Master's Thesis

Identification of Direct Substrates for CMGC Kinases with ^{18}O-ATP-based Kinase Assay and LC-MS/MS

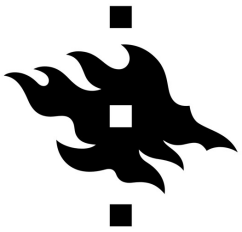
Vera Varis

Molecular Biosciences

Faculty of Biological and Environmental Sciences

UNIVERSITY OF HELSINKI

February 2020



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Faculty	Degree Programme
Faculty of Biological and Environmental Sciences	Molecular Biosciences

Author
Vera Amanda Varis

Title
Identification of Direct Substrates for CMGC Kinases with ^{18}O-ATP-based Kinase Assay and LC-MS/MS

Subject/Study track
Biochemistry

Level	Month and year	Number of pages
Master's thesis	February 2020	58

Abstract

Protein kinases are signaling molecules that regulate vital cellular and biological processes by phosphorylating cellular proteins. Kinases are linked to variety of diseases such as cancer, immune deficiencies and degenerative diseases.

This thesis work aimed to identify direct substrates for protein kinases in the CMGC family, which consists of the cyclin-depended kinases (CDK), mitogen activated protein kinases (MAPK), glycogen synthase kinase-3 (GSK3) and CDC-like kinases (CLK). CMGC kinases have been identified as cancer hubs in interactome studies, but large-scale identification of direct substrates has been difficult due to the lack of efficient methods.

Here, we present a heavy-labeled ^{18}O-ATP-based kinase assay combined with LC-MS/MS analysis for direct substrate identification. In the assay, HEK and HeLa cell lysates are treated with a pan-kinase inhibitor FSBA which irreversibly blocks endogenous kinases. After the removal of FSBA, cell lysates are incubated with the kinase of interest and a heavy-labeled ATP, which contains ^{18}O isotope at the γ-phosphate position. Resulting phosphopeptides are enriched with Ti^{4+}- IMAC before the LC-MS/MS analysis, which distinguishes the desired phosphorylation events based on a mass shift caused by the heavy ^{18}O.

With this pipeline of methods, we managed to quantify and identify direct substrates for 26 members of CMGC kinase family. A total of 1345 substrates and 3841 interacting kinase-substrate pairs were identified in cytosolic cell lysates, from which 165 were annotated in the PhosphoSitePlus® database. To identify substrates for kinases with nuclear localization, ten kinases were tested with nuclear HEK cell lysate. We identified 194 kinase-substrate pairs, 141 of which were unique to the nuclear fraction and 27 annotated in the PhosphoSitePlus® database. Finally, kinases with outstandingly high amounts of novel substrates were subjected to gene ontology analysis. We were able to link the gene ontology classifications of novel substrates to the biological processes regulated by the kinase of interest. These results indicate that heavy-labeled ^{18}O-ATP-based kinase assay linked LC-MS/MS is a useful tool for large-scale direct kinase substrate identification.

Keywords
protein kinase, kinase, kinase assay, CMGC, Ti4+-IMAC, LC-MS/MS, mass spectrometry

Supervisors
Tiina Öhman, Salla Keskitalo

Where deposited
Viikki Science library, University of Helsinki

Additional information

Tiedekunta	Koulutusohjelma
Bio- ja ympäristötieteellinen tiedekunta	Molekyylibiotieteet

Tekijä
Vera Amanda Varis

Työn nimi
Suorien substraattien tunnistaminen CMGC-kinaaseille ^{18}O-ATP-kinaasitestin ja LC-MS/MS avulla

Oppiaine/Opintosuunta
Biokemia

Työn laji	Aika	Sivumäärä
Pro gradu	Helmikuu 2020	58

Tiivistelmä

Proteiinikinaasit ovat signaalimolekyylejä, jotka säätelevät tärkeitä biologisia prosesseja fosforyloimalla solun proteiineja. Kinaasit on linkitetty moniin sairauksiin kuten syöpään, immuunipuutostauteihin ja rappeumasairauksiin.

Tämän Pro gradu tutkimuksen tarkoituksena oli tunnistaa suoria substraatteja CMGC-kinaasiperheen jäsenille, joihin kuuluu muun muassa CDK-, MAPK-, GSK-3 ja CLK-kinaaseja. Proteiinien vuorovaikutussuhteita tutkiessa CMGC-kinaasiperheen jäseniä on tunnistettu syöpäkeskittymiksi, mutta suoria kinaasien substraatteja ei ole tunnistettu laajassa mittakaavassa, pääsyynä tehokkaiden menetelmien puuttuminen.

Tässä tutkimuksessa esittelemme uuden, ^{18}O-leimattuun ATP:hen ja nestekromatografiaan yhdistettyyn massaspektrometriaan (LC-MS/MS) perustuvan menetelmän, jonka tarkoituksena on tunnistaa kinaasien suoria substraatteja. Menetelmässä HEK ja HeLa solulysaatit käsitellään kinaasi-inhibiittori FSBA:lla, joka estää endogeenisten kinaasien toiminnan. FSBA:n poistamisen jälkeen tutkittavaa kinaasia inkuboidaan käsitellyssä solulysaatissa ATP:n kanssa, joka sisältää ^{18}O-isotoopin sen γ-fosfaatissa. Syntyvät fosfopeptidit rikastetaan Ti4+-IMAC materiaalin avulla ennen LC-MS/MS analyysiä, joka tunnistaa toivotut fosforylaatiotapahtumat painavan ^{18}O-isotoopin aiheuttaman massaeron avulla.

Tämän metodin avulla onnistuimme kvantifioimaan ja tunnistamaan suoria substraatteja 26:lle CMGC-perheen kinaasille. Yhteensä tunnistimme 1345 substraattia ja 3841 kinaasi-substraattiparia sytosolisesta fraktiosta, joista 165 oli listattu PhosphoSitePlus® tietokantaan. Jotta tunnistaisimme paremmin substraatteja kinaaseille jotka lokalisoituvat tumaan, kymmen kinaasia testattiin HEK solujen tumafraktiolla. Tunnistimme yhteensä 194 substraattia, joista 141 oli uniikkeja tumafraktiolle ja joista 27 löytyi PhosphoSitePlus® tietokannasta. Lopuksi, kinaasit joilla oli huomattava määrä uusia tunnistettuja substraatteja valittiin geeniontologia-analyysiin. Pystyimme linkittämään näiden substraattien geeniontologialuokitukset tunnettuihin kinaasien säätelemiin biologisiin prosesseihin. Nämä tulokset osoittavat, että ^{18}O-ATP-kinaasitesti yhdistettynä massaspektrometriaan on toimiva laajan mittakaavan menetelmä tunnistamaan kinaasien suoria substraatteja.

Avainsanat
proteiinikinaasi, kinaasi, CMGC, Ti4+-IMAC, LC-MS/MS, massaspektrometria

Ohjaaja tai ohjaajat
Tiina Öhman, Salla Keskitalo

Säilytyspaikka
Viikin tiedekirjasto, Helsingin Yliopisto

Muita tietoja

TABLE OF CONTENTS

ABBREVIATIONS

CDK	cyclin dependent kinase
CLK	CDC2-like kinases
CMGC	kinase family consisting of CDK, MAPK, GSK-3, and CDC-like kinases
DYRK	dual-specificity tyrosine (Y)-phosphorylation-regulated kinase
ERK	extracellular signal-regulated kinase
FSBA	5′ -[p-(fluorosulfonyl)benzoyl]adenosine
GSK-3	Glycogen synthase kinase-3
HIPK	homeodomain-interacting protein kinase
JNK	c-Jun N-terminal kinase
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry
MAPK	mitogen-activated protein kinase
MS	mass spectrometry

I. INTRODUCTION

1. Protein Kinases

Protein kinases are enzymes important for extracellular and intracellular signaling. Eukaryotic protein kinases catalyze reversible reactions of protein phosphorylation, where γ-phosphate from ATP (adenosine triphosphate) molecule is transferred to a serine, threonine or tyrosine hydroxyl group on a protein substrate (Kannan and Neuwald, 2004). Phosphorylation events are strictly controlled, dynamic and labile. Kinases react to stimuli to induce protein phosphorylation, which can be reversibly controlled by phosphatases that dephosphorylate their substrates. Kinase-mediated phosphorylation events often change protein's conformation, leading to changes in protein activity, cellular location and/or association with other proteins (Kannan and Neuwald, 2004). These phosphorylation events are one type of post-translational modifications (PTM) and they mediate most of the signal transduction in eukaryotic cells. By controlling substrate activity, kinases control many cellular processes, such as metabolism, transcription, cell cycle progression, apoptosis, and differentiation (Manning *et al.* 2002).

Kinases have highly regulated functions and their structure is well conserved among species. The human kinome consists of over 500 kinases, which means that nearly 2% of the human genome is constructed from kinases alone, highlighting the importance of this group of proteins (Manning *et al.* 2002; Taylor and Kornev, 2011). It has been estimated that around one third of eukaryotic proteins become phosphorylated at any one time (Mann *et al.* 2002). However, only around 5% of human phosphorylation sites have an annotated regulatory role (Hornbeck *et al.* 2015; Strumillo *et al.* 2019).

1.1. Structure and Function of Protein Kinases

Protein kinases are highly regulated. They function as molecular switches that are turned on and off by precise biological signals which can be triggered by molecules such as hormones or neurotransmitters, but also stress, for example nutrient deprivation may function as a trigger (Taylor and Kornev, 2011). Kinases have a highly dynamic molecular structure, and the regulatory mechanisms of kinases rely on a large Activation Segment that can be folded and unfolded (Taylor and Kornev, 2011). The internal structure of active kinases is well conserved, but inactive kinase structure has more variation enabling high substrate specificity (Taylor and Kornev, 2011) and kinase-specific functions (Nolen *et al.* 2004). All kinases consist of two distinct lobes, N-lobe and C-lobe, which differ structurally and functionally. N-lobe is smaller, it is located in the amino terminal end (N-terminal) of the protein and it consists of a five stranded β-sheet and a helical structure with a conserved αC-helix. Between β1 and β2 is a glycine-rich loop that shields the ATP and positions the γ-phosphate for transfer. The αC-β4 loop is functionally anchored to the larger C-lobe, located in the carboxy terminal end (C-terminal) of the protein. C-lobe consists mainly of α-helices and a β-sheet. A catalytic loop is positioned between β6- and β7-strands and it has conserved residues that have catalytic activity (Taylor and Kornev, 2011; Nolen *et al.* 2004) (Figure 1).

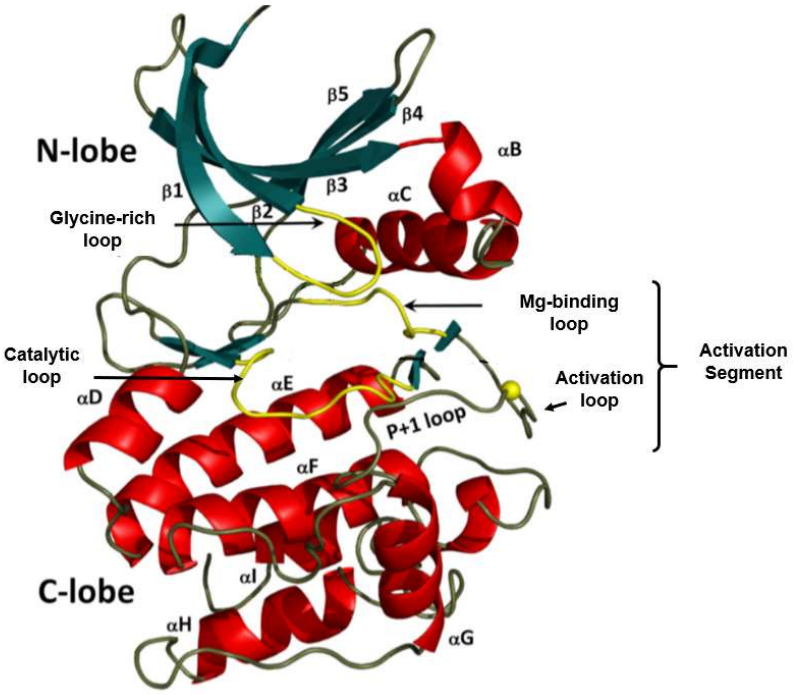


Figure 1. Structure of Protein Kinases. Kinases are bilobal as they are composed of two parts, N-lobe and C-lobe. N-lobe consists of a five stranded β-sheet (β1- β5) and an α-helix (αB and αC). Glycine-rich loop is positioned between β1 and β2 strands and it shields the ATP molecule upon γ-phosphate transfer. C-lobe consists mainly of α helices. Activation Segment is located between the two lobes and it consists of a Mg-binding loop, β9 strand, activation loop and P + 1 loop. (Modified from Taylor and Kornev, 2011)

The Activation Segment is a region between two conserved tripeptide motifs (DFG…APE) (Nolen *et al.* 2004), and it is the most variable region in the kinase core (Nolen *et al.* 2004; Taylor and Kornev, 2011). The Activation Segment starts with a Mg-binding loop which includes the DFG motif, followed by a β9-strand, the activation loop, and the P+1 loop (Nolen *et al.* 2004). The DFG motif has two conformations, DFGin and DFGout, classified by the Phe residue position respective to the αC-helix (Modi and Dunbrack, 2019). In the inactive or active DFGin conformation, the Phe residue is in contact with the αC-helix (Modi and Dunbrack, 2019), engaging two hydrophobic residues in the helix (Nolen *et al*. 2004). The DFGout is an inactive form where the Phe residue occupies the ATP binding site, leaving the αC-helix site exposed (Modi and Dunbrack, 2019). The DFGin-DFGout conformation change influences kinase's ability to bind ATP, magnesium and/or substrate and is important for kinase inhibitor design, as inhibitors only bind to specific conformations (Modi and Dunbrack, 2019). Activation loop is the regulatory site for kinases and often this region has sites for phosphorylation or other activity modulator interactions while P+1 loop is important for kinase-substrate interaction (Nolen *et al.* 2004). Activation loop and P+1 loop are integral parts of the peptide binding groove and their diverse sequences and conformations have evolutionary driven the kinase-specific functions and substrate specificity (Nolen *et al.* 2004). Canonical residues within kinase groups appear to be linked in kinase activation and substrate recognition (Kannan and Neuwald, 2004), highlighting the importance of understanding kinase structure in predicting its function and substrate specificity. 3D and crystal structures of kinases have revealed the structural

basis for distinct kinase activation mechanisms and substrate recognition for different kinase classes. For instance, serine and threonine specific kinases have a distinctive structure in their active state when compared to the kinases that phosphorylate tyrosine residues on their substrates (Krupa *et al.* 2004).

Even protein kinases themselves are often regulated by phosphorylation. This can happen by a direct phosphorylation of the Activation Segment, via an indirect prephosphorylation of the substrate (Kannan and Neuwald, 2004) or via autophosphorylation (Manning *et al.* 2002). Activation of one kinase often takes multiple kinases and/or other modulators (Taylor and Kornev, 2011) and when kinase has the desired Activation Segment structure, specific substrate binding is possible.

1.2. Kinase Families - CMGC Kinases

Kinases can be characterized and grouped based on the conserved structures in the Activation Segment. In 2002, Manning and his co-workers identified 518 human protein kinase genes and grouped them into families and subfamilies based on their similarity in sequence at the catalytic domain. Information of sequence and structure outside of the catalytical domain, known biological functions, and similar classifications among other species were also used for the hierarchical grouping (Manning *et al.* 2002). Similarities in structure and function allowed the creation of a phylogenetic tree displaying the relationships between kinase families and subfamilies (cellsignal.com; Figure 2). Tyrosine kinases form one distinct family (TK; Figure 2) whereas other kinase groups phosphorylate primarily serine and threonine residues. However, some kinases have dual-specificity and can phosphorylate serine and threonine as well as tyrosine residues. Other kinase families are TKL (tyrosine-kinase like), STE (Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases), CK1 (Casein kinase-1), AGC (Protein kinase A, G, and C subfamilies), CAMK (Calcium/calmodulin-dependent protein kinase), and CMGC (subfamilies CDK, MAPK, GSK, CDK containing family) (Figure 2A).

Within a kinase family, also subfamilies are separated from each other based on their sequence differences. One group of human protein kinases is CMGC kinase family, and its name comes from four different kinase groups: Cyclin-dependent kinases (CDK), mitogen-activated protein kinases (MAPK), glycogen synthase kinases (GSKs), and CDK-like kinases (CLKs). All the family members share similar functions and conserved amino acid residues. CMGC kinases share specific features which are important for coprotein binding. Especially the residues between the N-terminal end of the activation segment, and a so-called CMGC arginine close to the substrate phosphorylation site are specific for the CMGC kinases. These features are a strong evidence of structural linkage between substrate recognition, coprotein binding, and kinase activation. (Kannan and Neuwald, 2004). Importantly, individual subfamilies have specific functions. Relations between subfamilies and individual CMGC kinases can be observed in Figure 2B, where the distance between branches is proportional to the differences in their sequences (www.cellsignal.com). CDK, MAPK (with ERK, JNK, and p38 groups), and GSK-3 subfamilies are closely related, and these kinase subfamilies have serine/threonine specific kinases. A group of more divergent kinases consists of dual-specificity kinases and this branch includes kinase subfamilies such as CLKs, HIPKs and DYRKs.

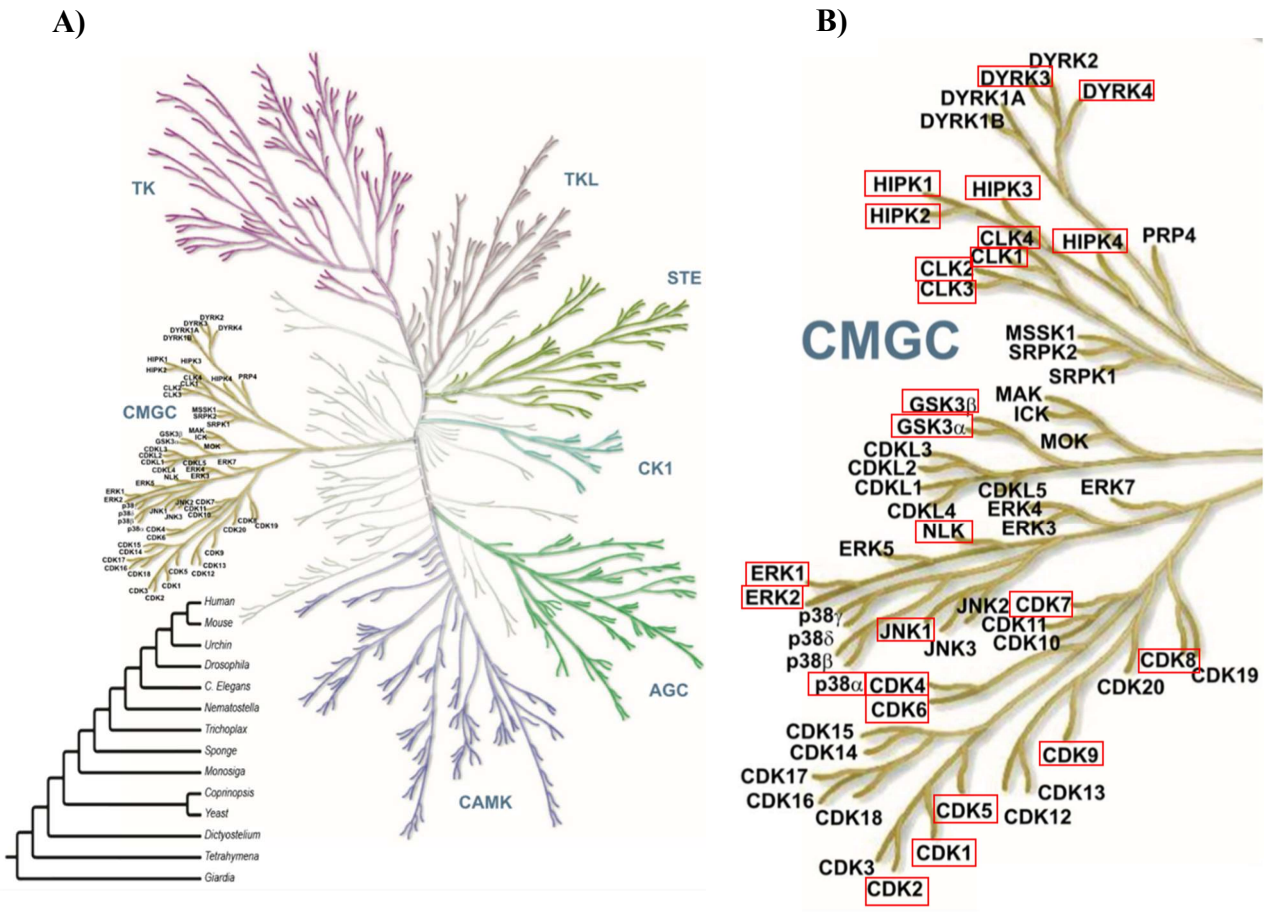


Figure 2. Overview of human protein kinases, conservation among species and relationships between CMGC kinases. A) Sequence similarity between the kinase families are shown in the kinase family tree. AGC contains PKA, PK and PKC families; CAMK, calcium/calmodulin-dependent protein kinas; CK1, casein kinase 1; CMGC contains CDK, MAPK, GSK-3, and CLK families; STE, homologs of yeast Sterile 7, sterile 11 and sterile 20 kinases; TK, tyrosine kinase; and TKL, tyrosine kinase-like. Conservation across species shown in the lower left corner. **B)** Family tree of CMGC kinases. Kinases represented in this study are marked with red. (Modified from www.cellsignal.com)

1.3. CMGC Kinase Subfamilies

Although CMGC kinases have some common, specific features, individual subfamilies have specialized functions and roles in the cell (Kannan and Neuwald, 2004). Some kinases are well-studied due to the vital cellular processes they control. CDKs regulate cell cycle (Malumbres *et al*. 2009), meanwhile MAPKs and GSK-3 kinases have a prominent role in cell signaling (Cargnello and Roux, 2011; Cohen and Frame, 2001). Other CMGC kinases, specifically dual-specificity subfamilies, DYRKs, HIPKs, and CLKs are not as extensively studied. Below, CMGC subfamilies studied in this research are shortly introduced.

Serine-Threonine Protein Kinases

1.3.1. CDK (cyclin dependent kinases)

The CDKs (cyclin dependent kinases) are involved in multiple cell functions and are best known for regulatory roles in cell cycle and transcription. CDKs are proline directed serine/threonine kinases and their enzymatic activity depends on a separate subunit called cyclin. There are approximately 30 types of cyclins (Malumbres, 2014), which belong to five different classes: A, B, C, D and E (Malumbres and Barcacid, 2009). Cyclins regulate the function of CDKs and different cyclin/CDK complexes have specified functions. Cyclins are often referred as links between the environment and the cell cycle machinery, because growth factors induce the expression of D-type cyclins that form cell cycle regulating complexes with CDK (Hydbring *et al.* 2016). In addition to their role in cell cycle regulation, mammalian cyclin/CDK complexes regulate transcription, DNA damage repair, cell death, differentiation, immune responses and metabolism (Hydbring *et al.* 2016). Apart from the typical cyclin associated CDKs, CDK5's activation depends on a non-cyclin protein CDK5R1 (p35) or CDK5R2 (p39) (Malumbres, 2014). Based on their primary functions, CDKs can be divided into two groups: cell cycle regulating CDKs and transcriptional CDKs.

Cell cycle regulating CDKs

Cell cycle is a tightly regulated process, which consists of five phases: G0 is a resting state, in G1 phase the cell grows and its cellular content is duplicated, in the S phase each of the chromosomes duplicate, G2 is the second growth phase which leads the cell to the M, mitotic, phase where the cell finally divides into two daughter cells. Cell cycle is controlled by checkpoints, where possible defects during DNA synthesis and chromosome segregation are detected. If checkpoints are activated, cell cycle arrest is induced through modulation of CDK activity. Cell cycle arrest has an important role in repairing detected defects and preventing their transmission to the daughter cells (Malumbres *et al*. 2009).

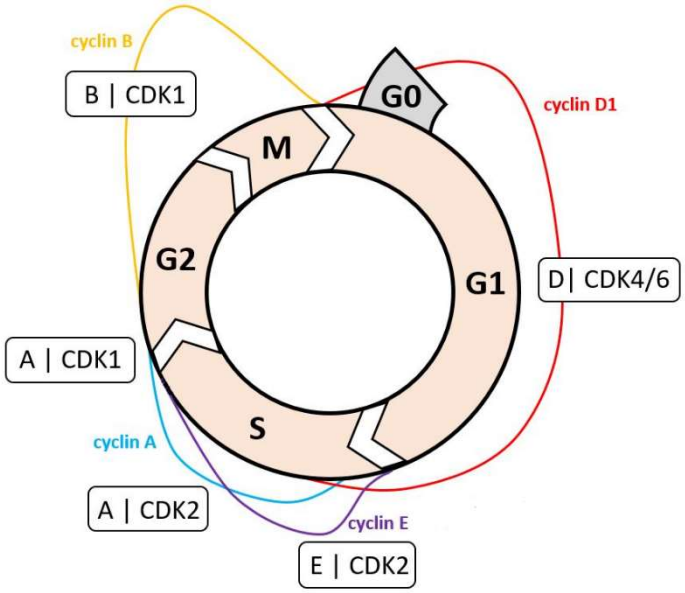


Figure 3. Cell cycle regulation with cyclin/CDK complexes. All phases of cell cycle, G0 (senescence), G1 (first growth phase), S (DNA replication), G2 (growth and preparation for mitosis), and M (mitotic phase, cell division) are represented. Approximate expression levels of cyclins shown as they fluctuate during cycle. Cyclin/CDK complexes regulating each phase of the cell cycle is marked, starting with D/CDK4 or D/CDK6 complex.

CDKs that regulate cell cycle can be divided into interphase CDKs (CDK2, CDK4 and CDK6) and to a mitotic CDK1 (Malumbres and Barbacid, 2009). There are multiple cyclin/CDK complexes, but only a few specific ones directly regulate cell cycle. Mitogenic signaling induces the expression of cyclin D and the correct folding and transport of CDK4 and/or CDK6 into the nucleus (Malumbres and Barbacid, 2005). Once expressed, cyclin D forms a complex with CDK4 or CDK6 in the G1 phase of cell cycle and these complexes phosphorylate transcription factors (Hydbring *et al.* 2016) and transcription regulating proteins, such as members of the retinoblastoma family: pRb, p107 and p130 (Malumbres and Barbacid, 2005). Upon transition from G1 to S phase, cyclin E becomes upregulated and it activates CDK2 (Hydbring *et al.* 2016). Cyclin E/CDK2 complex phosphorylates a larger population of cell cycle regulating proteins, as well as proteins involved in histone modification (NPAT, HIRA, CBP/p300), DNA replication (Cdt1) and repair (BRAC1, Ku70), centrosome duplication and maturation (CP110, Msp1), and its own inhibitor $p27^{Kip1}$ to subject it for degradation by proteasome (Malumbres and Barbacid, 2005). When cells enter S phase, cyclin A becomes expressed and it forms complexes with CDK2 and CDK1. A/CDK1 and A/CDK2 complexes share several substrates, such as DNA replicating licencing factors (MCMs) and cell cycle regulating proteins pRb, p53, BARD1 and BRCA. It remains unclear whether A/CDK1 and A/CDK2 complexes have differential roles in the S to G2 transition (Malumbres and Barbacid, 2005). Additionally, cyclin E/A/CDK2 has been shown to phosphorylate transcription factors MYC and B-MYB activating target genes during cell cycle (Hydbring *et al.* 2016). CDK3 is highly related to CDKs 1 and 2 and it is also hypothesized to inactivate pRb (Manning, 2005). CDK3 binds with cyclin C and participate in G0 to G1 progression, but the actual role of CDK3 in cell cycle remains unsolved (Hydbring *et al.* 2016; Malumbres and Barbacid, 2005). During G2 phase, B-type of cyclins are produced and CDK1 forms a complex preferably with cyclins B1 and B2. This complex phosphorylates a wide range of substrates and controls the progression of mitotic phase and rest of the cell cycle, influencing also DNA replication and centrosome and chromosomal function

(Hydbring *et al.* 2016). It is estimated that B/CDK1 complex phosphorylates over 70 different proteins in mammalian cells (Malumbres and Barbacid, 2005).

Due to the high number of substrates that cyclin/CDK complexes phosphorylate, these molecules play role in multiple cellular and biological processes. Non-canonical functions for cell cycle regulating cyclins include muscle differentiation, neuronal development, immune system differentiation and function, regulation of cell migration, and regulation of energy production and metabolism, both at the cellular and at the organismal level (Hydbring *et al.* 2016).

Transcriptional CDKs

Some CDKs regulate transcription by phosphorylating substrates involved in the process, such as RNA polymerase II. RNA polymerase II transcribes all protein-coding and nearly all non-coding RNA genes and is therefore a crucial part of transcription machinery (Allen and Taatjes, 2015). CyclinC/CDK8 and cyclinT1/CDK9 complexes regulate transcription by phosphorylating the carboxy-terminal domain of the larger subunit of RNA polymerase II (Malumbres and Barbacid, 2005). CyclinC/CDK8 targets also general transcription initiation factor IIH (TFIIH) to regulate transcription (Akoulitchev *et al.* 2000). TFIIH is involved in promoter clearance and transcription progression and it consists of six subunits in addition to CDK-activating kinase, CAK, which is composed of cyclin H, CDK7 and MAT1. CyclinC/CDK8 phosphorylates cyclin H to repress the function of CAK (Malumbers, 2005). CDK8 is also part of Mediator complex which regulates processes important for transcription, such as the organization of chromatin architecture and the regulation of RNA polymerase II pre-initiation, initiation, re-initiation, pausing and elongation (Allen and Taatjes, 2015).

Central mechanisms of transcription elongation regulation involve sequential phosphorylation of C-terminal domain of RNA polymerase II by transcription factors TFIIH and the positive transcription elongation factor b, P-TEFb (Peterlin and Price, 2006). In mammalian cells, CDK9 functions as a catalytic subunit for the P-TEFb complex, and it is combined with a cyclin T1 or T2 and other polypeptides (Fu *et al*. 1999). In addition to T-type cyclins, CDK9 also forms complexes with cyclin K, which has only 29% amino acid similarity at the cyclin box compared to cyclin T1 although no difference has been observed in substrate specificity between these two complexes (Fu *et al*. 1999).

CDK5

CDK5 does not form complexes with cyclins, instead it functions in combination with a non-cyclic molecule p35 or p39. CDK5 regulates multiple cellular processes within cytoarchitecture of the nervous system, such as neuronal migration, actin dynamics, microtubule stabilization and transport, cell adhesion, axon guidance, and synaptic structure and plasticity (Dhavan and Tsa, 2001).

Deregulation of CDK5 is involved in pathogenesis in cytoskeletal abnormalities and can cause neurodegenerative diseases (Patrick *et al.* 1999). One major factor for CDK5 involvement in diseases is the cleavage of p35 into p25 which is truncated and a more stable form of p35 (Patrick *et al.* 1999; Dhavan and Tsa, 2001). Cleavage results in p25/CDK5 complex translocation in the nucleus, where it has a prolonged activation (Liu, 2017) and drastic effects in neurons, such as neurite retraction, microtubule collapse and apoptosis (Dhavan and Tsa, 2001). The p25/CDK5 complex hyperphosphorylates Tau protein, reducing its ability to bind microtubules and causing microtubule destabilization. The p25/CDK5 kinase has shown to become accumulated in the neurons of brains of patients with Alzheimer's Disease (AD), suggesting a mechanism for the pathogenesis in AD (Patrick *et al.* 1999). In addition to CDK5 activity in neurons, the kinase affects biological processes in non-neuronal cells including gene expression, cell migration, apoptosis and myogenesis, and it is involved is some pathological processes, such as cancer, senescence, diabetes, immune dysfunction and inflammation (Liu *et al.* 2017). CDK5 regulates epithelial cell adhesion and migration through various pathways. For instance, CDK5 mediates phosphorylation of talin in epithelial cell lines, which increases adhesive strength and therefore regulates migration (Tripathi and Zelenka, 2010).

1.3.2. MAPK (mitogen-activated protein kinases)

In mammals, MAPK family consists of 14 different serine/threonine kinases. These can be divided into seven different groups – there are four types of conventional kinases and three types of atypical kinases. Conventional kinases include extracellular signal-regulated kinases 1 and 2 (ERK1/2), c-Jun amino-terminal kinases (JNKs), p38 MAPKs, and ERK5. Atypical MAPKs comprises ERK3/4, ERK7/8, and Nemo-like kinase (NLK). Conventional MAPKs become activated through three-tiered kinase cascades. Stimulus, such as mitogens, cytokines and cellular stress activate MAPKKK, leading to phosphorylation of MAPKK which stimulates MAPK activation by dual phosphorylating on threonine and tyrosine residues at the MAPK's conserved Thr-X-Tyr motif (Cargnello and Roux, 2011) (Figure 4). Atypical MAPKs are not organized in this manner, and less is known about the molecular mechanisms resulting in their activation. This may be due to the absence of Thr-X-Tyr motif in ERK3/4 and NLK, where tyrosine residue is replaced by either glycine or glutamine (Cargnello and Roux, 2011). Both atypical and conventional MAPKs phosphorylate their substrates in serine or threonine residues followed by proline. This is one factor resulting in their specificity and thus MAPKs are often described as proline-directed kinases (Cargnello and Roux, 2011). MAPKs phosphorylate a wide range of substrates and it has been estimated that each MAPK has 200-300 substrates (Cuadro and Nebrada, 2010).

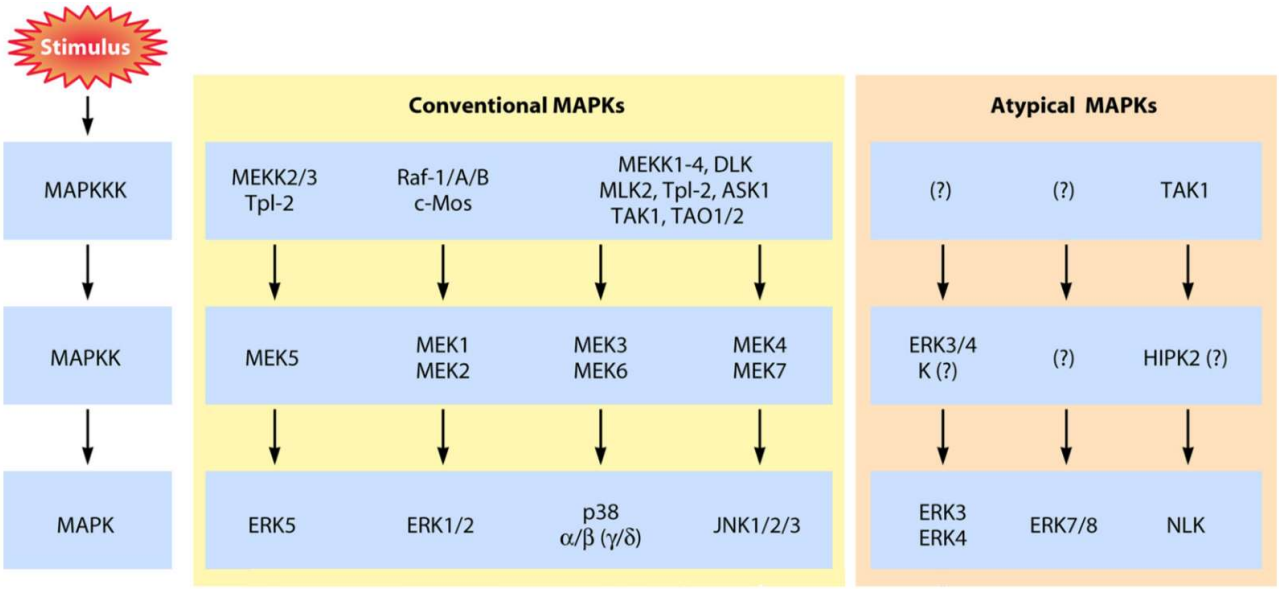


Figure 4. Summary of MAPK signaling cascades. Stimulus, such as mitogens, cytokines and cellular stress activate MAPKKK, leading to phosphorylation of MAPKK and MAPK. Conventional MAPKs (ERK5, ERK1/2, p38, JNK) function downstream of MEK proteins, which are phosphorylated by upstream MEKK or alternatively Raf-1 or c-Mos (upstream from ERK1/2). Kinases functioning upstream from atypical MAPKs is not well known. It is hypothesized that TAK1/HIPK2/NLK forms one signaling cascade. (Modified from Corgnello and Roux, 2011)

Conventional MAPKs

ERK1 (MAPK3) and ERK2 (MAPK1) are multifunctional serine/threonine kinases that share a similar structure and their amino acid sequences are at least 83% similar (Meloche and Pouysségur, 2007; Cargnello and Roux, 2011). All known activators of ERK1 activate ERK2 in parallel, and due to the same activation mechanisms and similar functions, these kinases are often referred as ERK1/2 (Roskoski, 2012). ERK1/2 module is activated mainly in response to extracellular stimulus, activating signaling pathways via cell surface receptors. These receptors respond to various stimuli, for instance insulin or different growth factors, such as platelet-derived growth factor (PDGF) and epidermal growth factor (EGF). They have also been shown to become activated by G-protein coupled receptor (GPCR) ligands, cytokines, osmotic stress and microtubule organization (Cargnello and Roux, 2011). Activation occurs via MAPK signaling cascade where ERK1 and ERK2 act downstream from their MAPKKs, MEK1 and MEK2. MEK1/2 activates ERK1/2 by phosphorylating its conserved Thr-Glu-Tyr motif in its activation loop (Cargnello and Roux, 2011).

ERK1/2 is located in the cytoplasm of quiescent cells but translocate in the nucleus upon activation by extracellular stimulus. Some ERK1/2 substrates localize in the cytoplasm, some in the nucleus, but membrane and cytoskeleton associated substrates have also been reported (Cargnello and Roux, 2011). ERK1/2 has an important role in cell cycle regulation and in the control of cell proliferation (Meloche and Pouysségur, 2007), as it induces positive regulators of cell cycle (Cargnello and Roux, 2011). ERK1/2 signaling regulates cell growth and promotes G1 progression by targeting D-type

cyclin pathway, stabilizing c-myc, regulating p21 and p27 expression, and by downregulating antiproliferative genes (Meloche and Pouysségur, 2007). It is still unknown if ERK1 and ERK2 have distinct functions in the cell and these functions would be very difficult to pinpoint (Roskoski, 2012).

There are three known JNK isoforms, which are highly homologous, being greater than 85% identical. JNK1 (MAPK8) and JNK2 (MAPK9) have similar locations and they are distributed in multiple tissues. JNK3's (MAPK10) tissue distribution is more specific and it is restricted into neuronal tissue, testis, and cardiac myocytes (Bode and Dong, 2007). JNKs have a conserved Thr-Pro-Tyr motif in their activation loop (Cargnello and Roux, 2011) and a complete JNK activation is dependent on dual phosphorylation of threonine and tyrosine residues within this motif (Bode and Dong, 2007). JNK phosphorylation depends on specific stimulus, cell type, and temporal aspects. JNK has been reported to induce apoptosis but also to be associated with cell proliferation and enhancement of cell survival (Bode and Dong, 2007). The most well-known JNK substrate is c-Jun but JNKs also phosphorylate a variety of transcription factors such as p53, Elk-1, c-Myc, HSF-1, and STAT3 (Cargnello and Roux, 2011) and apoptosis related proteins from the BCL-2 family (Bode and Dong, 2007).

MAPK p38 has different isoforms – p38α, p38β, p38γ, and p38δ. These isoforms have overlapping substrate specificities and therefore they share similar functions (Cuadro and Nebrada, 2010). Isoform p38α is expressed widely in various tissues, meanwhile other isoforms are tissue specific. p38 has a nuclear localization signal and inactive p38 can be found in both, nucleus and cytoplasm. However, there is incomplete information about localization of activated p38 (Cuadro and Nebrada, 2010). p38 pathway is activated in response to environmental stress, inflammatory cytokines such as interleukin-1 (IL-1) and tumor necrosis factor alpha (TNF-α), and growth factors (Cargnello and Roux, 2011; Cuadro and Nebrada, 2010). Activation of the signaling cascade leads to dual phosphorylation in a Thr-Gly-Tyr residue in p38 activation loop, changing kinase conformation and facilitating p38 for substrate binding (Cuadro and Nebrada, 2010).

p38 isoforms have a critical role in immune and inflammatory responses as their major function is to produce proinflammatory cytokines (Cargnello and Roux, 2011). They also affect cell proliferation and survival, cell-cycle progression through G1/S and G2/M transitions, and p38 can extend the MAPK pathway by phosphorylating MAPKAPK (MAPK activated protein kinases) family members such as MSKs, MNK1, MK2/3, and MK5 (Cargnello and Roux, 2011). The p38 isoforms phosphorylate a variety of substrates, including other protein kinases, nuclear proteins such as transcription factors and proteins that regulate chromatin remodeling, and cytoplasmic proteins that are involved in a variety of biological processes, including protein degradation and localization, mRNA stability, endocytosis, apoptosis, cytoskeleton dynamics and cell migration (Cuadro and Nebrada, 2010).

ERK5 is expressed with variable levels in all tissues. It is highly expressed in brain, thymus, and spleen (Cargnello and Roux, 2011). ERK5 is required for normal embryonic development and it is necessary for normal development of vascular system. ERK5 has been shown to regulate cell survival and apoptosis in mouse embryo (Yan *et al.* 2003).

Atypical MAPKs

The atypical MAPKs include ERK3/4, ERK7, and NLK. ERK3 and ERK4 are structurally very similar as their amino acid structure is 73% identical (Cargnello and Roux, 2011). ERK3/4 lack the Tyr residue in their activation loop, having a Ser-Glu-Gly motif instead, thus making them atypical MAPKs. ERK3/4 are shown to become activated by class I p21-activated kinases which phosphorylate Ser residue in ERK3/4 activation loop (Deléris *et al.* 2011). It is also known that ERK3/4 can be inhibited by dual-specificity MAPK phosphatase DUSP2-mediated dephosphorylation (Perander *et al.* 2017). The downstream signaling of ERK3/4 is not well known, as the only known substrate for ERK3/4 is MAPK activated protein kinase MK5 (Cargnello and Roux, 2011; Perander *et al.* 2017). The role of ERK4 in biological processes is unknown, but ERK3 has been shown to affect cell proliferation, cell cycle progression and cell differentiation (Cargnello and Roux, 2011).

ERK7/8 module is poorly understood. ERK7 can be autophosphorylated, but there are no known upstream kinases for the ERK7 (Lau and Xu, 2019). However, it has been shown that the activation loop of ERK7 motif includes Thr-Glu-Tyr motif, suggesting that a MAPKK may exist (Cargnello and Roux, 2011). Align with these findings, ERK8 has been shown to autophosphorylate in its activation loop (Cargnello and Roux, 2011).

Nemo-like kinase, NLK, modifies cellular signaling and it has been reported to have a role in nervous system development, bone formation and cancer activity (Ishitani and Ishitani, 2013). NLK is involved in the Wnt signaling pathway, where it is activated downstream of Wnt-1 and Wnt-5a (Cargnello and Roux, 2011). NLK can also be activated by various cytokines, including interleukin 6, granulocyte colony stimulating factor (G-CSF), and transforming growth factor β (TGF-β). However, the molecular mechanisms behind the activation process remain unclear (Cargnello and Roux, 2011; Ishitani and Ishitani, 2013). NLK has similar structure as ERK2 but does not have a Thr-X-Tyr motif in its activation loop, which is characteristic for conventional MAPKs. Instead, NLK has a Thr-Gln-Glu motif, similar to the Thr-X-Gln motif in CDK1. In fact, NLK and CDK1 have 38% similarity in their kinase domain (Cargnello and Roux, 2011). When overexpressed, NLK has been shown to autophosphorylate on threonine 286 in result of dimerization (Cargnello and Roux, 2011; Ishitani and Ishitani, 2013). There are no known MAPKK for NLK, however HIPK2 has been shown to either phosphorylate NLK or to induce its autophosphorylation (Cargnello and Roux, 2011; Kanei-Ishii *et al*. 2004). NLK has been shown to negatively regulate Notch-dependent transcriptional activation (Ishitani *et al.* 2010).

1.3.3. GSK-3 (Glycogen synthase kinase-3)

Glycogen synthase -3 (GSK-3) is a serine/threonine kinase that was originally recognized as a glycogen metabolism regulator. Later, it became associated with signaling by insulin, growth factors and nutrients, and as a regulator of cell fate in embryonic development. It has a role in control of cell division, apoptosis and microtubule function, being involved in diseases such as cancer, diabetes, and neurodegenerative disorders (Cohen and Frame, 2001). GSK-3 has high activity in resting cells

and its enzymatic activity typically decreases rapidly in response to extracellular signaling. GSK-3 can be regulated in multiple different ways including phosphorylation, protein complex association, priming and substrate specificity, subcellular localization, and proteolytic cleavage (Medina and Wandosell, 2011). It works downstream from some well-known protein kinases such as PKB and Akt, connecting it to multiple signaling pathways, and it is also an important kinase in Wnt signaling pathway (Cohen and Frame, 2001). GSK-3 regulates cell cycle by phosphorylating cyclin D and subjecting it for degradation, inhibiting D/CDK4/6 complex formation (Cohen and Frame, 2001). GSK-3 has been shown to phosphorylate multiple transcription factors, such as c-Myc and c-jun, as well as microtubule-associated protein Tau, which is linked to neurodegenerative disorders, such as Alzheimer's disease (Cohen and Frame, 2001). In mammals there are two isoforms of GSK-3; GSK-3α and GSK-3β. They are 98% identical within their ATP binding pocket but they differ in size (GSK-3α 51kDa and β 47kDa) and in the structure of N- and C-terminal ends (Medina and Wandosell, 2011). However, studies with mice have shown that they have distinct biological roles, because homozygous inactivation of GSK-3β is embryonically lethal but GSK-3-α null mice are viable (Hoeflich *et al.* 2000).

CMGC Kinases with Dual Specificity

Dual specificity kinases can autophosphorylate or phosphorylate substrates on both types of residues, serine/threonine and tyrosine residues. CMGC kinase family have subfamilies with dual specificity, and the subfamilies studied here are CLK, DYRK, and HIPK. They are phylogenetically diverged from serine/threonine directed kinases (Figure 2A) and therefore have slightly different structure and unique functions when compared to the serine-threonine directed kinases within the same family. However, they are less studied when compared to the well-known kinase subfamilies such as MAPKs and CDKs.

1.3.4. CLK (cdc2-like kinase)

CLKs are dual specificity kinases that have an ability to autophosphorylate at tyrosine residues, but phosphorylation of their substrates is restricted to serine and threonine residues (Nayler *et al.* 1997). Cdc2-like kinases form a group of four protein kinases, CLK1-4. They are considered to belong in a LAMMER family due to their conserved 'EHLAMMERILGPLP' motif in their kinase domain. CLKs share some of the features common with serine/threonine kinases within the CMGC family. Particularly, CLK1 has proline-directed activity similarly to MAPKs and CDKs (Bullock *et al.* 2009). CLKs locate in the nucleus and they regulate mRNA splicing in mice (Nayler *et al.* 1997). Together with another CMGC family dual-specificity kinase subfamily SPRK, CLKs phosphorylate serine/arginine rich (SR) proteins that are components of the spliceosomal complex (Bullock *et al.* 2009; Nayler *et al.* 1997). CLK1, CLK2, and CLK4 prevents chromatin breakage by phosphorylating Aurora B at the serine 331 residue, activating it in late cytokinesis (Petsalaki and Zachos, 2015). In addition, CLK2 has been shown respond to insulin and to directly phosphorylate PSG-1α,

suppressing hepatic gluconeogenesis and affecting the insulin/Akt metabolic pathway (Rodgers *et al.* 2010).

1.3.5. DYRKs (Dual-specificity tyrosine (Y)-phosphorylation-regulated kinase)

DYRK, Dual-specificity tyrosine (Y)-phosphorylation-regulated kinase, is a kinase family with three different subfamilies: DYRK, HIPK (homeodomain-interacting protein kinases), and PRP4 (pre-mRNA processing protein 4 kinase).

Kinases in the DYRK subfamily regulate cellular functions, such as cell survival, cell differentiation, gene transcription and endocytosis (Aranda *et al.* 2011). DYRK subfamily has 5 members in mammals: DYRK1A, DYRK1B, and DYRK2-4. These family members have two conserved regions in their genome: a kinase domain and a DYRK-homology box (DH) (Figure 5). Phylogenetically, DYRKs can be divided into two subgroups: DYRK1A and DYRK1B belong to class I and DYRKs 2-4 belong to class II. Class I DYRKs have a nuclear localization signal at their N-terminal end and within the C-terminal region they have a PEST motif which is rich in proline, glutamic acid, serine, and threonine residues. The three class II DYRKs share a N-terminal autophosphorylation accessory region, which is required for DYRK2-4 autophosphorylation (Aranda *et al*. 2011; Papadopoulos *et al.* 2011).

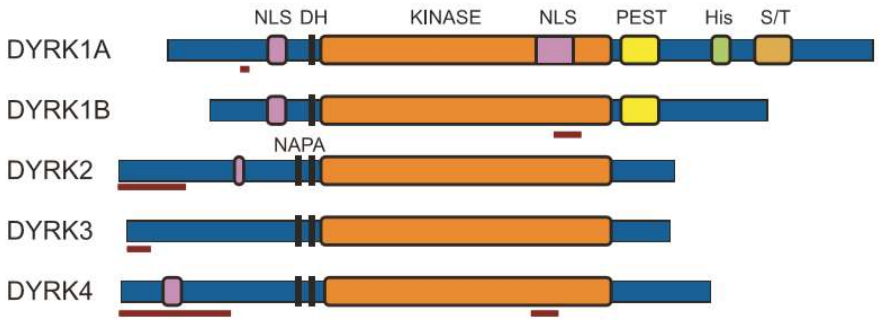


Figure 5. Structural relationships between DYRK family members. DYRK homology box (DH) is located upstream from the kinase domain (orange) and is present in every DYRK family member. DYRK1A and DYRK1B have a nuclear localization signal (NLS) upstream from the DH box and a PEST motif downstream from the kinase domain. DYRK2, DYRK3 and DYRK4 are class II DYRKs because they have a N-terminal autophosphorylation accessory region (NAPA-signal). (Aranda *et al.* 2011)

The activation loop of DYRK kinase domain has a YxY sequence, which tyrosine residues have been shown to autophosphorylate *in vivo* (Aranda *et al.* 2011). Phosphorylation of the tyrosine residues is restricted only to autophosphorylation, and despite of DYRKs dual specificity, there are no known DYRK substrates with tyrosine phosphorylation (Aranda *et al.* 2011). DYRK activity is mainly controlled by autophosphorylation at the active site, but also other control mechanisms have been suggested. Phosphorylation outside of the activation loop, regulatory proteins and subcellular localization might affect DYRK activity. DYRK1A and DYRK1B have a nuclear localization because of the localization signal but DYRK2, DYRK3 and DYRK4 are characterized as cytoplasmic. However, DYRK3's localization is predominantly nuclear (Aranda *et al.* 2011) and DYRK4 localization can depend on the splice variant (Papadopoulos *et al.* 2011). DYRKs function often as priming kinases, meaning that the phosphorylation by DYRK at a specific residue allows another kinase to phosphorylate in a nearby motif with a more effective response. DYRKs for instance prime phosphorylation by GSK-3 kinases (Aranda *et al.* 2011). DYRK1A, DYRK1B and DYRK3 are characterized anti-apoptotic, as DYRK1A and DYRK3 phosphorylate a p53 inhibitor SIRT1, DYRK1A phosphorylates and inactivates apoptotic caspase 9 and phosphorylates HIP1, blocking cell death in neuroprogenitor cells, and DYRK1B phosphorylates p21 and p27, which are inhibitors of CDKs (Aranda *et al.* 2011). DYRK2, on the other hand, is proapoptotic and it phosphorylates p53 on its S46 residue (Taira *et al*. 2007). DYRKs also mediate endocytosis, as DYRK3 was identified in genome-wide studies of human kinases in clathrin- and caveolae/raft-mediated endocytosis (Pelkmans *et al*. 2005) and DYRK1A has been shown to phosphorylate clathrin coated vesicle associated proteins in rat brain extracts, including proteins such as clathrin heavy chain, endophilin 1 and dynamin 1 (Murakami *et al*. 2009). Other endocytosis related proteins, such as synaptojanin1 and the scaffold protein amphyphisin1 are also substrates of DYRK1A (Aranda *et al.* 2011).

Homeodomain interacting protein kinases (HIPKs) are one of the subfamilies of DYRK kinases. DYRKs and HIPKs become activated through same mechanisms. All HIPKs have dual-specificity and they autophosphorylate on tyrosine residues in the activation loop to gain full activity (van der Laden *et al.* 2015). HIPK subfamily consist of four members, HIPK1-3, which have about a 90% similar domain structure, and HIPK4 which is more distantly related (van der Laden *et al.* 2015). HIPK2 is the most studied member of HIPK subfamily and it has been shown to promote β-catenin stabilization in Wnt signaling pathway (Verheyen and Gottardi, 2010) and to promote apoptosis via downregulation of a transcriptional corepressor CtBP (Zhang *et al.* 2003). It is suggested that HIPK2 functions upstream in an atypical MAPK signaling cascade, where it directly phosphorylates NLK (Corgnello and Roux, 2011). Overall, HIPK2 have been shown to have regulatory functions in cell differentiation, proliferation, apoptosis, and stress responses to DNA damage and hypoxia (van der Laden *et al.* 2015). Although other HIPKs are less studied, HIPK1 is associated with the AIP1-ASK1 complex and it is identified as a novel signal transducer involved in TNFα-induced ASK1-JNK/p38 activation (Li *et al.* 2005). These findings indicate, that HIPKs are involved in cell signaling and cellular processes, such as apoptosis.

2. Clinical Relevance of Protein Kinases

As presented in detail above, kinases are strictly regulated, highly specific and they function in various cellular processes. Chromosomal mapping of kinases revealed that at least 164 kinases were in the same loci associated with tumors and an additional 80 kinases were linked with other major diseases (Manning *et al.* 2002), indicating that kinase genes are potentially disease-modifying. Kinase genes have also been identified as druggable, meaning that they have the capability to bind a small molecule with suitable chemical properties (Hopkins and Groom, 2002), which makes kinases promising drug targets. Indeed, over 20% of all putative drug targets are kinases (Hopkins and Groom, 2002) and by 2014, a total of 260 kinases have been targeted by 19,000 kinase inhibitors (Sugiyama and Ishihama, 2016) and by 2017 a total of 38 kinase inhibitors were approved by the United States Food and Drug Administration (FDA) (Ferguson and Gray, 2018). One major disease class treated with kinase inhibitors is cancer, and interaction analysis has identified CMGC kinases as cancer hubs (Varjosalo *et al.* 2013), highlighting the importance of this kinase family. Previously, targeting receptor tyrosine kinase signaling has been studied successfully, but now the focus in cancer treatment is shifting towards two interesting areas of kinase inhibition: transcriptional kinases and kinases that regulate immune responses (Ferguson and Gray, 2018). Cancers require a high production of proteins and therefore transcription in cancer cells is accelerated. By inhibiting this transcriptional activity, it might be possible to tackle the proliferation and survival of cancer cells. Transcriptional CDKs are potential targets for inhibition in so-called transcription-addicted cancers, and CDK7 especially is a potential drug target to fight advanced solid tumors. There are also developments for CDK8 and CDK9 inhibitors, some of them already in Phase I and II clinical trials by 2018 (Ferguson and Gray, 2018).

Outside of oncology, kinases are possible drug targets for diseases in autoimmunity and inflammation, infection, degenerative conditions, and unfolded protein response causing prion diseases (Ferguson and Gray, 2018). Autoimmune and inflammatory diseases are often regulated by intracellular kinase signaling pathways and especially MYD88 related, JAK-STAT, Bruton tyrosine kinase (BTK) – spleen tyrosine kinase (SYK) and p38 MAPK pathways have been studied with a goal to target them with inhibitors (Ferguson and Gray, 2018). Targeting p38 MAPK subfamily member p38α with ATP competitors have shown anti-inflammatory effects in mice models, however clinical trials have failed (Ferguson and Gray, 2018; Zhang *et al.* 2007). The p38 MAPK pathway is still a point of interest and both upstream kinases and downstream effectors of p38 are considered as potential targets for inhibition (Zhang *et al.* 2007), highlighting the importance of comprehensive understanding of components involved in the pathway. For instance, p38 substrate MK2 has been identified as a potential drug target and both ATP-competitive and non-ATP-competitive inhibitors have been studied (Zhang *et al.* 2007), however such inhibitors are yet to be approved by FDA (Ferguson and Gray, 2018). Production of JAK inhibitors have been more successful and a pan-JAK inhibitor tofacitinib has been approved by the FDA for the treatment of rheumatoid arthritis (Ferguson and Gray, 2018). CDK5 and leucine-rich repeat serine/threonine-protein kinase 2 (LLRK2) have been identified as possible targets to tackle neurodegenerative diseases, however the possibilities remain unexplored (Ferguson and Gray, 2018). When in complex with p25, CDK5 hyperphosphorylates Tau protein, causing Alzheimer's disease among other neurodegenerative

disorders (Patrick *et al.* 1999). GSK-3 has been shown to have similar functions in Tau phosphorylation, and it is hypothesized that GSK-3 inhibition could reduce the abnormal Tau phosphorylation, enabling it to bind to microtubules and therefore prevent microtubule destabilization (Cohen and Frame, 2001). Additionally, GSK-3 inhibitors have been shown to prevent the apoptosis of neuronal cells, indicating that GSK-3 inhibition might prevent neuronal apoptosis after a stroke (Cohen and Frame, 2001).

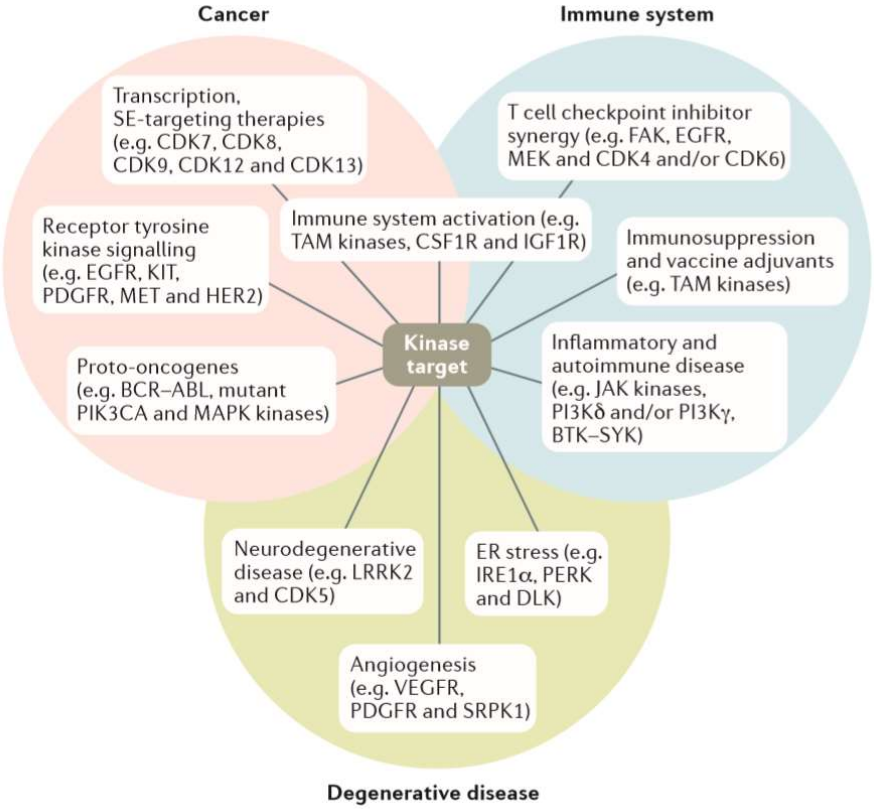


Figure 6. Possible therapeutic targets for kinase inhibition. Kinases are used as drug targets and their inhibition often is used as therapeutics in cancer, immune system disorders and degenerative diseases. Well-known targets in cancer treatment include receptor tyrosine kinases, such as EGFR and HER2 kinases. Proto-oncogenes, such as BCR-ABL and CMGC family members, MAPKs, drive abnormal proliferation of cancer cells and therefore are used as drug targets. Other members of CMGC family, transcriptional CDKs (7, 8, 9, 12, 13) are also potential targets in cancer treatment. Oncogenic kinases, CDK4 and CDK6 among others, aid cancer to evade the immune system. These kinases could be used in synergy with T cell checkpoint inhibition to fight cancer. TAM kinases could be used to treat immunosuppression and JAK kinases, for instance, are an option to help treat autoinflammatory diseases. Multiple kinases could be inhibited in degenerative diseases, where kinases causing ER stress and promoting angiogenesis could be blocked. Notably, CMGC family member CDK5 is a possible target for inhibition in neurogenerative diseases. (Ferguson and Gray, 2018)

A druggable molecule is not necessarily a good drug target – the molecule must also be linked with a disease. Kinases are identified as druggable molecules, as they can bind small molecules, such as inhibitors at the required affinity (Hopkins and Groom, 2002). Importantly, many kinases in the

CMGC family are linked with pathological processes, making them promising drug targets. However, the direct substrates for all CMGC kinases are unknown, suggesting that the important biological and disease linked functions of these kinases are not fully understood. Therefore, identification of kinase substrates is important in order to gain a comprehensive understanding of kinases.

3. Proteomic Approach to Substrate Identification

Protein kinases have on average hundreds of substrates each, forming extremely complex phosphorylation networks. Comprehensive understanding of these networks is important for exemplifying the global regulation of cellular events, as well as drug development, therapy and diagnosis (Sugiyama and Ishihama, 2016). Because of the complex nature of biological samples, phosphorylation events have been difficult to study for a couple of reasons. It is difficult to identify phosphorylated proteins in a sample because many signaling molecules tend to have a low abundance in cells and the stoichiometry of phosphorylated peptides in a sample is generally low (Mann *et al*. 2002; Sugiyama and Ishihama, 2016). Also, recognizing the phosphorylated site within a protein or peptide is difficult because of three main reasons. First, phosphorylation events are dynamic, and phosphatases may dephosphorylate sites prior to the identification. Second, phosphoproteins are heterogenous, meaning that proteins become phosphorylated at multiple sites and a biological sample may include heterogenous population of the same protein with different phosphorylated forms. Third, most analytical techniques have a limited dynamic range and it may be difficult to identify minor phosphorylation sites (Mann *et al.* 2002).

Previously, kinase-substrate interactions have been studied *in vivo* with enzyme inhibitors and functional knockouts, where the amino acid bearing the modification is deleted (Mann and Jensen, 2003; Meng *et al.* 2008). However, it is difficult to determine if the substrate is directly phosphorylated by the kinase of interest or if the phosphorylation occurred by downstream molecules or by other kinases (Sugiyama and Ishihama, 2016). To overcome problems with *in vivo* studies, *in vitro* kinase profiling using peptide libraries became useful, as the method provides an easier and wider high-throughput profiling of sequence preference of kinases (Sugiyama and Ishihama, 2016). *In vitro* profiling method includes two main approaches: random peptide libraries or array-based approaches and MS-based methods.

3.1. Peptide Libraries and Array-based Methods

Substrates can be identified with synthetic peptide libraries, where synthetic penta- and heptapeptides are expressed on solid-phase beads, phosphorylated by an *in vitro* kinase in the presence of radiolabeled [γ-^{32}P]-ATP, which can be detected based on its radioactivity, and the phosphopeptides can be identified with amino acid sequencing referred as Edman degradation (Sugiyama and Ishihama, 2016; Mann *et al*. 2002). Peptide libraries are also used in array-based profiling of kinase substrates. This approach can be done in-solution, where kinase reaction is

conducted before peptide immobilization on an avidin-coated membrane array, or on a microarray format, where peptides are immobilized on a glass-slide before the kinase reaction (Sugiyama and Ishihama, 2016). Because the sequence and position of immobilized peptides is known, array-based methods do not require amino acid sequencing, like the traditional use of peptide libraries (Sugiyama and Ishihama, 2016). Array method is very sensitive, however, gaining a full coverage of the phosphorylated peptides is difficult because the population of chosen peptides does not represent the whole proteome of a cell (Mann *et al.* 2002). In addition to radioisotype labeling, there are several other reporter mechanisms for phosphorylation recognition, such as fluorescence-labeled anti-pY antibody, a fluorescent phosphosensor dye, and surface plasmon resonance with a zinc(II)-chelating compound (Sugiyama and Ishihama, 2016). Peptide libraries provide a straightforward method, but it does not conclusively identify physiological substrates. Many factors affect substrate selection by a kinase, such as subcellular localization, docking domain and scaffold proteins (Sugiyama and Ishihama, 2016) and therefore the short primary sequence information of peptides is not enough for *bona fide* kinase substrate identification (Meng *et al.* 2008). For physiological substrate identification, kinase assays using protein arrays can be used. In this approach, candidate proteins are immobilized on an array instead of peptides and the *in vitro* profiling of kinase-substrate relationships can predict *in vivo* substrates, although not conclusively identify them (Sugiyama and Ishihama, 2016). However, proper phosphorylation by a kinase requires that the substrate has the correct secondary structure, as well as correct PTMs, such as priming phosphorylation or glycosylation, which causes that denatured proteins in arrays may not get phosphorylated in the correct way (Meng *et al.* 2008). This method is laborious and costly on large scale and therefore not suitable for high-throughput profiling of kinase substrates (Meng *et al.* 2008; Sugiyama and Ishihama, 2016).

3.2. Identification of Phosphopeptides with MS

Mass spectrometry (MS) based proteomic methods are used for high-throughput protein sequencing and its different applications such as kinome profiling. In mass spectrometry-based proteomics, a sample (often cell lysate or tissue sample) is treated with a protease for protein digestion and then analyzed with liquid-chromatography combined with MS (LC-MS) (Sugiyama and Ishihama, 2016). With MS-based methods it is possible to localize PTMs, such as phosphorylation, on large scale. Although MS is used for phosphorylation site identification, this approach often fails to identify phosphorylated peptides in complex peptide samples generated with protein digestion (Zhou *et al.* 2013). This is due to the low abundance of signaling molecules within cells, generally low stoichiometric amount of phosphopeptides in biological samples, and the low sensitivity of negatively charged phosphopeptides in positive-mode MS, when compared to unmodified peptides in the sample (Sugiyama and Ishihama, 2016; Mann *et al*. 2002).

Therefore, phosphopeptide enrichment methods are important to increase the fraction of phosphorylated peptides in the peptide mixture analyzed by MS. Phosphopeptides can be enriched by immunoprecipitation but it is only applicable for phosphotyrosines (Zhou *et al.* 2013). Currently, global phosphopeptide enrichment is commonly done by chromatographic methods (Zhou *et al.* 2013). Some common prefractionation methods include hydrophilic interaction liquid

chromatography (HILIC), electrostatic repulsion hydrophilic interaction chromatography (ERLIC), and strong cation exchange (SCX) chromatography (Tape *et al.* 2014), from which SCX became very popular due to the negative charge of phosphopeptides which is retained also in acidic conditions, enabling the separation of phosphorylated and non-phosphorylated peptides (Zhou *et al.* 2013). Chromatographic methods using chelation/coordination chemistry, such as metal oxide chromatography (MOAC) and immobilized metal ion affinity chromatography (IMAC) became the first truly successful strategies for phosphopeptide enrichment (Zhou *et al.* 2013). Selectivity of IMAC is based on Fe^{3+} or Ga^{3+} which bind phosphopeptides in the presence of chelating ligands such as iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA). However, aspartate- and glutamate-rich negatively charged non-phosphorylated peptides and electron donor amino acids, such as histidine, tend to bind IMAC to material, causing nonspecific adsorption and therefore interfere with the effective identification of phosphopeptides (Zhou *et al.* 2013; Sugiyama and Ishihama, 2016; Mann *et al*. 2002). To address this issue, acidic residues can be blocked using methyl esterification before the enrichment (Zhou *et al.* 2013, Mann *et al.* 2002), pH and solvent can be optimized, or additives such as acidic amino acids can be used (Sugiyama and Ishihama, 2016). Another solution is to use metal oxide affinity chromatography instead of IMAC and especially titanium oxide TiO_2 has been proven effective for enrichment, because chelated titanium ions bind specifically to phosphopeptides (Tape *et al*. 2013). Recently, Zhou *et al.* (2013) developed a Ti^{4+}- IMAC material, which uses monodisperse microspheres, flexible linkers between the microspheres and the chelating ligand, and Ti^{4+} for the coordination bond instead of the traditional Fe^{3+} or Ga^{3+}. Ti^{4+}- IMAC material binds phosphopeptides with the help of coordination interactions of the titanium ion and oxygens in phosphate groups, as shown in Figure 7 (Zhou *et al.* 2013).

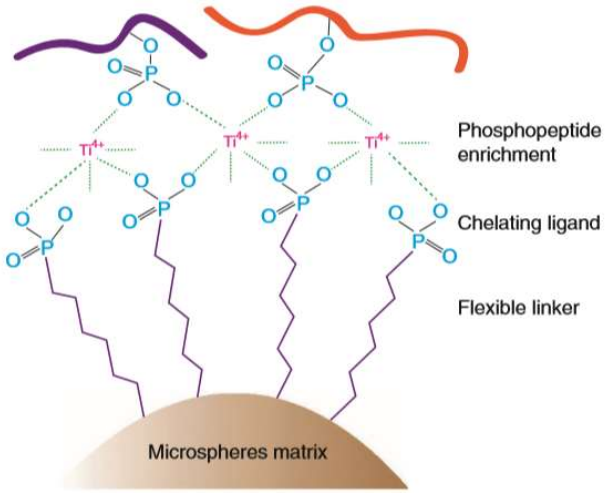


Figure 7. Phosphonate groups are used as chelating ligands that are linked to microsphere matrix. Ti^{4+} forms coordination between the oxygens in the phosphonate groups and the oxygens in the phosphates present in the phosphopeptides. With this interaction, phosphopeptides bind to the matrix and can be enriched from the complex peptide mixture.

3.3. Kinase Assay Linked with MS

Identification of *in vitro* kinase substrates using mass spectrometry has been widely used. These methods commonly use samples from cell lysate or tissue, which are phosphorylated by a purified kinase *in vitro*, followed by mass spectrometric identification of phosphopeptides as described above. The challenge with this method is the identification of phosphorylation events conducted by the kinase of interest, compared to the background phosphorylation caused by endogenous kinases (Sugiyama and Ishihama, 2016). Radiolabeling with [γ-^{32}P]-ATP, excessive concentration of purified kinase, inactivation of endogenous kinases by heating, and quantitative proteomics are used to distinguish the wanted phosphorylation events from the background (Xue e*t al*. 2012). Recently, an *in vitro* kinase assay method which uses stable isotopically labeled ATP has been developed, where ATP molecules' γ-phosphate has ^{16}O atoms replaced with a heavy isotope oxygen ^{18}O (Müller *et al.* 2016). In this method, kinase of interest and the 'heavy' ATP is added into a cell lysate which is treated with a pan-kinase inhibitor FSBA (5′ -[p-(fluorosulfonyl)benzoyl]adenosine), which irreversibly blocks the endogenous kinases so that the only kinase functioning in the assay is the kinase of interest (Müller *et al.* 2016). The kinase of interest transfers the heavy phosphate group to its substrates that are present in the lysate and the resulting mass shift from the heavy oxygen atoms can be detected by MS (Müller *et al.* 2016).

In this study, two methods are combined into a pipeline with the goal of large-scale direct substrate identification. In-solution kinase assay with heavy ^{18}O-ATP is used to label the proteins that are phosphorylated by the kinase of interest. After the assay, proteins are purified, and digested by trypsin. The formed peptides are subjected into phosphopeptide enrichment with Ti^{4+}- IMAC material before MS analysis and substrate identification.

II. AIMS OF THE STUDY

The overall aim of this study was a large-scale identification of direct substrates for CMGC kinases, using a novel method of heavy-labeled ^{18}O-ATP-based kinase assay combined with LC-MS/MS. To validate the findings and to evaluate the method, identified substrates were compared to known kinase-substrate pairs annotated at PhosphoSitePlus® database. The goal was to draw conclusions on which proteins are directly phosphorylated by the kinases of interest and to predict the biological processes affected by the phosphorylation events.

III. MATERIALS AND METHODS

1. Sample Preparation

HEK 293 and HeLa cells were cultivated in DMEM (GE Healthcare), supplemented with 10% foetal bovine serum (FBS) and antibiotics (penicillin, 50 ug/mL and streptomycin, 100 µg/mL). Cells on a plate were washed with PBS, dislodged with PBS and EDTA, transferred to a falcon tube and centrifuged 1,400 x g for 5 min. Cells were washed with PBS and pelleted with centrifugation at 1,400 x g for 5 min before lysis. Cytosolic cell lysate was prepared by lysing the pelleted cells with buffer containing 50 mM Tris-HCL, pH 7.5, 150 mM NaCl, 5 mM EDTA, 1% NP-40 (Invitrogen, Thermo Fisher Scientific), and protease inhibitors cocktail (Sigma) on ice. The cell debris was cleared by centrifugation at 16,000 x g for 10 min. To prepare the nuclear fraction, cells were first lysed with cell lysis buffer (Buffer I) containing 10 mM Hepes, pH 7.5, 10 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol (DTT), 0.5% IGEPAL® (Sigma-Aldrich) and 1 x protease inhibitor cocktail (Sigma) and incubated 20 min on ice. Tubes were vortexed and then centrifuged at 12,000 x g at 4 °C for 10 min. The pelleted nuclei were washed with the cell lysis buffer (Buffer I), centrifuged at 12,000 x g at 4°C for 10 min and resuspended in the nuclear extraction buffer (Buffer II) containing 20 mM Hepes, 400 mM NaCl, 1 mM EDTA, 1 x protease inhibitor cocktail (Sigma) and 1 µl Benzonase® (Merck). Suspended cells were incubated on ice for 15 min, then sonicated 3 x 3 min with icing for 3 min in between sonications. Nuclear extract was collected by centrifugation at 12,000 x g at 4 °C for 15 min. The protein contents were measured using a BCA protein assay kit (Pierce, Thermo Scientific) and the cell fractions were stored at -80 °C.

Cell fractions were thawed on ice and endogenous kinases were inhibited with 5′ -[p-(fluorosulfonyl)benzoyl]adenosine (FSBA; Sigma-Aldrich) in DMSO at a final concentration of 1 mM FSBA and 10% DMSO in Tris-HCL, pH 7.5 for 1 h at 30 °C. Excess FSBA reagent was removed by ultracentrifugation with 15 mL 10 K MWCO Amicon® Ultra-4 centrifugal filter units (Merck) at 3,500 × g at RT. Proteins were washed 4× the initial volume with kinase assay buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT), adjusted to 400 µl volume and stored on ice. For kinase reaction, 200 µg of FSBA-treated cell lysate was incubated with 1 µg of kinase (Life Technologies) and 1 mM γ[18O4]-ATP (Cambridge Isotope Laboratory) in 30 °C for 1 hour. For negative control, 200 µg of FSBA-treated cell lysate was incubated with 1 mM γ[18O4]-ATP in the absence of kinase. Kinase assay reactions were stopped with 100 µl of 8 M urea.

Before digestion, the proteins in kinase reaction samples were reduced with 5 mM Tris(2-carboxyethyl)phosphine (TCEP; Sigma-Aldrich) in the ratio of 1:10 for 20 minutes in 37 °C and then alkylated with 10 mM iodoacetamide (IAA; Sigma-Aldrich) in the ratio of 1:20 for 20 min in room temperature in the dark. 600 µl of ammonium bicarbonate (AMBIC; Sigma-Aldrich) was added to dilute urea before trypsin digestion. Sequencing Grade Modified Trypsin (Promega) was used to get a 1:100 enzyme:substrate ratio and the samples were incubated overnight at 37 °C. Samples were then desalted with C18 macrospin columns (Nest Group). Phosphopeptide enrichment was performed using Ti4+-IMAC. The IMAC material was prepared and used essentially as described (Zhou *et al*. 2013). Briefly, Ti4+-IMAC beads were loaded onto GELoader tips (Thermo Fisher Scientific) and conditioned with loading buffer (80% acetonitrile (ACN), 6% trifluoroacetic acid

(TFA)). The protein digests were dissolved in loading buffer and added into the spin tips. The columns were washed first with 50% ACN, 0.1% TFA with 200 mM NaCl and then without salt. The bound phosphopeptides were eluted into a new tube with 10% ammonia, and then purified with C18 microspin columns (Nest Group) before LC-MS/MS analysis.

2. MS-analysis

The LC-MS/MS analysis was performed on a Q Exactive ESI-quadrupole-orbitrap mass spectrometer coupled to an EASY-nLC 1000 nanoflow LC (Thermo Fisher Scientific), using Xcalibur version 3.1.66.10 (Thermo Fisher Scientific). The phoshopeptide sample was automatically loaded from an autosampler into a C18-packed precolumn (Acclaim PepMap™100 100 µm x 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) in buffer A (1 % ACN / 0.1 % formic acid (FA)). Peptides were transferred onward to a C18-packed analytical column (Acclaim PepMap™100 75 µm x 15 cm, 2 µm, 100 Å) and separated with a 120-minute linear gradient from 5 % to 35 % of buffer B (98 % ACN and 0.1 % FA in HPLC grade water) with a constant flow rate of 300 nl/min. The total measurement time was 140 min per sample. The mass spectrometry analysis was performed in data-dependent acquisition in positive ion mode. MS spectra were acquired from m/z 300 to m/z 2,000 with a resolution of 70,000 with Full AGC target value of 3,000,000 ions and a maximal injection time of 120 ms in profile mode. The 10 most abundant ions with charge states from 2+ to 7+ were selected for subsequent fragmentation (higher energy collisional dissociation, HCD) and MS/MS spectra were acquired with a resolution of 17,500 with AGC target value of 5,000, a maximal injection time of 120 ms, and the lowest mass fixed at m/z 120 in centroid mode. Dynamic exclusion duration was 30 s.

3. Data Processing and Analysis

Raw data were processed with MaxQuant version 1.6.0.16 (Cox and Mann, 2008). MS spectra were searched against the human component of the UniProtKB database (release 2017_12 with 20192 entries) using the Andromeda search engine (Cox *et al*. 2011). Carbamidomethylation (+57.021 Da) of cysteine residues was used as static modification. Heavy phosphorylation of serine/threonine/tyrosine (+85.966 Da) and oxidation (+15.994 Da) of methionine was used as dynamic modification. Precursor mass tolerance and fragment mass tolerance were set to less than 20 ppm and 0.1 Da, respectively. A maximum of two missed cleavages was allowed. The results were filtered to a maximum false discovery rate (FDR) of 0.05.

Processed data was analyzed manually and filtered based on localization probability with a cut-off at 0.75 and proteins identified from the negative control were filtered out. Common substrates between two datasets were calculated using VENNY 2.1 online tool (Oliveros, 2015). Identified phosphorylation sites and substrates were compared to kinase-substrate data extracted from PhosphoSitePlus® (08182019; www.phosphosite.org) and visualized using RStudio version 1.1.463.

Gene Ontology classes for biological processes (GO_BP) were acquired from UniProtKB database and visualized with Cytoscape version 3.7.1.

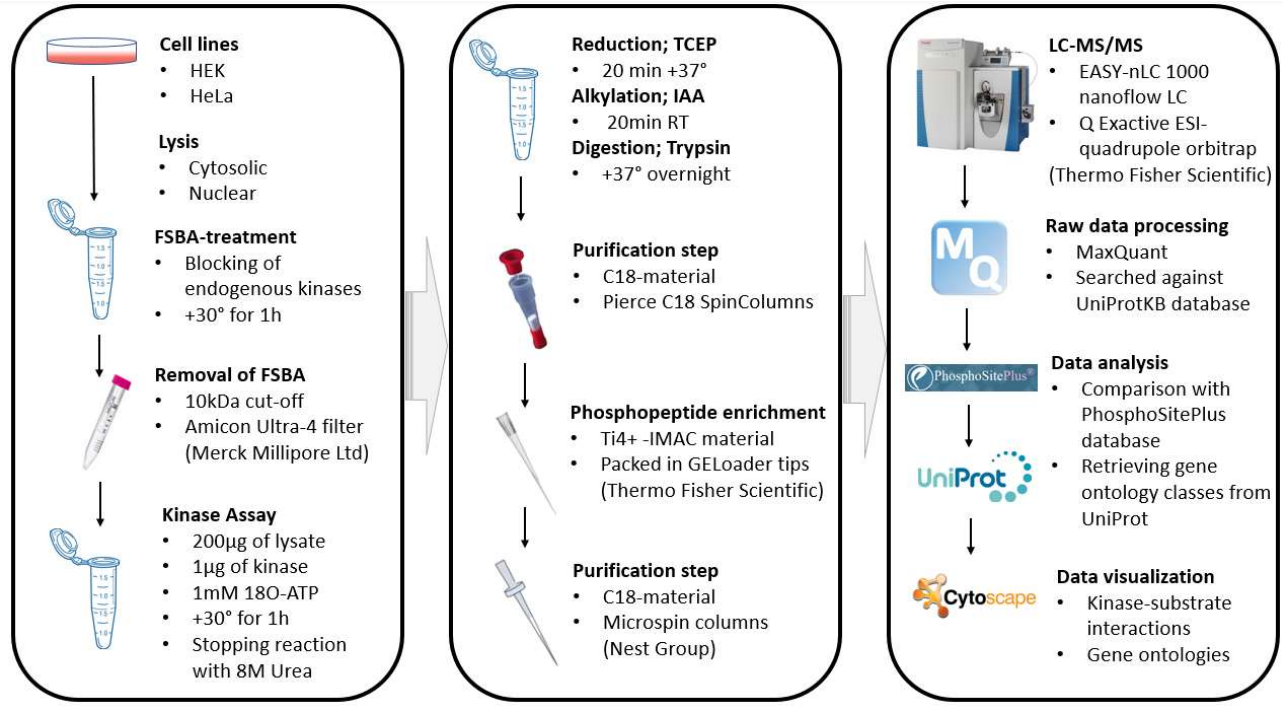


Figure 8. Schematic diagram of the methods. HEK and HeLa cells are lysed with two protocols to produce cytosolic and nuclear extracts. The lysates are treated with pan-kinase inhibitor FSBA to block endogenous kinases. FSBA is removed with ultrafiltration with a 10 kDa cut-off and the buffer is exchanged to kinase assay buffer. Kinase assay, where kinase of interest phosphorylates its substrates with heavy-labelled ^{18}O ATP is done with 200 µl lysate, 1 µg of purified kinase of interest and 1 mM of ^{18}O-ATP and incubated in +30 °C for an hour until reaction is stopped with urea. Samples are reduced, alkylated and digested before purification with C18-columns. Phosphopeptides are enriched with Ti4+ -IMAC material and further purified with C18-columns. Samples are analyzed with LC-MS/MS using Q Exactive ESI-quadrupole orbitrap (Thermo Fisher Scientific). Raw data is processed with MaxQuant, searched againt UniProt database and compared to known phosphorylation data from PhosphoSitePlus® database. Gene ontology classes for identified substrates are retrieved from UniProt and visualized with Cytoscape.

IV. RESULTS

1. Quantification of Direct Substrates of CMGC Kinases

On systematic level, CMGC kinases have previously been studied with interaction studies which reveal indirect and transient interactions. Here, kinase assays combined with LC-MS/MS is used for identification of direct substrates for 26 members of the CMGC kinase family. The results are analyzed at two different levels: at phosphorylation site level and at substrate level. The quantity of phosphorylated sites may not match to the number of substrates, because some substrates are phosphorylated at multiple sites. Importantly, formed kinase-substrate pairs, also referred here as kinase-substrate interactions, do not necessarily match with the number of substrates identified per each kinase group, because some kinases share the same substrates. Simplified schematic of the networks and quantifications of substrates, interaction partners and phosphorylation sites and -events are presented in Figure 9.

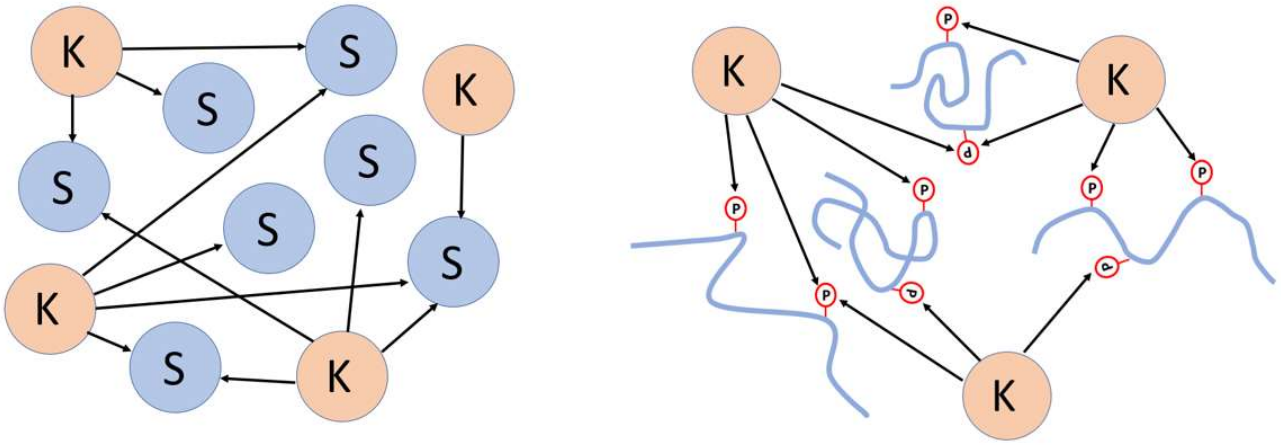


Figure 9. Simplification of quantifying kinase-substrate interactions and phosphorylation events. **A)** Kinase-substrate interactions are marked as arrows. This schematic presents 12 interactions between kinases and substrates, but only a total of 4 kinases and 7 substrates. In this figure, a single kinase has from one to four direct substrates. **B)** Phosphorylation events are marked as arrows, phosphorylation is shown as a letter P attached to a substrate (in blue). Three kinases phosphorylate a total of four substrates. There are nine phosphorylated sites within these four substrates which can be phosphorylated by 11 different phosphorylation events.

1.1. Number of Direct Substrates in the Cytosolic Fraction

In the cytosolic fraction of HEK and HeLa cells, a total of 1345 substrates were identified resulting in a total of 3841 kinase-substrate interactions. Numbers of phosphorylated sites and identified substrates for individual kinases and numbers of identified substrates and kinase-substrate interactions for CMGC subfamilies are summarized in Table 1.

Table 1: Quantification of phosphorylation sites and substrates for individual kinases, and substrates and kinase-substrates interaction pairs for CMGC kinase subfamilies in cytosolic fraction

Kinase	**HEK**	**HELA**	**Common**	**Total**	**Total by Subfamily**
	p-sites (substrates)	p-sites (substrates)	p-sites (substrates)	p-sites (substrates)	Substrates/ Kinase-Substrate Interactions
CDKs					
CDK1	123 (109)	415 (345)	52 (60)	486 (394)	901 / 1720
CDK2	239 (199)	692 (545)	115 (118)	817 (626)	
CDK4	7 (7)	29 (28)	0 (2)	36 (33)	
p25/CDK5	110 (97)	345 (290)	46 (51)	409 (336)	
p35/CDK5	57 (49)	152 (136)	20 (25)	189 (160)	
CDK6	27 (24)	10 (8)	2 (1)	35 (31)	
CDK7	23 (21)	39 (37)	2 (3)	60 (55)	
CDK8	10 (9)	10 (8)	0 (0)	20 (17)	
cyclinK/CDK9	15 (13)	20 (19)	0 (0)	35 (32)	
cyclinT1/CDK9	26 (23)	15 (14)	1 (1)	40 (36)	
MAPKs					
MAPK1 (ERK2)	143 (125)	37 (33)	11 (15)	168 (143)	234 /365
MAPK3 (ERK1)	62 (56)	30 (26)	5 (8)	87 (74)	
MAPK8 (JNK1)	44 (38)	71 (61)	14 (15)	99 (84)	
MAPK14 (P38a)	38 (35)	20 (13)	5 (5)	53 (43)	
MAP2K1 (MEK)	10 (10)	13 (12)	1 (1)	22 (21)	
NLK	117 (102)	193 (150)	32 (34)	218 (218)	218 / 218
GSK-3					
GSK-3α	35 (33)	70 (62)	13 (14)	92 (81)	117 /158
GSK-3β	52 (44)	42 (40)	6 (7)	88 (77)	
Dual-specificity *CLKs*					
CLK1	45 (41)	28 (26)	5 (7)	68 (60)	305 / 410
CLK2	214 (176)	162 (135)	48 (54)	328 (257)	
CLK3	40 (32)	20 (18)	3 (6)	57 (44)	
CLK4	33 (27)	32 (26)	2 (4)	63 (49)	
DYRKs					
DYRK3	458 (356)	423 (325)	172 (168)	709 (513)	565 / 710
DYRK4	119 (106)	149 (134)	39 (43)	229 (197)	
HIPKs					
HIPK1	28 (26)	34 (28)	4 (4)	58 (50)	167 / 260
HIPK2	31 (28)	42 (39)	10 (10)	63 (57)	
HIPK3	35 (31)	71 (62)	17 (16)	89 (77)	
HIPK4	48 (38)	66 (53)	15 (15)	99 (76)	

1.1.1. CDKs

From the largest CMGC subfamily, CDK1-2 and CDK4-9 were studied, and these kinases phosphorylated a total of 901 substrates, composing a network of 1720 direct kinase-substrate interactions (Table 1). Cell cycle regulating CDKs 1 and 2 phosphorylate more sites and substrates compared to the other cell cycle regulating CDKs 4 and 6 and transcriptional CDKs 7-9. CDK1 phosphorylated 123 sites and 109 substrates in HEK cell line and the respective numbers for HeLa cell line was 415 p-sites and 345 substrates, showing that phosphorylation was more robust in the HeLa cell lysate. In total of 486 p-sites and 394 substrates were phosphorylated by CDK1. CDK2 phosphorylated 239 sites and 199 substrates in HEK cell line, 692 sites and 545 substrates in the HeLa cell line, and the total number of phosphorylated sites was 817 with a total of 626 identified substrates. This was the largest number of sites and substrates phosphorylated by the CDKs studied and importantly, samples from HEK and HeLa cell lines shared 115 p-sites and 118 substrates. CDK1 and CDK2 shared a total of 281 substrates and 33 substrates were common between both kinases and both cell lines (Figure 10). 508 substrates were uniquely found in HeLa cell line and 133 of them were shared between CDK1 and CDK2.

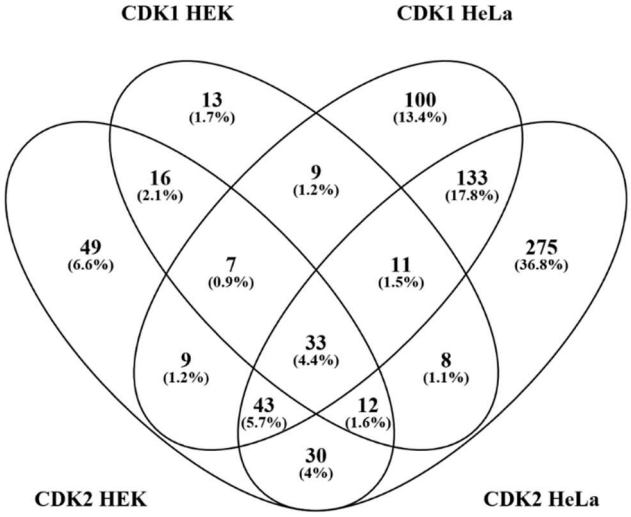


Figure 10. Quantification of common substrates between CDK1, CDK2 and two cell lines. CDK1 and CDK2 share 33 substrates that are expressed in both HEK and HeLa cell lines. In both CDK1 and CDK2 kinase assays, phosphorylation was more robust in HeLa cell line. 133 HeLa-specific substrates were common between CDK1 and CDK2 kinases.

CDK5 works in a combination with a non-cyclic regulatory subunit, which has two forms, p35 and a truncated p25. Kinase assays here were made with both complexes p35/CDK5 and p25/CDK5. The p25/CDK5 complex phosphorylated a total of 409 sites and 336 substrates, where only 110 p-sites and 97 substrates were phosphorylated in HEK cell line compared to 345 sites and 290 substrates phosphorylated in HeLa. Phosphorylation by p35/CDK5 complex was not as extensive, with only a total of 189 sites and 160 substrates phosphorylated. As the p25/CDK5 complex, p35/CDK5 complex phosphorylated a much larger number of sites and substrates in HeLa cell line. The substrates phosphorylated by p25/CDK5 complex uniquely in HeLa cell lysate composed a 42.6% fraction of all substrates identified. A fraction of 68.3% of all the substrates phosphorylated by CDK5 were phosphorylated in HeLa cell line only, meaning that 258 proteins were identified exclusively in HeLa cell lysate (Figure 11A).

Two cyclin/CDK9 complexes were studied: T1/CDK9 and K/CDK9. The number of identified p-sites for these complexes across both cell lines ranged from 15 to 26 and number of substrates from 13 to 23. Interestingly, K/CDK9 complex phosphorylated more p-sites and substrates in HeLa cell lysate, meanwhile T1/CDK9 phosphorylated more robustly in HEK cell lysate. There were no shared substrates with all four samples, and only one shared substrate with T1/CDK9 HEK and HeLa samples (Figure 11B). The relatively low number of p-sites and substrates identified in CDK9 compared to CDKs 1 and 2 is consistent with other transcriptional CDKs studied here (CDKs 7-9), where the number of identified p-sites ranged from 10 (CDK8, HEK and HeLa) to 39 (CDK7, HeLa), and number of substrates ranged from 8 (CDK8, HeLa) to 37 (CDK7, HeLa) (Table 1). However, two of the cell-cycle regulating CDKs 4 and 6 also had very low number of identifications when compared to the CDK1 and 2.

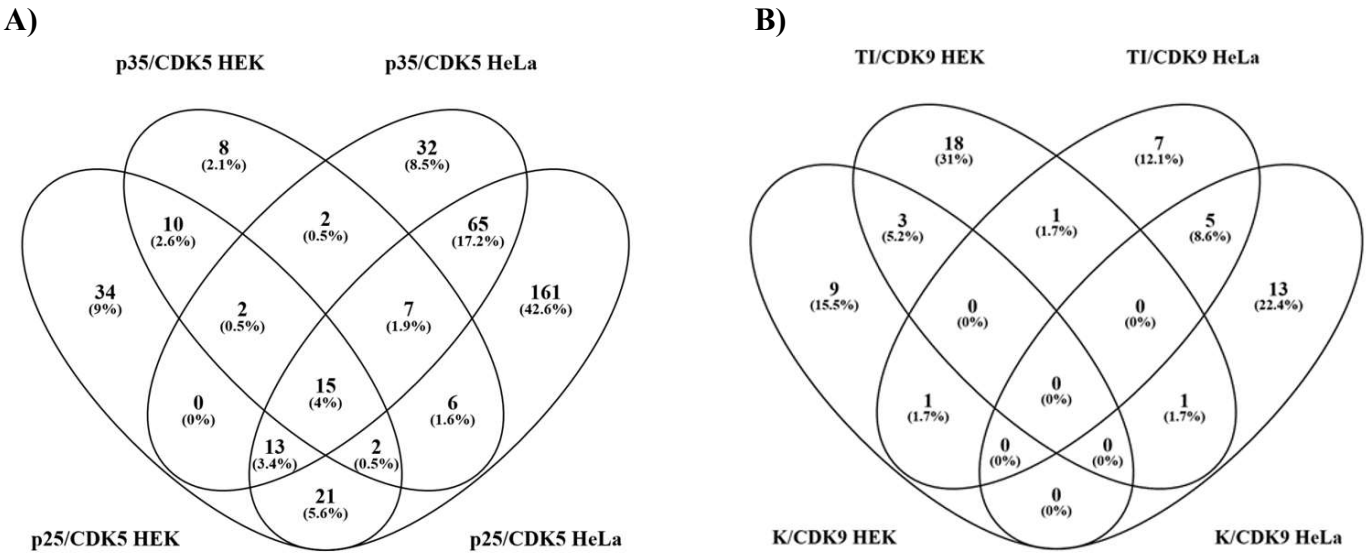


Figure 11. Quantification of CDK5 and CDK9 substrates with different regulatory subunits and different cell lines. A) 15 substrates expressed in HEK and HeLa cell lines are shared between p25/CDK5 and p35/CDK5 complexes. Overall, p25/CDK5 complex resulted in more robust phosphorylation, specifically in HeLa cells. 65 HeLa-cell line specific substrates were phosphorylated by both complexes, while the respective number for HEK-specific substrates was 10. **B)** No common substrates between T1/CDK and K/CDK complexes were identified that would have been phosphorylated in both cell lines. A total of 10 substrates were common between T1/CDK9 and K/CDK9 complexes, 3 of which were restricted to HEK cell line only and 5 to HeLa cell line.

1.1.2. MAPK and NLK

For individual subfamilies, MAPKs studied (MAPK 1, 3, 8, 14) and one MAPKK (MAP2K1) phosphorylated a total of 234 substrates adding up to 365 MAPK-substrate interactions. MAPK1 phosphorylated a total of 168 sites in 143 substrates, which was the largest set of substrates

phosphorylated by MAPKs. MAPK1 phosphorylated more sites (143) and substrates (125) in HEK cells compared to the HeLa cell line (37 sites, 33 substrates). Other MAPKs studied did not have as a drastic change between the cell lines. The one MAPKK studied, MAP2K1 (MEK1), phosphorylated less sites compared to the other MAPK-family members, with a total of 21 sites phosphorylated, meanwhile number of sites phosphorylated by MAPKs ranged from 58 to 168. NLK phosphorylated a total of 218 substrates in the two cell lines, with 117 p-sites and 102 substrates phosphorylated in the HEK cell line and 193 sites and 150 substrates in HeLa cell line.

1.1.3. GSK-3

GSK-3α and β phosphorylated a total of 117 substrates and a total of 158 kinase-substrate interactions were observed. Identified phosphorylation sites and substrates were distributed relatively evenly between the α and β forms, with a total of 92 p-sites and 88 substrates that were phosphorylated by GSK-3α and for GSK-3β the respective numbers were 88 and 77 (Table 1). Phosphorylation of the HEK cell lysate was more robust in the GSK-3α kinase assay when compared with the assays with β isoform, which phosphorylated more proteins in the HeLa cell lysate (Table 1; Figure 12). There were a total of 40 substrates unique to the GSK-3α that were not phosphorylted by GSK-3β and four of them were shared between the two cell lines. In the GSK-3β assays, 36 unique substrates were identified, 23 of them in HEK cell line and 13 of them in HeLa. No substrates phosphorylated solely by GSK-3β were identified in both cell lines. The α and β isoforms shared a total of 41 substrates but only 2 of them were shared between all four samples, meanwhile 6 of them were common only in the HEK cell lysate and 14 of them in HeLa (Figure 12).

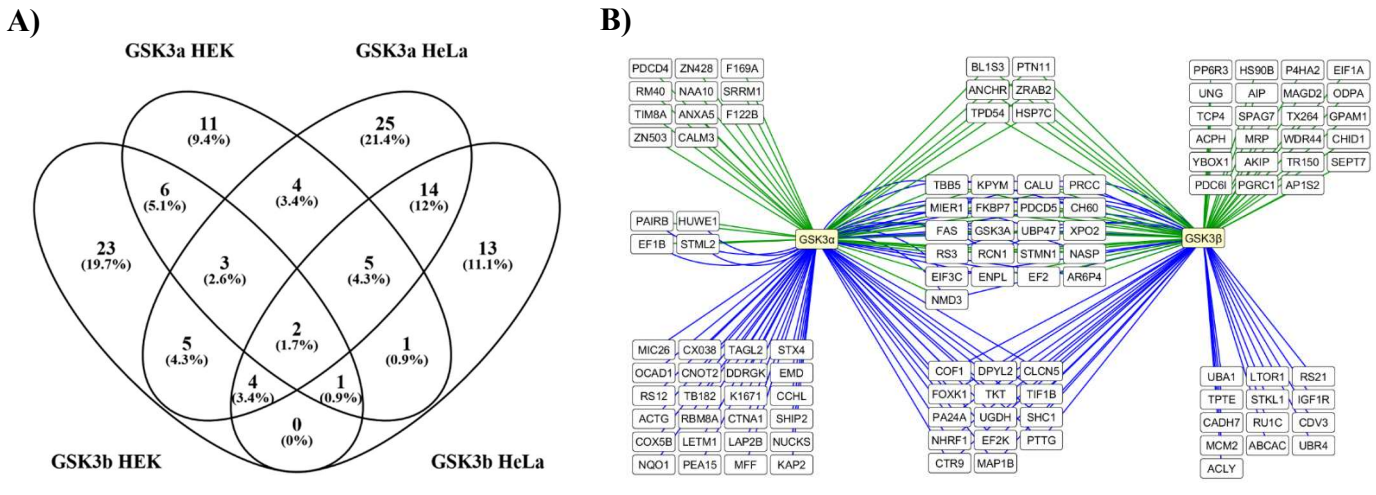


Figure 12. Common substrates between GSK-3α and GSK-3β kinases A) Venn-diagram showing the number of common substrates between kinases and the two different cell lines, HEK and HeLa. **B)** Identifications of GSK-3 substrates. Nodes present the substrates, labeled as UniProt entry names. Substrates connected to green edges were identified from HEK samples and substrates identified from HeLa samples are connected to the respective kinase with a blue edge.

1.1.4. CLK

One of the dual-specificity kinase groups studied here, CLKs, composed a kinase-substrate interaction network with 410 kinase-substrate interactions. From the 305 substrates phosphorylated by CLKs, CLK2 phosphorylated the largest fraction with 257 substrates and a total of 328 phosphorylation sites. The total number of phosphorylated substrates by other CLKs ranged from 44 to 60 and number of phosphorylated sites ranged from 57 to 68, which is far less compared to phosphorylations conducted by CLK2.

1.1.5. DYRK

DYRKs phosphorylated a total of 565 substrates, 513 of which were phosphorylated by DYRK3 (Table 1). DYRK4 phosphorylated a total of 197 substrates and 229 phosphorylation sites, which is 316 substrates and 480 p-sites less than DYRK3. For both kinases, only a slight difference was observed in the level of phosphorylation between the two cell lines. However, the identities of substrates phosphorylated in the cell lines differed in both kinases. DYRK3 HEK and HeLa samples shared 168 substrates, meanwhile DYRK4 had only 43 common substrates between both cell lines. A total of 34 substrates were shared between both DYRKs and both cell lines, showing that the two kinases have shared functions in regards of phosphorylating the same proteins (Figure 13). Additionally, as many as 86 substrates were shared between DYRK3 HEK and HeLa cells, increasing the confidence of the substrates being true, direct substrates for DYRK3. Across DYRKs, there were a total of 177 substrates phosphorylated only in HeLa cells, of which 123 were phosphorylated exclusively by DYRK3 and 29 exclusively by DYRK4. A total of 194 substrates were identified from HEK cell lines only, 18 of them phosphorylated only by DYRK4 and as many as 155 exclusively by DYRK3, showing that DYRK3 has multiple substrates that are not phosphorylated by DYRK4.

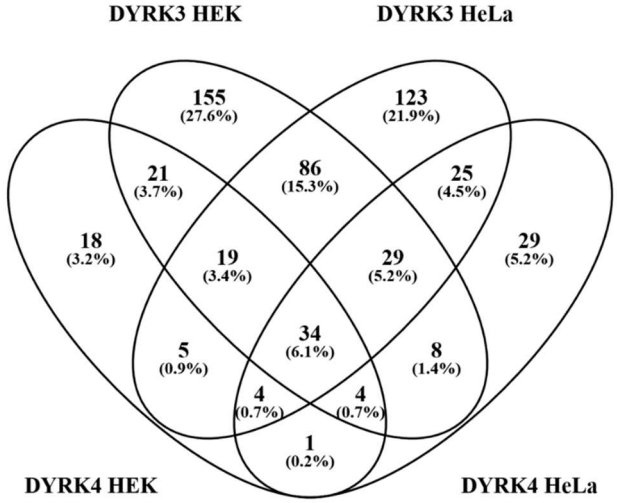


Figure 13. Venn diagram of the number of common substrates between DYRK3 and DYRK4. Number of kinases that are identified in kinase assays with HEK and HeLa cell lines. Percentages are calculated based on the pool of substrates identified in all four samples.

The total number of substrates phosphorylated of the HIPKs 1-4 was 167, and the kinases formed a kinase-substrate network with a total of 260 kinase-substrate pairs (Table 1). The substrates phosphorylated by a single HIPK ranged from 50 HIPK1 substrates to 77 HIPK3 substrates. All the HIPKs phosphorylated more proteins in HeLa cells when compared to the HEK cells.

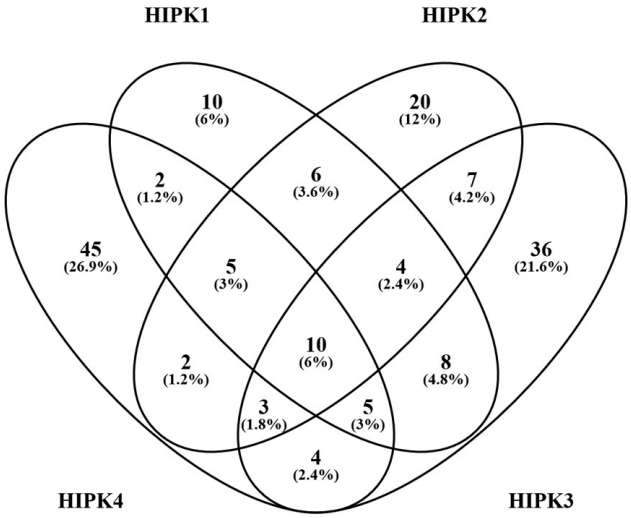


Figure 14. Venn diagram of the number of common substrates between HIPK1, HIPK2, HIPK3, and HIPK4. Comparison between the total amounts of substrates identified for each HIPK kinase. Number of substrates is calculated based on kinase assay results from both HEK and HeLa samples.

Only 10 of all the 50 identified HIPK1 substrates were unique to HIPK1 (Figure 14). Meanwhile other HIPKs had more unique substrates and specifically HIPK4 had as many as 45 unique substrates and HIPK3's respective number of substrates was 36 (Figure 14). A total of 10 substrates were shared by all four HIPKs and approximately a third of all kinases identified (56 kinases; 33,5%) were shared at least by two HIPKs.

1.2. Number of Direct Substrates in the Nuclear fraction

Some of the studied kinases localize mainly in the nucleus and therefore cytosolic cell lysate may not include all the substrates that reside naturally in nucleus. Therefore, HEK cells were lysed following a protocol for nuclear extracts and kinases with nuclear localization, CDK2, CDK4, and CDKs 6-9 as well as CLK1-4, were subjected for kinase assays. In total, we identified 194 kinase-substrate interactions, 141 of which were unique to the nuclear fraction and not identified in the cytosolic samples. The number of substrates identified for individual kinases are presented in Table 2.

Table 2. **Quantity of phosphorylation sites and proteins identified from nuclear fraction**. A set of 10 kinases (two forms of cyclin/CDK9 complexes) were assigned to kinase assay with a nuclear fraction of a HEK cell lysate. Number of phosphorylated sites and proteins are shown, as well as the number of substrates matching with the results from cytoplasmic fraction of HEK cell line and the number of substrates unique to the nuclear fraction.

Kinase	HEK/Nuclear		Kinase	HEK/Nuclear	
	p-sites (proteins)	Match cytosolic/ Nuclear only	***Dual-specificity***	p-sites (proteins)	Match cytosolic/ Nuclear only
CDKs			*CLKs*		
CDK2	145 (102)	50/52	CLK1	28 (20)	5/15
CDK4	9 (9)	1/8	CLK2	87 (59)	34/25
CDK6	10 (9)	1/8	CLK3	37 (20)	7/13
CDK7	3 (3)	0/3	CLK4	46 (32)	3/29
CDK8	6 (5)	0/5			
CDK9_cyclinK	20 (16)	0/16			
CDK9_cyclinT1	28 (23)	3/20			

CDK2 phosphorylated a total of 145 phosphorylation sites and 102 proteins, which is less than the amount of phosphorylation events happening in the cytosolic fraction, where 239 sites and 199 proteins were phosphorylated (Table 1). CDK2 had 50 substrates that matched between the cytosolic and nuclear fractions and 52 proteins were phosphorylated in the nuclear fraction only. Other CDKs phosphorylated relatively low amounts of p-sites and proteins. CDK4 phosphorylated nine proteins, one site on each, meanwhile CDK6 phosphorylated nine proteins and ten p-sites. Both kinases had eight new substrates identified that did not match the cytosolic fraction. CDK7 phosphorylated only three proteins, one site each, which is the lowest amount of phosphorylation events observed for a kinase across the whole study. CDK9 phosphorylated 16-20 substrates in the nuclear fraction, which is very similar to the 13-23 substrates phosphorylated in the cytosolic samples. CyclinK/CDK9 phosphorylated 16 substrates that were all unique for the nuclear fraction. CyclinT1/CDK9 phosphorylated 3 substrates that matched the cytosolic fraction and 20 identified substrates were uniquely nuclear.

CLKs phosphorylated less substrates in the nuclear fraction when compared to the cytosolic fraction presented in Table 1. All four CLKs phosphorylated substrates identified in both, nuclear and cytosolic fraction, but also substrates unique for the nuclear extract. CLK4 had the highest number of unique substrates to nuclear fraction with 29 substrates, followed by CLK2 with 25 substrates, CLK1 with 15 substrates and CLK3 with 13 substrates. CLK2 mediated phosphorylation was more robust when compared to other CLKs, which was characteristic for CLK2 also in the cytosolic fraction.

2. Comparison to PhosphoSitePlus® Database - Novel Substrates of CMGC Kinases

Identified phosphorylation sites and substrates were compared to PhosphoSitePlus® (PSP) database. Kinase-substrate dataset was downloaded from PSP website (www.phosphosite.org; Hornbeck *et al.* 2015) and the datasheet information of kinase- substrate pairs and phosphorylation sites of substrates was used and filtered based on kinase and organism. Well studied kinase groups such as CDKs and MAPKs are well represented in the database, meanwhile substrate information of dual-specificity kinases are annotated much less. CLKs, DYRKs and HIPKs had only few annotated substrates and for some kinases in this group (CLK3, DYRK4, HIPK3, and HIPK4) there was no information of the sites or substrates they phosphorylate.

2.1. Quantification of Novel Substrates Identified in the Cytosolic Fraction

CDKs 1 and 2 have the most sites annotated in the PSP database from all the kinases studied here. With CDK1 kinase assay 56 known substrates were identified that matched with the 268 substrates annotated in the database. In addition, 338 novel substrates were identified that were not annotated. Similarly, CDK2 had 65 substrates matching with the annotated 215 substrates and a high number of novel substrates with 561 identifications. From other cell-cycle regulating kinases, CDK4 and CDK6 kinase assay results had no matches to the database, so all 33 CDK4 substrates and 31 CDK6 substrates can be considered novel. Compared to the 70 known substrates annotated at PSP, CDK5 had 5 matches when combined with p25 molecule and 4 matches when in combination with p35 molecule. The p25/CDK5 complex had 331 novel substrates which is nearly double compared to the 156 novel substrates phosphorylated by p35/CDK5. In total for this kinase group, we identified 788 novel substrates and 1588 novel interacting kinase-substrate pairs (Figure 15).

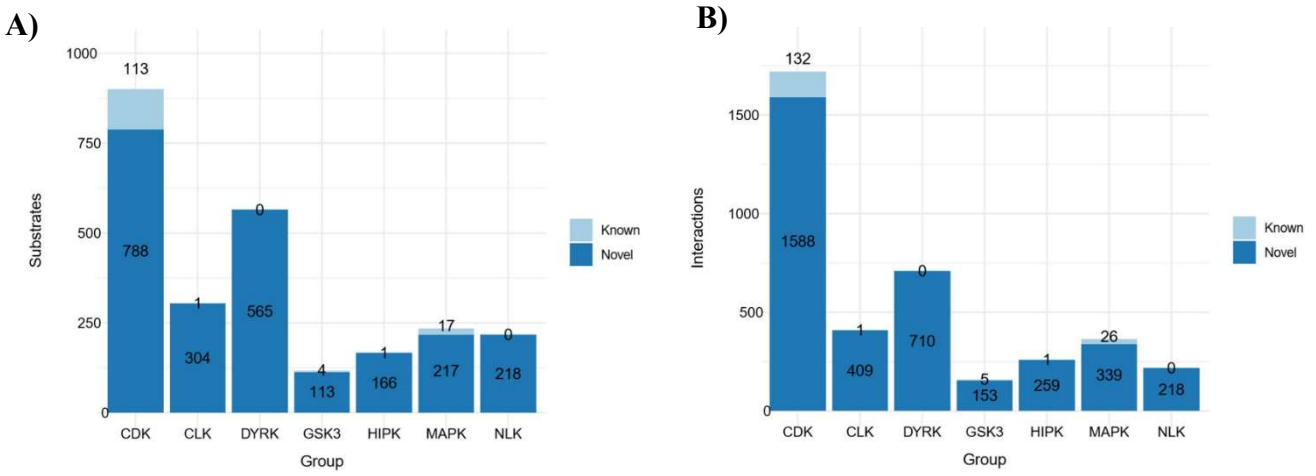


Figure 15. Known and novel substrates and kinase-substrate interactions for CMGC subfamilies
A) Stacked bar plot diagram of known and novel substrates, when kinase assay data is compared to PhosphoSitePlus® database. Kinases are grouped by their CMGC subfamilies. **B)** Known and novel kinase-substrate interactions in each of the CMGC kinase subfamily.

MAPKs have been extensively studied and they are also well cited in the PSP database. There are 192 annotated substrates for MAPK1, and the kinase assay results matched with 11 of them. The remaining 132 substrates identified were not annotated in the database, and therefore considered novel. Similarly, MAPK3 has 157 annotated substrates in the database and our results matched with 7 of them, resulting in 67 novel substrates. MAP2K1, which is positioned upstream from MAPK1 and MAPK3 in the MAPK pathway, has 7 annotated phosphorylation sites, 2 of which were matched with our results and identified as MAPK1 and MAPK3. In HEK and HeLa cells MAP2K1 phosphorylated MAPK1 at the Y-187 position, but not on the other cited position of T-185. MAPK3 was phosphorylated at the Y-204 site in HEK cells only. No phosphorylation was observed in the cited p-sites T-202, T-207, Y-210. In total, MAPKs phosphorylated 217 novel substrates which compose a network with 339 novel MAPK-substrate interactions (Figure 15). PSP did not have any annotated substrates for NLK, and therefore NLK had a high number of 218 novel substrates.

GSK-3α kinase assay results had two substrates matching with the 24 annotated substrates and 78 novel substrates. GSK-3β had 124 annotated substrates, but only three of them matched the kinase assay results, leaving 74 substrates unmatched. In total, GSK-3 group phosphorylated 117 substrates, 113 of which were not cited in the PSP database (Figure 15A), resulting in a total of 153 novel kinase-substrate pairs (Figure 15B).

Substrates and phosphorylation sites for dual-specificity kinases are not comprehensively annotated in the PhosphoSitePlus® database. For CLKs, there were 0-3 substrates annotated per kinase. Only one kinase, PTPN1, matched between the CLKs and the PSP data. CLK2 phosphorylated PTPN1 at the position S-50, but the phosphorylation event was absent in CDK1, even though it was annotated in PSP. CLKs phosphorylated a total of 304 substrates that were not annotated at the PhosphoSitePlus® database, resulting in 409 novel kinase-substrate interacting pairs (Figure 15). With 256 phosphorylated substrates not matching the database, CLK2 had significantly more novel substrates compared to other CLKs, which phosphorylated from 44 to 60 substrates each. DYRK3 had only 2 annotated substrates in the database and no substrates identified in the kinase assays matched to these. DYRK4 had no annotated phosphorylation sites or substrates. Therefore, all the identified 513 DYRK3 substrates and 197 DYRK4 substrates can be considered novel. PRAS40 was one of the annotated DYRK3 substrates but it was phosphorylated only by DYRK4 in the kinase assay. For the HIPK group, HIPK2 is the most annotated kinase with 20 substrates annotated in PhosphoSitePlus®. Autophosphorylation of HIPK2 was the only match between kinase assay results and the database. In kinase assay, HIPK2 was autophopshorylated at the S-441 site, which is not marked as a phosphorylation site in the PSP database. According to PSP, autophosphorylation should occur at positions Y-361, S-364, T-880 and S-882. HIPK1 had two substrates annotated in PSP but no matched with the kinase assay results. In total, 166 novel substrates and 259 kinase-substrate pairs were identified for the HIPK group (Figure 15).

2.2. Quantification of Novel Substrates Identified in the Nuclear Fraction

There was a total of 194 kinase-substrate pairs identified in the nuclear fraction samples of ten kinases, 27 of which matched to the PhosphoSitePlus® database, resulting in 167 novel kinase-substrate pairs. A total of 134 kinase-substrate pairs were novel and unique to the nuclear fraction samples. CDK2 phosphorylated 24 novel substrates identified only in the nuclear samples. the respective number for K/CDK9 was 13 and for T1/CDK9 19. CDK4, CDK6, CDK7 and CDK8 had no matches to the database and they phosphorylated only eight substrates or fewer each. CDK4 and CDK6 had one substrate matching the HEK or HeLa samples of cytosolic fractions, indicating that all majority of the substrates were novel and unique to the nuclear fraction. CLKs had no matches with PhosphoSitePlus® database, so all the identified substrates can be considered novel. Number of CLK substrates unique to the nuclear fraction ranges from 13 to 29 (Table 2).

3. Gene Ontology Analysis of Direct Substrates

To further investigate how the identified substrates affect cellular functions, substrates were grouped based on their gene ontology classifications (GO Biological Processes; GO_BP). Gene ontology information for each identified substrate was acquired from UniProt, analyzed and visualized with Cytoscape 3.7.1. Analysis of GO classes were conducted with kinases with outstandingly high amounts of novel substrates. These kinases include CDK1, CDK2, CDK5, CLK2, DYRK3, DYRK4 and NLK.

3.1. CDKs Phosphorylate a Variety of Substrates

CDK1 phosphorylated a total of 343 novel substrates in HEK and HeLa cells and 47 of them were shared between the two cell lines (Figure 16A). Gene ontology analysis of the substrates phosphorylated in both cell lines revealed that many of these substrates are involved in biological processes such as apoptotic processes, immune responses, metabolic processes, mRNA processing and transport, among others (Figure 16C). CDK2 had 99 novel substrates common between the two cell lines (Figure 16B) and they have similar functions with CDK1 substrates, with a slight increase in GO classes related to cell-cycle, proliferation, cell division and cytoskeleton organization (Figure 16D). CDK2 phosphorylates some cell-cycle regulating substrates that were not identified in the CDK1 samples, such as RPTOR, TBB4A and TBB4B, TBG1 and TBG2. RPTOR is involved in TOR and TORC1 signaling and it positively regulates G1/S transition of mitotic cell cycle. Meanwhile both, tubulin beta-4A and 4B as well as tubulin gamma-1 and 2 chains are involved in the G2/M transition. LTOR1 and PSA5 were identified in both HEK and HeLa cell lines in CDK2 but only in HeLa line in CDK1, and therefore they are missing from the Figure 16C. LTOR1 positively regulates TOR signaling and is involved in cell cycle arrest and PSA5 is involved in many signaling pathways such as Wnt signaling, MAPK cascade, T cell signaling, NIK/NF-kappaB signaling and it negatively regulates G2/M transition. From the substrates mentioned above, only TBG1 and TBG2 are known substrates for CDK2, the rest were not annotated at PhosphoSitePlus® database.

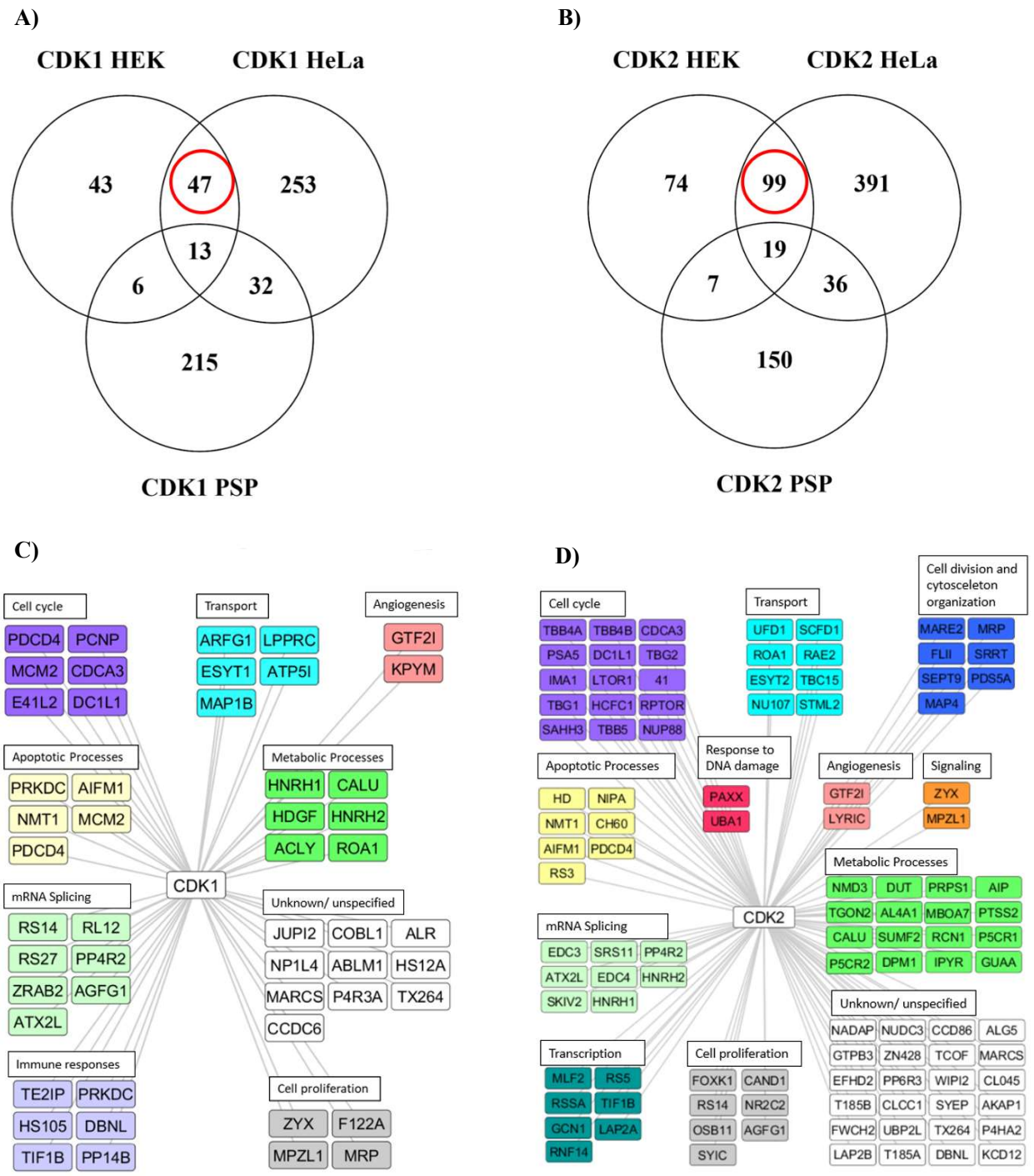


Figure 16. GO classes for novel substrates identified from CDK1 and CDK2 cytosolic fractions.
A) Venn-diagram for the substrates identified in HEK and HeLa cell lines, and the PhosphoSitePlus® database for CDK1. CDK1 had 49 common substrates between the two cell lines that were not annotated in the database and therefore they are considered novel and used in GO analysis. **B)** The respective number for CDK2 substrates was 99 and these substrates were used in GO analysis. **C)** Gene ontology analysis (GO; Biological Processes) of the novel substrates visualized with Cytoscape. Substrates are grouped based on their GO classifications. CDK1 have substrates involved in cell cycle, apoptotic processes, transport, angiogenesis, metabolic processes, mRNA processing and cell proliferation. **D)** GO analysis of CDK2 reveals that the CDK2 substrates have additional GO classifications compared to CDK1: transcription, response to DNA damage and cell-cell signaling. Some substrates may have GO classifications subject to multiple of the groups presented, so the grouping here is not conclusive and does not rule out other biological processes.

CDK2 kinase assay was also done with HEK nuclear fraction (Table 2). In total, 52 substrates were unique to the nuclear extract and not observed in the cytosolic fraction. Mainly, the CDK2's nuclear substrates were involved in transcription, as 27 of them had roles in transcription regulation, especially in positive or negative regulation of cellular DNA-templated transcription ([GO:0045893] and [GO:0045892]). In addition, 15 substrates had roles in mRNA splicing, 7 were involved in apoptotic processes and 7 in DNA repair, often double-strand- or base-excision repair.

CDK5 was combined with two different non-cyclic molecules: p25 and p35. The p25/CDK5 complex is a cleaved from of p35/CDK5, and it is more stable, has prolonged activity and ability to hyperphosphorylate substrates, such as Tau (Liu *et al.* 2017). The p25/CDK5 and p35/CDK5 complexes had 26 common substrates that were novel, and these substrates are presented in Figure 17A. These substrates were involved in biological processes related to transport, apoptotic processes, metabolism and biosynthesis, angiogenesis and brain development. Stathmin 1 (STMN1) was phosphorylated in HeLa cells by both, p25/CDK5 and p35/CDK5 complexes and in HEK cells by p35/CDK5. In HeLa cells, phosphorylation occurred at position S38 and in HEK phosphorylation at the position S25 was identified.

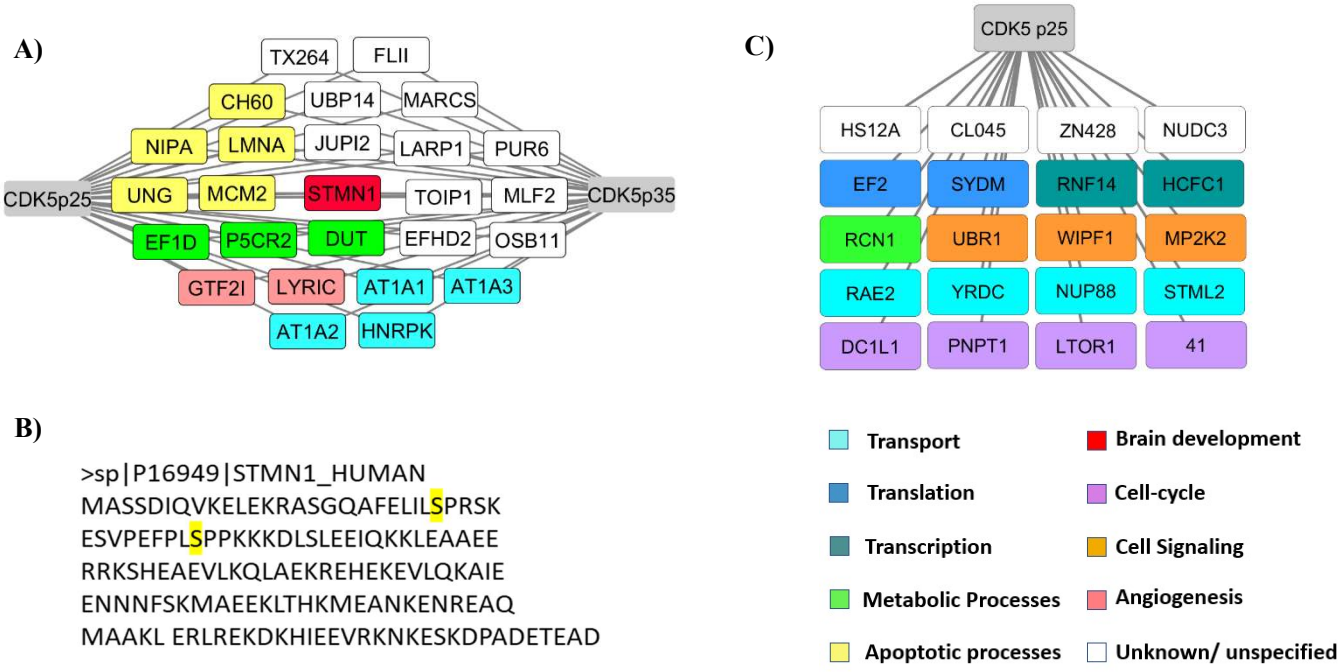


Figure 17. GO classifications of novel substrates phosphorylated by both CDK5 complexes and p25/CDK5 only A) Common substrates between p25/CDK5 and p35/CDK5 complexes and their gene ontology classifications. The two complexes share substrates with functions in angiogenesis, apoptotic processes, transport and metabolic processes. Both complexes phosphorylated Stathmin, which amino acid sequence is presented in B), with highlighted serines at phosphorylated positions 25 and 38. C) Gene ontology classes for substrates phosphorylated in both HEK and HeLa cell lysates exclusively by the p25/CDK5 complex.

Novel CDK5 substrates AT1A1, AT1A2 and AT1A3 are different subunits of sodium-potassium transporting ATPases, which regulate ion transmembrane transport. They mediate cardiac muscle function by regulating cardiac conduction [GO:1903779] and cellular communication by electrical coupling involved in it [GO:0086064]. In addition, AT1A1 and AT1A2 help regulating blood pressure [GO:0008217] and control the responses to clycoside [GO:1903416] among other functions.

CDK5 combined with p25 phosphorylated more substrates than p35/CDK5 complex, as shown in table 1. Substrates that were phosphorylated in both cell lines but only by p25/CDK5-complex affected biological processes related mainly in cell cycle, cell signaling and transport (Figure 17C). MP2K2, also called ERK activator kinase 2 or MEK2, regulates cell cycle by its downstream effects in the MAPK cascade. LTOR1 regulates MAPK cascade [GO:0043410] and TOR signaling [GO:0032008] positively, meanwhile URB1 is a negative regulator of TOR signaling [GO:0032007]. The substrates involved in transport-related biological processes had GO classifications of for example protein export from and import into the nucleus ([GO:0006611], [GO:0006606]; NUP88), intracellular protein transport ([GO:0006886]) and vesicle-mediated transport ([GO:0016192]) (RAE2) and negative regulation of transport ([GO:0051051]; YRDC), so that no specific pattern emerged within the transport-classification of gene ontologies.

3.2. NLK Phosphorylates Substrates Involved in Cellular Signaling

All 218 identified NLK substrates were novel, as none of them matched to the PhosphoSitePlus® database. NLK's substrates play role in multiple biological processes (Figure 21). Majority of the substrates identified were involved in cell signaling, as a minimum of 107 substrates that had GO classes related to signaling. 44 substrates were involved in antigen presentation and 41 of them were different chains of HLA class I histocompatibility antigens (both, HLA-B and HLA-C). These antigens were phosphorylated at the position S66 and they were only detected in HeLa cell lines. NLK has an apparent role in Wnt signaling as it phosphorylates 12 substrates involved in the pathway. WNK1, UBP8 and CCAR2 have shown to positively regulate the canonical Wnt pathway [GO:0090263], whereas TLE3 and TLE4, DAB2, EMD, GSK-3α and GSK-3β regulate the pathway negatively [GO:0090090]. A total of eight NLK's substrates were also involved in NF-kappaB signaling. Five of them, RS3, HS71B and HS71A, DDRGK and TE2IP have been shown to positively regulate NF-kappaB transcription factor activity [GO :0051092]. In addition to Wnt signaling and NF-kappaB signaling, substrates of NLK are involved in JAK-STAT cascade (CTR9; [GO:0007259]), signal transduction by p53 class mediator (PTTG; [GO:0043518]) which results in a cell cycle arrest (TB182, CNOT2, CNO6L; [GO:0006977]), and in cellular responses to interferon gamma (ACTG, STX4, VIME, ZYX; [GO:0071346]).

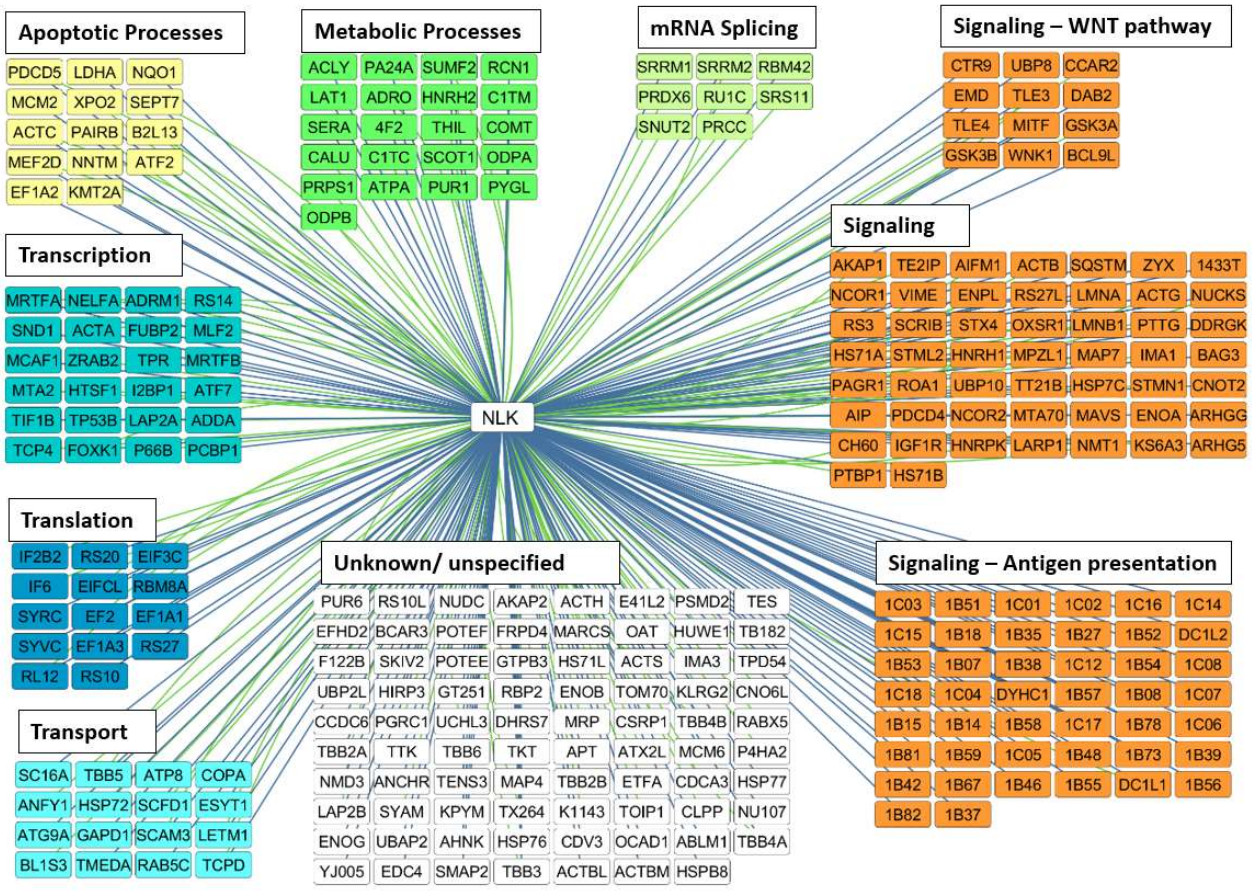


Figure 21. GO classifications of NLK substrates grouped and visualized

Nodes represent the substrates phosphorylated by NLK, labeled as UniProt entry names and colored by their groups based on their GO classifications. The edges represent cell lines, green edged substrates were phosphorylated in HEK cell line and blue edged substrates were phosphorylated in HeLa cell line. NLK phosphorylates substrates involved in the regular biological processes, such as transport, translation, transcription, apoptotic processes and metabolism. NLK has multiple substrates that affect cell signaling pathways, such as Wnt pathway. NLK also phosphorylates substrates involved in antigen presentation, mainly in HeLa cell lysate.

3.3. CLK2 Substrates Have Roles in Cytoskeleton Organization and mRNA Splicing

CLK2 kinase assay results from HEK and HeLa cell lines had 54 common and novel substrates which were assigned to gene ontology analysis to identify the biological processes they affect. CLK2's substrates influence many events in the cell, including translation, transcription, apoptotic processes, DNA repair and damage response, cell proliferation, cytoskeleton organization and cytokinesis (Figure 18). Substrates involved in actin cytoskeleton organization (DIAP, 41; [GO:0030036] and ACTB; [GO:0098974]) and reorganization (PTN1; [GO:0031532]) are directly phosphorylated by CLK2. RAD18 and DDB2 have a role in DNA repair [GO:0006281], meanwhile

HUWE1 is involved in base-excision repair [GO:0006284] and NUCKS helps repairing DNA damage by homologous recombination of double stranded breaks [GO:0000724]. From the substrates involved in apoptotic processes, PKN2 is also associated with cytokinesis, ENPL with actin rod assembly, and SCRIB and DNM1L with morphogenesis. In addition, apoptosis regulating proteins MCM2 and ASNS, as well as the cell proliferation associated protein KHDR1, have roles in the cell cycle.

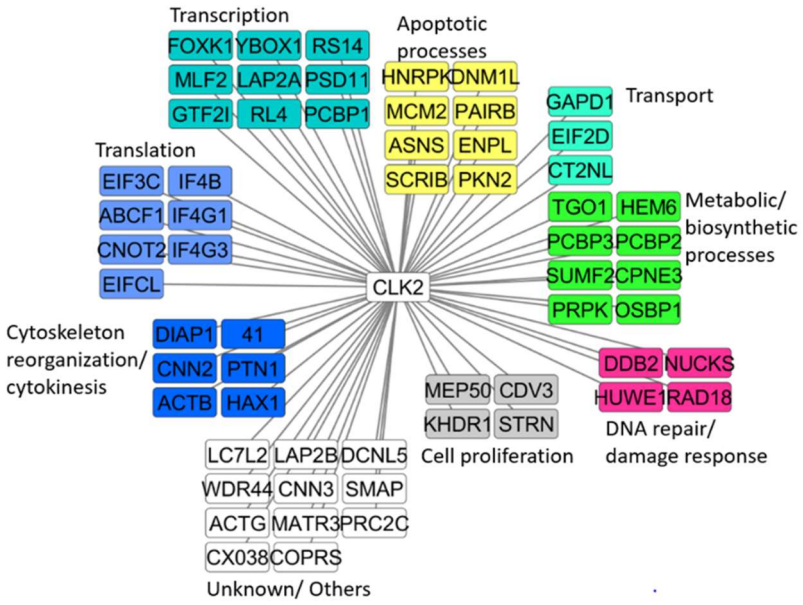


Figure 18. GO classes of novel CLK2 substrates. CLK2 phosphorylates proteins that are involved in multiple biological processes such as cytoskeletal organization and cytokinesis, translation, transcription, apoptotic processes, transport, metabolic processes, DNA repair and cell proliferation.

Some CLK2 substrates that were only phosphorylated in the HEK cell lysate include Serine/arginine-rich splicing factors 7 and 11 (SRSF) and Serine/arginine repetitive matrix proteins 1 and 2 (SRRM), which are involved in mRNA slicing via spliceosome ([GO:0000398]). SRSF7, SRSF11, and SRRM1 affect also mRNA 3'-end processing ([GO:0031124]) and mRNA export from the nucleus ([GO:0006406]). SSRM1, SSRM2 and SRSF11 were identified from nuclear extract sample. In addition, a total of 20 substrates identified only from HEK lysate had roles in spliceosome functions, including RNA binding proteins (RBM7, RBM8A, RMXL1 and RBMX), ribonucleoproteins (HNRPD, HNRPK and HNRPU) and splicing factors as described above. In conclusion, CLK2 substrates have many functions within the cytoskeleton organization, DNA repair, and mRNA splicing. However, many substrates affect multiple biological processes in the cell, and there is evidence that also the cell cycle may be affected by the functions of direct CLK2 substrates.

3.4. DYRK3 and DYRK4 Share Similar Functions

DYRK3 phosphorylated a total of 513 novel substrates which have a variety of functions in the cell, some of them presented in Figure 19. Major biological processes that DYRK3's substrates mediate in cells are transcription, translation, metabolic processes, cell differentiation and apoptotic processes. The substrates have multiple biological functions and would not belong to only one gene ontology group presented here. Therefore, the number of substrates per group is a bare minimum of substrates having these biological processes in the cell. At least 45 substrates are involved in transcription, 49 in translation, 49 in metabolic processes and 88 in transport. Five proteins of the 88 involved in transport are involved in COPII vesicle coating [GO:0048208], 11 mediated protein transport [GO:0015031], and five of them are involved in endocytosis [GO:0006897]. More specific gene ontology groups, such as signaling, mRNA splicing, cell cycle, angiogenesis and proliferation have less substrates, ranging from 9 to 14 substrates per group. For these proteins, multiple gene ontology classes were assigned and many of them had functions also in apoptotic processes, cell differentiation and transcription.

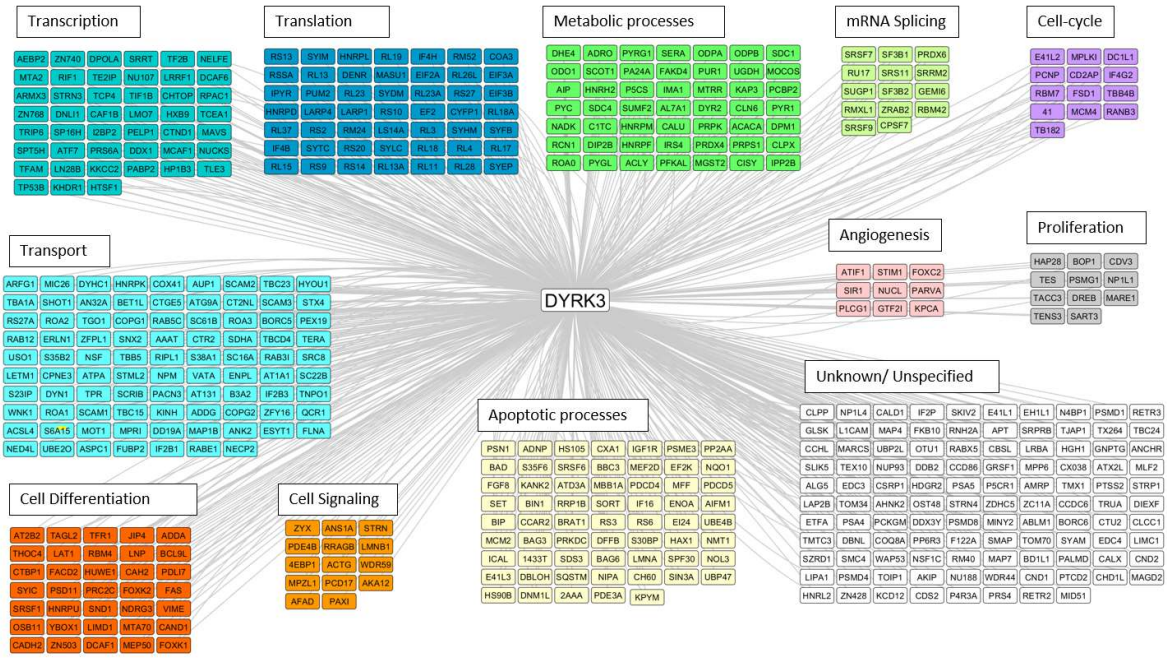


Figure 19. Grouping of DYRK3 substrates based on GO classifications

DYRK3 substrates from both cell lines are grouped based on their GO classifications. Majority of DYRK3 substrates were involved in cell differentiation, transport, transcription, translation, metabolic processes and apoptotic processes. Also, proteins with function related to mRNA splicing, cell cycle, angiogenesis, proliferation and cell signaling were observed. Some substrates have multiple functions and could belong to multiple groups.

DYRKs 3 and 4 had hundreds of substrates identified and 34 of them were shared by both kinases and both cell lines. The substrates are presented in Figure 20, with their biological process gene ontology classes. DYRKs 3 and 4 phosphorylate substrates involved in transport, translation, metabolic processes, cell cycle and apoptosis, among others. The substrates involved in transport regulated for example amino acid transport (AAAT; [GO:0006865]), ER to Golgi vesicle-mediated transport (BET1L, TGO1 and DC1L1; [GO:0006888]) and vesicle-mediated transport (CPNE3; [GO:0016192]). Importantly, this grouping is not conclusive as for instance DC1L1 has multiple different GO classes listed and it has functions in both, transport and cell cycle. Translational initiation [GO:0006413] is a biological process well represented in the substrate group shared by DYRK3 and DYRK4 as EIF3B, RS14, RL23A, and LARP all play a role in it. EF2 and RS14 are involved in cell differentiation, as EF2 regulates specifically hematopoietic progenitor cell differentiation [GO:0002244] and RS14 erythrocyte differentiation [GO:0030218]. The cell type which differentiation is driven by HUWE1, NDRG3 and FOXK1 was not specified by the GO class [GO:0030154].

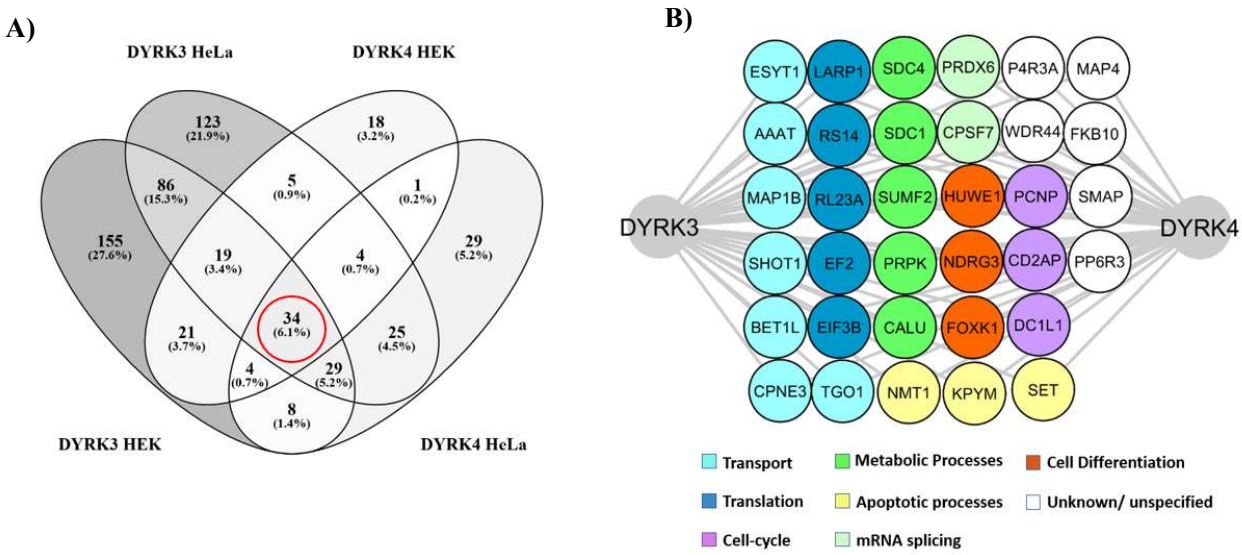


Figure 20. Common substrates of DYRK3 and DYRK4 and their GO classifications
A) Venn-diagram of DYRK substrates identified in both cell lines. A total of 34 substrates were common between both cell lines and both DYRK3 and DYRK4 kinases and these substrates were chosen for GO classification analysis. **B)** GO classes for the selected 34 kinases. DYRK3 and DYRK4 share substrates with functions in transport, translation, metabolic- and apoptotic processes, cell differentiation, cell cycle and mRNA splicing.

V. CONCLUSIONS

The overall aim of this study was to identify direct substrates for protein kinases in the CMGC family. Protein kinases are signaling molecules that regulate cellular and biological processes by phosphorylating cellular proteins. Phosphorylation event often changes protein's conformation resulting in alterations for example in protein activation or localization. Members of the CMGC kinase family regulate critical cellular functions that are involved in pathological processes, and aberrant signaling in for instance cell cycle is characteristic for many cancers (Malumbres and Barbacid, 2005; Ferguson and Gray, 2019). Inhibition of CMGC kinases is used to prevent transcription and signaling of proto-oncogenes in cancer treatment, but also immune deficiencies and degenerative diseases can be addressed by CMGC kinase inhibition (Ferguson and Gray, 2019). The knowledge of kinase-substrate pairs is crucial for mapping cellular signaling events and for understanding regulatory mechanisms in cells, for example in case of drug development. However, large-scale identification of direct substrates for kinases is difficult and time consuming. Mass spectrometry-based methods are used to identify *in vitro* substrates but identifying the phosphorylation events conducted by the kinase of interest is challenging due to background phosphorylation and mass spectrometer's difficulties to recognize phosphopeptides in a complex peptide mixture (Sugiyama and Ishihama, 2016). Here, a kinase assay combined with LC-MS/MS analysis is presented. In the method, a pan-kinase inhibitor FSBA is used to block endogenous kinases in a cell lysate before the kinase of interest is added to phosphorylate its substrates with heavy-labeled ATP, which contains an ^{18}O isotope at the γ-phosphate (Müller *et al*. 2016). Phosphopeptides are enriched with Ti^{4+}- IMAC before the LC-MS/MS analysis (Zhou *et al*. 2013), which distinguishes the desired phosphorylation events based on a mass shift caused by the heavy ^{18}O (Müller *et al.* 2016).

In this study, cell lysates from HEK and HeLa cell lines were used for cytosolic fractions. As many of the kinases studied have previously been identified as cancer hubs (Varjosalo *et al.* 2013), we wanted to analyze kinase interactome in HeLa cell line and to compare the results to HEK cell line. Compared to normal human tissues, HeLa cells express over 2,000 proteins higher than the physiological range and these proteins are mainly related in proliferation, transcription and DNA-repair (Landry *et al.* 2013). Therefore, cancer derived HeLa cell line expresses very different proteins than HEK cell line, which is originally derived from human embryonic kidney, indicating that the use of two cell lines provides a larger population of proteins to be tested. Some identified substrates overlapped between the two cell lines, but hundreds of substrates were exclusively phosphorylated in either of the two cell lines. As expected, more proteins were phosphorylated in the HeLa cell line, which expresses more proteins susceptible for phosphorylation by CMGC kinases. In addition to the cytosolic fraction, nuclear fraction of HEK cells was used for kinases that primarily have a nuclear localization to gain a better recognition of direct substrates.

With this pipeline of methods, we were able to quantify and identify direct substrates for 26 members of CMGC kinase family. A total of 1345 unique substrates and 3841 interacting kinase-

substrate pairs were identified in cytosolic cell lysates. A total of 165 kinase-substrate pairs were previously annotated at the PhosphoSitePlus® database, indicating that the assay provides reliable data of true kinase-substrate pairs. The remaining identified 3676 kinase-substrate interactions were previously not annotated in the database, highlighting the importance of large-scale analysis for direct kinase substrate identifications. In addition, we identified a total of 194 kinase-substrate pairs for ten kinases tested with nuclear extract of HEK cells, 141 of them being unique to the nuclear fraction. Quantification of substrates provides valuable information of the method, as well as of the kinases.

Kinase assay results show that phosphorylation by p35/CDK5 complex was not as extensive as phosphorylation by the p25/CDK5 complex. This difference was expected, as the non-cyclic compound p35 is less stable than its cleavage product p25 which has prolonged activity and ability to hyperphosphorylate substrates when forming a complex with CDK5 (Liu *et al.* 2017). This indicates that the method is suitable for distinguishing between two kinase complexes, replicating their biological functions. We were able to identify substrates uniquely phosphorylated by the p25/CDK5 complex, but also substrates phosphorylated by both p25/CDK5 and p35/CDK5. Notably, one of these substrates was Stathmin1 which has been shown to become phosphorylated by CDK1 and CDK2 but not CDK5 (PhosphoSitePlus®). Stathmin levels are decreased in adult brains with Down syndrome and Alzheimer's disease (Cheon *et al.* 2001), meanwhile deregulation of CDK5 is involved in pathogenesis in cytoskeletal abnormalities, possibly causing neurodegenerative diseases (Patrick *et al.* 1999). Our results show that Stathmin1 is directly phosphorylated by CDK5, indicating that the phosphorylation event may affect biological processes in neurons. However, the mechanisms remain unknown.

We analyzed gene ontology classes of substrates and we were able to link the biological processes of novel substrates to the processes regulated by the kinase of interest. For instance, NLK has been shown to work as a cellular signaling modifier, as it is involved in multiple well-characterized signaling pathways, such as Wnt/β-catenin and Notch signaling (Kanei-Ishii *et al*. 2004; Ishitani and Ishitani, 2013). Kinase assay results show that NLK's substrates function is cell signaling, as approximately half of identified NLK's substrates play a role in signaling processes, twelve of them regulating Wnt/β-catenin pathway. However, we did not identify substrates involved in Notch signaling, even though NLK has been shown to suppress Notch signaling by interfering with formation of the Notch active transcriptional complex (Ishitani *et al.* 2010). Rather, our kinase assay results show that NLK phosphorylates five substrates that positively regulate NF-kappaB transcription activity, suggesting that NLK may have a role in the regulation of NF-kappaB signaling. Similarly, identified DYRK3 substrates had gene ontology classifications corresponding to the known DYRK3 functions. DYRK3 has been shown to affect long-range vesicular transportation from the cell membrane, mainly via controlling the dynamics of caveolar vesicles (Aranda *et al.* 2011). DYRK3 was also identified in studies looking into kinases involved in clathrin- and caveolae/raft-mediated endocytosis (Pelkamans *et al.* 2005), but the direct substrates of DYRK3 affecting these biological functions were still unknown (Aranda *et al.* 2011). Here, analysis of gene ontology classes revealed that at least 88 DYRK3 substrates are involved in transport, including COPII vesicle coating, protein transport, and endocytosis, supporting the hypothesis that DYRK3 mediated phosphorylation has a role in vesicular transportation and endocytosis. In addition, DYRK3 shares same or similar substrates with DYRK1A and their functions seem to be similar in regards of

involvement in endocytosis. In rat models, DYRK1A has been shown to phosphorylate dynamin1 (DYN1) which is required in the formation of endocytic vesicles and amphiphysin I (AMPH) which regulates synaptic endocytosis (Murakami *et al.* 2009). Our results show that DYN1 and Dynamin-1-like protein (DNM1L), as well as an endocytosis regulating protein BIN1 were phosphorylated by DYRK3, indicting a similar substrate specificity within the DYRK family. Additionally, comparison between DYRK3 and DYRK4 substrates revealed that both DYRK3 and DYRK4 are likely to have similar roles in transport events in the cell. Within the DYRK subfamily, kinases have high structural similarity, differentiated mainly based on signals outside of the kinase domain, such as nuclear localization signals in the N-terminal end (Aranda *et al*. 2011; Figure 5). Therefore, even if the kinases phosphorylate same substrates in *in vitro* kinase assays, they may not phosphorylate exact same substrates *in vivo,* if their localizations differ. Likewise, substrates of CLK2 had gene ontology classifications linked to known functions of the kinase. It has been shown that CLKs phosphorylate serine/arginine-rich (SR) proteins, which are components of the spliceosomal complex (Bullock *et al.* 2009; Nayler *et al.* 1997). Kinase assay results show that in cytosolic HEK cell lysate, CLK2 phosphorylates Serine/arginine-rich splicing factors 7 and 11 (SRSF) and Serine/arginine repetitive matrix proteins 1 and 2 (SRRM), which are involved in mRNA slicing via spliceosome, which supports previous findings on CLK2's role in splicing. However, these proteins were not annotated as direct CLK2 substrates in the PhosphoSitePlus® database. Importantly, CLK2 has been shown to phosphorylate Aurora-b kinase, which regulates the abscission phase of cytokinesis (Petsalaki and Zachos, 2016), but in kinase assay, Aurora-b phosphorylation by CLK2 was not observed. Although, CLK2 phosphorylated six other proteins involved in cytoskeleton reorganization or cytokinesis, supporting the hypothesis that CLK2 has a role in these biological processes.

Kinase assay linked with LC-MS/MS can identify hundreds of direct substrates for kinases. However, false positives are possible if FSBA has not worked optimally, leaving some endogenous kinases unblocked. Using a negative control where the FSBA treated cell lysate is tested without an added kinase alleviates this problem. Another possible pitfall is that the removal of FSBA may be incomplete before the kinase assay, so that FSBA blocks the kinase of interest and only few substrates would be identified. Therefore, having more technical replicates would be desired, if it is feasible for the scope of the study. Here, the two cell lines worked as pseudoreplicates, increasing the confidence of true kinase-substrate pairs, when the same phosphorylation event was found in both cell lines. However, this is not enough to draw unambiguous conclusions on single kinase-substrate pairs. Nevertheless, this study provides a great amount of information on which kinases are interesting to study further and gives an overlook over the biological processes affected by protein phosphorylation.

It is necessary to acknowledge the importance of kinase-substrate interactions, as they regulate many biological, possibly pathological, processes in the cell. The method presented here is suitable for large-scale substrate screening and our results provide valuable information about direct substrates of CMGC kinases, especially for less-studied kinases such as the dual-specificity kinases in CLK, DYRK and HIPK subfamilies. The method is also applicable for specific kinase screens more closely related to clinical applications, for instance kinases considered as potential drug targets. For specific kinases, the method could be used in combination with suitable cell lines so that the *in vitro* assay results would mimic *in vivo* conditions as closely as possible. Knowing the direct substrates of kinases would make it easier to predict the outcome of kinase inhibition and how it affects cellular

signaling and biological processes. Importantly, there are multiple kinases in the CMGC family that are not extensively studied even though they regulate vital cellular processes. These kinases should be studied further, because they appear to have more prominent roles in cells as previously known and thus, they may become new potential targets for drug development. Therefore, it can be concluded that large-scale identification of direct kinase substrates with kinase assay linked LC-MS/MS is a useful tool to gain a complete overview of the human kinome.

VI. AKNOWLEDGEMENTS

I want to thank Markku Varjosalo for taking me into his research group first as a summer trainee and for providing me the opportunity to continue as a master's student. I am thankful for the great amount of responsibility that I got as a young and enthusiastic, but also a naïve student, taking my first steps into the world of academic research. I was lucky to have amazing supervisors, Tiina Öhman and Salla Keskitalo, who helped me throughout the process. I learned a lot about science and proteomics – but even more about life in general, especially about how to be patient if everything does not go my way with the first try. Thank you for understanding my desire to pursue a dual career with soccer and science and for being only one email, call or text away when I needed advice, no matter where in the world I was playing. Salla, special thanks for encouraging me to take this project and believing that I can carry through it regardless of the short period of time spent in Finland. Thank you Tiina for being my go-to person in the lab and for all the guidance in data-analysis and writing.

I want to thank my friends and family, in Finland and in the US. Mary, for keeping me in check and encouraging me to keep on working, Tim and Tiff for your love and support. To friends and family in Finland, I wouldn't have come this far without your support throughout the years.

VII. REFERENCES

Akoulitchev, S., Chuikov, S., & Reinberg, D. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature*, *407*(6800), 102. (2000).

Allen, B. L., & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature reviews Molecular cell biology*, *16*(3), 155. (2015).

Aranda, S., Laguna, A., & de la Luna, S. DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. *The FASEB Journal*, *25*(2), 449-462. (2011).

Bode, A. M., and Dong, Z. The functional contrariety of JNK. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, *46*(8), 591-598. (2007).

Bullock, A. N., Das, S., Debreczeni, J. É., Rellos, P., Fedorov, O., Niesen, F. H., ... & Turk, B. E. Kinase domain insertions define distinct roles of CLK kinases in SR protein phosphorylation. *Structure*, *17*(3), 352-362. (2009).

Cargnello, M., and Roux, P. P. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiology and molecular biology reviews*, *75*(1), 50-83. (2011).

Cheon, M. S., Fountoulakis, M., Cairns, N. J., Dierssen, M., Herkner, K., & Lubec, G. Decreased protein levels of stathmin in adult brains with Down syndrome and Alzheimer's disease. In *Protein Expression in Down Syndrome Brain* (pp. 281-288). Springer, Vienna. (2001).

Cohen, P., and Frame, S. The renaissance of GSK3. *Nature reviews Molecular cell biology*, *2*(10), 769. (2001).

Cox, J., and Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, *26*(12), 1367-1372. (2008).

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*, *10*(4), 1794-1805. (2011).

Cuadrado, A., & Nebreda, A. R. Mechanisms and functions of p38 MAPK signalling. *Biochemical Journal*, *429*(3), 403-417 (2010).

Déléris, P., Trost, M., Topisirovic, I., Tanguay, P. L., Borden, K. L., Thibault, P., & Meloche, S. Activation loop phosphorylation of ERK3/ERK4 by group I p21-activated kinases (PAKs) defines a novel PAK-ERK3/4-MAPK-activated protein kinase 5 signaling pathway. *Journal of Biological Chemistry*, *286*(8), 6470-6478. (2011).

Dhavan, R., and Tsai, L. H. A decade of CDK5. *Nature reviews Molecular cell biology*, *2*(10), 749. (2001).

Ferguson, F. M., and Gray, N. S. Kinase inhibitors: the road ahead. *Nature reviews Drug discovery*, *17*(5), 353. (2018).

Fu, T. J., Peng, J., Lee, G., Price, D. H., & Flores, O. Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. *Journal of Biological Chemistry*, *274*(49), 34527-34530. (1999).

Guo, X., Williams, J. G., Schug, T. T., & Li, X. DYRK1A and DYRK3 promote cell survival through phosphorylation and activation of SIRT1. *Journal of Biological Chemistry*, jbc-M110 (2010).

Hoeflich, K. P., Luo, J., Rubie, E. A., Tsao, M. S., Jin, O. U., & Woodgett, J. R. Requirement for glycogen synthase kinase-3β in cell survival and NF-κB activation. *Nature*, *406*(6791), 86. (2000).

Hopkins, A. L., and Groom, C. R. The druggable genome. *Nature reviews Drug discovery*, *1*(9), 727 (2002).

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res. 2015 43:D512-20. (2015)*

Hydbring, P., Malumbres, M., & Sicinski, P. Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. *Nature reviews Molecular cell biology*, *17*(5), 280 (2016).

Ishitani, T., Hirao, T., Suzuki, M., Isoda, M., Ishitani, S., Harigaya, K., ... & Itoh, M. Nemo-like kinase suppresses Notch signalling by interfering with formation of the Notch active transcriptional complex. *Nature cell biology*, *12*(3), 278. (2010).

Ishitani, T., & Ishitani, S. Nemo-like kinase, a multifaceted cell signaling regulator. *Cellular signalling*, *25*(1), 190-197. (2013).

Kanei-Ishii, C., Ninomiya-Tsuji, J., Tanikawa, J., Nomura, T., Ishitani, T., Kishida, S., ... & Matsumoto, K. Wnt-1 signal induces phosphorylation and degradation of c-Myb protein via TAK1, HIPK2, and NLK. *Genes & development*, *18*(7), 816-829. (2004).

Kannan, N., and Neuwald, A. F. Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2α. *Protein Science*, *13*(8), 2059-2077. (2004).

Krupa, A., Preethi, G., & Srinivasan, N. Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. *Journal of molecular biology*, *339*(5), 1025-1039 (2004).

Lau, A. T., & Xu, Y. M. Regulation of human mitogen-activated protein kinase 15 (extracellular signal-regulated kinase 7/8) and its functions: A recent update. *Journal of Cellular Physiology*, *234*(1), 75-88. (2019).

Landry, J. J., Pyl, P. T., Rausch, T., Zichner, T., Tekkedil, M. M., Stütz, A. M., ... & Gagneur, J. The genomic and transcriptomic landscape of a HeLa cell line. *G3: Genes, Genomes, Genetics*, *3*(8), 1213-1224. (2013).

Li, X., Zhang, R., Luo, D., Park, S. J., Wang, Q., Kim, Y., & Min, W. Tumor necrosis factor α-induced desumoylation and cytoplasmic translocation of homeodomain-interacting protein kinase 1

are critical for apoptosis signal-regulating kinase 1-JNK/p38 activation. Journal of Biological Chemistry, 280(15), 15061-15070. (2005).

Liu, W., Li, J., Song, Y. S., Li, Y., Jia, Y. H., and Zhao, H. D. Cdk5 links with DNA damage response and cancer. *Molecular cancer*, *16*(1), 60. (2017).

Malumbres, M., and Barbacid, M. Mammalian cyclin-dependent kinases. *Trends in biochemical sciences*, *30*(11), 630-641. (2005).

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J. M., Manning, G., ... and Wolgemuth, D. J. Cyclin-dependent kinases: a family portrait. *Nature cell biology*, *11*(11), 1275. (2009)

Malumbres, M., and Barbacid, M., Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews cancer*, *9*(3), 153 (2009).

Malumbres, M. Cyclin-dependent kinases. *Genome biology*, *15*(6), 122. (2014).

Mann, M., Ong, S. E., Grønborg, M., Steen, H., Jensen, O. N., & Pandey, A. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology*, *20*(6), 261-268. (2002).

Mann, M., & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature biotechnology*, *21*(3), 255. (2003).

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. The protein kinase complement of the human genome. *Science*, *298*(5600), 1912-1934. (2002).

Medina, M., and Wandosell, F. Deconstructing GSK-3: the fine regulation of its activity. *International Journal of Alzheimer's Disease*, (2011).

Meloche, S., and Pouysségur, J. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1-to S-phase transition. *Oncogene*, *26*(22), 3227. (2007).

Meng, L., Michaud, G. A., Merkel, J. S., Zhou, F., Huang, J., Mattoon, D. R., & Schweitzer, B. Protein kinase substrate identification on functional protein arrays. *BMC biotechnology*, *8*(1), 22. (2008).

Modi, V., and Dunbrack, R. L. Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences*, *116*(14), 6818-6827 (2019).

Müller, A. C., Giambruno, R., Weißer, J., Májek, P., Hofer, A., Bigenzahn, J. W., ... & Bennett, K. L. Identifying kinase substrates via a heavy ATP kinase assay and quantitative mass spectrometry. *Scientific reports*, *6*, 28107 (2016).

Murakami, N., Bolton, D., & Hwang, Y. W. Dyrk1A binds to multiple endocytic proteins required for formation of clathrin-coated vesicles. *Biochemistry*, *48*(39), 9297-9305. (2009).

Nayler, O., Stamm, S., and Ullrich, A. Characterization and comparison of four serine-and arginine-rich (SR) protein kinases. *Biochemical Journal*, *326*(3), 693-700. (1997).

Nolen, B., Taylor, S., and Ghosh, G. Regulation of protein kinases: controlling activity through activation segment conformation. *Molecular cell*, *15*(5), 661-675. (2004)

Oliveros, J.C. Venny. An interactive tool for comparing lists with Venn's diagrams. (2007-2015). https://bioinfogp.cnb.csic.es/tools/venny/index.html

Papadopoulos, C., Arato, K., Lilienthal, E., Zerweck, J., Schutkowski, M., Chatain, N., ... & de la Luna, S. Splice variants of the dual specificity tyrosine phosphorylation-regulated kinase 4 (DYRK4) differ in their subcellular localization and catalytic activity. *Journal of Biological Chemistry*, *286*(7), 5494-5505. (2011).

Patrick, G. N., Zukerberg, L., Nikolic, M., de La Monte, S., Dikkes, P., & Tsai, L. H. Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration. *Nature*, *402*(6762), 615. (1999).

Pelkmans, L., Fava, E., Grabner, H., Hannus, M., Habermann, B., Krausz, E., and Zerial, M. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* 3571, 1–9 (2005).

Perander, M., Al-Mahdi, R., Jensen, T. C., Nunn, J. A., Kildalsen, H., Johansen, B., ... & Seternes, O. M. Regulation of atypical MAP kinases ERK3 and ERK4 by the phosphatase DUSP2. *Scientific reports*, *7*, 43471. (2017).

Peterlin, B. M., & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Molecular cell*, *23*(3), 297-305. (2006).

Petsalaki, E., & Zachos, G. Clks 1, 2 and 4 prevent chromatin breakage by regulating the Aurora B-dependent abscission checkpoint. *Nature communications*, *7*, 11451. (2016).

Rodgers, J. T., Haas, W., Gygi, S. P., & Puigserver, P. Cdc2-like kinase 2 is an insulin-regulated suppressor of hepatic gluconeogenesis. *Cell metabolism*, *11*(1), 23-34. (2010).

Roskoski Jr, R. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacological research*, *66*(2), 105-143 (2012).

Strumillo, M. J., Oplova, M., Vieitez, C., Ochoa, D., Shahraz, M., Busby, B. P., ... & Beltrao, P. Conserved phosphorylation hotspots in eukaryotic protein domain families. *Nature communications*, *10*(1), 1977. (2019).

Sugiyama, N., and Ishihama, Y. Large-scale profiling of protein kinases for cellular signaling studies by mass spectrometry and other techniques. *Journal of pharmaceutical and biomedical analysis*, *130*, 264-272. (2016).

Sun, B., Mason, S., Wilson, R. C., Hazard, S. E., Wang, Y., Fang, R., ... & Zhao, J. J. Inhibition of the transcriptional kinase CDK7 overcomes therapeutic resistance in HER2-positive breast cancers. *Oncogene*, 1-14. (2019).

Taira, N., Nihira, K., Yamaguchi, T., Miki, Y., and Yoshida, K. DYRK2 is targeted to the nucleus and controls p53 via Ser46 phosphorylation in the apoptotic response to DNA damage. *Mol. Cell*. 25, 725–738 (2007).

Tape, C. J., Worboys, J. D., Sinclair, J., Gourlay, R., Vogt, J., McMahon, K. M., ... & Jørgensen, C. Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC. *Analytical chemistry*, *86*(20), 10296-10302. (2014).

Taylor, S. S., and Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences*, *36*(2), 65-77. (2011).

Tripathi, B. K., & Zelenka, P. S. Cdk5: A regulator of epithelial cell adhesion and migration. *Cell adhesion & migration*, 4(3), 333-336. (2010).

van der Laden, J., Soppa, U., & Becker, W. Effect of tyrosine autophosphorylation on catalytic activity and subcellular localisation of homeodomain-interacting protein kinases (HIPK). *Cell Communication and Signaling*, *13*(1), 3. (2015).

Varjosalo, M., Keskitalo, S., Van Drogen, A., Nurkkala, H., Vichalkovski, A., Aebersold, R., & Gstaiger, M. The protein interaction landscape of the human CMGC kinase group. *Cell reports*, *3*(4), 1306-1320 (2013).

Verheyen, E. M., & Gottardi, C. J. Regulation of Wnt/β-catenin signaling by protein kinases. *Developmental dynamics*, *239*(1), 34-44. (2010).

Xue, L., Wang, W. H., Iliuk, A., Hu, L., Galan, J. A., Yu, S., ... & Tao, W. A. Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proceedings of the National Academy of Sciences*, *109*(15), 5615-5620. (2012).

Yan, L., Carr, J., Ashby, P. R., Murry-Tait, V., Thompson, C., & Arthur, J. S. C. Knockout of ERK5 causes multiple defects in placental and embryonic development. *BMC developmental biology*, *3*(1), 11. (2003).

Zhang, Q., Yoshimatsu, Y., Hildebrand, J., Frisch, S. M., & Goodman, R. H. Homeodomain interacting protein kinase 2 promotes apoptosis by downregulating the transcriptional corepressor CtBP. *Cell*, *115*(2), 177-186. (2003).

Zhang, J., Shen, B., & Lin, A. Novel strategies for inhibition of the p38 MAPK pathway. *Trends in pharmacological sciences*, *28*(6), 286-295. (2007).

Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A. J., ... & Mohammed, S. Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nature protocols*, *8*(3), 461. (2013).