Phosphoproteomics Analyses to Identify the Candidate Substrates and Signaling Intermediates of the Non-Receptor Tyrosine Kinase, SRMS

A Thesis Submitted to the College of
Graduate and Postdoctoral studies
In Partial Fulfillment of the Requirements For
The Degree of Doctor of Philosophy
In the Department of Biochemistry
University of Saskatchewan,
Saskatoon

By
RAGHUVEERA KUMAR GOEL

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for the Doctoral degree in Biochemistry from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was completed. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan for any scholarly use of the findings described in my thesis.

Requests for permission to copy or to make other use of any material in this thesis, in whole or part, should be addressed to:

Head of the Department of Biochemistry, Microbiology and Immunology
Health Sciences Building, 107 Wiggins road, University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5
Canada

OR

Dean, College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

SRMS (Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylaton sites) is a non-receptor tyrosine kinase that belongs to the BRK family kinases (BFKs) and is evolutionarily related to the Src family kinases (SFKs). Like SFKs and BFKs, the SRMS protein comprises of two domains involved in protein-protein interactions, namely, the Src-homology 3 domain (SH3) and Src-homology 2 domain (SH2) and one catalytic kinase domain. Unlike members of the BFKs and SFKs, the biochemical and cellular role of SRMS is poorly understood primarily due to the lack of information on the substrates and signaling intermediates regulated by the kinase. Previous biochemical studies have shown that wild type SRMS is enzymatically active and leads to the tyrosine-phosphorylation of several proteins, when expressed exogenously in mammalian cells. These tyrosine-phosphorylated proteins represent the candidate cellular substrates of SRMS which are largely unknown. Further, previous studies have determined that the SRMS protein displays a characteristic punctate cytoplasmic localization pattern in mammalian cells. These SRMS cytoplasmic puncta are uncharacterized and may provide insights into the biochemical and cellular role of the kinase.

Here, we utilized mass spectrometry-based quantitative label-free phosphoproteomics to (a) identify the candidate SRMS cellular substrates and (b) candidate signaling intermediates regulated by SRMS, in HEK293 cells expressing ectopic SRMS. Specifically, using a phosphotyrosine enrichment strategy we identified 663 candidate SRMS substrates and consensus substrate-motifs of SRMS. We used customized peptide arrays and performed the high-throughput validation of a subset of the identified candidate SRMS substrates. Further, we independently validated Vimentin and Sam68 as bonafide SRMS substrates. Next, using Titanium dioxide (TiO2)-based phosphopeptide enrichment columns, we identified multiple signaling intermediates of SRMS. Functional gene enrichment analyses revealed several common and unique cellular processes regulated by the candidate SRMS substrates and signaling intermediates. Overall, these studies led to the identification of a significant number of novel and biologically relevant SRMS candidate substrates and signaling intermediates, which mapped to a number of cellular and biological processes primarily involved in cell cycle regulation, apoptosis, RNA processing, DNA

repair and protein synthesis. These findings provide an important resource for future mechanistic studies to investigate the cellular and physiological functions of the SRMS.

Studies towards characterizing the SRMS cytoplasmic puncta showed that the SRMS punctate structures do not colocalize with some of the major cellular organelles investigated, such as the mitochondria, endoplasmic reticulum, golgi bodies and lysosomes. However, studies investigating the involvement of the SRMS domains in puncta-localization revealed that the SRMS SH2 domain partly regulates this localization pattern. These results highlight the potential role of the SRMS SH2 domain in the localization of SRMS to these cytoplasmic sites and lay important groundwork for future characterization studies.

ACKNOWLEDGMENTS

I deem it invaluable to thank my supervisor and mentor, Dr. Kiven Erique Lukong for his guidance, and perseverance towards the completion of my Doctoral program. I extend my gratitude to the present and former members of our lab, Yetunde Ogunbolude, Joshua MacAusland-Berg, Aditya Mandapati, Chenlu Dai, Dr. Sayem Miah, Dr. Edward Bagu and Lexie Martin. I would like to express my special and heartfelt gratitude to Akanksha Baharani for her great friendship and help all these years. I greatly appreciate the assistance of the former and present administrative secretaries of our Department, Mrs. Margaret Strohan, Mrs. Lori Lisitza and Mrs. Deana Thunderchild. I also thank the members of my supervisory committee, Dr. William J. Roesler (Department Head), Dr. Stanley Moore, Dr. Deborah Anderson (Director of Research, Saskatchewan Cancer Agency), Dr. Jeremy Lee (Department Graduate Chair) and Dr. Vikram Misra (Department of Veterinary Microbiology and Cognate member of the advisory committee). And it is but in no less measure that I also extend my gratitude to all personnel of the Health Sciences Supply center, College of Medicine.

I cherish and laud the priceless support and faith bestowed upon me by my family and Dr. Lukong, alike.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABL	Abelson Murine Leukemia viral oncogene homolog 1
ACK	Activated Cdc42-associated Kinase
AGC	Protein kinase A, protein kinase G and protein kinase C
ALK	Anaplastic Lymphoma Kinase
APS	Ammonium Persulfate
BCR-ABL	Breakpoint Cluster Region- Abelson Murine leukemia viral oncogene homolog 1
BRK	Breast Tumor Kinase
c-ABL	Cellular Abelson murine Leukemia viral oncogene homolog 1
CaMK	Calcium/Calmodulin-dependent Kinase
Cdc2	Cell Division Control protein 2
Cdc42	Cell Division Control protein 42
CDK	Cyclin-Dependent Kinase
CDK1/2/3/5/9	Cyclin-Dependant Kinase 1/2/3/5/9
CKAP2	Cytoskeletal-Associated Protein 2
CLK	CDC-like kinase
CRK	CT-10 Regulator of Kinase
CSK	c-Src Kinase
CK	Casein kinase
CK1	Casein Kinase 1
CK2α	Casein kinase 2 alpha
CMGC	CDK, MAPK, GSK, CLK
DAPI	4,6'-diamidino-2-phenylindole
DDR	Discoidin Domain Receptor
Dok1	Downstream of Tyrosine kinases 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
DTT	Dithiothreitol

EDTA	Ethylenediaminetetraacetic Acid
EPH	Ephrin
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EGTA	Ethylene Glycol Tetraacetic Acid
EIF4A3/1	Eukaryotic Initiation Factor 3/1
ER	Endoplasmic Reticulum
ERM	Ezrin Radixin Moesin
FAK	Focal Adhesion Kinase
FDR	False Discovery Rate
FES	Feline Sarcoma
FGFR	Fibroblast Growth Factor Receptor
FRK	Fyn-Related Kinase
GST	Glutathione S-Transferase
GFP	Green Fluorescent protein
GNL3L	G-protein Nucleolar 3-Like
GRB10	Growth factor Receptor-Bound protein 10
GSK	Glycogen Synthase Kinase
GSK3 beta	Glycogen Synthase Kinase 3 Beta
HEK293	Human Embryonic Kidney 293
HMGA1	High mobility group protein A1
HNRNPM	Heterogenous Nuclear Ribonucleoprotein M
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Insulin Receptor
JAK	Janus Kinase
JEF	JAK, ERM, FAK
KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1
KRT18	Keratin 18
LC-MS	Liquid Chromatography- Mass Spectrometry

LMR	Lemur
MAPK	Mitogen Activated Protein Kinase
MCM7	Minichromosome maintenance complex component 7
MET	Mesenchymal Epithelial Transition
MOPS	3-(N-morpholino) propanesulfonic acid
MuSK	Muscle-Specific Kinase
MYH9	Myosin heavy chain 9
mTOR	Mammalian Target of Rapamycin
NRTK	Non-Receptor Tyrosine Kinase
NUCL	Nucleolin
PBS	Phosphate Buffered Saline
PDGFR	Platelet-Derived Growth Factor Receptor
PEI	Polyethyleneimine
PFA	Paraformaldehyde
PHOSIDA	Phosphorylation Site Database
PI3K	Phosphatidylinositol 3-kinase
PKA/G/C	Protein kinase A/G/C
PTEN	Phosphatase and Tensin Homolog deleted on chromosome 10
PTK	Protein Tyrosine Kinase
PTK5/6/7/70	Protein Tyrosine Kinase 5/6/7/70
PTM	Post-Translational Modification
PTPN13	Protein Tyrosine Phosphatase, Non-receptor type 13
PMSF	Phenylmethanesulfonyl fluoride
RET	Rearranged during Transfection
RTK	Receptor Tyrosine Kinase
ROR	RAR-related Orphan Receptor
RYK	Related to tyrosine Kinase
ΔN SRMS	N-terminal deletion SRMS
Sam68	Src-Associated substrate during Mitosis of 68 kDa
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

SH2	Src-homology 2
SH3	Src-Homology 3
SRMS	Src-Related kinase lacking C-terminal regulatory tyrosine and N-terminal Myristoylation Sites
STE	Sterile Kinase
STE7/11/20	Sterile Kinase 7/11/20
STYK1	Serine/Threonine and tyrosine receptor Kinase 1
SYK	Spleen Tyrosine kinase
TAM	Tyro3-Axl-Mer
TFA	Trifluoroacetic Acid
TEC	Transient Erythroblastopenia Of Childhood
TIE	Tyrosine kinase with Immunoglobulin-like and EGF-like domains
TNF	Tumor Necrosis Factor
TRK	Tropomycin Receptor Kinase
TK	Tyrosine Kinase
TKL	Tyrosine Kinase-Like
VEGFR	Vascular Endothelial Growth Factor Receptor
UMPS	Uridine Monophosphate Synthetase

1. Review of the literature

1.1 Tyrosine kinases

Tyrosine kinases (TKs) or protein tyrosine kinases (PTKs) are enzymes that catalyze the transfer of a phosphate group from a molecule of ATP to the tyrosine residues of proteins (Figure 1) (Knighton *et al.*, 1993). The phosphorylation of tyrosine residues on proteins by tyrosine kinases entails many biologically important consequences in a cellular and physiological context (Hunter, 2014). By phosphorylating tyrosine residues on proteins, tyrosine kinases elicit conformational changes in these proteins which affects their biochemical function and/or intermolecular binding to other proteins, which together modulates intracellular signaling (Groban *et al.*, 2006; Kimura *et al.*, 1996). Such phosphorylation is reversible where specific tyrosine phosphatases catalyze the removal of the phosphate group(s) from the tyrosine residue(s) of these proteins (Figure 1) (Frearson and Alexander, 1997; Stoker, 2005). In this manner, the biochemical function of tyrosine kinases is regulated in cells (Frearson and Alexander, 1997; Stoker, 2005). With the culmination of the human genome project, a total of 90 tyrosine kinases were found to

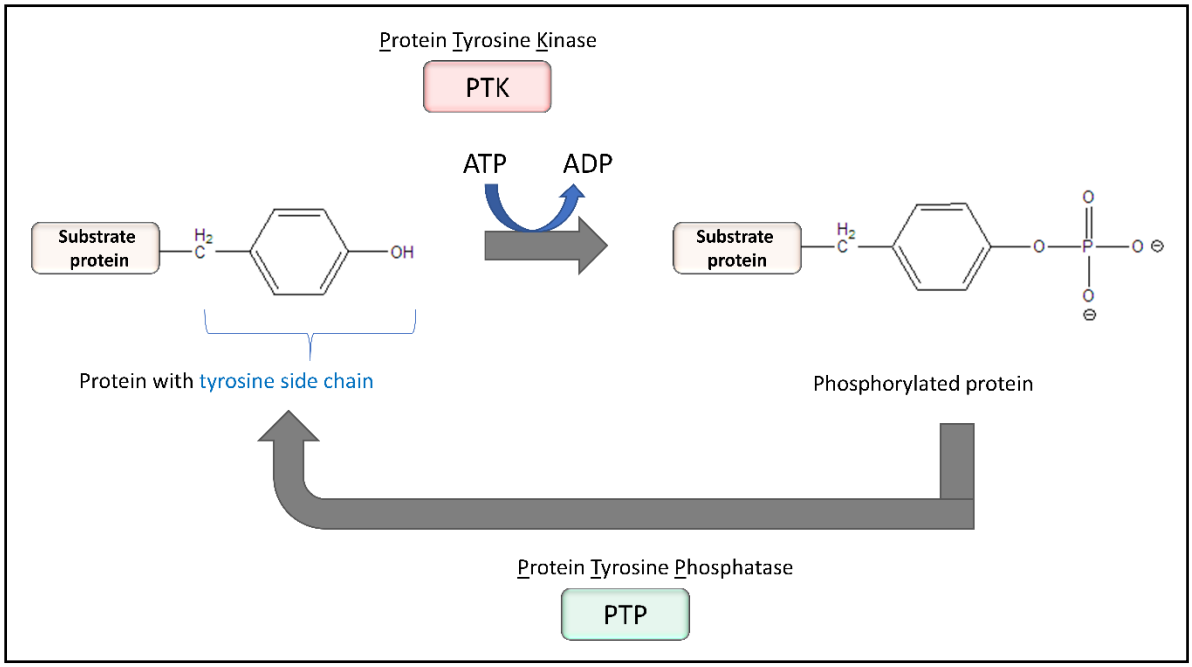


Figure 1.1: Overview of protein tyrosine phosphorylation and dephosphorylation: Protein tyrosine kinases (PTKs) covalently attach a phosphate group, derived from an ATP moiety, to the nucleophilic hydroxyl group of a tyrosine residue on a protein. Protein tyrosine phosphatases (PTPs) in turn catalyze the de-phosphorylation of these residues by removing the phosphate group.

be coded by the human genome (Manning *et al.*, 2002; Robinson *et al.*, 2000). Of these, 58 are receptor type and 32 are non-receptor type tyrosine kinases (Manning *et al.*, 2002; Robinson *et al.*, 2000).

1.1.1 Receptor tyrosine kinases

Receptor tyrosine kinases (RTKs) are embedded in the cell membrane and possess both, an extracellular domain and an intracellular domain (Lemmon and Schlessinger, 2010; Manning *et al.*, 2002). The extracellular domain is involved in binding to ligands whereas the intracellular portion of the RTK is comprised of the kinase domain which catalyzes the phosphorylation of tyrosine residues (Hubbard and Miller, 2007; Lemmon and Schlessinger, 2010). A total of 58 different receptor tyrosine kinases have been identified to date in the mammalian kinome (Figure 2). These are grouped into 20 sub-families, namely, the ErbB, IR (Insulin receptor), PDGFR (Platelet-derived growth factor receptor), VEGFR (Vascular endothelial growth factor receptor), FGFR (Fibroblast growth factor receptor), PTK7 (Protein tyrosine kinase 7), TRK (Tropomycin receptor kinase), ROR (RAR-related orphan receptors), MuSK, (Muscle-specific kinase), MET (Mesenchymal Epithelial Transition), TAM (Tyro3-Axl-Mer), TIE (Tyrosine kinase with immunoglobulin-like and EGF-like domains), EPH (Ephrin), RET (Rearranged during transfection), RYK (Related to tyrosine kinase), DDR (Discoidin domain receptor), ROS (c-Ros1), LMR (Lemur), ALK (Anaplastic lymphoma kinase) and STYK1 (Serine/threonine and tyrosine receptor kinase 1) (Figure 2) (Lemmon and Schlessinger, 2010). The extracellular domains of receptor tyrosine kinases bind to their respective ligands, triggering receptor dimerization which results in the autophosphorylation of tyrosine residues on the intracellular kinase domains of the receptor (Lemmon and Schlessinger, 2010). These phosphotyrosine residues serve as binding sites for specific cytoplasmic proteins which relay signaling downstream to elicit specific cellular functions (Pawson, 1995). Notably, the LMR family, STYK1 family and ErbB2 kinase represent the only known receptor tyrosine kinases that do not possess an extracellular ligand-binding domain (Lemmon and Schlessinger, 2010). Ligand stimulation-based activation of receptor tyrosine kinases leads to the activation of specific intracellular signaling cascades that involve a variety of cytoplasmic proteins. Such proteins include, but are not limited to, cytoplasmic kinases and adapter or scaffolding proteins (Lemmon and Schlessinger, 2010). The signals eventually culminate in the transcriptional regulation of specific genes in the nucleus and/or the modulation

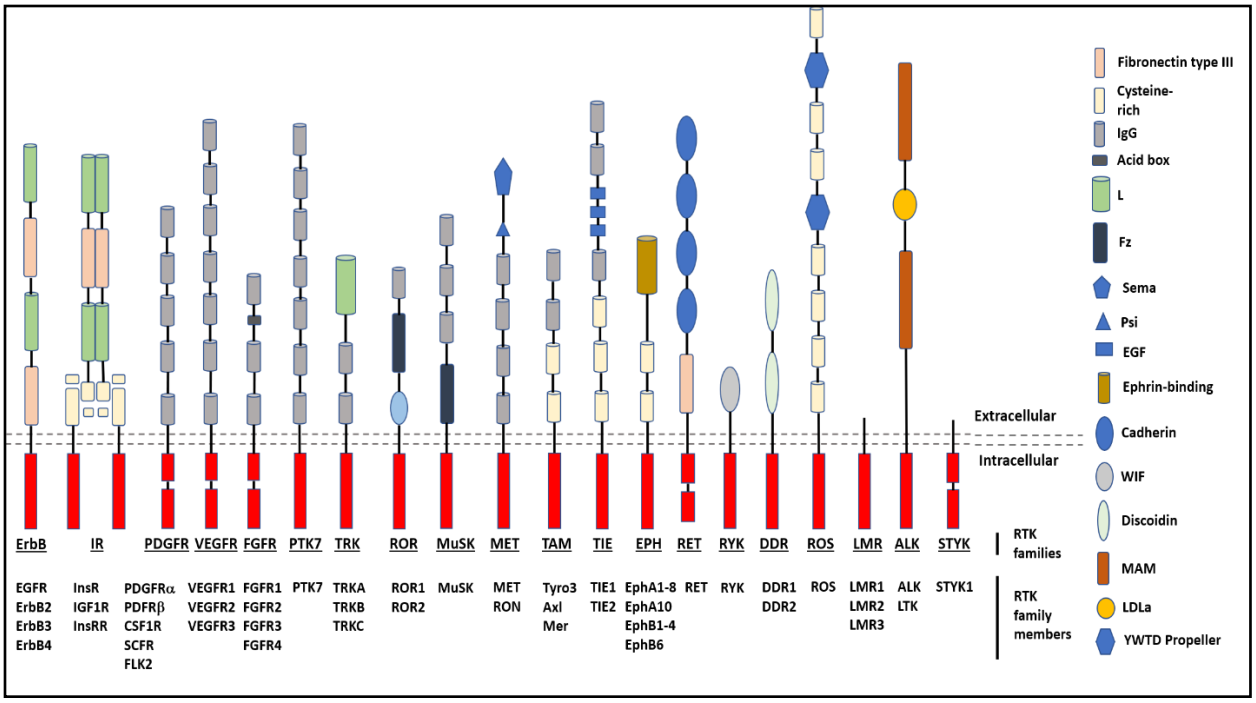


Figure 1.2: The mammalian receptor tyrosine kinase (RTK) complement: Shown here is a schematic representation of the 20 receptor tyrosine kinase families encoded by the mammalian kinome (Lemmon and Schlessinger, 2010). The kinase members of each family are indicated. The intracellular side of these transmembrane receptor tyrosine kinases comprises the catalytic/kinase domain (highlighted in red). The extracellular side of these kinases represent various subunits that form the ligand-binding domain. These include the L (Leucine-rich), Cysteine-rich, Fibronectin type III, Cadherin, Discoidin, Ig (Immunoglobulin), EGF (Epithelial Growth Factor), Psi, WIF (Wnt Inhibitory Factor), Ephrin-binding, Fz (Frizzled), LDLa (Low-Density Lipoprotein class a) YWTD (tyrosine-tryptophan-threonine-aspartic acid), Acid box, Sema and MAM (Meprin/A5-protein/PTPmu) subunits.

of protein function in the cell thereby regulating important biological processes such as cell proliferation, migration, apoptosis and senescence, to name a few (Lemmon and Schlessinger, 2010; Zwick *et al.*, 2001).

1.1.1.1 EGFR signaling pathway

The EGFR (Epithelial growth factor receptor) signaling pathway is one of the most well-characterized and biologically important receptor tyrosine kinase-mediated signal transduction pathways that regulates growth, proliferation and differentiation in mammalian cells (Oda *et al.*, 2005). The EGFR receptor belongs to the ErbB family of RTKs, whose other three members include ErbB2, ErbB3 and ErbB4, all of which display conserved structural and functional characteristics. Several ligands such as EGF (Epidermal growth factor), TGFα (Transforming

growth factor alpha), epiregulin, amphiregulin, epigen and betacellulin, are known to bind to EGFR. EGF, however, is the major high-affinity ligand of EGFR (Oda *et al.*, 2005).

Upon ligand-binding, EGFR undergoes homo- or heterodimerization with another EGFR monomer or monomers of other members of the ErbB family, respectively (Figure 1.3) (Oda *et al.*, 2005; Yarden and Shilo, 2007). Dimerization stimulates the activation of the cytoplasmic kinase domains of EGFR which results in the autophosphorylation of tyrosine residues within the cytoplasmic kinase domain (Yarden and Shilo, 2007). These autophosphorylation residues are Y1092, Y1172, Y1197, Y1110 and Y1016 (Nogami *et al.*, 2003). These autophosphorylated residues serve as docking sites for the recruitment of various cytoplasmic proteins to the receptor (Figure 1.3) (Oda *et al.*, 2005). These proteins bind to the docking sites on EGFR via their SH2 and/or PTB domains (Figure 1.3).

Following binding to EGFR, these proteins relay signaling through various intracellular pathways to elicit distinct cellular functions. Notably, EGFR heterodimerization with other ErbB members recruit multiple other signaling proteins which considerably adds to the complexity of EGFR-signaling (Oda *et al.*, 2005). Figure 1.3 shows some of the major EGFR activation-dependent signaling pathways. For instance, Grb2 (Growth factor receptor-bound protein 2) is recruited to autophosphorylated EGFR via the Grb2 SH2 domain (Figure 1.3A) (Capuani *et al.*, 2015). Grb2 then binds to the guanine nucleotide exchange factor SOS (Son of sevenless) via the Grb2 SH3 domains (Zarich *et al.*, 2006). Upon binding to Grb2, SOS becomes activated (Katz *et al.*, 2007; Zarich *et al.*, 2006). Activated SOS leads to the activation of Ras which in turn activates Raf kinase activity (Katz *et al.*, 2007). Raf kinase phosphorylates and activates Mek1/2 thereby activating a MAPK (Mitogen-activated protein kinase) such as Erk1/2 (Extracellular signal-regulated kinase 1/2) (Katz *et al.*, 2007). Activated Erk1/2 translocates to the nucleus to phosphorylate transcription factors such as Elk1 and c-myc which transactivate target genes implicated in cell survival and proliferation (Figure 1.3A) (Schuhmacher and Eick, 2013; Shin *et al.*, 2011).

Likewise, JAK1/2 (Janus kinase1/2) is recruited to activated EGFR resulting in the phosphorylation and activation of the JAK proteins (Figure 1.3B) (Andl *et al.*, 2004). Activated JAK1/2 then phosphorylate STAT3 (Signal transducer and activator of transcription3) resulting in STAT3-dimerization (Andl *et al.*, 2004). Activated STAT3 dimers translocate to the nucleus

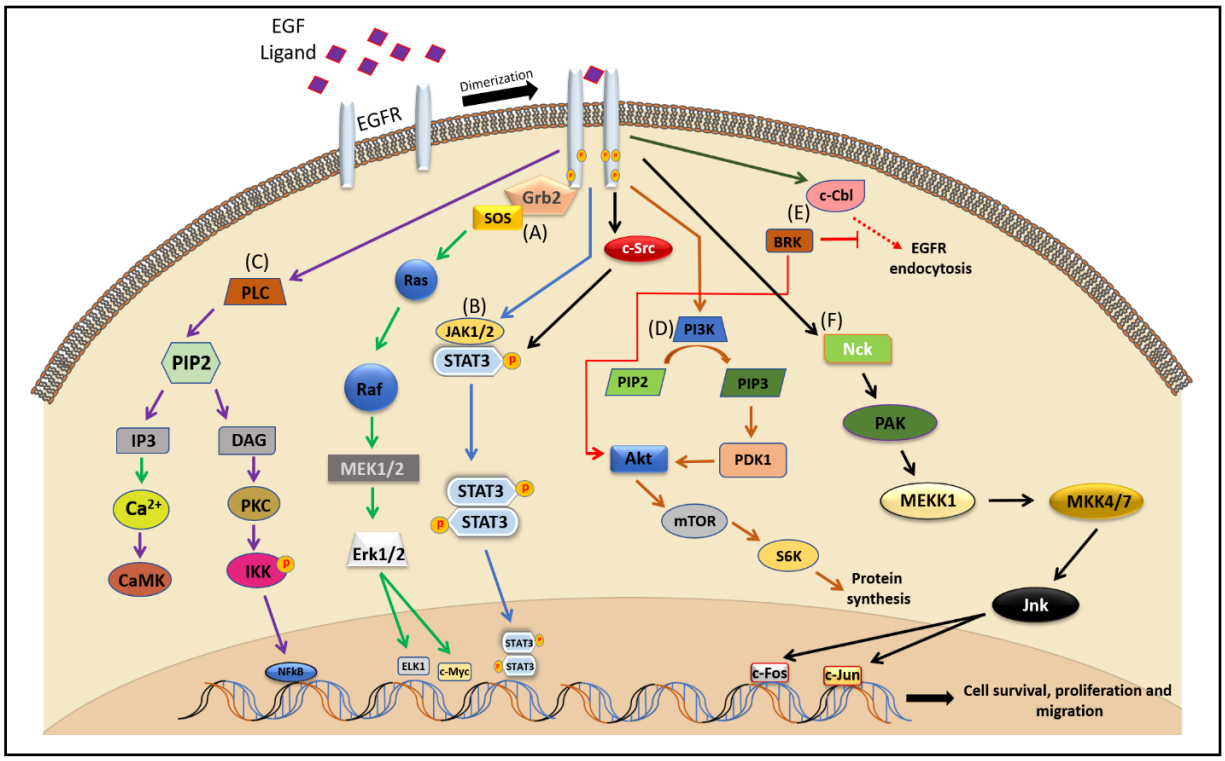


Figure 1.3: The EGFR-mediated signaling pathways. A schematic representation of the major signaling pathways propagated downstream of ligand-induced activation of EGFR. Following stimulation with EGF, the EGFR monomers undergo dimerization which activates the cytoplasmic kinase domains of the receptor. As a result of enzymatic activation, EGFR undergoes autophosphorylation resulting in the formation of docking sites for the recruitment of various cytoplasmic adapter proteins and kinases. These cytoplasmic proteins relay downstream signaling leading to the activation of various types of signaling cascades. **A.** Recruitment of Grb2 to docking sites on activated EGFR leads to signaling through Ras, Raf and Mek1/2 resulting in Erk1/2 activation which enhances the transcriptional activity of ELK1 and c-Myc. **B.** Activated EGFR recruits and phosphorylates JAK1/2. Active JAK1/2 phosphorylates STAT3, causing STAT3 dimerization, nuclear translocation and transactivation of STAT3-target genes. c-Src, activated downstream of EGFR, also directly phosphorylates STAT3 leading to STAT3-dimerization and activation. **C.** Activation of PLC downstream of EGFR leads to the hydrolysis of PIP2 into IP3 and DAG. IP3 stimulates Calcium ion release from the endoplasmic reticulum thereby stimulating CaMK signaling. Increase in cellular DAG concentration stimulates the activation of PKC which activates NFκB via IKK leading to the transactivation of NFκB-target genes. **D.** Activated PI3K catalyzes the conversion of PIP2 to PIP3. PIP3 promotes the PDK1-mediated activation of Akt which triggers the mTOR signaling via S6K resulting in enhanced protein synthesis. **E.** c-Cbl attenuates membrane-bound EGFR expression and downstream signaling by promoting EGFR endocytosis, a process which is negatively regulated by BRK. Further, BRK also direcly phosphorylates AKT leading to AKT activation. **F.** Nck-binding to EGFR activates PAK which potentiates Jnk signaling through MEKK1 and MKK4. Activated Jnk translocates to the nucleus to phosphorylate and activate, c-Fos and c-Jun. Jnk-mediated activation of c-Fos and c-Jun results in the transactivation of c-Fos and c-Jun-target genes. All EGFR-regulated transcriptional factors are involved in the transcriptional activation of genes broadly involved in cell survival, proliferation and migration.

where these transactivate target genes involved in cell proliferation and migration (Figure 1.3B) (Carpenter and Lo, 2014; Sherry *et al.*, 2009). c-Src is also known to be activated downstream of EGF signaling where activated c-Src directly phosphorylates STAT3 leading to STAT3 dimerization (Figure 1.3) (Cao *et al.*, 1996; Osherov and Levitzki, 1994; Schreiner *et al.*, 2002).

EGFR also activates Phospholipase C which results in the hydrolysis of PIP2 (Phosphatidylinositol-4,5-bisphosphate) to IP3 (Inositol trisphosphate) and DAG (1,2-Diacylglycerol) (Iwabu *et al.*, 2004; Xie *et al.*, 2010). IP3 causes the release of calcium ions (Ca^{2+}) from the endoplasmic reticulum which activate calcium calmodulin-dependent kinase (CaMK)-mediated signaling (Sun *et al.*, 2011). An increase in the cellular concentration of DAG triggers the activation of PKC (Protein kinase C) which in turn phosphorylates and activates IKK (IκB kinase) (Lallena *et al.*, 1999). IKK activates the transcription factor NFκB leading to the transactivation of NFκB-target genes such as those involved in cell survival and growth (Figure 1.3C) (Cheng *et al.*, 2011; Israel, 2010).

One of the major downstream signaling modules of EGFR is PI3K (Phosphatidylinositide 3-kinases) (Figure 1.3D) (Freudlsperger *et al.*, 2011; Slomovitz and Coleman, 2012). PI3K is recruited to and activated by EGFR upon ligand stimulation. Active PI3K catalyzes the conversion of PIP2 to PIP3 (Phosphatidylinositol-3,4,5-trisphosphate). PIP3 recruits Akt and PDK1 (Phosphoinositide-dependent kinase1) together which allows PDK1 to phosphorylate and activate Akt (Freudlsperger *et al.*, 2011; Slomovitz and Coleman, 2012). Akt is a key regulator of several signaling proteins, one of which includes mTOR (Mammalian target of rapamycin) (Laplante and Sabatini, 2012). Akt phosphorylates and inhibits the activity of the mTOR-inhibitor, TSC2 (Tuberous sclerosis complex2), thereby relieving the inhibitory constraints on mTOR. Activated mTOR phosphorylates and activates S6K1/2 (Ribosomal protein S6 kinase1/2) thereby promoting protein synthesis (Figure 1.3D) (Laplante and Sabatini, 2012).

EGFR signaling is also negatively regulated by the ubiquitin ligase c-Cbl, which promotes EGFR endocytosis and subsequent degradation via the lysosomal pathway (Ravid *et al.*, 2004; Visser Smit *et al.*, 2009). In this context, BRK has been shown to antagonize c-Cbl-mediated EGFR endocytosis (Figure 1.3 E) (Li *et al.*, 2012). Further, EGF-induced activation of EGFR has been shown to promote EGFR-BRK association and BRK activation. Activated BRK further potentiates EGFR signaling by directly phosphorylating and activating EGFR, thereby forming a positive feedback loop for EGFR signaling (Li *et al.*, 2012).

Other downstream effectors of EGFR include Nck (Non-catalytic region of tyrosine kinase) which is recruited to the activated receptor via its SH2 domain (Figure 1.3F) (Tomar and Schlaepfer, 2010). Nck then binds to Pak1 via the Nck SH3 domain (Tomar and Schlaepfer, 2010). Nck-bound PAK1 is then activated at the membrane by binding to Rac1 and cdc42 (Lu and Mayer, 1999). Activated PAK1 in turn activates Jnk (c-Jun kinase) which is mediated via MEKK1 and MKK4 (Tomar and Schlaepfer, 2010). Active Jnk undergoes nuclear translocation where it phosphorylates and activates the transcription factors c-Jun and c-Fos (Tomar and Schlaepfer, 2010). Both, c-Jun and c-Fos then transactivate target genes involved in cell proliferation and migration (Figure 1.3F) (Ye and Field, 2012). In this way, ligand-induced activation of EGFR triggers multiple cytoplasmic signaling cascades that are essential for mammalian growth and development (Sibilia *et al.*, 2007; Zhang *et al.*, 2013).

1.1.2 Non-receptor tyrosine kinases

Non-receptor tyrosine kinases, as the name suggests, are not membrane-bound and are present in the cytoplasm of cells (Neet and Hunter, 1996). There are 32 different non-receptor tyrosine kinases encoded by the mammalian genome and are categorized into 10 sub-families (Manning *et al.*, 2002; Neet and Hunter, 1996). These include the ABL, ACK, CSK, FAK, FES, SRC, BRK, JAK, SYK and TEC families (Neet and Hunter, 1996). Non-receptor tyrosine kinases are composed of a kinase domain that imparts catalytic activity and other specific binding domains that mediate interaction with various binding partners, substrates and other molecules such as DNA (Neet and Hunter, 1996).

1.1.2.1 Src family kinases

The Src family kinases (SFKs) represents one of the most well-characterized non-receptor tyrosine kinase families encoded by the mammalian genome (Amata *et al.*, 2014; Parsons and Parsons, 2004). The family comprises nine different non-receptor tyrosine kinases, namely c-Src, Yes, Fyn, Fgr, Yrk, Lyn, Blk, Hck, and Lck (Amata *et al.*, 2014; Parsons and Parsons, 2004). v-Src is an oncogenic orthologue of c-Src and one of the first constitutively active tyrosine kinases to be discovered with oncogenic properties (Hunter, 2015; Hunter and Sefton, 1980). While v-Src was first identified from Rous Sarcoma virus, c-Src was discovered in mammalian cells and reported as a protooncogene (Hunter, 2015; Hunter and Sefton, 1980; Oppermann *et al.*, 1979). c-

Src maps to chromosomal loci 1p34-p36 and 20q12-q13, both of which are pre-disposed to genetic re-arrangements in multiple malignancies (Le Beau *et al.*, 1984; Parker *et al.*, 1985).

SFKs possess an N-terminal myristoylation or palmitoylation signal that allows these kinases to anchor to the plasma membrane (Koegl *et al.*, 1994; Patwardhan and Resh, 2010). They also possess three evolutionarily conserved functional domains, namely, the SH3 (Src-homology 3), SH2 (Src-homology 2) and kinase domains (Figure 1.4) (Eck *et al.*, 1994; Parsons and Parsons, 2004). The SH3 and SH2 domains are intermolecular binding domains where the SH3 domains bind to polyproline motifs whereas the SH2 domains bind to phosphotyrosine-containing motifs, on cellular proteins (Gan and Roux, 2009; Panni *et al.*, 2002; Pawson and Gish, 1992; Pawson and Schlessingert, 1993). The kinase domain imparts catalytic activity to the enzyme and is involved

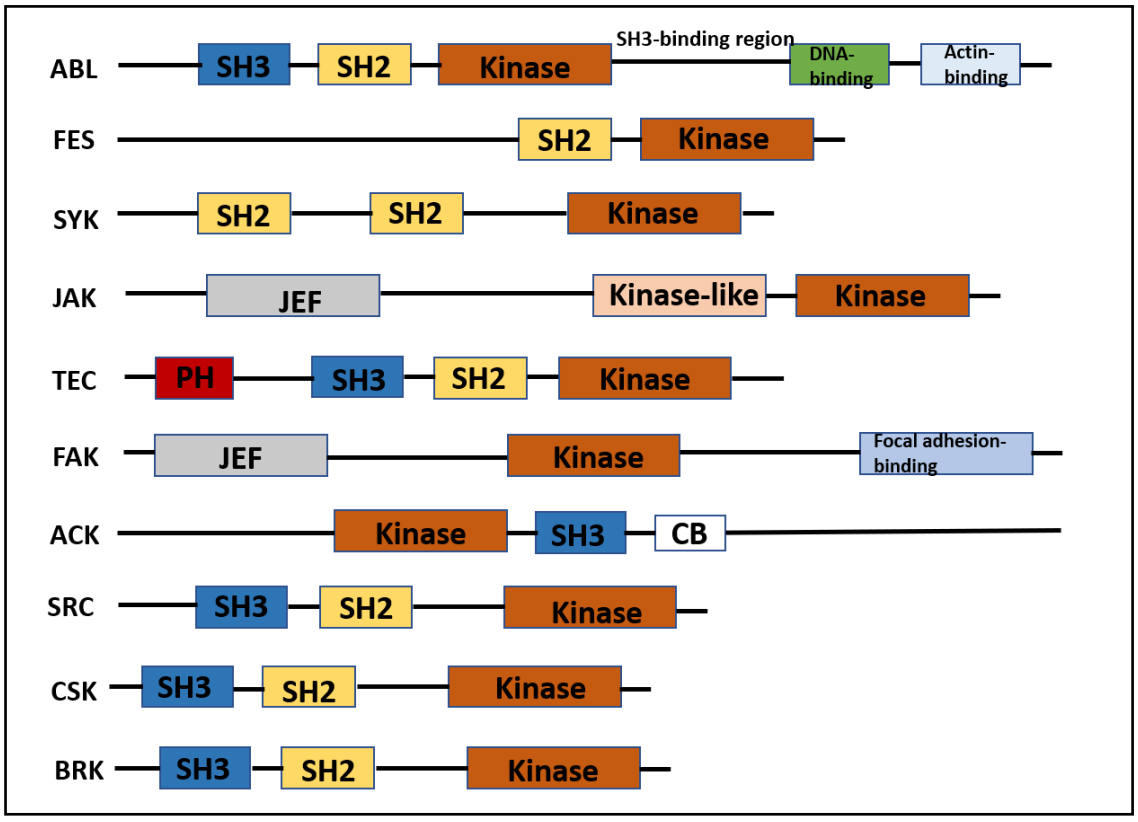


Figure 1.4: The mammalian non-receptor tyrosine kinase complement: Shown here is the general domain-structure organization of members belonging to the ten non-receptor tyrosine kinase families. The prototypical member of each non-receptor tyrosine kinase family possesses a kinase domain (highlighted in red) and one or several intermolecular-binding domains which include the SH3, SH2, DNA-binding, actin-binding, JEF (JAK, ERM, FAK), PH (Pleckstrin homology), Focal adhesion-binding, kinase-like and CB (Cdc42-binding) domains, as annotated in the figure.

in autophosphorylation as well as the phosphorylation of target proteins (Hubbard *et al.*, 1998; Miller, 2003). The SH3 domains in Src family kinases are known to bind to sequences displaying the X-P-X-X-P consensus motif while the SH2 domains generally exhibit strong specificity towards the P-Y-E-E-I motif (where Y represents the phosphorylated tyrosine residue) (Feng *et al.*, 1994; Lim *et al.*, 1994; Songyang *et al.*, 1993; Yu *et al.*, 1994). Further, high-throughput kinase assays using libraries of degenerate peptides as candidate substrates have shown that the kinase domains of Src family kinases preferably phosphorylate sequences comprising either an isoleucine, leucine or valine at the -1 position and a glutamic acid or glycine at the +1 position, with respect to the central phosphotyrosine residue (Deng *et al.*, 2014; Shah *et al.*, 2018; Songyang *et al.*, 1995). Both, motif specificities of the kinase domain as well as the SH3 and/or SH2 domains are involved in substrate-selection by members of the Src family kinases (Miller, 2003).

Enzymatic activation in Src family kinases is auto-regulated by intramolecular associations involving the SH3, SH2 and kinase domains (Cooper *et al.*, 1986; Liu *et al.*, 1993; Superti-Furga, 1995; Xu *et al.*, 1999). Key to this mode of enzymatic autoregulation is the phosphorylation of a highly conserved residue, Y530, in the C-terminal tail of Src- family kinases (Figure 1.5) (Cooper *et al.*, 1986; Liu *et al.*, 1993; Superti-Furga, 1995; Xu *et al.*, 1999). This C-terminal residue is a target for phosphorylation by c-Src kinase (CSK) (Nada *et al.*, 1991). Phosphorylation of this site by CSK triggers intramolecular rearrangements ultimately leading to the catalytical inhibition of SFKs (Figure 1.5) (Nada *et al.*, 1991). Extensive structural and cell-based studies have shown that phosphorylation of the conserved Y527 amino acid in chicken c-Src (Y530 in human c-Src) by CSK results in the intramolecular binding of Y530 with the c-Src SH2 domain, thereby locking the enzyme in an inactive conformation (Figure 1.5) (Cooper *et al.*, 1986; Liu *et al.*, 1993). This is followed by additional intramolecular rearrangements where the SH3 domain binds to the SH2-kinase linker region to stabilize the enzymatically inactive conformation (Figure 1.5) (Cooper *et al.*, 1986; Liu *et al.*, 1993; Superti-Furga, 1995; Xu *et al.*, 1999). Accordingly, studies have shown that deletion of either the c-Src SH2 or SH3 domain or introducing a point-mutation at Y527 (Y527F), disrupts the auto-inhibited conformation of the enzyme and results in enhanced catalytic activity of the kinase (Okada *et al.*, 1993). In the context of wild type c-Src however, the phosphatase-dependent dephosphorylation of the C-terminal tyrosine residue and/or availability of high-affinity SH3/SH2 interactors relieve the intramolecular binding constraints and promotes an active conformation of the kinase (Figure 1.5) (Roskoski, 2005; Xu *et al.*, 1999). A number of

phosphatases have been reported to target the C-terminal tyrosine residue in c-Src and Src family members leading to the enzymatic activation of these kinases. These phosphatases include PTP1B, SHP1, CD45, RPTPα, RPTPε, and LAR (Huntington and Tarlinton, 2004; Maksumova *et al.*, 2005; Ponniah *et al.*, 1999; Roskoski, 2005; Somani *et al.*, 1997; Su *et al.*, 1999; Vacaresse *et al.*, 2008; Yang *et al.*, 2006; Zheng *et al.*, 2000).

The active conformation of the Src family kinases is characterized by the autophosphorylation of a highly conserved activation loop-tyrosine residue (Y419 in c-Src) (Figure 1.5) (Xu *et al.*, 1999). This tyrosine residue lies in the catalytic cleft between the smaller N-terminal and larger C-terminal lobes of the kinase domain in c-Src. In the inactive state resulting

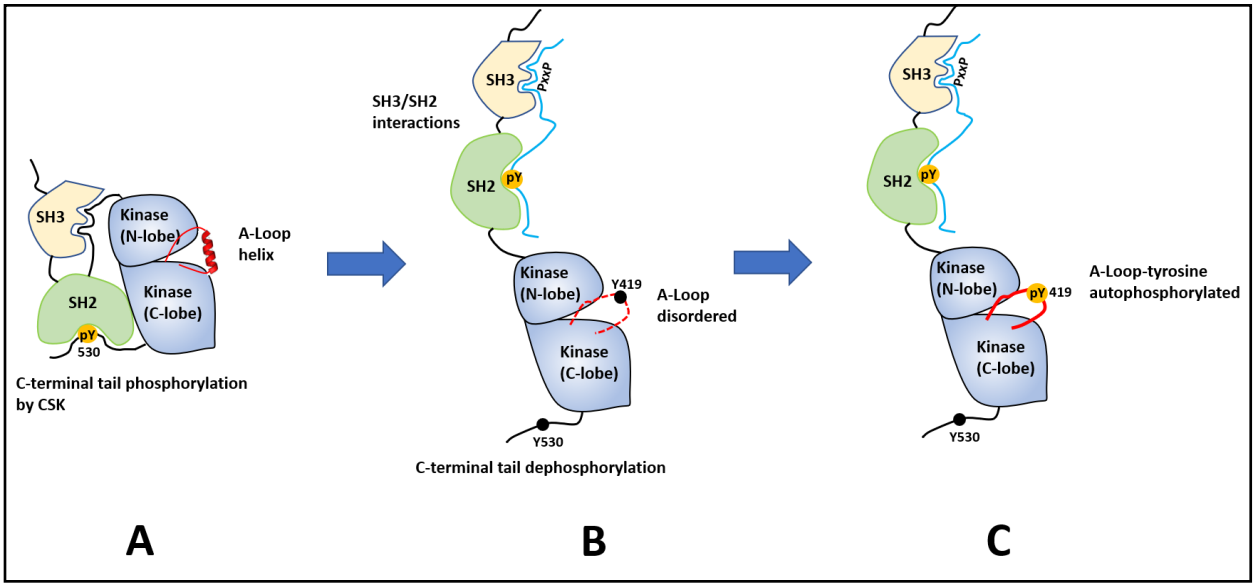


Figure 1.5: Regulation of c-Src kinase activity. Shown here is a schematic representation of the conformational changes involved in the activation of c-Src. **A.** The autoinhibited conformation is characterized by the phosphorylation of the C-terminal Y530 by CSK, leading to the intramolecular binding of phosphorylated Y530 (pY530) to the SH2 domain. Furthermore, the SH3 domain also binds to the SH2-kinase linker region thereby packing the SH3 and SH2 domains against the kinase domain lobes. This results in the activation loop adopting a partial helical structure which occludes access of the kinase domain to ATP for autophosphorylation of the activation loop tyrosine residue. **B.** The de-phosphorylation of the C-terminal Y530 together with the binding of the SH3 and SH2 domains with their cellular proteins targets, comprising poly-proline and phosphotyrosine motifs, respectively, results in an open-conformation of the enzyme. In this conformation, the activation loop helix becomes disordered which promotes access of the kinase domain to ATP. **C.** Access to ATP enables the autophosphorylation of the activation loop tyrosine residue, Y419 which results in enzymatic activation. Figure adapted from (Xu *et al.*, 1999).

from the intramolecular interactions of the SH2 domain with Y530, the kinase domain is oriented in an inflexible conformation that permits the activation loop to form a partial alpha helix. The partial helical structure of the activation loop obstructs access of the catalytic site to ATP, thereby preventing autophosphorylation (Figure 1.5). However, following Y530 dephosphorylation and/or binding of the SH3 and SH2 domains to target proteins, the activation loop becomes disordered which results in increased access of the catalytic cleft to the ATP substrate. This leads to the autophosphorylation and enzymatic activation of c-Src (Figure 1.5) (Xu *et al.*, 1999). By interacting with and phosphorylating target cellular proteins, c-Src and other members of the Src family are known to be involved in regulating key cellular processes such cell survival, proliferation, angiogenesis, migration and apoptosis (Erpel and Courtneidge, 1995; Fresno Vara *et al.*, 2001; Han *et al.*, 2014; Park *et al.*, 2007; Saigal and Johnson, 2008).

1.1.2.2 BRK family kinases

BRK or PTK6 (Breast tumor kinase/Protein tyrosine kinase 6), FRK/ PTK5 (Fyn-related kinase/ Protein tyrosine kinase 5) and SRMS/ PTK70 (Src-related kinase lacking N-terminal myristoylation sites and C-terminal regulatory tyrosine residue/ Protein tyrosine kinase 70) represent an evolutionarily conserved lineage of the Src-family kinases (Goel and Lukong, 2015; Serfas and Tyner, 2003). The genes encoding BRK and SRMS lie adjacent on the same chromosomal locus, 20q.13.33 whereas the FRK gene is distantly located on chromosomal loci 6q21-q22.3 (Goel and Lukong, 2015, 2016). At the genomic level, BRK, FRK and SRMS possess 8 exons whereas Src family kinases possess 12 exons (Figure 1.6) (Goel and Lukong, 2015; Serfas and Tyner, 2003). Given the distinct pattern of intron-exon splicing, BRK, FRK and SRMS were considered a distinct family of non-receptor tyrosine kinases, separate from the Src family (Goel and Lukong, 2015; Serfas and Tyner, 2003). Like c-Src and other members of the Src family kinases, BRK, FRK and SRMS possess all three functional domains, namely, the SH3, SH2 and kinase domains (Figure 1.7) (Goel and Lukong, 2015, 2016). Further, the mechanism of enzymatic autoregulation in BRK and FRK are conserved with the Src family. Specifically, the phosphorylation of the conserved C-terminal tyrosine residue in BRK (Y447) and FRK (Y497) leads to the enzymatic inactivation of these kinases whereas autophosphorylation of the activation

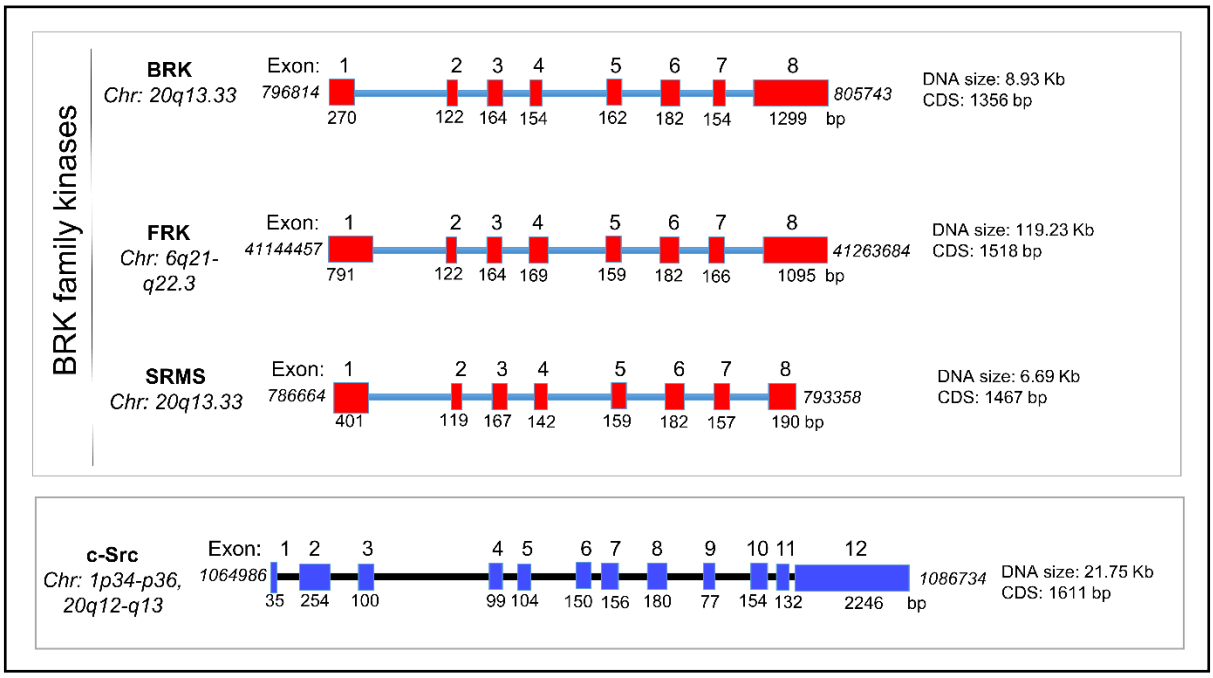


Figure 1.6: **Intron-Exon splicing pattern of the BRK family kinases and c-Src**. Shown here is a schematic representation of the intron-exon architecture of the BRK family (BRK, FRK and SRMS) and the prototypical member of the Src family kinases, c-Src. The chromosomal loci and corresponding genetic location on the chromosome is annotated. The exons (highlighted in red for BRK family kinases and blue for Src) are numbered and the length of each exon indicated. Also indicated is the length of the DNA and the corresponding coding sequence (CDS). Figure obtained from (Goel and Lukong, 2015), with the permission and rights to reuse/reproduce.

loop tyrosine residue (Y342 in BRK and Y387 in FRK) results in enzymatic activation (Figure 1.7) (Goel and Lukong, 2015, 2016). SRMS, however, lacks a C-terminal tail altogether, and the mode of enzymatic autoregulation was instead shown to be dependent on the 50-amino acid N-terminal region (Goel *et al.*, 2013). Further, the conserved activation loop autophosphorylation site, Y380 has been shown to be essential for SRMS enzymatic activity (Figure 1.7) (Goel *et al.*, 2013).

Both BRK and FRK represent well characterized members of the BRK family in the context of the signaling intermediaries regulated, and cellular role(s) played by these kinases. For instance, a number of studies have shown that BRK plays a growth promoting role whereas FRK plays a tumor suppressor role, in breast cancers (Bagu *et al.*, 2017; Brauer and Tyner, 2009, 2010; Goel and Lukong, 2015, 2016; Ogunbolude *et al.*, 2017). Both kinases have been shown to bind to and phosphorylate various cellular proteins to regulate these cellular functions (Brauer and Tyner, 2009, 2010; Goel and Lukong, 2015, 2016). SRMS, on the other hand, is an understudied

kinase of this family. The subsequent sections of this thesis shall review background information on SRMS.

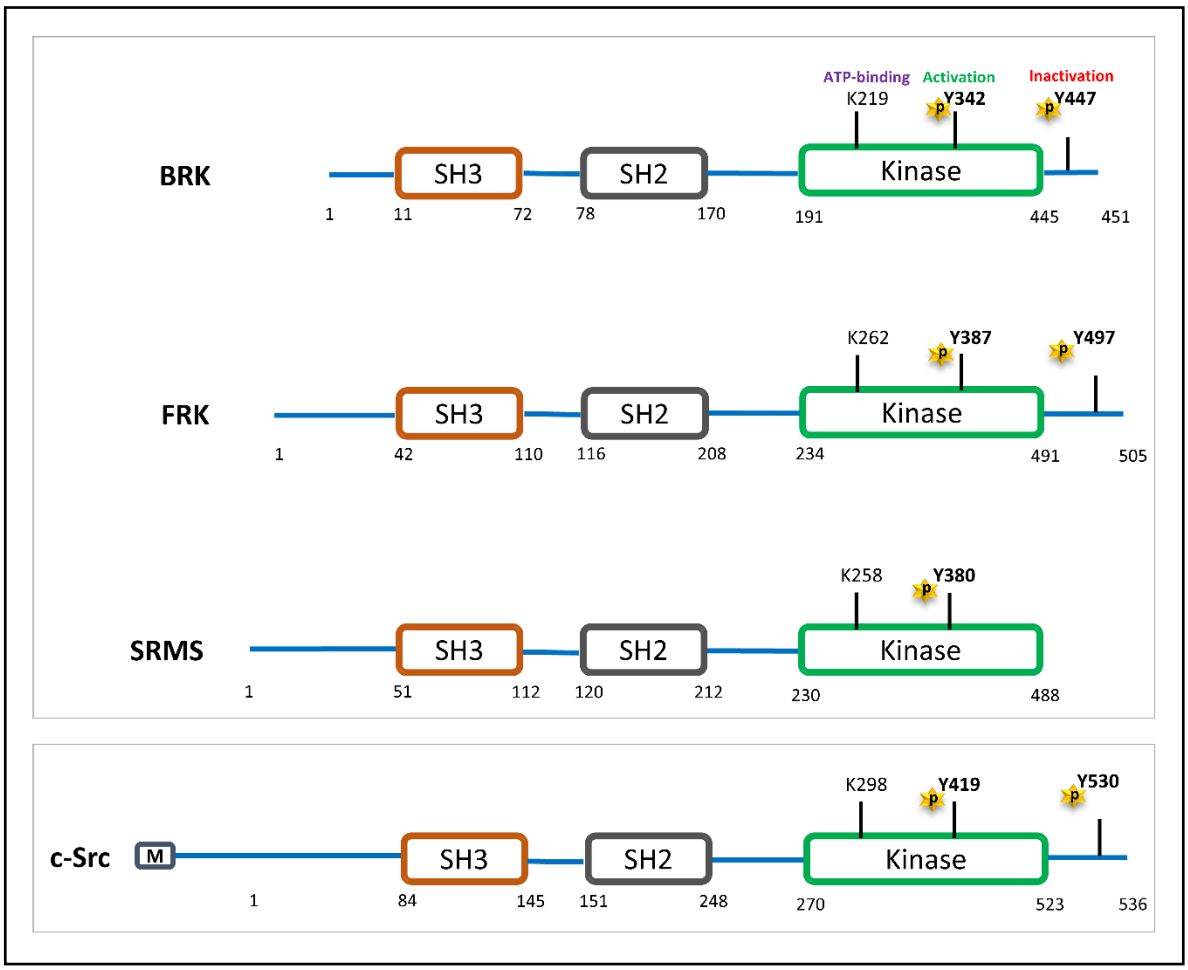


Figure 1.7: **Protein domain-structure organization of BRK family kinases and c-Src.** Schematic representation of the BRK family kinases (BRK, FRK and SRMS) and c-Src, showing the different functional protein domains present on each protein. Like c-Src, members of the BRK family possess all three functional domains, namely, the SH3, SH2 and the kinase domain. The amino acid positions of the various domains are indicated. Also shown are key conserved amino acid residues which are implicated in the the regulation of the enzymatic activity of these proteins. These include the ATP-contacting lysine (K219 in BRK, K262 in FRK and K258 in SRMS), activation-loop autophosphorylation tyrosine residue (Y342 in BRK, Y387 in FRK and Y380 in SRMS) required for enzymatic ativation and a C-terminal tyrosine residue present only in BRK (Y447), FRK (497) and c-Src (Y530) that negatively regulates enzymatic activity.

1.1.2.2.1 SRMS

SRMS (pronounced "*shrims*") is a non-receptor tyrosine kinase and a member of the BRK family kinases (BFKs) (Goel and Lukong, 2015). The gene encoding SRMS was first discovered by Kohmura et al. in 1994 in a screen aimed at identifying novel genes that may potentially be involved in neural precursor cell (NPC) differentiation (Kohmura *et al.*, 1994). The gene was named SRMS as an abbreviation to its descriptive nomenclature which essentially denotes the characteristic sequence-specific differences observed between SRMS and Src family kinases (Kohmura *et al.*, 1994). Thus, the abbreviation expands to imply that SRMS is a "Src-Related kinase", "lacking a C-terminal regulatory residue and N-terminal Myristoylation sites", both of which are hallmark biochemical features of the SFKs. The study also reported that the murine SRMS gene is located 2.1 cM distal to chromosomal locus D2Mit25 whereas the human *SRMS* orthologue is positioned adjacent to the BRK gene on the chromosomal locus 20q13.33 (Kohmura *et al.*, 1994). Kawachi et al. in 1996 reported on cloning the same gene from murine neonatal skin tissue. The gene was named PTK70 (Protein tyrosine kinase 70) (Kawachi *et al.*, 1997). Given the lack of a myristoylation signal and a conserved C-terminal tyrosine residue, Kawachi et al. proposed that SRMS is likely a part of another non-receptor tyrosine kinase family distinct from SFKs (Kawachi *et al.*, 1997). Later reports substantiated the classification of SRMS, BRK and FRK as a separate kinase family owing to the conserved intron-exon splicing pattern associated with these kinases (Goel and Lukong, 2015; Serfas and Tyner, 2003). Therefore, SRMS, BRK and FRK were considered a separate family of non-receptor tyrosine kinases and thereafter named the BRK family kinases (Goel and Lukong, 2015; Serfas and Tyner, 2003).

Since the discovery of SRMS in 1994, a few studies have investigated the biochemical enzymatic regulation and the potential cellular role of SRMS. The subsequent sections shall review information from the literature relevant to SRMS biology.

1.1.2.2.1.1 Biochemical regulation of SRMS enzymatic activity

The SRMS cDNA encodes a 54 kDa protein spanning 488 amino acids that harbors a SH3, SH2 and a kinase domain; a structural framework common to BRK, FRK and SFKs (Goel *et al.*, 2013) (Figure 1.6 and 1.7). As described above, the enzymatic activity of SFKs, BRK and FRK is primarily regulated by phosphorylation of a conserved C-terminal tyrosine residue (Goel and Lukong, 2015, 2016; Goel *et al.*, 2013; Li *et al.*, 2007; Yeatman, 2004). However, SRMS

lacks a C-terminal tyrosine residue implying that the mechanism of enzymatic regulation in SRMS is biochemically distinct (Goel *et al.*, 2013) (Figue 1.8). The highly conserved activation loop-tyrosine residue represents the primary site of autophosphorylation in BRK, FRK and members of the Src family kinases and is required for enzymatic activation (Goel and Lukong, 2015, 2016; Goel *et al.*, 2013). This residue corresponds to Y380 in SRMS and site-directed mutagenesis studies have confirmed this to be the primary autophosphorylation site in SRMS (Goel *et al.*, 2013) (Figure 1.8). The phosphatases PTPN13 and PTEN have been shown to inhibit the catalytic activity of c-Src and BRK by directly dephosphorylating the activating autophosphorylation sites on these kinases which correspond to Y419 (c-Src) and Y342 (BRK) (Glondu-Lassis *et al.*, 2010; Wozniak *et al.*, 2017). No phosphatase has been reported to dephosphorylate the conserved activating Y380 residue in SRMS. Therefore the role of phosphatases in regulating SRMS enzymatic activity is unknown.

Previous studies by our group characterized the enzymatic activity of SRMS and discovered that the presence of the 50 amino acid-long N-terminal region, preceding the SH3 domain, is essential for SRMS tyrosine kinase activity (Goel *et al.*, 2013) (Figure 1.8). Deletion of this N-terminal region was found to abolish SRMS cellular tyrosine kinase activity. Additionally, an SH2-deletion mutant of SRMS showed lower tyrosine kinase activity compared

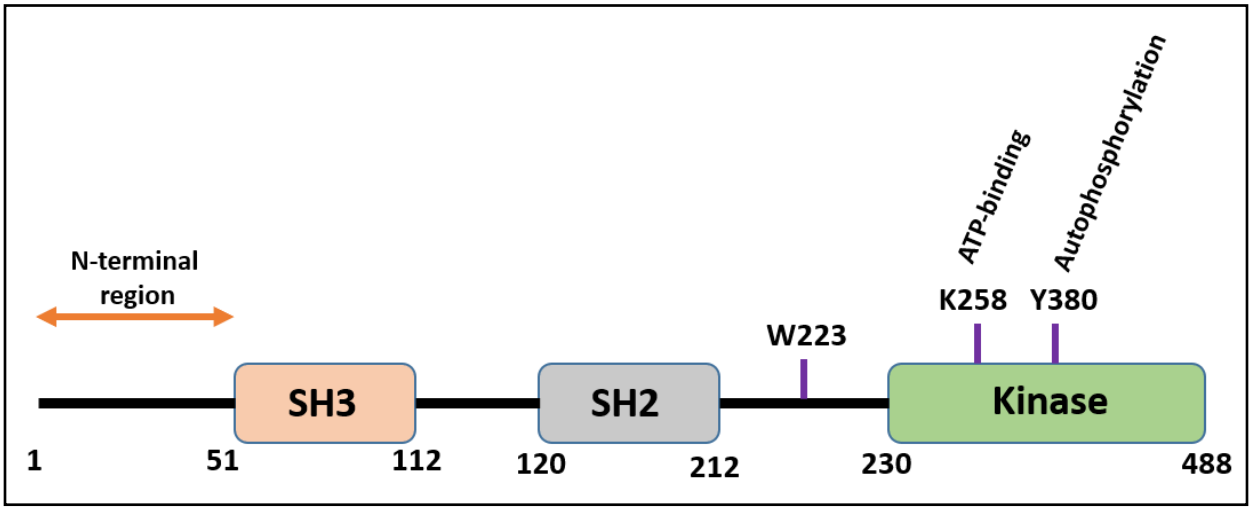


Figure 1.8: **Protein domain-structure organization of SRMS.** Schematic representation of the protein domain-structure of SRMS, showing the SH3, SH2 and kinase domains. The unique 50 amino acid-long N-terminal region essential for SRMS enzymatic activity is highlighted (Goel *et al.*, 2013). Also shown is the W223 residue in the SH2-kinase linker region and the ATP-binding K258 and the activation loop-autophosphorylation residue Y380.

to wild type SRMS, indicating that the SH2 domain is potentially also involved in the autoregulation of SRMS catalytic activity. The SRMS K258 residue forms part of a highly conserved ATP-binding pocket in tyrosine kinases (Figure 1.8). Indeed substituting K258 with a methionine residue was found to abolish SRMS catalytic activity. Furthermore, a SRMS W223A mutant was also generated to study the role of this residue on SRMS catalytic activity. The SRMS W223 residue is conserved in BRK (W184) as well as Src family kinases and lies in the SH2-kinase linker region (Goel *et al.*, 2013) (Figure 1.8). Structural studies on the Src family kinase-member Hck have shown that the analogous residue, W260 stabilizes a catalytically inactive conformation of the enzyme (LaFevre-Bernt *et al.*, 1998). Specifically, crystal structures of Hck revealed that the interaction of the SH3 domain with the SH2-kinase linker region allows the W260 residue to interact with the αC helix in the N-terminal lobe of the kinase domain which stabilizes the inactive conformation. Accordingly, the Hck W260A mutant showed increased tyrosine kinase activity compared to wild type Hck demonstrating the significance of this residue in the autoregulation of Hck enzymatic activity (LaFevre-Bernt *et al.*, 1998). A previous study on BRK showed that substituting the conserved W184 residue with alanine resulted in significantly reduced kinase activity compared with wild type BRK (Kim and Lee, 2005; Qiu and Miller, 2004). This indicated that unlike the Src family kinase Hck, this conserved residue in BRK likely serves to stabilize the active conformation of the kinase (Kim and Lee, 2005; Qiu and Miller, 2004). Interestingly, substituting the conserved W223 residue with alanine in SRMS was also found to result in significantly reduced tyrosine kinase activity compared to wild type SRMS (Goel *et al.*, 2013). This suggested that the tryptophan residue plays a similiar biochemical role in maintaining an active conformation in both BRK and SRMS (Goel *et al.*, 2013; Qiu and Miller, 2004). Analyses of the crystal structures of wild type SRMS and its inactivating mutants will further help in characterizing the underlying biochemical mechanisms involved in the autoregulation of SRMS enzymatic activity.

1.1.2.2.1.2 Protein expression in mammalian cells/tissues

To date, very few studies have examined the expression of SRMS protein/mRNA in mammalian cells and tissues. Kohmura *et al.*, credited with the discovery of the SRMS cDNA from mouse neural precursor cells (NPCs), identified two variants of the SRMS mRNA via Northern blotting analyses, bearing molecular sizes of 2 kb and 2.6 kb (Kohmura *et al.*, 1994).

Though it is not clear whether these variants represent products of alternative splicing events, the expression of these variants were found to differ temporally and spatially in certain murine tissues examined (Kohmura *et al.*, 1994). For instance, while the 2.6 kb variant was found to be abundantly expressed in tissues derived from the murine lung, liver, spleen, ovary, kidney and intestines, the expression of the 2 kb variant was limited to fewer organs including the testis and cerebral tissues. Lower expression levels of the 2.6 kb variant were also seen in murine heart, cerebrum, cerebellum and thymus tissues. This indicates that SRMS expression is likely spatially altered across different murine organs where the cellular roles of the kinase are yet to be examined. Additionally, whereas the expression of the 2.6 kb mRNA variant was detected in the embryonic day 15 (E_{15}) brain tissues, it was found to be relatively lower in the postnatal brain. However, the 2 kb mRNA variant was more abundant in the postnatal brain. Interestingly, both variants were found to be low in the E_{10} brain-derived NPCs, E_{11} brain and adult mouse brain tissues (Kohmura *et al.*, 1994). This therefore indicates that SRMS expression is not only spatially regulated but potentially also temporally regulated during mammalian development.

A study by Kawachi *et al*. also identified two mRNA variants in mouse tissues which was speculated to result from alternate splicing events (Kawachi *et al.*, 1997). As with the observations of Kohmura *et al*., Kawachi and colleagues noted organ-specific differences in SRMS mRNA expression, though these observations were not entirely identical to those noted by the Kohmura group (Kohmura *et al.*, 1994) (Kawachi *et al.*, 1997). Specifically, Kawachi *et al*. determined that the SRMS mRNA expression was more abundant in murine liver, lung, thymus and skin compared to brain, kidney, heart and spleen (Kawachi *et al.*, 1997). Furthermore, quantitative analyses of SRMS mRNA in selected skin-derived cell lines revealed a relatively higher level of SRMS mRNA expression in the Sq-1974 cell line compared to other cell lines tested such as the B16 melanoma 4A5, NFSaY83 and L cell lines. As well, SRMS mRNA levels were found to be significantly higher in the primary epidermal cells compared to the primary dermal fibroblast cells (Kawachi *et al.*, 1997).

In another study, Anam *et al*. quantitatively assessed the mRNA expression levels of various receptor tyrosine kinases, non-receptor tyrosine kinases and other key signaling proteins in murine bone marrow cell (BMC)-derived hematopoietic stem/progenitor cells (HPSCs) as well as bone marrow mesenchymal stromal cells (BMMSCs) (Anam and Davis, 2013). It was reported that certain members of SFKs such as Lck, Lyn, Hck and Fyn were found to be differentially

expressed in the HPSCs and BMMCs. However, the expression of SRMS and another tyrosine kinase, Tnk1 was not detected in either cell type (Anam and Davis, 2013). The underlying reasons behind the absence of SRMS expression were not reported and may potentially indicate a functionally redundant role of SRMS in the differentiation of bone-marrow stem cells.

The expression pattern of SRMS in normal or cancer-derived human cells/tissues is not well-characterized. A previous study by our group reported the expression of SRMS in breast tissue microarrays representing breast carcinoma samples and the adjacent normal mammary tissue (Goel *et al.*, 2013). The array used comprised a small cohort of 24 tumor grade-specific invasive ductal carcinoma samples and matched adjacent normal samples, derived from 6 breast cancer patients. Immunohistochemical analyses of the array indicated that SRMS expression was generally higher in the breast carcinoma tissues compared to matched adjacent normal tissues. Furthermore, SRMS expression was found to potentially correlate with tumor grade, with higher grade tumor tissue samples showing increased expression of SRMS compared with low-grade carcinoma tissues. A similar observation was also made in immunoblotting studies using a panel of eight immortalized breast cancer cell lines. The results showed that SRMS expression was relatively higher in the breast cancer cell lines compared to the immortalized normal mammary epithelial cell line 184B5 (Goel *et al.*, 2013). Breast cancer is a heterogeneous disease with different clinical, histopathological, and molecular characteristics (Perou *et al.*, 2000; Sorlie *et al.*, 2001). Therefore, an evaluation of SRMS mRNA and protein expression in a broader panel covering the various breast cancer pathological and molecular subtypes would be necessary to understand if SRMS expression can be statistically correlated between normal-tumor pairs.

1.1.2.2.1.3 Subcellular localization

Unlike SFKs, SRMS lacks a myristoylation/palmitoylation signal known to dictate plasma-membrane anchoring of SFKs or a nuclear localization signal such as that embedded within the FRK SH2 domain (Cance *et al.*, 1994; Goel and Lukong, 2016; Goel *et al.*, 2013; Koegl *et al.*, 1994; Patwardhan and Resh, 2010). However, both the endogenous and ectopically expressed SRMS localize predominantly to punctate cytoplasmic structures, as studied in HEK293, HeLa and certain breast cancer cell lines such as MDA-MB 231, AU565 and SKBR3 (Goel *et al.*, 2013). This subcellular localization pattern has not been reported for SFKs, BRK or FRK. SFKs and BRK are known to exhibit varying subcellular localization patterns depending on the activation

status of the enzymes. For instance, endogenous and ectopically expressed active c-Src has been shown to localize at the plasma membrane as well as the perinuclear region of Cos-1 cells (Donepudi and Resh, 2008). In the perinuclear region, active c-Src was found to co-traffic with EGFR in endosomes. On the other hand, kinase-dead Src was found to localize only to the inner plasma membranes but not the endosomes, in Cos-1 cells (Donepudi and Resh, 2008). Furthermore, in mouse embryonic fibroblasts active c-Src has been shown to localize to focal adhesion sites on the plasma membrane (Tu *et al.*, 2010). Consistently, constitutively active v-Src has also been shown to localize predominantly in focal adhesions in rat fibroblast cells (Hayes and Moss, 2009). In patient-derived breast cancer tissues, active c-Src has been reported to predominantly localize to the plasma membrane compared to inactive c-Src, which was largely found in the cytoplasm (Anbalagan *et al.*, 2012).

In the case of BRK, previous studies have shown that ectopically expressed BRK displays a diffuse nucleo-cytoplasmic localization in HeLa and HEK293 cells (Lukong *et al.*, 2005; Lukong and Richard, 2008). Endogenous BRK, on the other hand, has been reported to colocalize with Sam68 in Sam68-like nuclear bodies (SNBs) as well as AKT in the plasma membrane in MCF7 cells (Derry *et al.*, 2000; Ostrander *et al.*, 2010). Furthermore, like c-Src, active BRK has been found to be predominantly localized to the plasma membrane of immortalized prostate cancer cells, breast cancer cells and patient-derived breast carcinoma tissues (Derry *et al.*, 2003; Ostrander *et al.*, 2010; Peng *et al.*, 2014). Therefore, the subcellular localization of Src and BRK seems to be regulated in a kinase-dependent manner. Similarly, the catalytically deficient mutants of SRMS, namely the N-terminal-deletion or the SRMS K258M (kinase-dead) mutants, have been reported to display altered subcellular localization compared to wild type SRMS (Goel *et al.*, 2013). Specifically, these kinase-inactive mutants show a partly diffused cytoplasmic localization as opposed to the predominant punctate pattern displayed by wild type SRMS (Goel *et al.*, 2013). These SRMS cytoplasmic punctae are uncharacterized and it is not known whether these punctate structures are formed by the colocalization of SRMS with specific cytoplasmic proteins or those in cellular organelle(s).

1.1.2.2.1.4 Substrates and binding partners

A number of cellular bindings partners and substrates of SFKs, BRK and FRK have been idenified and characterized to date. However, little is known about the cellular targets of

SRMS. Presently, Dok1 and BRK remain the only two substrates validated for SRMS (Fan *et al.*, 2015; Goel *et al.*, 2013).

Takeda et al. described an *in vitro* proteomics study where several potential substrates of Src and BRK family kinases were reported (Takeda *et al.*, 2010). The proteomics study involved screening of a panel of 519 unique proteins against SFKs and BFKs in a high-throughput *in vitro* kinase assay. CRK (p38) was identified as a common phosphorylated target of the kinases and used to rank the substrate-specificity of the other kinases. A fair number of potential substrates were identified for each kinase by the study (Table 1.1). Dok1 (Downstream of tyrosine kinase 1) was one of the potential SRMS substrates identified by this study (Takeda *et al.*, 2010). Dok1 was subsequently validated and characterized as the first bonafide SRMS substrate (Goel *et al.*, 2013). Notably, Dok1 has previously been characterized as a cellular substrate of BRK, Src and Abl kinases (Janas and Van Aelst, 2011; Miah *et al.*, 2014; Ng *et al.*, 2015; Woodring *et al.*, 2004).

A recent study identified BRK as a substrate of SRMS (Fan *et al.*, 2015). SRMS was shown to phosphorylate the C-terminal Y447 residue in BRK via immunoprecipitation assays as well as direct *in vitro* kinase assays. Further, SRMS was found to phosphorylate only BRK Y447 but not the conserved Y530 in the c-Src C-terminal tail (Fan *et al.*, 2015). Phosphorylation of BRK Y447 plays an important role in the regulation of BRK kinase activity (Goel and Lukong, 2015; Miah *et al.*, 2012; Ostrander *et al.*, 2010). However, whether phosphorylation of BRK Y447 by SRMS leads to the inhibition of BRK activity was not reported by the study and warrants further investigation.

The motif-specificity of the SRMS SH2 domain has been reported previously which has shed light on the the type of potential SRMS SH2 domain-dependent interactions (Liu *et al.*, 2015; Zhao *et al.*, 2013). SH2 domains comprise a highly conserved phosphotyrosine-binding pocket which, in Src family kinases, is formed by the amino acid sequence F-L-V-R-E-S (Eck, 1995; Ottinger *et al.*, 1998; Wagner *et al.*, 2013). Mutant SH2 domains carrying point-mutations in this pocket have been shown to exhibit significantly reduced phosphotyrosine-binding ability (Campbell and Jackson, 2003; Mayer *et al.*, 1992). This phosphotyrosine-binding pocket in SRMS corresponds to the sequence, F-L-I-R-P-S. Whereas the role of these amino acids in the SRMS SH2 domain have not been well-characterized in the context of phosphotyrosine-binding, high-throughput studies have nonetheless identified SRMS SH2 domain consensus motifs (Liu *et al.*, 2015; Zhao *et al.*, 2013). Zhao et al. reported the identification of the SRMS SH2-domain

consensus motifs using a high-throughput peptide library-screening approach (Zhao *et al.*, 2013). The study used a library of phosphotyrosine peptides randomized for amino acid residues flanking the central phosphotyrosine residue and screened binding-affinities to various recombinant GST-tagged SH2 domains. The study found that the SRMS SH2 domain selectively bound phosphotyrosine peptides displaying hydrophobic residues at the P+3 position (where P is the phosphotyrosine). Specifically, it was noted that the SRMS SH2 domain favours motifs possessing a phenylalanine, methionine or a cysteine residue at the P+3 position. Whereas at the P+1 position

Table 1.1: List of potential SRMS substrates identified by Takeda et al.. Shown in this table are the potential SRMS substrates identified by Takeda et al., 2010 via high-throughput *in vitro* kinase assays. The potential substrates were ranked based on substrate-specificity towards the commonly phosphorylated substrate, CRK.

S.No	Genbank Accession no.	Protein name	Category	Relative activity (CRK =1)	Rank
1.	AK303912	Tom1 like 1	Adapter protein	2.05	1
2.	AK055944	Dok1	Adapter protein	1.68	2
3.	AK315002	Tom1 like 1	Adapter protein	1.52	3
4.	BC013200	Lck	Tyrosine Kinase	1.22	4
5.	AK314619	CRK-associated substrate	Adapter protein	1.10	5
6.	BC053532	CRK-associated substrate	Adapter protein	1.05	6
7.	BC001718	CRK	Adapter protein	1.00	7
8.	BC065912	ARG	Tyrosine Kinase	0.84	8
9.	AK075020	SYK	Tyrosine Kinase	0.43	9
10.	AK095513	Disabled 1	Adapter protein	0.39	10
11.	AK295613	HEF-like protein	Adapter protein	0.26	11
12.	AK027148	NEK11	Ser/Thr kinase	0.26	12

the SRMS SH2 domain exhibited specificity towards a histidine or glutamic acid residue, the P-1 position favoured either a histidine or a tyrosine residue (Zhao *et al.*, 2013). Characterization of

the SRMS SH3/SH2 ligands may provide vital cues on SRMS substrate specificity and the regulation of associated signaling pathways.

A similar study by Liu *et al*. examined the *in vitro* binding specificities of 78 different SH2 domain-containing proteins towards tyrosine-phosphorylated peptides derived from various receptors proteins implicated in immune response signaling (Liu *et al.*, 2015). Specifically, compared to Src, BRK or FRK, the SRMS SH2 domain was found to exhibit generally minimal binding specificity towards either the LIRB4 pY360, DSCAM pY1708, PGRFB pY797, PGRFB pY904, CD3E pY199, SLAF pY304, STAM1 pY371, LAY1 pY71, LY9 pY626 or CD79A pY199 synthetic peptides (Liu *et al.*, 2015). This may imply that SRMS SH2-mediated intermolecular interactions and consequently its role as a signaling intermediate in the immune receptor network may be limited. Further, the results of the study also suggest that unlike other BFK and SFK members, the SRMS SH2 domain potentially interacts with a highly limited and specific cohort of tyrosine phosphorylated proteins, which may have implications on the SRMS SH2 domain-dependent cellular interactome and SRMS cellular functions (Liu *et al.*, 2015).

1.1.2.2.1.5 Cellular and physiological roles

Information on the function and cellular role of SRMS is sparse. The absence of characterized bonafide endogenous targets/subtrates of SRMS has limited our understanding of the function and role of the enzyme. The identification of BRK Y447 as a target phosphorylation site of SRMS indicates that SRMS may potentially function as a negative regulator of BRK catalytic activity *in vivo* (Fan *et al.*, 2015). This may highlight the role of SRMS in the potential regulation of kinase-dependent cellular and physiological functions of BRK (Fan *et al.*, 2015). In one of the early discoveries relevant to SRMS, Kohumura et al. reported varying expression levels of SRMS in murine brains across the different embryonic developmental stages (Kohmura *et al.*, 1994). Given the apparent temporal regulation of SRMS expression in the murine brain, it is possible that SRMS may potentially play a developmental role towards the differentiation of murine brain cells (Kohmura *et al.*, 1994). Similarly, due to the differential expression of SRMS mRNA in the murine epidermal and fibroblast cells, speculations of a possible role of SRMS in the proliferation and/or differentiation of murine keratinocytes were also reported (Kawachi *et al.*, 1997).

Potts *et al.* reported the involvement of SRMS as a potential negative regulator of autophagy in mammalian cells (Potts *et al.*, 2013). Autophagy is a cellular process in which proteins and organelles are degraded within eukaryotic cells (Kristensen *et al.*, 2008; Sica *et al.*, 2015). In this process, the cytoplasmic material is engulfed in autophagosomes which then fuse with lysosomes for degradation (Kristensen *et al.*, 2008; Sica *et al.*, 2015). During the induction of autophagy, the cytosolic form of the microtubule-associated protein 1A/1B-light chain 3 (LC3-I) is conjugated to phosphatidylethanolamine to yield LC3-phosphatidylethanolamine conjugate (LC3-II) (Hansen and Johansen, 2011; Tanida *et al.*, 2008). LC3-II is then recruited to the outer membrane of autophagosomes (Hansen and Johansen, 2011; Tanida *et al.*, 2008). The turnover of LC3-II, therefore, is often regarded as a marker of autophagic induction (Tanida *et al.*, 2008). In the study by Potts *et al.* the human colorectal cancer cell line HCT116 was used in high-throughput siRNA and miRNA screens to identify genes that regulate the expression of a group of six well-characterized reporter genes namely, ACSL5, BNIP3L, ALDOC, LOXL2, BNIP3, and NDRG1 (Potts *et al.*, 2013). These genes served as reporters or proxies for different cellular conditions including autophagy. The genes identified from the screens were clustered based on the level of regulation of the proxy reporter genes. SRMS was consequently identified in a cluster of genes with a well-documented role in the regulation of autophagy. It was further shown that knockdown of SRMS in the human osteosarcoma cell line U2OS, engineered to express GFP-LC3 as the autophagic reporter, resulted in an increase in the number of LC3-positive puncta. On the other hand, overexpression of SRMS was found to decrease the number of LC3-positive puncta. Additionally, modulating SRMS expression levels did not have implications on the mTOR pathway as determined by assessing p70S6K pT389 levels. This implied that the SRMS-dependent regulation of autophagic flux is not dependent on the mTOR pathway (Potts *et al.*, 2013). Another study showed that EGFR directly phosphorylated Beclin on Y229, Y233 and Y352 and increased Beclin1 association with Rubicon and Bcl2 but decreased Beclin1 association with the VPS34 kinase (Wei *et al.*, 2013). In this way EGFR was found to suppress nutrient depletion-induced autophagy. Interestingly, the same study reported that unlike EGFR, recombinant SRMS did not phosphorylate Beclin1 in *in vitro* kinase assays, indicating that the SRMS does not regulate autophagy by phosphorylating Beclin1 (Wei *et al.*, 2013).

In an effort to understand the physiological significance of SRMS, Kohumura et al. generated a homozygous SRMS-knock out ($SRMS^{-/-}$) mouse model (Kohmura *et al.*, 1994). These

transgenic mice were found to display no differences in overall phenotype compared to control mice ($SRMS^{+/+}$). The mice were viable, fertile and showed no apparent external or internal phenotypic deficits or abnormalities. The study therefore concluded that SRMS may be functionally redundant and the SRMS-associated cellular and physiological roles may likely be compensated by other tyrosine kinases (Kohmura *et al.*, 1994). It may be noted that a similar observation was made earlier in FRK knock-out mice where no obvious differences could be observed in the overall murine phenotype compared to control mice (Chandrasekharan *et al.*, 2002). However, characterization of the FRK knock-out mice showed that these mice had significantly lesser T3 thyroid hormones levels in the blood compared to control mice. Further, gene expression analyses revealed increased expression of c-Src and Yes kinases in the intestine and colon, respectively (Chandrasekharan *et al.*, 2002). Likewise, other studies demonstrated that FRK knock-out mice displayed a significant decrease in the number of pancreatic β-cells in the embryonic and post-natal stages (Akerblom *et al.*, 2007). It is therefore possible that the physiological role of SRMS is not entirely redundant and further *in vivo* investigations may be warranted in this regard.

1.2 Phosphoproteomics analyses by mass spectrometry

Phosphorylation represents an important post-translational modification (PTM) that is integral to the regulation of signal transduction in mammalian cells (Corwin *et al.*, 2017; Day *et al.*, 2016). Phosphoserine, phosphothreonine and phosphotyrosine represent the most studied phosphorylation events in the mammalian proteome (Santos and Lindner, 2017; Wang *et al.*, 2015). A number of studies have also reported the identification of a small number of eukaryotic phosphohistidine, phosphoarginine and phospholysine sites on cellular proteins (Ciesla *et al.*, 2011; Klumpp and Krieglstein, 2002). However, consensus among various phosphoproteomics studies indicate that phosphoserine constitutes nearly 80% of the total phosphoproteome followed by phosphothreonine (15%) and phosphotyrosine (4%) (Cesaro and Pinna, 2015).

Given the importance of phosphorylation signal transduction, a paramount task in cell biology remains to identify substrates of the different kinases in mammalian cells (Corwin *et al.*, 2017). This is especially important in the context of tyrosine kinases where the vast pool of cellular substrates for the majority of the tyrosine kinases have been unexplored. Non-receptor tyrosine kinases represent roughly only 6% of the mammalian kinome (Manning *et al.*, 2002). Several

studies have shown that non-receptor tyrosine kinases like SFKs and BRK, for instance, regulate a number of essential cellular processes linked to growth and proliferation by interacting with and phosphorylating target cellular proteins (Abram and Lowell, 2008; Brauer and Tyner, 2010; Goel and Lukong, 2015; Yeatman, 2004). The identification of the cellular substrates and signalling intermediates of the other non-receptor tyrosine kinases are therefore key to understanding the greater role and significance of these enzymes in mammalian cells.

Mass spectrometry (MS) is an analytical technique that measures the mass to charge (m/z) ratio and intensity of ions in the gas-phase (Grebe and Singh, 2011). This principle can be applied to identify and quantify unknown peptides in any biological sample (Grebe and Singh, 2011). Tandem mass spectrometry (referred to as MS/MS) involves the determination of the mass/charge ratio of an intact peptide (referred to as the MS1 scan) followed by fragmentation of the peptide (referred to as the MS2 scan) to determine its amino acid sequence (McLafferty, 1981; Syka *et al.*, 2004; Wells and McLuckey, 2005). The mass of an intact peptide alone does not provide enough information to precisely and unambiguously identify a peptide (Olsen and Mann, 2004; Peng and Gygi, 2001). Therefore, fragmentation of the precursor peptide in an MS2 scan is necessary to correctly and unambiguously assign sequence-specific identity to the peptide (Olsen and Mann, 2004; Peng and Gygi, 2001). In modern mass spectrometry analyses, the selection of peptides for fragmentation is data-driven and completely automated through the use of dedicated MS instrument software (Bauer *et al.*, 2014; Hu *et al.*, 2016). In the data-dependent acquisition (DDA) mode, a fixed number of highly abundant precursor peptides from MS1 are selected by the software for fragmentation to obtain the MS2 profile of the peptides (Bateman *et al.*, 2014; Mann *et al.*, 2001). This is helpful in eliminating noise from undesired MS1 profiles corresponding to non-peptide moieties in the sample.

In a typical proteome-wide (global) phosphoproteomics experiment, the total cellular proteins are first extracted from the sample of choice (cells/tissues) and subjected to digestion using endoproteinases such as trypsin and/or Lys-C (Figure 1.9). From the crude tryptic peptides, phosphorylated peptides are selectively enriched using either metal-ion affinity-based techniques such Titanium dioxide (TiO_2) or antibody affinity-based enrichment techniques utilizing immunoprecipitation assays (Fila and Honys, 2012; Macek *et al.*, 2009; Nilsson, 2012) (Figure 1.9). The TiO_2-based enrichment technique utilizes immobilized TiO2-coated resin for binding phosphopeptides in the mobile phase (Roux and Thibault, 2013; Thingholm *et al.*, 2009). This

technique is suited for primarily enriching phosphoserine and phosphothreonine peptides and to a lesser extent, phosphotyrosine peptides (Possemato *et al.*, 2017; Thingholm *et al.*, 2009). The enrichment of phosphotyrosine peptides is ideally performed using well-characterized phosphotyrosine antibodies via immunoprecipitation assays (Thingholm *et al.*, 2009; van der Mijn *et al.*, 2015). After performing the enrichment steps, the phosphopeptides are separated by liquid chromatography (LC) and injected into the mass spectrometer where these are ionized and subsequently fragmented to obtain the MS1 and MS2 profiles (Xie *et al.*, 2011) (Figure 1.9). The

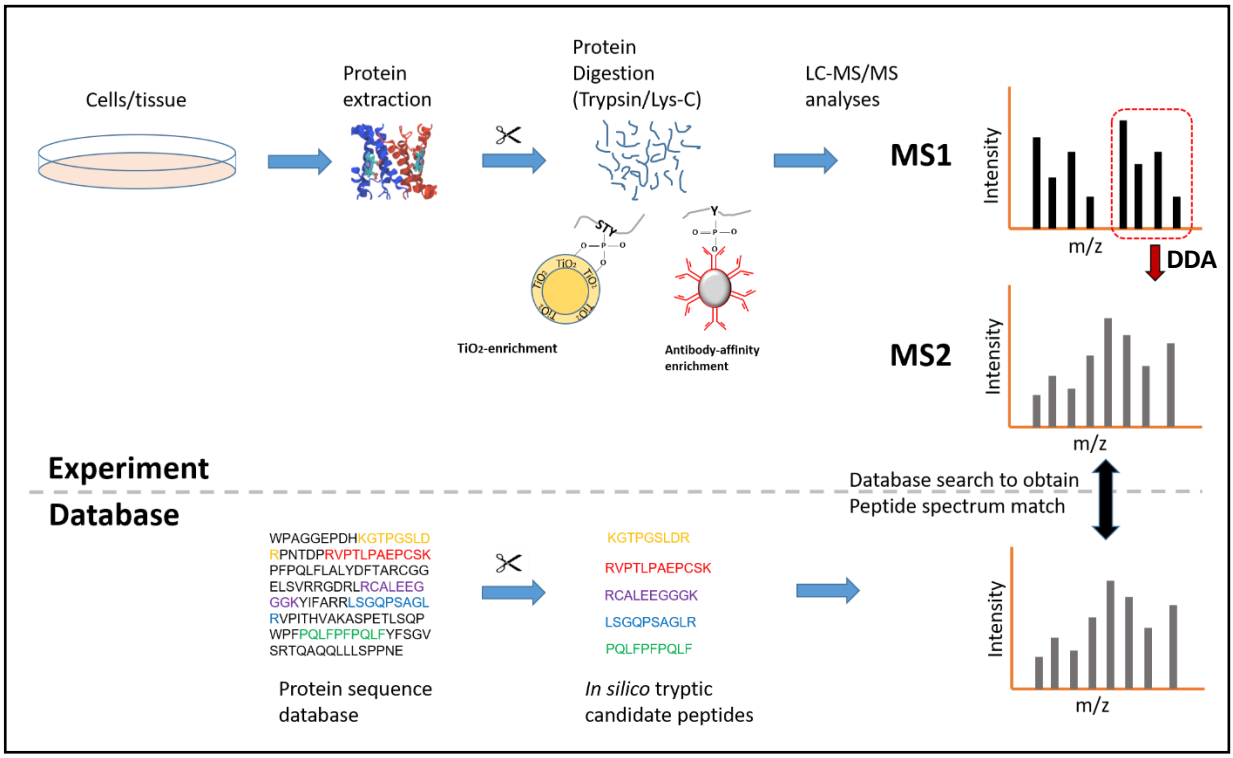


Figure 1.9: **Outline of a general LC-MS/MS-based phosphoproteomics workflow.** Total proteins, extracted from cells/tissue, are digested with endoproteinases such as Trypsin and/or Lys-C and subjected to phosphopeptide enrichment, via enrichment techniqes such as TiO_2-based phosphopeptide enrichment or phospho-specific antibody-based enrichment. The enriched phosphopeptides are separated by liquid chromatography followed by mass spectrometry analysis. Through data-dependent acquisition (DDA), user-defined precursor ions from the MS1 scan are selected for fragmentation in the MS2 scan. The MS2 spectra of the peptide is used for assigning amino acid sequence information to the identified peptide using database-search software. The experimentally identified spectra is searched within a database of *in silico*-generated theoretical spectra, derived from *in silico*-digested peptides. The database used for this search comprises of all species-specific protein entries. A peptide spectrum match is obtained when the theoretically-derived spectra matches the experimentally-obtained spectra. A peptide spectrum match is used to assign peptide sequence and protein identity.

MS1 and MS2 spectral information obtained from the mass spectrometer is then used by specific "search"software to determine the identity of the peptides (Wang and Wilson, 2013) (Figure 1.9). Popular examples of such software include Sequest, Mascot and Andromeda (Cox *et al.*, 2011; Koenig *et al.*, 2008; Sadygov, 2015). Here, the software refers to a database of species-specific protein entries (Beavis and Fenyö, 2000; Ulintz *et al.*, 2006; Yates, 1998). From this database, depending on the endoproteinase used during the experimental workflow, the software generates a library of theoretical spectra corresponding to the mass/charge ratio of *in silico*-digested peptides. This theoretical database of spectra is matched to the experimentally obtained spectra from the mass spectrometry analyses (Beavis and Fenyö, 2000; Ulintz *et al.*, 2006; Yates, 1998) (Figure 1.9). Consequently, a probabilistic peptide-spectrum match (PSM) is obtained from such analyses and used to infer identity of the experimentally obtained peptide (Frank, 2009). In the case of a phosphosite-containing peptide, the presence of a phosphate group increases the mass of the precursor ion (in the MS1 scan) and in most cases, that of the fragmented ions (in the MS2 scan) (Kim *et al.*, 2016). This information is used to identify phosphopeptides from their non-phosphorylated counterparts. Importantly, PTMs other than phosphorylation are particularly challenging to distinguish as these exhibit highly similar mass (Kim *et al.*, 2016). Further, PTMs in general may add very little to the overall mass of a given peptide and may be difficult to distinguish from peptides that do not contain PTMs (Kim *et al.*, 2016). Conventional high-resolution mass spectrometers such as Orbitraps offer an unprecedented sensitivity to distinguish between the different PTMs to assign the correct PTM to a peptide (Kim *et al.*, 2016; Olsen *et al.*, 2005). A challenge here, however, is the ability to precisely identify the correct location of the PTM especially in peptides containing two or more PTMs, as in the case of multiply-phosphorylated peptides (Kim *et al.*, 2016). To this end, conventional software incorporate iteration-based scoring methodologies to score the probability for a PTM to exist at a given site on a given peptide (Kim *et al.*, 2016). For example, the Andromeda software implements an iterative algorithm to derive a "Localization probability" for high-confidence PTM-site assignment (Cox *et al.*, 2011). Overall, the entire experimental workflow in mass spectrometry-based phosphoproteomics is integrated to automated computational analyses for the identification of phosphoproteins and the corresponding phosphosites on these proteins.

1.2.1 Quantitation approaches in mass spectrometry analyses

In addition to providing information on the mass to charge ratio of the precursor and fragment ions, mass spectrometry also provides information on the intensity of the ions (Han *et al.*, 2008; Pechan and Gwaltney, 2012). Whereas the m/z ratio gives information about the identity of the peptide, the intensity provides quantitative information on the abundance of the parent peptide in the sample (Han *et al.*, 2008; Pechan and Gwaltney, 2012). In mass spectrometry analyses, relative quantitation of protein/peptide abundance can be accomplished using either labelling-based approaches such as SILAC (Stable isotope labelling of amino acids in culture) and iTRAQ (Isobaric tags for relative and absolute quantification) or label-free approaches (Mann, 2006; Neilson *et al.*, 2011; Wiese *et al.*, 2007). In the case of label-based approaches such as SILAC, the cells are metabolically labelled *in vivo* with amino acids comprising "heavy" isotopes such as ^{13}C and ^{15}N (Mann, 2006) (Figure 1.10). In the other experimental condition, the cells are left unlabelled and therefore, comprise amino acids representing the corresponding natural or "light" isotopes such as ^{12}C and ^{14}N (Mann, 2006). The proteins are then extracted from both experimental conditions, digested with endoproteinases and combined prior for multiplexed analyses on the mass spectrometer (Figure 1.10). The peptides derived from the "heavy" isotope-labelled samples result in a mass shift when analysed by mass spectrometry. This is leveraged to obtain a relative quantitation of peptides between the "heavy" and "light" experimental conditions by comparing the 'heavy" and "light" peptide precursor ion intensities in the MS1 scan (Mann, 2006) (Figure 1.10).

The iTRAQ method also involves a differential chemical-labelling approach (Unwin, 2010; Wiese *et al.*, 2007). However, in this case, amino acid labelling occurs *in vitro* and on peptide-digests instead of intact proteins (Rauniyar and Yates, 2014; Wiese *et al.*, 2007) (Figure 1.10). The chemical labels used in iTRAQ are isobaric, implying that these chemical groups share a similar mass (Rauniyar and Yates, 2014; Thompson *et al.*, 2003). However, upon fragmentation in the mass spectrometer, these chemical groups yield reporter ions of different mass (Thompson *et al.*, 2003; Unwin, 2010; Wiese *et al.*, 2007). The LC-MS/MS-based measurement of the abundance of the reporter ions is reflective of the abundance of the labelled peptide from which these ions were produced (Figure 1.10). In this way, proteins from different samples can be labelled with different isobaric chemical groups and relative quantification of each protein

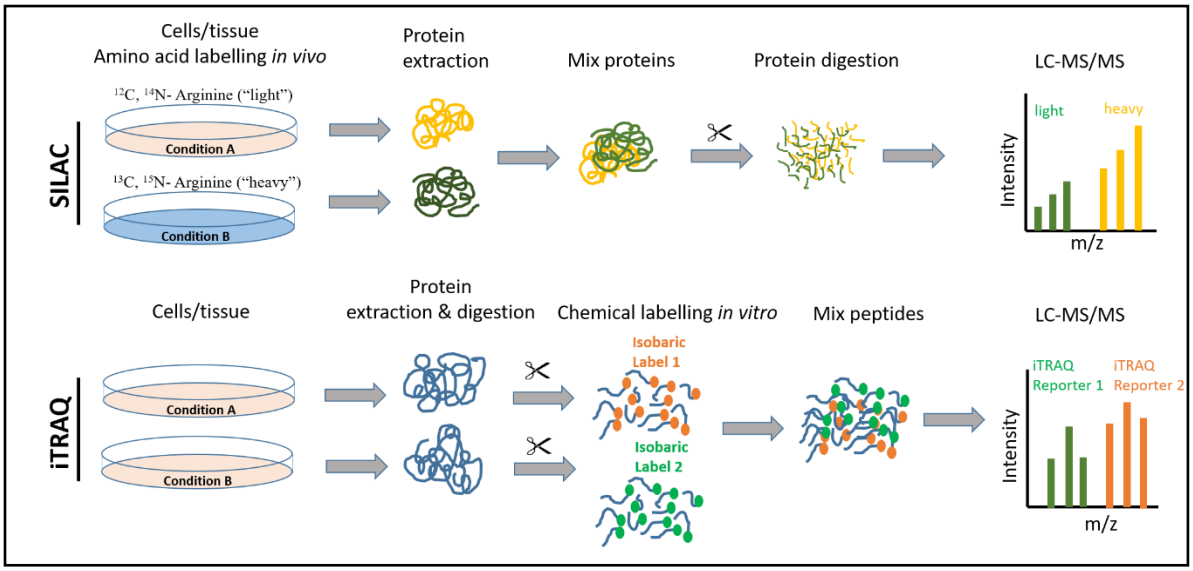


Figure 1.10: **Overview of metabolic/chemical labelling of amino acids as a quantification strategy in mass spectrometry analyses.** The upper panel depicts a schematic representation of the SILAC labelling workflow for two experimental conditions. In one experimental condition, the cellular proteins are labelled *in vivo* during cell culture using amino acids comprised of "heavy" isotopes of carbon (C) and nitrogen (N) such as ^{13}C and ^{15}N. In the other experimental condition, the cellular proteins are left unlabelled and comprise naturally ocuring "light" ^{12}C and ^{14}N. Total proteins are extracted from both experimental samples and mixed prior to LC-MS/MS analyses. Quantification is achieved by comparing the MS1 intensities of the "light" (unlabelled) and "heavy" (labelled) peptides. The lower panel presents a schematic workflow of the iTRAQ quantification approach. In this case, the cellular proteins are first digested and the peptides are labelled *in vitro* with different chemical isobaric groups tags for both experimental conditions. Quantification is similarly achieved by comparing the MS1 intensities of the different isobaric reporter ions between the two experimental conditions.

determined based on the abundance of the corresponding reporter ions produced (Thompson *et al.*, 2003; Wiese *et al.*, 2007) (Figure 1.10).

Label-free quantitation, as the name suggests, does not involve chemical or metabolic labelling of proteins. The label-free quantification approach is divided into two types: (i) quantitation based on spectral-counting or (ii) quantitation based on summed ion intensities (Nahnsen *et al.*, 2013; Neilson *et al.*, 2011) (Figure 1.11). Spectral-counting involves measuring the number of MS/MS spectra identified for a given protein or peptide in two or more experimental conditions (Neilson *et al.*, 2011) (Figure 1.11). The approach assumes that high-abundance peptides will result in a greater number of MS/MS spectra than peptides at lower abundance, between two experimental samples being assessed (Neilson *et al.*, 2011). The other label-free quantitation approach involves measuring the peptide precursor ion abundances across the

corresponding retention times to generate a chromatographic elution profile of these peptides (Figure 1.11). For a given peptide, the entire precursor ion signal across specific retention times referred to as the "area under the curve" (AUC), is measured between two or more experimental conditions for relative quantitation (Neilson *et al.*, 2011) (Figure 1.11). Overall, for both, label-based and label-free approaches, dedicated software are available which incorporate raw LC-MS/MS data to perform peptide/protein quantitation. Examples of such software include MaxQuant, Progenesis QI and Proteome Discoverer™ (Qi *et al.*, 2012; Tyanova *et al.*, 2016a; Veit *et al.*, 2016). These software integrate peptide-identification data obtained from the "search" software to ultimately provide consolidated information on the name and abundance of a specific peptide/protein in a given sample. For the LC-MS/MS-based phosphoproteomics analyses described in this thesis, label-free quantitation based on precursor ion intensity or AUC-measurement, was used.

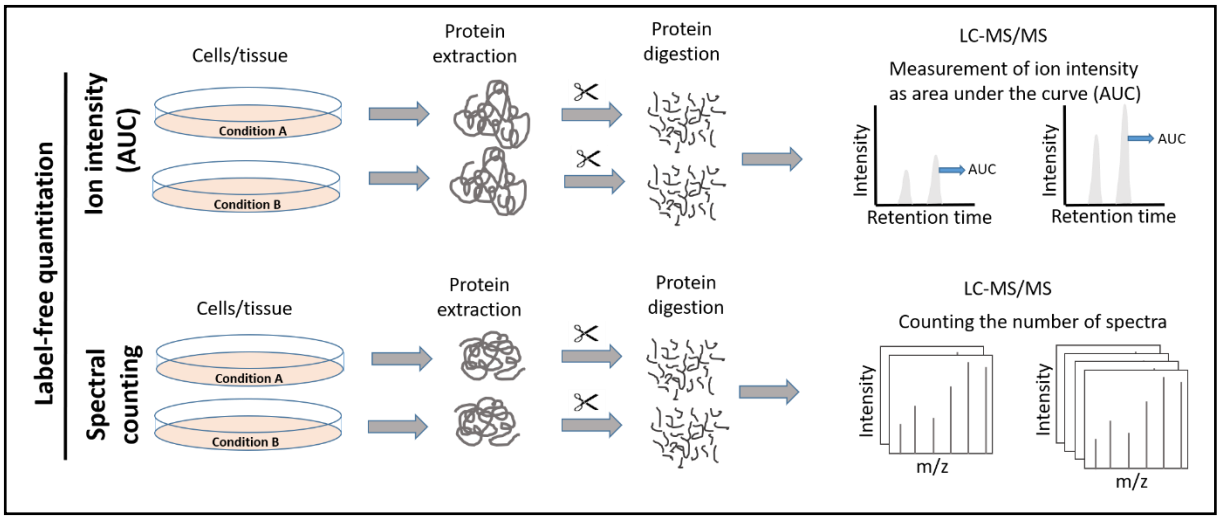


Figure 1.11: **Overview of label-free quntification strategies for mass spectrometry analyses.** Shown here is a schematic workflow of the two different types of label-free quantification strategies for LC-MS/MS analysis. The upper panel depicts the approach of quantifying peptides based on their precursor ion intensities over specific retention times which corresponds to the area under the curve (AUC) of their chromatographic elution profiles. The samples from different experimental conditions are processed in parallel and analysed sequentially via LC-MS/MS. The AUC or precursor ion intensities of corresponding peptides in the different experimental conditions are measured. The lower panel depicts the spectral-counting approach of label-free quantification. Samples from the different experimental conditions are processed in parallel and analysed via LC-MS/MS. Peptide quantification is achieved by counting the corresponding MS1 spectra in the different experimental conditions.

2.0 Hypothesis and Objectives

2.1 Rationale and Hypothesis

The non-receptor tyrosine kinase SRMS belongs to the BRK family kinases whose other two members include BRK and FRK (Goel and Lukong, 2015, 2016). Both BRK and FRK represent well-studied kinases in the context of the biochemical and cellular roles played by these enzymes in mammalian cells (Goel and Lukong, 2015, 2016). The identification of a number of substrates, binding partners and/or signalling intermediates of BRK and FRK has led to a broad mechanistic understanding of the cellular and/or physiological roles played by these kinases (Brauer and Tyner, 2010; Goel and Lukong, 2015, 2016). SRMS, on the other hand, is a relatively understudied member of the BRK family. Though discovered in 1994, little or no information on the biochemical, cellular and physiological role of SRMS exists in the literature. Previous biochemical studies on SRMS have shown that the wild-type kinase is catalytically active and can phosphorylate several cellular proteins when expressed ectopically in mammalian cells (Goel *et al.*, 2013). The identity of these potential SRMS cellular substrates and other signalling intermediates are largely unknown. Additionally, the SRMS protein is known to display a characteristic punctate cytoplasmic localization in mammalian cells (Goel *et al.*, 2013), indicating that SRMS biology may be regulated by its intracellular localization. Kinases are generally known to play specific cellular functions by phosphorylating specific cellular proteins and the intracellular localization of kinases plays a key role in this process.

We hypothesize that SRMS regulates specific cellular processes by phosphorylating its cellular substrates and/or regulating specific signaling intermediates. We further hypothesize that the SRMS protein is sequestered to specific cytoplasmic organelle(s) resulting in the characteristic cytoplasmic punctate localization pattern displayed by the kinase.

2.2 Objectives

The overall aim of this work is to obtain a better understanding of the cellular and biochemical significance of SRMS by identifying the cellular substrates and signaling intermediates of the kinase via quantitative mass spectrometry-based phosphoproteomic analyses. This work further aims to characterize the intracellular localization pattern of SRMS. Therefore, the specific objectives of this work are:

1. To identify and validate the potential cellular substrates of SRMS.
2. To identify the potential signaling intermediates regulated indirectly by SRMS.
3. To characterize the cytoplasmic punctate localization displayed by SRMS.

3. Materials and methods

3.1 Reagents and chemicals

The reagents and chemicals used towards the experiments described in this thesis have been compiled in Table 3.1. Table 3.2 lists the addresses of the manufacturers/suppliers of these reagents/chemicals.

Table 3.1: List of reagents and/or chemicals. This table lists all commercially procured reagents and chemicals used to prepare solutions or other material to perform the experiments described in this thesis. The names of the supplier and catalogue information related to the reagents/chemicals are also indicated.

Reagents/ Chemicals	Supplier and Catalogue number.
2x Laemmli	Sigma-Aldrich, S3401
Acrylamide	Sigma-Aldrich, A5934
Acetic Acid	Fisher Scientific, A465250
Acetonitrile, LC-MS-grade	ThermoFisher Scientific, 51101
Acetonitrile	EMD-Millipore, AX0145
Agar	Fisher Scientific, BP1425-2
Agarose	Fisher Scientific, BP1360
Bovine serum albumin	Amresco, 0332
Bradford's assay dye reagent concentrate	Bio-Rad, 500-0006
Aprotinin	Sigma-Aldrich, A6279
Ampicillin	EMD Millipore, 171255
Ammonium Persulfate (APS)	Sigma-Aldrich, A3678
Adenosine triphosphate (ATP) 10 mM	New England Biolabs, P0756S
Bovine calf serum	Thermo Scientific, SH30087.02
BCA protein assay kit	ThermoFisher Scientific, 23225
N'N-methylenebisacrylamide	Amresco, 0172
Boric acid	EMD Millipore, 203-667
Chloramphenicol	Sigma-Aldrich, C0378
4',6-diamidino-2-phenylindole (DAPI)	ThermoFisher Scientific, P36941
Disodium phosphate (Na_2HPO_4)	Fisher Scientific, S375-500

Dimethyl sulfoxide (DMSO)	Fisher Scientific, TS-20684
Dithiothreitol (DTT)	ThermoFisher Scientific, R0861
Dulbecco's modified Eagle's medium (DMEM)	Hyclone, SH300243.01
EconoTaq polymerase	Lucigen, 30031-1
Ethylenediaminetetraaceticacid (EDTA)	EMD Millipore, 324503
Epidermal growth factor (EGF)	Millipore-Sigma, E9644
Fetal bovine serum	ThermoFisher Scientific, 10438026
Formic acid, LC-MS-grade	ThermoFisher Scientific, 28905
GST-bind resin	Novagen, 70541
Glutathione	Sigma-Aldrich, G4251
Gel Red	Biotium, 41002
Glycerol	Sigma-Aldrich, G5516
Glycine	Fisher Scientific, S80028
Ethyl alcohol (EtOH)	Fisher Scientific, S25310
Hydrochloric acid (HCl)	Fisher Scientific, A508-P500
HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	Sigma-Aldrich, H3375
Iodoacetamide, LC-MS-grade	Sigma-Aldrich, I6125
Kanamycin	EMD Millipore, 420311
Lys-C endoproteinase (LC-MS-grade)	Wako, 129-02541
Methanol	EMD Millipore, MX0475
Magnesium chloride ($MgCl_2$)	Fisher Scientific, 232-094-6
Manganese chloride ($MnCl_2$)	Fisher Scientific, M87-500
Monopotassium phosphate (KH_2PO_4)	Fisher Scientific, BP-362-500
Nitrocellulose membrane	Pall Life sciences, 66489
Phosphatase Inhibitor cocktail	ThermoFisher Scientific,78420
Potassium chloride (KCl)	Fisher Scientific, AC193780000
Polyethyleneimine (PEI)	Polysciences Inc., 23966-1
Pre-stained protein ladder	New England Biolabs (NEB), P7711S
Protease inhibitor cocktail	ThermoFisher Scientific, 78430

Protein A beads	Santa Cruz Biotech, sc-2001
Protein G beads	Santa Cruz Biotech, sc-2002
PRO-Q, Diamond Phosphoprotein Stain	ThermoFisher Scientific, P33300
Phenylmethylsulfonylfluoride (PMSF)	Fisher Scientific, 329-98-6
QuikChange site-directed mutagenesis kit	Agilent, 600250
Restriction enzymes for molecular cloning	New England Biolabs
Retention time calibration mixture	ThermoFisher Scientific, 88321
RNAiMAX transfection reagent	ThermoFisher Scientific, 13778030
Sodium acetate	Sigma-Aldrich, S2889
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich, L3771
Sodium orthovanadate (Na_3VO_4)	EMD Millipore, 567540
Sodium pyrophosphate	Sigma-Aldrich, S6422
Sodium azide (NaN_3)	Fisher Scientific, S227-100
T4 DNA ligase	New England Biolabs (NEB), M0202
Terrific Broth	Fisher Scientific, BP2468-500
N,N,N´,N´-tetramethylethylenediamine (TEMED)	Sigma-Aldrich, T9281
Trifluoroacetic acid (TFA), LC-MS-grade	ThermoFisher Scientific, 28904
Tris base	Fisher Scientific, BP152-5
Triton X-100	Fisher Scientific, AC215680000
Trypsin (Cell culture grade)	Fisher Scientific, SV3003101
Trypsin (LC-MS grade)	Worthington, LS-003744
Tween-20	Fisher Scientific, BP337500
Water (LC-MS grade)	Fisher Scientific, W-5

Table 3.2: Reagent supplier addresses. This table lists the names and addresses of all reagent/chemical suppliers.

Supplier	Address
Agilent	Mississauga, Ontario, Canada
Amresco	Solon, Ohio, USA
Biotium	Burlington, Ontario, Canada
Bio-Rad	Hercules, California, USA

Cell Signaling Technologies	Danvers, Massachusetts, USA
Clontech	Mountain view, California, USA
EMD Millipore	Danvers, Massachusetts, USA
Fisher Scientific	Walton, Massachusetts, USA
Hyclone	Pittsburgh, Pennsylvania, USA
Invitrogen Life Technologies	Green Island, New York, USA
LI-COR, Odyssey	Lincoln, Newark, USA
Lucigen	Middleton, Wisconsin, USA
New England Biolabs	Ipswich, Massachusetts, USA
Novagen	Madison, Wisconsin, USA
Paul Life Sciences	Port Washington, New York, USA
Polysciences Inc.	Warrington, Pennsylvania, USA
Sigma-Aldrich	St. Louis, Missouri, USA
Santa Cruz Biotechnologies	Santa Monica, California, USA
ThermoFisher Scientific	Logan, Utah, USA
Waters corporation	Wilford, Massachusetts, USA
Worthington	Lakewood, New Jersey, USA

3.2 Expression vectors and siRNA

3.2.1 Expression vectors

PCR amplifications were performed using EconoTaq polymerase supplied with *Taq* buffer and dNTP mix with the following reaction and cycling conditions:

Reactions Components (50 µl reaction):

10X Standard *Taq* Reaction Buffer with $Mg2^{+}$ (1x), 10 mM dNTPs (200 µM), 10 µM Forward Primer (0.2 µM), 10 µM Reverse Primer (0.2 µM), Template DNA (<100 ng), *Taq* DNA Polymerase (1.25 units), Nuclease-free water (to 50 µl). Cycling Conditions:

1. Initial Denaturation	95 °C (30 seconds)
30 cycles of	
2. Final Denaturation	95 °C (30 seconds)
3. Annealing	45-68 °C (30-60 seconds)

4. Extension	72 °C (1 minute/kb)
5. Final extension	72 °C for 5 minutes
6. Hold	4°C

All primers used towards PCR reactions were purchased from Invitrogen life technologies, (New York, USA). The full length SRMS cDNA was originally amplified from the pANT7-cGST-SRMS plasmid, purchased from DNASU Plasmid Repository (Tempe, AZ) via PCR and cloned C-terminal to the GFP sequence into the *Hind*III and *Bam*HI sites of the pEGFP-C1 plasmid (Goel *et al.*, 2013). In this way, plasmids encoding GFP-SRMS wild type, GFP-SRMS K258M and GFP-ΔN SRMS were constructed and described in a previous study (Goel *et al.*, 2013). N-terminal FLAG-tagged wild type SRMS or SRMS K258M cDNA sequences were cloned into the *Not*1 and *Bam*H1 restriction sites of the pQCXIH vector (Clontech, USA) using the primer pair: 5'-AAAAGCGGCCGCGCCACCATGGACTACAAAGACGATGACGACAAGGAGCCGTTCCT CAGGA-3' (FLAG-tag nucleotide sequence underlined) and 5'-AAAAGGATCCTCAGGGGTGGCATCTGTGGATGGCGTG-3'. An alternate forward primer was paired with the same reverse primer to generate an untagged wild type SRMS construct using the pQCXIH vector backbone: 5'-AAAAGCGGCCGCATGGAGCCGTTCCTCAGGA-3'. cDNA encoding the SRMS kinase was amplified using the primers: 5'-AAAAAGGATCCATGCCCCAGAAGGCCCCGAGG-3' and 5'-AAAAACTCGAGTCAGGGGTGGCATCTGTG-3' and cloned into the *Bam*H1 and *Xho*1 restriction sites in-frame with the GST sequence of the pGEX-6p1 vector. EGFP-Vimentin-7 (referred to as GFP-Vimentin herewith) was procured from Addgene (plasmid # 56439, Addgene, Massachesetts, USA). The Vimentin cDNA sequence was amplified from the EGFP-Vimentin construct using the primers: 5'- AAAAAAAGCTTCGATGTCCACCAGGTCCGTGTCC-3' and 5'-AAAAAGGATCCTCATTCAAGGTCATCGTGATGCTG-3' and cloned into the *Hind*III and *Bam*HI restriction sites in the pmCherry-C1 vector (Clontech, USA). Vectors encoding Myc-Sam68 and GFP-Sam68 have been described before (Lukong *et al.*, 2005). To generate the GST-Vimentin construct, the Vimentin cDNA sequence was amplified using the primers: 5'-AAAAAGGATCCATGTCCACCAGGTCCGTG-3' and 5'-AAAAAGTCGACTCATTCAAGGTCATCGTGATGCTG-'3 and cloned into the *Bam*HI and *Sal*I restriction sites of the pGEX-6P1 VECTOR. Similarly, cDNA sequence encoding Sam68 was

amplified using primers: 5'-AAAAAGTCGACTCATGCAGCGCCGGGACGACCCCGCCGCG-3' and 5'-AAAAAGCGGCCGCTCAATAACGTCCATATGGGTG-3 and cloned into the *Sal*I and *Not*I sites of the pGEX-6p1 vector. The amplified sequences were ligated into the respective digested plasmids using the T4 DNA ligase by following the manufacturer's instructions. The ligated products were used for transforming the commercially procured X-Gold ultra-competent cells (#200315, Agilent Technologies). The following day, independent bacterial colonies were isolated, cultured and DNA extracted from the respective clonal bacterial cultures. The presence of the ligated sequences was confirmed by restriction digestion and via sequencing (sequencing facility: NRC-PBI, University of Saskatchewan, Saskatoon).

3.2.1.1 Site-directed mutagenesis

Site directed mutagenesis was employed to introduce point mutations in the "FLIRPS-ESS" phosphotyrosine-binding pocket of the SRMS SH2 domain. The primer pair used to generate this mutant was:

5'-GAACCAGGGGCCGCAGCAGCAGCAGCAGCAGCAGCAGCACTCGGGGGCTAC-3'

and 5'GTAGCCCCCGAGTGCTGCTGCTGCTGCTGCTGCTGCTGCGGACCCTGGTCC-3'

The commercially available QuikChange site-directed mutagenesis kit was used according to the manufacturer's protocol outlined below:

Reaction Components (50 µl reaction):

Standard Reaction Buffer with $Mg2^{+}$ (1x), 20 ng Template DNA, 125 ng forward primer, 125 ng reverse primer, 200 µM dNTP mix, nuclease free water (to 50 µl) and *Pfu turbo* DNA polymerase (1.25 units). Cycling Conditions:

1. Initial Denaturation	95 °C (30 seconds)
20 cycles of	
2. Final Denaturation	95 °C (30 seconds)
3. Annealing	55 °C (30-60 seconds)
4. Extension	68 °C (1 minute/kb)
5. Final Extension	68 °C (5 minutes)
6. Hold	4 °C

The PCR-mutagenesis products were incubated with the restriction enzyme *Dpn*I at 37 °C for 1 hour to digest the parental cDNA and the digested PCR product directly used for transforming the commercially procured X-Gold ultra-competent cells (#200315, Agilent, USA). The following day, bacterial colonies were picked, DNA extracted from the clones and screened via sequencing (sequencing facility: NRC-PBI, University of Saskatchewan, Saskatoon).

3.2.2 SRMS siRNA

The SRMS siRNA (sc-63066) targeting human SRMS mRNA and scrambled siRNA (sc-37007) were procured from Santa Cruz Biotechnologies (Santa Cruz Biotechnologies, USA). The SRMS siRNA was purchased as a pool of two duplexed siRNAs, sc-63066A (sense- CUUGCGCUCUAUGACUUCAtt and antisense- UGAAGUCAUAGAGCGCAAGtt) and sc-63066B (sense- GCGAUCAAGGUCAUCAAGUtt and antisense- ACUUGAUGACCUUGAUCGCtt).

3.3 Cell lines and cell culture:

HEK293, HeLa and MDA-MB 231 cells were purchased from the American Tissue Type Culture Collection (ATCC, USA). All cell lines were maintained in DMEM High glucose (4.5g/L) Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum and cultured at 37 °C in 5% CO_2.

The HEK293 cell line was selected for the proteomics studies described in this thesis. HEK293 is a normal cell line isolated from the human embryogenic kidneys and expresses very low level of SRMS (Goel *et al.*, 2013). The cell line is also relatively more amenable to exogenous gene expression experiments compared with other mammalian cell lines (Thomas and Smart, 2005). Therefore, we used this cell line as an unbiased and suitable model to investigate the cellular substrates of SRMS in human cells. Further, validation studies involving ectopic expression of SRMS were carried out in the HEK293 and HeLa cell lines. Both cell lines are known to display low endogenous/background SRMS expression and were therefore found ideal for ectopic SRMS expression studies purpose (Goel *et al.*, 2013).

3.3.1 Plasmid/ siRNA transfections and EGF stimulation:

All transfections were carried out in the HEK293 and HeLa cell lines which were cultured as described above. Cells cultured in 10 cm tissue culture dishes or 6 well tissue culture plates were transiently transfected with 10 µg or 2.5 µg DNA, respectively, using 1% Polyethyleneimine (PEI) as the transfection reagent. For each well in the 6-well plate, 2.5 µg of the appropriate DNA was mixed with 107.5 µl of 0.15 M sterile NaCl via gentle vortexing for 10 seconds. 15 µL of PEI was then added to this mixture followed by another 10 seconds of gentle vortexing. For transfections in a 10 cm tissue culture dish, 10 µg DNA was mixed with 430 µl of 0.15 M sterile NaCl. 60 µl of PEI was added and the solution vortexed for mixing. DNA-PEI complex formation was allowed to take place by incubating the mixture at room temperature for 10 minutes followed by dispensing it dropwise into the wells or 10 cm culture dishes. The culture dishes were gently swirled to allow even distribution of the DNA-PEI complex in the culture dish and incubated overnight in the cell culture incubator. 24 hours post transfection, transfection efficiencies were determined via fluorescence microscopy using the Olympus 1x51 inverted fluorescence microscope and cells harvested for use towards the experiment. SRMS or scramble siRNAs were transfected using the RNAiMAX transfection reagent in cells cultured to 70-80% confluency in 10 cm culture dishes. 36 µl RNAiMAX solution was mixed with 600 µl serum-free media in a microcentrifuge tube by gentle vortexing for 10 seconds. In a separate microcentrifuge tube, 100 pmol siRNA was mixed with 600 µl serum-free media by vortexing gently for 10 seconds. The diluted siRNA mix was then added to the diluted RNAiMAX solution and vortexed briefly again for 10 seconds. This mix was allowed to incubate for 5 minutes at room temperature and then dispensed dropwise in the 10 cm tissue culture dish. The dish was gently swirled to allow even distribution of the siRNA-RNAiMAX complex across the dish. 24 hours post-transfections, the cells were lysed and immunoblotting performed to confirm SRMS-knockdown using antibodies against SRMS (Table 3.5). For stimulating cells with EGF, the cells were serum starved overnight and treated with 100 ng/mL EGF for various times and lysed for use as total cell lysates or for immunoprecipitation.

3.4 SDS-PAGE and Western Blotting

3.4.1 SDS-PAGE

For preparing whole cell lysates, the media in the tissue culture plates was first aspirated and residual media removed by gently rinsing the culture plates once with 1x PBS. Whole cell lysates were then prepared by harvesting the cells directly in 2x Laemmli (sample buffer). Total cell lysates were prepared using freshly prepared lysis buffer (Table 3.3). 2x Laemmli was added to the total cell lysates or immunoprecipitates in a 1:1 ratio. Protein samples derived from either whole cell lysates or immunoprecipitates, were resolved via sodium dodecyl sulphate (SDS) Polyacrylamide gel electrophoresis (PAGE). SDS-PAGE was performed using the Mini-Protein 4 gel electrophoresis system (#165800FC, Bio-Rad, USA). 10% polyacrylamide gels with a 1.5 mm thickness were cast using the appropriate glass plates provided with the gel electrophoresis pack. The resolving gel comprised 10% acrylamide, 0.8% bis-acrylamide, 0.4% SDS, 375 mM Tris HCl pH 8.8, 0.16% (w/v) APS, 0.1% TEMED and H_2O. The stacking gel comprised 4% acrylamide, 0.8% bis-acrylamide, 0.4% SDS, 125 mM Tris HCL pH 6.8, 0.24% (w/v) APS, 0.1% (w/v) TEMED and H_2O.

The protein samples and the pre-stained protein ladder were boiled at 100 °C for 5 minutes prior to loading onto the gels (10 wells or 15 wells). The gels were run in 1xSDS running buffer at a constant voltage of 150 volts for 1.5 hours or until the bromophenol blue dye front passed through the gel.

3.4.2 Western Blotting:

Following electrophoresis, the gels were removed carefully from the glass plates. The stacking gels were discarded. The resolving gel was overlaid on five 3 mm filter paper sheets. The filter paper sheets were cut to the size of the gel and pre-soaked in transfer buffer. Likewise, nitrocellulose membranes were also cut to the size of the gel and soaked in transfer buffer. The wet membranes were then overlaid on the gels. Another five sheets of 3 mm filter paper were placed on top of the nitrocellulose membrane to give a “sandwich”-like assembly. This assembly was placed into the gel-holder cassette and the cassette inserted into the Western blotting electroblotting apparatus (Bio-Rad). The apparatus was filled with Western blotting transfer buffer (Table 3.4) up to the manufacturer’s indicated mark. Protein transfer was allowed to proceed for 1 hour at a constant voltage of 100 volts at 4 °C.

After the completion of protein-transfer, the nitrocellulose membranes were removed from the cassette and subjected to blocking by incubating with 5% skim milk on a rocking platform for 40 minutes at room temperature. After blocking, the membranes were washed thoroughly in 1x PBS to remove residual skim milk. Next, primary antibodies diluted in primary antibody buffer (Table 3.3) and prepared at appropriate dilutions (listed below), were added to the membrane for overnight incubation on a rocker at 4 °C. The following day, the membrane was washed three times for 15 minutes with 1x PBST buffer. The membranes were then incubated with the appropriate secondary antibody (LICOR goat anti mouse/ goat anti rabbit, as mentioned below), diluted in secondary antibody buffer at a dilution factor of 1:10,000. Incubation of the membranes with secondary antibodies was carried out for 40 minutes on a rocker at room temperature. The membranes were subsequently washed three times with 1x PBST (Table 3.4) and used for analyses on the LI-COR Odyssey infrared scanning instrument (LI-COR, USA). Images were acquired using the Image Studio™ Lite software (LI-COR, USA) linked to the LI-COR instrument at the appropriate laser intensities. All immunoblotting images acquired were transformed to greyscale and exported in the .tiff image format. The Western blotting experiments were repeated at least 3 times to ensure reproducibility.

3.5 Primary and secondary antibodies

Phosphotyrosine antibodies, 4G10 (#05-321) and pTyr 1000 (#8954) were procured from EMD Millipore and Cell Signaling Technologies, respectively. FLAG antibodies (#MA1-91878) were purchased from Thermo-Scientific, USA. The well characterized Sam68 antibody, AD1, has been described before (Lukong *et al.*, 2005) and was used for immunoprecipitation experiments. All other primary antibodies were purchased from Santa Cruz Biotechnologies. These include antibodies against GFP (#sc-101525), GST (#sc-459), SRMS (#sc-374524), Myc (#sc-40), Sam68 (#sc-1238) and β-actin (#sc-47778). Secondary antibodies for Western blotting analyses namely, goat anti-rabbit (IR Dye-680RD IgG, #926-68071) and goat anti-mouse (IR Dye-800CW IgG, #926-32210), were purchased from LI-COR Odyssey, USA. Secondary goat anti-mouse IgG-Texas red (#sc-2781) antibodies were purchased from Santa Cruz Biotechnologies (Santa Cruz, USA). Secondary antibodies for Western blotting analyses namely, goat anti-rabbit (IR Dye-680RD IgG, #926-68071) and goat anti-mouse (IR Dye-800CW IgG, #926-32210), were purchased from LI-COR Odyssey, USA. The lyophilized secondary antibodies were diluted in

double distilled H_2O at a final concentration of 1 mg/mL and aliquots of these stock antibodies stored at -20 °C. For immunofluorescence analyses, secondary goat anti-mouse IgG-Texas red (#sc-2781) antibodies were purchased from Santa Cruz Biotechnologies at a dilution of 1:200. The working concentrations used for primary antibodies are described in the table 3.2.

TABLE 3.3: List of primary antibodies used. Shown in this table are the names and working concentrations/dilutions of primary antibodies used in Western blotting (WB), immunoprecipitation (IP) and immunocytochemistry (IC) experiments.

Primary antibody	**Working Concentration/Dilution**
Anti-SRMS (sc-374524)	0.2 µg/mL (WB), 2 µg/mL (IC)
Anti-GST (sc-33613)	0.2 µg/mL (WB)
Anti-β-actin (sc-4778)	0.2 µg/mL (WB)
Anti-GFP (sc-101525)	0.2 µg/mL (WB)
Anti-FLAG (MA1-91878)	4 µg (IP)
Anti-Myc (sc-40)	4 µg (IP)
Anti-phosphotyrosine (sc-508)	0.2 µg/mL (WB)
Anti-Sam68 (rabbit) (AD1)	1:1000 (WB), 4 µg (IP)
Anti-Sam68 (sc-1238)	1:1000 (WB)

3.6 Expression of recombinant GST-fused proteins

All recombinant GST-fused proteins (GST-SRMS kinase, GST-Vimentin and GST-Sam68) were expressed in the *E. coli* Rosetta™ (DE3) strain (#7094, EMD Millipore, USA). Briefly, the Rosetta strains transformed with the respective plasmids were cultured in ampicillin and chloramphenicol-containing Terrific broth (BP-2468-2, Fisher Scientific, USA) until $Abs._{600} \sim 0.8$. The expression of the GST-fused proteins was induced with 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) (Table 3.5) at 37 °C for 3 hours. Bacteria were then harvested via centrifugation at 7000 X g and the pellets lysed in freshly prepared lysis buffer (Bacterial cell lysis buffer composition described in Table 3.4). The recombinant proteins were purified from the

TABLE 3.4 List of media/buffers and their composition. This table lists the names and composition of all media/buffers used for the experiments described in this thesis.

BUFFER/ MEDIA	COMPOSITION
1x PBS (Phosphate buffered saline)	137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 1.46 mM KH_2PO_4, pH 7.4 .
1x PBST	1x PBS, 0.1% Tween-20 (v/v)
Blocking buffer	5% Skim milk in 1x PBS buffer (5 g Skim milk in 100 mL PBS buffer)
Primary antibody buffer	0.1% Tween-20, 5% BSA in 1x PBS
Secondary antibody buffer	0.1% Tween-20, 5% BSA in 1x PBS
Western blotting Transfer buffer	25 mM Tris, 192 mM glycine, pH 8.3, 20% methanol.
1X SDS PAGE Running buffer	25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3.
Mammalian cell lysis buffer for Immunoprecipitation assays	50 mM TrisCl pH 7.5, 1 % Triton X-100, 150 mM NaCl, protease inhibitors: Aprotinin 0.6 µg/mL, 1 mM PMSF and phosphatase inhibitor: 1 mM sodium orthovanadate.
RIPA lysis buffer	50 mM Tris-HCl, 150 mM NaCl, 0.1% SDS, 1% Triton X-100 and 0.5% sodium deoxycholate. Buffer supplemented with commercial protease and phosphatase inhibitor cocktails (Refer to Table 3.1 for commercial source information).
Kinase reaction buffer	250 µM unlabeled ATP, 5 mM MOPS, pH 7. 2, 2.5 mM β-glycerol-phosphate, 4 mM $MgC1_2$, 2.5 mM $MnC1_2$, 1 mM EGTA (Ethylene glycol tetraacetic acid), 0.4 mM EDTA and 0.5 mM DTT
Bacterial cell lysis buffer	50 mM Tris 7.5, 150 mM NaCl, 1% Triton X-100, protease inhibitors: Aprotinin 5 mg/L and 0.1 mM PMSF
Terrific broth for bacteria culture	50.8 g of Terrific broth powder in 1 L double-distilled water. Media is autoclaved before use. Kanamycin or ampicillin is added to obtain a working concentration of 1 µg/mL. Additionally, chloramphenicol is added at a working concentration of 12 µg/mL.

TABLE 3.5 List of stock solutions of chemicals/reagents. Shown in in this table is a list of stock solutions of specific chemicals/reagents and their composition.

CHEMICALS/REAGENTS STOCK	COMPOSITION
10 mg/mL Kanamycin	0.5 g 50 mL ddH_2O (double distilled H_2O).
50 mg/mL Ampicillin	1.7 g in 50 mL ddH_2O.
34 mg/mL Chloramphenicol	0.34 g in 10 mL 100% Ethanol
1 M IPTG (Isopropyl β-D-1-thiogalactopyranoside)	1 g IPTG dissolved in 4.2 mL ddH_2O, filtered through 0.22 μm.
100 mM PMSF (Phenylmethanesulfonylfluoride)	1.74 g PMSF in 100 mL Isopropanol.
0.2 M Iodoacetamide	95 mg Iodoacetamide in 5 mL LC-MS-grade H_20
100 mM HEPES pH 8.0	23 g HEPES in 450 mL LC-MS-grade H_2O, adjust pH to 8.0 using NaOH
1.25 M DTT	19.25 g DTT in 100 mL LC-MS-grade H_2O
1 mg/mL Trypsin	20 mg Trypsin dissolved in 1 mM HCl.

crude lysates by incubating with GST-bind resin overnight at 4 °C. The beads were washed six to eight times with lysis buffer and GST-fusion proteins eluted with 10 mM glutathione, 50 mM Tris-Cl pH 7.4 and 1 mM DTT.

3.7 Immunoprecipitation and *in vitro* kinase assays

For immunoprecipitation assays, cells from a 10 cm dish were lysed in 1 mL lysis buffer comprising 50 mM Tris-Cl, 150 mM NaCl, 1% Triton X-100, 0.6 μg/mL Aprotinin, 1 mM PMSF and 1 mM Na_3VO_4, incubated on ice for 30 minutes and centrifuged at 20,000 X g at 4 °C for 10 minutes. The supernatant was collected and divided in three equal volumes for use towards protein immunoprecipitation, mock/control immunoprecipitation with IgG, or input. For immunoprecipitation of the desired protein, 4 μg of the appropriate primary antibody was added to 330 μL of clarified lysates. To confirm the specificity of the primary antibody-

immunoreactivity, 4 µg of control isotype-matched IgG sera was added to an equal volume of clarified lysates in a separate immunoprecipitation assay. The lysates were incubated with the antibody overnight at 4 °C on a gyrorotator. The beads were centrifuged at 3500 X g for 1 minute, washed twice with lysis buffer and the immunoprecipitates eluted with the addition of 30 µL 2x Laemmli. The samples were boiled at 100 °C for 5 minutes and resolved via SDS-PAGE.

For *in vitro* kinase assays, GST-Sam68 or GST-Vimentin were incubated with GST-SRMS kinase *in vitro* with kinase reaction buffer (250 µM unlabeled ATP, 5 mM MOPS, pH 7. 2, 2.5 mM β-glycerol-phosphate, 4 mM $MgC1_2$, 2.5 mM $MnC1_2$, 1 mM EGTA, 0.4 mM EDTA and 0.5 mM DTT) for 30 minutes at 30 °C. The *in vitro* kinase reactions were quenched by the addition of 2x sample buffer, boiled and loaded on SDS-PAGE gels.

3.8 Confocal microscopy

For fluorescence microscopy analyses of cells expressing ectopic GFP or mCherry-fused proteins, cells seeded on glass coverslips were cultured in 6-well culture dishes and transfected as described in the previous section. The following day, the culture media was aspirated and the cells fixed with 3% paraformaldehyde (PFA) for 10 minutes. The fixed cells were then directly mounted on glass slides using mounting media supplemented with DAPI (4',6-diamidino-2-phenylindole) for imaging. For immunofluorescence analyses, cells seeded on coverslips were fixed with 3% PFA and permeabilized with 0.5% Triton X-100 for 5 minutes. The fixed cells were then stained with the appropriate primary antibody at a concentration of 2 µg/mL in 1x PBS for 1 hour at room temperature. The cells were subsequently washed with 0.1% Triton X-100 in 1x PBS for 5 minutes followed by staining with secondary antibodies at a 1:200 dilution in 1x PBS. The coverslips were then mounted on glass slides using mounting media-supplemented DAPI and used for imaging. For lysosomal-staining of live cells using LysoTracker™-Red DND-99 dye (#L7528, ThermoFisher Scientific, USA), cells seeded on cover slips were transfected with the indicated plasmids. 24-hours-post transfection, the cells were fixed with 3% PFA and mounted on glass slides for imaging. Fluorescence imaging of all samples was performed using the Zeiss confocal laser scanning microscope (LSM) system (LSM-700, Carl Zeiss, USA). The LSM 700 system was equipped with the inverted Axio Observer.Z1 SP microscope. All images were acquired using the Zen 2012 SP1 (64-bit, black edition) software (Carl Zeiss, USA). Fluorescence from the EGFP protein was excited using the 488 nm laser (4x Pigtailed laser with single-mode fibers, Carl Zeiss,

USA). Fluorescence from the mCherry protein, Texas-red (sulforhodamine 101 acid chloride)-conjugated antibodies or the LysoTracker™-Red DND-99 dye, was excited using the 555 nm laser (4x Pigtailed laser with single-mode fibers, Carl Zeiss, USA). Fluorescence from the DAPI dye was excited using the 405 nm laser (4x Pigtailed laser with single-mode fibers, Carl Zeiss, USA). All images were acquired at 40x magnification (40x oil-immersion objective) with a resolution of 1024 x 1024 pixels and a pixel-bit depth of 16 bits. Where necessary, images were acquired using at least two averaging scans to enhance the signal-to-noise ratio. Colocalization between two proteins with different fluorescence emissions was assessed by employing the z-stack function of the LSM 700 confocal microscope. The number of z-stack scans and the spatial interval between each scan was appropriately defined to scan the region of interest. All fluorescence and immunofluorescence microscopy experiments were performed at least three times to ensure reproducibility.

For counting cells displaying the punctate localization of ectopic SRMS, cells seeded on coverslips were transfected with GFP-SRMS and viewed by fluorescence microscopy 24 hours-post transfection. The cells were visualized at 20x magnification and total transfected cells (cells showing punctate or diffusive localization) counted from four random microscopic fields. Fifty transfected cells were counted in each field for a total of 200 transfected cells per slide. Then the number of cells showing exclusively punctate localization of SRMS were counted from the same field. Finally, the cells with SRMS punctae were quantified as the percentage of cells showing punctae in the total population of transfected cells [(Number of cells with SRMS punctae in a specific field/ total number of transfected cells in that field) X 100]. Quantifications obtained from all microscopic fields was averaged. Data from three experimental replicates were obtained and presented as Mean ± SEM (Standard Error of the Mean). Statistics were performed using the Graphpad Prism software (Version 5.01).

3.9 Sample preparation and data analysis procedures for the identification of candidate SRMS substrates by LC-MS/MS-based phosphotyrosine-profiling.

Note: All solutions described in the following procedures were prepared using LC-MS-grade water.

3.9.1 Protein digestion and purification of tryptic peptides

HEK293 cells were cultured to nearly 85% confluency in a total of 24 x 10 cm dishes. 12 x 10 cm dishes were transfected with plasmids encoding either GFP-SRMS wild type (SRMS-WT) or GFP alone as control. 24 hours post-transfection, the cells were lysed in lysis buffer comprising 9 M urea and 20 mM HEPES pH 8.0. The respective lysates were pooled in sterile 50 mL conical tubes and sonicated four times with 10-second bursts using the Virsonic 100 ultrasonic cell disruptor (Boston Laboratory Equipment, USA). These were then centrifuged at 15,000 X g and the supernatants transferred to sterile 50 mL conical tubes. Total protein concentration was determined via Bradford's assay. 12 mg total lysate protein, each from GFP and GFP-wild type SRMS-expressing cells, was taken for further processing. Proteins were reduced with dithiothreitol (DTT) and alkylated with iodoacetamide. The reduced and alkylated proteins were diluted four-fold with 20 mM HEPES pH 8.0 and digested with LC-MS-grade trypsin at a 1:100 enzyme to protein ratio (w/w) overnight at 37°C. Crude tryptic digests were purified via reversed-phase (hydrophobic) solid-phase extraction using Sep-Pak cartridges (#WAT051910, Waters Corporation, USA). In this step, the purification cartridges were first equilibrated by passing 6 mL of 0.1% TFA. Next, the tryptic digests were loaded on the cartridges and the bound peptides washed three times by passing 5 mL of 0.1% trifluoroacetic acid (TFA). The washed peptides were then eluted with 5 mL of a 0.1% TFA and 40% acetonitrile solution. The purified tryptic peptides were lyophilized overnight and used towards enrichment of phosphotyrosine peptides.

3.9.2 Enrichment of phosphotyrosine peptides

The enrichment of tyrosine-phosphorylated peptides was performed using the PTMScan p-Tyr 1000 phosphotyrosine enrichment kit (#8803, Cell Signaling Technologies, USA). The lyophilized peptide samples were re-suspended in 1x Immunoaffinity purification (IAP) buffer provided in the kit. The diluted peptides were then centrifuged at 10,000 X g for 5 minutes at 4 °C. The clear supernatant was retained. The pTyr-1000 antibody-bead conjugates were then washed with 1x PBS and incubated with the peptide mix overnight at 4 °C on a gyrorotator. The following day, the immunoprecipitates were washed twice with LC-MS-grade water and once with 1x PBS. The immunoprecipitated peptides were then eluted by incubating the beads with 100 µL of 0.1% TFA at room temperature for 10 minutes. The eluted peptides were desalted using the Empore C18 solid phase extraction plate (#66875, Sigma-Aldrich, USA) as per the manufacturer's

protocol. For peptide-desalting, the C18 columns were first equilibrated by passing 50 µL of 0.1% TFA and 50% acetonitrile solution followed by 50 µL of 0.1% TFA alone. Eluates from the phosphotyrosine immunoprecipitates were then passed through the equilibrated C18 columns. Next, the columns were washed twice with 50 µL of 0.1% TFA and the peptides eluted using 0.1% TFA and 40% acetonitrile solution. The desalted peptides were dried in a speedvac centrifuge for 3 hours and stored at -20 °C until ready for analyses via LC-MS/MS.

3.9.3 Reversed-phase Chromatography and Mass Spectrometry

Samples were analyzed by nano UPLC-MS/MS with a Proxeon EASY-nLC 1000 system interfaced to a Thermo-Fisher Q Exactive HF mass spectrometer (Sample analyses performed by MSBioworks Inc., Ann Arbor, Michigan, USA). Peptides were loaded on a trapping column and eluted over a 75 µm x 25 cm analytical column (Thermo-Fisher Scientific, USA) at 300 nL/min using a 2-hour reversed-phase gradient. Both columns were packed with PepMap C18, 3 µM resin (Thermo-Fisher Scientific, USA). Mobile phases A and B consisted of 0.1% formic acid in water and 0.1% formic acid in 90% acetonitrile, respectively. Peptides were eluted from the column at 300 nL/minute using the following linear gradient: from 2 to 25% B in 100 minutes, from 25 to 50% B in 110 minutes, from 50 to 90% B in 112 minutes, from 90 to 2% B in 113 minutes and held at 2% B for an additional 7 minutes. The spray voltage was 2.2 kV. The mass spectrometer was operated in the data-dependent acquisition mode with the Orbitrap operating at 60,000 FWHM (Full width at half maximum) and 17,500 FWHM for MS and MS/MS respectively. Full scans were acquired at a resolution of 60,000 FWHM with a maximum injection time of 120 milliseconds in the Orbitrap analyzer. The 15 most abundant ions with charge states ≥ 2 were selected for MS/MS fragmentation by high collision-induced dissociation (HCD) and analyzed at a resolution of 17,500 FWHM with a maximum injection time of 60 milliseconds.

3.9.4 Data processing and analysis

All acquired MS/MS spectra were searched against the Uniprot human complete proteome FASTA (Fast all) database v.02/2013 (81,213 entries) using the MaxQuant software (Version 1.5.3.17) that integrates the Andromeda search engine. Label-free quantitation of the data was performed using MaxQuant. Enzyme specificity was set to trypsin and up to two missed cleavages were allowed. Cysteine carbamidomethylation was specified as fixed modification whereas serine,

threonine and tyrosine phosphorylation were specified as variable modifications. Peptide precursor ions were searched with a maximum mass deviation of 6 ppm and fragment ions with a maximum mass deviation of 20 ppm. Peptide and protein identifications were filtered at 1% FDR using the target-decoy database search strategy (Elias and Gygi, 2010). Proteins that could not be differentiated based on MS/MS spectra alone were grouped to protein groups (default MaxQuant settings). Following the principle of parsimony, each protein group was considered a single identified protein for further data analyses since sufficient evidence was unavailable to distinguish between proteins from the same protein group. A threshold Andromeda score of 40 and a threshold delta score of 8 was applied to the identified phosphopeptides in accordance with parameters described previously (Cox and Mann, 2008; Sharma *et al.*, 2014; Tyanova *et al.*, 2016a). The Andromeda score is a search-engine score computed by the Andromeda search software. In order to assign peptide and protein identity the Andromeda search software matches the mass spectrometry-based experimentally derived spectra with the database/ *in silico*-generated peptide spectra and scores the peptide spectrum matches (Cox *et al.*, 2011; Sharma *et al.*, 2014). This score, called the Andromeda score, provides a quantitative estimate of the probability that the LC-MS/MS-identified spectra matches its corresponding database spectra. The higher the score, the more likely that a given peptide spectra matches the database spectra and therefore the higher the accuracy in the assignment of peptide identity. The Delta score adds to the accuracy of the peptide identification process (Cox *et al.*, 2011; Sharma *et al.*, 2014). A given experimentally-obtained peptide spectra may have multiple peptide spectrum matches in a database search. The Delta score is defined as the difference in the Andromeda score of one PSM and the second-best PSM. The Delta score therefore gives a measure of how well a top-scoring PSM compares with the next best PSM. The higher the Delta score, the greater the probability that the PSM is correct and therefore accurate in the context of peptide identity (Cox *et al.*, 2011; Sharma *et al.*, 2014). Pearsons's correlation coefficients were determined in R using $\log_2$ transformed phosphopeptide intensities, as described previously (Grimes *et al.*, 2013; Mertins *et al.*, 2012). The MaxQuant file output designated "Phospho(STY)sites" was used for further analysis on the Perseus software (Version 1.5.0.15). Potential contaminants and reverse hits were filtered. The resulting list of modified peptides was further filtered at the level of phosphosite localization using a localization probability threshold of 0.75 to derive all class I phosphosites, as described previously (Tyanova *et al.*, 2016b). For downstream analyses, all peptides with modifications other than phosphotyrosine (pTyr) were

discarded. Annotated spectra for class I phosphosites identified as statistically significant are provided in the supplementary material.

3.9.5 Data availability in public repository

The mass spectrometry proteomics data (including the files for viewing the annotated spectra) have been deposited to the ProteomeXchange Consortium (Deutsch *et al.*, 2017; Vizcaino *et al.*, 2014) via the PRIDE (Vizcaino *et al.*, 2016) partner repository with the dataset identifier PXD006809. The data can be accessed on https://www.ebi.ac.uk/pride/archive/login using the following reviewer account details:

Username: reviewer22933@ebi.ac.uk

Password: xu7lPykO

3.9.6 Peptide array analyses

Customized peptide microarrays (PepStar™) comprising 102 peptides were manufactured by JPT peptide technologies GmbH (Germany). The peptides were immobilized on modified glass slides at a concentration of 100 fmol/mm^2 with a maximum diametrical spot size of 150 µM on the glass surface. Each peptide was printed on the array in nine replicates. The array peptides and the corresponding phosphosites are listed in Supplementary Table 6. HEK293 cells cultured in four 10 cm dishes were transfected with vectors encoding either GFP-SRMS K258M or GFP-SRMS wild type. 24 hours post-transfection, the cells were harvested and lysed in lysis buffer [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 2.5 mM sodium pyrophosphate, 1 mM Na_3VO_4, 1 mM NaF, leupeptin (1 µg/ml), aprotinin (1 µg/mL) and 1 mM PMSF]. The lysates were incubated on ice for 10 minutes and centrifuged in a microcentrifuge at 14,000 X g for 10 minutes at 4 °C. Total protein concentration in the clarified lysates was determined using the BCA assay kit and normalized using the appropriate volume of lysis buffer. An 80 µL aliquot of the supernatant was mixed with 10 µL of the activation mix [50% glycerol, 500 µM ATP, 60 mM $MgCl_2$, 0.05% v/v Brij-35, and 0.25 mg/mL BSA] and incubated on the peptide arrays for 2 hours at 37 °C in a humidity chamber. The slides were then washed once in PBS-Triton (1% Triton X-100 in 1x PBS) and then stained with the PRO-Q phosphoprotein staining solution with agitation for 1 hour. The arrays were then washed three times with the de-staining solution [20% acetonitrile and 50 mM sodium acetate, pH 4.0] for 10 minutes each. The

de-staining solution was replaced between each wash. The peptide arrays were then washed once with double-distilled water and scanned on the GENEPIX professional 4200A microarray scanner (MDS Analytical Technologies, USA Analytical Technologies USA) at 532 to 560 nm with a 580 nm filter to detect dye fluorescence. Phosphopeptide signals were collected using the GENEPIX 6.0 software (MDS). Data processing, including signal normalization and relative fold-change differences in phosphorylation intensity associated with each peptide, was performed using the PIIKA2 platform (http://saphire.usask.ca/saphire/piika/) (Trost *et al.*, 2013).

3.9.7 Motif enrichment analyses

Phosphosites derived from candidate SRMS substrates were used to identify SRMS kinase-specific motifs. 13 amino acid-long phosphopeptide sequences comprising a central phosphosite were searched for consensus motifs using the Motif-x (http://motif-x.med.harvard.edu/) and MMFPh (http://www.cs.dartmouth.edu/~cbk/mmfph/) motif analyses tools (Chou and Schwartz, 2011; Wang *et al.*, 2012). The data was searched against a background dataset derived from the IPI (International Protein Index) human proteome in both motif-analyses tools (Kersey *et al.*, 2004). For a sequence to be considered as a motif, the minimum number of times that the sequence was required to occur in the dataset, was set to 20. Furthermore, the required motif significance was set to 10E-06.

3.9.8 Functional gene enrichment analyses

Pathway enrichment analysis of the putative SRMS substrates was performed using the g:Profiler software (Ver. r1741_e90_eg37) (Reimand *et al.*, 2016). A background set of 13,505 unique phosphorylated proteins, comprising phosphorylated proteins identified in our MS analyses as well as those derived from proteomics datasets available with PhosphositePlus.org (Hornbeck *et al.*, 2015), was used to compute the enrichment of statistically significant cellular pathways or processes. We retrieved gene sets corresponding to biological processes from Gene Ontology(GO) with statistical enrichment among input proteins (FDR<0.05 according to the internal multiple testing correction method of g:Profiler). Gene sets were filtered by size (minimum size of functional gene set 10, maximum size of functional gene set 500, at least 3 common genes of pathway and input protein list). We included high-confidence annotations of biological processes and excluded genes with only electronic annotations in GO (Inferred from Electronic Annotations,

IEA). The resulting enriched pathways were visualized with the Enrichment Map software (Merico *et al.*, 2010) in Cytoscape (Ver. 3.0). We used a combined overlap and Jaccard coefficient of 0.75 for network granularity and manually annotated the resulting subnetworks of pathways for associated biological themes.

3.10 Sample preparation and data analyses procedures for the identification of candidate SRMS-regulated signaling intermediates by LC-MS/MS-based global phosphopeptide profiling.

Note: All solutions described in the following procedures were prepared using LC-MS-grade water.

3.10.1 Protein digestion and peptide purification

HEK293 cells were cultured to approximately 85% in six 10 cm culture dishes. Three 10 cm dishes were transfected with plasmids encoding GFP-SRMS wild type and the other three dishes were transfected with plasmids encoding GFP alone as control. 24-hours post-transfection the cells were observed under a fluorescent microscope to ensure that similar and over 80-90% transfection efficiencies were achieved. The transfected cells were trypsinized, washed with 1x PBS and lysed in RIPA lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% SDS, 1% Triton X-100 and 0.5% sodium deoxycholate) supplemented with protease and phosphatase inhibitor cocktails. For complete lysis, the cells were sonicated four times with 10-second bursts using the Virsonic 100 ultrasonic cell disruptor (Boston laboratory equipments, USA). The lysates were centrifuged at 12,000 X g for 10 minutes and the clarified lysates collected. 100 µL of total proteins were purified by chloroform:methanol:water precipitation in a microcentrifuge tube where 400 µL of methanol followed by 100 µL chloroform and 300 µL of double-distilled water was added with intermittent vortexing. Next, this tube was centrifuged at 14,000 X g for 2 minutes. The top aqueous layer was carefully removed by pipetting. 400 µL of methanol was added to the mix and the tube centrifuged again for 14,000 X g for 3 minutes, resulting in an insoluble protein pellet. The methanol was carefully pipetted and the residual methanol allowed to evaporate by air-drying. The precipitated protein pellet was then resuspended in 8 M urea pH 8.0 containing 400 mM ammonium bicarbonate. Next, the proteins were reduced with 10 mM DTT and alkylated with 40 mM iodoacetamide followed by digestion at 37 °C with Lys-C (1:100 enzyme to protein ratio) for 6 hours and trypsin (1:100 enzyme to protein ratio) overnight. The digestion reactions

were quenched with 0.1% formic acid and desalted using C_{18} MacroSpin columns (The Nest Group, USA) as described in the previous section. The desalted samples were dried in a speedvac and dissolved in a buffer comprising 3.5% formic acid and 0.1% trifluoroacetic acid (TFA).

3.10.2 Phosphopeptide enrichment

The samples were subjected to TiO_2-based phosphopeptide enrichment using TopTip MicroSpin columns (Cat. #TT1TIO, Glygen Corp., USA), as described previously (Krishnan *et al.*, 2012; Rich *et al.*, 2016). For this, the samples were first acidified with 0.5% TFA and 50% acetonitrile and loaded onto the TopTip columns. The flowthrough from each sample was collected, dried in a speedvac and resuspended in 0.1% TFA for LC-MS analysis. The columns were then washed first with 100% acetonitrile, then with 0.2 M sodium phosphate pH 7.0, 0.5% TFA, and finally with 50% acetonitrile. The enriched phosphopeptides were eluted with 28% ammonium hydroxide, dried in a speedvac centrifuge for 3 hours and resuspended in 0.1% TFA. Peptide concentrations in the phosphopeptide-enriched and flowthrough samples were estimated from A_{280} absorbance using the NanoDrop 2000 instrument (ThermoFisher Scientific, USA). The peptide concentrations were adjusted to 0.05 µg/µL with 0.1% TFA. A 1:10 dilution of the retention Time Calibration Mixture was further added to each sample prior to LC-MS/MS analyses.

3.10.3 Mass spectrometry analyses

All samples were analysed on the Q-Exactive Plus mass spectrometer (ThermoFisher Scientific) connected to a waters nanoACQUITUPLC system equipped with a Waters Symmetry® C18 180 µM x 20 mm trap column and a 1.7 µm, 75 µm x 250 mm nanoACQUITY UPLC column. A total of three replicates for each sample were analysed by LC-MS/MS. 5 µL of each replicate sample (0.05 µg/µL) was injected into the mass spectrometer. Peptide trapping was carried out for 3 minutes at 5 µL/min in 99% Buffer A (0.1% formic acid in water) and 1% Buffer B (0.075% formic acid in acetonitrile) prior to eluting with linear gradients that reached 30% Buffer B at 140 minutes, 40% Buffer B at 155 minutes, and 85% Buffer B at 160 minutes. Two blanks (1st blank comprising 100% acetonitrile and 2nd blank comprising Buffer A) followed each injection to ensure against sample carry over. The mass spectrometer was operated in the data-dependent acquisition mode with the Orbitrap operating at 60,000 FWHM and 17,500 FWHM for MS and

MS/MS, respectively. Full scans were acquired at a resolution of 60,000 FWHM with a maximum injection time of 120 milliseconds in the Orbitrap analyzer. The fifteen most abundant ions with charge states ≥ 2 were selected for fragmentation by HCD (MS/MS) and analyzed at a resolution of 17,500 FWHM with a maximum injection time of 60 milliseconds.

3.10.4 Data processing and analyses

Label-free quantitation of the raw LC-MS/MS data was performed using the Progenesis QI software (Version 3.0, Nonlinear Dynamics, USA) as described before (Bordner *et al.*, 2011; Rich *et al.*, 2016). All MS/MS spectra were searched against the SWISSProt database (taxonomy restricted to Homo sapiens, database version dated November 2015) using the Mascot search engine (Version 2.0). Data from the TiO_2-enriched and flow through fractions were analyzed separately, and the resulting quantitative analyses were combined for the respective experimental conditions. Chromatographic/spectral alignment, mass spectral peak-picking, data filtering and statistical analyses for the protein and peptide quantitation was performed on Progenesis QI. Raw mass spectral features were aligned based on their retention time using a randomly selected reference run. All other runs were automatically aligned to the reference run to minimize retention time variability between the runs. Only spectra with ion signals quantified at or above three times the standard deviation of the noise were selected for subsequent analyses. A normalization factor for each run was calculated to account for the variability in sample loading and ionization. This normalization factor was determined by calculating the quantitative ratio of the reference run to the run being normalized, with the assumption that most proteins/peptides are not changing in the experiment. The experimental set-up grouped replicates of each condition for comparative analyses. The algorithm then calculated and tabulated raw and normalized abundances and ANOVA p-values (calculated as the mean difference and variance associated with the replicate data in both experimental conditions) for each quantified peptide in the dataset. The MS/MS spectra was exported as a .mgf file (Mascot generic files) for database searching. The Mascot search algorithm was used for database searching. Carbamidomethylation (Cysteine), phosphorylation (Serine, Threonine, Tyrosine), deamidation (Asparagine, Glutamine), acetylation (Lysine) and oxidation (Methionine) were specified as variable modifications. Two missed tryptic cleavages were allowed. The precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was set to 0.2 Da. Stringent conditions were set in Mascot to filter out low scoring

peptides by imposing a confidence probability score (p) of <0.05. Peptide and protein identifications were filtered at 1% FDR using the target-decoy strategy (Elias and Gygi, 2010). Only phosphopeptides mapping to proteins with at least two tryptic peptides were considered for analyses. The Mascot search results was exported as .xml files and then imported into the processed dataset in Progenesis QI software where the Mascot-generated peptide identifications were assigned to the corresponding quantified features. Scatterplots illustrating Pearson's correlation coefficients and fold-change phosphopeptide abundance were generated in R using in-house-developed scripts.

3.1.0.5 Data availability in public repository

The mass spectrometry proteomics data have been deposited to the ProteomeXchange consortium (Deutsch *et al.*, 2017; Vizcaino *et al.*, 2014) via the PRIDE partner repository (Vizcaino *et al.*, 2016) with the dataset identifier PXD009234. The data can be accessed on https://www.ebi.ac.uk/pride/archive/login using the following reviewer account details:
Username: reviewer12016@ebi.ac.uk
Password: HKaGSifi

3.10.6 Motif enrichment analyses

Phosphopeptide sequences of 13 amino acids in length that were significantly altered between the experimental conditions (p-value = 0.05 and 3-fold upregulation) were searched for consensus sequences using the Motif-x (Chou and Schwartz, 2011) and PHOSIDA (Gnad *et al.*, 2007) algorithms. Motifs were searched using the IPI human proteome database as the background dataset (Kersey *et al.*, 2004). Motif-x, the minimum number of occurrences for a motif in the dataset and the required motif significance was set to 20 and 10E-06, respectively. On PHOSIDA, the minimum score for phospho-Ser/Thr/Tyr was specified as 10 and the minimum proportion of matching sites was set to 5%.

3.10.7 Functional gene enrichment analysis

Genes mapped from significantly upregulated phosphopeptides were used for identifying cellular and molecular processes, pathways and upstream regulators using the Ingenuity Pathway Analysis (IPA) software (QIAGEN Redwood City, http://www.qiagen.com/ingenuity). Genes

were queried against the Ingenuity knowledge database as the reference set. The analysis was restricted to the documentation of experimentally observed findings on human genes. The Benjamini-Hochberg (B-H) multiple testing correction (B-H corrected p-value <0.05) scoring method was used in the IPA tool to compute the significance for the functional enrichment analysis. Activation z-scores were calculated by IPA's z-score algorithm to predict the overall activation or inhibition of the identified functional cellular processes/pathways and upstream regulators. A positive z-score (z-score > 0) implies an overall predicted activation of the process/pathway/upstream regulator whereas a negative z-score (z-score < 0) implies an overall predicted inhibition or downregulation of the pathway/process/upstream regulator. z-scores ≥ 2 or ≤ -2 are considered significant by IPA for predicted activation or inhibition, respectively. Cellular processes/upstream regulators with no z-scores imply that IPA was unable to generate prediction states for these functionalities.

3.10.8 NetworKIN analyses

The NetworKIN tool (www.networkin.info) is well described for use towards predicting kinase-substrate relationships (Linding *et al.*, 2007). The tool has been previously used to map candidate kinases targeting phosphosites on known proteins identified from LC-MS/MS-based phosphoproteomics analyses (Durnberger *et al.*, 2014; Hoffman *et al.*, 2015; Karayel *et al.*, 2018). The NetworKIN tool integrates the NetPhorest database which comprises information on various kinase-specific and phospho-binding domain-specific motifs (Linding *et al.*, 2007; Miller *et al.*, 2008). We used the NetworKIN (Version 3.0) platform to identify candidate kinases upstream of the phosphosites identified in our global phosphoproteomics analyses. All upregulated phosphosites [Log2 (SRMS/control) ≥ 1.584] identified in the SRMS-regulated phosphoproteome were used for analyses. The resulting predicted annotations were filtered using a NetworKIN confidence score cut-off of 3.0 and a NetworKIN score-difference cut-off of 4.0. The score-difference defines the maximum difference between the best prediction and the second-best prediction. The NetworKIN data was presented in an Excel file. The identified kinases were plotted on a mammalian kinome dendogram generated using the KinMap tool (www.kinhub.org) (Eid *et al.*, 2017).

4.0 Results

4.1 Identification of novel candidate substrates of SRMS

The results described in this section have been published in the journal Molecular and Cellular Proteomics (MCP). As part of the "License to Publish" agreement between the authors and the Journal, ownership of the copyright remains with the authors. The authors reserve the right to use all or part of the published work for non-commercial purposes including theses/dissertations. As per the "License to Publish" agreement, I am required to cite the publication, as below, for use in this thesis.

This research was originally published in *Molecular and Cellular Proteomics*.

Raghuveera Kumar Goel, Marta Paczkowska, Jüri Reimand, Scott Napper and Kiven Erique Lukong. Phosphoproteomics analysis identifies candidate substrates of the non-receptor tyrosine kinase, SRMS. Molecular and Cellular Proteomics, 2018, Paper in Press. © Goel et al.

NOTE: The Supplementary Tables (Supplementary Tables 1-8) referenced in this section as well as legends to these tables can be accessed using the following weblink, as mentioned in Appendix A:

http://www.mcponline.org/content/early/2018/03/01/mcp.RA118.000643/suppl/DC1

4.1.1 Summary

SRMS (Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites), also known as PTK 70 (Protein tyrosine kinase 70), is a non-receptor tyrosine kinase that belongs to the BRK family of kinases (BFKs). To date little is known about the cellular role of SRMS primarily due to the unidentified substrates or signaling intermediates regulated by the kinase. In this study, we used phosphotyrosine antibody-based immunoaffinity purification in large-scale label-free quantitative phosphoproteomics to identify novel candidate substrates of SRMS. Our analyses led to the identification of 1258 tyrosine-phosphorylated peptides which mapped to 663 phosphoproteins, exclusively from SRMS-expressing cells. DOK1, a previously characterized SRMS substrate, was also identified in our analyses. Functional enrichment analyses revealed that the candidate SRMS substrates were enriched in various biological processes including protein ubiquitination, mitotic cell cycle, energy metabolism and RNA processing, as well as Wnt and TNF signalling. Analyses of the sequence surrounding the phosphosites in these

proteins revealed novel candidate SRMS consensus substrate motifs. We utilized customized high-throughput peptide arrays to validate a subset of the candidate SRMS substrates identified in our MS-based analyses. Finally, we independently validated Vimentin and Sam68 as bona fide SRMS substrates through *in vitro* and *in vivo* assays. Overall, our study identified a number of novel and biologically relevant SRMS candidate substrates, which suggests the involvement of the kinase in a vast array of unexplored cellular functions.

4.1.2 Introduction

Though non-receptor tyrosine kinases constitute only about 6% of the total kinases encoded by the human genome (Manning *et al.*, 2002), these kinases play physiologically significant roles linked to mammalian growth and development (Casaletto and McClatchey, 2012; Nurmio *et al.*, 2011). Like serine/threonine kinases, tyrosine kinases phosphorylate various substrate proteins to regulate specific intracellular signaling pathways which ultimately elicit specific cellular and physiological functions (Hunter, 2009). The catalytic activity of tyrosine kinases is highly regulated in eukaryotic cells and involve biochemically distinct mechanisms of autoregulation (Hubbard, 2002; Hubbard *et al.*, 1998; Nolen *et al.*, 2004). Further, these kinases are known to exhibit distinct biochemical substrate-motif specificities which are believed to impart specificity to the cellular functions regulated by these kinases (Amanchy *et al.*, 2011; Mok *et al.*, 2010; Ubersax and Ferrell, 2007). To date, high-throughput phosphoproteomics approaches have been applied to identify the cellular substrates of very few non-receptor tyrosine kinases (Amanchy *et al.*, 2009; Amanchy *et al.*, 2008; Xue *et al.*, 2013). Thus, key questions remain unanswered regarding the cellular roles and biochemical specificities of other non-receptor tyrosine kinases. We previously noted that the expression of wild type SRMS in HEK293 cells induced the tyrosine phosphorylation of several endogenous proteins (Goel *et al.*, 2013). These proteins represent the potential cellular substrates of SRMS and are largely unidentified. The identity of these candidate cellular substrates of SRMS would be essential for a better understanding of the role of SRMS in mammalian cell biology.

Mass spectrometry-based interrogation of the cellular phosphoproteome has afforded high-throughput and robust identification of various serine/threonine and tyrosine kinase substrates and associated signaling intermediates (Amanchy *et al.*, 2009; Amanchy *et al.*, 2008; Bian *et al.*, 2013; Courcelles *et al.*, 2013; Hsu *et al.*, 2011; Xue *et al.*, 2013). Over the years, the technology has

been used to precisely quantify multiple phosphorylation events occurring dynamically in the cellular phosphoproteome (Larance and Lamond, 2015; Roux and Thibault, 2013).

In the present study, we quantitatively probed the tyrosine-phosphoproteome of HEK293 cells upon the exogenous introduction of wild type SRMS and identified 663 candidate substrates of the kinase. Motif analyses revealed novel SRMS substrate consensus sequences among the candidate SRMS substrates. We used customized high-throughput peptide arrays and validated a subset of the candidate SRMS substrates. We finally independently validated Vimentin and Sam68, as bona fide SRMS substrates via immunoprecipitation analyses as well as direct *in vitro* kinase assays.

4.1.3 Results

4.1.3.1 SRMS induces kinase-dependent tyrosine phosphorylation of cellular proteins

We previously reported wild type SRMS as a catalytically active tyrosine kinase whose enzymatic activity is primarily regulated by the first 50 amino acids in the N-terminal region of the protein (Goel *et al.*, 2013). We observed that the exogenous expression of wild type SRMS and not ΔN-SRMS (N-terminal deletion) or SRMS-K258M mutants resulted in increased tyrosine phosphorylation of cellular proteins (Goel *et al.*, 2013). We confirmed this finding by ectopically expressing GFP-tagged constructs encoding either wild type SRMS, ΔN-SRMS or SRMS-K258M in HEK293 cells (Figure 4.1A). We used general phosphotyrosine antibodies to probe the levels of tyrosine-phosphorylated proteins in the whole cell lysates derived from these cells. Our results confirmed that the ectopic expression of wild type SRMS but not ΔN-SRMS or SRMS-K258M lead to the tyrosine-phosphorylation of several endogenous proteins in these cells (Figure 4.1B). The lysates were also probed with antibodies against GFP to ensure that the SRMS variants were expressed at similar levels in these cells (Figure 4.1B). The tyrosine-phosphorylated proteins observed in wild type SRMS-expressing cells represent the potential substrates of SRMS. We previously reported DOK1 as one of the proteins that is tyrosine phosphorylated in SRMS-overexpressing HEK293 cells (Goel *et al.*, 2013). However, the identity of the other tyrosine-phosphorylated proteins in these cells and the specific tyrosine residues potentially targeted by SRMS remains unknown.

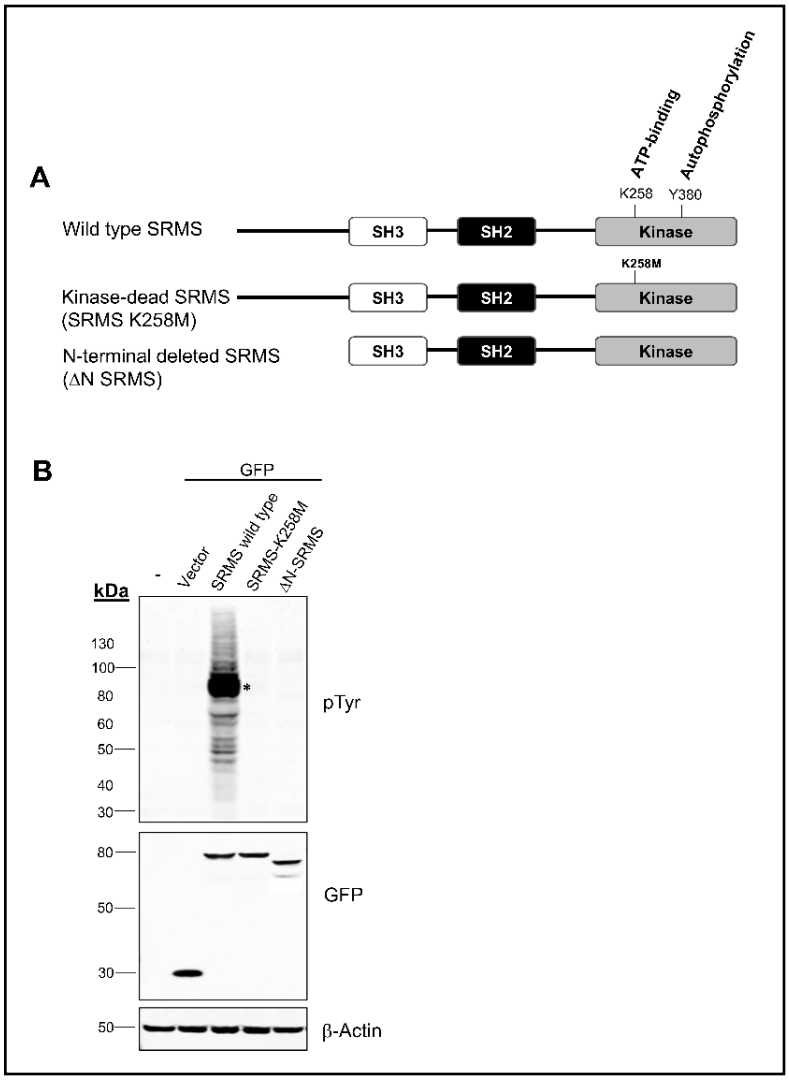


Figure 4.1: SRMS induces the tyrosine phosphorylation of several cellular proteins. **A.** Shown here are the domain-structure representations of the wild type SRMS protein, the kinase-dead SRMS mutant (K258M) and the N-terminal-deleted SRMS mutant (ΔN-SRMS) which lacks the first 50 amino acids in the N-terminal region of SRMS. Y380 represents the highly conserved autophosphorylation site within the activation loop of the SRMS kinase domain and K258 represents the conserved ATP-contacting residue in the ATP-binding pocket of the SRMS kinase domain. **B.** GFP-tagged SRMS constructs, namely, wild type SRMS, kinase-dead SRMS (SRMS-K258M) and ΔN-SRMS were transiently overexpressed in HEK293 cells. Untransfected HEK293 cells or HEK293 cells transfected with the empty vector encoding GFP alone were used as controls. Post-transfection, the cells were lysed and the lysates probed with antibodies against total tyrosine phosphorylated residues (pTyr), GFP and β-actin. Immunoblotting for β-actin was used as a control for protein loading. The asterisk ("*") represents the approximate position of autophosphorylated GFP-wild type SRMS.

4.1.3.2 Phosphoproteomics workflow to enrich tyrosine-phosphorylated targets by immuno-affinity purification

In order to identify the cellular substrates of SRMS, we performed a label-free quantitative phosphoproteomics approach to profile tyrosine-phosphorylated proteins in HEK293 cells ectopically expressing enzymatically active wild type SRMS (Figure 4.2). We transiently transfected either GFP-tagged wild type SRMS or the empty vector as a control, in HEK293 cells in a large-scale experiment (Figure 4.2A). We used the empty vector-expressing cells as a control to represent all steady state tyrosine phosphorylation events occurring natively in cells. A total of twelve 10 cm dishes were transfected with either GFP-wild type SRMS or the control vector backbone. Transfection efficiencies were determined to be over 90% (Figure 4.2A). We harvested the cells directly in a strong denaturing lysis buffer and performed in-solution trypsin digestion of total proteins (Figure 4.2B and 4.2C). The crude tryptic peptides were desalted via reversed-phase solid-phase extraction, and anti-phosphotyrosine antibodies were used to enrich for tyrosine phosphorylated peptides from cells expressing either GFP-SRMS or the empty vector control (Figure 4.2C and 4.2D). The enriched tyrosine-phosphorylated peptides from both populations of cells were then sequenced via LC-MS/MS (Figure 4.2E) and protein identification and quantitation performed using the Andromeda and MaxQuant tools, respectively (Figure 4.2F).

4.1.3.3 Identification of candidate SRMS substrates

Using the MaxQuant-Andromeda integrated platform and the Perseus quantitative analyses framework, we applied quality thresholds (Cox and Mann, 2008; Cox *et al.*, 2011; Tyanova *et al.*, 2016a; Tyanova *et al.*, 2016b) to allow the identification of high-confidence phosphopeptides (Figure 4.3A- 4.3D). The search parameters on MaxQuant-Andromeda were adjusted using 1% FDR and quality Andromeda search and Delta scores were used as minimum thresholds (Figure 4.3A and 4.3B, Supplementary Tables 1 and 2). Using Pearson's correlation analysis, we observed good reproducibility of the phosphopeptides quantified from both biological replicates of each sample ($R \geq 0.8$) (Figure 4.4A and 4.4B). We finally based subsequent analyses exclusively on class I phosphosites that matched a reasonably stringent t-test-based statistical threshold (p-value < 0.05, Student's t-test) (Figure 4.3C and 4.3D, Supplementary Table 3). Using these criteria, our analyses led to the identification of a total of 1450 tyrosine- phosphorylated peptides pooled from both, SRMS-expressing and control cells (Figure 4.5A and Supplementary Table 3). Of these,

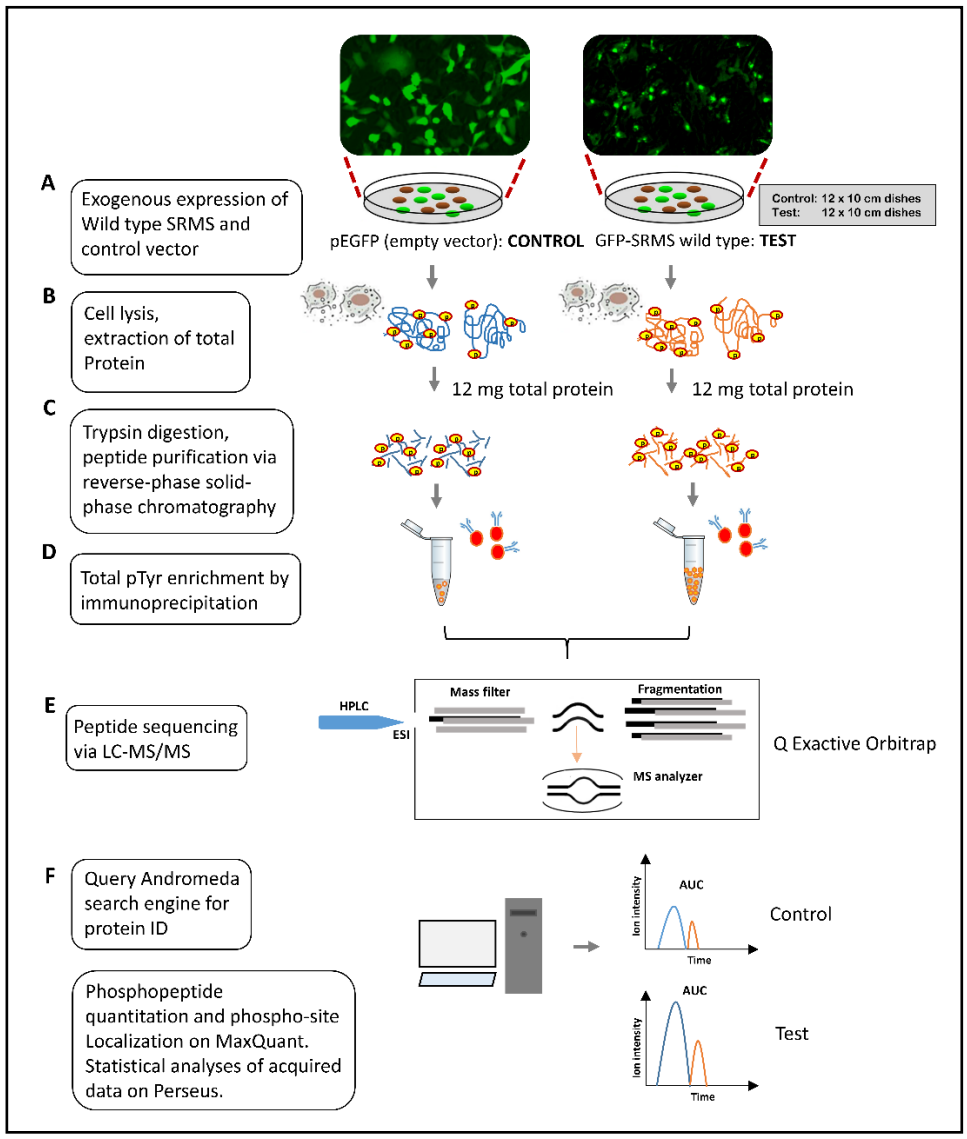


Figure 4.2: Experimental workflow. A label-free quantitation-based mass spectrometry approach was adopted to identify the candidate cellular substrates of SRMS. **A.** Plasmid constructs encoding either GFP-tagged wild type SRMS or GFP alone (empty vector) were exogenously introduced in HEK293 cells. **B.** 24 hours post-transfection the cells were lysed under denaturing conditions. **C.** 12 mg total protein, each from GFP alone and GFP-SRMS conditions, was trypsin-digested and the crude tryptic peptides were desalted via reversed-phase solid-phase chromatography. **D.** The desalted peptides were then used towards immunoprecipitation of total tyrosine-phosphorylated peptides. **E.** The enriched tyrosine-phosphorylated peptides from each condition were sequenced via LC-MS/MS. All fractions were separated using a 2-hour a reversed-phase gradient and electro-sprayed into a Q Exactive HF quadrupole instrument. **F.** Protein IDs and phosphosite localization were determined using the MaxQuant platform which integrates the Andromeda search engine. The entire workflow was performed in 2 biological replicates.

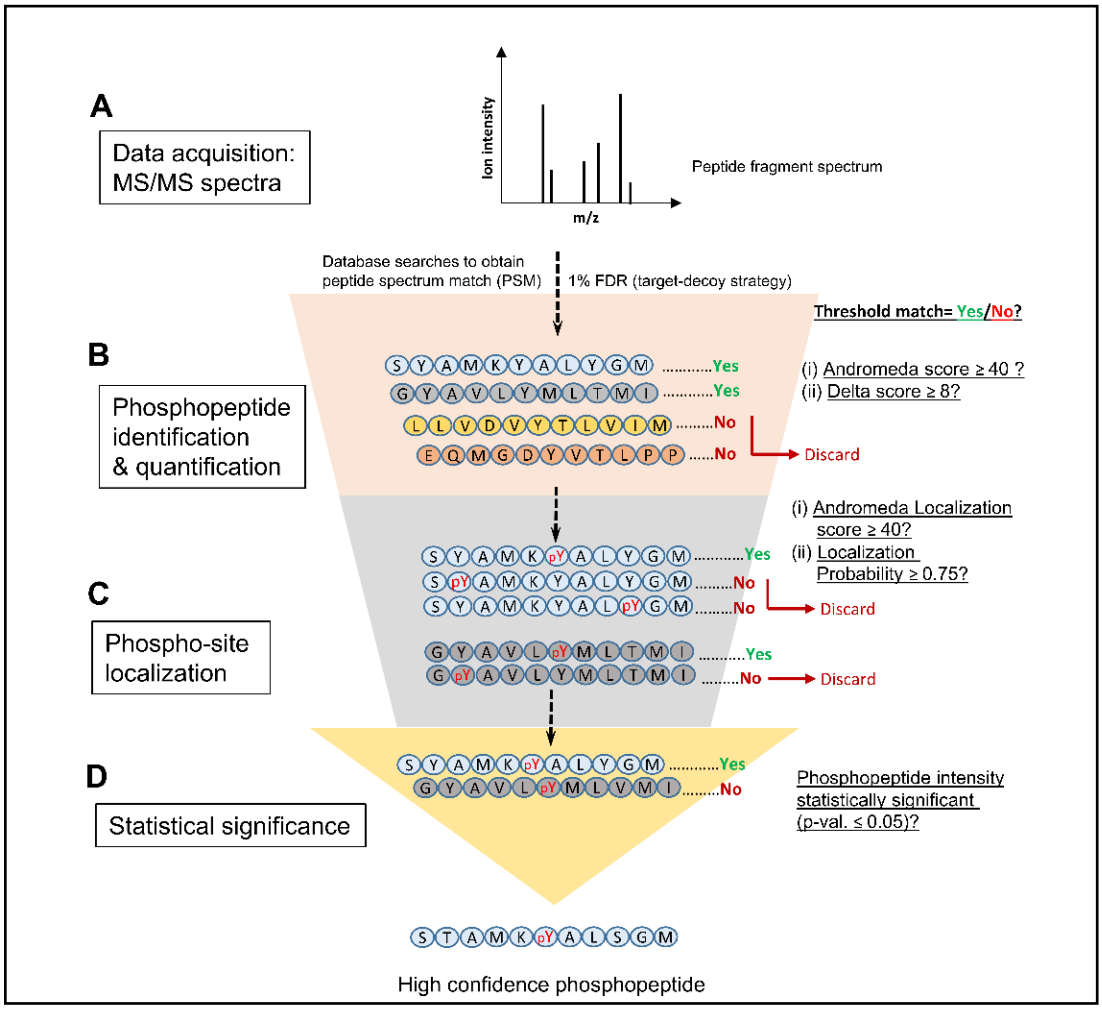


Figure 4.3: Flowchart depicting the computational pipeline used to identify high confidence phosphopeptides. A. The acquired raw instrument data were processed on the MaxQuant-Andromeda platform that integrates the Perseus software for data quantitation. Each peptide spectrum obtained was queried using the Andromeda search engine to generate reliable peptide-spectrum matches (PSM) against the queried database. Peptide sequences and protein IDs were assigned at 1% False Discovery Rate (FDR) based on target-decoy searches. **B.** Phosphopeptides were quantified based on their precursor ion intensities, calculated as the area of the associated peak or area-under-the curve (AUC). Phosphopeptides with an associated "Andromeda score" < 40.0 and Delta score < 8 were discarded. **C.** Class I phosphosites were selected with an "Andromeda "localization score" and "localization probability" threshold of 40.0 and 0.75, respectively. **D.** Finally, a statistical t-test threshold was applied to the identified phosphopeptides. Only phosphopeptides with corresponding precursor ion intensities that were statistically significant, with a p-value ≤ 0.05 between both biological replicates, were retained. "Threshold match= Yes" indicates that the phosphopeptide(s) matched the threshold for the above parameters and were accepted. "Threshold match=No" indicates that the phosphopeptides failed to match the threshold for the parameters and were discarded.

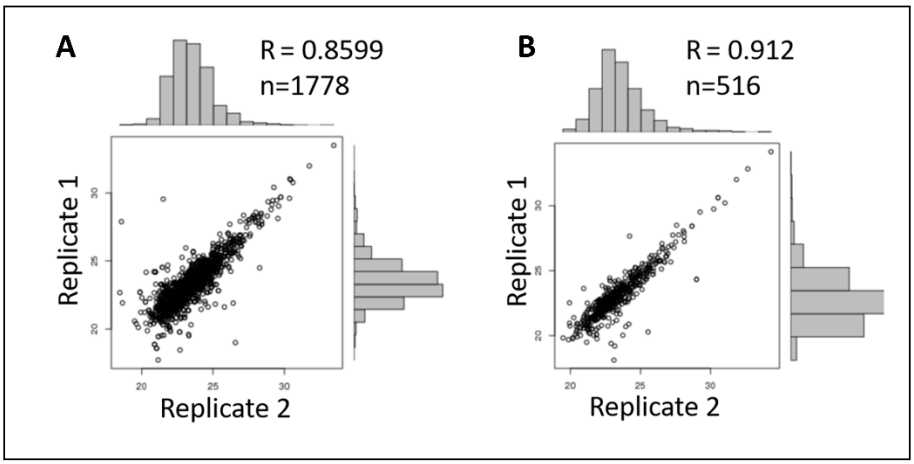


Figure 4.4: Reproducibility of phosphopeptide identifications: Pearson's correlation coefficients (R score) were determined for Log_2-transformed phosphopeptides (matching the MaxQuant-Andromeda search parameters and criteria) quantified from both replicates of **A.** SRMS-expressing cells (n= 1778 phosphopeptides) and **B.** control (empty vector-expressing) cells (n=516 phosphopeptides).

1258 peptides were unique and mapped to 771 distinct proteins (Figure 4.5A and Supplementary Table 3). A total of 141 peptides were found to be conserved in more than one protein and such proteins were consequently clustered in the respective protein groups by MaxQuant (Supplementary Table 3). As expected, we identified a far greater number of tyrosine-phosphorylated peptides in SRMS-overexpressing cells compared to control cells (Figure 4.5B, left panel). This observation was consistent with our findings from Figure 4.1B as well as additional immunoblotting analyses where we used general phosphotyrosine antibodies to assess total tyrosine-phosphorylated proteins in the same pool of lysates which were used towards mass spectrometry analyses (Figure 4B, right panel). While SRMS-overexpressing cells readily exhibited a significant number of tyrosine-phosphorylated proteins, longer exposures of the immunoblot revealed a fair but comparatively lower number of tyrosine-phosphorylated proteins in the control cells as well (Figure 4.6). This is expected since tyrosine phosphorylation occurs natively in cells although at a far lower biological stoichiometry compared to serine/threonine phosphorylation events (Conrads and Veenstra, 2005; Hunter, 1998).

Of the 1258 unique tyrosine-phosphorylated peptides identified between both samples, 1129 phosphopeptides were identified exclusively in SRMS-expressing cells whereas only 117 phosphopeptides were enriched in the control cells (Figure 4.5B, left panel and Supplementary Table 3). Additionally, 99 phosphopeptides were found common to both SRMS-expressing and

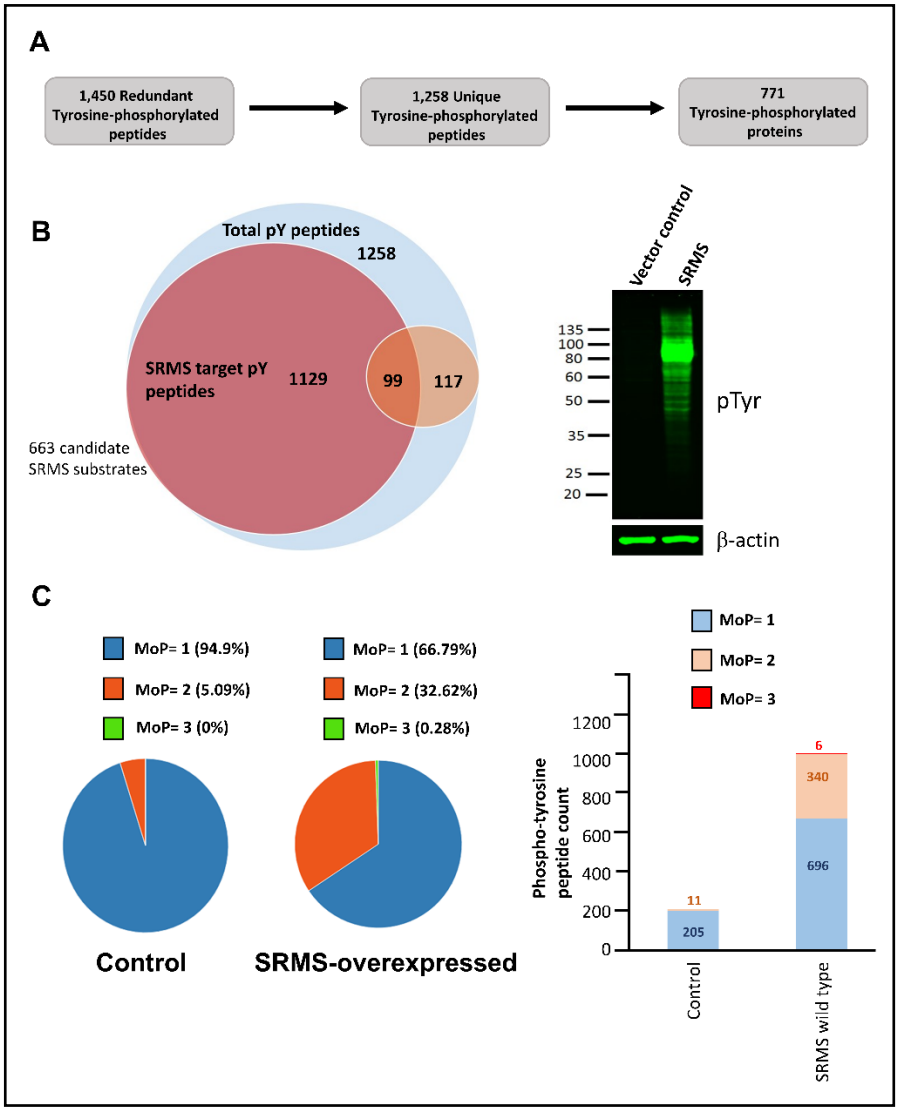


Figure 4.5: Overview of the identified tyrosine-phosphoproteome. A. The MaxQuant-Andromeda platform quantified a total of 1450 redundant tyrosine-phosphorylated peptides from both control and SRMS-overexpressing cells. Of these, 1258 tyrosine-phosphorylated peptides were unique and mapped to 771 unique proteins. **B.** A Venn diagram depicts the number of tyrosine-phosphorylated peptides either unique to SRMS-overexpressing cells, control cells or common to both. Of the 1258 unique tyrosine-phosphorylated peptides identified globally, 1129 peptides were enriched exclusively in SRMS-overexpressing cells while 117 peptides were identified exclusively in control cells. 99 peptides were found common to both samples (left panel). The 1129 phosphopeptides identified exclusively in SRMS-expressing cells mapped to 663 proteins. Prior to trypsinization, 25 ug whole cell lysates derived from large-scale transfections of HEK293 cells, were probed with general phosphotyrosine antibodies via immunoblotting analyses to confirm the expression of tyrosine-phosphorylated proteins resulting from SRMS kinase activity (right panel). **C.** The tyrosine-phosphorylated peptides were segregated based on multiplicity of phosphorylation (MoP). Singly- (MoP=1), doubly- (MoP=2) or triply- (MoP=3)-phosphorylated peptides were quantified and represented either as percent-abundance of total tyrosine-phosphorylated peptides in each sample (pie-chart, left panel) or in absolute numbers (right panel).

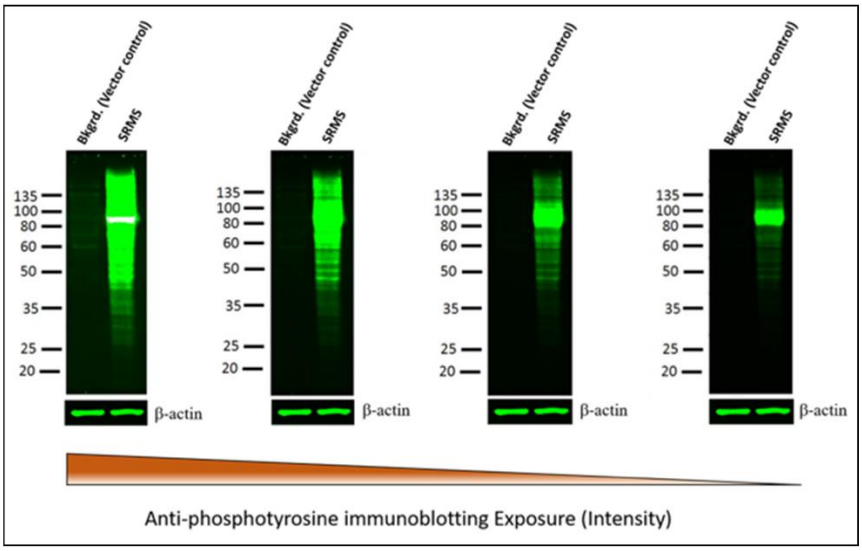


Figure 4.6: SRMS-induced tyrosine phosphorylation of endogenous proteins. cDNA encoding either GFP alone or GFP-tagged wild type SRMS were exogenously expressed in HEK293 cells in large-scale (12 x 10 m dishes each) and lysates harvested for use towards LC-MS analyses. 25 µg total protein from each condition was used in quantitative immunoblotting analyses using general phosphotyrosine antibodies. Antibodies against β-actin were used to detect total β-actin levels in the lysates and used as loading control. Near-infrared fluorescent dye-conjugated secondary antibodies (800 nm) were used in immunoblotting analyses. The phosphotyrosine immunoblots were scanned at various exposure levels to detect the expression of both, high and low-abundance tyrosine-phosphorylated proteins in the samples.

control cells (Figure 4.5B, right panel and Supplementary Table 3**)**. We considered the tyrosine-phosphorylated targets as candidate substrates of SRMS only if the corresponding phosphopeptides were exclusively enriched in the SRMS overexpressing cells at a p-value threshold of 0.05 between the replicates. All tyrosine-phosphorylated peptides identified in the control cells were considered background phosphorylation events. Since none of the phosphopeptides identified as common to SRMS-overexpressing and control cells displayed at least two-fold differential abundance in phosphopeptide intensities, we reasoned that proteins mapped from these peptides would be less likely to represent candidate SRMS substrates and were therefore filtered. Using this approach, we identified 663 candidate SRMS substrates across two biological replicates with a p-value <0.05 (Figure 4.5B, left panel and Supplementary Table 3). A subset of the SRMS candidate substrates is listed in Table 4.1. Importantly, our phosphoproteomics analyses identified DOK1 as a hyperphosphorylated target in wild type SRMS-expressing cells. This was an important observation since we previously validated DOK1 as a direct substrate of SRMS and showed that the exogenous expression of SRMS led to the

phosphorylation of endogenous DOK1 in HEK293 cells (Goel *et al.*, 2013). Our MS analyses therefore validated our previous findings. The SRMS Y380 residue represents a key autophosphorylation site within the kinase domain activation loop and is essential for SRMS

Table 4.1: List of candidate SRMS substrates identified using quantitative phosphoproteomics. List of top-ranked candidate SRMS substrates identified using the label-free quantitative phosphoproteomics approach in LC-MS/MS analyses. The candidate substrates are ranked based on associated mean log_2 phosphorylation intensities from both biological replicates. The gene name, protein name and the corresponding phosphosite(s) identified in the proteins are listed. The mean log_2 intensities for each phosphosite are indicated in parenthesis. Phosphosites identified in SRMS are indicated. Although not identified as a top-ranked candidate substrate, KHDRBS1/Sam68, an important Src and BRK substrate is also listed in the table. Refer to Supplementary Table 3 using the journal's weblink provided (Appendix A), to access the complete list of candidate SRMS substrates.

S. no.	Gene names	Protein names	Amino Acid Position (Mean log2 Intensity)
1	SRMS	Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites	87 (28); 121 (29); 156 (29); 169 (30); 180 (30); 196 (28); 197 (25); 299 (28); 340 (25); 380 (25); 422 (24); 427 (24); 443 (28); 456 (28)
2	VIM	Vimentin	61 (28); 383 (26)
3	DNAJB6	DnaJ homolog subfamily B member 6	4 (25); 5 (25); 23 (26); 53 (22); 65 (28); 68 (28)
4	LDHA	L-lactate dehydrogenase A chain	83 (27)
5	C8orf33	UPF0488 protein C8orf33	186 (27)
6	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	71 (23); 681 (27)
7	ABCC3	Canalicular multi-specific organic anion transporter 2	233 (27)
8	CKAP2	Cytoskeleton-associated protein 2	592 (25); 599 (25); 677 (27); 678 (27)
9	EIF4G2	Eukaryotic translation initiation factor 4 gamma 2	128 (21); 185 (27)
10	EIF2S3	Eukaryotic translation initiation factor 2 subunit 3	89 (27); 102 (27)
11	RPS3	40S ribosomal protein S3	34 (24); 87 (27); 166 (27); 167 (27)
12	DNAJB1	DnaJ homolog subfamily B member 1	5 (26); 6 (26); 67 (25); 176 (25)
13	EIF4A3	Eukaryotic initiation factor 4A-III	48 (26); 50 (26); 155 (25)
14	PSMA7	Proteasome subunit alpha type-7	71 (23); 79 (23); 102 (26)
15	PHB2	Prohibitin-2	34 (25); 81 (23); 121 (26)
16	DSTN	Destrin	82 (26); 85 (26)
17	MAGED1	Melanoma-associated antigen D1	92 (24); 126 (23); 161 (26); 487 (23); 505 (26); 509 (25); 609 (26)
18	RBM4	RNA-binding protein 4	25 (24); 37 (26); 101 (24); 113 (24); 190 (23); 194 (23); 345 (23)
19	DDX3X	ATP-dependent RNA helicase DDX3X	200 (26); 243 (23); 602 (22)
20	BUB3	Mitotic checkpoint protein BUB3	141 (26)
21	KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1	167 (22)

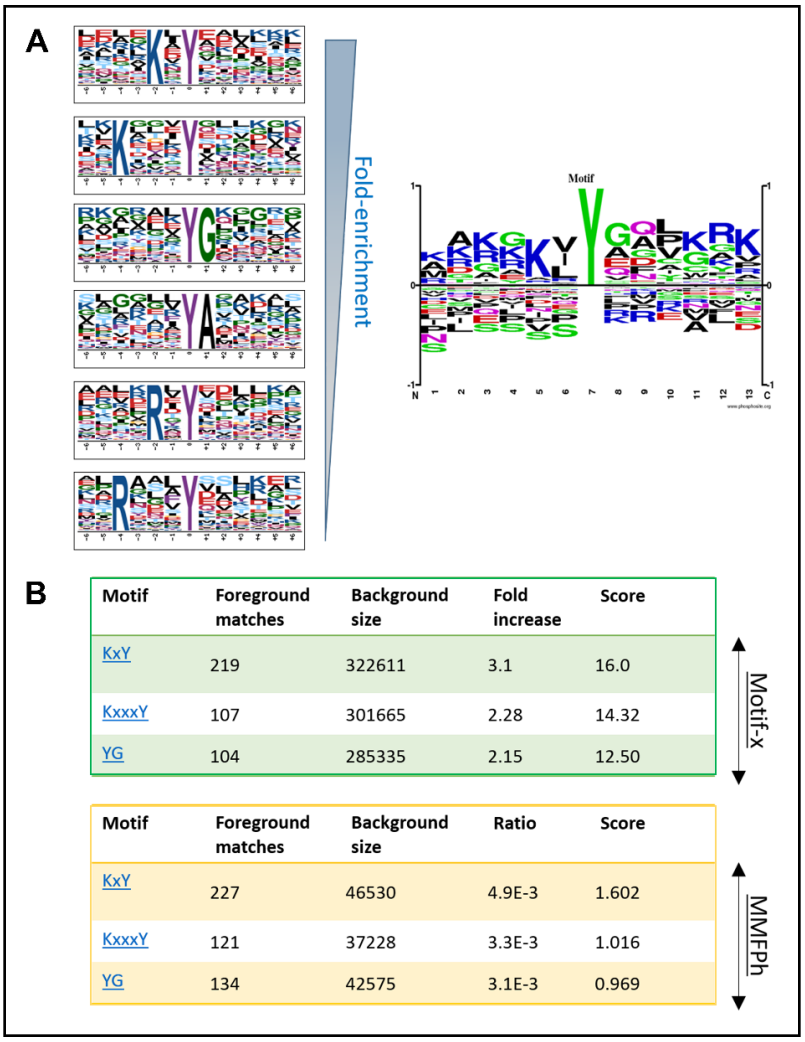
enzymatic activity (Goel *et al.*, 2013). As expected, the peptide corresponding to SRMS Y380 was identified as one of the major phosphopeptides from our MS analyses. Additionally, we identified 14 other tyrosine- phosphorylated sites in SRMS that were not previously known (Table 4.1).

We also assessed the multiplicity of phosphorylation associated with tyrosine residues in the identified tyrosine-phosphorylated peptides. We found that the majority of the tyrosine-phosphorylated peptides, identified between both samples, were singly phosphorylated (Figure 4.5C). However, the singly-phosphorylated peptides were significantly more abundant (>3-fold) in SRMS-expressing cells compared to control cells (Figure 4.5C). Interestingly, doubly-phosphorylated peptides were over 30-fold more abundant in SRMS-expressing cells compared to control cells (Figure 4.5C). Additionally, though detected only as a minor proportion, triply-phosphorylated peptides were identified in SRMS-expressing cells but not in the control cells (Figure 4.5C). The greater multiplicity of phosphorylation observed on individual peptides in SRMS-expressing cells, indicates that SRMS likely phosphorylates its substrates on several tyrosine residues.

4.1.3.4 *In silico* analyses identifies candidate SRMS substrate motifs

As described above, our analyses revealed 1129 non-redundant tyrosine-phosphorylated peptides in SRMS-overexpressing cells. We were interested in identifying candidate SRMS substrate motifs to understand the pattern of substrate-specificity exhibited by the SRMS kinase domain. Motif-specificities exhibited by tyrosine kinase domains are less understood compared to the other well-characterized modular domains such as the SH3, SH2 and PTB domains, primarily since the cellular substrates of several of these kinases are unknown. To identify the consensus sequence(s) of SRMS substrates, we used the entire pool of unique phosphopeptides (n=1129) derived from the 663 candidate SRMS substrates, enriched exclusively from SRMS-overexpressing cells. We performed motif-analyses using two well described and popular *in silico* tools, namely, Motif-x (Chou and Schwartz, 2011) and MMFPh (Maximal motif finder for phosphopeptides) (Wang *et al.*, 2012). The active site of kinases is known to recognize residues that are immediately N-terminal and/or C-terminal to the phospho-acceptor site, implying that consensus sequences are generally short (Ubersax and Ferrell, 2007). We therefore refined the

length of our query phosphopeptides to 13 amino acids. Using Motif-x, we identified six major classes of SRMS substrate consensus motifs at a significance threshold of 10^{-6} (Figure 4.7A and

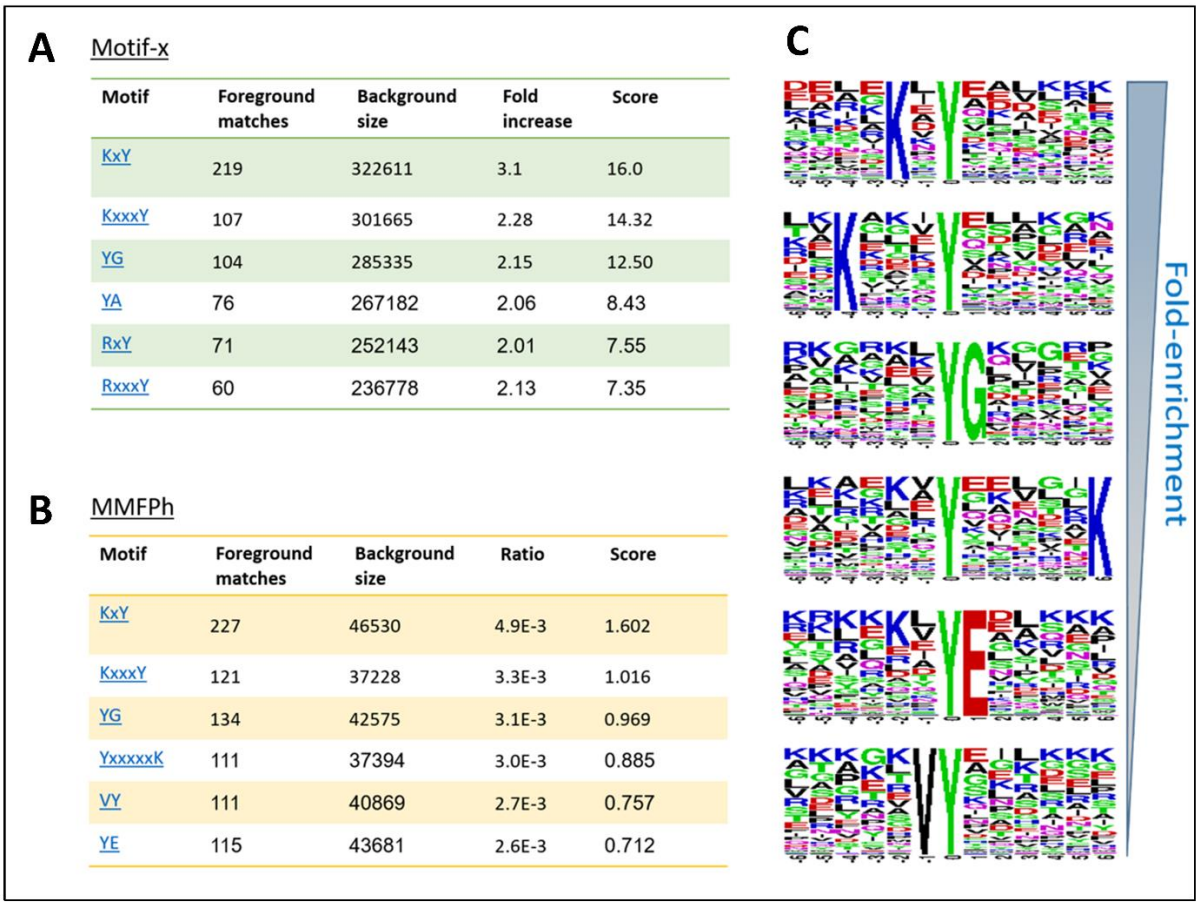


Motif	Foreground matches	Background size	Fold increase	Score
KxY	219	322611	3.1	16.0
KxxxY	107	301665	2.28	14.32
YG	104	285335	2.15	12.50

Motif	Foreground matches	Background size	Ratio	Score
KxY	227	46530	4.9E-3	1.602
KxxxY	121	37228	3.3E-3	1.016
YG	134	42575	3.1E-3	0.969

Figure 4.7: Top candidate SRMS consensus motifs identified by Motif-x and MMFPh. A. The *in silico* motif analyses tool motif-x (Chou and Schwartz, 2011) was used to assess overrepresented motif patterns in tyrosine-phosphorylated peptides (n=1129) derived from 663 candidate SRMS substrates. Six candidate consensus motifs are represented in individual motif logos and in the order of their fold-enrichment (left panel). A representative motif logo of all candidate SRMS consensus motifs was generated using the Phosphosite motif analyses tool (http://www.phosphosite.org/sequenceLogoAction) (right panel). The size of the amino acid letter, in the motif logo (left and right panels), is proportional to the magnitude of overrepresentation of this residue in the pool of queried peptides. **B**. The tables represent statistical information on the top three SRMS candidate consensus motifs identified using either the Motif-x (Chou and Schwartz, 2011) or MMFPh (Wang *et al.*, 2012) tools. Foreground matches indicate the number of queried peptides that correspond to the identified motif. Background size indicates the background pool of phosphopeptides used as reference. The algorithmically determined scores associated with each motif are indicated.

Supplementary Table 4). Interestingly, the top three consensus sequences identified by Motif-x were also reported as overrepresented by the other *in silico* tool, MMFPh (Figure 4.7B). Additionally, both Motif-x and MMFPh identified 3 unique consensus motifs among the phosphosites derived from candidate SRMS substrates (Figure 4.8A- 4.8C and Supplementary Table 4).



Motif	Foreground matches	Background size	Fold increase	Score
KxY	219	322611	3.1	16.0
KxxxY	107	301665	2.28	14.32
YG	104	285335	2.15	12.50
YA	76	267182	2.06	8.43
RxY	71	252143	2.01	7.55
RxxxY	60	236778	2.13	7.35

Motif	Foreground matches	Background size	Ratio	Score
KxY	227	46530	4.9E-3	1.602
KxxxY	121	37228	3.3E-3	1.016
YG	134	42575	3.1E-3	0.969
YxxxxxK	111	37394	3.0E-3	0.885
VY	111	40869	2.7E-3	0.757
YE	115	43681	2.6E-3	0.712

Figure 4.8: All candidate SRMS consensus motifs identified by Motif-x and MMFPh. Tyrosine-phosphorylated peptides derived from 662 candidate SRMS substrates were queried using two *in silico* motif analyses tools, namely, **A.** Motif-x (Chou and Schwartz, 2011) and **B.** MMFPh (Wang *et al.*, 2012). All identified motifs and the associated statistical information is presented. Foreground matches indicate the number of queried peptides that correspond to the identified motif. Background size indicates the background pool of phosphopeptides used as reference. The algorithmically determined scores associated with each motif is indicated. **C.** Six candidate SRMS substrate consensus motifs, identified as significantly enriched by the MMFPh tool, are represented in individual motif logos. The size of the amino acid letter, in the motif logo is proportional to the overrepresentation of this residue in the pool of queried peptides.

Overall, motif-analyses using these two programs revealed that top SRMS candidate consensus sequences predominantly displayed a lysine residue at either the -2 (*KxY*, where Y represents the phosphorylated tyrosine residue at position "0" and "x" is any amino acid residue) or -4 position (*KxxxY*) (Figure 4.7 and 4.8). Another overrepresented candidate consensus motif displayed a glycine at the +1 position (*YG*) (Figure 4.7 and 4.8). Other enriched motifs contained either a lysine residue at +6 (*YxxxxxK*), valine at -1 (*VY*) or a glutamic acid at +1 (*YE*) positions, respectively (Figure 4.8). We also examined the possibility that the more rigorous SRMS substrate consensus motifs would likely be represented by a subset of phosphopeptides enriched at significantly higher relative abundances. To address this, we performed motif-analyses using 86 phosphopeptides corresponding to candidate SRMS substrates, that were quantified with average log2 ion intensities of at least 24.0. Using the Motif-x tool, we consistently identified the *KxY* motif as the most overrepresented consensus sequence among the queried phosphopeptides, implying that this may likely represent one of the major SRMS kinase consensus motifs (Figure 4.9).

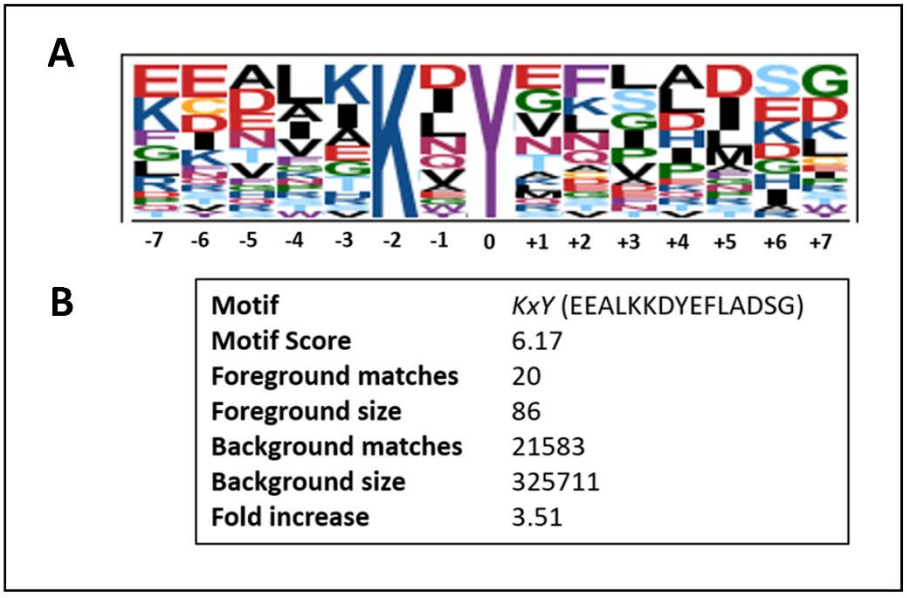


Figure 4.9: Analyses of candidate SRMS consensus motifs from phosphopeptides identified at higher relative abundance: A. The *in silico* motif analyses tool Motif-X (http://motif-x.med.harvard.edu/) was used to assess overrepresented motif patterns in 86 tyrosine-phosphorylated peptides corresponding to candidate SRMS substrates with LC-MS/MS ion intensities ≥ 24.0. The overrepresented motif, as identified by Motif-x, is shown in the motif logo. **B.** Statistical information relevant to the identified consensus motif is presented.

4.1.3.5 Functional gene enrichment analysis of candidate SRMS substrates

The cellular and physiological role of SRMS is poorly understood. To gain insights into the biological function of SRMS, we performed pathway enrichment analyses on the candidate SRMS substrates using the g:Profiler software (Reimand *et al.*, 2016). In order to complete a high-confidence pathway analysis and restrict our enrichment analysis solely to phosphoproteins, we used a custom background set of all reported phosphoproteins curated with the PhosphositePlus database (Hornbeck *et al.*, 2015). This increased the stringency at the level of the expected number of proteins associated with every Gene Ontology (GO)-annotated pathway identified. We further filtered low-confidence GO annotations of genes that had been only derived from automated computational analysis. Our analysis revealed 176 significantly enriched biological pathways and processes (FDR<0.05) that associated with a large fraction (411/663 proteins or 62%) of the candidate SRMS substrates (Figure 4.10 A and 4.10B and Supplementary Table 5). We constructed an enrichment map visualization (Merico *et al.*, 2010) and manually annotated subnetworks of similar pathways as functional themes (Figure 4.10B). We found that the top pathways and processes associated with the candidate SRMS substrates included themes such as proteolysis and protein ubiquitination (e.g., negative regulation of protein ubiquitination, FDR=$5x10^{-11}$, 31 proteins), mitotic cell cycle (e.g., cell cycle phase transition, FDR=$3x10^{-11}$, 66 proteins), cellular energy metabolism (e.g., ATP metabolic process, FDR=$4x10^{-8}$, 36 proteins) and RNA processing (e.g., RNA catabolic process, FDR=$8x10^{-12}$, 54 proteins) (Figure 4.10A and 4.10B). Several other interesting themes with fewer pathways were also apparent, such as regulation of stem cell differentiation (FDR=0.03, 16 proteins) and cellular response to hypoxia (FDR=0.004, 21 proteins). Two themes corresponding to established signal transduction pathways were also significantly enriched: Wnt signaling pathway (FDR=$9x10^{-4}$, 18 proteins) and tumor necrosis factor (TNF)-mediated signaling pathway (FDR=$4x10^{-3}$, 20 proteins) (Figure 4.10A and 4.10B). Overall, pathway enrichment analyses of our phosphoproteomics dataset revealed several distinct and related cellular processes potentially regulated by SRMS in a kinase-dependent manner.

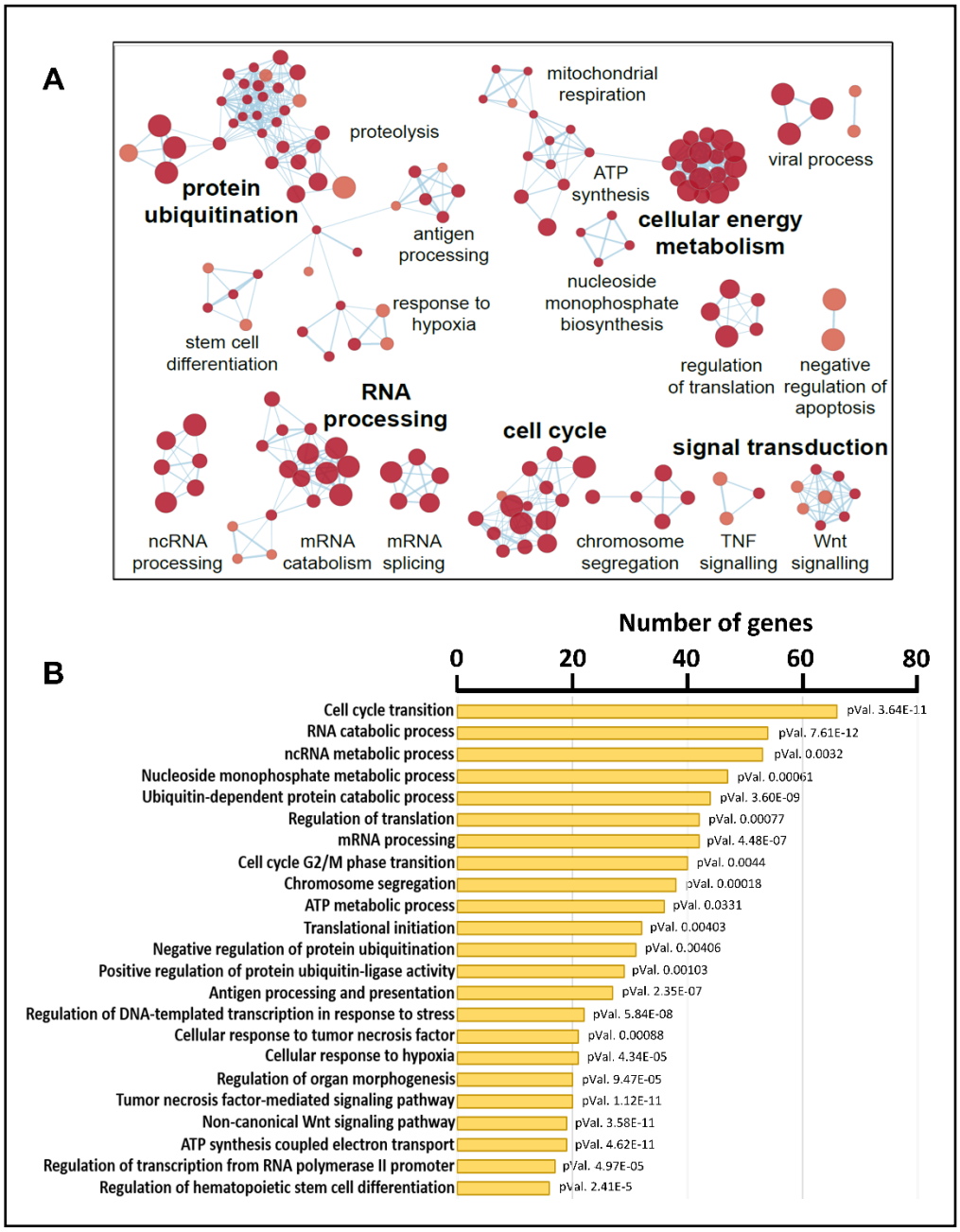


Figure 4.10: Functional gene enrichment analyses of candidate SRMS substrates. A. Enrichment map of significantly enriched cellular processes and pathways enriched by candidate SRMS substrates (FDR<0.05). Network nodes represent enriched pathways and edges connect pathways with many shared genes. Biological themes associated with groups of nodes represent a simplified summary of multiple related pathways and were assigned through manual curation. The size of each node is proportional to the number of genes involved in the given pathway or process. **B.** Barplot shows the most significant pathways and processes and number of candidate SRMS substrates enriched in each pathway/process. The p-value associated with each annotated process/pathway is indicated.

4.1.3.6 Validation of candidate SRMS substrates using peptide arrays

Our mass spectrometry-based phosphoproteomics analyses identified 663 candidate SRMS substrates (Figure 4.5B). We decided to utilize peptide microarrays in a complementary approach to perform high-throughput validation of a subset of the candidate SRMS substrates. Peptide microarrays are robust and versatile tools that have been previously used for screening kinase-substrates (Parikh *et al.*, 2009), identifying kinase substrate-motifs (Leung *et al.*, 2007) and investigating phosphoproteome-based signaling in eukaryotic cells (Diks *et al.*, 2004). We generated a customized peptide array comprising 102 unique peptides (Supplementary Table 6). Of these, 100 peptides were derived from the tyrosine-phosphorylated peptides identified in our MS analyses (Supplementary Table 6). These 100 peptides included 80 test peptides corresponding to 58 candidate SRMS substrates that were identified exclusively in SRMS-expressing cells at log2 intensities >20 and 20 negative control peptides that were selected randomly (Supplementary Table 6). Of these 20 negative control peptides, 9 were identified in our MS analyses from control cells alone and 11 were common to both SRMS-expressing and control cells (Supplementary Table 6). The remaining two are control peptide, a negative control peptide corresponding to Beclin1 Y352 that was previously reported as a direct phosphorylation site for EGFR but not SRMS (Wei *et al.*, 2013), and a peptide corresponding to the BRK Y447 residue since this was recently reported as a site that is directly phosphorylated by SRMS (Fan *et al.*, 2015). Each peptide was printed in three triplicate sets (9 peptide replicates in total) on each array to allow for intra-array normalization (Scholma *et al.*, 2016; Trost *et al.*, 2013) of the peptide phosphorylation intensities.

We used lysates derived from HEK293 cells overexpressing either GFP-SRMS wild type or GFP-kinase-dead SRMS to assess the phosphorylation status of the peptides on the microarrays. Data from our peptide microarray analyses showed that compared to arrays treated with kinase-dead SRMS, wild type SRMS-treated arrays displayed increased phosphorylation of 52 peptides derived from candidate 44 candidate SRMS substrates (fold change phosphorylation > 1) (Supplementary Table 7). Of these, 26 peptides, derived from 24 candidate substrates showed a SRMS kinase-dependent increase in tyrosine phosphorylation across all technical replicates (p-value = 0.003-0.09) (Supplementary Table 8). A subset of these validated targets is represented in Figure 4.11 and Table 2.

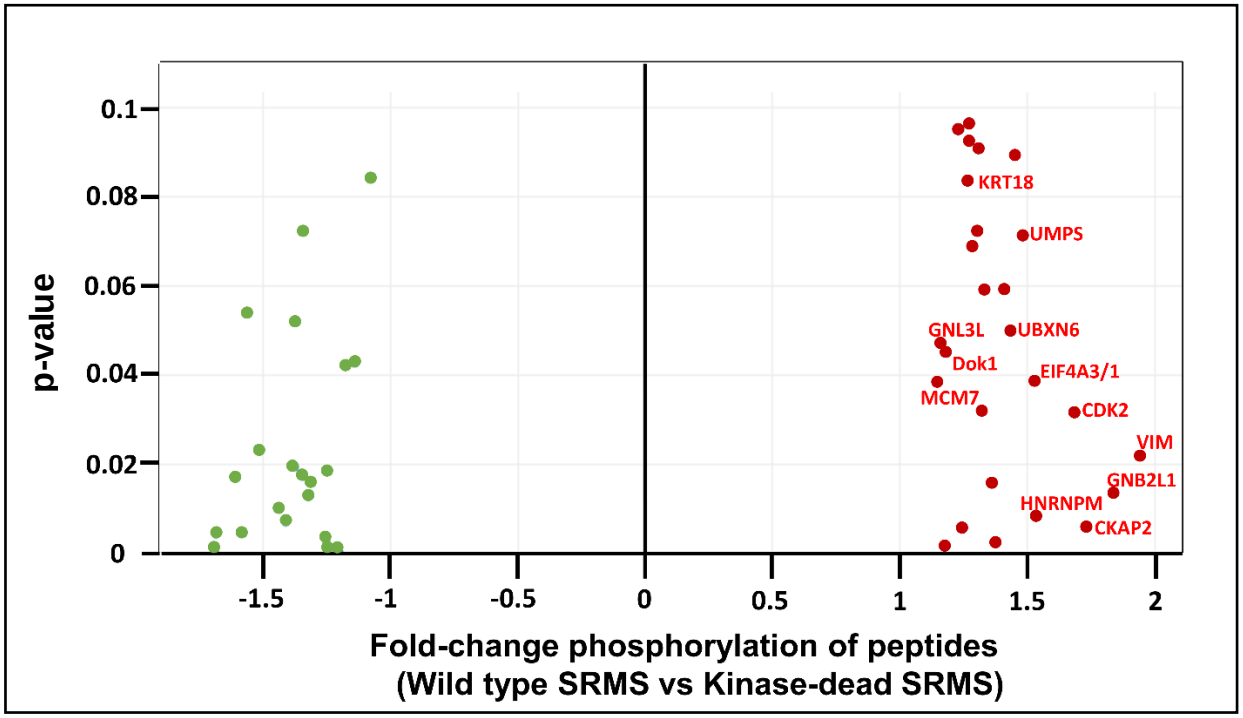


Figure 4.11: Validation of target peptides using peptide microarrays. A scatter plot of differentially phosphorylated peptides between wild type SRMS-treated and kinase-dead SRMS-treated conditions. Peptides highlighted in red represent hyperphosphorylated targets identified in arrays treated with wild type SRMS (p-value range 0.003-0.09). Peptides highlighted in green represent hypophosphorylated targets identified in arrays treated with wild type SRMS. Selected hyperphosphorylated peptides are labelled in the scatter plot.

As expected, none of the 21 control peptides, including Beclin1 Y352, showed increased phosphorylation in the wild type SRMS-treated arrays compared to kinase-dead SRMS-treated arrays (Supplementary Table 7 and 8). Importantly, we observed an increase in the phosphorylation of the peptide corresponding to DOK1 Y337, in wild type SRMS-treated arrays compared to kinase-dead SRMS-treated arrays (Table 4.2 and Supplementary Tables 7 and 8). This again corroborates our previous findings where we identified DOK1 as a *bona fide* SRMS substrate (Goel *et al.*, 2013). We also observed a modest increase in the phosphorylation of the BRK Y447 peptide in arrays treated with lysates of cells overexpressing wild type SRMS. The BRK Y447 peptide showed 1.2-fold increased phosphorylation in wild type SRMS-treated arrays compared to kinase-dead SRMS-treated arrays (Supplementary Tables 7 and 8). Additionally, we found that the top target peptides validated in our phosphorylation assays displayed motifs that matched the candidate consensus motifs for SRMS (Table 4.2), as determined through *in silico*

analyses (Figure 4.7, 4.8 and 4.9). These included motifs, *RxY* (VIM Y383 and MAGED1 Y505), *VY* (CDK2/3 Y19), *KxY* (UMPS Y326 and GRB10 Y354), *KxxY* (DOK1 Y337), *YA* (HNRNPM Y681 and RPS3 Y87) and *YG* (EIF4A3 Y50 and MCM7 Y33).

Table 4.2: List of top SRMS targets validated on peptide arrays. List of peptides that displayed increased phosphorylation on the central phospho-acceptor tyrosine residue (p-Tyr. site) upon incubation with lysates derived from HEK293 cells expressing ectopic wild type SRMS, compared to lysates of cells expressing the kinase-dead SRMS mutant. Fold-change phosphorylation of each peptide is presented as the ratio of phosphorylation intensity corresponding to wild type SRMS-treated versus kinase-dead SRMS-treated conditions. The statistical significance (P-value) of phosphorylation was determined by PIIKA2 (Trost *et al.*, 2013) using phosphorylation intensities from 9 peptide replicates on the array. Candidate SRMS consensus motifs are indicated for the phosphopeptides whose phosphosite sequences match these motifs. "-" The motifs associated with these peptides were not enriched in our *in silico* motif analyses.

S. no.	Peptide corresponding to protein	UniProtKB ID	p-Tyr. site	Fold-change phosphorylation	pValue	Candidate SRMS substrate Motif
1	Vimentin (VIM)	P08670	Y383	1.94436	0.021	*xx**RxY**xx*
2	Guanine nucleotide-binding protein subunit beta-2-like 1, N-terminally processed (GNB2L1)	P63244	Y246	1.84406	0.013	--
3	Cytoskeleton-associated protein 2 (CKAP2)	Q8WWK9	Y592	1.73666	0.005	--
4	Cyclin-dependent kinase 2/3 (CDK2/3)	G3V5T9	Y19	1.68668	0.031	*xx**VY**xx*
5	Heterogeneous nuclear ribonucleoprotein M (HNRNPM)	P52272	Y681	1.53786	0.008	*xx**YA**xx*
6	Eukaryotic initiation factor 4A-III/ 4A-I (EIF4A3/1)	P38919	Y50	1.53058	0.038	xx**YG**xx
7	Uridine 5-monophosphate synthase (UMPS)	P11172	Y326	1.48503	0.071	*xx**KxY**xx*
8	Keratin, type I cytoskeletal 18 (KRT18)	F8VZY9	Y94	1.45291	0.089	--
9	UBX domain-containing protein 6 (UBXN6)	Q9BZV1	Y181	1.43765	0.050	--
10	Guanine nucleotide-binding protein-like 3-like protein (GNL3L)	Q9NVN8	Y243	1.41008	0.059	--
11	Melanoma-associated antigen D1 (MAGED1)	Q9Y5V3	Y505	1.37761	0.002	*xx**RxY**xx*
12	40S ribosomal protein S3 (RPS3)	P23396	Y87	1.36634	0.015	*xx**YA**xx*
13	Docking protein 1 (Dok1)	Q99704	Y337	1.33743	0.059	*xx**KxxY**xx*
14	DNA replication licensing factor MCM7	P33993	Y33	1.32199	0.032	*xx**YG**xx*
15	Growth factor receptor-bound protein 10 (GRB10)	Q13322	Y354	1.31135	0.090	*xx**KxY**xx*
16	Guanine nucleotide exchange factor VAV2	P52735	Y552	1.28324	0.068	--

4.1.3.7 Validation of Vimentin and Sam68 as SRMS substrates

Our peptide array analyses revealed vimentin as a top candidate substrate of SRMS (Figure 11 and Table 4.2). Vimentin (VIM) is a key cytoskeletal protein and a substrate of c-Src and various serine/threonine kinases (Cheng *et al.*, 2003; Eriksson *et al.*, 2004; Ferrando *et al.*, 2012). Interestingly, KHDRBS1/Sam68 was also identified as a candidate SRMS substrate from our MS analyses although this was not tested in our peptide microarrays. Sam68 is a key RNA-binding protein implicated in cellular RNA-metabolic processes whose functions are regulated by phosphorylation (Feracci *et al.*, 2016). We have previously shown that the SRMS family member, BRK phosphorylates Sam68 and regulates its RNA-binding properties (Lukong *et al.*, 2005). To verify vimentin and Sam68 as *in vivo* SRMS substrates, we performed immunoprecipitation analyses using GFP-vimentin and Myc-Sam68 constructs. We first co-expressed GFP-vimentin with either Flag-tagged wild type SRMS or SRMS K258M in HEK293 cells. Vimentin was immunoprecipitated using antibodies against GFP and the immunoprecipitates and input probed with antibodies against GFP, Flag or total phospho-tyrosines (Figure 4.12A and 4.12B). Firstly, our immunoblotting analyses using antibodies against Flag-tag showed that both wild type SRMS and SRMS K258M co-precipitated with ectopically expressed vimentin indicating that SRMS associates with vimentin in a kinase-independent manner (Figure 4.12A). We observed the tyrosine-phosphorylation of vimentin only in wild type SRMS-transfected cells and not SRMS K258M-transfected cells, confirming that the phosphorylation of vimentin is SRMS kinase-dependent (Figure 4.12A). Likewise, we co-expressed Myc-Sam68 with either GFP-tagged wild type SRMS or SRMS-K258M. We immunoprecipitated Sam68 using antibodies against Myc and probed the immunoprecipitates and input using antibodies against Myc, GFP and total phosphotyrosines (Figure 4.13A and 4.13B). As seen with vimentin, we found that both wild type SRMS as well as SRMS K258M co-precipitated with Sam68, indicative of kinase-independent pattern of intermolecular association (Figure 4.13A). Importantly, Sam68 was found to be tyrosine phosphorylated only upon co-expression with wild type SRMS but not kinase-dead SRMS (Figure 4.13A). Together, our findings concur with the results of our peptide array analyses and suggest that both vimentin and Sam68 are substrates of SRMS.

To test if vimentin and Sam68 are direct substrates of SRMS, we performed *in vitro* kinase assays using purified recombinant GST-SRMS kinase protein and GST-vimentin or GST-Sam68

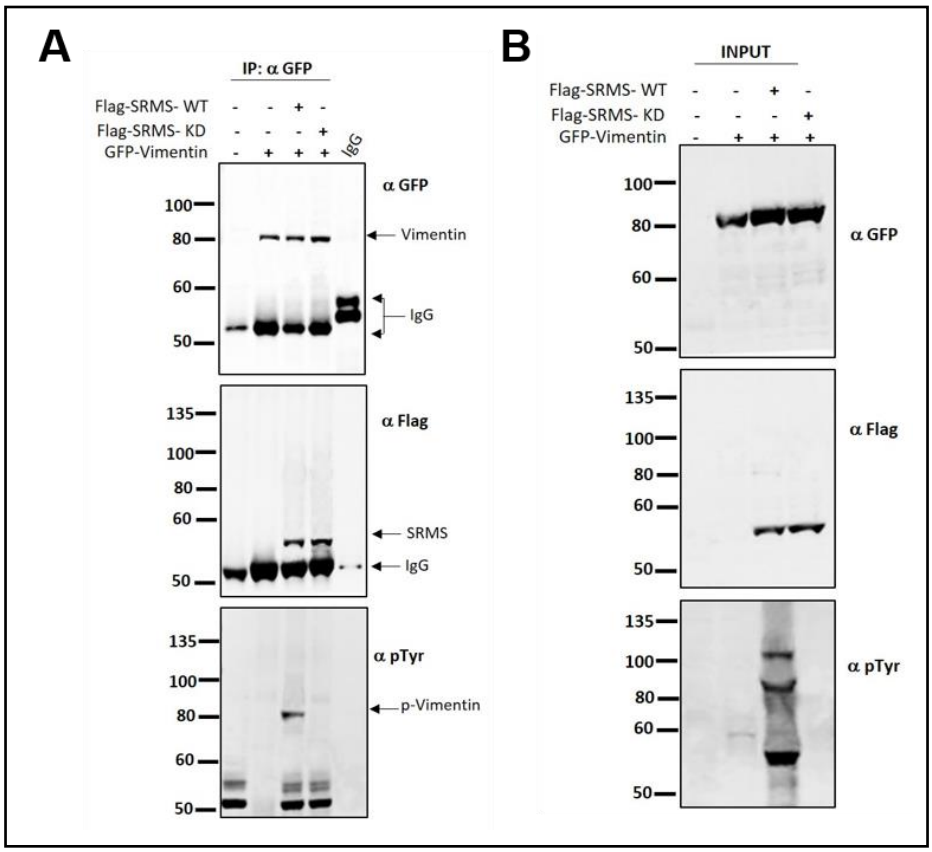


Figure 4.12: Validation of vimentin as a SRMS substrate. A. HEK293 cells were transiently transfected with constructs encoding GFP-vimentin alone or together with Flag-wild type SRMS (SRMS WT) or Flag-Kinase-dead SRMS (SRMS KD). GFP-vimentin was immunoprecipitated using antibodies against GFP. The immunoprecipitates (right panel) were probed with antibodies against GFP (top), Flag (Middle) or total phosphotyrosines (Bottom). **B.** Input from the corresponding lysates were probed with the same antibodies and are shown in the same order (Right panel).

proteins. Our results showed that in the presence of ATP, SRMS kinase directly phosphorylated both vimentin and Sam68 *in vitro* (Figure 4.14A and 4.14B). As expected, in the absence of SRMS, vimentin and Sam68 were not phosphorylated. As an additional control, we incubated the GST protein alone with GST-SRMS kinase and observed no phosphorylation of the GST protein confirming that SRMS directly phosphorylates both, vimentin and Sam68 (Figure 4.14A and 4.14B).

Since we found that vimentin and Sam68 co-precipitated with SRMS in our immunoprecipitation analyses, we asked if SRMS colocalizes with vimentin or Sam68 *in vivo*. To examine this, we co-transfected GFP-vimentin or GFP-Sam68 with either mCherry-wild type

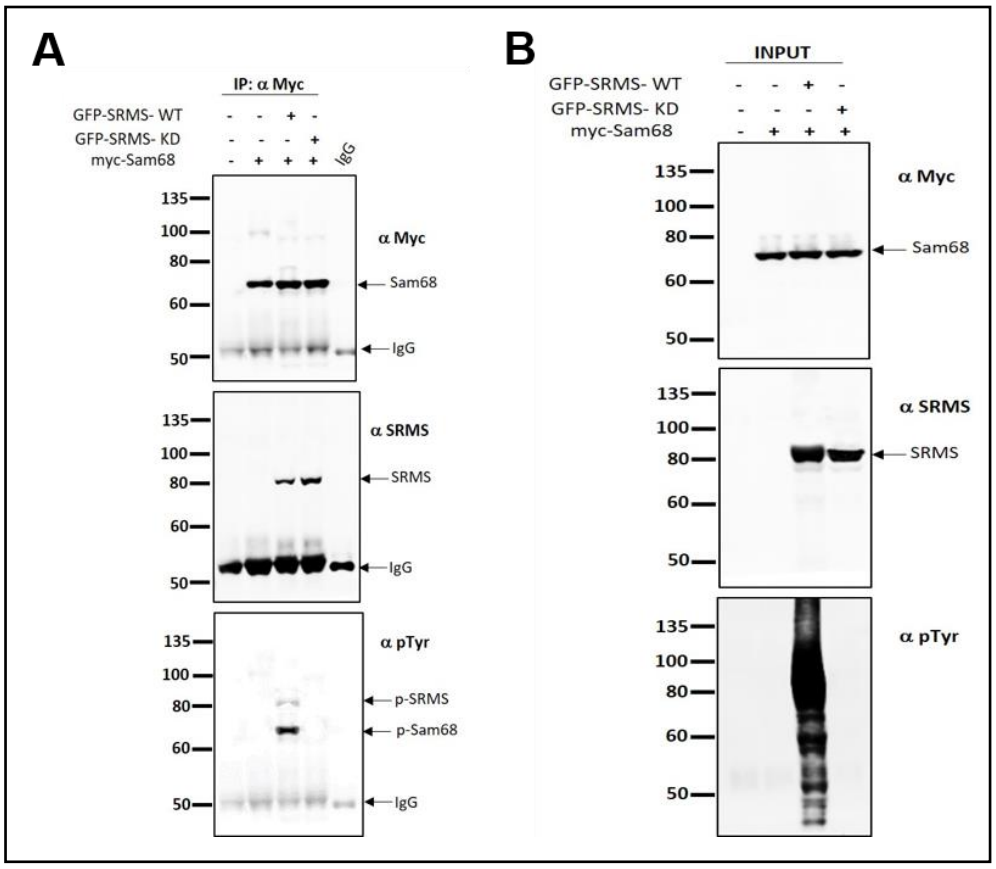


Figure 4.13: Validation of Sam68 as a SRMS substrate. A. HeLa cells were transfected either individually with Myc-Sam68 or co-transfected with GFP-wild type SRMS or GFP-kinase-dead SRMS. Myc-Sam68 was immunoprecipitated using anti-Myc antibodies. The immunoprecipitates (left panel) were probed by Western blotting with antibodies against Myc (top), GFP (middle) or total phosphotyrosines (bottom). **B.** Input from the corresponding lysates were probed with the same antibodies and are shown in the same order (Right panel).

SRMS or SRMS K268M, in HEK293 cells and assessed colocalization by confocal microscopy analyses. Firstly, we observed that GFP-tagged or untagged wild type SRMS localized predominantly to cytoplasmic punctae (Figure 4.15A and 4.15B), as we previously reported (Goel *et al.*, 2013). Vimentin was predominantly found to localize to cytoplasmic tubular networks (Figure 4.16A), consistent with previous reports describing these as vimentin intermediate filaments (VIFs) (Flynn *et al.*, 2016; Mendez *et al.*, 2010). Importantly, we observed that GFP-vimentin and mCherry-wild type SRMS colocalized in the cytoplasm of HEK293 cells (Figure 4.16B). We tested this co- localization reciprocally using GFP-wild type SRMS and mCherry-vimentin constructs. We observed that both mCherry-vimentin and GFP-SRMS colocalized in the cytoplasm (Figure 4.16C). We also noted that SRMS colocalized with vimentin in VIFs

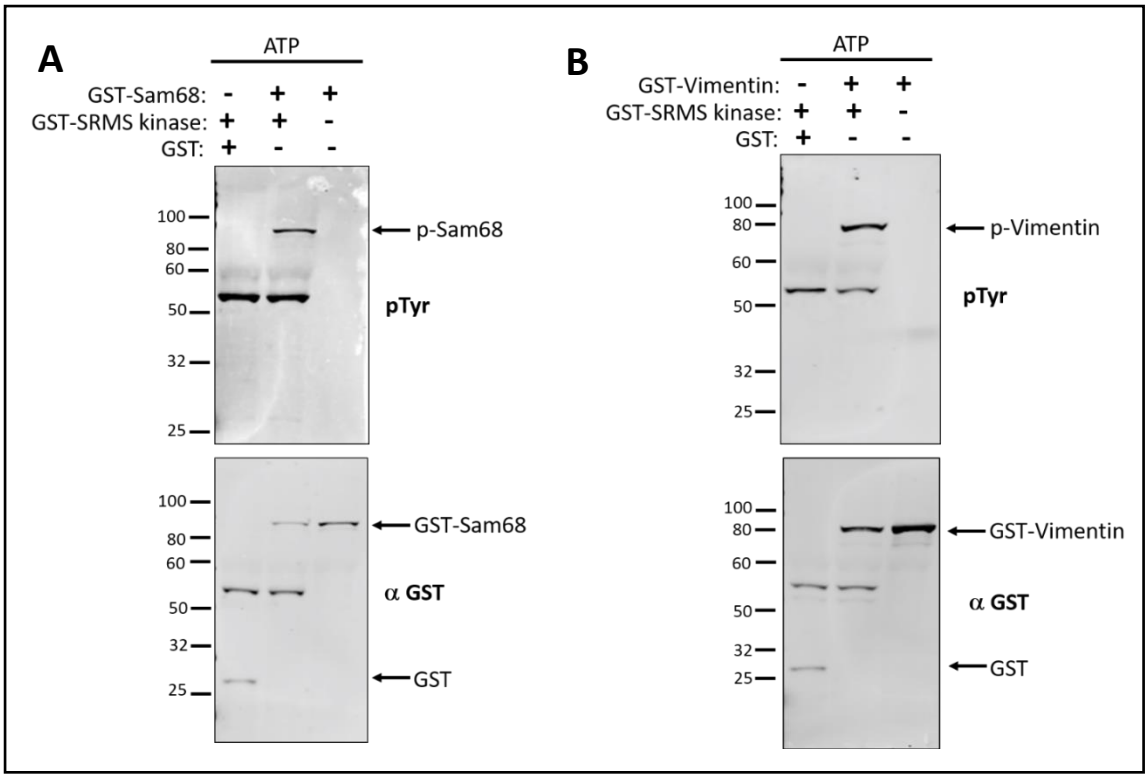


Figure 4.14: Validation of Sam68 and Vimentin as direct substrates of SRMS. A. *In vitro* kinase assays were performed using purified recombinant GST-SRMS kinase and GST-Sam68. GST-Sam68 was incubated with or without GST-SRMS kinase and in the presence of ATP. GST-SRMS kinase was also incubated with GST alone as a control. Phosphorylation of Sam68 by SRMS was assessed by Western blotting using general phosphotyrosine antibodies. The immunoblots were also probed with antibodies against GST to show the expression of the GST-Sam68, GST-SRMS kinase and GST alone. **B.** Purified recombinant GST-vimentin was incubated with or without GST-SRMS kinase and in the presence of ATP. Phosphorylation of vimentin by SRMS was assessed by Western blotting using general phosphotyrosine antibodies. GST-SRMS kinase was also incubated with GST alone as a control. The immunoblots were also probed with antibodies against GST to show the expression of the respective proteins.

rather than in the characteristic SRMS-cytoplasmic puncta, indicating that vimentin likely sequesters SRMS to VIFs (Figure 4.16B and 4.16C). No colocalization was observed between mCherry alone and GFP-vimentin, implying that the observed colocalization between SRMS and vimentin does not depend on the fluorescent protein tags used (Figure 4.16D). Although Sam68 is known to localize predominantly to the nucleus (Lukong *et al.*, 2005; Wrighton, 2016), certain studies have reported a cytoplasmic localization of the RNA-binding protein (Huot *et al.*, 2009; Liao *et al.*, 2013; Paronetto *et al.*, 2006). Consistent with previous reports, we found that Sam68 localized predominantly to the nucleus in cells co-expressing control vector encoding mCherry

alone (Figure 4.17A). However, in a few cells we also observed a partial cytoplasmic localization of Sam68 (Figure 4.17B). Co-expression of GFP-Sam68 and mCherry-SRMS revealed partial colocalization of both proteins where SRMS colocalized with cytoplasmic Sam68 (Figure 4.17C). No colocalization was observed between GFP-Sam68 and mCherry alone (Figure 4.17A and 4.17B). Together, our data show that both vimentin and Sam68 colocalize with SRMS in the cytoplasm.

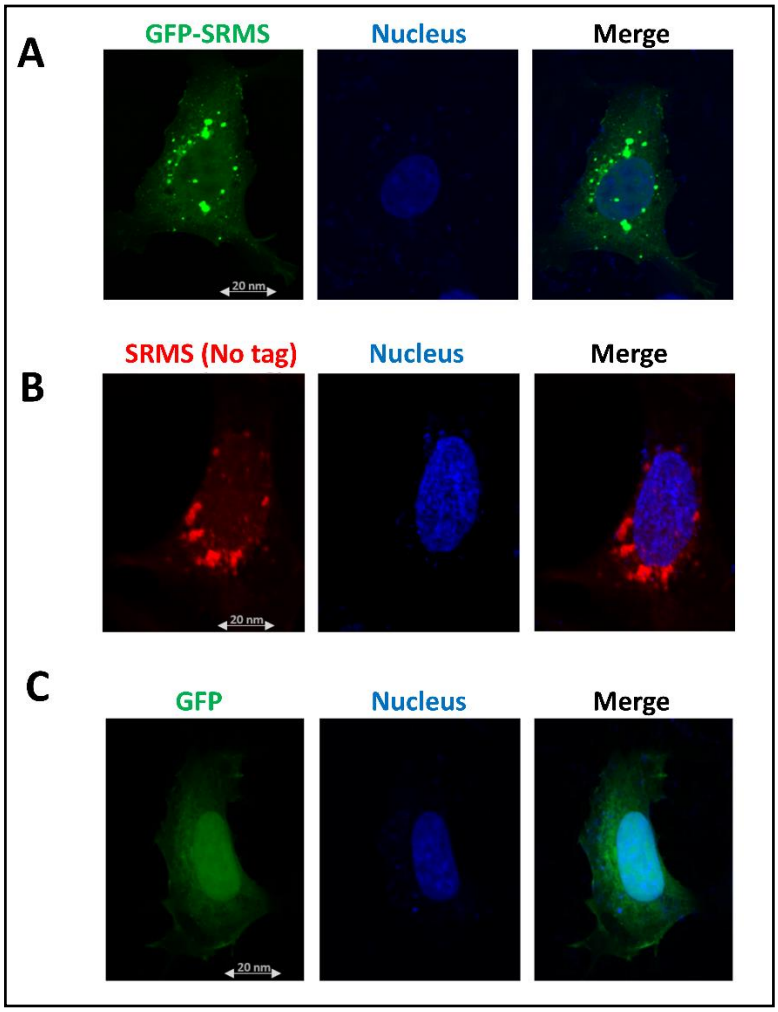


Figure 4.15: Subcellular localization of wild type SRMS in cytoplasmic punctae. HEK293 cells were seeded on glass coverslips in 6-well culture dishes. Vectors encoding: **A.** GFP-wild type SRMS, **B.** untagged wild type SRMS or **C.** GFP alone were exogenously introduced in the cells. Transfected cells were fixed, permeabilized and ectopic untagged SRMS detected using primary antibodies against SRMS (mouse IgG) and secondary anti-mouse Texas red antibodies. Nuclei were stained with DAPI. Images were acquired on a confocal microscope at 40x magnification. Scale bar = 20 nm.

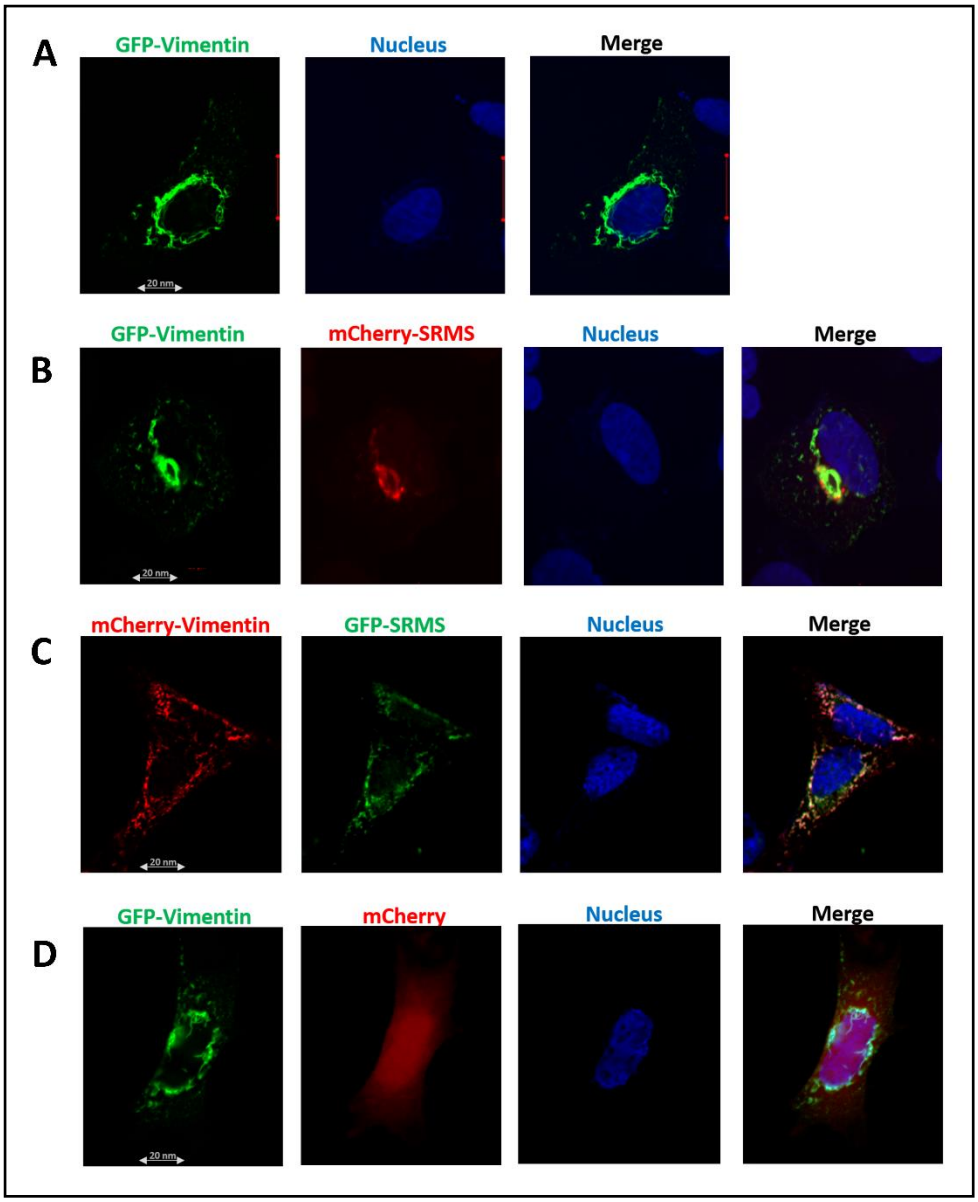


Figure 4.16: Colocalization of SRMS with vimentin. **A.** HEK293 cells, seeded on glass coverslips and cultured in 6-well culture dishes, were transfected with a vector encoding GFP-vimentin alone or **B.** co-transfected with vectors encoding GFP-vimentin and mCherry-SRMS (wild type), **C.** mCherry-vimentin and GFP-wild type SRMS or **D.** GFP-vimentin and mCherry alone. Colocalization between SRMS and vimentin was determined via confocal microscopy. Images were acquired at 40x magnification. Scale bar = 20 nm.

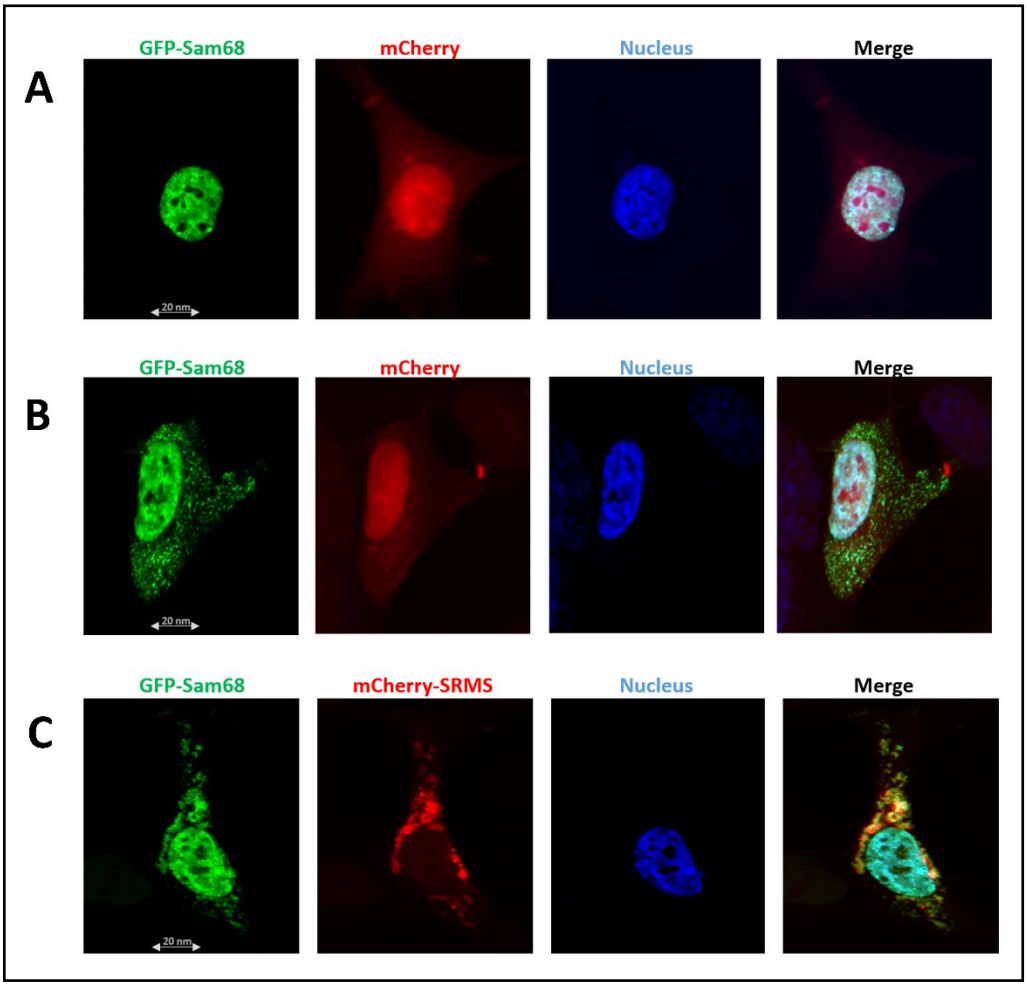


Figure 4.17: Colocalization of SRMS with Sam68. A. and B. HEK293 cells, seeded on glass coverslips and cultured in 6-well culture dishes, were co-transfected with either GFP-Sam68 and mCherry alone or **C.** GFP-Sam68 and mCherry-wild type SRMS. Colocalization between Sam68 and SRMS or mCherry alone was determined via confocal microscopy. Images were acquired at 40x magnification. Scale bar = 20 nm.

4.1.3.8 SRMS induces the tyrosine phosphorylation of Sam68 in an EGF-dependent manner

Sam68 is known to be frequently tyrosine phosphorylated, which is essential for the regulation of the RNA-binding function of the protein (Richard, 2010). Since we found that Sam68 is a SRMS substrate, we checked if SRMS was required for the tyrosine-phosphorylation of Sam68 at the endogenous level. We used MDA-MB 231 cells since this cell line is known to express both proteins at the endogenous level (Goel *et al.*, 2013). To determine whether endogenous SRMS promotes the phosphorylation of Sam68, we transiently knocked-down SRMS and immunoprecipitated endogenous Sam68 from either parental, scrambled siRNA or SRMS siRNA-

transfected cells. Firstly, we observed a significant knockdown of endogenous SRMS in the SRMS siRNA-transfected cells (Figure 4.18A, lower panel). However, immunoblotting the Sam68 immunoprecipitates with general phosphotyrosine antibodies revealed no significant difference in Sam68 tyrosine-phosphorylation between SRMS-knockdown and parental or scramble siRNA-transfected cells (Figure 4.18A, top panel). This implied that under steady-state conditions endogenous SRMS has no impact on the tyrosine-phosphorylation of endogenous Sam68. Previous studies have shown that Sam68 is a downstream target of the EGFR-signaling pathway where Sam68 undergoes significant tyrosine-phosphorylation upon stimulation of cells with EGF (Lukong *et al.*, 2005). Additionally, BRK has been shown to modulate the EGF-stimulated tyrosine-phosphorylation of Sam68 (Lukong *et al.*, 2005). We therefore tested if SRMS affects Sam68-phosphorylation in response to EGF-stimulation. We first stimulated serum-starved MDA-MB 231 cells with EGF at intervals of 5 minutes for up to 15 minutes. We then immunoprecipitated endogenous Sam68 from either serum-starved or EGF-stimulated cells and assessed the tyrosine-phosphorylation status of Sam68 via immunoblotting with phosphotyrosine antibodies. As expected, we found that EGF-stimulation led to increased tyrosine-phosphorylation of endogenous Sam68 which peaked at 10 minutes post-stimulation (Figure 4.18B). Next, we stimulated control and SRMS-depleted MDA-MB 231 cells with EGF for 10 minutes and immunoprecipitated Sam68 to probe with phosphotyrosine antibodies. We found that knocking down endogenous SRMS led to reduced EGF-stimulated tyrosine-phosphorylation of Sam68 compared to scramble siRNA-transfected cells (Figure 4.18C). Further, upon increased exposure of the immunoblot we observed a faint but noticeable amount of SRMS co-precipitating with Sam68 implying that both proteins are able to associate in these cells (Figure 4.18C). Our results therefore show that SRMS regulates the phosphorylation of Sam68 in an EGF-dependent manner in MDA-MB 231 cells. This further underscores the potential physiological significance of Sam68 as a substrate of SRMS.

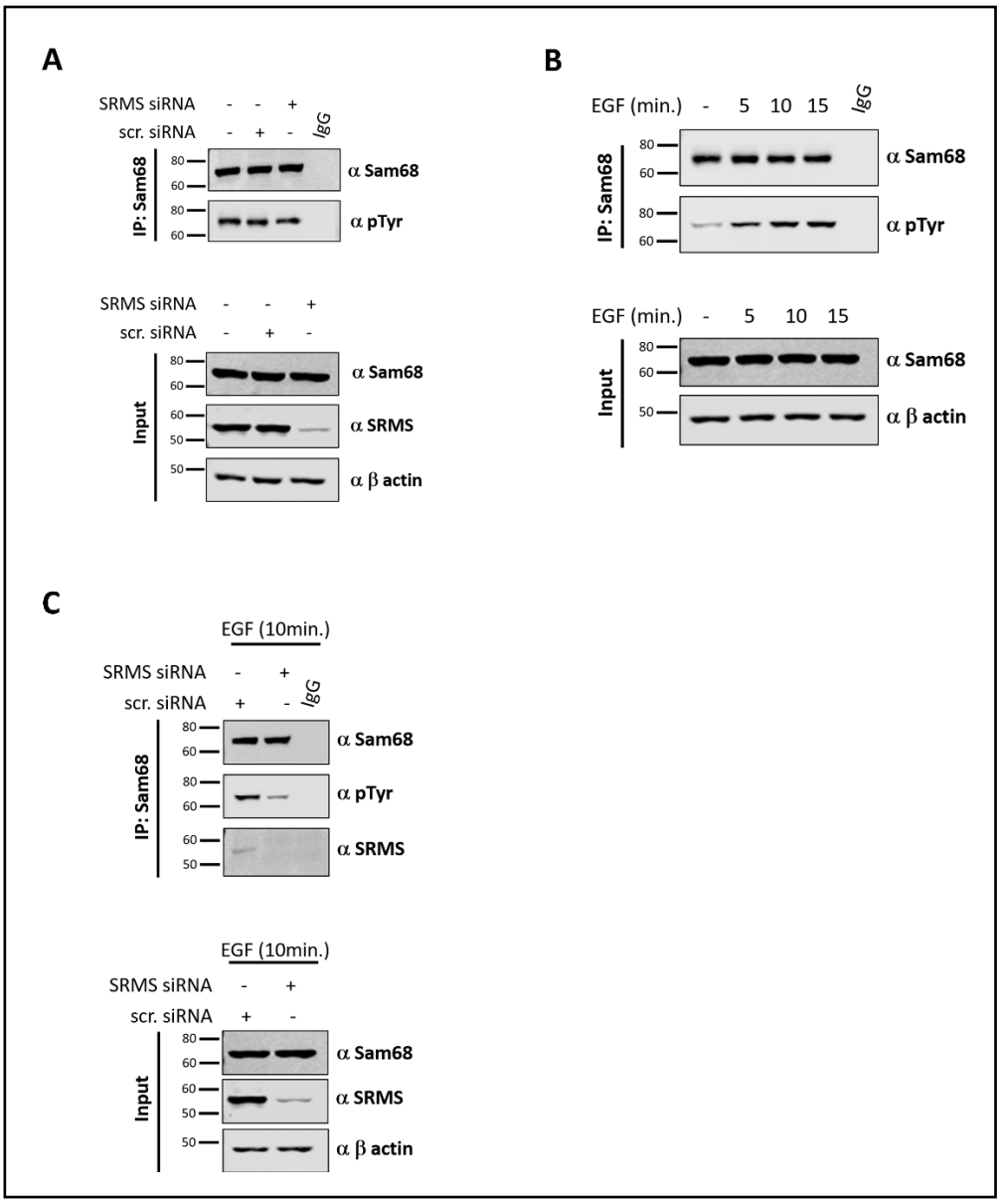


Figure 4.18: EGF-induced Sam68 phosphorylation mediated by SRMS. A. MDA-MB 231 cells were transfected with either scramble siRNA or siRNA against SRMS. Parental and siRNA-transfected cells were lysed and Sam68 immunoprecipitated using Sam68 antibodies. The immunoprecipitates were probed via Western blotting with antibodies against Sam68 and general phosphotyrosines. Input from the corresponding lysates were probed with antibodies against SRMS and Sam68. β-actin served as a loading control. **B.** Serum-starved MDA-MB 231 cells were either treated with 100 ng/mL EGF for the indicated duration or left untreated. Sam68 was immunoprecipitated from the lysates derived from the untreated and treated cells. The immunoprecipitates were probed with antibodies against Sam68 and phosphotyrosines. Input from the corresponding lysates were probed with antibodies against Sam68 and β-actin. **C.** MDA-MB 231 cells, transfected with either scramble siRNA or SRMS siRNA, were stimulated with 100 ng/mL EGF for the indicated duration and lysed. Sam68 was immunoprecipitated from the lysates and the immunoprecipitates probed with antibodies against Sam68, phosphotyrosine and SRMS. Input from the corresponding lysates were probed with antibodies against Sam68, SRMS and β-actin.

4.1.4 Discussion

SRMS is a non-receptor tyrosine kinase of the BRK family kinases. Unlike BRK and FRK, SRMS remains a largely understudied member of the family. We previously characterized the enzymatic activity of SRMS and reported that wild type SRMS is enzymatically active and induces the tyrosine-phosphorylation of several proteins when exogenously expressed in mammalian cells (Goel *et al.*, 2013). Ours represents the first quantitative phosphoproteomics study to identify a vast pool of novel candidate SRMS substrates from mammalian cells ectopically expressing the wild type SRMS kinase. Our unbiased analyses resulted in the identification of 1129 unique tyrosine-phosphorylated peptides in wild type SRMS-expressing cells. These were mapped to 663 unique proteins which were determined to be candidate SRMS substrates since the corresponding phosphopeptides were identified exclusively in cells expressing ectopic wild type SRMS. Importantly, we identified DOK1 as a hyperphosphorylated protein in SRMS-overexpressing cells. This confirmed our previous findings where we characterized DOK1 as a bona fide SRMS substrate and demonstrated that endogenous DOK1 was phosphorylated by ectopic SRMS in HEK293 cells. Further, as expected, the peptide corresponding to SRMS Y380 was identified as one of the major phosphopeptides from our MS analyses (Table 4.1). We also identified 14 novel tyrosine-phosphorylated sites in SRMS (Table 4.1) which may either represent SRMS autophosphorylation sites or target sites for other tyrosine kinases. The biochemical and functional significance associated with these phosphosites may be worth investigating in future studies. Intriguingly, we found 117 tyrosine-phosphorylated peptides in control cells but not in cells expressing ectopic SRMS (Figure 4.3B). Notably, of the 663 identified tyrosine-phosphorylated proteins in SRMS-expressing cells, 6 are tyrosine-phosphatases (PTPRA, PTPRE, PTPN6, PTPN11, PTPRK, PTPN18) (Supplementary Table 3). It is possible that SRMS-mediated tyrosine phosphorylation may either directly or indirectly regulate the tyrosine-phosphatase activity of one or several of these phosphatases potentially leading to the de-phosphorylation of specific targets. This may at least partly explain why the 117 tyrosine-phosphorylated peptides were identified in the control cells but not the SRMS-expressing cells.

It may be noted that similar earlier phosphoproteomics studies have identified candidate substrates from cells overexpressing specific kinases (Amanchy *et al.*, 2008; Leroy *et al.*, 2009; Wu *et al.*, 2010; Xue *et al.*, 2012). However, it can be argued that such approaches offer limited insights into the physiological context of kinase-substrate interactions. Native phosphorylation

events occurring at the cellular level, either in the presence or absence of relevant biological stimuli, are a more appropriate representation of physiological kinase activities. To this end, alternative approaches involve the modulation of either the expression level or catalytic activity of native kinases to identify stoichiometric differences among phosphorylated proteins between the control and test conditions (Amanchy *et al.*, 2009; Hsu *et al.*, 2011; Poss *et al.*, 2016; Xue *et al.*, 2012). However, screening for physiological substrates of uncharacterized kinases can be particularly challenging. In the case of SRMS, the expression of the kinase has been detected in various breast cancer cell lines (Goel *et al.*, 2013; Potts *et al.*, 2013). Other groups have reported the expression of SRMS in various murine organs (Kawachi *et al.*, 1997; Kohmura *et al.*, 1994). However, the activation status of endogenous SRMS in these cells is not known. The phosphorylation of the activation loop tyrosine is a primary indicator of enzymatic activation in tyrosine kinases (Hunter, 2009). Unlike c-Src (Dong *et al.*, 2016) and BRK (Jiang *et al.*, 2017), no commercial antibodies exist to detect the phosphorylation of the SRMS activation-loop tyrosine residue (Y380). Additionally, whereas various specific chemical inhibitors of Src (Yeatman, 2004) and BRK (Jiang *et al.*, 2017) have been characterized, to our knowledge, no such specific inhibitors of SRMS enzymatic activity have been reported. Considering these technical shortcomings, we reasoned that the overexpression of the catalytically active wild type SRMS kinase in human cells may represent a reasonably viable alternative strategy to identify the cellular substrates of the kinase. Additionally, we have previously reported that both endogenous and ectopic SRMS localizes to cytoplasmic punctate structures. This implies that the localization and therefore the biochemical and cellular functions of ectopic SRMS may likely be similarly regulated. This further highlights the potential biological significance of the candidate substrates that were identified in our study.

Other groups have reported the identification of high-confidence kinase substrates via similar phosphoproteomics approaches by correlating the identified phosphosites with known or well-characterized consensus substrate motifs of the kinase (Courcelles *et al.*, 2013; Francavilla *et al.*, 2017; Kettenbach *et al.*, 2011). Deng et al. identified *xIYx* and *YxxV* as two potential consensus substrate motifs of SRMS using an arrayed positional scanning peptide library (Deng *et al.*, 2014). Our *in silico* analyses of the phosphosites derived from candidate SRMS substrates uncovered various overrepresented motif patterns (Figure 4.7 and 4.8). Surprisingly, the *xIYx* and *YxxV* motifs were not among those identified in our analyses. It is not clear whether the motifs we

identified in the present study were among the positional peptide variants tested by Deng et al. (Deng *et al.*, 2014). Further, though we identified *KxY* as one of the top candidate SRMS consensus motifs (Figure 4.7, 4.7 and 4.9), our validation analyses using peptide arrays revealed that the Vimentin Y383 peptide corresponding to the *RxY* motif displayed relatively increased phosphorylation compared to the other motifs (Table 4.2 and Figure 4.11). It may be noted that various factors govern enzyme-substrate specificity in a cellular context. The subcellular localization as well as the intermolecular interactions between the kinase and substrate proteins represent key factors that contribute to the enzyme-substrate specificity (Hunter, 2009). Additionally, when performed exclusively, *in vitro* studies involving peptide arrays may not accurately recapitulate physiologically relevant conditions favoring enzyme-substrate interactions. Moreover, the phosphorylation of a peptide is not entirely representative of the phosphorylation of the same sequence within a globular protein. All this may partly explain why our motif analyses of the candidate cellular SRMS substrates did not identify the two motifs described previously (Deng *et al.*, 2014). We therefore reason that the identification of candidate kinase motifs from cell-based phosphoproteomics studies would likely provide rational opportunities for motif-validation in *in vitro* studies, such as with peptide arrays.

Our analyses using peptide arrays revealed that the phosphorylation of a number of peptides could not be validated since these were rendered statistically insignificant. Importantly, it has been noted that due to technical reasons data normalization measures tend to account for the loss of information from nearly 50% to 70% of peptides printed on such arrays (Scholma *et al.*, 2016). Therefore, we normally use quasi-stringent t-testing (p-value < 0.1) in our analyses to limit the bias associated with stringent statistical thresholds (Kindrachuk *et al.*, 2012; Napper *et al.*, 2015). This method of applying a quasi-stringent t-testing threshold in peptide array analyses has also been described by Scholma *et al.* (Scholma *et al.*, 2016).

Using the g:Profiler platform (Reimand *et al.*, 2016), we identified various biological processes/pathways enriched by the candidate SRMS substrates (Figure 4.10A and 4.10B). RNA processing emerged as one of the major functional clusters associated with the candidate SRMS substrates (Figure 4.10A and 4.10B). Additionally, some of the top candidate SRMS substrates included translation initiation factors such as EIF4G2, EIF2S3 and EIF4A3, implying a potential role of SRMS in protein translation. Sam68 was one of the proteins annotated in both the RNA processing and regulation of protein translation clusters (Supplementary Table 5). Sam68 is a KH

domain-containing RNA-binding protein (RBP) shown to regulate several processes including pre-mRNA splicing, mRNA translation, nuclear export, RNA stability, signal transduction, and cell cycle progression (Lukong *et al.*, 2005; Lukong and Richard, 2003). Despite its predominantly nuclear localization, Sam68 is a major substrate of Src in the cytoplasm during mitosis (Taylor *et al.*, 1995), and of BRK in the nucleus (Lukong *et al.*, 2005). Both the subcellular localization Sam68 and its RNA-binding functions are modulated by phosphorylation (Lukong *et al.*, 2005; Lukong and Richard, 2003). Other tyrosine kinases known to target Sam68 include Fyn, Lck, Tec, Jak3, BRK, Zap70, Btk or the insulin receptor (Di Fruscio *et al.*, 1999; Lukong *et al.*, 2005; Sanchez-Jimenez and Sanchez-Margalet, 2013). We validated Sam68 as a bona fide SRMS substrate via co-immunoprecipitation analyses and direct *in vitro* analyses (Figure 4.13A, 4.13B and 4.14A). We further demonstrated that Sam68 displays a partial cytoplasmic localization (Figure 4.17A and 4.17B) and that cytoplasmic Sam68 colocalizes with SRMS *in vivo* (Figure 4.17C). More importantly, we determined that SRMS plays an important role in the EGF-induced tyrosine-phosphorylation of Sam68 (Figure 4.18). In a previous study it was shown that BRK expression is also needed for the efficient phosphorylation of Sam68 upon EGF-stimulation (Lukong *et al.*, 2005). Specifically, the study demonstrated that Sam68-Y440 was a major site targeted by BRK *in vivo* (Lukong *et al.*, 2005). Our MS analysis identified Sam68-Y167 as a hyperphosphorylated site in cells expressing ectopic SRMS implying that Sam68-Y167 may potentially be a major SRMS target site. Taken together, since kinase functions are known to be tightly regulated in a temporal and spatial context (Bergeron *et al.*, 2016; Cai *et al.*, 2008), our results indicate that the biochemical and cellular functions of SRMS, with respect to Sam68, may be partly regulated under specific stimuli such as in response to EGF.

Vimentin was identified as one of the top candidate SRMS substrates in our phosphoproteomics analyses. We validated vimentin as a *bona fide* SRMS substrate via *in vitro* and *in vivo* assays and showed that both proteins colocalize in the cytoplasm (Figure 4.12A, 4.12B, 4.14B and 4.16). Vimentin is the most common intermediate filament (IF) and an essential component of the cytoskeleton (Jiu *et al.*, 2017). Vimentin is post-translationally modified by a variety of kinases on both Tyr (Chen *et al.*, 2007) and Ser/Thr residues (Eriksson *et al.*, 2004) leading to the structural re-arrangements of vimentin IF filaments (Hyder *et al.*, 2008). To our knowledge, the phosphosites we identified in vimentin in the present study have not been functionally characterized before and may warrant future studies.

A recent study by Potts *et al.* showed that knockdown of SRMS resulted in increased light chain 3 (LC3)-positive puncta per cell and decreased abundance of GFP-LC3 in an autophagy-related 5 (ATG5)–dependent manner (Potts *et al.*, 2013). Overexpression of SRMS decreased the number of LC3-positive puncta per cell, suggesting that SRMS potentially inhibits autophagy upstream of autophagosome formation (Potts *et al.*, 2013). However, in our study, functional gene enrichment analyses of these candidate substrates did not identify any autophagy-linked cellular process or pathway. Therefore, it is possible that the regulation of autophagy by SRMS, as reported by Potts *et al.*, is not kinase-dependent or may likely be influenced in a tissue or cell line-specific context. It is possible that examining the SRMS-interacting partners may reveal proteins that are linked to autophagy.

In the present study we used ectopic SRMS to identify the candidate SRMS cellular substrates. A few or several of these substrates are likely to represent physiologically important substrates of SRMS. This will certainly lead to a better understanding of the biochemical and functional specificity associated with SRMS in a physiological context. In conclusion, our phosphoproteomics analyses helped uncover a vast array of novel candidate SRMS substrates which provides insights into the biochemical substrate motif specificities and the potentially diverse cellular roles played by the tyrosine kinase. Given that SRMS is the least studied member of the BRK family kinases, our findings provide an important resource for future mechanistic studies to investigate the cellular and physiological functions of the kinase.

4.2 Identification of candidate signaling intermediates of SRMS

The results described in this section are part of a research manuscript that has been accepted for publication by the journal Proteome Science. As part of the "Open access" policy (https://www.biomedcentral.com/about/policies/access-to-articles) implemented by the journal, the ownership of the copyright remains with the authors. The authors reserve the right to use all or part of the published work for non-commercial purposes including theses/dissertations. The manuscript and authorship details are:

Raghuveera Kumar Goel, Mona Meyer, Marta Paczkowska, Jüri Reimand, Frederick Vizeacoumar, Franco Vizeacoumar, TuKiet T. Lam and Kiven Erique Lukong. Global phosphoproteomic analysis identifies SRMS-regulated signaling intermediates.

NOTE: The Supplementary Tables referenced in this section (Supplementary Tables 9, 10 and 11) have been compressed in a single file and uploaded as an "Additional File" to the Usask Thesis and Dissertation submittal system. Accessibility to this file is governed by the University of Saskatchewan. The legends to Supplementary Tables 9, 10, and 11 are available in Appendix A.

4.2.1 Summary

The non-receptor tyrosine kinase, SRMS (Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites) is a member of the BRK family kinases (BFKs) which represents an evolutionarily conserved relative of the Src family kinases (SFKs). Tyrosine kinases are known to regulate a number of cellular processes and pathways via phosphorylating substrate proteins directly and/or by partaking in signaling cross-talks leading to the indirect modulation of various signaling intermediates. In the study described in the previous section, we profiled the tyrosine-phosphoproteome of SRMS and identified multiple candidate substrates of the kinase. The broader cellular signaling intermediates of SRMS are unknown. In order to uncover the broader SRMS-regulated phosphoproteome and identify the SRMS-regulated indirect signaling intermediates, we performed label-free global phosphoproteomics analysis on cells expressing wild-type SRMS. Our analyses identified 60 hyperphosphorylated (phosphoserine/phosphothreonine) proteins mapped from 140 hyperphosphorylated peptides. Bioinformatics analyses identified a number of significantly enriched biological and cellular processes among which DNA repair pathways were found to be upregulated while apoptotic pathways were found to be downregulated. Analyses of motifs derived from the upregulated

phosphosites identified Casein kinase 2 alpha (CK2α) as one of the major potential kinases contributing to the SRMS-dependent indirect regulation of signaling intermediates. Overall, our phosphoproteomics analyses identified serine/threonine phosphorylation dynamics as important secondary events of the SRMS-regulated phosphoproteome with implications in the regulation of cellular and biological processes.

4.2.2 Introduction

Kinase-substrate interactions are known to modulate the formation of secondary and tertiary protein complexes in cells (Kapoor *et al.*, 2015; Krapivinsky *et al.*, 2004; Wegener and Krappmann, 2008). Such protein complexes are implicated in the regulation of other kinases thereby potentiating a concerted mode of signaling, leading to the modulation of various cellular and biological functions (Besset *et al.*, 2000; Ochoa *et al.*, 2016). Importantly, serine/threonine kinases constitute the majority of the mammalian kinome and evidence suggests that serine/threonine and tyrosine kinases are interconnected by several protein-protein interactions and cross-phosphorylation events, indicative of global signaling cross-talks (Bhavanasi *et al.*, 2015; Colinge *et al.*, 2014; Emkey and Kahn, 1997). This inevitably suggests that a tyrosine kinase, for instance, may likely have an impact on the greater cellular phosphoproteome, extending beyond the direct phosphorylation of its substrates. The identification of such secondary or effector phosphorylation events on signaling intermediaries is essential towards understanding the broader mechanisms of action of kinases.

Mass spectrometry has emerged as a powerful technique to identify and quantify proteins as well as associated post-translational modifications (PTMs) in almost any biological cell/tissue type (Larance and Lamond, 2015; Walther and Mann, 2010). Label-free quantitation techniques offer a straightforward and cost-effective approach to reliably quantify peptide and proteins abundances in cells (Cox *et al.*, 2014; Kauko *et al.*, 2015; Neilson *et al.*, 2011). The technique has gained popularity in recent years and successfully applied to several phosphoproteomics studies (Bensimon *et al.*, 2010; Dazert *et al.*, 2016; Sharma *et al.*, 2014; Steger *et al.*, 2016).

Overall, in this study, using label-free quantitation we investigated the SRMS-regulated global phosphoproteome to identify potential secondary phosphorylation events involving phosphoserine and phosphothreonine sites. We used the TiO_2-based phosphopeptide enrichment strategy as a tool to preferentially identify serine/threonine phosphorylation events in cells

expressing wild-type SRMS. Our analyses identified multiple significantly upregulated phosphosites in the SRMS phosphoproteome. Using bioinformatics analyses, the upregulated proteins and phosphosites were mapped to various cellular and biological processes and cognate serine/threonine kinases, respectively. Overall, our findings provide important insights on the cross-talks between SRMS and serine/threonine kinases in the orchestrated modulation of SRMS-regulated cellular functions.

4.2.3 Results

4.2.3.1 Global analysis of the SRMS-regulated phosphoproteome

Tyrosine kinases and serine/threonine kinases are known to exhibit dynamic cross-talks leading to a concerted mode of regulation of signaling networks (Bhavanasi *et al.*, 2015; Colinge *et al.*, 2014; Emkey and Kahn, 1997). Evidence from large-scale database curation and high-throughput experimental observations indicate that such cross-talks are modulated through secondary and tertiary protein-protein interactions around primary kinase-substrate interaction and phosphorylation events (Colinge *et al.*, 2014; Ochoa *et al.*, 2016). We therefore sought to study the SRMS-regulated serine/threonine phosphoproteome to infer cues on its contribution to the cellular signaling network potentially regulated SRMS. To survey these global phosphoproteomic changes we performed metal-ion enrichment-based quantitative phosphoproteomics analysis on cells ectopically expressing wild-type SRMS. Cells expressing the empty vector backbone served as control for all background phosphorylation events occurring natively in HEK293 cells. We confirmed the expression of wild-type SRMS in the cells by immunoblotting the lysates with antibodies against GFP and SRMS (Figure 4.19). As expected, immunoblotting with phosphotyrosine antibodies confirmed the enzymatic activation of wild-type SRMS in these cells (Figure 4.19). The total proteins were then subjected to dual enzymatic digestion with Lys-C and trypsin followed by phosphopeptide enrichment using titanium dioxide (TiO_2) resin prior to LC-MS/MS analyses (Figure 4.20). The experiment was performed in three independent replicates to allow assessment of the statistical reproducibility of our analysis. Collectively, we identified 995 unique phosphosites from 1459 redundant phosphopeptides which mapped to 439 unique phosphoproteins, at an estimated False Discovery Rate (FDR) of 1% (Figure 4.21A, Supplementary Table 9). We found significant reproducibility between the replicate datasets, as

reflected by a Pearson's correlation coefficient in the range of 0.84 to 0.96 (Figure 4.21B). We quantified the number of phosphoserine, phosphothreonine and phosphotyrosine sites to assess the

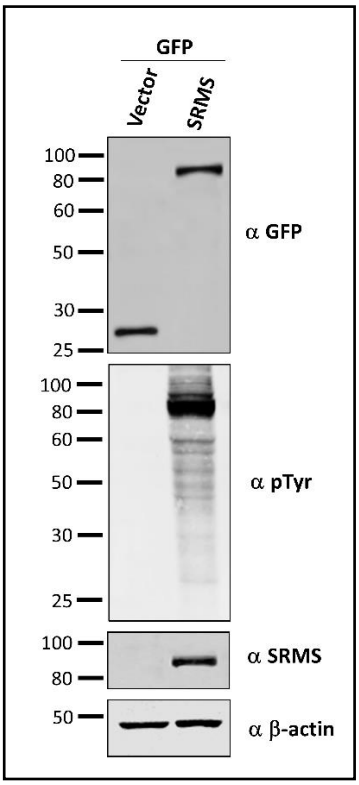


Figure 4.19: Phosphoproteomics analyses of cells expressing ectopic wild-type SRMS. Vector encoding either GFP alone or GFP-SRMS wild type was exogenously introduced in HEK293 cells. Prior to phosphopeptide enrichment, immunoblotting analyses was performed on a portion of total cell lysates derived from the transfected cells. The lysates were probed with antibodies against GFP, total phosphotyrosines and SRMS. Immunoblotting with antibodies against β-actin was used to assess the loading of total proteins.

distribution of the identified phosphoproteome. Consistent with previous studies employing metal ion-based phosphopeptide enrichment techniques (Gnad *et al.*, 2016; Kuzmanov *et al.*, 2016; Possemato *et al.*, 2017), we observed that the majority of the identified phosphosites were represented by phosphoserine (85.2%) followed by phosphothreonine sites (13.3%) (Figure 4.21C). Phosphotyrosine sites represented a minor fraction (0.76%) of the identified phosphoproteome (Figure 4.21C). This was an expected observation since previous studies have reported a similar phosphotyrosine enrichment profile using metal ion enrichment chromatography despite significant induction of cellular tyrosine phosphorylation, for instance, via pervanadate treatment (Di Palma *et al.*, 2013; Possemato *et al.*, 2017). Additionally, singly phosphorylated

peptides were more strongly represented than doubly or triply phosphorylated peptides in our dataset, consistent with previous reports (Courcelles *et al.*, 2013; Gnad *et al.*, 2016; Pan *et al.*, 2008) (Figure 4.21D).

To gain a better understanding of the SRMS-regulated phosphoproteome, we first applied a p-value threshold of ≤ 0.05 to focus on phosphopeptides that displayed statistically significant differential regulation (up or downregulation) compared to control cells (Supplementary Table 9).

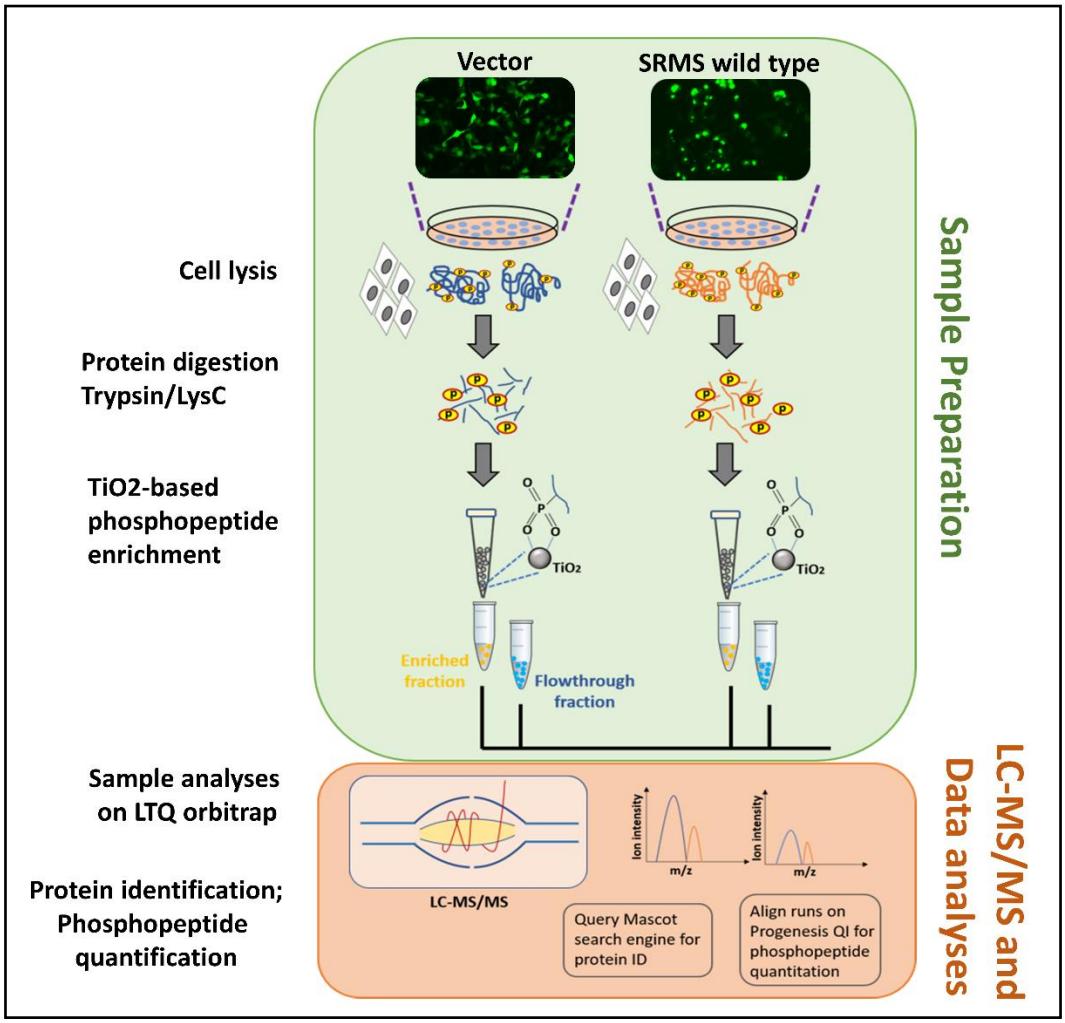


Figure 4.20: Phosphoproteomics workflow. Schematic representation of the label-free quantitation-based phosphoproteomics workflow using cells expressing GFP alone (the empty vector control) or cells expressing GFP-SRMS wild type. The cells were lysed, and total proteins subjected to dual enzymatic digestion (Trypsin/Lys-C). The crude tryptic peptides were used towards phosphopeptide enrichment using TiO_2 resin. Both enriched and flowthrough fractions were analysed by LC-MS/MS and data analyses performed using the Mascot search engine (for protein identification) and Progenesis QI tool (for phosphopeptide quantitation).

We then determined significantly regulated phosphopeptides in the wild-type SRMS phosphoproteome based on the differential abundance of the corresponding average precursor ion intensities in the control cells. To ensure stringent analyses, phosphopeptides quantified at average pre-cursor ion intensities $\geq$ 3-fold [Log_2 (SRMS/control) $\geq$ 1.584] compared to corresponding

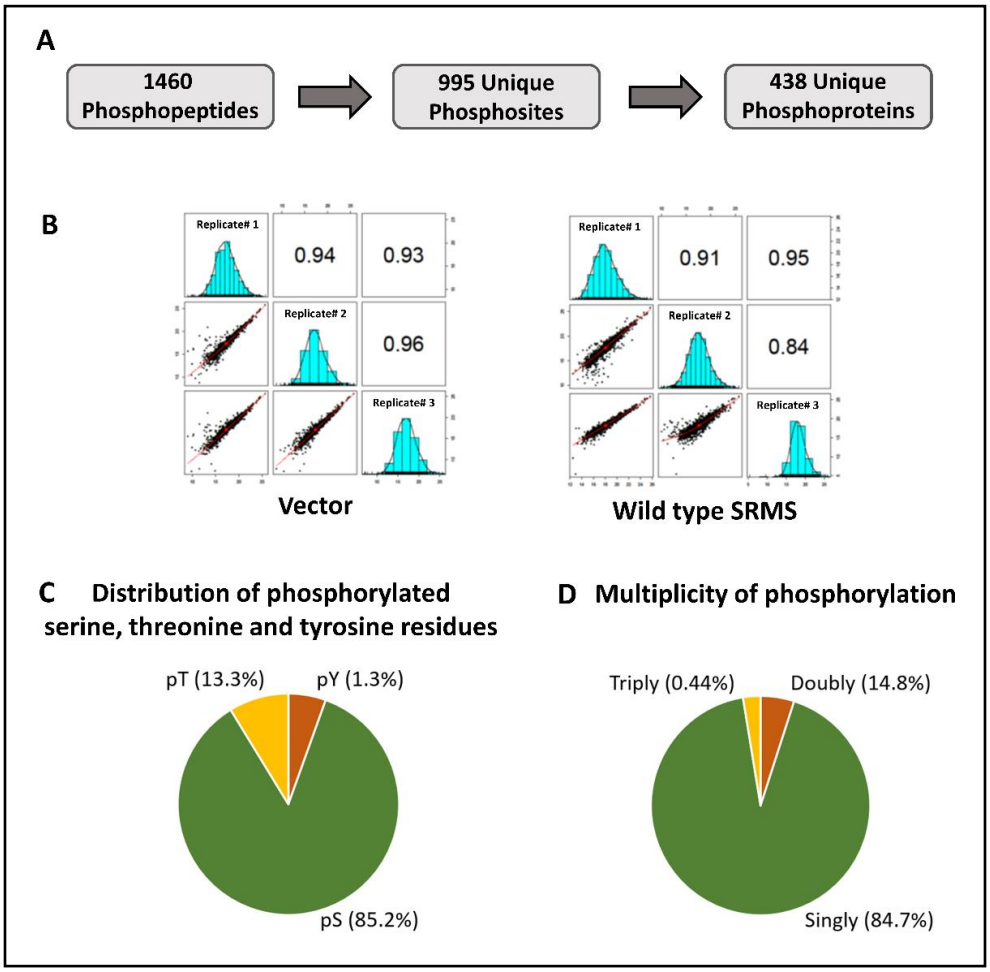


FIGURE 4.21: Overview of the identified phosphoproteome. A. The coverage of the phosphoproteome showing the total number of identified phosphopeptides, unique phosphosites and unique proteins mapping from the phosphopeptides. **B.** Pearson's correlation analyses of the 3 replicate datasets corresponding to the vector control and wild type SRMS phosphoproteome. The Pearson's correlation coefficients and associated distribution curve histograms between Replicate 1, 2 and 3 for both experimental conditions are reflected in a matrix format. **C.** Pie-chart representation of the multiplicity of phosphorylation of the identified phosphopeptides indicating the percentage of phosphopeptides carrying either a single phosphosite, double phosphosites or triple phosphosites. **D.** Pie-chart depiction of the proportion of phosphoserine, phosphothreonine and phosphotyrosine sites in the identified phosphoproteome.

average intensities in control cells were considered hyperphosphorylated or upregulated (Figure 4.22A and Supplementary Table 9). Similarly, phosphopeptides with average intensities ≤ 0.5-fold [Log_2 (SRMS/control) ≤ -1; conversely equal to 2-fold upregulation in control cells] were considered hypophosphorylated or downregulated (Figure 4.22A and Supplementary Table 9). Using these criteria, we identified 140 upregulated and 1 downregulated (NUCKS S19) phosphopeptides in the wild-type SRMS-regulated phosphoproteome (Figure 4.22 and Supplementary Table 9). Some of the significantly upregulated phosphosites included NUCL S206, NUCL S184, MYH9 S1943, HS90A S263 and MARCS S147. Overall, these phosphopeptides mapped to 60 upregulated and 1 downregulated proteins, respectively (Table 4.3 and Supplementary Table 9). We used the 0.5-fold cut-off for downregulation since our dataset comprised only 37 phosphopeptides below the median 1-fold differential abundance (p-value ≤ 0.05) and most of these peptides displayed only marginal downregulation in phosphorylation

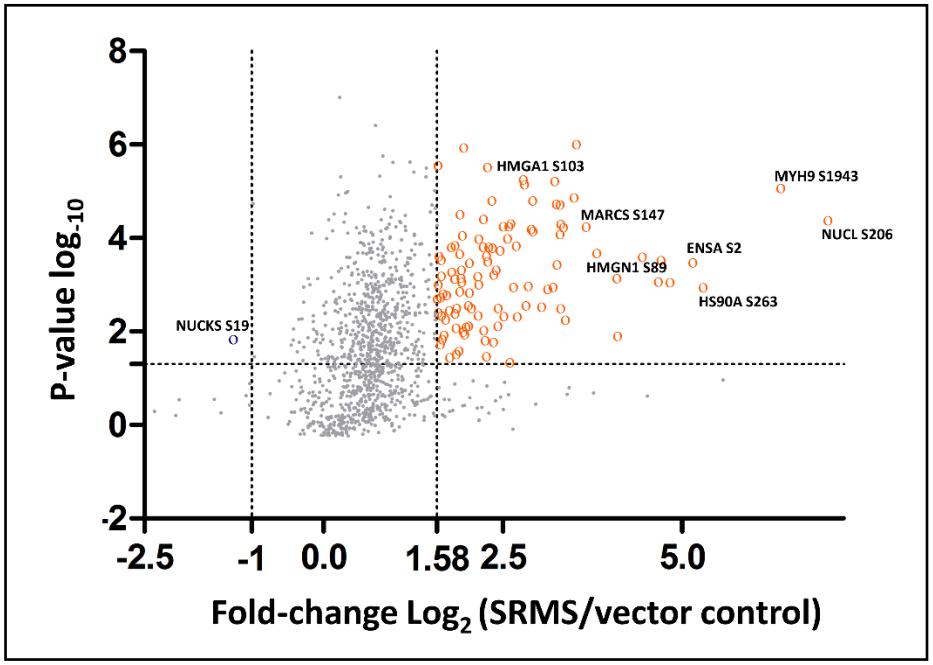


FIGURE 4.22: Identification of SRMS-regulated phosphosites. A scatter plot showing the the phosphopeptide log_2 fold-change (SRMS/control) plotted against the -Log_{10} p-value, highlighting the significantly regulated phosphopeptides (ANOVA p-value ≤ 0.05; upregulation fold-change cut-off = Log_2 ≥ 1.58 and downregulation fold-change cut-off = Log_2 ≤ -1). Upregulated phosphopeptides are highlighted in orange while the downregulated phosphopeptide is highlighted in blue. Selected upregulated phosphoproteins and corresponding phosphosites, namely HMGA1-S103, MARCS-S147, HMGN1-S89, ENSA-S2, HS90A-S263, MYH9-S1943 and NUCL-S206, are annotated in the scatter plot.

(SRMS/Control ~ 0.7-fold). Additionally, we reasoned that phosphopeptides that are at least two-fold more abundant in control than in the SRMS-regulated phosphoproteome would be likely to represent genuine SRMS-regulated hypophosphorylation events.

Table 4.3: List of SRMS-dependent upregulated phosphoproteins. Shown here is a subset of the top hyperphosphorylated/upregulated proteins and corresponding phosphosites, as identified in the wild-type SRMS phosphoproteome. Refer to Supplementary Table 1, using the weblink provided in the Bibliography section, to access the complete dataset comprising details on all regulated phosphosites and phosphoproteins.

Protein Accession (UniProt ID)	Phosphosite(s)	Protein Description
AHNK	S5731, S5739	Neuroblast differentiation-associated protein
AKA12	S248, S381, 732, S743, S749	A-kinase anchor protein 12
BAD	S118, S75	Bcl2-associated agonist of cell death
CALX	S554, S564	Calnexin
DKC1	S451, S453, S455, S458, S494	H/ACA ribonucleoprotein complex subunit 4
FOXO3	S55, T51	Forkhead box protein O3
HS90	S263, S252	Heat shock protein HSP 90-alpha
HMGA1	S102	High mobility group protein HMG-I/HMG-Y
IF2P	S107, S113	Eukaryotic translation initiation factor 5B
MAP1B	S1016, S1154, S1869, T1879	Microtubule-associated protein 1B
MARCS	S147, T150	Myristoylated alanine-rich C-kinase substrate
NUCL	S206, S184, S145	Nucleolin
NUDC	S136, S139, S145	Nuclear migration protein nudC
SFR19	Y305, T335, S998, T994, T1001	Splicing factor, arginine/serine-rich 19
SRRM1	S465, S775	Serine/arginine repetitive matrix protein 1
SRRM2	S1404, T1413	Serine/arginine repetitive matrix protein 2
TCP4	S9, S17, S19	Activated RNA polymerase II transcriptional coactivator p15

4.2.3.2 Functional annotation of the SRMS-regulated phosphoproteome

To better understand the cellular and biological processes mapped from the SRMS-regulated phosphoproteins, we performed functional gene enrichment analyses using the Ingenuity Pathway analyses (IPA) tool. We increased the stringency of our analyses by restricting functional annotations inferred exclusively from experimental observations. This led us to identify 30 molecular and cellular processes (Benjamini-Hochberg multiple testing p-value < 0.05) mapped from the majority of the hyperphosphorylated proteins (48/60 proteins or 80%) identified in the SRMS phosphoproteome (Figure 4.23A and Supplementary Table 10). These cellular processes broadly represented key functional categories that included protein synthesis (p-value range 1.11E-07 to 9.87E-03), cell cycle (p-value range 8.49E-03 to 1.03E-02), RNA post-transcriptional modification (p-value range 3.49E-08 to 1.95E-04), cell death and survival (p-value range 6.30E-06 to 2.76E-02), cell growth and proliferation (p-value range 1.35E-04 to 1.29E-02) and DNA replication and repair (p-value 7.05E-08 and 1.02E-04). Importantly, our analyses also led us to determine the predicted activation states of four molecular and cellular processes at a significant activation z-score threshold of ± 2. Specifically, cellular processes related to viral infection (20 proteins, z-score = 3.8, p-value = 5.45E-06) and DNA repair (10 proteins, z-score = 2.18, p-value = 7.05E-08) were predicted with a significantly increased activation state while apoptosis (17 proteins, z-score = -2.11, p-value = 3.94E-03) and necrosis (15 proteins, z-score = -2.48, p-value = 2.76E-02) were predicted to display an overall decreased activation state (Figure 4.23A and Supplementary Table 10). Some of the other interesting cellular/biological themes enriched from the hyperphosphorylated proteins included cell movement (p-value = 1.35E-04) and catabolism of protein (p-value = 9.87E-03) (Figure 4.23A and Supplementary Table 10). Analyses of the upstream regulators identified 5 proteins, namely, PI3K (Phosphoinositol-3-kinase), CDK9 (Cyclin-dependent kinase 9), CK2 (Casein kinase 2), CK2α (Casein kinase 2 catalytic subunit alpha) and FN1 (Fibronectin) (Figure 4.23B and Supplementary Table 10). Importantly, among these, two kinases were associated with significantly increased predicted activation- the tetrameric kinase complex, CK2 (z-score = 3.54, p-value = 7.41E-05), its monomeric catalytic subunit, CK2α (z-score = 4.87, p-value = 9.32E-19), and CDK9 (z-score = 2.81, p-value = 7.33E-05)) (Figure 4.23B and Supplementary Table 10). Further, our analyses also revealed that a small number of the hyperphosphorylated proteins were enriched in specific signaling pathways which included Telomerase signaling (-Log p-value = 2.33), PI3K/AKT signaling (-Log p-value = 2.18), Sirtuin

signaling (-Log p-value = 1.93) and Phospholipase C signaling (-Log p-value = 1.44) (Supplementary Table 10).

4.2.3.3 Analyses of phosphorylation motifs and predicted kinases

The amino acids surrounding a phosphosite constitute important recognition motifs for cognate kinases (Amanchy *et al.*, 2007; Turk, 2008; Ubersax and Ferrell, 2007). Our phosphoproteomics

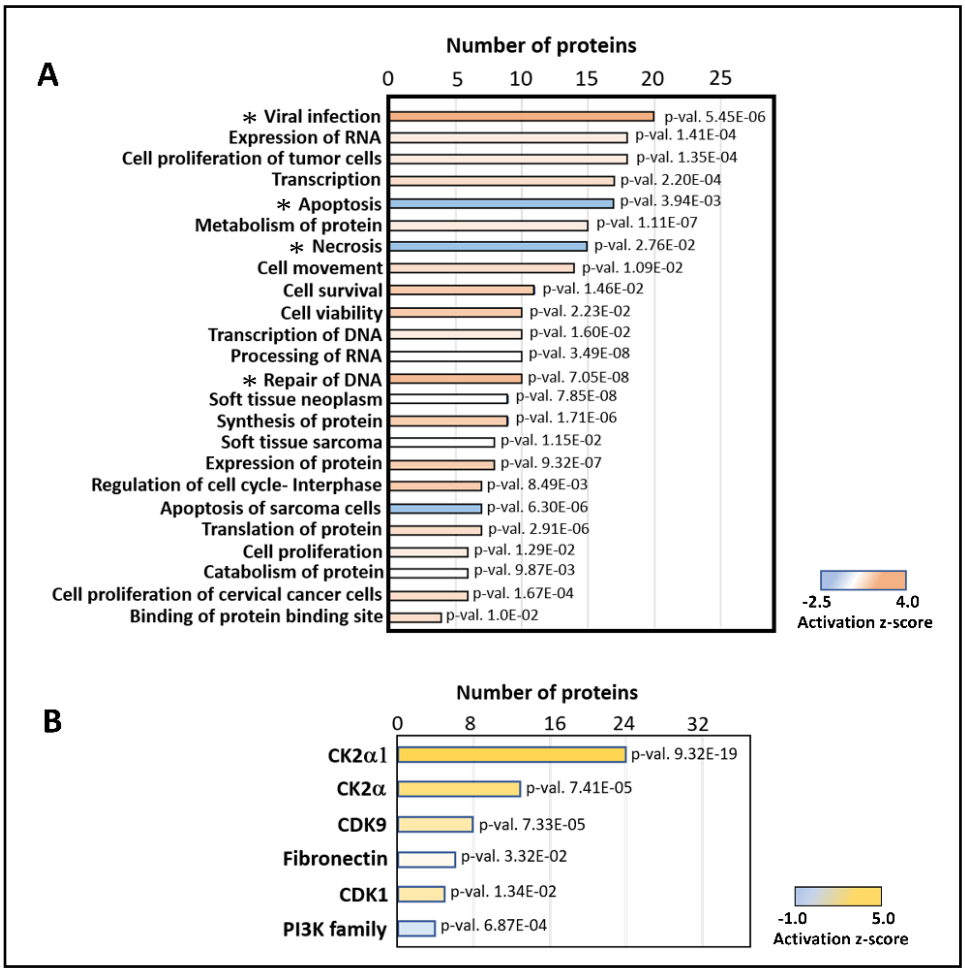
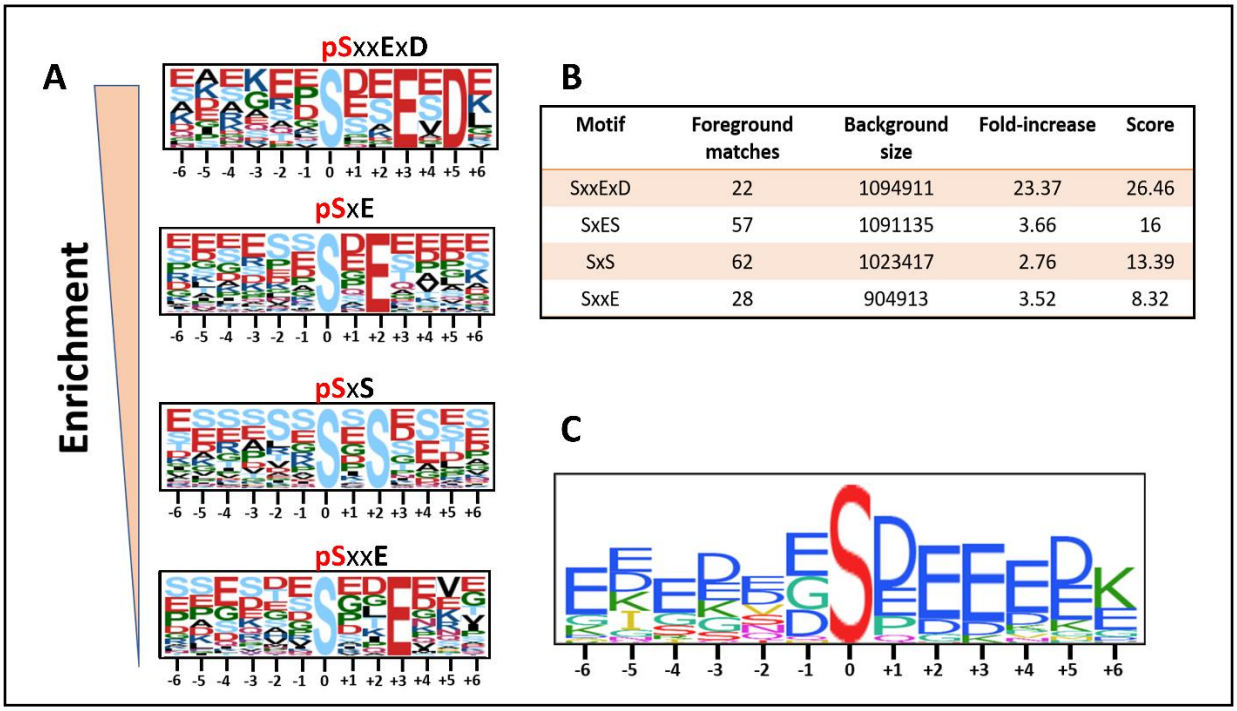


FIGURE 4.23: Functional enrichment analyses of the significantly altered phosphoproteins. A. IPA analyses of the top cellular and molecular processes enriched from upregulated phosphoproteins identified in the SRMS-regulated phosphoproteome (n=60; corresponding to upregulated phosphopeptides; SRMS/Control $Log_2 \geq 1.58$-fold). z-score indicates the predicted upregulation (z-score > 1) or downregulation (z-score < 0) of specific cellular and molecular processes. **B.** IPA analyses of upstream regulators for the upregulated phosphoproteins (n=60). z-score indicates the predicted activation (z-score > 1) or inactivation (z-score < 0) of the indicated upstream proteins. An asterisk ("*") mark indicates cellular processes with significant z-scores (z-score ≥ 2 or ≤ -2)

analysis identified 140 upregulated phosphopeptides that mapped to 60 proteins in the SRMS-regulated phosphoproteome. We assessed the enrichment of motifs among these upregulated phosphopeptides using the Motif-x and PHOSIDA (Phosphorylation Site Database) motif-enrichment tools (Chou and Schwartz, 2011; Gnad *et al.*, 2007). Our analyses using both the tools resulted in the enrichment of similar motifs (p-value = 10E-06) which primarily corresponded to phosphoserine sites (Figure 4.24A, 4.24B and 4.24C). No significant phosphothreonine or phosphotyrosine motifs were found due to their lower abundances in our dataset. Collectively, we identified four major motifs, all comprising positional variations in amino acids following the central phosphoserine residue. One significantly overrepresented motif comprised a glutamic acid and aspartic acid residue in the +3 and +5 positions (*SxxExD*) (Figure 4.24A, 4.24B and 4.24C).



Motif	Foreground matches	Background size	Fold-increase	Score
SxxExD	22	1094911	23.37	26.46
SxES	57	1091135	3.66	16
SxS	62	1023417	2.76	13.39
SxxE	28	904913	3.52	8.32

FIGURE 4.24: Motif-enrichment analyses of SRMS-dependent upregulated phosphopeptides. A. Motif-logos showing the significantly enriched motifs (p-value < 10E-06) identified by Motif-x (Chou and Schwartz, 2011). The positions of the amino acid residues C-terminal or N-terminal to the central phosphoresidue (position "0") are shown in the logos. The height of the amino acid residues is proportional to their enrichment at the specific position in the pool of the queried phosphopeptides. **B.** Table representing the scoring information relevant to the enriched motifs identified by Motif-x. **C.** A consolidated motif-logo generated by PHOSIDA (Gnad *et al.*, 2007) showing the enrichment of various amino-acid residues at specific positions relative to the central phosphoresidue (position "0").

Other top motifs comprised either a glutamic acid and a serine residue at the +2 and +3 positions (*SxES*), a serine at +2 position (*SxS*) or a glutamic acid residue at +3 position (*SxxE*) (Figure 4.24A, 4.24B and 4.24C).

We further analyzed the upregulated phosphopeptides using the NetworKIN tool (Linding *et al.*, 2007) to identify candidate kinases responsible for the site-specific phosphorylation of the associated motifs. The NetworKIN algorithm exploits information based on consensus motif recognition as well as context-specific factors such as the physical association and co-expression of kinases and substrates, to predict corresponding kinase-substrate relationships (Linding *et al.*, 2007; Van Hoof *et al.*, 2009). NetworKIN analyses of the hyperphosphorylated motifs identified 25 candidate serine/threonine kinases, corresponding to 16 kinase subfamilies, at a NetworKIN score > 3 (Figure 4.25 and Supplementary Table 11). Kinases found to potentially target multiple phosphosites included CDK1 (5 phosphosites, NetworKIN score range: 9.62-17.93), CK2α (19 phosphosites, score range: 3.17-35.63), GSK3 β (5 phosphosites, score range: 4.75-10.43), CK2α 2 (3 phosphosites, score range: 8.47-9.61), GRK2 (3 phosphosites, score range: 4.30-9.76) and CDK5 (2 phosphosites, score range: 3.42-14.02) (Figure 4.25 and Supplementary Table 11). Importantly, Casein Kinase 2 α (CK2α) was the among the highest scoring candidate kinases and found to target the maximum number of phosphosites in our dataset (Figure 4.25 and Supplementary Table 11). CK2α is a functionally independent catalytic subunit of the tetrameric holoenzyme, CK2 (Zhou *et al.*, 2016). The consensus recognition motifs for the acidophilic kinase, CK2α, are characterized by the presence of either glutamic acid or aspartic acid residues near the phosphorylation site (St-Denis *et al.*, 2015; Van Hoof *et al.*, 2009). Indeed, the motifs we identified by motif enrichment analyses (*SxxExD, SxE and SxxE*) conform to the CK2α consensus recognition sequences, thereby confirming CK2α as a major upstream kinase of the upregulated phosphosites. This also corroborated findings from our functional gene enrichment analyses using IPA where the holoenzyme CK2 and its catalytic module, CK2α were identified as major upstream regulatory kinases of the upregulated phosphoproteins (Figure 4.23B). Taken together, our findings suggest that CK2α represents a key serine/threonine kinase that is potentially activated in the SRMS phosphoproteome. This further indicates that CK2α and/or CK2 may likely represent key downstream targets of SRMS.

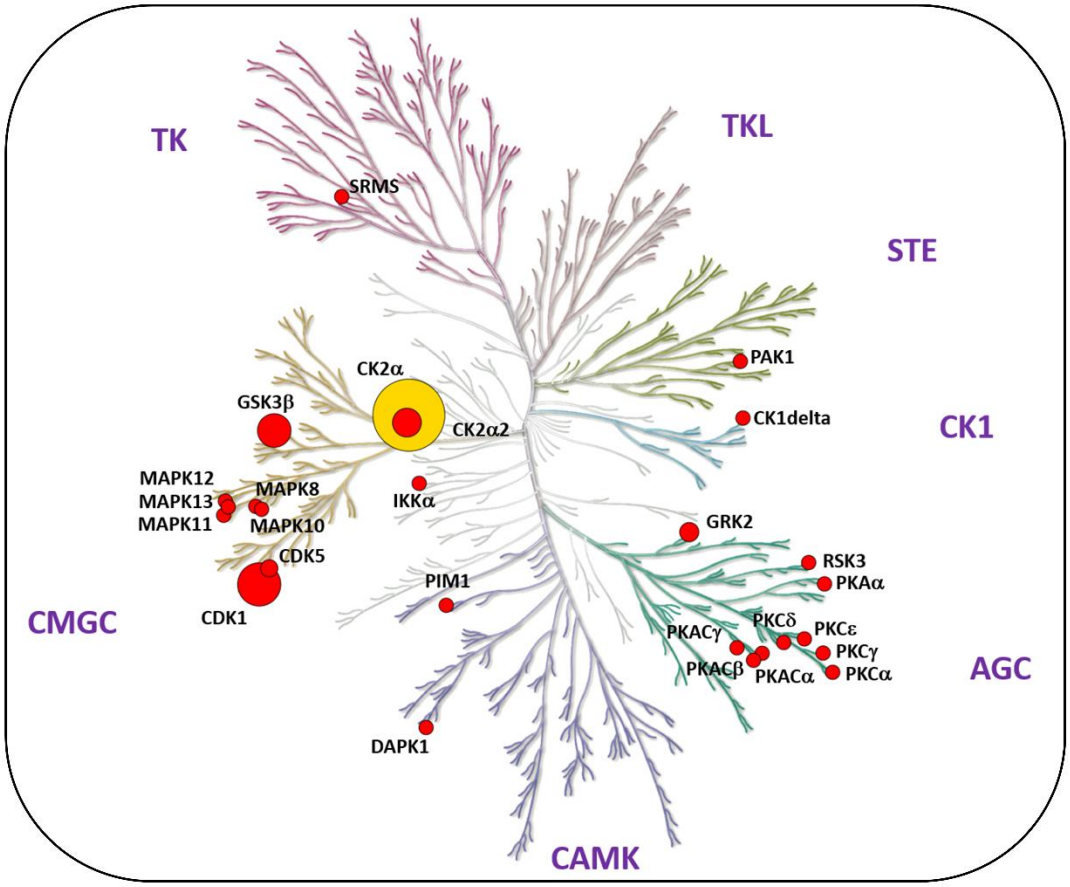


FIGURE 4.25: Identification of candidate kinases predicted to target the upregulated phosphosites in the SRMS-regulated phosphosites. Dendogram of the human kinome, constructed using KinMap (Eid *et al.*, 2017), highlighting the candidate kinases predicted to target the upregulated phosphosites identified in the SRMS-regulated phosphoproteome. Candidate kinases were identified by NetworKIN (Linding *et al.*, 2007) analysis. Node size is proportional to the number of the phosphosites targeted by the kinase. CK2α (node highlighted in yellow) represents the largest node in the kinome dendogram. All major kinase families are annotated in the dendogram which include: TK (Tyrosine Kinases), TKL (Tyrosine Kinase-Like), STE (Sterile kinases; homologs of the yeast STE7, STE11 and STE20 kinases), CK1 (Casein Kinase 1), AGC (comprising Protein kinase A/ PKA, PKG and PKC kinase sub-families), CAMK (Calcium/Calmodulin-dependent kinases) and CMGC (comprising cyclin-dependent kinase (CDK), mitogen-activated protein kinase (MAPK), glycogen synthase kinase (GSK) and CDC-like kinase (CLK)).

4.2.4 Discussion

Here we used quantitative mass spectrometry analyses to identify cognate serine/threonine phosphorylation events significantly altered in the SRMS-regulated phosphoproteome. We identified 60 phosphoproteins mapped from 140 phosphopeptides which were upregulated by at

least 3-fold in the SRMS-regulated phosphoproteome. To our knowledge, this represents the first study to investigate global serine and threonine phosphorylation changes induced by a bonafide non-receptor tyrosine kinase in the eukaryotic phosphoproteome.

Despite the ectopic expression of SRMS in cells, we identified only 1.3% phosphotyrosine peptides in contrast to our previous study where we reported multiple phosphotyrosine peptides (Goel *et al.*, 2018) (Figure 4.21C). The lysis of cells under less stringent conditions, such as by RIPA lysis buffer used in this study, and the use of TiO_2 resin, known to less favourably bind to phosphotyrosines, are factors that may explain the lower enrichment of tyrosine-phosphorylated peptides. The former may be an equally important factor since tyrosine phosphorylation is far more stringently regulated by phosphatase activity than serine/threonine phosphorylation (Sharma *et al.*, 2014) and as such the use of denaturing conditions during cell lysis have proved useful in such studies (Sharma *et al.*, 2014; Yoshida *et al.*, 2015; Zhang *et al.*, 2017). The relatively lower selectivity of TiO_2 resin towards phosphotyrosine peptides can also be demonstrated by the fact that we identified only 2 major phosphorylation sites on SRMS- Y299 and Y456, in this study. Our previous phosphotyrosine enrichment-based approach enabled the identification of several other phosphotyrosine sites on wild-type SRMS (Goel *et al.*, 2018). However, since the focus of this study was to identify phosphoserine and phosphothreonine events, the relatively poor enrichment of phosphotyrosine peptides did not pose a concern.

The cellular roles of SRMS have not been well characterized to date. In the present study, functional enrichment analyses of the upregulated phosphoproteins in the SRMS-regulated phosphoproteome identified apoptosis and necrosis as potentially important downregulated cellular processes in cells expressing ectopic SRMS. A previous study by Potts et al. implicated SRMS in the potential negative regulation of autophagy (Potts *et al.*, 2013), implying that SRMS may potentially be a positive regulator of stress-induced cell survival. This is in line with our present findings which not only projected a downregulation of apoptotic processes but also a potential upregulation of cell survival and proliferative processes in cells expressing ectopic SRMS, as determined by a positive z-score from IPA analyses (Figure 4.23A and Supplementary Table 10). Interestingly, processes related to DNA repair were also predicted to be upregulated (Figure 4.23A and Supplementary Table 10). As an example, HMGA1 was among the upregulated proteins enriched in DNA repair processes (Supplementary Table 10). We identified HMGA1 S102 as a hyperphosphorylated site in cells expressing ectopic SRMS (Supplementary Table 9).

HMGA1 S102 is a target site for phosphorylation by CK2 and studies have shown that the hyperphosphorylation of HMGA1 S102 impairs the DNA-binding ability of the protein (Domizio *et al.*, 1990; Palvimo and Linnala-Kankkunen, 1989; Schwanbeck *et al.*, 2001; Wang *et al.*, 2017). The dissociation of HMGA1 from the DNA in turn promotes efficient DNA repair, presumably by allowing various DNA repair factors to be recruited to the sites of DNA lesions, as reported previously (Adair *et al.*, 2005; Maloney *et al.*, 2007). Taken together, our data points towards a potential role of SRMS in modulating cellular DNA repair processes by regulating HMGA1 S102 phosphorylation. Further, HMGA1 has also been shown to inhibit apoptosis by suppressing p53-mediated transcriptional repression of apoptosis-related genes like Mdm2, Bax and p21 (Esposito *et al.*, 2010). This therefore also explains the cross-talk involving HMGA1 between cellular processes lined to DNA repair, apoptosis and cell proliferation in our dataset (Supplementary Table 10).

Analyses of the phosphosites using NetworKIN identified CK2 as one of the major candidate upstream kinases for the upregulated phosphosites identified in the SRMS-regulated phosphoproteome (Figure 4.25, Supplementary Table 11). This raises the possibility that SRMS may potentially function upstream of CK2 and as a candidate regulator of CK2 activity. CK2 is an active tetrameric serine/threonine kinase composed of two catalytic subunits (α and α') and two regulatory β-subunits (Lee *et al.*, 2013). Activation of CK2 is primarily regulated *in vivo* by inositol phosphates, phospholipase D2 (PLD2) and protein kinase C (PKC) (Korolchuk *et al.*, 2005; Lee *et al.*, 2013; Solyakov *et al.*, 2004). However, previous studies have identified c-Abl, BCR-Abl and Src-family kinases, Lyn and c-Fgr as regulators of CK2 activity (Donella-Deana *et al.*, 2003; Heriche and Chambaz, 1998). Specifically, the catalytic subunit of CK2, CK2α, was identified as a substrate of these kinases where phosphorylation by c-Abl or BCR-Abl was shown to inhibit CK2α activity while phosphorylation by Lyn or c-Fgr was shown to increase CK2α activity (Donella-Deana *et al.*, 2003; Heriche and Chambaz, 1998). Interestingly, our previous phosphotyrosine enrichment-based phosphoproteomics analysis found CK2α to be significantly tyrosine-phosphorylated exclusively in cells expressing ectopic wild type SRMS, projecting CK2α as a candidate SRMS substrate (Supplementary Table 11) (Goel *et al.*, 2018). Specifically, CK2α Y182, Y188 and Y323 were identified as hyperphosphorylated sites (Supplementary Table 11). Of these, Y182 represents a key residue lying in the activation loop of the CK2α subunit and the trans-autophosphorylation of this site has been shown to increase CK2α activity (Donella-Deana

et al., 2001). Additionally, CK2 Y188 has also been reported as another activation loop phosphosite contributing to CK2α activation, albeit to a lesser extent than CK2 Y182 (Donella-Deana *et al.*, 2001; Sarno *et al.*, 2002). However, findings from our previous phosphoproteomics study revealed that CK2α Y182/Y188 are candidate target sites of SRMS which imply that CK2 enzymatic activity may potentially be modulated by SRMS. This may highlight SRMS as a possible modulator of CK2α kinase-dependent functions.

CDK1/cdc2 is another upstream candidate kinase identified by NetworKIN analysis which was also identified as an upstream regulator by IPA analyses. Kinases Wee1, Myt1 and Lyn are known to regulate the activity of CDK1 by phosphorylating the inhibitory Y15 site on CDK1 (Fattaey and Booher, 1997; Uckun *et al.*, 1996; Watanabe *et al.*, 1995). Our previous phosphoproteomics study also identified CDK1 as a candidate target of SRMS where CDK1 Y19, Y270 and Y286 were found to be hyperphosphorylated (Supplementary Table 11) (Goel *et al.*, 2018). However, these sites on CDK1 have not been functionally characterized. Therefore, the mechanism by which CDK1 may be activated in the presence of SRMS, leading to the phosphorylation of the CDK1 consensus phosphosites, is not known.

It is important to note that the overall cellular and molecular processes mapped from the SRMS-regulated signaling intermediates identified in the present study were consistent with the major functional themes mapped from the candidate SRMS substrates in our previous study (Goel *et al.*, 2018). Some of the major functional themes enriched by the candidate SRMS substrates included RNA processing, viral processes, negative regulation of apoptosis, cell cycle regulation and protein ubiquitination (Figure 4.10 and Supplementary Table 5) (Goel *et al.*, 2018). These correlate well with the broad functional categories identified in the present study such as RNA post-transcriptional modifications (processing of RNA), viral infection, cell death and survival (apoptosis), cell cycle, and protein degradation, respectively (Supplementary Table 10). Collectively, our findings here indicate that some of the major cellular processes regulated by SRMS involve additional signaling factors characterized by serine/threonine phosphorylation events. Our study therefore presents key evidence that serine/threonine phosphorylation likely forms part of important secondary signaling events triggered by SRMS.

Overall, our global phosphoproteomic profiling reveals that the non-receptor tyrosine kinase, SRMS can indirectly regulate multiple signaling intermediates which are characterized by an altered serine/threonine phosphorylation status. The regulation of serine/threonine

phosphorylation events in the presence of SRMS is likely a result of SRMS-dependent direct or indirect modulation of serine/threonine kinase activity. In this context, the serine/threonine kinase, Casein kinase 2 (CK2) may likely represent one of the major downstream targets of SRMS. Overall, findings from our study form an important mechanistic resource for characterizing the cellular role(s) of SRMS in mammalian cells.

4.3 Characterizing the SRMS cytoplasmic punctae

4.3.1 Summary

Previous studies from our laboratory revealed that endogenous as well as ectopic SRMS protein localize predominantly to the cytoplasm and in distinct cytoplasmic punctate structures in various immortalized human cell lines (Goel *et al.*, 2013). The reasons underlying the characteristic punctate localization pattern of SRMS are unknown. SRMS possesses functional protein-protein binding SH3 and SH2 domains known to associate with its cellular substrates (Goel *et al.*, 2013). Therefore, it is possible that SRMS is localizes to specific cytoplasmic sites by interacting with and/or phosphorylating specific cellular proteins. These cellular proteins may be constituents of specific cytoplasmic organelles, implying that SRMS may be sequestered to and therefore colocalize with certain cytoplasmic organelles. To address this, colocalization of ectopic SRMS with known markers of specific cellular organelles, such as the endoplasmic reticulum, mitochondria, lysosomes, and golgi bodies, was examined. Further, whether the SRMS SH3, SH2 and/or kinase domains play any role in the formation of the cytoplasmic puncta was also investigated by fluorescence microscopy analyses. Overall, the data obtained showed that SRMS did not display significant colocalization with any of the organelles examined. However, the SRMS SH2 domain was found to partly contribute to the formation of the SRMS cytoplasmic puncta.

4.3.2 Results

4.3.2.1 Colocalization analyses of SRMS with cellular cytoplasmic organelles

To determine whether the SRMS cytoplasmic punctae results from the colocalization of the protein with specific cytoplasmic organelles, we performed immunofluorescence analyses using confocal microscopy. We selected well-characterized markers of four cytoplasmic organelles, namely the endoplasmic reticulum (ER), mitochondria, lysosomes and golgi bodies. GFP-tagged wild type SRMS or GFP alone were ectopically introduced in HEK293 cells via transient transfection. 24 hours-post transfection the cells were prepared for immunofluorescence analyses using the appropriate primary and secondary antibodies.

4.3.2.1.1 Colocalization analysis with endoplasmic reticulum (ER)

Primary antibodies against calnexin were used to label endoplasmic reticuli in HEK293 cells. Calnexin is an endoplasmic reticulum-based transmembrane chaperone protein that regulates the folding of glycoproteins within the ER (Ellgaard and Helenius, 2003). Firstly, in our immunofluorescence analyses the cytoplasmic pattern of localization of Calnexin observed was consistent with that reported by others as well as immunofluorescence data provided by various manufacturers' on their antibody product website (Fernandez-Valdivia *et al.*, 2006) (Figure 4.26). This therefore helped confirm the immunoreactive specificity of the antibody used. Further, the pattern of localization of ER was reproducibly observed with each experimental replicate we

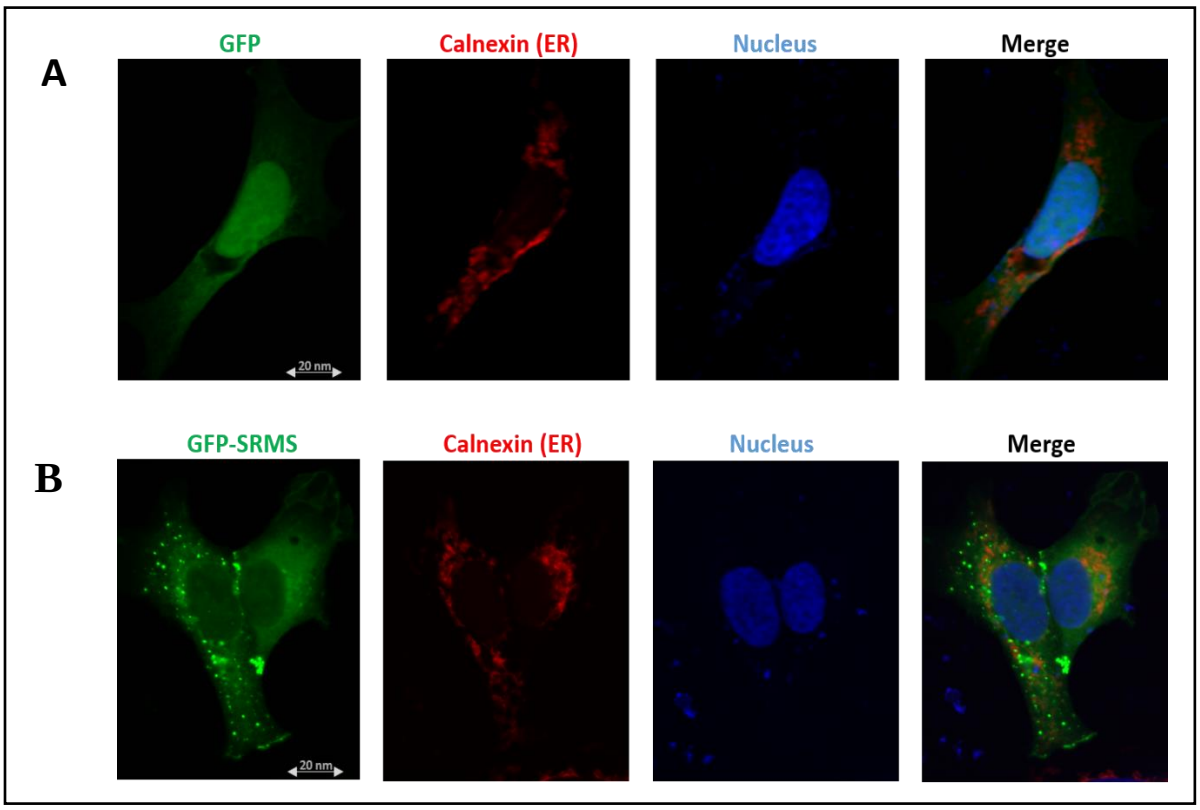


Figure 4.26: Colocalization analyses of Calnexin (endoplasmic reticulum) and wild type SRMS. HEK293 cells were seeded in glass coverslips in 6-well tissue culture dishes and transfected with vectors encoding: **A.** GFP alone (top panel) and **B.** GFP-SRMS. 24 hours post-transfection the cells were fixed and processed for immunofluorescence analyses. Primary antibodies against Calnexin (mouse IgG) and secondary goat anti-mouse IgG-Texas red (594 nm) antibodies were used to label ER. DAPI was used to counter-stain the nucleus. The cells were imaged at 40x magnification by confocal microscopy. Scale bar = 20 nm.

performed. As expected, GFP-SRMS localized to characteristic cytoplasmic punctae (Figure 4.26). However, no obvious or significant colocalization was observed between the ER and ectopically expressed GFP-SRMS (Figure 4.26).

4.3.2.1.2 Colocalization analysis with mitochondria

Colocalization analysis was performed between GFP-SRMS and mitochondria using antibodies against the mitochondrial marker mitofilin. Mitofilin is a transmembrane protein of the inner mitochondrial membrane and controls the mitochondrial cristae morphology and integrity (Ding *et al.*, 2015; von der Malsburg *et al.*, 2011). A characteristic and reproducible cytoplasmic localization of the mitochondria was observed which was consistent with the manufacturer's immunofluorescence data (Figure 4.27). No obvious or significant colocalization was observed

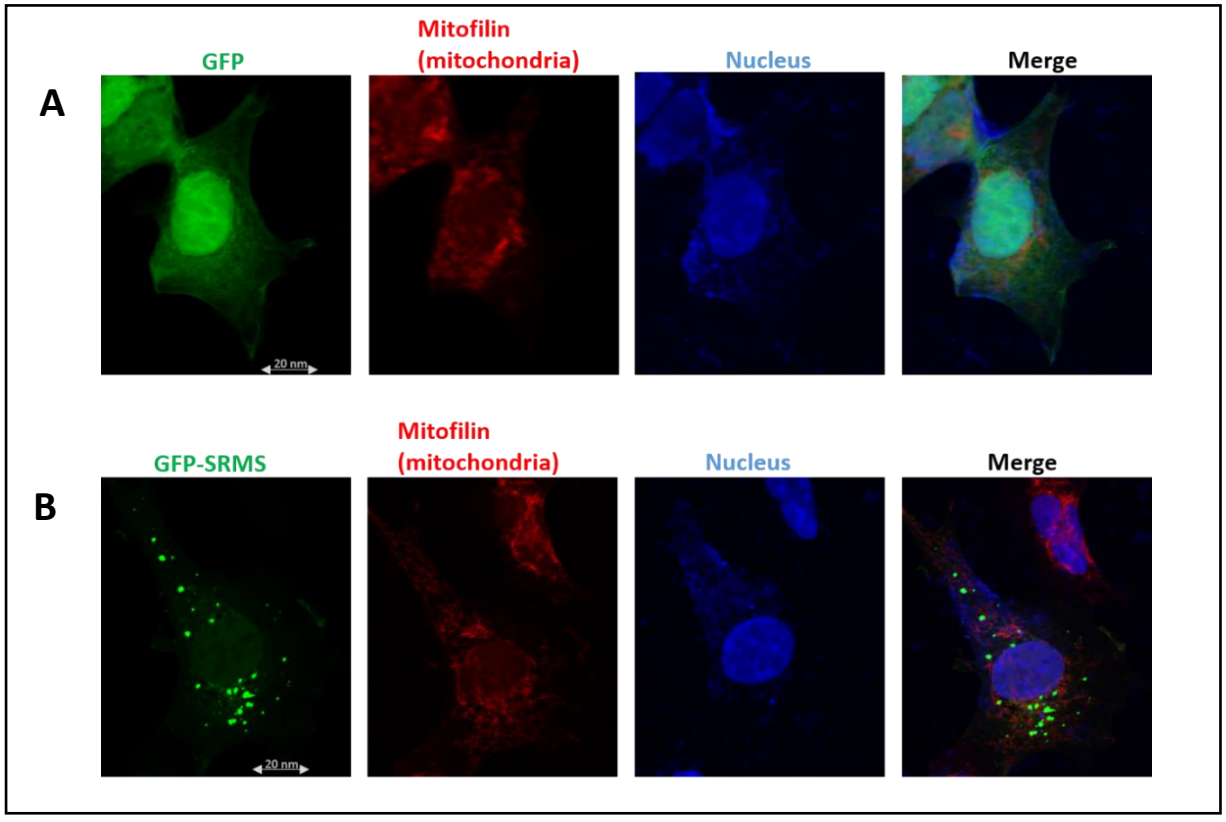


Figure 4.27: Colocalization analyses of Mitofilin (mitochondria) and wild type SRMS. HEK293 cells were seeded in glass coverslips in 6-well tissue culture dishes and transfected with vectors encoding: **A.** GFP alone (top panel) and **B.** GFP-SRMS. 24 hours post-transfection the cells were fixed and processed for immunofluorescence analyses. Primary antibodies against Mitofilin (rabbit IgG) and secondary goat anti-rabbit IgG-Texas red (594 nm) antibodies were used to label the mitochondria. DAPI was used to counter-stain the nucleus. The cells were imaged at 40x magnification by confocal microscopy. Scale bar = 20 nm.

between mitofilin and GFP-SRMS (Figure 4.27).

4.3.2.1.3 Colocalization analysis with golgi bodies

Colocalization analysis was performed between GFP-SRMS and golgi bodies using antibodies against golgin p97, a marker of golgi bodies. Golgin p97, also referred to as GOLGA A1, is a protein that is targeted to the trans-golgi network (TGN) of the golgi bodies and is involved in regulating the localization of other TGN-resident proteins (Luke *et al.*, 2005). From our immunofluorescence analyses, we observed a pattern of cytoplasmic localization of golgin p97 that was consistent with previous studies as well as with immunofluorescence data available with various manufacturers (Alzhanova and Hruby, 2006) (Figure 4.28). Our analyses did not reveal

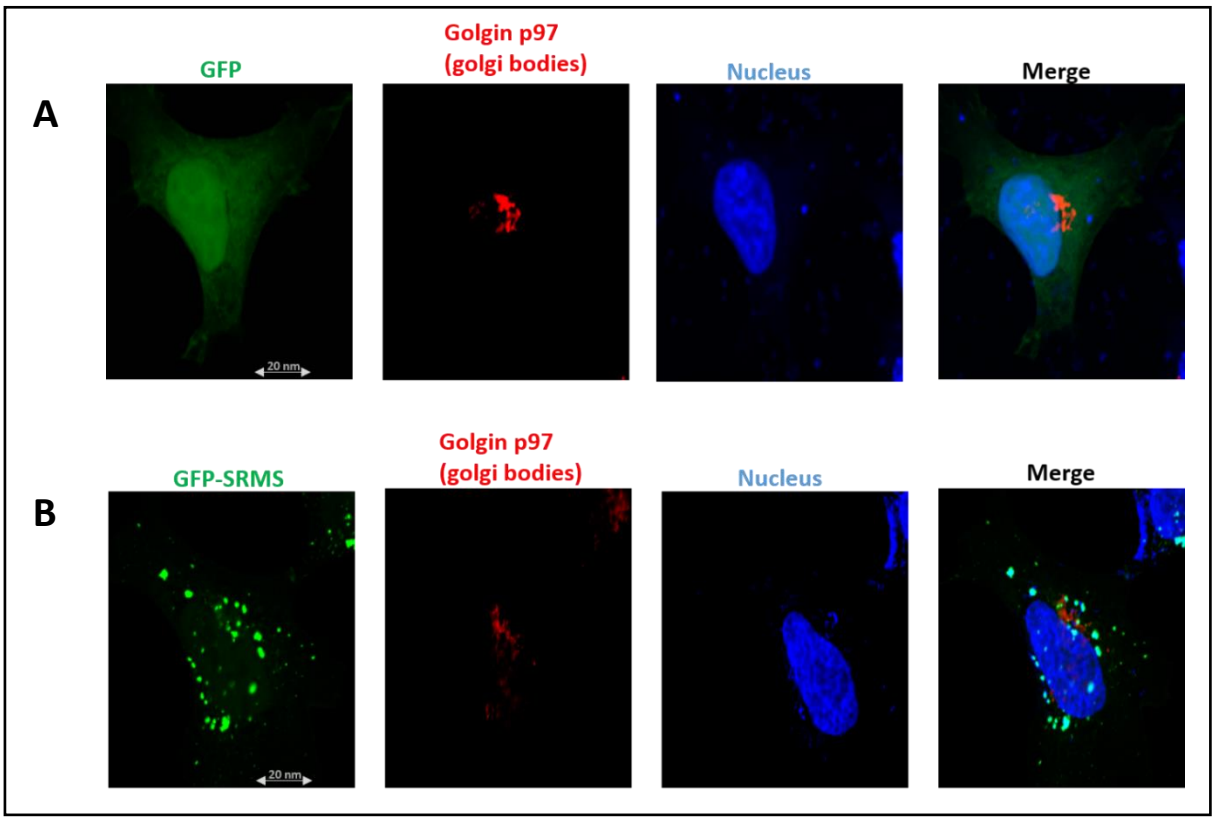


Figure 4.28: Colocalization analyses of Golgin p97 (golgi bodies) and wild type SRMS. HEK293 cells were seeded in glass coverslips in 6-well tissue culture dishes and transfected with vectors encoding: **A.** GFP alone (top panel) and **B.** GFP-SRMS. 24 hours post-transfection the cells were fixed and processed for immunofluorescence analyses. Primary antibodies against Golgin p97 (mouse IgG) and secondary goat anti-mouse IgG-Texas red (594 nm) antibodies were used to label the Golgi bodies. DAPI was used to counter-stain the nucleus. The cells were imaged at 40x magnification by confocal microscopy. Scale bar = 20 nm.

significant colocalization of GFP-SRMS with golgin p97 indicating that SRMS does not colocalize with golgi bodies (Figure 4.28).

4.3.2.1.4 Colocalization with lysosomes

SRMS colocalization with lysosomal bodies was tested using a commercially available LysoTracker™ stain that specifically and reversibly stains the cellular lysosomes of live cells. Weakly basic amines are known to accumulate in low-pH vesicles such as lysosomes. The LysoTracker™ stain essentially comprises weakly basic fluorescent amines that selectively permeate the membranes of live cells and accumulate in the lysosomes. The LysoTracker™ probe was used to stain lysosomes of live HEK293 cells expressing either GFP alone or GFP-SRMS (Figure 4.29). We observed a punctae-like localization of the lysosomes which was consistent

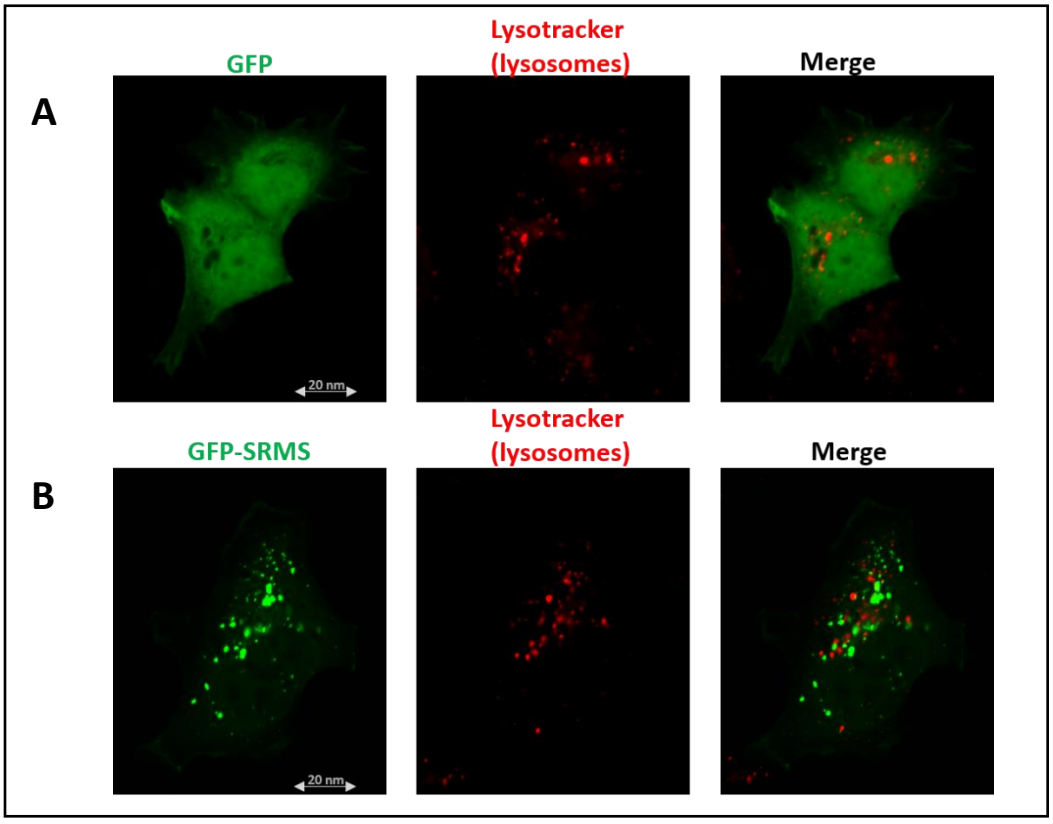


Figure 4.29: Colocalization analyses of lysosomes and wild type SRMS. HEK293 cells were seeded on glass coverslips in 6-well tissue culture dishes and transfected with vectors encoding: **A.** GFP alone (top panel) and **B.** GFP-SRMS. 24 hours post-transfection 75 nM LysoTracker™ stain was added directly to the culture media and the cells cultured for 2 hours. The cells were then fixed, mounted on glass slides and imaged by confocal microscopy at 40x magnification. Scale bar = 20 nm.

with the manufacturer's fluorescence microscopy data as well as other published studies (Figure 4.29) (Fossale *et al.*, 2004; Hashiguchi *et al.*, 2002). However, no significant colocalization was observed between GFP-SRMS and the LysoTracker™-stained lysosomal bodies (Figure 4.29).

Overall, our immunofluorescence analyses revealed that the SRMS cytoplasmic punctae did not display significant colocalization with either the mitochondria, endoplasmic reticulum, golgi bodies or lysosomes. This implied that other factors may be involved in establishment of the characteristic SRMS cytoplasmic punctae.

4.3.2.2 Role of SRMS functional domains in the formation of SRMS

Since we were unable to clearly establish the role of specific subcellular organelles in sequestering SRMS to cytoplasmic punctae, we asked whether the SRMS functional domains play a role in sequestering the kinase to these punctae, possibly via intermolecular interactions with other cellular proteins. We assessed the involvement of the three functional domains in SRMS, namely the SH3 (Src-homology 3), SH2 (Src-homology 2) and the kinase domain in promoting the formation of the characteristic cytoplasmic puncta displayed by the full-length protein. cDNA sequences corresponding to either the SRMS SH3, SRMS SH2 or the SRMS kinase domain were amplified to generate GFP-fusion constructs using the pEGFP-C1 vector (Figure 4.30A). These constructs were first exogenously expressed in HEK293 cells and expression of the respective proteins confirmed via immunoblotting analyses using antibodies against GFP (Figure 4.30B). The constructs were then ectopically introduced in HEK293 and subcellular localization determined via fluorescence microscopy (Figure 4.31). Our analyses demonstrated that cells transfected with either the SRMS SH3 domain or the SRMS kinase domain displayed a predominantly diffuse localization pattern (Figure 4.31B, 4.31C and Figure 4.32). This diffuse pattern of localization was comparable to the localization of the control GFP protein (Figure 4.31A). However, the SRMS SH2 domain displayed the characteristic cytoplasmic puncta-type localization in a significant proportion (55-60%) of transfected cells (Figure 4.31D and Figure 4.32). We also tested if the puncta localization displayed by the SRMS SH2 domain would be affected by the presence of other domains in SRMS. Since the SRMS SH3 domain alone produced a diffuse localization (Figure 4.31B and Figure 4.32), we tested the influence of this domain on

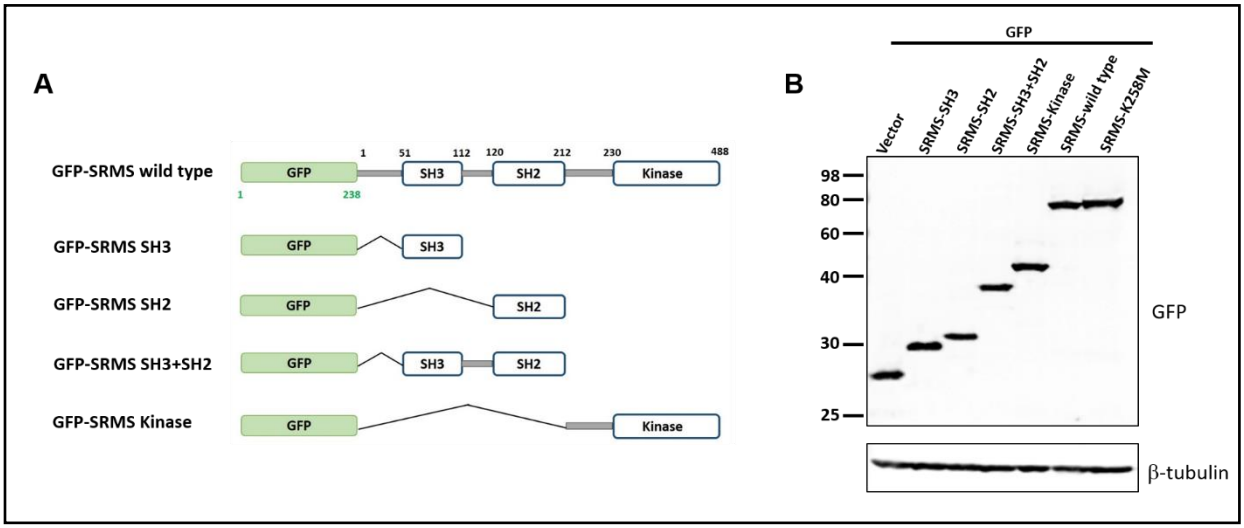


Figure 4.30: Generation of GFP-SRMS domain constructs. A. Shown here is a schematic representation of the various GFP-tagged constructs generated from the wild type SRMS cDNA sequence. cDNA sequences encoding either the SRMS SH3, SRMS SH2, SRMS SH3-linker-SH2 or the SRMS kinase domains were cloned into the pEGFP-C1 vector to generate the indicated GFP-tagged constructs. **B.** The indicated constructs were exogenously introduced in HEK293 cells and lysates probed via immunoblotting analyses using antibodies against GFP. β-tubulin was used as the loading control.

the localization of the SRMS SH2 domain. We generated a construct that encoded GFP-tagged SRMS SH3-linker-SH2 protein (the linker indicates the amino acid sequence between the SH3 and SH2 domains in SRMS). Through fluorescence microscopy analyses, we found that the GFP-SH3+SH2 protein also localized to cytoplasmic puncta in roughly the same proportion of transfected cells as the SRMS SH2 domain (Figure 4.32E and Figure 4.33). This implied that the SRMS puncta formation is at least partially driven by the SRMS SH2 domain and less likely by the other SRMS domains. Further, whereas wild type SRMS displayed the puncta localization in 90% of transfected cells, the SRMS SH2 domain alone displayed puncta localization in only 55-60% cells. This indicates that wild type SRMS-mediated tyrosine phosphorylation of target proteins may possibly serve as binding sites for the SRMS SH2 domain which may partly explain the absence of a similar proportion of cytoplasmic puncta with the SH2 domain alone.

SH2 domains characteristically bind to tyrosine-phosphorylated sequences on proteins (Liu *et al.*, 2012; Pawson *et al.*, 2001). We asked whether the cytoplasmic puncta localization displayed by the SRMS SH2 domain is due to the intermolecular interactions of this domain with native tyrosine-phosphorylated proteins in cells. To test this possibility, colocalization of the SRMS SH2 domain with endogenous tyrosine-phosphorylated proteins was assessed via immunofluorescence

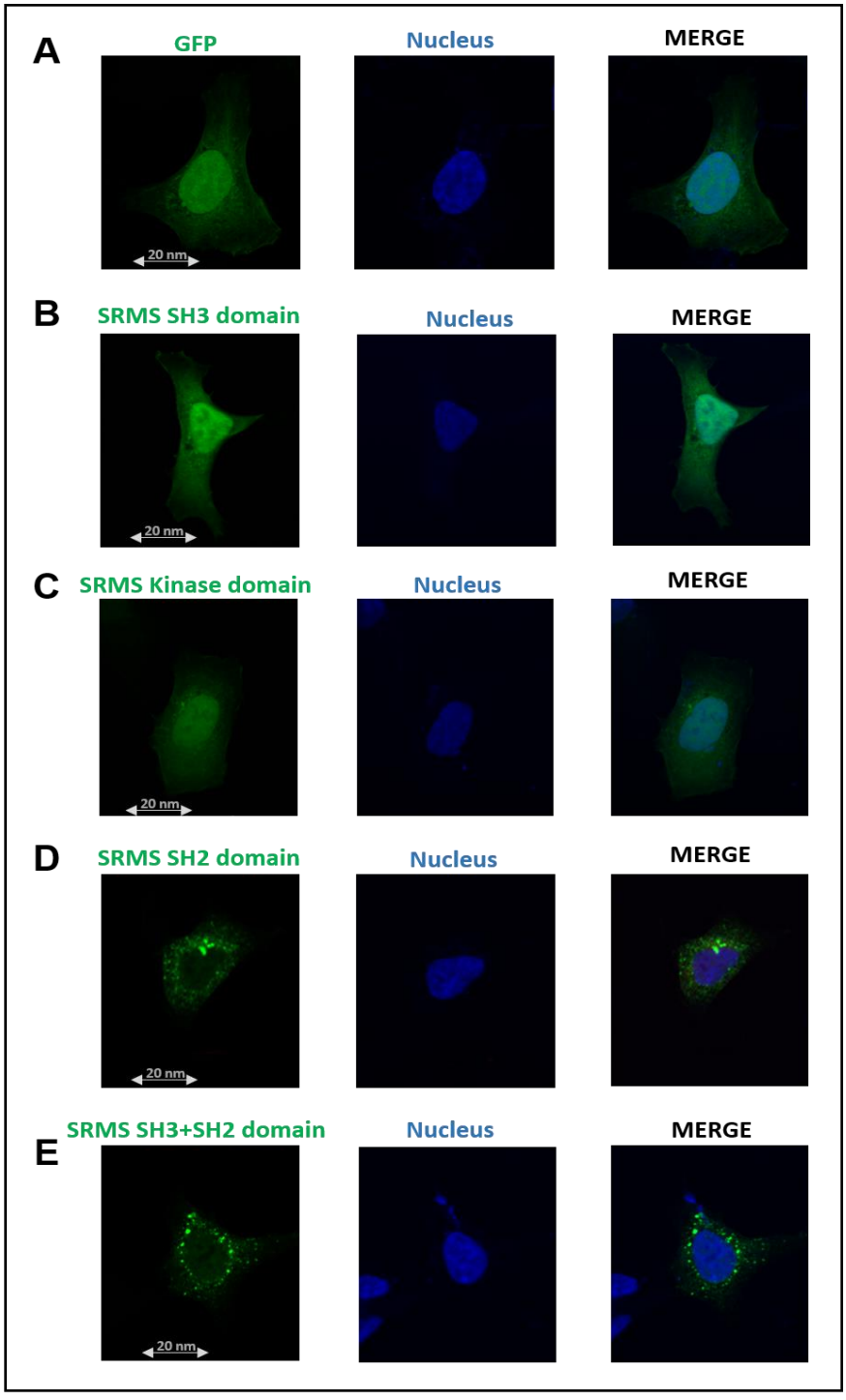


Figure 4.31: Subcellular localization of SRMS functional domains. HEK293 cells were seeded in glass coverslips in 6-well tissue culture dishes and transfected with vector encoding: **A.** GFP alone; **B.** GFP-SRMS SH3 domain, **C.** GFP-SRMS Kinase domain, **D.** GFP-SRMS SH2 domain and **E.** GFP-SRMS SH3-linker-SH2. Transfected cells were fixed, permeabilized and the nuclei stained with DAPI. Images were acquired at 40x magnification on a confocal microscope. Scale bar = 20 nm.

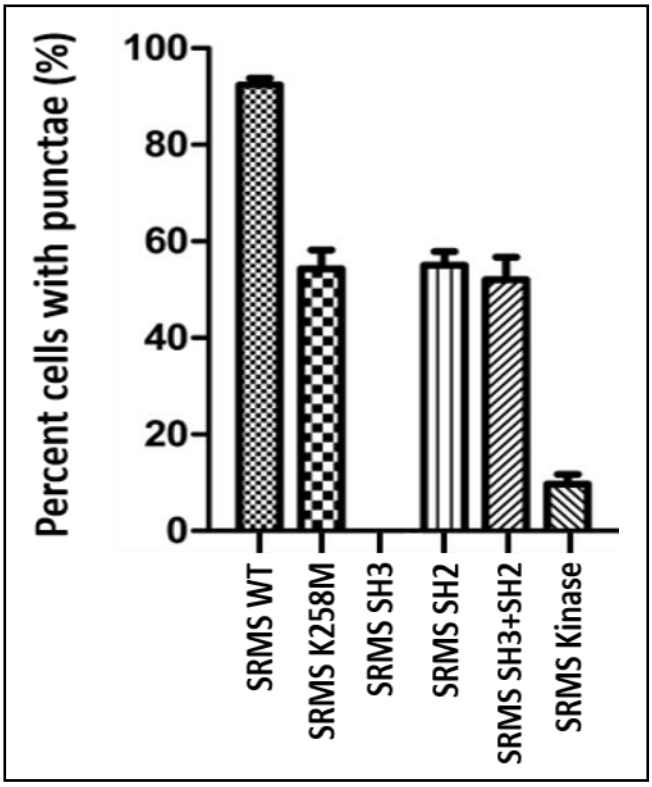


Figure 4.32: Quantification of puncta localization displayed by SRMS variants. HEK293 cells were seeded in glass coverslips in 6-well tissue culture plates and transfected with vectors encoding either the GFP-tagged wild type SRMS (SRMS WT), GFP-SRMS K258M mutant, GFP-SRMS SH3 domain, GFP-SRMS SH2 domain, GFP-SRMS SH3+SH2 domains or GFP-SRMS kinase domain. Transfected cells were fixed, permeabilized and the nuclei stained with DAPI. The percent transfected cells displaying puncta-localization, in each transfected condition, are plotted on the y-axis. 200 transfected cells were counted and the number of cells displaying puncta-localization calculated as the percentage of total transfected cells and presented as mean ± SEM. Data were obtained from 3 experimental replicates (n=3).

analyses using HEK293 cells expressing ectopic GFP-SRMS SH2. General phosphotyrosine antibodies were used to detect the localization of the total endogenous tyrosine-phosphorylated proteins in these cells. Our immunofluorescence analyses revealed a generally diffuse localization pattern of total tyrosine-phosphorylated proteins in HEK293 cells indicating the presence oftyrosine phosphorylated proteins in both, the nucleus and cytoplasm (Figure 4.33A- 4.33C). Further, we observed that the cytoplasmic puncta produced by the SRMS SH2 domain only partly colocalized with total tyrosine-phosphorylated proteins in transfected cells (Figure 4.33C). We next asked if the SRMS-SH2 domain-driven cytoplasmic puncta formation is exclusively dependent on the ability of the domain to bind tyrosine-phosphorylated proteins. To do this, we sought to generate a mutant form of the SRMS SH2 domain that is defective in phospho-tyrosine

binding. SH2 domains possess a highly conserved signature phosphotyrosine recognition motif namely, the "FLVRES" motif, that forms an integral part of the SH2 domain phosphotyrosine-binding pocket (Campbell and Jackson, 2003; Liu *et al.*, 2012). Previous studies have reported the involvement of various residues within this motif in contacting phosphotyrosyl residues (Mayer *et al.*, 1992; O'Reilly *et al.*, 2000; Safari *et al.*, 2011). Additionally, other residues, usually

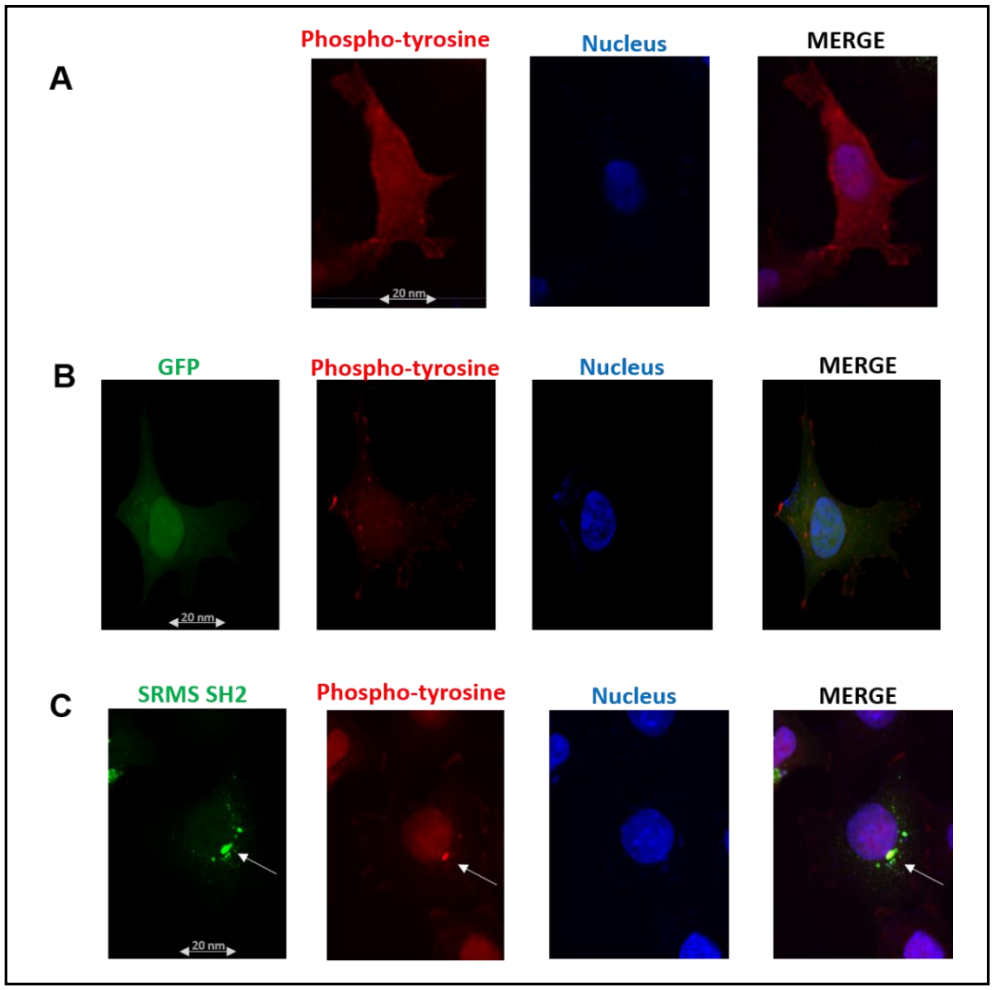


Figure 4.33: Colocalization analyses of SRMS SH2 domain and native tyrosine-phosphorylated proteins. HEK293 cells were seeded in glass coverslips in 6-well tissue culture dishes. **A**. Untransfected cells, or cells transfected with **B.** GFP alone or **C.** GFP-SRMS SH2 were fixed and processed for immunofluorescence analyses. General phosphotyrosine antibodies and secondary goat anti-mouse IgG-Texas red (594 nm) antibodies were used to label total tyrosine phosphorylated proteins in cells. DAPI was used to counter-stain the nucleus. The cells were imaged at 40x magnification by confocal microscopy. The SRMS cytoplasmic punctae are indicated in white arrows. Scale bar = 20 nm.

E and S/T, positioned immediately after the FLVRES motif are also known to play a role in mediating interactions with the phosphotyrosyl residue (Amanchy *et al.*, 2007; Campbell and Jackson, 2003). The conserved phosphotyrosine-recognition motif in the SRMS SH2 domain corresponds to the sequence, FLIRPS (amino acids 144-149). The SRMS S149 residue in the FLIRPS is succeeded by the residues ESS which are usually conserved and in other SH2 domains have been shown to be important for phosphotyrosyl-binding (Amanchy *et al.*, 2007; Campbell and Jackson, 2003). We generated a mutant SRMS SH2 domain where all residues in the FLIRPS-ESS sequence were substituted with alanine residues. We next determined the localization pattern of this SRMS SH2 domain mutant together with the wild type SRMS SH2 domain. Vectors encoding GFP-tagged wild type SRMS SH2 or SH2 mutant were ectopically introduced in HEK293 cells and subcellular localization determined via fluorescence microscopy analyses. We reasoned that if the cytoplasmic puncta localization displayed by the SRMS SH2 domain was dependent on the FLIRPS motif-mediated phosphotyrosine-protein binding, then disrupting the FLIRPS motif should impair the cytoplasmic puncta pattern of localization. Contrary to our expectations, we observed that like the wild type SRMS SH2 domain, the SRMS SH2 domain mutant also localized to cytoplasmic punctae in the same proportion of transfected cells (Figure 4.34A and 4.34B). In both cases roughly 50-55% transfected cells displayed cytoplasmic puncta localization of the proteins (Figure 4.34 B). This leads to the possible conclusions that (a) there may be structural cues, other than those involving the "FLVRES" phosphotyrosine-recognition motif, that dictate phosphotyrosine binding in the SRMS SH2 domain, or (b) there may be factors completely independent of the phosphotyrosine-binding ability of the SRMS SH2 domain that contribute to the formation of the cytoplasmic puncta. Further studies on the characterization of the SRMS cytoplasmic puncta will form part of the future experiments in our laboratory.

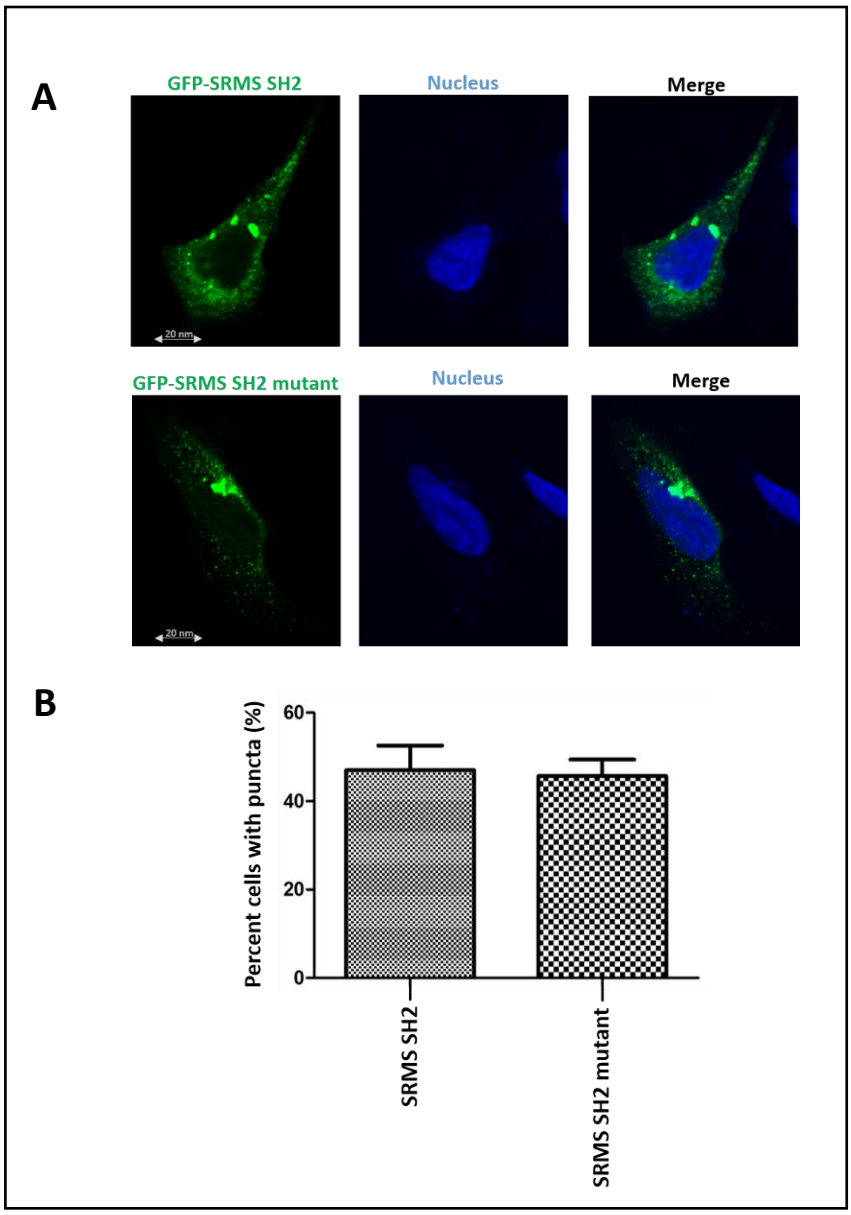


Figure 4.34: Quantification of punctae localization displayed by the SRMS SH2 domain and the SH2-domain mutant. **A.** HEK293 cells were seeded in glass coverslips in 6-well tissue culture plates and transfected with vectors encoding either the GFP-SRMS SH2 domain or the GFP-SRMS SH2 mutant harboring mutations in the FLIRPS-ESS sequence. Transfected cells were fixed, permeabilized and the nuclei stained with DAPI. Images were acquired at 40x magnification on a confocal microscope. Scale bar = 20 nm. **B.** Cells displaying puncta-localization were quantified. The percent transfected cells displaying puncta-localization, in each transfected condition, are plotted on the y-axis. 200 transfected cells were counted and the number of cells displaying puncta-localization calculated as the percentage of total transfected cells and presented as mean ± SEM. Data were obtained from 3 experimental replicates (n=3).

5.0 General Discussion

5.1 Conclusions

Non-receptor tyrosine kinases are key enzymes implicated in the regulation of mammalian growth and development. These enzymes engage in interactions with and/or phosphorylation of target cellular proteins to regulate essential cellular processes in mammalian cell biology such as cell growth, proliferation and differentiation. SRMS is a member of the BRK family of non-receptor tyrosine kinases whose other two members include BRK and FRK. BRK and FRK represent well-studied kinases in the context of their cellular substrates, signaling intermediates and overall cellular roles. SRMS on the other hand is the least studied member of the BRK family. Less is known about the cellular substrates, signaling intermediates and cellular functions played by the kinase. Previous studies have shown that SRMS is, biochemically, an active tyrosine kinase which is able to associate with and phosphorylate its two known substrates, Dok1 and BRK(Goel *et al.*, 2013; Miah *et al.*, 2014). Other studies have reported SRMS protein expression in various mammalian tissues and immortalized cell lines (Goel *et al.*, 2013). Identification of the cellular substrates and signaling intermediates of SRMS would be important to understand the possible cellular roles played by the kinase in mammalian cells.

We employed quantitative mass spectrometry-based phosphoproteomics analyses to identify the candidate cellular substrates and signaling intermediates of SRMS. Firstly, using an antibody-based phosphotyrosine-enrichment approach, we identified 663 candidate SRMS substrates, from cells expressing ectopic wild type SRMS. Analyses of the sequences surrounding the phosphosites in the candidate substrates revealed novel SRMS substrate motifs, among which the *KxY* and *KxxY* motifs (where Y represents the phosphosite in the candidate SRMS substrates). were identified as the major consensus motifs. We confirmed findings from our mass spectrometry analyses by using a high-throughput validation approach, employing customized peptide arrays comprised of peptides derived from the candidate SRMS substrates. We further independently validated Vimentin and Sam68 as bonafide SRMS substrates by demonstrating that SRMS colocalized with and phosphorylated both targets *in vitro* and *in vivo*, through ectopic-expression studies in mammalian cell lines. At the endogenous level, we also showed that SRMS promoted Sam68-tyrosine phosphorylation in EGF-dependent manner in MDA-MB 231 cells, thereby highlighting the significance of Sam68 as a potential physiological substrate of SRMS.

Various studies have shown that by engaging in substrate-phosphorylation and/or protein-protein interactions, tyrosine kinases can potentially influence the activation of serine/threonine kinases as part of the greater mammalian kinome. Such events occurring downstream of tyrosine kinases constitute key tyrosine kinase signaling intermediaries. We were interested in identifying the potential phosphoserine/phosphothreonine changes occurring in cells expressing ectopic SRMS. Using a global phosphoproteomic-profiling approach, we identified 60 proteins that were hyperphosphorylated on phosphoserine/phosphothreonine sites. Motif-enrichment analyses revealed that the phosphosites on these hyperphosphorylated proteins were candidate targets of 25 different serine/threonine kinases. Casein kinase 2 was identified as a major serine/threonine kinase. Importantly, Casein kinase 2 was also identified as a candidate SRMS substrate from our phosphotyrosine enrichment-based MS analyses. We specifically identified the activation loop Y182 site, the phosphorylation of which promotes CK2 activation, as a candidate target of SRMS. This implies that SRMS may potentially be involved in the positive regulation of Casein kinase activity and Casein kinase-dependent cellular functions in cells. Further experiments are necessary to characterize the possible role of SRMS in the regulation of Casein kinase activity and functions.

Importantly, bioinformatics-based functional pathway analyses of both, the candidate SRMS substrates and signaling intermediates, revealed a major overlap in the types of cellular processes regulated by these proteins. This indicates that the targets we identified in our mass spectrometry experiments at large, are not part of random cellular events but instead a part of a SRMS-regulated concerted group of proteins partaking in functionally-related cellular processes.

Our findings related to the characterization of the SRMS cytoplasmic puncta ruled out the possibility that SRMS is sequestered to specific subcellular organelles such as the mitochondria, endoplasmic reticulum, golgi bodies and lysosomes. Our data, however, point to the possibility that the SH2 domain in the SRMS protein partially regulates the formation of the SRMS cytoplasmic punctae. The composition of the SRMS cytoplasmic punctae and how the SRMS SH2 domain potentially sequesters the SRMS protein to these cytoplasmic punctate structures is not clearly understood and will require further investigation.

Overall, by employing advanced mass-spectrometry techniques this project has identified numerous biologically-relevant mammalian cellular targets of SRMS. This will serve as an important resource for future mechanistic studies aimed at understanding the cellular and physiological roles of SRMS.

5.2 Future Directions

5.2.1 Investigating the cellular function of SRMS

5.2.1.1 Characterizing the functional link between SRMS, Sam68 and vimentin

Our phosphoproteomics analyses identified multiple candidate substrates and signaling intermediates of SRMS. We identified and validated Sam68 as a novel SRMS substrate. Sam68 is an RNA-binding protein and tyrosine phosphorylation of Sam68 by BRK, for instance, has been shown to negatively regulate the RNA-binding properties of the protein (Derry *et al.*, 2000; Lin *et al.*, 1997). It would be interesting to investigate whether SRMS affects the RNA-binding properties of Sam68 upon phosphorylation. This is especially relevant since the SRMS-targeted site on Sam68 (Y167) lies in the nucleic acid-binding, KH domain of Sam68 and phosphorylation of the Y167 residue within the Sam68 KH domain may play a role in regulating the RNA-binding properties of this domain (Valverde *et al.*, 2008). Other studies have shown that Sam68 sequesters BRK to the nucleus in Sam68 nuclear bodies (SNBs) (Lukong *et al.*, 2005). The phosphorylation of Sam68 by BRK in the nucleus was shown to suppress cell cycle progression, as studied in rat astrocytes (Lukong *et al.*, 2005). Our findings here showed that SRMS colocalizes with Sam68 in the cytoplasm. Cytoplasmic Sam68 has been previously found to be associated with polysomes indicating that in the cytoplasm, Sam68 is likely implicated in protein synthesis by promoting the translation of specific RNAs (Paronetto *et al.*, 2006). It may be possible that in contrast with BRK, SRMS plays a role in promoting the cytoplasmic localization of Sam68 where SRMS-targeted phosphorylation of Sam68 is potentially implicated in mRNA translation. This may be worth investigating in future studies.

We also identified vimentin as another novel SRMS substrate. Vimentin intermediate filaments are known to play a key role in cell migration and cell mitosis (Chou *et al.*, 1990; Ivaska *et al.*, 2007). Site-specific phosphorylation of vimentin primarily on serine residues is functionally well characterized in the context of intermediate filament dynamics (Eriksson *et al.*, 2004; Izawa and Inagaki, 2006). For instance, phosphorylation of vimentin on S38 and S72 by PAK1 kinase leads to impaired intermediate filament assembly (Eriksson *et al.*, 2004). Tyrosine-phosphorylation of vimentin, however, is not well-documented in the literature. Given our findings here that SRMS colocalization with vimentin likely occurs in vimentin intermediate filaments, it would be interesting to investigate whether SRMS plays a role in the regulation of vimentin intermediate filament assembly/disassembly.

5.2.1.2 Studying the potential role of SRMS in the regulation of casein kinase activation

Casein kinase 2 was identified in this work as a candidate SRMS target from two independent phosphoproteomic screens. Specifically, the activating-autophosphorylation tyrosine residue, Y182 and the partial activating residue Y188, in casein kinase 2 were found to be hyperphosphorylated in SRMS-transfected cells (Donella-Deana *et al.*, 2001; Sarno *et al.*, 2002). Further, candidate casein-kinase substrates were also found to be hyperphosphorylated in these cells, collectively indicating that SRMS may function as an upstream activator of casein kinase. Casein kinase 2 is a ubiquitously-expressed protein implicated in essential cellular functions linked to, for instance, cell division, cell viability and damage-induced DNA-repair (Lebrin *et al.*, 2001; Olsen *et al.*, 2012; Ruhs *et al.*, 2017; Yefi *et al.*, 2011). Further studies aimed at investigating Casein kinase 2 as a SRMS substrate and a key downstream target, would be important. In the context of its role in cellular DNA-repair, Casein kinase 2-dependent regulation of DNA-repair through HMGA1 S102 phosphorylation (one of the CK2 targets identified in our phosphoproteomic screen) would be worth investigating under conditions of modulated SRMS expression.

5.2.1.3 Characterizing the role of SRMS as a regulator of BRK activity *in vivo*

Though BRK is not directly linked to the findings reported in this work, a previous published study identified BRK Y447 as a target site for phosphorylation by SRMS (Fan *et al.*, 2015). Given that pY447 is the c-Src pY530-equivalent negative regulatory site in BRK, it would be important to characterize the effect of SRMS-targeted Y447-phosphorylation on BRK enzymatic activation (Goel and Lukong, 2015; Okada, 2012). This will help understand the potential role of SRMS as a regulator of BRK kinase-dependent functions *in vivo*. BRK is highly expressed in breast cancer cells and we previously also reported SRMS expression in various breast cancer cell lines and tissues (Goel and Lukong, 2015; Goel *et al.*, 2013; Miah *et al.*, 2012). Future studies aimed at investigating the effect of modulated SRMS expression on BRK activation *in vivo* may highlight a new mechanistic paradigm of BRK activation in breast cancer cells. As such, by potentially negatively regulating BRK kinase activity, SRMS may play a tumor suppressor role in breast cancer cells where the enzyme suppresses the kinase-dependent oncogenic functions of BRK. Such impact of SRMS in breast cancer cells will require further investigations. Additionally,

identifying the molecular cues underlying the potential spatio-temporal enzymatic regulation of BRK by SRMS in mammalian cells would also be important to study.

5.2.2 Characterizing the SRMS cytoplasmic puncta

Our findings in this work showed that a significant proportion of cells expressing ectopic SRMS SH2 domain displayed the punctate localization, characteristic of the wild-type full length SRMS protein. This implied that the SRMS puncta may be driven at least in part by the SRMS SH2 domain. Future work would be necessary to characterize the mechanism by which SRMS is sequestered to these punctae. To this end, an important direction of research may lie in studying the SRMS interactome. It is possible that specific SRMS-interacting partners may represent key constituents of the SRMS cytoplasmic punctae and modulating the binding of SRMS with these interacting proteins may help characterize these punctate structures further. In this context, the use of mass spectrometry-based proximity-dependent biotin identification (BioID) method may be well suited to identify the SRMS cellular interactors (Lambert *et al.*, 2015). The method involves the exogenous expression of a modified bait protein, fused to a promiscuous bacterial biotin ligase (BirA R118G mutant), that tethers to bonafide protein interactors *in vivo* upon biotinylation (achieved by adding biotin to the culture media). The bait protein can then be immunoprecipitated and the bound proteins identified via LC-MS analyses. Using this approach protein-protein associations/interactions, that are either weak or transient in nature, can be identified (Lambert *et al.*, 2015). It is expected that this approach may be better able to define the cellular constituents of the SRMS cytoplasmic punctate structures and further our understanding of how and perhaps why SRMS localizes to these punctate structures.

6.0 Bibliography

Abram, C.L., and Lowell, C.A. (2008). The diverse functions of Src family kinases in macrophages. Front Biosci *13*, 4426-4450.

Adair, J.E., Kwon, Y., Dement, G.A., Smerdon, M.J., and Reeves, R. (2005). Inhibition of nucleotide excision repair by high mobility group protein HMGA1. J Biol Chem *280*, 32184-32192.

Akerblom, B., Anneren, C., and Welsh, M. (2007). A role of FRK in regulation of embryonal pancreatic beta cell formation. Mol Cell Endocrinol *270*, 73-78.

Alzhanova, D., and Hruby, D.E. (2006). A trans-Golgi network resident protein, golgin-97, accumulates in viral factories and incorporates into virions during poxvirus infection. J Virol *80*, 11520-11527.

Amanchy, R., Kandasamy, K., Mathivanan, S., Periaswamy, B., Reddy, R., Yoon, W.H., Joore, J., Beer, M.A., Cope, L., and Pandey, A. (2011). Identification of Novel Phosphorylation Motifs Through an Integrative Computational and Experimental Analysis of the Human Phosphoproteome. J Proteomics Bioinform *4*, 22-35.

Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S.G., and Pandey, A. (2007). A curated compendium of phosphorylation motifs. Nat Biotechnol *25*, 285-286.

Amanchy, R., Zhong, J., Hong, R., Kim, J.H., Gucek, M., Cole, R.N., Molina, H., and Pandey, A. (2009). Identification of c-Src tyrosine kinase substrates in platelet-derived growth factor receptor signaling. Mol Oncol *3*, 439-450.

Amanchy, R., Zhong, J., Molina, H., Chaerkady, R., Iwahori, A., Kalume, D.E., Gronborg, M., Joore, J., Cope, L., and Pandey, A. (2008). Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. J Proteome Res *7*, 3900-3910.

Amata, I., Maffei, M., and Pons, M. (2014). Phosphorylation of unique domains of Src family kinases. Front Genet *5*, 181.

Anam, K., and Davis, T.A. (2013). Comparative analysis of gene transcripts for cell signaling receptors in bone marrow-derived hematopoietic stem/progenitor cell and mesenchymal stromal cell populations. Stem Cell Res Ther *4*, 112.

Anbalagan, M., Moroz, K., Ali, A., Carrier, L., Glodowski, S., and Rowan, B.G. (2012). Subcellular localization of total and activated Src kinase in African American and Caucasian breast cancer. PLoS One *7*, e33017.

Andl, C.D., Mizushima, T., Oyama, K., Bowser, M., Nakagawa, H., and Rustgi, A.K. (2004). EGFR-induced cell migration is mediated predominantly by the JAK-STAT pathway in primary esophageal keratinocytes. Am J Physiol Gastrointest Liver Physiol *287*, G1227-1237.

Bagu, E.T., Miah, S., Dai, C., Spriggs, T., Ogunbolude, Y., Beaton, E., Sanders, M., Goel, R.K., Bonham, K., and Lukong, K.E. (2017). Repression of Fyn-related kinase in breast cancer cells is associated with promoter site-specific CpG methylation. Oncotarget *8*, 11442-11459.

Bateman, N.W., Goulding, S.P., Shulman, N.J., Gadok, A.K., Szumlinski, K.K., MacCoss, M.J., and Wu, C.C. (2014). Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). Mol Cell Proteomics *13*, 329-338.

Bauer, M., Ahrne, E., Baron, A.P., Glatter, T., Fava, L.L., Santamaria, A., Nigg, E.A., and Schmidt, A. (2014). Evaluation of data-dependent and -independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. J Proteome Res *13*, 5973-5988.

Beavis, R.C., and Fenyö, D. (2000). Database searching with mass-spectrometric information. Trends in Biotechnology *18*, 22-27.

Bensimon, A., Schmidt, A., Ziv, Y., Elkon, R., Wang, S.Y., Chen, D.J., Aebersold, R., and Shiloh, Y. (2010). ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. Sci Signal *3*, rs3.

Bergeron, J.J., Di Guglielmo, G.M., Dahan, S., Dominguez, M., and Posner, B.I. (2016). Spatial and Temporal Regulation of Receptor Tyrosine Kinase Activation and Intracellular Signal Transduction. Annu Rev Biochem *85*, 573-597.

Besset, V., Scott, R.P., and Ibanez, C.F. (2000). Signaling complexes and protein-protein interactions involved in the activation of the Ras and phosphatidylinositol 3-kinase pathways by the c-Ret receptor tyrosine kinase. J Biol Chem *275*, 39159-39166.

Bhavanasi, D., Badolia, R., Manne, B.K., Janapati, S., Dangelmaier, C.T., Mazharian, A., Jin, J., Kim, S., Zhang, X., Chen, X., *et al.* (2015). Cross talk between serine/threonine and tyrosine kinases regulates ADP-induced thromboxane generation in platelets. Thromb Haemost *114*, 558-568.

Bian, Y., Ye, M., Wang, C., Cheng, K., Song, C., Dong, M., Pan, Y., Qin, H., and Zou, H. (2013). Global screening of CK2 kinase substrates by an integrated phosphoproteomics workflow. Sci Rep *3*, 3460.

Bordner, K.A., George, E.D., Carlyle, B.C., Duque, A., Kitchen, R.R., Lam, T.T., Colangelo, C.M., Stone, K.L., Abbott, T.B., Mane, S.M., *et al.* (2011). Functional genomic and proteomic analysis reveals disruption of myelin-related genes and translation in a mouse model of early life neglect. Front Psychiatry *2*, 18.

Brauer, P.M., and Tyner, A.L. (2009). RAKing in AKT: a tumor suppressor function for the intracellular tyrosine kinase FRK. Cell Cycle *8*, 2728-2732.

Brauer, P.M., and Tyner, A.L. (2010). Building a better understanding of the intracellular tyrosine kinase PTK6 - BRK by BRK. Biochim Biophys Acta *1806*, 66-73.

Cai, X., Lietha, D., Ceccarelli, D.F., Karginov, A.V., Rajfur, Z., Jacobson, K., Hahn, K.M., Eck, M.J., and Schaller, M.D. (2008). Spatial and temporal regulation of focal adhesion kinase activity in living cells. Mol Cell Biol *28*, 201-214.

Campbell, S.J., and Jackson, R.M. (2003). Diversity in the SH2 domain family phosphotyrosyl peptide binding site. Protein Eng *16*, 217-227.

Cance, W.G., Craven, R.J., Bergman, M., Xu, L., Alitalo, K., and Liu, E.T. (1994). Rak, a novel nuclear tyrosine kinase expressed in epithelial cells. Cell Growth Differ *5*, 1347-1355.

Cao, X., Tay, A., Guy, G.R., and Tan, Y.H. (1996). Activation and association of Stat3 with Src in v-Src-transformed cell lines. Mol Cell Biol *16*, 1595-1603.

Capuani, F., Conte, A., Argenzio, E., Marchetti, L., Priami, C., Polo, S., Di Fiore, P.P., Sigismund, S., and Ciliberto, A. (2015). Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells. Nat Commun *6*, 7999.

Carpenter, R.L., and Lo, H.W. (2014). STAT3 Target Genes Relevant to Human Cancers. Cancers (Basel) *6*, 897-925.

Casaletto, J.B., and McClatchey, A.I. (2012). Spatial regulation of receptor tyrosine kinases in development and cancer. Nat Rev Cancer *12*, 387-400.

Cesaro, L., and Pinna, L.A. (2015). The generation of phosphoserine stretches in phosphoproteins: mechanism and significance. Mol Biosyst *11*, 2666-2679.

Chandrasekharan, S., Qiu, T.H., Alkharouf, N., Brantley, K., Mitchell, J.B., and Liu, E.T. (2002). Characterization of mice deficient in the Src family nonreceptor tyrosine kinase Frk/rak. Mol Cell Biol *22*, 5235-5247.

Chen, Y., Choong, L.Y., Lin, Q., Philp, R., Wong, C.H., Ang, B.K., Tan, Y.L., Loh, M.C., Hew, C.L., Shah, N., *et al.* (2007). Differential expression of novel tyrosine kinase substrates during breast cancer development. Mol Cell Proteomics *6*, 2072-2087.

Cheng, J., Montecalvo, A., and Kane, L.P. (2011). Regulation of NF-kappaB induction by TCR/CD28. Immunol Res *50*, 113-117.

Cheng, T.J., Tseng, Y.F., Chang, W.M., Chang, M.D., and Lai, Y.K. (2003). Retaining of the assembly capability of vimentin phosphorylated by mitogen-activated protein kinase-activated protein kinase-2. J Cell Biochem *89*, 589-602.

Chou, M.F., and Schwartz, D. (2011). Biological sequence motif discovery using motif-x. Curr Protoc Bioinformatics *Chapter 13*, Unit 13 15-24.

Chou, Y.H., Bischoff, J.R., Beach, D., and Goldman, R.D. (1990). Intermediate filament reorganization during mitosis is mediated by p34cdc2 phosphorylation of vimentin. Cell *62*, 1063-1071.

Ciesla, J., Fraczyk, T., and Rode, W. (2011). Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochim Pol *58*, 137-148.

Colinge, J., Cesar-Razquin, A., Huber, K., Breitwieser, F.P., Majek, P., and Superti-Furga, G. (2014). Building and exploring an integrated human kinase network: global organization and medical entry points. J Proteomics *107*, 113-127.

Conrads, T.P., and Veenstra, T.D. (2005). An enriched look at tyrosine phosphorylation. Nat Biotechnol *23*, 36-37.

Cooper, J.A., Gould, K.L., Cartwright, C.A., and Hunter, T. (1986). Tyr527 is phosphorylated in pp60c-src: implications for regulation. Science *231*, 1431-1434.

Corwin, T., Woodsmith, J., Apelt, F., Fontaine, J.F., Meierhofer, D., Helmuth, J., Grossmann, A., Andrade-Navarro, M.A., Ballif, B.A., and Stelzl, U. (2017). Defining Human Tyrosine Kinase Phosphorylation Networks Using Yeast as an In Vivo Model Substrate. Cell Syst *5*, 128-139 e124.

Courcelles, M., Fremin, C., Voisin, L., Lemieux, S., Meloche, S., and Thibault, P. (2013). Phosphoproteome dynamics reveal novel ERK1/2 MAP kinase substrates with broad spectrum of functions. Mol Syst Biol *9*, 669.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics *13*, 2513-2526.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res *10*, 1794-1805.

Day, E.K., Sosale, N.G., and Lazzara, M.J. (2016). Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process. Curr Opin Biotechnol *40*, 185-192.

Dazert, E., Colombi, M., Boldanova, T., Moes, S., Adametz, D., Quagliata, L., Roth, V., Terracciano, L., Heim, M.H., Jenoe, P., *et al.* (2016). Quantitative proteomics and phosphoproteomics on serial tumor biopsies from a sorafenib-treated HCC patient. Proc Natl Acad Sci U S A *113*, 1381-1386.

Deng, Y., Alicea-Velazquez, N.L., Bannwarth, L., Lehtonen, S.I., Boggon, T.J., Cheng, H.C., Hytonen, V.P., and Turk, B.E. (2014). Global analysis of human nonreceptor tyrosine kinase specificity using high-density peptide microarrays. J Proteome Res *13*, 4339-4346.

Derry, J.J., Prins, G.S., Ray, V., and Tyner, A.L. (2003). Altered localization and activity of the intracellular tyrosine kinase BRK/Sik in prostate tumor cells. Oncogene *22*, 4212-4220.

Derry, J.J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A.W., Chen, T., and Tyner, A.L. (2000). Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability. Mol Cell Biol *20*, 6114-6126.

Deutsch, E.W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D.S., Bernal-Llinares, M., Okuda, S., Kawano, S., *et al.* (2017). The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. Nucleic Acids Res *45*, D1100-D1106.

Di Fruscio, M., Chen, T., and Richard, S. (1999). Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc Natl Acad Sci U S A *96*, 2710-2715.

Di Palma, S., Zoumaro-Djayoon, A., Peng, M., Post, H., Preisinger, C., Munoz, J., and Heck, A.J. (2013). Finding the same needles in the haystack? A comparison of phosphotyrosine peptides enriched by immuno-affinity precipitation and metal-based affinity chromatography. J Proteomics *91*, 331-337.

Diks, S.H., Kok, K., O'Toole, T., Hommes, D.W., van Dijken, P., Joore, J., and Peppelenbosch, M.P. (2004). Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. J Biol Chem *279*, 49206-49213.

Ding, C., Wu, Z., Huang, L., Wang, Y., Xue, J., Chen, S., Deng, Z., Wang, L., Song, Z., and Chen, S. (2015). Mitofilin and CHCHD6 physically interact with Sam50 to sustain cristae structure. Sci Rep *5*, 16064.

Domizio, P., Liesner, R.J., Dicks-Mireaux, C., and Risdon, R.A. (1990). Malignant mesenchymoma associated with a congenital lung cyst in a child: case report and review of the literature. Pediatr Pathol *10*, 785-797.

Donella-Deana, A., Cesaro, L., Sarno, S., Brunati, A.M., Ruzzene, M., and Pinna, L.A. (2001). Autocatalytic tyrosine-phosphorylation of protein kinase CK2 alpha and alpha' subunits: implication of Tyr182. Biochem J *357*, 563-567.

Donella-Deana, A., Cesaro, L., Sarno, S., Ruzzene, M., Brunati, A.M., Marin, O., Vilk, G., Doherty-Kirby, A., Lajoie, G., Litchfield, D.W., *et al.* (2003). Tyrosine phosphorylation of protein kinase CK2 by Src-related tyrosine kinases correlates with increased catalytic activity. Biochem J *372*, 841-849.

Donepudi, M., and Resh, M.D. (2008). c-Src trafficking and co-localization with the EGF receptor promotes EGF ligand-independent EGF receptor activation and signaling. Cell Signal *20*, 1359-1367.

Dong, L., Yu, W.M., Zheng, H., Loh, M.L., Bunting, S.T., Pauly, M., Huang, G., Zhou, M., Broxmeyer, H.E., Scadden, D.T., *et al.* (2016). Leukaemogenic effects of Ptpn11 activating mutations in the stem cell microenvironment. Nature *539*, 304-308.

Durnberger, G., Camurdanoglu, B.Z., Tomschik, M., Schutzbier, M., Roitinger, E., Hudecz, O., Mechtler, K., and Herbst, R. (2014). Global analysis of muscle-specific kinase signaling by quantitative phosphoproteomics. Mol Cell Proteomics *13*, 1993-2003.

Eck, M.J. (1995). A new flavor in phosphotyrosine recognition. Structure *3*, 421-424.

Eck, M.J., Atwell, S.K., Shoelson, S.E., and Harrison, S.C. (1994). Structure of the regulatory domains of the Src-family tyrosine kinase Lck. Nature *368*, 764-769.

Eid, S., Turk, S., Volkamer, A., Rippmann, F., and Fulle, S. (2017). KinMap: a web-based tool for interactive navigation through human kinome data. BMC Bioinformatics *18*, 16.

Elias, J.E., and Gygi, S.P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol *604*, 55-71.

Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. Nat Rev Mol Cell Biol *4*, 181-191.

Emkey, R., and Kahn, C.R. (1997). Cross-talk between phorbol ester-mediated signaling and tyrosine kinase proto-oncogenes. I. Activation of protein kinase C stimulates tyrosine phosphorylation and activation of ErbB2 and ErbB3. J Biol Chem *272*, 31172-31181.

Eriksson, J.E., He, T., Trejo-Skalli, A.V., Harmala-Brasken, A.S., Hellman, J., Chou, Y.H., and Goldman, R.D. (2004). Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments. J Cell Sci *117*, 919-932.

Erpel, T., and Courtneidge, S.A. (1995). Src family protein tyrosine kinases and cellular signal transduction pathways. Curr Opin Cell Biol *7*, 176-182.

Esposito, F., Tornincasa, M., Chieffi, P., De Martino, I., Pierantoni, G.M., and Fusco, A. (2010). High-mobility group A1 proteins regulate p53-mediated transcription of Bcl-2 gene. Cancer Res *70*, 5379-5388.

Fan, G., Aleem, S., Yang, M., Miller, W.T., and Tonks, N.K. (2015). Protein-tyrosine Phosphatase and Kinase Specificity in Regulation of SRC and Breast Tumor Kinase. J Biol Chem *290*, 15934-15947.

Fattaey, A., and Booher, R.N. (1997). Myt1: a Wee1-type kinase that phosphorylates Cdc2 on residue Thr14. Prog Cell Cycle Res *3*, 233-240.

Feng, S., Chen, J.K., Yu, H., Simon, J.A., and Schreiber, S.L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241-1247.

Feracci, M., Foot, J.N., Grellscheid, S.N., Danilenko, M., Stehle, R., Gonchar, O., Kang, H.S., Dalgliesh, C., Meyer, N.H., Liu, Y., *et al.* (2016). Structural basis of RNA recognition and dimerization by the STAR proteins T-STAR and Sam68. Nat Commun *7*, 10355.

Fernandez-Valdivia, R., Zhang, Y., Pai, S., Metzker, M.L., and Schumacher, A. (2006). l7Rn6 encodes a novel protein required for clara cell function in mouse lung development. Genetics *172*, 389-399.

Ferrando, I.M., Chaerkady, R., Zhong, J., Molina, H., Jacob, H.K., Herbst-Robinson, K., Dancy, B.M., Katju, V., Bose, R., Zhang, J., *et al.* (2012). Identification of targets of c-Src tyrosine kinase by chemical complementation and phosphoproteomics. Mol Cell Proteomics *11*, 355-369.

Fila, J., and Honys, D. (2012). Enrichment techniques employed in phosphoproteomics. Amino Acids *43*, 1025-1047.

Flynn, M.P., Fiedler, S.E., Karlsson, A.B., Carr, D.W., Maizels, E.T., and Hunzicker-Dunn, M. (2016). Dephosphorylation of MAP2D enhances its binding to vimentin in preovulatory ovarian granulosa cells. J Cell Sci *129*, 2983-2996.

Fossale, E., Wolf, P., Espinola, J.A., Lubicz-Nawrocka, T., Teed, A.M., Gao, H., Rigamonti, D., Cattaneo, E., MacDonald, M.E., and Cotman, S.L. (2004). Membrane trafficking and mitochondrial abnormalities precede subunit c deposition in a cerebellar cell model of juvenile neuronal ceroid lipofuscinosis. BMC Neurosci *5*, 57.

Francavilla, C., Lupia, M., Tsafou, K., Villa, A., Kowalczyk, K., Rakownikow Jersie-Christensen, R., Bertalot, G., Confalonieri, S., Brunak, S., Jensen, L.J., *et al.* (2017). Phosphoproteomics of Primary Cells Reveals Druggable Kinase Signatures in Ovarian Cancer. Cell Rep *18*, 3242-3256.

Frank, A.M. (2009). A ranking-based scoring function for peptide-spectrum matches. J Proteome Res *8*, 2241-2252.

Frearson, J.A., and Alexander, D.R. (1997). The role of phosphotyrosine phosphatases in haematopoietic cell signal transduction. Bioessays *19*, 417-427.

Fresno Vara, J.A., Caceres, M.A., Silva, A., and Martin-Perez, J. (2001). Src family kinases are required for prolactin induction of cell proliferation. Mol Biol Cell *12*, 2171-2183.

Freudlsperger, C., Burnett, J.R., Friedman, J.A., Kannabiran, V.R., Chen, Z., and Van Waes, C. (2011). EGFR-PI3K-AKT-mTOR signaling in head and neck squamous cell carcinomas: attractive targets for molecular-oriented therapy. Expert Opin Ther Targets *15*, 63-74.

Gan, W., and Roux, B. (2009). Binding specificity of SH2 domains: insight from free energy simulations. Proteins *74*, 996-1007.

Glondu-Lassis, M., Dromard, M., Lacroix-Triki, M., Nirde, P., Puech, C., Knani, D., Chalbos, D., and Freiss, G. (2010). PTPL1/PTPN13 regulates breast cancer cell aggressiveness through direct inactivation of Src kinase. Cancer Res *70*, 5116-5126.

Gnad, F., Ren, S., Cox, J., Olsen, J.V., Macek, B., Oroshi, M., and Mann, M. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol *8*, R250.

Gnad, F., Wallin, J., Edgar, K., Doll, S., Arnott, D., Robillard, L., Kirkpatrick, D.S., Stokes, M.P., Vijapurkar, U., Hatzivassiliou, G., *et al.* (2016). Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network. Proteomics *16*, 1992-1997.

Goel, R.K., and Lukong, K.E. (2015). Tracing the footprints of the breast cancer oncogene BRK - Past till present. Biochim Biophys Acta *1856*, 39-54.

Goel, R.K., and Lukong, K.E. (2016). Understanding the cellular roles of Fyn-related kinase (FRK): implications in cancer biology. Cancer Metastasis Rev *35*, 179-199.

Goel, R.K., Miah, S., Black, K., Kalra, N., Dai, C., and Lukong, K.E. (2013). The unique N-terminal region of SRMS regulates enzymatic activity and phosphorylation of its novel substrate docking protein 1. FEBS J *280*, 4539-4559.

Goel, R.K., Paczkowska, M., Reimand, J., Napper, S., and Lukong, K.E. (2018). Phosphoproteomics analysis identifies novel candidate substrates of the non-receptor tyrosine kinase, SRMS. Mol Cell Proteomics.

Grebe, S.K., and Singh, R.J. (2011). LC-MS/MS in the Clinical Laboratory - Where to From Here? Clin Biochem Rev *32*, 5-31.

Grimes, M.L., Lee, W.J., van der Maaten, L., and Shannon, P. (2013). Wrangling phosphoproteomic data to elucidate cancer signaling pathways. PLoS One *8*, e52884.

Groban, E.S., Narayanan, A., and Jacobson, M.P. (2006). Conformational changes in protein loops and helices induced by post-translational phosphorylation. PLoS Comput Biol *2*, e32.

Han, X., Aslanian, A., and Yates, J.R., 3rd (2008). Mass spectrometry for proteomics. Curr Opin Chem Biol *12*, 483-490.

Han, X., Zhang, W., Yang, X., Wheeler, C.G., Langford, C.P., Wu, L., Filippova, N., Friedman, G.K., Ding, Q., Fathallah-Shaykh, H.M., *et al.* (2014). The role of Src family kinases in growth and migration of glioma stem cells. Int J Oncol *45*, 302-310.

Hansen, T.E., and Johansen, T. (2011). Following autophagy step by step. BMC Biol *9*, 39.

Hashiguchi, N., Kojidani, T., Imanaka, T., Haraguchi, T., Hiraoka, Y., Baumgart, E., Yokota, S., Tsukamoto, T., and Osumi, T. (2002). Peroxisomes are formed from complex membrane structures in PEX6-deficient CHO cells upon genetic complementation. Mol Biol Cell *13*, 711-722.

Hayes, M.J., and Moss, S.E. (2009). Annexin 2 has a dual role as regulator and effector of v-Src in cell transformation. J Biol Chem *284*, 10202-10210.

Heriche, J.K., and Chambaz, E.M. (1998). Protein kinase CK2alpha is a target for the Abl and Bcr-Abl tyrosine kinases. Oncogene *17*, 13-18.

Hoffman, N.J., Parker, B.L., Chaudhuri, R., Fisher-Wellman, K.H., Kleinert, M., Humphrey, S.J., Yang, P., Holliday, M., Trefely, S., Fazakerley, D.J., *et al.* (2015). Global Phosphoproteomic Analysis of Human Skeletal Muscle Reveals a Network of Exercise-Regulated Kinases and AMPK Substrates. Cell Metab *22*, 922-935.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res *43*, D512-520.

Hsu, P.P., Kang, S.A., Rameseder, J., Zhang, Y., Ottina, K.A., Lim, D., Peterson, T.R., Choi, Y., Gray, N.S., Yaffe, M.B., *et al.* (2011). The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. Science *332*, 1317-1322.

Hu, A., Noble, W.S., and Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. F1000Res *5*.

Hubbard, S.R. (2002). Protein tyrosine kinases: autoregulation and small-molecule inhibition. Curr Opin Struct Biol *12*, 735-741.

Hubbard, S.R., and Miller, W.T. (2007). Receptor tyrosine kinases: mechanisms of activation and signaling. Curr Opin Cell Biol *19*, 117-123.

Hubbard, S.R., Mohammadi, M., and Schlessinger, J. (1998). Autoregulatory mechanisms in protein-tyrosine kinases. J Biol Chem *273*, 11987-11990.

Hunter, T. (1998). The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. Philos Trans R Soc Lond B Biol Sci *353*, 583-605.

Hunter, T. (2009). Tyrosine phosphorylation: thirty years and counting. Curr Opin Cell Biol *21*, 140-146.

Hunter, T. (2014). The genesis of tyrosine phosphorylation. Cold Spring Harb Perspect Biol *6*, a020644.

Hunter, T. (2015). Discovering the first tyrosine kinase. Proceedings of the National Academy of Sciences *112*, 7877-7882.

Hunter, T., and Sefton, B.M. (1980). Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proc Natl Acad Sci U S A *77*, 1311-1315.

Huntington, N.D., and Tarlinton, D.M. (2004). CD45: direct and indirect government of immune regulation. Immunol Lett *94*, 167-174.

Huot, M.E., Brown, C.M., Lamarche-Vane, N., and Richard, S. (2009). An adaptor role for cytoplasmic Sam68 in modulating Src activity during cell polarization. Mol Cell Biol *29*, 1933-1943.

Hyder, C.L., Pallari, H.M., Kochin, V., and Eriksson, J.E. (2008). Providing cellular signposts--post-translational modifications of intermediate filaments. FEBS Lett *582*, 2140-2148.

Israel, A. (2010). The IKK complex, a central regulator of NF-kappaB activation. Cold Spring Harb Perspect Biol *2*, a000158.

Ivaska, J., Pallari, H.M., Nevo, J., and Eriksson, J.E. (2007). Novel functions of vimentin in cell adhesion, migration, and signaling. Exp Cell Res *313*, 2050-2062.

Iwabu, A., Smith, K., Allen, F.D., Lauffenburger, D.A., and Wells, A. (2004). Epidermal growth factor induces fibroblast contractility and motility via a protein kinase C delta-dependent pathway. J Biol Chem *279*, 14551-14560.

Izawa, I., and Inagaki, M. (2006). Regulatory mechanisms and functions of intermediate filaments: a study using site- and phosphorylation state-specific antibodies. Cancer Sci *97*, 167-174.

Janas, J.A., and Van Aelst, L. (2011). Oncogenic tyrosine kinases target Dok-1 for ubiquitin-mediated proteasomal degradation to promote cell transformation. Mol Cell Biol *31*, 2552-2565.

Jiang, J., Gui, F., He, Z., Li, L., Li, Y., Li, S., Wu, X., Deng, Z., Sun, X., Huang, X., *et al.* (2017). Targeting BRK-Positive Breast Cancers with Small-Molecule Kinase Inhibitors. Cancer Res *77*, 175-186.

Jiu, Y., Peranen, J., Schaible, N., Cheng, F., Eriksson, J.E., Krishnan, R., and Lappalainen, P. (2017). Vimentin intermediate filaments control actin stress fiber assembly through GEF-H1 and RhoA. J Cell Sci *130*, 892-902.

Kapoor, P., Bao, Y., Xiao, J., Luo, J., Shen, J., Persinger, J., Peng, G., Ranish, J., Bartholomew, B., and Shen, X. (2015). Regulation of Mec1 kinase activity by the SWI/SNF chromatin remodeling complex. Genes Dev *29*, 591-602.

Karayel, O., Sanal, E., Giese, S.H., Uretmen Kagiali, Z.C., Polat, A.N., Hu, C.K., Renard, B.Y., Tuncbag, N., and Ozlu, N. (2018). Comparative phosphoproteomic analysis reveals signaling networks regulating monopolar and bipolar cytokinesis. Sci Rep *8*, 2269.

Katz, M., Amit, I., and Yarden, Y. (2007). Regulation of MAPKs by growth factors and receptor tyrosine kinases. Biochim Biophys Acta *1773*, 1161-1176.

Kauko, O., Laajala, T.D., Jumppanen, M., Hintsanen, P., Suni, V., Haapaniemi, P., Corthals, G., Aittokallio, T., Westermarck, J., and Imanishi, S.Y. (2015). Label-free quantitative phosphoproteomics with novel pairwise abundance normalization reveals synergistic RAS and CIP2A signaling. Sci Rep *5*, 13099.

Kawachi, Y., Nakauchi, H., and Otsuka, F. (1997). Isolation of a cDNA encoding a tyrosine kinase expressed in murine skin. Exp Dermatol *6*, 140-146.

Kersey, P.J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004). The International Protein Index: an integrated database for proteomics experiments. Proteomics *4*, 1985-1988.

Kettenbach, A.N., Schweppe, D.K., Faherty, B.K., Pechenick, D., Pletnev, A.A., and Gerber, S.A. (2011). Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. Sci Signal *4*, rs5.

Kim, H., and Lee, S.T. (2005). An intramolecular interaction between SH2-kinase linker and kinase domain is essential for the catalytic activity of protein-tyrosine kinase-6. J Biol Chem *280*, 28973-28980.

Kim, M.S., Zhong, J., and Pandey, A. (2016). Common errors in mass spectrometry-based analysis of post-translational modifications. Proteomics *16*, 700-714.

Kimura, T., Sakamoto, H., Appella, E., and Siraganian, R.P. (1996). Conformational changes induced in the protein tyrosine kinase p72syk by tyrosine phosphorylation or by binding of phosphorylated immunoreceptor tyrosine-based activation motif peptides. Mol Cell Biol *16*, 1471-1478.

Kindrachuk, J., Arsenault, R., Kusalik, A., Kindrachuk, K.N., Trost, B., Napper, S., Jahrling, P.B., and Blaney, J.E. (2012). Systems kinomics demonstrates Congo Basin monkeypox virus infection selectively modulates host cell signaling responses as compared to West African monkeypox virus. Mol Cell Proteomics *11*, M111 015701.

Klumpp, S., and Krieglstein, J. (2002). Phosphorylation and dephosphorylation of histidine residues in proteins. Eur J Biochem *269*, 1067-1071.

Knighton, D.R., Cadena, D.L., Zheng, J., Ten Eyck, L.F., Taylor, S.S., Sowadski, J.M., and Gill, G.N. (1993). Structural features that specify tyrosine kinase activity deduced from homology modeling of the epidermal growth factor receptor. Proc Natl Acad Sci U S A *90*, 5001-5005.

Koegl, M., Zlatkine, P., Ley, S.C., Courtneidge, S.A., and Magee, A.I. (1994). Palmitoylation of multiple Src-family kinases at a homologous N-terminal motif. Biochem J *303 (Pt 3)*, 749-753.

Koenig, T., Menze, B.H., Kirchner, M., Monigatti, F., Parker, K.C., Patterson, T., Steen, J.J., Hamprecht, F.A., and Steen, H. (2008). Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. J Proteome Res *7*, 3708-3717.

Kohmura, N., Yagi, T., Tomooka, Y., Oyanagi, M., Kominami, R., Takeda, N., Chiba, J., Ikawa, Y., and Aizawa, S. (1994). A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption. Mol Cell Biol *14*, 6915-6925.

Korolchuk, V.I., Cozier, G., and Banting, G. (2005). Regulation of CK2 activity by phosphatidylinositol phosphates. J Biol Chem *280*, 40796-40801.

Krapivinsky, G., Medina, I., Krapivinsky, L., Gapon, S., and Clapham, D.E. (2004). SynGAP-MUPP1-CaMKII synaptic complexes regulate p38 MAP kinase activity and NMDA receptor-dependent synaptic AMPA receptor potentiation. Neuron *43*, 563-574.

Krishnan, N., Lam, T.T., Fritz, A., Rempinski, D., O'Loughlin, K., Minderman, H., Berezney, R., Marzluff, W.F., and Thapar, R. (2012). The prolyl isomerase Pin1 targets stem-loop binding protein (SLBP) to dissociate the SLBP-histone mRNA complex linking histone mRNA decay with SLBP ubiquitination. Mol Cell Biol *32*, 4306-4322.

Kristensen, A.R., Schandorff, S., Hoyer-Hansen, M., Nielsen, M.O., Jaattela, M., Dengjel, J., and Andersen, J.S. (2008). Ordered organelle degradation during starvation-induced autophagy. Mol Cell Proteomics *7*, 2419-2428.

Kuzmanov, U., Guo, H., Buchsbaum, D., Cosme, J., Abbasi, C., Isserlin, R., Sharma, P., Gramolini, A.O., and Emili, A. (2016). Global phosphoproteomic profiling reveals perturbed signaling in a mouse model of dilated cardiomyopathy. Proc Natl Acad Sci U S A *113*, 12592-12597.

LaFevre-Bernt, M., Sicheri, F., Pico, A., Porter, M., Kuriyan, J., and Miller, W.T. (1998). Intramolecular regulatory interactions in the Src family kinase Hck probed by mutagenesis of a conserved tryptophan residue. J Biol Chem *273*, 32129-32134.

Lallena, M.J., Diaz-Meco, M.T., Bren, G., Paya, C.V., and Moscat, J. (1999). Activation of IkappaB kinase beta by protein kinase C isoforms. Mol Cell Biol *19*, 2180-2188.

Lambert, J.P., Tucholska, M., Go, C., Knight, J.D., and Gingras, A.C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. J Proteomics *118*, 81-94.

Laplante, M., and Sabatini, D.M. (2012). mTOR signaling in growth control and disease. Cell *149*, 274-293.

Larance, M., and Lamond, A.I. (2015). Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol *16*, 269-280.

Le Beau, M.M., Westbrook, C.A., Diaz, M.O., and Rowley, J.D. (1984). Evidence for two distinct c-src loci on human chromosomes 1 and 20. Nature *312*, 70-71.

Lebrin, F., Chambaz, E.M., and Bianchini, L. (2001). A role for protein kinase CK2 in cell proliferation: evidence using a kinase-inactive mutant of CK2 catalytic subunit alpha. Oncogene *20*, 2010-2022.

Lee, W.K., Son, S.H., Jin, B.S., Na, J.H., Kim, S.Y., Kim, K.H., Kim, E.E., Yu, Y.G., and Lee, H.H. (2013). Structural and functional insights into the regulation mechanism of CK2 by IP6 and the intrinsically disordered protein Nopp140. Proc Natl Acad Sci U S A *110*, 19360-19365.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117-1134.

Leroy, C., Fialin, C., Sirvent, A., Simon, V., Urbach, S., Poncet, J., Robert, B., Jouin, P., and Roche, S. (2009). Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates SRC invasive activity in advanced colon carcinoma cells. Cancer Res *69*, 2279-2286.

Leung, G.C., Ho, C.S., Blasutig, I.M., Murphy, J.M., and Sicheri, F. (2007). Determination of the Plk4/Sak consensus phosphorylation motif using peptide spots arrays. FEBS Lett *581*, 77-83.

Li, L., Hisamoto, K., Kim, K.H., Haynes, M.P., Bauer, P.M., Sanjay, A., Collinge, M., Baron, R., Sessa, W.C., and Bender, J.R. (2007). Variant estrogen receptor-c-Src molecular interdependence and c-Src structural requirements for endothelial NO synthase activation. Proc Natl Acad Sci U S A *104*, 16468-16473.

Li, X., Lu, Y., Liang, K., Hsu, J.M., Albarracin, C., Mills, G.B., Hung, M.C., and Fan, Z. (2012). Brk/PTK6 sustains activated EGFR signaling through inhibiting EGFR degradation and transactivating EGFR. Oncogene *31*, 4372-4383.

Liao, W.T., Liu, J.L., Wang, Z.G., Cui, Y.M., Shi, L., Li, T.T., Zhao, X.H., Chen, X.T., Ding, Y.Q., and Song, L.B. (2013). High expression level and nuclear localization of Sam68 are associated with progression and poor prognosis in colorectal cancer. BMC Gastroenterol *13*, 126.

Lim, W.A., Richards, F.M., and Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375-379.

Lin, Q., Taylor, S.J., and Shalloway, D. (1997). Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. J Biol Chem *272*, 27274-27280.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jorgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., *et al.* (2007). Systematic discovery of in vivo phosphorylation networks. Cell *129*, 1415-1426.

Liu, B.A., Engelmann, B.W., and Nash, P.D. (2012). The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. FEBS Lett *586*, 2597-2605.

Liu, H., Li, L., Voss, C., Wang, F., Liu, J., and Li, S.S. (2015). A Comprehensive Immunoreceptor Phosphotyrosine-based Signaling Network Revealed by Reciprocal Protein-Peptide Array Screening. Mol Cell Proteomics *14*, 1846-1858.

Liu, X., Brodeur, S.R., Gish, G., Songyang, Z., Cantley, L.C., Laudano, A.P., and Pawson, T. (1993). Regulation of c-Src tyrosine kinase activity by the Src SH2 domain. Oncogene *8*, 1119-1126.

Lu, W., and Mayer, B.J. (1999). Mechanism of activation of Pak1 kinase by membrane localization. Oncogene *18*, 797-806.

Luke, M.R., Houghton, F., Perugini, M.A., and Gleeson, P.A. (2005). The trans-Golgi network GRIP-domain proteins form alpha-helical homodimers. Biochem J *388*, 835-841.

Lukong, K.E., Larocque, D., Tyner, A.L., and Richard, S. (2005). Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression. J Biol Chem *280*, 38639-38647.

Lukong, K.E., and Richard, S. (2003). Sam68, the KH domain-containing superSTAR. Biochim Biophys Acta *1653*, 73-86.

Lukong, K.E., and Richard, S. (2008). Breast tumor kinase BRK requires kinesin-2 subunit KAP3A in modulation of cell migration. Cell Signal *20*, 432-442.

Macek, B., Mann, M., and Olsen, J.V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol *49*, 199-221.

Maksumova, L., Le, H.T., Muratkhodjaev, F., Davidson, D., Veillette, A., and Pallen, C.J. (2005). Protein tyrosine phosphatase alpha regulates Fyn activity and Cbp/PAG phosphorylation in thymocyte lipid rafts. J Immunol *175*, 7947-7956.

Maloney, S.C., Adair, J.E., Smerdon, M.J., and Reeves, R. (2007). Gene-specific nucleotide excision repair is impaired in human cells expressing elevated levels of high mobility group A1 nonhistone proteins. DNA Repair (Amst) *6*, 1371-1379.

Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol *7*, 952-958.

Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem *70*, 437-473.

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

Mayer, B.J., Jackson, P.K., Van Etten, R.A., and Baltimore, D. (1992). Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming activity in vivo. Mol Cell Biol *12*, 609-618.

McLafferty, F.W. (1981). Tandem mass spectrometry. Science *214*, 280-287.

Mendez, M.G., Kojima, S., and Goldman, R.D. (2010). Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. FASEB J *24*, 1838-1851.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One *5*, e13984.

Mertins, P., Udeshi, N.D., Clauser, K.R., Mani, D.R., Patel, J., Ong, S.E., Jaffe, J.D., and Carr, S.A. (2012). iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. Mol Cell Proteomics *11*, M111 014423.

Miah, S., Goel, R.K., Dai, C., Kalra, N., Beaton-Brown, E., Bagu, E.T., Bonham, K., and Lukong, K.E. (2014). BRK targets Dok1 for ubiquitin-mediated proteasomal degradation to promote cell proliferation and migration. PLoS One *9*, e87684.

Miah, S., Martin, A., and Lukong, K.E. (2012). Constitutive activation of breast tumor kinase accelerates cell migration and tumor growth in vivo. Oncogenesis *1*, e11.

Miller, M.L., Jensen, L.J., Diella, F., Jorgensen, C., Tinti, M., Li, L., Hsiung, M., Parker, S.A., Bordeaux, J., Sicheritz-Ponten, T., *et al.* (2008). Linear motif atlas for phosphorylation-dependent signaling. Sci Signal *1*, ra2.

Miller, W.T. (2003). Determinants of substrate recognition in nonreceptor tyrosine kinases. Acc Chem Res *36*, 393-400.

Mok, J., Kim, P.M., Lam, H.Y., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., *et al.* (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci Signal *3*, ra12.

Nada, S., Okada, M., MacAuley, A., Cooper, J.A., and Nakagawa, H. (1991). Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60c-src. Nature *351*, 69-72.

Nahnsen, S., Bielow, C., Reinert, K., and Kohlbacher, O. (2013). Tools for label-free peptide quantification. Mol Cell Proteomics *12*, 549-556.

Napper, S., Dadgar, S., Arsenault, R.J., Trost, B., Scruten, E., Kusalik, A., and Shand, P. (2015). Induction of tissue- and stressor-specific kinomic responses in chickens exposed to hot and cold stresses. Poult Sci *94*, 1333-1345.

Neet, K., and Hunter, T. (1996). Vertebrate non-receptor protein-tyrosine kinase families. Genes Cells *1*, 147-169.

Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C., and Haynes, P.A. (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics *11*, 535-553.

Ng, K.Y., Yin, T., Machida, K., Wu, Y.I., and Mayer, B.J. (2015). Phosphorylation of Dok1 by Abl family kinases inhibits CrkI transforming activity. Oncogene *34*, 2650-2659.

Nilsson, C.L. (2012). Advances in quantitative phosphoproteomics. Anal Chem *84*, 735-746.

Nogami, M., Yamazaki, M., Watanabe, H., Okabayashi, Y., Kido, Y., Kasuga, M., Sasaki, T., Maehama, T., and Kanaho, Y. (2003). Requirement of autophosphorylated tyrosine 992 of EGF receptor and its docking protein phospholipase C gamma 1 for membrane ruffle formation. FEBS Lett *536*, 71-76.

Nolen, B., Taylor, S., and Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. Mol Cell *15*, 661-675.

Nurmio, M., Joki, H., Kallio, J., Maatta, J.A., Vaananen, H.K., Toppari, J., Jahnukainen, K., and Laitala-Leinonen, T. (2011). Receptor tyrosine kinase inhibition causes simultaneous bone loss and excess bone formation within growing bone in rats. Toxicol Appl Pharmacol *254*, 267-279.

O'Reilly, A.M., Pluskey, S., Shoelson, S.E., and Neel, B.G. (2000). Activated mutants of SHP-2 preferentially induce elongation of Xenopus animal caps. Mol Cell Biol *20*, 299-311.

Ochoa, D., Jonikas, M., Lawrence, R.T., El Debs, B., Selkrig, J., Typas, A., Villen, J., Santos, S.D., and Beltrao, P. (2016). An atlas of human kinase regulation. Mol Syst Biol *12*, 888.

Oda, K., Matsuoka, Y., Funahashi, A., and Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol *1*, 2005 0010.

Ogunbolude, Y., Dai, C., Bagu, E.T., Goel, R.K., Miah, S., MacAusland-Berg, J., Ng, C.Y., Chibbar, R., Napper, S., Raptis, L., *et al.* (2017). FRK inhibits breast cancer cell migration and invasion by suppressing epithelial-mesenchymal transition. Oncotarget *8*, 113034-113065.

Okada, M. (2012). Regulation of the SRC family kinases by Csk. Int J Biol Sci *8*, 1385-1397.

Okada, M., Howell, B.W., Broome, M.A., and Cooper, J.A. (1993). Deletion of the SH3 domain of Src interferes with regulation by the phosphorylated carboxyl-terminal tyrosine. J Biol Chem *268*, 18070-18075.

Olsen, B.B., Wang, S.Y., Svenstrup, T.H., Chen, B.P., and Guerra, B. (2012). Protein kinase CK2 localizes to sites of DNA double-strand break regulating the cellular response to DNA damage. BMC Mol Biol *13*, 7.

Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics *4*, 2010-2021.

Olsen, J.V., and Mann, M. (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc Natl Acad Sci U S A *101*, 13417-13422.

Oppermann, H., Levinson, A.D., Varmus, H.E., Levintow, L., and Bishop, J.M. (1979). Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). Proc Natl Acad Sci U S A *76*, 1804-1808.

Osherov, N., and Levitzki, A. (1994). Epidermal-growth-factor-dependent activation of the src-family kinases. Eur J Biochem *225*, 1047-1053.

Ostrander, J.H., Daniel, A.R., and Lange, C.A. (2010). Brk/PTK6 signaling in normal and cancer cell models. Curr Opin Pharmacol *10*, 662-669.

Ottinger, E.A., Botfield, M.C., and Shoelson, S.E. (1998). Tandem SH2 domains confer high specificity in tyrosine kinase signaling. J Biol Chem *273*, 729-735.

Palvimo, J., and Linnala-Kankkunen, A. (1989). Identification of sites on chromosomal protein HMG-I phosphorylated by casein kinase II. FEBS Lett *257*, 101-104.

Pan, C., Gnad, F., Olsen, J.V., and Mann, M. (2008). Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors. Proteomics *8*, 4534-4546.

Panni, S., Dente, L., and Cesareni, G. (2002). In vitro evolution of recognition specificity mediated by SH3 domains reveals target recognition rules. J Biol Chem *277*, 21666-21674.

Parikh, K., Diks, S.H., Tuynman, J.H., Verhaar, A., Lowenberg, M., Hommes, D.W., Joore, J., Pandey, A., and Peppelenbosch, M.P. (2009). Comparison of peptide array substrate phosphorylation of c-Raf and mitogen activated protein kinase kinase kinase 8. PLoS One *4*, e6440.

Park, S.I., Shah, A.N., Zhang, J., and Gallick, G.E. (2007). Regulation of angiogenesis and vascular permeability by Src family kinases: opportunities for therapeutic treatment of solid tumors. Expert Opin Ther Targets *11*, 1207-1217.

Parker, R.C., Mardon, G., Lebo, R.V., Varmus, H.E., and Bishop, J.M. (1985). Isolation of duplicated human c-src genes located on chromosomes 1 and 20. Mol Cell Biol *5*, 831-838.

Paronetto, M.P., Zalfa, F., Botti, F., Geremia, R., Bagni, C., and Sette, C. (2006). The nuclear RNA-binding protein Sam68 translocates to the cytoplasm and associates with the polysomes in mouse spermatocytes. Mol Biol Cell *17*, 14-24.

Parsons, S.J., and Parsons, J.T. (2004). Src family kinases, key regulators of signal transduction. Oncogene *23*, 7906-7909.

Patwardhan, P., and Resh, M.D. (2010). Myristoylation and membrane binding regulate c-Src stability and kinase activity. Mol Cell Biol *30*, 4094-4107.

Pawson, T. (1995). Protein modules and signalling networks. Nature *373*, 573-580.

Pawson, T., and Gish, G.D. (1992). SH2 and SH3 domains: from structure to function. Cell *71*, 359-362.

Pawson, T., Gish, G.D., and Nash, P. (2001). SH2 domains, interaction modules and cellular wiring. Trends Cell Biol *11*, 504-511.

Pawson, T., and Schlessingert, J. (1993). SH2 and SH3 domains. Current Biology *3*, 434-442.

Pechan, T., and Gwaltney, S.R. (2012). Calculations of relative intensities of fragment ions in the MSMS spectra of a doubly charged penta-peptide. BMC Bioinformatics *13 Suppl 15*, S13.

Peng, J., and Gygi, S.P. (2001). Proteomics: the move to mixtures. J Mass Spectrom *36*, 1083-1091.

Peng, M., Emmadi, R., Wang, Z., Wiley, E.L., Gann, P.H., Khan, S.A., Banerji, N., McDonald, W., Asztalos, S., Pham, T.N., *et al.* (2014). PTK6/BRK is expressed in the normal mammary gland and activated at the plasma membrane in breast tumors. Oncotarget *5*, 6038-6048.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Ponniah, S., Wang, D.Z., Lim, K.L., and Pallen, C.J. (1999). Targeted disruption of the tyrosine phosphatase PTPalpha leads to constitutive downregulation of the kinases Src and Fyn. Curr Biol *9*, 535-538.

Poss, Z.C., Ebmeier, C.C., Odell, A.T., Tangpeerachaikul, A., Lee, T., Pelish, H.E., Shair, M.D., Dowell, R.D., Old, W.M., and Taatjes, D.J. (2016). Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. Cell Rep *15*, 436-450.

Possemato, A.P., Paulo, J.A., Mulhern, D., Guo, A., Gygi, S.P., and Beausoleil, S.A. (2017). Multiplexed Phosphoproteomic Profiling Using Titanium Dioxide and Immunoaffinity Enrichments Reveals Complementary Phosphorylation Events. J Proteome Res *16*, 1506-1514.

Potts, M.B., Kim, H.S., Fisher, K.W., Hu, Y., Carrasco, Y.P., Bulut, G.B., Ou, Y.H., Herrera-Herrera, M.L., Cubillos, F., Mendiratta, S., *et al.* (2013). Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. Sci Signal *6*, ra90.

Qi, D., Brownridge, P., Xia, D., Mackay, K., Gonzalez-Galarza, F.F., Kenyani, J., Harman, V., Beynon, R.J., and Jones, A.R. (2012). A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. OMICS *16*, 489-495.

Qiu, H., and Miller, W.T. (2004). Role of the Brk SH3 domain in substrate recognition. Oncogene *23*, 2216-2223.

Rauniyar, N., and Yates, J.R., 3rd (2014). Isobaric labeling-based relative quantification in shotgun proteomics. J Proteome Res *13*, 5293-5309.

Ravid, T., Heidinger, J.M., Gee, P., Khan, E.M., and Goldkorn, T. (2004). c-Cbl-mediated ubiquitinylation is required for epidermal growth factor receptor exit from the early endosomes. J Biol Chem *279*, 37153-37162.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res *44*, W83-89.

Rich, M.T., Abbott, T.B., Chung, L., Gulcicek, E.E., Stone, K.L., Colangelo, C.M., Lam, T.T., Nairn, A.C., Taylor, J.R., and Torregrossa, M.M. (2016). Phosphoproteomic Analysis Reveals a Novel Mechanism of CaMKIIalpha Regulation Inversely Induced by Cocaine Memory Extinction versus Reconsolidation. J Neurosci *36*, 7613-7627.

Richard, S. (2010). Reaching for the stars: Linking RNA binding proteins to diseases. Adv Exp Med Biol *693*, 142-157.

Robinson, D.R., Wu, Y.M., and Lin, S.F. (2000). The protein tyrosine kinase family of the human genome. Oncogene *19*, 5548-5557.

Roskoski, R., Jr. (2005). Src kinase regulation by phosphorylation and dephosphorylation. Biochem Biophys Res Commun *331*, 1-14.

Roux, P.P., and Thibault, P. (2013). The coming of age of phosphoproteomics--from large data sets to inference of protein functions. Mol Cell Proteomics *12*, 3453-3464.

Ruhs, S., Stratz, N., Quarch, K., Masch, A., Schutkowski, M., Gekle, M., and Grossmann, C. (2017). Modulation of transcriptional mineralocorticoid receptor activity by casein kinase 2. Sci Rep *7*, 15340.

Sadygov, R.G. (2015). Using SEQUEST with theoretically complete sequence databases. J Am Soc Mass Spectrom *26*, 1858-1864.

Safari, F., Murata-Kamiya, N., Saito, Y., and Hatakeyama, M. (2011). Mammalian Pragmin regulates Src family kinases via the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif that is exploited by bacterial effectors. Proc Natl Acad Sci U S A *108*, 14938-14943.

Saigal, B., and Johnson, F. (2008). Pathways mediating apoptosis and cell cycle arrest downstream of Src family kinases in aerodigestive tumors. Cancer Research *68*, 4864-4864.

Sanchez-Jimenez, F., and Sanchez-Margalet, V. (2013). Role of Sam68 in post-transcriptional gene regulation. Int J Mol Sci *14*, 23402-23419.

Santos, A.L., and Lindner, A.B. (2017). Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease. Oxid Med Cell Longev *2017*, 5716409.

Sarno, S., Ghisellini, P., and Pinna, L.A. (2002). Unique activation mechanism of protein kinase CK2. The N-terminal segment is essential for constitutive activity of the catalytic subunit but not of the holoenzyme. J Biol Chem *277*, 22509-22514.

Scholma, J., Fuhler, G.M., Joore, J., Hulsman, M., Schivo, S., List, A.F., Reinders, M.J., Peppelenbosch, M.P., and Post, J.N. (2016). Improved intra-array and interarray normalization of peptide microarray phosphorylation for phosphorylome and kinome profiling by rational selection of relevant spots. Sci Rep *6*, 26695.

Schreiner, S.J., Schiavone, A.P., and Smithgall, T.E. (2002). Activation of STAT3 by the Src family kinase Hck requires a functional SH3 domain. J Biol Chem *277*, 45680-45687.

Schuhmacher, M., and Eick, D. (2013). Dose-dependent regulation of target gene expression and cell proliferation by c-Myc levels. Transcription *4*, 192-197.

Schwanbeck, R., Gymnopoulos, M., Petry, I., Piekielko, A., Szewczuk, Z., Heyduk, T., Zechel, K., and Wisniewski, J.R. (2001). Consecutive steps of phosphorylation affect conformation and DNA binding of the chironomus high mobility group A protein. J Biol Chem *276*, 26012-26021.

Serfas, M.S., and Tyner, A.L. (2003). Brk, Srm, Frk, and Src42A form a distinct family of intracellular Src-like tyrosine kinases. Oncol Res *13*, 409-419.

Shah, N.H., Lobel, M., Weiss, A., and Kuriyan, J. (2018). Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. Elife *7*.
Sharma, K., D'Souza, R.C., Tyanova, S., Schaab, C., Wisniewski, J.R., Cox, J., and Mann, M. (2014). Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep *8*, 1583-1594.
Sherry, M.M., Reeves, A., Wu, J.K., and Cochran, B.H. (2009). STAT3 is required for proliferation and maintenance of multipotency in glioblastoma stem cells. Stem Cells *27*, 2383-2392.
Shin, S.Y., Kim, C.G., Lim, Y., and Lee, Y.H. (2011). The ETS family transcription factor ELK-1 regulates induction of the cell cycle-regulatory gene p21(Waf1/Cip1) and the BAX gene in sodium arsenite-exposed human keratinocyte HaCaT cells. J Biol Chem *286*, 26860-26872.
Sibilia, M., Kroismayr, R., Lichtenberger, B.M., Natarajan, A., Hecking, M., and Holcmann, M. (2007). The epidermal growth factor receptor: from development to tumorigenesis. Differentiation *75*, 770-787.
Sica, V., Galluzzi, L., Bravo-San Pedro, J.M., Izzo, V., Maiuri, M.C., and Kroemer, G. (2015). Organelle-Specific Initiation of Autophagy. Mol Cell *59*, 522-539.
Slomovitz, B.M., and Coleman, R.L. (2012). The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. Clin Cancer Res *18*, 5856-5864.
Solyakov, L., Cain, K., Tracey, B.M., Jukes, R., Riley, A.M., Potter, B.V., and Tobin, A.B. (2004). Regulation of casein kinase-2 (CK2) activity by inositol phosphates. J Biol Chem *279*, 43403-43410.
Somani, A.K., Bignon, J.S., Mills, G.B., Siminovitch, K.A., and Branch, D.R. (1997). Src kinase activity is regulated by the SHP-1 protein-tyrosine phosphatase. J Biol Chem *272*, 21113-21119.
Songyang, Z., Carraway, K.L., 3rd, Eck, M.J., Harrison, S.C., Feldman, R.A., Mohammadi, M., Schlessinger, J., Hubbard, S.R., Smith, D.P., Eng, C., *et al.* (1995). Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. Nature *373*, 536-539.
Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., *et al.* (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767-778.
Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001). Gene expression patterns of breast carcinomas

distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A *98*, 10869-10874.

St-Denis, N., Gabriel, M., Turowec, J.P., Gloor, G.B., Li, S.S., Gingras, A.C., and Litchfield, D.W. (2015). Systematic investigation of hierarchical phosphorylation by protein kinase CK2. J Proteomics *118*, 49-62.

Steger, M., Tonelli, F., Ito, G., Davies, P., Trost, M., Vetter, M., Wachter, S., Lorentzen, E., Duddy, G., Wilson, S., *et al.* (2016). Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. Elife *5*.

Stoker, A.W. (2005). Protein tyrosine phosphatases and signalling. J Endocrinol *185*, 19-33.

Su, J., Muranjan, M., and Sap, J. (1999). Receptor protein tyrosine phosphatase alpha activates Src-family kinases and controls integrin-mediated responses in fibroblasts. Curr Biol *9*, 505-511.

Sun, L., Yu, F., Ullah, A., Hubrack, S., Daalis, A., Jung, P., and Machaca, K. (2011). Endoplasmic reticulum remodeling tunes IP(3)-dependent Ca(2)+ release sensitivity. PLoS One *6*, e27928.

Superti-Furga, G. (1995). Regulation of the Src protein tyrosine kinase. FEBS Lett *369*, 62-66.

Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A *101*, 9528-9533.

Takeda, H., Kawamura, Y., Miura, A., Mori, M., Wakamatsu, A., Yamamoto, J., Isogai, T., Matsumoto, M., Nakayama, K.I., Natsume, T., *et al.* (2010). Comparative analysis of human SRC-family kinase substrate specificity in vitro. J Proteome Res *9*, 5982-5993.

Tanida, I., Ueno, T., and Kominami, E. (2008). LC3 and Autophagy. Methods Mol Biol *445*, 77-88.

Taylor, S.J., Anafi, M., Pawson, T., and Shalloway, D. (1995). Functional interaction between c-Src and its mitotic target, Sam 68. J Biol Chem *270*, 10120-10124.

Thingholm, T.E., Jensen, O.N., and Larsen, M.R. (2009). Analytical strategies for phosphoproteomics. Proteomics *9*, 1451-1468.

Thomas, P., and Smart, T.G. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. J Pharmacol Toxicol Methods *51*, 187-200.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel

quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem *75*, 1895-1904.

Tomar, A., and Schlaepfer, D.D. (2010). A PAK-activated linker for EGFR and FAK. Dev Cell *18*, 170-172.

Trost, B., Kindrachuk, J., Maattanen, P., Napper, S., and Kusalik, A. (2013). PIIKA 2: an expanded, web-based platform for analysis of kinome microarray data. PLoS One *8*, e80837.

Tu, C., Ortega-Cava, C.F., Winograd, P., Stanton, M.J., Reddi, A.L., Dodge, I., Arya, R., Dimri, M., Clubb, R.J., Naramura, M., *et al.* (2010). Endosomal-sorting complexes required for transport (ESCRT) pathway-dependent endosomal traffic regulates the localization of active Src at focal adhesions. Proc Natl Acad Sci U S A *107*, 16107-16112.

Turk, B.E. (2008). Understanding and exploiting substrate recognition by protein kinases. Curr Opin Chem Biol *12*, 4-10.

Tyanova, S., Temu, T., and Cox, J. (2016a). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc *11*, 2301-2319.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016b). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods *13*, 731-740.

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

Uckun, F.M., Tuel-Ahlgren, L., Waddick, K.G., Jun, X., Jin, J., Myers, D.E., Rowley, R.B., Burkhardt, A.L., and Bolen, J.B. (1996). Physical and functional interactions between Lyn and p34cdc2 kinases in irradiated human B-cell precursors. J Biol Chem *271*, 6389-6397.

Ulintz, P.J., Zhu, J., Qin, Z.S., and Andrews, P.C. (2006). Improved classification of mass spectrometry database search results using newer machine learning approaches. Mol Cell Proteomics *5*, 497-509.

Unwin, R.D. (2010). Quantification of proteins by iTRAQ. Methods Mol Biol *658*, 205-215.

Vacaresse, N., Moller, B., Danielsen, E.M., Okada, M., and Sap, J. (2008). Activation of c-Src and Fyn kinases by protein-tyrosine phosphatase RPTPalpha is substrate-specific and compatible with lipid raft localization. J Biol Chem *283*, 35815-35824.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J *275*, 2712-2726.

van der Mijn, J.C., Labots, M., Piersma, S.R., Pham, T.V., Knol, J.C., Broxterman, H.J., Verheul, H.M., and Jimenez, C.R. (2015). Evaluation of different phospho-tyrosine antibodies for label-free phosphoproteomics. J Proteomics *127*, 259-263.

Van Hoof, D., Munoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L., and Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell *5*, 214-226.

Veit, J., Sachsenberg, T., Chernev, A., Aicheler, F., Urlaub, H., and Kohlbacher, O. (2016). LFQProfiler and RNP(xl): Open-Source Tools for Label-Free Quantification and Protein-RNA Cross-Linking Integrated into Proteome Discoverer. J Proteome Res *15*, 3441-3448.

Visser Smit, G.D., Place, T.L., Cole, S.L., Clausen, K.A., Vemuganti, S., Zhang, G., Koland, J.G., and Lill, N.L. (2009). Cbl controls EGFR fate by regulating early endosome fusion. Sci Signal *2*, ra86.

Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., *et al.* (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res *44*, D447-456.

Vizcaino, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., *et al.* (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol *32*, 223-226.

von der Malsburg, K., Muller, J.M., Bohnert, M., Oeljeklaus, S., Kwiatkowska, P., Becker, T., Loniewska-Lwowska, A., Wiese, S., Rao, S., Milenkovic, D., *et al.* (2011). Dual role of mitofilin in mitochondrial membrane organization and protein biogenesis. Dev Cell *21*, 694-707.

Wagner, M.J., Stacey, M.M., Liu, B.A., and Pawson, T. (2013). Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. Cold Spring Harb Perspect Biol *5*, a008987.

Walther, T.C., and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. J Cell Biol *190*, 491-500.

Wang, M., Jiang, Y., and Xu, X. (2015). A novel method for predicting post-translational modifications on serine and threonine sites by using site-modification network profiles. Mol Biosyst *11*, 3092-3100.

Wang, P., and Wilson, S.R. (2013). Mass spectrometry-based protein identification by integrating de novo sequencing with database searching. BMC Bioinformatics *14 Suppl 2*, S24.

Wang, T., Kettenbach, A.N., Gerber, S.A., and Bailey-Kellogg, C. (2012). MMFPh: a maximal motif finder for phosphoproteomics datasets. Bioinformatics *28*, 1562-1570.

Wang, Y.T., Pan, S.H., Tsai, C.F., Kuo, T.C., Hsu, Y.L., Yen, H.Y., Choong, W.K., Wu, H.Y., Liao, Y.C., Hong, T.M., *et al.* (2017). Phosphoproteomics Reveals HMGA1, a CK2 Substrate, as a Drug-Resistant Target in Non-Small Cell Lung Cancer. Sci Rep *7*, 44021.

Watanabe, N., Broome, M., and Hunter, T. (1995). Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle. EMBO J *14*, 1878-1891.

Wegener, E., and Krappmann, D. (2008). Dynamic protein complexes regulate NF-kappaB signaling. Handb Exp Pharmacol, 237-259.

Wei, Y., Zou, Z., Becker, N., Anderson, M., Sumpter, R., Xiao, G., Kinch, L., Koduru, P., Christudass, C.S., Veltri, R.W., *et al.* (2013). EGFR-mediated Beclin 1 phosphorylation in autophagy suppression, tumor progression, and tumor chemoresistance. Cell *154*, 1269-1284.

Wells, J.M., and McLuckey, S.A. (2005). Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol *402*, 148-185.

Wiese, S., Reidegeld, K.A., Meyer, H.E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics *7*, 340-350.

Woodring, P.J., Meisenhelder, J., Johnson, S.A., Zhou, G.L., Field, J., Shah, K., Bladt, F., Pawson, T., Niki, M., Pandolfi, P.P., *et al.* (2004). c-Abl phosphorylates Dok1 to promote filopodia during cell spreading. J Cell Biol *165*, 493-503.

Wozniak, D.J., Kajdacsy-Balla, A., Macias, V., Ball-Kell, S., Zenner, M.L., Bie, W., and Tyner, A.L. (2017). PTEN is a protein phosphatase that targets active PTK6 and inhibits PTK6 oncogenic signaling in prostate cancer. Nat Commun *8*, 1508.

Wrighton, K.H. (2016). Nuclear organization: Building nuclear bodies with RNA. Nat Rev Mol Cell Biol *17*, 463.

Wu, F., Wang, P., Zhang, J., Young, L.C., Lai, R., and Li, L. (2010). Studies of phosphoproteomic changes induced by nucleophosmin-anaplastic lymphoma kinase (ALK) highlight deregulation of tumor necrosis factor (TNF)/Fas/TNF-related apoptosis-induced ligand signaling pathway in ALK-positive anaplastic large cell lymphoma. Mol Cell Proteomics *9*, 1616-1632.

Xie, F., Liu, T., Qian, W.J., Petyuk, V.A., and Smith, R.D. (2011). Liquid chromatography-mass spectrometry-based quantitative proteomics. J Biol Chem *286*, 25443-25449.

Xie, Z., Chen, Y., Liao, E.Y., Jiang, Y., Liu, F.Y., and Pennypacker, S.D. (2010). Phospholipase C-gamma1 is required for the epidermal growth factor receptor-induced squamous cell carcinoma cell mitogenesis. Biochem Biophys Res Commun *397*, 296-300.

Xu, W., Doshi, A., Lei, M., Eck, M.J., and Harrison, S.C. (1999). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Mol Cell *3*, 629-638.

Xue, L., Geahlen, R.L., and Tao, W.A. (2013). Identification of direct tyrosine kinase substrates based on protein kinase assay-linked phosphoproteomics. Mol Cell Proteomics *12*, 2969-2980.

Xue, L., Wang, W.H., Iliuk, A., Hu, L., Galan, J.A., Yu, S., Hans, M., Geahlen, R.L., and Tao, W.A. (2012). Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. Proc Natl Acad Sci U S A *109*, 5615-5620.

Yang, T., Massa, S.M., and Longo, F.M. (2006). LAR protein tyrosine phosphatase receptor associates with TrkB and modulates neurotrophic signaling pathways. J Neurobiol *66*, 1420-1436.

Yarden, Y., and Shilo, B.Z. (2007). SnapShot: EGFR signaling pathway. Cell *131*, 1018.

Yates, J.R., 3rd (1998). Database searching using mass spectrometry data. Electrophoresis *19*, 893-900.

Ye, D.Z., and Field, J. (2012). PAK signaling in cancer. Cell Logist *2*, 105-116.

Yeatman, T.J. (2004). A renaissance for SRC. Nat Rev Cancer *4*, 470-480.

Yefi, R., Ponce, D.P., Niechi, I., Silva, E., Cabello, P., Rodriguez, D.A., Marcelain, K., Armisen, R., Quest, A.F., and Tapia, J.C. (2011). Protein kinase CK2 promotes cancer cell viability via up-regulation of cyclooxygenase-2 expression and enhanced prostaglandin E2 production. J Cell Biochem *112*, 3167-3175.

Yoshida, T., Zhang, G., Smith, M.A., Lopez, A.S., Bai, Y., and Li, J. (2015). Correction: Tyrosine Phosphoproteomics Identifies Both Codrivers and Cotargeting Strategies for T790M-Related EGFR-TKI Resistance in Non-Small Cell Lung Cancer. Clin Cancer Res *21*, 3571.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W., and Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. Cell *76*, 933-945.

Zarich, N., Oliva, J.L., Martinez, N., Jorge, R., Ballester, A., Gutierrez-Eisman, S., Garcia-Vargas, S., and Rojas, J.M. (2006). Grb2 is a negative modulator of the intrinsic Ras-GEF activity of hSos1. Mol Biol Cell *17*, 3591-3597.

Zhang, X., Maity, T., Kashyap, M.K., Bansal, M., Venugopalan, A., Singh, S., Awasthi, S., Marimuthu, A., Charles Jacob, H.K., Belkina, N., *et al.* (2017). Quantitative Tyrosine Phosphoproteomics of Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor-treated Lung Adenocarcinoma Cells Reveals Potential Novel Biomarkers of Therapeutic Response. Mol Cell Proteomics *16*, 891-910.

Zhang, X., Zhu, J., Li, Y., Lin, T., Siclari, V.A., Chandra, A., Candela, E.M., Koyama, E., Enomoto-Iwamoto, M., and Qin, L. (2013). Epidermal growth factor receptor (EGFR) signaling regulates epiphyseal cartilage development through beta-catenin-dependent and -independent pathways. J Biol Chem *288*, 32229-32240.

Zhao, B., Tan, P.H., Li, S.S., and Pei, D. (2013). Systematic characterization of the specificity of the SH2 domains of cytoplasmic tyrosine kinases. J Proteomics *81*, 56-69.

Zheng, X.M., Resnick, R.J., and Shalloway, D. (2000). A phosphotyrosine displacement mechanism for activation of Src by PTPalpha. EMBO J *19*, 964-978.

Zhou, B., Ritt, D.A., Morrison, D.K., Der, C.J., and Cox, A.D. (2016). Protein Kinase CK2alpha Maintains Extracellular Signal-regulated Kinase (ERK) Activity in a CK2alpha Kinase-independent Manner to Promote Resistance to Inhibitors of RAF and MEK but Not ERK in BRAF Mutant Melanoma. J Biol Chem *291*, 17804-17815.

Zwick, E., Bange, J., and Ullrich, A. (2001). Receptor tyrosine kinase signalling as a target for cancer intervention strategies. Endocr Relat Cancer *8*, 161-173.

7.0 APPENDIX A

SUPPLEMENTARY TABLES 1- 8:

Supplementary Tables 1 to 8 are available as part of the published article Goel et al., MCP 2018. These Supplementary Tables and legends to these tables can be accessed freely on the Journal's website as "Supplemental Data", using the following weblink: http://www.mcponline.org/content/early/2018/03/01/mcp.RA118.000643/suppl/DC1

SUPPLEMENTARY TABLES 9-11:

Supplementary Tables 9 to 11 are part of a manuscript that has been accepted for publication by the journal, *Proteome Science*. These Supplementary Tables have been compressed into a single file and uploaded as an "Additional File" to the University of Saskatchewan's Thesis and Dissertation submittal system. Accessibility to this file is governed by the University of Saskatchewan.

The legends to Supplementary Tables 9 to11 are appended below:

SUPPLEMENTARY TABLE 9: The table is organized in the following spreadsheets:

- Spreadsheet "All Phosphopeptides": This table lists all identified phosphopeptides from three independent replicates, via LC-MS/MS. Additional information corresponding to each phosphopeptide is also shown and includes peptide ID# (numerical ID), fraction (enriched; 1 or flowthrough; 2 fractions), Retention time, m/z, charge, measured mass, mass error, score, peptide sequence, modifications (type of PTM), UniProt accession, Phosphosite(s), sequence window, description of protein, use in quantitation (at least 2 unique tryptic peptides were identified for the corresponding protein), ANOVA p-values corresponding to each phosphopeptide identified across the three independent replicates, normalized abundance (phosphopeptide intensity) of the identified phosphopeptides across each replicate, average phosphopeptide intensities of the peptide in the control cells or cells

expressing wild type SRMS, fold-change abundance of phosphopeptide intensities in SRMS-expressing cells (SRMS/Control) and Log_2-scaled fold-change of SRMS/control values.

- Spreadsheet "All unique phosphoproteins": This table lists all the unique phosphoproteins mapped from the phosphopeptides quantified via LC-MS/MS in the Vector and SRMS-expressing cells. Also shown are all associated phosphosites identified for each protein.
- Spreadsheet "Phosphopeptides p-value ≤ 0.05": This table lists all phosphopeptides filtered at an ANOVA p-value threshold of 0.05. Also shown is other relevant information corresponding to each phosphopeptide, as in Spreadsheet "All Phosphopeptides".
- Spreadsheet "Hyperphosphorylated proteins": This table lists all the hyperphosphorylated/upregulated proteins identified in the SRMS-regulated phosphoproteome. These proteins were mapped from the hyperphosphorylated/upregulated peptides (Log_2 SRMS/Control ≥ 1.58 and ANOVA p-value ≤ 0.05). Therefore, only phosphosites identified at Log_2 SRMS/Control ≥ 1.58 and ANOVA p-value ≤ 0.05, are shown for every protein.
- Spreadsheet "Hypophosphorylated proteins": Shown here is the hypophosphorylated/downregulated protein mapped from the hypophosphorylated/downregulated peptide (Log_2 SRMS/Control ≤ -1 and ANOVA p-value ≤ 0.05) identified in SRMS-expressing cells.

SUPPLEMENTARY TABLE 10: The file is organized into the following spreadsheets:

- Spreadsheet "Cell. & Biol. processes": This table shows all the cellular and biological processes identified by IPA analysis (www.ingenuity.com) enriched from the

upregulated/hyperphosphorylated proteins identified in the SRMS-regulated phosphoproteome. To minimize redundancy, the cellular and biological processes are categorized in specific functional "categories". Also shown for every cellular and biological process are P-values (Fisher's exact t-test), activation z-score, predicted activation state (based on a cut-off z-score= +/-2), protein names and the number of proteins enriched in the corresponding cellular and biological process.

- Spreadsheet "Upstream regulators": This table lists all the upstream regulators identified by IPA analysis for the SRMS-dependent upregulated/hyperphosphorylated proteins. Also shown are p-values, activation z-score, predicted activation state (based on a cut-off z-score= +/-2), protein names and number of proteins for every upstream regulator identified by IPA analyses.
- Spreadsheet "Signaling pathways": Shown here are the canonical signaling pathways corresponding to the upregulated/hyperphosphorylated proteins, as identified by IPA analysis. Also shown are the -log p-values, ratio (the ratio estimates the representation of the proteins in each canonical pathway. This is calculated as the ratio of the number of proteins that are enriched in the pathway to the number of proteins in the Ingenuity database reference dataset that make up the pathway) and the protein names corresponding to each canonical pathway.

SUPPLEMENTARY TABLE 11: This table lists all the candidate upstream kinases identified by NetworKIN analysis, for the upregulated phosphosites. Shown in the table is the substrate (phosphoprotein identified in the SRMS-regulated phosphoproteome), position (corresponding upregulated phosphosite), upstream kinase ID (candidate kinase known to target the corresponding

phosphosite), NetworKIN score, kinase group (NetPhorest-annotated kinase group), NetPhorest score, STRING identifier (the corresponding STRING-annotated identifier/ID for the interaction between the kinase and its substrate phosphoprotein), STRING score, Motif sequence (the specific motif corresponding to the phosphosite that is targeted by the candidate kinase) and STRING path.