, 1374-1377 (1999).
10. M. K. Saba-El-Leil *et al.*, An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep* **4**, 964-968 (2003).
11. N. Hatano *et al.*, Essential role for ERK2 mitogen-activated protein kinase in placental development. *Genes Cells* **8**, 847-856 (2003).
12. Y. Yao *et al.*, Extracellular signal-regulated kinase 2 is necessary for mesoderm differentiation. *Proc Natl Acad Sci U S A* **100**, 12759-12764 (2003).
13. M. Aouadi, B. Binetruy, L. Caron, Y. Le Marchand-Brustel, F. Bost, Role of MAPKs in development and differentiation: lessons from knockout mice. *Biochimie* **88**, 1091-1098 (2006).
14. S. F. Krens *et al.*, Distinct functions for ERK1 and ERK2 in cell migration processes during zebrafish gastrulation. *Dev Biol* **319**, 370-383 (2008).

15. S. F. Krens, M. Corredor-Adamez, S. He, B. E. Snaar-Jagalska, H. P. Spaink, ERK1 and ERK2 MAPK are key regulators of distinct gene sets in zebrafish embryogenesis. *BMC Genomics* **9**, 196 (2008).
16. F. Bost *et al.*, The extracellular signal-regulated kinase isoform ERK1 is specifically required for in vitro and in vivo adipogenesis. *Diabetes* **54**, 402-411 (2005).
17. S. Guihard *et al.*, The MAPK ERK1 is a negative regulator of the adult steady-state splenic erythropoiesis. *Blood* **115**, 3686-3694 (2010).
18. C. Bourcier *et al.*, p44 mitogen-activated protein kinase (extracellular signal-regulated kinase 1)-dependent signaling contributes to epithelial skin carcinogenesis. *Cancer Res* **66**, 2700-2707 (2006).
19. C. Fremin, M. K. Saba-El-Leil, K. Levesque, S. L. Ang, S. Meloche, Functional Redundancy of ERK1 and ERK2 MAP Kinases during Development. *Cell Rep* **12**, 913-921 (2015).
20. E. T. Richardson *et al.*, ERK Signaling Is Essential for Macrophage Development. *PLoS One* **10**, e0140064 (2015).
21. Y. He *et al.*, Erk1 positively regulates osteoclast differentiation and bone resorptive activity. *PLoS One* **6**, e24780 (2011).
22. Y. Satoh *et al.*, ERK2 contributes to the control of social behaviors in mice. *J Neurosci* **31**, 11953-11967 (2011).
23. Y. Otsubo *et al.*, Mechanical allodynia but not thermal hyperalgesia is impaired in mice deficient for ERK2 in the central nervous system. *Pain* **153**, 2241-2252 (2012).
24. W. B. Hamilton, K. Kaji, T. Kunath, ERK2 suppresses self-renewal capacity of embryonic stem cells, but is not required for multi-lineage commitment. *PLoS One* **8**, e60907 (2013).
25. C. Pan, J. V. Olsen, H. Daub, M. Mann, Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* **8**, 2796-2808 (2009).
26. T. S. Lewis *et al.*, Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol Cell* **6**, 1343-1354 (2000).
27. H. Kosako *et al.*, Phosphoproteomics reveals new ERK MAP kinase targets and links ERK to nucleoporin-mediated nuclear transport. *Nat Struct Mol Biol* **16**, 1026-1035 (2009).
28. F. Gnad *et al.*, Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms. *Proteomics* **16**, 1998-2004 (2016).
29. S. M. Carlson *et al.*, Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. *Sci Signal* **4**, rs11 (2011).
30. U. Marklund, G. Brattsand, O. Osterman, P. I. Ohlsson, M. Gullberg, Multiple signal transduction pathways induce phosphorylation of serines 16, 25, and 38 of oncoprotein 18 in T lymphocytes. *J Biol Chem* **268**, 25671-25680 (1993).
31. U. Marklund, G. Brattsand, V. Shingler, M. Gullberg, Serine 25 of oncoprotein 18 is a major cytosolic target for the mitogen-activated protein kinase. *J Biol Chem* **268**, 15039-15047 (1993).
32. Z.-Y. Zheng *et al.*, STAT3β disrupted mitochondrial electron transport chain enhances chemosensitivity by inducing pyroptosis in esophageal squamous cell carcinoma. *Cancer Letters* **522**, 171-183 (2021).
33. M. I. Linder *et al.*, Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. *Dev Cell* **43**, 141-156 e147 (2017).
34. E. B. Unal, F. Uhlitz, N. Bluthgen, A compendium of ERK targets. *FEBS Lett* **591**, 2607-2615 (2017).
35. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520 (2015).

36. D. O. Debets, J. M. Valverde, M. Altelaar, Capturing the signalling dynamics of the MAPK-AKT-mTOR pathway in a single targeted phosphoproteomics assay. *bioRxiv*, 2022.2001.2017.476555 (2022).
37. T. Schmidlin *et al.*, High-Throughput Assessment of Kinome-wide Activation States. *Cell Syst* **9**, 366-374 e365 (2019).
38. K. S. Kovacina *et al.*, Identification of a proline-rich Akt substrate as a 14-3-3 binding partner. *J Biol Chem* **278**, 10189-10194 (2003).
39. E. Vander Haar, S. I. Lee, S. Bandhakavi, T. J. Griffin, D. H. Kim, Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40. *Nat Cell Biol* **9**, 316-323 (2007).
40. D. R. Alessi, F. B. Caudwell, M. Andjelkovic, B. A. Hemmings, P. Cohen, Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. *FEBS Lett* **399**, 333-338 (1996).
41. M. I. Dobrikov, E. Y. Dobrikova, M. Gromeier, Dynamic regulation of the translation initiation helicase complex by mitogenic signal transduction to eukaryotic translation initiation factor 4G. *Mol Cell Biol* **33**, 937-946 (2013).
42. K. Inoki, Y. Li, T. Zhu, J. Wu, K. L. Guan, TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nat Cell Biol* **4**, 648-657 (2002).
43. Y. Li, K. Inoki, P. Vacratsis, K. L. Guan, The p38 and MK2 kinase cascade phosphorylates tuberin, the tuberous sclerosis 2 gene product, and enhances its interaction with 14-3-3. *J Biol Chem* **278**, 13663-13671 (2003).
44. B. L. Jacobs *et al.*, Identification of mechanically regulated phosphorylation sites on tuberin (TSC2) that control mechanistic target of rapamycin (mTOR) signaling. *J Biol Chem* **292**, 6987-6997 (2017).
45. C.-H. Chen *et al.*, Autoregulation of the Mechanistic Target of Rapamycin (mTOR) Complex 2 Integrity Is Controlled by an ATP-dependent Mechanism*. *Journal of Biological Chemistry* **288**, 27019-27030 (2013).
46. P. Creixell *et al.*, Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* **163**, 202-217 (2015).
47. C. Vantaggiato *et al.*, ERK1 and ERK2 mitogen-activated protein kinases affect Ras-dependent cell signaling differentially. *J Biol* **5**, 14 (2006).
48. J. Lovric, S. Dammeier, A. Kieser, H. Mischak, W. Kolch, Activated raf induces the hyperphosphorylation of stathmin and the reorganization of the microtubule network. *J Biol Chem* **273**, 22848-22855 (1998).
49. Z. Wen, Z. Zhong, J. E. Darnell, Maximal activation of transcription by statl and stat3 requires both tyrosine and serine phosphorylation. *Cell* **82**, 241-250 (1995).
50. Y. Shindo *et al.*, Conversion of graded phosphorylation into switch-like nuclear translocation via autoregulatory mechanisms in ERK signalling. *Nat Commun* **7**, 10485 (2016).
51. M. Marchi *et al.*, The N-terminal domain of ERK1 accounts for the functional differences with ERK2. *PLoS One* **3**, e3873 (2008).
52. P. Langlais, Z. Yi, L. J. Mandarino, The identification of raptor as a substrate for p44/42 MAPK. *Endocrinology* **152**, 1264-1273 (2011).
53. A. Carriere *et al.*, ERK1/2 phosphorylate Raptor to promote Ras-dependent activation of mTOR complex 1 (mTORC1). *J Biol Chem* **286**, 567-577 (2011).
54. P. P. Roux, J. Blenis, ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev* **68**, 320-344 (2004).
55. M. Frodin, S. Gammeltoft, Role and regulation of 90 kDa ribosomal S6 kinase (RSK) in signal transduction. *Mol Cell Endocrinol* **151**, 65-77 (1999).

56. C. J. Richardson *et al.*, SKAR is a specific target of S6 kinase 1 in cell growth control. *Curr Biol* **14**, 1540-1549 (2004).
57. D. Gao *et al.*, Rictor forms a complex with Cullin-1 to promote SGK1 ubiquitination and destruction. *Mol Cell* **39**, 797-808 (2010).
58. D. Sarbassov Dos, A. Guertin David, M. Ali Siraj, M. Sabatini David, Phosphorylation and Regulation of Akt/PKB by the Rictor-mTOR Complex. *Science* **307**, 1098-1101 (2005).
59. Y. Kato *et al.*, Bmk1/Erk5 is required for cell proliferation induced by epidermal growth factor. *Nature* **395**, 713-716 (1998).
60. N. Mody, D. G. Campbell, N. Morrice, M. Peggie, P. Cohen, An analysis of the phosphorylation and activation of extracellular-signal-regulated protein kinase 5 (ERK5) by mitogen-activated protein kinase kinase 5 (MKK5) in vitro. *Biochem J* **372**, 567-575 (2003).
61. T. Honda *et al.*, Phosphorylation of ERK5 on Thr732 is associated with ERK5 nuclear localization and ERK5-dependent transcription. *PLoS One* **10**, e0117914 (2015).
62. S. R. Desai *et al.*, Regulation of Cdk7 activity through a phosphatidylinositol (3)-kinase/PKC-iota-mediated signaling cascade in glioblastoma. *Carcinogenesis* **33**, 10-19 (2012).
63. S. Larochelle *et al.*, Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nat Struct Mol Biol* **13**, 55-62 (2006).
64. Y. Zhou, J. K. Shen, F. J. Hornicek, Q. Kan, Z. Duan, The emerging roles and therapeutic potential of cyclin-dependent kinase 11 (CDK11) in human cancer. *Oncotarget* **7**, 40846-40859 (2016).
65. K. Naidoo *et al.*, Evaluation of CDK12 Protein Expression as a Potential Novel Biomarker for DNA Damage Response-Targeted Therapies in Breast Cancer. *Mol Cancer Ther* **17**, 306-315 (2018).
66. K. M. Ekumi *et al.*, Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res* **43**, 2575-2589 (2015).
67. Y. M. Wu *et al.*, Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell* **173**, 1770-1782 e1714 (2018).
68. L. Polzien *et al.*, Identification of novel in vivo phosphorylation sites of the human proapoptotic protein BAD: pore-forming activity of BAD is regulated by phosphorylation. *J Biol Chem* **284**, 28004-28020 (2009).
69. B. Li *et al.*, Proteomics and Phosphoproteomics Profiling of Drug-Addicted BRAFi-Resistant Melanoma Cells. *J Proteome Res* **20**, 4381-4392 (2021).
70. H. Post *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **16**, 728-737 (2017).
71. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
72. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
73. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
74. H. Mi *et al.*, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res* **49**, D394-D403 (2021).
75. J. Douglass *et al.*, Identifying protein kinase target preferences using mass spectrometry. *Am J Physiol Cell Physiol* **303**, C715-727 (2012).
76. B. MacLean *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).

CHAPTER IV



A CDK-mediated phosphorylation switch of disordered protein condensation

Juan Manuel Valverde[1,2†], Geronimo Dubra[3,4†], Henk van den Toorn[1,2], Guido van Mierlo[5], Michiel Vermeulen[5], Albert J.R. Heck[1,2], Carlos Elena-Real[6], Aurélie Fournet[6], Emile Al Ghoul[7], Dhanvantri Chahar[3,4], Austin Haidar[8], Matteo Paloni[6], Angelos Constantinou[7], Alessandro Barducci[6], Kingshuk Ghosh[8], Nathalie Sibille[6], Pau Bernado[6], Puck Knipscheer[9], Liliana Krasinska[3,4‡], Daniel Fisher[3,4‡], Maarten Altelaar[1,2‡]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, 3584 CH Utrecht, Netherlands.
[2]Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, Netherlands.
[3]IGMM, University of Montpellier, CNRS, Inserm, Montpellier, France.
[4]Equipe Labellisée LIGUE 2018, Ligue Nationale Contre le Cancer, Paris, France.
[5]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands.
[6]CBS, University of Montpellier, INSERM, CNRS, Montpellier, France.
[7]IGH, University of Montpellier, CNRS, Montpellier, France.
[8]Department of Physics and Astronomy, and Department of Molecular and Cellular Biophysics, University of Denver, Denver, Colorado 80208, USA.
[9]Oncode Institute, Hubrecht Institute–KNAW and University Medical Center, Utrecht, 3584 CT, Netherlands.

[†‡] Equal contributions

Manuscript under review

Abstract

Cell cycle transitions arise from collective changes in protein phosphorylation states triggered by cyclin-dependent kinases (CDKs), but conceptual and mechanistic explanations for the abrupt cellular reorganisation that occurs upon mitotic entry are lacking. Specific interactions between distinct CDK-cyclin complexes and sequence motifs encoded in substrates might result in highly ordered phosphorylation, while bistability in the mitotic CDK1 control network can trigger switch-like phosphorylation. Yet the dynamics of mitotic phosphorylation has not been demonstrated *in vivo*, and the roles of most cell cycle-regulated phosphorylations are unclear. Here, we show evidence that switch-like phosphorylation of intrinsically disordered proteins (IDPs) by CDKs contributes to mitotic cellular reorganisation by controlling protein-protein interactions and phase separation. We studied protein phosphorylation in single *Xenopus* embryos throughout synchronous cell cycles, performed parallel assignment of cell cycle phases using egg extracts, and analysed dynamics of mitotic phosphorylation using quantitative targeted phosphoproteomics. This provided a high-resolution map of dynamic phosphosites from the egg to the 16-cell embryo and showed that mitotic phosphorylation occurs on entire protein complexes involved in diverse subcellular processes and is switch-like *in vivo*. Most cell cycle-regulated phosphosites occurred in CDK consensus motifs and located to intrinsically disordered regions. We found that substrates of CDKs and other cell cycle kinases are significantly more disordered than phosphoproteins in general, a principle conserved from yeast to humans, while around half are components of membraneless organelles, whose assembly is thought to involve phase separation. Analytical modelling predicts modulation of homotypic IDP interactions by CDK-mediated phosphorylation, which was confirmed by biophysical and biochemical analysis of a model IDP, Ki-67. These results highlight the dynamic control of intrinsic disorder as a conserved hallmark of the cell cycle and suggest a mechanism for CDK-mediated mitotic cellular reorganisation.

Introduction

Eukaryotic cell cycle progression depends on the CDK1-subfamily of CDKs and is presumed to arise from the collective behaviour of altered protein phosphorylation states. With the notable exception of CDK1, most CDK and cyclin genes are dispensable for cell proliferation in the majority of cell types in the mouse (*1-3*), while in fission yeast, oscillating activity of CDK1 alone can drive the entire cell cycle (*3, 4*). This suggests that a rather limited core network of CDKs can drive the eukaryotic cell cycle, and that changes in overall CDK activity somehow determine the sequence of the complex processes required to duplicate the genome and distribute cellular components during cell division. This "quantitative model"(*4*), implies that there exist low and high overall CDK activity thresholds for entry into S-phase and mitosis, respectively, determined by the CDK-regulatory network. This network involves positive and double-negative feedback loops, as well as futile cycles of CDK and CDK-opposing phosphatase activity (*5*). Mathematical modelling shows that such features of network organisation can generate ultrasensitivity and hysteresis in CDK1 activation (*6*), while the resulting bistability of CDK1 activity leads to a switch-like G2/M transition (*7*). These theoretical concepts are supported by experimental evidence in *Xenopus* egg extracts and mammalian cells (*6, 8-12*).

The presumed switch-like dynamics of the CDK1 regulatory network is consistent with the abrupt morphological reorganisation of the cell at mitosis. In metazoans, the nuclear envelope and lamina breaks down and many cellular structures are rapidly disassembled. These include nuclear pore complexes, nucleoli, pericentriolar material, splicing speckles, Cajal bodies, promyelocytic leukaemia (PML)-nuclear bodies and stress granules (*13-16*), which have been collectively referred to as membraneless organelles (MLO). Thus, MLO assembly and disassembly occurs in a cell cycle-dependent manner. MLOs are thought to assemble by mechanisms involving multivalent interactions between intrinsically-disordered regions (IDR) of proteins(*13*), and this process can be regulated by protein kinases, including CDKs (*17-20*). Protein phosphorylation in general is enriched in IDRs (*21*) and this also appears to be true for CDKs (*22-24*). As such, an attractive model is that CDK-mediated IDR phosphorylation might trigger rapid dissolution of many MLOs at mitosis. This would be consistent with the fact that CDK1-family CDKs can phosphorylate hundreds of sites on diverse proteins (*25-28*), and regulate DNA replication, mitosis, transcription, chromatin remodeling, DNA repair, the cytoskeleton, nuclear transport, protein translation, formation of a mitotic spindle and even ciliogenesis (*29-31*).

Direct evidence for switch-like dynamics of cell cycle-regulated phosphorylation *in vivo* is currently lacking. Single-cell proteomics studies (*32, 33*) have insufficient sensitivity and reproducibility for low stoichiometry and highly dynamic targets such as phosphosites. Therefore, studies analysing cell cycle phosphorylation have generally used cells blocked at different stages of the cell cycle to generate "snapshots" of the phosphorylation landscape (*34*). However, highly dynamic phosphorylation states cannot readily be determined from populations of cells (*35*). Moreover, whole-culture synchronisation methods generate artefacts due to cell cycle perturbation (*36, 37*). This might explain why, in an *in vivo* phosphoproteomics study in fission yeast synchronised by chemical block and release of CDK1, overall cell cycle phosphorylation dynamics appeared progressive rather than switch-like (*38*). Alternative phosphoproteomics approaches on unsynchronised cells selected with centrifugal elutriation (*39*) or FACS (*40*), lack the temporal resolution to determine the dynamics of protein phosphorylation throughout the cell cycle.

Here, we overcame these obstacles by using an extremely sensitive phosphopeptide enrichment strategy (*41*) to perform quantitative phosphoproteomics on the highly synchronous early cell cycles of *Xenopus laevis* embryos, which consist solely of S and M-phase (*42, 43*). By performing parallel phosphoproteomics using synchronously replicating or mitotic egg extracts we could attribute cell cycle behaviour of individual sites. This allowed us to investigate the general features of cell cycle-regulated phosphorylation compared to the entire phosphoproteome, revealing the importance of intrinsic disorder. We next compiled high-confidence CDK substrates in human and yeast, and analysed disorder on a proteome-wide scale in all three species. Our data provide evidence for switch-like mitotic phosphorylation of multiple subunits of protein complexes involved in diverse biological processes, and suggest that CDKs control these by phosphorylating IDRs. The latter are key drivers of protein phase separation (PS), which is thought to drive the formation of MLOs. Biophysical and biochemical analysis of a model IDP, namely Ki-67, showed that CDK-mediated phosphorylation regulates homotypic interactions and PS. Altogether, our results suggest that CDKs can drive rapid reorganisation of the cellular architecture by phosphorylation of IDRs and modulation of protein PS.

Results & Discussion

A high-resolution map of *in vivo* cell cycle phosphorylation

We took advantage of the naturally synchronous early cell cycles of *Xenopus laevis* embryos to perform quantitative phosphoproteomics *in vivo*, using a sensitive automated phosphopeptide enrichment protocol (*41*). We collected single embryos at 15-minute intervals while recording visual cues of cell divisions. Phosphopeptides from each embryo were purified, separated by nano-LC and analysed by mass spectrometry (Figure 1A). We identified 4583 high-confidence phosphosites mapping to 1843 proteins (Figure S1A), most being phosphoserines (Figure S1B). Individual embryo phosphorylation states strongly correlated (Figure S1C). We thus generated a dynamic map of protein phosphorylation from an unfertilised egg to a 16-cell embryo.

We focused on 1032 sites whose variation in phosphorylation over time was statistically significant (hereafter denoted "dynamic phosphosites") which occurred on 646 proteins. Gene ontology (GO) and network analysis revealed high functional association and interconnectivity between groups of proteins involved in RNA binding and the nuclear pore complex (NPC), DNA replication and chromatin remodeling, and microtubule regulation (Figure 1B). Hierarchical clustering uncovered

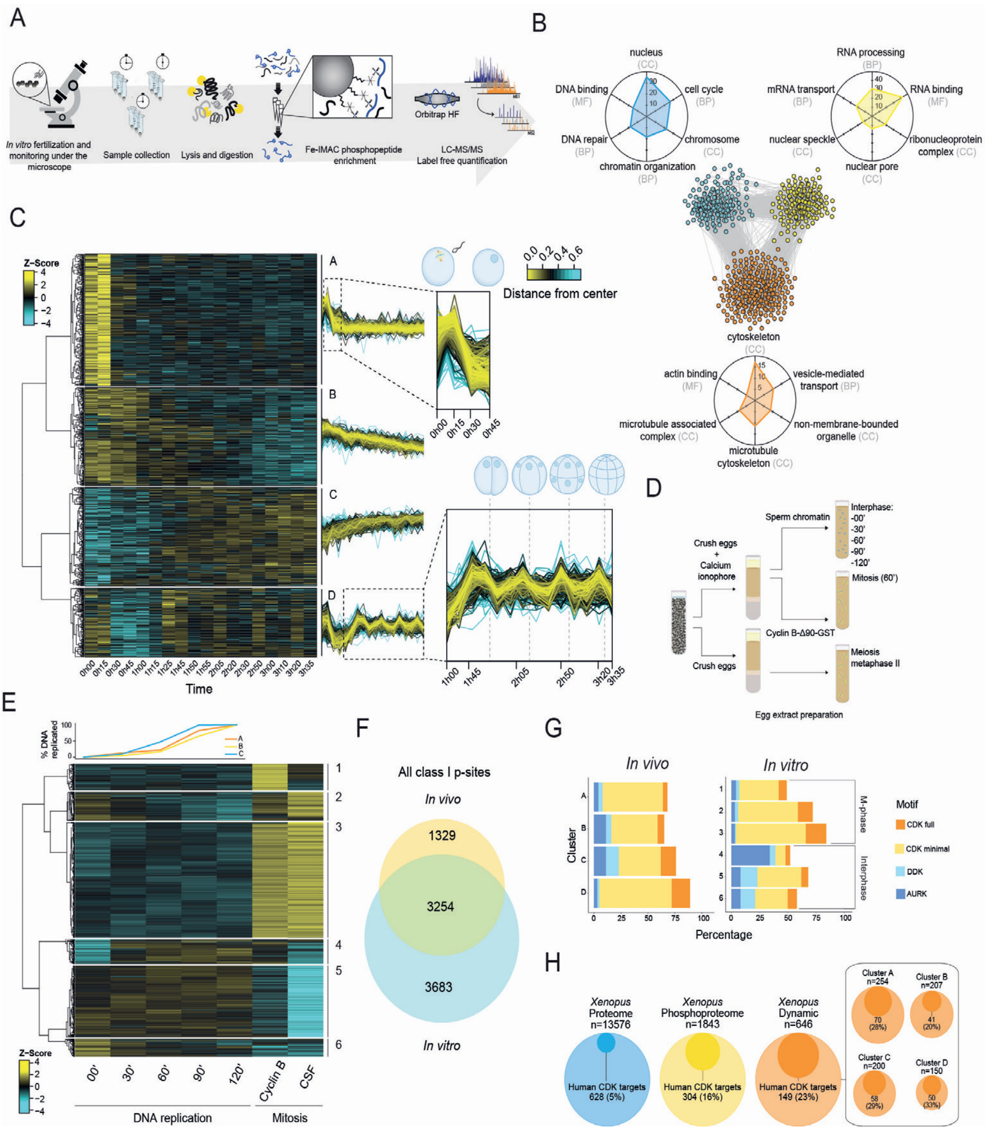


Figure 1. The time-resolved phosphoproteome from a single-cell to a 16-cell embryo and its cell cycle assignment. (A) Schematic representation of the workflow. Single *Xenopus* eggs and embryos were collected followed by cell lysis, protein digestion, phosphopeptide enrichment and high-resolution proteomics analysis. (B) STRING network of functionally associated proteins undergoing dynamic phosphorylation (each node represents a protein). Vicinity clustering reveals three main groups (yellow, blue and orange) with a high degree of association. Radar plots show the corresponding GO terms (adjusted p value <0.05) for each group (axes show -Log$_{10}$(adj p value) for each GO term). **Legend continues on next page.**

Figure 1 legend continued. (C) Hierarchical clustering of significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05), reveals 4 clusters with distinct regulation (A-D). Dashed boxes in clusters A and D are zoomed-in to highlight dynamic phosphorylation patterns (dashed lines depict the time points of cell division). (D) Scheme of the experiment in the *Xenopus* egg extract. (E) Top: quantification of DNA replication in each biological replicate. Below: Hierarchical clustering of dynamic phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05) reveals differential regulation of phosphosites during S-phase and mitosis. (F) Overlap between *in vivo* (embryo) and *in vitro* (egg extract) phosphoproteomics. (G) Proportion of phosphosites according to their potential upstream kinase for each cluster in the *in vivo* (top) and *in vitro* (bottom) experiments. (H) Circle plots presenting enrichment of homologues of human CDK substrates among *Xenopus* phosphoproteins detected *in vivo* and those with dynamic phosphosites.

four distinct groups that reflect cell cycle-regulated behaviour (Figure 1C). The levels of clusters A and B phosphosites were highest in eggs and post-fertilisation, and decreased during the first round of DNA replication, suggesting that dephosphorylation of these sites may prepare the zygote for upcoming cell divisions (*44*). GO analysis for group A highlighted proteins involved in RNA regulation and nuclear organisation, including the NPC and nuclear transport, chromosomal structure and segregation (Figure S1D), as also observed in a recent study on meiosis exit (*45*). Cluster B phosphosites were enriched in regulators of RNA biosynthesis and stability, translation, actin, DNA replication and repair (Figure S1D). Cluster C phosphosites progressively increased after meiotic exit, while cluster D phosphosites had a clear oscillating signature with upregulation preceding each cell division. GO analysis of cluster C shows dominance of interphase cell cycle processes including DNA replication, RNA-related processes and chromosome organisation (Figure S1D), and included phosphosites displaying a reciprocal oscillating trend and a lower amplitude compared to cluster D sites. Several such sites, *e.g.* S31 of the replication licensing protein MCM4, were from monophosphorylated peptides, while the multiphosphorylated forms were found in cluster D (Figure S1E). Thus, cluster C contains the earliest phosphorylations of proteins that are highly phosphorylated at mitosis. Cluster D shows coordinated phosphorylation of multiple members of protein complexes involved in diverse processes, suggesting a common mechanism of regulation (Figure S1F). Importantly, phosphoproteome changes were not simply a reflection of changes in abundance of the corresponding proteins (Figure S2), which are generally negligible during *Xenopus* early development (*46*).

We assigned *in vivo* embryo phosphosites to different cell cycle stages by comparing with phosphorylation patterns of replicating or mitotic egg extracts (Figure 1D). Replication was initiated by adding purified sperm chromatin to interphase egg extracts and quantified over time (Figure 1E, top), while mitosis was triggered by adding recombinant cyclin B and verified microscopically. We also used egg extracts arrested at meiotic metaphase II (Cytostatic Factor, CSF-arrested). Overall, we identified 6937 phosphosites, which included 71% of the sites identified *in vivo* (Figure 1F). 1728 sites varied between S and M-phase, including 693 sites upregulated in S-phase and 1035 in mitosis (Figure 1E). GO analysis of interphase and mitotic sites revealed processes enriched in *in vivo* cluster C and cluster D, respectively (Figure S3A). Several DNA-replication factors, including MCM4 and RIF1, showed multi-site phosphorylation specifically in S-phase (Figure S3B). This phosphoproteomics dataset greatly increases the known repertoire of phosphorylation sites upregulated during S-phase (*34*).

We next analysed the cell cycle behaviour of dynamic phosphosites that we found *in vivo* (Figure S3C). Most embryo cluster A sites were upregulated in both CSF-arrested meiotic extracts and mitotic extracts*,* highlighting the global similarities of regulation of meiotic and mitotic M-phase, despite the additional activity of the Mos/MEK/MAP kinase pathway in meiosis. Around half of embryo cluster B sites were present only in interphase, while the rest showed a minimum phosphorylation in late S-phase, confirming their dephosphorylation during the first round of DNA replication. As expected, most sites from embryo clusters C and D were part of the *in vitro* S-phase and mitotic groups, respectively. Therefore, single embryo data can successfully identify cell cycle-dependent phosphorylation. In mitosis, as expected, monophosphorylated species are reduced because multisite phosphorylation emerges (Figure S3D; Figure S1E).

Predominance of CDK targets

Analysis of kinase consensus motifs showed that proline-directed (S/T-P) sites, which conform to the minimal consensus for CDKs, comprise 51% of all detected phosphosites *in vivo* and 60% of dynamic sites (Figure S4A). Around 10% of all phosphosites matched the full CDK1-family consensus site: S/TPxK/R. Replicating and mitotic extracts displayed a similar trend (Figure S4A). Putative CDK targets dominated all clusters, with 80% of sites in cluster D *in vivo* and mitotic clusters *in vitro* conforming to the minimal CDK motif (Figure 1G, Figure S4B, C). Consensus sites of other kinases such as Aurora, Polo-like kinase (PLK), DBF4-dependent kinase (DDK) and Casein kinase I and II were present to a lesser extent (Figure S4B, D). In meiotic M-phase, MAP kinases, which have the same consensus motif as CDKs, are likely responsible for sites specific to embryo cluster A or CSF extracts, but these kinases are inactivated during early embryonic cell cycles (*47*), suggesting that most of the other dynamic proline-directed phosphorylations are due to CDKs.

Although few direct CDK substrates have been characterised in *Xenopus*, they are likely conserved between vertebrates. We therefore manually curated a set of 654 human CDK1-subfamily targets (see Methods for sources). 303 of these have *Xenopus* homologues among the 1843 phosphoproteins we detected*,* and 149 were present among the 646 proteins with dynamic phosphosites in *Xenopus* embryos (Figure 1H). Thus, the predominance of CDK motifs among dynamic phosphosites reflects a high proportion of *bona fide* CDK substrates. This is a conservative estimate, since we only considered proline-directed sites as CDK motifs, although we found that 10-20% of human and yeast CDK substrates (see Methods for sources) were non-proline-directed (Figure S4E), confirming a recent finding (*48*). These data reinforce the dominant role of CDKs in cell cycle-regulated phosphorylation.

Mitotic phosphorylation is switch-like in vivo

We next determined whether mitotic phosphorylation of individual phosphosites is progressive or switch-like *in vivo*. We analysed dynamics of 64 cluster D sites from diverse protein complexes in single embryos every 180-seconds using quantitative targeted phosphoproteomics (*49-51*) by parallel reaction monitoring (*52*), thereby obtaining a quantitative description of mitotic phosphorylation *in vivo* at extremely high-time resolution (Figure 2A). This revealed parallel and abrupt upregulation of all phosphosites preceding each cell division (Figure 2B, C), indicating switch-like phosphorylation of diverse protein complexes at mitotic onset. This was not due to oscillation of CDK1-Y15 inhibitory phosphorylation, which was downregulated over time (Figure 2D), as previously reported (*53*), consistent with lack of corresponding phosphorylation of the CDK1-Y15-regulatory enzymes, CDC25 and WEE1. In contrast, oscillating phosphorylations on NIPA and the APC/C, which regulate mitotic cyclin accumulation (*54, 55*), as well as Greatwall kinase, which activates the PP2A inhibitors Arpp19/ENSA (*56-58*), were apparent (Figure S5A). These data suggest that control of mitotic cyclin levels and PP2A activity, and therefore the overall CDK/phosphatase activity ratio (*59*), suffices for switch-like mitotic phosphorylation whereas regulated CDK1-Y15 phosphorylation is not essential (Figure S5B). This is consistent with the self-sufficiency of futile cycles of opposing enzymes in generating switch-like network output in the absence of allosteric regulation (*60*).

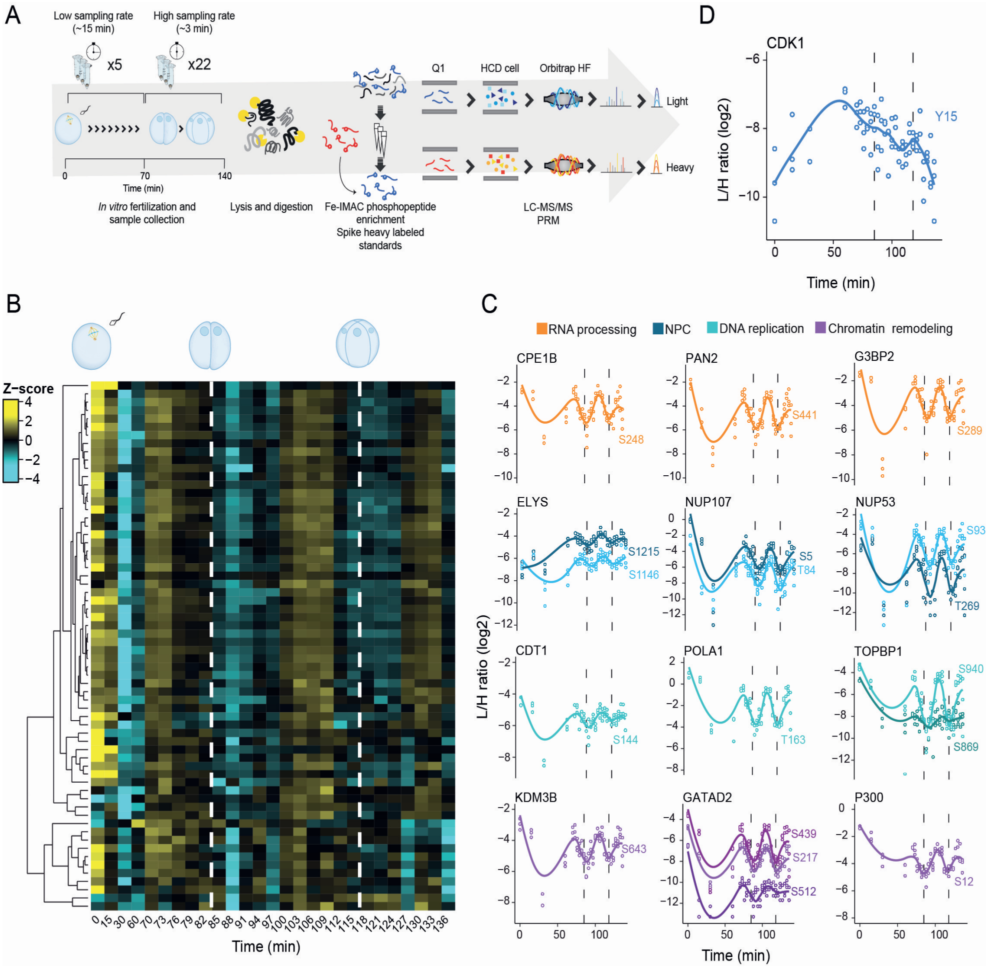


Figure 2. Switch-like mitotic phosphorylation *in vivo*. (A) Schematic representation of the workflow. Samples were collected over two cell divisions and enriched phosphopeptides were subjected to targeted proteomics analysis. (B) Heat map shows a highly synchronous wave of phosphorylation preceding each of the two cell divisions. Dashed lines depict times when cell divisions were recorded. (C) Single phosphosite plots from selected proteins. Each dot represents a biological replicate (n=3). Dashed lines depict times when cell divisions were recorded. (D) Single phosphosite plot of CDK1 inhibitory phosphorylation (Y15).

The cell cycle phosphoproteome is intrinsically disordered

We wondered whether the diverse dynamic phosphoproteins share common structural features facilitating switch-like CDK-mediated phosphorylation. Phosphosites in general are often located in intrinsically disordered regions (IDRs) of proteins (*21*), which is also true for yeast and mouse CDK sites (*22-24*). Yet previous analyses did not exclude the possibility that this is an artefact due to the enrichment of serine, threonine and proline in disordered regions, which is consistently predicted across the entire proteome of *Xenopus*, human and yeast (Figure S6A). We corrected for this compositional bias, and found that phosphorylatable residues in IDR are indeed more highly phosphorylated than those in ordered regions (Figure 3A-C). This enrichment was increased for proteins with at least one site displaying dynamic phosphorylation; the same was true for human CDK substrates (Figure 3B, C). To estimate the differential phosphorylation of disordered sites globally, we calculated the ratio of dynamically phosphorylated (*Xenopus*) or CDK-phosphorylated (yeast, human) to non-phosphorylated serine and threonine in both disordered and structured regions (Figure S6B; see Methods). This confirmed that cell cycle-regulated phosphorylation is largely skewed towards disordered regions and that CDKs preferentially phosphorylate disordered sites (Figure 3D, Figure S6C). We then asked whether this is also true for substrates of other protein kinases. We analysed the mitotic PLK and Aurora kinases, DYRK kinases, which promote mitotic phosphorylation of several IDPs (*18*), NEK kinases, which have roles in centrosome duplication and various stages of mitosis, and MAP kinases, which share the proline-directed S/T consensus site. For each kinase, documented phosphosites were strongly enriched in IDRs (Figure S6C, D), supporting the idea that phosphorylation of residues in IDRs is kinetically favoured (*21*).

To explain the dominance of CDK-mediated phosphorylation in the cell cycle, we surmised that their substrates might be more disordered than phosphoproteins in general. We therefore determined the percentage of disordered residues of proteins in our datasets, compared to the rest of their respective phosphoproteomes. This revealed that, on average, both *Xenopus* dynamic phosphoproteins and human and yeast CDK substrates contain approximately twice the proportion of disordered amino acids as other phosphoproteins (Figure 3E, Figure S6E), putting them among the top quartile of proteins with the most disorder in the proteome. If this reflects the importance of disordered proteins for the cell cycle generally, then substrates of other cell cycle kinases might also be more disordered than other phosphoproteins. Indeed, targets of most cell cycle kinases are significantly more disordered than targets of MAP kinase (Figure 3F), whose phosphosites are also proline-directed and preferentially located in IDRs (Figure S6D).

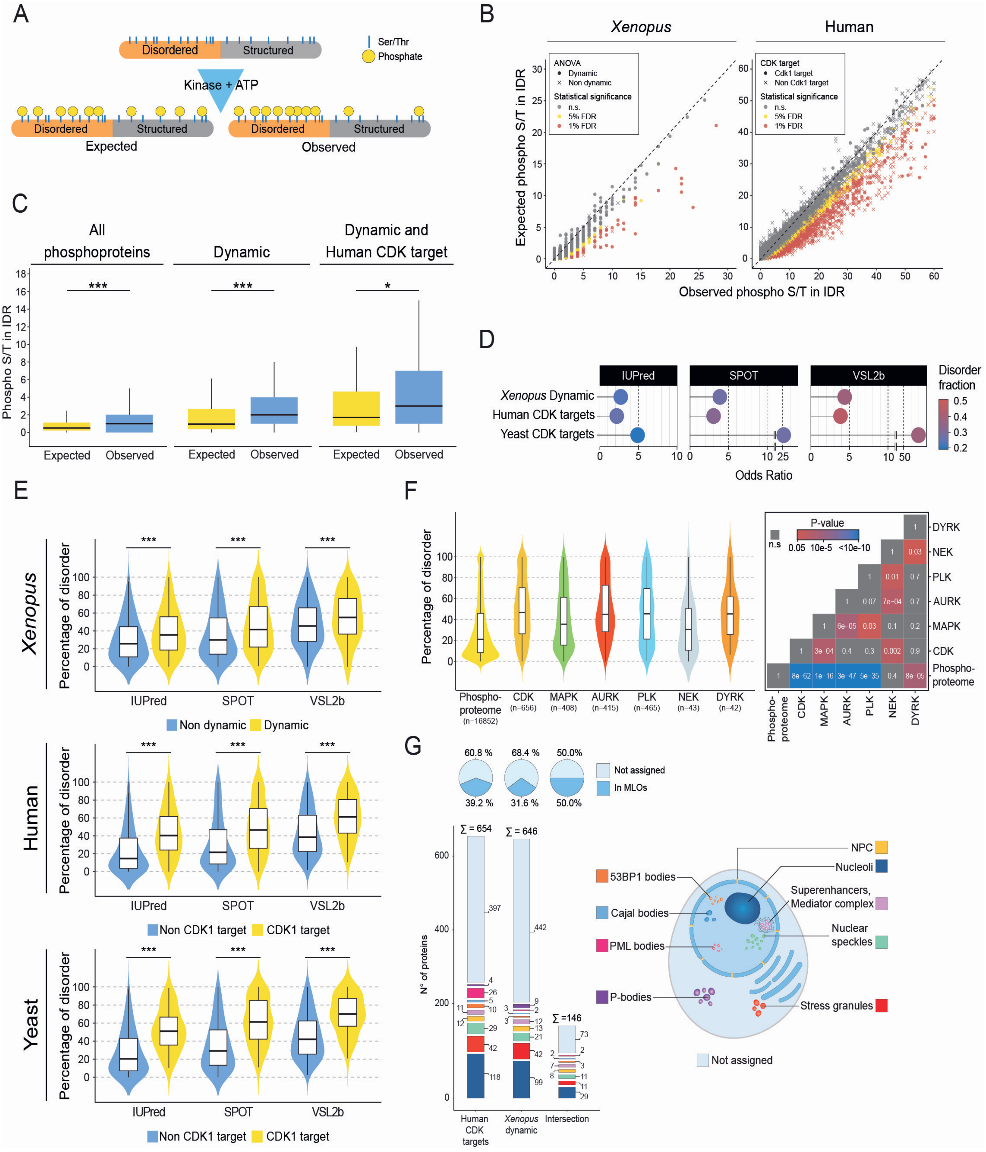


Figure 3. The cell cycle phosphoproteome is characterised by intrinsic disorder and MLO components. (A) Scheme illustrating hypothetical enrichment of phosphorylation in disordered regions when taking into account amino acid compositional bias. (B) Scatter plot of expected vs observed phosphorylated Ser/Thr for each protein of human and *Xenopus* phosphoprotein datasets. FDR thresholds of 5% and 1% are marked in yellow and red respectively. Circles: proteins with at least one dynamic phosphorylation in *Xenopus*, or human CDK1 subfamily substrates, respectively. **Legend continues on next page.**

Figure 3 legend continued. (C) Boxplots showing expected vs observed phosphorylated Ser/Thr among all phosphoproteins detected (left), phosphoproteins with at least one dynamic phosphosite (middle), and dynamic phosphoproteins also detected as CDK1 subfamily targets in humans (right). Distributions were compared with the Wilcoxon signed-rank test. *$p<0.05$, **$p<0.01$, ***$p<0.001$. (D) Plots showing the common Odds Ratio of Ser/Thr phosphorylation in structured and ordered regions calculated with the Fisher's test (see Figure S6b, c). For all organisms, the disordered regions were calculated with three different disorder predictors. The disordered fraction is presented in a colour scale. (E) Violin plots of the distribution of disordered residues per protein for CDK targets vs the rest of the phosphoproteome for human and yeast, and dynamic phosphoproteins vs the rest of the phosphoproteome for *Xenopus*. Intrinsic disorder was calculated with three different predictors (IUPred, SPOT, and VSL2b). Statistical significance was evaluated with the Wilcoxon–Mann–Whitney test; ***$p<0.001$. (F) Violin plot (left) showing the distribution of disordered residues per protein for CDK, MAPK, Aurora, PLK, NEK and DYRK kinase targets vs the rest of the phosphoproteome for human targets. Statistical significance was assessed by Kruskal-Wallis ANOVA, and pairwise comparisons were performed with Dunn's post-hoc tests. The adjusted p-values (Benjamini-Hochberg) are shown in a tile plot (right). (G) Human CDK1 subfamily targets, *Xenopus* dynamic phosphoproteins, and the intersection of both sets, that are present in our manually curated proteome of membraneless organelles.

Enrichment of MLO components among CDK substrates

We thus reasoned that phosphorylation may have been selected to regulate the functions of IDPs during the cell cycle. IDPs are key components of membrane-less organelles (MLO), many of which (*e.g.* Cajal bodies, nucleoli, nuclear pore complexes, splicing speckles) are thought to arise by phase separation (PS) (*61*), are disassembled in mitosis, and can be regulated by phosphorylation (*17-19*). To corroborate our hypothesis, we analysed available data on cellular localisation for each of our curated human CDK substrates. We found that 257 (39.2%) are present in MLOs, including key IDPs such as coilin (Cajal bodies), nucleophosmin, nucleolin and Ki-67 (nucleoli), 53BP1 (53BP1 bodies), nucleoporins (NPC) and PML (PML bodies) (Figure 3G). We then manually curated an MLO proteome from human proteomics studies (See Methods for sources). Homologues of 204 dynamic *Xenopus* phosphoproteins (31.6%) localise to MLOs, as do 73 of the 149 proteins (50%) that show dynamic phosphorylation in *Xenopus* and are CDK substrates in human (Figure 3G). The vast majority of proline-directed phosphosites and confirmed CDK sites in these proteins were located in predicted IDRs (Figure S7).

CDKs regulate IDR phase separation

Both stochastic and specific interactions between IDPs contribute to PS and MLO assembly (*61-63*). We hypothesised that cell cycle kinase-mediated phosphorylation might modulate such interactions. We first applied a machine learning classifier (*64*) to predict whether cell cycle-regulated phosphoproteins have an increase in average propensity for PS (PSAP score). Indeed, we observed a sharp increase in the PSAP score, from the proteome to the phosphoproteome, and a further increase for dynamic phosphoproteins, with the highest score for mitotic cluster D (Figure S8A). Similarly, the propensity for PS is far higher amongst targets of most cell cycle kinases (CDK, Aurora, PLK, but not NEK) and DYRK kinases than the overall phosphoproteome, but less so for MAP kinase substrates.

Next, to better understand the biochemical effects of their cell cycle-regulated phosphorylation, we analysed a selection of IDRs from CDK substrates. We applied a general heteropolymer theory that uses sequence charge decoration matrices (SCDM), based on electrostatic interactions only, to identify intra-chain interaction topology (*65, 66*). Since this should correlate with inter-chain interactions that promote PS, SCDMs provide indirect insights to propensity to phase separate. Of the 12 IDPs tested, 7 (nucleolin, nucleophosmin, NUP53, ELYS, MCM4, 53BP1 and the splicing factor SF3B1) had SCDM maps showing visibly decreased self-association propensity (increased red regions in Figure S8B), implying reduced propensity to phase separate, upon CDK-site phosphorylation. Conversely, for SRRM2, CDK-mediated phosphorylation is predicted to increase intra-chain attraction (Figure S8B) and hence PS tendency. For 4 proteins (MDC1, TICRR, COILIN, and CDT1), SCDM maps were inconclusive. To further analyse these trends, we calculated radius of gyration of several IDRs using all-atom simulation. Effects of phosphorylation on CDT1 (28.4Å to 30.3Å), TICRR (56.2Å to 57.3Å) and coilin (39 Å to 37.9 Å) were minor, while MCM4 IDR expands upon phosphorylation (21.9Å to 26.3Å), consistent with SCDM analysis. Overall, these data suggest that phosphorylation is a key regulator of homotypic interactions, an important element of PS propensity, of most IDRs.

To test this hypothesis, we focused on a model CDK substrate, Ki-67, an IDP that organises heterochromatin structure (*67*) and perichromosomal layer formation from nucleolar components in mitosis (*68, 69*). Ki-67 contains a multivalent Ki-67 repeat domain that is highly phosphorylated in mitosis by CDKs (Figure 4A), which regulates its perichromosomal localization (*70*). SCDM analysis predicted that phosphorylation of full-length Ki-67 should promote self-interaction and thus PS, but this cannot be attributed to interactions of its repeat motif alone, since phosphorylation of the latter is predicted to reduce homotypic interactions (Figure 4B). In agreement, coarse-grained (CG) molecular dynamics (MD) simulations showed that the radius of gyration of full-length Ki-67 decreased upon phosphorylation (Figure 4C, left) while that of a single consensus repeat motif increased (Figure 4C, right). MD simulations also showed that PS is enhanced by increasing repeat valency and counteracted by phosphorylation (Figure 4D), consistent with SCDM analysis. To test these predictions experimentally, we first used the optogenetic Cry2 "optodroplet" system (*71*) with full length Ki-67 or a series of deletion mutants. Full-length Ki-67 localised to the nucleolus, as expected, but exposure to blue light caused

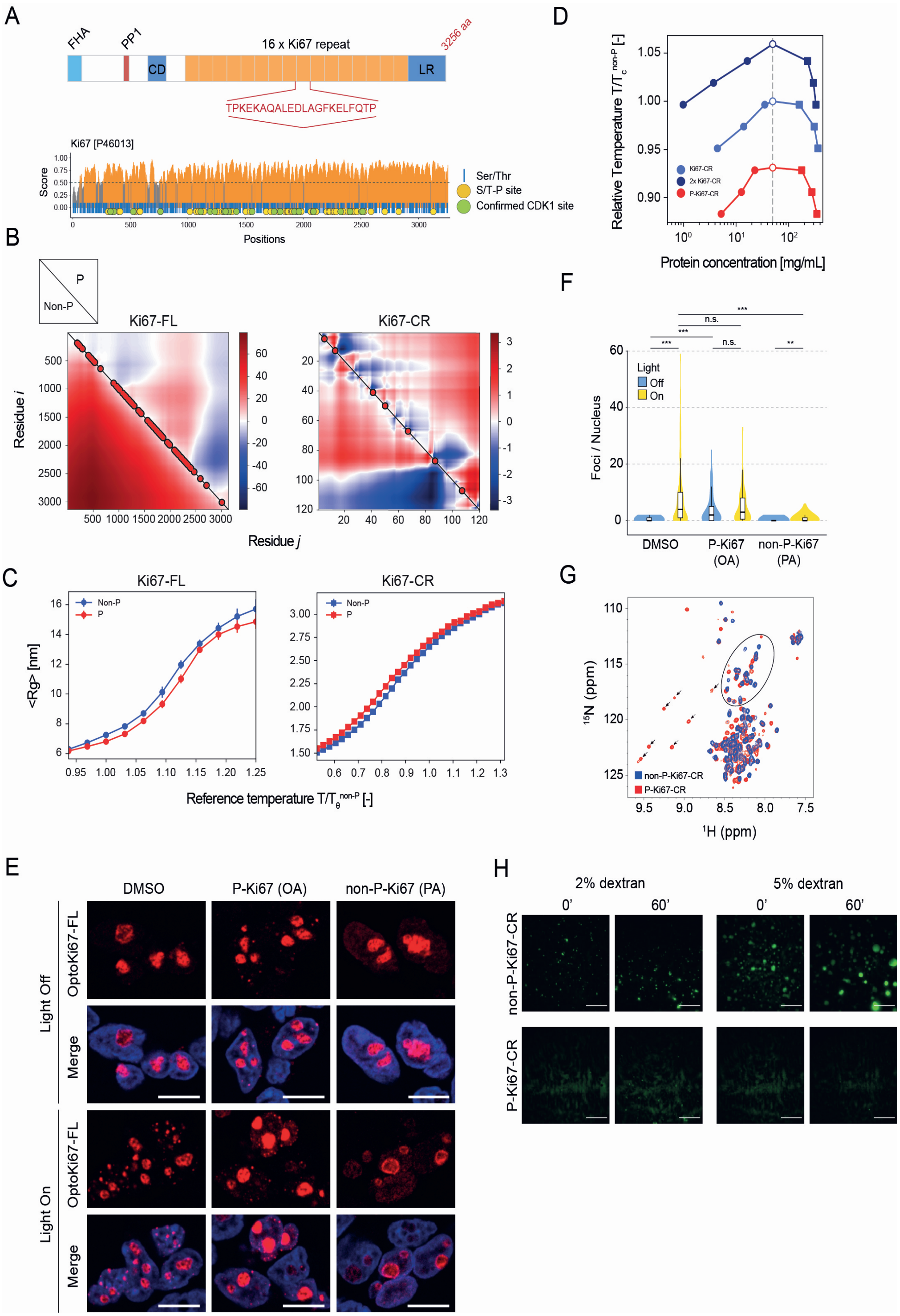

A
FHA
PP1
16 x Ki67 repeat
3256 aa
CD
LR
TPKEKAQALEDLAGFKELFQTP
Ki67 [P46013]
Score
1.00
0.75
0.50
0.25
0.00
0
500
1000
1500
2000
2500
3000
Positions
Ser/Thr
S/T-P site
Confirmed CDK1 site
B
P
Non-P
Ki67-FL
Ki67-CR
Residue i
Residue j
C
Ki67-FL
Ki67-CR
Non-P
P
<Rg> [nm]
Reference temperature T/T_θ^non-P [-]
D
Relative Temperature T/T_c^non-P [-]
Ki67-CR
2x Ki67-CR
P-Ki67-CR
Protein concentration [mg/mL]
E
DMSO
P-Ki67 (OA)
non-P-Ki67 (PA)
Light Off
Light On
OptoKi67-FL
Merge
F
Light
Off
On
Foci / Nucleus
***
n.s.
***
***
n.s.
**
DMSO
P-Ki67 (OA)
non-P-Ki67 (PA)
G
¹⁵N (ppm)
¹H (ppm)
non-P-Ki67-CR
P-Ki67-CR
H
2% dextran
5% dextran
0'
60'
non-P-Ki67-CR
P-Ki67-CR

Figure 4. CDK-mediated phosphorylation regulates phase separation of a model IDP. (A) Top, scheme of the human Ki-67 protein (FHA, forkhead-associated domain; PP1, PP1 phosphatase-binding domain; CD, conserved domain; LR, leucine arginine-rich domain). Highlighted, Ki-67 repeat consensus motif. Bottom, diagram of IUPred score over the length of human Ki-67. Regions with scores >0.5 (orange) are considered to be disordered, and <0.5 (grey) structured. Blue vertical lines indicate Ser and Thr residues; yellow circles, known Ser/Thr-Pro phosphosites; green circles, confirmed CDK1 subfamily phosphorylations. (B) Sequence Charge Decoration Matrix (SCDM) maps for full length Ki-67 (FL, left) and Ki-67 consensus repeat (CR, right), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using colour schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles. (C) Dependency of the radius of gyration (Rg) on the simulation temperature in single-chain MD simulations for full chain Ki-67 (left) and consensus repeat (right). The reference temperature is the θ temperature of the non-phosphorylated molecule for full chain and consensus repeat, respectively. Reported error bars are obtained by block analysis over 10 blocks. (D) Binodal curves from phase coexistence simulations of the Ki-67 consensus repeat sequence. For each temperature, filled circles indicate the dilute phase density and squares indicate the coexisting dense phase density. Empty circles indicate the fitted critical temperature (T_c) of each system. The T_c of the non-phosphorylated monomer (light blue empty circle) was the reference for the normalisation of the temperature values. The light gray dashed line indicates the total concentration used in the simulations. The reference temperature is the θ temperature of the non-phosphorylated molecule for full chain and consensus repeat, respectively. Reported error bars are obtained by block analysis over 10 blocks. (E) Representative fluorescent images of HEK-293 cells expressing opto-Ki-67 (FL) construct before (Light Off) and after (Light On) exposure to blue light. Cells were pretreated for 1h with either vehicle (DMSO), 0.5 µM okadaic acid (OA), to inhibit protein phosphatase 2, or 5 µM purvalanol A (PA), to inhibit CDKs. DNA was stained with Hoechst 33258; scale bars, 10µm. (F) Violin plot presenting quantification of results from (E); the number of foci per nucleus was counted. Statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise *post-hoc* comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. (G) Overlaid NMR ^{1}H-^{15}N HSQC of unphosphorylated (blue) and CDK-phosphorylated (red) GFP-tagged Ki-67 consensus repeat. Each cross-peak corresponds to one residue. The seven new deshielded cross peaks (highlighted by a black flag) appearing above 8.5 ppm in ^{1}H correspond to phosphorylated serines or threonines (^{1}H downfield chemical shift perturbation on phosphorylated Ser/Thr residues due to phosphate electronegativity). Non phosphorylated Ser/Thr residues are surrounded by a black oval. (H) Representative fluorescence images of *in vitro* phase separation assay with purified GFP-tagged Ki-67 consensus repeat (CR), non-phosphorylated (non-P) or *in vitro* phosphorylated with recombinant CDK1-cyclin B-CKS1 (P), at indicated dextran concentrations and time points; scale bars, 10µm.

rapid appearance of small round foci in the nucleoplasm, which was dependent on the level of induced Ki-67 expression, consistent with PS (Figure S9A). Importantly, promoting CDK-mediated phosphorylation by inhibiting PP2A with okadaic acid (*59*) led to foci formation in the absence of blue light, while pan-CDK inhibition with purvalanol A prevented induction of foci upon light (Figure 4E, F). These results indicate that, as predicted by SCDM and MD, phosphorylation of full-length Ki-67 promotes PS. Results were similar for constructs lacking the C-terminal LR domain, that binds chromatin, or the N-terminal domain, which is required for the nucleolar localisation of Ki-67 (Figure S9B). Finally, we purified a consensus repeat polypeptide (Figure S10A) and phosphorylated it *in vitro* with recombinant CDK complexes. Nuclear Magnetic Resonance spectroscopy showed a reduced amide proton spectral dispersion typical for an IDP, and confirmed appearance of 7 phosphorylated residues upon incubation with purified CDKs and ATP (Figure 4G). We mapped phosphorylation sites and intensity by phosphoproteomics and Phos-Tag-SDS-PAGE, indicating stoichiometric phosphorylation (Figure S10B, C). Purified GFP-tagged Ki-67 repeat motif could phase-separate *in vitro*, and, as predicted, this was abolished upon full phosphorylation by CDK (Figure 4H). Taken together, these results confirm that CDK-mediated phosphorylation is able to both promote or inhibit homotypic interactions that contribute to PS, and suggest that Ki-67 may have several competing modes of PS that are differentially regulated by phosphorylation. Our data suggest a mechanism for Ki-67-mediated mitotic targeting of nucleolar components to the perichromosomal layer (*67, 68*) via CDK-mediated phosphorylation, which reduces PS of several major nucleolar IDPs, thus triggering nucleolar disassembly, while simultaneously promoting PS of Ki-67 bound to chromatin to recruit nucleolar components.

In conclusion, this work reveals *in vivo* that CDK-dependent mitotic phosphorylation occurs in a switch-like manner on diverse proteins whose common denominators are a high level of disorder and localisation to MLOs. Furthermore, our data show that CDK-mediated phosphorylation regulates homotypic interactions between IDPs, which may coordinate diverse cellular processes during the cell cycle. While this is not incompatible with models in which high-affinity interactions contribute to MLO formation by PS (*72, 73*), it suggests that cell cycle control may be less specific than previously thought.

Methods

Egg collection and *in vitro* fertilisation

Female *X. laevis* frogs were primed with 50 international units (IU) of human chorionic gonadotropin at least 2 days, and no more than 7-8 days before a secondary injection with 625 IU to induce ovulation. Roughly 16 hours after the second injection, fresh eggs were collected by pelvic massage and kept in 1x Marc's Modified Ringer's (MMR). Next, eggs were placed in a petri dish and checked under the microscope to keep only those that exhibited the healthy pigment pattern (dark animal pole and white vegetal pole).

To perform the *in vitro* fertilisation, around 1/3 of a full testis was cut into fine pieces and mixed with in 500 µl of 1x MMR. The suspension was pipetted up and down until big clumps were dissolved. Next, buffer was removed from the petri dish and the eggs were collected. Once eggs were well dispersed across the dish, the sperm suspension was added. The dish was then flooded with 0.1x MMR to induce fertilisation.

Sample collection

The first time point was collected immediately before adding the sperm suspension (time 0′ corresponds to the unfertilised egg). Eggs were kept at room temperature (18-20°) and under a dissecting microscope after fertilisation. At approximately 15 minutes, fertilised eggs underwent shrinkage of the animal hemisphere and rotation within the vitelline membrane, so that the animal hemisphere faced upwards. These changes are known indicators of successful fertilisation, so only the eggs that underwent these changes were used for the experiment.

Samples were collected approximately every 15 minutes. Eggs were rapidly placed in individual tubes and snap froze in liquid nitrogen, trying to preserve the phosphorylation events occurring at that specific time. Since the eggs were monitored under the microscope, we were able to determine if samples were collected before or after a cell division had occurred.

Xenopus egg extracts

Interphase *Xenopus* egg extracts were prepared, and DNA replication time courses performed, as described previously (*74*). Mitosis was induced in extracts by adding recombinant GST-Cyclin BΔ90 (40 ng/ml).

Cell lysis

For the cell lysis, we used a similar approach to that described by Lindeboom *et al (75)*. Briefly, each sample was thawed and homogenised with 15 µL of ice-cold cell lysis buffer (20mM Tris-HCl pH 8.0, 70mM KCl, 1mM EDTA, 10% glycerol, 5mM DTT, 0.125% Nonidet P-40, 1mM PMSF, 1x complete EDTA-free protease inhibitor, 1xPhoStop). Samples were subsequently centrifuged at max speed on a benchtop Eppendorf centrifuge. 10 µL of soluble material was recovered and snap-frozen in liquid nitrogen. Samples were stored at -80°C until further processing.

Protein digestion and phosphopeptide enrichment

Cell lysates were digested using the FASP method (*76*). Briefly: proteins were thawed and immediately reduced and alkylated with 10mM DTT and 0.05M iodoacetamide. Next, proteins were digested with Lys-C (overnight) at 37°C in a wet chamber, followed by addition of trypsin and further incubation under the same conditions for 4 hours. Both enzymes were used at 1:50 enzyme to protein ratio (protein quantification by Bradford assay showed that each individual egg provides ~20 µg of yolk free protein). For the egg extract experiment, each FASP filter was loaded with 200µg of protein. Peptides were cleaned using the Oasis HLB 96 well plates (Waters Corporation) and consequently subjected to phosphopeptide enrichment using Fe(III)-NTA 5µL cartridges in the automated AssayMAP Bravo Platform (Agilent Technologies), as described by Post *et al (41)*. Both flow through (peptides) and eluates (phosphopeptides) were dried down and stored at -80°C until further use.

LC-MS/MS analysis

All samples for label-free shotgun proteomics were analysed using a UHPLC 1290 system (Agilent Technologies) coupled to an Orbitrap Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Nano flow rate was achieved using a split flow setup aided by an external valve as described by Meiring *et al (77)*. Peptides were first trapped onto a pre-column (inner diameter [ID] of 100 µm and 2 cm length; packed in-house with 3µm C18 ReproSil particles [Dr. Maisch GmbH]) and eluted for separation into an analytical column (ID of 75 µm and 50cm length; packed in-house with 2.7 µm Poroshell EC-C18 particles (Agilent Technologies). The latter was done using a two-buffer system, consisting of buffer A (0.1% formic acid [FA] in water) and buffer B (0.1% FA in 80% ACN). Peptides were trapped during 5 minutes at 5 µL/min flow-rate with solvent A. For the measurement of the full proteome, we used a 155 min gradient from 10 to 36% of solvent B. For the phosphoproteome, we used a 95 min gradient from 8 to 32% of solvent B. Both methods included a wash with 100% solvent B for 5 minutes followed by a column equilibration with 100% solvent A for the last 10 minutes.

The mass spectrometer was operated in data dependent acquisition (DDA) mode. Full scan MS was acquired from 375-1600 m/z with a 60,000 resolution at 200 m/z. Accumulation target value was set to 3e6 ions with a maximum injection time of 20 ms. Up to 15 (12 for the phosphoproteome) of the most intense precursor ions were isolated (1.4m/z window) for fragmentation using high energy collision induced dissociation (HCD) with a normalised collision energy of 27. For MS2 scans an accumulation target value of 1e5 ions and a maximum injection time of 50 ms were selected. Scans were acquired from 200-2000m/z with a 30,000 resolution at 200m/z. Dynamic exclusion was set at 24s for the proteome and 12 s for the phosphoproteome.

For targeted proteomics, an EASY-nLC 1200 System (Thermo Fisher Scientific) coupled to an Orbitrap Q Exactive HF was used. Peptides were separated using an EASY-Spray analytical column (ID of 75µm and 25cm length; packed with 2µm C18 particles with a 100 Å pore size) (Thermo Fisher Scientific). Gradient lengths were shortened to 60 minutes. Phosphopeptides of interest from the previous experiment were selected and heavy-labeled versions were synthesised (JPT

Peptide Technologies). These synthetic standards were used during method development for retention time scheduling and instrument ion fill-time optimisation. Additionally, synthetic heavy peptides were pooled and combined with synthetic retention time peptide standards (iRT, Biognosys) to generate a spectral library, measured in DDA mode using the same LC-MS setup. This spectral library provided fragment intensity and retention time information for quality control assessment of targeted measurements. Samples were reconstituted in 2% FA containing ~200 fmol of each synthetic standard. The mass spectrometer was operated in data independent acquisition mode with an inclusion list of targets for parallel reaction monitoring (PRM). The list included the m/z values for the heavy and light versions of the phosphopeptides. Optimal measurement parameters were determined using test samples spiked with the heavy-labeled standards in order to guarantee optimal sensitivity for detection of endogenous phosphopeptides. We measured the targets of interest in a scheduled fashion, during a four-minute window with a 120,000 resolution, maximum injection time of 246ms and an accumulation target value of 2e5 ions, to ensure maximum specificity and sensitivity.

Data processing

DDA raw files were processed with MaxQuant (*78*) (v1.6.0.1) using a false discovery rate (FDR) <0.01. The default settings were used, with the following exceptions: variable modifications, specifically methionine oxidation, protein N-term acetylation and serine, threonine and tyrosine phosphorylation were selected. Cysteine carbamidomethylation was selected as a fixed modification. We also enabled the 'match between runs' option with the default values. Fractions were set so that matching was done only among biological replicates and samples of consecutive time points. The database search was conducted against a database generated by Temu *et al (79)*. This was particularly helpful since other publicly available databases contained several incomplete and/or poorly annotated sequences, which proved to be impractical for further data analysis.

The data was uploaded to the Perseus platform (*80*) for further analysis. Briefly: decoy sequences and potential contaminants were filtered out. Only high confidence localisation (>0.75 localisation probability) phosphosites were conserved for further analysis. Intensities were $\log_2$ transformed and then normalised by subtracting the median intensity of each sample. Biological replicates were grouped accordingly by time point; this grouping allowed us to filter the data and keep only those phosphosites that could be detected in at least two out of three biological replicates in any of the time points. Missing values were imputed from a random normal distribution applying a downshift of 1.8 times the standard deviation of the dataset, and a width of 0.3 times the standard deviation. This effectively replaced missing values at the lower end of the intensity distribution. We then performed an ANOVA (Benjamini-Hochberg FDR <0.05) to determine which phosphosites displayed statistically significant changes through the time course. Average phosphosite intensities were grouped using a combination of k-means and hierarchical clustering using the ComplexHeatmap package (*81*) in R. Protein intensities were processed in a similar fashion, removing proteins that were only identified by peptides that carry one or more modified amino acids.

Next, the full list of proteins with significantly changing phosphosites were matched against the human Uniprot database using the Basic Local Alignment Search Tool (BLAST), to render Uniprot identifiers that were compatible with different Gene Ontology (GO) analysis tools. We used the STRING web tool (*82*) to gain insight into the relation amongst dynamically phosphorylated proteins. The full list of phosphoproteins was uploaded and analysed using default settings. Next, the protein network was loaded into Cytoscape for clustering and visualisation. Proteins were clustered using GLay community clustering (*83*) and enrichment of GO terms per cluster was obtained using BiNGO (*84*) (shown in Figure 1B).

GO term enrichment was also acquired individually for each group (A-D) obtained after hierarchical clustering of dynamic phosphosites. For this we used STRING and filtered the enriched GO terms to keep only those with $p < 0.01$, fold enrichment >5 and a minimum of 5 proteins per term. The list of terms was further condensed by removal of redundant terms using the Revigo web tool (*85*). Remaining GO terms (including BP, MF and CC) were manually curated to further avoid redundancy. The final set of GO terms per cluster are shown in Figure S1. Following the same strategy, we analysed GO term enrichment for the interphase and mitotic clusters from the *in vitro* dataset separately (shown in Figure S3).

PRM raw files were analysed with Skyline software (*86*). Signal quality for each target of interest was assessed visually for all samples. Quality control of endogenous signals was done by confirming the perfect co-elution of both peptide forms (heavy and light), assessing their retention time and peak shape. We also used the similarity of the relative intensity of fragment ions (rdotp > 0.9) between light and heavy to exclude signals that showed poor correlation. Quantifications were done with a minimum of three fragments per phosphopeptide. The data was loaded into R for data cleanup and visualisation, using the Complex-Heatmap and ggplot2 packages.

Motif analysis

Obtained phosphopeptides were aligned by centering them around the phosphosite detected and the conserved motifs for the different kinases were determined using regular expressions by applying the following rules:

PLK: [D/N/E/Y]-X-[S/T]-[Hydrophobic / ^P]

AURA/AURB: [K/R]-X-[S/T]*[^P]

NEK: [L|M|F|W]-X-[S|T]*-[A|V|I||L|F|W|Y|M]-[K|R]

Casein kinase 1: [D/E]-[D/E]-[D/E]-X-X-[S/T]* or [S/T]-X-X-[S/T]*

Casein kinase 2:[S/T]-[S/T]*-X-[E/D/S]

DDK: [S/T]*-[E/D]-X-[E/D] or [S/T]*-[S/T]-P

PKA: R-[R/K]-X-[S/T]*-[Hydrophobic]

Cdk full consensus motif: [S/T]*-P-X-[K/R]

Cdk minimal consensus motif: [S/T]*-P

Where [ ] groups multiple amino acids for one position, ^ at the left of a certain amino acid informs that it is forbidden for that position, and X represents any amino acid.

Data collection for human and yeast CDK1 targets

Data of CDK1 substrates for *S. cerevisiae* were downloaded from online supplementary information of papers describing two different studies using *in vitro* (*28*) and *in vivo (22)* approaches, respectively. We defined high confidence yeast CDK1 targets as the intersection of both datasets. Other phosphorylations detected in both studies for which there was no evidence for CDK1 involvement were considered as the non-CDK1-mediated phosphoproteome (universe). For human CDK1 subfamily targets, we extracted information available in the PhosphoSitePlus database (*87*). An additional step of manual curation from the following studies (*25, 26, 70, 88-103*) was performed to obtain a high confidence human CDK1 subfamily targets dataset. The phosphoproteome universe was constructed with all the phosphorylated proteins deposited in the PhosphoSitePlus database after subtraction of the CDK1 subfamily targets.

Data collection for MLO proteomes

Data from proteomics studies of the composition of MLOs characterised by liquid-liquid phase separation was obtained from the following sources: stress granules (*104, 105*), nuclear speckles (*104, 106*), PML nuclear bodies (*104, 107*), P-bodies (*108*), nucleoli (*109, 110*), nuclear pore complexes (*111*), Cajal bodies (*104, 112*), Super-enhancer-Mediator condensates (*113*).

Prediction of intrinsically disordered regions

For the UniProt proteomes of human, yeast and *Xenopus laevis*, disorder information was fetched from MobiDB (*114*) with the exception of SPOT disorder predictor, which was calculated for all the proteins of each dataset. For *Xenopus* proteomics studies, we used the available standalone software of IUPred, VSL2B, and SPOT to predict IDRs in all the proteins of the database.

Differential disorder composition

For the three organisms analysed (*Xenopus,* human and yeast), the amino acid composition for the entire phosphoproteome and for the disordered regions of the phosphoproteome was calculated. For each amino acid, we estimated the differential disorder composition with the equation: (Comp. Disorder – Comp. Phosphoproteome) / Comp. Phosphoproteome

Positives values show amino acids enriched in disordered regions while negative values represent amino acids depleted in disordered regions.

Bioinformatic and statistical analysis of disorder and phosphorylation

All the statistical analysis was performed with the R programming language (https://www.r-project.org/) using R studio as an integrated development environment (available at https://rstudio.com/). The packages Tidyverse and Bioconductor (*115*) were used for cleaning, manipulation, and graphical representation of the data. Sequence logos were generated using the information content as described by Douglass *et al (116)*. IUPred scores were plotted with an *ad hoc* designed script, available upon request.

For the contingency table analysis, the disordered regions of CDK targets and dynamic phosphoproteins were calculated with three predictors (IUPred, VSL2B, and SPOT). For each combination of disorder predictor and phosphorylation dataset, a two-by-two table with the counts of phosphorylatable residues (Ser/Thr) phosphorylated or not, either located in IDRs or in structured regions (Figure S6B), was generated. Each table was then analysed with the Fisher test for obtaining the odds ratio and the associated P-value.

The source code for all the analysis conducted in this publication is available upon request.

Calculation of SCDMs

Elements (i,j) of SCDM were calculated using equation 1 of Huihui and Ghosh (*66*). In this coarse grain model, each amino acid is considered a point with a charge q = -1 for Aspartic and Glutamic acid, charge q = 1 for Arginine and Lysine, charge q = 0 for all other amino acids. Phosphorylation is modeled by replacing neutral charge of Serine, Threonine to q = -2 to mimic the effect of double negative charge of phosphate groups. SCDM maps were made visually continuous between neighboring (i,j) pairs using spline-16 interpolation.

All atom simulation

We performed all-atom Monte Carlo simulation of MCM4, TICRR, CDT1, Coilin (unphosphorylated and phosphorylated forms) to compute radius of gyration by using CAMPARI (version 2) based on the ABSINTH implicit solvation paradigm (*117, 118*). Simulations were carried out at 298 K. For CDT1, 16 trajectories were generated with each trajectory having 10 million steps. For MCM4, 27 trajectories were generated with each trajectory having 6.5 million steps. For Coilin, 54 trajectories were generated with each trajectory running for 4 million steps. For TICRR, 90 trajectories were generated with each trajectory having 3 million steps. For each trajectory, irrespective of the sequence, first 1.5 million steps were discarded due to equilibration yielding a cumulative (over all trajectories) of at least 135 million steps for each sequence. Each of the simulations were run using zero salt with only neutralizing Na^+ and/or Cl^- ions added in a droplet of 400 Angstrom. PDB files were generated every 5000 Monte Carlo steps. Phosphorylated sequences were modeled by replacing Serine or Threonine by two Aspartic acids. ACE cap and NME tail were added to each protein sequence.

Coarse-grained force field

We adopted a one-bead-per-residue coarse-grained (CG) model that has been shown to capture the structural and phase separation properties of flexible proteins as a function of their sequence (*119*) and phosphorylation pattern (*120*). In this framework, bonded interactions between neighboring residues were modeled using a harmonic potential with a constant of 1000 kJ/mol/nm2 and a bond length of 0.38 nm. Electrostatic interactions between charged residues, i.e. Asp, Glu (-1e); Lys, Arg (+1e), His (+0.5e); ph-Ser, ph-Thr (-2e), were computed with a screened Coulomb potential, using a Debye-Huckel length of 1 nm. Short-range interactions were modeled by Lennard-Jones (LJ) potentials defined by

$$V_{ij}(r) = 4\lambda_{ij}\epsilon\left((\sigma_{ij}/r)^{12}-(\sigma_{ij}/r)^{6}\right)$$

where values of λ_{ij}, ϵ, and σ_{ij} used reported values (*119, 120*) with the exception of λ_{ArgArg} that was set equal to 0.01, instead of 0.00, in order not to neglect excluded volume effects while using LJ functional form. In the simulation of full-length Ki-67, the conformation of the N-terminal folded domain (1-128) was restrained by an elastic network based on experimental structure (PDB: 1R21) with a distance cutoff of 2 nm and an elastic constant of 5000 kJ/mol/nm2. All MD simulations were performed with GROMACS 2018.3 (*121*). Simulations were run in the NVT ensemble, controlling the temperature by means of a Langevin thermostat with a friction constant of 25 ps^{-1} and a time step of 10 fs.

Single chain MD simulations of full-length Ki-67

The structure with PDB id 1R21 was used to model the conformation of the first 128 residues of the full-length Ki-67, while TraDES (*122, 123*) was used to generate the initial conformation of the remaining disordered part of the non-phosphorylated molecule. The models were then joined using UCSF Chimera 1.14 (*124*) . The initial configuration for the phosphorylated full-length Ki-67 was obtained by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-

Thr) the amino acids in the non-phosphorylated structure. Initial configurations were inserted in a cubic simulation box with a side of 100 nm with periodic boundary conditions. We employed a Parallel Tempering (PT-MD) scheme, with eleven replicas in the 300-400K range in order to enhance the conformational sampling and obtain a reliable estimate of the radius of gyration of the protein as a function of the temperature. PT-MD simulation ran for 1e8 steps, attempting Monte Carlo exchanges between neighboring replicas every 1e2 steps and saving snapshots every 1e4 steps for further analysis.

Single chain MD simulations of consensus repeats

The initial extended configuration of the non-phosphorylated consensus repeat of Ki-67 was generated with the tLEAP tool available in AmberTools18 (*125*), while the phosphorylated structure was generated by 'mutating' (i.e. using parameters that have been tuned specifically for ph-Ser/ph-Thr) the residues in the non-phosphorylated configuration. PT-MD simulation followed the same protocol employed for the simulations of full-length Ki-67, with the only difference being the range of temperatures, between 200K and 500K, with intervals of 10K.

Phase Coexistence MD simulations

Phase coexistence simulations of monomers and dimers of the non-phosphorylated consensus repeat and of monomers of the phosphorylated consensus repeat were carried out employing the slab method (*119*). Initial configurations of the slab simulations for each system were generated by inserting 200 copies of the respective molecules in a 20x20x30nm box in random positions and orientations with the *gmx insert-molecules* tool available in GROMACS 2018.3 (*121*), and extending the simulation box in the z direction to 200 nm. Simulations were run for 1.5e8 steps, saving frames every 10000 steps for further analysis, discarding the first 0.5e8 steps of the simulations as equilibration.

Estimation of theta temperatures

Coil-to-globule transition temperatures (T_θ) were estimated for non-phosphorylated full-length Ki-67 and the monomer of the non-phosphorylated consensus repeat from single chain PT-MD simulations, following a reported approach (*126*). Intramolecular distances as a function of the chain separation |i-j| were computed for each temperature replica and fitted with the following expression:

$$R_{ij} = 0.55|i\text{-}j|^{\nu}$$

where the scaling exponent ν is the fitting parameter. We then determined T_θ as the temperature corresponding to ν = 0.5

Estimation of binodal curves

Density profiles from phase coexistence simulations were estimated by means of the gmx density tool available in GROMACS 2018.3, centering the dense phase in the middle of the z-axis. The concentration of the diluted phase was computed by averaging the density in the box at 0nm<z<60nm and 140nm<z<200nm, while the concentration of the dense phase was obtained by averaging the density at 90nm<z<110nm. Following the approach of Dignon *et al* (*119*) the critical temperatures, T_c, for the three systems investigated with phase coexistence simulations, have been evaluated by fitting the following equation:

$$\rho_H\text{-}\rho_L = A(T_c\text{-}T)^{0.325}$$

Human opto-Ki-67 plasmid construction

pCDNA5_FRT_Ki67-FL-mCherry-Cry2 was generated by PCR amplification of human Ki67. The fragment was cloned into the AflII/KpnI digested pCDNA5_FRT_TO_TurboID-mCherry-Cry2 (Addgene plasmid # 166504) using the In-Fusion HD Cloning Kit protocol.

Generation of Flp-In T-REx 293 opto-Ki67 cell lines

Flp-In T-REx 293 (Termo Fisher Scientifc, Darmstadt, Germany) cell line was grown under standard conditions (37˚C, 5% CO2) in Dulbecco's modified Eagle's medium (Merck-Sigma-Aldrich, D5796). The medium was supplemented with 10% fetal bovine serum (FBS), 100 µg/ ml Zeocin and 15 µg/ml Blasticidin. One million HEK-293 T-REx cells were plated in a 6-well plate 24 hours before transfection. The next day, 500 ng of each optoKi67 expression plasmid is combined with 3.5 µg pOG44 encoding the Flp recombinase. Transfection was made with 8 µl Lipofectamine 2000 according to the manufacturer's instructions. 48 hours post transfection, cells were transferred to a 100 mm petri dish. On the next day, selection was performed by adding hygromycin B at a final concentration of 50 µg/mL. Around 14 days after selection, clones were pooled and expanded. The cells were tested for the expression of the construct by immunofluorescence. Flp-InT-Rex 293 opto-Ki67 stable cell lines were maintained with 15 µg/mL Blasticidin and 15 µg/mL Hygromycin.

Opto-Ki-67 activation

Cells were plated in DMEM on coverslips a day prior to activation. Expression of opto-Ki67 was induced with 2 µg/ml doxycycline for 16 hours. For light activation, plates were transferred into a custom-made illumination box containing an array of 24 LEDs (488nm) delivering 10 mW/cm2 (light intensity measured using a ThorLabs-PM16-121-power meter). Cry2 activation was induced using 4 min of blue light cycles: 4s On followed by 10s Off. Cells were fixed with 4% paraformaldehyde (PFA) for 15 min at RT, counterstained with Hoechst (Invitrogen, Cat H21491) in PBS-TritonX (0.2%), and mounted on glass slides using Prolong Gold antifade reagent (Invitrogen, Cat P36930). Images were captured using a 63x objective (NA 1.46 oil).

To check the effect of inhibitors on Ki-67 foci formation, the cells were incubated for 1h with 0.5 µM okadaic acid, 5 µM purvalanol A or vehicle (DMSO) for 1 hour prior to light activation and fixation. Foci number was analysed using FIJI Software and statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise post-hoc comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method. Plots were generated using the ggplot2 library in R.

Ki-67 consensus repeat DNA construct

The cDNA sequence coding for Ki-67 consensus repeat (Ki67-CR) was ordered from IDT® gene synthesis. It was subsequently cloned into pDB-GFP plasmid between HindIII and XhoI sites to obtain the pDB-GFP-Ki-67-CR vector. In this construct, Ki-67-CR was fused with a (his)6-GFP N-terminal tag. GFP sequence is followed by the HRV 3C (3C) protease recognition site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro). Specific cleavage can occur between Gln and Gly, with Gly-Pro remaining at the C terminus Ki-67-CR without any tag.

Ki-67 consensus repeat expression and purification

The pDB-GFP-Ki-67-CR plasmid was transformed into *E. coli* BL21(DE3); transformed cells were grown overnight at 25°C in N-5052 auto-induced medium (*127*), supplemented with 50 µg/ml kanamycin and 15N NH4Cl. Cells were harvested by 20 min centrifugation at 6000g at 4°C. The pellet was resuspended in 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 2mM DTT (buffer A) and stored at -80°C. Cells were supplemented with a Complete® EDTA free tablet (Roche), lysed by sonication, insoluble proteins and cell debris were sedimented by centrifugation at 40000g at 4°C for 30 min. Supernatant was supplemented with imidazole to 5 mM final and loaded onto 5ml gravity affinity columns (Ni Sepharose Excel 5ml, Cytiva), equilibrated with buffer A. Columns were washed with 50 ml of buffer A and proteins were eluted with a one-step gradient of buffer B (buffer A containing 500 mM imidazole). The peak fractions were analysed by SDS-PAGE. Fractions containing tagged Ki-67-CR were pooled and dialysed overnight at 4°C against buffer C (50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT). The dialysed protein was then loaded on a Superdex S200 16/60 (HiLoad 16/600 Superdex 200pg, Cytiva) equilibrated with buffer C. Fractions containing the protein of interest were analysed by SDS-PAGE and pooled. The purified GFP-Ki-67-CR protein was concentrated to 5 mg/ml with Vivaspin centrifuge concentrator (Sartorius Stedim Biotech).

In vitro Ki-67 peptide phosphorylation assay

Protein was desalted by using PD10 Mini-Trap column in 50 mM Hepes 7.5, 50 mM NaCl, 2 mM DTT. Phosphorylation reaction was performed in a total volume of 220 µl, and contained 140 µl of GFP-Ki-67-CR (400 µg), 5mM $MgCl_2$, 500 µM ATP, 50mM β-glycerophosphate, recombinant CDK1-cyclin B/CKS1 (5 µg; gift from Jane Endicott and Tony Ly) in 50 mM Hepes pH 7.5, 50 mM

NaCl, 2 mM DTT. Reaction was incubated at 30°C for 18 hours and stopped by adding 10 mM EDTA. It was subsequently desalted by using PD10 Mini-Trap column in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT, and used to perform NMR experiments, Phos-tag SDS-PAGE and phosphoproteomics.

NMR experiments and data analysis

All NMR samples contained final concentrations of 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a 1H frequency of 700 and 800 MHz. 15N-HSQC was acquired for each sample in order to determine amide (1HN and 15N) chemical shifts of non-phosphorylated and phosphorylated GFP-tagged Ki-67-CR. 15N-HSQC spectra were acquired for respectively non-phosphorylated (and phosphorylated) proteins at 800 (700) MHz using 32 (128) scans, 128 (256) increments and a spectral width of 22.5 ppm in the indirect dimension. All spectra were processed with TopSpin v3.5 (Bruker Biospin) and analysed using CCPN-Analysis software (*128*). Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the 1H/X frequency ratio of the zero point according to Markley *et al* (*129*).

Total internal reflection fluorescence (TIRF) experiments

In order to measure the ability of the phosphorylated and non-phosphorylated Ki-67 repeat to liquid-liquid phase separate, 50 µM protein was prepared in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT. Dextran was added just before preparing samples on circular glass coverslips (2.5 cm, 165 µm thick, Marienfeld). Coverslips were cleaned with a 15 min cycle of sonication with ultrasounds in 1M KOH, followed by a second cycle of sonication in deionized water. Samples were deposited into wells of Press-to-seal silicone isolater with adhesive (Invitrogen), and covered with a second coverslip to avoid evaporation. Images were acquired with a custom-made TIRF microscope using a LX 488-50 OBIS laser source (Coherent). Oil immersion objective with a 1.4 numerical aperture (Plan-Apochromat 100x, Zeiss) was used. Fluorescence was collected with an EmCCD iXon Ultra897 (Andor) camera. The setup includes a 1.5x telescope to obtain a final imaging magnification of 150-fold, corresponding to a camera pixel size of 81.3 nm. Fluorescence images at different time points were obtained by averaging 150 individual images, each acquired over 50 ms exposure time.

Phos-tag SDS-PAGE

For Phos-tag SDS-PAGE (12.5% Phos-tag^{TM} SuperSep^{TM} pre-cast gel 50 µmol/L; Fujifilm Wako Chemicals #193-16571), 500 ng of unphosphorylated and 500 ng of CDK1-cyclin B-CKS1 phosphorylated GFP-Ki-67-CR protein were loaded. SDS-PAGE was performed following standard protocol, with the exception of two additional washes 20 min each in transfer buffer containing

10 mM EDTA, followed by a wash in transfer buffer, preceding wet transfer. For Western blot, anti-GFP antibody (rabbit polyclonal Chromotek PABG1; 1:10 000) was used.

Mapping of Ki-67-CR phosphorylation sites by mass spectrometry

6.5 µg of GFP-Ki-67-CR protein phosphorylated by CDK1-cyclin B-CKS1 were digested in a FASP filter as described earlier for the other samples. Peptides were subsequently cleaned using C18 cartridges and phospho-enriched using Fe(III)-NTA cartridges in the AssayMAP Bravo. Phosphopeptides were measured in a technical duplicate, acquiring in DDA mode with an Ultimate 3000 uHPLC system coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific) during a 60 minutes gradient. Raw files were searched against the Ki-67-CR sequence using MaxQuant with the same parameters as applied to the other phosphoproteomics experiments.

Supplementary material

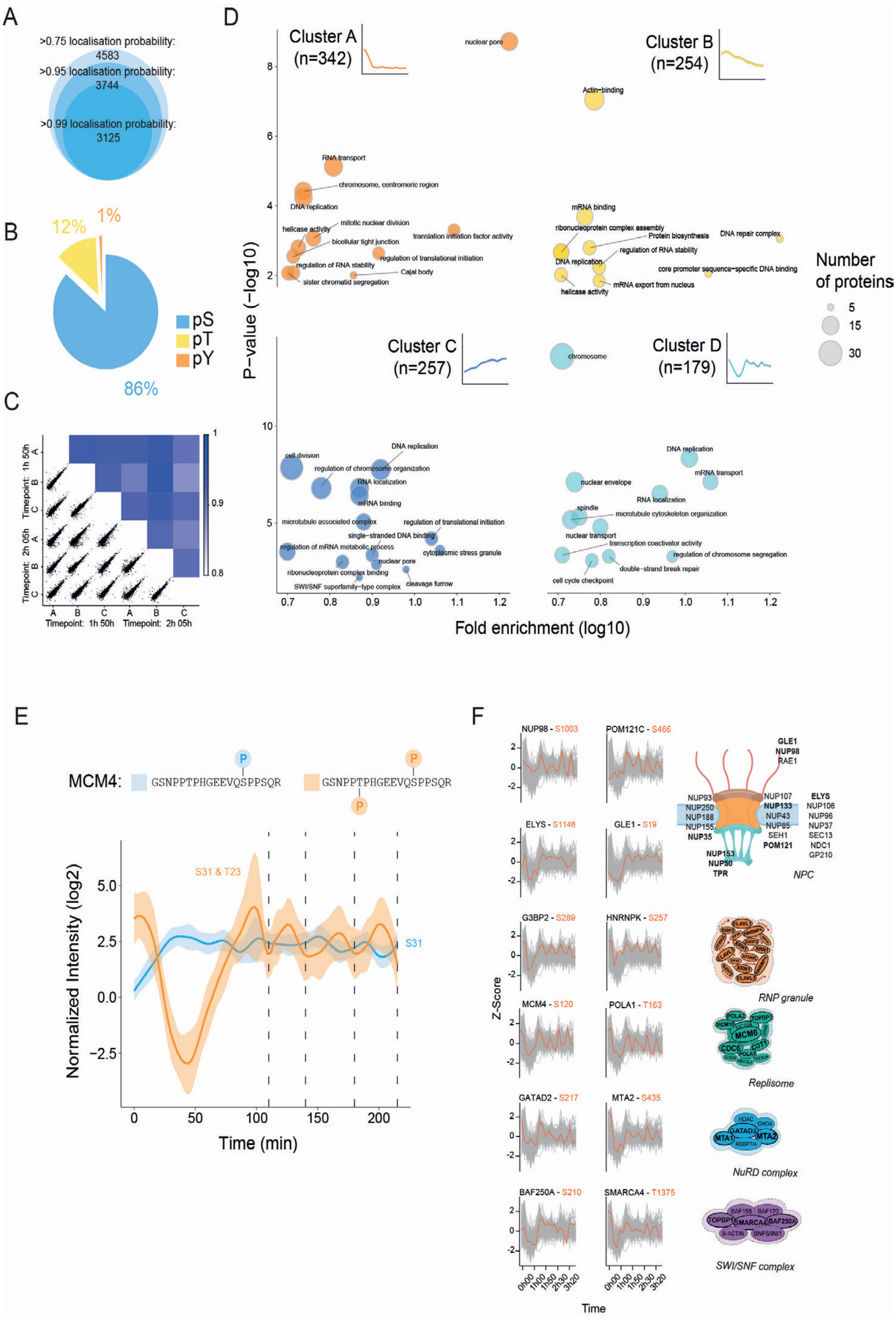


Figure S1. Phosphosite dynamics correlates with cell cycle phases; related to Figure 1. (A) Total number of phosphosites detected and their distribution according to the site localisation probability score. (B) Distribution of phosphosites identified among serine, threonine and tyrosine residues. (C) Correlation coefficients for two randomly selected time points. Scatter plots of significantly enriched (Fisher's exact test with Bonferroni correction, $p<0.05$) GO (BP, MF, CC, Uniprot keywords) terms for all dynamic phosphosites per cluster in the *in vivo* experiment, presenting the fold-enrichment of specific terms vs statistical significance. The size of the circles correlates with the number of proteins associated with the specific term. More details and the full list of enriched GO terms per cluster is found in Data S1. (D) *In vivo* reciprocal trends of singly- and multi-phosphorylated peptides carrying phosphorylated T23 and S31 of MCM4 (dashed lines depict the time points of cell division): orange curves, the trend of T23 and S31 in the multi-phosphorylated peptide; blue curve, the trend of S31 in the singly-phosphorylated peptide. (E) Examples of proteins with known association showing similar oscillating phosphorylation. Plots highlight the dynamic trend of the cluster (grey) and selected phosphosites (orange) over time. Right, illustrations of protein complexes formed by the proteins undergoing dynamic phosphorylation. Proteins highlighted in bold show at least one oscillating phosphosite in our dataset.

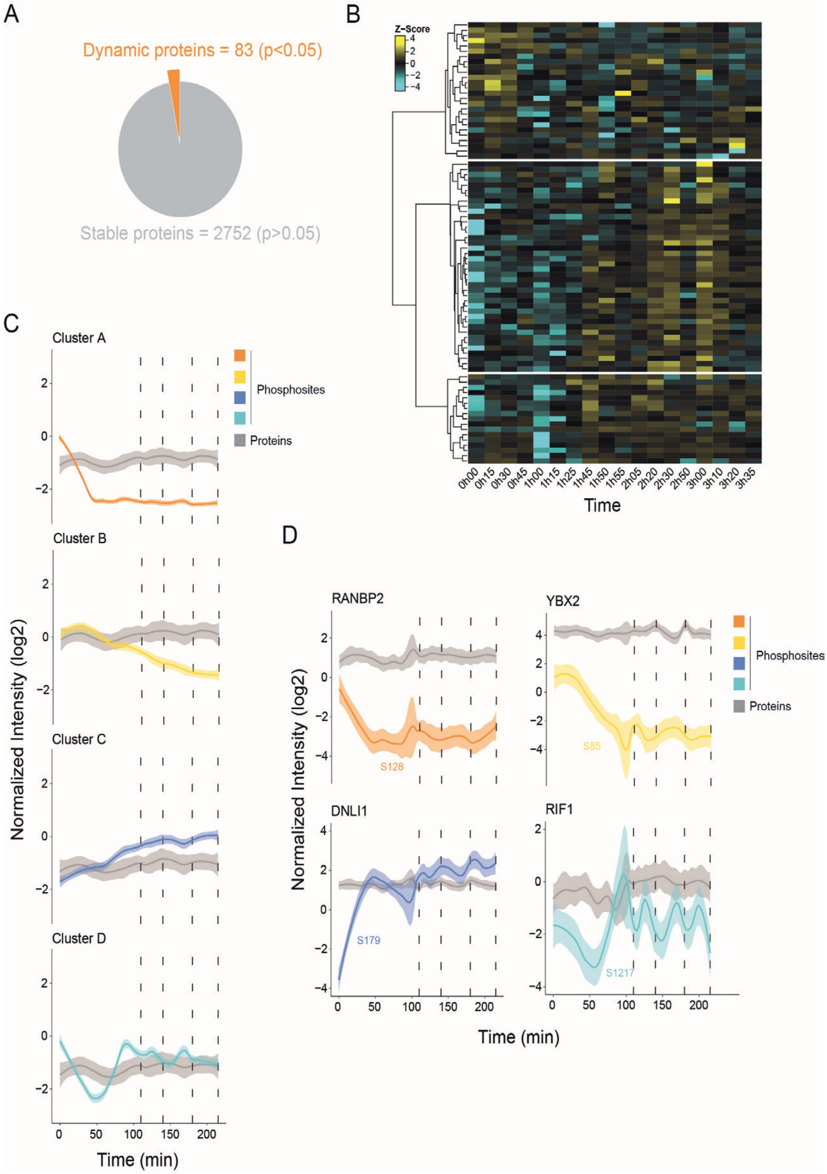


Figure S2. Variation in the total proteome versus the phosphoproteome in early embryonic cell cycles; related to Figure 1. (A) Total proteome analysis reveals 83 proteins out of 2835 showing significant changes in abundance (ANOVA, Benjamini-Hochberg correction, FDR 0.05) over the time course. (B) Heat map showing abundance of the variable proteins over the time course. (C) Comparison of dynamic variations in total protein compared to total phosphosites from the four clusters shown in Figure 1C (dashed lines depict the time points of cell division). (D) Examples of dynamics of individual phosphosites from the four clusters shown in Figure 1C and levels of the corresponding protein (dashed lines depict the time points of cell division).

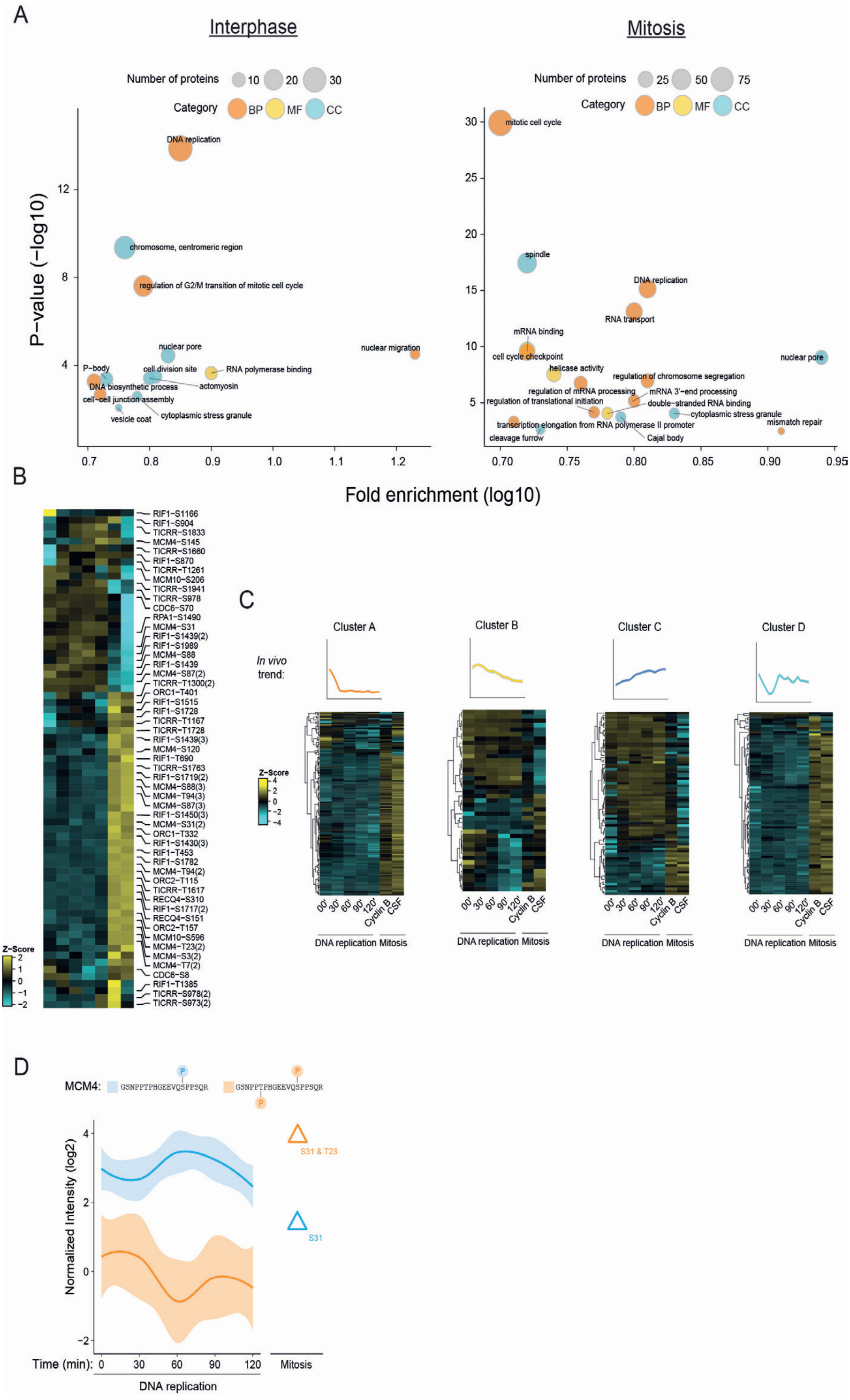
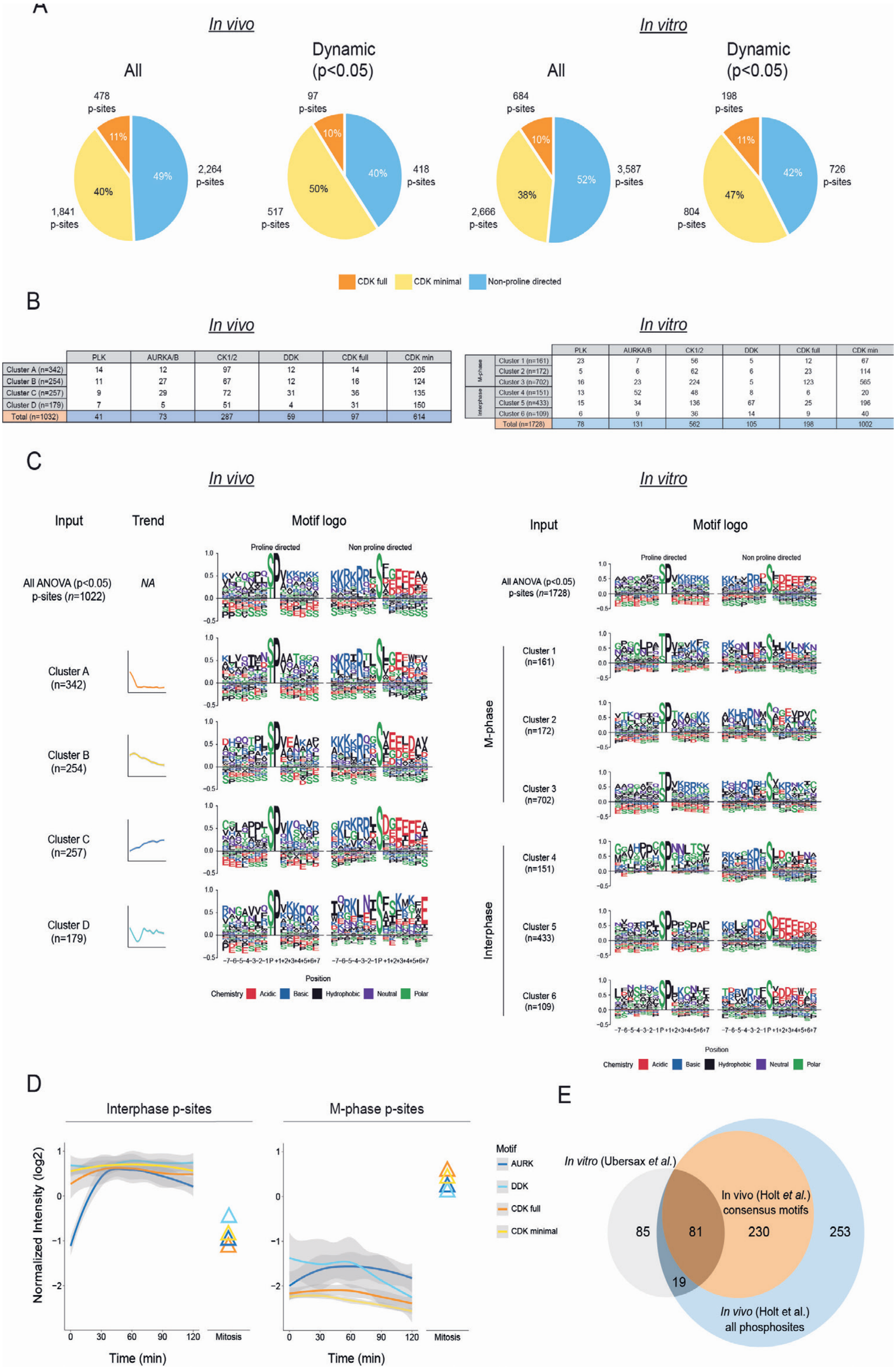

A
Interphase
Number of proteins 10 20 30
Category BP MF CC
P−value (−log10)
DNA replication
chromosome, centromeric region
regulation of G2/M transition of mitotic cell cycle
nuclear pore
nuclear migration
P−body
cell division site
RNA polymerase binding
DNA biosynthetic process
actomyosin
cell−cell junction assembly
vesicle coat
cytoplasmic stress granule
4
8
12
0.7 0.8 0.9 1.0 1.1 1.2
Mitosis
Number of proteins 25 50 75
Category BP MF CC
mitotic cell cycle
spindle
DNA replication
RNA transport
mRNA binding
nuclear pore
cell cycle checkpoint
helicase activity
regulation of chromosome segregation
regulation of mRNA processing
mRNA 3'−end processing
regulation of translational initiation
double−stranded RNA binding
cytoplasmic stress granule
transcription elongation from RNA polymerase II promoter
mismatch repair
cleavage furrow
Cajal body
5 10 15 20 25 30
0.70 0.75 0.80 0.85 0.90 0.95
Fold enrichment (log10)
B
RIF1−S1166
RIF1−S904
TICRR−S1833
MCM4−S145
TICRR−S1660
RIF1−S870
TICRR−T1261
MCM10−S206
TICRR−S1941
TICRR−S978
CDC6−S70
RPA1−S1490
MCM4−S31
RIF1−S1439(2)
RIF1−S1989
MCM4−S88
RIF1−S1439
MCM4−S87(2)
TICRR−T1300(2)
ORC1−T401
RIF1−S1515
RIF1−S1728
TICRR−T1167
TICRR−T1728
RIF1−S1439(3)
MCM4−S120
RIF1−T690
TICRR−S1763
RIF1−S1719(2)
MCM4−S88(3)
MCM4−T94(3)
MCM4−S87(3)
RIF1−S1450(3)
MCM4−S31(2)
ORC1−T332
RIF1−S1430(3)
RIF1−T453
RIF1−S1782
MCM4−T94(2)
ORC2−T115
TICRR−T1617
RECQ4−S310
RIF1−S1717(2)
RECQ4−S151
ORC2−T157
MCM10−S596
MCM4−T23(2)
MCM4−S3(2)
MCM4−T7(2)
CDC6−S8
RIF1−T1385
TICRR−S978(2)
TICRR−S973(2)
Z−Score 2 1 0 −1 −2
C
In vivo trend:
Cluster A
Cluster B
Cluster C
Cluster D
Z−Score 4 2 0 −2 −4
00' 30' 60' 90' 120' Cyclin B CSF
DNA replication Mitosis
D
MCM4: GSNPPTPHGEEVQSPPSQR GSNPPTPHGEEVQSPPSQR
P
S31 & T23
S31
Normalized Intensity (log2)
−2 0 2 4
Time (min): 0 30 60 90 120
Mitosis
DNA replication

Figure S3. *In vitro* phosphoproteomics discriminates interphase and mitotic phosphorylation; related to Figure 1. (A) Scatter plots of significantly enriched (Fisher's exact test with Bonferroni correction, $p<0.05$) GO terms for all dynamic phosphosites upregulated during interphase (left) and mitosis (right), presented as fold-enrichment of specific terms vs statistical significance. The size of the circles correlates with the number of proteins associated with the specific term, while the color corresponds to the GO term category. (B) Heatmap of dynamic phosphosites detected in DNA replication factors. (C) Behaviour of *in vivo* dynamic phosphosites (top) in *in vitro* experiments. (D) *In vitro* dynamics of T23 and S31 of MCM4. Orange curve shows upregulation of the multi-phosphorylated peptide (T23 and S31) during mitosis, while the blue curve shows the opposite trend for the singly-phosphorylated peptide (S31), as observed *in vivo*, in Figure S1D.



In vivo

	PLK	AURKA/B	CK1/2	DDK	CDK full	CDK min
Cluster A (n=342)	14	12	97	12	14	205
Cluster B (n=254)	11	27	67	12	16	124
Cluster C (n=257)	9	29	72	31	36	135
Cluster D (n=179)	7	5	51	4	31	150
Total (n=1032)	41	73	287	59	97	614

In vitro

		PLK	AURKA/B	CK1/2	DDK	CDK full	CDK min
M-phase	Cluster 1 (n=161)	23	7	56	5	12	67
	Cluster 2 (n=172)	5	6	62	6	23	114
	Cluster 3 (n=702)	16	23	224	5	123	565
Interphase	Cluster 4 (n=151)	13	52	48	8	6	20
	Cluster 5 (n=433)	15	34	136	67	25	196
	Cluster 6 (n=109)	6	9	36	14	9	40
	Total (n=1728)	78	131	562	105	198	1002

Figure S4. CDK consensus phosphosites dominate the early embryo phosphoproteome; related to Figure 1. (A) Distribution of potential CDK targets among all detected phosphosites and dynamic phosphosites, *in vivo* (embryo, left) and *in vitro* (egg extract, right). (B) Observed phosphorylation motifs in the dynamic phosphoproteome *in vivo* (left) and *in vitro* (right). See methods for details. Note: in some cases, the sum of consensus sites exceeds the number of phosphosites due to redundancy between motif predictions. (C) Sequence motif logo for all dynamic phosphosites and for each of the clusters shown in Figure 1c, for the *in vivo* and *in vitro* experiments. Motifs are shown separately for proline-directed and non-proline directed phosphosites. (D) Dynamic trend of phosphorylations of potential kinase targets in egg extract S-phase clusters (4-6, left) and in mitotic clusters (1-3, right). (E) Venn diagram of observed *in vitro* (grey) and *in vivo* (blue) yeast CDK targets. *In vivo* targets showing CDK minimal consensus motif phosphorylations are highlighted in orange.

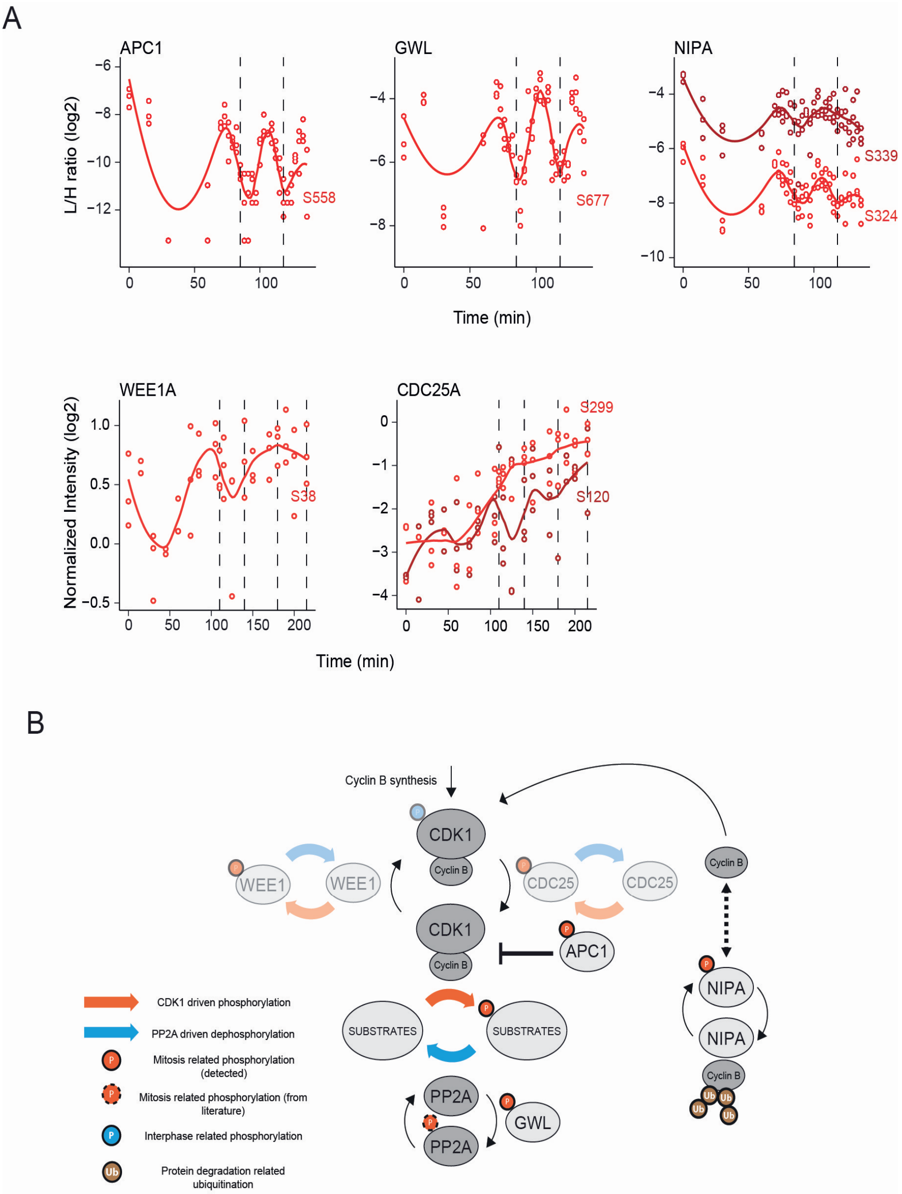
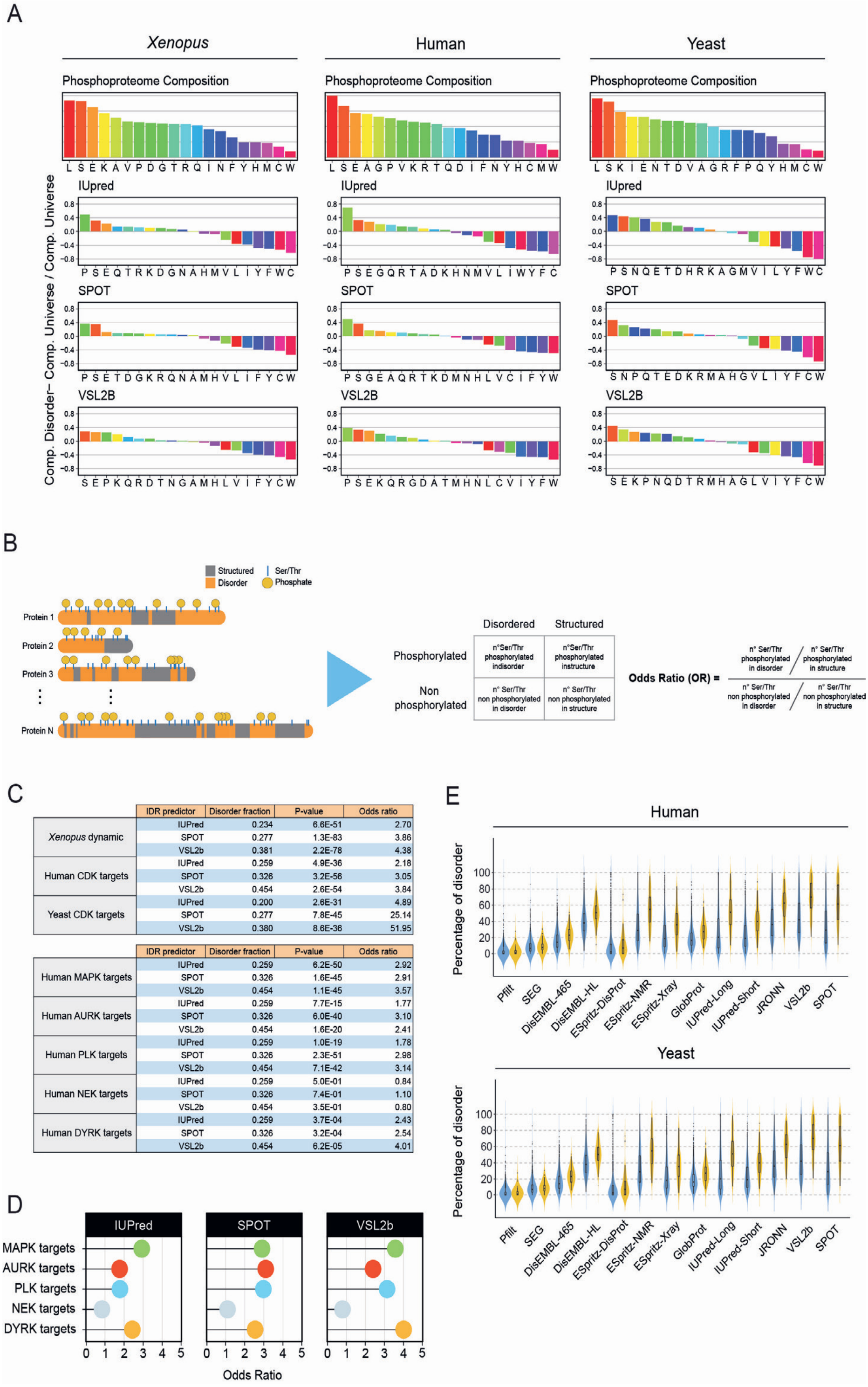


Figure S5. Phosphorylation dynamics of the CDK1-oscillator network; related to Figure 2. (A) Single phosphosite plots of CDK1 regulators measured by targeted (top) or shotgun (below) phosphoproteomics. (B) CDK1-oscillator network: our data suggests that control of cyclin levels via positive (*e.g.* NIPA ubiquitin ligase) and negative (*e.g.* APC) feedback loops, accompanied by PP2A inactivation via GWL, can generate oscillation of CDK1 activity during early cell divisions. CDK1-Y15 regulation via feedback loops consisting of CDC25 and WEE1A (greyed out) seems to be less important for switch-like mitotic phosphorylation after the first cell division.



	Disordered	Structured
Phosphorylated	n°Ser/Thr phosphorylated indisorder	n°Ser/Thr phosphorylated instructure
Non phosphorylated	n° Ser/Thr non phosphorylated in disorder	n° Ser/Thr non phosphorylated in structure

$$\textbf{Odds Ratio (OR)} = \frac{\text{n° Ser/Thr phosphorylated in disorder} \,/\, \text{n° Ser/Thr phosphorylated in structure}}{\text{n° Ser/Thr non phosphorylated in disorder} \,/\, \text{n° Ser/Thr non phosphorylated in structure}}$$

	IDR predictor	Disorder fraction	P-value	Odds ratio
Xenopus dynamic	IUPred	0.234	6.6E-51	2.70
	SPOT	0.277	1.3E-83	3.86
	VSL2b	0.381	2.2E-78	4.38
Human CDK targets	IUPred	0.259	4.9E-36	2.18
	SPOT	0.326	3.2E-56	3.05
	VSL2b	0.454	2.6E-54	3.84
Yeast CDK targets	IUPred	0.200	2.6E-31	4.89
	SPOT	0.277	7.8E-45	25.14
	VSL2b	0.380	8.6E-36	51.95

	IDR predictor	Disorder fraction	P-value	Odds ratio
Human MAPK targets	IUPred	0.259	6.2E-50	2.92
	SPOT	0.326	1.6E-45	2.91
	VSL2b	0.454	1.1E-45	3.57
Human AURK targets	IUPred	0.259	7.7E-15	1.77
	SPOT	0.326	6.0E-40	3.10
	VSL2b	0.454	1.6E-20	2.41
Human PLK targets	IUPred	0.259	1.0E-19	1.78
	SPOT	0.326	2.3E-51	2.98
	VSL2b	0.454	7.1E-42	3.14
Human NEK targets	IUPred	0.259	5.0E-01	0.84
	SPOT	0.326	7.4E-01	1.10
	VSL2b	0.454	3.5E-01	0.80
Human DYRK targets	IUPred	0.259	3.7E-04	2.43
	SPOT	0.326	3.2E-04	2.54
	VSL2b	0.454	6.2E-05	4.01

Figure S6. Cell cycle phosphorylation occurs preferentially in IDRs; related to Figure 3. (A) Differential amino acid composition (see methods) in disordered regions for *Xenopus*, human and yeast determined with three IDR predictors. Amino acids are coloured in a rainbow pattern according to their relative abundance in each phosphoproteome. Disruptions of the rainbow pattern show specific compositional signatures for IDRs. (B) Scheme of the Odds Ratio analysis using contingency tables. The counts of phosphorylated Ser/Thr for all the proteins for each set (CDK-mediated in yeast/human, or dynamic in *Xenopus*, and other cell cycle-related kinases in human), in disordered and structured regions, are stored in a 2x2 table. (C) Tables showing results of statistical analysis of Odds Ratio with Fisher's test, using three disorder predictors. (D) Plots of the Odds Ratio for human cell cycle-related non-CDK kinases and MAPK. (E) Violin plots of the distribution of percentage of disordered residues per protein for CDK targets *vs* the rest of the phosphoproteome for human and yeast. Intrinsic disorder information of 13 different predictors was obtained from MobiDB, except for SPOT (calculated).

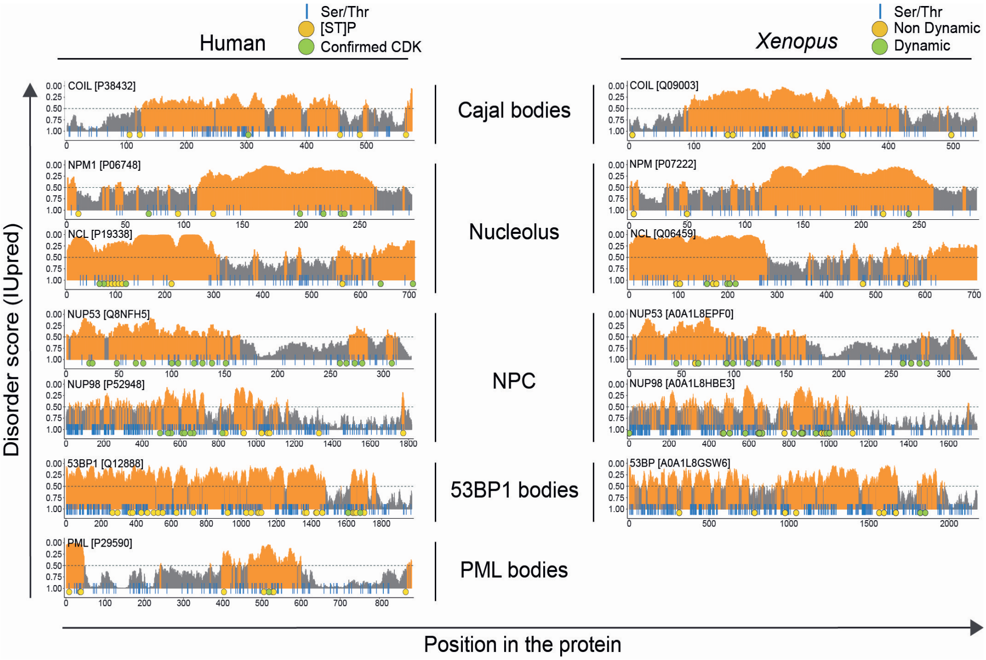


Figure S7. Cell cycle phosphorylation of key MLO proteins; related to Figure 3. Diagrams of IUPred scores over the length of human CDK targets identified as primary components of MLOs in different studies, and their *Xenopus* homologues in this study. Regions with scores >0.5 (orange) are considered to be disordered, and <0.5 (grey) structured. Blue vertical lines indicate Ser and Thr residues; yellow circles, known Ser/Thr-Pro phosphosites (human) and non-dynamic phosphosites (*Xenopus*); green circles, confirmed CDK1 subfamily phosphorylations (human) and dynamic phosphorylations (*Xenopus*), from both embryos and egg extracts.

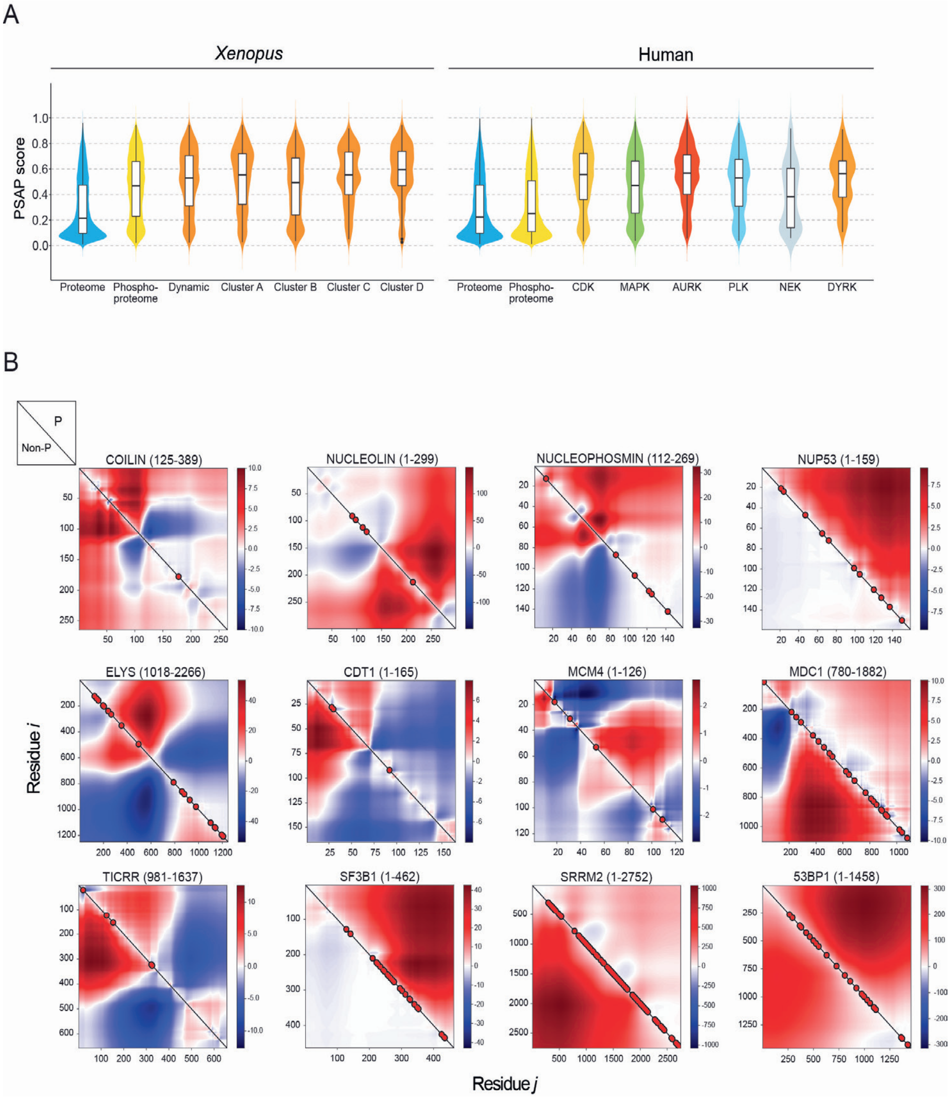


Figure S8. CDK-mediated phosphorylation regulates IDR phase separation propensity; related to Figure 4. (A) Violin plots presenting PSAP score for *Xenopus* dynamic phosphoproteins (left), and human kinase targets, in comparison with total proteome and phosphoproteome (right). (B) Sequence Charge Decoration Matrix (SCDM) maps for a selection of human CDK targets and major MLO components (IDRs analysed are indicated), depicting the contribution of electrostatic interaction dictating the distance between two amino acid residues i and j (shown in x and y axes). The values of SCDM for different residue pairs (i,j) are shown using colour schemes with red and blue denoting positive (repulsive) and negative (attractive) values, respectively. The lower and upper triangles indicate SCDM map for the unphosphorylated (non-P) and phosphorylated (P) sequences, respectively. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are indicated with red circles.

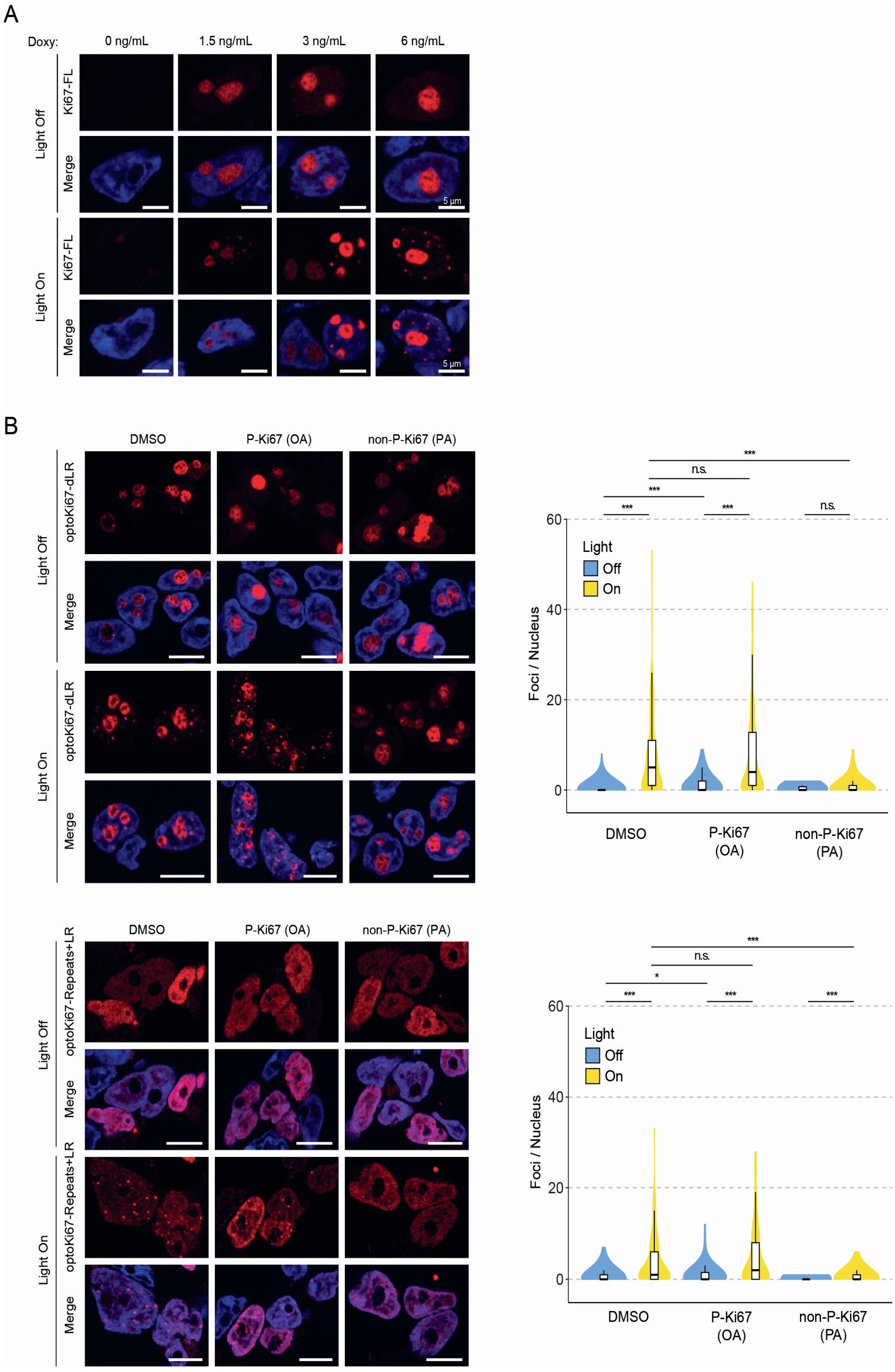

A
Doxy:
0 ng/mL
1.5 ng/mL
3 ng/mL
6 ng/mL
Light Off
Light On
Ki67-FL
Merge
5 µm
B
DMSO
P-Ki67 (OA)
non-P-Ki67 (PA)
optoKi67-dLR
Merge
Light Off
Light On
***
n.s.
***
***
***
n.s.
60
40
20
0
Light
Off
On
Foci / Nucleus
DMSO
P-Ki67 (OA)
non-P-Ki67 (PA)
optoKi67-Repeats+LR
***
n.s.
*
***
***
***

Figure S9. CDK-mediated phosphorylation regulates Ki-67 condensation; related to Figure 4. (A) Representative fluorescent images of HEK-293 cells expressing opto-Ki-67 (FL) construct, induced by the indicated concentrations of doxycycline (Doxy), before (Light Off) and after (Light On) exposure to blue light. DNA was stained with Hoechst 33258. (B) Left, representative fluorescent images of cells expressing opto-Ki-67 constructs (with deleted LR domain, dLR, top; or with deleted N-terminus, Repeats+LR, bottom), before (Light Off) and after (Light On) exposure to blue light. Cells were pretreated for 1h with either vehicle (DMSO), 0.5 µM okadaic acid (OA), to inhibit protein phosphatase 2, or 5 µM purvalanol A (PA), to inhibit CDKs. DNA was stained with Hoechst 33258; scale bars, 10 µm. Right, quantification of results; the number of foci per nucleus was counted. Statistical significance was assessed by one-way ANOVA on ranks (Kruskal–Wallis test) and pairwise *post-hoc* comparisons using the Mann–Whitney test. P-values were adjusted by the Benjamini-Hochberg method.

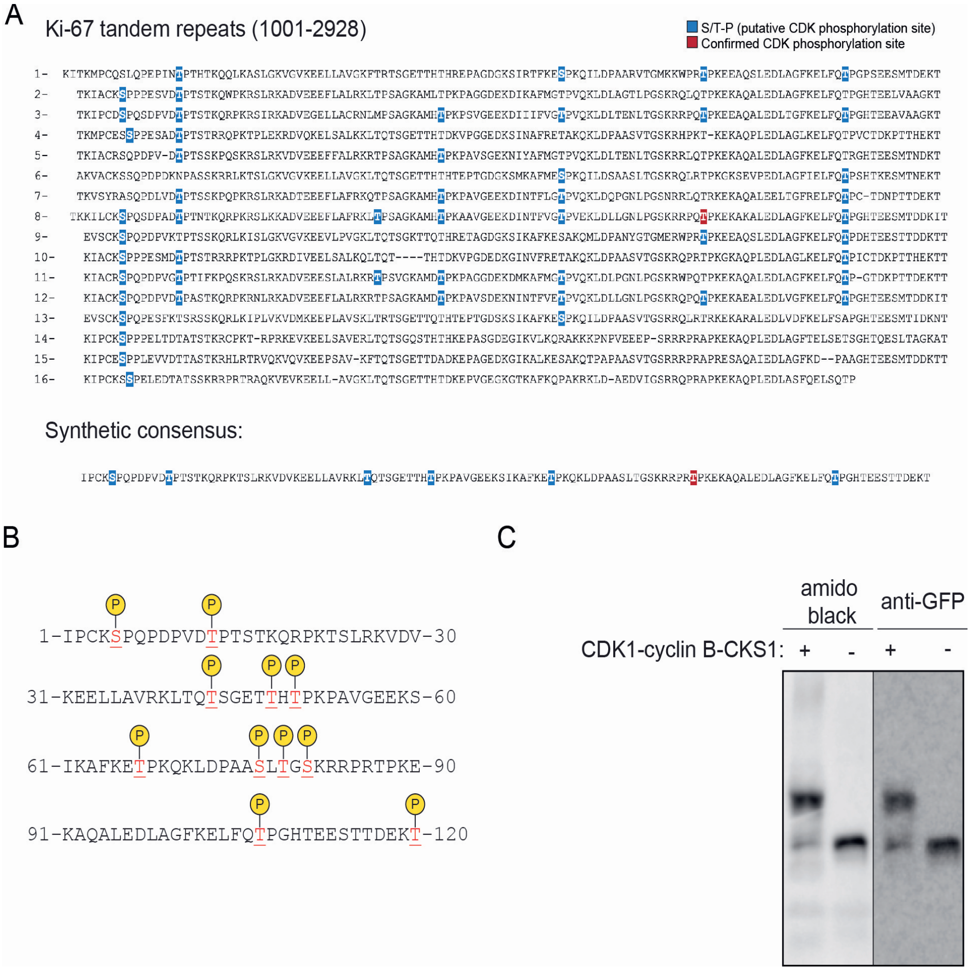


Figure S10. Ki-67 consensus repeat is stoichiometrically phosphorylated by CDK1 *in vitro*; related to Figure 4. (A) Alignment of human Ki-67 repeats, top, with the sequence of the consensus repeat depicted at the bottom. Confirmed and putative (Ser/Thr-Pro) CDK phosphorylation sites are highlighted in red and blue, respectively. (B, C) GFP-Ki-67 consensus repeat was phosphorylated *in vitro* using recombinant CDK1-cyclin B-CKS1 protein. Subsequently, the phosphosites were mapped by mass-spectrometry (B), and the stoichiometry of phosphorylation was analysed by Phos-Tag SDS-PAGE (C; amidoblack staining was used as loading control).

Acknowledgments

We thank Merlijn Witte for technical assistance with the *Xenopus laevis* egg fertilization experiments, Ariane Abrieu for a gift of CSF egg extracts, and Markus Raschle from the Technical University of Kaiserslautern for providing the *Xenopus laevis* protein database. Funding: AJRH and MA acknowledge support from the Horizon 2020 program INFRAIA project Epic-XS (Project 823839) and the NWO funded Netherlands Proteomics Centre through the National Road Map for Large-scale Infrastructures program X-Omics (Project 184.034.019) of the Netherlands Proteomics Centre. JMV is supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR). PK and MV are funded by the Oncode Institute which is financed by the Dutch Cancer Society and by the gravitation program CancerGenomiCs.nl from the Netherlands Organisation for Scientific Research (NWO). DF and LK are Inserm employees. GD is funded by the Institut National de Cancer, France (INCa) PRT-K programme (PRT-K17 n° 2018-023). The Fisher lab is funded by the Ligue Nationale Contre le Cancer, France (EL2018.LNCC/DF) and INCa (PLBIO18-094). The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05).

References

1. L. Liu *et al.*, G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells. *Nature Cell Biology* **19**, 177-188 (2017).
2. D. Santamaría *et al.*, Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* **448**, 811-815 (2007).
3. D. Coudreuse, P. Nurse, Driving the cell cycle with a minimal CDK control network. *Nature* **468**, 1074-1079 (2010).
4. D. L. Fisher, P. Nurse, A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *EMBO J* **15**, 850-860 (1996).
5. D. Fisher, L. Krasinska, D. Coudreuse, B. Novak, Phosphorylation network dynamics in the control of cell cycle transitions. *J Cell Sci* **125**, 4703-4711 (2012).
6. B. Novak, O. Kapuy, M. R. Domingo-Sananes, J. J. Tyson, Regulated protein kinases and phosphatases in cell cycle decisions. *Current Opinion in Cell Biology* **22**, 801-808 (2010).
7. J. J. Tyson, B. Novak, Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions. *Journal of Theoretical Biology* **210**, 249-263 (2001).
8. S. Rata *et al.*, Two Interlinked Bistable Switches Govern Mitotic Control in Mammalian Cells. *Current Biology* **28**, 3824-3832.e3826 (2018).
9. N. B. Trunnell, A. C. Poon, S. Y. Kim, J. E. Ferrell, Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Molecular Cell* **41**, 263-274 (2011).
10. S. Y. Kim, J. E. Ferrell, Jr., Substrate Competition as a Source of Ultrasensitivity in the Inactivation of Wee1. *Cell* **128**, 1133-1145 (2007).
11. J. R. Pomerening, E. D. Sontag, J. E. Ferrell, Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biology* **5**, 346-351 (2003).
12. W. Sha *et al.*, Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. *Proceedings of the National Academy of Sciences* **100**, 975-980 (2003).
13. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
14. A. A. Hyman, C. A. Weber, F. Julicher, Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39-58 (2014).
15. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
16. J. B. Woodruff, A. A. Hyman, E. Boke, Organization and Function of Non-dynamic Biomolecular Condensates. *Trends Biochem Sci* **43**, 81-94 (2018).
17. D. Berchtold, N. Battich, L. Pelkmans, A Systems-Level Study Reveals Regulators of Membrane-less Organelles in Human Cells. *Mol Cell* **72**, 1035-1049 e1035 (2018).
18. A. K. Rai, J. X. Chen, M. Selbach, L. Pelkmans, Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211-216 (2018).
19. W. Hur *et al.*, CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. *Developmental Cell* **54**, 379-394.e376 (2020).
20. G. Yahya *et al.*, Stress granules display bistable dynamics modulated by Cdk. *Journal of Cell Biology* **220**, e202005102 (2021).
21. L. M. Iakoucheva *et al.*, The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037-1049 (2004).
22. L. J. Holt *et al.*, Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682-1686 (2009).

23. W. Michowski *et al.*, Cdk1 Controls Global Epigenetic Landscape in Embryonic Stem Cells. *Molecular Cell* **78**, 459-476.e413 (2020).
24. A. M. Moses, J. K. Heriche, R. Durbin, Clustering of phosphorylation site recognition motifs can be exploited to predict the targets of cyclin-dependent kinase. *Genome Biol* **8**, R23 (2007).
25. J. D. Blethrow, J. S. Glavy, D. O. Morgan, K. M. Shokat, Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A* **105**, 1442-1447 (2008).
26. Y. Chi *et al.*, Identification of CDK2 substrates in human cell lysates. *Genome Biol* **9**, R149 (2008).
27. A. Errico, K. Deshmukh, Y. Tanaka, A. Pozniakovsky, T. Hunt, Identification of substrates for cyclin dependent kinases. *Adv Enzyme Regul* **50**, 375-399 (2010).
28. J. A. Ubersax *et al.*, Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859-864 (2003).
29. P. Hydbring, M. Malumbres, P. Sicinski, Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. *Nat Rev Mol Cell Biol* **17**, 280-292 (2016).
30. L. Krasinska, D. Fisher, Non-Cell Cycle Functions of the CDK Network in Ciliogenesis: Recycling the Cell Cycle Oscillator. *Bioessays* **40**, e1800016 (2018).
31. S. Lim, P. Kaldis, Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development* **140**, 3079-3093 (2013).
32. B. Budnik, E. Levy, G. Harmange, N. Slavov, SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**, 161 (2018).
33. C. Lombard-Banek, S. A. Moody, M. C. Manzini, P. Nemes, Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus laevis and Zebrafish Embryos. *Anal Chem* **91**, 4797-4805 (2019).
34. V. Olsen Jesper *et al.*, Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Science Signaling* **3**, ra3-ra3 (2010).
35. J. E. Purvis, G. Lahav, Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945-956 (2013).
36. S. Cooper, The synchronization manifesto: a critique of whole-culture synchronization. *The FEBS Journal* **286**, 4650-4656 (2019).
37. T. Ly, A. Endo, A. I. Lamond, Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**, e04534 (2015).
38. M. P. Swaffer, A. W. Jones, H. R. Flynn, A. P. Snijders, P. Nurse, CDK Substrate Phosphorylation and Ordering the Cell Cycle. *Cell* **167**, 1750-1761 e1716 (2016).
39. T. Ly *et al.*, A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *eLife* **3**, e01630 (2014).
40. T. Ly *et al.*, Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *Elife* **6**, (2017).
41. H. Post *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **16**, 728-737 (2017).
42. J. Newport, M. Kirschner, A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675-686 (1982).
43. J. W. Newport, M. W. Kirschner, Regulation of the cell cycle during early Xenopus development. *Cell* **37**, 731-742 (1984).
44. D. Clift, M. Schuh, Restarting life: fertilization and the transition from meiosis to mitosis. *Nat Rev Mol Cell Biol* **14**, 549-562 (2013).
45. M. Presler *et al.*, Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the Xenopus egg. *Proc Natl Acad Sci U S A* **114**, E10838-E10847 (2017).

46. E. H. Peuchen *et al.*, Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. *Sci Rep* **7**, 15647 (2017).
47. J. E. Ferrell, M. Wu, J. C. Gerhart, G. S. Martin, Cell cycle tyrosine phosphorylation of p34cdc2 and a microtubule-associated protein kinase homolog in Xenopus oocytes and eggs. *Molecular and Cellular Biology* **11**, 1965-1971 (1991).
48. K. Suzuki *et al.*, Identification of non-Ser/Thr-Pro consensus motifs for Cdk1 and their roles in mitotic regulation of C2H2 zinc finger proteins and Ect2. *Scientific Reports* **5**, 7929 (2015).
49. T. Schmidlin *et al.*, High-Throughput Assessment of Kinome-wide Activation States. *Cell Syst* **9**, 366-374 e365 (2019).
50. T. Schmidlin *et al.*, Assessment of SRM, MRM3, and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *PROTEOMICS* **16**, 2193-2205 (2016).
51. R. T. Lawrence, B. C. Searle, A. Llovet, J. Villén, Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nature Methods* **13**, 431-434 (2016).
52. A. C. Peterson, J. D. Russell, D. J. Bailey, M. S. Westphall, J. J. Coon, Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics *. *Molecular & Cellular Proteomics* **11**, 1475-1488 (2012).
53. T. Y. Tsai, J. A. Theriot, J. E. Ferrell, Jr., Changes in oscillatory dynamics in the cell cycle of early Xenopus laevis embryos. *PLoS Biol* **12**, e1001788 (2014).
54. F. Bassermann *et al.*, Multisite phosphorylation of nuclear interaction partner of ALK (NIPA) at G2/M involves cyclin B1/Cdk1. *J Biol Chem* **282**, 15965-15972 (2007).
55. C. Kraft *et al.*, Mitotic regulation of the human anaphase-promoting complex by phosphorylation. *The EMBO Journal* **22**, 6598-6609 (2003).
56. S. Vigneron *et al.*, Greatwall maintains mitosis through regulation of PP2A. *The EMBO Journal* **28**, 2786-2793 (2009).
57. A. Gharbi-Ayachi *et al.*, The Substrate of Greatwall Kinase, Arpp19, Controls Mitosis by Inhibiting Protein Phosphatase 2A. *Science* **330**, 1673-1677 (2010).
58. S. Mochida, L. Maslen Sarah, M. Skehel, T. Hunt, Greatwall Phosphorylates an Inhibitor of Protein Phosphatase 2A That Is Essential for Mitosis. *Science* **330**, 1670-1673 (2010).
59. L. Krasinska *et al.*, Protein Phosphatase 2A Controls the Order and Dynamics of Cell-Cycle Transitions. *Molecular Cell* **44**, 437-450 (2011).
60. A. Goldbeter, D. E. Koshland, An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences* **78**, 6840-6844 (1981).
61. S. F. Shimobayashi, P. Ronceray, D. W. Sanders, M. P. Haataja, C. P. Brangwynne, Nucleation landscape of biomolecular condensates. *Nature* **599**, 503-506 (2021).
62. Y.-H. Lin, H. Wu, B. Jia, M. Zhang, H. S. Chan, Assembly of model postsynaptic densities involves interactions auxiliary to stoichiometric binding. *Biophysical Journal* **121**, 157-171 (2022).
63. K. Ghosh, Stoichiometric versus stochastic interaction in models of liquid-liquid phase separation. *Biophysical Journal* **121**, 4-6 (2022).
64. G. van Mierlo *et al.*, Predicting protein condensate formation using machine learning. *Cell Reports* **34**, 108705 (2021).
65. J. Huihui, K. Ghosh, An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *The Journal of Chemical Physics* **152**, 161102 (2020).
66. J. Huihui, K. Ghosh, Intrachain interaction topology can identify functionally similar intrinsically disordered proteins. *Biophysical Journal* **120**, 1860-1868 (2021).

67. M. Sobecki *et al.*, The cell proliferation antigen Ki-67 organises heterochromatin. *eLife* **5**, e13722 (2016).
68. D. G. Booth *et al.*, Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *eLife* **3**, e01641 (2014).
69. Y. Hayashi, K. Kato, K. Kimura, The hierarchical structure of the perichromosomal layer comprises Ki67, ribosomal RNAs, and nucleolar proteins. *Biochemical and Biophysical Research Communications* **493**, 1043-1049 (2017).
70. N. Hégarat *et al.*, Cyclin A triggers Mitosis either via the Greatwall kinase pathway or Cyclin B. *The EMBO Journal* **39**, e104419 (2020).
71. Y. Shin *et al.*, Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171.e114 (2017).
72. Z. Feng, B. Jia, M. Zhang, Liquid–Liquid Phase Separation in Biology: Specific Stoichiometric Molecular Interactions vs Promiscuous Interactions Mediated by Disordered Sequences. *Biochemistry* **60**, 2397-2406 (2021).
73. A. Musacchio, On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal* **41**, e109952 (2022).
74. N. Parisis *et al.*, Initiation of DNA replication requires actin dynamics and formin activity. *The EMBO Journal* **36**, 3212-3231 (2017).
75. R. G. H. Lindeboom, A. H. Smits, M. Perino, G. J. C. Veenstra, M. Vermeulen, Mass Spectrometry-Based Absolute Quantification of Single Xenopus Embryo Proteomes. *Cold Spring Harb Protoc*, (2018).
76. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
77. H. D. Meiring, E. van der Heeft, G. J. ten Hove, A. P. J. M. de Jong, Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* **25**, 557-568 (2002).
78. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
79. T. Temu, M. Mann, M. Raschle, J. Cox, Homology-driven assembly of NOn-redundant protEin sequence sets (NOmESS) for mass spectrometry. *Bioinformatics* **32**, 1417-1419 (2016).
80. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
81. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
82. D. Szklarczyk *et al.*, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
83. G. Su, A. Kuchinsky, J. H. Morris, D. J. States, F. Meng, GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135-3137 (2010).
84. S. Maere, K. Heymans, M. Kuiper, BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448-3449 (2005).
85. F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
86. B. MacLean *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).
87. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520 (2015).

88. A. Orthwein *et al.*, Mitosis Inhibits DNA Double-Strand Break Repair to Guard Against Telomere Fusions. *Science* **344**, 189-193 (2014).

89. Haley D. M. Wyatt, S. Sarbajna, J. Matos, Stephen C. West, Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Molecular Cell* **52**, 234-247 (2013).

90. M. I. Linder *et al.*, Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. *Dev Cell* **43**, 141-156 e147 (2017).

91. J. Liu *et al.*, Cell cycle-dependent localization of the CDK2-cyclin E complex in Cajal (coiled) bodies. *Journal of Cell Science* **113**, 1543-1552 (2000).

92. Y. Chi *et al.*, A novel landscape of nuclear human CDK2 substrates revealed by in situ phosphorylation. *Science Advances* **6**, eaaz9899.

93. U. R. Klein, M. Haindl, E. A. Nigg, S. Muller, RanBP2 and SENP3 Function in a Mitotic SUMO2/3 Conjugation-Deconjugation Cycle on Borealin. *Molecular Biology of the Cell* **20**, 410-418 (2008).

94. H. Goto *et al.*, Complex formation of Plk1 and INCENP required for metaphase–anaphase transition. *Nature Cell Biology* **8**, 180-187 (2006).

95. O. Bartsch, S. Horstmann, K. Toprak, K.-H. Klempnauer, S. Ferrari, Identification of cyclin A/Cdk2 phosphorylation sites in B-Myb. *European Journal of Biochemistry* **260**, 384-391 (1999).

96. M. Curtis, S. N. Nikolopoulos, C. E. Turner, Actopaxin is phosphorylated during mitosis and is a substrate for cyclin B1/cdc2 kinase. *Biochemical Journal* **363**, 233-242 (2002).

97. A. Fourest-Lieuvin *et al.*, Microtubule Regulation in Mitosis: Tubulin Phosphorylation by the Cyclin-dependent Kinase Cdk1. *Molecular Biology of the Cell* **17**, 1041-1050 (2005).

98. R. E. Milner, J. L. Busaan, C. F. Holmes, J. H. Wang, M. Michalak, Phosphorylation of dystrophin. The carboxyl-terminal region of dystrophin is a substrate for in vitro phosphorylation by p34cdc2 protein kinase. *Journal of Biological Chemistry* **268**, 21901-21905 (1993).

99. M. Lowe *et al.*, Cdc2 Kinase Directly Phosphorylates the cis-Golgi Matrix Protein GM130 and Is Required for Golgi Fragmentation in Mitosis. *Cell* **94**, 783-793 (1998).

100. J. Yun *et al.*, Cdk2-dependent Phosphorylation of the NF-Y Transcription Factor and Its Involvement in the p53-p21 Signaling Pathway*. *Journal of Biological Chemistry* **278**, 36966-36972 (2003).

101. M. Kitzmann *et al.*, cdk1- and cdk2-Mediated Phosphorylation of MyoD Ser200 in Growing C2 Myoblasts: Role in Modulating MyoD Half-Life and Myogenic Activity. *Molecular and Cellular Biology* **19**, 3167-3176 (1999).

102. D. A. Thiel *et al.*, Cell Cycle-Regulated Phosphorylation of p21-Activated Kinase 1. *Current Biology* **12**, 1227-1232 (2002).

103. M. Li, B. Stefansson, W. Wang, E. M. Schaefer, D. L. Brautigan, Phosphorylation of the Pro-X-Thr-Pro site in phosphatase inhibitor-2 by cyclin-dependent protein kinase during M-phase of the cell cycle. *Cellular Signalling* **18**, 1318-1326 (2006).

104. K.-w. Fong *et al.*, Whole-genome screening identifies proteins localized to distinct nuclear bodies. *Journal of Cell Biology* **203**, 149-164 (2013).

105. S. Jain *et al.*, ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487-498 (2016).

106. J. Dopie, M. J. Sweredoski, A. Moradian, A. S. Belmont, Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *Journal of Cell Biology* **219**, e201910207 (2020).

107. J. Liu *et al.*, Functional proteomic analysis of promyelocytic leukaemia nuclear bodies in irradiation-induced MCF-7 cells. *The Journal of Biochemistry* **148**, 659-667 (2010).

108. A. Hubstenberger *et al.*, P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell* **68**, 144-157.e145 (2017).

109. L. Stenström *et al.*, Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. *Molecular Systems Biology* **16**, e9469 (2020).

110. L. Tafforeau *et al.*, The Complexity of Human Ribosome Biogenesis Revealed by Systematic Nucleolar Screening of Pre-rRNA Processing Factors. *Molecular Cell* **51**, 539-551 (2013).

111. D. H. Lin, A. Hoelz, The Structure of the Nuclear Pore Complex (An Update). *Annual Review of Biochemistry* **88**, 725-783 (2019).

112. M. Machyna, P. Heyn, K. M. Neugebauer, Cajal bodies: where form meets function. *WIREs RNA* **4**, 17-34 (2013).

113. M. Quevedo *et al.*, Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nature Communications* **10**, 2669 (2019).

114. D. Piovesan *et al.*, MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Research* **49**, D361-D367 (2021).

115. R. C. Gentleman *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80 (2004).

116. J. Douglass *et al.*, Identifying protein kinase target preferences using mass spectrometry. *Am J Physiol Cell Physiol* **303**, C715-727 (2012).

117. A. Vitalis, R. V. Pappu, in *Annual Reports in Computational Chemistry,* R. A. Wheeler, Ed. (Elsevier, 2009), vol. 5, pp. 49-76.

118. A. Vitalis, R. V. Pappu, ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry* **30**, 673-699 (2009).

119. G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Computational Biology* **14**, e1005941 (2018).

120. R. M. Regy, J. Thompson, Y. C. Kim, J. Mittal, Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **30**, 1371-1379 (2021).

121. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

122. H. J. Feldman, C. W. Hogue, A fast method to sample real protein conformational space. *Proteins* **39**, 112-131 (2000).

123. H. J. Feldman, C. W. Hogue, Probabilistic sampling of protein conformations: new hope for brute force? *Proteins* **46**, 8-23 (2002).

124. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612 (2004).

125. D. A. Case *et al.*, The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, 1668-1688 (2005).

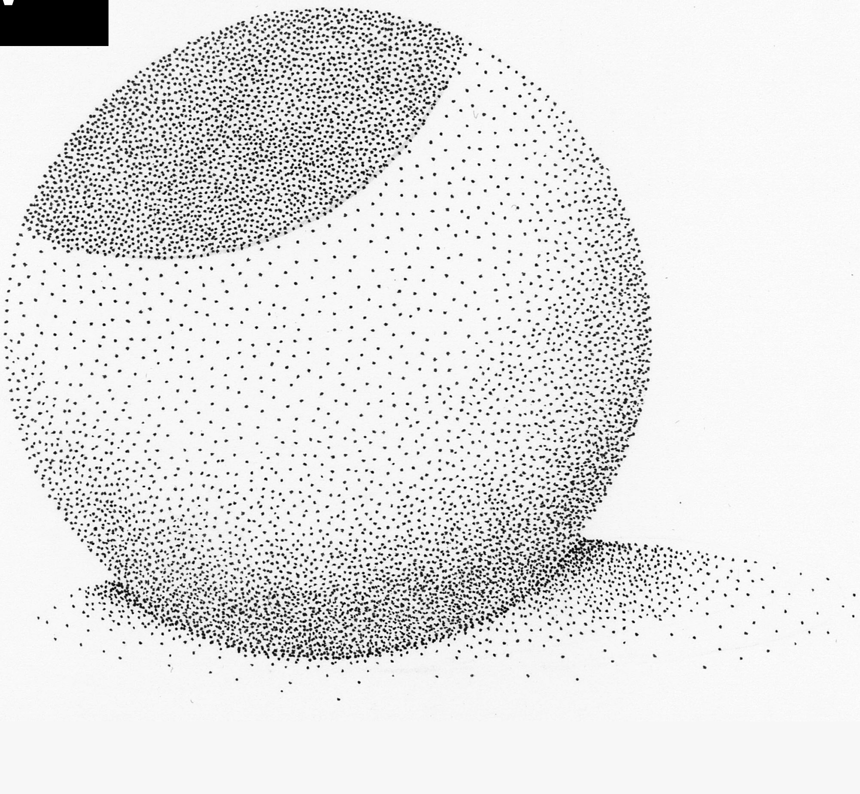
126. L. Dignon Gregory, W. Zheng, B. Best Robert, C. Kim Young, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **115**, 9929-9934 (2018).

127. F. W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234 (2005).

128. W. F. Vranken *et al.*, The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics* **59**, 687-696 (2005).

129. J. L. Markley *et al.*, Recommendations for the presentation of NMR structures of proteins and nucleic acids. *European Journal of Biochemistry* **256**, 1-15 (1998).

CHAPTER V



Combined quantitative (phospho) proteomics and mass spectrometry imaging reveal temporal and spatial protein changes in human intestinal ischemia-reperfusion

Anna M. Kip[1], Juan Manuel Valverde[3], Maarten Altelaar[3], Ron M.A. Heeren[2], Inca H.R. Hundscheid[1], Cornelis H.C. Dejong[1,4], Steven W.M. Olde Damink[1,4], Benjamin Balluff[2], Kaatje Lenaerts[1]

[1]Department of Surgery, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre, Maastricht, The Netherlands.
[2]Maastricht Multimodal Molecular Imaging institute (M4i), Maastricht University, Maastricht, The Netherlands
[3]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, Utrecht 3584 CH, Netherlands.
[4]Department of General, Visceral- and Transplantation Surgery, RWTH Aachen University Hospital, Aachen, Germany

Published in Journal of Proteome Research 2022, Volume 21, Issue 1, pages 49-66

Abstract

Intestinal ischemia-reperfusion (IR) injury is a severe clinical condition, and unraveling its pathophysiology is crucial in order to improve therapeutic strategies and reduce the high morbidity and mortality rates. Here, we studied the dynamic proteome and phosphoproteome in the human intestine during ischemia and reperfusion, using LC-MS/MS analysis to gain quantitative information of thousands of proteins and phosphorylation sites, as well as MS imaging (MSI) to obtain spatial information. We identified a significant decrease in abundance of proteins related to intestinal absorption, microvillus and cell junction, whereas proteins involved in innate immunity, in particular the complement cascade, and extracellular matrix organization increased in abundance after IR. Differentially phosphorylated proteins were involved in RNA splicing events and cytoskeletal and cell junction organization. In addition, our analysis points to MAPK and CDK families to be active kinases during IR. Finally, MALDI-TOF MSI presented peptide alterations in abundance and distribution, which resulted, in combination with FTICR-MSI and LC-MS/MS, in the annotation of proteins related to RNA splicing, the complement cascade and extracellular matrix organization. This study expanded our understanding of the molecular changes that occur during IR in human intestine and highlights the value of the complementary use of different MS-based methodologies.

Introduction

Intestinal ischemia-reperfusion (IR) is a clinical phenomenon carrying high morbidity and mortality, and can occur in various conditions. Based on etiology, intestinal ischemia is classified into chronic ischemia – *e.g.* due to atherosclerosis – and acute ischemia. The latter is further divided into occlusive disease, caused by obstruction of the mesenteric blood flow, or non-occlusive disease, caused by hypoperfusion for example due to major surgical procedures, trauma, hemorrhagic shock or sepsis (*1-3*). The lack of oxygen during ischemia obviously leads to cell injury, and hence rapid reperfusion is crucial. However, reperfusion can also aggravate injury as the sudden oxygen supply to the ischemic intestine initiates a cascade of events, including Ca^{2+} influx and the production of reactive oxygen species (ROS) which damages cellular structures and activates an inflammatory response (*4*). The intestinal epithelium serves as an important barrier that protects the body from the hostile luminal environment. Disruption of this barrier, which can be caused by IR injury, allows entry of harmful luminal microorganisms and toxins into the sterile inner mucosa, which may cause a severe inflammatory response. Observations from a human experimental model of intestinal IR showed that the intestine was relatively resistant to short periods of ischemia (*5, 6*), whereas prolonged ischemia (>45 min) followed by reperfusion disrupted the epithelial lining and induced inflammation (*7*). Cell death occurs initially at the villi tips and progresses towards the crypt with increasing duration of the ischemic period (*7, 8*). Severe IR can eventually lead to bowel necrosis and severe systemic inflammation. The high mortality rates (60-80%) of acute intestinal ischemia (*1-3*) can be attributed to the difficulty to diagnose acute mesenteric ischemia at an early stage, as well as the lack of effective therapeutic options (*9-11*). Further unraveling the molecular mechanisms underlying ischemia-reperfusion is crucial in order to improve therapeutic strategies and patient outcome.

Identification of changes in (local) protein abundance that occur following ischemia and during reperfusion in the human experimental model is of great interest for understanding the biological processes involved in IR injury and tissue repair. As many cellular functions are regulated by the dynamic phosphorylation of proteins, we were interested in investigating changes in protein phosphorylation as well. Furthermore, given the aforementioned differences in function and susceptibility to IR across the layers and cell types of the intestinal wall, spatial information of protein changes is of particular interest in the context of intestinal IR injury.

Liquid chromatography coupled to mass spectrometry (LC-MS)-based proteomics has emerged as an important tool to study cell biology and disease mechanisms, and enables untargeted identification and quantification of several thousand proteins (*12*). A multi-step process, including protein extraction, digestion and separation by LC, is currently used to obtain this thorough protein coverage and accuracy. However, with homogenization of the tissue during sample preparation, spatial information on protein distribution gets lost. A powerful technology to complement LC-MS-based proteomics is matrix-assisted laser desorption/ionization (MALDI) MS imaging (MSI) of tissue sections. MALDI MSI allows analyzing hundreds of molecules simultaneously, providing spatial distribution and local abundance of these molecules in biological tissues in a label-free manner. Combining MALDI MSI with the histological information of these tissue enables a histology-driven analysis. Intact proteins can be detected by MSI, however, their detection is limited by low ionization efficiency that limits sensitivity in the high mass range of the instrumentation. Recent studies have shown imaging of proteins up to 200 kDa (*13, 14*). MSI of trypsinized proteins can therefore – at least theoretically – increase the coverage of the proteome beyond this sensitivity limit. Moreover, the on-tissue digestion approach allows a better integration of MSI data with LC-MS/MS of the same or an adjacent tissue section to enable the identification of the observed local peptide signals, especially when using high-mass resolution MSI instrumentation (*15, 16*). However, these high-mass resolution MSI methods are limited by long acquisition time and consequently suffer from a low throughput. This issue can be overcome by the use of high-speed MALDI-Time-Of-Flight (TOF) MSI complemented by high-mass resolution MSI data (*17*), which is the approach that we applied in our study.

The main objective of this study is to analyze the dynamic proteome and phosphoproteome in human intestine exposed to ischemia and reperfusion. To this end, we used two complementary MS-based technologies: LC-MS/MS (phospho)proteomics to gain in-depth quantitative information, and MALDI MSI of tryptic peptides to obtain spatial information and study location-specific protein changes.

Results

Changes in protein expression and protein phosphorylation during IR of the human intestine were studied by combining a quantitative MS-based (phospho)proteomics approach with imaging MS. The dynamic (phospho)proteome during IR was investigated by analyzing tissue samples collected after 45 min of ischemia (45I), 30 and 120 min of reperfusion (30R and 120R) and in control tissue (Ctrl). A schematic of the experimental design and data analysis workflow are depicted in Figure 1.

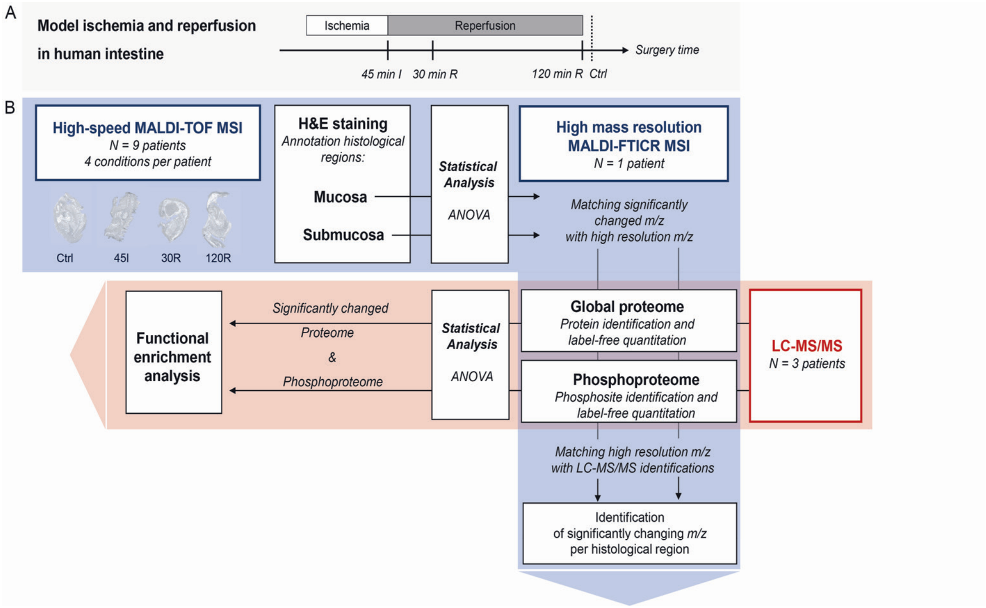


Figure 1. Experimental design and data analysis workflow. (A) Experimental model of ischemia-reperfusion in human intestine, with tissue collection after 45 min of ischemia (45I), 30 min of reperfusion (30R) and 120 min of reperfusion (120R), and tissue not exposed to ischemia-reperfusion (Ctrl). (B) Mass spectrometry and data analysis workflow. The red area shows the workflow of LC-MS/MS measurement and subsequent data analysis. The blue area shows the MS imaging (MSI) measurements.

Proteomics

LC-MS/MS analysis resulted in the identification and quantification of 2,562 proteins. The correlation between biological replicates was high (Pearson r >0.9), with the exception of one 30R sample (r <0.85), which was therefore excluded from further analysis (Figure S1). Cluster analysis of the complete proteome showed that the two main clusters were represented by 45I and Ctrl samples on the one hand, and 30R and 120R samples on the other hand (Figure S2). Analysis of dynamic changes in the proteome revealed that the abundance of 239 proteins was significantly altered during IR. Hierarchical clustering of these differentially expressed proteins resulted in four main clusters reflecting distinct temporal expression profiles (Figure 2A) and clearly distinguished the proteins decreasing (cluster 1 and 2), and proteins increasing in abundance (cluster 3 and 4) during reperfusion.

Functional analysis of the dynamic proteome

To get global functional insight in the changing proteome during IR, we performed GO term enrichment analysis of differentially expressed proteins. These proteins were predominantly involved in processes such as intestinal absorption and digestion, cell junction organization, and innate immune responses (Figure 2B). In line with this observation, overrepresented Cellular Component GO terms included brush border and microvillus, actin filament and cell-cell contact zone. Proteins in cluster 1 of the changing proteome, showing a decrease in abundance at 30R which restored at 120R, were related to various processes such as protein translation, including EIF2S2, EIF4EBP1, EIF5B, and the cellular response to stress, such as HSPA4, TNIK, Nup50, LAMTOR1.

Protein abundances in cluster 2 decreased during reperfusion (Figure 2C) and were significantly enriched for GO terms related to intestinal digestion/absorption, microvillus and cell junction. Moreover, network analysis of the total changing proteome showed interactions of proteins involved in these GO terms. The majority of interacting proteins in these networks exhibited the same temporal profile (cluster 2; Figure 2D, blue fill). Proteins involved in digestion and absorption included IFABP, LCT, ANPEP, NAALADL1, and SLC5A1 (Figure 2D, left). Proteins associated with microvillus organization included actin bundling proteins such as VIL1, ESPN, and PLS1, motor protein MYO1A, anchoring protein EZR, and microvillus-microvillus adhesion molecules CDHR2 and MYO7B. Other proteins playing a role in actin filament organization were COBL and MAP7. In addition, various proteins playing an important role in cell-cell junction organization were decreased such as CDH1, CDH17, CDHR5, CDHR2, AFDN, NECTIN-1, F11R, EpCAM and CD2AP and CGN (Figure 2D, right; Table S1).

Cluster 3 showed an opposite expression profile with increased protein levels during reperfusion. Overrepresented GO terms in this cluster were predominantly associated with the innate immune response, in particular the complement pathway (Figure 2C). Both regulatory and effector proteins of the complement cascade were significantly increased at reperfusion and included C3, C5, C6, C8B, PROS1 and F2, which were all among the top 20 proteins showing the highest fold change amongst conditions, and exhibited very similar temporal profiles. In addition, various proteins in cluster 3 were involved in extracellular matrix organization, including COL15A1, COL18A1, FBN1, NID2, ITGA and JAM3. One of the protein interaction networks, resulting from network analysis of the complete changing proteome, contained mostly proteins of cluster 3 and showed interactions of complement proteins and proteins related to extracellular matrix organization (Figure 2E, red fill).

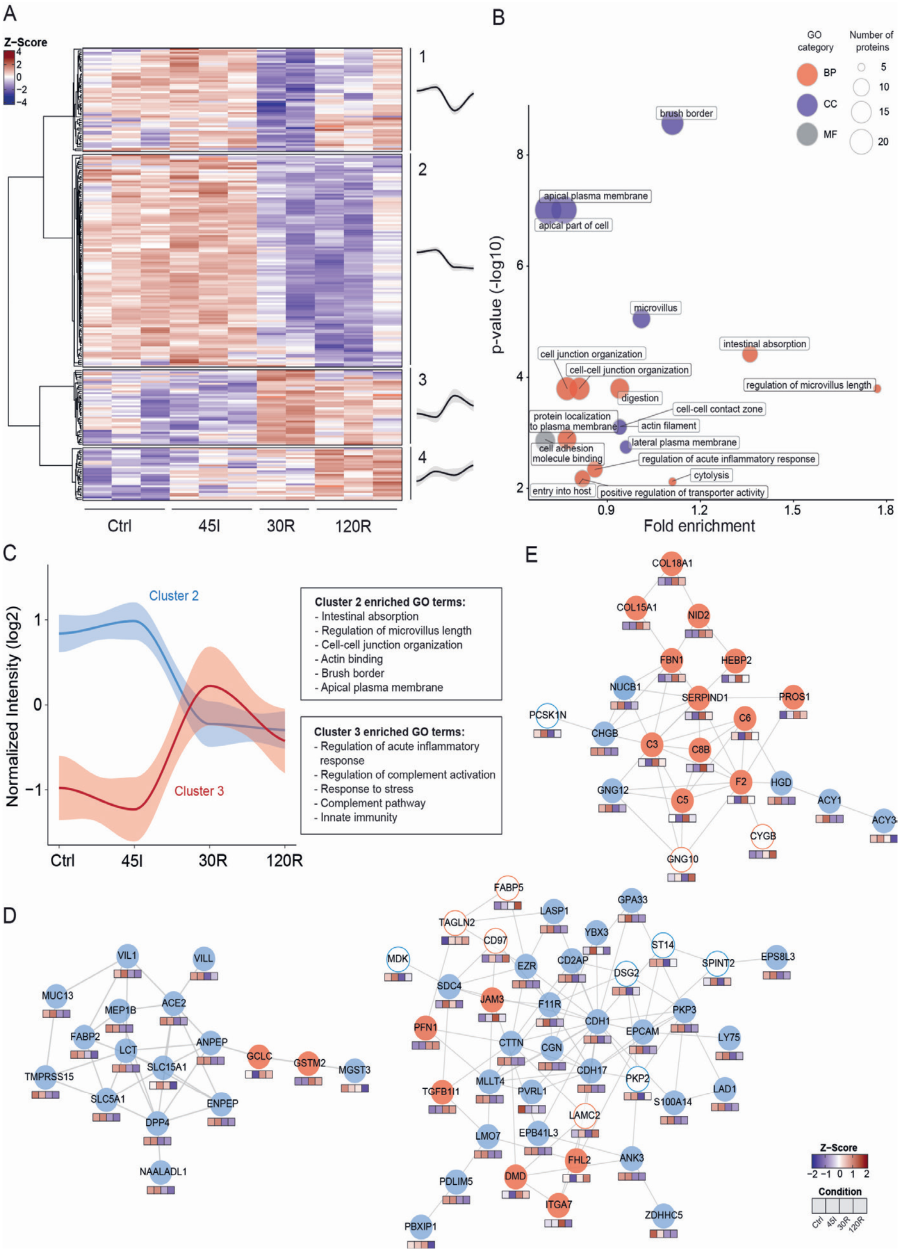


Figure 2. The dynamic proteome during ischemia-reperfusion in human intestine. (A) Heatmap visualizing clustering of differentially expressed proteins. Hierarchical clustering was based on Z scores of the log2 values of differentially expressed proteins. ANOVA test was used and P<0.05 was considered statistically significant. Average temporal profiles are shown for every cluster (grey area represents CI). (B) Functional enrichment analysis of differentially expressed proteins. Overrepresented Gene Ontology (GO) terms are shown (P<0.05). **Legend continues on next page.**

Figure 2 legend continued. GO term fold enrichment (log10[observed/expected]) is plotted against p-value (-log10). The size of the dot correlates with the number of proteins linked that GO term, as indicated in the legend. Red dot, GO Biological Processes; Blue dot, GO Cellular component, Grey dot, GO Molecular Function. (C) The dynamic profile for proteins in cluster 2 (blue) and 3 (red) is shown, as well as corresponding overrepresented GO terms. (D, E) Protein networks showing interacting proteins. Networks were generated using STRING. Color of the circle indicates the temporal profile (cluster) of the protein. Blue line, cluster 1; Blue fill, cluster 2; Red fill, cluster 3, Red line, cluster 4. Color cubes below each protein indicates Z-score intensity in Ctrl, 45I, 30R and 120 R respectively. The majority of proteins in the networks in panel D are located in cluster 2 (blue fill). The majority of proteins in the network in panel E are located in cluster 3 (red fill).

Proteins in cluster 4 gradually increased in abundance during ischemia and reperfusion and were involved in a variety of biological processes without a clear overrepresentation, and included the metabolism of amino acids, metabolism of nucleotides, post-translational protein modification and cellular response to stress. Interesting proteins in this cluster include CYGB, SQSTM1 and CASP1.

In addition to the protein interaction networks that could be linked to cluster 2 and 3, another network showed interactions of proteins located in all four clusters and thus showing distinct temporal profiles (Figure S3). These proteins were associated with the cellular response to stress, protein and RNA metabolism.

Phosphoproteomics

LC-MS/MS analysis resulted in identification and quantification of 1,802 phosphosites derived from 1,214 proteins (whole phosphoproteome). The observed distribution of phosphosites was 90% phosposerine, 9.8% phosphothreonine, and 0.2% phosphotyrosine (Figure S4A). The majority of these phosphosites had a high localization probability score, indicative of accurate localization of the phosphorylated residue in the peptide backbone (Figure S4B). In total 305 phosphosites on 162 proteins showed a significant change during IR (dynamic phosphoproteome).

Prediction of the kinases responsible for detected phosphorylation

A kinase enrichment analysis for all detected phosphosites was performed to get an overview which kinases were potentially active during IR. Here, 279 detected phosphosites were mapped to putative effector kinases (Figure S4C). Some of these kinases, such as GSK3B and Casein kinase II, are constitutively active and participate in a myriad of cellular processes. We also predicted activity of cyclin dependent kinases, namely CDK1 and CDK2, both involved in cell cycle control. Finally, we detected putative targets of the MAP kinases MAPK9 (JNK2) and MAPK10 (JNK3), part of the JNK signaling pathway, and MAPK14 (p38α), part of the p38 MAP kinase pathway, which are both activated in response to cellular stress.

Kinase enrichment analysis that focused on the 305 significantly regulated phosphosites, mapped 47 phosphosites to potential effector kinases (Figure S4D). Notably present were putative substrates of several MAP kinases, like MAPK9 (JNK2), MAPK3 (ERK1), MAPK13 (p38δ) and MAPK8 (JNK1). Dynamic phosphosites linked to Casein kinase II and CDKs were detected as well.

Clustering and motif enrichment of the dynamic phosphoproteome

Hierarchical clustering of altered phosphosites resulted in six distinct groups and revealed a highly dynamic regulation of protein phosphorylation during the course of ischemia and reperfusion (Figure 3A). In contrast to the global proteome, phosphorylation changes occur rapidly, already following ischemia. Kinases recognize their substrate partly through certain sequence motifs near the phosphorylation site, and some of these motifs are associated with specific kinases. Motif analysis revealed differences in its composition amongst clusters (Figure S5). First, we observed that proline-directed phosphorylation comprised almost half of the significantly changed phosphosites (144 out of 305). These were spread across all six clusters. Amongst the non-proline-directed motifs, the clusters 1, 2, 5 and 6 showed predominantly acidic motifs, characterized by the presence of aspartic and glutamic acid. Cluster 3 showed basic residues upstream of the phosphorylation site, and cluster 4 was mostly comprised of proline directed phosphosites. These results show how the activity of different kinases changes during the different stages of IR injury.

Functional analysis of the proteins with significantly regulated phosphorylation

In general, differentially phosphorylated proteins were involved in the regulation of mRNA processing and RNA splicing, supramolecular fiber organization/cytoskeleton and cell junction organization (Figure 3B). Related molecular functions – *e.g.* RNA polymerase binding and actin binding - and cellular components – *e.g.* spliceosomal complex, actin cytoskeleton and adherence junction - were overrepresented as well. In contrast to the changing proteome, the GO analysis per cluster revealed that differentially phosphorylated proteins related to the same biological process appeared in different clusters. Moreover, changing phosphosites from the same protein were represented in different clusters, for instance phosphosites on MISP (cluster 2, 3, 4, 5) or SRRM2 (cluster 1, 3, 5).

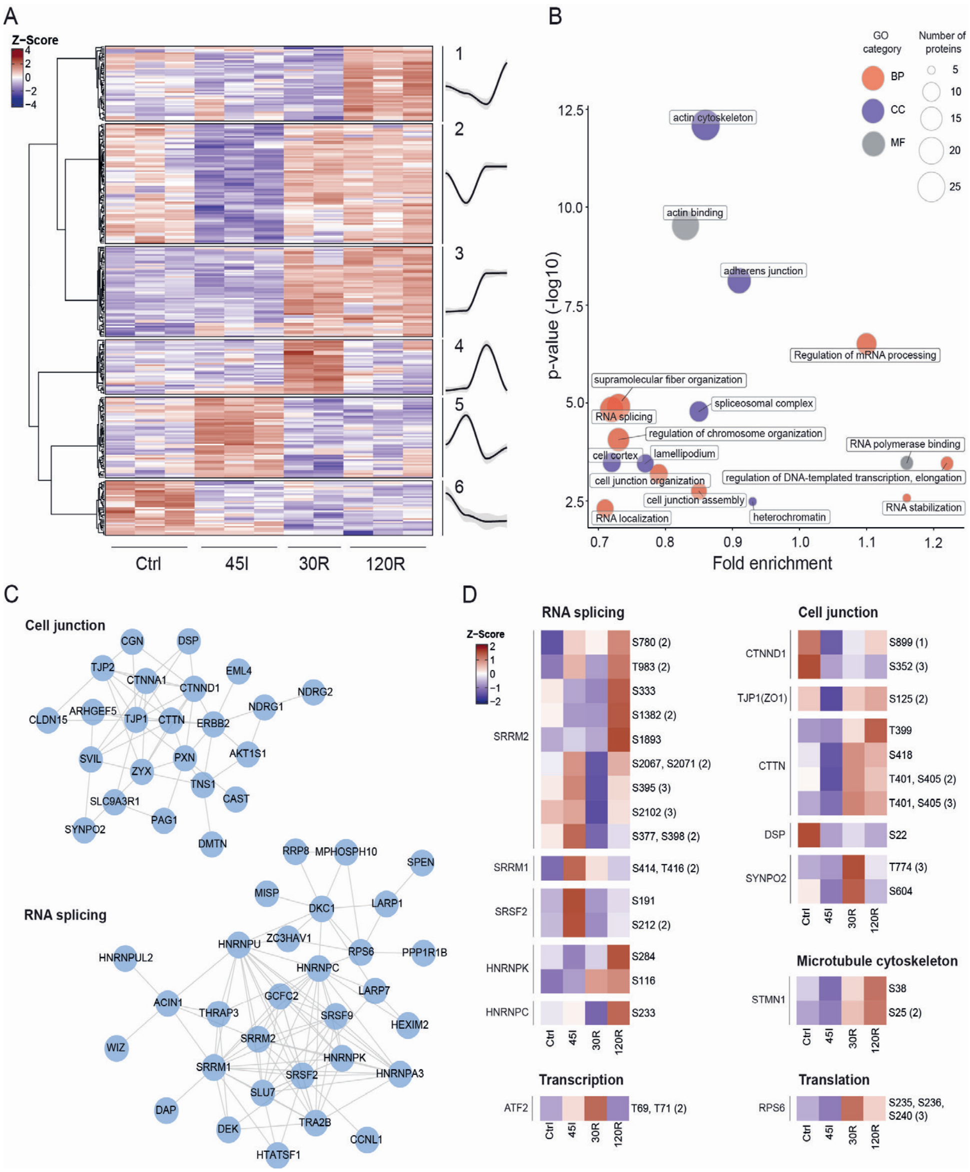


Figure 3. The dynamic phosphoproteome during ischemia-reperfusion in human intestine. (A) Heatmap visualizing clustering of changing phosphosites, and average temporal profiles for every cluster (grey area represents CI). Hierarchical clustering was based on Z scores of the log2 values of differentially expressed proteins. ANOVA test was used and P<0.05 was considered statistically significant. (B) Functional enrichment analysis of differentially phosphorylated proteins. Overrepresented Gene Ontology (GO) terms are shown (P<0.05). The size of the dot correlates with the number of proteins linked that GO term, as indicated in the legend. Red, GO Biological Processes; Blue, GO Cellular component, Grey, GO Molecular Function. **Legend continues on next page.**

Figure 3 legend continued. (C) Networks of highly interconnected phosphoproteins related to overrepresented terms of Cell junction and RNA splicing. (D) Heat maps of single phosphosites related to different biological processes. Showing the Z scores of the averaged log2 intensities for each condition.

Next, similar to the proteome data, we explored the connectivity and association of dynamically phosphorylated proteins. The network analysis showed that many of the phosphorylated proteins were related to either cell junctions, or RNA splicing, form discrete networks of highly interconnected proteins (Figure 3C). This highlights that proteins related to these two biological processes are highly regulated by phosphorylation during IR.

Altered phosphorylation of RNA splicing factors was observed for various serine/arginine proteins, including SRRM2, SRRM1 and SRSF2 (Figure 3D). In addition, several heterogenous nuclear ribonucleoproteins were found to show changes in phosphorylation and included HNRNPK and HNRNPC (Figure 3D). Differentially phosphorylated proteins related to cell junction organization included cadherin-associated CTNND1, scaffolding proteins such as ZO-1 (TJP1), and other proteins contributing to cell adhesion and related cytoskeleton organization including CTTN, DSP and SYNPO2 (Figure 3D). Phosphorylated proteins associated with supramolecular fiber organization included STMN1 and ESPN.

Exploration of specific phosphosites with known functionality

After focusing on the proteins that displayed dynamic phosphorylation during IR, we next explored our data for phosphosites with a previously studied functionality. By applying Post-translational modification- signature enrichment analysis (PTM-SEA) to all the phosphosites detected, various signatures were found to be regulated during IR injury (Figure S4E). Overall, control and ischemic samples showed a lower expression of phosphosites related to growth factor response (*e.g.* EGF treatment and ERK2/MAPK1 signature) and cell division (*e.g.* CDK1 and CDK2 signatures), with the lowest intensities observed during ischemia. In contrast, 30 min reperfusion presented a high intensity of phosphosites related to growth factor stimulus. A similar trend was observed for CDK targets, which were highly phosphorylated during the reperfusion conditions when compared to the control and ischemic samples.

When looking at the individual trend of biologically relevant phosphosites belonging to these signatures, we encountered phosphorylation on transcription factor ATF2, namely T69 and T71 to be upregulated upon ischemia, peaking at 30 min reperfusion and dropping at 120 min (Figure 3D). On the other hand, phosphorylation on the ribosomal protein RPS6, a known marker of active translation, increased drastically at 30 min reperfusion. Potential targets of MAP kinases and CDKs were upregulated during reperfusion times, such as S405 and S418 of CTTN, and S25 and S38 of STMN1 (Figure 3D). The latter two proteins are related to cytoskeleton organization.

Proteins showing regulated phosphosites as well as significant alterations in abundance

When looking at the overlap between the proteome and the phosphoproteome, we found that twenty-two of the differentially phosphorylated proteins were also found to be significantly changed in abundance at the protein level. Among these overlapping proteins, a major part (16/22) was located in cluster 2 of the changing proteome, and hence these proteins showed a decrease in abundance during reperfusion and the majority was related to cell junction and cytoskeleton (*e.g.* CDHR5, CTTN, CGN, MYO7B and ESPN) and digestion/absorption (*e.g.* LCT, SLC9A3R). For these 16 proteins, almost half of the phosphorylation changes occurred already at 45I (27 changing phosphosites in clusters 1, 2, 5, and 6 versus 29 altered phosphosites in clusters 3 and 4), indicating that alterations in phosphorylation preceded a decrease in their abundance.

On-tissue imaging of tryptic peptides using MALDI-TOF MSI

Next to IR-induced proteomic changes in whole tissues, we explored histological region-specific proteins changes. High-speed MALDI-TOF MSI enabling bottom-up tissue proteomics experiments were performed on a total of 36 tissue sections belonging to nine patients (four experimental conditions per patient: Ctrl, 45I, 30R, 120R). Prior to imaging, proteins underwent tryptic on-tissue proteolysis. As digestion efficiency greatly influences signal intensities of tryptic peptides, a spot of cytochrome C was added to each slide as a quality control (*18*), and its digestion profile was used to detect outliers. Based on PCA analysis of the average cytochrome C mass spectrum, one out of nine patient data sets was excluded from analysis (Figure S6). Figure 4A shows the average peptide spectrum across all control tissues.

MALDI-TOF MSI of distinct histological structures in human small intestine

We first compared peptide distributions with the tissue's histology in order to evaluate the potential of MALDI MSI to detect region-specific IR-induced protein changes. After MALDI MSI, tissue sections were H&E-stained, and optical scans were co-registered to the MSI images. Histological regions were annotated in the H&E images (Figure 4B, Figure S7) and heatmaps of average mass spectra acquired from mucosa, submucosa and muscle regions showed distinct peptide profiles for the different histological regions (Figure 4C). In addition, individual peptide images showed specific localization in distinct intestinal tissue structures. For example, *m/z* 866.5 was located in the mucosa region (Figure 4D) and *m/z* 1198.7 in muscle (Figure 4E). In addition, *m/z* 840.5 was found to be associated with connective tissues of the submucosa layer but also surrounding muscle tissue (Figure 4F), and *m/z* 1274.6 was specifically located in blood vessels (Figure 4G). In order to identify region-specific IR-induced protein changes, we performed subsequent analysis for each histological layer separately. Since 10 out of 32 tissues did not contain muscle in the analyzed section (Figure S7), statistical analysis could not be performed for the muscle layer.

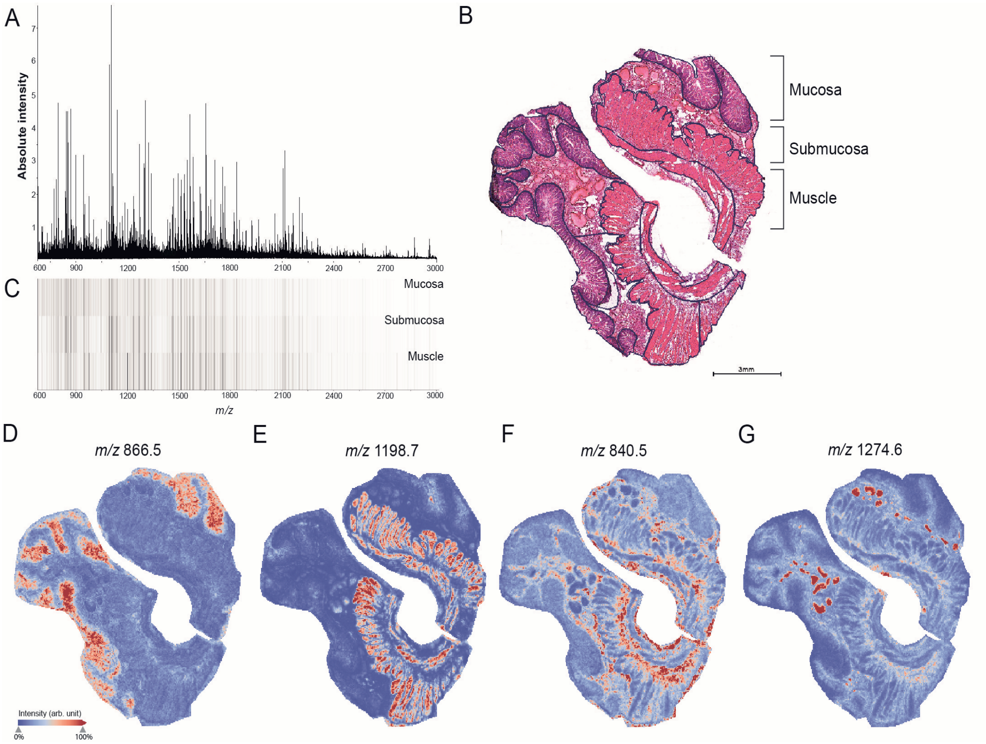


Figure 4. MALDI-TOF MSI of distinct histological structures in human small intestine. (A) Average mass spectrum of whole tissues (Ctrl). (B) H&E staining showing annotation of histological layers mucosa, submucosa and muscle. See also Figure S7 for region annotations in all tissues. (C) Heatmap visualizing intensity differences for average mass spectra obtained from mucosa, submucosa and muscle layer of small intestine. Individual peptide images of m/z values with specific localization in (D) mucosa (m/z 886.5 ± 0.2Da), (E) muscle (m/z 1198.7 ± 0.2Da), F) submucosa (m/z 840.5 ± 0.2Da), and blood vessels (m/z 1274.6 ± 0.2Da). All peptide images were generated from TOF-MSI data.

IR-induced changes in mucosa and submucosa

In the context of IR injury, we are particularly interested in the mucosa layer as this is the most susceptible to damage. PCA analysis of mucosa regions, revealed the highest similarity between Ctrl and 45I conditions on one hand, and reperfusion conditions (30R, 120R) on the other hand (Figure 5A), which is congruent with the clustering of the LC-MS/MS proteomics data. Peak picking resulted in 319 peptide signals to be included for statistical analysis, of which 154 *m/z* values were found to be significantly changing in intensity during IR. In general, signal intensities were either gradually correlated or anti-correlated to the IR sequence (Ctrl-45I-30R-120R). Remarkably, a decreasing intensity gradient was observed predominantly for peptides in the lower mass range (*m/z* <1500), whereas peptides in the higher mass range (*m/z* >1500) showed an increasing intensity gradient. Images of the peptides with the ten highest fold changes showed mucosa-specific localization and a decreasing abundance during IR in most of the patients (Figure 5C-II, all tissues in Figure S8A-D). Two of these were among the top 10 peaks with highest intensity (Figure S8C, D). Interestingly, some peptides show a distribution shift from whole mucosa in Ctrl towards the villus tips after 120R, as shown in Figure 5C-III (Figure S8A). The peaks with an increasing intensity gradient appeared to be mostly low intensity peptides, expressed in all histological layers.

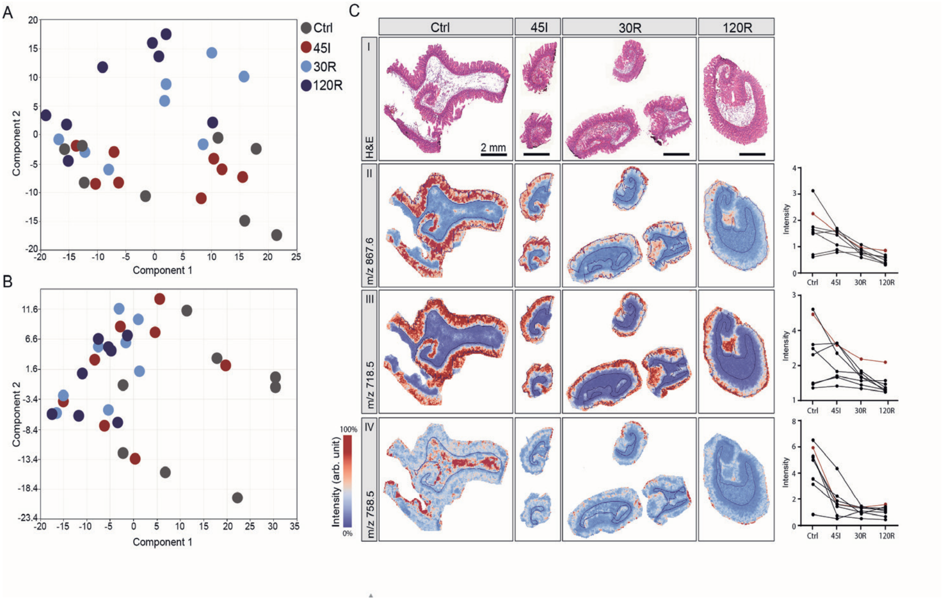


Figure 5. Region-specific changes in response to ischemia-reperfusion. Principle component analysis of (A) mucosa regions and (B) submucosa regions. (C) H&E staining (I) and peptide images showing high-fold change peptides in mucosa (II, III) and submucosa (IV). **Legend continues on next page.**

Figure 5 legend continued. Corresponding graphs show intensity data for all patients (presented peptide images correspond to red line in the graph). All peptide images were generated from TOF-MSI data for the indicated m/z values ±0.2Da.. See also Figure S8 and Figure S9 for peptide images from all samples.

PCA analysis of submucosa regions resulted in slightly different grouping of samples compared to mucosa (Figure 5B). In total, 185 peptide signals were found to be significantly changed between IR conditions and all showed a decreasing intensity gradient. Looking at peptides exhibiting a high fold change, only a few showed specific localization in the submucosa layer (Figure 5C-IV, all tissues in Figure S9A, B). Images of high intensity peptides better represented submucosa-specific peptide changes (Figure S9C, D).

Protein identification of relevant peptides obtained from MALDI-TOF MSI analysis

In order to link the MALDI-TOF MSI data (R=15,000 at *m/z* 1000*)* of peptides to the LC-MS/MS protein identifications, we used MALDI-FTICR MSI measurements (R=200,000 at *m/z* 1000) as an intermediate step to obtain more accurate mass descriptions of the peptides of interest. With regard to the mucosa, 75 out of 154 significantly changing *m/z* values from TOF-MSI could be matched to one or more peaks in the high mass resolution FTICR data. Matching these accurate *m/z* values to the LC-MS/MS (phospho)proteome data enabled annotation of ten proteins. Performing the same analysis for the submucosa data resulted in 96 matched FTICR *m/z* values and in the annotation of 11 proteins. None of the annotations were based on matched phosphopeptides. We next highlight a selection of proteins of interest, showing layer-specific localization (peptide images can be found in Figure S10).

Remarkably, among annotated proteins significantly changing in the mucosa, three proteins (ELAV1, SNRNP70, HNRNPC) were associated with mRNA splicing and processing. Peptides signals belonging to these proteins were mucosa-specific and had highest intensity in the control sample which decreased during IR sequence (Figure 6-II, Figure S10A-C). Images of complement C5 showed localization specifically in the mucosa layer. The decreasing intensity during IR was accompanied by a distribution shift towards the villus tips (Figure 6-III) and expression in villus debris in reperfusion samples (Figure S10D). Interestingly, for complement C7 a decreasing intensity ($P<0.01$) in submucosa was accompanied with an increase in mucosa layer ($P=0.02$). Peptide images confirm this distribution shift (Figure 6-IV, Figure S10E). Collagen-alpha-2(I) chain, a structural constituent of the extracellular matrix, exhibited a significant downregulation in the submucosa (Figure 6-V, Figure S10F).

Imaging MS and quantitative MS-based proteomics are complementary methods

When comparing imaging MS with LC-MS/MS proteomics results, only two (Nup50, complement C5) of the 22 annotated proteins overlapped with the list of significantly changed proteins based on LC-MS/MS analysis. Nevertheless, the proteins found to be changed in MSI experiments were involved in processes that were over-represented among differentially expressed proteins in LC-MS/MS analysis. Figure 7 summarizes and connects the most important results of this study.

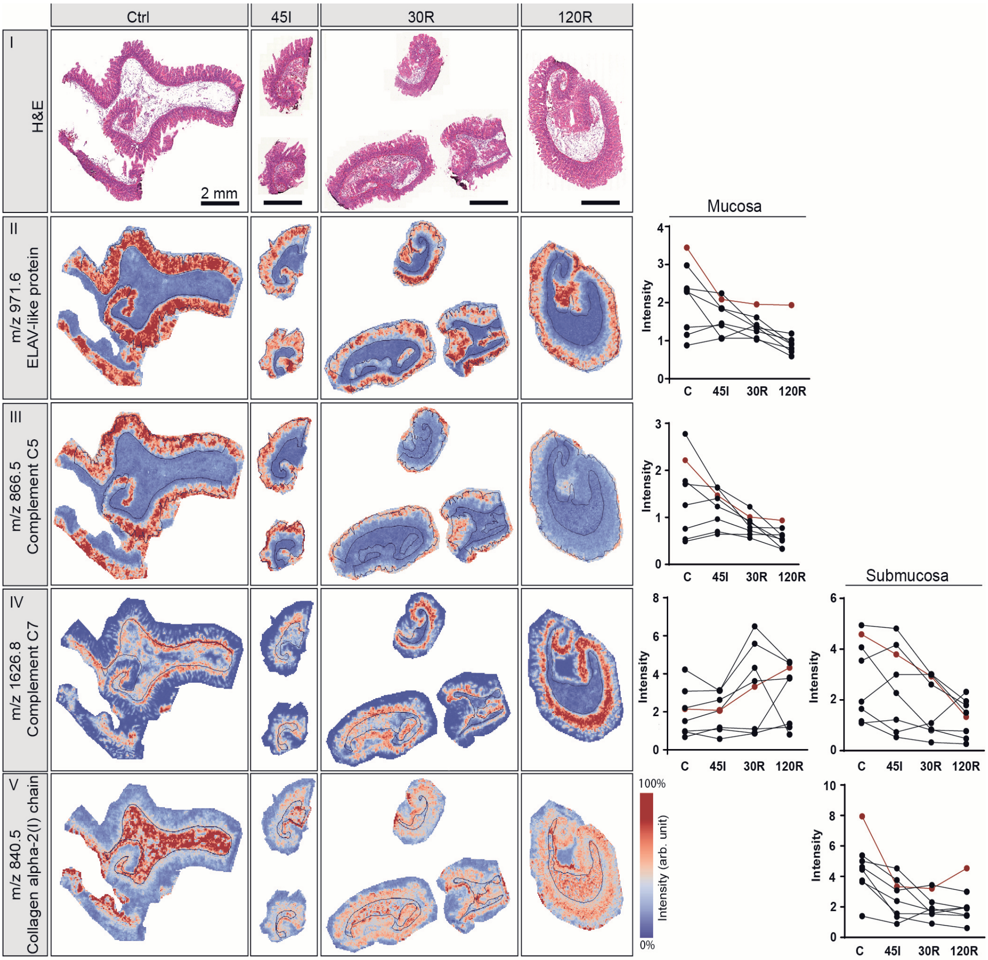


Figure 6. A selection of annotated proteins using MALDI-TOF MSI in combination with FTICR and LC-MS/MS. H&E staining (I) and peptide images from one patient are shown for indicated m/z values, which were annotated as ELAV-like protein (II), complement C5 (III), complement C7 (IV) and collagen alpha2(I)chain (V). **Legend continues on next page.**

Figure 6 legend continued. Graphs show intensities for all patients in mucosa (left) and/or submucosa (right) in case the m/z was differentially expressed in the respective layer. Intensities corresponding to the presented peptide images are shown in red. All peptide images were generated from TOF-MSI data for the indicated m/z values ±0.2Da. See also Figure S10 for peptide images from all patient samples.

Overrepresented processes in the changing proteome (blue), phosphoproteome (red), or both (blue-red) are highlighted, and images of significantly changed peptides and corresponding annotated proteins are shown with arrows pointing towards the corresponding process (red, mucosa; blue, submucosa). The overview shows that the applied methods complement each other; insight into the different regulated processes in the global (phospho)proteome is accompanied by partially overlapping spatial information provided by MSI.

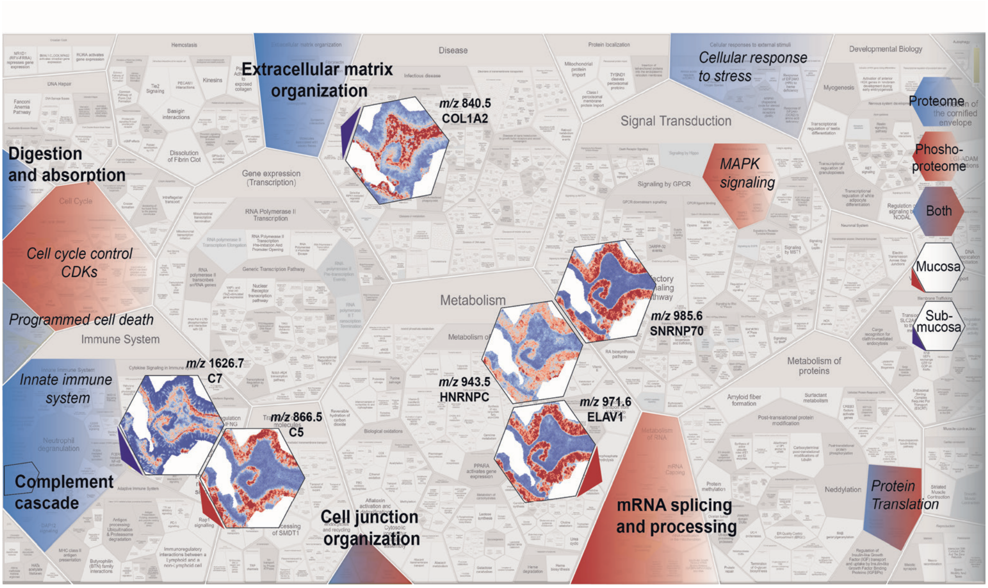


Figure 7. Summary of MS-based proteome and phosphoproteome functional analysis and imaging MS data showing that these methods complement as well as support each other. Overview summarizing regulated processes during IR based on the dynamic proteome (blue) and phosphoproteome (red) or both (blue-red). Processes that were shown to be significantly enriched are indicated in bold. Peptide images show significantly changing peptides, specifically located in mucosa (red arrow) or submucosa (blue arrow). The annotated proteins were related to indicated processes (arrow) Overview image was adjusted from ReacFoam format (Reactome.org). C5, complement 5; C7, complement 7; SNRNP70, small nuclear ribonucleoprotein U1 subunit 70; HNRNPC, heterogeneous nuclear ribonucleoprotein C; ELAV1, ELAV like protein 1; COL1A1, collagen 1 alpha2 (I) chain.

Discussion

Here, we present a comprehensive study of the proteome and phosphoproteome in combination with MALDI MSI to unravel protein alterations during IR of the human intestine. LC-MS/MS-based (phospho)proteomics resulted in the identification and quantification of thousands of proteins and phosphosites, and enabled thorough functional enrichment and interaction network analyses. We showed that proteins related to intestinal absorption, microvillus structure and cell junction decreased in abundance after IR, whereas proteins involved in innate immunity were increased in abundance. Phosphoproteome analysis revealed regulation of RNA splicing events and cytoskeletal/cell junction organization, and suggested MAPK and CDK families to be active kinases during IR. In addition, MALDI MSI enabled identification of mucosa-specific protein changes as well as alterations in protein distribution, for instance a shift in localization of complement C5 during the course of IR.

Functional interpretation of the dynamic proteome in intestinal IR

Functional enrichment analysis of the dynamic proteome during IR revealed that proteins showing a decrease in abundance during reperfusion were overrepresented for GO terms related to microvilllus/cell junction/cytoskeleton. Proteins showing an increase during reperfusion were related to the innate immune response. Functional interpretation of the changing proteome will be discussed per cluster.

The downregulation of various translation initiation factors (EIF2S2, EIF4EBP1, EIF5B) suggests that protein translation is inhibited in the early reperfusion phase. Inhibition of translation initiation is one of the cytoprotective mechanisms of the unfolded protein response (*19*), which is induced in response to proteotoxic stress in the ER, and known to play an important role in IR injury (*8*). Interestingly, we observed a decrease in abundance of HspA4, a member of the Hsp70 family, which act as chaperones and is known to be induced in response to proteotoxic stress and protect cells from its harmful effects (*20*). We speculate that acute IR-induced/oxidative stress depleted Hsp70 protein at 30R, which was rapidly recovered by IR stress-induced transcription of Hsp70s (*21*). Another stress-response related protein showing decreased abundance was Nup50, which has a direct role in nuclear transport and is known to be sensitive to different stressors including oxidative stress (*22*). A reduction in TNIK, which acts as a critical activator of Wnt targets (*23*), suggests that this kinase may play a role in inhibition of proliferation during IR-induced cellular stress.

The decreased abundance of proteins involved in intestinal digestion/absorption and related to microvillus and cell junction organization likely reflects the loss of villus tips as a consequence of reperfusion injury. This group of proteins included many structural proteins of intestinal microvilli, but also brush border enzymes and transporters for nutrient absorption. Previous work described how IR-induced destruction of the intestinal epithelium led to contraction of the epithelial sheets and shedding of the damaged villus tips which resulted in reduced length of the villi (*6*). Reorganization of the actin cytoskeleton and accumulation of F-actin at the basal side of enterocytes enabled this protective mechanism. The actin cytoskeleton is tightly anchored to the lateral membranes by cell-cell junction complexes. This interaction between the cytoskeleton and cell junction is crucial for the integrity of epithelial barrier and changes in organization of either one affects the other and may contribute to gut barrier disruption (*24*), which is known to occur in inflamed and injured intestine (*25*) and cardiac IR (*26*). A decreased abundance of cytosolic IFABP protein in the villus is a well-known consequence of intestinal IR (*5*). The loss of enterocyte membrane integrity results in the release of IFABP into the circulation and has been shown to be a useful serological biomarker for intestinal IR injury (*27-29*).

Functional analysis of the cluster of proteins exhibiting increased expression during reperfusion (cluster 3), strongly indicates activation of the innate immune response upon reperfusion of the ischemically damaged intestine, and points in particular to a crucial role of the complement system, as also demonstrated by the presented network of interconnected proteins in this cluster. Activation of the complement system has been well-documented in animal models of intestinal IR (*30-33*). In addition, complement activation after IR in human intestine has been previously reported (*7*). In that study, high amounts of complement activation product C3c were detected in the luminal debris of shed enterocytes but not in mucosal tissue, whereas native C3 was present in the tissue (*7*). Interestingly, our MSI data show that C5 expression shifted towards the villus tips and that C5 was almost absent in the mucosal tissue itself during reperfusion. The decrease of complement C5 in the mucosa layer seems contradictory to the quantitative proteomics data, showing an overall increase in complement proteins. This discrepancy may be explained by homogenization of whole tissue, including the luminal debris that may contains complement proteins, for proteomics analysis. Images of C7, on the other hand, showed increasing abundance in the mucosa layer together with a decrease in submucosal expression. It should be noted that we cannot verdict on actual complement activation, as our proteomics analysis could not distinguish the native and active forms of complement factors. Our findings shed new light on the importance of the complement system in IR of human intestine and could give rise to further studies investigating activation of complement and its role in human intestinal IR. To date, the potentially protective effects of complement inhibition during intestinal IR have been investigated in animal models only (*31-33*). Moreover, these data underline the strength of combining LC-MS/MS data giving robust and reliable quantitative data and MS images exposing changes in protein localization. An alternative spatial proteomics approach that could be very useful to identify and quantify peptides in a specific tissue area, is laser capture microdissection of the area of interest followed by LC-MS/MS analysis (*34, 35*).

Furthermore, both LC-MS/MS and MSI approaches point to changing ECM organization during IR, which is in accordance with previous proteomics analysis of hypoxia-reoxygenation in a human intestinal organoid model (*36*). Interestingly, network analysis indicated interconnection of proteins directly related to innate immunity and ECM proteins. Unbalanced ECM remodeling, and associated altered expression of ECM proteins, are a well-known feature in inflammatory bowel disease (IBD) (*37*). Immune activation and inflammation are known to induce both the degradation and synthesis of the ECM. The interplay between inflammation and the ECM is a dynamic process, in which ECM alterations can also actively promote inflammation and contribute to disease progression in IBD (*38*). The observed alterations in expression of ECM proteins, both quantitatively and in peptide images, may reflect remodeling of the ECM as a result of IR-induced inflammation.

The small cluster of proteins that gradually increased during IR (cluster 4) contained proteins involved in a variety of processes. Cytoglobin, which has an important role in oxygen transport, was significantly increasing during IR, and has been shown to be protective against IR injury in other organs (*39, 40*). Interestingly, the increase in caspase-1 protein may point to promotion of pyroptosis, a pro-inflammatory form of programmed cell death, which is initiated by caspase-1, and has recently been shown to play a role in murine intestinal IR injury (*41*).

The dynamic phosphoproteome

Protein phosphorylation and its regulation by kinases and phosphatases play a key role in the regulation of cellular functions, and changes in phosphorylation can be a cause as well as a consequence of a variety of diseases (*42*). By analyzing our data with a combination of phosphosite- and protein centric approaches, we identified dynamic protein phosphorylation events that regulate specific biological processes, and hence are expected to play an important role in the cellular response to IR.

Hierarchical clustering accompanied with motif analysis per cluster revealed the dynamic nature of protein phosphorylation during IR. The majority of phosphosites are proline directed, which suggests that they are potential targets of a variety of kinases, from CDKs to MAP kinases (*43, 44*). In the case of IR, it is very likely that many of the regulated phosphosites are targets of MAP kinases such as JNK and p38, which are activated by cellular stress and thus their targets are expected to locate to clusters showing upregulation after IR (cluster 1, 3, 4, 5). Indeed, phosphosites of STMN1 (S25, S38), ATF2 (T69, T71) and CTTN (S405, S418), which are proposed targets of either JNK or p38, locate to clusters 3 and 4, which contain phosphosites that appear upregulated during reperfusion. Proline-directed phosphosites could also be targets of MAP kinases related to growth and survival, like ERK1/ERK2. We found that phosphorylation of ERK2 on T185 and Y187, which are indicative of kinase activation (*45*), were located in cluster 4, which suggests that ERK2 is highly active shortly after reperfusion. Cluster 5, which exhibits upregulated phosphosites at 45I, is predominantly comprised of non-proline directed phosphosites, which suggests that other kinases than MAPK are likely to be active during ischemia. Overall, the differences in dynamics and motif composition amongst clusters suggests that protein phosphorylation response to IR is complex and comprises different effector kinases acting at different stages.

Phosphorylation dynamics and its relation with signaling pathways

We further examined which kinases may be responsible for the protein phosphorylations, by analyzing how phosphosite dynamics relate to signatures of specific kinases or signaling pathways. As expected, we observed that most potential MAPK and CDK targets tend to decrease during ischemic conditions, followed by a drastic increase during reperfusion. However, when looking into specific phosphosites with known biological function, we highlighted some examples displaying interesting trends. One of these is the phosphorylation of ATF2, a transcription factor regulating cell growth and survival, on T69 and T71, which are known targets of MAP kinases in response to both cellular stress and growth factors (*46*). Both phosphosites showed increased intensity upon ischemia and peaked at 30R, which equals induction of ATF2 transcription activity (*47-49*). This is in line with previously reported increased ATF2 binding activity in renal IR (*50*). It is conceivable that stress responsive MAP kinases (JNK and/or p38) induce phosphorylation during ischemia, which is subsequently boosted by other MAP kinase activity (*e.g.* ERK1/ERK2) in response to growth factors upon reperfusion (*51*).

In addition, we identified protein phosphorylation events known to regulate protein translation, namely S235 and S236 phosphorylation on RPS6, an important ribosomal protein that is regulated by kinases responsive to growth factors (*52, 53*). As these phosphosites on RPS6 promote assembly of the preinitiation complex (*54*), our data indicate translation seemed to be inhibited during ischemia and resumed upon reperfusion. The latter supports our findings in the proteome data, where proteins involved in translation appeared to decrease in abundance early in reperfusion and then recover after 120R.

Functional interpretation of changes in the phosphoproteome

Through GO enrichment analysis, we found that most of the phosphorylation-regulated proteins were involved in RNA splicing and cell junction/cytoskeleton organization. This is consistent with a previous phosphoproteome study in a swine model for cardiac IR, reporting that the majority of phosphoprotein alterations were involved in RNA processing and cell junction (*55*).

Pre-mRNA splicing is executed by the spliceosome. The phosphorylation state of splicing factors is crucial for correct regulation of their function and organization, and for the formation of the spliceosome complex (*56*). Splicing factors exhibited increasing as well as decreasing phosphorylation (cluster 1 and 5), suggesting that both phosphorylation and dephosphorylation events play a role in the regulation of RNA splicing, which has been shown previously (*57*). In addition to the role in constitutive splicing, phosphorylation acts as a major player in regulation of alternative splicing (*58*), a process that 95% of human genes undergo, and is considered an important mechanism in pathological cellular processes (*59*). Accumulating evidence indicates that pre-mRNA splicing plays an essential role in the adaptation to hypoxic stress (*60-62*). Hypoxia-induced alternative splicing, for example of VEGF, Bcl-x, BNIP3 and CAIX, changes gene expression patterns in order to enhance proliferation and survival (*61*). This may explain the ischemia-induced changes in phosphorylation of splicing-related proteins in our model. Furthermore, stress-activated MAP kinases such as JNK and p38, which are activated during IR, can indirectly modulate splicing by phosphorylation of selected splicing factors (*63*). In addition to the phosphorylation changes in splicing factors, MSI results showed a locally decreased mucosal intensity of peptides annotated as spliceosome component SNRNP70 and RNA binding proteins ELAV1 and HNRNPC which play a role in RNA splicing. This further supports regulation of splicing events during IR.

A substantial part of phosphorylation-regulated proteins was related to cell junction and cytoskeletal organization. In contrast to the decrease in abundance of these proteins following reperfusion, changes in phosphorylation exhibited various temporal profiles, including phosphosites that were regulated immediately after ischemia. Regulation of cell junction proteins by phosphorylation plays an important role in the assembly and disassembly of adherence junctions (*64*). The latter are known to interact with the actin cytoskeleton of adjacent cells, suggesting that the dynamic phosphorylation of junctional and cytoskeletal proteins during IR likely affects its organization and integrity. Interestingly, we detected some specific phosphosites with a known function in the regulation of cytoskeleton organization, which showed an increase during reperfusion. Cortactin (CTTN) phosphorylation on S405 and S418 is associated with cytoskeleton reorganization (*65, 66*), and STMN1 phosphosites S16, S25 and S38 are linked to polymerization of the microtubule cytoskeleton (*67*). Together, these findings suggest that increased phosphorylation of junctional proteins upon reperfusion may be related to the reorganization of the cytoskeleton during IR of the human intestine (*6*).

MSI data interpretation and limitations

The clustering of control and ischemia samples on the one hand and reperfusion samples on the other hand, was comparable for MSI and proteome data. However, the fact that for MSI the peptides in the lower mass range exhibited a decreasing gradient, and those in the higher mass range an increasing intensity gradient may suggest that enzymatic digestion efficiency was affected negatively by ischemia and an increasing reperfusion time. While these results have to be interpreted with care, we can speculate that IR-induced changes in endogenous proteolytic enzymes may affect digestion efficiency. This is supported by animal studies which have shown that during IR, pancreatic enzymes from the intestinal lumen leak into the intestinal wall, resulting in self-digestion(*68, 69*).

Where LC-MS/MS proteomics is a well-established and robust technology, MSI methodologies are evolving rapidly. One of the existing challenges of tryptic peptide MSI is the identification of the corresponding proteins. Here, we assigned MSI-detected peptides to protein data from LC-MS/MS analysis of the same tissue samples. A trade-off between speed, sensitivity and mass resolution made us decide to use MALDI-TOF MSI for the screening of the 36 tissue samples, and to perform additional MALDI-FTICR measurements only as an intermediate step to aid in the identification process because of their higher mass accuracy. When compared to phospho- and global proteomics, MSI identification results were sparse. There are several explanations for this. First, the number of peptide signals is relatively low in MSI analysis due to the mass resolution of the TOF system and the lack of an additional dimension of separation. This low mass resolution was also the major limiting factor for identification of the peptide signals since about 30% of the TOF peaks could be further resolved into at least two peptide signals in the FTICR spectrum. We therefore assume that the use of high-resolution MALDI MSI for all tissue sections would have significantly increased the number of identified (phospho)peptides, although it is more time consuming. Another factor that limits identification is the use of different ionization methods (ESI for LC-MS/MS versus MALDI in MSI), which inherently limits the overlap between the detected peptides. Despite these detrimental factors, the proteins annotated to significantly changing peptides with our MSI analysis were related to the same processes that were shown to be altered in our quantitative proteomics analysis, supporting the coherence between the different types of data used. For further studies, MSI- and/or morphology guided laser microdissection could be performed and analyzed by LC-MS/MS (*34*).

Conclusion

Altogether, we identified IR-induced alterations in abundance, phosphorylation and distribution of proteins, which expanded our understanding of the molecular events that occur during IR in human intestine. In addition, the study highlights the strength of the complementary use of different MS-based methodologies.

Methods

Patients and experimental procedure

Human intestinal tissues exposed to ischemia and reperfusion were obtained using a controlled *in vivo* experimental model. The study was approved by the Medical Ethical Committee of the Maastricht University Medical Center+ and written informed consent was obtained from all patients. Nine patients (Sex 3M:6F; median age 66 years, range 43–84 years) undergoing pancreatoduodenectomy were included in the study. The experimental procedure was performed as described previously(*5*). In short, a 6–cm jejunal segment, which is routinely resected as part of the surgical procedure, was isolated and subjected to ischemia by placing vascular clamps across the mesentery. After 45 min, one third of the ischemic segment was resected (45I). Next, clamps were removed to start reperfusion. Another segment of isolated jejunum was removed after 30 min (30R) and 120 min of reperfusion (120R). Finally, a jejunal segment which was not exposed to IR, but underwent similar surgical handling, was resected (Control - Ctrl). Jejunal tissue samples were immediately snap-frozen and stored at -80 °C.

LC-MS/MS analysis

Cell lysis and protein digestion

Tissue samples were treated with sodium deoxycholate (SDC) 1% to induce cell lysis. The buffer also contained 10 mM tris(2-carboxyethyl)-phosphinehydrochloride (TCEP), 40 mM chloroacetamide, 100 mM TRIS pH 8.0, further supplemented with protease inhibitor (cOmplete mini EDTA-free; Roche, Basel, Switzerland) and phosphatase inhibitor (PhosSTOP, Roche). Samples were sonicated with a Bioruptor Plus (Diagenode, Liège, Belgium) for 15 cycles of 30 seconds. Protein amount in each sample was quantified by a Bradford protein assay. Next, proteins were digested overnight at 37 °C with Lys-C (FUJIFILM Wako pure chemical corporation, Osaka, Japan) and Trypsin (Sigma-Aldrich, Zwijndrecht, The Netherlands), with enzyme to protein ratios of 1:75 and 1:50, respectively. SDC was precipitated by addition of 2% formic acid (FA) and peptides were desalted using Sep Pak C18 cartridges (Waters Corporation, Milford, Massachusetts, USA), to subsequently be dried down and stored at -80 °C.

Phosphopeptide enrichment

Phosphopeptides were enriched by Ti(IV)-IMAC; 500 µg of beads were packed into microtip columns and washed with methanol and a loading buffer made of 80% acetonitrile (ACN) and 6% trifluoroacetic acid (TFA). 200 µg of peptides per sample were dissolved in loading buffer and subsequently loaded into the columns. Peptides were washed with 50% ACN/ 0.5% TFA in 200 mM NaCl, followed by a second wash with 50% ACN/ 0.1% TFA. Phosphopeptides were eluted with 10% ammonia and 80% ACN/ 2% FA directly into 10% FA. Samples were dried down and stored at -80 °C until LC-MS/MS analysis. A detailed description of this protocol was published elsewhere (*70*).

Data acquisition by LC-MS/MS

Samples were analyzed using a UHPLC 1290 system (Agilent, Santa Clara, California, USA) coupled to an Orbitrap Q Exactive HF (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Nano flow rate (~300 nl/min) was achieved by passively splitting the flow using an external valve (*71*). Peptides were first trapped into a pre-column (inner diameter [ID] of 100 µm and 2 cm length; packed in-house with 3 µm C18 ReproSil particles [Dr. Maisch GmbH]) and eluted into an analytical column (ID of 75 µm and 50 cm length; packed in-house with 2.7 µm Poroshell EC-C18 particles [Agilent]). We used a two-system buffer consisting of solvent A (0.1% FA in water) and B (0.1% FA in 80% ACN). Peptides were trapped during 5 min at 5 µL/min flow-rate with solvent A, before switching to a nano flow of ~300 nL/min. For the measurement of the full proteome we used a 155 min gradient from 10 to 36% of solvent B. On the other hand, for the phosphoproteome we used a 95 min gradient from 8 to 32% of solvent B. Both methods included a wash with 100% solvent B for 5 min followed by a column equilibration with 100% solvent A during the last 10 min.

The mass spectrometer was operated in data dependent acquisition mode. For the proteome analysis, full scan MS was acquired from *m/z* 375-1600 with a 60,000 resolution at *m/z* 200. Accumulation target value was set to $3e^6$ ions with a maximum injection time of 20 ms. Up to 15 of the most intense precursor ions were isolated (*m/z* 1.4 window) for fragmentation using high energy collision induced dissociation (HCD) with a normalized collision energy of 27. For MS2 scans an accumulation target value of $1e^5$ ions, a maximum injection time of 50 ms and a dynamic exclusion time of 24 seconds were selected. Scans were acquired from *m/z* 200-2000 with a 30,000 resolution at *m/z* 200. For the phosphoproteome the same settings were used with the exception of the dynamic exclusion window, which was set to 16 seconds. The electrospray voltage was set to 1.9 kV during the measurement of all samples.

Data processing

Raw files were processed with MaxQuant (version 1.6.17.0) using a false discovery rate (FDR) <0.01. The default settings were used, with the following exceptions: variable modifications, specifically methionine oxidation, protein N-term acetylation and serine, threonine and tyrosine phosphorylation were selected. Cysteine carbamidomethylation was selected as a fixed modification. Label free quantification was performed and we enabled the ‘match between runs’ option with the default values. Database search was conducted against the human reviewed Swiss-Prot database (October, 2020).

The results were uploaded to Perseus (version 1.6.0.2) for subsequent analysis. For the proteome only proteins identified by more than one unique peptide were considered, and for the phosphoproteome only phosphosites with a localization probability score >0.75 were kept for further analysis. Decoys and potential contaminants were removed. The intensities were log2 transformed and normalized by median subtraction. Finally, all values were filtered to keep only those proteins (or phosphosites) that were detected in a minimum of 2 out of 3 replicates for at least one condition. Missing values were replaced using a normal distribution applying a downshift of 1.8 times the standard deviation of the dataset, and a width of 0.3 times the standard deviation.

An ANOVA test ($p < 0.05$) was used to keep only significantly changing proteins (or phosphosites) among the different conditions. Z-scored intensities were visualized using the Complex Heatmap package in R, applying a combination of k-means and hierarchical clustering. K-means clusters were set to 4 and 6, for proteome and phosphoproteome respectively. This number was chosen based on the gap statistic method, which estimates the optimal number of groups for a given dataset (*72*). Next, we used the Pearson correlation distance with average linkage for clustering. Gene ontology (GO) analysis was done using the STRING web tool (*73*); enriched GO terms were filtered in order to keep only those with $p < 0.01$, fold enrichment (Log10[observed/expected]) >5 and a minimum of 5 proteins per term. List of terms was further condensed by removal of redundant terms using Revigo web tool (*74*).

Sequence logos for motif analysis were obtained implementing previously described algorithms in R (*75*). Briefly, each sequence logo displays over- and underrepresented residues in each position of the sequence window centered on the phosphorylated residue. Calculations are based on the frequency change between a foreground (phosphosites from each cluster) and a background (all detected phosphosites). For kinase enrichment analysis we used the online tool KEA2, to look for phosphosites previously linked to effector kinases (*76*). In addition, we applied post-translational modification-signature enrichment analysis (PTM-SEA), which looks for enrichment of phosphosite-specific 'signatures' related to specific kinases, signaling pathways or perturbations previously reported in literature (*77*).

MALDI MSI analysis

Tissue preparation for MALDI MSI analysis

Fresh frozen tissues were sectioned at 10 μm thickness at -20 °C using a cryostat (Leica, Leica CM1860, Leica biosystems) and thaw-mounted onto clean indium tin oxide (ITO)-coated glass slides (Delta Technologies LTD, Loveland, USA). A within-subjects experimental design was pursued, i.e. all four different tissue sections (Ctrl, 45I, 30R and 120R) from one patient were always mounted on the same ITO slide. Mounted tissue sections were dried in a vacuum desiccator for 20 min, followed by three 2-min washes in 100% ethanol and then two 5-min washes in water. Fresh ethanol/water was used in every wash, and sections were not dried between steps. Antigen retrieval was performed in a 10 mM citric acid buffer (Sigma-Aldrich) (pH 6.0) for 20 min using the Antigen Retriever 2100 (Aptum Biologics, Rownhams, UK). Sections were allowed to cool down for 20 min, rinsed with water, and dried in a vacuum desiccator. Prior to trypsin digestion, 1 µL of 1 mg/mL cytochrome C (from equine heart, Sigma-Aldrich) was applied on the slide, away from the tissue, to evaluate digestion efficacy. Water-dissolved porcine trypsin (20 µg/mL) was sprayed onto the tissue samples using a SunCollect pneumatic sprayer device (Sunchrom GmbH, Friedrichsdorf, Germany) in 15 layers (flowrate: 10 µL/min, speed: 900 mm/min, track spacing: 1 mm, spray head distance: 25 mm). Afterwards, samples were incubated at 37 °C for 17 h in an airtight box containing 50% methanol. Finally, slides were coated with 5 mg/mL α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich) in 50% acetonitrile and 0.2% trifluoroacetic acid using the SunCollect sprayer device. The matrix was applied in a series of 7 layers with increasing flowrate starting at 10 µL/min followed by 20, 30 and 40 µL/min for all subsequent layers (speed: 1390 mm/min, track spacing: 2 mm, spray head distance: 25 mm). Prior to matrix application, slides were scanned (Super Coolscan 5000 ED, Nikon) to obtain high-quality optical images.

MALDI MSI data acquisition

High-speed MALDI-TOF MSI analysis was performed on a RapifleX MALDI Tissuetyper (Bruker Daltonics GmbH, Bremen, Germany) equipped with a 10 kHz Nd:YAG (355 nm) laser. The instrument was operated in positive ionization reflectron mode, and peptide spectra were acquired in a mass range *m/z* 620–3000 with a spatial raster with of 50 µm and 500 averaged laser shots per pixel. An experimental mass resolution of 15,000 was achieved at *m/z* 1000. High mass resolution MALDI-Fourier-transform ion cyclotron resonance (FTICR) MSI experiments were performed with a Solarix 9.4 Tesla (Bruker Daltonics), achieving an experimental mass resolution of 200,000 at *m/z* 1000. MSI data were acquired within a mass range of *m/z* 800–3000 (1E6 data points) in positive ionization mode with a transient time of 2.94 sec. The spatial raster width was 70 µm. At each pixel, 600 shots were accumulated with a laser frequency of 500 Hz. Data acquisition was controlled using ftmsControl and FlexImaging 4.1 (Bruker Daltonics). All MSI measurements were preceded by an instrument calibration using Red phosphorus. MALDI-TOF analysis was performed on 36 tissue samples from 9 patients. MALDI-FTICR MSI measurements were performed on selected samples (4 conditions from 1 patient) to improve identification of the proteins behind the relevant peptides obtained from TOF analysis.

Histological staining and tissue annotation

After MSI analysis, the matrix was removed by submersion in 70% ethanol and tissues were stained with hematoxylin and eosin (H&E). Optical images of H&E-stained tissues were obtained with a MIRAX desk scanner (Sysmex, Etten-Leur, The Netherlands). The MSI images were co-registered with the corresponding histological images in the FlexImaging software (v5.0, Bruker Daltonics), which allows the annotation of the histological regions of interest: mucosa, submucosa, and muscle layer.

MSI data pre-processing

MALDI-TOF MSI data were recalibrated using FlexAnalysis v3.4 (Bruker Daltonics). Cubic-enhanced calibration function was performed with a 500 ppm peak assignment tolerance and using *m/z* 868.5, 1138.6, 1562.8, 2115.2, 2567.3 and 2869.3 as calibrants. A total of 36 (9 patients with 4 conditions each) MALDI-TOF MSI data sets were imported into SCiLS 2019c (Bruker Daltonics) where mass spectra were normalized to their total ion count. Peak picking was performed on the overall mean spectrum in mMass (*78*) using the following settings: 35 precision baseline correction, deisotoping with an isotope mass tolerance of *m/z* 0.1 and isotope intensity tolerance of 50%), and a signal to noise (S/N) ratio of 7. The peak list was then imported back into SCiLS to create a data matrix containing for every annotated region and sample the maximum intensity in each peak interval (*m/z* ±0.2).

The average spectra of the MALDI-FTICR MSI datasets were recalibrated in mMass with linear correction using the tryptic peptides of histone H2A (*m/z* 944.5312; pos. 22-30 "AGLQFPVGR") and histone H4 (*m/z* 1325.7535; pos. 25-36 "DNIQGITKPAIR") and known trypsin autolysis products at *m/z* 842.5094 (pos. 108-115 "VATVSLPR") and *m/z* 1045.5637 (pos. 98-107 "LSSPATLNSR"). Peak picking on those recalibrated spectra was performed in MATLAB R2018 (Mathworks, Natick, Massachusetts, USA) using the following settings: TopHat filter (window: 30 dp) for baseline correction, Gaussian smoothing (window: 20 dp), a minimum intensity for peak picking of 500, deisotoping with an isotope mass tolerance of 0.02 *m/z* and isotope intensity tolerance of 50%.

Statistical analysis MALDI-TOF MSI data

An outlier detection was performed based on a cytochrome C spot applied onto every slide prior to trypsin digestion. For this, peak picking was limited to known peptides of cytochrome C, which resulted in the consideration of four signals (HKTGPNLHGLFGR, *m/z* 1433.77; HKTGPNLHGLFGRK, *m/z* 1561.87; TGPNLHGLFGRK, *m/z* 1296.72; TGPNLHGLFGR, *m/z* 1168.62). Using these 4 features, Principal Component Analysis (PCA) was performed on the average intensity in the cytochrome C spot from each slide (N=9) in SCiLS. Any measurement outside the 95% confidence ellipse in the PC1-PC2 score plot was considered an outlier.

The data matrix from the included patients was exported for statistical analysis in R (*v3.5.1*). Trypsin-related peaks were determined by Pearson correlation > 0.9 to the *m/z* 842.5, and excluded from analysis (removal of 6 peaks). Statistically significant differences in peak intensities between the conditions (Ctrl, 45I, 30R, 120R) were tested for each histological layer separately (mucosa, submucosa, muscle). The intensities for every peak were compared using a repeated-measurements ANOVA. P-values were corrected for multiple-testing by Benjamini-Hochberg and p-values <0.01 were considered statistically significant. As several tissue sections lacked the muscle layer, statistical analysis could not be performed for the muscle layer.

Protein identification strategy

MALDI-FTICR MSI experiments were performed on selected samples to obtain high mass-resolution data of tryptic peptides, which was used to identify the proteins behind the significantly changed peptides from MALDI-TOF MSI statistics. These significantly changed *m/z* values were matched with a tolerance of ±80 ppm to the peaks in the MALDI-FTICR MSI measurements. Next, these accurate peptide masses were matched with the masses of the (phospho)peptides detected using LC-MS/MS and corresponding identified protein with a tolerance of ±6 ppm. As there is no alkylation and reduction step in the MSI workflow, the peptide masses of the LC-MS/MS were adapted by subtracting the mass shift (*m/z* 57.02146) caused by the formation of S-carboxyamidomethylcysteine for every cysteine in the peptide.

The matching and identification process was performed using four FTICR MSI data sets (data was obtained from one patient across all four conditions). The annotation of a protein was accepted if mass matching from TOF to FTICR (with ±80 ppm tolerance) and FTICR to LC-MS/MS (with ±5 ppm tolerance) resulted in one matching protein ID and this in at least three out of the four FTICR data sets. In cases where two FTICR peaks were detected in the analogous mass range in the TOF spectrum, the most intense FTICR peak (>10 fold higher) within this window was selected for mass matching with the LC-MS/MS data.

Data availability

The mass spectrometry proteomics and phosphoproteomics datasets have been deposited to the ProteomeXchange Consortium via the PRIDE (*79*) partner repository with the data set identifier PDX026076.

Supplementary material

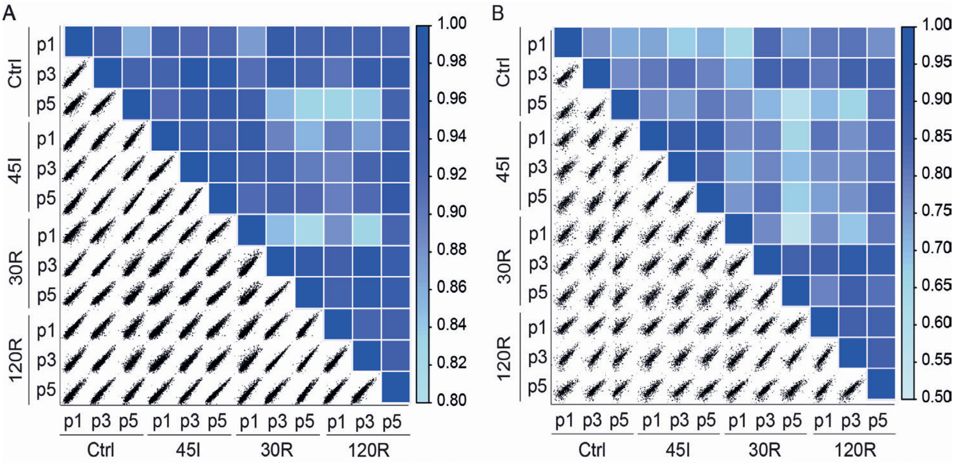


Figure S1. Correlation plots between all samples for proteome (A) and phosphoproteome (B). Good correlation between biological replicates, except for 30R-p1, which was therefore excluded from analysis. Patient p1, p3 and p5 correspond to patient numbers in MSI data. Intensity scale indicates the correlation coefficient.

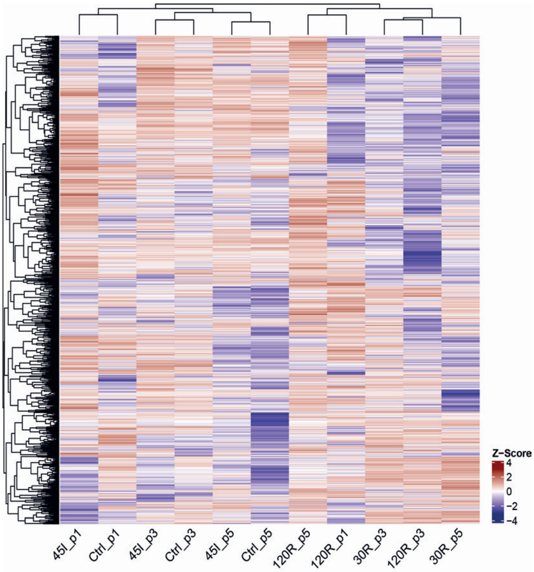


Figure S2. Hierarchical clustering of the complete proteome. Heatmap visualizes Z-scored intensities, and was generated using a combination of k-means and hierarchical clustering (Complex Heatmap package in R). Two main clusters are Ctrl and 45I samples on the one hand and 30R and 120R on the other hand.

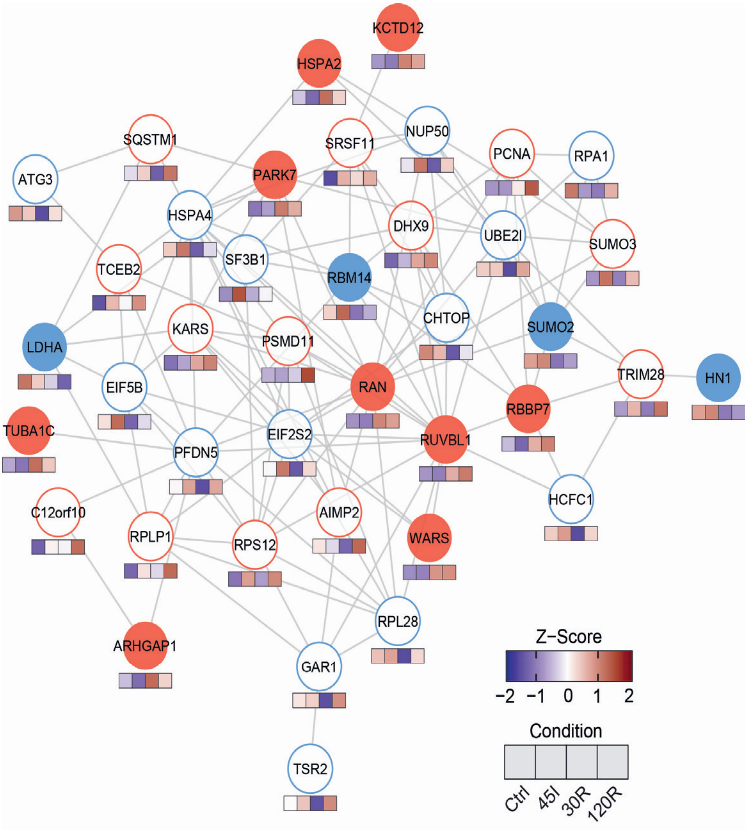


Figure S3. Protein interaction network. Network is generated using STRING. Color of the circle indicates the temporal profile (cluster) of the protein. Blue line, cluster 1; Blue fill, cluster 2; Red fill, cluster 3, Red line, cluster 4. Color cubes below each protein indicates Z-score intensity in Ctrl, 45I, 30R and 120 R respectively. This network shows interactions between proteins located in all clusters, with a majority in cluster 1 and 4.

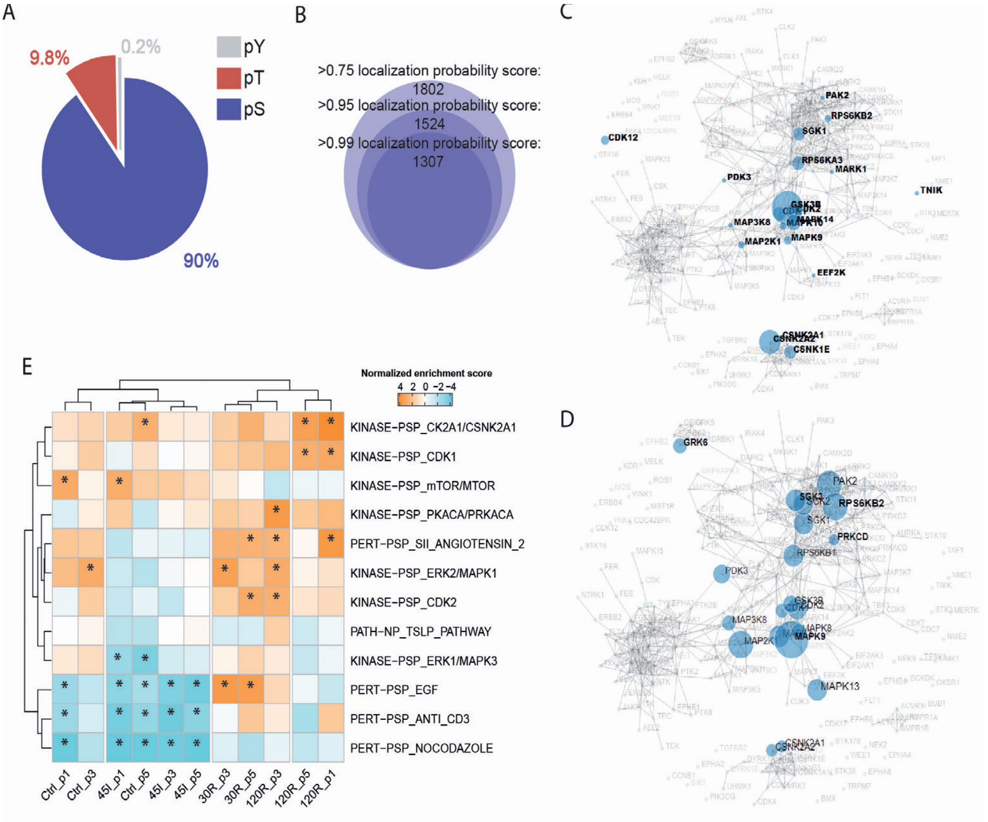


Figure S4. Global overview of the phosphoproteome. (A) Pie chart showing the proportion of phosphoserines, phosphothreonines and phosphotyrosines detected. (B) Concentric circles with proportional size corresponding to the number of phosphosites above the given localization probability scores. (C) Kinase enrichment analysis using KEA2 tool, showing a kinase network highlighting those with targets overrepresented in the whole phosphoproteome and (D) in the dynamic phosphoproteome (ANOVA, $p<0.05$). (E) Post-translational modification- signature enrichment (PTM-SEA) analysis. Heat map shows normalized enrichment scores of phosphorylation signatures detected in the whole dataset. Arterisks indicate enriched or depleted signatures enriched in the given sample (FDR<0.01).

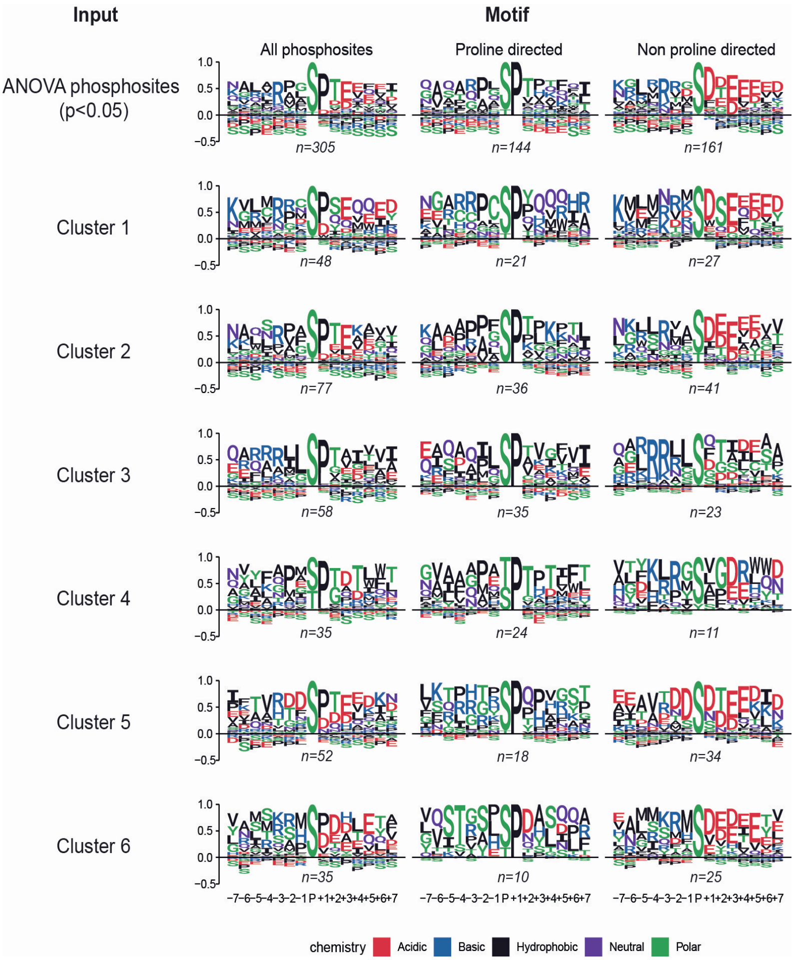


Figure S5. Sequence motif logos for dynamic phosphosites overall and per cluster. Motifs for all phosphosites is shown per cluster on the left; separate motifs for proline and non-proline directed phosphosites are shown in the middle and to the right respectively. The number of phosphosites from which the motif logo was obtained, is indicated below the logo.

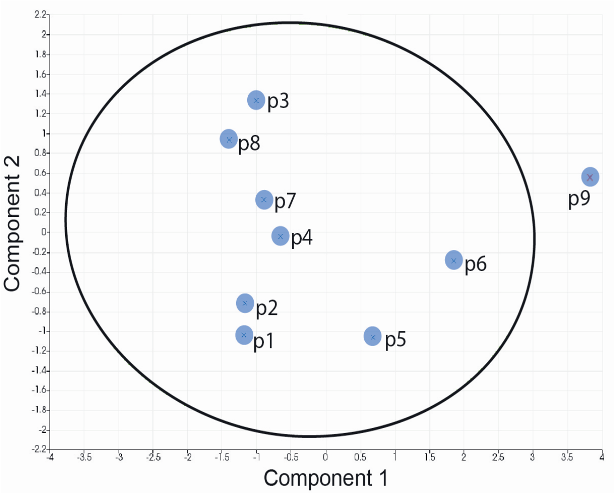


Figure S6. Principle component analysis of the average cytochrome C mass spectrum. Every dot represents the cytochrome C mass spectrum from one slide, and is thus used as a measure for digestion efficiency of the data sets from one patient. Respective patient numbers are indicated (p1 – p9). The ellipse indicates the 95% confidence interval, and any measurement outside the ellipse is considered an outlier.

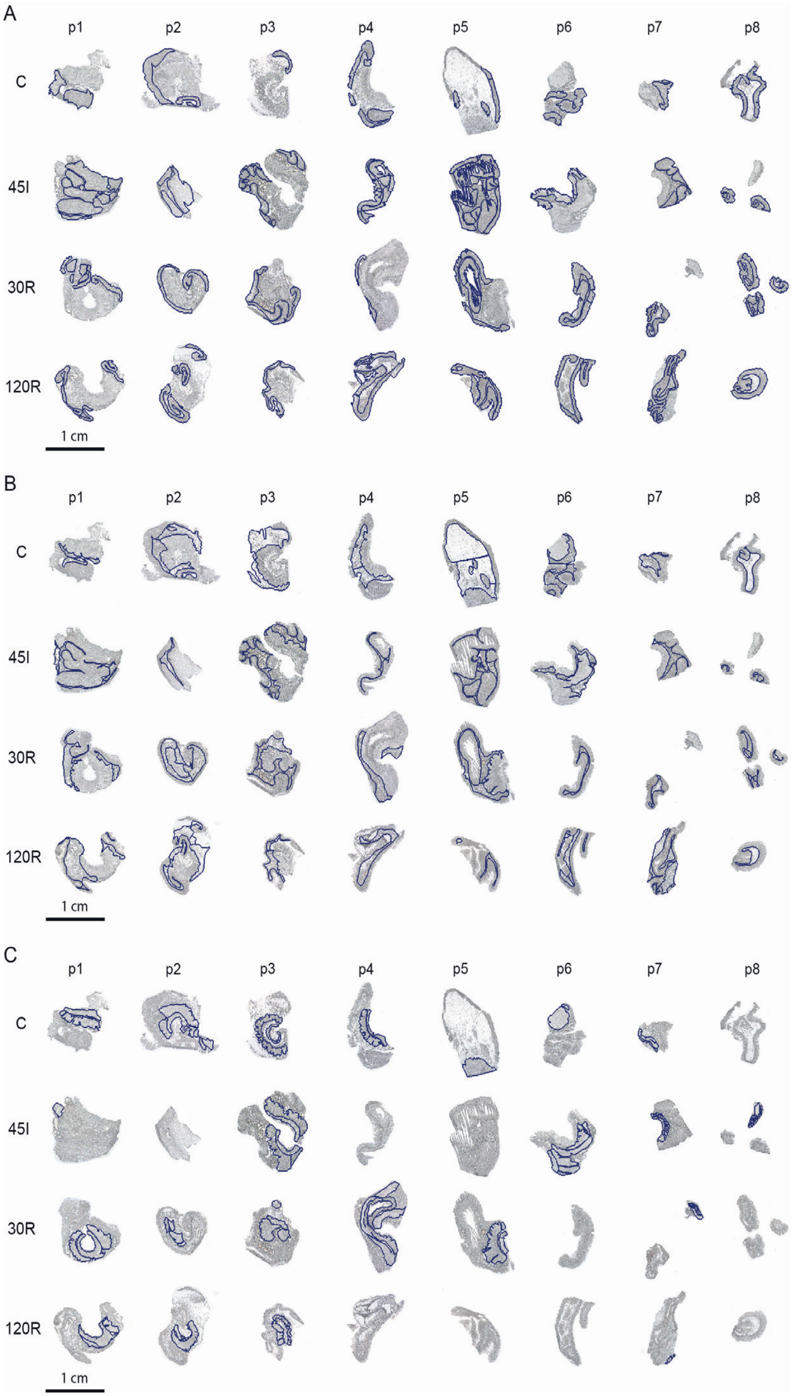


Figure S7. Annotation of histological regions in all samples analyzed with MALDI-TOF MSI. (A) Mucosa regions (B) Submucosa Regions (C) Muscle regions.

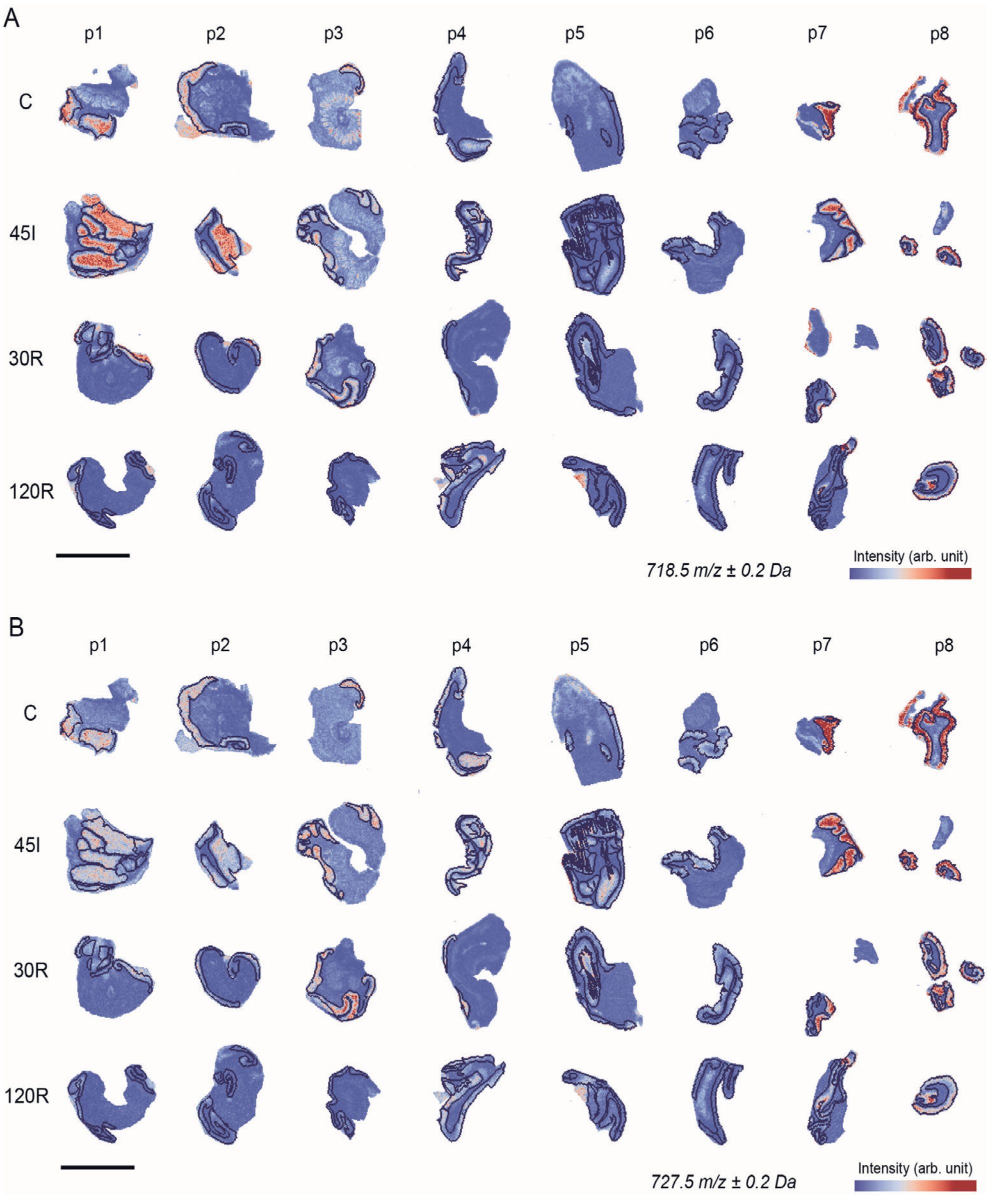


Figure S8. Individual peptide images showing examples of significantly changed m/z in mucosa layer with a high fold change (all examples in top 10 highest fold change) (A, B). Figure continues on next page

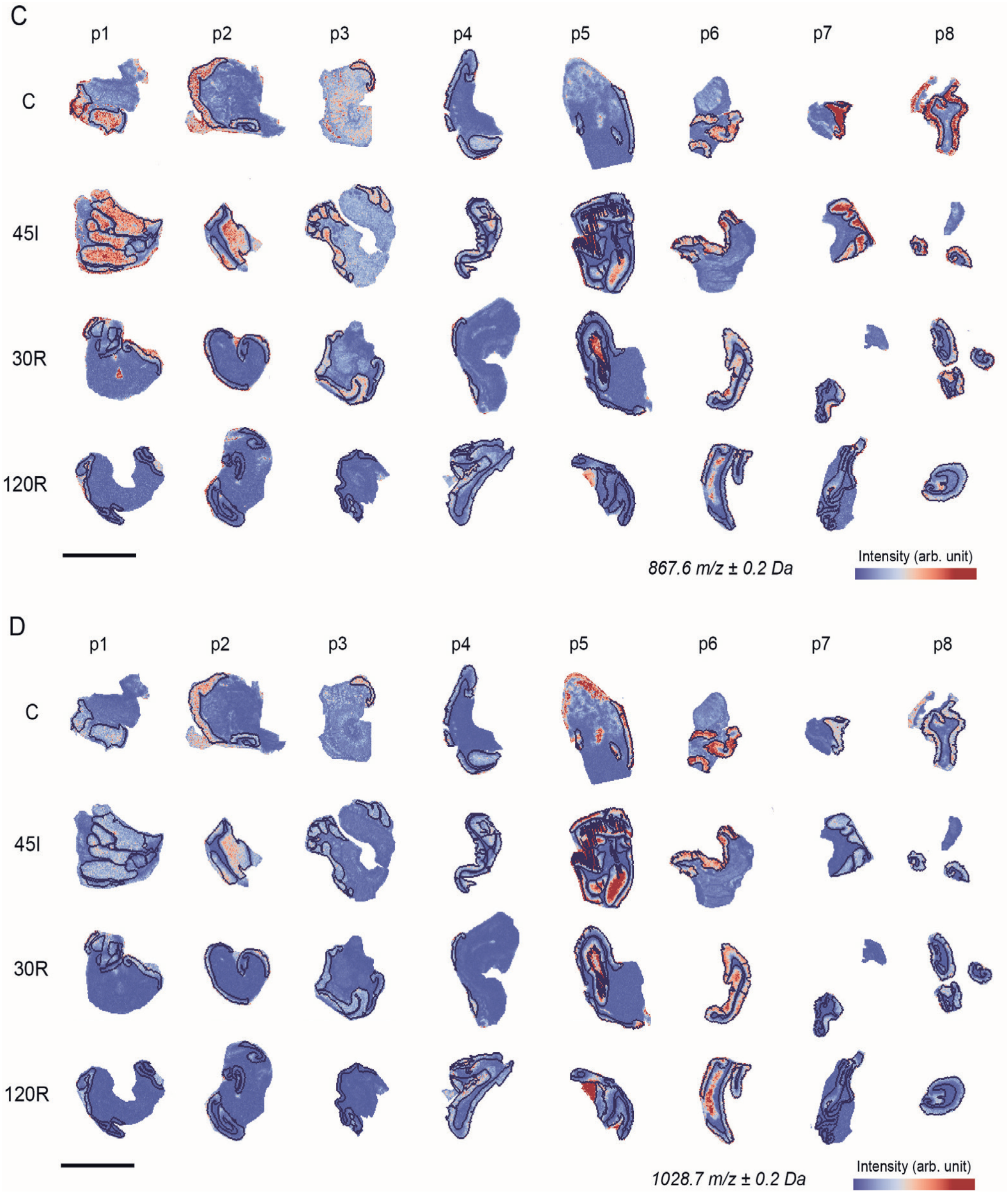


Figure S8 continued. (C, D) Individual peptide images showing examples of significantly changed m/z in mucosa layer with a high fold change (all in top 10 highest fold change), and also among top 10 highest intensity m/z values.

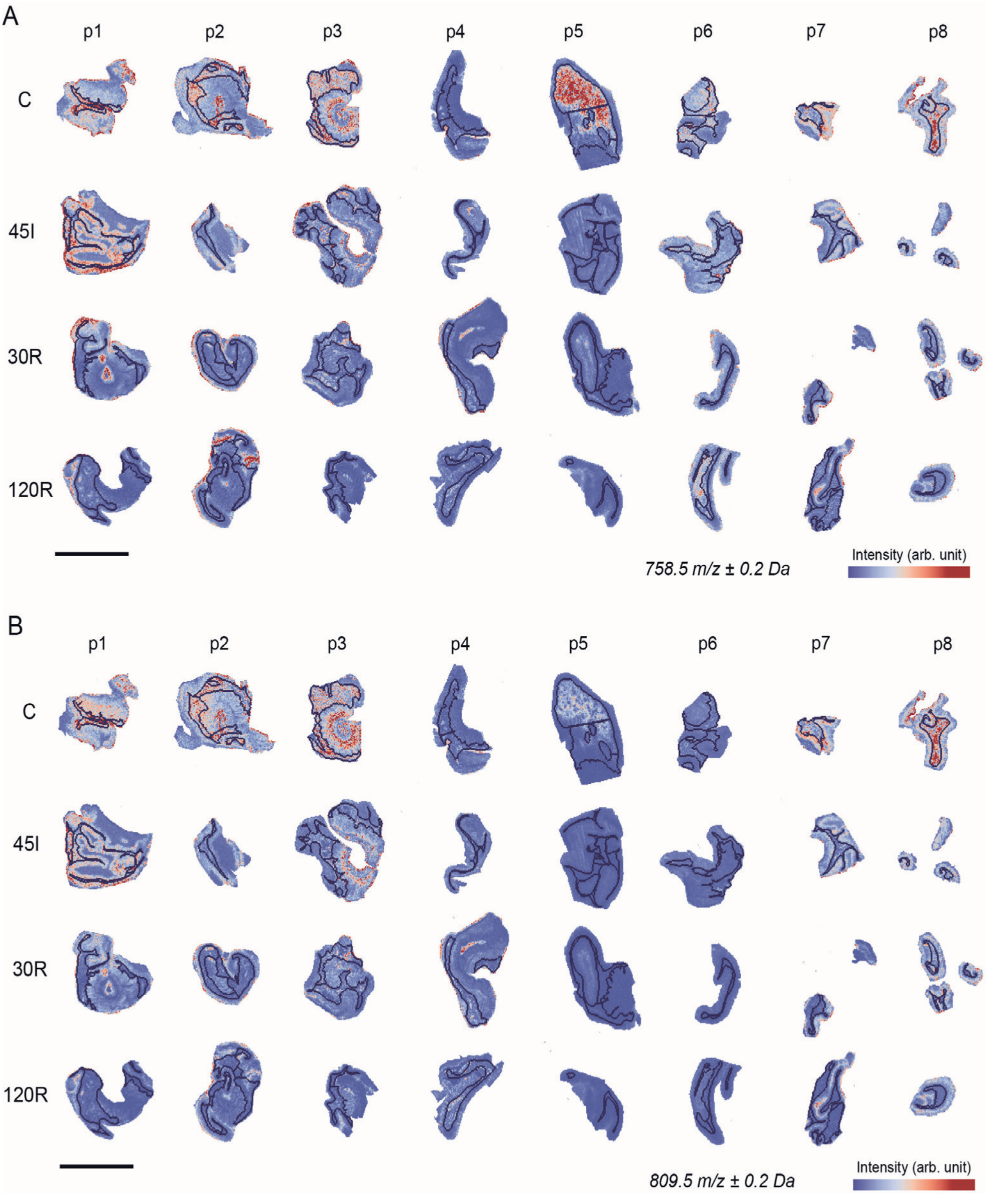


Figure S9. Individual peptide images showing examples of significantly changed m/z in submucosa layer with a high fold change (A, B); in top 10 highest fold change. Figure continues on next page

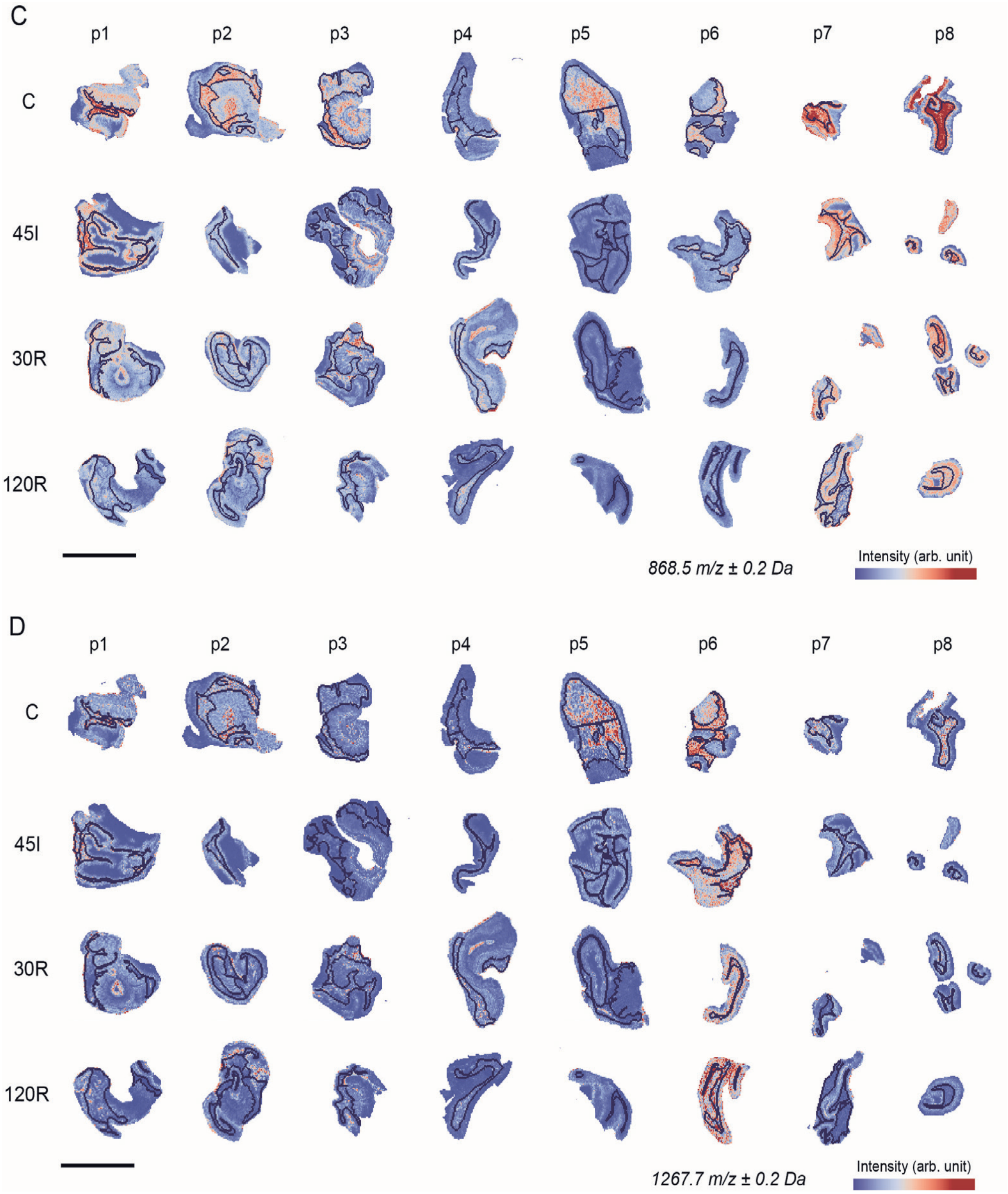


Figure S9 continued. (C, D) Individual peptide images showing examples of significantly changed m/z in submucosa layer with a high abundance (top 3 highest intensity in Ctrl sample among samples with a fold change >2).

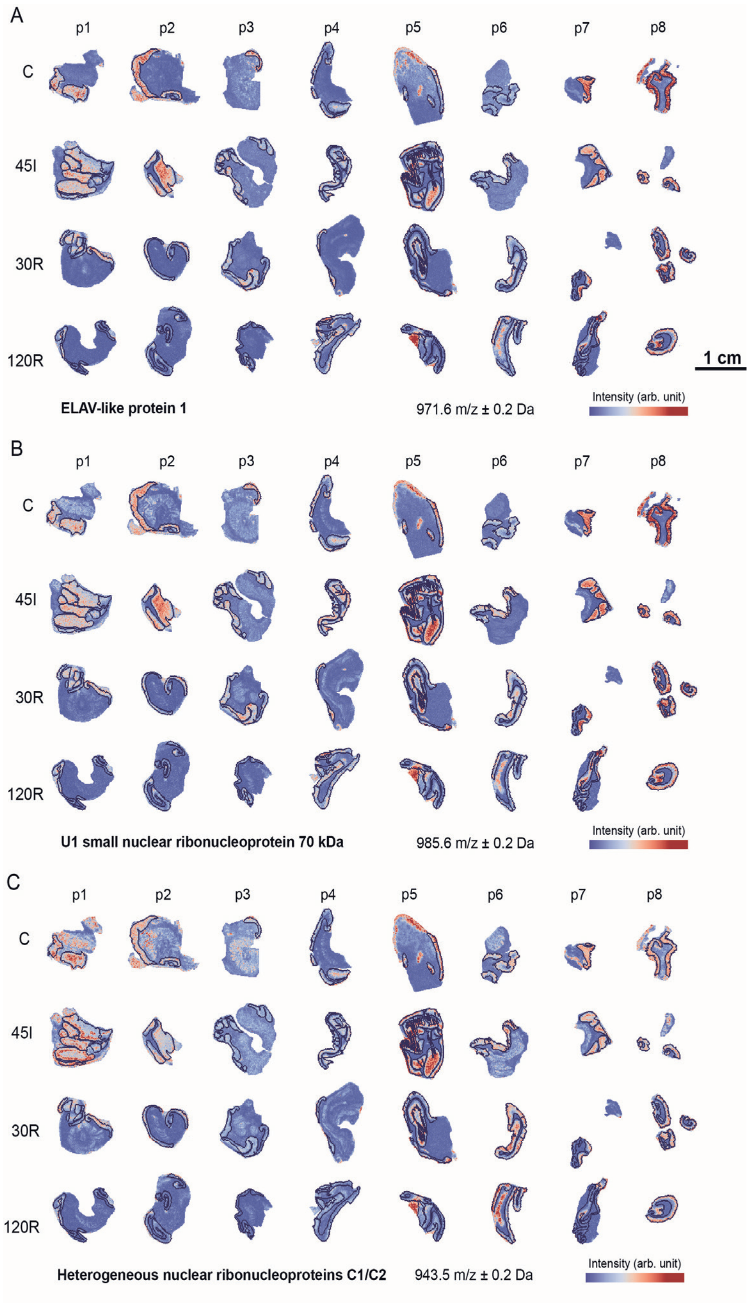


Figure S10. Individual peptide images for a selection of annotated proteins using MALDI-TOF combined with FTICR and LC-MS/MS. Proteins related to RNA splicing and processing are located in the mucosa layer (A) ELAV like protein 1 (ELAV1), (B) small nuclear ribonucleoprotein U1 subunit 70 (SNRNP70), (C) heterogeneous nuclear ribonucleoprotein C (HNRNPC). Figure continues on next page

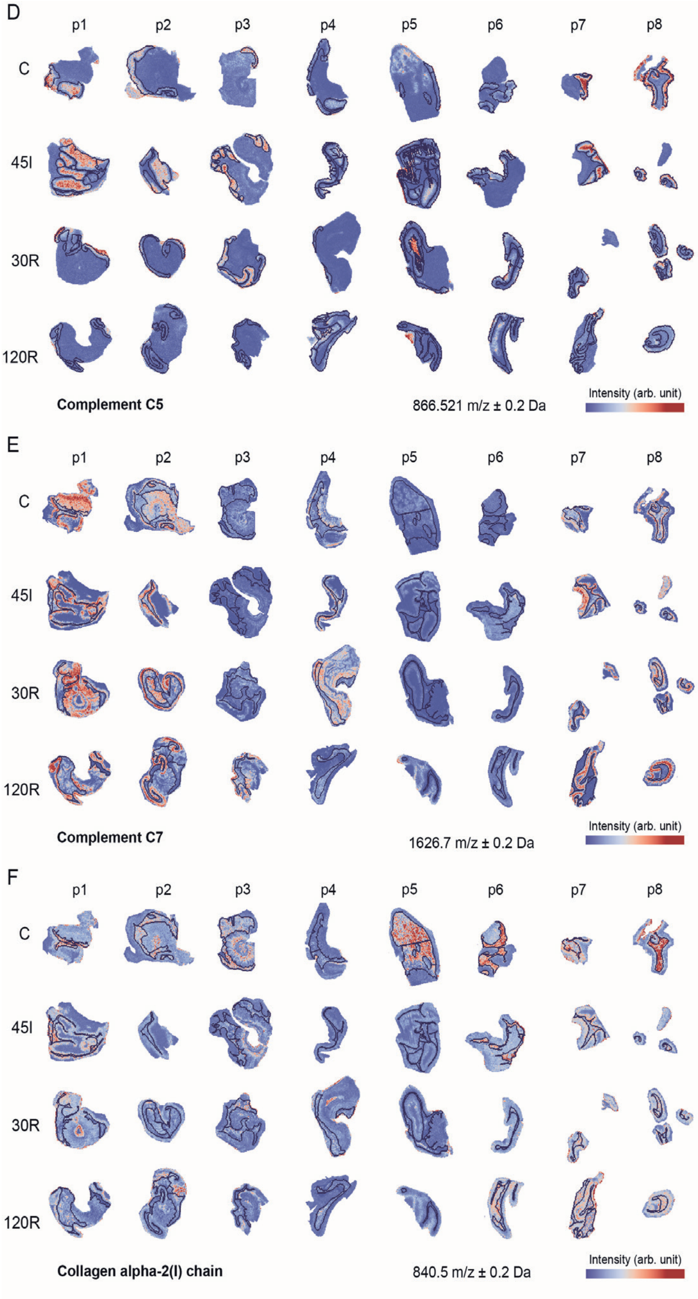


Figure S10 continued. (D, E) Complement factors C5 and C7 located in mucosa and submucosa layer, and (F) collagen 1 alpha2 (I) chain (COL1A2) located in the submucosa layer.

Acknowledgments

This work was funded by NUTRIM, Maastricht University Medical Centre+ (NUTRIM Graduate Program grant to Anna M. Kip), NWO (Aspasia grant 015.010.046 to Kaatje Lenaerts), NWO (project 184.034.019), and was part of the Netherlands X-omics Initiative. Juan Manuel Valverde was supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR). We thank the surgical team of Maastricht University Medical Centre+ for their excellent surgical assistance.

References

1. American Gastroenterological Association Medical Position Statement: guidelines on intestinal ischemia. *Gastroenterology* **118**, 951-953 (2000).
2. M. Bala *et al.*, Acute mesenteric ischemia: guidelines of the World Society of Emergency Surgery. *World journal of emergency surgery : WJES* **12**, 38 (2017).
3. W. A. Oldenburg, L. L. Lau, T. J. Rodenberg, H. J. Edmonds, C. D. Burger, Acute mesenteric ischemia: a clinical review. *Archives of internal medicine* **164**, 1054-1062 (2004).
4. T. Kalogeris, C. P. Baines, M. Krenz, R. J. Korthuis, Ischemia/Reperfusion. *Compr Physiol* **7**, 113-170 (2016).
5. J. P. Derikx *et al.*, Rapid reversal of human intestinal ischemia-reperfusion induced damage by shedding of injured enterocytes and reepithelialisation. *PloS one* **3**, e3428 (2008).
6. J. Grootjans *et al.*, Rapid lamina propria retraction and zipper-like constriction of the epithelium preserves the epithelial lining in human small intestine exposed to ischaemia-reperfusion. *J Pathol* **224**, 411-419 (2011).
7. J. Grootjans *et al.*, Human intestinal ischemia-reperfusion-induced inflammation characterized: experiences from a new translational model. *The American journal of pathology* **176**, 2283-2291 (2010).
8. J. Grootjans *et al.*, Level of activation of the unfolded protein response correlates with Paneth cell apoptosis in human small intestine exposed to ischemia/reperfusion. *Gastroenterology* **140**, 529-539.e523 (2011).
9. W. T. Kassahun, T. Schulz, O. Richter, J. Hauss, Unchanged high mortality rates from acute occlusive intestinal ischemia: six year review. *Langenbeck's archives of surgery* **393**, 163-171 (2008).
10. A. Mastoraki *et al.*, Mesenteric ischemia: Pathogenesis and challenging diagnostic and therapeutic modalities. *World journal of gastrointestinal pathophysiology* **7**, 125-130 (2016).
11. P. Kougias *et al.*, Determinants of mortality and treatment outcome following surgical interventions for acute mesenteric ischemia. *Journal of vascular surgery* **46**, 467-474 (2007).
12. A. F. Altelaar, J. Munoz, A. J. Heck, Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35-48 (2013).
13. H. Liu *et al.*, A Caffeic Acid Matrix Improves In Situ Detection and Imaging of Proteins with High Molecular Weight Close to 200,000 Da in Tissues by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *Analytical chemistry* **93**, 11920-11928 (2021).
14. J. Han *et al.*, Imaging of protein distribution in tissues using mass spectrometry: An interdisciplinary challenge. *TrAC Trends in Analytical Chemistry* **112**, 13-28 (2019).

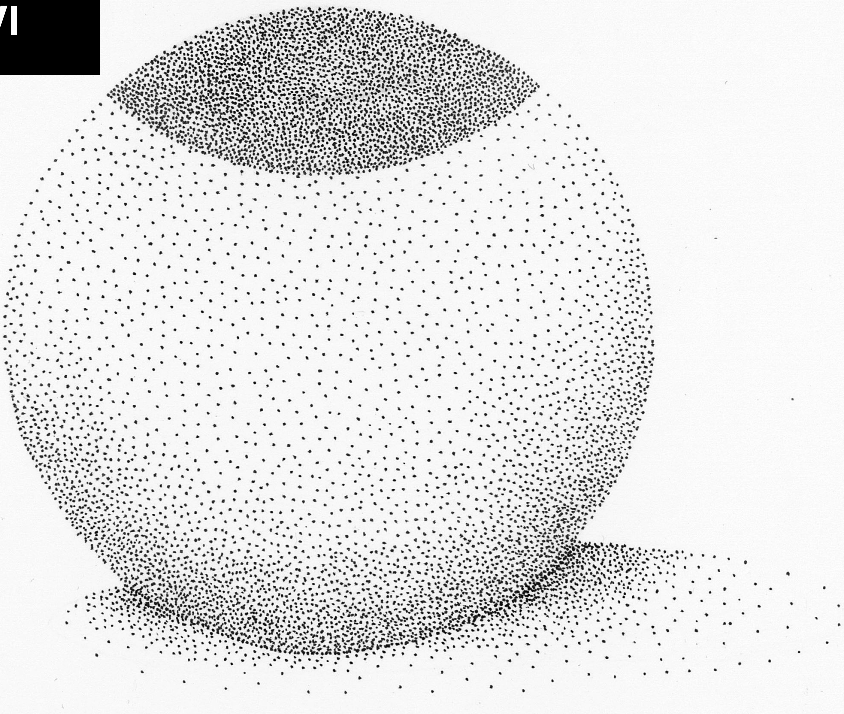
15. K. Huber *et al.*, Approaching cellular resolution and reliable identification in mass spectrometry imaging of tryptic peptides. *Analytical and bioanalytical chemistry* **410**, 5825-5837 (2018).
16. Y. Schober, T. Schramm, B. Spengler, A. Rompp, Protein identification by accurate mass matrix-assisted laser desorption/ionization imaging of tryptic peptides. *Rapid communications in mass spectrometry : RCM* **25**, 2475-2483 (2011).
17. J. M. Spraggins *et al.*, Next-generation technologies for spatial proteomics: Integrating ultra-high speed MALDI-TOF and high mass resolution MALDI FTICR imaging mass spectrometry for protein analysis. *Proteomics* **16**, 1678-1689 (2016).
18. D. R. N. Vos *et al.*, Strategies for managing multi-patient 3D mass spectrometry imaging data. *J Proteomics* **193**, 184-191 (2019).
19. G. D. Pavitt, D. Ron, New insights into translational regulation in the endoplasmic reticulum unfolded protein response. *Cold Spring Harb Perspect Biol* **4**, a012278 (2012).
20. R. Rosenzweig, N. B. Nillegoda, M. P. Mayer, B. Bukau, The Hsp70 chaperone network. *Nature Reviews Molecular Cell Biology* **20**, 665-680 (2019).
21. A. M. Kip *et al.*, Temporal transcript profiling identifies a role for unfolded protein stress in human gut ischemia-reperfusion injury. *Cell Mol Gastroenterol Hepatol.*, https://doi.org/10.1016/j.jcmgh.2021.1011.1001 (in press).
22. M. Kodiha, A. Chu, N. Matusiewicz, U. Stochaj, Multiple mechanisms promote the inhibition of classical nuclear import upon exposure to severe oxidative stress. *Cell Death Differ* **11**, 862-874 (2004).
23. T. Mahmoudi *et al.*, The kinase TNIK is an essential activator of Wnt target genes. *Embo j* **28**, 3329-3340 (2009).
24. D. Delacour, J. Salomon, S. Robine, D. Louvard, Plasticity of the brush border - the yin and yang of intestinal homeostasis. *Nat Rev Gastroenterol Hepatol* **13**, 161-174 (2016).
25. S. Lechuga, A. I. Ivanov, Disruption of the epithelial barrier during intestinal inflammation: Quest for new molecules and mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1864**, 1183-1194 (2017).
26. E. E. Tansey *et al.*, Reduction and redistribution of gap and adherens junction proteins after ischemia and reperfusion. *Ann Thorac Surg* **82**, 1472-1479 (2006).
27. T. Kanda *et al.*, Intestinal fatty acid-binding protein as a sensitive marker of intestinal ischemia. *Dig Dis Sci* **37**, 1362-1367 (1992).
28. T. Kanda *et al.*, Diagnosis of ischemic small bowel disease by measurement of serum intestinal fatty acid-binding protein in patients with acute abdomen: a multicenter, observer-blinded validation study. *J Gastroenterol* **46**, 492-500 (2011).
29. D. H. Schellekens *et al.*, Plasma intestinal fatty acid-binding protein levels correlate with morphologic epithelial intestinal damage in a human translational ischemia-reperfusion model. *J Clin Gastroenterol* **48**, 253-260 (2014).
30. T. V. Arumugam, I. A. Shiels, T. M. Woodruff, D. N. Granger, S. M. Taylor, The role of the complement system in ischemia-reperfusion injury. *Shock* **21**, 401-409 (2004).
31. K. Wada, M. C. Montalto, G. L. Stahl, Inhibition of complement C5 reduces local and remote organ injury after intestinal ischemia/reperfusion in the rat. *Gastroenterology* **120**, 126-133 (2001).
32. S. D. Fleming *et al.*, C5a causes limited, polymorphonuclear cell-independent, mesenteric ischemia/reperfusion-induced injury. *Clin Immunol* **108**, 263-273 (2003).
33. T. V. Arumugam *et al.*, Protective effect of a new C5a receptor antagonist against ischemia-reperfusion injury in the rat small intestine. *J Surg Res* **103**, 260-267 (2002).
34. F. Dewez *et al.*, MS Imaging-Guided Microproteomics for Spatial Omics on a Single Instrument. *Proteomics* **20**, e1900369 (2020).

35. S. T. P. Mezger, A. M. A. Mingels, O. Bekers, R. M. A. Heeren, B. Cillero-Pastor, Mass Spectrometry Spatial-Omics on a Single Conductive Slide. *Analytical chemistry* **93**, 2527-2533 (2021).
36. A. M. Kip *et al.*, Proteomics analysis of human intestinal organoids during hypoxia and reoxygenation as a model to study ischemia-reperfusion injury. *Cell Death Dis* **12**, 95 (2021).
37. J. H. Mortensen *et al.*, The intestinal tissue homeostasis - the role of extracellular matrix remodeling in inflammatory bowel disease. *Expert Rev Gastroenterol Hepatol* **13**, 977-993 (2019).
38. A. C. Petrey, C. A. de la Motte, The extracellular matrix in IBD: a dynamic mediator of inflammation. *Curr Opin Gastroenterol* **33**, 234-238 (2017).
39. S. F. Tian *et al.*, Mechanisms of neuroprotection from hypoxia-ischemia (HI) brain injury by up-regulation of cytoglobin (CYGB) in a neonatal rat model. *J Biol Chem* **288**, 15988-16003 (2013).
40. H. Nishi *et al.*, Cytoglobin, a novel member of the globin family, protects kidney fibroblasts against oxidative stress under ischemic conditions. *The American journal of pathology* **178**, 128-139 (2011).
41. Y. Jia *et al.*, Metformin protects against intestinal ischemia-reperfusion injury and cell pyroptosis via TXNIP-NLRP3-GSDMD pathway. *Redox Biol* **32**, 101534 (2020).
42. P. Cohen, The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur J Biochem* **268**, 5001-5010 (2001).
43. M. Malumbres, Cyclin-dependent kinases. *Genome Biology* **15**, 122 (2014).
44. Z. Chen *et al.*, MAP kinases. *Chem Rev* **101**, 2449-2476 (2001).
45. R. Roskoski, Jr., ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacol Res* **66**, 105-143 (2012).
46. S. Morton, R. J. Davis, P. Cohen, Signalling pathways involved in multisite phosphorylation of the transcription factor ATF-2. *FEBS Lett* **572**, 177-183 (2004).
47. S. Gupta, D. Campbell, B. Derijard, R. J. Davis, Transcription factor ATF2 regulation by the JNK signal transduction pathway. *Science* **267**, 389-393 (1995).
48. C. Livingstone, G. Patel, N. Jones, ATF-2 contains a phosphorylation-dependent transcriptional activation domain. *EMBO J* **14**, 1785-1797 (1995).
49. H. van Dam *et al.*, ATF-2 is preferentially activated by stress-activated protein kinases to mediate c-jun induction in response to genotoxic agents. *EMBO J* **14**, 1798-1811 (1995).
50. H. Morooka, J. V. Bonventre, C. M. Pombo, J. M. Kyriakis, T. Force, Ischemia and reperfusion enhance ATF-2 and c-Jun binding to cAMP response elements and to an AP-1 binding site from the c-jun promoter. *J Biol Chem* **270**, 30084-30092 (1995).
51. D. M. Ouwens *et al.*, Growth factors can activate ATF2 via a two-step mechanism: phosphorylation of Thr71 through the Ras-MEK-ERK pathway and of Thr69 through RalGDS-Src-p38. *EMBO J* **21**, 3782-3793 (2002).
52. S. Ferrari, H. R. Bandi, J. Hofsteenge, B. M. Bussian, G. Thomas, Mitogen-activated 70K S6 kinase. Identification of in vitro 40 S ribosomal S6 phosphorylation sites. *J Biol Chem* **266**, 22770-22775 (1991).
53. H. R. Bandi, S. Ferrari, J. Krieg, H. E. Meyer, G. Thomas, Identification of 40 S ribosomal protein S6 phosphorylation sites in Swiss mouse 3T3 fibroblasts stimulated with serum. *J Biol Chem* **268**, 4530-4533 (1993).
54. P. P. Roux *et al.*, RAS/ERK signaling promotes site-specific ribosomal protein S6 phosphorylation via RSK and stimulates cap-dependent translation. *J Biol Chem* **282**, 14056-14064 (2007).
55. D. Ledee *et al.*, Quantitative cardiac phosphoproteomics profiling during ischemia-reperfusion in an immature swine model. *Am J Physiol Heart Circ Physiol* **313**, H125-h137 (2017).
56. T. Misteli, RNA splicing: What has phosphorylation got to do with it? *Curr Biol* **9**, R198-200 (1999).

57. J. Prasad, K. Colwill, T. Pawson, J. L. Manley, The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol Cell Biol* **19**, 6991-7000 (1999).
58. C. Naro, C. Sette, Phosphorylation-mediated regulation of alternative splicing in cancer. *Int J Cell Biol* **2013**, 151839 (2013).
59. E. T. Wang *et al.*, Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476 (2008).
60. S. Natua, C. Ashok, S. Shukla, Hypoxia-induced alternative splicing in human diseases: the pledge, the turn, and the prestige. *Cell Mol Life Sci* **78**, 2729-2747 (2021).
61. A. Kanopka, Cell survival: Interplay between hypoxia and pre-mRNA splicing. *Exp Cell Res* **356**, 187-191 (2017).
62. K. Nakayama, N. Kataoka, Regulation of Gene Expression under Hypoxic Conditions. *Int J Mol Sci* **20**, 3278 (2019).
63. A. M. Al-Ayoubi, H. Zheng, Y. Liu, T. Bai, S. T. Eblen, Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion. *Mol Cell Biol* **32**, 2880-2893 (2012).
64. C. Bertocchi, M. Vaman Rao, R. Zaidel-Bar, Regulation of adherens junction dynamics by phosphorylation switches. *J Signal Transduct* **2012**, 125295 (2012).
65. N. Martinez-Quiles, H. Y. Ho, M. W. Kirschner, N. Ramesh, R. S. Geha, Erk/Src phosphorylation of cortactin acts as a switch on-switch off mechanism that controls its ability to activate N-WASP. *Mol Cell Biol* **24**, 5269-5280 (2004).
66. L. C. Kelley, K. E. Hayes, A. G. Ammer, K. H. Martin, S. A. Weed, Cortactin phosphorylated by ERK1/2 localizes to sites of dynamic actin regulation and is required for carcinoma lamellipodia persistence. *PLoS One* **5**, e13847 (2010).
67. N. Larsson, U. Marklund, H. M. Gradin, G. Brattsand, M. Gullberg, Control of microtubule dynamics by oncoprotein 18: dissection of the regulatory role of multisite phosphorylation during mitosis. *Mol Cell Biol* **17**, 5530-5539 (1997).
68. M. Chang, E. B. Kistler, G. W. Schmid-Schönbein, Disruption of the mucosal barrier during gut ischemia allows entry of digestive enzymes into the intestinal wall. *Shock* **37**, 297-305 (2012).
69. M. Chang, T. Alsaigh, E. B. Kistler, G. W. Schmid-Schönbein, Breakdown of mucin as barrier to digestive enzymes in the ischemic rat small intestine. *PloS one* **7**, e40087 (2012).
70. H. Zhou *et al.*, Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* **8**, 461-480 (2013).
71. H. D. Meiring, E. van der Heeft, G. J. ten Hove, A. P. J. M. de Jong, Nanoscale LC-MS(n): technical design and applications to peptide and protein analysis. *J Sep Sci* **25**, 557-568 (2002).
72. R. Tibshirani, G. Walther, T. Hastie, Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 411-423 (2001).
73. D. Szklarczyk *et al.*, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
74. F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
75. V. Vacic, L. M. Iakoucheva, P. Radivojac, Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* **22**, 1536-1537 (2006).
76. A. Lachmann, A. Ma'ayan, KEA: kinase enrichment analysis. *Bioinformatics* **25**, 684-686 (2009).

77. K. Krug *et al.*, A Curated Resource for Phosphosite-specific Signature Analysis*[S]. *Molecular & Cellular Proteomics* **18**, 576-593 (2019).
78. M. Strohalm, M. Hassman, B. Kosata, M. Kodicek, mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid communications in mass spectrometry : RCM* **22**, 905-908 (2008).
79. Y. Perez-Riverol *et al.*, The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-d450 (2019).

CHAPTER VI



Conclusions and outlook

Protein phosphorylation is the most studied PTM, known to influence almost every cellular process studied up to date. MS-based technologies enabled the mapping of hundreds of thousands of phosphorylation sites, of which only a very small proportion have a known function and a defined upstream kinase (≈2%) (*1*). Gaining insight into the functionality of this 'dark phosphoproteome' (as coined by Needham and colleagues) is a major challenge for the scientific community (*1*). Here, I argue that even though MS prompted an explosion in the mapping of phosphosites, a change in strategy is needed to transform all the information gathered so far into mechanistic insight on how phosphorylation regulates biological processes. First, I will address the upside of combining shotgun and targeted MS approaches for the study of phosphoproteomes, and how this can greatly benefit data acquisition and interpretation. Next, I will highlight the need to migrate to experimental models that are 'single-cell-type', and the relevance of studying protein phosphorylation throughout time. Moreover, I will address the importance of assessing protein phosphorylation at the subcellular level and the efforts currently being made to tackle this challenging hurdle. In addition, I will focus on the importance of considering the structural features of phosphorylated regions, and how this is a useful guide when assessing the potential functionality of phosphorylation sites. Finally, I propose that MS specialists in the field of phosphoproteomics should unite forces with experts in other areas, yielding the multi-disciplinary approach needed to advance the frontiers of knowledge in the study of protein phosphorylation.

The best of both worlds: combining shotgun and targeted proteomics to study protein phosphorylation

Shotgun proteomics has established itself as a widely used technique in the field of cellular and molecular biology (*2*). When it comes to its application in the study of protein phosphorylation, it has shown unparalleled ability for global identification of phosphosites. The latter comes accompanied by drawbacks, such as the prevalence of missing values due to the stochastic nature of precursor ion selection methods, and the intrinsically transient nature of protein phosphorylation. These features combined hamper robust analyte quantification across samples. Here, targeted proteomics proves to be a great alternative to study specific phosphorylation sites with enhanced sensitivity and reproducibility. First, by using heavy labeled standards we add another layer of confidence to the measurements, aiding in the interpretation of missing values and improving analyte quantification. Furthermore, targeting specific phosphosites with a known biological function greatly improves the extraction of biological insight from the data. An example of this approach was **chapter II**, in which we developed a targeted proteomics assay to measure phosphorylation sites from the MAPK-AKT-mTOR pathway. With a very simple starvation experiment, we showed that our assay is able detect upregulation of AKT signaling in response to amino acid or glucose starvation. The latter shows how phosphosites with known functionality can be used as proxies to assess activation of specific signaling pathways.

Next, we combined targeted and shotgun phosphoproteomics, which provided us with another layer of information that improved data interpretation. In **chapter IV**, we confirmed the switch-like nature of CDK1 derived phosphorylation at mitotic onset. We achieved this by measuring some of the phosphosites detected with shotgun proteomics by PRM, which provided greater reproducibility and it allowed us to measure at a higher time-resolution. In **chapter III**, we used SRM to determine the phosphorylation state of the MAPK-AKT-mTOR pathway, and to assess the kinome activation state across cell lines that lacked either ERK1 or ERK2. The latter showed that ERK2 depletion leads to activation of several kinases as a compensatory response, and partially explained the phosphorylation signature observed in the shotgun proteomics data. Overall, both of these chapters show that these acquisition methods complement each other, and using them in parallel has the potential to circumvent some of the limitations observed when using shotgun phosphoproteomics alone. This strategy has already been implemented by others, for example, the Villen lab used shotgun phosphoproteomics to study downstream effectors of mTORC2. Depletion of the regulating subunit RICTOR, revealed changes in the phosphoproteome that could be linked to downregulation of AKT signaling, which was assessed by PRM measurement of relevant phosphosites (*3*). Similarly, a colleague in our lab used shotgun phosphoproteomics to study the cellular response to stimulation of glutamate receptors, followed by SRM measurements that could pinpoint ERK2, PKCa and CAMKIIA as the potential effectors of upregulated phosphorylation (*4*). The previous examples show the benefits of combining these acquisition methods; shotgun proteomics allows global evaluation of the phosphoproteome, while targeted measurements aid in the further confirmation and interpretation of the results.

Capturing real phosphorylation dynamics

The rise of single cell technologies for the study of transcriptomes has produced a vast amount of data, displaying the relevance of cell heterogeneity and the need to perform single cell measurements to better understand cellular processes (*5, 6*). The field of proteomics has lagged behind this technological surge, partially because MS-based techniques do not make use of an amplification step, as sequencing techniques of nucleic acids do (*7*). This has led to a sensitivity hurdle, which is currently being tackled mainly by improvements in the sample preparation workflow (*8, 9*), but also by enhancing the data acquisition methods (*10-12*). Some of these improvements enabled the measurement of the first proteomes originating from very small groups of cells, and even from single cells (*13*). These efforts have allowed mapping of changes in protein expression throughout the cell cycle (*10, 14*) and during monocyte differentiation (*15*) at a single cell level.

Despite these advances, the measurement of protein phosphorylation on this type of samples remains elusive, mainly due to the relatively low stoichiometry of this PTM. This is likely to be circumvented in the near future, when specialized phosphopeptide enrichment techniques are incorporated into the sample preparation workflow of specimens that yield low protein amounts (e.g. single somatic cell). An important hurdle remaining is the manipulation of single cells (*7*). The latter is still challenging, requiring the use of fluorescence activated cell-sorting (FACS) or laser capture micro-dissection for cell isolation (*10, 16, 17*). These not only require direct manipulation of the cells which might introduce artifacts, but entail handling time during which one might miss the real phosphorylation dynamics. Single cell sorters and specialized liquid handling equipment that enable to perform experiments on multi-well plates are needed to address these issues.

For the previous reasons, we decided to use an experimental model that got us as close as possible to the measurement of real phosphorylation dynamics on single cells *in vivo*. In **chapter IV** we made use of the *Xenopus laevis* eggs, which yield a protein amount (~ 20 µg) compatible with our current sample preparation workflows for phosphoproteomics (*18*). *In vitro* fertilization of these eggs leads to embryonic cell divisions, which are highly synchronous, allowing us to capture phosphorylation waves as the embryo divided. Others have used similar models, in which cell synchronization is easily achieved, especially in the case of gamete activation (*19-21*). Still, according to our knowledge, our report is the first to use a single specimen (e.g. one egg or embryo) as sample, and not a population of specimens. This way, we managed to circumvent issues linked to the use of populations of cells, where the averaged behavior of the population translates into a distorted version of the real single cell dynamics (*22*). The latter causes phosphorylation events that are switch-like, to be detected as gradual increments, due to the differences in timing amongst the population of cells (Figure 1). Moreover, by previously knowing the timescale of the events we were interested in studying (fertilization and cell divisions), we could acquire samples at relevant time points, which allowed us to capture the real dynamic behavior of phosphosites. We think this type of experimental models provide a glimpse into what phosphorylation dynamics in single cells look like and should be exploited until technological improvements allow phosphoproteomics on single somatic cells.

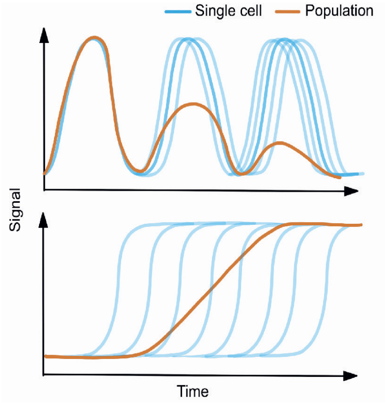


Figure 1. Averaged signal from a cell population can obscure real single cell dynamics. Unsynchronized pulses upon average give the appearance of damped oscillations (top). Variable delay from cell to cell in a population can be interpreted as a gradual increase, while the real dynamics are switch-like (bottom). Modified from Purvis & Lahav, 2013 (*22*).

The need for spatial resolution

As we showed in **chapter V**, proteins show differential spatial distribution within a tissue. Furthermore, we know that proteins also compartmentalize differently inside the cell, mainly due to divisions mediated by phospholipid membranes or by formation of membrane-less organelles (see next section). If the majority of proteins are prone to this type of segregation, one can only speculate the same happens to kinases and phosphatases, delimiting their corresponding activities to certain sections of the cell. To explore this, the Olsen group recently applied a cell fractionation workflow followed by dynamic phosphoproteome profiling, which revealed protein phosphorylation signatures on different subcellular compartments. When applying their workflow to study osmotic stress, they observed relocation of large ribosomal subunit proteins from the cytosol to the nucleus, which also correlated with changes in their phosphorylation status (*23*).

Despite finding interesting differences in the distribution of phosphorylation signals, the previous approach relies on large cell populations subjected to a fractionation protocol. On the other hand, an ideal approach would assess differences in protein phosphorylation at a subcellular level directly in single cells. For instance, the Mann group is currently developing an elegant pipeline termed Deep Visual Proteomics (*24*), combining high-end microscopy and AI-based image analysis, which enables identification and isolation of single cells from tissues that are subsequently subjected to ultrasensitive MS analysis (*10*). They managed to isolate nuclei from single cells and showed how microscopic differences correlated with quantitative changes in protein expression detected by MS (*24*). Even though the authors do not report analysis of protein phosphorylation, we envision that this type of pipelines will be improved in the future, by coupling phosphopeptide enrichment techniques that enable assessment of protein phosphorylation at a subcellular level on single somatic cells.

Despite the promising potential of Deep Visual Proteomics, this type of workflow requires cutting-edge instrumentation, from high-end microscopes to state-of-the-art mass spectrometers. Since the resources and expertise required to apply this pipeline are not widespread, other approaches that are simpler, yet equally resourceful should be exploited. For instance, the Nemes lab also adopted the *Xenopus laevis* embryo as an experimental model, on which they have applied *in situ* subcellular capillary microsampling, followed by one-pot extraction and digestion of proteins. Peptides are subsequently separated by capillary electrophoresis and measured by high-resolution MS (*25, 26*). This microsampling technique permits to retrieve small samples (~5 ng) from specific locations within the cell, and can be done in a swiftly manner (<5 seconds). This makes it a promising sampling technique to assess phosphorylation dynamics on single cells at a subcellular level. For instance, it could be exploited to measure the dynamics of mitotic phosphorylation waves as they spread throughout the cell (*27*).

Protein phosphorylation, intrinsic disorder and phase separation

Protein phosphorylation is biased towards intrinsically disordered regions (IDRs) of proteins, as reported in **chapter IV** and previously by others. (*28, 29*). IDRs are defined as those regions that do not acquire a defined three-dimensional, globular structure under physiological conditions (*30-32*). When using nuclear magnetic resonance (NMR) to study proteins in solution, IDRs are described as a dynamic ensemble of different conformational states, which prompts to think that these regions are highly flexible and mobile (*33-35*). Many speculate that this is the reason why IDRs are so heavily phosphorylated; mobile parts of the protein are more likely to be exposed to kinases, hence are more prone to undergo modification. Even though this statement is potentially accurate, this has also led scientists to think that many of these phosphorylations are spurious, caused by kinase promiscuity and not necessarily functional (*36*). The latter is supported by the fact that IDRs tend to be less conserved than structured regions.

So, are phosphorylation sites in IDRs functional? Here I argue that many of them are. First, it is important to compare the conservation of phosphorylatable amino acids inside IDRs. Residues known to be phosphorylated show a lower evolution rate than those that have not been reported to be modified (*37, 38*). The latter suggests that phosphorylated residues are more than simple bystander victims to kinase promiscuity. Having said that, IDRs still have a higher evolutionary rate than structured regions. So why would functional phosphosites locate to these rapidly evolving disordered regions? To answer this, it is important to consider that phosphorylation can modulate protein function by different means. In the most classic of examples, addition of a phosphate group can have an allosteric effect, changing protein conformation. Such is the case for glycogen phosphorylase, a highly structured protein that upon phosphorylation undergoes a conformational change that causes substrate access to the previously buried catalytic site (*39*). This type of regulation is highly dependent on protein folding, meaning that mutation of the modified site would drastically hamper protein functionality.

In contrast, protein phosphorylation is also able to modulate functionality by controlling bulk electrostatic interactions, which is usually dependent on the concerted regulation of multiple phosphorylation sites (*40*). An elegant example was presented by Strickfaden *et al* in 2007, in which they showed that CDKs regulate the binding of Ste5 (scaffolding protein of the MAPK pathway) to the plasma membrane. They nicely showed that binding of this protein to the membrane is abrogated after phosphorylation of eight residues in the disordered N-terminus of the protein (*41*). Phosphate groups introduce negative charges that disrupt the interactions between the lysine and arginine rich region and the negatively charged phospholipids in the membrane. Mutational analysis also demonstrated that binding depended on the number of phosphate groups, but not on the exact positions of the modified residues (*41*). The latter exhibits a type of protein regulation via phosphorylation drastically different from the one observed in glycogen phosphorylase, since it does not depend on one solely phosphorylation site, and does not involve allosteric regulation.

As mentioned before, phosphorylation sites are largely enriched in IDRs, with disordered stretches frequently presenting multisite phosphorylation. The latter poises IDRs to function as 'electrostatic switches', which could be regulated by kinase and phosphatase activities. This would explain why phosphorylation sites tend to always be present in certain IDRs throughout evolution, without being constrained to specific positions ('conserved'). Furthermore, it would also explain why despite having poorly conserved primary structure, IDRs have preserved specific physicochemical features throughout evolution (*42*). For example, if the net charge of an IDR were to change drastically from one species to another, the effect of the phosphorylation sites could be different, leading to a loss or change in functionality.

David Morgan's Lab first showed this for CDK1 targets in yeast; the vast majority of phosphosites under regulation by CDK1 locate to IDRs, are poorly conserved, and most substrates exhibit multisite phosphorylation (*43*). Interestingly, in this report, the authors highlight that many of the substrates are involved in processes like nuclear transport and chromatin remodeling, which at that moment were not considered to be under cell cycle control. By now, we know that proteins related to these processes locate to nuclear pore complexes (NPCs), splicing speckles, nucleoli and Cajal bodies, all of which are found in the nuclei. These bodies disassemble during mitosis and are thought to form by means of biomolecule phase separation (PS) (*44*). This mechanism involves a variety of intra- and intermolecular interactions driving the formation of subcellular compartments, also known as membrane less organelles (MLOs). The interactions driving PS are varied, but unspecific interactions between IDRs, and especially electrostatic interactions, are considered key drivers in the formation of these compartments (*45, 46*). In **chapter IV**, we used a combination of techniques, showing that phosphorylation of IDRs is likely to regulate PS of CDK substrates, which we think is one of the mechanisms behind the rapid cellular reorganization between interphase and mitosis. For example, multisite phosphorylation of nucleoporins drives the dissolution of NPCs during mitosis (*47, 48*); very likely by disturbing the molecular interactions that held the IDRs of nucleoporins together during interphase (*49-51*). Similar switches are likely to locate in members of the nucleoli, splicing speckles and Cajal bodies, all of which are MLOs that form via PS and rapidly dissolve during mitosis (*52-56*).

Here it is important to stress that during mitosis, PS can also induce the assembly of certain structures like the centromere, the perichromosomal layer and the mitotic spindle (*57-60*). Many of these compartments have proteins that are highly phosphorylated in mitosis, suggesting that phosphorylation can also induce PS of these proteins (*61*). Importantly, protein phosphorylation is unlikely to exclusively abolish or enhance PS, but rather adds a layer of control in a more nuanced way. For example, in chapter IV we suggest that CDK driven phosphorylation could be modulating Ki-67 partitioning between two competing modes of PS, the nucleoli during interphase and the perichromosomal layer during mitosis. A similar model has been described for RNA polymerase II, which upon phosphorylation of its C-terminal domain by regulatory CDKs, switches between two different phase-separated compartments, transcriptional condensates and splicing condensates (*62*).

This type of regulation is likely not to be constrained to CDKs. Since other kinases also present a strong bias towards IDRs, it is expected that many of them also regulate bulk electrostatic interactions that partly drive PS. This has been shown for DYRK3 which acts as a dissolvase of MLOs during mitosis (*63*), but also regulates stress granule assembly (*64*). Similar functions have been observed for MBK-2, the homologue of DYRK3 in *C. elegans*, which controls the disassembly of specialized RNP granules known as germline P-granules in early embryos (*65*). In contrast, mTORC1 kinase has been shown to phosphorylate members of P-granules at more advanced developmental stages promoting PS (*66*). The kinase DNA-PK has been shown to phosphorylate FUS disordered domain, disrupting PS, and more importantly, aberrant protein aggregation, which is known to have cytotoxic effects linked to amyotrophic lateral sclerosis (ALS) (*67*). Interestingly, TDP-43, which is another protein linked to ALS, has been shown to undergo reduced PS and protein aggregation when hyperphosphorylated (*68*), highlighting the potential therapeutic benefits derived from understanding the role of protein phosphorylation in PS. Finally, other cellular bodies suggested to form via PS like postsynaptic densities, tight junctions and signaling clusters have been shown to be under control of protein phosphorylation (*69-71*). Overall, this body of evidence strongly supports a model in which protein phosphorylation (and possibly other PTMs) (*72*), exerts spatiotemporal control over cellular processes by regulating protein-protein interactions responsible for PS.

Final words

The study of protein phosphorylation and its role in cellular biology needs a multi-disciplinary approach. Phosphoproteomics has already carved its niche, applying the powerful technology of MS to identify thousands of phosphorylation sites in a vast ocean of proteins spanning a wide array of different organisms. However, the ceaseless identification of newer phosphosites is proving to be insufficient to derive mechanistic biological insight explaining how this PTM regulates a myriad of diverse cellular processes. Hence, a change in strategy is necessary: first, it is of paramount importance to study protein phosphorylation dynamics on single-cell-type models. Next, in a not so distant future, we should obtain spatial resolution that enables to determine in which exact location of the cell are specific phosphorylation events taking place. In addition, MS specialists interested in protein phosphorylation should build bridges with cellular biologists, bioinformaticians, structural biologists and biophysicists, which can aid in the analysis and interpretation of phosphoproteomics data. The latter will boost the hypothesis-generating process, getting us closer to the actual mechanisms that describe how protein phosphorylation regulates almost every single domain of the cellular life

References

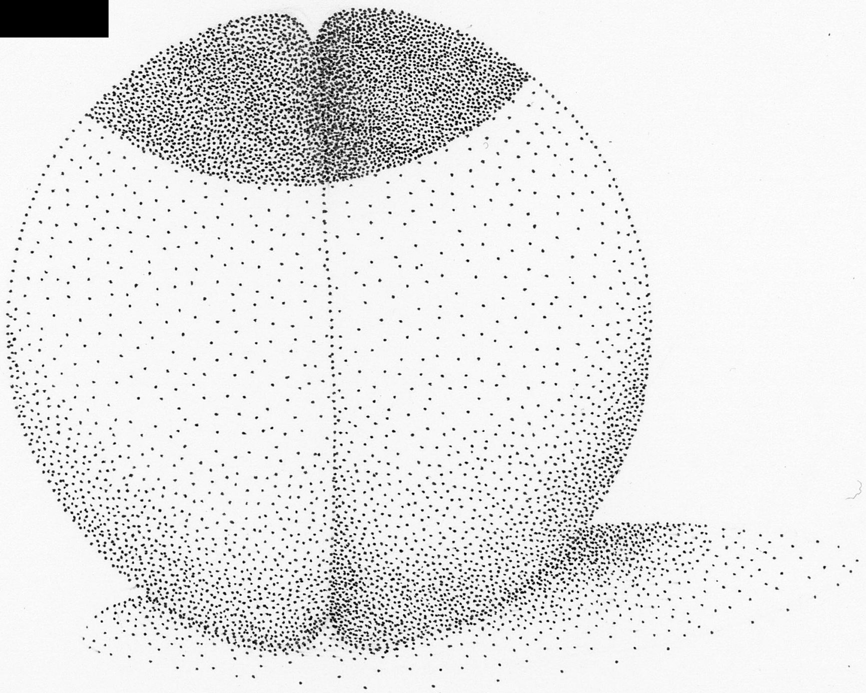
1. E. J. Needham, B. L. Parker, T. Burykin, D. E. James, S. J. Humphrey, Illuminating the dark phosphoproteome. *Sci Signal* **12**, (2019).
2. R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
3. S. W. Entwisle *et al.*, Proteome and Phosphoproteome Analysis of Brown Adipocytes Reveals That RICTOR Loss Dampens Global Insulin/AKT Signaling*. *Molecular & Cellular Proteomics* **19**, 1104-1119 (2020).
4. C. A. G. H. van Gelder *et al.*, Temporal Quantitative Proteomics of mGluR-induced Protein Translation and Phosphorylation in Neurons. *Molecular & Cellular Proteomics* **19**, 1952-1968 (2020).
5. S. Aldridge, S. A. Teichmann, Single cell transcriptomics comes of age. *Nature Communications* **11**, 4307 (2020).
6. E. Shapiro, T. Biezuner, S. Linnarsson, Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics* **14**, 618-630 (2013).
7. V. Marx, A dream of single-cell proteomics. *Nat Methods* **16**, 809-812 (2019).
8. Y. Zhu *et al.*, Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature Communications* **9**, 882 (2018).
9. C. Ctortecka *et al.*, An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. *bioRxiv*, 2021.2004.2014.439828 (2022).
10. A.-D. Brunner *et al.*, Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv*, 2020.2012.2022.423933 (2021).
11. R. G. Huffman, A. Chen, H. Specht, N. Slavov, DO-MS: Data-Driven Optimization of Mass Spectrometry Methods. *Journal of Proteome Research* **18**, 2493-2500 (2019).
12. A. T. Chen, A. Franks, N. Slavov, DART-ID increases single-cell proteome coverage. *PLoS Comput Biol* **15**, e1007082 (2019).
13. B. Budnik, E. Levy, G. Harmange, N. Slavov, SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**, 161 (2018).
14. A. Leduc, R. G. Huffman, N. Slavov, Droplet sample preparation for single-cell proteomics applied to the cell cycle. *bioRxiv*, 2021.2004.2024.441211 (2021).
15. H. Specht *et al.*, Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22**, 50 (2021).
16. E. M. Schoof *et al.*, Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nature Communications* **12**, 3341 (2021).
17. A. F. M. Altelaar, A. J. R. Heck, Trends in ultrasensitive proteomics. *Current Opinion in Chemical Biology* **16**, 206-213 (2012).
18. H. Post *et al.*, Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. *J Proteome Res* **16**, 728-737 (2017).
19. B. M. Invergo *et al.*, Sub-minute Phosphoregulation of Cell Cycle Systems during Plasmodium Gamete Formation. *Cell Rep* **21**, 2017-2029 (2017).
20. M. Presler *et al.*, Proteomics of phosphorylation and protein dynamics during fertilization and meiotic exit in the Xenopus egg. *Proc Natl Acad Sci U S A* **114**, E10838-E10847 (2017).
21. E. H. Peuchen *et al.*, Phosphorylation Dynamics Dominate the Regulated Proteome during Early Xenopus Development. *Sci Rep* **7**, 15647 (2017).
22. J. E. Purvis, G. Lahav, Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945-956 (2013).

23. A. Martinez-Val *et al.*, Spatial-proteomics reveals phospho-signaling dynamics at subcellular resolution. *Nature Communications* **12**, 7113 (2021).
24. A. Mund *et al.*, AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. *bioRxiv*, 2021.2001.2025.427969 (2021).
25. C. Lombard-Banek *et al.*, In Vivo Subcellular Mass Spectrometry Enables Proteo-Metabolomic Single-Cell Systems Biology in a Chordate Embryo Developing to a Normally Behaving Tadpole (X. laevis)**. *Angewandte Chemie International Edition* **60**, 12852-12858 (2021).
26. C. Lombard-Banek, S. A. Moody, M. C. Manzini, P. Nemes, Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus laevis and Zebrafish Embryos. *Anal Chem* **91**, 4797-4805 (2019).
27. J. B. Chang, J. E. Ferrell Jr, Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle. *Nature* **500**, 603-607 (2013).
28. L. M. Iakoucheva *et al.*, The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037-1049 (2004).
29. A. L. Darling, V. N. Uversky, Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* **9**, 158 (2018).
30. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* **16**, 18-29 (2015).
31. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**, 321-331 (1999).
32. V. N. Uversky, Introduction to Intrinsically Disordered Proteins (IDPs). *Chemical Reviews* **114**, 6557-6560 (2014).
33. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **6**, 197-208 (2005).
34. R. Schneider, M. Blackledge, M. R. Jensen, Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Current Opinion in Structural Biology* **54**, 10-18 (2019).
35. M. R. Jensen, M. Zweckstetter, J.-r. Huang, M. Blackledge, Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chemical Reviews* **114**, 6632-6660 (2014).
36. G. E. Lienhard, Non-functional phosphorylations? *Trends in Biochemical Sciences* **33**, 351-352 (2008).
37. C. R. Landry, E. D. Levy, S. W. Michnick, Weak functional constraints on phosphoproteomes. *Trends Genet* **25**, 193-197 (2009).
38. F. Gnad *et al.*, Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms. *Proteomics* **16**, 1998-2004 (2016).
39. D. Barford, S. H. Hu, L. N. Johnson, Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *Journal of Molecular Biology* **218**, 233-260 (1991).
40. Z. Serber, J. E. Ferrell, Tuning Bulk Electrostatics to Regulate Protein Function. *Cell* **128**, 441-444 (2007).
41. S. C. Strickfaden *et al.*, A Mechanism for Cell-Cycle Regulation of MAP Kinase Signaling in a Yeast Differentiation Pathway. *Cell* **128**, 519-531 (2007).
42. T. Zarin *et al.*, Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife* **8**, (2019).
43. L. J. Holt *et al.*, Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682-1686 (2009).

44. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
45. Clifford P. Brangwynne, P. Tompa, Rohit V. Pappu, Polymer physics of intracellular phase transitions. *Nature Physics* **11**, 899-904 (2015).
46. A. Bremer *et al.*, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature Chemistry* **14**, 196-207 (2022).
47. E. Laurell *et al.*, Phosphorylation of Nup98 by multiple kinases is crucial for NPC disassembly during mitotic entry. *Cell* **144**, 539-550 (2011).
48. M. I. Linder *et al.*, Mitotic Disassembly of Nuclear Pore Complexes Involves CDK1- and PLK1-Mediated Phosphorylation of Key Interconnecting Nucleoporins. *Dev Cell* **43**, 141-156 e147 (2017).
49. G. Celetti *et al.*, The liquid state of FG-nucleoporins mimics permeability barrier properties of nuclear pore complexes. *Journal of Cell Biology* **219**, e201907157 (2019).
50. H. B. Schmidt, D. Gorlich, Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *Elife* **4**, (2015).
51. P. Denning Daniel, S. Patel Samir, V. Uversky, L. Fink Anthony, M. Rexach, Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proceedings of the National Academy of Sciences* **100**, 2450-2455 (2003).
52. D. Berchtold, N. Battich, L. Pelkmans, A Systems-Level Study Reveals Regulators of Membrane-less Organelles in Human Cells. *Mol Cell* **72**, 1035-1049 e1035 (2018).
53. Y. S. Mao, B. Zhang, D. L. Spector, Biogenesis and function of nuclear bodies. *Trends in Genetics* **27**, 295-306 (2011).
54. P. Brangwynne Clifford, J. Mitchison Timothy, A. Hyman Anthony, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences* **108**, 4334-4339 (2011).
55. A. R. Strom, C. P. Brangwynne, The liquid nucleome – phase transitions in the nucleus at a glance. *Journal of Cell Science* **132**, jcs235093 (2019).
56. E. M. Courchaine, A. Lu, K. M. Neugebauer, Droplet organelles? *The EMBO Journal* **35**, 1603-1612 (2016).
57. J. Y. Ong, J. Z. Torres, Phase Separation in Cell Division. *Molecular Cell* **80**, 9-20 (2020).
58. X. Liu *et al.*, Phase separation drives decision making in cell division. *Journal of Biological Chemistry* **295**, 13419-13431 (2020).
59. S. Cuylen *et al.*, Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature* **535**, 308-312 (2016).
60. H. Jiang *et al.*, Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell* **163**, 108-122 (2015).
61. J. Soding, D. Zwicker, S. Sohrabi-Jahromi, M. Boehning, J. Kirschbaum, Mechanisms for Active Regulation of Biomolecular Condensates. *Trends Cell Biol* **30**, 4-14 (2020).
62. Y. E. Guo *et al.*, Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548 (2019).
63. A. K. Rai, J. X. Chen, M. Selbach, L. Pelkmans, Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211-216 (2018).
64. F. Wippich *et al.*, Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* **152**, 791-805 (2013).
65. J. T. Wang *et al.*, Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in C. elegans. *Elife* **3**, e04591 (2014).
66. G. Zhang, Z. Wang, Z. Du, H. Zhang, mTOR Regulates Phase Separation of PGL Granules to Modulate Their Autophagic Degradation. *Cell* **174**, 1492-1506.e1422 (2018).

67. Z. Monahan *et al.*, Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J* **36**, 2951-2967 (2017).
68. L. G. da Silva *et al.*, Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation. *bioRxiv*, 2021.2004.2030.442163 (2021).
69. O. Beutel, R. Maraspini, K. Pombo-García, C. Martin-Lemaitre, A. Honigmann, Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. *Cell* **179**, 923-936.e911 (2019).
70. J. Emperador-Melero *et al.*, PKC-phosphorylation of Liprin-α3 triggers phase separation and controls presynaptic active zone structure. *Nature Communications* **12**, 3057 (2021).
71. X. Su *et al.*, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599 (2016).
72. M. Hofweber, D. Dormann, Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics. *J Biol Chem* **294**, 7137-7150 (2019).

CHAPTER VII



Summary
Nederlandse samenvatting
List of publications
About the author
Acknowledgements

Summary

The work presented in this thesis focuses on the study of protein phosphorylation in different experimental models, hence trying to answer different technical and/or biological questions. Still, all chapters share a common feature, and that is the study of changes in protein phosphorylation across time. This work shows that adding the temporal dimension to phosphoproteomics experiments greatly aids in data interpretation and extraction of biological insight.

In **chapter II,** we used targeted phosphoproteomics to explore the phosphorylation dynamics of the MAPK-AKT-mTOR signaling pathway. Here we developed an SRM assay to measure functional phosphosites from this protein network, known to be a major player in determining cellular fate. Furthermore, by using an alternative protease for protein digestion we could access relevant phosphorylation sites that cannot be detected when applying tryptic digestion only. In a proof of principle experiment, we provide evidence of temporal differences in phosphorylation events responsible for the activity of this signaling pathway. Next, we used the assay to explore the cellular response to starvation of glucose or amino acids. This revealed an increase in AKT derived signaling, which functions as a pro-survival mechanism when cells are deprived of these nutrients. We think this chapter nicely shows the upside of targeted phosphoproteomics, since this technique allows to measure predefined phosphorylation sites of interest with high sensitivity and reproducibility.

In **chapter III,** we used MS-based phosphoproteomics to study the differential role of the mitogen activated protein kinases (MAPK), ERK1 and ERK2 in the cellular response to growth factor treatment. These two enzymes locate at the core of the MAPK signaling pathway, and they are known drivers of cellular growth, proliferation and differentiation. Still, up to date is not known if they are redundant or if each kinase has specific functions. We show that depletion of either enzyme drastically affects the proteome and the phosphoproteome, albeit the absence of ERK2 has a more pronounced effect, with clear downregulation of phosphorylation sites in members of the nuclear pore complex. Furthermore, we used targeted phosphoproteomics to show that in the absence of ERK2, cells compensate with upregulation of alternative kinases. A time-course experiment revealed that cells lacking ERK2 show a slower phosphorylation activity after treatment with growth factors, indicative that ERK2 is a more efficient kinase when compared to ERK1. These findings correlate with previous reports that suggest ERK2 has a more relevant role in cellular signaling than ERK1.

In **chapter IV** we combined shotgun and targeted phosphoproteomics to measure protein phosphorylation dynamics in single embryos of *Xenopus laevis*. This way, we acquired what we think is the first map of cell cycle related phosphorylation in an unperturbed biological system *in vivo*. Using an *in vitro* model of the cell cycle, we show that our *in vivo* measurements can discriminate between cell cycle phases (interphase and mitosis). Targeted phosphoproteomics at

high-temporal resolution showed that mitotic phosphorylation is switch-like *in vivo*, and it occurs on highly interconnected proteins that display larger levels of intrinsic disorder when compared to other phosphoproteins. Large-scale bioinformatics analysis revealed that this feature is conserved from yeast to human, suggesting that phosphorylation of highly disordered proteins is a hallmark of the cell cycle. Next, we used different molecular modeling tools, which showed that phosphorylation regulates protein homotypic interactions, which is one of the driving forces behind protein phase separation. In addition, we found that a high proportion of the dynamically phosphorylated proteins are found in membrane-less organelles, which are known to assemble-disassemble throughout the cell cycle. Since the vast majority of the detected phosphosites were potential CDK targets, we validated our findings using a known CDK substrate, namely Ki-67. We provide evidence showing that CDK driven phosphorylation regulates the phase separation of this protein, which suggests that cellular reorganization at mitosis is mediated by CDK driven phosphorylation and modulation of protein phase separation.

Finally, in **chapter V** we combined LC-MS based proteomics and phosphoproteomics with mass spectrometry imaging (MSI). We used these tools to study the ischemia-reperfusion (IR) injury in human intestinal tissue. Proteome measurement showed that upon IR there is a decrease in abundance of proteins related to intestinal absorption, microvillus and cell-cell junctions. On the other hand, proteins that displayed an increase in abundance were mostly related to the innate immune response, namely proteins of the complement cascade. Phosphoproteomics showed activation of cellular stress response upon ischemia, very likely mediated by MAPKs like JNK or p38. Upon reperfusion, there was a clear increase of potential targets of other MAPKs, such as ERK1/ERK2, which are mainly linked to growth factor response. Finally, MSI showed the varied distribution of proteins to different types of tissue, highlighting the importance of combining LC-MS with other techniques that provide spatial information.

Samenvatting

Het werk in dit proefschrift richt zich op het onderzoeken van eiwit fosforylering in verschillende experimentele modellen, waarbij gepoogd wordt verscheidene technische en biologische vragen te beantwoorden. De hoofdstukken in dit proefschrift hebben allen gemeenschappelijk dat eiwit fosforylering wordt bestudeerd door de tijd heen. Dit werk laat zien dat het in acht nemen van de dimensie tijd in *phosphoproteomics* experimenten van groot belang is voor het interpreteren van de data en het verkrijgen van biologische inzichten.

In **Hoofdstuk 2** maken we gebruik van *targeted phosphoproteomics* om signaaltransductie door middel van eiwitfosforylering in de MAPK-AKT-mTOR signaleringsroute in kaart te brengen. De gepresenteerde *SRM* methode stelt ons in staat om functionele fosforyleringen te meten die een cruciale rol spelen in de lotsbestemming van een cel. Door middel van het gebruik van verschillende enzymen voor eiwit digestie, hebben we de methode kunnen uitbreiden met fosforyleringen die niet detecteerbaar zijn met trypsine alleen. In dit hoofdstuk wordt aan de hand van een voorbeeld duidelijk dat fosforyleringen in deze signaaltransductieroute tijdsafhankelijk zijn. Verder hebben we deze methode toegepast om te onderzoeken op welke manier deze signaleringsroute een rol speelt bij de cellulaire reactie op een tekort aan glucose en aminozuren. Onthouding van deze voedingsstoffen leidde in dit experiment tot een verhoogde AKT activiteit als overlevingsmechanisme. De voordelen van *targeted phosphoproteomics* komen in dit hoofdstuk duidelijk naar voren; deze techniek stelt ons in staat om belangrijke fosforyleringen met hoge sensitiviteit en reproduceerbaarheid te meten.

In **Hoofdstuk 3** hebben we de rol van ERK1 en ERK2 in signaaltransductie na behandeling met groeifactoren onderzocht. Deze enzymen spelen een cruciale rol in de MAPK signaalroute en drijven celgroei, proliferatie en differentiatie aan. Het is echter onduidelijk of deze kinases elk een specifieke rol hebben in de cel, of dat hun functie overtollig is. Wij laten in dit hoofdstuk zien dat het uitschakelen van beide enzymen een groot effect heeft op het *proteome* en *phosphoproteome*, maar dat afwezigheid van ERK2 een sterker effect bewerkstelligt, leidend tot verminderde fosforylering van eiwitten in het nuclear pore complex. Met behulp van *targeted phosphoproteomics* hebben we verder laten zien dat andere kinases verhoogd actief zijn in afwezigheid van ERK2. In afwezigheid van ERK2 reageren cellen minder snel op toevoeging van groeifactoren, wat impliceert dat ERK2 een efficiëntere kinase is dan ERK1. Deze bevindingen zijn in lijn met eerdere studies die suggereren dat ERK2 een relevantere rol speelt in signaaltransductie dan ERK1.

In **Hoofdstuk 4** combineren we *shotgun* en *targeted phosphoproteomics* om de plasticiteit van eiwit fosforylering in *Xenopus leavis* embryo's te meten. Dit heeft het eerste overzicht van celcyclus-gerelateerde fosforylering *in vivo* opgeleverd. We hebben onze *in vivo* metingen gevalideerd door middel van een *in vitro* model, waarbij we hebben aangetoond dat we in staat zijn om *in vivo* de verschillende fasen van de celcyclus te onderscheiden. Daarnaast hebben we met behulp van *targeted phosphoproteomics* aangetoond dat fosforylering tijdens de mitose zich gedraagt als een aan-uit schakelaar. De eiwitten waarop deze fosforyleringen voorkomen, zijn onderling sterk verbonden en bevatten meer intrinsiek ongestructureerde regio's dan gemiddeld. Grootschalige bio-informatica analyses hebben aangetoond dat deze fosforylering van ongestructureerde eiwitten niet alleen in humane cellen een rol speelt, maar geconserveerd is in evolutie. Dit impliceert dat dit een cruciaal kenmerk is van de celcyclus. Verder hebben we met behulp van moleculaire modellering laten zien dat fosforylering homotypische eiwit interacties reguleert. Deze interacties zijn de drijvende kracht achter eiwit fasescheiding. Aanvullend hebben we in dit hoofdstuk beschreven dat een groot deel van de dynamische fosfo-eiwitten onderdeel zijn van membraan-loze organellen die vormen en uiteenvallen tijdens de celcyclus. Aangezien het grootste deel van de gedetecteerde fosforyleringen potentiële CDK substraten zijn, hebben we deze bevindingen gevalideerd met behulp van het bekende CDK substraat, Ki-67. We hebben bewijs geleverd dat CDK-gedreven fosforylering de fasescheiding van dit eiwit reguleert, wat impliceert dat cellulaire rerorganisatie tijdens mitose tot stand komt via CDK-gedreven fosforylering en veranderingen in eiwit fasescheiding.

Tenslotte hebben we in **Hoofdstuk 5** LC-MS *proteomics* en *phosphoproteomics* gecombineerd met *mass spetrometry imaging* (IMS). We hebben deze technieken gebruikt om ischemie- en reperfusieschade (IR) in humaan darmweefsel te bestuderen. De *proteomics* data laten een vermindering van eiwitten zien die betrokken zijn bij darmabsorptie, microvillus en cel-cel verbindingen na IR. Eiwitten betrokken bij het aangeboren immuunsysteem, voornamelijk eiwitten van de complement cascade, waren daarentegen verhoogd aanwezig na IR. *Phosphoproteomics* heeft activatie van de cellulaire stressreactie als gevolg van ischemie aangetoond, hoogstwaarschijnlijk bemiddeld door MAPKs, zoals JNK of p38. Een duidelijke verhoging van potentiële substraten van andere MAPKs, zoals ERK1/ERK2 werd gevonden na reperfusie. Dit is waarschijnlijk verbonden met signaaltransductie door groeifactoren. MSI heeft tenslotte de verdeling van eiwitten over verschillende typen weefsels aangetoond. Dit markeert het belang van het combineren van LC-MS met andere technieken die informatie verschaffen over de ruimtelijke verdeling van eiwitten.

List of publications

Valverde, J.M.* , Dubra, G*., van den Toorn, H., van Mierlo, G., Vermeulen, M., Heck, A.J., Elena-Real, C., Fournet, A., Al Ghoul, E., Chahar, D., Haider, A., Paloni, M., Constantinou, A., Barducci, A., Ghosh, K., Sibille, N., Bernardo, P., Knipscheer, P., Krasinska, L.†, Fisher, D. † & Altelaar, M. † A CDK-mediated phosphorylation switch of disordered protein condensation. Manuscript under review.

Valverde, J.M.*, Debets, D.O.* & Altelaar, M. Capturing the signalling dynamics of the MAPK-AKT-mTOR pathway in a single targeted phosphoproteomics assay. Manuscript under review.

Valverde, J.M., van Breugel, P. & Altelaar, M. Phosphoproteome profiling reveals differences in ERK1 and ERK2 dependent phosphorylation. Manuscript in preparation.

Kip, A.M., **Valverde, J.M.**, Altelaar, M., Heeren, R.M.A., Hundscheid, I.H.R., Dejong, C.H.C., Olde Damink, S.W.M., Balluff, B. & Lenaerts, K. Combined quantitative (phospho)proteomics and mass spectrometry imaging reveal temporal and spatial protein changes in human intestinal ischemia-reperfusion. Journal of Proteome Research 2022, 21(1), 49-66.

Valverde, J.M., Rodríguez, K., Herrera, M., Segura, A., Vargas, M., Villalta, M., Montero, M., Gutiérrez, J.M., León, G. Comparisson of the adjuvant activity of emulsions with different physicochemical properties on the antibody response towards the venom of West African carpet viper (Echis ocellatus). Toxicon 2017, 127(1), 106-111.

*, † Equal contributions

About the author

I was born on November 23rd of 1988 in San José, Costa Rica. I attended high school at Colegio Monterrey, in Vargas Araya, Montes de Oca. After finishing high school, I started studies in Microbiology and Clinical Chemistry at Universidad de Costa Rica in 2007. I did my thesis at Instituto Clodomiro Picado under the supervision of Dr. Guillermo León, where I studied the adjuvant properties of emulsions used to produce anti-venoms against snakebite. I did a clinical practice at Hospital Mexico just before culminating the program in 2012. Next, I started a Master's program with an emphasis on Cellular Biology at Universidad de Costa Rica. The latter was under the supervision of Dr. Steve Quirós, and we were interested in studying the cellular DNA damage response after treatment with genotoxic drugs. During this time, I took the course 'Introduction to Proteomics' given by Dr. Bruno Lomonte. The latter was key in sparking my interest in mass spectrometry based proteomics. In 2016, I obtained a scholarship from the Ministry of Science and Technology of Costa Rica to undertake PhD studies abroad. This led me to The Netherlands, specifically to the labs of Maarten Altelaar and Albert Heck at Utrecht University. Here I studied protein phosphorylation dynamics in different biological contexts by using mass spectrometry. The results of this research are presented on this thesis.

Acknowledgements

This book contains the product of my work and effort for the last years. Having said that, I am not the solely responsible for the results shown here. Many of you have influenced my life in a variety of ways. In some cases, your involvement in this thesis has been direct, by participating or helping me in any of the projects shown here. In other cases, your support has been indirect: you are likely to have shared with me one or many moments in which you have listened to me talk about the work on this thesis or about the struggles of the "PhD student life". Even if you knew it or not, many of those moments are what kept me pushing towards the finish line. Moreover, if you listened to one of my monologues on 'phosphorylation, protein disorder and phase separation', I do thank you for all the patience!

First, I would like to thank Maarten. I will always appreciate you for taking a chance with me. I was a completely unknown guy from Costa Rica and still you decided to take me into your lab and allow me to pursue my PhD under your supervision. As you know, at times I really felt I was struggling with all the freedom and lack of structure intrinsic to working on research. In addition, you have a relative 'hands off' approach that didn't make it easier for me at the beginning. Despite all this, I will always be grateful for those moments in which you casually stepped into my office and said "let's do phosphoproteomics on *Xenopus* eggs" or "let's explore the differences between ERK1 and ERK2". Little did I knew that when you casually pitched those ideas, we were deciding the different chapters of my thesis. Even though it took me some time to realize it, now I can appreciate that all the freedom you gave me is what allowed me to pursue my curiosity and to come up with hypotheses to explain the phosphorylation trends we observed in the *Xenopus* project. This 'freedom' is what eventually led us to our current hypothesis around protein phosphorylation and the control of phase separation. From all this time working with you I take many teachings for which I will always be grateful. The best wishes for you and the next steps in your career!!

Next, I would like to thank Albert. I am grateful for the valuable input you provided during the first stages of my PhD, especially when we were beginning with the *Xenopus* project. Every time we had a chat, either in your office or in the hallways you managed to drop some nuggets of wisdom that have helped me develop as a scientist. I hope the next time we meet is in Costa Rica, a country I have learned you are very fond of.

Puck: I would also like to thank you for receiving me in your lab. Even though my time there was brief, you provided me with all the things I needed to conduct the experiments that were key to the main chapter of this thesis. Whenever it was possible, you and other members of your lab (Merlijn, Koichi, Nerea, Roxanne) took the time to help me and answer my questions, and for that I am grateful. Mirjam Kip: even though things were slow for us at the beginning of our collaboration, once things picked up, we managed to work quite swiftly and get the work published. To work with you was a joy: you were very efficient, and we communicated clearly during the whole process. It was a very nice experience to work alongside you.

Next, I would like to thank those in France. Dan and Liliana: even though we have not met in person yet (hopefully will change in the future), I am so happy that our roads crossed. Without you guys chapter IV of this thesis would not be what it currently is. Not only because of all the work you've put into it, but also because of all the things I've learned from you two along the way. To learn about cell cycle, mitosis and others subjects from you guys has been a great experience for me. You also helped me to get in contact with Tony Ly and Adrian Saurin at University of Dundee, which led to a new position and the start of a new adventure for me and my wife in the UK. For all this I am very grateful! Gero: mi compa, que buena sorpresa fue toparme con un hermano latinoamericano tan pura vida como vos. La colaboración con vos fluyó desde el inicio. Gracias por todo lo que me enseñaste sobre R, por las discusiones acerca de ciclo celular, fosforilación y phase separation, entre otros! Pero más importante aún, gracias por la amistad y por los buenos momentos que hemos compartido durante la elaboración de este proyecto. Un abrazo!!

Donna! What can I say, so many moments shared in front of the Bravo and/or the Altis! Thanks a lot for being the best colleague, it was always great to work alongside you. Teaming up with you was one of the highlights of my PhD. I learned a lot from you. Continue to be the kind person you already are, the world needs more of your kind! I can say that after all my years in The Netherlands I have left one of my best friends here. I am so happy to have you in my life and I really hope you continue to be a part of it.

And off course I want to thank my office mates, of which there were many: the first generation, Anja and Lindsey: I will always appreciate your help during my first days in the lab. In the case of Lindsey, thanks a lot for giving me my first introduction into proteomics data analysis, you were patient and took your time teaching me some of the things I will later continue to use during my PhD. Next is the second generation. Nicholas: thanks a lot for all the fun moments we shared. Even though your time in the lab was brief, you were always kind to me and we for sure shared many laughs in the office. Nadine: thanks for listening to me talk/complain about the PhD struggles. You always had wise advice or at least were kind enough to hear me express my thoughts. It was great to share an office with you during the majority of my PhD, we had just the perfect balance of hard work and having a laugh every now and then. I cannot think of a better office mate! Franziska: thank you for being a great listener, with you I could always discuss scientific results, get help for my scripting in R, but most importantly, you listened to me whenever things in the lab/life were not going well. Thank you for that! Finally, Laura: you were the last one to join our office, but quite an addition you were! It was always fun to hear your (many) rants. You for sure had strong opinions but could also be a great listener whenever I needed one. Oh, and you were a great provider of candy during all these years, so thank you for that as well!!

Harm: man what can I say, thank you for all the help and knowledge you shared with me during these years. So many times did I walk into your office with a question or problem in the lab, and you were always kind to help me out. You taught me many things and I will always be grateful for that. The best wishes for you and your family! Mirjam: thanks for all your support in the lab during all these years. I know you put in a lot of effort to make our jobs in the lab easier, and believe me this doesn't go unnoticed! Arjan: thanks for the help in the lab during all these years. Whenever I had issues with a mass spec I knew you were the person to ask. You were always patient and took your time answering my questions. Pieter: what can I say, only you and I know the struggle it was to get our stable cell lines lacking ERK1 or ERK2. Thanks for all your help and support with this project, without your work none of it would've been possible. I also want to thank you for the chats we had, I always felt better after sharing some of my frustrations in the lab with you. Best of lucks for you in your new job! Soenita: even though our overlap in the lab was brief, I will always remember your kind words after a Wednesday presentation and your help during my early days in the cell culture lab. Ceri: thanks for all your help troubleshooting LCs and for taking over the cell culture lab. You were always eager to help out with any issues in the lab and for that I am grateful. Overall, all of you guys helped me in one way or the other during my time in the lab, and for that I thank you!

Corine: thanks for your help during all my years in the lab. Starting with all the paperwork that you helped me do while I was still in Costa Rica, to all the times I had to renew my contract or residence permit. All these things were overwhelming, especially during my early days in the lab, but with your help I managed to sail through it. Geert: only you and I know how many times have I asked for your help. From my R studio not working, to my computer just completely crashing, you swiftly fixed most of the 'computer problems' I had. Big thanks for that! Henk: thanks to you for implementing the IUPred disorder predictor in R for me. This first step would end up being key in the development of chapter IV. You also helped me doing other analyses about the molecular features of disordered regions, and even though some of them did not make it to this thesis, I still thank you for all that work!

Thierry: you introduced me to the world of targeted proteomics. Thanks for all the time invested in front of your pc teaching me how to work with Skyline. It feels like just yesterday when we tried to hook up a Proxeon to the QTrap, just to end up realizing this was not going to work. Still, I learned a lot about LC-MS setups just by playing around with those instruments. Thanks for all those moments and teachings! Matina: what can I say, it was always fun to have a chat with you. Thanks for hearing me out when things were not going great in the lab. Also, thanks a lot for all the help you gave me at different points during my PhD. Every time I saw the box of media additives to grow MCF10A that you inherited me I thought of you. I was very happy for you when you successfully defended your thesis. Now that I get to defend mine, I can only remember those moments in which we wondered if we were going to see the light at the end of the tunnel or not. Well, looks like we did it! David: mae siempre le voy a agradecer el haberme puesto en contacto con Maarten. Fue un paso clave en un largo proceso que inició en Costa Rica y ahora culmina acá. Gracias por los buenos momentos vividos en Utrecht, por las idas al cine, y todas las habladas que nos pegamos intentando arreglar el mundo. Gracias por lo que fue una despedida de soltero muy graciosa en Ámsterdam y muchos otros buenos momentos vividos durante nuestro tiempo en Holanda. Un abrazo y la mejor de las suertes para lo que sea que venga! Fabio: you were probably one of the first friends I made in Utrecht. With you it felt as if we had met years ago, it was that easy talking to you (and believe me, during my early days in Utrecht, that meant a lot). You sparked the interest of cycling in me, and I will always be grateful for all those hours we spent in the bike routes around Utrecht. This really helped me during my first days in The Netherlands, which at times felt very overwhelming. Even though our overlap was brief, I will always remember you as a great guy!

Johannes: my boy! What can I say that you already don't know?? Really mate, how cruel is life that we managed to spark our friendship just as it was almost time for me to leave Utrecht?? Well, even though our friendship is somewhat new, it has been very intense (in the best way possible!). Even with your incessant need to have me chugging beers, I still manage to always have a great time around you (damn, maybe is because of the beer chugging??). Getting to know you and Morgane was a highlight of our last months in Utrecht, and I cannot wait to spend more time with you guys in the future. Visits to Scotland and Costa Rica are pending! Gadi: thanks for the quick introductory lesson you gave me on ggplot2. I would eventually end up using this package for the rest of my PhD. But more importantly, thank you for all the moments shared in front of Orbi14, for all the chats and for always being such a nice and kind person! Julia: Aka Jules, Juliana. Gracias por las risas y por los buenos momentos, siempre se pasaba bonito con vos! Fujia: man, even though our overlap in the lab was brief, you have always showed me genuine interest and kindness. It was great to find another fan of Liu Cixin in the lab. The best of lucks for you finishing your PhD. Big hug! Tim: thanks for helping me out with the scratch assay and other things in the lab. Especially when it came to the Altis, you were the expert and you always helped me when I had questions. Good luck finishing your PhD. Oleg: thanks for all your help during my brief adventure with crosslinking MS. Even though we did not manage to continue the project/idea we had in mind, I am grateful for all the discussions and time we shared. Kelly S: you always helped me whenever I had questions about lab stuff and also gave me a lot of input during the subgroup meetings. Many of your suggestions really helped me improve my presentations during all these years. Thanks for that! Sem: thanks for all the moments and laughs shared during our time in the lab. No matter if it was in the lab, having a beer or in the padel court, it was always easy to have a good time with you mate!

A los amigos de Holanda que conocí fuera del laboratorio: Felipe y Jessica, aunque no fue mucho el tiempo que coincidimos en Holanda, cada momento y salida que tuvimos estuvo bien gozada. Siempre era refrescante hablar con un par de ticos, especialmente después de tanto tiempo lejos de casa. Me alegro montones de haberlos conocidos y de que nuestra amistad haya crecido durante este tiempo. Jorge Bustos: mae que bueno que fue recibirte en la casa y vacilar al lado tuyo y de Marlen. Ya que te gustó tanto Costa Rica espero que eventualmente nos veamos por allá, o mejor aún, que nos enseñes un poco de Málaga! Armin, my friend thanks for all the good times in the padel court. Those moments are part of what kept me sane when I thought this thesis was going to drive me insane. I hope that we get to play again sometime in the future!

Hay compas en Costa Rica que siempre han estado pendientes de mi durante todo este tiempo. Y por supuesto que tengo que empezar por este par: Tavo y Oscar. Maes ustedes son mis compas del alma, ya son muchos los años, y pese al tiempo y la distancia a ustedes siempre los siento cerca. Gracias por acompañarme, por los mensajes y las llamadas, por las habladas de deportes (NBA, champions, la sele.. lo que sea!). Los quiero muchísimo y siempre les deseo lo mejor! Jaime: mae no sabe lo feliz que me ha hecho revivir mi amistad con usted. Por algún motivo esto sucedió cuando nos encontrábamos a miles de kilómetros de distancia, pero lo importante es que sucedió. Mae hablar paja con usted siempre es interesante. Gracias por el interés genuino con el que siempre me pregunta por mi trabajo, espero que me cuente que le parece toda la paja escrita en esta tesis y nos podamos sentar a hablar del tema con una birra en el futuro. Un abrazo mi compa! Keno: mae, gracias por las dos visitas. Especialmente en la segunda, fue genial tenerlo todos esos días en la casa. Créame que para Marlen y para mi fue refrescante recibir a un compa como usted y hablar todo lo que hablamos en esos días. Un abrazo y los mejores deseos siempre! Y por supuesto que sigo con los microbios (beios).

Paradójicamente soy la persona menos activa en nuestro grupo de whatsapp porque no me gusta lo 'impersonal' de estos grupos, y hoy me refiero a ustedes en grupo acá. Pese a todo el tiempo estando largo, ustedes han sido de gran apoyo para mí. Ya sea con un mensaje de vez en cuando, con una llamada o con uno que otro favor, los he sentido cerca. Desde que me fui de Costa Rica el nivel de cariño que ustedes me mostraron fue exagerado. Siempre, siempre voy a recordarlo. Es más, con el nivel de despedida que me hicieron, como no iba a terminar el doctorado?? Era inaceptable no concluir la misión! Tengo que mencionar a Vane, Adrims y Nacho. Con ustedes es con quien más he hablado durante todo este tiempo. Gracias por sacar el rato para escucharme hablar de las mil y una penurias vividas por acá, créanme que toda esa catarsis me ayudó a mantenerme cuerdo para llegar al final de esta travesía. Mención especial para Caro y Nacho: ustedes dos vinieron a dar hasta Holanda. Créanme que los momentos vividos acá los voy a atesorar siempre en mi memoria. Nacho: mae gracias por el interés que siempre ha mostrado en mi brete y vida acá. Su visita y el viaje que hicimos a Escocia (será que se repite pronto?) siempre van a ser recuerdos increíbles. Abrazo enorme dude!

Seguidamente quiero agradecer a varias personas en la UCR. Especialmente a los miembros del DAC (Walter, Manuel, Nela, Ileana y Eugenia), que siempre me han apoyado, incluso desde antes de que saliera de Costa Rica. Puntualmente a Steve: muchísimas gracias por todo su apoyo en este proceso. Gracias a su consejo fue que apliqué a la beca del MICITT, y gracias en parte a su apoyo es que se me concretó la reserva de plaza en la U. Después está toda su gestión para facilitar que yo siga mi formación ahora en Escocia. Siempre le voy a estar agradecido por todo este apoyo. Meli: gracias por siempre enviarme apoyo y buenas vibras a la distancia en nuestros 'catch up' habituales. Los mejores deseos para con su doctorado. Con su talento y dedicación estoy seguro de que las cosas van a salir bien. Cynthia: aprovecho para decirle que estoy súper feliz por usted al saber que culminó su doctorado con éxito! Siempre le voy a agradecer por motivarme a conseguir el trabajo en la UCR. Ese es uno de muchos pasos que me traen hoy acá. Gracias también por toda la motivación y el apoyo que recibí de usted cuando me iba de Costa Rica y por la visita a Holanda! Jorge: mae, aunque usted no lo sepa, mucha de la inspiración que me llevó a optar por un doctorado viene de usted. Empezando cuando era estudiante de licenciatura y usted sembró en mi la idea de realizar estudios de posgrado, hasta todos los momentos que compartimos en cultivo celular del CIET, donde usted me enseñó muchas prácticas que sigo usando hasta el día de hoy. Gracias por todo eso.

Si decidiera mencionar a todos los miembros de mi familia que de alguna forma me han ayudado a llegar hasta este punto en mi vida, llenaría un libro igual a este de sólo agradecimientos. A continuación, hago mi mejor intento para condensar tanta gratitud en un par de párrafos. Tengo que comenzar por mis tíos, tías y primas: si bien no puedo mencionarlos a todos, muchos de ustedes fueron pieza instrumental en mi formación como persona. Además, sin el apoyo que me brindaron especialmente durante mi educación universitaria, no habría conseguido alcanzar muchos de los logros que tengo hasta el día de hoy (incluida la obtención de este grado académico). Tía Estela, Tía Rosa y Guille: ustedes fueron segundas madres para mí, gracias por todo el apoyo y cariño que me han dado. Especialmente para Guille que ya nos abandonó: gracias por enseñarme de resiliencia y disciplina, usted fue todo un ejemplo a seguir en mi vida. Gracias por tanto! A Tío Carlos y Tío Rigo: gracias por el apoyo y cariño, ustedes han sido grandes influencias en mi vida y siempre les voy a agradecer tanto amor. A Tío Coca: gracias por toda su ayuda y apoyo con trámites legales de distinta índole que tantas veces me facilitaron la vida a miles de kilómetros de distancia. A mis primas María, Carol y Jessica que tanto cariño e interés me han demostrado desde que tengo uso de razón. Gracias por todo su apoyo. Me sobran los nombres para poner en este párrafo, pero me falta el espacio. Si no le incluí en esta lista pero usted está leyendo esta tesis, es probable que tenga algo que agradecerle a usted también, así que de corazón, gracias!

Tengo muchísimo que agradecer a la familia que me vio nacer y que me hizo la persona que soy hoy. A mis hermanos, Daniel y Mariana. Gracias por todas las llamadas que tuvimos los domingos, por aparecer un rato en la cámara aunque tal vez no tuvieran ganas. Muchas veces el verlos allí e intercambiar unas cuantas palabras me llenaron de energía para la semana que venía. Gracias a ambos por cuidar a Juno y a Mia que nos dejaron mientras yo estaba acá, y gracias también por ayudarme a sobrellevar la pérdida. A Mariana en especial, gracias por las visitas y los buenos momentos vividos acá. Y en el caso de Daniel, aunque su venida a Holanda no se dio, gracias por hacer posible la visita de Mami y Papi en algún momento. Espero que el viaje a Escocia si se concrete! A mis padres, Hilda y Manuel, es mucho lo que tengo que agradecerles. Nada de lo que he logrado hasta el día de hoy habría sucedido sin el apoyo y el esfuerzo de ustedes. Gracias por todo su interés en mi trabajo, por todas las llamadas de fin de semana y por las visitas que me hicieron. Créanme que me llenaban de vida para continuar con mi trabajo, especialmente en los momentos más difíciles, cuando la distancia se sentía mucho y el tiempo se hacía largo. Espero que se sientan parte de este logro, porque créanme que es tan mío como de ustedes. Los amo con todo el corazón.

Quiero seguir con los miembros de la familia que hice durante mi tiempo en Holanda. Empiezo por Doña Gaby: gracias por todo el apoyo y cariño que me ha brindado durante estos años. Le agradezco muchísimo por estar siempre tan pendiente de mi progreso con el doctorado. Hoy le puedo decir que afortunadamente ya lo terminamos! A mi cuñado y cuñadas: Shirley, gracias también por siempre mostrar interés en mi trabajo y en nuestra vida en Holanda. Gracias por venir a visitarnos, por todos los buenos momentos compartidos, y esperando que vengan más pronto! Esteban, Kelli y Gael: gracias por el interés tan auténtico con que siempre me preguntaban por mi trabajo y el progreso con el doctorado. Todavía recuerdo cuando los visitamos en Hong Kong y pasaron poco más de una hora escuchando un monólogo mío sobre proteínas y espectrometría de masas. Ese día me di cuenta de lo mucho que me querían. Espero verlos más seguido para poder seguir entreteniéndolos (o aburriéndolos?) cuando les hablo de ciencia. Un abrazo fuerte!!

Finalmente tengo que agradecer a la persona que ha estado siempre a mi lado desde el inicio de esta aventura. Marlen. Sólo usted y yo sabemos las cosas por las que hemos pasado desde que me vine a vivir a Holanda en 2016. El separarnos trajo momentos difíciles que eventualmente no hicieron más que cementar y fortalecer nuestra relación. Cada llamada, cada visita y cada abrazo contó. Créame que desde el inicio su apoyo marcó toda la diferencia, y no quiero imaginar donde estaría sin él. Y ni que decir de todo lo que vino después. Su valentía al decidir mover su vida a Holanda es algo que hasta el día de hoy me asombra y me llena el corazón. Todavía recuerdo toda la zozobra y estrés que fue realizar el papeleo para la visa, sumado al organizar una boda entre los dos a la distancia. Fueron días intensos, pero ahora cuando veo mi vida con usted, puedo disfrutar el resultado de todo el trabajo y amor que hemos puesto en nuestra relación. Gracias por traer todo lo increíble que vino con usted a Holanda, por llenar mis días de tanto cariño y por enseñarme tanto en el camino. Gracias por traer su arte a este libro. Por escuchar mi idea para la portada, pero sobre todo por crear algo aún mejor de lo que yo pude imaginar. Gracias por escucharme, por darme su apoyo, pero también su honestidad siempre que la he necesitado. La amo con todo el corazón y no puedo estar más feliz de haber compartido todo este trayecto a su lado.